A Mixed Methods Investigation on Perceived Quality in Crowdsourced Localisation

Thesis Submitted for the Degree of
Doctor of Philosophy
by

Tabea Margret De Wille
Department of Computer Science and Information Systems,
University of Limerick

Supervisors: Dr Chris Exton and Reinhard Schäler

Submitted to the University of Limerick, April 2017
Abstract

With thousands of language communities worldwide and ever growing access to technology and internet communication, the demand for translation of content has been growing steadily. Traditional forms of language translation cannot cope with the future needs and demands of humanity. One of the solutions described and practiced has been localisation crowdsourcing.

However, one of the concerns that have been expressed is that crowdsourced localisation might not provide sufficiently high quality. One perspective on quality is that of perceived quality, which can be influenced by extrinsic cues rather than the actual, objectively observable attributes of a product.

This research asks the following questions:

1. How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?
2. What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?
3. What is the impact of 'naive theories' on perceived quality in localisation?

This research uses a mixed methods approach that includes a survey as well as a set of experiments. It targets a range of groups in localisation, from translators and buyers to students and lecturers of translation studies. As a special focus, it includes volunteers and organisations that collaborate with the non-profit organisation The Rosetta Foundation\(^1\) (TRF) to provide voluntary, crowdsourced localisation for not-for-profit projects via the platform Trommons\(^2\).

Based on the data collected and analysed, this research concluded that accurate representation of meaning was of primary concern for all groups. In addition, specific aspects were raised for each group that were congruent with their main interests and perspectives. Overall, attitudes towards crowdsourced localisation were positive. However,

\[^1\] https://www.therosettafoundation.org/
\[^2\] https://trommons.org/
crowdsourcing also introduced quality uncertainty, which was especially prevalent in respondents with lower levels of training, experience and specialisation.
Declaration

I hereby declare that this thesis is entirely my own work except for collaborations as outlined in the acknowledgements section of this thesis, and that it has not been submitted as an exercise for a degree at any other university.

Parts of this thesis have been presented at conferences and published in conference proceedings. A full list of the publications and conference presentations can be consulted in appendix A.
Acknowledgements

I am grateful and indebted to many people who have supported me over the years. To name just a few:

First and foremost, I would like to thank my supervisors Dr. Chris Exton and Reinhard Schäler for their invaluable guidance, kind support and generous mentorship over the past years. I would also like to thank Dr. Montserrat Bermudez Bausela at the University Alfonso X el Sabio in Madrid (UAX) for the opportunity to collaborate on design and execution of Experiment 1 and the associated questionnaire. Questionnaires 1 and 2 were conducted in collaboration with staff at The Rosetta Foundation and researchers at the Localisation Researcher Centre³. I would like to thank Paulina Abzieher, Alba Pi Barris, Rubén Escobar Capilla, Angela Daly, Geraldine Exton, Konstantinos Koutsikouris, Alison Nolan, and Soroush Saadatfar for the collaboration.

Many thanks to Konstantinos Koutsikouris for conducting the coding that served as foundation for the inter coder analysis calculations, as well as for extensively testing the questionnaire. I would also like to thank Geraldine Exton for helping with proofreading this thesis.

Special thanks go to my parents Eberhard and Martina De Wille who have over many years enabled me to pursue my education. I would also like to thank my siblings and friends for their support over the years. Especially Susanne De Wille, Geraldine Exton, Konstantinos Koutsikouris, and my partner Paul Broe have been my rocks and cheer leaders along this process. Thank you! Finally, I would like to thank all those that have agreed to be interviewed, fill in questionnaires, participate in experiments, provide translations/reviews and pilot questionnaires. Without you, this research would have not been possible and I am very grateful for your time and your generosity.

This research is supported by Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre (www.adaptcentre.ie) at the University of Limerick

³ http://www.localisation.ie/
Dedication

To Emilia, Johanna, Katharina, Magdalena, Sarah-Marie and all those yet to come.
# Table of Contents

Abstract .............................................................................................................................................. 3

Declaration ........................................................................................................................................ 5

Acknowledgements .......................................................................................................................... 6

Dedication .......................................................................................................................................... 7

Table of Contents ............................................................................................................................ 8

List of Tables .................................................................................................................................... 11

List of Figures .................................................................................................................................. 14

List of Abbreviations ...................................................................................................................... 21

Part I: Research Context .................................................................................................................. 22

Chapter 1: Introduction .................................................................................................................. 22

1.1. Introduction and Overview ....................................................................................................... 22

1.2. Research Questions .................................................................................................................. 30

1.3. Thesis Structure ....................................................................................................................... 33

Chapter 2: Theoretical Framework .................................................................................................. 34

2.1. The Concept and Study of Quality ......................................................................................... 34

2.2. Core Assumptions of this Thesis ............................................................................................ 40

2.3. The Quality Production Process .............................................................................................. 45

2.4. The Quality Experience Process ............................................................................................. 48

2.5. The Quality Evaluation Process ............................................................................................. 53

2.6. Summary .................................................................................................................................. 59

Chapter 3: Quality in Localisation ................................................................................................... 60

3.1. Introduction ............................................................................................................................. 60

3.2. The Theoretician's Perspective ............................................................................................... 61

3.3. The Business of Localisation ................................................................................................. 64
3.4 Applying the Integrative Quality Framework ................................................................. 79
3.5. Summary .......................................................................................................................... 92
Chapter 4: Quality in Crowdsourced Localisation ............................................................... 94
4.1. Crowdsourced and Volunteer Localisation ................................................................. 94
4.2. Summary ........................................................................................................................ 127
Part II: Methodological Approach ..................................................................................... 128
Chapter 5: Methodology and Mixed Methods Research Design ....................................... 128
5.1. Introduction .................................................................................................................... 128
5.2. Overview and Overall Study Design ............................................................................. 131
5.3. Outline of the Research Design ..................................................................................... 132
5.4. Rationale for Using Mixed Methods Research ............................................................. 146
5.5. Summary and Overview of the Research ..................................................................... 148
5.6. Methodological Considerations ..................................................................................... 151
Chapter 6: Data Collection .................................................................................................. 167
6.1. Introduction .................................................................................................................... 167
6.2. Interviews ..................................................................................................................... 167
6.3. Content Analysis .......................................................................................................... 176
6.4. Questionnaires ............................................................................................................ 189
6.5. Experiments .................................................................................................................. 201
Part III: Data Analysis and Research Findings .................................................................. 233
Chapter 7: Survey Results ................................................................................................... 233
7.1. Demographics ............................................................................................................... 233
7.2. Attitudes ....................................................................................................................... 249
Chapter 8: Experiment Results ............................................................................................ 343
8.1. Experiment 1: Professional Status, Spanish ................................................................. 343
8.2. Experiment 2: Price of Translation, German Linguists ...................................................... 394
8.3. Experiment 3: Price of Translation, Buyers/Clients ............................................................ 409

Chapter 9: Summary and Discussion ......................................................................................... 418
  9.1. Summary and Discussion ................................................................................................. 418

Chapter 10: Conclusion ............................................................................................................. 464
  10.1. Original Contributions of the Research ........................................................................ 469
  10.2. Limitations and Future Research .................................................................................. 470
  10.3. Implications of the Research ......................................................................................... 471

Bibliography ............................................................................................................................... 474

Appendices ................................................................................................................................. 488
  Appendix A ............................................................................................................................. 488
  Appendix B ............................................................................................................................. 490
  Appendix C ............................................................................................................................. 494
  Appendix D ............................................................................................................................. 497
  Appendix E ............................................................................................................................. 508
  Appendix F ............................................................................................................................. 542
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Motivations for volunteering, Trommons volunteers</td>
<td>115</td>
</tr>
<tr>
<td>2</td>
<td>Overview of questionnaires included in survey part of this research</td>
<td>141</td>
</tr>
<tr>
<td>3</td>
<td>Overview of experiments conducted for this research</td>
<td>144</td>
</tr>
<tr>
<td>4</td>
<td>Research overview</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>Overview of interview participant groups</td>
<td>168</td>
</tr>
<tr>
<td>6</td>
<td>Derivations, synonyms and subsets of Multidimensional Quality Metrics used in semi-automated checks</td>
<td>185</td>
</tr>
<tr>
<td>7</td>
<td>MQM label deviations</td>
<td>188</td>
</tr>
<tr>
<td>8</td>
<td>Overview of undergraduate student participants</td>
<td>192</td>
</tr>
<tr>
<td>9</td>
<td>Experiment overview</td>
<td>201</td>
</tr>
<tr>
<td>10</td>
<td>Overview of reviewers</td>
<td>210</td>
</tr>
<tr>
<td>11</td>
<td>Initial distribution of headers and paragraphs per scenario type</td>
<td>216</td>
</tr>
<tr>
<td>12</td>
<td>Initial translator allocation per scenario</td>
<td>216</td>
</tr>
<tr>
<td>13</td>
<td>Final distribution of headers and paragraphs per scenario type</td>
<td>217</td>
</tr>
<tr>
<td>14</td>
<td>Final translator allocation per scenario</td>
<td>217</td>
</tr>
<tr>
<td>15</td>
<td>Student participant distribution</td>
<td>218</td>
</tr>
<tr>
<td>16</td>
<td>Translator price point overview</td>
<td>222</td>
</tr>
<tr>
<td>17</td>
<td>Initial price point allocation per scenario</td>
<td>224</td>
</tr>
<tr>
<td>18</td>
<td>Final price point allocation per scenario</td>
<td>225</td>
</tr>
<tr>
<td>19</td>
<td>Translator price point overview</td>
<td>226</td>
</tr>
<tr>
<td>20</td>
<td>Sample comments, translations rated 1</td>
<td>230</td>
</tr>
<tr>
<td>21</td>
<td>Sample comments, translations rated 5</td>
<td>230</td>
</tr>
<tr>
<td>22</td>
<td>Participant categories UAX</td>
<td>234</td>
</tr>
<tr>
<td>23</td>
<td>Grouping of participants at UAX</td>
<td>234</td>
</tr>
</tbody>
</table>
Table 24: Education/training in localisation related areas of non-profit client participant group. 244

Table 25: Common labels emerging from definitions of localisation quality across all groups. 251

Table 26: Quality aspects ranked by importance by professional industry buyer participant group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9. 259

Table 27: Quality aspects ranked by importance by linguist participant group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9. 262

Table 28: Results of Trommons reviewer comments content analysis - labels with fewer than 10 mentions. 264

Table 29: Results of Trommons reviewer comments content analysis - labels with more than 10 mentions. 265

Table 30: Ranking of payment vs. experience for professional industry buyer group. 285

Table 31: Ranking of payment vs. training for professional industry buyer group. 289

Table 32: Ranking of reasons why a translator might ask for low rates for professional industry buyer group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9. 295

Table 33: Ranking of reasons why a translator might ask for low rates for linguist respondent group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9. 301

Table 34: Ranking of payment vs. experience for linguist respondent group. 303

Table 35: Ranking of payment vs. training for linguist respondent group. 307

Table 36: Common labels emerging from definitions of 'language professional' status across all groups. 316

Table 37: Relationship between information provided to participants vs. actual translator background for BadInfo scenarios in Experiment 1. 345

Table 38: Example of selections of 'expert' translator in percentage, lecturers at UAX participant group. 379

Table 39: Overview of source language word counts per scenario. 386
Table 40: Average selection of ‘expert’ translation relative to word count of scenarios, student and lecturers UAX and Spanish linguists as well as those indicating that professional status was a deciding factor for quality vs. those who didn’t think so. ................................................................. 388

Table 41: Average mention of quality aspects across all participant groups ................................................. 434
List of Figures

Figure 1: Integrative Quality Framework as presented in Golder et al. (2012, p.3) ......................... 44
Figure 2: Visualisation of quality production process ........................................................................ 45
Figure 3: Different types of attributes, as presented by Golder et al. (2012) in web appendix ........... 47
Figure 4: Visualisation of experienced vs. perceived attributes in quality experience process ......... 49
Figure 5: Visualisation of quality evaluation process ..................................................................... 54
Figure 6: Visualisation of translator to end-user link ..................................................................... 68
Figure 7: Visualisation of localisation process chain ..................................................................... 68
Figure 8: Table of common standards in localisation, as provided in Txabarriaga 2009 (p.6) ........... 75
Figure 9: Data collection methods in relation to research questions ............................................. 133
Figure 10: Example of MQM parent-child category relationship (Lommel and Melby 2014) .......... 186
Figure 11: Rates quoted by translators in response to proz.com recruitment call ....................... 208
Figure 12: Example screen; Experiment 1 (no information provided) ........................................... 213
Figure 13: Example screen; Experiment 1 (information provided) ............................................... 214
Figure 14: Example screen; Experiment 2 (no information provided) .......................................... 223
Figure 15: Example screen; Experiment (information provided) ................................................... 224
Figure 16: Sample screen; Experiment 3 (no information provided) ............................................. 228
Figure 17: Sample screen; Experiment 3 (Information provided) .................................................. 229
Figure 18: Professional experience of UAX participant groups ...................................................... 235
Figure 19: Reviewer training of UAX lecturer participant group .................................................... 236
Figure 20: Highest level of education in Spanish linguist participant group .................................. 237
Figure 21: Primary role of Spanish linguist participant group ....................................................... 238
Figure 22: Time involved in localisation of Spanish linguist participant group .............................. 239
Figure 23: Highest level of education of German linguist participant group ................................. 240
Figure 24: Primary role German linguist participant group ......................................................... 241
Figure 25: Time involved in localisation of German linguist participant group ......................... 242
Figure 26: Years of experience managing localisation in non-profit client participant group ...... 243
Figure 27: Highest level of education in industry buyer participant group ............................... 245
Figure 28: Localisation experience in industry buyer participant group .................................. 246
Figure 29: Highest level of education in 'other' linguists participant group ................................. 247
Figure 30: Commonly mentioned aspects of quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups. ..................................................................................................................... 252
Figure 31: Criteria used for evaluating quality in linguist participant group .............................. 272
Figure 32: Expected level to which majority of reviewers would agree with change requests in linguist participant group .......................................................................................................................... 273
Figure 33: Ranking of factors impacting quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups. .................................................................................................................. 278
Figure 34: Impact of factors on quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups .................................................................................................................. 279
Figure 35: Quality rankings based on price for professional industry buyer group .................. 283
Figure 36: Ranking of voluntary translation with 5 years vs. 3 months of experience for professional industry buyer group .................................................................................................................. 286
Figure 37: Ranking of 5 cent per word translation with 5 years vs. 3 months of experience for professional industry buyer group ............................................................................................................ 287
Figure 38: Ranking of 15 cent per word translation with 5 years vs. 3 months of experience for professional industry buyer group ............................................................................................................ 288
Figure 39: Ranking of voluntary translation with formal training vs. no formal training for professional industry buyer group .................................................................................................................. 290
Figure 40: Ranking of 5 cent per word translation with formal training vs. no formal training for professional industry buyer group .................................................................................................................. 291
Figure 41: Ranking of 15 cent per word translation with formal training vs. no formal training for professional industry buyer group .................................................................................................................. 292
Figure 57: Overall view of translation crowdsourcing for linguist respondent group, broken down by primary role of respondents ................................................................. 335

Figure 58: Example of high vs. low agreement results in experiment scenarios .................... 347

Figure 59: Best translation selected by student and lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .......................................................................................................................... 350

Figure 60: Standard deviation in selections for student and lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .......................................................................................................................... 351

Figure 61: Average standard deviation in selections for student and lecturer UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator .......................................................................................................................... 352

Figure 62: Best translation selected by lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................. 354

Figure 63: Standard deviation in selections for lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. 355

Figure 64: Average standard deviation in selections for lecturer UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator .......................................................................................................................... 356

Figure 65: Best translation selected by 3rd/4th year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .......................................................................................................................... 359

Figure 66: Standard deviation in selections for 3rd/4th year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .......................................................................................................................... 360

Figure 67: Average standard deviation in selections for 3rd/4th year student UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator .......................................................................................................................... 361

Figure 68: Best translation selected by 1st/2nd year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .......................................................................................................................... 363
Figure 69: Standard deviation in selections for 1st/2nd year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 364

Figure 70: Average standard deviation in selections for 1st/2nd year student UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 365

Figure 71: Best translation selected by Spanish linguist respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. . 367

Figure 72: Standard deviation in selections for Spanish linguist respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. . 368

Figure 73: Average standard deviation in selections for Spanish linguist respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 369

Figure 74: Best translation selected by respondents who thought professional status of translator had impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 371

Figure 75: Standard deviation in selections by respondents who thought professional status of translator had impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 372

Figure 76: Average standard deviation in selections by respondents who thought professional status of translator had impact on quality, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 373

Figure 77: Best translation selected by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 375

Figure 78: Standard deviation in selections by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 376

Figure 79: Average standard deviation in selections by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator. .................................................................................................................. 377

Figure 80: Average selection of 'expert' translation relative to mean of all respondent groups, student and lecturers UAX and Spanish linguists as well as those indicating that professional status was a deciding factor for quality vs. those who didn't think so. .................................................................................................................. 381
Figure 81: Mean of standard deviation for student and lecturer UAX and Spanish linguist participant groups as well as those indicating that professional status was a deciding factor for quality vs. those who didn’t think so. BadInfo refers to the information provided to participants, not the actual translator. ................................................................. 384

Figure 82: Standard deviation vs. source word count for each scenario, lecturers UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator...... 389

Figure 83: Standard deviation vs. source word count for each scenario, Spanish linguist respondent group. BadInfo refers to the information provided to participants, not the actual translator...... 391

Figure 84: Standard deviation vs. source word count for each scenario, 3rd/4th year student UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator. ................................................................. 392

Figure 85: Standard deviation vs. source word count for each scenario, 1st/2nd year student UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator. ................................................................. 393

Figure 86: Selections made when no information on cost of translation was provided, German linguist participant group........................................................................................................... 396

Figure 87: Selections made when correct information on cost of translation was provided, German linguist participant group........................................................................................................... 397

Figure 88: Selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the information the participants were given........................................................................................................... 399

Figure 89: Selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the actual cost of translation. (Note: Since the chart is sorted with ascending values, the order of labels has been changed relative to figure 88)............................................................................................................ 400

Figure 90: Average selections made when no information on cost of translation was provided, German linguist participant group........................................................................................................... 401

Figure 91: Average selections made when correct information on cost of translation was provided, German linguist participant group........................................................................................................... 402

Figure 92: Average selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the information the participants were given........................................................................................................... 403
Figure 93: Average selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the actual cost of translation. ........................................................................................................................................404

Figure 94: Ranking of quality expectations, German linguists who participated in experiment 2. 406

Figure 95: Standard deviation in selections for German linguist, broken down by scenario ..........408

Figure 96: Average standard deviation in selections for the German linguist respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator. ........................................................................................................409

Figure 97: Average selections when no information on cost of translation was provided, non-profit client respondent group .................................................................413

Figure 98: Average selections when information (correct and incorrect) on cost of translation was provided, non-profit client respondent group ..................................................414

Figure 99: Average selections when no information on cost of translation was provided, professional industry buyer respondent group ........................................................................416

Figure 100: Average selections when information (correct and incorrect) on cost of translation was provided, professional industry buyer respondent group ..................................417
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATA</td>
<td>American Translators Association</td>
</tr>
<tr>
<td>BLEU</td>
<td>Machine Translation quality evaluation metric</td>
</tr>
<tr>
<td>CEO</td>
<td>Chief Executive Officer</td>
</tr>
<tr>
<td>CNG</td>
<td>Centre for Next Generation Localisation</td>
</tr>
<tr>
<td>CV</td>
<td>Curriculum Vitae</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>MQM</td>
<td>Multidimensional Quality Metrics</td>
</tr>
<tr>
<td>MLV</td>
<td>Multi-Language Vendor</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-Governmental Organisation</td>
</tr>
<tr>
<td>NVivo</td>
<td>Brand name for qualitative data analysis software</td>
</tr>
<tr>
<td>OASIS</td>
<td>Organization for the Advancement of Structured Information Standards</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>ROM hacking</td>
<td>Modification of data on a read-only memory chip</td>
</tr>
<tr>
<td>SLV</td>
<td>Single-Language Vendor</td>
</tr>
<tr>
<td>SOLAS</td>
<td>Service-Oriented Localisation Architecture Solution</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Sciences</td>
</tr>
<tr>
<td>ST</td>
<td>Source Text</td>
</tr>
<tr>
<td>TAUS</td>
<td>Translation Automation User Society</td>
</tr>
<tr>
<td>TRF</td>
<td>The Rosetta Foundation</td>
</tr>
<tr>
<td>TT</td>
<td>Target Text</td>
</tr>
<tr>
<td>UAX</td>
<td>University Alfonso X el Sabio in Madrid</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>XLIFF</td>
<td>XML Localisation Interchange File Format</td>
</tr>
</tbody>
</table>
Part I: Research Context

Chapter 1: Introduction

1.1. Introduction and Overview

The concept of quality is a central issue in the localisation field, both in academia and in industry. Current efforts are primarily focused on the question of how quality can be achieved and how it should be evaluated. Ideally, these efforts would eventually lead to a situation where all aspects of quality are well defined, and a clear pathway on how to achieve it for all types of content and projects is laid out. Evaluation would then allow the commercial side of localisation to determine a price appropriate for the value to a buyer of the level of quality delivered.

In reality, debate continues around how to determine the best strategies for managing quality and indeed on how to define it. This situation is not unique to localisation, and as later sections will show, central questions around defining and managing quality remain without conclusive answers in the literature. However, literature outside the field of localisation and translation offers a view on quality that the localisation field so far has paid little attention to: The concept of perceived quality, which has been extensively studied in areas outside of localisation, like marketing, economics and business management.

Golder et al. (2012) describe quality as a

"set of three distinct states of an offering's attributes' relative performance generated while producing, experiencing, and evaluating the offering." (p.4).

During the quality experience process a set of attributes is delivered and the customer perceives these attributes through their own individual lens. This lens is composed of their measurement knowledge, motivation, emotions and expectations (p.4). Due to this individual lens only an individual sub-set of attributes is perceived by the customer, which results in the customer missing information about a product's attributes. This leads to a state of information asymmetry, where the seller knows more about the quality of a product than the buyer (Akerlof 1970), and the customer will use indirect measures outside of the actual product's attributes to assess quality. These indirect measures
can include aspects like images, advertising or brand names (Garvin 1984, p.32) as well as the price of the product (Akerlof 1970, pp.489–490). This effect commonly leads to prices and quality dropping for an entire market (Akerlof 1970) and could partially explain the downward trend of prices paid for localisation services and the increased commoditisation that has been a frequent topic of discussion within the localisation industry over the past years (Sargent and DePalma 2007, p.16).

While this thesis will discuss the concept of quality in general and for localisation specifically in later chapters, the focus of this research is on perceived quality in the context of not-for-profit crowdsourced localisation, in relation to the for-profit localisation industry. Here the purpose of this research is to gain an understanding of indirect measures that are applied within the localisation field, for-profit professionally sourced and not-for-profit crowdsourced, to cope with information asymmetry, as well as the effect those indirect measures have on how quality is perceived.

This research uses a mixed methods approach that includes a survey of different groups in localisation to gain insights into such indirect measures as well as a set of experiments that were designed to observe the effect the indirect measures have on perceived quality. It targets a range of groups in localisation, from translators and buyers to students and lecturers of translation studies. In addition, and as a special focus, it includes volunteers and organisations that collaborate with the non-profit organisation The Rosetta Foundation (TRF)⁴ to provide voluntary, crowdsourced localisation for not-for-profit projects via the platform Trommons⁵, which is administered by TRF.

1.1.1. Crowdsourcing and Localisation Crowdsourcing

Crowdsourcing as a concept is not a recent invention. It is defined in recent times as "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" (Howe 2006).

However, as early as the mid-19th century, the Oxford English Dictionary was compiled using selections submitted by volunteers (Mugglestone 2000, p.23).

---

⁴ https://www.therosettafoundation.org/
⁵ https://trommons.org/
Similarly, it is not evident that the volume of non-professional translation compared to paid-for translation is growing. Indeed, it may be the case that non-professional translation has always been more common than professional translation. This seems especially likely when considering the large number of people who speak two or more languages worldwide, with over half the population of the European Union alone claiming to speak more than one language (European Commission 2012, p.12). As a result, there are a range of potential scenarios where non-professional translations would be conducted in everyday life, also bearing in mind the ongoing, centuries old history of migration that spans continents, cultures and languages.

However, over the past years, technology has enabled more and more organisations and individuals to initiate and contribute to crowdsourcing projects that allow for a concerted, organised effort in non-professional translations. Platforms like Amazon Mechanical Turk\(^6\) and Crowdin\(^7\) offer a low effort entry into harnessing the power of the crowd and numerous initiatives and organisations like Facebook\(^8\), Duolingo\(^9\), Translators Without Borders\(^10\), LinkedIn\(^11\), and Twitter\(^12\) have implemented their own technology and workflows to solicit translations at low cost or free of charge from ‘the crowd’.

As will be described in more detail in section 4.1.2, these initiatives vary on different levels: project sizes, number of participants, for-profit or not-for-profit, paid or unpaid, solicited or unsolicited, as well as field, recruitment, technology and processes involved. For this research, the focus was on the Trommons community and platform, which will be introduced in more detail in section 4.1.4.

1.1.2. Quality in Localisation Crowdsourcing

While an initiative like TRF with a platform like Trommons can be considered success stories where ‘the crowd’ comes together to improve the lives of others,

\(^6\) https://www.mturk.com/mturk/welcome
\(^7\) https://crowdin.com/
\(^8\) https://www.facebook.com/
\(^9\) https://www.duolingo.com/
\(^10\) http://translatorswithoutborders.org/
\(^11\) https://www.linkedin.com/
\(^12\) https://twitter.com/twitter?lang=en
crowdsourcing projects in general commonly face criticism and scepticism.

Depending on the type of initiative and its aims and approaches, the way these initiatives are received varies and different concerns may be raised. Those, which receive the most attention, are:

- ethical concerns (Hosaka 2008; Dolmaya 2011);
- a perceived threat to the translation profession (Dolmaya 2011);

Criticism and scepticism towards localisation crowdsourcing based on anticipated or observed quality is commonly founded on the basic assumption that ‘professional’ translation is of superior quality compared to those translations provided by ‘volunteers’. This is reflected in the concern about LinkedIn’s move to crowdsourcing translation voiced by Jiri Sejskal, President of the American Translators Association who writes:

"The crowdsourcing model cannot work unless volunteers are competent. Companies seeking global reach regularly lose face and money by relying on amateurs to translate key materials, and the embarrassing results are generally set right by professional translators."

(American Translators Association 2009, p.3)

In the accompanying press release, this move is described as

- "disappointing and Highly Unprofessional" (American Translators Association 2009, p.1);
- compromising of LinkedIn’s professionalism (American Translators Association 2009, p.1);
- "hazardous waters" (American Translators Association 2009, p.2);
- and the article makes reference to the poor quality of the early years of Wikipedia and error-prone results encountered at Facebook (American Translators Association 2009, p.1).

Crowdsourced translations on Facebook in particular are commonly criticised for sloppiness and skimping which result in grammatical, spelling and usage problems (Hosaka
2008, p.1). They are also described negatively, for example as "garbage translations" (McDonough Dolmaya 2012, p.186 - citing a study participant).

These assumptions can also be observed in the setup of research experiments in an academic context.

Zaidan and Callison-Burch (2011) base their experiments on "professionally-produced reference translations" (p.1220) allowing for objective comparison of quality of professional and non-professional translations (p.1220). They assert that due to the lack of information provided about the translators on Amazon Mechanical Turk "soliciting translations from anonymous non-professionals carries a significant risk of poor translation quality. Whereas hiring a professional translator ensures a degree of quality and care, it is not very difficult to find bad translations provided by Turkers" (p.1221).

And indeed, examples provided in the paper would suggest that the translations provided through crowdsourcing on Amazon Mechanical Turk are of inferior quality, although it is not clear how many of the translations sourced through that channel had such issues and whether the professionally produced translations had fewer issues. The basic assumption made is that "(t)he more the selected translations resemble the professional translations, the higher the quality" (Zaidan and Callison-Burch 2011, p.1223).

Similar attitudes could be observed when TRF was first launched. Anecdotal evidence based on conversations with founding member and former CEO Reinhard Schäler suggests that organisations were generally sceptical about the quality of localisation they would receive from crowdsourced volunteers. When conversations were first started to get different organisations engaged in the initiative, common concerns included that investing in the initiative, be it a time or a financial investment, would be too risky for organisations because volunteer translations posed a risk to quality. This was even still the case after the initial weaknesses in the process had been addressed and TRF received primarily positive feedback from those organisations that did decide to participate, with rarely any quality issues arising.

However, these and similar arguments brought forth by professional translation bodies, academia and potential non-profit collaborators make several crucial assumptions that are insufficiently supported:
1. **It is not clear whether most paid-for, professional translations are of higher quality than most crowdsourced, voluntary translations.**

   As far as this researcher is aware,
   - there is no comprehensive body of research that evaluates and compares the output of either;
   - and indeed, it can't be safely assumed that paid-for translations by professional translators are on average of high quality.

2. **It is not clear what distinguishes a 'professional' translator from a 'volunteer' translator and what impact on quality any differences might have.**

   While there are professional bodies for translation as well as academic degree programmes and professional training programmes,
   - the profession of the translator is insufficiently protected, allowing anyone who wishes to do so to call themselves a translator, independent of their educational background or professional experience.
   - In addition, it is conceivable that trained and experienced translators, who earn their living by providing professional translations, also contribute to crowdsourcing projects, which would place them in both categories, 'professional' and 'volunteer', making the distinction between these two groups meaningless.
   - As a result, it is not clear whether the distinction would be indicative of the quality that can be expected from either groups, or to what extent other factors like differences in processes and technology might play a role.
3. It is not clear what is meant by quality, and whether different quality expectations apply to voluntary, crowdsourced vs. paid-for, professional translations.

As following sections will discuss, quality in translation and localisation is notoriously hard to define (Chapter 3), in part because there are different types of quality that should be considered.

1.1.3. Perceived Quality

Assumptions like the three described in the previous section could be described to fall into the category of 'naive theories'. Naive theories are

"informal, common-sense explanations people use in their everyday lives to make sense of their environment and often diverge from formal, scientific explanations of what actually happens"


Through marketing and advertising, customers receive partial information about a product, and then have to fill in the missing information based on information that is available (Garvin 1984, p.32; Deval et al. 2012, p.1186), leading to quality being perceived through indirect measures like images, advertising or brand names, rather than the objective characteristics of the product (Garvin 1984, p.32).

Previous research suggests that the extent to which customers are susceptible to cues outside of a product's objective attributes depends on their level of expertise (Deval et al. 2012; Golder et al. 2012) and commonly, buyers know less about a product than the seller does, leading to a state of information asymmetry (Akerlof 1970, p.489). In the context of localisation this can be applied to scenarios where the person making a recommendation or the decision on whether a service that has been delivered - primarily translated text - is of an acceptable quality does not have sufficient information, skills or expertise to be able to fully judge the quality of the work. Situations like these can, for example, arise under the following conditions:

- The decision maker (buyer) does not know localisation quality best practices, definitions, standards and processes well enough to be able to
come to an informed judgment. This could for example arise if their area of expertise is in a different field and managing localisation within their organisation is a secondary task. Based on market research conducted by the Common Sense Advisory in 2004, less than one-third of localisation buyers carried a localisation related job title, while the majority of buyers held other roles, from marketing experts to C-level executives (DePalma and Beninatto 2004, pp.5–6).

- The buyer might not speak the language(s) their content has been translated into to a sufficiently high level to be able to come to an informed judgment. This could arise if they are managing localisation projects that include a large number of languages.

While there are numerous ways that situations like these are addressed by professional localisation buyers in the industry, some of which will be elaborated on throughout this thesis paper, those approaches require not only knowledge of common issues encountered in localisation projects but also the skills and tools to address those and visibility into the processes by which the localised versions of a text are created. In addition, many aspects of assessing quality still rely on the judgment of individuals who speak the target language(s). This poses the additional challenges:

- The buyer might not know the reviewer tasked with assessing the quality of the work well enough to be able to judge whether their assessment can be trusted.
- The reviewer may or may not have sufficient training and experience to be able to assess the quality of the work.
- In situations where quality expectations are not well defined, the reviewer will have to make decisions on quality definitions and standards to evaluate the work against, which may or may not conform with the buyer's quality expectations.
As the sections on quality (Chapter 2) and quality in localisation (Chapter 3) will discuss in greater detail, perceived quality is only one part of the bigger quality picture, yet one worth considering in greater detail. Even if the localisation industry and academia collectively came to an agreement on how to define, produce and evaluate quality in a comprehensive and universal way, without addressing the state of information asymmetry between buyers and sellers of localisation, difficulties with being able to sell quality would likely persist. Investigating how quality is perceived and which factors influence these perceptions can therefore not only provide insights that are relevant to the area of localisation crowdsourcing, but to the localisation industry as a whole.

1.2. Research Questions

With the above background in mind, this thesis sets out to address three research questions, which are closely related and building on each other:

1. How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?
2. What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?
3. What is the impact of 'naive theories' on perceived quality in localisation?

1.2.1. How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?

The research first draws on existing literature in localisation and translation studies, as well as industry reports to gain an understanding of prevalent theories of quality. It then surveys multiple groups in localisation crowdsourcing and for-profit localisation to gain insights into how these theories and definitions of quality are utilised in real world settings. This is achieved through semi-structured interviews, content analysis and questionnaires and includes volunteers and non-profit organisations on Trommons, professional translators, localisation buyers/decision makers in a professional environment as well as students and lecturers of translation studies.
By including different groups in the survey stage, the research adds to existing knowledge of localisation quality by examining perspectives that are currently not widely researched and lays the foundation for the following two questions by establishing to what extent expectations are shared. Without a shared understanding of quality expectations, attempts at producing and evaluating quality would likely fall short, because different players in the process would work towards different goals.

Furthermore, because quality definitions and evaluation need to consider a range of attributes of the final localisation product that differ in how easily they can be described, this part of the research provides insights into the extent to which different groups are able to articulate localisation quality expectations and evaluation comprehensively.

The survey part of the research also asks the following question:

1.2.2. What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?

Deval *et al.* (2012) describe naive theories as inferences that inform product evaluation. These inferences are external of the actual products and inform how quality is perceived in the absence of full information on the product's attributes. This research surveys translators and buyers/decision makers in non-profit and for-profit settings to gain an understanding of what mechanisms these different groups apply to infer quality. Here the question is whether different groups approach quality inferences differently and whether different naive theories are active depending on the individual's background.

In addition to surveying for different naive theories, this research also specifically seeks to understand to what extent the translator's background and the price of the translation play a role as theories that impact perceived quality.

These two variables were selected with an awareness that in the current discourse on translation crowdsourcing, voluntary translators are commonly contrasted with professional translators (Hosaka 2008; American Translators Association 2009; Zaidan and Callison-Burch 2011; McDonough Dolmaya 2012). By definition, a professional engages in an activity to earn their livelihood, so payment for services rendered plays a major role in what defines a professional. In addition, a professional is "characterized by or conforming to the technical or ethical standards of a profession" (Merriam-Webster 2017) - these
attributes are commonly linked to how well-trained and experienced an individual is, so considering the professional background of a translator as a potential factor that could guide perceived quality is appropriate. Since crowdsourcing by definition involves an open call (Howe 2006), anyone - professionals and non-professionals - is able to participate. In addition, crowdsourcing projects are commonly conducted without payment or tangible rewards.

Therefore, while other naive theories are collected throughout the survey stage, special attention is given to these two variables, price and professional status of the translator.

The research then moves away from surveying opinions, attitudes and theories and turns towards investigating the impact these have on the decisions made in localisation by asking:

1.2.3. What is the impact of 'naive theories' on perceived quality in localisation?

This question is addressed through a series of experiments that were conducted with students and lecturers of translation studies, volunteer and for-profit translators, as well as non-profit and for-profit buyers/decision makers in localisation. As with the survey conducted to address the second research question, the experiments focused on two variables to test their impact on perceived quality:

- price of translations (voluntary, low cost, medium cost and high cost translations);
- the professional background of the translator who has provided the translation (novice translators, trained translators, trained and experienced translators).
1.3. Thesis Structure

This thesis is divided into three parts, each including several chapters:

1.3.1. Part I, Research Context

This includes chapter 1, which is the current chapter and outlines the rationale and aims of the research as well as the research question and thesis structure. Chapter 2 introduces the theoretical framework applied and introduces the concept of perceived quality in greater detail. Chapter 3 discusses quality in localisation, which is followed by a review of the literature on crowdsourcing in translation and localisation in chapter 4. Chapter 4 also introduces the Trommons community.

During the structuring of this thesis, one point to be considered was in which order to present the background of the research with the literature review and the theoretical framework. Since the theoretical framework chapter provides structure for the data collection and analysis, it would ordinarily be well positioned before the methodology chapter (Chapter 5). However, as the literature review on quality in localisation and localisation crowdsourcing will show, the topic of localisation quality in particular can benefit from being presented using similar terminology as is used in the literature on quality outside of localisation, as described in chapter 2. The decision was therefore to place the theoretical framework ahead of the subsequent background chapters.

1.3.2. Part II, Methodological Approach

This includes chapter 5, which introduces the overall research design and methodology used, along with methodological considerations, followed by the specific approach to data collection and analysis described in chapter 6.

1.3.3. Part III, Research Findings

This final part first describes the results of the survey section of the research in chapter 7. Chapter 8 then describes the results of the experiments conducted for this thesis, followed by chapter 9 with summary and discussion and finally the conclusion in chapter 10.

The thesis then includes a list of references as well as appendices.
Chapter 2: Theoretical Framework

2.1. The Concept and Study of Quality

The ability to define quality and distinguish good from bad quality is commonly recognised as a major challenge but also of great importance to the business world (Akerlof 1970, p.500; Golder et al. 2012, p.1). Especially in the area of perceived quality, competitive advantages can be gained when the connection between how quality is perceived and the behavioural intentions of customers is examined (Hansen and Bush 1999, p.119; Cronin et al. 2000, p.193).

However, quality as a concept has proven to be notoriously difficult to define and there are wide ranging differences in the perspectives taken on the topic. Hansen and Bush (1999) say:

"Quality has become something of a cliché. Advertisers would have us believe that every company produces a high-quality product. In fact, the term "quality" has been so overused that, to many, it has begun to lose its meaning. This has resulted in confusion among customers, as well as within companies. What is quality? Many people have difficulty answering this. A common response is, 'I can't define quality, but I know it when I see it.' Inability to define quality does not, however, preclude interest in, and strong opinions regarding the concept."

(Hansen and Bush 1999, pp.119–120).

The following sections outline the concept of quality in literature outside the field of localisation, with a focus on perceived quality and the factors that play into the quality perception process. The first part provides an overview of different approaches to quality in use, followed by an introduction to perceived quality and the elements that contribute to the concept.

2.1.1. Approaches to Quality

A major challenge in talking about quality is that many intuitively understand what quality is from their perspective, but without being able to consider the full picture. A common issue is that definitions of quality depend on the perspective taken, and different, inconsistent theories exist in the literature regarding how quality should be viewed. Approaches include quality being defined as value, conformance to specifications, fitness
for use, loss avoidance and meeting/exceeding of customers' expectations (Reeves and Bednar 1994, p.419).

Garvin (1984) outlines five approaches to defining quality (Transcendent, Product-based, User-based, Manufacturing-based and Value-based) and eight dimensions of quality (Performance, Features, Reliability, Conformance, Durability, Serviceability, Aesthetics and Perceived Quality). These five approaches have been applied to the translation industry by Fields et al. (2014).

### 2.1.1.1. Transcendent Approach

It has been argued that quality is absolute and universally recognisable, but that it can only be recognised through experience but not be defined precisely (Garvin 1984, p.25). This transcendent approach to quality, or viewing it as excellence, offers the benefit of being uncompromising in its standards and expectations of achievements. However, it offers little practical guidance to the realities of business, since it is hard to assess quality with this definition and it is not clear who determines the standards of excellence. As an abstract and subjective assessment approach, it is then vulnerable to individual judgments. In addition, aiming to achieve excellence is not always appropriate for all products or services, and would require a customer to be willing to pay for the highest standards (Reeves and Bednar 1994, p.429).

This approach lies on one side of the extreme on a spectrum of being able to measure quality (Ophuis and Trijp 1995, p.177).

Applied to the translation industry, the transcendent approach would suggest that there is not only innate beauty and artistic excellence that can be achieved in a translation, but also that measuring quality of a translation might not be possible or necessary if the translator is an expert (Fields et al. 2014, p. 408).
2.1.1.2. Manufacturing-based Approach

The opposite end of the spectrum from the transcendent approach measures quality against predetermined standards (Zeithaml 1988 in Ophuis and Trijp 1995, p.177). This manufacturing centric approach is primarily concerned with engineering and manufacturing practice, and conformance with requirements is a large focus. Deviation from designs or specifications implies a reduction in quality (Garvin 1984, pp.27–28; Crosby 1979 in Parasuraman et al. 1985, pp.41–42).

Internally, conformance can be measured as the number of defects encountered or the proportion of units that require rework because they fail to meet specifications (Garvin 1984, p.31), where the highest level of quality can be defined as zero defects or getting it right the first time (Garvin 1984, p.27; Parasuraman et al. 1985, p.41).

Externally, the manufacturing-based approach also considers the number of service calls or repairs under warranty and can be thought to include reliability in addition to conformance (Garvin 1984, p.31). In this case, reliability can be described as the "probability of a product's failing within a specified period of time." (Garvin 1984, p.30)

A manufacturing-based approach focused on conformance can provide multiple advantages, especially for practical applications in a business setting. Here, quality is relatively easy to measure and can help an organisation increase efficiency and a consistent strategy. This is particularly beneficial in industries where standards are available and important (Reeves and Bednar 1994, p.430). The customer's interest is considered in the process of establishing the specifications (Reeves and Bednar 1994, p.431) and the approach can be considered to be ultimately beneficial to the customer's interest since it can offer cost reductions through the reduction in the number of deviations and a greater focus on performance during the manufacturing process, as well as the subsequent avoiding of rework and warranty expenses (Garvin 1984, pp.27–28).

However, the approach is not applicable to all situations. While it benefits some types of industries, it is not possible to evaluate all goods in this way (Reeves and Bednar 1994, pp.431–432). The customer's interests are represented as far as they can be included in specifications, but the approach doesn't consider situations where the customer does not know or care about conformance to internal specifications. In addition, the customer will
evaluate the totality of the product which may not be limited to the specifications that the product met at some point along the chain (Reeves and Bednar 1994, pp.431–432).

In the translation industry, this manufacturing or production-based approach has had substantial impact and is reflected in process oriented standards like EN 15038 and ISO 17100 (Fields et al. 2014, p. 409).

2.1.1.3. User-based Approach

The limitations of the manufacturing-based approach with a focus on conformance to specifications could lead to the conclusion that the primary focus when defining quality should be the perspective of the user or customer. And indeed, a pervasive definition of quality is the extent to which a product or service meets and/or exceeds a customer's expectations (Buzzell and Gale 1987; Grönroos 1990; Zeithaml et al. 1990 - all in Reeves and Bednar 1994, p.423).

This school of thought has been described as the opposite of a product-based view, since quality depends on the person judging it and the highest quality is achieved when an individual's different wants or needs are satisfied (Garvin 1984, p.27). Here researchers assert that "quality is whatever the customers say it is, and the quality of a particular product or service is whatever the customer perceives it to be" (Buzzell and Gale 1987 in Reeves and Bednar 1994, p.427), as well as that "only customers judge quality; all other judgments are essentially irrelevant" (Zeithaml et al. 1990 in Reeves and Bednar 1994, p.427).

These definitions strongly emphasise the wants and needs of the individual. This puts this approach in contrast with the transcendent approach which assumes the highest level of excellence to be the indicator for quality, while a user-centred approach also considers the price of the product or service, and by extension cost of production (Zeithaml 1988, p.5). As a result, it considers the 'fitness for use' of a product (Garvin 1984, p.27; Reeves and Bednar 1994, p.425) by acknowledging that customers have different conditions, wants, needs and budgets.

Zeithaml (1988) argues that for the user-based approach, quality evaluations can be a result of context or comparisons where a product
"is evaluated as high or low depending on its relative excellence or superiority among products or services that are viewed as substitutes by the consumer."

(Zeithaml 1988, p.5).

This perspective can be viewed as closely related to the aesthetics dimension, addressing the question of how a product looks, feels, sounds, tastes or smells as judged by an individual (Garvin 1984, p.32).

Taking a user-based approach to quality provides multiple benefits. It allows customers to articulate their expectations, while they are not always able to relate to specifications or standards (Reeves and Bednar 1994, p.432). These expectations might include subjective factors like helpfulness, courtesy, confidence and appearance, which are critical to customers' judgments but can be difficult to quantify and assess. By including different attributes and weights, companies can compete on different dimensions of quality (Reeves and Bednar 1994, p.433).

However, this view on quality is highly subjective and not without its problems. Since different customers place different weights on various attributes, it can be difficult to aggregate preferences, or reach a meaningful definition of quality (Garvin 1984, p.27; Reeves and Bednar 1994, p.433). On a more fundamental level, setting quality equal to maximum satisfaction could be a fallacy since it can't be assumed that a product that is preferable to a different one is also of higher quality (Garvin 1984, p.27).

While the user-based approach is very relevant for the translation industry, it is not clear to what extent the end-user’s perspective is helpful to translation quality assessment (Fields et al. 2014, p. 408).

### 2.1.1.4. Product-based Approach

Where the user-based approach provides challenges in measuring and defining quality, the product-based view offers precise and measurable attributes that can be evaluated to determine quality.

The product-based view relates to differences in ingredients or attributes possessed by the product and therefore allows for goods to be ranked according to attributes they possess. In this view, there is a correlation between cost and quality, where quality is considered to be an inherent characteristic of the goods and not something ascribed to them.
Therefore it can be assessed objectively and is not just based on preference (Garvin 1984, pp.25–27).

A product-based view can include several dimensions, like performance, features, durability and serviceability. Performance relates to the primary operating characteristics of a product which are both measurable but also to an extent a matter of individual preferences. Differences can also be ascribed to different performance classes rather than differences in quality. Similarly, features relate to secondary characteristics of a product that supplement the product's basic functions, which are measurable but also affected by individual preferences (Garvin 1984, p.30).

Finally, durability describes the "amount of use one gets from a product before it physically deteriorates" and serviceability describes the speed, competence and courtesy of repair (Garvin 1984, pp.31–32).

In the translation industry, the product-based perspective emphasises metrics to quantify aspects of accuracy and fluency (Fields et al. 2014, p. 408).

2.1.1.5. Value-based Approach

The value-based approach defines quality in terms of cost and price, where performance is provided at an acceptable price and conformance at an acceptable cost (Garvin 1984, p.28).

This approach includes multiple attributes of a product like excellence, price and durability and requires firms to consider both, internal efficiency and as a result cost implications as well as external effectiveness in meeting customer expectations. It also allows for the comparison of disparate objects and experiences in different performance categories (Reeves and Bednar 1994, p.28).

The difficulty here is that two related but distinct concepts are mixed, where quality as a measure of excellence is equated with value as a measure of worth. This concept of affordable excellence is difficult to apply in practice (Garvin 1984, p.28). In addition, it can be difficult to understand what components are important and which weights should be applied to them. There is also disagreement on whether quality is a subcomponent of value, value a definition of quality, or if value is a subcomponent of quality although they are
commonly treated as separate constructs (Reeves and Bednar 1994, p.430).

In the translation industry, the value-based approach takes the perspective of the requester. Where the provider might hold a transcendent view of quality, and quality is unaffected by factors like price, the provider perspective sees quality as a result of many factors like price, timeliness and end-users’ reactions (Fields et al. 2014, p. 410).

2.2. Core Assumptions of this Thesis

A central issue of the approaches outlined is their focus on either the side of the producer or the user, while not sufficiently considering the point where the two sides meet. While approaches to producing quality or gathering the customer's requirements certainly merit in-depth consideration, this current research focuses on the transaction between producers and consumers of quality, to address the topic of perceived quality as it relates to different parties in a transaction.

At its core, this research assumes that quality is produced by one party (producer or seller) for a second party (consumer, customer or buyer), while potentially being evaluated by a third party who might make recommendations to either the producing or the consuming party. Going forward, this thesis will use the terms ‘producer’ or ‘seller’ interchangeably to mean the party producing, selling or providing a product or service and ‘buyer’, ‘consumer’, ‘client’ or ‘customer’ to mean the party consuming, purchasing, or accepting a product or service independent of whether monetary values are also exchanged or whether this buyer is the end-user of the product.

This research also assumes that a major interest of the seller is to achieve customer satisfaction in the transaction, with the goal of securing the current and potential future transactions, while also providing this product or service in a manner most favourable to maximising profit. The buyer however wishes to receive a product or service that provides them with the most favourable balance between their investment and quality and therefore wishes to understand the balance of these two factors for the product or service the seller is providing, relative to competing options.

The question is then what happens when these two interests, which are potentially directly opposed, meet and the potential conflict needs to be resolved. However, this is not
meant to assume that either party is dishonest or trying to gain unfair advantages over the other party. The central issue is not the ethics of such a transaction but rather how the quality the seller has to offer is then perceived by the buyer, or the concept of perceived quality.

2.2.1. Perceived Quality Dimension

Perceived quality has been defined in differing, yet similar ways. While some researchers argue that objective quality may not exist but that in fact all quality is based on perception (Zeithaml 1988, p.4) and specifications are based on what managers perceive to be important (Zeithaml 1988, p.5), others draw a clear distinction between perceived quality and objective quality and consider it just one of many dimensions of quality (Dodds and Monroe 1985; Holbrook and Corfman 1985 - all in Zeithaml 1988; Golder et al. 2012, p.1).

Several components can be observed with slight variations in a range of definitions of perceived quality in the literature.

Quality here is perceived by the customer or consumer. This customer perceives or judges the overall quality, superiority, excellence and characteristics of the product or service and considers an offering relative to alternatives and relative to the intended purpose (Aaker 1991 in Vantamay 2007, p.2; Garvin 1984, p.32; Zeithaml 1988, p.5; Ophuis and Trijp 1995, p.178). Customer assessment will be generally subjective and separate of the actual product quality. Customers commonly lack information about a product's attributes (Garvin 1984, p.32) and the perceived quality varies according to the individual consumer's perceptual abilities, experience level and personal preferences (Ophuis and Trijp 1995, p.178). As a result, customers commonly use surrogate or indirect quality indicators to come to a judgment of perceived product quality when they cannot evaluate all dimensions of quality (Ophuis and Trijp 1995, p.178).
2.2.2. Intrinsic vs Extrinsic Cues

An important distinction to consider when discussing the different elements that play into the concept of perceived quality is the difference between intrinsic product attributes and extrinsic cues.

Intrinsic product attributes are product-specific, cannot be changed without changing the nature of the attribute itself (Zeithaml 1988, pp.6–7) and provide intrinsic cues regarding the composition of the product (Sirohi et al. 1998, p.226). Intrinsic cues are product specific but they can be generalised across dimensions of quality to product classes or categories. For example, thickness of juice can be positively related to quality for tomato juice but not for children's orange juice (Zeithaml 1988, p.7). Intrinsic product attributes will be described in more detail in section 2.3.1.

In addition to intrinsic attributes and cues, products and services generally also carry extrinsic cues. Such extrinsic cues are related to the product but not part of the product itself and can refer for example to price, brand name and level of advertising (Bhuian 1997; Schiffman and Kanuk 2000; Lin and Kao 2004 – all in Vantamay 2007, p.114; Zeithaml 1988, p.6; Sirohi et al. 1998, p.226). Extrinsic cues will be described in more detail in section 2.5.2.

2.2.3. The Integrative Quality Framework

To better understand the relationship between the product attributes or intrinsic cues and the extrinsic cues, it is beneficial to consider an integrative quality model that includes both the seller's and the buyer's perspective.

One framework that not only generalises across different types of offerings (products and services) but also types of firms (for-profit and non-profit) and customers (businesses and individuals) as well as multiple dimensions of quality was developed by (Golder et al. 2012) and will be described in more detail as the foundation for the conceptualisation of perceived quality for this research.

Golder et al. (2012) define quality as "a set of three distinct states of an offering's attributes" (p.2), where quality refers to the relative performance during the production, evaluation and experience of the attributes of a product or service. These states are not
combined into an overall concept of quality but instead each state is treated separately (p.2).

The three states of quality include the quality production process, the quality experience process and quality evaluation process, represented by the three rectangular fields in Figure 1. Each of these states influence and are influenced by the other two states through the interaction between the producing firm and the customer, which is represented by the arrows in Figure 1. Customers are impacted by but also impact each state.

1. Quality Production Process: The firm produces attributes in a product which are then delivered to the customer.

2. Quality Experience Process: The customer then experiences these attributes and perceives some of them.

3. Quality Evaluation Process: The customer comes to an evaluated aggregate quality and satisfaction by comparing the perceived attributes with expectations.

   (Golder et al. 2012)

However, the customer can also influence attribute design through the evaluated aggregate quality and satisfaction as well as perceived attributes and emotions while the customer's expectations also influence the perceived attributes (Golder et al. 2012, p.3).

The following sections outline the three distinct states in more detail.
Figure 1: Integrative Quality Framework as presented in Golder et al. (2012, p.3)
2.3. The Quality Production Process

During the quality production process (Figure 2), a producer uses resource inputs and converts them into produced attributes via their attribute design and process design specifications. Resource input can be human resources like physical labour, knowledge and insights, as well as material inputs like raw materials and intermediate products supplied by other producers. In some cases, the customer supplies the resource input by either supplying material goods or, through the process of co-production, human resources (Golder et al. 2012, p.4).

During the quality production process, quality control consists of monitoring of produced attribute quality and maintenance or improvement of design specifications. Here the producer will conduct quality evaluations that are aimed at determining whether the produced attributes meet the expectations set out in the attribute design specifications, or whether the process design requires adjustments since the process design is created to implement the attribute design. Attributes can be universal, preference or idiosyncratic, when considered along the two variables of whether they can be measured and whether preferences differ, which will be further outlined in section 2.3.1. During the quality production process, both resource inputs and produced attributes are inspected and those that do not meet specifications rejected. Therefore, during the quality production process, produced attribute quality is the attribute performance of an offering relative to the design specifications (Golder et al. 2012, p.4).

Figure 2: Visualisation of quality production process
While the quality state and measurement during the quality production process resides with the producer, the customer can influence the attribute design through different means, like co-production and contributing to attributes or resource inputs. The customer's involvement allows the producer to monitor the customer's emotions and adjust produced attributes, and if the customer contributes to the production of attributes, they are more likely to meet his expectations. In addition, co-production can have positive effects on evaluated aggregate quality and satisfaction by shifting the customer's expectations, reducing uncertainty, and increasing the customer's ability to perceive a product's attributes (Golder et al. 2012, p.13). These concepts will be described in greater detail at later points in the chapter.

2.3.1. Product Attributes

The attributes of a service or product can be classified according to how easily they can be measured (ambiguous and unambiguous) as well as customer preferences (homogeneous and heterogeneous). Unambiguous attributes are generally measured with instruments like thermometers, scales or seismometers, while ambiguous attributes are measured through human senses (Hotelling 1929; Salop 1979; Desai 2001; Kahn and Wansink 2004; Sutton and Staw 2007 - all in Golder et al. 2012, p.8).
Figure 3 shows how these two variables result in Preference, Universal or Idiosyncratic attributes. Preference attributes are those which are unambiguous in measurement but customers have heterogeneous preferences (Golder et al. 2012, p.8). An example could be room temperature, which can be measured unambiguously but different individuals would prefer a warmer or cooler room to feel comfortable.

Universal attributes are both unambiguous in measurement and customer preferences are generally homogeneous. Examples include attributes like legroom or flight delays during aeroplane travel, which can be measured unambiguously and general preference would be homogeneous in that most customers would prefer as much legroom and as few delays as possible (Golder et al. 2012, p.8).

Finally, idiosyncratic attributes are ambiguous in measurement and customer preferences are heterogeneous. Those attributes include elements like art, beauty or social interactions like the appearance of service staff or their level of friendliness (Golder et al. 2012, p.8).

2.4. The Quality Experience Process

Once the product with its set of attributes has been produced, the firm delivers these attributes for the customer to perceive (Figure 4). Generally, the producer will be able to measure a product's attributes better since they have more access to measurement instruments. In addition, the customer only perceives some of the delivered attributes, since they experience and perceive the product through their individual lens. This lens is influenced by their measurement knowledge, motivation and emotions. Measurement knowledge and measurement motivation refer to the customer's ability and desire to assess attribute performance while minimising bias and variance. In combination with the nature of an attribute and the associated degree of difficulty, these factors result in a situation where in many cases, delivered and perceived attributes or what the buyer and seller perceive are not the same (Golder et al. 2012, pp.4–5).

In his seminal and Nobel-prize winning work on the effects of quality uncertainty and information asymmetry on the market, Akerlof (1970) described the detrimental effects this quality uncertainty can have on the seller-buyer relationship. Since both high and low quality products have to sell at the same price in the case that the buyer can't tell the difference, the low quality products will drive the good quality products out of the market (Akerlof 1970, p.490). As an example, for the localisation context, efforts towards producing greater quality localisation could be considered wasted if the decision-making buyer isn't able to distinguish the product that is of higher quality from that of lower quality, and is therefore not willing to pay for a higher quality output. Understanding which signals buyers perceive to be indicators of quality is therefore important for a successful transaction that satisfies both sides.
The following sections will discuss the three intervening factors in attribute perception (measurement knowledge, measurement motivation and emotions) in more detail.

2.4.1. Measurement Knowledge

In section 2.3.1, factors that would make an attribute easier or more difficult to measure have already been described. The classification of product attributes as universal, preference or idiosyncratic, can provide a first explanation of why measurement knowledge might vary.

Golder et al. (2012) define measurement knowledge as the customer's ability to assess attribute performance with minimal bias or variance relative to more objective measures, which could be instrument reading or expert consensus (p.9) and propose that depending on the nature of the attribute, customers will be more or less able to perceive and evaluate its performance (p.8). They also suggest that the importance of measurement
knowledge may be related to the type of attributes, where it might be less important for universal and preference attributes since they can be measured and communicated unambiguously (p.9).

An attempt at explaining why measurement knowledge differs between buyers and sellers, depending on the types of attributes, might be made when considering literature on knowledge management and the acquisition of skills.

The literature on knowledge management commonly distinguishes between codified knowledge, explicit and tacit knowledge.

Codified knowledge is subject to quality control by editors, peer review and debate. It is commonly incorporated into educational programmes and given status through examinations and courses (Eraut 2000, p.114).

On the other hand, personal knowledge is "the cognitive resource which a person brings to a situation that enables them to think and perform." (Eraut 2000, p.114) This personal knowledge may be explicit or tacit (Eraut 2000, p.114). While tacit knowledge can refer to both knowledge which is not communicated or knowledge which cannot be communicated (Eraut 2000, p.118), there is evidence that some kinds of knowledge are easier to communicate than others (Eraut 2000, p.119). Making tacit knowledge explicit can mean that the person who has the knowledge learns to talk about it (Eraut 2000, p.118). Among other factors that influence how well people can talk about what they know, their prior experience in doing so and a climate that encourages regular mutual consultation are important (Eraut 2000, p.120).

This suggests that those attributes that are ambiguous to measure might also be more difficult to talk about, since the vocabulary to describe them is insufficiently defined. Where unambiguously measurable attributes have been clearly defined, and are part of a codified body of knowledge, ambiguous attributes, perceived by human senses, might be more difficult to describe due to their being personal knowledge. As a result, it might for example be easier for the majority of individuals to measure and describe which factors make an offering too hot or too cold relative to their expectations, than it would be to measure and describe the aesthetics of another product or pleasantness of an interaction.
The Dreyfus and Dreyfus (1980) model of skill acquisition describes the first three stages (novice, competence and proficiency) as requiring rules and guidelines in order to be able to connect general situations to specific actions. These guidelines are needed to an increasingly lesser degree the further the learner moves through the stages (Dreyfus and Dreyfus 1980, pp.7–12). Only from the fourth stage (expertise) does previous experience allow for an intuitively grasped appropriate action for each situation (Dreyfus and Dreyfus 1980, p.12).

In the process of moving from novice to expertise, guidelines would likely be acquired from different sources like education through primary and secondary schooling, tertiary education, knowledge sharing with other professionals, standardisation, certification, educational and professional bodies as well as on-the-job instruction. With increasing experience, the expertise stage is then reached. A customer is likely to share some knowledge with the seller, based on equal or similar experiences like primary and secondary schooling as well as the buyer having potentially gained further knowledge through a range of activities like self-education or gaining certifications. However, since the producer has obtained the skills to actually produce and provide the product, their range of training, education and self-education is likely to be more extensive than that of the customer, in addition to having gained experience through more intensive interaction with the product. This would then likely place the seller on a higher stage of skill acquisition than the buyer, putting the seller in a position to more easily and intuitively grasp attributes that cannot be measured with instruments, such as idiosyncratic attributes, while the buyer would be able to do so only on a lower level. In addition, the seller would likely have more expertise in using instruments that allow universal and preference attributes to be measured than the buyer.

This is also supported by the observation that experts in a domain have fewer information gaps and are therefore less susceptible to external cues (Deval et al. 2012, p.1186). In addition, Golder et al. (2012) suggest that involvement in co-production might increase a customer's measurement knowledge allowing them to perceive the delivered attributes more accurately (p.13). This would also include customers with experience in the service category, which as a result have higher levels of expertise (Bettman and Park 1980;
Brucks 1985 - all in Eisingerich and Bell 2008, p.2).

2.4.2. Measurement Motivation

In addition to measurement knowledge, measurement motivation, as in "the customer's desire to assess attribute performance with minimal bias and variance relative to more objective measures" (Golder et al. 2012, p.9) or the customer's level of interest and time to evaluate the product (Zeithaml 1988, p.9) plays a role in the assessment of product quality.

Since measuring attribute performance involves devoting resources to that activity, a high desire to assess attribute performance implies a high level of resources being invested (Golder et al. 2012, p.9). Measurement knowledge and motivation likely also go hand in hand, since low measurement knowledge implies that it would require a higher investment in time and energy to be able to evaluate the product.

2.4.3. Emotions

In the context of the study of perceived quality, emotions refer to a set of feelings which are an outcome of perceiving certain attributes and evoke feelings in the customer during the quality experience process. They then act as filter for further attribute perception and evaluations (Golder et al. 2012, p.9).

Emotions can impact the quality experience primarily in three ways:

1. They may lead to confirmatory bias through which those attributes that validate feelings are perceived and evaluated and those that disconfirm them are neglected (Cohen et al. 2008 in Golder et al. 2012, p.9).
2. Negative emotions may result in the failure to perceive delivered attributes (Loewenstein 1996 in Golder et al. 2012, p.9) and
3. Positive emotions may result in some attributes being under weighed during quality evaluation (Arnould and Price 1993 in Golder et al. 2012, p.9).
Examples of the impact of emotions on the customer experience can be found in the literature, for example when employees display positive emotions, this can have a positive effect on customers (Arnould and Price 1993; Hennig-Thurau et al. 2006 - all in Golder et al. 2012, p.13). In some cases, value perceptions trigger feelings like shame or guilt among customers of discount stores (Zielke 2014).

2.5. The Quality Evaluation Process

The final process, the quality evaluation process (Figure 5), has as a result the summary judgment of quality and customer satisfaction, which are reached by the customer comparing the perceived attributes of an offering with their expectations.

Expectations serve as reference levels which can come from the customer's experiences or external sources like third parties, competing offerings or the firm itself. Golder et al. (2012) describe three types of expectations:

- 'Will' expectations (predicted attribute performance);
- 'Should' expectations (attribute performance a product ought to deliver relative to competing offerings) and
- 'Ideal' expectations (ideal attribute performance across all offerings in a category)

(p.4).

Expectations are also associated with expectation uncertainty, or "the variance in a customer's expectation of an attribute's performance." (Golder et al. 2012, p.10).
Relative to these reference levels, customers come to an aggregation of perceived attribute performance. This differs depending on the attribute type and is relative to the 'should' expectation.

With 'Should' expectations, norms like perceived equity and fairness come into play. For universal attributes, exceeding the 'should' expectation results in a positive perceived attribute performance, while falling below the 'should' expectation results in a negative perceived attribute performance.

For preference and idiosyncratic attributes, positive perceived attribute performance is linked to lower and upper bounds of a 'should' expectation (Golder et al. 2012, p.11).

While evaluation is also part of the production process, the focus of this research is on the evaluation of quality on the side of the buyer, customer or consumer.

2.5.1. The Role of 'Ideal' and 'Should' Expectations in Perceived Quality

'Ideal' expectations refer to the customer's ideal preferences across all offerings. Relative to these preferences, perceived attributes are then assessed. For universal attributes, 'ideal' expectations are universal and ideal uncertainty close to zero, while for preference and idiosyncratic attributes, 'ideal' expectations differ and expectation uncertainty is higher (Golder et al. 2012, p.11).

'Should' expectations refer to what different offerings ought to deliver, according to the customer. These are relative to price, competing offering and purchases of other customers - so 'should' expectations differ from 'ideal' expectations insofar that 'should'
expectations consider these additional factors. 'Should' expectations consider a narrow set of competing offerings, while 'ideal' expectations are the same for all offerings in a category (Golder et al. 2012, p.11).

### 2.5.2. The Role of 'Will' Expectations in Perceived Quality

'Will' expectations are "the attribute performance levels a customer predicts or believes an offering is going to deliver" (Golder et al. 2012, pp.10–11). These 'will' expectations result from prior experience and/or prior beliefs based on external information sources. These expectations can influence the customer's perception in three ways:

1. It may lead them to shift the observation towards their expectation (confirmatory bias);
2. In the case of low measurement knowledge or motivation a customer may not directly observe attributes but instead retrieve 'will' expectations;
3. In the case of high 'will' uncertainty, perceived attributes can be emphasised instead of stored expectations to deal with the uncertainty.

(Golder et al. 2012, pp.10–11).

Research suggests a positive impact on 'will' expectations if customers are able to accumulate information and reduce their 'will' uncertainty if they participate in the production process or provide resource inputs (Kelley et al. 1990; Bettencourt 1997 – all in Golder et al. 2012, p.13).

A priori beliefs or 'will' expectations based on external information sources have also been termed

- extrinsic cues (for example, (Zeithaml 1988)) or
- indirect quality indicators (for example, (Ophuis and Trijp 1995)).

As outlined in previous sections, commonly in the quality experience process the buyer won't perceive all internal attributes of a product due to limitations in measurement knowledge and measurement motivation, as well as the role emotions play.

In this event, the buyer will then use indirect quality indicators or extrinsic cues to

Extrinsic cues can serve as a shorthand or surrogate to signal product quality (Zeithaml 1988, p.7). Examples of extrinsic cues include a range of different factors like the price, brand name and levels of advertising. Price here can mostly be considered a surrogate, while brand name can be considered a bundle of information presented as a shorthand to customers (Zeithaml 1988, p.8).

However, other extrinsic cues are also in use, such as high levels of advertising, which signal quality for goods whose attributes are determined during use. The argument has been made that a high level of advertising signals that the firm considers the product worth advertising for (Schmalensee 1978 - in Zeithaml 1988, p.8).

Other factors like the country of origin of a product have also been shown to impact perceived quality. While consumers like the low price of goods from some countries like China, they are often sceptical about the durability of those goods (Khan and Ahmed 2012, p.120).

The reliance on intrinsic or extrinsic cues as well as the effect of extrinsic cues is also dependent on situation and the consumer’s background. Some evidence suggests that intrinsic cues have a stronger effect on quality perception than extrinsic cues (Szybillo and Jacoby 1974, p.77). However, dependence on extrinsic attributes is primarily strong when intrinsic cues are not available or the consumer, has little or no experience with the product. In this case, when evaluating intrinsic cues requires more time and effort than is worthwhile for the consumer or when quality is difficult to evaluate (Zeithaml 1988, p.9). Intrinsic attributes become important at the point of consumption where they allow the consumer to evaluate such attributes. In addition, intrinsic attributes can be used by consumers to search for specific products based on those attributes (Zeithaml 1988, p.9).

Different extrinsic cues will also not always have the same effect on all buyers. Research suggests that the way a cue like low price is perceived may be linked to the naive theory that is active at any given point (Deval et al. 2012, pp.1186–1187). Naive theories are

"informal, common-sense explanations people use in their everyday lives to make sense of their environment and often diverge from formal, scientific explanations"
The example of price as an extrinsic cue would mean that individuals might have theories as to why a product might be low or high price. There are complex interplays between competing naive theories, where one can dominate the other (Deval et al. 2012, p.1186). Experiment research has shown that individuals can be influenced, or primed, in their product evaluation depending on which naive theory is activated by the researcher. However, with increasing expertise the individual is less susceptible to being influenced by priming (Deval et al. 2012, p.1185).

Examples of primed theories affecting consumers' beliefs include price being associated with quality or exclusivity, where a low price might signal good value for discounted promotional prices or a high price signalling that a product might be exclusive. The consumer may then interpret the extrinsic cue in different ways depending on which naive theory is active (Deval et al. 2012, p.1189). In the context of discount stores, such attributions have also been associated with positive or negative associations and emotions. Attributions are causes that customers ascribe to observed phenomena (Heider 1958; Kelley 1973 - all in Zielke 2014, p.328). Positive and negative associations and emotions can manifest as guilt when low prices are associated with low quality due to use of inferior ingredients or lower quality management (Zielke 2014, p.328). In some contexts low prices are associated with lower customer enjoyment (Diallo et al. 2015, p.1142).

2.5.2.1. Price as Extrinsic Cue

Overall, price as an extrinsic cue and the relationship between price and perceived quality has been the topic of a relatively large body of research in the literature, with conflicting results.

Strong evidence suggests that consumers tend to overestimate the price-quality relationship and that depending on their a priori beliefs on the general relationship between price and quality their subjective perceptions will vary, with the maxim that 'you get what you pay for' strengthening the association between high price and high quality (for example: Peterson 1970, p.525; Kardes et al. 2004, p.368; Cronley et al. 2005, p.159; Judd 2015, p.301). Commonly the explanation offered is that quality offerings are more
expensive to produce.

However, relative to other factors like brand name or store name, the price effect on perceived quality may be smaller (Rao and Monroe 1989, p.355) and the availability of other cues may weaken the importance of price as an extrinsic cue (Sirohi *et al.* 1998, pp.227–228).

It has also been argued that methodological deficiencies in previous research point towards the price-perceived quality relationship being not firmly supported by evidence (Riesz 1980, p.261; Rao and Monroe 1989, p.351) In fact the relationship may be nonlinear and variable across individuals and products (Peterson 1970, p.526; Reeves and Bednar 1994, p.436)) as well as there being no general price-perceived quality relationship (Zeithaml 1988, p.11).

The variables that have been suggested include:

- price awareness of the consumer as well as awareness of quality variations (Zeithaml 1988, p.12);
- the brand-quality relationship (Gardner 1971, p.243);
- which hypothesis is considered first and whether it is plausible;
- information load (how much information is available to serve as basis for judgment);
- whether the information is belief-consistent or inconsistent;
- how the information is organised and
- how big the desire for a definite opinion rather than ambiguity or inconsistency is.


### 2.5.2.1.1. Free as a Special Price

As a final aspect, the notion of 'free' as a special price, and how this it might impact perceptions, needs to be considered for this research, since volunteer translation or crowdsourcing are generally provided for free.

If, as outlined in section 2.5.2.1., high price is indeed correlated with high perceived quality, but low price is preferable, then a free product or service might simply be placed at
the most extreme end of the linear relationships, since if performance is positively related to effort, then a price of zero would suggest the lowest effort being invested in the product. However, some researchers suggest that 'free' holds a special position in the cost-benefit difference influencing choices (Shampanier et al. 2007, p.742), where zero price does not only mean low cost but also increased valuation and abnormally raise the demand (Shampanier et al. 2007, p.743; Nicolau and Sellers 2012, p.243).

Potential explanations could be cognitive dissonance theory which states that giving an individual no reward for an activity is better than giving them a small reward (Deci et al. 1999; Gneezy and Rustichini 2000, p.791; Shampanier et al. 2007, pp.742–743) or that different norms may apply for low cost offerings (market norms) and free offerings (social norms) (Shampanier et al. 2007, p.743). Under a social contract, individuals are then more likely to exert higher effort than under a low-payment market contract (Shampanier and Ariely 2006, p.17). While a buyer may not be aware of those concepts, they might intuitively grasp them through their own experiences, resulting in a more positive evaluation of free offerings compared to low-cost ones.

2.6. Summary

In summary, while multiple models of describing quality have been presented in the literature, the focus of this research is on the concept of perceived quality. Here the central issues are the discrepancy between the set of attributes a seller delivers to the buyer, and the sub-set of attributes the buyer perceives. Perception of attributes may be impacted by several factors, like the type of attribute in question as well as the individual's measurement knowledge, motivation and emotions. This discrepancy leads to quality uncertainty, which can be resolved through several mechanisms. For this research, the role of extrinsic cues to replace measurement of intrinsic attributes is primarily of interest, with a focus on price and the professional status of the producer.
Chapter 3: Quality in Localisation

3.1. Introduction

As outlined in the theoretical framework in chapter 2, different concepts of quality can be found across the literature. Similarly, a range of perspectives exist within the localisation field. For the purposes of this research, perceived quality in crowdsourced localisation is the primary focus. Section 4.1.5. will discuss empirical evidence on quality in objectively measured crowdsourced localisation. However, to the best of this researcher's knowledge no previous studies on perceived quality in crowdsourced localisation have been published. While studies of perceived quality and factors impacting it in other fields have been conducted (for example: Bucaria and Chiaro 2007; Havumetsä 2012; Gaspari et al. 2014; Teixeira 2014), the areas of study and resulting extrinsic cues were considered too loosely related to the present research to merit in-depth discussion.

To approach the topic of perceived quality in crowdsourced localisation and factors impacting perception, the following chapter will first situate perceived quality into the wider context. It first provides an overview of common models of describing and measuring quality in localisation. Here, a lack of models with universal agreement, and the need to consider perceived quality in the context of localisation becomes apparent when considering the commonly used outsourcing model in localisation. Inherent in this is the central issue of information asymmetry, stemming from differences in individual parties' measurement knowledge and motivation. Here, the primary focus is on the subject side of the quality experience and evaluation processes; the person perceiving the quality.

Chapter 4 then provides an overview of volunteer and crowdsourced localisation as well as different models for categorising crowdsourced initiatives. It then focuses on two extrinsic cues being available to the person perceiving quality: the professional status of the translator and the price of the translation. As chapter 3 describes, external cues are of greater importance with lower measurement knowledge on the side of the person perceiving quality.
3.2. The Theoretician's Perspective

With translation being an integral part of localisation, the field of translation studies provides a large body of literature in relation to localisation quality.

Additionally, while this research focuses on the practitioner side of localisation, academia plays two roles in the concept and study of quality in localisation: that of the observer and that of the informer (Koskinen and Dam 2016, p.259).

In the observer role, academics conduct research into practices within the industry, some of which will be featured later in this section. The second role, that of the informer, involves the formation of, support of and participation in different organisations, events and bodies that enable academics and members of the localisation industry to collaborate to improve on different aspects of managing and producing quality at different stages of the localisation process. Examples of such collaboration include, but are by far not limited to, the committee at OASIS\(^\text{\textsuperscript{14}}\) for the XLIFF\(^\text{\textsuperscript{15}}\) standard, which is continuously working towards improving interoperability within the process and as a result process and product quality, or the advisory team at TAUS\(^\text{\textsuperscript{16}}\), a resource centre aimed at enabling better translations.

In addition, the informer role involves the education or training of large numbers of individuals involved in the process, including engineers, project managers and especially linguists (translators and reviewers). It is therefore beneficial to consider the theoretician's perspective on translation quality since the theoretical views in academia will have likely been imparted to different members of the industry on different occasions.

Within the field, researchers frequently comment on the fact that "the concept of quality varies greatly, within and outside translation studies." (Colina 2008, p.98) and that there is a lack of a general framework. It is noted that instead there are multiple partial approaches with generally limited applicability to practice (Al-Qinai 2000, p.498; Brunette 2000, p.169; Colina 2008, p.99; García 2014, p.432; Mateo 2016, p.147).

\(^{14}\) Organization for the Advancement of Structured Information Standards: https://www.oasis-open.org/

\(^{15}\) XML Localisation Interchange File Format: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff

\(^{16}\) Translation Automation User Society: https://www.taus.net/team
Colina (2008) describes models for assessing quality as broadly falling into two categories:

1. Theoretical approaches based on research. These then include reader-oriented or text-oriented approaches;
2. Anecdotal or experiential approaches, which include approaches commonly used by practitioners.

(pp.237–238)

Other distinctions between quality assessment models that have been made are ‘top-down’ vs. ‘bottom-up’ (Drugan 2013), as well as models with quantitative dimension and non-quantitative models (Williams 2004). Mossop (2014) describes ‘philosophies’ of revision that consider the goals of satisfying clients, whether a translation is ‘fit for purpose’, doing what is necessary to protect or promote a language, as well as procedural concepts of quality that focus on how the translation is produced (p. 22-23).

O’Brien (2012) groups translation quality models into those that view quality from the point of view of errors and those that view translation quality in a holistic model and consider service provision, competences, tools and procedures (p. 57).

House (2014) groups theoretical stances related to translation evaluation into those that take into account the relation between the source text and its translation, those that consider the relationship between features of the text(s) and how they are perceived as well as those that take account of the consequences views of these relationships have (p. 241).

3.2.1. Theoretical approaches: Text-oriented

Text-oriented approaches focus on detailed analysis and comparison of source text and target text, or only on the target text (Drugan 2013, p.46). Such approaches consider the question whether the source text has been rendered in a way that is "as accurate as possible, as economical as possible, in denotation and in connotation, referentially and pragmatically" (Nerudová 2012, p.10). The focus here is on the source text type and function (House 2001; Colina 2008, p.101, 2009, p.238) and whether the target text and its units reach 'optimum equivalence' with the source text considering the author's intention,
the context of the text and its linguistic and stylistic level (Drugan 2013, p.65).

These approaches have been criticised for focusing too much on the source text (Colina 2009, p.238) as well as being too subjective and founded on intuitive criteria (Al-Qinai 2000, p.498; House 2001 - in Nerudová 2012). In addition, this approach can be problematic when a translation is required to have a different function from that of the original text (Colina 2008, p.101).

3.2.2. Theoretical approaches: Reader-response

Response-oriented models are rooted in psycholinguistic and behavioural approaches, and measure the response of the reader to the target text, relative to the source text (Nida 1964; Carroll 1966; Nida and Taber 1969 - all in Colina 2009, p.237; Drugan 2013, p.46), thereby recognising the role the reader plays in translation and the effects a translation has.

Reader-response models have been criticised for relying on the 'black box' of the human mind and for being reductionist (Drugan 2013, p.46). In addition, it can be difficult to determine and measure responses, especially considering that readers of translated text will usually belong to a different group from the target group of the original text (Colina 2008, p.100). Also, these responses only address the equivalence of the effect on the reader, while ignoring other aspects like the purpose of the translation (Colina 2009, p.237) as well as being primarily concerned with specific text types (Al-Qinai 2000, p.498).

Independent of the criticism of the individual models, a major shortcoming has been a lack of integration into the practitioner's localisation processes.

Models proposed by theorists have been criticised for

- having a narrow view of what quality is, commonly conflating quality with revision and editing (Drugan 2013, p.37);
- being too complex or too vague (Colina 2008, p.103);
- unrealistic application recommendations (Drugan 2013, p.40);
- being only tested on a select few cases (Drugan 2013, p.49), and
- not considering professional concerns (Colina 2008, p.103; Drugan 2013, p.39).
In addition, some of the models don't consider the complex and multi-component reality of evaluation (Colina 2008, p.103). According to Drugan (2013), 
"during hundreds of interviews and research visits to LSPs [Localisation Service Providers - addition by author] for this book, not a single academic model was mentioned as a way of assessing translation quality in the real world." (p.36)

As a result of the gap between academia and industry, the need for research applicable to industry and end-user needs has been highlighted (Bucaria and Chiaro 2007, p.93; García 2014, p.431), also with a view of preventing the translation field from being a "playground for amateurs" (Drugan 2013, p.44). This need for closer collaboration between theoreticians and practitioners becomes especially apparent when considering that it is the theoretician who will commonly be responsible for training future practitioners in university programmes.

While some advances in this regard have been made, notably with the inclusion of technology like computer-assisted translation tools in translation programmes. Translation approaches have changed due to these and other new technologies like machine translation and collaborative approaches. However, they require broad adoption and adaptation in the way translation is taught in order to close the gap between theory and practice (García 2014, p.433).

The localisation industry has reacted to the lack of applicability of theoretical approaches by developing its own quality models, which will be discussed in section 3.3.2. However, before being able to discuss quality in the localisation industry, it is first necessary to detail what is meant by 'the localisation industry' in the context of this research.

3.3. The Business of Localisation

The following section outlines the parts of a basic workflow that are being considered for this research. Since organisations in the localisation industry tend to be protective of their business interests, getting in-depth looks 'behind the scenes' can be
difficult and literature on some topics is somewhat sparse. However, several studies have been conducted by the independent market research company Common Sense Advisory\textsuperscript{17}. With a few exceptions, the research conducted by Common Sense Advisory is generally published through their own platform rather than peer-reviewed publications. It addition, it could, due to the approach to recruiting and sampling of respondents, be considered problematic in terms of generalisability. However, commonly the information provided through those studies is currently the best information available and is also commonly referenced throughout the peer-reviewed literature. Therefore, Common Sense Advisory research has been included in this and following sections while bearing in mind its potential limitations.

For the purpose of this research, 'localisation\textsuperscript{18}' will be defined as

"the linguistic and cultural adaptation of digital content to the requirements and the locale of a foreign market; it includes the provision of services and technologies for the management of multilingualism across the digital global information flow. Thus, localization activities include translation (of digital material as diverse as user assistance, websites and videogames) and a wide range of additional activities" (Schäler 2010, p.209).

However, for this thesis a challenge was whether to use the term ‘localisation’ or ‘translation’ to describe the concept being investigated. As Melby et al. 2014 outline, the line between translation and localisation can be difficult to define and it is possible to view translation as part of localisation, localisation as part of translation or both concepts independently but overlapping. In addition, depending on how much content is adapted to fit the target culture and requirements, the term ‘transcreation’ might apply (p. 395). This researcher holds the view that translation and localisation are two separate, but overlapping, concepts. This suggests that typically, in order to be able to speak about localisation, content adaptation would involve translation plus other elements, for example the adaptation of images, audio or code. While the focus of this thesis is on the translation element of localisation, there are nonetheless additional elements present throughout. For

\textsuperscript{17} https://www.commonsenseadvisory.com/AboutUs/CompanyOverview.aspx
\textsuperscript{18} In the literature, localisation is often spelt with a ‘z’. However, in this thesis it is referred to using the UK spelling conventions, except in direction quotations.
example, interview participants were describing localisation specific elements like variables and markup, as well as localisation projects. In addition, projects requested on Trommons can be either ‘pure’ translation or involve localisation elements, for example when volunteers provide translations for a non-profit organisation’s website. Therefore, the decision was made to use the term ‘localisation’ throughout with the understanding that for the purposes of this thesis, the terms ‘translation’ and ‘localisation’ are to a large degree interchangeable.

While this aspect is not explicitly mentioned in Schäler’s definition, this research will assume that localisation primarily is provided by one party for the benefit of another party. A person localising content would in most cases make this effort primarily not for their own use, but for the content to be used by another person or group of people.

Furthermore, localisation is also a part of a multi-billion USD service industry (Sargent and DePalma 2007, p.1) that comprises a range of different types of vendors. While according to one study a small portion of translation work is conducted by internal staff, 85.9% of the buyers surveyed outsourced most or all of their translation work (Sargent and DePalma 2007, p.4). Most content is localised by multi-language vendors (MLVs), followed by freelancers and single-language vendors (SLVs). In addition, studies suggest that the outsourcing market in localisation is likely to grow due to the trend going towards more outsourcing of such services (Kelly and Stewart 2010, p.8).

Outsourcing localisation services can be beneficial for companies, since it allows them to avoid fixed costs in favour of variable costs as well as buying competencies that are not a core competency of the organisation purchasing services (DePalma 2012, p.1). This aspect of outsourcing non-core competency will play a key role in later sections.

In addition, vendors not only help with managing tasks but also implicitly promise to provide added value by managing quality: by vetting translators; doing checks; and providing standardised quality processes (Kushner 2013, p.1246). They often act as a broker "keenly interested in ensuring that the translator and client do not circumvent the agency" (Kushner 2013, p.1249), while providing translators with regular access to work (Kushner 2013, p.1249).

On the vendor side, responsibility for localisation projects commonly lies with
project managers, whose tasks include execution of project plans, acting as client contacts, providing clients with quotes, tackling production and client-related issues (and to a lesser degree language tasks like editing), handling of translation memories, selecting linguists, negotiating rates with vendors and providing them with feedback. In addition, project managers might perform a final review on projects or use automated quality assurance tools (Pielmeier and DePalma 2013, pp.6–9).

Despite this heavy involvement in localisation projects, survey results seem to indicate that project managers at times struggle with a lack of control over their projects and clients, in addition to being under pressure to earn a healthy profit margin for their company (Geller et al. 2011, pp.8–9). Project managers also commonly deal with stressful numbers of projects to handle at the same time (DePalma and Beninatto 2004, p.7).

In section 2.2., the assumption was made that at its core, quality is produced by a producer or seller for a consumer, customer or buyer, with potentially a third party being involved in evaluation. However, with the above description of localisation tasks being outsourced to a multi-language vendor in mind, additional parties need to be considered. Not only do project managers at MLVs commonly act as a link between the customer and the translator, but in addition, the customer is commonly not the end-user of the localised content and may act on behalf of the end-user or another internal customer. As a result, localisation projects can involve several parties in different combinations, with Figure 6 showing the shortest possible chain and Figure 7 a realistic long one, which could be extended by further steps of outsourcing or internal clients. In addition, localisation projects commonly involve additional resources like reviewers, engineers and terminologists. However, for the purpose of this research, the focus will be on two connecting points in the localisation service chain:

1. The transaction between a translator (filling the role of a seller) and a project manager at the MLV (filling the role of a buyer);
2. The transaction between the project manager at the MLV (now filling the role of a seller) and the contact person on the client side (filling the role of a buyer) (Figure 7).
As described in the section on quality outside of localisation (Section 2.2), in more general terms, this research assumes for these transactions that sellers wish to achieve customer satisfaction in the transaction to secure the current and future business, while the buyer wants to receive a product that is of the highest possible quality. At the same time, both parties also wish to maximise their profit by selling/buying the service at the rate that is the most favourable to themselves, relative to the provided quality. Assuming that the sellers in the transaction are competent and all parties are honest, it is therefore in the seller's interest that the buyer perceives the service product they are buying as high quality, while it is in the interest of the buyer to understand whether the service product is of sufficiently high quality to justify the price asked. These competing interests - most favourable rate for both seller and buyer, relative to quality of the product - then meet and need to be resolved, in part through coming to a shared understanding of the quality of the product.

While the vendor makes an initial decision on which translator, process and
technology to use for any given project, the client will make the final decision on whether the resulting product is appropriate and acceptable, possibly in collaboration with the vendor, end-user and additional evaluators.

3.3.1. Localisation Service and Product

One aspect to consider when talking about localisation is whether it should be thought of primarily as a service or as a good. In the quality literature, different theories exist, depending on whether goods or services are being discussed. While in chapter 2 the primary focus was on the quality of goods, there exists a separate body of literature on service quality specifically, (for example Parasuraman et al. 1985; Brady and Cronin 2001; Eisingerich and Bell 2008). These were not further discussed due to the following considerations. Based on different attributes assigned to goods and services, the researcher proposes to view localisation as a service that turns an intangible good into another intangible good (or service product).

In the literature, goods are generally defined as physical objects for which a demand exists, and whose physical attributes are preserved over time. It is possible to establish and exchange ownership since goods exist independently of the owner, and unit ownership rights can be exchanged between institutions. Goods can be traded on markets (Parry et al. 2011, p.20). While the definition is not uncontested, services are commonly described as intangible, heterogeneous, inseparable and perishable (Parry et al. 2011, pp.20–21).

'Intangible' refers to the concept of services not being physical objects but only existing in connection with other things. According to Parasuraman et al. (1985) "most services cannot be counted, measured, inventoried, tested, and verified in advance of sale to assure quality" (p.42). This however poses the challenge of agreeing whether things such as music, books or films are a product or a service (Parry et al. 2011, p.21). While they have no physical dimension of their own, "Hill (1999) argues that in their saleable form these intangible products have the salient economic characteristics of goods and little in common with services" (Hill 1999 - in Parry et al. 2011, p.21).

This raises the question whether a localised product like a video game, website or e-learning video should also be considered an intangible good as argued by Hill above. On
the surface, there appears to be no significant difference between those products and the above-mentioned music, books or films. However, for many localised products the elements that are localised are inseparable from the language agnostic elements that form part of the product in terms of saleability. As an example, while a video game as a whole could be considered an intangible good, the parts that are localised, which would generally involve text, audio, images/graphics and potentially small portions of code, cannot be considered saleable by themselves, without being embedded in the original intangible good. From the perspective of saleability and being intangible, localisation would therefore be most suitably described as a service that turns one intangible good into another intangible good by altering some of its elements.

Other factors associated with services and goods also point localisation in either one or the other direction. Services have been described as heterogeneous in the sense that they vary according to the context, nature and requirements of each customer (Parasuraman et al. 1985, p.42; Parry et al. 2011, p.21), although this is also possible for tangible goods like cars which might be customised to the individual user's needs (Parry et al. 2011, p.21). Localisation is generally also heterogeneous in that it is provided for each customer according to their needs, and based on their individual source language content so the results are not exchangeable.

Services being inseparable implies that they are inextricably linked with customers in terms of production and consumption. While a service provider may not provide their service without the customer, a manufacturing company can still produce and deliver their goods without the customer being engaged or present (Parasuraman et al. 1985, p.42; Parry et al. 2011, p.22).

Localisation does not necessarily require the customer or end-user to be present or engaged. Once requirements have been clarified and source content provided, the localisation team can create a localised version to be delivered as a final product. However, a level of customer engagement is generally necessary because many localisation projects require access to pre-compilation versions and resource files of the content to be localised to avoid difficult and costly workarounds. In addition, the provision of source content and requirements, although not strictly necessary in cases where the source language content is
otherwise accessible, is generally part of the localisation process in some form. This is the case for fan-generated localisation described in section 4.1.2.

Finally, perishability of services refers to the concept of services perishing once they are performed and cannot be stored. An example used are aeroplane seats where if the seat is not sold, it cannot be stored for the future once the plane has departed. However, not all services are perishable and for many, the result is visible after the service has been provided, as with the example of a house cleaning or hair cutting service (Parry et al. 2011, p.22). Similarly, in localisation the result of the localisation team's efforts is visible after it has been completed and is not perishable.

It becomes clear that the distinction between goods and services isn't perfectly possible for all products, based on the current state of the discussion (Parry et al. 2011, p.25). While there are some clear-cut cases like commodity items (e.g. oil or wheat) on the one hand and purely intangible services (e.g. a back massage) on the other, a wide range of products fall into categories that could be considered either a good or a service. The argument has also been made that products and services are linked, and therefore their quality should be considered together (Sirohi et al. 1998, p.225; Hansen and Bush 1999, p.120; Golder et al. 2012, p.2), as is the case with the quality framework described in chapter 2.

While the question of whether localisation should primarily be considered a service or a good cannot be definitively resolved within the scope of this research, and indeed a resolution may not even be necessary or valuable, it should be acknowledged that in the context of this thesis only a narrow aspect of perceived localisation quality is being considered. Localisation at a minimum includes service aspects in large parts of the process of delivering an altered intangible good. While the study of the different service elements of the production process and how different service attributes like reliability, responsiveness, assurance and empathy (Parry et al. 2011, p.23) or trust (Eisingerich and Bell 2008, p.7) would be likely to yield valuable insights into the concept of perceived quality in practical application in the localisation industry, the scope of this present research has been defined only to investigate perceived quality in relation to the final product of the service. However, the fact that the service element of localisation cannot be entirely
separated from the end product was also noticeable especially in the qualitative elements of this research. This was then included peripherally in the analysis and discussion of the data.

### 3.3.2. Quality in the Localisation industry

Section 3.2. briefly outlined the different approaches to defining and evaluating quality in translation studies, as well as the lack of adoption by practitioners in the localisation industry.

A core issue here is that academics and industry are pursuing different goals when considering how to define and assess quality (Drugan 2013, p.37). From the academic perspective, the question is: 'When is a translation good?'. From the industry perspective, it is of greater importance what aspects of the translation and service will satisfy the customer's needs, making the question for the industry primarily one of when the translations are 'good enough' (Drugan 2013, pp.39–42).

As a result, quality assessment in the industry has to take a broader view, and fulfil several purposes. Apart from knowing whether a localised product is of an acceptable quality level for release, goals also include the prevention of future quality issues, being able to predict quality and quality related business goals.

#### 3.3.2.1. Prevention of future issues

One goal of quality assessment in the industry is the identification of quality problems post-translation with the view of preventing problems in the future (Drugan 2013, p.79). However, not all types of evaluation seem equally suited to this goal.

Chesterman (2007) describes five categories of Translation Quality Assessment (TQA), which include:

1. retrospective assessment (relation between source text and target text);
2. prospective assessment (effect of target text on the reader);
3. lateral assessment (comparison of translated text with original);
4. introspective assessment (evaluating decision making during translation process);
5. pedagogical assessment (translator education)

(Chesterman 2007 - in Drugan 2013, p.48).

Where retrospective, prospective and lateral assessment are primarily focused on after-the-fact evaluation of the translation and play an important role in the localisation process, introspective and pedagogical assessment approaches seem especially suited to analysing root causes of quality issues which could then be used to pre-empt issues in the future.

3.3.2.2. Prediction of quality

This goal of preventing quality issues also goes hand in hand with the need in the industry to be able to anticipate the quality they might receive from one resource or the other:

"Clients, of course want to know before investing significantly in translation that quality levels will be sufficient. (...) TQA, the focus of academic interest in this area, is necessarily performed post-translation, whereas clients require assurances in advance that suppliers can produce the goods, then ongoing project updates on how targets are being met."

(Drugan 2013, pp.70–71)

Dunne (2006) describes this discrepancy as a shift away from Quality Control towards Quality Assurance, as observed in the manufacturing industry in the 1970s, where the focus changed from inspection and correction towards risk management strategies in order to be able to identify and mitigate future issues proactively with the goal of ensuring final product quality (p.96).

This manufacturing-based approach to quality (Section 2.1.1.2.) that has at its core the conformance with specifications, would require in a localisation context that client requirements are formulated and communicated, which are then implemented by the vendor through quality project and product plans. When the product is produced, verifications then ensure compliance with those requirements and finally, client feedback identifies areas for improvement (Brunette 2000, p.173; Dunne 2006, p.98). In the context of localisation
quality, such an implementation would have benefits also for perceived localisation quality, since it would involve the client strongly in attribute design (Golder et al. 2012, p.3).

However, this model as detailed in the ISO 9001\(^{19}\) standard presupposes several points which are difficult to achieve in localisation projects:

1. The client is requesting a product that involves the transformation of raw materials or parts into the requested product (Dunne 2006, p.99), where the quality of those inputs can be measured quantitatively and objectively, which would give the vendor the opportunity to reject inadequate input materials (Dunne 2006, p.100). However, a crucial difference when applying this model to localisation is that it essentially deals with the adaptation of a pre-existing product rather than the conversion of raw material/parts into a new product. This pre-existing product is generally supplied by the client and rarely would a vendor be able to reject a project due to inadequate source content (Dunne 2006, p.100), despite quality implications for the localised end product;

2. The model relies on the client providing specifications and requirements to the vendor. However, often in localisation the client may not be able to do so, due to a lack of familiarity with target language, culture, legal and technical requirements.

"Indeed, this lack of knowledge and expertise is a major factor in many clients' decision to outsource localization in the first place"


Arguably, the onus to define specifications should then be on the experts, in this case the localisation service providers. And indeed, the localisation industry has reacted to the gap between theoretical approaches and practitioners' needs as well as the need to more clearly define its processes and requirements by devising and adopting its own quality definitions and evaluation strategies. These include, among others (Figure 8),

- the LISA QA model;
- the J2450 standard;

\(^{19}\) https://www.iso.org/iso-9001-quality-management.html
• EN 15038, ISO 12616;
  (Kelly and DePalma 2008a, p.1; Jiménez Crespo 2011, p.134; O’Brien 2012, pp.59–57)
• and recently ISO 17100\(^{20}\) in addition to a wide range of tailor-made models in use at individual companies (O’Brien 2012; Drugan 2013, p.36).

<table>
<thead>
<tr>
<th>General Standards and Guidelines</th>
<th>Translation Guidelines and Standards</th>
</tr>
</thead>
</table>
  International standard listing generic requirements for a quality management system  
  
  ASTM F2575  
  A quality of language translation services guide. |
| **STP800** | ISO TR 10013  
  Guidelines for quality management system documentation.  
  
  IEC 60601 (standard for electrical medical devices)  
  ISO 13485:2003  
  Medical devices -- Quality management systems - - Requirements for regulatory purposes  
  
  ISO TC 37  
  Terminology and other language resources  
  
  ISO 15223-1:2007  
  Medical devices -- Symbols to be used with medical device labels, labeling and information to be supplied.  
  
  ICH  
  International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.  
  
  FDA translation regulations (ex., 21 CFR 56.111)  
  SAE J2450, a set of translation guidelines which can be adapted to different fields.  
  
  CAT (Computer Assisted Translation)-related open standards, and QA tools like ITR Blackjack, are becoming standards themselves because of their role in terminology management and consistency.  
  
  The European Union’s (EU) Medical Device Directive (93/42/EEC) (MDD) applies specifically to translation of materials related to medical devices. |
  Codes for the representation of names of languages |
| **ISO TR 10013** | ISO 12616:2002  
  Translation-oriented terminography |

Figure 8: Table of common standards in localisation, as provided in Txabarriaga 2009 (p.6)

\(^{20}\) https://www.iso.org/standard/59149.html
However, while doubtless great progress has been made in this direction over the past years, the wide range of models and standards in use leaves the localisation industry without a single, objective, agreed-upon way to measure quality (Drugan 2013, p.35). O’Brien (2012) has conducted a benchmarking evaluation of several quality evaluation models like the LISA QA model v. 3.1, J2450 or EN15038 and models that are used exclusively within individual companies. While these generally show a high level of consensus in some categories like ‘language’, which commonly include grammar, syntax, punctuation and spelling, as well as, with less consensus, ‘typos’, fluency and cultural references (pp.60–61), agreement is lower for other areas. Categories like ‘terminology’ and ‘accuracy’ are generally in use, but with less agreement on the definition of those categories (p.61).

The ‘style’ category is still mostly present, but "of the top four 'Language' errors, 'Style' is the one with the least consensus across models" (O’Brien 2012, p.61). While 4 out of 7 models refer to a lack of adherence to a client style guide, otherwise style could include tone, register, language variants, slang, literal translations or awkward syntax (p.61).

This lack of consensus likely stems in part from the fact that the short hand 'localisation industry' really stands for a diverse, global and fast-paced set of organisations, individuals, contents and projects with diverse needs and requirements (Drugan 2013, p.37), and that, as Jiménez-Crespo (2011) argues, it may be theoretically and methodologically impossible to predefine quality for all texts (p.134).

This is also reflected in the certification criteria of ISO 17100, which describe processes for translation service providers. Here the need to receive project specifications through client instructions is explicitly acknowledged:

"A.4.4 The TSP [Translation Service Provider – added by author] endeavours to obtain additional information required with regards to difficulties in the source language content and other project specifications by contacting the client for instructions and passing that information on to all relevant parties involved in the project."

(Jonas 2015, p.8)

Results from small-scale studies suggest that the general adoption rates appear to be low on the vendor side. Additionally, standards appear to have low visibility and
importance on the client side (Kelly and DePalma 2008a, pp.3–4; Ray et al. 2013, p.17). This would imply that even if a vendor was to be certified by a standards organisation, clients don't necessarily understand what those standards include and why they would indicate quality.

Finally, models in use in the industry have been criticised for being anecdotal and subjective. Since they are devised by practitioners but not based on theoretical approaches with the aim to establish general principles (Colina 2008, p.103; Drugan 2013, p.46), they can be limited in transferability, as well as replicability. Additionally, inter-rater reliability can be difficult to assess (Brunette 2000, pp.99–100; Colina 2009, p.237).

The third and final observation to make on the manufacturing-based model of quality control is:

3. Implicitly, the model relies on unambiguously measurable product attributes, since ambiguity in measurement would prevent specification of requirements and confirmation of whether they have been adhered to. This is likely impossible to achieve for all aspects of localised products, as following sections in this chapter will discuss in further detail.

While quality of source content appears to be a common source of quality issues (Kelly and DePalma 2009a, p.6), this aspect will not be considered further here since according to the literature, no evidence exists suggesting a connection between source content quality and perceived quality of the localised product or differences between source content quality for traditionally sourced and crowdsourced localisation projects.

3.3.2.3. Quality related business goals

Another goal of translation quality assessment can be linked to business goals. Drugan (2013) notes that

"the drive to outsource translations has also led to a greater focus on measuring quality because in-house divisions increasingly need to be able to demonstrate they add value, by providing higher quality than outsourcing" (Drugan 2013, p.73).

On the vendor side, sellers need to be able to demonstrate why their services add
value relative to free offerings like freely available machine translation (p.74).

Other business related areas can include the selection of contractors as well as dealing with complaints related to quality issues and financial penalties (Mossop 2001 - in Drugan 2013, p.72). Another business related goal can be the differentiation of service offerings from competing ones and being able to justify rates (Drugan 2013, p.41).

This suggests that a primary goal of vendors and in-house localisation teams would be to achieve and demonstrate 'perfect quality', which according to DePalma (2013) is the aim of the traditional translation model. However, this goal is in conflict with the goal of 'maximum availability' relative to the speed and cost of localisation (Kelly and DePalma 2008, p.11), stemming from the tension between large volumes of content to be localised into a wide range of languages and availability of budgets. Against these considerations, "quality matters less than simply making information available" (DePalma 2013, pp.1–2). Or, as described previously in this section, what is important, is knowing when the localised content is 'good enough' (Drugan 2013, pp.39–42).

With this approach to 'maximum availability', the client perspective may reflect the end-user's needs to a degree.

"While poor quality output is not acceptable to translators, it is acceptable to most of the rest of the population, if they want immediate information, and the on-line 'culture' demands rapid access to and processing of information."


Differing expectations of quality were also observed by Bowker (2009), who reported on the findings of a study comparing recipient evaluations of machine translated content. Here, 88% of surveyed language professionals insisted that only human translations were acceptable, while participants not working in the language industry were willing to accept post-edited machine translation nearly half of the time (p.143). However, the findings of this study also differentiated the results depending on whether the purpose of the translations was functional or aimed at preserving or promoting a culture (p.148).

In summary, as section 3.2. and this section have outlined, there is no shortage of quality models in localisation and translation - be it in industry or academia. However, so far, the challenges of clarifying what those models are aiming to achieve, as well as how to achieve those goals, have yet to be overcome, and no universally or even widely applicable approach has emerged.
3.4 Applying the Integrative Quality Framework

Previous sections have outlined a typical chain in the localisation outsourcing process, which for the purpose of this thesis involves

- a seller (linguist),
- a buyer/seller (project manager at multi-lingual vendor) and
- a buyer (client purchasing localisation services, but not end-client).

In addition, a brief overview has been provided indicating some of the challenges the localisation industry faces in relation to producing quality.

However, so far absent from this discussion has been information asymmetry (Akerlof 1970) in the process chain and the resulting implications for perceived quality in localisation.

As described, within the integrative quality framework, the three quality states encompass:

- the quality production process,
- the quality experience process and
- the quality evaluation process.

During the quality production process, resource inputs are converted into produced attributes via attribute design and process design specifications.

In the context of localisation, the two most basic resource inputs would be the source content and previously localised content as well as human resources in shape of the time, efforts, skills and talents of the individual(s) performing the necessary tasks during the production process. For the purpose of this research, the resources primarily considered are linguists (translators and reviewers).

Technology like computer-assisted translation tools or terminology databases could be classified as part of both resource input and the process design, since they provide previously generated translations and reference material in addition to dictating and supporting workflows. Additional processes in quality production would include, among others, a range of quality control and assessment steps like review and evaluation.
conducted by humans as well as use of quality control technology like Errorspy\textsuperscript{21}, ApSic Xbench\textsuperscript{22} and quality control options included in a range of computer-assisted translation tools.

Quality assessment plays a role both in the quality production process, and in the quality evaluation process. Quality assessment as a method of monitoring produced attribute quality is commonly organised by the seller (linguist and then again vendor) prior to delivering the localised content to the buyer (vendor and client) (Section 3.3.).

Independent of seller-side quality assessment during production, the buyer also engages in a quality evaluation process after having received and experienced the localised content with its set of attributes. As previous sections have outlined in general terms outside of localisation, the outcome of the quality experience and evaluation processes and therefore how quality is perceived depends on several factors, which will be addressed in the context of localisation in the following sections:

- The nature of the attribute that is being perceived
- Measurement knowledge, motivation and emotions
- 'Will' expectations and 'will' uncertainty
- Intrinsic and extrinsic cues

3.4.1. Product Attributes

In section 2.3.1., a classification distinguishing between how easily an attribute can be measured (ambiguous and unambiguous) as well as customer preferences (homogeneous and heterogeneous) was introduced (Golder \textit{et al.} 2012, p.8). Three types of attributes result from these two variables:

- Preference attributes: unambiguous measurement and heterogeneous preferences
- Universal attributes: unambiguous measurement and homogeneous preferences

\textsuperscript{21} \url{http://www.errorspy.com/}
\textsuperscript{22} \url{https://www.xbench.net/}
preferences

- Idiosyncratic attributes: ambiguous measurement and heterogeneous preferences

(Golder et al. 2012, p.8)

These distinctions are relevant for localisation in two ways:

1. Universal applicability
2. Measurement knowledge required

3.4.1.1. **Universal Applicability**

Homogeneous/heterogeneous preference suggests that some attributes will be universally applicable while others are project, client and content dependent. Examples for heterogeneous preferences could include among others:

- Limitations on character count for a text string, for example if the string is used in a small menu item. This is heterogeneous because under other circumstances, the same source content might be subject to no restrictions on character count or to even more stringent limitations.

- Compliance with previously defined terminology, for example in the context of brand specific language. While other terms might also be correct and therefore acceptable for other contexts, specific terms might be defined as part of a project glossary. Since terminology requirements as described here are context specific, they are heterogeneous attributes.

Both examples (character count and brand terminology) so far have been examples for preference attributes since they are not only heterogeneous in preference but also unambiguous in measurement in that it is possible to unambiguously say whether a string satisfies the criteria in character limits or whether one term or the other term was used in a translation. These attributes can also be measured using automated processes like character counts and terminology verification software.

Since preference attributes are heterogeneous yet can be unambiguously measured,
they lend themselves especially well to being defined in client requirement documentation and are also likely to be the type of attributes clients are especially concerned with. One common quality related complaint on the client side in a survey of localisation buyers includes vendors and freelancers not using resources provided like glossaries, style guides, translation memories and terminology databases (Kelly and DePalma 2009a, p.2). In addition, having their change requests implemented 'graciously and without complaint' appears to be of great importance to buyers (Kelly and DePalma 2008b, p.4).

The distinction between homogeneous and heterogeneous preferences also includes the idiosyncratic category, which means attributes that are ambiguous in measurement and heterogeneous in preference. Examples include:

- How to deal with culture or language specific expressions
- What style or register applies for the content in question
- What 'sounds good'

Part of the issue here is that such

"discourse elements (rhetoric, style, register, ambiguity, implication, allusion, metaphor) or textual properties (coherence, cohesion, [sic] however, do not lend themselves to being codified into precise rules."

(García 2014, p.432).

This suggests that resulting from a lack of codification, attributes in this category are difficult to teach to novices and are instead primarily experience based, tacit knowledge (Dreyfus and Dreyfus 1980, pp.7–12). Without the individual linguist being able to make this experience based tacit knowledge explicit (Eraut 2000), these idiosyncratic attributes would appear to be what Fawcett describes as quality assessment proceeding "according to the lordly, but completely unexplained whimsy of 'It doesn't sound right'" (Drugan 2013, p.37) and are as a result a potential source of disagreement between different linguists. This observation is also supported by the analysis of different quality evaluation models conducted by O’Brien (2012), where 'style' was the category with the least consensus across the models (p.61) regarding how it should be defined.
3.4.1.2. **Measurement Knowledge Required**

Ambiguous/unambiguous measurement has implications for the measurement knowledge that is required to undertake the measurement, at least to the extent where knowledge has been codified or included in automated verifications:

"Errors of form, from spelling to syntax, can be detected by contrasting against a specific set of rules that every language speaker learns, and are easy to agree with; being rules, machines can often learn them too."

(García 2014, p.432).

Examples include:

- Spelling errors in a target text, which could potentially be detected by a child after having completed primary school and having learned the codified spelling rules of the language. Spelling errors can also be detected by automated spell checkers to a large extent, although neither of the measurement instruments (school child or spell checker) in this example are 100% reliable;

- The already mentioned character limitations which require no knowledge of the target language or localisation processes beyond being able to distinguish one character from another. Here, too, measurement is possible with automation.

For these examples and similar ones, a knowledge of technology that can detect issues is of great importance, and often part of a localisation vendor's quality checks.

3.4.2. **Measurement Knowledge, Motivation and Emotions**

The previous section outlined the different attributes in localisation projects. In the following sections, measurement knowledge and measurement motivation in the context of perceived localisation quality will be considered in more detail.

The role of emotions in quality perception was described briefly in section 2.4.3. However, a search for related literature in the area of localisation has yielded no results.
While there is no evidence to suggest that emotions don't play a role in perceived quality in localisation, it is difficult to estimate which emotions are prevalent and what role they play.

### 3.4.2.1. Measurement Knowledge in Localisation

Measurement knowledge according to Golder et al. (2012) is the customer's ability to assess attribute performance with minimal bias or variance relative to more objective measures (p.9). The ability to measure attributes is also dependent on the nature of the attribute (p.8), as discussed in section 2.3.1. The following section discusses factors that are likely to impact measurement knowledge in localisation to evaluate which participants in the localisation chain would have higher or lower measurement knowledge, as well as the types of measurement knowledge that can reasonably be expected to be present. By understanding measurement knowledge for different groups, it is then possible to predict broadly to what extent those groups would be likely to be reliant on extrinsic cues. Such cues could be professional status and price of localisation, based on the integrative quality framework proposed by Golder et al. (2012) and discussed in the theoretical framework (chapter 2).

Golder et al. (2012) state that

"relative to a typical customer, firms are more capable of measuring attribute performance accurately because firms have greater access to measurement instruments."

(Golder et al. 2012, p.5).

This suggestion can also be observed in literature on localisation (Pym et al. 2016, p.35) where buyers are quoted to say that they are unsure "how to know you are getting quality when you don't speak the language" (Sargent and DePalma 2007, p.19). Additionally, client-side localisation managers have been described as

- not qualified for the task and having to rely on hearsay and opinions (DePalma et al. 2006, pp.13–15),
- needing to be taught how to buy better (Ray and Hegde 2012, p.14) and
- 'most clients' being described not seeming to have "any idea of what's involved with translation/localization, and are too impatient to take the time
to learn." (Sargent and DePalma 2007, p.18- quoting a language service provider).

This information asymmetry (Akerlof 1970) is also implied in the fact that a lack of knowledge and expertise in an area that is not part of an organisation's core competency is commonly a major factor in the decision to outsource localisation (DePalma 2012, p.1; Shreve 2000, p.226 - in Dunne 2006, p.101).

And indeed, when considering the complex set of skills involved in measuring the different attributes of a localised product, it would imply that clients are at a disadvantage compared to specialist project managers on the vendor side as well as linguists. A range of models describe the different competences involved in translation (see Orozco and Hurtado Albir 2002, p.376 for a collection of models) The Dreyfus and Dreyfus (1980) model of competence including the

- novice,
- advanced beginner,
- competence,
- proficiency and
- expertise states

has also been applied to translation competence (Orozco and Hurtado Albir 2002, p.376). In addition to translation skills, the suggestion has been made that

"in order to objectively and professionally assess the quality of translations one needs to have acquired a certain amount of expertise and knowledge”


Additionally, relative to translation experts, lay evaluators could be expected to be less detailed and systematic in their approach, as well as being unable to use the linguistic terminology of translation quality assessment. As a result, a lay person might have difficulty expressing their opinions and assessment of a translation's quality (Havumetsä 2012, p.56).

However, from the assumption that measurement knowledge in localisation is improved by increased expertise it does not automatically follow that localisation vendors hold the keys to assessing quality, while clients are left outside.
One aspect to consider is whether client-side buyers are localisation specialists or whether they are experts in another area, and only handle localisation 'on the side'. In an analysis of 177 client-side buyers 69% didn't have 'localisation' in their title but instead seemed, based on their titles, focused on areas like marketing, product management, procurement, legal and engineering (DePalma and Beninatto 2004, pp.5–6). Additionally, respondents had performed the function of purchasing localisation services for fewer than six years in 'nearly three-quarters' of the individuals studied (DePalma and Beninatto 2004, p.7). However, that still leaves 31% whose position is focused enough for them to have 'localisation' in their job title as well as about a quarter of the study participants who have been purchasing localisation services for more than six years.

In addition, an increased tendency to centralise localisation vendor management on the client side has been reported (Kelly and Stewart 2010, p.45). This could imply greater specialisation on localisation for purchasers. Combined with an observed increase in maturity in handling content and localisation processes in some organisations (DePalma 2012, p.1), individuals switching from vendor to client side (Geller et al. 2011, p.4) as well as buyers' increased abilities in defining, measuring and assessing quality also improve with improvements around methods, technology and knowledge (Alba and Hutchinson 1987, p.411; Kelly et al. 2008, p.3). This suggests that rather than being able to draw a line between 'vendor-side buyer' and 'client-side buyer' when anticipating measurement knowledge of these groups, a much more complex set of factors like level of specialisation, educational and training background, and experience would have to be considered.

This uncertainty in predicting measurement knowledge for the three parties in the localisation chain can be further exemplified when considering some of the knowledge areas involved in localisation:

- domain knowledge,
- linguistic knowledge and
- production knowledge.

3.4.2.1.1. Domain knowledge

The importance of domain expertise, or subject knowledge in relation to localisation
has been highlighted in the literature (Sargent and DePalma 2007, p.11; Muñoz Martín 2009, p.25). Some differentiate between domain experts and linguistic experts (Allman 2006; Neather 2012, p.266), suggesting that students in translation schools would rarely have been exposed for long periods to the vocabulary and textual conventions of a domain (Shreve 2002, p.163). However, Öner (2013) argues that the translator's training will enable them to handle a range of different domains and texts, while also acknowledging that specialisation isn't unnecessary (p.77).

Independent of whether a translator needs to develop subject expertise or at what point this is acquired, it is reasonable to assume that a localisation manager embedded on the client side would be likely to know more about the product's specific domain by virtue of proximity to the individuals who have developed the source product. For example, even if a translator is specialised in games translations, the localisation manager who is employed at that particular games publisher would be likely to know more about this specific game than the domain-expert translator.

However, for this domain knowledge to also transfer to measurement knowledge, the localisation manager on the client side would also require knowledge of the target language and cultural conventions.

### 3.4.2.1.2. Linguistic knowledge

Regarding linguistic knowledge, it can be reasonably assumed that the translator has greater measurement knowledge than any of buyers, vendor and client side. While a project manager at a Multi-Language Vendor (MLV) might have linguistic knowledge of one or several language combinations, it is unlikely that this project manager would have measurement knowledge of all language combinations in a project. In a survey of 409 project managers,

- only 8% were language specialists, handling projects in specific language combinations,
- while 64% were generalists handling any type of project (Pielmeier and DePalma 2013, p.11) and
- only 17% always or frequently handle language tasks (Pielmeier and
A similar situation could present itself on the client side, where the client might deal with more language combinations than they have measurement knowledge for. However, it is also possible that a client requests localisation services for a language combination in which they also possess linguistic skills.

3.4.2.1.3. Production knowledge

The third example of probable imbalance in measurement knowledge between different parties in the localisation chain is what Muñoz Martín (2009) describes as a part of domain knowledge, and will be referred to here as production knowledge for easier distinction. Production knowledge for the purpose of this discussion includes translation-specific declarative knowledge such as, for example, conversion of units and measures, using the tools of the trade, converting units and measures (p.25). Commonly, although not exclusively, such skills are acquired through formal training programmes and deliberate practice (Pérez-González and Susam-Saraeva 2012, p.151). However, Shreve (2002) argues that translators "certainly do not have the basis for developing domain-specific production rules, strong methods, for accomplishing the translation task" (p.163) through translator training programmes. Whether such skills are acquired primarily through training or through working experience, at least linguists who are formally educated in addition to being experienced are likely to have high levels of production knowledge.

On the vendor side, in the case of a project manager who doesn't speak the language(s) in question, attribute measurement that a project manager can conduct personally will be limited to unambiguously measurable attributes, which have commonly been included in automated quality tools. Examples include spell-checking, terminology compliance checks, consistency checks and character counters, although deciding on whether issues highlighted by the tool are really issues or false positives can be difficult without linguistic knowledge. In the previously mentioned survey of 409 project managers, 67% of the respondents indicated that they routinely perform final project review, although 56% never or rarely use automated tools for quality assurance and only 8% systematically do so (Pielmeier and DePalma 2013, pp.6–7). It is unclear whether buyers on the client side
routinely personally perform quality checks although there is nothing to suggest that a specialised, experienced buyer on the client side would possess less measurement knowledge on production-related factors than an equally specialised and experienced vendor-side project manager.

3.4.2.1.4. The role of third-party review

It should be mentioned that, in addition to the goal of improving the quality of the localised product, a common attempt to address a lack of measurement knowledge involves requesting review from a second linguist. This step might be included by the translator (primarily with the goal to improve on quality), the vendor or the client.

However, even though this practice is widely in use, it is fraught with difficulties, commonly stemming from a lack of agreement on what attributes to measure and how. This can then lead to clients complaining about overzealous reviews which result in consuming buyers’ time (Kelly and DePalma 2009b, p.9; Drugan 2013, p.42) due to overly granular reviews and “academic nitpicking” (Kelly and DePalma 2009b, pp.4–5). At the same time, buyers may conduct informal quality reviews without giving the vendor advance notice. As a result, vendors express concerns over logistical issues and post-delivery change requests through delayed feedback and overzealous reviewers on the client side, providing undeserved feedback on the quality of the translations (Txabarriaga 2009, pp.1–2).

This practice likely stems from the clients’ need to judge what they are investing in without linguistic knowledge of the source and/or target language.

While translators will approach this issue by explaining, justifying and defending quality levels (Drugan 2013, p.69) in an attempt to build trust in clients, it seems likely that clients who are pressed for time will try to circumvent the need to engage with the finer details of the translation by asking a presumably trusted colleague to provide an evaluation of the translation.

In addition to the difficulties inherent in the review process, arguably, the best-case result third party review for perceived quality can achieve is increased confidence in the localised product, but not actually increasing measurement knowledge of the individual making the purchasing decision. As a result, a localisation manager would still need to rely
on extrinsic cues in lieu of being able to measure intrinsic attributes.

### 3.4.2.2. Measurement Motivation

The discussion on third party client review (Section 3.4.2.1) already touched on the possibility that different parties in the localisation chain might have differing measurement motivation.

Havumetsä (2012) suggests that some clients may not monitor quality at all and comments that this "gives a clue as to what s/he thinks is important in translation." (Havumetsä 2012, p.55). However, apart from the question of whether a client has sufficient measurement knowledge, a lack of monitoring of quality doesn't necessarily imply a lack of interest in quality but could also be due to time limitations (Kelly and DePalma 2009a, p.1). Especially for large projects which can range in the millions of source words and multiple target language combinations, in-depth quality monitoring is a time intensive task.

In addition, multiple concerns raised in a study of 73 client-side buyers suggest that at least some buyers measure and are interested in quality aspects important to their product. Those rated as 'very important' included:

- translation memory use (92%);
- file format standards adherence (76.6%);
- terminology database consultation (75.9%), and
- style guide compliance (66.1%)

(Kelly and Stewart 2010, p.30).

Measurement motivation would likely also differ depending on where an organisation is situated on what DePalma (2011) describes as the 9 stages of the localisation maturity model (p.2). This model, analogous to established capability maturity models (p.1) outlines different phases in maturity ranging from -3 (discouraging) to 5 (transparent). In this model, each progression towards 5 shows improvements in processes and forward planning to replace reactivity and ad-hoc approaches (DePalma et al. 2006, pp.7–9). In a client organisation where the approach to localisation has reached the final,
'transparent' phase, localisation concepts are a natural part of planning, development and quality life cycles as well as continuous process improvement (DePalma et al. 2006, p.8). In a mid-range phase, 'reactive' localisation is handled as an ad hoc response to business demands where a lack of defined processes is compensated for with individual heroics (DePalma et al. 2006, p.7).

3.4.3. Expectations

Sections 2.5.1. and 2.5.2. described the role of expectations in the quality evaluation process, where the expectations serve as reference levels against which perceived attributes are compared, with the result being the customer's evaluated aggregate quality or satisfaction.

For this research, of the three types of expectations (ideal, should and will), the 'will' expectations play an especially central role. 'Will' expectations are those attribute performance levels that a customer believes a product will possess. When this customer has low measurement knowledge or motivation, the 'will' expectations then may be retrieved to replace the product's attributes, which the customer in this situation doesn't directly observe (Golder et al. 2012, pp.10–11).

In the context of localisation, this could for example mean that a buyer expects that the seller will deliver a high-quality Japanese website. When the buyer is then faced with having to decide whether the Japanese website is of acceptable quality, the previously formed expectation that it will be of high quality then impacts the way the quality of the website is perceived by the client.

If the client is unable to observe all the intrinsic attributes of the Japanese website's content, for example due to a lack of Japanese language skills, then extrinsic cues might be used to form an expectation and resulting perception of the website's quality.

Zeithaml (1988) refers to extrinsic cues as shorthands (for example, brand name) and surrogates (for example, price) (Zeithaml 1988, p.8). In localisation, several extrinsic cues might serve as either a shorthand or surrogate, like the reputation of the vendor, the advertising material provided, or the demeanour of the sales staff. However, for the purpose of this research, price and professional status cues were selected as focus due to their being
primary distinguishing factors between 'professional' and 'crowdsourced' or 'volunteer' localisation. Their role in localisation and crowdsourced/volunteer localisation will be discussed in section 4.1.3.1., following an introduction to the concept of crowdsourced and volunteer localisation in section 4.1.2.

3.5. Summary

This chapter has outlined different approaches to quality in localisation as well as the localisation process chain underlying this research. In addition, the importance of the study of perceived quality in the localisation context has been highlighted by describing the application of the integrative quality framework (Golder et al. 2012) to localisation.

The distinction between universal, preference and idiosyncratic attributes also applies to localisation. As a result, some attributes of a localisation product will be unambiguously measurable and universally preferred; while for others, individuals will have different preferences or unambiguous measurement is not possible.

In addition, sections 3.4.2.1 and 3.4.2.2. have outlined the differences in measurement knowledge and motivation that can be predicted, albeit not reliably and only in broad strokes, for different parties in the localisation chain. Here it was highlighted that predicting measurement knowledge based on the roles individuals play in the localisation process is difficult to generalise. While client-side buyers would likely tend towards lower measurement knowledge than vendor-side buyers, and linguists could be in general expected to have the highest levels of measurement knowledge, this assumption cannot be made for all cases. Instead, specific knowledge areas like domain, linguistic and production knowledge need to be considered, in addition to level of specialisation, education and experience. Finally, it should be noted that even when these factors can be compared for individuals, it does not necessarily mean that measurement knowledge can reliably be predicted, since knowledge can be obtained from a wide range of sources that would be difficult to measure and classify, like freely available information on the internet, or through colleagues and conferences.

However, for the purpose of this research, specialisation, education and experience will serve as primary factors for predicting measurement knowledge, while bearing those
limitations in mind.

Chapter 4 will discuss, after an outline of crowdsourcing in localisation, the two extrinsic cues that could serve as a surrogate or short-hand in cases of low measurement knowledge: professional status and price of localisation.
Chapter 4: Quality in Crowdsourced Localisation

4.1. Crowdsourced and Volunteer Localisation

Howe (2006) is frequently credited with giving the first description and definition of crowdsourcing as

"the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call"

(Howe 2006).

While the concept of crowdsourcing in general is not new, the topic has gained traction with increased technological capabilities and has received increasing attention in the literature.

4.1.1. Terminology

One of the first points of discussion in literature on the topic is generally the question of what to call the concept as a whole or the concept that is the focus of that particular research (Perrino 2009, p.63; O’Hagan 2011a, pp.13–14; McDonough Dolmaya 2012, p.169; Fernández Costales 2013, p.93; Jiménez Crespo 2015, p.60), while different labels are used more or less synonymously (O’Hagan 2011a, p.13).

Terminology includes, among others:

- User-generated translation (Perrino 2009; O’Hagan 2011a)
- *CT3 (community, crowdsourced and collaborative translation (DePalma and Kelly 2008)
- Crowd-sourced translation/Translation crowdsourcing (Perrino 2009; Baer 2010; Cronin 2010; Dodd 2011; McDonough Dolmaya 2012)

Different terms have different weighting and suggest different focal points. Community translation, crowdsourcing and collaborative translation have been ascribed as carrying activist and ideological connotations (Pym 2011 - in O’Hagan 2011; Cronin 2010).
O’Hagan (2011a) sees this as justified in the context of fan translations, but not for translation in a commercial, for-profit context (pp.13–14). In addition, according to Perrino (2009), crowdsourced translation could suggest disorganisation and absence of rules (p.63). Volunteer translation is appropriate in contexts where the distinguishing factor is remuneration, although "a lack of monetary reward is not an essential characteristic of community translation" (O’Hagan 2011a, pp.13–14). Other terms in use include fan-based or fan translation as well as amateur translation, although 'fan' is strongly associated with fan culture for manga, anime and video games (O’Hagan 2009; Perrino 2009, p.63; Fernández Costales 2013, p.93). The term 'amateur' carries negative connotations, suggesting potential quality issues (Perrino 2009, p.63).

Since the majority of terms in use focus on the individuals participating in the localisation/translation effort, the decision on which terms to use is especially difficult in a thesis paper that primarily focuses on the 'naive theories' (Deval et al. 2012) research participants have about those that participate in such crowdsourced/voluntary localisation initiatives and the impact those theories then have on their perception of quality. Any decision on terminology also implies an assumption on those that participate in the initiative.

Therefore, the decision was made to use ‘crowdsourcing’, in line with McDonough Dolmaya (2012), due to its popular use in the literature in order to describe the concept being studied. By using crowdsourcing as a term, the open call nature of recruiting individuals to contribute is being acknowledged, and through that also the potential diversity of individuals responding.

Since the second focus of this research is remuneration of translators, the terms 'volunteer' or 'voluntary' are used interchangeably with 'crowdsourcing' to indicate a lack of monetary reward.

4.1.2. Crowdsourcing Models

Difficulties in the discussion also arise from the wide range of models that can be broadly included in the concept. Different attempts at classification have been made, generally depending on the focus of individual studies.
**Paid vs. Unpaid:** While some crowdsourcing involves financial rewards (Zaidan and Callison-Burch 2011; Garcia 2015; Flanagan 2016), other initiatives' incentives range from badges, leader boards, mentions in company blogs, free services and software to marketing opportunities, event invitations and commonly, no explicit reward at all (Mesipuu 2012; Risku *et al.* 2016 - all in Flanagan 2016; Ray and Kelly 2011, p.28; Pym *et al.* 2016).

**Solicited vs. Unsolicited:** Other classifications focus on whether the impulse for the initiative came from the individuals conducting the localisation (unsolicited) or from an organisation who wishes to have the content localised (solicited) (O’Hagan 2009; Fernández Costales 2013, p.98; Jiménez Crespo 2015, p.62).

Unsolicited or user-initiated (Dombek 2014, p.27) usually revolve around content where the crowd has a high interest in the domain like Wikipedia content (Flanagan 2016, p.152), illegal translations of books, for example the translation into Spanish of *Harry Potter and the Order of the Phoenix* in Venezuela (O’Brien and Schäler 2010, p.2), and a range of fan-driven initiatives around audio-visual media (Fandubbing or Fansubbing; (O’Hagan 2009, p.99; Bold 2011, pp.3–4)), Comic books/Manga (Scanlation; (O’Hagan 2009, p.100; Bold 2011, pp.3–4; Fernández Costales 2013, p.31)) and video games (ROM hacking; (O’Hagan 2009, pp.107–108; Bold 2011, pp.3–4).

Solicited (Jiménez Crespo 2015) or content-owner initiated (Dombek 2014) crowdsourcing has an impulse initiating from the company or organisation (Dombek 2014, p.27) and can be either for-profit or non-profit (Flanagan 2016, p.152).

Examples include crowdsourcing platforms like Amazon Mechanical Turk (Zaidan and Callison-Burch 2011; Goto *et al.* 2014; Yan *et al.* 2014), the crowdsourced localisation of Facebook (O’Brien and Schäler 2010, p.2; Jiménez Crespo 2013; Dombek 2014) or LinkedIn (American Translators Association 2009) and a range of other products owned by Symantec, Sun and Microsoft (O’Brien and Schäler 2010, p.2). Additionally, initiatives like TRF/Trommons, which is featured in this research, and initiatives like Translators Without Borders (O’Brien and Schäler 2010, p.3) fall in this category.

**Cause vs product vs outsourcing-driven:** Another classification to distinguish between different types of initiatives focuses on the factor that drives the model, while not
being necessarily linked to the initial impulse. These drives are described as the 'cause', the 'product' and 'outsourcing':

- Product-driven initiatives focus on the actual product, like a software product, video or comic book.
- Cause-driven initiatives are focused on non-profit, humanitarian areas, while
- outsourcing-driven have no specific social or humanitarian mission but are focused on the translation needs of a for-profit organisation (O’Hagan 2009, pp.111–112; Dolmaya 2011, p.99; Ray and Kelly 2011; Pérez-González and Susam-Saraeva 2012, p.155; Flanagan 2016, p.152; Orrego-Carmona 2016, p.1).

A challenge with any of the existing classifications of localisation crowdsourcing is that the groupings are proposed around discussions other than quality or perceived quality in localisation. Within the different categories, a range of factors could contribute to increased or diminished quality, as well as perceptions of quality. For example, solicited translations could be done by either a highly specialised, trained, experienced and motivated translator or alternatively by a bored, barely bilingual amateur with no previous experience or training in translation and very little motivation. The outcome in quality, as well as how the quality would be perceived depending on what the person evaluating the product knows about the translator would likely differ quite significantly, but both would fall in the category of 'solicited' crowdsourcing.

For this reason, this research will consider initiatives based on a taxonomy proposed by Hosseini et al. (2014), which describes four pillars of crowdsourcing:

1. The crowd: the people taking part in the crowdsourcing activity;
2. The crowdsourcer: the entity (person, organisation etc.) who wishes to utilise the crowd's power and wisdom;
3. The crowdsourcing task: the activity in which the crowd participates;
4. The crowdsourcing platform: the system (software or non-software) in
which the task is performed

(3.2–3)

As mentioned previously, the focus of this research is on the two primary factors distinguishing 'crowdsourced localisation' from 'industry' or 'professional' localisation: the individuals conducting the localisation; and the remuneration they receive. For this research, the topic of payment will also be focused primarily around the question of how payment impacts 'the crowd'.

For this reason, the focus of the following sections will be on the first pillar 'the crowd' and will consider monetary rewards to be a factor that impacts 'the crowd' positively or negatively. Pillars 2 through 4 play a peripheral role, mainly when considering the aspects that impact 'the crowd', like motivational factors inherent in who 'the crowdsourcer' and what 'the crowdsourcing task' is. Although the question of how 'the crowdsourcing platform' enables or hinders quality in and perception of quality in crowdsourced localisation is worth investigating further, it will be set aside as out of scope for this thesis.

4.1.3. The Crowd

Before discussing 'the crowd' it is necessary to consider the translation profession for the following two reasons:

1. Some groups of translators earning their living with translations have been critical of some crowdsourced localisation initiatives (American Translators Association 2009; Flanagan 2016, pp.157–158). While there is no broad evidence to suggest that this is also the opinion of most professional translators, or merely the opinion of 'the vocal minority', those that are vocally against it might be able to alter 'will' expectations of decision makers by suggesting that the crowdsourced output would be of poor quality.

2. When taken as extrinsic cues, 'the crowd' being considered an entity that is different and separate from 'the professionals' would likely have an impact on perceived quality if the assumption is that one signals a different set of
quality related attributes than the other.

The following sections will approach the question of how different and separate 'the crowd' is from 'the professionals' by outlining the status of the translation profession and the relationship between the professional and non-professional translator. It will then consider scenarios in which one or the other group might be better suited, as well as information known on some parts of 'the crowd', as well as their motivations to volunteer.

4.1.3.1. The Translation Profession

4.1.3.1.1. Translation as a 'Failed' Profession

Pym et al. (2016) argue that especially in translation "the value of a translation is commonly not easily attested by the person paying for the service" (Pym et al. 2016, p.33), which has also been argued in section 3.4.2.1. on measurement knowledge in localisation.

As a result, clients largely depend on external status of the translators (p.33) to determine trustworthiness. Signals that indicate whether someone should be considered a 'professional' could include attested experience, evidence of reliability (p.33) as well as academic or professional training (Koskinen and Dam 2016, p.263; Pym et al. 2016, p.33).

However, one of the challenges clients as well as translators face here is that the translation field could be considered as an example of a 'failed professionalising occupation' (Elsaka 2005 - in Sela-Sheffy 2015, secs3468-7806). Similarly to other under professionalised or marginalised professions like journalists, craft-artists, therapists or teachers (Sela-Sheffy 2015, secs3476-3480), the translation and interpretation professions are here in contrast with so-called 'success stories' of professionalisation which have traditionally been more strongly institutionalised and protected, like law, medicine and accounting (Sela-Sheffy 2015, secs3474-3476).

As a result of this lack of demarcation that would distinguish top specialists from non-elite manpower as well as qualified and amateur practitioners (Sela-Sheffy 2016, p.55), it is unclear whether a 'professional' translator is for example

- an individual with training and formal qualifications or
• an individual who merely earns a living with translations

(Paloposki 2016, p.16).

This suggests that theoretically anyone who wishes to do so could call themselves a 'professional translator' and attempt to earn money with translations, regardless of whether they are sufficiently skilled, trained and experienced to provide adequate results (Chesterman 2001, p.146).

This is not to say that the professionalisation process has been stagnant. Research around translation as a profession within translation studies has been increasing (See Sela-sheffy 2016, p.55 for a collection). Translation ethics have been proposed around representation, service, communication and norms (Chesterman 2001, pp.139–141) as well as a hieronymic oath for translation with an eye on strengthening international accreditation of translators (Chesterman 2001, p.153). Further examples for efforts to increased professionalisation include the accreditation and ranking functionality on one of the biggest freelance translator platforms in the industry, ProZ.com (Pym et al. 2016, pp.39–40). In addition, the provision that ISO 17100 certified translation service providers obtain evidence that shows that a translator meets one of the following criteria:

"a) a recognized graduate qualification in translation from an institution of higher education;
b) a recognized graduate qualification in any other field from an institution of higher education plus two years of fulltime professional experience in translating;
c) five years of full-time professional experience in translating."

(Jonas 2015, p.7)

However, Pym et al. (2016) argue that while ordinarily the process towards professionalisation is linear in other areas, the technological shift in recent years has caused a relapse for the translation profession (p.36). Through the increasing internationalisation of the translation market, free online machine translation, and increased electronic communication, complications in form of difficulties with checking credentials, increased de-professionalisation and the fact that through machine translation, everyone can produce 'something that looks like a translation' have been introduced (p.36). One example here includes the rise in scamming and identity theft that involves the copying and altering of CVs and professional credentials, which are then used by such identity thieves in order to get translation clients (pp.47–48).
This may be further complicated by the high rate of freelance translators in the industry (Kushner 2013, p.1248; Koskinen and Dam 2016, p.259) as well as the global distribution of professionals and their clients, meaning that buyers often deal with individuals rather than established organisations who protect their reputation by vetting and training the individuals they employ, in addition to national differences in training and professional bodies that could play a role.

Another commonly mentioned challenge to segments of the market comes from volunteers (or amateurs/non-professionals) (Cronin 2010; O’Hagan 2011b - all in Jimenez-Crespo 2013, p.24; Pym et al. 2016, p.36), who are said to threaten the livelihood of professionals (Austermühl 2011, p.15; Fernández Costales 2013, p.104) and blur the boundary between the professional translator and the amateur (O’Hagan 2009 - in McDonough Dolmaya 2012, p.168; Dolmaya 2011, p.104) while primarily serving corporate interests through exploitation of unpaid workers (Dodd 2011 - in Dolmaya 2011, p.98; Flanagan 2016, p.159).

Unsurprisingly, those that have already established their positions in the translation profession would wish to control and restrict entry (Koskinen and Dam 2016, p.263) and might be vocally against threats to the profession (Flanagan 2016, p.153). This may take the form of (maybe rightfully) condemning the quality of crowdsourced output or warning against the dangers to quality (American Translators Association 2009; O’Hagan 2009, p.114; Désilets and Meer 2011, p.32; Dombek 2014, p.4; Flanagan 2016, pp.160–161; Orrego-Carmona 2016, p.4), as well as taking offence at the suggestion that they work for free (Flanagan 2016, p.158), in parts to the threat free labour presents to the status of the profession (Dolmaya 2011, pp.99–100).

However, there is no clear evidence to support the idea that 'non-professional' translation threatens the status of 'professional' translation. For one, it cannot be taken as given that content that is being crowdsourced would have been otherwise translated by professional translators, since it is also possible that it would have not been translated at all.

In some cases, fan projects have been described as willing to stop the project when being asked to do so by the game’s intellectual property owner (O’Hagan 2009, p.109). However, such projects could either highlight the need for localisation and motivate
companies to hire professionals to create an official version, or it could give the impression
that professional localisation isn't needed since the crowd has already provided a version.

Crowdsourced initiatives may also serve as a training ground for novice translators
(O’Hagan 2009, p.110; O’Brien and Schäler 2010, p.6; McDonough Dolmaya 2012, p.183)
as well as providing new employment opportunities in project consulting, management and
quality assurance (Austermühl 2011, p.16; Désilets and Meer 2011, p.33) and giving more
visibility to the topic of localisation through the recognition given to volunteer translators
(Dolmaya 2011, p.103).

4.1.3.1.2. Distinguishing Features of 'The Crowd' and 'The Professionals'

One central question here has yet to be addressed: Even if it was possible to clearly
demarcate 'professionals' from 'non-professionals' in localisation, would this distinction
then actually be a reliable indicator of quality that can be expected and hence serve as a
suitable extrinsic cue? After all, the argument could be made that translation capability
primarily requires an individual to be bilingual with monolingual communicative and
linguistic abilities (Pérez-González and Susam-Saraeva 2012, p.150). This notion is also
supported by older theories of translation that argue for the notion of 'natural translation',
conducted by individuals with no training and little exposure to translated contents, or the
concept of 'native translation' which suggests that individuals are capable of picking up
translation competence from their environment, similarly to their linguistic competence,
and in the public perception it doesn't appear to be clear whether translators are seen to
have an 'exclusive' body of knowledge (Pérez-González and Susam-Saraeva 2012, p.150).

Also against the background of the vast amounts of bilingual speakers (European
Commission 2012, p.12) it would then be possible to argue

"that it is the professional - rather than the non-professional - translation that should
be taken as the exception within the wider context of translation"
(Pérez-González and Susam-Saraeva 2012, p.157),

while others draw the line between the professional and the non-professional based
on remuneration (Pym 2011 - in O’Hagan 2011a, p.14; Pérez-González and Susam-Saraeva
2012, p.151; Orrego-Carmona 2016, p.3), which would indicate little quality difference
between the two groups.
Yet, the assumption and argument that there is a difference between what a 'professional' can deliver compared to a 'non-professional' has been made numerous times in literature.

Examples include

- the practice of using professional translations as the gold standard in an experiment to evaluate quality provided by the crowd (Zaidan and Callison-Burch 2011; Berg 2013; Deriemaeker 2014),
- the expectation that there is a better output via ProZ.com recruitment than Amazon Mechanical Turk (Kushner 2013, p.1246),
- the suggestion that non-professional translations on LinkedIn would lead to quality issues (American Translators Association 2009),
- the practice of employing professionals to check translations done by the crowd (Moran et al. 2012, p.160; Dombek 2014, p.70),
- professional translators suggesting that non-professionals can't translate social media platforms to the level that a professional could (Flanagan 2016, p.162),
- and the distinction being made between three translation worlds (industrial, craft and amateur) (Gouadec 2007 - in Fernández Costales 2013, p.105).

Several suggestions have been made why a 'professional' translator might be able to provide better quality than a 'non-professional'. These include among others, a code of ethics, which those in crowdsourcing don't necessarily know of or follow. Part of such a code of ethics prohibits translators from carrying out work they don't have the required competences for (Flanagan 2016, p.160), suggesting that the crowd would be more likely to take on tasks they don't have the competences for than professional translators. Professional translators must also not only strive for excellence, fairness, truthfulness, trustworthiness, empathy, courage and determination (Chesterman 2001, p.146), but also be able to find and evaluate alternatives during translation through language skills (contrastive linguistic and cultural), technical and research skills (Chesterman 2001, p.147). Competencies suggested for professional translators include translation pedagogy, translation quality assessment and
criticism, professional ethics and norms (Pérez-González and Susam-Saraeva 2012, p.150), as well as "linguistic, extralinguistic, instrumental, and strategic knowledge about translation", which is not knowledge possessed by all bilinguals (Jiménez Crespo 2013, p.26).

However, researchers have also acknowledged that while "training, expertise, association or a combination of these criteria can work as a signal of trustworthiness and could be used to distinguish between professional and non-professional translators"

(Orrego-Carmona 2016, p.2), it is not clear how and to what extent any of these signals are linked to quality of the individual's output. This is also reflected in the observation that non-professional translators improve as translators over time (Barra 2009; Dwyer 2012; Massidda 2012 - all in Orrego-Carmona 2016, p.2) as well as the suggestion that for any translator, expertise isn't static, but instead something that changes throughout a translator's lifespan (Muñoz Martín 2009, p.26).

Added to this is the lack of clarity regarding whether or not translator training is necessary for being a professional translator (Koskinen and Dam 2016, pp.258–259) or whether experience might have more market value than academic qualifications (Koskinen and Dam 2016, p.259).

In fact, Pérez-González and Susam-Saraeva (2012) suggest that non-professionals may have an advantage, since due to their lack of training they are able to innovate rather than being 'indoctrinated' with the need to stay close to the target text. Additionally, they are part of the audience in many cases (p.158). The concept that non-professionals may have an advantage in some situations could also be observed on Facebook, where for some translations the crowd outperformed the professional translators due to knowledge of the Facebook interface (O’Hagan 2009, pp.113–114). In addition, in the area of fan translations, fans might also have knowledge that compensates for their lack of professional experience, like linguistic, pragmatic and discursive features that the professional might lack (Jiménez Crespo 2013, p.26) and the non-conformity of fans’ approaches to translation seems to gain acceptance, leading to fan translation groups being hired for official versions (O’Hagan 2009, p.101). Also in game translation, a not uncommon practice is where
individuals get hired by organisations based on their experience as a gamer and knowledge of the gaming world, as well as 'some' translation experience, even when they do not necessarily have previous localisation experience in that field (Munoz Sanchez 2009, p.182). This then raises the question regarding the fact that if individuals who were previously fan translators are now being paid by an organisation to perform the same task, should they be considered a 'paid amateur' or a 'professional'? In addition, does this change in status suggest a change in quality output?

With this discussion, it is not the researcher's intention to suggest that there is no difference in the quality a 'professional' translator will be able to provide, compared to a 'non-professional'. In fact, especially in cases where individuals occupy extreme ends of the spectrum (the professional being a domain expert, experienced and trained vs. the non-professional being not trained, experienced or having specific domain knowledge), it is reasonable to assume that there would be a difference in quality output. However, as Neather (2012) states, "who qualifies as an expert and in what sense often depends on where one is standing and with whom one is interacting" (pp.265–266). If the assumption is that the translator's status sends a set of signals to indicate relative trustworthiness (Pym et al. 2016, p.33) and is therefore used as an extrinsic cue for perceived quality judgments, then the fact that the definition of 'professional translator' vs. 'non-professional translator' as well as associated capabilities is fuzzy suggests that the individual calling themselves 'professional' and the individual using this declaration as a cue might associate different features and capabilities with this short-hand. This suggests that the reliability of this extrinsic cue is questionable at best.

4.1.3.1.3. Who are the Volunteers?

So far, the assumption in the discussion has been made that, while with fuzzy boundaries, there are boundaries between 'the crowd' and 'the professionals' in that one group engages in crowdsourced projects, while the other group conducts professional work. However, as already indicated with the example of fan translators being hired by organisations (Munoz Sanchez 2009, p.182) and crowdsourcing serving as a training ground for novice translators (O’Hagan 2009, p.110; O’Brien and Schäler 2010, p.6;
McDonough Dolmaya 2012, p.183), the boundaries are not only fuzzy, but also porous, and the distinction

"further complicated by the fact that any of these mediators can provide their services on an amateur or professional basis, as acknowledged by the remuneration they receive for their work"

(Pérez-González and Susam-Saraeva 2012, p.151).

In the literature on the topic, two suggestions have been made:

- that crowdsourcing is conducted primarily by non-professionals or amateurs (Cintas and Sanchez 2006, p.38; O’Hagan 2009, p.96; Austermühl 2011, p.2; Dolmaya 2011, p.103; O’Brien 2011, p.19; Tatsumi et al. 2012, p.2; Fernández Costales 2013, p.91; Flanagan 2016, p.151)
- and that it also involves individuals who otherwise work as professional translators (O’Hagan 2011a, p.13, 2011b, p.13; McDonough Dolmaya 2012, p.178; Flanagan 2016, pp.151–152).

Evidence to support either can be found through several small-scale studies that also reported on the demographics of their sample.

For a study on crowdsourcing on Wikipedia, McDonough Dolmaya (2012) reports that 68% of respondents had no formal training in translation, while approximately 30% had some sort of formal training (p.173). Of the 75 respondents, 68% had never worked as professional translators and most were not directly involved in translation-related professions, while approximately 15% were primarily working in translation or a translation related-field (McDonough Dolmaya 2012, p.174). The results of this case study also indicate that former and current translators are primarily engaged in project management, editing and revision rather than translation, which was covered primarily by non-professionals (McDonough Dolmaya 2012, p.178).

O’Brien and Schäler (2010) report on a survey conducted among the volunteers translating at TRF in 2010. Here, 222 volunteers declared themselves to be 'professional translators', while 35 said they were 'amateurs'. 42% said they had a lot of work experience, while 18% described themselves as 'beginners', and of the 35 amateurs, 23% said they had no experience (p.4).
These two examples with conflicting outcomes, one showing a crowd that consists primarily of non-professionals and the other mostly of professionals, underline the uncertainty of just how porous the boundaries between 'the crowd' and 'the professionals' are. This is highlighted by the original definition of crowdsourcing, which speaks of an "undefined, generally large group of people" (Howe 2006) and can involve anonymity in the sense that the crowd doesn't know the crowdsourcer and others in the crowd, as well as a lack of defined borders (Hosseini et al. 2014). This anonymity and lack of defined borders has also been commented on by researchers in the context of the crowdsourcing platforms Amazon Mechanical Turk, which "does not provide any information about Turkers." (Zaidan and Callison-Burch 2011, p.1221) and Cucumis, which is described as looking "like the home of the empowered anonymous amateur." (Perrino 2009, pp.69–70. With this, such platforms stand in contrast with platforms where professional translators can be sourced, like ProZ.com and Translators Café, where translators are asked to take responsibility for their work as well as encouraged to fill in all details and use their real names (Perrino 2009, p.73).

A key factor here is the recruitment and selection strategy of individual initiatives, or whether a community should be closed or open. Open communities are generally not restricted other than requiring registration, while closed communities vet participants before they are able to join the project (Mesipuu 2012 - in Flanagan 2016, p.151). Recruitment and vetting strategies for either type can include, depending how open or closed the community is:

- announcements through blogs and social media,
- providing preferred profiles of translators for a project,
- voting mechanisms,
- review and validation (Ray and Kelly 2011, p.7),
- active engagement to become a decision-making member trusted with training new members and ensuring quality (Pym et al. 2016, p.44),
- needing to have reached a specific level of language learning before being

23 http://www.cucumis.org/
able to proceed to translations (Deriemaeker 2014, p.27),

- careful selection and management of experts (Anastasiou and Gupta 2011, p.638) or
- the absence of any strategy, in cases where users don't even have to register in order to participate (Perrino 2009, p.67).

### 4.1.3.1.4. Motivation to Volunteer

Inherent in the question of who participates in crowdsourcing initiatives is also the question of what motivates them to do so. This is linked to:

- pillar two (the crowdsourcer, which can be for example a for-profit or non-profit organisation)
- and pillar three (the task or the activity the crowd participates in)

(Hosseini et al. 2014), in that both imply different motivations for participation, which is then linked with the rewards (whether tangible or intangible) crowdsourcing participants receive.

While motivation is not a primary concern of this research, it is nonetheless of interest to briefly consider two very narrow questions in relation to motivation:

1. Which types of participants can be attracted to initiatives based on their motivations?
2. How well do motivational goals support good quality output?

Research around motivation in crowdsourced localisation commonly references self-determination theory as described by Ryan and Deci (2000). Here, a distinction is made between intrinsic and extrinsic motivation, where intrinsic motivation is characterised by "the inherent tendency to seek out novelty and challenges, to extend and exercise one's capacities, to explore, and to learn." (p.70). This intrinsic motivation requires supporting and can also be hindered by extrinsic factors like tangible rewards, which were found to undermine intrinsic motivation (p.70).
Against this background, research on motivation in crowdsourced localisation then distinguishes between extrinsically motivating and intrinsically motivating factors.

Examples of intrinsic motivations include
- making information available to others (Bold 2011, p.7; McDonough Dolmaya 2012, p.181),
- supporting organisations or an initiative,
- projects being intellectually stimulating (McDonough Dolmaya 2012, p.181),
- a sense of community (Dolmaya 2011, p.101),
- humanitarian causes,
- pride in native tongue and
- interest in the content (Désilets and Meer 2011, pp.32–33).

Examples of extrinsic motivations include
- the translator's wish to enhance their reputations (McDonough Dolmaya 2012, p.183),
- gaining experience (O’Brien and Schäler 2010, p.6; McDonough Dolmaya 2012, p.183)
- and improving language skills (O’Brien and Schäler 2010, p.6; Désilets and Meer 2011, p.33).

Overall, there seems to be a combination of motivations in volunteer translators (Jiménez Crespo 2014, p.30). While in some cases both, professionals and non-professionals are motivated by the same factors (here altruistic motivations as well as gaining experience) (McDonough Dolmaya 2012, pp.181,183), in other cases there seems to be a stronger focus on extrinsic motivations for professional translators than for non-professionals (McDonough Dolmaya 2012, p.188).

Although a lack of monetary reward is not a distinguishing factor of crowdsourcing since at least as can be estimated from the literature most projects are not financially rewarded, it would likely play a role in the quality of crowdsourced localisation when
considering motivation theory.

One example of paid crowdsourcing is Amazon Mechanical Turk. Yan et al. (2014) suggest that the motivation for workers hired through this platform is financial, as opposed to other initiatives (p.1135). Whether this is the original motivation to participate, self-determination theory suggests that any intrinsic motivation participants might have would be hindered by the monetary rewards offered (Ryan and Deci 2000, p.70). Additionally, rewards paid tend to be very low (Callison-Burch 2009, p.287; Zaidan and Callison-Burch 2011, p.1223; Yan et al. 2014, p.1135). The detrimental effect of low monetary rewards described by Ryan and Deci (2000), may then partially explain why researchers working on those platforms report considerable variations in quality (Callison-Burch 2009, p.288; Zaidan and Callison-Burch 2011, p.1221) and participants ignoring instructions or cutting and pasting machine translation rather than providing their own translations (Callison-Burch 2009, p.289).

This suggests that different motivations for engaging in a project - be it paid or unpaid - can attract different individuals, as well as having implications for the level at which they engage. In the case of intrinsically motivated work, for example supporting social justice by participating in humanitarian crowdsourcing initiatives, intrinsic motivation would be likely to be higher than for projects where a low monetary reward is offered to work on a topic that primarily furthers a company's financial interests, which might also show signs of less dedication and lower quality outputs.

4.1.4. The Trommons Community and Platform

The present research was conducted on broadly two groups: the Trommons community (volunteer translators and non-profit clients); and the wider localisation industry (students and lecturers, for-profit buyers, translators). Section 3.3. described the localisation industry as it relates to this thesis. The following section will describe the organisation and platform in more detail, and introduce the Trommons community.

Trommons is an open-source, web-based platform which is administered by TRF. TRF is a registered charity in Dublin, Ireland, which has worked since 2009 to eradicate the information gap faced by communities under-served by commercial localisation through
making information accessible in the languages of those communities. Due to the commercial focus of the localisation industry, only language combinations and projects that are potentially profitable are covered by commercial localisation providers. However, this means that vital information around justice, education, healthcare, and economic wellbeing is not accessible to millions of people around the world, which is where TRF and the Trommons community (language volunteers and non-profit organisations) provide an important service with the goal of alleviating this situation (Centre for Global Intelligent Content 2013; The Rosetta Foundation 2014, 2016a, 2016b).

As of March 2017, Trommons connects more than 16000 volunteers with more than 360 organisations. Trommons runs on open-source technology called SOLAS (Service-Oriented Localisation Architecture Solution), which was originally developed through research within CNGL I\textsuperscript{24} and CNGL II\textsuperscript{25} at the Localisation Research Centre at the University of Limerick. Currently, only one part of SOLAS (SOLAS match) is in use on Trommons. Volunteers primarily provide translations and reviews free of charge to non-profit organisations and projects.

### 4.1.4.1. Trommons volunteers

Through its volunteers, Trommons can cover a set of languages that many translation service providers might find challenging, allowing for content to be localised for under-served or hard to source language communities. However, common languages in localisation like French, Italian, German and Spanish are also in the top 10 of the languages the volunteers identify as their native language. When asked about their native language in the survey questionnaire which will be described in section 6.4.1. of this thesis,

- 26% of the respondents identified as Spanish speakers,
- 13.14% English (Position 3),
- 6.99% Vietnamese (Position 4),
- 5.72% French (Position 5),
- 4.66% Italian (Position 6) and

\[\text{\footnotesize \textsuperscript{24}http://www.adaptcentre.ie/}\]
\[\text{\footnotesize \textsuperscript{25}http://www.adaptcentre.ie/}\]
German only on position 10, after
Arabic, Russian and Greek.

Position 2 is held by the ‘other’ category which includes 47 languages with fewer than 5 native speakers within the survey respondents and accounts for 17.80%.

9.4% of the respondents identified as bi- or trilingual when asked about their native language, with 22.4% of those that named more than one language identifying with the combination Spanish and Catalan, followed by Ukrainian and Russian (8.2%).

The results of this survey also indicate that the average Trommons volunteer is
female (77%)
and between 18 and 49 years old (89.5%)
with the age category 25-34 being represented strongest (39.9%),
followed by the groups 35 to 49 (27.4%)
and 18 to 24 (22.2%).

When asked about their highest level of education, none of the respondents selected ‘No studies’ or ‘Primary education’. The highest number of respondents selected
Bachelor or equivalent as their highest level of education (45.8),
39.2% Master or equivalent,
6% Third level,
.1% Doctorate or equivalent,
3.1% Secondary education and
1.7% Short-cycle third level education.

75.7% of respondents said that they had studied translation or a related discipline.

Most of the respondents had been registered on Trommons for less than a year, with 20.2% being registered for between 6 months and 1 year, 21% between three and 6 months, and 27.4% less than three months. Only 31.4% had registered more than one year prior to the time of answering the survey, which is also in line with the rapid growth of the
volunteer database through new registrations reported by TRF staff.

The responses might suggest that the volunteer community on Trommons is fundamentally different from other crowdsourcing initiatives like Wikipedia (see section 4.1.3.1.3) in that survey responses could be interpreted to mean that Trommons attracts mainly professional translators, while others like Wikipedia mainly attracts non-professionals.

However, the limitation of the data gathered in this survey, due to errors in question wording, is that it is unclear if the highest level of education respondents indicated might be in a non-translation subject in some cases or, in other words, whether the respondents indicating they had studied translation or a related discipline had done this to Bachelor's or Master's level or whether some respondents may have understood the word ‘studied’ to include informal studies or lower level certifications and similar. In addition, the question of what is a related discipline to translation, is left open. This means that a very wide range of potential subjects could be considered related by respondents. Another issue is that it is unclear whether volunteers on Trommons currently work as professional translators or have done so in the past.

Nonetheless, while the level of formal translation knowledge in the community should not be overstated based on this survey questionnaire, the results suggest a highly-educated community with a high level of formal knowledge in translation and related disciplines.

This is also partially supported by the questions aimed at understanding the volunteers’ motivations to contribute on Trommons. Of 399 volunteers that answered the question "How much do you agree with the following statements about volunteering?",

- 256 (64%) strongly agreed or agreed with the statement "I want to gain experience to get a job",
- 364 (91%) strongly agreed or agreed with the statement "I like to learn through hands-on experience" and
- 320 (80%) respondents strongly agreed or agreed that it was important to them to receive evaluations of their work.

Similar results are shown by O’Brien and Schäler (2010), who report that a
relatively high number of volunteers self-identified as having a beginner level or only some expertise in translation (p.4) and after wishing to support TRF’s cause, a highly motivating factor for volunteering was the wish to gain professional translation experience (p.6).

In addition, among factors motivating future volunteering, the highest ranking was feedback from both TRF’s clients and professional translators, which supports the observation gathered from the 2014 survey (p.7).

When considered side by side, this seems to indicate that while the average volunteer on Trommons is highly educated and has studied translation or a related field, they may be lacking or feeling they are lacking in hands-on experience in the field of translation and so could not be considered an experienced expert since they are likely to be lacking the experience to have progressed from a lower level of skill acquisition to the expertise stage (Dreyfus and Dreyfus 1980, p.12).

This suggests that Trommons is seen as a training environment by volunteers, although altruistic motivation is also a major factor (Table 1). This is in line with the results of other initiatives like Wikipedia (McDonough Dolmaya 2012, p.183).
<table>
<thead>
<tr>
<th>How much do you agree with the following statements about volunteering?</th>
<th>Disagree or Strongly disagree</th>
<th>Strongly agree or Agree</th>
<th>Neither agree nor disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel it is important to help others</td>
<td>0%</td>
<td>95%</td>
<td>4%</td>
</tr>
<tr>
<td>I like to learn through hands-on experience</td>
<td>1%</td>
<td>91%</td>
<td>7%</td>
</tr>
<tr>
<td>I feel better about myself when I help others</td>
<td>2%</td>
<td>82%</td>
<td>14%</td>
</tr>
<tr>
<td>I want to gain experience to get a job</td>
<td>10%</td>
<td>64%</td>
<td>24%</td>
</tr>
<tr>
<td>I like to be part of a group that shares an interest in community service</td>
<td>4%</td>
<td>71%</td>
<td>24%</td>
</tr>
<tr>
<td>It is a way to focus on something other than my own life</td>
<td>16%</td>
<td>53%</td>
<td>28%</td>
</tr>
<tr>
<td>The reason behind my volunteering varies according to the type of task</td>
<td>17%</td>
<td>46%</td>
<td>32%</td>
</tr>
<tr>
<td>I always pick my tasks for the same reasons</td>
<td>23%</td>
<td>28%</td>
<td>45%</td>
</tr>
<tr>
<td>It is important to me to receive feedback on my work</td>
<td>2%</td>
<td>80%</td>
<td>15%</td>
</tr>
<tr>
<td>I regularly see tasks that fulfill my most important reason for volunteering</td>
<td>14%</td>
<td>34%</td>
<td>45%</td>
</tr>
</tbody>
</table>

Table 1: Motivations for volunteering, Trommons volunteers
4.1.4.2. Trommons organisations

Trommons provides organisations with a web form to upload files which they can request to be translated. This form requests some basic information like the tasks that are needed, word count and language pairs. In addition, organisations can provide some non-mandatory background and reference information. Organisations are also able to add tags to their projects which volunteers can follow to get notified if a task with this tag becomes available. However, the tag system is currently not widely and consistently in use. The organisations are vetted by staff at TRF to ensure that the platform and its volunteers aren’t exploited. Staff also monitor the activities on the platform and supports both volunteers and organisations, in case of issues arising.

The non-profit organisations and charities registered on Trommons represent a diverse variety of non-profit and charity groups, ranging from causes related to culture and recreation, education, health and social services to environment, advocacy, religion and politics. According to the 2014 survey (Section 6.4.1.), most organisations have between one and 10 members of staff (64.1%), followed by 11 to 50 members of staff (23.1%). When asked about their motivation for working with volunteer translators, the majority of respondents indicated that saving money for their organisation was a motivating factor, followed by the quality being better than machine translation. Other strong motivating factors were the importance of getting their message out quickly, wanting to involve others in their community work and there being no other way to translate their content.

4.1.4.3. Trommons platform and workflows

Trommons runs on open-source technology called SOLAS which was originally developed through research within CNGL I and CNGL II. Currently, only one part of SOLAS (SOLAS match) is in use on Trommons.

SOLAS match allows volunteers to select a language combination, preferred tasks and types of projects. Based on their selections it then matches them with projects they might be interested in working on. Volunteers can then look for projects they want to work

---

26 http://www.adaptcentre.ie/
27 https://github.com/TheRosettaFoundation/SOLAS-Match
on through a task stream and claim those projects and the files attached to them. On sign-up, volunteers have to provide only minimal information and there is no general selection or qualification process. Only in some instances, for example when an organisation requires a non-disclosure-agreement to be signed, are specific volunteers approached, who will then be sole collaborators with that organisation. Similar to other volunteer efforts, translators generally receive no financial reward but can request a confirmation of the projects they have worked on through Trommons, which can then serve as a reference for potential career development.

On the platform itself, entire files are uploaded and in some cases split into chunks of several pages to lower the workload of the individual volunteer. These individual chunks are then combined and edited generally for consistency. Volunteers therefore work with unsegmented text and can choose whether they wish to use a computer-assisted translation tool or any other technology for their work. This approach offers the benefit of the translator being able to see the context of individual sentences which provides an advantage regarding the production of quality. After translation, translators upload their work back to Trommons, and it can then be claimed by a proof-reader if this step was requested on project creation. Proof-readers as well as organisations are presented with a form which allows them to provide the volunteer translator and proof-reader with feedback. This feedback form includes four-star rating options as well as a free-form field. However, it is unclear whether the amount and quality of feedback volunteers receive from other volunteers and non-profit customers on Trommons or for other initiatives is suited to help them improve their skills or help them understand what they are already doing well (De Wille 2014, p.76).

In addition, generally low levels of interaction may have a negative effect on the volunteers’ desire to gain experience and improve their translation skills, which may then have implications for the quality of localisation provided by Trommons volunteers. While respondents to the survey (Section 6.4.1.) expressed interest in being part of a community, collaboration within the community is rather low. On a project level, volunteers can interact with the representative of the organisation who has requested the task, as well as staff at TRF. In addition, indirect interaction through reviewer comments and project documents
can take place. Trommons also incorporates a discussion board which would give volunteers the opportunity to interact with the larger community, but this discussion board is only sporadically used.

A need for increased support was also expressed by volunteers in the responses to the question: "Assuming a hypothetical scenario without resource restrictions, in which Trommons became a ‘quality paradise’ for translators and proof-readers, which of the following features would you find most helpful? Please pick THREE items from the list below and give reasons for your choice."

- 67.9% (Position 1) of the 305 respondents selected "Linguistic resources such as translation memories, glossaries, and style guides.",
- 40% (Position 2) "Tutor-system for inexperienced translators";
- 37.7% (Position 3) "A project-specific communication platform for organisations, translators and proofreaders".

It should be noted that two of the top three responses (Positions 2 and 3) indicate that volunteers would not only like to have more support for their translation work, but that they would also prefer to receive that support from the community, suggesting a need for increased connectedness and to learn from a larger group of experts.

This can also be observed in the lack of feedback and opportunities to improve on their translations offered to volunteers by the organisations requesting the translations. In the previously mentioned survey (further outlined in section 6.4.1.), representatives of non-profit organisations registered on Trommons were asked about their experience related to quality of the projects they had requested through the platform. 38.2% of the 34 respondents on the non-profit organisation side indicated that they had encountered problems in the past when it came to quality problems, with the majority (61.54%, Position 1) of issues occurring very often, often or sometimes being "Linguistic problems (e.g. terminology, grammar, style, register) with translated (target language) file” (13 respondents).

While on first look this appears to be a large number of issues, it is unclear how the quality was assessed for these projects, what the background was of those who did the
assessment and whether issues observed were related to previously provided guidelines. However, 72.7% of the organisation representatives stated that they reviewed quality after the translation/proofreading task on Trommons had been completed, indicating that there may be low confidence in the quality of the output on Trommons. 53.8% (Position 1, 7 respondents) said the issues were fixed on the organisation side ("We fixed it ourselves"), while only 7.7% contacted the volunteer about issues ("We contacted the volunteer"; Position 4, tied with "We ignored it" and "We posted the task on Trommons again"). None of the respondents selected "We used the Trommons Forum to ask a question / discuss the issue with volunteers."

This low level of interaction, though it does not necessarily indicate a difference in objective or perceived quality, distinguishes the Trommons community from other cause or product driven crowdsourcing initiatives, which tend to have a stronger focus on the community aspect and collaborative approaches to translations (Section 4.1.2.). Where other crowd sourcing efforts will see hierarchies emerging either naturally or by them being formally implemented (Perrino 2009; McDonough Dolmaya 2012, p.178; Dombek 2014). Due to the lack of interaction between volunteers and the absence of a formal structure that relates to evaluation of their work, no hierarchies can be observed to have emerged within the Trommons community except for some very active volunteers having been featured in the regular newsletter published by TRF28. Other community-related activities of a small number of volunteers include attending and organising events related to Trommons, donating funds to TRF or more regular interaction with staff at TRF.

This lack of interaction does not necessarily exclude the group of users interacting with the Trommons platform from being considered a community due to the shared goal of providing access to information to underserved language communities (Geyer et al. 1996, p.63). However, the community aspect is nonetheless relatively weak and the overall group of users on Trommons could also be considered a large umbrella group that contains a set of informal, loosely defined sub-groups based on individual projects, as well as individual volunteers working separately from other volunteers.

28 https://www.therosettafoundation.org/blog/
4.1.4.4. Trommons: Managing and Producing Quality

Previous sections described Trommons, its community, technology and goals based on the researcher's observations as well as literature and survey questionnaires. However, this only allows for a superficial perspective from the outside. To gain a deeper understanding of processes and the community on Trommons in relation to managing and producing quality, in-depth interviews were conducted with staff at TRF. The following sections report on the outcomes of those interviews and the themes that emerged from them. The two members of staff interviewed had been at TRF for 1.5 and 4 years. One had a background in translation studies while the other did not, but had previously worked at other non-profit organisations. Both were engaged, among other responsibilities, in the day-to-day operations of TRF and Trommons. The interview process is further described in chapters 5 and 6.

4.1.4.4.1. The day-to-day of Trommons

Early in the interviews, the informants were asked to describe their day-to-day activities at TRF and how Trommons was managed on a regular basis. One of the interviewees described that the original idea with Trommons had been to simplify the way projects were managed and to cut out the project manager role.

However, both informants described their involvement in projects on Trommons to include:

- Advising organisations on how to create a project or creating it for them;
- Advising organisations on using the platform and project aspects like expected turnaround time;
- Checking projects, including the information provided and amending it if it was needed, source files to ensure they can be downloaded and 'look okay', number of languages and number of words are correct;
- Sometimes assign a project to a volunteer if the volunteer encounters technical difficulties with Trommons or the projects don't get picked up, which includes emailing volunteers directly or finding volunteers through other platforms.
4.1.4.4.2. Managing quality on Trommons

When asked about how quality is managed on Trommons, interviewees also described high levels of involvement from their side.

As a first step, staff tend to check many of the files that are uploaded for translation to ensure that the file is of suitable quality for translation, although this is not a rigorous process. One of the main aspects that is being checked here is that the files can be edited by the translators without major technical difficulties.

Another mechanism utilised for managing quality is encouraging organisations not only to create a translation task, but also a proofreading task, to filter out potential issues with the translations.

While proofreading is done primarily by volunteers, staff at TRF also commonly check translations after they have been uploaded on Trommons, if they speak either the language combination or the target language. Based on their assessment of aspects like grammar, spelling, consistency and fluency, they will then add another round of proofreading if needed although this intervention is not strictly part of their job description and Trommons was originally designed for minimal intervention by a project coordinator.

Another quality control mechanism involves the selection of volunteers. While anyone who wishes to do so can sign up to volunteer and claim tasks that are available on Trommons, without there being tests or other checks, for several projects staff have contacted volunteers by email to encourage them to volunteer for a task.

TRF Staff 2 estimated that approximately 30 to 35% of the projects involved directly contacting pre-selected volunteers, while TRF Staff 1 estimated that for up to 80% of the projects on Trommons, there was some kind of intervention relating to quality by staff at TRF.

4.1.4.4.3. Quality on Trommons

Apart from the previously described aspects of grammar, spelling, consistency and fluency that TRF Staff 1 described as factors being evaluated when translations were checked by TRF staff, TRF Staff 2 also described priorities in quality for the non-profit organisations on Trommons:
"I would say that in that case that normally like NGOs wants to spread a message, so I would say that you can read the content, it's easily understandable, then of course some would be like the deadline involved aswell in that cases because some of them made it quite quick and they are long documents. So yeah, I would say that they are aware that the quality, if they need - you know - in a short deadline, some of them might be aware that the quality might degrade." [TRF Staff 2]

Both interview participants described the translations on Trommons as generally good:

"I would say it's mostly good. Sometimes excellent, and sometimes awful."  
[TRF Staff 1]

"To be honest, I mean, sometimes there is some bad feedbacks, but in general there's not a lot of bad feedbacks from the proofreaders or from the organisation. Actually, all the bad feedbacks are from the volunteers who proofread some document because the organisations never complain"  
[TRF Staff 2]

On the topic of whether the low number of complaints from organisations indicated a low number of quality issues or merely that organisations didn't complain about quality issues that did exist, opinions were mixed. TRF Staff 2 thought that organisations would let TRF know if there were issues but that in many cases it would be difficult for them to evaluate the quality due to their not speaking the language combinations for projects they requested. However, this interviewee also thought that the organisations are generally very grateful for the help they receive with translations and might therefore not feel comfortable complaining about low quality or that they might not check for quality because the quality wasn't very important to them as long as the texts could be understood.

TRF Staff 1 agreed that there was little insight into how happy organisations were with the translations due to lack of feedback from them, but emphasised the need to gain more insights into this.

However, TRF Staff 1 also explained that this lack of feedback was understandable when considering that engaging with translations was only a small part of what the NGOs had to accomplish and that this would make it difficult for them to invest the time to engage
more.

While staff at TRF have contacted volunteers in the past if feedback on their work was poor, the overall approach is that reviewer feedback is the main mechanism for helping volunteers improve and there are no steps taken to exclude individuals from volunteering in the future.

Another topic discussed during the interviews was what the informants thought the organisations on Trommons knew about the platform, its processes and volunteers. TRF Staff 2 thought that the organisations were generally familiar with the processes due to creating tasks themselves, and that they generally didn't have a need to know more about processes. However, this interviewee also mentioned that some organisations might be a bit suspicious of Trommons since the concept of translations being provided for free via an organisation that specialises in this topic might seem alien to some.

4.1.5. Empirical Evidence: Quality in Crowdsourced Localisation

Unfortunately, only little empirical evidence on the question of quality in crowdsourced localisation can be found in the literature, an observation also made by Jiménez Crespo (2014, p.29).

One area of study has been on professional subtitling relative to non-professional subtitling (Bold 2011; Berg 2013; Orrego-Carmona 2016).

Through an analysis of a fansubbing website, Legendas.tv29 observed translations rated by 353 users to be 'of top quality' (Bold 2011, p.8).

In an unpublished Bachelor's thesis, Berg (2013) compares fan translations and official translations of the subtitles for the anime 'Neon Genesis Evangelion' by considering foreignisation, domestication and 'quality and/or accuracy' (p.9). Foreignisation and domestication are especially of interest in the fan translation area since goals and approaches between official and fan versions differ (Cintas and Sanchez 2006, p.47; Munoz Sanchez 2009, pp.178–179; O’Hagan 2009, p.105; Bold 2011, p.15; Berg 2013, p.49;

29 http://legendas.tv
Based on a sample selected from the subtitles where sampling was conducted based on the "special interest to the questions posed" (Berg 2013, p.10), the observations are made that:

- There is little difference between the two texts in regard to foreignisation and familiarisation;
- There is an apparent lack of overarching translation strategy, clear aim or purpose and resulting "odd word choices, strange omissions and inconsistent familiarization" in the fan version;
- The fan version contains "typos and syntactic irregularities while the official translation does not".

(Orrego-Carmona 2016) collected data on the reception of subtitled versions of the TV show 'The Big Bang Theory', comparing a professional version extracted from an official DVD with two non-professional versions produced by different subtitling communities through questionnaires, eye-tracking and interviews (p.163). Here the focus of the study was non-professional subtitling that aims to imitate professional study (p.3) instead of the previously mentioned examples where fans apply different conventions to their work than official versions. The conclusion of this study was that there was not enough evidence to confirm the main hypothesis (p.15), which was that participants' comprehension would be higher with professional subtitles compared to non-professional ones (p.5). However, some differences in attention shifts and mean fixations were observed, suggesting that "on average, professional subtitles allow participants to have a more consistent reading behaviour" (Orrego-Carmona 2016, p.15).

Zaidan and Callison-Burch (2011), Yan et al. (2014) and Tatsumi and Aikawa (2012) have all conducted research into the performance of 'the crowd' in combination with machine translation/machine learning techniques.

Zaidan and Callison-Burch (2011) investigated the possibility of combining crowdsourced translations and obtaining 'professional quality from non-professionals' by "soliciting multiple translations, redundantly editing them and then selecting the best of the
Comparisons were made by comparing the crowdsourced translations to a professional translation gold standard by using a BLEU, which is a metric originally designed for the evaluation of machine translation (p.1225). Here, they observed that translators sourced via the crowdsourcing platform Amazon Mechanical Turk, who were paid relatively low amounts of money for the task ($0.10: translating a sentence; $0.25: editing a set of ten sentences; $0.06: ranking a set of four translations) (p.1223) commonly contained linguistic issues and reflected non-native English, "but [were] generally done conscientiously (in spite of the relatively small payment)." (Zaidan and Callison-Burch 2011, p.1221). They conclude that it is possible to obtain high quality from non-professional translators (p.1228), by selecting the best output out of a number of redundant translations.

Similarly, Yan et al. (2014) apply a two-step process involving non-professional translators and post-editors with the goal of achieving professional quality translations by selecting the best crowdsourced translation from different candidates through a co-ranking method (p.1142).

Tatsumi and Aikawa (2012) approach the combination of crowd and machine from the opposite direction, by investigating crowd post-editing of machine translated content. Here, 200 foreign students were invited to participate in a crowd post-editing exercise and were given instructions to avoid over-editing, ignore stylistic differences and in the event that the machine translated output was too low, provide their own translations (p.71). The same instructions were then given to professional translators, who were asked to score randomised translations that were either machine translation output or post-edited by the crowd, without knowing how the translations had been sourced (p.72). In the event that translators provided their own translations due to not being able to find a satisfactory version in the translations offered, which was the case in 7 to 25% of the cases depending on language (p.74), differences in edits made by the crowd and professionals were also evaluated. Tatsumi and Aikawa (2012) conclude that crowd post-editing produces 'good enough' translation around 50 to 70% of the time, although differences can be observed depending on the length of the sentences as well as languages. In addition, in cases where crowd post-edited versions were not considered good enough, only small amounts of
revision were necessary (p.75).

Deriemaeker (2014) and Jimenez-Crespo (2013) compare crowdsourced translations with texts produced by professionals (Deriemaeker 2014), or found in text corpora (Jiménez Crespo 2013).

In his unpublished master's dissertation, Deriemaeker (2014) investigates the quality of the output produced by users of the language learning/crowdsourcing platform Duolingo\(^\text{30}\) with that of professional translators. While the justification of the weighting method applied for this study is at times somewhat problematic, it is nonetheless interesting to note that even though the crowd in this instance primarily consisted of language learners who were translating from the language they were trying to learn into their native language (p.38), errors were mostly target-language specific, like spelling, punctuation, and grammar issues (p.44) and not limited to issues with meaning (p.44). He concludes that although total amounts of errors for the crowdsourced and professional translations was almost identical, the error weighting suggests that the crowd's quality output is significantly lower than that of the professional translations, although "crowdscourers produced a more than adequate result." (Deriemaeker 2014, p.44)

Jimenez-Crespo (2013) compares crowdsourced translations of navigation menus and dialogue boxes on Facebook with translations in a corpus containing similar content from 25 social networks in Spain, focusing on verbal use and terminological conventions (p.23). Here the purpose of the study is to explore how natural-sounding translations produced on Facebook are and concludes that "both verbal use and terminological conventions based on navigation interaction are quite similar in both textual populations" (Jiménez Crespo 2013, p.40). He also suggests that one factor may be that through the widespread use of Facebook, the crowdsourced translations found there might have also impacted the conventions in use for such contents (p.40-41).

These studies suggest that the output of 'the crowd' relative to 'the professionals' differs on various levels and to various degrees. While in some studies the output was found to be 'quite similar' (Jiménez Crespo 2013; Orrego-Carmona 2016), and in others the crowdsourced versions 'good enough' (Tatsumi et al. 2012), or 'more than adequate'\(^\text{30}\) https://www.duolingo.com/
(Deriemaeker 2014), other studies suggest greater differences (Berg 2013) or that quality is best achieved through machine learning intervention (Zaidan and Callison-Burch 2011; Yan et al. 2014).

Nearly all the studies mentioned found the crowdsourced versions to be of lower quality than professional output. However, due to the limited number of studies as well as different approaches to gathering data, variations in outcomes, and different crowds being observed, it is not possible to conclude with certainty that crowdsourced outputs are of lower quality than professional outputs, and if so, to what extent and in which aspects.

### 4.2. Summary

As outlined in chapter 3, differing levels of measurement knowledge and motivation can be predicted for the individual parties in the localisation process chain. This chapter has outlined the current state of crowdsourcing in localisation as well as two extrinsic cues that could serve as surrogate or short-hand when an individual in the localisation process has low measurement knowledge but still has to make a purchasing decision. Here, the focus was on the professional status of the translator as well as the role price pays in impacting the translator.

Boundaries between 'professional translators' and 'the crowd' are fuzzy and appear to be porous, indicating that while the distinction might be indicative of quality in some cases, it can't be used as a reliable extrinsic cue to predict quality.

Similarly, price appears to be an unreliable indicator. While better paid localisation projects could attract more experienced or educated linguists as well as allowing the linguist to dedicate more time and care to the project, cognitive dissonance theory suggests that if a linguist is intrinsically motivated to contribute to a cause without payment, offering a low amount of money could negatively impact intrinsic motivation.
Part II: Methodological Approach

Chapter 5: Methodology and Mixed Methods Research Design

5.1. Introduction

Before outlining the detailed approach to collecting and analysing the data for this research, this section will first consider the suitability of different participant groups and justification for their inclusion in the research.

Deval et al. (2012) showed a link between the theories individuals have about the way something was created and the way they perceive it. This implies that we should not only learn which version of localised content would be preferred depending on the circumstances in which it was created. We also need to ask whether these preferences are based on objective features of this content, or whether the perceived quality might be influenced by factors like naive theories (Deval et al. 2012) or extrinsic cues (Golder et al. 2012). Therefore, the purpose of this research is to gain insights into prevalent naive theories about crowdsourced translations and how these theories impact the perception of the quality of such translations. Finally, it seeks to measure the extent to which these theories and perceptions then impact the choices made during the selection and review process (Figure 9).

5.1.1. Researching the impact of 'naive theories' on perceived quality in localisation

Section 2.4.1. described measurement knowledge as a distinguishing factor in perceptions of quality. High measurement knowledge of the intrinsic attributes of localised content would suggest that extrinsic cues might be less dominant. To gain a deeper understanding of the relationship between measurement knowledge and the impact of extrinsic attributes on perceived quality, a range of participant groups with different anticipated levels of measurement knowledge in localisation were invited. In addition, demographic data on educational background and experience was collected. Participation was invited from:
1. Students at the University Alfonso X el Sabio in Madrid (UAX)

While conducting experiments with student groups can be convenient since student groups are easily accessible, this approach can also be problematic when considering population validity since population groups with specific characteristics can think and act differently than others (Oates 2006, p.99; Sani and Todman 2008, p.35).

In the case of this research, inviting participation from students offers opportunities beyond convenience that make this choice a suitable approach. For one, students who have chosen to study translation and localisation related subjects are likely invested in the concept of the translation profession and the benefits of formal translation education since they have taken steps to enrol and study in such courses. In addition, by inviting student participants it is possible to estimate relatively precisely what the individual participant should know about translation and localisation quality and to what extent they will have already conducted translations themselves. While there is an element of variability in that students might have gained additional knowledge and experience elsewhere or indeed might not have studied all content as part of their courses attentively, there is nonetheless a greater level of certainty possible in this regard than for entirely anonymous participants.

2. Lecturers at the University Alfonso X el Sabio in Madrid (UAX)

Lecturers at UAX were invited to participate because they can be expected to be experienced quality evaluators due to their teaching activities. In addition, lecturers were also expected to be highly educated and potentially also experienced translators.

By conducting the experiment with undergraduate students from different years as well as experienced and educated lecturers, the participant groups represent different stages of a translator’s training and experience progression. In addition, experiments here were conducted in the participants’ native language. Thereby, the impact of the participant's professional background when evaluating translations based on the translator's professional background can be traced without considering language knowledge itself.

3. Buyers of localisation services in the localisation industry:

Chapter 3 outlined the challenges in relation to quality that professional buyers face
- client or vendor side - when making purchasing decisions on localisation. For this group, measurement knowledge is uncertain, and might be of a variable nature. For example, experienced buyers would be likely to have developed strategies to address quality uncertainty in the process while less experienced buyers might have yet to do so.

4. Clients on Trommons:

Clients on Trommons essentially fill a buyer role, without providing monetary rewards. While it is difficult to estimate measurement knowledge for industry buyers, based on previous discussions with staff at TRF, client contacts at Trommons are generally experts in areas other than localisation, and also have job titles and responsibilities primarily not related to making decisions on localised content. Therefore, this group represents client-side buyers with low measurement knowledge and the inclusion of this group provides the opportunity to compare strategies to address quality uncertainty with those of professional localisation buyers.

5. Linguists:

While linguists would not be likely to be making purchasing decisions on localised content, they can nonetheless play a role if they make recommendations as part of third party review. In addition, in the case of language-specific experiments, linguists should have high levels of measurement knowledge for intrinsic attributes and therefore provide comparison values for the participant groups with low anticipated measurement knowledge.

5.1.2. Researching quality definitions of different participant groups

Chapter 3 described that there are not only different definitions of quality, but also different goals that quality evaluation needs to satisfy. The inclusion of the diverse range of participants outlined above also provides an opportunity to investigate whether different groups have different definitions of quality, and how well they can articulate these definitions. While a spontaneous definition may not be fully thought out and covering all aspects of an individual's concept of quality, the comparison across different group allows for broad observations on differences. In addition, the inclusion of different types of buyers
or clients provides an opportunity to compare strategies used to address quality uncertainty.

5.1.3. Researching prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field

By including linguists as well as clients both sourced on Trommons and through an open call, it might be possible to observe different attitudes, 'naive theories' and opinions on crowdsourcing and quality in localisation. While individuals involved in volunteering likely have a more positive attitude towards crowdsourcing, it is difficult to predict what the attitude of participants who are not involved in crowdsourcing might be. However, by observing declared opinions it is then possible to relate those to the impact they might have on perceived quality.

The following sections explain the overall research design and offer a rationale for mixed methods research as well as methodological considerations. This is then followed by an outline of the two main components of the research and the four methods employed, together with a description of data collection and analysis procedures.

5.2. Overview and Overall Study Design

From the literature on crowdsourced localisation, no single clear methodology or framework that could guide this research has emerged. While multiple methods were considered and identified as suitable for this research, none of the known methods alone are enough to approach the collection of research questions. For this reason and other considerations that will be detailed further in the following sections, a mixed methods approach was used.

5.2.1. Mixed Methods Definition

Prior to offering a rationale for a mixed methods approach and a detailed description of the different elements of the research, as well as further consideration, the researcher must consider what defines mixed methods research: Mixed methods studies generally involve as their main features the collection and analysis of both, quantitative and

For the purpose of this research, data was collected by utilising survey methods as well as an experiment phase. While the distinction between qualitative and quantitative data collection and analysis is not always clear-cut, (Creswell et al. 2008, pp.165–166) or for that matter useful for the purposes of this research, the data collection and analysis techniques utilised nonetheless lie on a continuum from ‘more qualitative’ to ‘more quantitative’, thereby justifying the label of a mixed methods research.

Specifically, the survey includes qualitative as well as quantitative data collection and analysis methods with a set of semi-structured interviews, content analysis with data generation and analysis being approached from a qualitative and quantitative angle, and four separate but closely linked questionnaires that include closed and open-ended question formats. The experiment phase includes three separate but similar experiments, where largely quantitative data was collected. A more detailed overview of the elements of the research will be provided in the following sections.

In mixed methods research, data can be collected concurrently or sequentially and is integrated in the process of research (Creswell et al. 2008, p.165; Onwueguzie and Johnson 2008, p.281; Creswell 2015, p.2).

For this research, data collection was approached in a convergent design structure (Creswell 2015, p.6). Different methods of data collection and analysis were conducted separately and the results later merged in a shared analysis and discussion stage. This approach offered the benefit of providing different angles on the topic which could then provide explanations and details for the others (Creswell 2015, p.36).

5.3. Outline of the Research Design

The following sections briefly describe the overall research design to provide an overview and a foundation for the discussion of the following rationale and methodological considerations. They provide an overview of survey and experiment research as well as interviewing, content analysis and questionnaires. Each element is described in further detail in chapter 6.
Figure 9: Data collection methods in relation to research questions
5.3.1. Survey

Survey research is characterised by information that is structured in a variable by case grid, where information is collected from several cases on the same variables. This research commonly involves questionnaires but can also include other data collection techniques (de Vaus 2014, p.3). This is the case for this research, where questionnaires, interviews and content analysis are utilised. By structuring the information in a variable by case grid, it is then possible to compare different cases, and ideally describe causal relationships (de Vaus 2014, p.5) by matching the variation in one variable with variations in other variables (de Vaus 2014, p.7). In the survey method the aim is to observe a naturally occurring variation in variables without attempting to intervene (de Vaus 2014, p.5).

One of the goals of survey research is to get standardised or systematic data from a large group of people or events, which would potentially, depending on sampling strategy utilised, allow for generalisation of the results to a larger population than the targeted group (Oates 2006, p.93).

Survey research allows for a wide range of people to be included and can provide a lot of data in short time at low cost and such larger data sets are commonly suitable for quantitative analysis (Oates 2006, p.104).

5.3.1.1. Interviews

In-depth interviews as qualitative research techniques where a researcher seeks information, perspectives, insights, feelings on behaviours, experiences or phenomena that cannot be observed by questioning or guiding participants (Salmons 2010, p.40) can be conducted with different goals in mind, which inform the interview style adopted (Salmons 2010, p.37). Goals can range from constructing meaning and knowledge with participants or getting information from respondents, depending on the world view taken (Salmons 2010, p.45).

By considering interview participants to be informants with a view of supplementing quantitative data with details and further themes, interviews as a data collection technique also suited the positivist stance the researcher wanted to take for this
In semi-structured interviews, the interviewer prepares a list of themes to which he or she would like to know the interviewee’s response. There must be a willingness to change the order of questions, ask additional questions if they result from the respondent's answers, and allow the respondent to raise new issues and themes if they fit in the overall research interest (Oates 2006, p.188).

Interviews provide the benefit of allowing the researcher to get detailed information by asking complex and open-ended questions. Through these questions it is possible to explore emotions, experiences or feelings as well as sensitive issues (Oates 2006, p.187).

However, since participants are aware of the purpose of the interview and the fact that it is a research-oriented conversation, there is a risk that their responses will be influenced by this knowledge to a larger or lesser degree. This can especially also be the case depending on how respondents perceive the interviewer. Their role and identity including factors like age gap, sex, ethnic background, social status, education and appearance might have an impact on the responses (Oates 2006, p.189). While the interviewer was not visible to respondents during the interviews conducted for this study, it would have been possible to find information on the researcher by searching on-line, such as the researcher’s education, professional background, publications and age as well as factors of less importance to the research area like sex, ethnic background, social status and appearance. Some of these factors are difficult to avoid (Oates 2006, p.188), and can play a role on the interview outcome. In addition, while the interviewer should take care to not influence the respondent during the actual interview, the presence of the interviewer, the respondents knowing they are participating in a research study, the way questions are phrased, and which responses are further pursued can have an impact on the respondent and the interview outcome.

5.3.1.1.1. Applying the interview technique

The first set of survey data collected consisted of in-depth interviews conducted by the researcher in the role of the interviewer. The goal of this phase was to gain deeper insights into previously defined themes to complement the other parts of the research and
partially inform their design.

For this study, the goal was to understand the social reality of how crowdsourced translations and quality in localisation are perceived, thereby primarily addressing research questions 1 (How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?) and 2 (What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?) (Figure 9). The researcher sought to learn about this reality by utilising semi-structured interviews in order to provide structure to the thought process of the participants (Salmons 2010, p.52). At the same time, the semi-structure gave them the freedom to explore new themes that are relevant to the research area.

For this research, interview recordings were transcribed and transcriptions were initially analysed using a thematic analysis technique. After this, further, cross-method analysis integrated the results of further data collection. The resulting data also informed subsequent questionnaire design as well as providing explanations and richer details for the results of the other elements of this research.

5.3.1.2. Content Analysis

Content analysis as a technique for making replicable and valid inferences from texts, art, images, maps and other contents while considering the contexts of their use (Krippendorff 2004, pp.18–19) utilises content, in the case of this study texts, that was originally produced for readers other than the researcher (Krippendorff 2004, p.22). This aspect of content analysis allows the researcher to investigate the topic with data generated from existing texts, which were created independent of the researcher or the study and many of which were created prior to the start of the study. This provides an opportunity to gain insights from what we could consider fully naturalistic conversations about the topic, without researcher interference. The lack of interference is of importance when considering Heisenberg's uncertainty principle which states that acts of measurement lead to contaminated observations due to their interference with the phenomena being assessed. This interference can stem from the subject's awareness of being tested, artificiality of or lack of experience with the tasks, expectations and stereotypes held by subjects as well as
the experimenter's interaction effects on subjects (Krippendorff 2004, p.40).

However, while the researcher and measurement don't directly influence the subjects, the researcher still has influence on the results in a later stage of data generation. In content analysis data are the result of the procedures the researcher has chosen rather than representing observations or readings. This means that while the analyst works with found texts, in fact data should be considered made, not found (Krippendorff 2004, p.81). This is amplified by some inherent features of texts included in content analysis, in that they have no objective, reader independent qualities with single meaning (Krippendorff 2004, p.22) which means that the meanings invoked aren't necessarily shared with others since the background of the individual plays into the meaning (Krippendorff 2004, p.23) and that meanings of texts are relative to a particular context (Krippendorff 2004, p.24). Therefore, to obtain a clear picture of the validity of the research conducted through content analysis, it is necessary to consider the personal background of the researcher, as well as the communities whose individuals generated the texts analysed.

An overview of the Trommons platform and community was provided in section 4.1.4. The researcher is a German native speaker with experience in the localisation industry as a quality assurance tester, test lead, as well as holding project management and management positions. In addition, the researcher holds university degrees in theoretical linguistics and localisation. As such, the researcher's analysis of the contents was both, guided and likely impacted by this background.

Another benefit of conducting content analysis on found texts is that content analysis can handle unstructured matter as data (Krippendorff 2004, p.41).

"The chief advantage of the unstructuredness of content analysis data is that it preserves the conceptions of the data's sources, which structured methods largely ignore."

(Krippendorff 2004, p.41).

By being context sensitive, unstructured data allow texts that are significant and informative to others (Krippendorff 2004, pp.41–42) and dealing with large volumes of data through small samples is common in academic research (Krippendorff 2004, p.42).

However, there was also a downside to using content analysis for this research: By analysing texts that were generated for other purposes, prior to research start, it was not possible to ask participants for confirmation or clarification of their statements. While it
would in some cases be possible to identify the authors, and contact them, this is not always an option if they have since left the platform to which they originally contributed the content to. In addition, for the ones that could be contacted, they would not necessarily remember having written the comment or what they wanted to express at the time, if sufficient time has passed since.

5.3.1.2.1. Applying the content analysis technique

For this research, one content analysis study was conducted. The sample consists of reviewer comments provided to translators on Trommons.org. The Trommons community has been described in more detail in section 4.1.4. Commonly, after the initial translator has completed and uploaded the requested translation to the platform, a review task is created and a different volunteer then reviews the translation. Upon providing the reviewed version of the translation, the second volunteer is then given the opportunity to also leave a rating of the quality of translation, as well as a comment that can be entered into a free form field. These comments were provided to the researcher by TRF staff and were analysed using manual and semi-automated analysis methods. The goal was to investigate how the Trommons community defines and describes quality in localisation, and how this relates to definitions in the wider localisation industry. The focus of the investigation was on the questions of which quality related issues the reviewers were primarily interested in as well as the consistency of terminology used, thereby addressing research question 1: “How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?”, specifically for the Trommons community (Figure 9).

5.3.1.3. Questionnaire

Questionnaires as a technique for collecting information are closely associated with survey research (de Vaus 2014, p.3). During the questionnaire design phase, the researcher clarifies concepts to be measured as well as which indicators would be suitable to measure these concepts prior to collecting information from individual cases (de Vaus 2014, p.41). Participants are then asked to respond to questions in an open-ended or closed format,
where in a closed format a limited number of responses are presented.

This provides the researcher with the advantage of being able to ask participants very specific questions that are aimed at answering the research question, which gives the potential of great clarity due to the ability to control variables to a large extent. In addition, due to the relatively easy distribution and data collection, especially for web-based questionnaires, it is possible to collect large volumes of data (Oates 2006, p.50) which can later be analysed with less time investment than is the case for data collection techniques that give the informant greater levels of freedom to express their thoughts, like interviewing.

However, questionnaire-based research poses the challenge that great care must be taken to design it in a way that minimises misunderstandings or bias and allows the informant to express their views while allowing the researcher to still analyse the data with relative ease (Oppenheim 2000, pp.128–129). For this study, these challenges were addressed in an extensive testing and piloting phase which will be outlined further in section 6.4.3.

5.3.1.3.1. Applying the questionnaire technique

For this research, a series of self-administered, web-based questionnaires were developed and participation invited from different groups of informants (Table 2). The goal of the questionnaires was to examine whether insights gathered during interviewing and content analysis also applied to a wider group of individuals on Trommons.org as well as the wider localisation field by gathering perspectives from a larger number of sources. With the questionnaires, primarily research questions 1 (How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?) and 2 (What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?) were addressed. In addition, some elements of the questionnaires informed the analysis of the associated experiments, to address research question 3 (What is the impact of 'naive theories' on perceived quality in localisation?) (Figure 9).

One questionnaire (‘collaborative questionnaire’ in Table 2) was self-contained and
developed in collaboration with other researchers, as a result of which it included a large number of questions not relevant to this research. All other questionnaires were based on a master questionnaire developed specifically for this research, and variations were derived from this master questionnaire. Questions were collected and worded based on several different sources, including literature, semi-structured interviews conducted for this research as well as previous questionnaires.

To avoid asking unnecessary questions or confusing informants, questionnaires were further scoped to suit informant groups. This resulted in a series of 6 questionnaires, mostly with only slight variations, which were used to gather information from a diverse set of target audience groups and were partially linked to experiments, which will be further outlined in sections 5.3.2. and 6.5. (Table 2).
Table 2: Overview of questionnaires included in survey part of this research

<table>
<thead>
<tr>
<th>Questionnaire Nr.</th>
<th>Target audience</th>
<th>Subset of</th>
<th>Linked with</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trommons.org volunteers</td>
<td>Collaborative questionnaire</td>
<td>none</td>
<td>This questionnaire was developed in collaboration with other researchers</td>
</tr>
<tr>
<td>2</td>
<td>Trommons.org non-profit organisations</td>
<td>Collaborative questionnaire</td>
<td>none</td>
<td>This questionnaire was developed in collaboration with other researchers</td>
</tr>
<tr>
<td>3</td>
<td>UCL students and lecturers</td>
<td>Targeted questionnaire</td>
<td>Experiment 1 (quality perceptions linked to professional status, Spanish)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Linguists (Trommons.org volunteers + wider localisation/translation industry)</td>
<td>Targeted questionnaire</td>
<td>Experiment 1 (quality perceptions linked to professional status, Spanish), Spanish - limited scope) and Experiment 2 (quality perceptions linked to cost of translation, German)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Buyers wider translation industry</td>
<td>Targeted questionnaire</td>
<td>Experiment 3 (quality perceptions linked to cost of translation, buyers)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Trommons.org non-profit organisations</td>
<td>Targeted questionnaire</td>
<td>Experiment 3 (quality perceptions linked to cost of translations, buyers)</td>
<td>For this questionnaire, questions taken from the master questionnaire were re-worded and additional instructions and definitions provided in order to aid non-expert informants.</td>
</tr>
</tbody>
</table>
Questionnaires 1 and 2 were sent to the user base on Trommons.org, with one version targeting volunteers and another version the non-profit organisations. These questionnaires were developed in collaboration with other researchers at the Localisation Research Centre, as well as staff at TRF. They covered among other topics some attitudes towards and definitions of quality, as well as demographics. This questionnaire informed section 4.1.4., describing the Trommons community.

Questionnaire 3 was developed in collaboration with Dr. Montserrat Bermudez Bausela at the University Alfonso X el Sabio in Madrid (UAX) and combined with Experiment 1 (quality perceptions linked to professional status, Spanish), which was also designed in collaboration with Dr. Bermudez Bausela. Participation was invited from students and lecturers at UAX. The goal of the questionnaire was to gain insights between the experiment results and the participants' professional and educational background, as well as their views on the link between quality and professional status and how to define quality.

Questionnaires 4 and 5 were administered together and branched off a set of shared demographic questions, where branching was dependent on the professional background of responses. Questionnaires 4 and 5 were combined with

- Experiment 1 (quality perceptions linked to professional status, Spanish), inviting participation from Spanish linguists in the industry through an open call, as well as
- Experiment 2 (quality perceptions linked to cost of translation, German), and
- Experiment 3 (quality perceptions linked to cost of translation, buyers),

as shown in Table 2 and further described in section 6.5. These questionnaires were the most extensive ones of the study and aimed at gathering demographic data to aid the analysis of experiment data, as well as views on the topics of crowdsourced translation, professionalism and quality. The goal of these questionnaires in combination with the experiments was to relate the demographic with the way quality perceptions are impacted by the information provided on the cost of the translation or professional status of
translators, as well as views expressed by participants.

**Questionnaire 6** was a simplified version of Questionnaire 5 and combined with Experiment 3. Since the target audience of Questionnaire 6 was a non-expert group of buyers of localisation services, it was necessary to simplify wording and to add further clarifications. In addition, the number of questions was reduced compared to Questionnaire 5. The goal of this questionnaire, in combination with the experiment, was to relate demographic information and views expressed by participants with the way quality perceptions are impacted by information provided on the cost of the translation.

### 5.3.2. Experiments

As highlighted in section 5.3.1., experiment research is similar to the survey method, in that the experimental method also allows information to be structured in a variable by case grid, where information is collected on the same variables for several cases. However, it differs from survey research fundamentally in that the variation in the experimental method is caused by intervention (de Vaus 2014, p.5).

Experiment research allows researchers to investigate and prove or disprove a cause and effect link between a factor and an observed outcome. This approach is thereby suited to empirical testing by allowing the observer to measure a factor, manipulate the circumstances and then re-observe and re-measure the factor to identify changes (Oates 2006, p.127). Utilising experiment techniques for data collection and analysis allows for a high level of precision to show causal relationships between variables (Oates 2006, p.137).

However, as for any method it is also necessary to consider shortcomings of the approach. Experiments are commonly artificial situations that are not comparable with real-world situations (Oates 2006, p.137), which is indeed the case for this research in that generally the scenario encountered by participants where they would have to choose between three different translations or reviewer comments would be unlikely to occur in a real-world situation.

It can also be difficult to control all relevant variables (Oates 2006, p.137), though as following sections will discuss, care has been taken to control as many variables as
possible so as not to skew the results unduly. Finally, another challenge of experiment research can be the necessity to conceal from participants what the purpose of the research is (Sani and Todman 2008, p.8) to avoid them performing in a way that is expected from them by the researcher or due to social pressure. This can be problematic since it is generally considered unethical to deceive research subjects (Oates 2006, p.137). As part of these experiments it was necessary to mildly deceive participants, which will be discussed in section 5.6.2.4.

### 5.3.2.1. Applying the experiment technique

In the context of this research, the goal of the experiments was to investigate whether attitudes towards and opinions on crowdsourcing and quality would have an impact on preferences of participants, and whether the impact would be mediated by factors like their professional background, education and experience. All three experiments were designed to address research question 3 (What is the impact of 'naive theories' on perceived quality in localisation?) (Figure 9).

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Topic</th>
<th>Language</th>
<th>Target audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Professional status</td>
<td>Spanish</td>
<td>University of Madrid; Linguists</td>
</tr>
<tr>
<td>2</td>
<td>Price of translations</td>
<td>German</td>
<td>Linguists</td>
</tr>
<tr>
<td>3</td>
<td>Price of translations</td>
<td>None (reviewer comments)</td>
<td>Buyers</td>
</tr>
</tbody>
</table>

**Table 3: Overview of experiments conducted for this research**

For this research, three different but similar experiments were conducted (Table 3). Each of the experiments followed a repeated measures or within-subjects design where the performance of a single group of participants with the same conditions was measured, then the independent variables were changed and the performance measured again. In the
analysis stage, the performance pre- and post-manipulation of independent variables was compared (Oates 2006, p.135). With choosing this design, the goal was to ensure that the change observed in dependent variables pre- and post-manipulation of independent variables could be attributed only to the independent variables (Oates 2006, pp.134–135). This approach increases internal validity of the experiment setup since participant variables that would otherwise be difficult to control can be accounted for by asking all participants to complete the experiment scenario with and without manipulation of independent variables. While it is possible that in such an approach the duration spent on the experiment can impact the participants' performance (Oates 2006, p.135), this was not of great concern for the experiment setups for this research since the duration spent on completing all experiment scenarios was approximately 30 minutes per participant and there would have been little opportunity for time and external influences to have a great impact.

For all three experiments, participants were presented with a list of options in a web-based form (either multiple translations of the same text or several reviewer comments) and were asked to select their preferred option. This was repeated about a dozen times (depending on the experiment), with one of the following provided together with each of the options:

- Scenario A: No information on professional status of translator or the cost of translation provided;
- Scenario B: Correct information on professional status of translator or the cost of translation provided;
- Scenario C: Incorrect information on professional status of translator or the cost of translation provided;

For this basic design that was then varied depending on target group of participants,

- Scenario A was always administered first and served as a baseline condition (Sani and Todman 2008, p.12).
- Scenarios B and C were the experimental conditions (Sani and Todman 2008, p.12), where the focus of analysis was on the differences in
performance between Scenario A and Scenarios B and C.

The fundamental assumption for this design was that across all participants’ preferences, patterns would emerge, where some options provided would generally be chosen more often than others in Scenario A. By comparing these patterns to those in Scenarios B and C it would then be possible to measure the effect that the manipulated variables, in the form of the information provided, would have on such patterns. However, since the expectation was that at least some participant groups would display a relatively high level of agreement on preferred options, the experimental condition included both, correct (Scenario B) and incorrect (Scenario C) information to eliminate, as far as possible, the variability stemming from objective quality aspects.

Sections 6.5.2., 6.5.3. and 6.5.4. provide a more detailed description of how this general experiment design was applied to the individual three experiment setups.

5.4. Rationale for Using Mixed Methods Research

This research investigates a topic with multiple layers of complexities. A central layer is the attitudes and opinions of a wide range of potential groups with a range of interests and backgrounds that have formed these attitudes. As outlined in the discussion on the theoretical framework for this research (Chapter 2), these attitudes can be impacted by naïve theories and result in changes in perceived quality, which then result in biased choices being made. An additional complexity is that individuals may not be fully aware of their attitudes, naïve theories and biases or may not wish to explicitly state them.

This level of complexity, which is not unusual for social phenomena, can be best understood by applying different methods (Creswell et al. 2008, p.164). For example, one of the disadvantages of survey research can be a lack of depth and detail, since focus tends to be on the breadth of coverage (Oates 2006, p.105). This aspect of depth and detail vs. breadth of coverage is addressed in this research by utilising different data collection techniques. Where semi-structured interviews provide depth and detail but are lacking in breadth of coverage, questionnaires provide insights from a greater number and more diverse range of people, but don't offer the depth and details of the interviews.
The strength of mixed methods research lies especially in the integration of different data sources and perspectives. Rich qualitative data allows us to investigate personal experiences and stories. Through the combination with a quantitative approach we are then able to find more generalisable trends. Through this combination, we can gain a better understanding of a topic than we could through only one approach alone (Johnson and Onwuegbuzie 2004, p.18; Creswell 2015, pp.2–3) and with the complementary approach to using different types of data gain insights into outcomes as well as the processes that lead to these outcomes (Bazeley 2009, p.90).

While the distinction between qualitative and quantitative research in the context of mixed methods research provides an initial categorisation of data collection and analysis methods, the perspectives gained on the topic go beyond this distinction. As outlined in section 5.1., this research includes a wide range of different participants. These have different levels of experience and training in translation and localisation as well as an involvement or lack of involvement in translation crowdsourcing were included. By including a wide range of participants, it is possible to gain insights into a range of attitudes, theories and opinions on the topic, which aids conceptualisation by providing initial trends and themes for further investigation.

This research also considers different types of data and the impact the presence of a researcher has on data and subjects. Fully naturalistic data (Scholfield 1995, p.33) gathered from reviewer comments, where the texts analysed were created for a purpose other than to be analysed for research and for an audience other than the researcher are considered alongside reactive data (Scholfield 1995, pp.33–34) gathered through interviews, questionnaires and experiments. This results in differing degrees and varieties of researcher influence on the data collected. Applying content analysis to texts generated for other purposes is initially least obtrusive (Ackland 2013, p.24) since the texts were generated prior to and independent of their relevance for research. Therefore, due to the absence of researcher-subject contact, the possibility of behaviour change is minimised (Neuman 2006 - in Ackland 2013, pp.24–25). However, as Krippendorff (2004) notes, in content analysis data is generated by the researcher, based on existing texts, so while researcher influence on subjects is non-existent, the researcher's individual perspective and perceptions can play a
significant role on the interpretation of the texts.

In the interview process, the interviewer is continuously present and the interaction between interviewer and participant can lead to the participant being strongly influenced by the interviewer. However, semi-structured interviews allow participants to elaborate on and explain their perspectives, as well as introducing new themes if they wish.

Questionnaires and experiments, in particular web-based ones as utilised in this research, minimise researcher involvement during the actual data collection stage, but depend on categories and concepts that were previously defined by the researcher. So, while the researcher will not directly influence subjects, the constraints of the questionnaire and experiment formats limit participants to a much larger degree than semi-structured interviews. By gathering data through questionnaires and experiments, it is possible to control different variables more strongly than in the previous approaches, while additional insights can be gained through free form fields in the questionnaires.

Finally, as outlined in sections 5.3.1. and 5.3.2., both, survey and experiment methods allow for the observation of variations in cases. However, in surveys the aim is to observe naturally occurring variation, while in the experimental method variations are caused by intervention (de Vaus 2014, p.5). Therefore, experiments can establish cause and effect, which can't be achieved through survey research (Oates 2006, p.105). Due to these similar and complementary approaches, using survey and experimental methods in combination ideally complement each other since they allow for the investigation of attitudes and opinions as well as how those affect actions.

While all four methods of data collection and analysis have their individual weaknesses, they were therefore chosen because of their complementary strengths.

5.5. Summary and Overview of the Research

In summary, this research seeks to understand attitudes towards and opinions on crowdsourcing, quality and professional status in relation to translation and localisation. This information is gathered by conducting interviews, content analysis and questionnaire research with different groups in translation and localisation. The second area of interest is how these affect how quality is perceived, which is investigated through a series of
experiments. Table 4 shows a summary of the individual elements of the research.
<table>
<thead>
<tr>
<th>Respondent group</th>
<th>Semi-structured Interview</th>
<th>Content Analysis</th>
<th>Questionnaire</th>
<th>Experiment</th>
</tr>
</thead>
</table>
| **Linguists (Trummons)**   | Individual observations on Trummons initiative, localization quality, factors impacting localization quality and localisation crowdsourcing. | Analysis of reviewer comments to understand aspects of quality that are important to reviewers as well as shared vocabulary to express quality aspects. | Perspectives on defining and evaluating quality, factors impacting quality, the language profession as well as translation crowdsourcing. | Experiment 1 - Spanish: Impact of extrinsic cues (professional status) on perceived quality.  
Experiment 2 - German: Impact of extrinsic cues (price of translation) on perceived quality. |
| **Linguists (Industry)**   | n/a                                                                                       | n/a                                                                                         | Perspectives on defining and evaluating quality, factors impacting quality, the language profession as well as translation crowdsourcing. | Experiment 1 - Spanish: Impact of extrinsic cues (professional status) on perceived quality.  
Experiment 2 - German: Impact of extrinsic cues (price of translation) on perceived quality. |
| **Non-profit client (Trummons)** | Individual observations on Trummons initiative, localization quality, factors impacting localization quality and localisation crowdsourcing. | n/a                                                                                         | Perspectives on defining and evaluating quality, factors impacting quality, the language profession as well as translation crowdsourcing. | Experiment 3: Impact of extrinsic cues (price of translation) on perceived quality. |
| **Professional buyer (Industry)** | Individual observations on localization quality, factors impacting localization quality and localisation crowdsourcing. | n/a                                                                                         | Perspectives on defining and evaluating quality, factors impacting quality, the language profession as well as translation crowdsourcing. | Experiment 3: Impact of extrinsic cues (price of translation) on perceived quality. |
| **Student and lecturers UAX** | n/a                                                                                       | n/a                                                                                         | Perspectives on defining and evaluating quality and the language profession.     | Experiment 3: Impact of extrinsic cues (professional status) on perceived quality. |
| **Staff at The Rosetta Foundation** | Details on Trummons producing and managing quality; Individual observations on Trummons initiative | n/a                                                                                         | n/a                                                                             | n/a                                                                                                     |

Table 4: Research overview
5.6. Methodological Considerations

Further to the discussion on the overall research design, some additional questions arise.

In addition to considerations that arise from more traditional, offline, design of research methodology, consideration must be given to issues that are specific to web-based environments due to the heavy focus of web-based research for this research and finally general ethical considerations as applied to this research.

5.6.1. Web-based Research

As mentioned in the introduction to this thesis (Chapter 1), crowdsourcing is not a new phenomenon but in its current forms it is largely facilitated and made possible through the world-wide web (web) and its participants. From being able to reach a wider group of potential participants, to increased speed of delivery of content and efficient utilisation of the crowdsourced contributions, localisation crowdsourcing would not be possible to the degree it is now being utilised without web technology.

5.6.1.1. Challenges of web-based research

A major consequence of the decision to conduct the research entirely on the web is that by definition, it excludes the entire population that is not currently connected to the internet. However, with translation crowdsourcing being a largely web-based phenomenon, it is also appropriate to conduct this research largely using web-based technology and in web environments.

In addition, it is reasonable to expect that most localisation professionals regularly access internet platforms when considering the global distribution and heavily outsourcing focused nature of the industry (Section 3.3.).

Another challenge of web-based research can be that users have an offline and an online existence, and that the offline existence can’t be controlled or monitored by the researcher (Oates 2006, p.136). However, since localisation and translation crowdsourcing largely take place on-line, the on-line existence of participants is of greater interest for this
research than their offline one. While it is possible that they would express different attitudes or opinions on the topics covered in this research when asked face to face in an offline setting, their actions and opinions expressed on the web have greater impact on the area of localisation crowdsourcing.

Further, potential difficulties to consider for web-based research stem from the heavy reliance on technology, both on the side of the researcher and the participants. This is especially relevant when considering questions like connectivity and network availability (Oates 2006, p.136). Since the researcher doesn't have information on the quality of the internet connection for the individual participant, it is not possible to know whether a participant who did not finish an interview, questionnaire or experiment did so because they decided to withdraw or because of technical issues.

This aspect could also play a role when it comes to slow connections. If participants had difficulties using the user interface during an experiment or filling in a questionnaire due to a slow internet connection, it might influence the way they interact with the content and therefore the results.

Finally, technical issues might stem from the hardware and operating system of the participants. It is not possible for the researcher to know how many participants took part in the experiment using, for example, mobile devices or what the hardware/operating system/browser combination of the individual participant was. Therefore, it is not possible to estimate whether any of the results might have been influenced by variability in technical setup or technical difficulties.

However, the researcher considers this a low risk and acceptable trade-off considering the benefits of using a web-based approach which will be outlined in the following paragraphs. Concessions to potential technical issues as well as ethical considerations were made by excluding any participants who did not complete the questionnaire or experiment after they had started either, since this should be interpreted as them either having decided to withdraw from the study or having technical issues that would influence the data.
5.6.1.2. *Opportunities of web-based research*

In addition to the nature of the topic and the target population for the different studies, multiple practical considerations make web-based research the favoured approach for this research.

Web-based research allows the researcher to interview populations that are otherwise hard-to-reach due to geographic boundaries, (Ackland 2013, p.34) and cheaply reach a dispersed and diverse audience with web-based, self-administered survey questionnaires (Ackland 2013, p.25) as well as experiments (Oates 2006, p.136).

These aspects are especially beneficial for the current research, since the translation and localisation industry is dispersed globally, and to a certain extent culturally specific attitudes might come into play when considering perceived quality and attitudes towards payment and professional status. While the current research does not specifically consider cultural differences as a variable, it is nonetheless beneficial to get a wide range of participants from different countries and cultural backgrounds to lower the impact of culturally specific attitudes and opinions.

However, conducting web-based research also presents challenges that are unique to each data collection and analysis method, which will be discussed in the following sections.

5.6.1.3. *Interviewing on the Internet*

Conducting interviews using internet technology can have several benefits but also challenges and disadvantages. Before considering these aspects a distinction between two different types of on-line interviewing techniques needs to be made:

- Synchronous communication, which provides the ability to send, receive and respond to messages at the same time (Salmons 2010, p.2) and
- asynchronous communication where researcher and participant send and receive messages at different times (Salmons 2010, p.2).

Synchronous communication can include technology like typed chat in real-time, or Voice over IP (VoIP) technology, while asynchronous communication could for example include an e-mail exchange of questions, answers and follow-ups or forum posts.
For this study, all interviews, with one exception, were conducted in a synchronous fashion using VoIP technology, but without video camera technology. The real time on-line interview was chosen because it approximates some aspects of an in-person interview and provides some unique advantages (Salmons 2010, p.28). The decision not to send or receive images of researcher and respondent was made to avoid a loss of quality for the audio transmission due to demands on bandwidth. One interview was conducted asynchronously at the request of the respondent, by providing them with questions and themes via e-mail and receiving written responses.

5.6.1.3.1. Opportunities of on-line interviews

While the traditional face-to-face conversation where the researcher is able to observe nonverbal signals and listen to verbal expressions (Salmons 2010, p.1) could be considered the ideal scenario since it provides the maximum amount of information to the researcher, conducting interviews on-line removes traditional constraints of in-person interviews.

By conducting interviews on-line, a greater pool of study participants at lower cost (Salmons 2010, pp.8–9) and at greater convenience for participants can be accessed. In the case of this study, it allowed the researcher to gain perspectives from respondents with very demanding time schedules for whom a meeting in person would have been too time consuming, as well as respondents located in several different countries. Getting input from respondents from a mix of countries and cultures is especially important for research in localisation, since it is one of the key features of the field, that individuals and organisations all over the world are part of it.

5.6.1.3.2. Challenges of on-line interviews

However, while conducting interviews for this study on-line removed a number of constraints, it also introduced new ones. During on-line interviews, the researcher is unable to observe the full range of verbal and nonverbal communication (Salmons 2010, p.10) which can be challenging for the actual interview process as well as result in a loss of
information that would have been available during a traditional face-to-face interview.

During the interviews for this study, it was at times difficult for researcher and respondent to understand when the other person had finished speaking, which at times resulted in the researcher unintentionally interrupting respondents. It then repeatedly became necessary for the interviewer to take a step back and encourage the respondent to continue speaking.

For a small number of interviews, internet connections on the participants' sides were slow, which resulted in distorted audio which was difficult to transcribe and resulted in a loss of information.

While one of the respondents was not comfortable with conducting a verbal, synchronous interview and instead elected to provide information in writing, as a general approach the researcher chose synchronous, spoken interviews as the preferred method.

While a written approach can be beneficial since it allows interviewees to consider their responses and provide them at their own convenience (Ackland 2013, p.31), it can also result in a loss of information since by providing all talking points in writing, there is little deep interaction between interviewer and interviewee.

In addition, written responses hold the danger of being more socially desirable responses than spontaneous ones and generally the quality of the data is possibly not as good as that from synchronous or in person interviews (Ackland 2013, p.31). However, in that case the decision was made that a written response would be preferential over no response, so the written interview was included for analysis.

Conducting synchronous interviews verbally as the main approach had the benefits of more closely resembling face-to-face interviews since it might be harder to construct socially desirable responses with a spontaneous reply and added interactivity (Ackland 2013, p.32). While some technical challenges like the ones outlined previously could have been avoided by conducting the interviews in a written chat format, which would have also generated automated transcripts rather than necessitating a transcription step based on audio recordings (Ackland 2013, p.32), the researcher decided against written chat interviews to give interviewees the greatest degree of spontaneity and naturalness possible in their responses, which could have been curtailed in a written chat setup, and might lead to
shorter, less detailed responses due to the effort involved in having to type a response.

A final consideration for conducting on-line interviews has to be the effect on the respondents of using this technology. Critics of internet research have argued that the use of the technology results in a loss of intimacy, immediacy and sense of presence (Salmons 2010, p.3) and it needs to be considered whether this can lead to respondents feeling uncomfortable or lacking the confidence to contribute valuable information to the research.

On this topic, the researcher agrees with Salmons (2010) who argues that
"the word Internet simply does not mean the same thing today as it did even a decade ago. As a result, many criticisms of electronic tools as inadequate for the nuanced, sometimes emotion-filled communication intrinsic to research interviews need to be re-examined."

(Salmons 2010, p.3).

Where information search and retrieval of information used to be more linear, today's on-line interactions are increasingly communicative and reciprocal (Salmons 2010, p.4) and activities that used to require physical proximity are now in many areas of life and work conducted via electronic communications (Salmons 2010, p.2). While this line of thought may not apply for all age groups, cultural backgrounds, professions and lifestyles, it does ring true for the groups invited to participate in the interviews for this study. The localisation industry is widely distributed across the globe. Staff at TRF, buyers from the industry, and also the representatives of non-profit organisations, and to a potentially slightly lesser extent, translators interviewed can be expected to be very familiar with this form of on-line communication from their daily work and personal life. It is also worth noting that during the interview process none of the participants mentioned difficulties with the technology used for the interviews or that they were uncomfortable with or unsure about speaking with the researcher via the internet.

5.6.1.4. Content Analysis on the Internet

While content analysis of web-based texts follows the general steps taken in traditional content analysis approaches, there are some specific characteristics of online contents that provide challenges as well as opportunities for research.

As with all data used for this research, the texts included in the content analysis
section were produced and then gathered on the web. While the nature of localisation, which deals with digital contents which have largely migrated to the web over the past decades, is one reason for this, another is the ease of accessing data through the platforms chosen. While it would have been theoretically possible to record views of different translators in an offline setting, these views would have had to have been elicited specifically for the purpose of studying them, and it would be logistically very difficult to get such a diverse range of opinions.

For this study, reviewer comments on Trommons.org were analysed (Section 6.3.).

While reviewer comments on Trommons.org were generated by users of the platform and could be considered user-generated content as defined by the Organisation for Economic Co-operation and Development (OECD)\(^\text{31}\) (Moens \textit{et al.} 2014, p.8), these texts were developed in a non-collaborative way, with a limited target audience. Individual users on Trommons are not connected to each other and there is a general disconnect between readers and writers of information. While the platform is shared among a group of volunteers, the content each of them contributes - be it translations or reviewer comments - is not accessible to the group of volunteers as a whole and as a result there is no dialogue stemming from the created content. Therefore, each comment represents its own, isolated unit and while some technical challenges from using web-based technology like connectivity, browser compatibility or the user's ability to use a web-based form might arise, the overall nature of how the reviewer comments were created, gathered and used doesn't raise major questions for data gathering and analysis.

\textbf{5.6.1.5. Questionnaires on the Internet}

By administering questionnaires using a web-based form and recruiting respondents through different web-based communication channels (e-mail, social media like Facebook, Twitter and LinkedIn, Trommons.org newsletter, ProZ.com forum thread), it was possible to easily distribute the questionnaire and gather data that was comparatively easy to analyse from a wide group of informants that would have been difficult and costly to reach.

\(^{31}\) http://www.oecd.org/
otherwise. In addition to informants being located in different countries and on different time zones, by offering them the opportunity to fill in a self-administered, web-based questionnaire, this also allowed them to do so at a time that was convenient for them and without major scheduling efforts.

However, while the use of on-line survey questionnaires is increasing due to the low cost but high efficiency of the method, some methodological problems need to be considered (Ackland 2013, p.25).

According to Manfreda et al. (2008), on-line surveys generally score six to 15 percentage points lower in response rates than traditional methods. To increase responses, several reminders were sent after the initial call for participation though care was taken to space those reminders to avoid overloading any of the channels utilised. These reminders were also limited depending on how intrusive the channel of communication was. While the researcher deemed it acceptable to use social media to remind potential participants that the questionnaire was still available, no e-mail reminder was sent after the initial invitation. No further potential approaches like offering incentives, using advanced graphics or multimedia or mixed-modes were utilised since those methods generally seem ineffective while increasing cost and time investment (Ackland 2013, p.30).

Since the questionnaire was self-administered, it would have been possible that one participant would respond to the same questionnaire multiple times, thereby skewing the results. This was partially prevented by restricting the number of responses that were possible based on the IP address of the respondent. IP addresses were only used in the Surveyonkey\textsuperscript{32} questionnaire platform to track responses, but were not exported together with the responses for analysis. While this approach could still lead to multiple responses from one person if they were to access the web form from different locations, or could indeed prevent individuals from responding if they shared an IP address with another person who had already filled in the form, this was deemed an acceptable risk since the total number of people who could not respond due to the IP restriction would likely be very low, and individuals responding multiple times would have to be highly motivated in order to make the effort to submit multiple responses from multiple locations.

\footnote{\textsuperscript{32} https://www.surveymonkey.com/}
Additional challenges that are not unique to conducting internet based research using questionnaires are ethical considerations and sampling, which will be discussed in more detail in sections 5.6.2.3. and 6.4.

5.6.1.6. Experiments on the Internet

Web-based experiments provide the advantage that the outcome of the experiment is only to a limited extent affected by the researcher or experimenter. It also allows the researcher to reach a large number of participants that might be difficult to reach otherwise, especially if they are geographically dispersed (Oates 2006, p.136).

Since the experiments were administered through the same web-based form (Surveymonkey) as the questionnaires, advantages of the approaches were similar in that it was possible to reach participants in a cost and time efficient way that would have otherwise been difficult to reach.

Similarly, challenges with participant credibility and the solutions applied were comparable. Difficulties with low response rates were compounded by the fact that only a limited number of respondents for the questionnaire were invited to participate in the experiment, and few of those invited agreed to do so. This was likely because the questionnaire already had low completion rates and was relatively long. However, since it was beneficial to get detailed information on the background, attitudes and opinions of experiment participants, this was deemed an acceptable trade-off.

5.6.2. Research Ethics

When conducting research that involves human participants it is necessary to carefully consider best practices for ethical conduct for research in general and the techniques for data collection and analysis specifically.

When asked to participate and/or agreeing to participate in research activities, individuals have the right not to participate or to withdraw at any point during research. They also have the right to give informed consent, to anonymity and to confidentiality (Oates 2006, p.56). As a general best practice, researchers should therefore provide an
explanation on what the research is about and debrief participants on findings, give participants the option to withhold data and anonymise names of individuals and organisations. Researchers should also discuss confidentiality issues before starting the research and avoid asking unnecessary questions (Oates 2006, pp.58–60).

When considering ethical issues for individual data collection and analysis techniques, aspects to investigate more closely are that of the differing levels of involvement for human participants depending on the type of research (Oates 2006, p.56) as well as the degree to which the research could potentially harm participants.

5.6.2.1. Interview Ethics

Conducting web-based interviews generally requires ethical considerations that are similar to those for face-to-face interviews with some additional elements like how participants can give or withdraw consent.

Before the interviews were arranged, all participants were provided with information on the purpose of the research as well as asked to sign a standard consent form provided by the Faculty of Science and Engineering Research Ethics Committee at the University of Limerick. This consent form can be found in appendix C.

When conducting web-based research, it can be difficult to tell whether an interviewee has withdrawn from the study (Ackland 2013, p.33) since the researcher can't be sure if they intended to withdraw or whether a non-response or withdrawal is due to technical issues or an oversight. In order to respect the individual's right not to participate and to withdraw at any point during research, in four cases where potential participants had initially indicated they were interested in being interviewed but then did not respond to the interviewer's introduction of the study and request for setting an interview date, no further contact was initiated in order not to send unwanted messages, since it was unclear whether they had decided to withdraw or whether their non-response was due to other, unknown factors.

33 http://scieng.ul.ie/research/research-ethics/
Once interviews had started, the participants were also reminded by the interviewer that they had no obligation to participate in the interview and that they could withdraw at any point. However, no interviewee indicated verbally that they wanted to withdraw during the interviews and no interview was aborted by an interviewee leaving the chat early.

Before being asked the first question, all respondents were reminded of the key elements of the consent form they had signed previously, namely:

- A brief introduction to the topic, paired with the information that the researcher would be happy to discuss the research in more detail after the interview if the participant wished to do so;
- The fact that the interview was being recorded;
- The fact that they could stop the interview at any time;
- The fact that the statements they provided would potentially be used in publications but that their name or other identifying information would be omitted from any reports. Staff at TRF were in addition made aware that the low number of current and past staff members could potentially lead to their responses being linked to them despite the researcher's efforts to prevent this. They were therefore asked to make the interviewer aware of any questions they would not like to answer due to that.

During the interviews, care was taken to not ask participants unnecessary questions. Questions on demographics were limited to training and professional experience as relevant to the research question.

To preserve the anonymity of respondents, information that could result in them being identified, like names of individuals or organisations, was edited in the transcripts and replaced with placeholders. In addition, each respondent was assigned a code name that indicated their role as well as an individual number that does not link to their identity. As a further measure, respondents are referred to with gender-neutral pronouns in the plural form (they, them, their) in this thesis. While not ideal from a stylistic perspective, it provides further anonymity.
During one interview, the participant also divulged information with the request that this information should not be shared with TRF. This information was omitted from the partial transcripts included in appendix D and not included in data analysis.

As a best practice for interview research, participants should be given the opportunity to check transcripts and confirm that the facts are correct and that they represent what they meant to express (Oates 2006, p.194). This was achieved by sending participants draft chapters of the analysis section of this research, asking them to confirm that they saw themselves and their opinions represented correctly and that they had no objections to any information given about them or their interviews. This also served as the debriefing step of this part of the research and provided participants with the option to withhold data if they wished. Of the 11 interview participants included in the thesis, three confirmed that they were satisfied with how they had been represented. One requested that it be made clearer that they had been talking about a hypothetical volunteer rather than the participant themselves and one corrected a factually incorrect statement that had been made by a different participant. The incorrect statement was excluded from the data results. The remaining participants did not respond to the researcher's request.

5.6.2.2. Content Analysis Ethics

As outlined in section 5.3.1.2., the texts included in the content analysis step were originally created for other purposes than research, and without the authors knowing that the content would be then later analysed. While this characteristic provides unique opportunities for gaining insights into the areas of interest, it also raises ethical questions that need to be considered. When dealing with digital trace data like the content analysed for this study, the three major ethical considerations for the research are informed consent, distinction between private and public individuals and participant anonymity (Ackland 2013, p.43).

When conducting research using reviewer comments on Trommons.org, the ideal scenario of seeking informed consent from all contributors in the sense of informing participants about the nature of the study, any risks it entails and ensuring they have understood these before agreeing to participate (Ackland 2013, pp.43–44) would have
entailed contacting a large number of users and tracking which user had given consent, which did explicitly not give consent and following up with non-responses.

However, Ackland (2013) argues that we need to consider the information extracted, how the information is used, what type of environment it is extracted from and whether the content is in the public domain when considering whether consent is required for on-line environments research, especially whether the environment is primarily used for information sharing or for social networking (Ackland 2013, pp.43–44). In the case of offline contexts, this can often be decided on whether data should be considered private or public, though for on-line environments the distinction might become blurred if users believe they are talking to a smaller audience and because there is little indication whether they perceive themselves to be talking to the general public, or only a few individuals (Oates 2006, p.64; Salmons 2010, pp.72–73; Ackland 2013, p.45).

Indicators for whether content should be considered public or not can include the size of the community, the terms of access and the question whether users would expect the content to be read by others.

Reviewer comments are not publicly available and users could reasonably assume that the comments can only be viewed by the original translator, the organisation that requested the translation and staff at TRF. In fact, the reviewer comments are only visible to users with administrative roles.

However, as is stated in the footer of the Trommons.org website,

"unless otherwise stated, all works contributed are licensed under the Creative Commons Attribution 3.0 Ireland License in accordance with the Terms and Conditions." 34

This disclaimer is visible on all major pages on Trommons.org, including the registration and profile pages.

In addition to this disclaimer, the Terms and Conditions state:

"Unless otherwise stated in the task details, all contributions by Volunteers are contributed under the terms of the Creative Commons Attribution 3.0 Ireland License. By claiming a task, you agree that your contribution will be bound by the terms under which the task is published." 35

Since this information is prominently visible in multiple locations on Trommons, it

34 http://trommons.org/
35 https://creativecommons.org/licenses/by/3.0/ie/
is reasonable to assume that volunteers would be aware of those policies and will have agreed to them by contributing their work, with work not only being a reviewed translation but also comments provided with those reviews.

However, it is difficult to estimate whether there would be any potential negative effects for volunteers if they were named as the authors of a specific reviewer comment. It was therefore decided to preserve anonymity by not revealing user names or any personal or professional information that could lead to the individual being identified.

5.6.2.3. Questionnaire Ethics

When conducting questionnaire-based research on the internet, it bears consideration how to handle the fundamental best practises of ethical research. For this series of questionnaires, all participants were presented with an introduction prior to being asked the first question which informed them about the purpose of the study, their right not to participate and withdraw at any point by closing the browser window with the questionnaire. It was also explained to them that any information they provided would potentially be used in academic publications but that they would remain anonymous. Participants were then asked to confirm they had understood these points and were happy to proceed by clicking the ‘next’ button on the web form.

As outlined in the discussion on conducting questionnaire-based research on the internet (Section 5.6.1.5.), it was necessary to collect IP addresses of respondents to limit the potential of one respondent submitting questionnaire answers more than once. However, these IP addresses were not exported together with the answers and were not used in the data analysis. No other identification like names, e-mail addresses, place of residence or work was collected and demographic questions were limited to generic information on professional experience and education, thereby limiting the potential of individual responses being traced back to the respondent.

To provide individuals the opportunity to withdraw from the study, they were instructed to close the browser window with the web form if they wished to withdraw. However, it was impossible to tell whether a participant had decided to withdraw and wished to not have their submitted responses up to that point considered, or whether they
had simply encountered technical difficulties or didn't have enough time to finish the questionnaire, but still wished to have their responses up to that point considered. Therefore, the researcher had to assume that any incomplete questionnaire forms signalled withdrawal and such incomplete questionnaire forms were excluded from subsequent analysis. In some cases, this resulted in a relatively high number of responses being excluded, with only 54% of the questionnaires having been completed in the study with the largest number of participants. However, the researcher considered this a necessary trade-off to respect the participants’ right to withdraw.

5.6.2.4. Experiments Ethics

Since experiments were combined with questionnaires for this research, considerations regarding the right not to participate, withdraw, anonymity and confidentiality apply as discussed in the previous section on questionnaire ethics. However, to avoid influencing data, it was necessary to partially deceive participants in the process of data collection.

In the process of generating translations to be used for the experiments, translators were not informed initially that the translations would be used for this purpose. Instead, translations were requested from them under the pretext of the Localisation Research Centre at the University of Limerick simply wishing to have the text translated. The subject matter of the text fits into the type of texts commonly requested from volunteers on Trommons. It is also related to Localisation Research Centre activities at the University of Limerick. This means that the text could be requested from volunteers and paid-for translators without giving cause for suspicion that their work might be intended for an experiment. This was necessary to avoid translators changing their approach to the translations due to knowing what they would be used for. For example, it is conceivable that they would have had the translations checked by another person or applied additional checks. However, for the purpose of this experiment it was necessary that the translations were completed in a way that was as close as possible to what they would have been for a real-life project.
Then, once they had provided the translations, translators received an explanation of the research. Permission to use the translations for the experiments without revealing the translators' identities was sought and given by all translators involved.

Another mild deception in the experiments was that participants evaluating translations or reviewer comments and selecting their preferred option were at times given the incorrect information about the cost of the translation or translator's background. This was necessary to account for potential variability in the results based on objective quality differences that could potentially be observed. Such cover stories or mild deceptions are commonly necessary in experiment research also in order to prevent participants from guessing the background (Sani and Todman 2008, p.8), thereby biasing data by adjusting their responses accordingly.

### 5.6.2.5. Ethics Approval

Approval for research was sought from the Faculty of Science and Engineering Research Ethics Committee and granted for all elements involving human participants directly. The individual applications were:

- **2016_04_05_S&E** Quality Perceptions and professional status in translation
- **2016_03_15_S&E** Perceived Quality in Localisation
- **2015_05_18_S&E** Cost-based Quality in Translations
- **2015_05_19_S&E** Quality Management in Volunteer Translations
Chapter 6: Data Collection

6.1. Introduction

With the overall outline of the research design and discussion of individual methods in mind, the following sections will describe the approach to data collection and analysis in more detail.

6.2. Interviews

As outlined in section 5.3.1.1., semi-structured interviews conducted synchronously via conferencing software (gotomeeting\(^{36}\)) and VoIP technology (Skype\(^{37}\)) were conducted, recorded, transcribed and later analysed using thematic analysis. The goal of the interviews was to gain deeper insights into the main themes of this research to inform the design of other data collection techniques, as well as to get rich views and descriptions on those topics to provide more detail on those areas than other survey methods could.

The following sections describe the sampling approach, the structure of and approach to the interviews as well as data analysis and validity, generalisability and reliability.

6.2.1. Selection of Participants

Research interviews as ‘conversations with a purpose’ (Salmons 2010, p.94) necessitate that the researcher locates individuals who can provide honest and robust information on the phenomena of interest as well as themselves and who are willing to fully participate in an interview and related communications (Salmons 2010, p.94).

The purpose of the interviews for this study was to gain rich insights into the wider aspects of perceived quality in relation to crowdsourced translation, by exploring the topics of professionalism, definitions and approaches to quality, attitudes towards crowdsourced translations and the perceived link between cost and quality of a translation. Participants were selected through purposive sampling (Salmons 2010, p.95) and invited from four different groups:

\(^{36}\) https://www.gotomeeting.com/en-ie
\(^{37}\) https://www.skype.com/en/
- Staff at TRF. The findings resulting from these interviews have been presented in section 4.1.4.
- Volunteer translators on Trommons/TRF who also provide translations for projects outside of Trommons,
- localisation industry professionals who make buying decisions and
- representatives of non-profit organisations working with Trommons/TRF.

This selection of participants provided a well-rounded set of perspectives on the topic from translator and client side both, in the industry and on the non-profit side as well as insights into the reality of Trommons and the services provided through it.

Participants were selected from existing sample frames through already existing records (Salmons 2010, p.102) and nominated based on predetermined criteria. When using a nomination approach to sampling, other participants or expert recommend potential participants which are identified based on common important patterns (Salmons 2010, p.99). Based on these recommendations, the following 13 respondents interviewed for the study fell into the categories outlined in Table 5.

<table>
<thead>
<tr>
<th>Number of respondents</th>
<th>Group</th>
<th>Sample frame</th>
<th>Nominated by</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>TRF staff</td>
<td>Group of staff at the time of study</td>
<td>Researcher</td>
<td></td>
</tr>
<tr>
<td>4 (+2 pilot)</td>
<td>Volunteer translators on Trommons</td>
<td>Registered volunteers on Trommons</td>
<td>TRF staff</td>
<td>Active volunteer translators on Trommons, also translates outside of Trommons</td>
</tr>
<tr>
<td>2</td>
<td>Representatives of non-profit organisations on Trommons</td>
<td>Registered organisations on Trommons</td>
<td>TRF staff</td>
<td>Main contact person of active organisation on Trommons</td>
</tr>
<tr>
<td>3</td>
<td>Buyers of commercial localisation services without connection to Trommons</td>
<td>Professional network of researcher</td>
<td>Researcher</td>
<td>Experienced buyer with several years in the localisation industry. No professional or individual involvement in crowdsourcing projects.</td>
</tr>
</tbody>
</table>

Table 5: Overview of interview participant groups
In total, 17 participants were approached, of whom 12 agreed to be interviewed in a spoken conversation and one preferred to answer questions in writing. The remaining four did not respond to the researcher's request for an interview. Two interviews were conducted as a pilot to refine the interviewing technique, as well as question phrasing, and were not included in the analysis.

A major challenge when conducting internet research can be the verification of the participant identity (Ackland 2013, p.33) and ensuring their credibility (Salmons 2010, p.111). By sampling based on existing sampling frames where individuals have verified their identities as well as asking for nomination by a trustworthy participant, issues around identity and credibility can be addressed (Salmons 2010, pp.111–112). For this study, all individuals who participated in the interviews were known to either the researcher or staff at TRF and their credibility had been established prior to the interviews through personal and professional interactions, which did however not include conversations on the research topics. However, this approach to sampling also introduces the potential of bias which might have implications for sampling validity. For example, it is possible that having been known to the researcher or staff at TRF prior to the study indicates a similar set of opinions due to a certain amount of shared experiences, while individuals from a different professional circle might have other experiences and opinions. Similarly, no claims to generalisability can be made from this sample.

In addition to the sampling approach, participant identity and credibility was managed by the researcher receiving preliminary information about their background via e-mail during the preparation stage and set up of the interview. Additionally, the participants signed a consent declaration after the researcher had explained the purpose of the study and the participants' rights if they decided to be interviewed. Participants were also asked at the beginning of the interview to provide background information about themselves, such as their connection to TRF and the wider translation/localisation industry.

6.2.2. Interview Data Gathering

Participants were invited via e-mail and individual times for the interview were arranged. Participants were also provided with a consent form which they returned with
signatures to satisfy the requirements of the Ethics Committee at the Faculty of Science and Engineering at the University of Limerick. Participants were also informed that the interviews would take between 20 and 30 minutes. However, during the interview process conversations were allowed to run over time if the interviewee indicated that they were still interested in talking more.

6.2.2.1. Structure of the interviews

Interviews were conducted individually, with only one respondent and the researcher, who also acted as the interviewer, present. The core part of the interviews, not including the explanations and discussion after the researcher had concluded the interview part, lasted between 20 and 40 minutes. Interviews were held entirely in English, though not all participants were native English speakers. As a result, one participant opted to answer questions in writing rather than being recorded in spoken English. For the participants that did agree to a recorded verbal interview, the language barrier did not cause major issues.

One of the tasks of the interviewer at the start of a research interview is to inform or remind the respondent of the purpose of the interview, get permission to record, give assurances about confidentiality and anonymity and put the respondent at ease (Oates 2006, p.191).

To achieve these goals, the start of all interviews followed a similar structure. Initially, the researcher only provided respondents with a general overview of the topic they were providing information on.

Before asking the first question, the interviewer reminded the respondents of the contents of the consent form as discussed in section 5.6.2.1.

Participants were then asked to introduce themselves, including their background in localisation, their connection to Trommons or crowdsourcing in general and their experience and position.

They were then asked a series of questions covering the themes previously defined as well as follow-up questions based on their responses. This allowed the participants to expand on thoughts and thereby provide richer data, while the interviewer was able to get
more information on points the interviewee had made.

Questions to guide the discussion had been defined in advance, but participants were given the freedom to deviate from these questions or to add to them. After the interview had covered the previously identified themes and additional questions stemming from participants' answers, the researcher then offered to tell the participants more about the research, which the majority of respondents were interested in hearing more about. This concluded the part of the interview that was later transcribed for analysis. While the detailed descriptions of the research yielded very interesting further conversations, it was difficult to estimate to what extent the respondent's thoughts were their own or whether they were mirroring the researcher's words, so the decision was made to exclude this section from the analysis.

Topics for discussion broadly included:

- Connection to Trommons and the localisation industry (Covering both, the role the individual plays in relation to Trommons and their education/training/work experience in the localisation industry);
- Experience with quality in localisation. Trommons specific and/or general;
- Definition of quality and ways to achieve and judge it;
- Quality expectations and perceptions in relation to cost in localisation (free vs. Paid-for/expensive translations);
- Theories on motivation and profile of volunteer translators;
- Theories on why some translators work for less money than others.

However, questions differed depending on the background of the respondents and were also adapted iteratively after each interview, as new points of interest arose, and others seemed to be of less relevance than originally thought. In addition, staff at TRF were asked to describe some aspects of processes on Trommons, especially in relation to quality
management.

6.2.2.2. Interview best practices

In addition to topics for discussion, some consideration had to be given to the actual interview process itself. It is in the nature of in-depth interviews for research purposes that during the interview process, it is clear to both, interviewer and respondent, that they are participating in a purposeful, research-oriented conversation that is on the record and can be used for research later unless otherwise specified (Oates 2006, p.187). The assumption, spoken or unspoken, is that the conversation primarily benefits the researcher who has the goal of gaining information from the respondent, and who wishes to discuss specific issues (Oates 2006, p.186). Researcher and participant also generally are engaged in an asymmetrical power relationship where the researcher is in a position of power due to recruiting participants and being responsible for the interview being conducted ethically and productively (Salmons 2010, p.55). Therefore, care needs to be taken to minimise those asymmetrical power relationships through mutual disclosure, promoting a sense of collaboration and affirmation that there is no hidden agenda (Salmons 2010, p.56).

The researcher followed many of the recommendations made in literature (Oates 2006, p.192; Salmons 2010) on interviewing strategies, with the goal of helping to put participants at ease about being interviewed. This was achieved by asking them questions where they could be expected to already have a well-formed view, like questions on how they handled certain aspects of quality in their environment, or what observations they had made, before asking them more difficult or sensitive questions that asked for speculations, judgements or definitions. In addition, the interviewer aimed to put respondents at ease by explaining that the purpose of the questions was not to see if they knew the right answers to the questions, but instead to get their opinions and perceptions as well as emphasising that there was no 'right or wrong' answer they were expected to provide. When respondents expressed doubts over whether their answers were useful for the researcher or were in line with expectations, they were assured that this was certainly the case.

The researcher also acknowledged in some cases that questions were very broad on purpose and that the respondent was free to respond in any way they could think of and
commonly commented on a response as being ‘interesting’ or ‘helpful’. While broad questions at times made it necessary for the researcher to clarify questions or re-phrase them, in other cases it gave respondents the opportunity to elaborate extensively on points and introduce new themes that had not been considered in the preparation of the interview themes. In some cases, participants were also given a hypothetical scenario and asked how they would act in such a scenario to support them with more difficult concepts.

While the researcher took care to not mix her own views and not lead the respondent (Oates 2006, p.193), it was at times necessary to ensure that the researcher had correctly understood the response by summarising it back to the respondent and asking whether that was what they had intended to say. The researcher then judged whether the summary matched what the respondent had intended to say both based on what their answer was (whether they agreed or disagreed with the summary) but also based on how emphatically they agreed or disagreed (did they repeatedly voice their agreement or disagreement, did they respond immediately and spontaneously).

After the interviewer had raised all major talking points and followed up on specific points the participants had made, the interviewer then asked participants whether they wanted to add or clarify anything of what they had previously said. Some participants used this opportunity to expand on or re-emphasise points previously made, while others didn't want to add or clarify anything.

Since volunteer translators formed the largest group of participants, out of the 6 respondents interviewed the first two interviews were used to pilot the interview approach to data gathering for this study. These interviews were not included for further analysis.

Piloting the interview phase was beneficial to the overall process since it allowed the interviewer to fine-tune the questions and reflect on whether interview best practice had been followed and if the approach was suitable as a whole.

6.2.2.3. Interview technology

All interviews were recorded using either built-in functionality in gotomeeting or
recording software for Skype. Recording interviews provides benefits like having a complete record of what has been said, allowing other researchers to listen to the recordings and analyse the data. However, it can also make respondents nervous about being recorded and more inhibited and audio recordings don't capture non-verbal communication (Oates 2006, p.191). Such non-verbal communication is commonly recorded in field notes taken during the interview itself. However, the researcher did not take field notes during the interviews, so as to be able to focus fully on the respondent and the conversation, except for summary points while listening to the responses to be able to ask follow-up questions. This decision was made since the researcher did not expect to gather any information during the interview itself that would not be available on the audio recording since the researcher was only able to hear the respondent during the interview and the later recording without any additional visual or situational information that would have required field notes.

After each interview had been conducted, it was manually transcribed by the researcher using Express Scribe transcription software. Transcribing interviews allows for easier analysis since text-based content can be searched through more easily than audio files (Oates 2006, p.193). However, during the transcription process some information like pitch will get lost and decisions have to be made on how to handle non-verbal sounds and interjections as well as incomplete sentences (Oates 2006, p.194).

Since spoken sentences don't indicate punctuation marks and are often not complete, audible sentences, care had to be taken to faithfully represent the conversation. This was done for the initial transcription used during analysis by commonly not placing any punctuation marks, allowing for run-on sentences, including interjections, expressions of emotion (laughter, sighs) and pauses. In addition, interruptions like a door bell ringing or coughing were included as a comment. To make quotes more appealing to read, non-verbal sounds, repetitions and interjections were removed prior to inclusion in the analysis section of this research, but the full transcription can be viewed in the appendix.

Unfortunately, in some cases the combination of poor audio recording quality and non-native pronunciation for some participants led to difficulty in understanding the audio for transcription. In cases where the audio could not faithfully be represented in writing

38 http://amolto.com/
with confidence, the passages were omitted from transcription and analysis and replaced with the insertion of the term [inaudible].

To preserve the anonymity of respondents, transcriptions were edited as outlined on the section on interview ethics (Section 5.6.2.1.).

Finally, all interviews were checked again by the researcher to ensure the transcriptions were complete and correct in preparation for the thematic analysis step.

6.2.3. Interview Analysis

For this research, coding the transcriptions was identified as the most suitable approach to analysing the data gathered. Codes, in the sense of

"a short phrase that symbolically assigns a summative, salient, essence-capturing and/or evocative attribute for a portion of language-based or visual data."

(Saldana 2009, p.3)

allow for data to be summarised, distilled, condensed and reduced (Saldana 2009, p.4) and thereby aid formation of categories and theories (Saldana 2009, p.12). As such, codes provide a link between data collection and the explanation of meaning (Saldana 2009, p.3) and could be considered an initial interpretive step in data analysis.

Saldana (2009) recommends that researchers consider formatting the content to be analysed into units, which are determined by topics or subtopics shifting focus ('stanzas') (Saldana 2009, p.18). This segmentation was completed together with a pre-coding and analytic memo writing step, which allowed the researcher to record initial thoughts and break down the content according to the codes that would most likely emerge during the coding process. Examples for stanzas were for example the distinction between an interviewee assessing a situation or making a broader statement (e.g.: "translation agencies can sell translations for very little") and the explanations or justifications offered for such statements (e.g.: "because uhm they pay their translators very little" or "and they pay their translators very little because there are loads of translators willing to work for very little")

By separating broader judgement or assessments from the explanations and justifications given, the researcher was then able to separate the overall opinions of interviewees from the theories that had contributed towards these opinions being formed.
Codes were then grouped into themes to match the overall structure of the topics being researched. These themes were then presented in this thesis together with other data to provide explanations and more detailed perspectives.

6.3. Content Analysis

As outlined in section 5.3.1.2., the content analysis part of this research included a sample of reviewer comments provided to translators on Trommons.org. These comments were aimed at providing information about the quality of the translation the volunteer translator had provided and were generated by either a representative of the organisation requesting the translation, or, for the majority of cases, a volunteer who had claimed the task of reviewing the translation previously completed by a different volunteer. The goal of this investigation was to gain insights into what aspects of quality the Trommons community focuses on, and to what extent it uses a shared vocabulary to describe those quality aspects. This was achieved by applying two different coding techniques: One, a fully manual round of coding with the aim of covering the full spectrum of quality aspects mentioned and two, a semi-automated coding round with the goal of discovering the specific terms used. The data resulting from this then allowed for comparison with quality definitions from the wider localisation industry to determine how the perspectives of the volunteer community relate to that of the wider industry regarding the nature of quality in translations.

The following sections describe the sampling approach, approach to the gathering of content and data generation as well as data analysis.

6.3.1. Trommons.org

As outlined in section 5.3.1.2., content analysis is generally conducted on texts that were produced for a purpose other than research and for an audience other than the researcher. This is also the case for the reviewer comments on Trommons.org, which were analysed for this part of the research.

The reviewer comments submitted on Trommons.org were produced with a limited
scope and audience. On the Trommons platform, reviewers can provide feedback on translations. This feedback is provided through a form which includes an option to award ratings in the categories ‘Corrections’, ‘Grammar’, ‘Spelling’ and ‘Consistency’ as well as a free form comment field.

When submitting a reviewed translation, reviewers are prompted to voluntarily provide an evaluation of the translation through this form. While it would technically be possible for the user to enter any written text into that field, the implied purpose and expectation is that the text entered is in relation to the quality of the presently submitted translation. Therefore, a large proportion of text submitted through this form can be expected to address the quality of the reviewed text.

For this investigation, only the comments entered in the free form comment field were analysed. These comments included both, positive and negative remarks since the focus was not on the actual quality of the translations, but rather which quality aspects reviewers would consider important enough to comment on. The following section describes the approach to sampling for these comments in more detail.

By including both, positive and negative remarks, this allowed the researcher to observe which aspects of quality reviewers considered important enough to put in writing as well as to what extent they shared vocabulary to describe those aspects.

There are some potential issues that arise from the text having been produced for this specific purpose.

Users might have differing opinions on the importance of providing comments after the review and might therefore invest different amounts of time and effort into producing the text. Since users don't specifically seek out the opportunity to provide such comments (for example, by directly contacting the translator via TRF through their own initiative or by creating a forum post - both of which would require significantly higher levels of motivation to produce and provide the feedback comments), but are instead offered a form to fill in after having completed a different, more administration-focused and technical activity (uploading a file to the platform). Potentially, this means that formulating a review comment is seen as a hurdle that stands in the way of finalising the project, and might come unexpected to the user if they haven't previously submitted a review.
The potential for the texts entered being low-effort and not reflecting the actual opinions of the user is mediated by several circumstances. The user is not only a volunteer for the project as a whole, but providing comments on the reviewed translation they submitted is also optional. It is possible to submit the file without entering any information into the field, so it can be assumed that the user is not entirely reluctant to enter the text they do provide. In addition, a large number of users of Trommons.org volunteer to improve their translation skills (Section 4.1.4.), which implies a general interest among the volunteers to receive feedback on their translations. While this might not be the case for all reviewers, this implies a general awareness of the importance of providing valuable and valid feedback.

Another aspect to consider is that of the assumed recipient for the comments. While the interface the text is entered in does not indicate or show traces of a recipient for the message provided, there are some basic assumptions the user could make.

Trommons does not have functionality to share reviewer comments with a wider audience. While the fact that the comments are only accessible by a limited number of people is not explicitly stated on the screen that the comments are entered into, the user interface of the platform as a whole is light weight enough for a user to be aware of this. However, without the explicit assurance that the comments will not be shared with a wider audience, the user can also not be entirely certain that this won't be the case.

In addition, there is no information on who will read the text. In the absence of a wider audience, the potential readers are staff at TRF, including anyone this staff provides with access to the comments, the volunteer translator who originally provided the translation, as well as the representative of the non-profit organisation who requested the translation and its review. There is also the potential for the comment not being read by anyone since its being submitted through a form does not provide immediate feedback on the message having been received.

Since translators create an account that is not linked to their offline identity, there is close to no information available to link the individual reviewer comment to the reviewer's background. Volunteers choose a username; their identities are not verified beyond e-mail verification and they also don't have to provide their credentials or curriculum vitae. For
many the express purpose of volunteering is to gain reputation and further their careers since many of them are also professional translators, so they generally would have an interest in appearing in a positive light. As a result, they would be likely to filter their comments accordingly.

On the other hand, if they do end up appearing in a negative light then they would be able to prevent this from being detrimental to them. It would only mean that would not receive the gains from having volunteered that they might have been hoping for.

The fact that this is a volunteer platform also allows for more courteous commenting, since the volunteers are not in direct competition as would be the case for a professional platform. In such a platform, it is beneficial for a reviewer to make the translator look bad to hopefully get the translation job in the future. On the other hand, appearances and impressions are also more important in a professional setting since reputation can be directly detrimental or advantageous for future earnings.

Based on these considerations and in the absence of information of the individual user's motivations and attitudes, the assumption is that as a general tendency, reviewer comments are created relatively spontaneously and unplanned based on previously formed opinions during the review process itself. While anonymity plays a role, and could have an impact on the nature and tone of the remarks, the relatively small number of potential recipients and the overall focus on reciprocity would likely mediate this effect.

In summary, the reviewer comments on Trommons could be considered as voluntary, spontaneous, with relatively low effort being put in, semi-anonymous, yet generally written with the aim of appearing professional and being helpful. They are likely to reflect the opinions of the person who has written them faithfully, but not necessarily fully.

6.3.1.1. Sampling Trommons.org

Prior to coding the content in order to create data for analysis, the questions of what represents a unit of content and what sample to draw from a specific sampling frame need to be considered (Krippendorff 2004, p.83). While unitising allows for a distinction of what constitutes a segment of text for analysis, sampling is then aimed at limiting the content to a
manageable, representative subset of all possible units (Krippendorff 2004, pp.83–84).

As outlined in section 5.6.1.4., the reviewer comments on Trommons.org are generated on a platform that is shared among a group of volunteers who are all potentially submitting content to the platform if there is a project available that they then choose to claim and work on. However, individual users are not connected to each other beyond the interaction of volunteers with the organisations they are providing translations or reviews for or the interaction between organisations and TRF and volunteers and TRF. Volunteers are generally not connected with each other, and neither are organisations connected with other organisations. Connections are usually project-based and the content contributed by each individual volunteer is generally not accessible to the volunteer community as a whole. As a result, all reviewer comments are provided in isolation of other reviewer comments and do not represent elements of a wider dialogue. None of the comments include hypertext and since each of the comments is a project-based unit by itself, linearity or the absence of it is not a concern. There are also no media other than text included. For the purposes of this research, each reviewer comment was treated as its own sampling unit of text and analysed in its entirety.

However, reviewer comments on Trommons.org could still be considered a moving target since the platform's user base and numbers of projects is ever growing. Therefore, the sample for analysis was limited to include only comments created from the launch of Trommons.org on March 18th, 2013 (http://www.therosettafoundation.org/blog/press-releases/) up to and including the entire year 2015.

For the analysis, relevance sampling of all reviewer comments submitted within the specified time frames was conducted. In relevance sampling, the units of text are intended to be the population of relevant texts, excluding the textual units that do not possess relevant information (Krippendorff 2004, p.119).

A total of 1112 lines of reviews were exported by staff at TRF and provided to the researcher. Part of the challenge of conducting content analysis is that as opposed to other data collection and analysis techniques in survey research, not all texts collected for content analysis are equally informative since the texts are generated for other purposes. The interest in content analysis is not an equal representation of the textual universe, but instead
a focus on those texts that are relevant to the research and reaching a fair answer to the research question (Krippendorff 2004, p.113). For this reason, the full set of 1112 lines of exported reviews was triaged to only include lines that contained comments on the quality of the translation. The triage steps were:

1. Marking and excluding of all lines (553 in total) that did not include a comment in the free form field
2. Marking and excluding of all lines (342 in total) that did not include details on specific quality aspects the reviewer had focused on (e.g. thanks to the translator without any quality related comments, generic remarks on the quality of the translation without details on specific aspects)

The remaining 219 lines were coded prior to analysis, as described in the following section. Due to the sampling approach of including all reviewer comments within a specific time frame on Trommons, the results can be considered generalisable for the Trommons community. This is potentially limited by the fact that only those reviewers who chose to post a comment were included and no follow-up conducted with those who didn't provide a comment.

6.3.1.2. Data Creation and Analysis Trommons.org

During the process of coding in content analysis the analyst observes and interprets the text being read according to observer-independent rules (Krippendorff 2004, p.126). Categories applied to texts must be exhaustive and mutually exclusive (Krippendorff 2004, p.132).

In order to facilitate comparison with quality definitions used in industry and academia, the decision was made to use a priori categories defined in the QT Launchpad Multidimensional Quality Metrics (MQM) (Lommel and Melby 2014) and their definitions in order to code Trommons reviewer comments, adding to those categories if needed.

While there are a wide range of quality evaluation frameworks established and in
use in translation and localisation (O’Brien 2012), many of those that are used to evaluate human translations are limited in the granularity of categories offered as well as in their precise definitions. The MQM, which were designed for the evaluation of machine translation and harmonised with the TAUS Dynamic Quality Framework (DQF) error typology (Translation User Automation Society 2015), allow for different levels of detail and dimensions and include descriptive listing "of over 100 types of translation quality issues derived from a careful study of existing translation quality evaluation metrics” (QT Launchpad 2014). These range in detail and are arranged hierarchically (QT Launchpad 2014). Therefore, the MQM were chosen as a framework due to this extensive set of categories with definitions and examples to aid understanding of the categories (QT Launchpad 2014). In addition, the MQM framework is laid out in a format that allows for it to be used as a decision scheme during the coding process. In such a decision scheme, each datum is the outcome of a predefined sequence of decisions. This approach allows for a larger number of categories to be handled while minimising unreliabilities as a result of different levels of generality and meaning overlap (Krippendorff 2004, p.135).

It should be noted that this use of a quality evaluation framework does not reflect the originally intended use. Where the usual process would be that a reviewer would be provided with a source text and its translation, and then describe issues observed by using a quality evaluation framework, in this case the researcher did not access the source text and translation but only categorised reviewer comments into those a priori categories (Burchardt and Lommel 2014). Essentially, the researcher was assigning codes the reviewer would have assigned had they used the MQM, based on the comments they submitted. Since the evaluation wasn't targeted at specific quality issues, as recommended for evaluation of machine translations, all MQM categories had to be included in the coding. This use of an established quality evaluation framework was considered conceptually appropriate since both, evaluation framework categories and reviewer comments remark on issues and quality of translations, with the difference that the framework categories express these concepts in a distilled version, whereby for this study the researcher takes on the task of distilling free-form evaluation comments into such categories in order to reduce them to manageable representations for ease of analysis.
However, from a practical perspective the approach proved difficult to work with. Since the framework includes more than 100 codes, it was difficult to assign the most suitable code. This was also reflected in the results of the analysis of inter-coder reliability. Here, a second coder received the MQM codes and an explanation of the process the researcher had used to apply the codes. This resulted in an average inter-coder agreement for manual coding of 0.624 (Cohen's Kappa). While the researcher considered this outcome acceptable to include the analysis results in this thesis by a narrow margin, it would be advisable to amend the approach for future research to improve inter-coder agreement as a threshold of 0.6 indicates tentative agreement for Cohen’s Kappa (Pfeil and Zaphiris 2010, p. 13).

As outlined in more detail in section 6.3.1.2., reviewer comments were coded in two separate steps: fully manual and semi-automated.

6.3.1.2.1. Semi-automated coding

In their evaluation comments, the reviewers used their own wording to describe specific quality concerns. As a result, the process involved comparing and matching the comments provided to the descriptions in the MQM descriptive listing, and from that assigning codes based on MQM-compliant labels to the individual comments in order to be able to quantify the occurrence of quality aspects mentioned (Lommel and Melby 2014).

A first check included searching for the label names or parts of them, as well as derivations of the names and synonyms or subsets (Table 6).

Whenever a match was found, the context the word appeared in was evaluated to ensure that the word had indeed been used to describe the quality issue as defined by MQM. If this was the case, the respective label was assigned and placed in a field separately from the results of the second check. The results of this evaluation will be referred to as ‘semi-automated’ going forward.

Example for false positive (search word ‘terms’):

There were quite a number of mistakes in terms of spelling but this may be a software error as it seemed to only relate to letters Ö" & Ä¶.
Example of positive (search word ‘terms’):

*This document is very thorough and needs to have attention to details, terminology research skills that the translator missed up.*

*I've kept the terms in other language but personally since they are not names from an institution/company... so I would have changed them.*
<table>
<thead>
<tr>
<th>MQM label</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>accurate</td>
</tr>
<tr>
<td>mistranslation</td>
<td>mistranslations, mistranslated</td>
</tr>
<tr>
<td>terminology</td>
<td>terms, terminology</td>
</tr>
<tr>
<td>overly-literal</td>
<td>too literal, bit literal, quite literal, very literally, word by word, word-to-word</td>
</tr>
<tr>
<td>no-translate</td>
<td>translated</td>
</tr>
<tr>
<td>date-time</td>
<td>date, time</td>
</tr>
<tr>
<td>omission</td>
<td>missed, missing, forgotten, omitted</td>
</tr>
<tr>
<td>untranslated</td>
<td>phrase forgotten, forgot to translate</td>
</tr>
<tr>
<td>addition</td>
<td>additional, additions</td>
</tr>
<tr>
<td>fluency</td>
<td>fluent, flow</td>
</tr>
<tr>
<td>variants-slang</td>
<td>slang</td>
</tr>
<tr>
<td>stylistics</td>
<td>style, stylistic, stylistically, style omissions</td>
</tr>
<tr>
<td>unidiomatic</td>
<td>idiomatic</td>
</tr>
<tr>
<td>inconsistency</td>
<td>consistent, consistency, consistency mistakes, consistency of formatting, formatting inconsistencies</td>
</tr>
<tr>
<td>spelling</td>
<td>misspellings, misspelling</td>
</tr>
<tr>
<td>diacritics</td>
<td>accent</td>
</tr>
<tr>
<td>punctuation</td>
<td>comma, commas, quotation mark (general quotation mark, not specific enough to be unpaired-marks)</td>
</tr>
<tr>
<td>unpaired-marks</td>
<td>quotes, quotation marks</td>
</tr>
<tr>
<td>grammar</td>
<td>grammatical, grammatically</td>
</tr>
<tr>
<td>word-order</td>
<td>word order</td>
</tr>
<tr>
<td>function-words</td>
<td>prepositions, article, articles, pronouns</td>
</tr>
</tbody>
</table>

Table 6: Derivations, synonyms and subsets of Multidimensional Quality Metrics used in semi-automated checks
6.3.1.2.2. Manual coding

A second check included a manual content analysis, with an initial count of the number of quality issues mentioned per evaluation comment to ensure that all issues had been assigned a label. Then, high-level (parent) labels were assigned to each of the issues mentioned as a first step of applying the decision tree provided in the MQM layout (see example in Figure 10). Finally, the decision tree was followed to the smallest branches in a third pass to replace parent labels with the lowest level of child label possible. However, in many cases, the comments were not specific enough to allow for that level of granularity. For this reason, parent labels and their children were initially used in parallel, then later children were added to their parent labels to aid analysis and discussion. The results of this check will be referred to as ‘manual’ going forward.

- **Accuracy (accuracy)**
  - **Addition (addition)**
  - Improper exact TM match (improper-exact-tm-match)
  - **Mistranslation (mistranslation)**
    - Ambiguous translation (ambiguous-translation)
    - Date/time (date-time)
    - Entity (such as name or place) (entity)
    - False friend (false-friend)
    - Mistranslation of technical relationship (technical-relationship)
    - Number (number)
    - Overly literal (overly.literal)
    - Should not have been translated (no-translate)
    - Unit conversion (unit-conversion)
  - **Omission (omission)**
    - Omitted variable (omitted-variable)
  - Over-translation (over-translation)
  - Under-translation (under-translation)
  - **Untranslated (untranslated)**
    - Untranslated graphic (untranslated-graphic)

*Figure 10: Example of MQM parent-child category relationship* (Lommel and Melby 2014)
One example of proof-reader comments read:

*It was clear that the translator had not made any research on the subject - he/she had for instance translated the name of the organization (Plant-for-the-Planet).*

Which means it belonged into the lowest child category (no-translate)

Another example of proof-reader comments:

*There seemed to be a few cases, within complex expressions, where word order could have altered the conveyed meaning.*

This was assigned to a category one level up, ‘Mistranslation’ which is defined in MQM as "The target does not accurately represent the source content". (Lommel and Melby 2014)

Initially, ‘no-translate’ was labelled as a category separately from ‘mistranslation’, though for later analysis the child category ‘no-translate’ was grouped into the parent category ‘mistranslation’.

Whenever applicable, MQM labels were assigned during the semi-automated and manual coding according to their definition. However, some labels were added or their definition amended since none of the MQM labels fitted the wording in the evaluation comments entirely (Table 7). These added labels were assigned to MQM parent labels, although some limitations arose. Some could be assigned to multiple parent labels due to the intended meaning of the original comment being unclear, or labels overlapping with several parent categories. In addition, some labels could not be assigned any MQM parent as their definition fell outside of the MQM scope.

If a comment mentioned a category multiple times, each mention was assigned a label except for examples that were given to specify the issue mentioned. This decision was made based on the assumption that repeated mentions of the same type of issue shows that the reviewer felt that the issue was of importance.

Most comments addressed the translated version of the text. If they concerned the original source text, labels were also assigned.
After coding, the number of occurrences for each code were summed up separately for the semi-automated and manual coding process. Then the lowest levels of child codes were grouped into their parent codes to show higher level patterns and aid the discussion of the results. Finally, codes were categorised in accordance with whether they occurred fewer or more than 10 times in the manual coding results.

<table>
<thead>
<tr>
<th>Label</th>
<th>MQM Parent</th>
<th>Nature of remarks</th>
<th>MQM Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>design-unclear</td>
<td>Design</td>
<td>Ambiguous, potentially related to design.</td>
<td>Introduced</td>
</tr>
<tr>
<td>inconsistency-format</td>
<td>Design or Fluency</td>
<td>Inconsistencies in formatting</td>
<td>Introduced</td>
</tr>
<tr>
<td>literal-unclear</td>
<td>Accuracy</td>
<td>Literal translation without indicating whether accurate representation of source content is affected.</td>
<td>Introduced</td>
</tr>
<tr>
<td>mistranslation-unclear</td>
<td>Accuracy</td>
<td>Ambiguous, potentially related to mistranslation.</td>
<td>Introduced</td>
</tr>
<tr>
<td>MT</td>
<td>Style, if implied. Otherwise, no MQM parent or whichever symptom is assigned to this root cause.</td>
<td>Assumed use of machine translation.</td>
<td>Introduced</td>
</tr>
<tr>
<td>national-language-standard</td>
<td>Locale convention</td>
<td>Conventions in addition to standards.</td>
<td>Altered</td>
</tr>
<tr>
<td>number-source</td>
<td>none, MQM offers number related issues in the 'mistranslation' or 'locale-convention' parents which are not appropriate here.</td>
<td>Incorrect number in the source text. Introduced</td>
<td></td>
</tr>
<tr>
<td>omission-unclear</td>
<td>Accuracy</td>
<td>Ambiguous, potentially related to omission.</td>
<td>Introduced</td>
</tr>
<tr>
<td>other-note</td>
<td>none, the labels offered in MQM don't include process related issues on this level.</td>
<td>Translator notes left in document.</td>
<td>Introduced</td>
</tr>
<tr>
<td>other-technical</td>
<td>Internationalization</td>
<td>Technical issues with document.</td>
<td>Introduced</td>
</tr>
<tr>
<td>overly-literal</td>
<td>Accuracy</td>
<td>Refers to mistranslation. Otherwise, unclear-literal or stylistics-literal</td>
<td>Altered</td>
</tr>
<tr>
<td>stylistics-comprehension</td>
<td>Style, as the issue is not severe enough to be categorised as 'Unintelligible' under 'Fluency'.</td>
<td>Indicates stylistics issue linked to how easily the text can be understood.</td>
<td>Introduced</td>
</tr>
<tr>
<td>stylistics-literal</td>
<td>Style</td>
<td>Refers to text being literal but without indicating mistranslation.</td>
<td>Introduced</td>
</tr>
<tr>
<td>stylistics-readability</td>
<td>Style</td>
<td>Indicates stylistics issue linked to how easily the text can be read.</td>
<td>Introduced</td>
</tr>
<tr>
<td>terminology-company</td>
<td>Terminology</td>
<td>Assumed organisation specific terminology without indication of guidelines provided.</td>
<td>Altered</td>
</tr>
<tr>
<td>unclear</td>
<td>none/other</td>
<td>Highly ambiguous remark</td>
<td>Introduced</td>
</tr>
</tbody>
</table>

Table 7: MQM label deviations
6.4. Questionnaires

As outlined in section 5.3.1.3., this research included 6 questionnaires which can be grouped into two main questionnaires with the following aims:

- To gather information on users registered on Trommons.org, their opinions on quality and motivations to volunteer (Questionnaires 1 and 2).

- To gather demographic information to supplement the experiments conducted for this research, and to get insights into attitudes towards and opinions on crowdsourcing, quality and professional status in translation and localisation (Questionnaires 3 through 6).

The following sections outline the design, participants and data analysis in more detail. A full overview of questions included in the questionnaires can be found in appendix E.

6.4.1. Questionnaires 1 and 2: Trommons.org

Questionnaires 1 and 2 were conducted in collaboration with staff at TRF as well as other researchers at the Localisation Research Centre at the University of Limerick. The following section focuses on the parts of the questionnaire that were relevant to this research. The two questionnaires differed slightly in that Questionnaire 1 was aimed at volunteers at Trommons.org, while Questionnaire 2 had organisations registered at Trommons.org as its target population.

The goal of the questionnaires was to get first insights into the views on quality of the Trommons volunteers and organisations, as well as demographic information and information on their motivations for using Trommons.org.

6.4.1.1. Questionnaire design

For this questionnaire, questions were generated based on initial review of the literature, but were mostly exploratory and open-ended.
The questionnaire targeting the volunteers included:

- demographic questions (e.g. age and gender),
- a section on the interaction with Trommons and TRF (e.g. "When did you register on Trommons?")
- motivation for volunteering and patterns of working on tasks on Trommons.org (e.g. "Have you ever claimed a task on Trommons?")
- a section on quality (e.g. "Which of the following criteria are the most important for you when it comes to translation quality?")
- a section on tools and technology (e.g. "Would a mobile version of a translation tool be useful? Please explain your answer.")
- and finally attitudes towards Trommons (e.g. "How likely are you to recommend Trommons to others as a place to volunteer?")

The organisation questionnaire covered:

- questions on the organisation (e.g. "Number of staff")
- a section on the interaction with Trommons and TRF (e.g. "Is your organisation following Trommons or TRF on any of the following social media channels?")
- a section on quality (e.g. "Have you encountered any problems in the past when it comes to quality on Trommons?")
- and attitudes towards Trommons (e.g. "How likely are you to recommend Trommons to other organisations as a place to have their projects translated?")

However, only the sections on motivation, demographics and quality were included in this thesis since the other sections were only marginally relevant. Primarily, the results were included in section 4.1.4.

Whenever applicable, care was taken to present the same questions to volunteers
and organisations, with slight wording adjustments. Both questionnaires were conducted using web-based questionnaire forms (Surveymonkey) and participants were invited via e-mail with one follow-up reminder e-mail being sent.

Participation was voluntary, as was answering any of the questions included in the survey.

6.4.1.2. Sampling

Questionnaire 1 was sent to approximately 6000 volunteers, questionnaire 2 to 150 representatives of the organisations on Trommons.org. The invitations were sent by TRF in September 2014.

To encourage the highest absolute number of responses possible, a census approach was taken without further sampling. Every user registered on Trommons.org at that time received one of the two questionnaires.

A total of 517 volunteer translators and 39 organisations participated.

6.4.2. Questionnaire 3: Students and Lecturers

Questionnaire 3 prefaced Experiment 1 (Spanish linguists, Professional status) with the goal of gathering demographic information on the experiment participants as well as their views on quality and professional status of translators.

6.4.2.1. Questionnaire design

Participants were asked about their professional status, educational background and professional experience, and their native language in the demographic section of the questionnaire. They were then asked to define translation quality in their own words, followed by two questions on the connection between professional status of a translator and quality, as well as how to define what a professional status is. This questionnaire was intentionally kept short since the accompanying experiment was relatively time intensive.
6.4.2.2. **Sampling**

Since this questionnaire was aimed at supplementing the data gathered with Experiment 1, the approach to sampling and the distribution of the participants will be discussed in more detail in section 6.5.2.2.

All participants in the experiment also filled in this questionnaire, the sample consisted of 34 undergraduate students (Table 8), one postgraduate student and 17 lecturers.

<table>
<thead>
<tr>
<th>Year</th>
<th>Invited</th>
<th>Participated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

*Table 8: Overview of undergraduate student participants*

6.4.3. **Questionnaires 4 and 5: Localisation Industry Questionnaire**

Questionnaires 4 and 5, which were administered with a common start and branched off into separate sets of questions depending on the responses to those initial questions, had two primary goals:

1. To identify potential participants for experiments;
2. To gain information on the topics of crowdsourcing in translation, quality perceptions and quality management.

Initially, all practitioners, students and academics in translation and localisation - paid-for or volunteers - except for those groups invited to fill in other questionnaires for this study, were invited to participate in the study. Based on the primary role they identified with as well as language proficiency, they were then directed towards different branches of the questionnaire and potentially one of the three connected experiments. However, only questionnaires 4 (aimed at linguists of all languages) and 5 (aimed at industry buyers) were included in this research to avoid overly complicating the interpretation of data.
Questions were largely the same for all branches and mainly differed in wording, with a few additional questions being asked of some groups. The following sections describe the design of the questionnaires as well as the approach to sampling, data gathering and analysis in more detail. Detailed descriptions on individual questions have been included for the individual data analysis and results sections in chapter 7.

6.4.3.1. Questionnaire Design

Questionnaires 4 and 5 consisted of an initial, shared demographic section, followed by sections on quality in localisation, localisation crowdsourcing and the professional status of translators.

The demographic section included questions on the education and professional background of the participants, including the question "Which role have you primarily filled in the localisation field?". Possible answers were either:

- linguistic roles (reviewer/translator),
- buyer,
- technical roles,
- academic researcher or
- student in the localisation or translation area.

Buyer, reviewer, translator and technical roles were also offered a choice between paid-for and volunteer as their primary role. In addition, participants could provide their own description of the role if none of those fit or indicate that they had no ties to localisation or translation. Definitions of ‘buyer’ and ‘reviewer’ were also provided:

By buyer we mean the person who decides on or recommends to decision makers whether a translation service will be requested and later the delivery accepted.

By reviewer we mean the person who assesses localisation quality and makes recommendations regarding improvements and potentially the purchasing decision.
6.4.3.2. Branching

Based on the answer selected, participants were then later asked to participate in an experiment and asked to answer further questions:

- Participants indicating that they had no work or study related ties to localisation or translation were disqualified.
- Participants selecting one of the ‘reviewer’ or ‘translator’ options were asked to fill in questionnaire 4 targeted at linguists (volunteer and wider translation/localisation industry). In addition, they were asked whether they spoke German or Spanish, and if they selected either of those languages, to what level they spoke the languages. Reviewers or translators who spoke neither of those languages or who indicated they were beginners or intermediate learners were not offered an experiment. Reviewers or translators who indicated they spoke German or Spanish as a native language or considered themselves proficient were asked to participate in the German or Spanish linguist experiment respectively. These were
  - Experiment 1 (quality perceptions linked to professional status, Spanish - limited scope) and
  - Experiment 2 (quality perceptions linked to cost of translation, German)
- Participants selecting one of the ‘buyer’ options were directed to the buyer questionnaire (questionnaire 5) and asked to participate in Experiment 3 (quality perceptions linked to cost of translations, buyers)
- Participants selecting technical work, academic work, student status and ‘Other’ were asked to answer questions similar to questionnaires 4 and 5, but were not offered the chance to participate in an experiment. Also, the responses were excluded from this thesis and will be included in future analyses.

After branching, all participants that had not been disqualified were asked a similar
set of questions, with slight variations on wording to match their professional background and experience.

The questions broadly covered the topics:

- Attitudes towards crowdsourced translation
- Theories on why people volunteer for translation, as well as theories on why some translators work for lower rates than others
- Definitions and aspects of translation and localisation quality
- Factors that impact quality with a section on a wider range of potential factors, followed by more in-depth questions focusing on payment, experience, training and professional status

6.4.3.3. Question design

While none of the questions asked were knowledge questions, some were more aimed towards affective or evaluative attitudes while others covered cognitive attitudes or the respondents' willingness to do something with regards to an object of attitude (Sudman and Bradburn 1982). Whenever appropriate, these differences were signalled by using phrases like "What do you think" or "In your opinion", and the attitude objects were defined, for example by describing the type of text respondents should assume for a set of questions or by defining concepts like translation, localisation and crowdsourcing. In addition, some attitudes were measured by asking a series of ranking questions or by asking respondents to select a limited number of options from a longer list of options.

Question formats included open-ended questions, which allowed participants to freely share their thoughts on a question, as well as more closed questions like multiple choice, ranking and matrix questions. Participants were offered to additionally provide their own thoughts in separate comment questions.

For some questions, participants were first asked to share their thoughts or definitions in an open-ended format, before being asked to answer a similar question in a more closed format. For example, the questionnaire first asked how participants would, in their own words, define translation or localisation quality. This question was then followed
up with a question where participants were asked to rank different criteria according to their impact on quality. Conversely, participants were first asked to give their opinion on whether a translator being a ‘professional’ was a deciding factor in the quality of a translation, with the follow-up question asking participants to describe in their own words what defines a translation or localisation professional. In all instances, such questions were placed on separate pages and participants were not able to go back to the previous page to alter their response after they had progressed to the follow-up question. The goal of this approach was to first elicit an impulsive response to a topic without unduly influencing the participant, and on the second step get more in-depth insights and provoke further thinking about the topic. Participants were then able to provide their additional thoughts, if they had any, in the open-end comment section. Through this approach, the researcher was able to gather a mix of data that provided both, the participants' initial thoughts with minimal researcher influence as well as their insights when being prompted with several options to choose from.

However, this approach of asking multiple questions about the same topic and follow-up patterns can also lead to fatigue, which could explain the relatively low completion rate for the questionnaires (approx. 54%, depending on respondent group). This potential of causing fatigue was anticipated but deemed acceptable during the design stage of the questionnaire since the researcher considered it preferable to get high-quality responses from fewer individuals over a larger number of responses, especially in difficult to measure attitude and opinion areas. In addition, due to the sampling approach taken, generalisability of the results could not be claimed regardless of the number of completed questionnaires, which also supported the decision to prioritise depth of responses over quantity.

Nonetheless, care was taken to improve completion rates as far as possible by utilising multiple techniques to lower fatigue and make it easier for respondents to follow the questionnaire. To ease respondents into the questionnaire, they were first asked easy and non-threatening demographic questions that required no written input from them. Question types were varied to keep participants interested, yet, as much as possible, topics were completed before a new topic was started. To avoid redundant questions for the
individual participants, branching and filtering was done automatically with redundant questions not being made visible to the individual.

Finally, extensive testing and piloting helped not only with ensuring as far as possible that questions and definitions were worded clearly, but also that the question format was unambiguous and did not cause confusion or frustration with respondents. Initially, a paper-based version was reviewed by the Statistical Consulting Unit at the University of Limerick, as well as a marketing researcher and finally three members of staff/steering committee at TRF. This version was then placed in a web-based form (Surveymonkey) and first tested by the researcher and then by a second tester. Finally, it was piloted by 5 initial participants remotely, who provided written feedback based on their impressions and one participant in person, who provided feedback in a think-aloud format while the researcher was taking notes. Based on these observations, final changes were made to make the questions more accessible and refine content, as well as with the aim of increasing completion rates by removing or altering sections that were confusing or too difficult to complete without causing frustration in participants.

6.4.3.4. Sampling

The goal of this set of questionnaires was to get insights on the research topic from a wide and varied group of participants in the localisation field, from different professions and backgrounds, rather than focusing on a specific group of people. By taking this exploratory approach, initial insights into factors that impact the opinion of individuals could be gained where currently little is known about those factors. These could include a range of aspects such as their:

- background in translation/localisation and crowdsourcing,
- training,
- professional experience or their professional connections and
- individual study of opinion-forming material like blogs, presentations, articles or books.

It is especially necessary to consider the possibility that any organisation or
platform potentially used for sampling would have previously influenced the opinions of participants through the dominant opinions expressed within the community or the official position of the organisation. As an example, if the researcher had contacted the American Translators Association (ATA)\(^{40}\) and asked them for contact details of individuals, which would have been difficult to achieve due to privacy and ethical considerations, the press release that was describing translation crowdsourcing in a negative way, released by the ATA, would have been a confounding factor. This would have made that particular group not representative enough of the wider industry since the opinions of the population would likely be influenced by the official stance of the ATA. This implies that to get a wider perspective on attitudes towards crowdsourcing, it is necessary to survey a wider cross-section of the industry than individual bodies or platforms can provide. However, when referring to ‘the industry’, this really refers to a diverse, distributed and large set of individuals, communities, companies and organisations where it is difficult to estimate the size of the overall population and impossible to fully and accurately catalogue all members of this population, together with contact information.

As a result, the researcher considered it not suitable to utilise probability-based sampling strategies for the broader localisation industry since this approach would have involved a limited number of potential participants, where a probability-based sample would have been determined and invited to participate (Ackland 2013, p.27). Instead, convenience sampling in combination with snowball sampling (Ackland 2013, p.27) was employed by setting up the web-based form utilised for the questionnaires and experiments to be unrestricted and open to the public for anyone to participate in (Ackland 2013, p.28). In practical terms, this meant that the link to the survey was sent to a wide range of contacts within the researcher's personal and professional network (e-mail and private chat conversations) as well as several social media platforms (Facebook groups, LinkedIn groups, Twitter, proz.com forum). In addition, it was advertised via the Trommons.org mailing lists and newsletter, with the request that the call for participation would be further shared through any networks of those who had seen this call for participation. In the demographic section of the questionnaire participants were then asked to indicate what their

\(^{40}\) https://www.atanet.org/
connection to localisation and translation was, and if participants indicated that they did not have a connection to localisation or translation, they were excluded from the survey. While this approach does not allow for statistical inferences about the larger population (Ackland 2013, p.27) or statements about response rates, it does provide an opportunity to get a wide range of perspectives as a foundation to build on with statistically more relevant approaches for future enquiries.

6.4.4. Questionnaire 6: Non-profit Questionnaire

Questionnaire 6, which was administered separately to the representatives of organisations registered on Trommons.org had similar goals to those outlined for questionnaire 5 and was also linked to experiment 3 (quality perceptions linked to cost of translations, buyers).

- One, it was aimed at confirming that the respondent was within the target group for experiment 3, and
- two, the goal was to gain information on crowdsourcing in translation, quality perceptions and quality management from the perspective of non-expert buyers working with crowdsourced translations and volunteer translators.

6.4.4.1. Questionnaire design

However, based on the profile of respondents and feedback received from staff at TRF the researcher saw it necessary to simplify and shorten the questionnaire considerably. Simplification was needed because the potential respondents were non-expert buyers, meaning they were experts in other fields but their core tasks and expertise were outside localisation and translation. As a result, terminology was additionally explained and jargon replaced with simpler words. For example, the term ‘localisation’ was omitted entirely and instead only the word ‘translation’ was used since translation was considered a concept the respondents would be more familiar with and since the distinction was not of great relevance to the questions. In some cases, respondents were offered alternate questions that
allowed them to provide their thoughts in a more open format to avoid the risk of them having to choose from a list of options where none applied or where they didn't fully understand all options.

In addition, since localisation and translation were not the core task, and were not likely to be of central interest to the group of individuals invited to participate, the decision was also made to lower the number of questions asked. This was also necessary because of the relatively small group of potential participants that were invited.

While this simplified and shortened questionnaire was as a result not as informative and rigorous in its approach as the larger questionnaires 4, 5 and 6, the researcher considered this an acceptable trade-off to encourage participation and increased completion rates from a limited group of informants with specific needs.

6.4.4.2. Sampling

For this questionnaire, the researcher was provided with a list of representatives of organisations registered on Trommons.org and all representatives were invited to participate in the study. While as a result it is not possible to draw generalisable conclusions from the results, this approach was nonetheless deemed the most appropriate approach.

Since the researcher had been provided with a list of individuals to contact, it would have been possible to draw a probabilistic sample from this list. However, as outlined in section 6.5.4.2., the group of individuals to be invited was not only small but also didn't have localisation and translation as their core interest and professional specialisation. Therefore, the expectation was that response rate would be very low and the decision made to prioritize receiving the largest number of responses to get as much in-depth information as possible over potential generalisability.

With this in mind, 430 representatives were invited to participate in the study via e-mail. 89 opened the invitation e-mail, and 17 respondents completed the full questionnaire, 8 the experiment.
6.5. Experiments

As outlined in section 5.3.2., three experiments were conducted (Table 9).

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Topic</th>
<th>Language</th>
<th>Target audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Professional status</td>
<td>Spanish</td>
<td>University of Madrid; Linguists</td>
</tr>
<tr>
<td>2</td>
<td>Price of translations</td>
<td>German</td>
<td>Linguists</td>
</tr>
<tr>
<td>3</td>
<td>Price of translations</td>
<td>none</td>
<td>Buyers (reviewer comments)</td>
</tr>
</tbody>
</table>

Table 9: Experiment overview

All three experiments were based on a shared design. Participant preferences for several variations of the same translation or reviewer comments were measured in a control condition without information provided on the cost of the translation or professional background of the translator. The experiment conditions then introduced correct or incorrect information on cost and professional background, depending on the specific experiment.

Experiment 1 (Spanish linguists, professional status) was designed and conducted in collaboration with Dr. Bermudez Bausela at the University Alfonso X el Sabio in Madrid. However, analysis and interpretation was then conducted by the researcher independently, with a view to preparing collaborative analysis and interpretation for future publication.

The following sections outline the preparation steps as well as detailed design and sampling approach for each of the experiments.

6.5.1. Experiment Preparation

Depending on the experiment, participants were presented with one of the following:
Experiment 1: English text and Spanish translations
Experiment 2: English text and German translations
Experiment 3: Reviewer comments generated based on the English to German translations for experiment 2

The following sections outline the text selected as well as how translations and reviewer comments were generated.

6.5.1.1. Text Selection

Prior to selecting a text for the experiments, several criteria had been defined that this text needed to satisfy:

- The text consists of short sections that would lend themselves to being placed into the experiment without requiring further segmentation. To not take up more of the participants' time than necessary and since there was no additional benefit to be expected from longer passages of text, each segment should have approximately one or two sentences. At the same time, a text that was already segmented had the benefit of avoiding issues with missing references, context or consistency which might have been the case if a single long passage had been split into shorter segments.

- The text is not overly simple, while at the same time not being highly specialised. This was deemed important since an overly simple text would not create enough controversy in the way it should be translated. On the other hand, a text from a highly-specialised field would favour some translators over others and would create an unrealistic challenge for novice translators, thereby potentially impacting the results if those novice translators would make obvious and frequent mistakes due to a high difficulty level.
• There is no translation already available for this text. This means that translators or participants couldn't search for an existing translation, assume it to be the ideal translation and modify their translation or evaluation accordingly.

Initially, several potential texts that had previously been translated on Trommons were evaluated as well as texts that were publicly available. However, none of the texts considered satisfied all the criteria outlined. Therefore, the decision was taken to create a text specifically for the experiments.

This medium-length piece of text created by the researcher was named ‘Succeeding in the MSc’ and consisted of 13 paragraphs with individual headings, where each paragraph was approximately two to three sentences long. The overall text length was 1145 words, with the shortest paragraph consisting of 49 words and the longest of 125. Headings were between two and 13 words long. The first paragraph served as an introduction, the following 12 paragraphs were a series of suggestions for students of the online programme MSc Multilingual Computing and Localisation\(^{41}\) at the University of Limerick. The full text used as foundation for the experiments can be found in appendix B.

Creating a text for this experiment could be considered a threat to ecological validity since the creation of the text might be influenced by the knowledge that it will be used for an experiment, introducing a degree of artificiality. However, using a text that was created for a different purpose limits the amount of control possible over the text. Not only would it have meant having to find a text that satisfied all other criteria, it also had to be possible to gain permission from the owner of the text while ensuring that they wouldn't in some way make it known that the text was being used for research purposes. Therefore, creating a text was considered an acceptable trade-off since it meant that other variables could be controlled directly. The text was also edited by a native speaker of English with a post-graduate degree in linguistics and teaching experience in English at secondary school level to minimise linguistic issues.

\(^{41}\) http://www.localisation.ie/education/
6.5.1.2. Text Translation

After the English text had been written and edited, translations were requested from German volunteer translators on Trommons.org, German paid-for translators on ProZ.com, Spanish translation students and Spanish translation lecturers. The translations and associated information presented can be found in appendix F.

6.5.1.2.1. Spanish Translator Sourcing

Six translations from English to Spanish were sourced from students and members of staff at the University Alfonso X el Sabio in Madrid. Translators were selected by collaborating researcher Dr Bermudez Bausela based on her knowledge of their performance in the case of students, and training and experience in the case of lecturers.

Translators belonged to the following three groups: Novice translators, Trained translators and Trained + Experienced translators (referred to as ‘Expert’ for brevity).

1. **Novice Translators** were first-year students in the degree of Translation and Interpreting. In the first year of their studies, students would not yet have received extensive training in translation. The only subject related to translation training as part of the first-year curriculum teaches English for future translation trainees. From this group, two students were selected in order to be able to compare results. One student could be described as highly-motivated with average grades, while the second student is a highly-motivated student with very good marks in general, and, in particular, in the above-mentioned subject. Both the students chosen from this first group were Spanish native speakers from Spain.

2. **Trained Translators** were fourth-year students in the degree of Translation and Interpreting. In the fourth year, students would have received in-depth training in translation and covered a range of relevant subjects but would not have gained a lot of real-life translation experience.

Again, two students were selected. Both students could be described as highly-motivated with one of them being an average-student regarding marks but having
exceedingly improved along the four years of the degree, and eager to get involved in any collaborative work involving translation. The other student had achieved very high marks in the degree. Both the students chosen from this first group were Spanish native speakers, one from Spain and the other one from Venezuela.

3. **Expert translators** were both lecturers at the University Alfonso X el Sabio in Madrid. Two lecturers were selected, one a lecturer of English with no translation background but more than 25 years of experience teaching English for General Purposes and English for Specific Purposes at tertiary level. The second lecturer is an experienced Translation teacher with over 15 years of experience in the field, working with English, French and Spanish. Both the lecturers were Spanish native speakers from Spain.

The 6 translators in this preparation step embraced the project with enthusiasm and were eager to participate by translating the English source text provided. However, the second expert translator did not provide translations for the last 8 segments requested due to time constraints. Since it was not possible to invent a cover story that would have prompted both, students and lecturers to translate the same text, all translators were made aware of the fact that their translations would be used for an experiment but not what the exact nature of the experiment was. They were also told that the reason for choosing multiple translators was to have multiple versions of the same text that could then be compared in the experiment results. However, none of the translators were given access to any of the other translations.

The translations were then collected and reviewed by Dr Bermudez Bausela to make sure that there was a minimum quality required present so that they were valid for the experiment.

**6.5.1.2.2. German Translator Sourcing: Paid-for**

Three translations were sourced on ProZ.com, through the translation job and bidding system on the platform. The request for quotes included the information that the
text contained success tips for the MSc Multilingual Computing and Localisation at the University of Limerick, that it was approximately two pages long, that a translation from English to German was required, but no Desktop Publishing was necessary. Those bidding for the job were also advised that they would need to fill in a form for the University of Limerick to be able to receive payment. Finally, a link to the MSc programme (www.localisation.ie/education) and a brief sample of the English text were provided, but without requesting sample translations.

33 translators submitted a bid, of which one was a representative of a localisation service provider rather than an individual translator. Several bids were made in currencies other than Euro, while others were submitted at a flat fee for the entire project rather than a per-word-rate. However, most bids were submitted at a per-word-rate (source words) and in Euro. To ease selection, non-Euro bids were converted to Euro using the www.xe.com and flat fees were broken down into per-word-rates.

The resulting list then went through a series of elimination steps to determine who should be contacted to provide the translations.

1. 5 bids were either without a rate or without a currency attached to the bid. These were excluded in the first step.
2. All translator profiles were checked to see whether they claimed to be native German speakers or whether they claimed more than one native language. To avoid native speaker status as a confounding factor, non-native speakers or those with multiple native languages were excluded. On this step, 7 bids were excluded from the list of potential translators.
3. From the remaining list, the most expensive, cheapest and median cost translations were identified.

The distribution was:

- two bids at 0.15 Euro (highest);
- one bid at 0.045 Euro and one at 0.05 Euro (lowest);
- three bids at 0.10 Euro (median).
Since there had been two highest bids, the two lowest were included in the next step.

4. For the next selection step the translators’ profiles on ProZ.com were evaluated to determine their level of experience and training in the translation or localisation field. Based on the highest level of training and experience, one translator from the highest, lowest and median group each were selected and the remaining four were informed that they might be approached at a later point to complete the translation.

5. The three selected translators received the full document and were asked to fill in a form to be set up in the University of Limerick’s financial system to receive payment after project completion. The translators received no additional information than what had been provided already in the ProZ.com job descriptions. None of the translators were informed that their work would be used for a research project; this was explained after they had sent the translation and permission was obtained to use those translations in this way. The translators were also not asked to adhere to a specific deadline, although interestingly all three of them tried to offer short deadlines on multiple occasions, even before they were set up in the financial system and had been given the go-ahead for the project. Once they received the go-ahead, they were told that the translation was not urgent and were asked to propose a reasonable deadline.

The decision to include the median price of 0.10 Euro in the experiment and expand the options to four was made to avoid bias due to the fact that 0.05 Euro and 0.15 represent extreme ends of cost per source word on the translation market (Gardner 1971, p.241). In fact, from the translations quoted, the 0.10 Euro per source word could still be considered a ‘high cost’ translation. Figure 11 shows the rates for all quotes received as far as rates had been indicated. Orange bars mark the translators that were selected for the experiment translations.
This means that a realistic ‘medium cost’ translation would have been closer to 0.08 or 0.09 Euro per source word based on these quotes. However, by opting to select the median price point between the two extremes at 0.10 Euro per source word, the researcher aimed to clearly signal to participants that this was the ‘medium cost’ translation. This decision was also supported by several considerations: Participants for Experiment 2 were linguists who would have a general idea about price points on the translation market based on their own decisions on what rates to quote for, but who would not likely have a full, detailed overview, allowing them to understand where exactly ‘medium’ was located. While buyers in Experiment 3 might have a better understanding of this, their ideas of what should be considered high, medium or low cost would also be influenced by the field they work in and the rates they are used to paying translators on a daily basis.

Therefore, the decision was made to clearly signal low, medium and high cost by using the median price point between the two extreme price points.

The remaining translators were informed that they had not been selected for the translation, except for the ones short-listed who were informed that they might be
approached for translations if the ones selected didn't end up delivering. In addition, translators shortlisted were later contacted with a request for them to work on the review task (Section 6.5.1.3.).

6.5.1.2.3. German Translator Sourcing: Volunteers

Two translations were requested via Trommons.org. Interviews with staff at TRF had shown that translators on Trommons were partially selected among the translators who were well-known to the staff and approached directly with translation requests, while others would simply pick up a translation task that appeared on the task stream on Trommons.org. Therefore, a similar approach was taken to solicit the volunteer translations. The organisation ‘Localisation Research Centre’ was created on Trommons.org, and this organisation posted the task request with the combination English UK to German Germany. The task contained the same description as the one posted on ProZ.com, and included the reference link to http://www.localisation.ie/education/

A volunteer translator claimed this task within a day and uploaded the translation after 12 days.

A second translation was requested from a volunteer translator directly by staff at TRF. This translation was uploaded after 17 days.

6.5.1.3. Text Review

Once all translations into German had been collected, they were split into segments of one to two sentences, following the natural segmentation of the source text with headers for each paragraph in its own segment. Translations were fully anonymised prior to the review step and each reviewer was only given instructions on how to conduct the review and the translations as well as the English source text. While experiment 2 (Cost of Translation, German) used the translations originally provided without review, reviewer comments and ratings gathered based on these translations were used for experiment 3 (Cost of Translation, Buyers).

In addition to a small set of questions on their professional background and
experience, reviewers were provided with the information that their review would be used for an experiment but would remain anonymous. They were then asked to read the English source text and all five translations, then rank each of the translations from one to 5, with 5 being the best rating for a translation. Further, reviewers were asked to comment on each translation for each segment, with the comment being written for an imaginary translation client who doesn't speak German. Finally, reviewers were also given the option to only evaluate a limited number of lines since evaluating all lines would have been very time intensive and the expectation was that not all the reviewer comments for all lines would be used in the experiments. None of the reviewers were given information on the professional background of the translator or the cost of the translations.

Reviewers with different backgrounds were invited to participate, some paid and some unpaid (Table 10). Unpaid reviewers were sourced from the researcher's personal network, while paid reviewers were selected from the original group of translators who had bid on conducting the English to German translations, but who had not been selected for the translation task.

<table>
<thead>
<tr>
<th>Reviewer No.</th>
<th>Paid?</th>
<th>Cost of review</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>N/a</td>
<td>Experience with volunteer translation, little professional translation experience</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>N/a</td>
<td>8+ years of experience in localisation industry (quality assurance and localisation management)</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>80 Euro</td>
<td>8+ years of experience in localisation industry (translation)</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>59.50 Euro</td>
<td>1 year of experience in localisation industry (translation)</td>
</tr>
</tbody>
</table>

Table 10: Overview of reviewers

None of the reviewers filled in all lines, and the comments varied for each reviewer and each individual line in level of detail, examples provided and number of issues observed.

6.5.2. Experiment 1: Professional Status, Spanish

Experiment 1 (Spanish linguists, professional status) was designed and conducted in
collaboration with Dr Montserrat Bermudez Bausela at the University Alfonso X el Sabio in Madrid, utilising Spanish translations and with translation students and lecturers as target population. The collaboration included all steps of the experiment, from initial conception and design of the study, via preparation steps and setting up of the experiment, to data gathering, pre-analysis preparation and initial analysis. Final analysis, discussion and write-up were conducted solely by the author of this thesis. A limitation was that the data gathering step was undertaken by Dr Bermudez Bausela, without the researcher being present. However, the researcher was able to discuss the approach to the gathering of data, as well as observe its progress through the Surveymonkey interface and was therefore closely involved also in this step, even if not physically present.

The goal of the experiment was to gain insights into the question of whether or not knowledge of the professional status (novice translator, trained translator, trained + experienced 'expert' translator) would have an impact on the perceived quality of a translation. This was measured by asking participants to select their preferred translation out of three versions offered.

The following section provides a description of the experiment design as well as the approach to sampling, data collection and data analysis.

6.5.2.1. Experiment Design

For this experiment, the Spanish translations described in section 6.5.1.2.1. were used. Once the translations had been sourced, they were separated into individual headers and paragraphs. These were then assigned codes, one for each of the segments. This resulted in a total of 26 segments. Each of those 26 segments had been translated by two first year students (‘Novice’), two fourth year students (‘Trained’) and two translators who were both formally trained and had several years’ experience in translations (‘Expert’). While the participants were given the information that the third group of translators was ‘Trained + Experienced’, the following will refer to these translators as ‘Expert’ for brevity.

The overall design of the experiment consisted of a Pre-test condition and a test condition with two variations. The experiment was set up in a questionnaire form (Surveymonkey), with one page per segment. All of the pages showed three translations,
each provided by a different translator belonging to the Novice, Trained or Expert category.

6.5.2.1.1. Pre-test condition

Participants were provided with a segment of English text and three Spanish translations. No information on the source of the translation was provided (NoInfo). They were then asked to 1) select the translation they thought was best and 2) make any changes they thought necessary to make the translation what they considered ideal. The results of participant changes to the translation were excluded from this thesis and will be considered for future publication without being further discussed going forward. Figure 12 shows the layout of the screen and the instructions provided to participants.
The following questions will present you with English content that has been translated into Spanish by different translators. Assuming all other things being equal, please select the translation you think is the best. Then copy that translation into the “corrections” field - copy the entire translation! Finally, make any changes to this translation that you think are needed to make it a great translation.

Please note: unfortunately, the form deselected the chosen translation when you highlight it for copying, so you will have to click on it again.

* 8. Original text: Succeeding in the MSc
   - Alcanzar el éxito con MSc
   - Exito en el master
   - Cómo tener éxito en el máster

* 9. Please copy the best translation into this field and make any corrections necessary.

10. Would you like to comment?

---

**Figure 12: Example screen; Experiment 1 (no information provided)**

**6.5.2.1.2. Test condition**

Participants were then asked to again evaluate translations from English to Spanish. Each of the translations into Spanish was prefaced with information on the source of the translation, namely whether the translator had been a) a novice, b) a trained translator or c) an expert translator. No additional definitions of the meaning or definition of these labels were provided, just the labels themselves.

In this test condition participants were asked to select which translation was best, and the fact that information on the translator's professional status had been included was mentioned. However, no instruction was given on whether participants should consider only the translation or also the information on professional status when making their
selections.

In some scenarios participants were provided with the correct information on the professional status of the translator (GoodInfo) and in other scenarios they were given incorrect information (BadInfo) on this status. Figure 13 shows an example screen with the test condition, in this case a GoodInfo condition, where participants were provided with the correct information on the background of the translators.

Figure 13: Example screen; Experiment 1 (information provided)

On all pages, participants were provided with the optional opportunity to comment before submitting their choice.

To check whether participants were still actively engaged in the task, rather than selecting options without reading the source text and translations in detail, they were asked
on one of the experiment pages to select a specific option and move on to the next page. This test task was excluded from analysis but will be considered again in the discussion chapter (section 9.2.2).

In summary, this means that for experiment 1, changes in the dependent variable (the percentage of participants who choose what should be the highest quality translation based on the actual professional status of the translator) was measured in relation to the independent variable (information on professional status provided, which can be correct or incorrect).

### 6.5.2.1.3. Text selection

To display the translations of the 26 segments, translated by 6 translators to participants, they were placed in a grid and assigned to the individual experiment steps with the following considerations:

1. The distribution of headers and paragraphs for each category, bearing in mind the properties of headers and paragraphs. Paragraphs consisted of several sentences and therefore provided participants with more intrinsic attributes for evaluation. Headers offered the benefit that they were short. As a result, participants didn't have to read through several long translations to decide, thereby data gathering was faster for headers while more in-depth for paragraphs. However, this brevity also meant that the headers by themselves commonly did not provide enough information to understand the meaning of the source text or judge the quality of the translation. This allowed the researchers to observe variability in the results depending on the information a participant could gain from the translation itself.

2. Translations should be as equally as possible distributed between translators.

3. A suitable number of NoInfo, GoodInfo and BadInfo categories should be created. This meant that NoInfo occurred slightly less often than GoodInfo and BadInfo,
since NoInfo was only included in the pre-test condition and served as a baseline to compare the test condition results with, so the decision was made to focus on test conditions.

4. For each combination, one Novice, one Trained and one Expert sample was selected.

Lines were assigned to each variable using a random number generator and for each of the cases, translators were alternated. This resulted in the following split for initial allocations as shown in Table 11, with the allocation per translator as shown in Table 12.

<table>
<thead>
<tr>
<th>Scenario type</th>
<th>Header</th>
<th>Paragraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoInfo</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>GoodInfo</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>BadInfo</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 11: Initial distribution of headers and paragraphs per scenario type

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>ES-First1</th>
<th>ES-First2</th>
<th>ES-Fourth1</th>
<th>ES-Fourth2</th>
<th>ES-Experienced1</th>
<th>ES-Experienced2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoInfo</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GoodInfo</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>BadInfo</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>16</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 12: Initial translator allocation per scenario

Since the translator ‘Experienced2’ didn’t complete the last 8 segments, those segments were chosen from the translations provided by translator ‘Experienced 1’. After all combinations had been assigned to translators, the final set of translated segments was selected from these combinations and placed in a questionnaire form. The final distribution of translations was as shown in Tables 13 and 14.
Table 13: Final distribution of headers and paragraphs per scenario type

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>Header</th>
<th>Paragraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Info</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Good Info</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bad Info</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 14: Final translator allocation per scenario

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>ES-First1</th>
<th>ES-First2</th>
<th>ES-Fourth1</th>
<th>ES-Fourth2</th>
<th>ES-Experienced1</th>
<th>ES-Experienced2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Info</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Good Info</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bad Info</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

In the initial set of experiment combinations, participants were presented with 13 pages of the experiment, each including one combination. After they had completed this initial set, they were given the opportunity to comment on the questionnaire as a whole.

6.5.2.2. Participant Sampling and Conducting the Experiment

For this experiment, four population groups were invited to participate:

1. Undergraduate students in the degree of Translation and Interpreting at the University Alfonso X el Sabio in Madrid;
2. Master students of Localisation and Project Management at the University Alfonso X el Sabio in Madrid;
3. Lecturers at the Faculty of Applied Languages of the University Alfonso X el Sabio in Madrid;
4. Spanish linguists from the wider localisation industry, open call.

The following section describes the participant groups in more detail.

1. Undergraduate students in the degree of Translation and Interpreting at the University Alfonso X el Sabio in Madrid

For this group, 1st to 4th year students studying for the degree of Translation and...
Interpreting were invited to participate. Participation was limited to students from this course since the expectation was that students from other courses would not necessarily be interested in translation as a topic and would possibly not have the same level of linguistic and cultural sensibility required for translation.

Participation was distributed as shown in Table 15.

<table>
<thead>
<tr>
<th>Year</th>
<th>Invited</th>
<th>Participated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4/Master's</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>

*Table 15: Student participant distribution*

Response rates varied between the years. Where only half of the second-year students decided to participate, all the students invited from years three and four also participated. While it took first, second and third year students approximately one hour and 15 minutes to fill in the questionnaire and complete the experiments, this time decreased significantly with fourth year students. A possible reason for this is that the students were in the last year of tuition and the further training might match their speed at reviewing the translations.

Completion of the optional, additional experiment section was low but all participants completed the main experiment.

All participants self-identified as Spanish native speakers, except for one first year student who was a German native speaker highly proficient in Spanish.

For this group of participants, the experiment was conducted in a classroom setting with Dr Bermudez Bausela present to answer participant questions.

2. Masters students of Localisation and Project Management at the University Alfonso X el Sabio in Madrid

This population group was invited to participate in the experiment via e-mail and during lectures. The experiment for this group was conducted outside the classroom to increase flexibility and convenience. However, out of 12 participants invited, only
3. Lecturers at the Faculty of Applied Languages of the University Alfonso X el Sabio in Madrid:

This population group was also invited to participate via e-mail and was given the opportunity to complete the experiment at their own convenience outside the classroom. Out of 32 participants invited, 17 responded and completed the experiment.

All lecturers invited were at the time of the experiment or previously members of staff at the Faculty of Applied Languages of the University Alfonso X el Sabio. They were involved in the following degrees and courses:

- Degree in Translation and Interpreting;
- Degree in Modern Languages and Management;
- Degree in International Relations;
- Expert course in Localisation, Project Management and Audio-visual Translation.

Some of the lecturers had extensive training and experience in Translation while others were experienced lecturers of English both as English for General Purposes and English for Specific Purposes. Lecturers specialised in International Relations also had specialist knowledge of linguistics. Participants from the group of lecturers took on average between one and 1.5 hours. This is similar to the time it took undergraduate students in years one through three to complete the experiment, suggesting that the lecturers took great care in evaluating the translations and making changes which could have offset time gains due to their experience and training.

Participants from all groups commented on the difficulty of judging text without additional context, especially for some of the shorter segments.

4. Spanish linguists from the wider localisation industry, open call

Participation for this experiment was also invited from Spanish linguists in an open
call. Respondents who chose to fill in questionnaire 4 were given a final set of questions asking them about their language combinations and proficiency to determine whether they qualified for the experiment. However, participants were not told that their response to this particular question would qualify or disqualify them for an experiment.

The first question was "Do you speak Spanish or German? If you speak both, please select the language where you have a higher level of proficiency". The options provided were:

1 I speak German;
2 I speak Spanish;
3 I speak neither German nor Spanish.

Participants who selected Option 3 were directed to the end of questionnaire screen, those who selected Option 1 (German) were asked a follow-up question on German and participants who selected Option 2 (Spanish) were asked the same follow-up question on Spanish: "Please select your level of knowledge for Spanish":

1 I do not speak Spanish;
2 Beginner (A);
3 Intermediate (B);
4 Proficient (C);
5 Native speaker.

Option 1 was offered in case participants had accidentally selected the option that they spoke Spanish on the previous question but did in fact not speak Spanish. Participants who selected Options 2 (Beginner) and 3 (Intermediate) were directed to the end of the questionnaire screen. Participants who selected Options 4 (Proficient) and 5 (Native speaker) were invited to participate in the experiment.

By inviting only translators and reviewers rather than just any participants who spoke Spanish to a high level, only those participants who were likely to have the skills to
evaluate translations were included.

42 respondents completed the experiment. Further demographic information on the respondents will be provided in section 7.1.1.

6.5.3. Experiment 2: Cost of Translation, German

Experiment 2 (quality perceptions linked to cost of translation, German) followed the structure of Experiment 1 in many aspects. This experiment was conducted using translations of the text ‘Succeeding in the MSc’ as outlined in section 6.5.1. from English into German. The goal of this experiment was to observe whether knowledge of the cost of a translation (low cost, median cost, high or free) would have an impact on the perceived quality of a translation. Participants were asked to select their preferred translation out of four versions offered, and as with Experiment 1, they were either given no information (NoInfo) on the cost of the translation, the correct information (GoodInfo) or the incorrect information (BadInfo). This experiment was also conducted using web-based questionnaire technology (Surveymonkey) and was offered to participants after they had completed questionnaire 4. By combining the experiment with a questionnaire, the researcher could observe the relationship between the participant's choices and their professional and educational background, as well as expressed opinions on quality in localisation/translation and the cost-quality link.

The following sections outline the design of the experiment in more detail, as well as the sampling approach.

6.5.3.1. Experiment Design

The design of this experiment followed the design of Experiment 1 (quality perceptions linked to professional status, Spanish), using the same pattern for selection of texts. German translations that had been sourced as described in section 6.5.1. were split into 26 segments consisting of headers and paragraphs. Each segment had been translated by five different translators and was assigned labels for the purpose of managing data and
to present the information to experiment participants as shown in Table 16.

<table>
<thead>
<tr>
<th>Data label</th>
<th>Presented in experiment</th>
<th>Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>LowCost</td>
<td>0.05 Euro per word</td>
<td>Paid translation, proz.com, 0.05 Euro per source word</td>
</tr>
<tr>
<td>MedCost</td>
<td>0.10 Euro per word</td>
<td>Paid translation, proz.com, 0.10 Euro per source word</td>
</tr>
<tr>
<td>HighCost</td>
<td>0.15 Euro per word</td>
<td>Paid translation, proz.com, 0.15 Euro per source word</td>
</tr>
<tr>
<td>VolApproached</td>
<td>Unpaid volunteer</td>
<td>Volunteer translator selected by TRF staff</td>
</tr>
<tr>
<td>VolSelfselect</td>
<td>Unpaid volunteer</td>
<td>Volunteer translator, via Trommons.org interface</td>
</tr>
</tbody>
</table>

Table 16: Translator price point overview

In the experiment, no distinction was made for participants whether a volunteer had been approached or whether they had claimed the translation via the Trommons.org interface.

As with Experiment 1, translations were placed into a questionnaire form (SurveyMonkey), although for Experiment 2 four translations were shown on each of the screens as discussed in section 6.5.1.

In the pre-test condition participants were shown a segment of English text and the four German translations, with no information on the source of the translation provided (NoInfo). They were then asked to select the translation they thought was best. Figure 14 shows the layout of the screen and the instructions provided to participants.
In the test condition, participants were also shown a segment of English text with four German translations, but each translation included information on the cost of the translation. No additional information on the background of the translators or the process of obtaining the translations was provided. Participants were asked to base their decision on the quality of the text itself and were made aware in the instructions that information on the cost of the translation had been included.

As with Experiment 1, some scenarios showed the correct information on the cost of the translation (GoodInfo) and others showed incorrect information (BadInfo). Figure 15 shows an example where participants were given the incorrect information on the cost of the translation.
Participants were given the option to comment before submitting each choice.

As with Experiment 1, one scenario had instructions to select a specific option in the source text and move on to the next scenario to be able to estimate whether participants were still paying attention to the source text and translations at this point. Here, too, this scenario was excluded from the data analysis but will be included in the discussion chapter (Section 9.2.2.).

The distribution between NoInfo, GoodInfo and BadInfo mirrored the distribution which had been generated for Experiment 1 exactly, resulting in the same headers and paragraphs of the source text being assigned the respective labels.

The allocation per translator is shown in Table 17.

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>LowCost</th>
<th>MedCost</th>
<th>HighCost</th>
<th>VolApproached</th>
<th>VolSelfselect</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoInfo</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GoodInfo</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>BadInfo</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 17: Initial price point allocation per scenario
After translated segments had been placed in the questionnaire form, the final distribution of translations was as shown in Table 18.

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>LowCost</th>
<th>MedCost</th>
<th>HighCost</th>
<th>VolApproached</th>
<th>VolSelfselect</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoInfo</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GoodInfo</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>BadInfo</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 18: Final price point allocation per scenario

6.5.3.2. Participant Sampling and Conducting the Experiment

Initially, respondents to the overall call for participation were directed into different branches of this main questionnaire, with linguists being directed to Questionnaire 4.

Participant selection mirrored the process for the Spanish linguist experiment (Section 6.5.2.2.), with participants being selected based on their self-declared level of German proficiency, based on the same considerations.

25 respondents completed the experiment. Further demographic information on the respondents will be provided in section 7.1.2.

6.5.4. Experiment 3: Cost of Translation, Buyers

Experiment 3 (quality perceptions linked to cost of translation, buyers) followed the structure of Experiment 2 closely. However, rather than utilising translations, this experiment was conducted by presenting participants with reviewer comments that had been gathered based on the English to German translations of the text ‘Succeeding in the MSc’ as outlined in section 6.5.1. The goal of this experiment was to observe whether knowledge of the cost of a translation (low cost, median cost, high cost or free) would have an impact on the perceived quality of a translation when participants did not have the translations available to them, but only second-hand information in the form of reviewer comments.

Participants were asked to indicate how likely they were to trust each of the
translations based on the reviewer comments. As with Experiment 2, they were either given no information (NoInfo) on the cost of the translation, the correct information (GoodInfo) or the incorrect information (BadInfo). This experiment was also conducted using web-based questionnaire technology (SurveyMonkey) and was offered to participants after they had completed Questionnaire 5.

As with the other experiments, the information on the professional and educational background, as well as expressed opinions on quality in localisation/translation and the cost-quality link allowed the researcher to observe the relationship between these factors and the participant's choices.

The following sections outline the design of the experiment in more detail, as well as the participants and data analysis steps.

### 6.5.4.1. Experiment Design

The basic structure of Experiment 3 followed that of the experiments previously described. Participants were presented with scenarios that represented either the pre-test (NoInfo) or the test condition (GoodInfo/BadInfo), where each of the scenarios included four reviewer comments. For the test condition, each comment was labelled with information on the cost of the translation that mirrored the labels described for Experiment 2 (Table 19).

<table>
<thead>
<tr>
<th>Data label</th>
<th>Presented in experiment</th>
<th>Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>LowCost</td>
<td>0.05 Euro per word</td>
<td>Paid translation, proz.com, 0.05 Euro per source word</td>
</tr>
<tr>
<td>MedCost</td>
<td>0.10 Euro per word</td>
<td>Paid translation, proz.com, 0.10 Euro per source word</td>
</tr>
<tr>
<td>HighCost</td>
<td>0.15 Euro per word</td>
<td>Paid translation, proz.com, 0.15 Euro per source word</td>
</tr>
<tr>
<td>VolApproached</td>
<td>Unpaid volunteer</td>
<td>Volunteer translator selected by TRF staff</td>
</tr>
<tr>
<td>VolSelfselect</td>
<td>Unpaid volunteer</td>
<td>Volunteer translator, via Trommons.org interface</td>
</tr>
</tbody>
</table>

*Table 19: Translator price point overview*
Here also there was no distinction made for participants between the self-selected volunteer translator and the volunteer translator who had been approached by TRF staff, and the individual scenarios were placed into the SurveyMonkey questionnaire form.

In the pre-test condition participants were shown four reviewer comments, with no information on the source of the translation provided (NoInfo). They were then asked to select how likely they were to trust or select a translation based on each of the comments (Figure 16).

In the test condition, participants were also shown four reviewer comments, but each included information on the cost of the translation with no additional information on the background of the translator or the translation process. Participants were asked to choose how likely they were to trust or select each translation based on the reviewer comments and were made aware that there was now also information on the cost of the translation available.

Some scenarios showed the correct information on the cost of the translation (GoodInfo) and others showed incorrect information (BadInfo) (Figure 17).
The following questions will present you with reviewer comments that were provided for translation projects. All other things being equal, please select the translation you would be most likely to trust or select, based solely on those comments.

Please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different.

29. How likely are you to select each of the translations based on the review provided?

<table>
<thead>
<tr>
<th>Comment</th>
<th>Likely</th>
<th>Neutral</th>
<th>Unlikely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar mistake, stiff and literal translation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very literal and formal, sounds unnatural and not really getting the points across.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Sounds awkward and „Zugang” is more like having access to resources like computers etc. Addresses both (all) genders in a correct and neutral way“</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“- Does not sound like an offer, an opportunity („Verantwortung“)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- „Verantwortung für“ not „Verantwortung über“</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

30. Would you like to comment?

Figure 16: Sample screen; Experiment 3 (no information provided)
The following questions will present you with reviewer comments that were provided for translation projects, as well as information on cost of the translation per source word. All other things being equal, please select the translation you would be most likely to trust or select, based solely on those comments.

Please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different.

39. How likely are you to select each of the translations based on the review and cost information provided?

<table>
<thead>
<tr>
<th>Unpaid Volunteer: Sounds awkward and &quot;Zugang&quot; is more like having access to resources like computers etc. Addresses both (all) genders in a correct and neutral way</th>
<th>Likely</th>
<th>Neutral</th>
<th>Unlikely</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15 Euro per word: Grammar mistake, stiff and literal translation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05 Euro per word: &quot;Subscription&quot; does not refer to &quot;library&quot; as can be seen easily by the following paragraph: Change of sentence structure is an improvement, good solution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10 Euro per word: Does not sound like an offer, an opportunity (&quot;Verantwortung&quot;) - &quot;Verantwortung für&quot; not &quot;Verantwortung über&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

40. Would you like to comment?

Figure 17: Sample screen; Experiment 3 (Information provided)

Selecting one of the options for each reviewer comment was mandatory within the form, while the comment field included in each of the scenarios could be filled in optionally. As with the experiments previously described, one option to check whether participants were still paying attention was included by embedding instructions to select a specific option in the reviewer comments, which was excluded from analysis but will be included in the discussion chapter (section 9.2.2.).

However, while the basic structure of the experiment was similar to that of Experiments 1 and 2, the design of this experiment posed unique challenges in that reviewer comments were included rather than translations themselves.

1. Participants weren't able to judge the translations themselves but could only see the reviewer comments. If a reviewer had commented on one translation in a very negative or very positive way, this could have biased the results since the reviewer's
opinion introduced an additional variable. This was resolved by changing the approach to selecting the reviewer comments rather than adopting the approach taken for Experiments 1 and 2.

For Experiment 3, all reviewer comments and the ratings that had been provided with the comments were placed in a spreadsheet, sorted according to the translator who had originally provided them. Then, based on the ratings and the wording of the comments, similar types of comments were combined. For example, the first scenario presented to participants included only reviewer comments that had been accompanied with the lowest rating ‘1’ and were worded in a negative way (see Table 20 for an example).

<table>
<thead>
<tr>
<th>MedCost</th>
<th>LowCost</th>
<th>HighCost</th>
<th>VolSelfselect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very literal and formal; grammar mistake, sounds unnatural and not really getting the points across.</td>
<td>Sounds awkward and „Zugang“ is more like having access to resources like computers etc.</td>
<td>Does not sound like an offer, an opportunity („Verantwortung“)</td>
<td></td>
</tr>
</tbody>
</table>

Table 20: Sample comments, translations rated 1

On the other hand, the third scenario included only reviewer comments that had been accompanied by the highest rating ‘5’ and were worded in a positive way (See Table 21 for an example).

<table>
<thead>
<tr>
<th>MedCost</th>
<th>LowCost</th>
<th>HighCost</th>
<th>VolSelfselect</th>
</tr>
</thead>
<tbody>
<tr>
<td>reflects source text in a good way, good adaption</td>
<td>Correct and short, sounds good. “Erfolgreich” would have been a closer translation, in my opinion (it can be read as an ellipsis of “Erfolgreich sein”= to be successful) but “Erfolg” has a better flow overall.</td>
<td>good translation; not too literal but getting everything across</td>
<td>Rather informal, but it is very clear and uses the students “everyday language”. It is an easily read introduction. Some expressions are not quite so good: “Arbeitsaufwand”, “das beste aus dem Studium zu machen.”</td>
</tr>
</tbody>
</table>

Table 21: Sample comments, translations rated 5

230
To further reduce variability due to the reviewer's opinions, the original allocation of comments to the scenarios was repeated from NoInfo to BadInfo and GoodInfo scenarios. This meant that the participants were shown the same scenario three times, with other scenarios between the repeats, with the independent variable of information on cost changing for each repeat. As an example, the comment "good translation; not too literal but getting everything across", which was based on a segment translated by the highly-paid translator, was included without information in scenario 3, with the correct information in scenario 8 and with the incorrect information (claiming the translation had been done by an unpaid volunteer) in scenario 13. GoodInfo and BadInfo scenarios were mixed, so that participants were sometimes shown the correct information first and sometimes the incorrect information.

2. Since this made it clear to the participants, in case they noticed the repetitions, that they were sometimes shown the correct information and sometimes the incorrect information about the cost of the translation, it was not possible anymore to simply ask them to select the preferred option. Instead, the focus shifted away from making one choice towards indicating their likelihood to trust or select each of the options based on the information presented.

All participants were presented with the same 15 scenarios.

6.5.4.2. Participant Sampling and Conducting the Experiment

All respondents who had been directed towards the ‘Buyer’ branch of the main questionnaire (Questionnaire 5), as well as non-profit clients on Trommons (Questionnaire 6) were invited to participate in this experiment.

Since they had self-declared as primarily buyers of localisation and translation services, and ‘buyer’ had been supplemented with the definition

"* By buyer we mean the person who decides on or recommends to decision makers whether a translation service will be requested and later the delivery accepted."

it was reasonable to expect that this group of respondents would have previously
made decisions on whether or not to accept a localisation/translation service or whether to request localisation/translation services from a particular individual or company.

As a result, participants who declared to be in the ‘buyer’ category could be expected to be well suited as a population group for this experiment.

7 respondents in the industry buyer group completed the experiment. Further demographic information on the respondents will be provided in section 7.1.3.2.

In addition, contacts at the non-profit organisations acting as clients on Trommons were invited to fill-in questionnaire 6 and were invited to participate in experiment 3. For this respondent group, the only criteria for inclusion were that they were on the client mailing list at TRF and had not indicated that they would not want to be contacted by e-mail.

8 respondents in the non-profit client group completed the experiment. Further demographic information on the respondents will be provided in section 7.1.3.1.
Part III: Data Analysis and Research Findings

Chapter 7: Survey Results

The following chapter reports on the findings of the different studies that were part of this thesis, as well as the data analysis techniques used. Section 7.1 describes the demographics of study participants, section 7.2 then reports on findings from the questionnaires, content analysis and interviews. Here, sections are grouped into different topics, which include

- 'Defining and evaluating quality',
- 'Factors impacting quality',
- 'The language professional and quality' and
- 'Volunteer translation and crowdsourcing'.

Questionnaires and experiments were analysed using a combination of frequency analysis and cross-tabulation in SPSS\textsuperscript{42} and Excel\textsuperscript{43}, interviews were coded using NVivo\textsuperscript{44} and content analysis of reviewer comments, and analysis of unstructured data from questionnaires, were analysed using coding in Excel.

A summary of the findings is included in the discussion chapter (Chapter 9).

7.1. Demographics

As described in Chapters 2 and 3, the background of the individual evaluating quality plays a role in how quality is perceived. This following section will first summarise the demographic information gathered on the study participants during data collection.

7.1.1. Experiment 1: Spanish Linguists

Experiment 1 (impact of professional status cue on perceived quality, section 6.5.2.) was administered to students and lecturers at the University Alfonso X el Sabio in Madrid,

\textsuperscript{43} https://products.office.com/en-ie/excel
\textsuperscript{44} http://www.qsrinternational.com/nvivo-product
as well as issuing an open call to Spanish speaking translation and localisation professionals (reviewers and translators). The following sections report on the results of the demographic section of the questionnaire for these two groups.

7.1.1.1. Students and Lecturers, University Alfonso X el Sabio in Madrid

As outlined in section 6.5.2.2., students and lecturers from the University Alfonso X el Sabio in Madrid were invited to participate in an experiment, combined with a survey questionnaire. During data preparation, those respondents who had not completed the full questionnaire and experiment were excluded from further analysis. This resulted in the exclusion of 7 responses, leaving 48 for analysis.

Participants fell into the following categories described in Table 22.

<table>
<thead>
<tr>
<th>Group</th>
<th>Numbers</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st year student</td>
<td>8</td>
<td>16.7</td>
</tr>
<tr>
<td>2nd year student</td>
<td>3</td>
<td>6.3</td>
</tr>
<tr>
<td>3rd year student</td>
<td>8</td>
<td>16.7</td>
</tr>
<tr>
<td>4th year and Master’s students</td>
<td>13</td>
<td>27.1</td>
</tr>
<tr>
<td>Lecturer</td>
<td>16</td>
<td>33.3</td>
</tr>
</tbody>
</table>

Table 22: Participant categories UAX

Participants were grouped into three categories, based on their level of training received at that point and their working experience (Table 23).

<table>
<thead>
<tr>
<th>Participant grouping</th>
<th>Numbers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st/2nd year students</td>
<td>11</td>
<td>23%</td>
</tr>
<tr>
<td>3rd/4th/Master’s students</td>
<td>21</td>
<td>43.80%</td>
</tr>
<tr>
<td>Lecturers</td>
<td>16</td>
<td>33.30%</td>
</tr>
</tbody>
</table>

Table 23: Grouping of participants at UAX

43 respondents self-identified as native speakers of Spanish, two as native speakers
of German and French respectively and three indicated that they were not native speakers of Spanish without providing further details. However, since all participants had demonstrated knowledge levels of the Spanish language that were sufficiently high to be enrolled into a translator programme, none were excluded from the following questionnaire and experiment based on their response.

On translation experience (Figure 18):

- 54% (the majority) indicated that they had no professional translation experience;
- 16% said they had less than one year of experience;
- 10% had between one and 5 years of experience and the remaining
- 20% were evenly distributed within the categories of the 5 to 30 years’ range.

None of the 1\textsuperscript{st} and 2\textsuperscript{nd} year students had professional translation experience, and the category 'less than 12 months of experience' was selected primarily by 3\textsuperscript{rd} and 4\textsuperscript{th} year
students, while those who had selected 5 to 30 years or up to 5 years of experience, were predominantly part of the lecturer respondent group.

Lecturers were also asked whether they had received training in reviewing translations (Figure 19). Students were omitted from this question since the assumption was that they had enrolled in the programme to receive training in reviewing translations, among other things.

![Review training; Lecturers](image)

**Figure 19: Reviewer training of UAX lecturer participant group**

### 7.1.1.2. Spanish Translators and Reviewers (Open Call)

74 of the participants who finished the questionnaire indicated that they were linguists (translators or reviewers for non-profit or for-profit projects), and that they spoke Spanish. This respondent group was then asked to indicate their level of Spanish knowledge and those who selected a high proficiency level (native or near-native knowledge) were invited to participate in the experiment as well after they had completed the questionnaire. 42 respondents also agreed to participate in this experiment.

15.1% of the 74 questionnaire respondents were currently a student at school,
college or university with the majority (54.5%) being enrolled as tertiary postgraduate students, and 18.2% tertiary undergraduate student. All but three of those 11 respondents currently enrolled in a programme were enrolled in a translation or localisation related course.

![Highest level of education; Spanish linguists](image)

**Figure 20: Highest level of education in Spanish linguist participant group**

Education levels of this respondent group were overall high (Figure 20) and 58.9% of the respondents indicated that their highest level of education was in a linguistics or translation related area. Of those respondents whose highest level of education was not in a linguistics or translation related field, only 5 had no training or education in this area, while the majority of respondents whose highest level of education was not in a translation or localisation related area had some linguistic or translator training or education.

All the respondents in this group had selected that the primary role they had held in the field was in translation or review.

59.5% had experience as for-profit translators or reviewers, 58.1% had volunteered on Trommons and 56.8% on initiatives other than Trommons. The majority identified their primary role as for-profit or paid-for translator/reviewer (Figure 21).
Of the 43 respondents who had experience volunteering on Trommons, 18 indicated that their primary role was paid-for translation or review, while of the 42 who had volunteered for other initiatives, 22 were primarily paid translators or reviewers.
While the largest group of respondents had been involved in localisation for one to 5 years, nearly half indicated more than 5 years of experience (Figure 22).

7.1.2. Experiment 2: German Linguists

In experiment 2 (the impact of price of translation cue on perceived quality, section 6.5.3.), 44 of the respondents who finished the questionnaire indicated that they were linguists (translators or reviewers for non-profit or for-profit projects), and that they spoke German. This respondent group was then asked to indicate their level of German knowledge and those who selected a high proficiency level (native or near-native knowledge) were invited to also participate in the experiment after they had completed the questionnaire. 25 respondents agreed to participate in the experiment.

27.9% of the 44 questionnaire respondents indicated they were currently a student at school, college or university with 58.3% being enrolled as a tertiary postgraduate student, and 33.3% a tertiary undergraduate student. 41.7% of the respondents currently enrolled in a programme were enrolled in a translation or localisation related course.
In this group, the majority of respondents were educated to Bachelor’s or Master’s level (Figure 23) and 62.8% of the respondents indicated that their highest level of education was related to linguistics or translation. Of those 14 respondents, whose highest level of education was not in a linguistics or translation related field, 8 (57%) had no other training or education in this area.

All the respondents in this group had selected translation or review as their primary role.
The majority of respondents were primarily for profit/paid-for translators or reviewers (Figure 24). Of the 19 respondents who had experience volunteering on Trommons, 10 (52%) indicated their primary role was paid-for translation or review, while of the 19 who had volunteered for other initiatives, 12 (63%) were primarily paid translators or reviewers.

The majority of respondents in this group had more than 5 years’ experience in localisation (Figure 25).
Figure 25: Time involved in localisation of German linguist participant group

7.1.3. Experiment 3: Buyers/Clients

Experiment 3 (impact of professional status cue on perceived quality, section 6.5.4.) was administered to contacts at organisations on Trommons (non-profit client) as well as for-profit buyers. The following sections report on the results of the demographic section of the questionnaire for this respondent group.

7.1.3.1. Non-profit Clients (Trommons)

To survey non-profit clients on Trommons, the call for participation in the questionnaire and experiment was sent out to 430 organisations, of which 89 opened the invitation e-mail. Response rates were low, with only 20 representatives of organisations on Trommons responding, of which 17 completed the full questionnaire. 8 of those that completed the questionnaire then also agreed to participate in the experiment.

While a higher response rate would have been preferable, the low numbers did not
come entirely unexpected, since the decision had been made to not send reminders or further invitations to avoid being intrusive, thereby protecting the interests of TRF in preserving good standings with the collaborating organisations. In addition, the majority of contact persons for organisations on Trommons are not specialists in localisation but in other areas, which could indicate a generally low interest in the topic of localisation.

This impression was also confirmed when participants were asked how much time they had spent managing localisation projects, including the time spent collaborating with TRF, which showed overall low levels of experience (Figure 26).

Figure 26: Years of experience managing localisation in non-profit client participant group

Regarding the relationship to TRF, 50% of the total respondents (including those that did not participate in the experiment, but only completed the questionnaire) self-reported as the main contact for translation projects, with 30% as one of several contacts. One participant indicated they had no ties to TRF or Trommons and was excluded from the rest of the questionnaire and experiments, while the remaining 15% indicated they maintained contact to TRF, but didn't manage translation projects.
Table 24: Education/training in localisation related areas of non-profit client participant group

<table>
<thead>
<tr>
<th>Localisation related training/education</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistics/Translation</td>
<td>17.60%</td>
</tr>
<tr>
<td>Project management/Business</td>
<td>11.80%</td>
</tr>
<tr>
<td>Content creation</td>
<td>5.90%</td>
</tr>
<tr>
<td>None</td>
<td>70.60%</td>
</tr>
</tbody>
</table>

The majority of the respondents had no training or education in a localisation related field (Table 24) and the training area they had selected was their main area of training for none of the participants. 11.8% of the respondents had received training in reviewing translations and assessing texts at an academic institution, and 29.4% in a non-profit organisation. The remaining had not received training in this area.

Overall, these responses indicate that the majority of participants were involved in translation projects on Trommons, but that translation was only a side project in terms of time spent on managing such projects, experience in related areas and training.

7.1.3.2. Industry Buyers (Open Call)

For industry buyers, 14 of the participants that finished the questionnaire indicated the primary role they had filled in translation or localisation was that of a for-profit (13) or non-profit (1) buyer of localisation/translation services. After the questionnaire, this group was also invited to participate in the experiment, to which 7 respondents agreed.

One respondent in this group indicated that they were currently a mature student studying in a field they had previously worked in, and that this field was translation or localisation related. None of the remaining respondents were at that point enrolled in school, college or university.
Education levels in this group were high (Figure 27),

- 15.4% of the respondents said their highest level of education was in linguistics or translation,
- 7.7% computer science or programming with a focus on localisation, and
- 7.7% content creation.

However, 8 respondents (61.5%) said their highest level of education was not in a translation or localisation related field. Apart from the highest level of education, 30.7% of respondents had no translation or localisation related training or education. 42.9% of respondents indicated that they had not received training in reviewing or assessing quality.

All the respondents in this group had selected ‘buyer of translation or localisation services’ as their primary role.

7.1% had volunteered on Trommons and 21.4% had volunteered as linguists on other voluntary initiatives. 28.5% had experience as for-profit/paid translators or reviewers.

Most of the respondents in this group had been involved in localisation for more than 5 years (Figure 28).
7.1.4. Demographics: Other Linguists

82 of the participants that finished the questionnaire indicated that they were linguists (translators or reviewers for non-profit or for-profit projects), but that they spoke neither Spanish nor German. They were therefore grouped into the 'Other Linguist' category and were not asked to participate in an experiment, but were included in the linguist questionnaire to gather insights into their attitudes and opinions on topics around quality and translation crowdsourcing.

9.9% of this respondent group were currently students at school, college or university with an even split across secondary students, tertiary postgraduate students and mature students studying a field they had previously worked in. All but one of those 8 respondents currently enrolled in a programme were not in a translation or localisation related course.
In this group, the majority of respondents were educated to bachelor's or master's level (Figure 29). 48.1% of the respondents indicated that their highest level of education was a linguistics or translation related field, 3.7% selected project management or business and 42% selected that their highest level of education was not in a translation or localisation related field. However, when asked about their training or education independent of their highest level of education, none of the respondents in this group selected that they had no education or training in the field.

Independent of whether they identified it as their primary role or not and being able to select multiple options, most frequently respondents indicated that their translation experience was:

- being a volunteer on Trommons (56.1%),
- linguistic volunteering on a platform other than Trommons (41.5%),
- for-profit linguistic experience (34.9%).

A relatively large number of respondents indicated they had been involved in
localisation or translation for fewer than 12 months (15.9%), while 29.6% fell in the ‘1 to 5 years’ category, 22.2% into the 6 to 10 years and 30.5% said they had been involved in the field for 11 or more years.

7.1.5. Demographics Interviewees

The semi-structured interviews included two contact persons at non-profit organisations that act as clients for translations on Trommons (OrgClient 1 and OrgClient 2), three buyers of localisation services in the industry (ProBuyer 1, ProBuyer 2 and ProBuyer 3) as well as 4 translators who volunteer on Trommons, as well as other volunteering initiatives, and being paid translators (Translator 1, Translator 2, Translator 3, Translator 4).

To protect the identity of the informants, no detailed, individual description of their backgrounds will be provided but instead only the most pertinent points.

The organisation clients were both experts in other fields and for both, translations were just a small part of their job at the respective organisation. Both had requested translations themselves, although one of them was primarily overseeing others who requested translations via Trommons. OrgClient 1 had requested texts primarily in the legal field, while OrgClient 2 was focused on technical content.

All four interviewees in the translator group were active volunteers on Trommons, with differing levels of experience and training. Translator 1 was the most experienced translator, while Translators 2, 3 and 4 had been volunteering between two years and a few months. Translators 1 and 2 were working as freelancers, while translators 3 and 4 were looking to start in the translation industry and were interested in gaining experience by volunteering.

The three buyers of localisation services in the industry had been working in localisation for a minimum of 8 years; one of them since the late 90s. All three had worked in different positions within the industry but were now decision makers regarding requesting and accepting localisation services, two in a project manager capacity and one in a general management position. All three were employed in localisation service provider organisations and were trained in localisation related areas like linguistics. For one of the
interviewees, video game related content was the focus of the work while the other two reported a mix of video games and other types of content.

7.2. Attitudes

The following sections describe the outcomes of data gathering and analysis for the questionnaires, content analysis and interviews. The sections are presented by topics, with different data collection and analysis techniques grouped into the themes. Section 7.2.1. describes the theme of defining and evaluating quality. Here the question how different groups define quality and which attributes of a localised end-product are valued higher than others in relation to quality was explored. In addition, techniques for determining or evaluating quality are included in this section.

Section 7.2.2. describes factors impacting quality with a focus on the price of translations, as well as training and experience of translators. Section 7.2.3. describes the outcomes around the topic of what defines a localisation or translation professional and how an individual being a professional might impact quality. Finally, section 7.2.4. addresses opinions and attitudes towards crowdsourced translations.

Interview excerpts are presented in edited versions for legibility, removing interjections, vocalized pauses, and false starts and duplicate words. In addition, punctuation marks were inserted.

Questionnaire responses were copied verbatim, including spelling and grammar errors. The researcher decided to refrain from marking such errors with the customary [sic] to not disrupt the flow of the text unduly. The only alterations made to responses were the occasional removal of parts of the response if they didn't relate to the theme being discussed, which was marked with '(...)'.

7.2.1. Defining and Evaluating Quality

As described in chapter 3, the question of how to define and by extension evaluate quality is widely discussed within academia and the localisation industry, but so far, no consensus has been found. To understand the expectations the different groups, who are
part of this research have, the question of how they define and evaluate quality was raised initially.

In the questionnaires, participants of all respondent groups were asked to describe translation quality in their own words. These responses resulted in unstructured, text-based data which were coded by the researcher into themes that emerged from the responses.

The themes assigned, and their definitions were as shown in Table 25, although not all themes emerged in the responses of all groups.

Number of mentions per category were then summed up (Figure 30). In addition, the number of individual categories that were assigned to each respondent were also summed up to evaluate the level of detail in the definition of quality provided by individual groups.

To determine reliability of the coding approach, 15% of responses were coded by a second coder and inter coder reliability (Cohen’s Kappa) calculated, which resulted in average reliability of 0.718, and was deemed sufficient as a value of 0.8 could be seen as strong agreement for Cohen’s Kappa (Pfeil and Zaphiris 2010, p. 13).

Questions varied slightly depending on respondent group. For example, students and lecturers at the University Alfonso X el Sabio in Madrid were asked only a small number of questions since they were also asked to participate in a more time-intensive experiment, which will be the focus of a forthcoming research publication.

In addition to being asked to provide a definition of translation quality, non-profit clients were also asked to describe which factors in translation were important to them or their organisation and how they decided if they were comfortable using a translation for their project when they didn't speak the language in question. These responses were grouped into themes which are presented together with examples.

Participants in the industry buyer group as well as industry linguists were asked to provide their own definition of quality. They were then offered a list of quality criteria and asked from their perspective to rank them according to importance, to gather more easily quantifiable data in addition to the participant's own descriptions.

Industry buyers were also asked to describe how they decide whether they can trust a translation with the question "When you receive a translation, how do you decide if you
are comfortable using that translation for your project assuming you don't speak the language in question?"

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>Specific mentions of additions</td>
<td>convey faithfully (without adding or subtracting) the message of the source language in a natural manner for the target</td>
</tr>
<tr>
<td>Attention to detail</td>
<td>Mention of attention to detail or thoroughness as a quality factor</td>
<td>Adaptation to content requirements in terms of style, tone, platform or technical specific needs. Attention to detail, thoroughness.</td>
</tr>
<tr>
<td>Client needs</td>
<td>Mention of end-user/customer/client requirements or needs</td>
<td>Meeting well defined requirements of customers who know what they want and what they need.</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Mention of how well a translation can be understood or has been understood</td>
<td>Translation/localization quality defines by understanding it easily from the reader.</td>
</tr>
<tr>
<td>Culture</td>
<td>Mention of culture or the target audience</td>
<td>Communication of concepts and skills in language that will be appropriate for a target group, culture, country or socio-economic group.</td>
</tr>
<tr>
<td>Faithful/Loyal/Accurate</td>
<td>Target texts convey the original author's intended meaning or reflect the original text's intention.</td>
<td>Being able to communicate an original text/source while being loyal to the writer's message.</td>
</tr>
<tr>
<td>Grammar</td>
<td>The word grammar is explicitly mentioned, or structure of the language.</td>
<td>Ensuring the final product is linguistically correct, stylistically appropriate and the expectations of all stakeholders are met.</td>
</tr>
<tr>
<td>Linguistics</td>
<td>General reference to linguistics or language without specifying an aspect.</td>
<td>You have to take in consideration : terminology, structures, source and target languages among others.</td>
</tr>
<tr>
<td>Misunderstood</td>
<td>The participant appears to have misunderstood the question and entered unrelated text or text that appears to answer a different question.</td>
<td>Books are quite good translated but translations of other texts like instructions for example aren't so good.</td>
</tr>
<tr>
<td>Naturalness (Idiomatic)</td>
<td>Mention of text sounding natural or like it had been written in the target language rather than being a translation.</td>
<td>In my opinion, translation quality is how well a text was translated into the target language. The text should be translated in such a way that the reader in the TL will not feel they are reading a translation.</td>
</tr>
<tr>
<td>Omission</td>
<td>Specific mentions of omissions</td>
<td>convey faithfully (without adding or subtracting) the message of the source language in a natural manner for the target</td>
</tr>
<tr>
<td>Other preferences</td>
<td>Mention of an aspect that doesn't fit in any other category.</td>
<td>You have to take in consideration : terminology, structures, source and target languages among others.</td>
</tr>
<tr>
<td>Process</td>
<td>Process related aspects that contribute to quality, like requirements, revision or using resources.</td>
<td>A translation that meets the clients specifications on quality. If I am not given any specification, I try to imagine who will receive the translation and what for, and I translate accordingly. I also try to write as correctly as possible.</td>
</tr>
<tr>
<td>Spelling</td>
<td>The word spelling is explicitly mentioned, or in case of linguistic issues being mentioned, this is counted both, towards Grammar and Spelling.</td>
<td>To have a text that is free of linguistic errors, that sounds natural and that is accurate based on the sources text.</td>
</tr>
<tr>
<td>Style</td>
<td>Mention of style, fluency or tone of the translation</td>
<td>Fluid writing in the target language, semantic accuracy in relation to the source text.</td>
</tr>
<tr>
<td>Technical</td>
<td>File formats, platforms and similar technical considerations</td>
<td>And thirdly, for localization purposes, it needs to be in a digital format that is familiar to the end-user.</td>
</tr>
<tr>
<td>Terminology</td>
<td>Mention of terminology, vocabulary or words</td>
<td>Translation quality consists in the achievement of the same meaning of both target and source text, with a proper quality on terminology, vocabulary...</td>
</tr>
</tbody>
</table>

Table 25: Common labels emerging from definitions of localisation quality across all groups
Figure 30: Commonly mentioned aspects of quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.


7.2.1.1. Defining Quality

7.2.1.1. UAX

By far the most frequently mentioned aspect to describe quality was the faithful or accurate representation of the text meaning (22). In the median range were

- process related aspects (7),
- style (7),
- grammar (5) and
- terminology (5).
- Naturalness (3),
- spelling (1),
- additions (1) and
- omissions (1) ranked at the bottom of the list and none of the respondents mentioned typography or untranslated text (Figure 30).

2 of the 49 respondents (4%) mentioned the client's requirements or the target audience in relation to quality:

"To me, translation quality is found when the text is intelligible, respects the original author's intentions and style, and it's adapted to the target audience."

[Questionnaire respondent]

"A translation that meets the clients specifications on quality. If I am not given any specification, I try to imagine who will receive the translation and what for, and I translate accordingly. I also try to write as correctly as possible."

[Questionnaire respondent]

In contrast, 20 (40.8%) described quality by referring to the text, while the remaining respondents focused on the message, the author or gave a response that didn't indicate either as focus:
"Fluid writing in the target language, semantic accuracy in relation to the source text." [Questionnaire respondent]

"I consider that a good translation must be faithful to the original text, respect its style, vocabulary choices, and sentence structures." [Questionnaire respondent]

The majority of respondents offered one (22) aspect to describe translation quality, while 9 respondents outlined two aspects and 7 respondents three or more aspects. The majority of quality descriptors were provided by 3rd/4th year students and lecturers, with 1st and 2nd year students nearly exclusively giving one aspect to describe quality. Among the 3rd/4th year students and lecturers half of the respondents provided two or more aspects of quality and no noticeable differences between those two groups could be observed.

7.2.1.1.2. Non-profit Clients

12 participants provided a definition of quality, of which 4 respondents seemed to have misunderstood the question. The majority of those respondents who did provide a definition that was included (5/62.5%), described quality including one aspect in their definition, three (37.5%) included two aspects and one (12.5%) three aspects.

The most frequently mentioned aspect to describe quality was the faithful or accurate representation of the text meaning (4/33.3%). This was followed by

- terminology (2/16.6%),
- style (2/16.6%) and
- cultural appropriateness (2/16.6%) in the median range.
- Grammar, spelling, naturalness all received one mentioned each and none of the respondents mentioned typography or process-related aspects (Figure 30).

The majority of participants provided a response both, to the question of which factors in translation were important to them and their organisations as well as to the question of how they defined translation quality. Here, too, the faithful representation of the source text received the most mentions with 6 out of 9 responses (66.6%) falling into the faithful/loyal/accurate category, and grammar, spelling, terminology or correctness of
language were each mentioned only once.

However, when asked to describe what was important to them rather than how they defined quality, nuances in what is important to the organisations regarding faithful representation of the source text could be observed.

One respondent differentiated between text for publication and text that the organisation only needed to be able to understand:

"For us, mostly the context and content of the translation is important when it is a document we just need to understand. When we need a translation of for example a press release, something that shall be published, the grammatical and orthographical correctness is also important." [Questionnaire respondent]

Four respondents referred to the meaning of the original needing to be preserved:

"to maintain the meaning of the original text" [Questionnaire respondent]

"That the text would make sense and not be a literal word-for-word translation which at times makes no sense. That the text would convey the meaning intended, if this is possible for the translator." [Questionnaire respondent]

"We need to be really accurate with our translations in order to be the same meaning with our team in India." [Questionnaire respondent]

"Keeping the meaning." [Questionnaire respondent]

Two respondents mentioned requirements beyond transfer of meaning, such as the spirit of the organisation and the context of the content.

"We hope that the volunteer translator will review our website, see the spirit of our organization, understand the context of our healing work with underserved populations--as a backdrop for the specific project they are translating."

[Questionnaire respondent]

"Someone familiar with general religious vocabulary will select terms that fit the context of the document. Ex. the one responsible for a religious community is not the community manager but superior. The religious specificity of certain
The importance of transferring meaning correctly was also stressed during the interviews conducted with non-profit clients. OrgClient 2, whose organisation has mostly requested technical content to be translated via Trommons, described it as follows:

"For us our main project it's actually a guide tool for activists so the important thing is that people get access to the core ideas and to the core tools and instruments we are trying to expand to other countries." [OrgClient 2]

However, OrgClient 2 also described the information being available as "the minimum quality", which indicates acknowledgement that other factors beyond transfer of meaning might be desirable.

OrgClient 1, who deals with projects where translated content is provided to government bodies focused on translations sounding natural. In addition, this respondent mentioned their readability, as well whether the translations contained 'typos' or 'don't seem right', showing more reference to stylistic issues than for OrgClient 2:

"For what is the readability like when you're reading a document. I know sometimes it is translated, of course you want to do it word for word but sometimes it seems like it was a little - you know - the grammar is a little off like they're trying to be precise with the word choice but it doesn't read well"

[OrgClient 2]

7.2.1.1.3. Professional Buyers

The 9 industry buyers who provided a definition of quality in their own words mentioned in 4 instances (44.4%) each aspect relating to the faithful, accurate or loyal transfer of meaning, as well as the translations needing to sound natural, or like they had been written in the target language. Target culture was mentioned three times, although for this respondent group this was also referred to as target 'market' or 'region'. Style, grammar and terminology received even fewer mentions and issues around process, spelling and technical issues were not mentioned at all (Figure 30).
Focus was also placed on satisfying end-user or client needs or customer requirements, which were mentioned 5 times in total out of 9 responses (55.5%):

"Meeting end user needs" [Questionnaire respondent]

"Meeting well defined requirements of customers who know what they want and what they need." [Questionnaire respondent]

"Quality is the extent to which a piece of (localized) content adheres to pre-agreed requirements that were (hopefully) put in place based on evidence/data with a goal to maximize the impact of this (localized) content." [Questionnaire respondent]

Respondents in this group provided a relatively high amount of detail in their quality descriptions, with

- three (33.3%) giving only one aspect of quality,
- four (44.4%) providing three aspects and
- one each (11.1%) two or four aspects.

The tendency to have a detailed concept of quality available could also be observed during the interviews with professional buyers in the industry. This group referred to:

- the use of standard quality metrics,
- the purpose of the translation needing to be considered (clarity of translation vs. style),
- the timelines for a project (quality requirements being lower for urgent projects),
- the target audience (some target groups being more concerned with issues like punctuation than others) as well as
- the subject matter expertise required to convey complex, highly specialised meaning accurately.

ProBuyer 1, who to a large extent deals with video game localisation, distinguishes between technical and linguistic quality, describing them as follows:
"(...) basically there's a technical side to quality and there's a purely linguistic side. (...) So the technical stuff is basically the stuff that a non-native person can check. (...) It's gonna be about the - you know - double spaces, wrong capitalisation compared to the source text, weird punctuation, numbers not matching, people making mistakes in variables and things like that. That basically a non-native person can check. And we have, we have a series of tools that allow us to run through files and those would pop up quite quickly. And so everything else is what we can't sort of judge in house so it's going to be about style and all that." [ProBuyer 1]

When questionnaire respondents were asked to rank quality criteria according to importance,

- the only criterion that ranked exclusively in positions 1 to 4 was the correct representation of meaning, where it also was in the top three for all four positions.
- Respondents also placed some importance on correct terminology, which ranked in the top three for positions 3 and 4.
- Less important were formatting which was placed primarily on position 8, as well as technical requirements on positions 6, 7 and 8.
- Consistency as well as style and register were placed in the mid-range with positions 4, 5 and 7 for consistency and style and register on positions 5 and 6.

Interestingly, opinions were split on the topics of linguistic errors (Spelling, Grammar and Punctuation) as well as adherence with customer specifications. Linguistic errors ranked in the top three for positions 1 and 2 as well as positions 5 and 6, while adherence with customer specifications was placed primarily in positions 2 and 3, as well as positions 7 and 8 (Table 26).
Table 26: Quality aspects ranked by importance by professional industry buyer participant group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9.
It should be noted that the pre-defined criteria the respondents were asked to rank did not include a category for target market or culture since the researcher had not considered this aspect when designing the questionnaires based on literature and interviews. However, this aspect was mentioned in the descriptions of quality respondents provided and should be considered for future questionnaire design.

7.2.1.4. Linguists

155 respondents in the linguist group provided a definition of quality in their own words (Figure 30).

- Most commonly, respondents mentioned that translations needed to convey meaning faithfully/loyally/accurately (42.6%).
- This was then followed by
  - style and register (31.6%) being common concerns and
  - that translations needed to sound natural (21.3%), or like they had been written in the target language originally, as well as
  - 'culture', or fitting the target audience/the user's need (21.3%).
- In the mid-range,
  - grammar (15.5%),
  - spelling (7.1%),
  - comprehension (6.4%) and
  - generally correct language (6.4%) were mentioned most frequently.

Finally, 31 respondents had misunderstood the question or provided an answer that wasn't suitable for analysis, while three felt they couldn't provide a definition of quality.

Further responses, with only few mentions each, included technical considerations like file formats, process related issues and attention to detail, the impact a text has, and whether it uses correct terminology or typography.

27.7% provided one aspect to when defining quality, 18.7% each mentioned two or three aspects. The remaining respondents mentioned 4 (8.4%), 5 (3.2%) or 6 (1.9%) aspects
to define quality.

As part of the interviews conducted for this research, informants were also asked to give their description of quality, or what aspects mattered regarding quality. A major topic highlighted during three of the interviews was that the message or facts had to be conveyed accurately. Other aspects mentioned were that the translations needed to have a nice style, sound natural and read well in the target language, be free of spelling and syntactic errors and maintain terminological consistency, especially as far as subject matter specific terminology was concerned while also using accurate terminology. There was general agreement that how well the target text was written was secondary to the original message being conveyed accurately.

Questionnaire participants were also provided with a list of quality criteria and asked to rank them by importance (Table 27).

Criteria most frequently listed in first, second or third position were

- "Accuracy, Omissions, Additions (the meaning is represented correctly)" (Top 3 for Positions 1 and 2),
- "Spelling, Grammar, Punctuation (there are no linguistic errors)" (Top 3 for Positions 1, 2, 3 and 4),
- Terminology (Top 3 for Positions 1, 2 and 3).

In the mid-range,

- Consistency (Top 3 for Positions 3, 4 and 5), as well as
- Style and Register (Top 3 for Positions 4, 5 and 6) were most frequently ranked.

In the lower-middle to lower range were

- Adherence with customer specifications (Top 3 for Positions 5 to 8),
- Formatting (Positions 6 to 8) and
- Technical requirements (Positions 6 to 8).
<table>
<thead>
<tr>
<th>Quality aspects</th>
<th>Pos. 1</th>
<th>Pos. 2</th>
<th>Pos. 3</th>
<th>Pos. 4</th>
<th>Pos. 5</th>
<th>Pos. 6</th>
<th>Pos. 7</th>
<th>Pos. 8</th>
<th>Pos. 9</th>
<th>Weighted avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy, Omissions, Additions (the meaning is represented correctly)</td>
<td>73</td>
<td>38</td>
<td>27</td>
<td>19</td>
<td>17</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>1412</td>
</tr>
<tr>
<td>Spelling, Grammar, Punctuation (there are no linguistic errors)</td>
<td>45</td>
<td>38</td>
<td>30</td>
<td>32</td>
<td>24</td>
<td>22</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1335</td>
</tr>
<tr>
<td>Terminology (correct Terminology)</td>
<td>22</td>
<td>49</td>
<td>44</td>
<td>31</td>
<td>18</td>
<td>11</td>
<td>17</td>
<td>5</td>
<td>0</td>
<td>1279</td>
</tr>
<tr>
<td>Consistency (the same terms, style etc. are used throughout the translation)</td>
<td>14</td>
<td>23</td>
<td>42</td>
<td>34</td>
<td>46</td>
<td>21</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>1162</td>
</tr>
<tr>
<td>Style, Register (the translation sounds good/appropriate)</td>
<td>18</td>
<td>23</td>
<td>26</td>
<td>38</td>
<td>30</td>
<td>32</td>
<td>15</td>
<td>11</td>
<td>4</td>
<td>1105</td>
</tr>
<tr>
<td>Adherence with customer specifications</td>
<td>18</td>
<td>12</td>
<td>18</td>
<td>21</td>
<td>32</td>
<td>42</td>
<td>26</td>
<td>22</td>
<td>6</td>
<td>966</td>
</tr>
<tr>
<td>Technical requirements</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>12</td>
<td>15</td>
<td>30</td>
<td>48</td>
<td>68</td>
<td>6</td>
<td>696</td>
</tr>
<tr>
<td>Formatting</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>30</td>
<td>66</td>
<td>69</td>
<td>6</td>
<td>615</td>
</tr>
<tr>
<td>Factors not listed here</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>168</td>
<td>295</td>
</tr>
</tbody>
</table>

Table 27: Quality aspects ranked by importance by linguist participant group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9.
7.2.1.5. Content Analysis

As described in section 6.3., reviewer comments provided by volunteers and partner organisations for volunteer translations on Trommons were labelled to categorise the quality issues described. Labels applied followed the Multidimensional Quality Metrics (MQM) with the definitions outlined there\footnote{Definitions available: http://www.qt21.eu/mqm-definition/definition-2015-12-30.html [last accessed April 2nd, 2017].} unless otherwise stated. The number of occurrences of those labels was then used for further quantitative analysis.

Reviewer comments that use label names, synonyms of the names and terminological subsets were labelled in a separate, semi-automated, step. The number of occurrences of such labels was then compared to the results of manual coding to show quality issues that reviewers tend to describe in more definitive terms vs. those where they use descriptive phrases or a wider range of terms instead.

The initial grouping of labels into more vs. fewer than 10 occurrences (Table 28) shows that the category ‘miscellaneous’, as well as all labels under the parent labels ‘verity’ and ‘design’ occur fewer than 10 times. Within this category, the labels ‘whitespace’, ‘accuracy’ and ‘duplication’ occurred in semi-automated checks as often or nearly as often as in manual checks.

In the group of labels with 10 or more occurrences (Table 29), the most frequent label assigned was ‘stylistics’, followed by ‘mistranslation’. However, the majority of quality evaluation models benchmarked by O’Brien (2012) treat terminology as a separate category (p.61). When using this approach, ‘stylistics’ stands out with 95 mentions, before a large gap to the next three labels: ‘grammar’ (56), ‘spelling’ (55) and ‘terminology’ (53). It should also be noted that within the ‘mistranslation’ parent label, more than half of the mentions that were manually tagged and nearly all of those found during the semi-automated analysis concerned terminology.
‘Stylistics’ also stands out because while it has the highest number of mentions in the manual coding results, it also has among the least mentions in the semi-automated coding results in relation to manual coding (18.95%). The only label with a lower percentage is ‘register’ (10%) and the next three higher up ‘locale-convention’ (19.23%), ‘untranslated’ (23.08%) and ‘mistranslation (no term)’ (25.81%) when treating terminology.
as a separate label.

On the other end of the spectrum, ‘grammar’ (76.79%) and ‘spelling’ (74.55%) are in the top five of semi-automated coding results in relation to manual coding, only topped by ‘typography’ (100%) and followed by ‘inconsistency’ (70.83%) and ‘terminology’ (39.62%).

Finally, both labels indicating that a translation had not been provided where one was expected (‘omission’, ‘untranslated’) were assigned in a relatively low number of cases.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>stylistics</td>
<td>18</td>
<td>95</td>
<td>18.9 Fluency-Content</td>
<td></td>
</tr>
<tr>
<td>mistranslation</td>
<td>29</td>
<td>83</td>
<td>34.9 Accuracy</td>
<td></td>
</tr>
<tr>
<td>grammar</td>
<td>43</td>
<td>56</td>
<td>76.8 Fluency-Mechanical</td>
<td></td>
</tr>
<tr>
<td>spelling</td>
<td>41</td>
<td>55</td>
<td>74.5 Fluency-Mechanical</td>
<td></td>
</tr>
<tr>
<td>terminology</td>
<td>21</td>
<td>53</td>
<td>39.6 Accuracy</td>
<td></td>
</tr>
<tr>
<td>mistranslation (no term)</td>
<td>8</td>
<td>31</td>
<td>25.8 Accuracy</td>
<td></td>
</tr>
<tr>
<td>locale-convention</td>
<td>5</td>
<td>26</td>
<td>19.2 Fluency-Mechanical</td>
<td></td>
</tr>
<tr>
<td>inconsistency</td>
<td>17</td>
<td>24</td>
<td>70.8 Fluency-Content</td>
<td></td>
</tr>
<tr>
<td>typography</td>
<td>21</td>
<td>21</td>
<td>100.0 Fluency-Mechanical</td>
<td></td>
</tr>
<tr>
<td>untranslated</td>
<td>3</td>
<td>13</td>
<td>23.1 Accuracy</td>
<td></td>
</tr>
<tr>
<td>omission</td>
<td>6</td>
<td>12</td>
<td>50.0 Accuracy</td>
<td></td>
</tr>
<tr>
<td>register</td>
<td>1</td>
<td>10</td>
<td>10.0 Fluency-Content</td>
<td></td>
</tr>
</tbody>
</table>

Table 29: Results of Trommons reviewer comments content analysis - labels with more than 10 mentions

### 7.2.1.2. Evaluating Quality

#### 7.2.1.2.1. Non-profit Clients

The non-profit clients on Trommons are generally relatively small organisations. The majority of organisations have

- between one and 10 members of staff (64.1%),
- 11 to 50 members of staff (23.1%).

The assumption in the design of the study was that the majority of the organisations
would have individuals handling requests via Trommons whose main expertise was not in localisation or translation, which would indicate a great amount of information asymmetry. This assumption was confirmed in the demographic section of the questionnaire, as outlined in section 7.1.3.1. To gain a better understanding of how those clients addressed the information asymmetry, they were asked how they decided on whether they were comfortable with using a translation for their project, assuming they didn't speak the language in question.

11 of the participants provided descriptions. These responses primarily fell into the categories of approaches where they checked the translations themselves, asked others to check the translations or simply trusted the translators. Some of the responses described a combination of the approaches. Three responses did not address the question itself and were not analysed further.

The three participants who described checking the translations themselves either didn't address the topic of not speaking the language themselves, or mentioned that they speak the languages they request translations into:

"If it makes sense when I read it through." [Questionnaire respondent]

"All of us speak the language, even we are not professional or is not our mother tongue we can read the language and understand it, but we are not enough fluent to be able to do proper translations." [Questionnaire respondent]

"I never submit a text in a language that I know nothing of. When reading a translation that seems doubtful, I too can use a dictionary. (...)" [Questionnaire respondent]

4 participants described asking others to check the translations for them, of which one mentioned that the volunteer translation serves "as a basis for the project":

"I send the translation to our coordinating team in the country and have them work over the translation. The volunteer translation serves as a basis for the project." [Questionnaire respondent]

3 participants mentioned they trusted the translators if they couldn’t ask someone to
check the translations for them and one that they either checked the translations themselves or asked someone else:

"I try to ask to some people (who are able to understand the language) to read some part of the text, in cases I don't know people that speak that languages I trust in translators" [Questionnaire respondent]

"At best, we find a partner in the project area who reads the text and is comfortable with the translation, otherwise we trust the Rossetta Foundation and its volunteers." [Questionnaire respondent]

"if we can't control the quality of a translation we use it, hoping it have been done in a good way" [Questionnaire respondent]

"I never submit a text in a language that I know nothing of. When reading a translation that seems doubtful, I too can use a dictionary. I am also fortunate to be able to check out a translation with other people who are native speakers of that language." [Questionnaire respondent]

And in two cases, participants only mentioned trusting the translators:

"I have to trust it's done well but I have never had any issue with the documents when I have used them in relation to cases" [Questionnaire respondent]

"I assume it is good because it has also been proofread." [Questionnaire respondent]

Only two responses indicated more sophisticated approaches to evaluating quality, with one participant describing mechanisms they applied to get an understanding of the volunteer's skills by considering previous work and how they used the shared language of communication:

"Generally, I would have some assurance from previous work with the volunteer and the feel for language that may be apparent from our use of common language. Also, some of our volunteers speak multiple languages." [Questionnaire respondent]

One respondent described a detailed approach that included a set of criteria and a standardized approach to evaluating quality in their workflow. Interestingly, this response
came from the participant who was also the most critical of volunteer translations and gave the only response among this respondent group to indicate that they did not have an overall positive opinion of volunteer translations. In addition, this respondent had indicated they had academic training in a linguistic field and was one of two respondents who selected quality checks to be a main factor in determining quality.

"I ask at least 3 people whose mother tongue it is to review with specific criterias I explain to them, and ask them to rate on a 1 to 5 scale within an excel sheet I designed." [Questionnaire respondent]

The interview participants were also divided on the topic, with OrgClient 1 describing a process where they themselves or someone else who knew the language would test a sample of the translation to look for issues like words missing as well as comparing the length of the original text with that of the translation. OrgClient 2 however indicated that they trusted in the system and that the translators, who they believed to be translators on other projects, in choosing to support the organisation with their work, wouldn't do bad work and that they would need to measure their professional principles themselves.

7.2.1.2.2. Professional Buyers

9 respondents offered an answer to the question of how they decided whether they were comfortable using a translation if they didn't speak the language in question.

1 respondent (11.1%) mentioned ISO certification as their main decision mechanism:

"The decision is before that Point: I only hire ISO 17100 certified Language Service Providers." [Questionnaire respondent]

5 respondents (55.5%) mentioned getting someone to review the translations as their primary decision mechanism. Some mentioned having faith or trusting translators or reviewers:

"Trust in translators, established by regular reviews. In particular cases, check by native speaker." [Questionnaire respondent]
“Feedback from a professional, native speaker and with experience within the system” [Questionnaire respondent]

“Have a native speaker review the translation within context and provide their feedback.” [Questionnaire respondent]

The remaining three respondents (33.3%) described a detailed, at times complex decision process that involved quality assurance and quality evaluation and a variety of techniques like setting quality goals, back translations, proofreading, contextual reviews and functional testing:

“Depending on the content type, language pair, translation workflow, resource selection, previous quality history, and other variables of a translation project, I would have an appropriate quality management program (incl. quality assurance and quality evaluation processes) in place that would ensure adherence to my quality goals with the most lean and economic means possible (and then verify that quality actually translates into business impact).” [Questionnaire respondent]

During the interviews with professional buyers in the industry participants were also asked to describe their mechanisms for quality evaluation and their decision process for trusting a translator or translation. The participants described a set of processes that involved pre- and post-translation mechanisms as well as a long-term perspective.

Pre-translation and long-term strategies primarily referred to the selection of vendors and freelancers. Selection criteria mentioned were the experience an individual had from other companies and projects and whether there had been problems with their work on previous projects as well as their performance during the hiring process, including a test translation, which would be evaluated and commented on by established and trusted resources. Selecting resources was also described as going hand in hand with long-term strategies, involving gathering information after project delivery on end user post-release feedback, feedback from direct clients, tracking bug types and numbers, as well as establishing a trusted in-house team for major languages.

Post-translation strategies involved evaluation based on quality metrics, which ProBuyer 3 described to include
Post-translation evaluation also tied into the long-term strategy with native speakers within the company being asked for their impressions. ProBuyer 1 described a process where nearly all translations were outsourced to external translators, but were then reviewed in-house. This would then result in what the participant described as the "signature" quality of the company.

If issues arise in post-translation evaluation - be it detected on client or vendor side - this would then result in a third party being asked to evaluate a sample and providing feedback on the translation. Based on this feedback, the issues would then be discussed with clients, translators and partners. However, interview participants also mentioned a lack of systematic third-party review due to financial constraints, as well as the question whether any review of translations would take place being subject to budget constraints in some cases. ProBuyer 1 described being asked to provide translations only, without proofreading, for cost reasons, but decided to reject the proposal, stating "we don't provide this type of service. If you want that you know, go somewhere else there are zillions of companies out there" [ProBuyer 1].

7.2.1.2.3. Linguists

As part of the questionnaire, respondents were asked how they evaluate quality if not provided with customer requirements. They were able to select multiple applicable options. The majority (67.5%) said that they used their own assessment criteria, while 22.5% used criteria developed by an individual or organisation other than themselves without adaptation, and 38% had adapted criteria developed by someone else to suit their needs (Figure 31).

However, even though the majority of respondents used their own assessment
criteria, 43.5% thought that for the majority of criteria the majority of reviewers would agree with their change requests if they were both to review the same translation. 17% thought it would be a matter of individual preference and 15.5% were unsure (Figure 32).
Figure 31: Criteria used for evaluating quality in linguist participant group
Figure 32: Expected level to which majority of reviewers would agree with change requests in linguist participant group
27 respondents also provided additional thoughts on this question. 8 responses described that the level of agreement would depend on the translator, reviewer, content or the type of issue:

"It would depend on the quality of the translation. If real errors (grammatical, typos, etc) are found then in general translators accept. However in the case of preferential changes it depends on the translator's outlook and openness to suggestions." [Questionnaire respondent]

"While there are "objective" criteria where probably most reviewers will agree on (spelling, terminology), style is always subject to discussion." [Questionnaire respondent]

9 highlighted the importance of justifying changes well or being careful not to make unnecessary or preferential changes:

"When reviewing, I am careful to only make corrections for actual errors of fact or language. And to back my corrections up with references and linguistic explanations. Of course, if a text is unacceptably full of errors, I will simply pick out some examples of the issues in the text, identify the types of issues, and refuse to continue the review." [Questionnaire respondent]

"There's two ways to review a text: Only change what's necessary and Change everything you don't like. I have clients for both ways. I usually go with the first and change *mistakes*. When asked, I will also change things I do not like." [Questionnaire respondent]

The remaining comments described individual experiences made with the review process, either positive or negative, where positive comments related to being part of a translator-reviewer team and negative comments described the difficulty of getting a translator to accept necessary changes.
The question of evaluation was also discussed with informants during the interviews. A common theme was that translation quality is subjective and that there was commonly disagreement between translators or translator and proof-reader on the topic.

Translator 1 described this subjectivity:

"Well, translation is very subjective - you know - so what I think is a good rendition, another translator might think it's not the best, but I think there's always room for disagreement on subtleties and things that are a matter of preference. But I think on being accurate and reflecting the original meaning and following the rules of the target language, I think everybody would agree on that."

[Translator 1]

Another explanation offered for disagreement was that there are different variants for some languages, where an expression might sound natural in one variant and awkward in another. In addition, subject specific terminology was mentioned as a source for disagreement. However, Translator 1 thought that "everyone would agree" on accurate representation of the original meaning and language rules being followed as areas with general agreement, while Translator 2 thought that the need to be consistent and accurate with terminology would be such an area.

On the question of how to handle such disagreements, Translator 1 suggested that it would depend on whether the client was a direct client or an agency, as well as the importance of being able to justify choices. When working with an agency, the agency's reviewer would be likely to make the final call on quality, while working with a client directly was, according to Translator 1, more difficult, since the client had less linguistic knowledge and knowledge of the language service market.

Finally, interviewees mentioned that while some agencies might increase their prices if proofreading was part of the service, it should be standard to deliver the final end product without needing more changes. In addition, providing feedback to translators was important to help them improve.
7.2.2. Factors Impacting Quality

The section on defining and evaluating quality (7.2.1.) already hinted at some factors that might impact quality, such as the short-term and long-term strategies employed by professional buyers in the localisation industry. In addition, the question of which factors impact quality were posed more explicitly in the questionnaires as well as interviews. The following sections report on the findings related to this topic for the different respondent groups.

In the questionnaires, participants (industry buyers, non-profit clients, as well as linguists) were given a broad list of factors that might impact translation quality and asked to pick the three they thought impacted translation quality most (Figure 33). Following that, respondents were then asked a series of questions that were aimed at understanding the relationship between key factors to distinguish a 'volunteer' and a 'professional' and the importance respondents assigned them in relation to expected quality. Focus was here on payment, experience and formal training as distinguishing factors (Figure 34). In addition, some respondent groups were asked about their theories as to why some translators might work at a lower rate than others.

In a set of three questions, participants (industry buyers and linguists) were presented with different scenarios, without knowing there would be a follow-up question after the one they were currently seeing. For each of the questions, they were asked to rank the scenarios by the quality they'd expect a translator to deliver, assuming that in all cases the translators knew the content equally well. In all three questions, one meant the highest quality that could be expected.

In the first set, four scenarios were given:

- unpaid volunteer;
- translator paid 5 cent per word;
- translator paid 10 cent per word;
- translator paid 15 cent per word.

In the second set, six scenarios were given, with added information on the
translator's experience (5 years vs. 3 months):

- unpaid Volunteer, 5 years’ experience;
- unpaid Volunteer, 3 months’ experience;
- translator paid 5 cent per word, 5 years’ experience;
- translator paid 5 cent per word, 3 months’ experience;
- translator paid 15 cent per word, 5 years’ experience;
- translator paid 15 cent per word, 3 months’ experience.

In the third set, another six scenarios were given, with information on the translator's training being added (formal training vs. no formal training):

- unpaid volunteer, WITH formal training;
- unpaid volunteer, NO formal training;
- translator paid 5 cent per word, WITH formal training;
- translator paid 5 cent per word, NO formal training;
- translator paid 15 cent per word, WITH formal training;
- translator paid 15 cent per word, NO formal training.

To lower the number of options with the introduction of the experience and training variables, sets two and three didn't include the 10 cent per word translator but only the two extremes - 5 cent and 15 cent - in addition to unpaid volunteer.

In the final set of questions, industry buyers were first asked whether they thought that they could tell from the comments of a reviewer whether the translation was expensive or cheap, and then provided with a list of potential reasons why translators would work for a low rate and asked to rank them by importance.

Linguists were asked whether they could tell from a translation whether it was expensive or cheap and were also asked to rank potential reasons why translators would work for a low rate.
Figure 33: Ranking of factors impacting quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.
Figure 34: Impact of factors on quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.
7.2.2.1. Non-profit Clients: Factors Impacting Quality

When asked about factors that impact quality the most, the 17 non-profit clients who completed the questionnaire

- most frequently selected
  - "Translator/reviewer works into their native language" (12/70.6%),
- followed by
  - Subject matter expertise (8/47%),
  - Experience (7/41.2%).
  - "Translator/reviewer level of motivation" was selected 5 times (29.4%), and
  - the source text quality 4 times (23.5%).
- At the bottom of the list with fewest selections were
  - "Translator/reviewer training or education",
  - technology and tools,
  - quality checks,
  - internationalisation and guidelines provided.
- Finally, none of the respondents selected
  - text type,
  - high or low payment and
  - "Time spent on translation" as factors most frequently impacting translation quality (Figure 33).

The relatively high value placed on experience, relative to payment, could also be observed when participants were presented with three aspects:

- Payment,
- Training and
- Experience

and were asked to estimate how important each of them were in their opinion for quality of translation one could expect to receive (Figure 34).
The following ratings were given to the elements as listed below:

- **High impact on quality:**
  - 70.6% experience;
  - 64.7% training;
  - 17.6% payment,
  thus showing a clear majority finding experience and training important and payment coming a distant third.

- **Some impact on quality:**

- **No or nearly no impact on quality:**
  - 35.3% payment;
  - 0% experience;
  - 11.8% training.

Across the conditions, payment is not seen to be a hugely influential factor on quality, but experience is clearly a defining aspect for most respondents.

### 7.2.2.2. Professional Buyers: Factors Impacting Quality

Among the 13 respondents in the professional industry buyer group that completed the questionnaire,

- most frequently, respondents picked 'source text quality' (61.5%),
- followed by
  - 'linguist experience',
  - 'Linguist familiarity with subject matter',
  - 'linguist works into their native language' and
  - 'quality checks performed' (38.5% for each of the four),
  when asked about the factors they thought influenced quality the most.

- In the mid-field ranked:
  - guidelines or information (30.77%), as well as
  - technology and tools (23.1%).

- Respondents selected only once or not at all:
o linguist training or education,
o time spent on translation,
o internationalisation,
o source text type,
o linguist motivation and
o linguist high or low payment (Figure 3).

When asked to indicate whether payment, formal training and experience had high impact, some impact or no/nearly no impact on quality, all but one respondent rated 'experience' a high impact factor, and most respondents placed 'payment' and 'formal training' in the 'some impact' category (Figure 34). While the importance placed on experience is consistent with the responses to the previous question, respondents seem to rank formal training and payment higher when asked about it directly than when these factors are provided as part of a longer list.

In the first question, participants ranked translators according to the quality they would expect to receive, going by the price of each translation (Figure 35). From the results, a correlation can be seen between the price of a translation and the quality participants would expect.

- Most frequently, 15 cent per word is ranked in position 1, 10 cent per word in position 2 and 5 cent per word in positions 3 and 4.
- The 15 cent per word price point was only selected once in position 4 and not at all in position 3.
- The 5 cent price point was selected not at all in position 1 and only once in position 2, showing a strong symmetry in the results for the paid translators.

However, the results for the unpaid volunteer are less clear. Participants assigned the unpaid volunteer to positions 1 and 4 in equal numbers and tended towards position 3 in the majority. From this it would appear that there is overall less agreement within the group of respondents what quality to expect from a volunteer as against that from paid translators.
Figure 35: Quality rankings based on price for professional industry buyer group
The second scenario detailed information about the price of translation and the translators’ level of experience (Table 30; Figures 36 to 38).

- The translator with 5 years of experience and paid 15 cent per word featured nearly exclusively in positions 1 and 2.
- Translators with three months of experience paid 5 cent per word and volunteer were primarily spread out over the bottom three positions.

From here, the results are more unexpected. Price is given high prominence relative to experience in the following cases:

- The translator paid 15 cent per word with three months of experience ranked in all five positions, equally in positions 2, 4 and 5.
- The translator paid 5 cent per word with 5 years of experience primarily ranks in position 3, with some selections in positions 2 and 4, but not in position 1 at all.
- The volunteer translator with 5 years of experience ranks primarily in positions 1 and 2, with some selections in positions 3 and 4.

This shows that while participants were less certain about the quality to expect from a volunteer, even when provided with the years of experience information, the trend was towards a higher quality expectation from a volunteer translator with 5 years of experience compared to a translator paid 5 cent per word with the same experience, while for the three months of experience no significant difference between a volunteer and translator paid 5 cent per word can be observed. It should be noted that when the translator was paid 15 cent per word but only had three months of experience, the expected quality ranking seemed to be slightly higher than even for a translator paid 5 cent per word but with 5 years of experience.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Pos. 1</th>
<th>Pos. 2</th>
<th>Pos. 3</th>
<th>Pos. 4</th>
<th>Pos. 5</th>
<th>Pos. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>unpaid Volunteer, 5 years experience</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>unpaid Volunteer, 3 months experience</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>translator paid 5 cents per word, 5 years experience</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>translator paid 5 cents per word, 3 months experience</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>translator paid 15 cents per word, 5 years experience</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>translator paid 15 cents per word, 3 months experience</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 30: Ranking of payment vs. experience for professional industry buyer group
Figure 36: Ranking of voluntary translation with 5 years vs. 3 months of experience for professional industry buyer group
Figure 37: Ranking of 5 cent per word translation with 5 years vs. 3 months of experience for professional industry buyer group.
Figure 38: Ranking of 15 cent per word translation with 5 years vs. 3 months of experience for professional industry buyer group.
Similar patterns were observed when participants were given information on whether the translator had received formal training or not, although slightly less pronounced (Table 31; Figures 39 to 41), indicating that the translator's experience may have slightly more influence than formal training.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Pos. 1</th>
<th>Pos. 2</th>
<th>Pos. 3</th>
<th>Pos. 4</th>
<th>Pos. 5</th>
<th>Pos. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>unpaid volunteer, WITH formal training</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>unpaid volunteer, NO formal training</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>translator paid 5 cents per word, WITH formal training</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>translator paid 5 cents per word, NO formal training</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>translator paid 15 cents per word, WITH formal training</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>translator paid 15 cents per word, NO formal training</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 31: Ranking of payment vs. training for professional industry buyer group

The majority of respondents thought they could tell from the comments of a reviewer whether the translation was expensive or cheap:

- always (14.3%) or
- in most cases (42.9%).

A little over one-third of the respondents said that they didn't think they could tell whether the translation was expensive or cheap from a reviewer's comments (Figure 42).
Figure 39: Ranking of voluntary translation with formal training vs. no formal training for professional industry buyer group
Figure 40: Ranking of 5 cent per word translation with formal training vs. no formal training for professional industry buyer group
Figure 41: Ranking of 15 cent per word translation with formal training vs. no formal training for professional industry buyer group
Figure 42: Ability to estimate price of translation based on reviewer comments for professional industry buyer respondent group
For this respondent group, when asked to rank a list of why a translator might work for a low price (Table 32), no clear pattern emerged, could have been due to the overall low number of responses for this group. In two cases, the results were clear:

- when the translator has not done good work and has lost clients, and
- the translator not having good tools.

Both of these reasons were in the top three for positions 7, 8 and 9. For all other explanations provided, only weak tendencies can be observed. The options: ‘translator not being well trained/qualified’ or ‘being new to the field and wanting to gain experience’ trended towards the upper positions, while the translator being naive and not understanding that the rates are too low tended towards the lower ones. Other reasons seem to be relatively evenly distributed or trending towards the middle positions.

The perceived relationship between quality and pricing was also discussed with the professional industry buyers during the interviews. Interviewees agreed that there was a link between price and quality, but also that the correlation wasn't always present since other factors also played a role. Informants also commented on having observed expensive translations that were low quality on occasion.

Some of the reasons respondents suggested as to why prices in localisation might vary, were: the language combination; the subject matter; and the country of residence of the translator. However, the country of residence was also linked to potential quality issues with there being a danger that the translator would gradually lose their native language ability.

Informants also suggested a range of reasons why some translators might accept lower prices than others that could indicate potential quality issues in the end results. Those were that translators might not always know the rates: they needed the job; or they needed to gain experience to enter the market. Two of the interviewees also described ending business relationships with vendors or freelancers who didn't deliver translations at sufficiently high quality, independent of whether they had been paid a high or low rate.
Table 32: Ranking of reasons why a translator might ask for low rates for professional industry buyer group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9.

<table>
<thead>
<tr>
<th>Reasons for low cost translation</th>
<th>Pos. 1</th>
<th>Pos. 2</th>
<th>Pos. 3</th>
<th>Pos. 4</th>
<th>Pos. 5</th>
<th>Pos. 6</th>
<th>Pos. 7</th>
<th>Pos. 8</th>
<th>Pos. 9</th>
<th>Weighted avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>... is new to the field and wants to gain experience</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>... works in a field or language pair where rates are usually low</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>... is not well trained/qualified</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>79</td>
</tr>
<tr>
<td>... works in a country where cost of living is low</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>79</td>
</tr>
<tr>
<td>... plans to work quickly without much care to justify the low rate</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>... urgently needs money and is working extra hours</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>... is naive and doesn’t understand that the rates are too low</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>42</td>
</tr>
<tr>
<td>... doesn’t do good work and already lost clients</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>... doesn’t have good tools so doesn’t have to cover an investment</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>34</td>
</tr>
</tbody>
</table>
However, informants also stressed that it was difficult to find good freelancers and that it was therefore important to pay the good ones a little more money to secure their services. Even though there were a few good translators who worked at low prices, if they got paid too little there was a danger that they would have to take too much work or work for which they didn't have sufficient subject matter knowledge, as a result of which quality might suffer. In addition, ProBuyer 3 observed that the best translators usually work at a reasonable rate, not too high and not too low.

Similarly, a vendor would review translations depending on the budget and clients would request translations without proofreading in order to lower cost. One of the reasons offered by ProBuyer 1 for this was that some clients see localisation as unnecessary cost and that they were also taking a risk because they didn't know whether their product would be successful. In addition, clients were often not aware of the importance of quality translations. These clients looking for low-cost localisations would then be able to find vendors in the group that tended to aim for the low-end of the market to secure this kind of business, some of whom were aiming to sell translations at low cost to secure business for additional services like audio recordings.

7.2.2.3. Linguists: Factors Impacting Quality

Among the 197 respondents in the linguist group that completed the questionnaire, the three most frequently picked factors impacting translation/localisation quality were:

- Linguist experience (69.5%),
- Linguist subject matter expertise (56.5%), and
- Linguists working into their native language (38.5%).

This was followed by:

- Linguist training or education in position 4 with 30% and
- Source text quality in position 5 with 26%.

Surprisingly, linguist high or low payment ranked in the bottom three, with only 5% selecting this as a factor impacting quality the most and only three of the participants selecting Internationalisation (Figure 33).
This trend was also confirmed when participants were asked to indicate whether Payment, Experience and Training would have

- 'High impact on quality',
- 'Some impact on quality' or
- 'No or nearly no impact on quality' (Figure 34).

The only option that was selected as having high impact by the majority of respondents was 'Experience' with 86.9% of the participants selecting 'high impact' and 12.2% 'some impact', while only 1% selected 'no or nearly no impact'.

The majority of participants indicated that they thought payment had 'some impact', with 30.5% saying it had 'high impact' and 10.7% thinking it had 'no or nearly no' impact on quality.

Training was overall rated slightly higher with a similar number of respondents indicating that training has 'high impact' (49.7%) on quality and those assigning it only 'some impact' (40.6%).

19 respondents selected the 'other' option and re-introduced other concepts to impact quality, such as motivation, subject-matter knowledge, time spent on translation and whether the translator is translating into their native language.

Respondents were also given the opportunity to add their thoughts to this question, where 45 respondents elaborated on their selections.

Comments on the relationship between payment and quality provided several reasons as to why quality would be impacted by payment. One theme that emerged was that the individual translator would not deliver better or worse quality depending on payment, but that it would be less likely for a good translator to work at a lower rate, which would result in jobs with less payment only being accepted by translators who would not be likely to be able to provide a good translation:

"Rates issue is a vicious circle which can impact quality since low paying agencies tend to hire low trained/experienced translators." [Questionnaire respondent]
"I chose "some impact on quality" for payment/rates because the lower the rates, the less likely the buyer is to attract a good translator, not because the lower the rate, the sloppier my own work. (...)" [Questionnaire respondent]

Other respondents felt that paying low rates would negatively impact the individual translator’s motivation or ability to provide a good translation, while some felt that the individual translator’s work ethic mattered the most, independent of payment:

"(...)Most of the time if you pay peanuts you get garbage, but a few will give real quality. Personally, I simply can’t reduce my level of quality: if I agreed to a peanut-paying job, I’d deliver quality. (...)" [Questionnaire respondent]

"(...)In regards to rates, these shouldn’t impact on quality as they are agreed on beforehand and one should strive to do one’s utmost regardless, but in my experience when the rate is quite low, the incentive to go the extra mile is somewhat lacking." [Questionnaire respondent]

"I think there are some people who will generally work hard only when they feel they are compensated adequately but there are those who have strong work ethic and work to please the client. (...)" [Questionnaire respondent]

"It is difficult to say with a great degree of certainty whether or not payment directly affects translation/localisation quality. I think it all really depends on the individual translator: his/her motivations, work ethic, sense of integrity and sense of pride attached to his/her work." [Questionnaire respondent]

On the topics of experience and training, the general theme emerging was that it again depends on the individual translator. However, some respondents placed more importance on formal training than others. Some thought that formal training was an important starting point upon which to build with experience, others thought that even with experience training was necessary for a good translator:

"(...) Formal training matters not a jot once you have solid experience; but it’s important to start with at least the basic techniques." [Questionnaire respondent]
"As regards formal training, the cases of really GOOD translator without previous studies exist, but they are too few exceptions. Training is important to cover the linguistic basics we need." [Questionnaire respondent]

Others acknowledged that formal training might be useful for a translator but thought that experience and subject matter expertise played a larger role:

"Formal training is important. But education in the field of specialization is just as important as linguistic education and both must be ongoing." [Questionnaire respondent]

"(...) Also, in this particular field, one can get abilities in both linguistic and subject matter aspects in other ways beyond formal training (although for some subject matters, as in technical, some level of formal training is highly desired)" [Questionnaire respondent]

To further understand what reasons respondents attributed to individual translators working at low rates, they were provided with a list of potential reasons and asked to rank them by importance (Table 33).

Reasons outside the individual translator's intentions or abilities generally ranked in the upper to upper-medium positions, with 'the translator working in a country where cost of living is low' being in the top three for positions 1 to 4 and 'the translator working in a field or language pair where rates are usually low' in top three for positions 2, 4 and 5. Additionally relatively high numbers of respondents ranked it in positions 1 and 3.

Also at the top of the list are 'lack of training or qualification' in top three for positions 1, 3 and 4 as well as 'the translator being new to the field and wanting to gain experience' in positions 1 and 2. In contrast, those reasons that imply that the translator knowingly intends to cut corners or generally doesn't provide good translations without further reasoning ranked towards the lower to lower-middle positions with 'doesn't do good work and already lost clients' being in the top three for positions 5, 6, 7, 8 and 9 while 'plans to work quickly without much care to justify the low rate' ranked in the top three for positions 6 to 8. More firmly at the bottom of the list are 'the translator not having good
tools’ and ‘the translator being naive, not understanding that the rates are too low’. Finally, the ‘translator urgently needing money’ is spread out in the top three across positions 3, 5 and 6. However, it should be noted that in some cases ranking into the top three for each of the positions was based on the difference between only few respondents selecting a category or not. Nonetheless, the general trends outlined above can be observed.

Again, respondents were given the opportunity to comment on their choices after the question. 31 took this opportunity, of whom 10 remarked on the question itself saying that they had found it difficult to rank the reasons, either because too many applied or because some didn't apply at all in their opinion. 7 remarked that the translators didn't set the rates but had to go along with what the market dictates, while 4 said that the list presented was missing the option that translators would work just for fun or as a hobby. The remaining responses repeated aspects that had been presented in the list.

Participants were then presented with three sets of scenarios and asked to rank the scenarios according to the quality they'd expect to receive from each translator.

Figure 43 shows the results of the first set, where no additional information apart from price points was provided. Here a strong majority of participants placed

- 15 cent per word in position 1, and
- 10 cent per word in position 2, with
- 20% of the 185 respondents to this question placing 10 cent in position 3.
- 5 cent per word was primarily ranked in positions 3 and 4,
- as was the unpaid volunteer.

However, it should be noted that while the unpaid volunteer was placed nearly as often in position 4 as 5 cent per word on position 3, more participants placed the volunteer translator in positions 1 and 2 (combined 22%) than the 5 cent translator (combined 8%).
Table 33: Ranking of reasons why a translator might ask for low rates for linguist respondent group. Weighted average assigns highest weighting to position 1, with subsequent positions being weighted less down to a weighting of 0 in position 9.
Figure 43: Quality rankings based on price for linguist group
When presented with the translator's experience (5 years vs. 3 months) in addition to price points (Table 34; Figures 44 to 46), the dynamics for the 15 cent price point with 5 years' experience remained roughly the same as it had been with no information on experience provided, with the 15 cent translator losing about 5 percentage points from position 1, which were distributed across positions 2 and 3. With the added information on the translator's 5 years' experience, opinions shifted more strongly for the volunteer translator and the translator at the 5 cent price point, with only 19.5% placing the volunteer and 23.2% the 5 cent translator in the bottom three positions, suggesting that both these groups were viewed more favourably than when no information on the translator's experience was provided.

When the information was added that the translator had three months of experience,

- most frequently participants placed the 15 cent translator in position 4,
- the 5 cent translator in position 5 and
- the volunteer translator in position 6.

The volunteer was placed in the top three positions by 8% of participants, the 5 cent translator by 8.1% and the 15 cent translator by 34.6%, which means that nearly half as many respondents placed a 15 cent translator with three months’ experience in a favourable position similar to that of a 5 cent translator with 5 years’ experience (75.7%) or a volunteer translator with 5 years of experience (78.8%). This suggests that even with the same amount of experience or less, a translator paid 15 cent per word is believed to deliver better quality than a translator paid 5 cent per word or a volunteer, with the volunteer being viewed slightly less favourably than the 5 cent translator.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Pos.1</th>
<th>Pos.2</th>
<th>Pos.3</th>
<th>Pos.4</th>
<th>Pos.5</th>
<th>Pos.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volunteer, 5 yrs experience</td>
<td>23</td>
<td>70</td>
<td>53</td>
<td>14</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>Volunteer, 3 mon experience</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>32</td>
<td>52</td>
<td>85</td>
</tr>
<tr>
<td>5c p. w., 5 yrs experience</td>
<td>4</td>
<td>65</td>
<td>71</td>
<td>20</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>5c p. w., 3 mon experience</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>36</td>
<td>76</td>
<td>56</td>
</tr>
<tr>
<td>15c p. w., 5 yrs experience</td>
<td>148</td>
<td>18</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>15c p. w., 3 mon experience</td>
<td>2</td>
<td>23</td>
<td>39</td>
<td>77</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

*Table 34: Ranking of payment vs. experience for linguist respondent group*
Figure 44: Ranking of voluntary translation with 5 years vs. 3 months of experience for linguist respondent group
Figure 45: Ranking of 5 cent per word translation with 5 years vs. 3 months of experience for linguist respondent group.
Figure 46: Ranking of 15 cent per word translation with 5 years vs. 3 months of experience for linguist respondent group.
Similar results can be observed when participants were provided with information on whether the translator had received formal training in addition to price points (Table 35; Figures 47 to 49). When the translator had received formal training, again the 15 cent price point was placed in first position by the majority of respondents, and opinions shifted to be more favourable for the 5 cent and volunteer translators.

As with the results obtained in the previous scenario, here too the volunteer translator was placed in 6th position, the 5 cent translator in 5th position and the 15 cent translator in 4th position most frequently. In the top three positions, the volunteer was placed by 12.4% of the participants, the 5 cent translator by 4.8% and the 15 cent translator by 42.7%. By comparison, 5 cent translators with formal training were placed in the top three by 68.6% of the participants and volunteer translators, illustrating the relatively high value placed on payment even when formal training or its absence is a factor to be considered.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Pos. 1</th>
<th>Pos. 2</th>
<th>Pos. 3</th>
<th>Pos. 4</th>
<th>Pos. 5</th>
<th>Pos. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volunteer, WITH formal training</td>
<td>24</td>
<td>55</td>
<td>64</td>
<td>12</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>Volunteer, NO formal training</td>
<td>3</td>
<td>14</td>
<td>6</td>
<td>33</td>
<td>29</td>
<td>98</td>
</tr>
<tr>
<td>5c.p.w., WITH formal training</td>
<td>3</td>
<td>52</td>
<td>72</td>
<td>24</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>5c.p.w., NO formal training</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>39</td>
<td>80</td>
<td>58</td>
</tr>
<tr>
<td>15c.p.w., WITH formal training</td>
<td>147</td>
<td>14</td>
<td>15</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>15c.p.w., NO formal training</td>
<td>8</td>
<td>46</td>
<td>25</td>
<td>75</td>
<td>12</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 35: Ranking of payment vs. training for linguist respondent group
Volunteer with vs. no formal training; Linguists

Figure 47: Ranking of voluntary translation with formal training vs. no formal training for linguist respondent group
Figure 48: Ranking of 5 cent per word translation with formal training vs. no formal training for linguist respondent group.
Figure 49: Ranking of 15 cent per word translation with formal training vs. no formal training for linguist respondent group
Figure 50: Ability to estimate price of translation based on reviewer comments for linguist respondent group
To understand how aware participants were of their potential bias towards high-price translations, they were also asked to estimate whether they could tell from a translation whether it was expensive or cheap (Figure 50). Here, 58% of the respondents said that they would be able to do so in most cases or always, while 33% said not in most cases and 7.5% never.

Factors influencing quality, in particular the price of a translation, were also discussed with interviewees in the linguist group.

While the linguists agreed that there was a link between price and quality, some saw a stronger correlation than others, with opinions ranging that there "absolutely" was a correlation to that "it depends" and "it's like in all other things in life".

Translator 1 argued that at a very low rate, there would be no expectation of a high-quality translation.

Reasons offered for price differences were, on the translator side, that translations that require technical knowledge or specialised expertise, experience and qualifications would come with a higher price tag, but that the quality would also be better. Translator 2 also thought that customers would be willing to pay more money for a translation if they knew they would be getting a good service.

In addition, Translator 4 argued that low rates would impact quality due to translators not being able to invest sufficient time in the project.

Other reasons given for lower rates were the market in different countries, as well as translators not knowing the market well enough to set appropriate rates.

Translator 1 also described the role agencies and clients play in the relationship between price and quality. On the agency side, an agency getting paid more would likely have higher criteria for translator selection and be able to afford an in-house proof-reader and second proof-reader. In addition, agencies who sell translations at low rates were in the interviewee's opinion able to do so because they pay translators very little, especially in language combinations with high competition rates. They could do this also because they commonly dealt with clients that didn't have high expectations, may not know if a translation was good or bad or weren't overly interested in the quality aspect of translation.
7.2.3. The Language Professional and Quality

In addition to more specific factors, like price, training and experience, that could impact quality, questionnaire respondents were also asked about their thoughts on the relationship between an individual being a language professional and quality. This section was included due to the observation from reviewing the literature (Section 4.1.3.1.) where the term 'professional' translator was used as a short hand to describe an individual in contrast to 'volunteer' or 'crowd'.

For this section, respondents were first asked to answer the question: "Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?". The possibilities to answer were: "Yes, in most cases", "Yes, in some cases", "Only sometimes", "No, not usually" and "I don't know/skip" (Figure 51).

Participants were intentionally not given a definition of what was meant with a professional translator. Instead, after this question, respondents were also asked to describe what in their own words defines a localisation or translation professional (Figure 52). This question was placed after the opinion question on whether a translator being a professional was a deciding factor in quality in order to get the most impulsive, initial responses prior to asking the participants to describe and thereby think about the concept of a translation professional in more detail.

Since responses to this question resulted in unstructured, text-based data, this data was coded by the researcher into themes that emerged from the responses.

After coding by the researcher, a random sample that included 20% of the lines was coded by a second coder in order to evaluate inter coder agreement (Scott 2009, p.347). As a first step, disagreements in coding were discussed between coder one and coder 2, and resolved as far as possible. This discussion led to some categories being merged, others eliminated and for some, definitions were improved. After this process, themes assigned and their definitions were as shown in Table 36.

The coding was then adapted and re-checked by the researcher, before another random sample (15%) was coded by coder 2. Based on this second round of coding, average inter-coder reliability was 0.989 (Cohen's Kappa) and deemed sufficient for the
purposes of this study. A value of 0.8 could be seen as strong agreement for Cohen’s Kappa (Pfeil and Zaphiris 2010, p. 13).
Figure 51: Link between professional status and quality across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.
<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>Mention of client/customer expectations, needs or relationship</td>
<td>Is able to produce an accurate text, using appropriate terminology as well as taking account of client requirements.</td>
</tr>
<tr>
<td>CultureKnowledge</td>
<td>Mentioning knowledge of culture</td>
<td>It's a person who has a perfect knowledge of the other language and of it culture</td>
</tr>
<tr>
<td>Experience</td>
<td>Mentioning experience with translation/localisation</td>
<td>Qualification, work experience</td>
</tr>
<tr>
<td>Gets Paid</td>
<td>Mentioning finances, livelihood or generally getting paid either explicitly or implicitly by referring to aspects like negotiations, paying taxes or rates.</td>
<td>A person who works as a translator on a regular basis and charges for it.</td>
</tr>
<tr>
<td>LanguageKnowledge</td>
<td>Mentioning linguistic knowledge, either general knowledge of language or specific aspects, linguistics and mastery of the language</td>
<td>It's a person who has a perfect knowledge of the other language and of it culture</td>
</tr>
<tr>
<td>Other</td>
<td>Aspects that didn't fit any other code</td>
<td>n/a</td>
</tr>
<tr>
<td>TechKnowledge</td>
<td>Mentioning knowledge of technology, technical skills or techniques for translation and localisation</td>
<td>A person who has studied not only languages, but translation techniques to deliver the same message and effect than the original text</td>
</tr>
<tr>
<td>Training</td>
<td>Mention of any type of translation/localisation related studies, whether formal, informal or unspecific.</td>
<td>Qualification, work experience; A translator who knows how and where to investigate new terms and who is willing to learn new things.</td>
</tr>
</tbody>
</table>

Table 36: Common labels emerging from definitions of ‘language professional’ status across all groups
Figure 52: Aspects defining a language professional across all participant groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.
7.2.3.1. UAX: Language Professional

In the students and lecturer group, 60.4% of the respondents selected that in most cases the translator being a professional was a deciding factor in the quality of a translation, 27.1% thought this was the case in some cases and only a combined 10.5% selected that the translator being a professional was only sometimes or not usually a deciding factor in the quality of the translation (Figure 51).

Responses were also correlated with the level of experience (Figure 53).

"Yes, in most cases" was selected

- primarily by lecturers (11) and
- the ‘Trained’ group of 3rd and 4th year students (14).

Novice translators (1st and 2nd year students) were less certain of the connection between quality and whether a translator was a professional.

When asked to describe the characteristics of a language professional, 48 respondents provided a description (Figure 52). The majority of respondents offered one (15) or two (26) aspects to define what described a translation or localisation professional, while 7 respondents outlined three or 4 aspects.

The most commonly mentioned aspect was language knowledge. In second position were mentions of training, with more than half of the mentions explicitly specifying a form of formal training. Experience and technical knowledge ranged in the mid-field and culture knowledge as well as payment came in at the bottom. None of the respondents mentioned client relationships.
Figure 53: Link between professional status and quality for student and lecturer UAX respondent group, break-down by sub-groups
7.2.3.2. Non-profit Clients: Language Professional

In the non-profit clients group, only 11.8% of the 17 respondents thought that a translator being a professional was a deciding factor of translation quality in most cases, 52.9% thought this was true in some cases and 35.3% thought that it was not usually a deciding factor (Figure 51).

12 respondents offered a definition of what a professional was in their opinion (Figure 52). This group most commonly mentioned language knowledge, followed by training and technical knowledge. Experience, culture knowledge and relationship with clients ranked towards the bottom of the list, while payment wasn't mentioned by any of the respondents.

7.2.3.3. Professional Buyers: Language Professional

When asked whether they thought that someone being a professional translator or not was a deciding factor in the quality of a translation one can expect to receive (Figure 51), the majority of the 13 professional buyers thought that this was true in most cases (42.9%) or in some cases (42.9%).

On the question of what defines a language professional, 9 participants provided an answer and offered a variety of perspectives with broad themes including the technical knowledge of an individual (knowledge of technology, processes and techniques, experience and training). Less commonly, participants mentioned payment, culture knowledge and client relationships (Figure 52). In addition, other factors mentioned were: utilising localisation services, soft skills, knowledge of target markets, appreciation of the complexity of localisation and subject matter expertise were mentioned.

7.2.3.4. Linguists: Language Professional

In the linguist group, the majority of the 197 respondents agreed that someone being a professional translator was a deciding factor in the quality of a translation one could expect to receive, with 53.5% thinking this was true in most cases and 32.5% in some cases (Figure 51).
129 provided descriptions of the factors that define a language professional (Figure 52). Most commonly mentioned were training, language knowledge and experience. Mentions of the individual getting paid or providing translations as their full-time job as well as technical knowledge ranged in the mid field, followed by client relationships and culture knowledge.

In addition, respondents mentioned

- some variations on the theme of dedication, motivation or passion for the profession,
- subject-matter knowledge,
- skills outside the core activities of a translator like business, marketing or soft skills, as well as
- professional ethics and the translation community,
- recognition from peers and
- being a member of a professional body.

### 7.2.4. Volunteer Translation and Crowdsourcing

The fourth topic covered by the questionnaires (non-profit clients, professional buyers, linguists) was aimed at uncovering attitudes towards volunteer/crowdsourced translations. For this questionnaire, a high-level approach was taken, treating volunteer translations, crowdsourcing, community translation and social translation as synonymous to get the respondent's initial reactions to the concept as a whole.

The first question asked was: "What is your general view on volunteer translations (also called crowdsourcing, community translation or social translation)?". Respondents could choose between

- 'I think it's overall positive','
- 'I think it's overall negative' and
- 'No opinion/neutral'.

Respondents were also given the opportunity to comment on this question.
Participants were then asked to describe the average volunteer translator before being given different options on which factors might motivate someone to volunteer their time and translate for free, and they were asked to pick which they felt were the three main motivations (Figure 54).
Figure 54: Ranking of reasons individuals might have for providing voluntary translations, all respondent groups. Percentages given are relative to the number of participants for each individual group, not the total number of participants across all groups.
7.2.4.1. Non-profit Clients: Volunteer Translation and Crowdsourcing

When asked about their general view of volunteer translations, crowdsourcing, community translation or social translation, 94.1% of the 17 non-profit clients selected that they thought they were overall positive, and only one respondent selected No opinion/neutral. None of the respondents indicated that they thought volunteer translations were overall negative. Some of the respondents also added a comment, with those that thought volunteer translations were generally positive including:

"I usually have the translations done by Trommons volunteers checked in the country where the materials will be used for training. This is usually a better way to check for local quality." [Questionnaire respondent]

"Over my many years I have done a lot of translations (written and verbal) and I was delighted to hear about the services of Trommons. I have already, within the last 12 months, submitted 4 texts for translation and I have found 3 of them almost perfect. One needed a good amount of re-working but the other 3 were almost perfect." [Questionnaire respondent]

However, the participant who had selected "No opinion/neutral" was more cautious with their assessment of volunteer translations:

"too many people think speaking a language means you can translate/interpret from and to it. This is simply not true. At the same time I think a poor quality TT from a volunteer translator is better than google translate." [Questionnaire respondent]

When asked to describe the average volunteer translator, 15 respondents provided a description, of which 8 were brief phrases or single words ranging from "I have never met one" to "Good", with half of them being positive descriptions.

The remaining 7 responses were longer descriptions, expressing not only the value that the volunteers provided to the organisations, but also an assessment of the quality and the volunteer's background. These ranged from very positive:
"Are quality translations that help us a lot in our work, as NGO" [Questionnaire respondent]

"As am member of an Organisation I think Rosetta Foundation helps a lot to carry out our tasks in a productive, efficient and fast way. There are many volunteers and not only translation is done also the revisions are really important for us, in order to have a confirmation of the translations done. It's really useful for NGO that have no high amount of money, specially the small ones for those tasks." [Questionnaire respondent]

"The average translation volunteer seems to be well versed in the languages used (in my case from French to English). The quality of the work received was excellent for most of my texts. The time taken to do the tasks was surprisingly short. I would love to be able to know who did what for me." [Questionnaire respondent]

To mixed:

"Most of the translator-volunteers who have done work for our organization have been very good in quality of work as well as in meeting deadline dates. Just a few volunteers have taken projects and then rejected them or just didn't respond after the deadline." [Questionnaire respondent]

"Most of the time the translation is of poor quality but sometimes it is great. Combining volunteer translation with an in-house or qualified person that can do a good proofreading works." [Questionnaire respondent]

Two respondents focused on describing the translators themselves rather than the value of their contribution or the results of their output. However, one of these felt unable to generalise, while the other had a generally favourable opinion about translators:

"There is a variety of volunteers, and it is difficult to describe an "average", as people can usually be assessed along technical skills and reliability." [Questionnaire respondent]

"engaged, motivated, social, intelligent - this is what I assume" [Questionnaire respondent]
These responses indicate that participants had an overall favourable opinion of volunteer translators and appreciated the value their contributions provided to their organisations, while also being more focused on the actual contributions than theories around the volunteers themselves.

Figure 54 shows the results of the question on motivations for volunteering. Most frequently, participants thought volunteers wanted to do good and help other people (15/88.2%) and wanted to get experience with translation (13/76.5%), while in the mid-range they selected that they thought translators want to be affiliated with a non-profit organisation for career purposes (7/41.2%) and want to contribute to a creative effort they are passionate about (7/41.2%). At the bottom were caring for a specific cause (3/17.6%), wanting to learn new things (2/11.8%), being on a break and bored (1/5.9%), not being a very good translator and needing to improve (1/5.9%) and finally, enjoying translation as a hobby which none of the participants selected.

Overall, there was a near equal split between altruistic or cause-driven motivations and those that are aimed at gaining experience, learning and improving career opportunities while those motivations that relate to translation for pleasure were selected very rarely.

The interviews with non-profit clients on Trommons also included the overall attitude towards volunteer translations and thoughts on who the volunteers were.
OrgClient 1 described the quality of the translations as good overall, but that there had been some translations that were not so great. In this case, the organisation resolved the issue by either re-contacting TRF or finding a new translator independent of TRF/Trommons to fix the issue. They described trying to re-contact TRF as difficult at times because the staff there then needed to get in touch with the translator, since the interviewee and their colleagues weren't necessarily in touch with the translator directly. Another aspect was that there was a general feeling of uncertainty ahead of the translations.

OrgClient 2 overall also thought of volunteer translators in a positive way, describing volunteers as driven by principles and therefore likely to do good, professional work since they did the work on a volunteer basis. This interviewee also described quality as a result of individual commitment and not a result of whether they get paid. They also described trusting that the quality of the translator is good without having to review them an important factor, with the trust being based on having found out how TRF work, their large pool of volunteers and that translators did the work voluntarily.

7.2.4.1.1. Describing the average volunteer

OrgClient 1 thought that translators volunteer to keep up their language or translation skills, and that they would want to have a marketable skill for a job so they were able to say they had translated as a volunteer. When asked who the informant thought the volunteers were, they weren't sure and speculated they might be students or other professionals who happen to be bilingual. Overall, a relatively big concern for this informant was the anonymity of working with Trommons, therefore not being able to know who would pick up a translation.

OrgClient 2 thought that volunteers were mostly professional translators, though some might also be enthusiastic amateurs who really like languages and want to work on translations. This informant thought that volunteers have a skill which they wish to use to give time to a cause:
"And I think it's something that they understand better than the rest of us; how important is to have grant access to people in other parts of the world." [OrgClient 2]

7.2.4.1.2. Reasons not to ask volunteers

OrgClient 1 then went on to describe that the main reason they wouldn't request a translation via Trommons was if they were under time pressure due to urgent deadlines and if confidentiality was a concern for their project, which is commonly the case. The interviewee also described a situation where TRF had been able, and very helpful, in finding translators who were willing to sign a confidentiality agreement and then work on such projects exclusively. However, the informant also thought that with the move from the system TRF had used before Trommons, or e-mailing TRF directly, this would not be a possibility anymore, and describes some misgivings with Trommons:

"It was definitely more personal when Rosetta was emailing because you thought okay like there's somebody connecting you" [OrgClient 1]

"Whereas the platform is kind of like you just kind of upload the project, hope somebody picks it up" [OrgClient 1]

OrgClient 1 also described that apart from Trommons, their organisation also sources translations among law students who are bilingual, and other volunteers, such as people they already know and trust because they have done work for them before, or by asking around for referrals. However, they also said that they mostly rely on Trommons for translations but review the translations separately.

OrgClient 2 however thought that the decision whether to ask a volunteer would be founded primarily on whether the organisation had resources to invest in a translation project, emphasising the foundation of trust between volunteers and organisations that also included the organisations not exploiting volunteers. While generally small errors would be acceptable for this client, the focus would be on expressing ideas and meanings correctly. Therefore, for very important projects this informant might request a translation or review
from someone they have worked with before due to their knowledge in the field and knowledge of communication conventions.

7.2.4.2. Professional Buyers: Volunteer Translations and Crowdsourcing

78.6% of the 13 professional buyer respondents thought that volunteer translations were overall positive. Only one respondent thought it was overall negative, and one respondent had no opinion/was neutral (Figure 55).

8 respondents also provided additional comments. These included overall positive remarks or those which were cautiously positive:

"Crowdsourcing is generally quite positive, with enthusiastic contributors."
[Questionnaire respondent]

"I think it is positive when the help provided is assisting a humanitarian cause"
[Questionnaire respondent]

As well as those that those that were mostly cautious, indicating that better ways of working with crowdsourced translations need to be found:

"My general feeling is that community/social translation doesn't have the same level of quality control as professional services and can lead to inconsistent results. That said, I've never personally used the services of a non-professional translator." [Questionnaire respondent]

"We need more of it and better ways to organize it and manage it"
[Questionnaire respondent]
Figure 55: Overall view of translation crowdsourcing for professional industry buyer respondent group
7.2.4.2.1. Describing the average volunteer

9 respondents offered a description of the average translation volunteer. More than half of these responses described volunteer translators as overall motivated activists, who were passionate about volunteering.

"Engaged activist with range of interests" [Questionnaire respondent]

"Not a skilled linguist, but highly motivated to do the job" [Questionnaire respondent]

"Someone passionate about languages and helping others" [Questionnaire respondent]

Other responses included those that thought volunteers were looking for experience as well as professionals who wanted to offer their efforts for good causes.

When offered a list of reasons why translators might provide translations voluntarily without payment, and asked to pick the three main motivations, respondents most frequently picked altruistic and cause-driven motivations like caring for a specific cause, wanting to do good and helping others and wanting to contribute to a creative effort. Career-focused motivations like wanting to gain experience, needing to improve and wanting to be affiliated with an organisation for career purposes ranked in the mid-range, while translating for enjoyment or for entertainment ranked lowest (Figure 54).

During the interviews with professional buyers in the localisation industry, crowdsourced or voluntary translations were also discussed with the informants. None of them had personal experience with crowdsourced projects except for ProBuyer 3, who had on occasion done free translations for a project. However, interviewees had heard of crowdsourced projects or observed the results.
ProBuyer 3 thought of volunteer translators as professionals who decided to provide translations for reasons other than financial gain, and that a professional should always provide the best quality, while at the same time it has to be made clear what is expected from volunteers. This informant also thought that individuals decided to volunteer in translation for a variety of reasons, including gaining experience, entering the market, an interest in the subject or translation as a hobby, as well as ethical or moral reasons, similarly to deciding to donate a percentage of one's income. However, since it wasn't possible to tell who the translators were due to individuals with different levels of experience being able to volunteer, the outcome might vary due to their differing backgrounds.

7.2.4.2.2. Attitudes towards volunteer translations

The basic assumptions the interviewees made about individuals who volunteer for translation projects were founded in projects they had heard about previously. ProBuyer 1 provided two anecdotes of crowdsourced translations, one of an individual who could sell translations below market value by crowdsourcing the translations and only proofreading them afterwards, as well as a game that had been translated by fans of the game who were selected individually based on their translation skills and then rewarded with in-game points.

ProBuyer 1 also described the example of the anime series 'Ghost in the Shell' and the great results that had been achieved through fan translations. They also suggested that it would not have been possible for professional translators to achieve the same level of quality as the fans had achieved.

However, when asked about projects like the user interface (UI) translations of Facebook, ProBuyer 1 expressed confusion as to why crowdsourcing was done in this case since UI translations did not require specific domain knowledge. In addition, this informant was unsure why a company like Facebook would not pay for professional translations or why translators would agree to provide translations for free in that case.

Similarly, ProBuyer 2 also associated crowdsourcing with projects where individuals passionate about the content, like games or subtitling, would get involved and
provide sometimes good, sometimes not so good quality. This informant distinguished between linguistic experts and subject matter experts, describing volunteers as subject matter experts and professional translators as linguistic experts, describing quality issues that may arise from volunteer translations as primarily formal language errors as well as stemming from a lack of process knowledge.

Based on these considerations, ProBuyer 2 thought that the ideal combination would be to have both volunteers as subject matter experts and professional translators as linguistic experts, on a project to benefit from the strengths of both groups. When asked about projects where volunteers were not necessarily subject matter experts, ProBuyer 2, as with ProBuyer 1, also expressed confusion why this would be done.

7.2.4.3. Linguists: Volunteer Translation and Crowdsourcing

In the linguist group, while the majority of the 197 respondents who completed the questionnaire in this group said they thought translation crowdsourcing was overall positive (61.5%), 13% thought it was overall negative and 24% were neutral or had no opinion (Figure 56).

It should also be noted that while nearly all respondents who had identified their primary role to be volunteer translator or reviewer thought that volunteer translation or crowdsourcing were overall positive, nearly 19% of those that identified primarily as paid-for translators thought it was overall negative, while 29% were neutral or had no opinion, meaning that while 87.7% of those that identified primarily as volunteer translators thought volunteer translations or crowdsourcing were overall positive, only 52% of primarily paid-for translators thought so (Figure 57).
Figure 56: Overall view of translation crowdsourcing for linguist respondent group
Figure 57: Overall view of translation crowdsourcing for linguist respondent group, broken down by primary role of respondents
Respondents were given the opportunity to elaborate on their opinion in their own words, and 101 provided additional comments. These comments ranged from positive or cautiously positive remarks to criticism.

41 respondents elaborated in a positive manner on their opinion on translation crowdsourcing:

"I started working for The Roseta Foundation, and I love being part of such a beautiful team. I get paid in other companies and I consider that the help we give in Rosetta is necessary to many." [Questionnaire respondent]

"I think it is a very positive experience for both sides. From the point of view of the translator you can choose the tasks you want in and in general they are very interesting and the feedback you get is very educational. Obviously the organisation benefits as can save money on translations and use it for other things. Volunteer translating is one of the few opportunities available that you can do from the comfort of your own home." [Questionnaire respondent]

Of the critical voices, the second most commonly mentioned problem was the potential of translation volunteering or crowdsourcing to be abused. The 28 respondents who elaborated on this, raised issues like the solicitation of free translations by big for-profit companies or even big non-profit companies that could afford to pay for translations, while others distinguished in a more general way between for-profit and non-profit organisations when judging whether volunteer translations were overall negative or positive:

"It's very frustrating as a professional from companies who can afford it like Microsoft, Google, Apple, Facebook etc. If an independent creator cannot financially afford translation but wishes to take a bet on volunteer work, that's OK." [Questionnaire respondent]

"I believe most translation work should be remunerated, with the exception of work done for non-profit organizations." [Questionnaire respondent]
"In any non-profit environment volunteer translation, if not necessary, is, at least, a good thing. On the other hand, I find voluntary work in any commercial environment despicable." [Questionnaire respondent]

The most commonly mentioned concern was quality of the crowdsourced output with 30 mentions, followed by the concern that volunteer translation might undermine the translation profession, which was raised twice. However, in addition to these two instances, the concern that organisations might abuse crowdsourcing could also be read to imply a concern for the translation profession since paid-for translations are getting replaced with free translations in a situation where the translations should be paid-for.

"I haven't seen many examples, but what I have seen on Wikipedia leaves a lot to be desired - source text vocabulary and structure is very evident." [Questionnaire respondent]

"The translations I performed are for non-profit organizations that can't afford professional translations. Some volunteers are less trained and that results in poor quality translations, hard to follow for the reader." [Questionnaire respondent]

"As long as it's done for legitimately pro-bono purposes (charities, NGOs) and not to undermine professional translators/cut corners on corporate translation, I support it." [Questionnaire respondent]

Some were also completely against volunteer translations:

"If it were good enough, somebody would have paid for that." [Questionnaire respondent]

"Professional translation work should be paid, like any other job." [Questionnaire respondent]

"Translators are highly educated professionals and should be paid appropriately for their work." [Questionnaire respondent]
Other, less frequently mentioned concerns included poorly managed crowdsourcing initiatives, and the expectation that volunteer translators should get something in return, like a reward or feedback to be able to improve their skills. Others expressed the opinion that only sufficiently qualified professionals should volunteer or that volunteer translations might undermine the translation profession.

7.2.4.3.1. Describing the average volunteer

Respondents were then asked to describe the average translation volunteer in their own words. 139 responses were provided to this question, again ranging from very positive to more negative or critical descriptions.

46 respondents described volunteer translators in a positive way, sometimes leaving only the positive description and sometimes adding a more critical element to the response. Positive descriptions focused generally on the motivation to volunteer, which included altruistic considerations, enthusiasm for a cause or language, and volunteer translators being generally nice people:

"Hard working, giving, humanist" [Questionnaire respondent]

"A very selfless person" [Questionnaire respondent]

Many responses included descriptions of the professional status of volunteer translators, which ranged from volunteer translators being described as bilingual amateurs (24) to professional translators who had retired or wanted to donate their time (12), or novice translators wanting to improve their skills (43). Some respondents focused on one of those descriptions, while others expressed the opinion that there were different types of volunteer translators:
"A person who speaks, of feels that they speak, a foreign language and wants
to contribute to a particular cause out of commitment or as a hobby."
[Questionnaire respondent]

"I believe that the average translation volunteer is a translation student who is
searching for his first approach to translation work." [Questionnaire
respondent]

12 responses also included rather negative descriptions of volunteer translators or
their work, while 19 felt unable to describe translation volunteers:

"Mostly bad and very literal (hello Google translation). It is no surprise since
translation is a job and requires technique, style and experience."
[Questionnaire respondent]

"Marginally competent" [Questionnaire respondent]

Respondents were then offered several reasons why a translator might volunteer
without payment and were asked to pick the three main motivations (Figure 54).

- The two most commonly selected motivations included
  - wanting to get experience with translation on first position (74.6%),
  - followed by wanting to do good and helping other people (60.9%).

- Also, other career-focused or altruistic motivations were selected frequently
  - with caring for a specific cause (38.1%), and
  - wanting to be affiliated with a non-profit organisation for career
    purposes (32%).

- More negative interpretations of motivations ranked at the bottom of the list,
  - with 6% of the respondents selecting that translators were not very good
    and needing to improve, and
  - 7.1% selecting translators were on a break and bored as a main
    motivation to volunteer.

During the interviews with the four translators, the topics discussed broadly fit into
the following categories:

- their own attitudes towards translation crowdsourcing,
- observations they had made in relation to translations on Trommons, as well as
- the informants’ thoughts on who the volunteers were and why they would volunteer, which was commonly informed by their own motivations for volunteering.

The interviewees generally thought that volunteers were a mix of amateurs or professionals with a slight tendency to assume they were amateurs. For example, Translator 4 describes it as follows:

"From what I saw I'd say most of them are amateurs but there were a few that were good so it's hard to tell. Some that could be professional, yeah"
[Translator 4]

Interviewees gave a wide range of reasons as to why individuals might decide to volunteer. Those included wanting to learn the language, helping people or charities, finding translations interesting, learning new vocabulary and improving language skills, improving employability, having a challenge and doing it for fun.

In addition, Translator 4 pointed out that some volunteers might decide to offer their services too early:

"I guess they want the experience and they want to learn French so they can start getting paid for it, I think maybe some people volunteer too soon, they should get maybe to a certain level before they take a type of task"
[Translator 4]

Translator 2 pointed out that it was difficult for beginners to find jobs because people preferred more experienced translators. Since everyone has to start somewhere and
makes mistakes at the beginning, there should, in Translator 2’s opinion, be a way of getting a standard to help translators be better and improve. Translator 2 also pointed out that there were different types of translators: some translate and then decide they want a degree while others speak both languages and study first, then start working in translations. Yet others have a subject matter degree like engineering or medicine and go into translation from there. Therefore, the experience and knowledge of language or terminology would differ between translators and their abilities and needs.

7.2.4.3.2. Attitudes towards volunteer translations

The general theme expressed by the informants was that overall, there shouldn't be a big difference between volunteer translations and those of a professional translator. While some quality aspects like style or spelling might be lacking, the overall level of service should be similar. Translator 4 suggested that the main difference might be that a volunteer needs more time for the translations.

Both Translator 3 and Translator 2 also expressed multiple times that they thought that the translations would be of high quality because volunteers would give their best or try hard to do a good job.

Translator 1 described that organisations could have a positive impact on the quality of the volunteer translations by selecting volunteers and expressed that they thought that TRF filter translators, selecting as volunteers only those that have studied translation and have experience to be able to volunteer. In addition, other proof-readers would review the translations. This informant also compared TRF with Translators Without Borders, who, according to the interviewee, have a higher selection standard than TRF, and were therefore more prestigious as a place for translators to volunteer. In contrast, organisations that request translations directly from volunteers as opposed to working with an organisation like TRF or Translators Without Borders would likely face quality issues.

Interview participants also relayed their observations on quality in volunteer translations, from the perspective of having accepted translation and proofreading tasks themselves. Translator 4 pointed out that a common issue was that there was little consistency between current and previous projects, giving the example of content that
referenced text that had previously been translated but instead of including the previously published translation, the volunteer had re-translated this text, therefore creating inconsistencies. This informant also described the translations as generally quite poor and explained that this might be because the volunteers might not have the necessary source or target language skills. Translator 1 also thought that the translations were almost always of poor quality when the interviewee did proofreading on volunteer translations, although in reference to translations outside of Trommons.

Translator 4 also highlighted the need of volunteers to get help with improving on quality, for example by receiving reference material and previous translations as well as more guidelines and reviews.
Chapter 8: Experiment Results

The following sections describe the outcomes of the three experiments conducted with different respondent groups.

Section 8.1. includes experiment results for the Spanish language experiment, which addressed the impact the information on the professional status of the translator has on the perceived quality of translations. This experiment was conducted with students and staff at the University Alfonso X el Sabio in Madrid as well as translators and reviewers through an open call.

Section 8.2. describes the outcome of experiments conducted in German, which addressed the impact the information on payment the translator received has on the perceived quality of translations. The experiment was conducted with translators and reviewers through an open call.

Section 8.3. describes the outcome of experiments conducted in English, addressing the impact the information on payment the translator received has on the perceived quality of translations with an added proxy, since here the evaluation wasn't done based on actual translations, but on comments provided by reviewers who had evaluated the original translations. The experiment was conducted with non-profit clients on Trommons as well as buyers in the localisation industry.

8.1. Experiment 1: Professional Status, Spanish

Experiment 1 presented participants with 12 scenarios (+ one control scenario, which was excluded from the main analysis but will be addressed in section 9.2.2.). In each scenario, participants were provided with an English sentence and three Spanish translations with the instruction to select the translation they thought was best.

For the first 4 scenarios, no information on the professional status of the translator was provided (NoInfo).

For the subsequent 8 scenarios, information on the professional status of the translator was added to each of the translations offered, with 4 scenarios providing correct information (GoodInfo) and 4 incorrect information (BadInfo). GoodInfo and BadInfo scenarios were mixed.
Professional status was indicated by the following:

- Novice translator ('Novice')
- Trained translator ('Trained')
- Trained + experienced translator ('Expert')

No further explanation on the translator's background apart from these phrases was provided.

Table 37 shows the relationship between the information given in the BadInfo scenarios and the professional status of the translator who had provided the translation.

Analysis of the results was conducted for each respondent group separately, followed by a comparison analysis for all groups.

As a first step, the number of all participants who selected each option for the different scenarios were added up to determine which translation was considered the best version by most of the participants, which came second and which third.
<table>
<thead>
<tr>
<th>Scenario Nr.</th>
<th>Information presented</th>
<th>Translation provided by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 9</td>
<td>Novice</td>
<td>Trained</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Trained + Experienced</td>
</tr>
<tr>
<td></td>
<td>Trained + Experienced</td>
<td>Novice</td>
</tr>
<tr>
<td>Scenario 10</td>
<td>Novice</td>
<td>Trained + Experienced</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Novice</td>
</tr>
<tr>
<td></td>
<td>Trained + Experienced</td>
<td>Trained</td>
</tr>
<tr>
<td>Scenario 11</td>
<td>Novice</td>
<td>Trained</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Trained + Experienced</td>
</tr>
<tr>
<td></td>
<td>Trained + Experienced</td>
<td>Novice</td>
</tr>
<tr>
<td>Scenario 12</td>
<td>Novice</td>
<td>Trained + Experienced</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Novice</td>
</tr>
<tr>
<td></td>
<td>Trained + Experienced</td>
<td>Trained</td>
</tr>
</tbody>
</table>

Table 37: Relationship between information provided to participants vs. actual translator background for BadInfo scenarios in Experiment 1
As a second step, standard deviation was calculated for each individual scenario as well as the three types of scenarios combined (All NoInfo, all GoodInfo, all BadInfo) to get a better understanding of the choices made by the participants as a whole and not skew the results to only consider the majority choices. Since each scenario provided participants with three options, it is conceivable that while one option 'won' by being selected more frequently than the other two options individually, the majority of participants might still have not selected that option. Figure 58 provides two examples.

- Here, options 'Example 1c' and 'Example 2c' both 'won' due to being selected more frequently than the other two options in their respective scenarios.
- However, in Example 1, participants agreed strongly that option C should be considered best.
- For Example 2, opinions were split with option B being selected by nearly as many participants as the 'winning' option C, and the combined options A and B being selected more frequently than option C.

Standard deviation for Example 1 is 33.7, for Example 2: 14.6. In the analysis of the experiment results, a greater standard deviation therefore indicates a higher level of agreement among participants, while low standard deviation indicates split opinions.

Finally, averaged standard deviation for each type of scenario will be presented to show broader trends across different participant groups.
Figure 58: Example of high vs. low agreement results in experiment scenarios
8.1.1. Students and Staff at UAX

8.1.1.1. Experiment: Selections (All UAX)

Figure 59 shows the results for the three scenario types, with

- those scenarios where no information on the translator's background was provided being on the very left (NoInfo),
- those where the correct information was given following (GoodInfo) and finally
- those scenarios where the incorrect information was provided last (BadInfo).

The results show that when no information was provided on translator background (NoInfo),

- the expert translator was selected most frequently in two scenarios (Expert; Scenarios 2 and 4),
- the trained translator in one scenario (Scenario 1) and
- the novice translator in one scenario (Scenario 3).

It is worth noting that in scenario 3, where the novice translator was selected most often, the expert translator was selected nearly as often as the first choice, the novice translator.

When provided with the correct information on the translators (GoodInfo), participants selected the expert translator in all four scenarios (Expert; Scenarios 5, 6, 7 and 8), with in three out of the four scenarios the novice translator being selected least frequently and the trained translator in second position (Scenarios 5, 6 and 8).

When participants were provided with the incorrect information about the translator's background (BadInfo), the results seem initially similar to that of the NoInfo scenarios. Based on the information provided, which was in fact not the real background of the translator, the ostensibly expert translator was selected in two scenarios (Expert; Scenarios 9 and 12). What was ostensibly the novice translator's version and that of the trained translator was selected most frequently in one scenario each (Scenarios 10 and 11).

When considering standard deviation for each individual scenario (Figure 60), as
well as the averaged standard deviation for each type of scenario (Figure 61), a difference between the NoInfo and BadInfo scenario group can be observed. This is not clear from only considering which option was selected most frequently for each of the scenarios. While there are some exceptions, the general trend is that participants showed lower levels of agreement in BadInfo scenarios than in NoInfo scenarios, with the highest level of agreement in GoodInfo.

The following sections describe the results of each individual participant group: Lecturers, 3rd/4th year students and 1st/2nd year students.
Figure 59: Best translation selected by student and lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 60: Standard deviation in selections for student and lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 61: Average standard deviation in selections for student and lecturer UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.
8.1.1.2. Experiment: Selections (Lecturers)

16 (33.3%) of the participants were lecturers at the University Alfonso X el Sabio. When provided with no information on the translator's background, the lecturer group selected

- the expert translators in three scenarios (Scenarios 2, 3 and 4) and
- the trained translator in one (Scenario 1).

When provided with correct information on translator background, the lecturer group selected

- the expert translator's version most frequently for all four scenarios (Scenarios 5, 6, 7 and 8).

When incorrect information was provided, the selections matched those of the group as a whole, with

- what was claimed to be the expert translator's version being selected in two scenarios (Scenarios 9 and 12), and
- the supposed trained and novice translator's versions once each (Scenarios 10 and 11) (Figure 62).

These results show that the lecturer group only differs from the group of all participants as a whole when provided with no information on the translator's background, in which case the more qualified translator's version is selected most frequently (3 expert, 1 trained for the lecturer respondent group compared to 2 expert, 1 trained and 1 novice for the group as a whole).

When considering standard deviation and thereby how strong the agreement is within the lecturer group, some additional differences can be observed (Figures 63 and 64).

Agreement among the lecturers on which translation was the best version was overall strong (Figures 63 and 64). In all except one scenario, the version selected most frequently also was selected by more than half of the participants. Similar to the analysis of all participant results (Section 8.1.1.1.), lecturers also showed lower levels of agreement for the BadInfo scenarios than for NoInfo and GoodInfo.
Figure 62: Best translation selected by lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 63: Standard deviation in selections for lecturer UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 64: Average standard deviation in selections for lecturer UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.
8.1.1.3. Experiment: Selections (3\textsuperscript{rd} and 4\textsuperscript{th} year Students)

As indicated in section 7.1.1.1., 21 (43.8\%) participants were 3\textsuperscript{rd} and 4\textsuperscript{th} year students at the University Alfonso X el Sabio, with the sole master's student being counted in this category. For the purpose of brevity, this group will be referred to as ‘advanced students’ in the following section.

When provided with no information on the translator's background, the advanced student group selected

- the expert translator most frequently in two scenarios (Scenarios 2 and 4),
- the trained translator in one scenario (Scenario 1) and
- the novice translator also in one (Scenario 3).

When provided with correct information on translator background, the advanced student group selected

- the expert translator's version most frequently for all four scenarios (Scenarios 5, 6, 7 and 8).

When incorrect information was provided, the selections deviated from the group as a whole, with

- what was claimed to be the expert translator's version being selected in one scenario (Scenario 12),
- the supposedly trained translator in two scenarios (Scenarios 9 and 11) and
- the supposed novice translator in one (Scenario 10) (Figure 65).

These results show that the advanced student group only differs from the group of all participants as a whole when provided with the incorrect information on the translator's background, in which case the more qualified translator's version is selected less frequently (1 expert, 2 trained, 1 novice for the advanced student group compared to 2 expert, 1 trained and 1 novice for the group as a whole).

Agreement among the advanced students on which translation was the best version was relatively weak (Figures 66 and 67). Most notably, while agreement was lower in
BadInfo scenarios compared to NoInfo, providing correct information on professional background of the translators increased agreement only slightly.
Figure 65: Best translation selected by 3rd/4th year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 66: Standard deviation in selections for 3rd/4th year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 67: Average standard deviation in selections for 3rd/4th year student UAX respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.
8.1.1.4. **Experiment: Selections (1\textsuperscript{st} and 2\textsuperscript{nd} year Students)**

As indicated in section 7.1.1.1., 11 (23\%) participants were 1\textsuperscript{st} and 2\textsuperscript{nd} year students at the University Alfonso X el Sabio in Madrid. This group will be referred to as ‘beginner students’ in the following section.

When provided with no information on the translator's background, the beginner student group selected

- the expert translator in one scenario (Scenario 4),
- the trained translator also in one (Scenario 1) and
- the novice translator in two scenarios (Scenarios 2 and 3).

When provided with correct information on translator background, the beginner translator group selected

- the expert translator in three scenarios (Expert; Scenarios 5, 7 and 8) and
- the trained translator in one (Scenario 6).

When provided with the incorrect information, this respondent group selected

- what was claimed to be the expert more frequently than the participant group as a whole (three for beginner students vs. two for group as a whole), and
- the supposedly novice translator in one of the scenarios (Scenario 11) (Figure 68).

These results show that beginner students deviate from the group as a whole in that they not only select the less experienced translators when provided with no information on professional status of the translator, but also are the only group that selected a trained translator in one of the scenarios when provided with correct information. However, when provided with the incorrect information on the translator's background, the beginner students group was more likely to select the ostensibly expert translator in the majority of scenarios, which deviates from the group as a whole.

Figures 69 and 70 show the standard deviation per scenario and averaged for each category.
Figure 68: Best translation selected by 1st/2nd year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 69: Standard deviation in selections for 1<sup>st</sup>/2<sup>nd</sup> year student UAX respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Agreement among beginner students on which translation was the best most closely matched that of the participants as a whole, despite being the smallest group (27%). Agreement was strongest for the GoodInfo scenarios, and nearly half that for BadInfo, with NoInfo situated in the middle. Overall, there is slightly less agreement on which translation is best in the beginner group compared to the participants as a whole.

**8.1.2. Spanish Linguists, Open Call**

42 respondents in the Spanish linguist category agreed to participate in the experiment.

Within this group, when participants were provided with no information on the translator's background,

- the expert translator (Scenarios 2, 3 and 4) was selected three times and
- the trained translator once (Scenario 1).
When correct information on translator background was provided,
- the expert translator was selected in three scenarios (Scenarios 5, 7 and 8) and
- the trained translator once (Scenario 6).

When incorrect information was provided,
- the ostensibly expert translator was selected once (Scenario 12),
- the ostensibly trained translator twice (Scenarios 8 and 11) and
- the translator marked as novice once (Scenario 10) (Figure 71).

Calculation of standard deviation and average standard deviation (Figures 72 and 73) shows slightly more agreement when no information is shown, compared to GoodInfo scenarios. NoInfo scenarios had the lowest levels of agreement.
Figure 71: Best translation selected by Spanish linguist respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 72: Standard deviation in selections for Spanish linguist respondent group, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 73: Average standard deviation in selections for Spanish linguist respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.

8.1.3. Comparison across Groups

The following sections describe the outcome of Experiment 1 (Spanish, Professional status) based on data already presented. It seeks to provide further insights by re-grouping information.

The following sections outline the results of analysis where all respondent groups were combined, regardless of their own background, and then split in two groups. Groups were formed based on the participant's response to a specific question in the questionnaire: "Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?"

Those that had selected 'Yes, in most cases' or 'Yes, in some cases' being placed in one group and those that had selected 'Only sometimes' or 'No, not usually' were placed in another group.
8.1.3.1. Experiment: Professional Status as Deciding Factor

Among the respondents who thought that the question of whether someone was a professional translator or not was a deciding factor in some or most cases, 70 agreed to participate in the experiment. Results are displayed in Figure 74.

When provided with no information on the translator’s professional status, this group chose

- the expert translator (Scenarios 2, 3 and 4) in three scenarios, and
- the trained translator once (Scenario 1).

When provided with the correct information, this group chose

- the expert translator in all four scenarios (Scenario 5, 6, 7 and 8).

When incorrect information was provided,

- the expert translator's version was selected in two scenarios (Scenarios 9 and 12), and
- the trained and novice translators once each (Scenarios 10 and 11).

Calculating standard deviation and average standard deviation (Figures 75 and 76) shows overall relatively low agreement, with agreement being lower relative to NoInfo for BadInfo and higher for GoodInfo.
Figure 74: Best translation selected by respondents who thought professional status of translator had impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 75: Standard deviation in selections by respondents who thought professional status of translator had impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 76: Average standard deviation in selections by respondents who thought professional status of translator had impact on quality, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.
8.1.3.2. Experiment: Professional Status not a Deciding Factor

10 respondents who thought that the question whether someone was a professional translator or not was a deciding factor in some or most cases agreed to participate in the experiment (Figure 77).

In NoInfo scenarios, this group chose
- the expert translator (Scenarios 2, 3 and 4) in three scenarios, and
- the trained translator once (Scenario 1).

In GoodInfo scenarios, this group chose
- the expert translator in three scenarios (Scenario 5, 7 and 8).

In BadInfo scenarios,
- the expert translator's version was selected in two scenarios (Scenarios 9 and 12),
- the trained and novice translators once each (Scenarios 10 and 11) (Figure 77).

Calculating standard deviation and average standard deviation (Figures 78 and 79) shows overall relatively low agreement, with agreement being lower relative to NoInfo for BadInfo and higher for GoodInfo. However, for this group agreement is only slightly lower when no information on the translator's professional status is provided than for those scenarios where correct information is provided.
Figure 77: Best translation selected by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 78: Standard deviation in selections by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario. BadInfo refers to the information provided to participants, not the actual translator.
Figure 79: Average standard deviation in selections by respondents who thought professional status of translator had no or little impact on quality, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.
8.1.3.3. Experiment: Comparison Across Groups

The following section compares the different groups on three different aspects:

- Overall performance in the Spanish experiments based on the professional status of the translator,
- levels of agreements as well as performance and
- levels of agreement relative to word counts of the source text in each individual scenario.

Groups included in the comparison were:

- All UAX: All participants at the University Alfonso X el Sabio in Madrid (combined Lecturers, 3\textsuperscript{rd}/4\textsuperscript{th} year students, 1\textsuperscript{st}/2\textsuperscript{nd} year students);
- Lecturers UAX: Only lecturers at the University Alfonso X el Sabio in Madrid;
- 3\textsuperscript{rd}/4\textsuperscript{th} UAX: Only 3\textsuperscript{rd} year, 4\textsuperscript{th} year and Master's students at the University Alfonso X el Sabio in Madrid;
- 1\textsuperscript{st}/2\textsuperscript{nd} UAX: Only 1\textsuperscript{st} year and 2\textsuperscript{nd} year students at the University Alfonso X el Sabio in Madrid;
- All Ling: Spanish linguist respondents, open call.

Unfortunately, only three respondents among the Spanish linguists who had never volunteered as translators or reviewers also participated in the experiment, making it not feasible to compare volunteers to non-volunteers.

8.1.3.3.1. Overall performance

The problem of what should be considered high or low performance needs to be resolved when attempting to evaluate the performance of each group relative to the information provided on the translator's background. At the same time, it is a worthwhile endeavour since it gives further insights into how reliant different groups were on extrinsic
information. For this analysis, the assumption will be that the translations completed by the expert translator ought to be of higher quality than that of the novice or trained (but inexperienced) translators. This implies that participant performance in the experiments would be linked to how often they selected the expert translator as the best translation. The ideal would be a 100% selection rate of the expert translator's version. However, as discussed in the review of the literature (Chapters 2 and 3), quality is difficult to define and evaluate and there are no universally agreed criteria in place. For this reason, performance of each individual participant group will be evaluated relative to the performance of all participants as a whole.

Figure 80 shows the results of the analysis, which included the following steps:

**Step 1:** For each type of scenario (NoInfo, GoodInfo, BadInfo), the percentage of respondents who selected the expert translator's version was averaged (mean) across the four scenarios.

As an example, the lecturers at UAX had selected the expert translator when no information was provided on the translator's background (NoInfo) as shown in Table 38.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>6.30%</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>81.30%</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>62.50%</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>62.50%</td>
</tr>
<tr>
<td>Mean</td>
<td>53.20%</td>
</tr>
</tbody>
</table>

*Table 38: Example of selections of 'expert' translator in percentage, lecturers at UAX participant group*

**Step 2:** The mean for all respondents at UAX and the linguists in the open call was calculated, thereby ensuring that each respondent was only included in this calculation once. This was done only for the NoInfo scenarios, resulting in a value of 48.2%. This shows that on average, 48.2% of all participants thought that what was the translation provided by the expert translator was also the best translation, when they were not provided with information on the translator's background. In Figure 80 this value is indicated marked
with the 'Mean NoInfo' line, and will be considered the reference value against which the performance of individual groups is evaluated, independent of the consideration of objective or absolute quality.

**Step 3:** For scenarios where incorrect information was provided, the information on translator background that participants were given was considered (BadInfo_Lie), as well as which translator had actually provided the translation (BadInfo_Truth).
Figure 80: Average selection of 'expert' translation relative to mean of all respondent groups, student and lecturers UAX and Spanish linguists as well as those indicating that professional status was a deciding factor for quality vs. those who didn’t think so.
The results indicate the following (Figure 80):

1. When no information was provided on translator background (NoInfo),
   - 3rd and 4th year students performed only slightly worse than the average performance of the whole group (46.1%);
   - Both lecturers (53.2%) and industry linguists (52.4%) performed slightly better than the group average;
   - 1st and 2nd year students’ performance was lowest relative to the group average (27.3%);
   - No difference can be observed between participants who thought translators being a professional had impact on quality (YesPro), compared to those who thought this wasn’t the case (NoPro).

2. When correct information was provided on translator background (GoodInfo),
   - All groups performed better than the NoInfo average;
   - Lecturers showed the greatest increase in performance (34.3%) relative to the NoInfo average;
   - 1st and 2nd year students showed the greatest increase in performance (29.5% - more than double!) relative to the NoInfo performance for that group;
   - Increases for 3rd and 4th year students (18.9%) and industry linguists (11.9%) were lowest;
   - Participants who thought that professional status and quality were correlated (YesPro; 22.1%) improved performance more than those who thought this wasn't the case (NoPro; 12.5%).

3. When incorrect information was provided on translator background (BadInfo),
   - All groups performed worse than the NoInfo average;
   - Lecturers, 3rd and 4th year students, and industry linguists selected what was
actually the expert translator (BadInfo_Truth) slightly more often than what was ostensibly the expert translator (BadInfo_Lie);

- Differences between these three groups were very small: Lecturers 2.9%, 3rd and 4th year students 3.8%, industry linguists 5.2%;
- The only group that picked the ostensibly expert translator over the actual expert more often, as well as more often than any other group were the 1st and 2nd year students (-25% difference);
- There was only a slight difference between those who thought professional status and quality are correlated (YesPro) and those who didn't (NoPro).

8.1.3.3.2. Level of agreement

Section 8.1 provided a rationale for using standard deviation to determine level of agreement, with high deviation indicating high agreement. Figure 81 shows mean standard deviation across all groups and scenario types.

The following can be observed:

1. For scenarios where no information on the translator's background was provided:
   - Lecturers and industry linguists have the highest levels of agreement on what should be considered the best translation;
   - Students show lower levels of agreement, with 3rd and 4th year students in the mid-range and 1st and 2nd year students in the lowest;
   - Agreement is slightly higher in the group who thought professional status is no deciding factor for expected quality (NoPro), compared to those who thought it is a deciding factor (YesPro).
Figure 81: Mean of standard deviation for student and lecturer UAX and Spanish linguist participant groups as well as those indicating that professional status was a deciding factor for quality vs. those who didn’t think so. BadInfo refers to the information provided to participants, not the actual translator.
2. For scenarios where correct information on the translator's background was provided:
   - Increases in agreement levels are greatest for lecturers (9.7%) and 1st/2nd year students (6.9%);
   - Only a slight increase can be observed for 3rd/4th year students (1.2%);
   - Industry linguists have slightly lower levels of agreement (-1.1%) and are the only group where agreement was lower than for NoInfo scenarios;
   - Participants who thought that the professional status of the translator was a deciding factor in expected quality had a greater increase in agreement (YesPro, 6.8%) than those who thought this wasn't the case (NoPro, 2%).

3. For scenarios where incorrect information on the translator's background was provided:
   - Agreement levels dropped for all groups, relative to NoInfo and GoodInfo scenarios;
   - Agreement levels were highest for lecturers (21.8%) and industry linguists (16.7%);
   - Agreement levels were lowest for students (14.5% and 13.2%);
   - Decreases in agreement relative to the NoInfo scenario were greatest in industry linguists (-9.6%) and 3rd/4th year students (7.4%);
   - Decreases in agreement relative to the NoInfo scenario were smallest in lecturers (6.9%) and 1st/2nd year students (5.6%);
   - Among the participants who thought that the professional status of the translator was a deciding factor in expected quality (YesPro), decrease in agreement was smaller (5.4%) than those who thought the opposite (NoPro; 8.1%).
8.1.3.3.3. Performance and agreement relative to word counts

One noteworthy observation during data analysis was that for some scenarios, performance and/or agreement were overall higher across all groups, compared to others. To understand why this may be the case, the experiment results were related to the source language word count of each of the scenarios since lower word counts could result in participants having less intrinsic information on the translation quality, thereby having to rely more heavily on extrinsic information.

For each of the information categories provided, two scenarios had a low word count (between two and 9 words) and two a higher word count (between 49 and 110 words); average word count for each type of scenario was similar (Table 39).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>WordCounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NoInfo</td>
<td>4</td>
</tr>
<tr>
<td>2-NoInfo</td>
<td>74</td>
</tr>
<tr>
<td>3-NoInfo</td>
<td>71</td>
</tr>
<tr>
<td>4-NoInfo</td>
<td>2</td>
</tr>
<tr>
<td>5-GoodInfo</td>
<td>68</td>
</tr>
<tr>
<td>6-GoodInfo</td>
<td>8</td>
</tr>
<tr>
<td>7-GoodInfo</td>
<td>89</td>
</tr>
<tr>
<td>8-GoodInfo</td>
<td>8</td>
</tr>
<tr>
<td>9-BadInfo_Lie</td>
<td>9</td>
</tr>
<tr>
<td>10-BadInfo_Lie</td>
<td>49</td>
</tr>
<tr>
<td>11-BadInfo_Lie</td>
<td>7</td>
</tr>
<tr>
<td>12-BadInfo_Lie</td>
<td>110</td>
</tr>
<tr>
<td>NoInfo Average</td>
<td>37.75</td>
</tr>
<tr>
<td>GoodInfo Average</td>
<td>43.25</td>
</tr>
<tr>
<td>BadInfo Average</td>
<td>43.75</td>
</tr>
</tbody>
</table>

Table 39: Overview of source language word counts per scenario

Table 40 shows the performance for each respondent group and scenario, combined with the source language word count for each scenario. The analysis here resulted in the observation that no trends or strong correlations between word count and performance could be observed. In addition, relating performance to the individual translators who had
provided each scenario translation showed no trends or strong correlations.

When considering agreement levels next to word counts of the scenarios, results vary depending on group.

1. **Lecturers (Figure 82):**
   - In NoInfo, no firm pattern between word count and the level of agreement between participants can be observed. The two scenarios with the low word count resulted in a high and low level of agreement each (1-NoInfo and 4-NoInfo), as did the two scenarios with high word count (NoInfo3 and NoInfo2);
   - In GoodInfo, a correlation between word count and level of agreement can be observed. The two scenarios with low word count also show lower levels of agreement among participants on which translation is the best (6-GoodInfo and 8-GoodInfo), while the two scenarios with high word count show higher levels of agreement (5-GoodInfo and 7-GoodInfo);
   - Finally, in NoInfo, these correlations are reversed. The two scenarios with high word count show less agreement (10-BadInfo and 12-BadInfo) than the two scenarios with low word count (9-BadInfo and 11-BadInfo). Here, the scenario with the highest word count is also the one with the greatest level of disagreement.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Word Counts</th>
<th>AllUAX</th>
<th>LecturersUAX</th>
<th>3rd/4thUAX</th>
<th>1st/2ndUAX</th>
<th>AllLing</th>
<th>YesPro</th>
<th>NoPro</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NoInfo</td>
<td>4</td>
<td>12.5</td>
<td>6.3</td>
<td>14.3</td>
<td>18.2</td>
<td>11.9</td>
<td>15.7</td>
<td>0</td>
</tr>
<tr>
<td>2-NoInfo</td>
<td>74</td>
<td>53.2</td>
<td>81.3</td>
<td>55</td>
<td>9.1</td>
<td>63.41</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>3-NoInfo</td>
<td>71</td>
<td>42.6</td>
<td>62.5</td>
<td>35</td>
<td>27.3</td>
<td>55.26</td>
<td>44.29</td>
<td>60</td>
</tr>
<tr>
<td>4-NoInfo</td>
<td>2</td>
<td>68.1</td>
<td>62.5</td>
<td>80</td>
<td>54.5</td>
<td>78.95</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>5-GoodInfo</td>
<td>68</td>
<td>80.9</td>
<td>93.8</td>
<td>70</td>
<td>81.8</td>
<td>86.5</td>
<td>87.1</td>
<td>70</td>
</tr>
<tr>
<td>6-GoodInfo</td>
<td>8</td>
<td>48.9</td>
<td>68.8</td>
<td>50</td>
<td>18.2</td>
<td>38.2</td>
<td>48.6</td>
<td>20</td>
</tr>
<tr>
<td>7-GoodInfo</td>
<td>89</td>
<td>72.3</td>
<td>100</td>
<td>60</td>
<td>54.5</td>
<td>76.5</td>
<td>74.3</td>
<td>70</td>
</tr>
<tr>
<td>8-GoodInfo</td>
<td>8</td>
<td>80.9</td>
<td>87.5</td>
<td>80</td>
<td>72.7</td>
<td>55.9</td>
<td>68.6</td>
<td>80</td>
</tr>
<tr>
<td>9-BadInfo_Lie</td>
<td>9</td>
<td>51.1</td>
<td>81.3</td>
<td>30</td>
<td>45.5</td>
<td>38.2</td>
<td>45.7</td>
<td>50</td>
</tr>
<tr>
<td>10-BadInfo_Lie</td>
<td>49</td>
<td>40.4</td>
<td>31.3</td>
<td>45</td>
<td>45.5</td>
<td>41.2</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>11-BadInfo_Lie</td>
<td>7</td>
<td>19.1</td>
<td>6.3</td>
<td>20</td>
<td>36.4</td>
<td>14.7</td>
<td>17.1</td>
<td>10</td>
</tr>
<tr>
<td>12-BadInfo_Lie</td>
<td>110</td>
<td>45.7</td>
<td>40</td>
<td>45</td>
<td>54.5</td>
<td>58.8</td>
<td>51.4</td>
<td>50</td>
</tr>
<tr>
<td>9-BadInfo_Truth</td>
<td>9</td>
<td>23.4</td>
<td>12.5</td>
<td>35</td>
<td>18.2</td>
<td>41.2</td>
<td>31.4</td>
<td>30</td>
</tr>
<tr>
<td>10-BadInfo_Truth</td>
<td>49</td>
<td>46.8</td>
<td>56.3</td>
<td>45</td>
<td>36.4</td>
<td>41.2</td>
<td>45.7</td>
<td>40</td>
</tr>
<tr>
<td>11-BadInfo_Truth</td>
<td>7</td>
<td>57.4</td>
<td>75</td>
<td>65</td>
<td>18.2</td>
<td>73.5</td>
<td>64.3</td>
<td>70</td>
</tr>
<tr>
<td>12-BadInfo_Truth</td>
<td>110</td>
<td>15.2</td>
<td>26.7</td>
<td>10</td>
<td>9.1</td>
<td>17.6</td>
<td>15.7</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 40: Average selection of 'expert' translation relative to word count of scenarios, student and lecturers UAX and Spanish linguists as well as those indicating that professional status was a deciding factor for quality vs. those who didn’t think so.
Figure 82: Standard deviation vs. source word count for each scenario, lecturers UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator.
2. **Industry Linguists (Figure 83):**

- In NoInfo, a correlation between word count and levels of agreement can be observed. For scenarios with low word count (1-NoInfo and 4-NoInfo), higher levels of agreement were achieved than for scenarios with high word count (2-NoInfo and 3-NoInfo);
- In GoodInfo, a correlation between word count and level of agreement that is similar to that of the lecturer group can be observed. Here, too, the two scenarios with low word count also show lower levels of agreement among participants (6-GoodInfo and 8-GoodInfo). The two scenarios with high word count show higher levels of agreement (5-GoodInfo and 7-GoodInfo);
- Finally, in NoInfo, no correlation between word count and levels of agreement can be observed. Both the two scenarios with low word count (9-BadInfo and 11-BadInfo) and those with high word counts (10-BadInfo and 12-BadInfo) are associated with one higher/lower level of agreement each.

3. **3rd and 4th year students (Figure 84):**

- In NoInfo, a correlation between word count and levels of agreement can be observed similar to that of the industry linguists. For scenarios with low word count (1-NoInfo and 4-NoInfo), higher levels of agreement were achieved than for scenarios with high word count (2-NoInfo and 3-NoInfo);
- In GoodInfo and NoInfo, no clear patterns can be observed.

4. **1st and 2nd year students (Figure 85):**

For this respondent group, no correlation between word count and levels of agreement can be observed for any of the scenario types.
Figure 83: Standard deviation vs. source word count for each scenario, Spanish linguist respondent group. BadInfo refers to the information provided to participants, not the actual translator.
Figure 84: Standard deviation vs. source word count for each scenario, 3rd/4th year student UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator.
Figure 85: Standard deviation vs. source word count for each scenario, 1st/2nd year student UAX respondent group. BadInfo refers to the information provided to participants, not the actual translator.
8.2. Experiment 2: Price of Translation, German Linguists

Experiment 2 (Cost of translation, German) was aimed at gaining insights into the impact that information on payment would have on perceived quality of translations. For this experiment, 17 German linguists (Section 7.1.2.) were presented with English segments and four German translations. Each of the translations was done by either

- a low cost (5 cent per source word),
- medium cost (10 cent per source word),
- high cost (15 cent per source word) or
- unpaid volunteer translator.

As with Experiment 1, participants were given 4 scenarios without information on who had provided the translation (NoInfo), followed by 9 scenarios where information on price of translation was provided. These 9 scenarios were mixed, with 4 providing correct information on price of translation (GoodInfo) and 5 incorrect information (BadInfo).

The following sections report on the results of this experiment. This includes a description of the selections made by the participants based on information provided, agreement levels and the results in relation to attitudes expressed in the questionnaire.

8.2.1. Selections and performance

As a first step, the percentage of selection for each option based on information on pricing was calculated for each type of scenario.

When no information was provided (Figure 86), participants selected

- the volunteer translation as the best in three scenarios (Scenarios 2, 3 and 4) and
- the low-cost translation in one scenario (Scenario 1).
- High cost translations were in second position in three of the four scenarios (Scenarios 1, 2 and 3).

When the correct information on pricing was provided (GoodInfo, Figure 87),

- high cost translations moved up in preference, and became the most
favoured option in three out of four scenarios (Scenarios 5, 7 and 8).

- However, in two of the scenarios the high cost option was tied with the volunteer option (Scenarios 5 and 7).
- In one scenario, low cost and volunteer were tied as the most frequently selected option (Scenario 6).

When incorrect information was provided (BadInfo_Lie, Figure 88; See Figure 89 for the correct source of the translations),

- what was supposedly the high cost translation was selected by the majority of participants in two scenarios (Scenarios 10 and 13), and
- the supposedly medium cost translation in three scenarios (Scenarios 9, 11, 12).
- In 4 of the 5 scenarios, the option that was favoured by participants had actually been provided by the volunteer translator (Scenarios 9, 10, 12 and 13) and one (Scenario 11) by a high cost translator.
- In addition, Scenario 9 was a near tie between the high cost and volunteer translators' versions.
Figure 86: Selections made when no information on cost of translation was provided, German linguist participant group
Figure 87: Selections made when correct information on cost of translation was provided, German linguist participant group
The same conclusions can be drawn when selection percentages are averaged across each scenario per scenario category (Figures 90 to 93).

- When no information was provided, participants clearly favoured the volunteer translation.
- When correct information was provided, there was a near tie in preference between volunteer and high cost translation.
- When incorrect translation was provided, the supposed medium cost translation was favoured, closely followed by the high cost translation.

However, when considering who had actually provided the translations, the volunteer translation was selected on average more often than any other option, followed by the high cost translation.

It is also worth noting that in none of the individual scenarios or the averages, the low-cost translation was favoured and in fact ranked towards the bottom in most of the results.
Figure 88: Selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the information the participants were given.
Figure 89: Selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the actual cost of translation. (Note: Since the chart is sorted with ascending values, the order of labels has been changed relative to figure 88).
Figure 90: Average selections made when no information on cost of translation was provided, German linguist participant group.
Figure 91: Average selections made when correct information on cost of translation was provided, German linguist participant group.
Figure 92: Average selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the information the participants were given.
Figure 93: Average selections made when incorrect information on cost of translation was provided, German linguist participant group. This figure shows the results based on the actual cost of translation.
These results are partially aligned with the questionnaire results and partially contradicting them. One question asked participants to rank translators with different price points according to the quality the participant would expect they deliver. Figure 94 shows the results of this question.

This analysis only includes those 17 respondents who participated in the questionnaire and the experiment. The results show a correlation between the highest level of quality that the participants expected from the high cost translator, placed the medium translator in second position, and expected the low-cost translator to deliver the lowest quality. This attitude can be partially observed in the GoodInfo scenarios, where the high cost translation was favoured more strongly than in the NoInfo scenarios.

However, in the questionnaire answer, participants were divided on where to place the volunteer translation. While none ranked the volunteer in first position, a near equal number placed the volunteer in positions 2, 3, 4. During the experiment, the volunteer translation ranked highest when no information on translation pricing was provided, indicating that the translations themselves appealed to a large number of participants. However, when the correct information was provided, volunteer translations nearly tied with the high cost translations. For BadInfo scenarios, when considering the information provided to participants, high cost translations again are favoured, together with medium cost translations. When considering who had actually provided the translation, volunteer translations and high cost translations were favoured.

It should also be noted that when presented with a list of factors that could impact quality and asked to pick the three they thought impact quality the most, none of the respondents selected price of the translation as one of their top three factors.

Unfortunately, the number of participants who had never provided voluntary translations was too small to make a feasible comparative analysis between groups of those who had volunteered and those who hadn’t.
Figure 94: Ranking of quality expectations, German linguists who participated in experiment 2
8.2.2. Levels of agreement

As already discussed in section 8.1., here, too standard deviation and the mean standard deviation were calculated to determine levels of agreement. High standard deviation would indicate high levels of agreement (Figures 95 and 96).

NoInfo and GoodInfo scenarios both show lower levels of agreement than BadInfo scenarios. This could also be observed in the break-down of the individual scenarios (Figures 86 to 89), with a difference that should be noted: Providing correct information on price of translation resulted in two 'winning' options being tied for several scenarios, whereas providing no information led to a clear 'winner' with the three 'not winners' following closely behind. This indicates that while agreement was similarly low for both, NoInfo and GoodInfo scenarios, the nature of the disagreement slightly differed.
Figure 95: Standard deviation in selections for German linguist, broken down by scenario
Figure 96: Average standard deviation in selections for the German linguist respondent group, broken down by scenario type. BadInfo refers to the information provided to participants, not the actual translator.

8.3. Experiment 3: Price of Translation, Buyers/ Clients

Experiment 3 (Price of translation, buyers/clients) was aimed at buyers of localisation services in the industry, as well as non-profit clients requesting services on Trommons. For this experiment, participants were shown a series of scenarios with four reviewer comments each, and were asked to indicate how likely or unlikely they were to trust or select the translation based on the comment.

- First, they were shown only scenarios with no information (NoInfo),
- then scenarios with information on the cost of the translations,  
  - some of which was correct (GoodInfo), and  
  - some incorrect (BadInfo).

The different reviewer comments were presented several times, combined with differing information on price point. Incorrect and correct information scenarios were mixed, which means that sometimes the participant saw the BadInfo scenario first and sometimes the GoodInfo one. Participants could select likely/neutral/unlikely once for each option, but were not restricted by having to select each once. As a result, they could
indicate that they were for example likely to use all four options provided in a scenario.

Experiments 1 and 2 benefited from the fact that participants were linguists who were able to evaluate the translated texts themselves, thereby relating the evaluation of internal attributes of the translations to the extrinsic information (professional status of translator or price of translations). In Experiment 3 however, the targeted participant group involved the added challenge for study design that participants could be expected to speak a wide range of language combinations, or indeed only one language, in addition to not being trained in evaluating translations, as the linguist participants in Experiments 1 and 2 would have been. Therefore, the experiment was designed to capture the realities of many decision makers for localisation services who, as described in chapter 3 have to rely on a third party to provide an evaluation of the translation. The buyer or client will then be able to base their decisions on the reviewer's comments in combination with extrinsic factors, one of them being the price of the translation.

This third-party evaluation approach adds variability in each of the different options per scenario which is difficult to control for. It can, however, have a great impact on the outcomes, since even slight differences in wording can signal different ideas to participants.

For this reason, the analysis of Experiment 3 had to remain on a higher, less detailed level than was possible Experiments 1 and 2 and involved the following adjustments:

1 - Results were averaged across scenarios for the scenario types instead of considering each scenario individually. Since scenarios were repeated across NoInfo, GoodInfo and BadInfo categories, averaging the results partially accounts for the variability introduced through the use of third party reviews.

2 - Another aspect of using third party reviews was that there was little benefit to be gained from distinguishing between scenarios where correct information or incorrect information was provided. Therefore, GoodInfo and BadInfo results are averaged to only 'Info', contrasting those scenarios where information was provided with those where no information was provided. NoInfo becomes the baseline for each respondent group and the analysis is focused on the extent to which the introduction of information on price points
changes the choices made by participants. However, having conducted the experiment to include both, GoodInfo and BadInfo variations is still beneficial since this lowers the possibility that individual wording of comments has impacted the results unduly.

3 - As outlined in section 6.5.1.3., each reviewer had also rated the translation between one and 5, with 5 being the highest rating. This meant that during experiment design it was possible to group only positive, neutral or negative comments into each of the scenarios. While the participants had been shown a total of 15 scenarios, only the 9 that were clearly negative/neutral/positive were included in analysis. While including the additional 6 scenarios in the experiment itself provided the benefit of not making the negative/neutral/positive combinations overly visible, their mixed nature (positive, neutral and negative reviewer comments within the scenario) adds variability that makes them not suitable for being included in analysis.

8.3.1. Experiment 3: Price; Non-profit Clients

8 respondents in the non-profit clients group on Trommons agreed to participate in the experiment. When provided with no information on pricing (Figure 97), the selections for 10 and 15 cent per word were relatively balanced across 'likely', 'neutral' and 'unlikely', with 'unlikely' being selected slightly less often than the other two options for both price points. Most participants selected 'likely' for the 5 cent translation, but 'unlikely' was relatively strong, at the expense of 'neutral', which was selected by only 8% of responses. The unpaid volunteer's version was the only one, that more participants indicated they were unlikely to trust or select based on the reviewer comments.

This picture changes in scenarios where information on the price of translation was provided (Figure 98). The number of participants who indicate they would 'likely' select the volunteer translator based on the reviewer comments more than doubles, while the numbers of those that are 'neutral' or think it 'unlikely' drops. Respondents are more neutral towards the 5 cent translator, with 'likely' and 'unlikely' selections dropping and 'neutral' increasing. Selections of 'likely' for the 10 cent translator however increase slightly, and selections of 'unlikely' more than double for the 15 cent translator. Considered together, given the price
of translations, this participant group is more likely to trust or select the volunteer translator, and less likely to do so for the 15 cent translator while being relatively neutral or uncertain towards the 5 and 10 cent translators.
Figure 97: Average selections when no information on cost of translation was provided, non-profit client respondent group.
Figure 98: Average selections when information (correct and incorrect) on cost of translation was provided, non-profit client respondent group
8.3.2. Experiment 3: Price; Industry Buyers

7 respondents in the industry buyer group who responded through the open call agreed to participate in the experiment.

Figure 99 shows the results of selection averages when no information was provided. Here, participants favoured the 5 cent and 15 cent translations, which received the majority of selections, and were relatively neutral towards the volunteer translation, based on reviewer comments. 10 cent translations received an equal selection of 'likely', 'neutral' and 'unlikely'.

When information on price of translation was provided (Figure 100), industry buyers strongly favoured the unpaid volunteer, which doubled in number of 'likely' selections, while selections of 'likely' for the 5 cent translator nearly halved in favour of 'neutral' which was increased from 9.5% to 40.5%. While 10 cent remained relatively stable, 'likely' selections of the 15 cent translator dropped by nearly three-quarter, while 'unlikely' selections were tripled. This suggests that when participants were told the price of translations, this group was more likely to trust or select the volunteer translator, less likely to trust or select the 15 cent translator and more neutral towards the 5 cent translator, while the 10 cent translator saw slight increases in likelihood of being selected but remained relatively stable.
Figure 99: Average selections when no information on cost of translation was provided, professional industry buyer respondent group
Figure 100: Average selections when information (correct and incorrect) on cost of translation was provided, professional industry buyer respondent group
Chapter 9: Summary and Discussion

9.1. Summary and Discussion

The following sections provide a summary and discussion of the findings, grouped into themes, followed by a section concluding this thesis and describing its original contributions as well as limitations. Four main points will be considered:

1. **Measurement knowledge of research participants.** As outlined in chapter 2 of this thesis, measurement knowledge is the customer’s ability to assess attribute performance relative to other, objective measures like instrument reading or expert consensus (Golder et al. 2012, p.9). Since measurement knowledge of intrinsic attributes is strongly linked in the literature to the extent to which extrinsic cues are used to evaluate quality (Zeithaml 1988, p.9; Deval et al. 2012, p.1185; Golder et al. 2012, pp.10–11), an understanding of measurement knowledge within the different participant groups is essential to be able to examine the results of the experiments conducted for this research meaningfully.

2. **Definitions of quality in different research participant groups.** As outlined in chapter 3, expectations towards quality not only differ between academia and practitioners, but also within different practitioner groups (O’Brien 2012; Drugan 2013, p.37). In addition, quality evaluation and strategies for managing and producing quality need to satisfy a range of goals, including prevention of future issues, prediction of quality and business goals. However, generalisation for different groups is not possible based on the literature. Therefore, differences in quality definitions between the different participant groups need to be considered in order to understand whether differences in perception might be connected to differing definitions of quality.

3. **Attitudes towards and opinions of factors impacting quality, including attitudes/opinions related to crowdsourced localisation.** As outlined in chapter 2, so-called ‘naive theories’ (Deval et al. 2012) can impact how a
product is evaluated, since they provide theories that explain how an extrinsic cue that is being used to evaluate quality, especially when measurement knowledge is low, is linked to quality. For example, low price could signal either good value or low quality. Low quality might be due to inferior ingredients or quality management (Deval et al. 2012, p.1189; Zielke 2014, p.328). Therefore, understanding the ‘naive theories’ of research participants in relation to distinguishing aspects between ‘professional’ and crowdsourced translations provides further insights into how quality is perceived.

4. Finally, this chapter will discuss the results of the experiments, bearing in mind the results of the first three points.

Key findings will be described and discussed in greater detail in the following sections.

- **KF 1:** ‘Expert buyers’ had greater measurement knowledge and detailed strategies available for lowering quality uncertainty than ‘non-expert buyers’.
- **KF 2:** Training increased ability to express quality concepts.
- **KF 3:** While there was wide agreement on some aspects of quality, buyers and linguists recognised that there are different types of quality.
- **KF 4:** Process-based quality bias could be observed in research participants, where the importance of factors impacting quality was given to those parts the individual added to the localisation process.
- **KF 5:** While for other extrinsic cues greater agreement on how to interpret those cues in relation to expected quality could be observed, the extrinsic cue ‘voluntary/crowdsourced localisation’ introduced uncertainty over where to place the translation on the ‘expected quality continuum’.
- **KF 6:** Despite this uncertainty, such initiatives were generally viewed in a positive light.
• **KF 7:** Participants with high measurement knowledge were less strongly influenced by extrinsic cues than those with lower measurement knowledge for extrinsic cues that were easy to place on the ‘expected quality continuum’.

• **KF 8:** If the extrinsic cue was harder to place on the ‘expected quality continuum’, interpretation uncertainty led to split results where some respondents were more heavily influenced than others.

• **KF 9:** Measurement knowledge was only an important factor in how strongly participants were influenced by extrinsic cues when sufficiently high amounts and quality of intrinsic attributes were available.

9.1.1. Measurement Knowledge

Section 3.4. in the literature review applied the concept of measurement knowledge to the localisation process chain by discussing factors that would increase or decrease measurement knowledge for different parties in the localisation process, and could therefore serve as a way of predicting to what extent those individual parties are susceptible to extrinsic cues when evaluating quality.

Here, the conclusion was that while tendencies could be predicted in that client-side vendors would likely tend towards lower measurement knowledge than vendors, and that linguists would likely have the highest levels of measurement knowledge on average, it is not possible to generalise across these groups without having further information. For the purpose of this research, specialisation (or lack thereof) in the field of localisation, localisation training/education and experience were considered primary factors in determining measurement knowledge, while also acknowledging that other factors, which were not considered further, might play a role.

Based on data collected in questionnaires and interviews during this research, the following sections will discuss the different levels of measurement knowledge for the respondent groups. The main points considered are:

• Demographic data on training, experience and level of specialisation

• Level of detail provided when describing quality
Strategies for addressing information asymmetry, especially in the participant groups without linguistic knowledge (professional industry buyers and non-profit clients)

9.1.1.1. Professional Industry Buyers vs. Non-profit Clients

Chapter 3 discussed expected differences in measurement knowledge based on training, specialisation and experience as well as the distinction between domain knowledge, linguistic knowledge, and production knowledge.

For this research, a clear distinction between the groups with linguistic knowledge (linguists, students and lecturers) and those without relevant linguistic knowledge (professional industry buyers and non-profit clients) can be made. This distinction is based on the different nature of activities undertaken by those groups: Where linguists and students/lecturers primarily engage in activities that involve translating texts or reviewing translations by comparing two language versions, the professional buyers and non-profit clients primarily engage in requesting and receiving localised content, commonly in multiple languages.

9.1.1.1.1. Demographic Data on Training, Experience and Level of Specialisation

This distinction between groups with linguistic knowledge and those without was partially confirmed for the two participant groups (industry buyers and non-profit clients) in this research, with only some respondents in the non-profit client group indicating that they only worked with languages they also spoke themselves (Sections 7.2.1.2.1. and 7.2.1.2.2.). This indicates that in the majority, both groups shared that they usually didn’t have linguistic knowledge for the projects they managed. This is congruent with both the observation that clients would commonly outsource language competencies they don’t have themselves and that project managers on the vendor side are commonly not specialised in one language combination (Pielmeier and DePalma 2013, p.11).

However, the two groups differed in their measurement knowledge based on training, specialisation and education:
In the group of professional industry buyers invited through an open call, (Section 7.1.3.2.) nearly a third of the respondents had no localisation related training or education, and 42.9% in this group had not received training in reviewing or assessing quality. For 15.4% the highest level of education was in linguistics or translation.

However, this group was highly experienced in the field (Figure 28), with nearly 85% of the respondents having 6 to 10 years of experience in localisation.

This validated the expectation prior to the study that this group would include at least some experienced and specialised vendor-side buyers who could then be compared with the anticipated less experienced clients among the Trommons non-profit organisations.

In the group of non-profit clients on Trommons, (Section 7.1.3.1.) half of the respondents had very little experience with managing localisation projects. Less than a quarter had one or more years of experience (Figure 26) and only 11.8% had worked in linguistics or translation.

Only 17.6% had training or education in linguistics/translation, but for none of the respondents in the group was this their main area of training. However, more than 40% had received training in reviewing translations and assessing texts.

In addition, the responses indicate that the majority of participants were involved in translation projects on Trommons, but that translation was only a secondary responsibility in terms of time spent on managing such projects, experience in related areas, training.

This also confirmed the expectations towards the non-profit clients on Trommons prior to research. Non-profit organisations on Trommons had been invited for two primary reasons: 1) due to the anticipation that this group would include decision makers with low measurement knowledge and 2) due to their involvement in a crowdsourcing initiative.

9.1.1.1.2. Level of Detail Provided when Describing Quality

Those differing levels of measurement knowledge based on demographic information were loosely supported in questions surrounding defining and evaluating quality.

As discussed in chapter 3, quality is not clearly and universally defined in the
localisation field. However, based on the assumption that with experience and practice, which could be gained also through training programmes, individuals can improve on making tacit knowledge explicit (Eraut 2000, pp.118–120), then the level of detail individuals provide when asked about quality definitions can serve as an indicator of how clearly the concept of quality is formulated in their own mind. This can then serve as a foundation for estimating measurement knowledge based on the assumption that more clearly defined expectations would allow for higher certainty on whether those expectations have been met.

In the professional buyer group (Section 7.2.1.1.3.), only one-third provided just one aspect to describe quality. During the interviews with professional buyers, interview participants provided very detailed and differentiated descriptions of quality, which included:

- considerations of the purpose of the project, timelines, target audience and the complexity of the content, indicating great ability to make tacit knowledge explicit (Eraut 2000, p.118) as well as
- high levels of expertise as displayed in the ability to apply knowledge to a range of situations (Dreyfus and Dreyfus 1980, p.12).

In contrast are the non-profit clients where 62.5% described quality using one aspect (Section 7.2.1.1.2.). This suggests that professional buyers have clearer definitions of quality spontaneously available to them than non-profit clients and supports the claim of lower measurement knowledge among non-profit clients made based on the demographic information.

A potential limitation of these findings is that participants might have been rushed while answering the questionnaire and therefore could have decided to provide only a brief description. However, there is no indication that one group would be more pressed for time than the other, although it is possible that lower interest levels among the non-profit clients, which were also reflected in low response rates, had an impact.
9.1.1.1.3. Strategies for Addressing Information Asymmetry

As described, research participants among professional industry buyers appear to have higher measurement knowledge than non-profit clients based on demographic information as well as the level of detail provided to define quality. However, in multilingual projects both of those groups would likely lack linguistic knowledge as outlined in chapter 3. When considering how the two groups address the information asymmetry inherent in the localisation process and having to make a decision on whether to accept this content, further differences were observed.

While both groups mentioned asking others to review the translations if they couldn’t review them themselves, a common theme among the non-profit clients (Section 7.2.1.2.1.) was that they would trust the translators or the process if they had no one to check the translations for them. In addition, only two of the respondents in this group (which included the only respondent who was also a trained linguist) described more sophisticated approaches to quality evaluation beyond asking another person for a review.

In contrast, professional buyers (Section 7.2.1.2.2.) described more complex and sophisticated approaches to quality like ISO certification, as well as long-term, pre-translation and post-translation strategies. Pre-translation and long-term strategies involved selection of translators who were vetted based on experience, reputation and previous performance while post-translation strategies involved detailed sets of criteria used for reviews and quality checks, conducted by project managers or trusted reviewers, with potentially another reviewing step added if there was disagreement. These strategies contribute to measurement knowledge in that the individual buyer can use technology to evaluate universal and preference attributes to an extent. In addition, it can lower ‘will’ uncertainty by relying on trusted resources.

9.1.1.3.1. Third-party review as an evaluation strategy

Those differences in approaches also provide insights into why third party review as a way of addressing information asymmetry can be difficult and the source of conflict between the different parties involved in the localisation process (Kelly and DePalma 2009b; Txabarriaga 2009; Drugan 2013, p.42).
As reported in section 7.2.1.1.4., linguist participants commonly use their own criteria for review. This would likely be especially common in a situation where a client requests review from a translator without providing further details on how the quality should be evaluated and holds potential for conflict since translator and reviewer would likely disagree on at least some of the attributes of the translation.

As a result, the client would then find themselves in a situation where they would have to decide which linguist to trust without being able to evaluate the translation themselves to form their own opinion. This situation would be especially difficult if disagreement is centred on idiosyncratic attributes since it can't be easily resolved by referring to codified bodies of knowledge or to the client's preferences.

On the other hand, professional buyers seem to be addressing this issue by not only having a more detailed concept of quality and sophisticated evaluation strategies, but also by requesting review from trusted and established linguists. This difference was also highlighted by one of the linguists interviewed who described that providing a direct client with review was often difficult since the client didn't have enough language and localisation knowledge.

In contrast, the same interview participant also said that it was easier when working with a vendor since their reviewer would often have the last word. This strategy was also reflected in an interview with a professional buyer who described their agency as providing 'signature quality' to clients by conducting in-house review with established linguists.

When using these strategies, professional buyers have not only recognised the information asymmetry that is built into the localisation process chain, but have also moved towards lowering uncertainty for themselves and improving their ability to decide whether to trust a translation despite their own low ability to measure all attributes of the localised content themselves. Through this, as well as other strategies like vetting freelancers and including quality checks in the process, vendors add value for the client regarding quality. By consistently applying and advertising those strategies, they are in a position to lower the customer’s ‘will’ uncertainty if they can establish themselves as a shorthand for quality, similarly to an established, trusted brand signalling product quality (Zeithaml 1988, p.7).
Observations are made in the literature that describe client-side buyers as unsure how to get quality when they don’t speak the language (Sargent and DePalma 2007, p.19). These are congruent with results of this research based on demographic information on training, experience and specialisation as well as levels of detail when describing quality and strategies to address information asymmetry.

Similarly, it has been suggested that greater information asymmetry would exist for buyers where localisation is not the core competency of their organisation (Dunne 2006, p.101; DePalma 2012, p.1). These results also support Golder et al. (2012) when they state that firms are more capable of measuring attribute performance accurately due to greater access to measurement instruments, or in the case of this research measurement strategies as described previously on the section of addressing information asymmetry, than the typical customer (p.5).

However, these results cannot be generalised to the wider group of professional industry buyers or client-side buyers and should instead be considered as indication that these two groups for this research represent two types of buyers: those with high measurement knowledge and those with lower ability to assess quality, express opinions and assess localisation quality, placing them more on the side of a localisation lay person (Havumetsä 2012, p.56). While these findings have implications for the results of this research, it needs to be acknowledged that there still exists variability within those participant groups and that in the wider context of the localisation industry, even greater variations likely exist.

**KF 1:** This finding was that ‘expert buyers’ had greater measurement knowledge and more detailed strategies for lowering quality uncertainty than ‘non-expert buyers’.

**9.1.1.2. Linguists vs. Students and Lecturers UAX**

**9.1.1.2.1. Demographic Data on Training, Experience and Level of Specialisation**

**9.1.1.2.1.1. Linguists invited through open call**

Linguists were invited through an open call to participate in the studies because of the expectation that they would have the highest measurement knowledge of all groups,
especially when presented with experiments for which they had the matching linguistic knowledge (Spanish or German). However, it could not be predicted whether primarily trained and/or experienced linguists or primarily novice linguists would respond to the open call, making a firm prediction on measurement knowledge for this group impossible, prior to conducting the research.

Based on demographic data collected through questionnaires for this research (Sections 7.1.1. and 7.1.2.), German and Spanish linguist respondents could be considered primarily ‘professional’ translators. In these groups, the majority of respondents had received tertiary education in a linguistics or translation related area, with only a small number having no education in the field, although the number of respondents with no formal education in the area was greater for the German respondent group than the Spanish.

The ‘other linguists’ group (Section 7.1.4.) included respondents who primarily identified as linguists but who didn't speak Spanish or German to a sufficiently high level to be included in one of the experiments. In this group, all respondents had training or education in localisation.

The Spanish and German linguists consisted primarily of experienced translators, with only small numbers having been involved in localisation for fewer than 12 months and nearly half (Spanish, Figure 22) or nearly two-third (German, Figure 25) having more than 6 years of experience. Similarly, in the ‘other linguists’ group, more than half had 6 or more years of experience in localisation. However, 15.9% had less than 12 months of experience.

The third aspect indicating measurement knowledge based on demographic data is that of specialisation. All participants in these groups had self-identified as primarily being linguists (for-profit or non-profit translators/reviewers), indicating high levels of specialisation.

As outlined in chapter 4 of this thesis, the challenge many professional linguists face is that in an industry where status boundaries for professional vs. non-professional translators are fuzzy, they need to be able to signal trustworthiness by being considered a ‘professional’ since the person paying for the service can commonly not easily evaluate the value of a translation (Pym et al. 2016, p.33). This can be achieved through attested
experience, reliability or academic and professional training (Koskinen and Dam 2016, p.263; Orrego-Carmona 2016, p.2; Pym et al. 2016, p.33). Based on the demographic data gathered, the three linguist respondent groups could in the majority therefore be considered to be ‘professional’ translators, at least to the extent they identify as primarily paid-for translators if the question of payment was to be considered an additional factor in determining status.

However, it is unclear what importance to assign the individual factors:

- specialisation,
- training,
- experience,
- payment.

While experience might have more market value than training (Koskinen and Dam 2016, pp.258–259), the Dreyfus and Dreyfus (1980) model of skill acquisition (p.12) suggests that training might provide an important foundation on which, through experience and specialisation, individuals can build in order to reach expertise. This is also supported by the observation that non-professional translators improve as translators over time (Barra 2009; Dwyer 2012; Massidda 2012 - all in Orrego-Carmona 2016, p.2), and none of those factors can be reliably linked to quality output.

This suggests that solely based on the self-reported data described it would be impossible to decide whether, for example, a translator with no formal training but several years of experience or one with formal training but less experience should be considered more of a ‘professional’ translator and whether either of those combinations would indicate higher measurement knowledge. In addition, it is not clear how many years of experience would be necessary to ‘make up’ for a lack of formal training and whether this would be a general principle or whether individual factors would play an additional role.

However, as the following section on the level of detail provided in describing quality suggests that formal training appears to play a role in at least some aspects of measurement knowledge.
9.1.1.2.1.2. Students and lecturers at UAX

As outlined in section 5.1., students and lecturers at the University Alfonso X el Sabio in Madrid (UAX) were invited to participate. The expectation was that this group would most purely reflect three main stages of a translator's career:

- from novice,
- to trained translator,
- to expert (trained + experienced) translator.

The novice translator group consisting of 1st/2nd year students had at that point limited training and none of the group members had professional experience. The trained translator group (3rd/4th year students) had higher levels of training or had nearly completed their training in the case of 4th year students, but had little or no experience in translation. Participants among the lecturers primarily selected the categories ‘1 to 5’ or ‘6 to 30’ years of experience in professional translation and more than three-quarter of the lecturers also had received training in reviewing translations, either at an academic institution, non-academic institution, or a for-profit company (Section 7.1.1.1.; Figures 18 and 19).

Based on these factors, measurement knowledge can be assumed to be high in lecturers, medium in trained translators and low in novice translators, which is also shown in the following sections on the level of detail participants provided when describing quality.

9.1.1.2.2. Level of Detail Provided when Describing Quality

The impact of training on the ability to make tacit knowledge explicit (Eraut 2000, p.118) could be observed in the differences between the 1st/2nd year students and the 3rd/4th year student respondent group. When asked to provide a definition of quality in their own words, the lowest amount of detail was provided by the 1st/2nd year students, who nearly exclusively provided one aspect to define quality. In contrast, for 3rd/4th year students, similarly to lecturers, half of the respondents provided more than one aspect of quality (Section 7.2.1.1.1.).

This suggests a higher impact on the ability to make tacit knowledge explicit of training rather than experience, since no noticeable difference between trained (3rd/4th year
students) and trained + experienced (lecturers) groups could be observed. However, due to the overall limitations of this study which included only a small number of participants and was only conducted in one single university department, those results should not be overstated and require further investigation to be confirmed.

This limitation might also explain why more than two-third of the linguists, who in the majority were both trained and experienced, provided more than one aspect when defining quality, while only half of the lecturers did so. In addition, the linguist group included respondents who provided up to 6 aspects to describe quality, while across all groups those who provided more than one aspect did generally not use more than three or 4 aspects to define quality. Since nearly ten times more linguists than lecturers provided a definition of quality, equal sized groups might have yielded similar results based on the observation that both groups were trained, experienced and specialised in translation. Since both groups can be expected to commonly describe quality issues, either by reviewing students’ assignment in the case of the lecturers or by reviewing translations in a commercial setting, no explanation immediately lends itself other than variability due to size of respondent group.

Whether training or experience should be considered to have greater impact on measurement knowledge and the ability to make tacit knowledge explicit, differences between those groups with overall low measurement knowledge (1\textsuperscript{st}/2\textsuperscript{nd} year students) and those with higher measurement knowledge (3\textsuperscript{rd}/4\textsuperscript{th} year students, lecturers, linguists) could be observed. These differences would likely play a role in the third-party review process, which will be discussed in more detail in the following section.

\textbf{9.1.1.2.3. Strategies for Addressing Information Asymmetry}

\textbf{9.1.1.2.3.1. Third-party review as an evaluation strategy}

Section 9.1.1.1.3. already described differences between industry buyers with high measurement knowledge and non-profit clients with low measurement knowledge, when considering the level of information asymmetry and strategies to address this asymmetry.

Interestingly, not only professional buyers have recognised and are addressing the
information asymmetry but also linguists. A recurring comment from respondents was that reviewers would have to be able to justify their changes and provide explanations, as well as evidence to back up their decisions. This would likely be beneficial in cases where a buyer isn’t able to evaluate quality, since it would help the reviewer appear more trustworthy if they provide such notes on their changes. In the context of justifying changes and providing explanations, the ability to make explicit the knowledge of what is good quality localisation would play an important role and would place groups like the industry linguists, trained students and lecturers in an advantaged position. This suggests that if the status of ‘professional’ translator is meant to signal quality (Zeithaml 1988, p.9), the factors of training and/or experience would be suited to make a distinction that indicates added value to the localisation process. An example could be improving the third party review process, which a distinction purely based on remuneration (Pérez-González and Susam-Saraeva 2012, p.151; Orrego-Carmona 2016, p.3) would likely not be able to provide.

However, buyers might not always want to get involved in overly granular reviews and "academic nitpicking" (Kelly and DePalma 2009b, p.5). While one interview participant in the linguist group described that depending on the client preferences, there were two ways to review a text. One included only changing what was necessary and the other also making preferential changes. The extent to which a client would like to get involved with the review appears to vary. In cases where this is not specified, it would then be difficult for a linguist to determine which approach to review to take and how much detail on justifications to provide.

A common theme among the linguist groups was that some aspects of quality are more easily measurable than others, which matches the classification of attributes as proposed by Golder et al. (2012) and discussed in section 2.3.1. Yet, more than two-third of the linguist respondents indicated they use their own criteria for evaluation, while 38% use criteria they have adapted to suit their needs. The respondent group was also split over the question of the extent to which others would agree with their evaluation. The prevalence of linguists’ own criteria for evaluation could partially explain the difficulties experienced by all parties involved in third party review experience (Kelly and DePalma 2009b, pp.4–5; Drugan 2013, p.9). It leaves the decision of
• what quality is,
• what aspects of quality should be considered most important, and
• how to describe quality

in the hands of the individual, which would likely lead to disagreements and communication issues.

KF 2: Training increased the ability to express quality concepts. This becomes especially clear when comparing participant groups with different levels of training.

9.1.1.3. Summary of Measurement Knowledge

Based on these results, the respondent groups can be placed on a continuum from low to high measurement knowledge, depending on whether training, specialisation or experience is considered more important. In addition, linguistic and domain knowledge need to be considered.

For the experiments, assumptions about linguistic knowledge were made. For the groups with assumed linguistic knowledge,
• the group with lowest measurement knowledge would be 1st and 2nd year students,
• followed by 3rd and 4th year students due to low levels of experience and no or partial training.
• On the other end of the spectrum the researcher would situate lecturers due to their high levels of training and experience.

However, the linguist groups were quite heterogeneous in their levels of training and experience. In addition, it is hard to estimate what additional informal training they might have acquired or to what extent their experience would have increased their measurement knowledge. Therefore, the linguist group is not placed on the proposed continuum.

For the two groups with no assumed linguistic knowledge,
• the Trommons clients occupy the low end of the measurement knowledge spectrum. Assuming projects with language combinations the client does not
speak, this group would be situated below the 1st/2nd year student group due to a lack of linguistic knowledge which would weight heavier than the students' lack of experience in localisation.

- Similarly, while the industry buyers are highly experienced and many of them trained, when placed in the context of projects where they did not speak the languages in question, they would rank in the mid-range, below the linguist group although it is difficult to definitively state whether their lack of linguistic knowledge would also place them below the 1st/2nd year students or whether their experience might balance out a lack of linguistic knowledge to a greater extent.

**9.1.2. Defining Quality**

As outlined in chapter 3, there is no universally agreed upon definition of quality in localisation and differences can be observed within different practitioner groups (O’Brien 2012; Drugan 2013, p.37). Additionally, quality evaluation and strategies for managing and producing quality need to satisfy a range of goals. Therefore, as part of the survey stage of this research, definitions of quality were gathered through questionnaires, interviews and content analysis.

The findings have been presented in Figure 30 (Section 7.2.1.) and suggest that across all respondent groups, the most important aspects of quality are faithful representation of meaning, style, naturalness, client needs, culture, grammar and terminology (Table 41).
<table>
<thead>
<tr>
<th>Category</th>
<th>Average % across all groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faithful/Loyal/Accurate</td>
<td>41.31</td>
</tr>
<tr>
<td>Style</td>
<td>21.20</td>
</tr>
<tr>
<td>Naturalness (idiomatic)</td>
<td>20.05</td>
</tr>
<tr>
<td>Client needs</td>
<td>13.89</td>
</tr>
<tr>
<td>Culture</td>
<td>13.66</td>
</tr>
<tr>
<td>Grammar</td>
<td>11.28</td>
</tr>
<tr>
<td>Terminology</td>
<td>10.79</td>
</tr>
<tr>
<td>Misunderstood</td>
<td>6.25</td>
</tr>
<tr>
<td>Process</td>
<td>4.86</td>
</tr>
<tr>
<td>Spelling</td>
<td>4.37</td>
</tr>
<tr>
<td>Culturally appropriate</td>
<td>4.17</td>
</tr>
<tr>
<td>Other</td>
<td>2.08</td>
</tr>
<tr>
<td>Comprehension</td>
<td>1.61</td>
</tr>
<tr>
<td>Linguistics</td>
<td>1.61</td>
</tr>
<tr>
<td>Technical</td>
<td>0.65</td>
</tr>
<tr>
<td>Addition</td>
<td>0.51</td>
</tr>
<tr>
<td>Omission</td>
<td>0.51</td>
</tr>
<tr>
<td>Attention to detail</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 41: Average mention of quality aspects across all participant groups

**9.1.2.1. Style vs. Meaning**

Golder *et al.* (2012) distinguish between attributes of a product that can be

- universal (unambiguously measurable and homogeneous in preference),
- preference attributes (unambiguously measurable and heterogeneous in preference)
- or idiosyncratic (ambiguous in measurement and heterogeneous in preference)

(p.8).

As discussed in chapter 3 of this thesis, this distinction can be directly applied to localised products. A similar distinction was made by one of the professional buyers interviewed, who described quality to include both 'technical' and 'linguistic' quality. The interviewee distinguished between the two by describing aspects of 'technical' quality to be aspects a non-native person could check by using tools to check for issues with:
• variables,
• terminology,
• typography,
• capitalisation,
• mismatching numbers and similar.

On the other hand, 'linguistic' quality includes aspects like style and how a person writes.

While linguists didn't distinguish between technical and linguistic knowledge, a recurring theme was nonetheless that some aspects of quality are objectively measurable, like grammar and spelling, while 'style' was more difficult. The fact that 'style' should be considered an idiosyncratic attribute is also supported when examining the results of the manual and semi-automated content analysis conducted on reviewer comments on Trommons. Here, stylistics stood out since the terminology used to describe such issues was to a great degree less consistent and precise than for grammar, spelling or terminology, which should be considered either universal or preference attributes, but generally heterogeneously measurable. This can be explained by the fact that such discourse elements don’t lend themselves to codification (García 2014, p.432), which would make them not only difficult to teach to novices due to their primarily experience based nature (Dreyfus and Dreyfus 1980, pp.7–12), but would likely be a source of disagreement between linguists due to a lack of shared terminology to describe such idiosyncratic attributes.

These findings were consistent with O’Brien (2012), who showed in her analysis of eleven Quality Evaluation models that the highest level of consensus was found for the category which commonly included grammar, syntax, punctuation and spelling. The ‘style’ category was present in seven out of eleven models and ruled out explicitly by one (J2450). O’Brien (2012) states that, "of the top four 'Language' errors, 'Style' is the one with the least consensus across models." (p.61). In addition, criteria for ‘style’ were frequently vague concepts like ‘tone’ or ‘awkward syntax’ (p.61)

This indicates that at least as far as the quality evaluation models benchmarked by O’Brien (2012) are concerned, such models primarily provide further structure for already codified and unambiguously measurable attributes. However, they appear to offer only
limited useful support for idiosyncratic attributes like ‘style’, which were also the attributes that based on this present research, pose the greatest challenge in the quality review process. This limited useful support might serve as a partial explanation as to why the majority of participants in the linguist group said that they use their own quality evaluation criteria, although this would require further research into reasons and approaches related to quality evaluation among linguists.

Based on these observations, the researcher suggests that the purpose of such quality evaluation needs to be defined clearly when metrics are designed. If the primary purpose is linked to business goals like demonstrating added value (Drugan 2013, pp.73–74) or being able to justify rates (Drugan 2013, p.41), including idiosyncratic attributes in the metric would introduce variability and lower the reliability of such metrics, while excluding the ‘style’ category as demonstrated by the J2450 metric (O’Brien 2012, p.61) would avoid such issues. However, if the goal is to increase quality then the inclusion of a ‘style’ category should be understood as a foundation for discussion between linguists with the goal to reach a consensus rather than a tool to gather business metrics.

Interestingly, a discrepancy emerged between quality aspects mentioned most frequently in reviewer comments compared to those observed when participants were asked directly. In reviewer comments on Trommons, stylistics were mentioned most frequently as aspects that were either positive or negative about a specific translation, while style was placed in second position or below that for the questionnaire respondents. By contrast, when respondents were asked directly how to define quality, accurate/faithful/loyal representation of meaning was mentioned most frequently by all respondent groups. In the content analysis, 'mistranslation' was then mentioned in second position. One explanation for this discrepancy could be found in the way one of the non-profit clients described quality in their interview. Here, the interviewee described the correct transfer of meaning to be the 'minimum level of quality'.

All groups that were asked to define quality gave top position to the accurate representation of meaning. This observation indicates that this aspect would be considered the 'minimum level of quality' also by these groups. In fact, this aspect of quality might be so fundamental that reviewers on Trommons didn't think it was necessary to point out
specifically when there were no issues for this category, which could explain the low number of mentions relative to other categories.

9.1.2.2. Further observations

Among both questionnaire respondents and the content analysis, aspects of quality such as grammar, spelling, terminology, process related aspects and similar generally ranked in the mid-range without any one feature showing a specific tendency. However, some interesting differences can be observed:

9.1.2.2.1. Disconnect between academic training and real-world industry needs

Definitions of quality suggested a disconnect between academic training and real-world industry needs. In the academic context (students and lecturers), nearly no mention of client requirements was made as a factor to define quality. In contrast, more than half of the professional industry buyers mentioned end-user needs or customer requirements, and ranked this factor in the upper-mid range, although opinions were somewhat split here with other professional buyers ranking customer specifications as low importance for quality. Among the linguists, customer specifications were ranked low in importance and not mentioned in the definitions of quality.

This suggests that customer requirements don't play a big role in the training environment of that particular group of students and lecturers. Although it is not possible to generalise to translator training at other institutions, the findings match Künzli (2004) who suggests that translation students need to be made more aware of the client’s role in the translation process as a partner (p.47). However, the low importance given to client requirements among the linguists in the present research is in contrast to the findings in Künzli (2004), where professional translators mitigated risk by making the client a partner in the translation process (p.44).

The finding that adherence is an important topic for the buyer-side is congruent with results reported by Kelly and Stewart (2010). Here, compliance with style guides (66.1%), terminology database consultation (75.9%) and file format standards adherence (76.8%)
were considered 'very important' by the majority of respondents (p.30).

The split opinions on where to rank client requirements within the professional buyer group, as well as the fact that half of the respondents mentioned this aspect, highlights that vendor-side clients essentially stand with one foot on either side of the debate of who 'owns' quality in localisation: On one side, linguists who, at least based on the results of this research, prioritise the text and their concepts of quality over customer requirements. On the other side, there are requirements-focused clients with greater concern for 'good enough' localisation than 'good' localisation (Section 3.3.2.).

The differences observed in the importance ascribed to client requirements in this research emphasises that not only definitions of quality, but also ownership of that definition is unclear in the localisation field.

9.1.2.2. Client distinctions between types of meaning

While all groups prioritised accurate transfer of meaning, non-profit clients often distinguished between different types of meaning important to them, indirectly highlighting the importance of domain expertise. Here, different types of faithful representation included:

- text that only needed to be understood,
- text that needed to represent original meaning,
- text that needed to represent the context of the content, and
- text that needed to represent the spirit of the organisation.

This suggests that this client group has recognised the importance of domain expertise for localisation which has also been highlighted in the literature (Sargent and DePalma 2007, p.11; Muñoz Martín 2009, p.25), and that the non-profit client group had quite clear ideas of what it is they expect from a translation, at least in these aspects. However, the descriptions of the different types of faithful representation were commonly vague in that they described concepts like ‘ensuring the spirit of the organisation is captured’, which would likely be difficult for a translator to put into practice. Here, limited ability to make such tacit knowledge explicit among the client group as discussed in section
9.1.1.1.2. would then likely lead to a situation where the client has expectations which the translator is not able to fulfil if these were not further specified in detail.

KF 3: There is not only wide agreement on some aspects of quality, but buyers and linguists have also recognised that there are different types of quality.

9.1.3. ‘Naive Theories’: Factors Impacting Quality

Previous sections in this chapter have discussed the findings that different participant groups had different levels of measurement knowledge, as well as different approaches to addressing information asymmetry and defining quality. As discussed previously, especially in situations where measurement knowledge is low, so-called ‘naive theories’ (Deval et al. 2012) provide explanations for the connection between extrinsic cues and quality. Different theories could then impact the signal of an extrinsic cue (Deval et al. 2012, p.1189; Zielke 2014, p.328). Therefore, this research not only included an investigation into measurement knowledge as previously discussed, but also into opinions of and attitudes towards extrinsic cues that were included in the experiment phase of this research.

As outlined previously, the focus of this research was on two extrinsic cues (price of localisation and professional status of translator) that could serve as a surrogate or short-hand in the absence of measurement knowledge for intrinsic attributes. However, it should not be ignored that other extrinsic cues might also play a role in perceived localisation quality. These factors were also explored in the survey part of this research which included non-profit clients, professional buyers and linguists.

9.1.3.1. The role of price in localisation

For all three groups included in this part of the research, price of translation appeared to be the least important factor to potentially impact quality (Section 7.2.2.; Figure 33). Among the non-profit clients and professional buyers, it didn't rank at all, indicating that none of the participants had selected this as an important factor impacting quality. Among the linguist group, 5% had selected it as a major factor impacting quality,
although that still placed it in the bottom three. This finding was unexpected and suggests that none of the respondent groups appear to place great weight on the maxim that ‘you get what you pay for’. This has been described in the literature as a strong factor leading consumers to overestimate the price-quality relationship on the one hand (Peterson 1970, p.525; Kardes et al. 2004, p.368; Cronley et al. 2005, p.159; Judd 2015, p.301), while at the same time allowing sellers to signal quality by using price as a surrogate to suggest quality (Zeithaml 1988, p.7).

However, while price ranked low when placed in the context of a larger list of factors that could impact quality for all groups, some greater differences between groups could be observed when only price, translator experience and translator training were ranked against each other (Figure 34). In addition, all groups seemed to place higher importance on price when fewer options were presented, which could be linked to the higher availability of other cues weakening the importance of price as an extrinsic cue (Sirohi et al. 1998, pp.227–228).

Professional buyers ranked payment slightly higher than non-profit clients. While a similar percentage of respondents thought that payment had ‘high’ impact on quality (15.38% for non-profit clients vs. 17.6% for professional buyers), just under a half of non-profit clients, but more than three-quarters of professional clients thought it had ‘some’ impact on quality. However, nearly twice as many linguists thought that payment had a high impact on quality compared to the client/buyer groups, and 59% thought it had ‘some’ impact.

The low importance given to payment among the non-profit clients is not surprising given that by definition all respondents in this group choose to work with unpaid volunteers. This likely biases the responses towards the low importance placed on payment side, and can't be assumed to also apply to client-side buyers in for-profit projects. However, it also indicates that at least the non-profit clients who responded to the questionnaire don't seem to expect low quality purely based on the fact that they aren't paying for the translations. This suggests that the non-profit client respondent group has recognised ‘free’ as a special price (Shampanier et al. 2007, p.742; Nicolau and Sellers 2012, p.243) and may have grasped the benefits of offering no rewards for a task over
offering small rewards (Deci et al. 1999; Gneezy and Rustichini 2000, p.791; Shampanier et al. 2007, pp.742–743), although this understanding may be informed by intuition and experience rather than empirical evidence.

The observation that linguists seem to place higher importance on price overall while for-profit buyers are situated between linguists and the non-profit clients can likely be explained when remembering the different positions in the localisation taken by the parties. Linguists always act as seller in this chain, which means that it is in their interest to signal a connection between quality and price of translation. This would allow them to ask for higher prices if they are able to signal quality, or conversely signal quality by asking for higher payment. In addition, the linguist would have less opportunity to compare different quality outputs depending on price than the professional buyer. Therefore, professional industry buyers would likely have a more realistic view on the connection between price and quality due to a greater awareness through regular purchasing of price ranges and their connections to quality variations, which may have an impact on the strength of price as an extrinsic cue (Zeithaml 1988, p.12). At the same time, vendors also fill a double-role where as a buyer, the goal is to purchase quality at as little cost as possible while as a seller, the goals would be similar to that of the linguist in that ideally, high price would signal high quality and high quality is able to demand higher payment.

9.1.3.2. Experience vs. Training

Across all groups, experience was ranked higher than training as an important factor impacting quality, which could indicate that the gap between academia and industry that has been described by critics of theoretical quality models (Section 3.3.) is also observed by practitioners in the field, leading them to place more value on hands-on experience. However, when considering the Dreyfus and Dreyfus (1980) model of skill acquisition which has as a first stage the novice who requires rules and guidelines, while with greater experience the individual then reaches the expertise level (p.12), this paints a less bleak picture of the perceptions of what academia has to offer practitioners in the industry. Instead, it could indicate that training is seen as a foundation, upon which experience then takes the individual to the next level towards expertise, which could explain the lower
position of training relative to experience in the perception of participants.

Additionally, based on the findings discussed in section 9.1.1.2.2., the results of this research suggest that translator training offers benefits in at least some quality production and management processes, like third party review. This offers an opportunity for translator training and education institutions to highlight the benefits of such programmes that go beyond the promise of increased employability, with a stronger focus of increased job performance and a solid foundation to build expertise on. This might be especially beneficial if, as suggested by Koskinen and Dam (2016), experience currently has indeed more market value than academic qualifications (p.259).

**9.1.3.3. Experience and training vs. Price**

The low importance given to price relative to training and experience was turned on its head when participants (professional buyers and linguists) were asked to rank different translators based on the quality the respondent would expect to receive from them, which also included voluntary translations as one of the ‘price points’. Here, the introduction of ‘voluntary’ or ‘crowdsourced’ as an extrinsic cue also introduced uncertainty in rankings, suggesting overall interpretation uncertainty due to conflicting naive theories being active.

Unsurprisingly, both, professional buyers and linguists placed highest quality expectations on highly paid translators and lowest quality expectations on low paid translators, when given no further information on the translator. While ranking of price against other factors consistently placed price towards the bottom of the list as outlined above, in a situation where no other information is provided, the prevalence of a perceived link between price and quality (Peterson 1970, p.525; Kardes et al. 2004, p.368; Cronley et al. 2005, p.159; Judd 2015, p.301) would make assigning the highest expected quality to the highest price point the most likely outcome.

However, results were more mixed for the volunteer translator. Both groups showed greater uncertainty on where to place the volunteer translator than for the paid-for translators, suggesting that a range of different naive theories were active for different respondents. This leads to differences in opinion on how the extrinsic cue ‘volunteer’ is linked to ‘quality’ (Deval et al. 2012, p.1186). This is supported by the observation that
linguists overall seemed to expect slightly lower quality from volunteer translators than professional buyers did, which is congruent with the findings outlined in later sections in this chapter that linguists view volunteer translation slightly more negatively than professional buyers do.

The results when participants were also given information on the translator's experience or training were more unexpected when considering the overall tendency to rank experience and training as more important than payment and the overall positive attitude towards volunteer translation expressed. This will be discussed in more detail in section 9.1.4.

When asked to rank translators based on payment as well as experience, both groups (professional buyer and linguist) showed great certainty about where to place the extreme ends, ranking the highly-paid translators with 5 years of experience highest, but the low paid translator and the volunteer translator with three months experience lowest. This indicates that at three months of experience, participants did not expect there to be a difference between an unpaid translator and a volunteer translator.

On the other hand, the trend was towards a higher quality expectation from a volunteer translator with 5 years of experience compared to a translator paid 5 cent per word with the same experience. This observation is congruent with Deci et al. (1999), which would support that an intrinsically motivated volunteer would perform better than a low-paid translator whose intrinsic motivation has been hindered by an extrinsic factor like low payment. At the same time it indicates that at a low experience level motivation would play a lower role in the opinion of the respondents than at a higher level of experience, suggesting that for translators at a lower level of expertise (Dreyfus and Dreyfus 1980, p.12) a perceived lack of skills might prevent a translator from providing a quality product even if they were highly motivated.

However, the observation could also be made that when the translator was paid 15 cent per word but only had three months of experience, professional buyers seemed to place this translator slightly higher than a translator paid 5 cent per word but with 5 years of experience, while linguists placed the 15 cent translator with less experience on a similar level as that of the more experienced volunteer and 5 cent translators. This suggests that at
least some respondents place higher importance on payment in connection to quality than experience, although the trend is not very strong overall.

Similar patterns were observed when participants were given information on whether the translator had received formal training or not, although slightly less pronounced, suggesting that the translator's experience may have slightly more influence than formal training.

Overall, experience seems to play a larger role than training and training a larger role than price across the participant groups. However, when voluntary translations are introduced as price point, then uncertainty of how to interpret this extrinsic cue in relation to quality expectations can be observed, placing the voluntary translation outside the price/quality continuum.

9.1.3.4. Other observations on factors impacting quality

Two additional observations on the ranking of factors impacting quality can be made:

9.1.3.4.1. Resource input ranking in top positions

Across all three groups, resource input ranked in the top positions of factors impacting quality. However, there were slight variations. Among the non-profit clients, ‘working into native language’ was ranked as the most important factor impacting quality, followed by ‘subject matter expertise’, then experience in third position and level of motivation in fourth. Interestingly, a motivated bilingual amateur with domain knowledge would check most boxes for being able to deliver quality based on these results. This supports the concept of translation as an occupation that has failed to professionalise (Chesterman 2001, p.146; Sela-Sheffy 2015, secs3476-3480) and indicates not only that the concept needs to be clarified but also increased efforts are necessary to educate non-specialist clients on the value a trained, experienced and/or specialised translator can add to a project compared to a bilingual amateur.

In contrast, professional buyers and linguists showed greater awareness of the value of training, experience and specialisation. Among the professional buyers, source text
quality ranked in first position, and in second position were factors concerning the linguist (experience, familiar with subject matter and native language - all three tied with 'quality checks performed').

Among the linguists, the top 5 factors impacting quality selected also concerned the resource input, although for this group the order was reversed. Here, greatest focus was placed on linguist experience, followed by subject matter familiarity, working into native language and in fourth position training or education. In fifth position followed source text quality

9.1.3.4.2. Different levels of importance assigned to the ‘tools of the trade’

Similarly, the three groups assigned different levels of importance to the ‘tools of the trade’ in localisation like technology and tools, quality checks, internationalisation and guidelines.

Among the non-profit clients, these all ranked in the bottom half, with only training/education and source text quality being placed slightly higher, which again indicates a lack of awareness of what specialist knowledge on localisation tools and techniques can add to a localisation project, while this awareness was greater among professional buyers and linguists.

Professional buyers positioned the process-oriented 'quality checks performed' in second position (along with three other factors that were tied for second position). In addition, in third position information and guidelines were considered important, followed by technology and tools used.

Linguists also ranked process oriented factors like quality checks performed, guidelines, tools and technology towards the mid-to bottom range of the list. In addition, linguists ranked time spent on translation in 7th position, while for non-profit clients it didn't rank at all and for for-profit clients it ranked in the 9th position.

Taken together, the observations made on the ranking of resource input and ‘tools of the trade’ suggest process-based quality bias in that each group views what is right in front of them as important in the way they manage localisation. Where linguists would believe that their input is of great importance to the quality produced, professional buyers tend
towards seeing the process as a whole. This also includes resource input and various 'tools of the trade', which are also factors where a vendor can add value to the localisation chain by managing these 'tools of the trade' as mediator between linguist and client-side buyer. Similarly, non-profit clients would likely have little knowledge of those 'tools of the trade' but would work with enthusiastic amateurs in the context of non-profit volunteering, and therefore see quality produced as part of that environment.

9.1.3.4.2.1. Professional status and quality

As described in section 4.1.3.1., the boundaries between professional and non-professional translators are fuzzy and porous. Yet, assumptions to predict localisation quality based on the distinction have been made (American Translators Association 2009; Zaidan and Callison-Burch 2011; Berg 2013; Kushner 2013; Deriemaeker 2014). While the label of ‘professional’ translator might serve as a shorthand to signal quality (Zeithaml 1988, p.7), some have also suggested that ‘non-professional’ translators might be able to provide quality as much or better than ‘professional’ translators (Pérez-González and Susam-Saraeva 2012, p.150) or that the main difference between a ‘professional’ and ‘non-professional’ would be remuneration (O’Hagan 2011a, p.14; Pérez-González and Susam-Saraeva 2012, p.151; Orrego-Carmona 2016, p.3), which would indicate little quality difference between a professional and a non-professional.

With this in mind, one part of the research had as its goal to determine the perceived connection between the professional status of a translator and quality, as well as which ‘naive theories’ (Golder et al. 2012) informed the perceived connection in different participant groups (Section 7.2.3.; Figures 51 and 52). Differences could be observed in the importance assigned to professional status: While in the academic context (students and lecturers) as well as among linguists greater importance was given to the translator’s professional status in connection to expected quality, this was less the case for professional buyers and non-profit clients.

Among the different groups, the students and lecturers were most likely to think that there was a correlation between whether an individual was a professional translator and quality, especially among the lecturers and 3rd/4th year students. This was then followed by
the linguists, who thought there was a correlation 'in most cases' slightly less often (60.4% for students and lecturers vs. 53.5% for linguists), but were nearly tied with students and lecturers when 'in most cases' and 'in some cases' were combined. 42.9% of the buyers thought there was a correlation 'in most cases', but this group was also tied with linguists and the lecturer/student groups when 'in most cases' and 'in some cases' were combined.

The only group that stood out were the non-profit clients, where only 11.8% thought that a translator being a professional was a deciding factor of translation quality in most cases, and only 64.7% thought so when 'in most cases' and 'in some cases' were combined, compared to the 80 to 85% for the other three groups. As with the factors impacting quality discussed in the previous section, this also supports the concept of translation as a failed professionalised occupation (Chesterman 2001, p.146; Sela-Sheffy 2015, secs3476-3480) and low awareness among non-specialist clients.

These findings also suggest that lecturers preparing students for a professional career as well as students who have decided to become a 'professional translator' would have strong incentives to believe that there is a correlation between quality and professional status. Similarly, linguists who themselves identify as professional translators would also be more inclined to see a connection, while non-profit clients who appear to be happy to work with volunteers who may or may not be 'professionals' seem to have gained their perspective from that experience.

The observation that overall professional buyers seem to think that there is a strong correlation between professional status and quality less often than linguists is surprising. While it is difficult to estimate why this might be the case, one explanation could be the wider range of experiences gathered through a large pool of individuals the buyers would have worked with over time.

Another potential explanation could be the differences in what, according to the respondent groups, defines a language professional. Here, only the professional buyers and the linguists rank 'getting paid' relatively high as a defining factor (4th position for each, with approx. 22% mentions). If payment is considered a defining factor of professional status (Chesterman 2001, p.146; Sela-Sheffy 2015, secs3476-3480), then it would seem likely that no correlation between professional status and expected quality is perceived.
While payment also ranks in 4th position for the linguist group, here the relatively loose connection between payment and quality might have been outweighed by the respondent's self-perception in connection to their own professional status and could be motivated by a perceived need to counter a potential increase in de-professionalisation of the translation occupation (Pym et al. 2016, p.36).

This is also supported by differences in how dispassionate or involved respondents appear in their answers. While the other groups provide more factual indicators of an individual being a professional, such as their training, experience, cultural knowledge, and process knowledge, the linguist group additionally mentions themes like dedication, motivation and passion. This suggests that the tendency in this group is to see translation more as a vocation, rather than merely a profession.

Another observation that can be made is the depth of factors mentioned in the descriptions. Students and lecturers mainly remain on the surface, mentioning language knowledge, training, process knowledge, experience, culture knowledge and, (very minimally) getting paid. This is mirrored also by the non-profit clients.

In contrast, additionally to the aspects mentioned by students, lecturers and non-profit clients, linguists also mentioned aspects that would likely be part of their daily job, including business, marketing and soft skills as well as professional ethics, being part of the community and recognition from peers.

Similarly, professional buyers added client relationship, soft skills, knowledge of target markets, and an appreciation of the complexities of localisation.

These differences appear to reflect, as with previous sections, process-based quality bias in the differences in the individual group's experiences. While students, lecturers and non-profit clients would have limited hands-on knowledge of the language profession, both, linguists and professional buyers were able to contribute to the definition by including their observations from their hands-on experience.

**KF 4:** Process-based quality bias could be observed in research participants. Here, importance was given to those factors in of the process which the individual participant groups had added to the process.

**KF 5:** The ‘voluntary/crowdsourced localisation’ extrinsic cue introduced
uncertainty over where to place the translation on the expected quality continuum.

9.1.4. ‘Naive Theories’: Volunteer and Crowdsourced Localisation

9.1.4.1. Views of volunteer translation and crowdsourcing

As already discussed in section 9.1.3., voluntary translations appear to have a special status in the perception of research participants in that results of ranking volunteer translators against paid-for translators shows a greater degree of uncertainty than is the case for low, medium or highly paid translators. The following section will discuss in more detail theories expressed around volunteer and crowdsourced localisation, based on the findings of the questionnaire and interview based studies. As with other factors impacting quality, different perspectives can be observed when considering the results of the questionnaire sections aimed at non-profit clients, professional buyers and linguists that were designed to elicit opinions on volunteer and crowdsourced localisation.

All three groups responded positively overall when asked directly about their opinion towards volunteer and crowdsourced translation, with linguists being overall more negative and showing a stronger personal connection than buyers and clients (Sections 7.2.4.1, 7.2.4.2, 7.2.4.3; Figures 55 and 56).

Non-profit clients, at 94.1%, were more certain than the professional buyers with 78.6% of a connection between professional status and quality, indicating they thought that crowdsourcing or volunteer translations were overall positive. However, this difference is likely also influenced by the low number of respondents in the groups, since in fact all but one of the non-profit clients (out of 17) and all but two of the professional buyers (out of 13) selected that category.

Linguists were overall slightly more negative, with only 61.5% selecting that they thought crowdsourcing was overall positive. These results were also connected to the experiences of the respondent group, with 87.7% of those who identified primarily as volunteer translators considering volunteer translations or crowdsourcing positively overall, whereas only 52% of primarily paid-for translators did. Whether these differences are because those who generally have a positive view of crowdsourcing or volunteering would
also be more inclined to participate in such an initiative, or whether the participation in an initiative shapes the opinions cannot be answered within the scope of this research.

Partial explanations for these differences could be found in the comment sections for the question. Non-profit clients frequently mentioned the benefits to their organisation from volunteer translation, focusing on the value crowdsourcing provided to their causes and their positive experiences. Professional buyers mentioned the greater benefits gained from local communities and activism. This perspective is reflected in the literature that highlights the positive effects of volunteer and crowdsourced localisation initiatives, for example for humanitarian purposes or in connection to open source software initiatives (Anastasiou and Schäler 2010; Exton et al. 2010, p.47; O’Brien and Schäler 2010, pp.1–2). These perspectives are also reflected in motivations for volunteering expressed by volunteers, which, among career-oriented goals, include social causes (O’Brien and Schäler 2010, p.6).

However, while all three groups included both positive and critical voices, the linguist group offered a range of detailed criticism that included

- the ethics of outsourcing-focused crowdsourcing conducted for large, for-profit organisations,
- the threats to the translation profession as well as
- being against crowdsourcing in principle since work should be paid for in all cases.

Similar concerns were voiced by researchers who consider volunteer translation

- a threat to the livelihood of professionals (Austermühl 2011, p.15; Fernández Costales 2013, p.104), and those who describe crowdsourcing as
- exploitation of unpaid workers (Dodd 2011 - in Dolmaya 2011, p.98; Flanagan 2016, p.159).

This indicates that to the buyer, crowdsourcing would seem more like just one of several options that are available, allowing the non-profit clients and professional buyers to
focus on the benefits in specific or abstract ways, while to the linguist crowdsourcing is more strongly connected to their personal experience, as well as needing to protect their livelihood against increased de-professionalisation (Pym et al. 2016, p.36). However, it should also be noted that the majority of the respondents also in the linguist group thought that volunteer/crowdsourced translations were overall positive. While it is not possible to generalise to all translators based on the sample for this research, at least the respondents to this survey appear to be overall more positive about localisation crowdsourcing than the general tenor of the literature might suggest.

9.1.4.2. Theories about the average volunteer

Knowing that participants overall thought of crowdsourcing in a positive way however does not necessarily indicate how they would perceive quality if the extrinsic cue of whether the translator had been a volunteer or not was to be provided. To gain more insights into this, participants were asked to describe the average volunteer.

Non-profit clients generally didn't seem to have a good concept of who the volunteers are, professional buyers were focused on activism and the volunteer as domain expert, while linguists thought of volunteers more often as novice translators or amateurs.

While only two respondents among the non-profit clients described the volunteer rather than the benefits they gained from volunteering, the descriptions here as well as from the interview participants were mixed on whether individuals were mostly professionals wanting to do good or mostly new translators wanting to improve. This low amount of detail, on who the volunteers might be gives the impression that the respondents hadn't considered this question before, at least not in much detail and that it might not be very important to them. This is also supported by the finding mentioned previously in section 9.1.4., that suggested that a bilingual amateur with domain knowledge and motivation might fit the profile for an individual who can provide quality translations from the non-profit client perspective and that the distinction between 'professional' and 'non-professional' as well as what either side might have to offer isn't as important as getting the content translated; demonstrated by the majority of comments focusing on the benefits rather than the translator. This is again in line with observations made previously that
individuals think of quality and who can provide it primarily from their own perspective. It also supports the suggestion made by Pérez-González and Susam-Saraeva (2012) that in the public perception it doesn’t appear to be clear whether translators have an exclusive body of knowledge (p.150) and that in the practice of organisations like those of the non-profit clients on Trommons, working with professionals rather than non-professionals might be the exception in the wider context (p.157).

What is more surprising however are the responses given by the professional buyers. Here there was a strong undercurrent of the activism side of volunteering, and no clear trend emerged on the question of whether volunteers were primarily professional translators or bilingual amateurs. This is possibly due to the low overall number of respondents.

During the interview, this focus on the cause or product rather than the outsourcing role that crowdsourcing can play also became apparent in the distinction made by informants between domain and linguistic experts. Informants made comments on managing projects by combining the two as well as how professional translators could not have done as well on a fansubbing project as the fans did. This distinction on the part of the interview participants can partially be explained with their own experiences in fan-centric contents like video games. As outlined in section 4.1.2., crowdsourcing has had a longer history in adaptation of content by fans, like fansubbing, fandubbing and ROM hacking. Additionally, the importance of domain expertise or subject knowledge has been highlighted in the literature (Sargent and DePalma 2007, p.11; Muñoz Martín 2009, p.85), including distinctions made between domain experts and linguistic experts (Allman 2006, p.18; Neather 2012, p.266). However, by also making this distinction between domain and linguistic experts, professional buyers seem to have recognised a limitation in their services and that non-professional translators can offer benefits as domain experts that a professional translator who is not a domain expert would not be able to provide. The suggestion that translator training would enable a translator to handle a range of domains and texts has however also been made (Öner 2013, p.77).

These observations are also supported by the ranking of the main assumed motivations for volunteering that questionnaire participants were asked to undertake. Here,
the top three positions were taken by cause, product and altruistic focused motivations while the first motivation that would directly benefit the translator, gaining experience with translation, came in fourth position.

To an extent, the division between the domain expert and the linguist as well as associating crowdsourcing/volunteering with products and causes rather than outsourcing and the translator's benefits, suggests that the participants view crowdsourcing as separate from their own services; two approaches that can live in parallel and that have different purposes, rather than being in competition. This is also reflected by the surprise interview participants expressed when asked about projects that were more outsourcing driven, like Facebook, or projects where the volunteer was not a domain expert. In these cases, informants expressed confusion as to why this would be considered beneficial by the requesting party or the volunteers.

The linguist group, however, focused strongly on the professional status of the volunteer, with the majority thinking that volunteers were primarily novice translators or bilingual amateurs. This was also reflected in the ranking question, where the main motivation in top position was getting experience. Doing good and caring for a cause only ranked in second and third position.

Interview participants also tended towards the idea of volunteer translators being primarily amateurs, but were mixed on the question. They also stated that there shouldn't be a big difference for quality, although some aspects such as style and turnaround time might differ. Another recurrent theme in the interviews was that volunteers would try hard, work to the best of their ability, or give the best they could.

This paints a picture of linguists seeing volunteer translators as lower skilled and potentially amateurs but motivated, linking back to the tendency to view translation as a vocation more than a profession (Section 9.1.4.). This reflects assumptions made in the literature that participants in voluntary or crowdsourced localisation are primarily non-professionals (Cintas and Sanchez 2006, p.38; O’Hagan 2009, p.96; Austermühl 2011, p.2; Dolmaya 2011, p.103; O’Brien 2011, p.19; Tatsumi et al. 2012, p.2; Fernández Costales 2013, p.91; Flanagan 2016, p.151). Interestingly, even the linguists who were interviewed, all volunteers themselves, thought volunteers were primarily non-professionals, contrary to
the findings reported by O’Brien and Schäler (2010), where 222 volunteers for TRF declared they were professional translators and only 35 said they were amateur translators (p.4). This suggests that even those who volunteer for the organisation might benefit from more information on the composition of the community.

9.1.4.3. Uncertainty around volunteering and crowdsourcing

This need for increased communication around the professional background of volunteers on Trommons could also be observed in the theme of uncertainty concerning the background of volunteers, which emerged from the interviews conducted as part of this research. As outlined, volunteers on Trommons can sign up without being vetted and only need to provide minimal information about themselves, without having to provide credentials. In combination with the interaction with a platform rather than individuals, this results in a relatively high level of anonymity, which was also expressed by one of the non-profit clients interviewed. Similarly, one of the translators described an expected quality difference between organisations with a more stringent vetting process as against one that with less restrictions on who could volunteer. On the surface, this places Trommons in the category of initiatives that work, in the spirit of the original definition of crowdsourcing with an “undefined, generally large group of people” (Howe 2006), with similarities to initiatives like Amazon Mechanical Turk and Cucumis (Perrino 2009, pp.69–70; Zaidan and Callison-Burch 2011, p.1221).

On the surface, this would suggest great uncertainty about the quality of translations provided by Trommons volunteers, due to high levels of openness and anonymity. However, as became clear during interviews with TRF staff, even on Trommons, staff involvement in managing quality is quite high, with direct intervention through quality checks as well as through selection of translators 'behind the scenes'. However, while this lowers uncertainty around quality for staff at TRF, without communicating this back to the non-profit clients these interventions would not be likely to lower uncertainty for those clients. Additionally, this places the burden of selection on the TRF staff in a similar way to the selection processes in for-profit localisation, while at the same time not fully harnessing the power of ‘the crowd’. Possible alternative models could include trusting new members
with training and ensuring quality (Pym et al. 2016, p.44), or community-based review and validation models that would involve greater collaboration among the members of each language community present on Trommons, fostered by the introduction of formal communities of practice (Kimble et al. 2001; Wenger 2009), which could not only aid in lowering uncertainty around the composition of the community and quality to be expected, but also aid members of the community in reaching their career and social interaction based motivations for volunteering.

Similarly, a recurring theme during the interviews with TRF staff was uncertainty about whether non-profit clients were satisfied with the translations provided. TRF staff also expressed that they were rarely getting feedback on the translations and weren't certain whether clients evaluate the quality of translations on their side. This was also reflected by the descriptions provided by one of the non-profit clients interviewed who described the translations generally as good, but that they would commonly resolve issues that they did encounter on their side without re-contacting TRF.

Among the translators, the general description was also that the translations on Trommons were generally good, but that there had been some rather bad ones. There were issues identified around consistency across the projects, suggesting the need for higher involvement of the organisations in the localisation process, by providing more background information and supporting documentation, as well as their involvement in the proposed communities of practice.

On the whole however, without conducting further empirical research into the quality provided by Trommons volunteers, which would also require defining quality expectations and a measurement scheme, quality on Trommons appears to be 'mostly good, sometimes excellent, and sometimes awful' [TRF Staff 1]. Even taking into account the limited data collected for this research, there appears to be a great amount of uncertainty around the topic.

When considering the uncertainty around the background of the volunteers, as well as what quality to expect from them, together with the overall positive attitude towards the concept, the research shows that uncertainty does not seem to detract from the positive. This suggests that quality might not always be seen as the most important factor in
localisation and that potentially quality risks might be seen to be outweighed by the benefits of voluntary and crowdsourced initiatives.

KF 6: Despite uncertainty over where to place volunteer/crowdsourced localisation on the ‘expected quality continuum’, such initiatives were generally viewed in a positive light by research participants.

9.1.5. Experiment Results

One of the goals of this research was to understand the impact extrinsic cues would have for different respondent groups. For this, a series of experiments was conducted where two conditions were compared:

- participant selections for scenarios with no information on the translator's background or price of translation with
- the selections where information was provided.

9.1.5.1. Experiment 1

Experiment 1 (Section 8.1) included groups with high levels of linguistic knowledge, since the experiment was conducted in the English - Spanish language combination and respondents invited were either native speakers or highly proficient in Spanish. However, as outlined in section 7.1.1., demographics indicate differing levels of measurement knowledge based on training and experience. This experiment was conducted to observe the impact that the knowledge of a translator's professional status might have on perceived quality in these groups.

Deval et al. (2012) demonstrated that individuals with higher levels of expertise could be less strongly influenced by 'naive theories'. Similarly, groups with higher measurement knowledge would be likely to be less reliant on extrinsic cues than those with lower measurement knowledge (Golder et al. 2012, p.5). This was also observed in the results of this research.

When no information on the translator's background was provided, differences in measurement knowledge could be observed, although only to the extent to which different
groups selected the most experienced translator (Figure 80). While lecturers, industry linguists and 3\textsuperscript{rd}/4\textsuperscript{th} year students selected the most experienced translator's version about half of the time, the group average for the 1\textsuperscript{st}/2\textsuperscript{nd} year students was 27.3\%, suggesting that at a minimum, the novice student group deviates from the consensus of the more trained/experienced translators. The fact that the expert translator was selected more frequently than the other translators might also indicate that there were differences in the intrinsic attributes that would generally be associated with higher quality. However, this would require further enquiry to be confirmed.

1\textsuperscript{st} and 2\textsuperscript{nd} year students were also more strongly influenced by the information on translator status provided for both scenarios where correct information was given, and those with incorrect information, which confirms the prediction that low measurement knowledge would be correlated with extrinsic cues.

Interestingly, both linguists and 3\textsuperscript{rd}/4\textsuperscript{th} year students were less strongly influenced by information on professional status, which would be congruent with the prediction based on the literature (Section 5.1.) that these groups would have higher measurement knowledge in this case obtained through training and/or experience. While 3\textsuperscript{rd}/4\textsuperscript{th} year students performed slightly worse than linguists when no information was provided, the outcomes were nearly the same when information on professional status was added. This is in spite of the linguist group primarily consisting of experienced and trained translators, while the student group had low levels of experience. This and the higher performance of the trained students, who also paid less attention to extrinsic cues when incorrect information was provided than novice students, suggests that at least in the context of this experiment, training played a greater role than experience in the extent to which participants relied on extrinsic cues. This is not consistent with the findings of the survey, where different participant groups estimated experience to be of greater importance for quality output than training. In section 9.1.1.2.2., the observation that trained translators (3\textsuperscript{rd}/4\textsuperscript{th} year students) performed similar to trained and experienced lecturers was also made in relation to the ability to make tacit knowledge of quality explicit (Eraut 2000, p.118).

Another observation that can be made is that extrinsic cues have a stronger impact on perceived quality when the 'naive theories' behind them match up with the intrinsic
attributes that the individual can observe than when they don't match. This can be most strongly observed with the lecturer group, whose quality evaluation appeared to be strengthened through extrinsic cues, determined by how often they picked the expert translator's version relative to how often they did so when no information on the translator was provided. However, when incorrect information is provided, then the intrinsic attributes appear to weigh stronger than extrinsic cues. This is also supported by the observation that those participants who thought that being a professional had an impact on quality selected the expert translator more frequently in scenarios where correct information was provided than those who didn't think so, while no differences could be observed for scenarios where no information or incorrect information was provided. These findings are congruent with literature that suggests that reliance on extrinsic cues is lower with higher measurement knowledge (Zeithaml 1988, p.9; Deval et al. 2012, p.1185). However, the observation in this research that extrinsic cues appear to support perceived quality when they match intrinsic attributes but are largely ignored when they don’t match adds a new dimension to previous findings, which supports the suggestion that intrinsic cues have a stronger impact than extrinsic cues (Szybillo and Jacoby 1974, p.77), at least in groups with sufficiently high measurement knowledge of intrinsic cues. While the results for this research are based on a small sample, future research could conduct further enquiries into the circumstances under which expert evaluators rely on extrinsic cues.

In summary, the prediction that high measurement knowledge would be correlated with lower reliance on extrinsic cues (Zeithaml 1988, p.9; Deval et al. 2012, p.1185) was confirmed. In addition, the observation was made that the respondents' training appeared to have a higher impact on the extent to which extrinsic cues influenced perceived quality than the respondents' experience. Finally, for all groups extrinsic cues that were congruent with observed internal attributes reinforced the perception of those attributes, while extrinsic cues that were contrary to intrinsic attributes appear to have been ignored more often among those groups with higher measurement knowledge.

This was also confirmed when the number of intrinsic attributes available to participants was considered. Here, higher numbers of words for each scenario, and by extension greater number of intrinsic attributes available for evaluation, had a higher
impact on performance for those groups with greater measurement knowledge than the 1st/2nd year students.

**KF 7:** Participants with high measurement knowledge were less strongly influenced by extrinsic cues, than those with lower measurement knowledge. This was the case especially for those extrinsic cues that were easy to place on the ‘expected quality continuum’.

### 9.1.5.2. Experiment 2

Experiment 2 also included a group with high levels of linguistic knowledge. Here, the experiment was conducted in the English - German language combination, with participants being native speakers of German or highly proficient. In addition, this group as a whole is likely to have high measurement knowledge since it consisted primarily of trained and experienced linguists as outlined in section 7.1.2. This experiment was conducted to observe the impact the knowledge of the price of the translation might have on perceived quality in this group. While preference for the volunteer translator’s version was high when no information was provided, uncertainty of how to interpret the extrinsic cue ‘voluntary/crowdsourced’, as previously discussed, appears to have led to split preferences in this group (Section 8.2.).

Interestingly, when no information was provided, participants selected the volunteer translator's version as the best most frequently (Figure 86). Medium cost and high cost were tied, and the low-cost translation was selected least frequently. This suggests that to some extent, there were differences between the translations' intrinsic attributes that would lead to relatively clear preferences, rather than an equal split, and that the intrinsic attributes of the low-cost translation were least associated with what the participants considered high quality translation.

However, when correct information was provided (Figure 87), the high cost translation gained in preference over the medium cost translation, and was placed near equal to the volunteer translation, which would suggest that participants' choices were impacted to some extent by the knowledge of how much the translation had cost. However, although it did not lead them to rate the volunteer translation much lower, indicating a split
between decisions informed by extrinsic cues (price) and intrinsic attributes of the translation or alternatively, a positive outlook on volunteer translations informing the selections. This matches the observation made in previous sections that different ‘naive theories’ on how to interpret the extrinsic cue ‘volunteer translator’ are active and, as shown by Deval et al. (2012), they impact perceptions. As previous sections have discussed, linguists more commonly thought of volunteer translators as amateurs than other groups and were split over what quality to expect from volunteers. These findings are reflected in the results of this experiment, despite the group as a whole having high levels of measurement knowledge.

This can also be observed when incorrect information was provided (Figures 88 and 89). Here, the translation labelled as medium cost was selected most frequently as the best translation, followed by the translation labelled as high cost. However, when considering who had actually provided the translations, the volunteer translation was selected on average more often than any other option, followed by the high cost translation.

This would suggest that in cases where participants perceived a mismatch between extrinsic cues and intrinsic attributes, they, too ignored extrinsic cues at least some of the time and continued to favour the translation they had favoured already when no information or correct information was provided. It is also interesting to note that when correct information was provided, this did not reinforce the preference of the volunteer translation (as it did for the selections of experiment 1 in the lecturer group), but instead the participants preferred the volunteer translation slightly less often than when no information was provided.

Whether the information that a translation was conducted by a high cost translator impacted perceived quality is difficult to confirm from these results. On the one hand, the percentage of those who preferred the high cost translation was consistently higher when they had been told that it was the high cost translation, relative to the scenarios where no information was provided. This would suggest that at least parts of the participant group were influenced by the information. On the other hand, it is also possible that the original results for the NoInfo category were an anomaly and the differences primarily occurred due to actual differences in intrinsic attributes.
However, when considering the levels of agreement based on standard deviation, then a clear difference between this experiment and experiment 1 emerges. In experiment 1 agreement was always highest when correct information was provided and lowest when incorrect information was provided (Section 8.1.3.3.2.). This suggests that correct information reinforced quality as it was perceived based on intrinsic attributes while creating uncertainty when extrinsic cues and intrinsic attributes didn't match, in experiment 2 agreement was nearly the same when no information and correct information was provided, while highest when incorrect information was provided (Section 8.2.2.).

This also suggests that incorrect information lowered uncertainty within the group as a whole, leading the researcher to conclude that selections were likely more strongly based on intrinsic attributes than on extrinsic cues.

This then indicates that in scenarios where intrinsic information is available to the person evaluating quality, professional status as an extrinsic cue has a higher impact on perceived quality than price as an extrinsic cue. However, this is the case to a lesser degree for groups with high measurement knowledge relative to those with low measurement knowledge, than price.

This is congruent with the findings of the survey where participants overall gave more importance to experience and training than to price. This can be explained when considering that training and experience are likely more suited to actually having an impact on quality than the relatively arbitrary distinction between 'paid' and 'unpaid', as was discussed in more detail in chapter 4.

KF 8: If the extrinsic cue is harder to place on the ‘expected quality continuum’, interpretation uncertainty led to split results where some respondents were more heavily influenced than others.

9.1.5.3. Experiment 3

Experiment 3 excluded linguistic knowledge and essentially deprived participants of intrinsic attributes by presenting them not with translations, but reviewer comments based on an evaluation of different translations. This experiment was conducted with professional buyers as well as non-profit clients, and based on demographic information, and therefore
included one group with high and one with low measurement knowledge.

As outlined in section 8.3., the results of this experiment should not be over stated since the fact that participants based their selections in part on a proxy (reviewer comments) rather than the actual product attributes (as was the case for experiments 1 and 2), added variability that was difficult to control. On the one hand, this approach was limited by the fact that buyers did not have the actual translations available and were therefore not able to perform checks that they might have in a real-world setting. On the other hand, the approach does have a high level of ecological validity in a scenario where a buyer might have low measurement motivation (due for example to time constraints) (Golder et al. 2012, p.9). This is also supported by the comments from professional buyers for individual scenarios in the experiment, that indicate that they made their choices based on what the reviewer had said by commenting on which of the factors mentioned by the reviewers was important in their opinion. Only one participant in the professional buyer group commented that they felt that the experiment was poorly designed and that they needed more context.

In contrast, non-profit buyers focused their comments around the payment information, either by stating that they didn't think the information had an influence on them, or by commenting that it did.

Based on the selections made by participants, the outcomes are relatively similar between the two groups, for some of the price points when no information was provided (Figures 97 and 99). Here, with only reviewer comments to base their decision on, both groups were likely to pick the 5 cent translator, and equally likely/neutral/unlikely to pick the 10 cent translator. However, while non-profit buyers were mostly unlikely to work with the volunteer and equally likely/neutral/unlikely for the 15 cent translator, professional buyers were mostly neutral towards the volunteer and mostly likely to work with the 15 cent translator.

Interestingly, these selections changed when information (whether correct or incorrect) was provided (Figures 98 and 100). In this case, both the professional buyers and non-profit clients changed to being

- likely to work with the unpaid volunteer and
• unlikely to work with the 15 cent translator, while being
• relatively indifferent towards the 5 and 10 cent translator with only a slight increase towards 'likely' for the 10 cent translator, while in both groups the selections of 'likely' dropped by nearly half for the 5 cent translator in favour of 'neutral'.

The fact that both groups were equally influenced by the information on payment suggests that reviewer comments alone are insufficient information for a buyer to base their evaluation on, independent of their measurement knowledge. Both groups favour unpaid volunteers over highly paid translators, which is less surprising for the non-profit clients, who could be expected to view volunteers favourably, than it is for the professional buyers. If the assumption had been that 'you get what you pay for' then professional buyers should have been more strongly in favour of the 15 cent translator, although it is also possible that they had higher expectations towards the highly-paid translator and were therefore less likely to work with them.

Based on these findings, buyers with low measurement motivation (or in this case being put in a situation that mimics low measurement motivation by providing only reviewer comments) appear to be equally influenced by price of translation as an extrinsic cue. In addition, the two respondent groups for this research were equally positive towards volunteer translation, indifferent towards mid-range cost translation and negative towards high-cost translation.

**KF 9:** Measurement knowledge appeared only to be an important factor in how strongly participants were influenced by extrinsic cues when sufficiently high amounts and quality of intrinsic attributes were available.
Chapter 10: Conclusion

This research set out to investigate perceived quality in crowdsourced localisation by asking the following three research questions:

1. How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?

2. What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?

3. What is the impact of 'naive theories' on perceived quality in localisation?

After providing background to the research by outlining an integrative quality framework and the concept of perceived quality, the concept of quality in localisation, a widely-used localisation process as well as applying the integrative quality framework to the localisation field, this research then aimed to answer the research questions through a mixed methods approach that included the following data collection and analysis techniques:

1. A survey consisting of multiple studies with semi-structured interviews, questionnaires and content analysis was conducted to address the ways quality in localisation is defined in different groups as well as to what extent these definitions and associated vocabulary are shared (Research question 1).

The survey part of this research was also aimed at research question 2, by gathering insights into prevalent attitudes, naive theories and opinions on crowdsourcing and quality. Naive theories were defined as inferences that inform product evaluation and are external of the actual product, but inform how quality is perceived in cases of low measurement knowledge (Deval et al. 2012). For this research, the focus was on price and professional status as extrinsic cues but also peripherally addressed other 'naive theories' about factors
impacting translation that might be used as extrinsic cues.

2 - A set of three experiments to investigate the impact of 'naive theories' relative to measurement knowledge on perceived quality in crowdsourced localisation (Research question 3).

This research involved several respondent groups, including translators (for-profit and non-profit), students and lecturers of translation as well as buyers/decision makers in localisation (for-profit and non-profit) and staff at TRF. By including a diverse set of respondent groups, the aim was to be able to observe the impact of measurement knowledge on the extent to which extrinsic cues would affect perceived quality.

The results of these studies were then analysed and presented in chapters 7 and 8, along with being summarised and discussed in previous sections of this chapter. Based on these findings, the following answers to the research questions are proposed:

**In answer to research question 1 (How is quality in localisation defined and to what extent do different groups agree on which aspects of quality are important?),** the findings of the studies conducted for this research suggest that overall, the most important aspect of quality across the groups is the accurate or faithful representation of meaning, followed by style, naturalness, client needs, culture, grammar and terminology (Figure 30; Table 41). Here, non-profit clients distinguished between different types of faithful representation of meaning, but expressed them in ways that would require further communication to be converted into requirements that could be satisfied by a linguist. The observation was made in several instances that different aspects of quality are more easily described and measured.

- One, the distinction between technical and linguistic quality or aspects of quality that individuals would find easier or harder to measure or agree on with universal or preference attributes being associated with higher agreement.
- Two, based on the observation that while style played a large role in how
quality was defined, it was also the quality attribute that was least consistently described and commonly referred to as difficult to agree on, placing it in the category of idiosyncratic attributes.

Additionally, the observation was made that there appears to be disagreement on who 'owns' quality in localisation. While linguists, students and lecturers were primarily focused on quality by considering the content, for-profit buyers gave more weight to client requirements when defining quality.

In answer to research question 2 (What are prevalent attitudes, 'naive theories' and opinions on crowdsourcing and quality within the localisation field?), the first area of consideration was the relationship between price, training and experience as perceived by research participants.

Here, when asked directly, price was considered least important for all groups, although it was given more importance by linguists and slightly more importance by professional buyers than non-profit clients. In addition, experience was overall ranked higher than training.

Except for non-profit clients, all groups perceived a strong correlation between the professional status of a translator and the quality that could be expected. However, different groups also had different concepts of what a professional linguist is. Distinguishing factors included ‘payment’ for professional buyers and linguists but not for non-profit clients. In addition, linguists at times seemed to think of translation more as a vocation than a profession and both, linguists and professional buyers included aspects related to their day-to-day experiences like soft skills, client relationship, target market knowledge, professional ethics etc.

Secondly, in relation to attitudes, theories and opinions on crowdsourcing/volunteer localisation the observation was made that all groups were overall more positive towards volunteer translation than available literature might suggest. However, linguists were somewhat more negative than professional buyers and non-profit clients, and those linguists who were primarily volunteers were more positive than those who were primarily paid-for translators.
Linguists also thought of volunteer translators primarily as bilingual amateurs and novice translators, while professional buyers seemed to be more split on the matter, with some assigning separate roles (domain expert vs. linguistic expert) to each group without perceiving them to be in competition. Non-profit clients however seemed to have a less clear concept of the identity of volunteer translators than linguists and professional buyers, while focusing on the benefits they gained from working with volunteers, which indicates that the focus for the clients here was more on getting translations than the background of the individuals providing them, or what to expect from them.

Linguists provided more detailed criticism of crowdsourced localisation including ethical considerations and the threat to the translation profession, while buyers were more focused on crowdsourcing as activism and the product/cause side, indicating that for buyers crowdsourcing tends to be just one option of many, while for linguists the topic appears to be more personal.

One recurring theme was that of uncertainty around who the volunteer translators are, especially on the non-profit client side. On the surface, Trommons provides high levels of anonymity to translators, with the associated uncertainty regarding quality. In reality, behind the scenes partial vetting and managing of the quality process takes place, although this is not necessarily known to the non-profit clients requesting translations on Trommons. In addition, there is uncertainty in whether non-profit clients are overall satisfied with the translations they receive overall, or whether they evaluate quality when they receive translations.

Uncertainty could also be observed as regards how volunteer translators were perceived relative to paid translators. While quality expectations were clearly high for high-cost translations and low for low-cost translations without additional factors, greater uncertainty could be observed on where to place volunteer translators. When combined with high experience, volunteer translators were perceived as more likely to deliver quality than low-paid translators, while in combination with low experience, these two groups were seen equally likely or unlikely.

For both research question 1 and 2, the observation was made repeatedly that concepts of quality, as well as attitudes, opinions and naive theories were commonly
influenced by the experiences the different groups would have likely had in their day-to-day activities in relation to localisation. This could also be observed in relation to factors other than payment, training and experience that might impact localisation quality. Here, non-profit clients mentioned aspects that could be covered by a bilingual amateur, suggesting little knowledge of the 'tools of the trade' and their importance to quality. Similarly, linguists focused on the role the linguist plays in producing quality, while for buyers a range of aspects all played a role from the source text input to linguist producing translations and quality improvement tools and techniques.

In answer to research question 3 (What is the impact of 'naive theories' on perceived quality in localisation?), a distinction was made between groups with high or low measurement knowledge, as well as different types of knowledge. Based on demographic information, different levels of measurement knowledge were observed in different respondent groups. This was loosely supported through observations of differences in the ability of different groups to make tacit knowledge explicit. In addition, the observation was made that strategies to address information asymmetry were more sophisticated and complex among industry buyers than non-profit buyers, which would be suited to partially increase measurement knowledge, but especially lower 'will' uncertainty among professional buyers.

Broadly, within the groups with linguistic knowledge, measurement knowledge appears to be lowest in 1st and 2nd year students, followed by 3rd and 4th year students while it is high in linguists and lecturers. In the groups without linguistic knowledge, measurement knowledge was observed to be lower for non-profit clients than for professional buyers.

In the experiment results then, differences in measurement knowledge could loosely be observed in the performance of the different groups, although with the limitation that here, performance was measured by the number of participants who selected the expert translator's version when no information was provided, without the researcher knowing the factors which informed that selection.

The experiment results also showed a correlation between greater reliance on
extrinsic cues and low measurement knowledge. Also, within the limitations of the research, the results suggest that training may have a stronger impact on measurement knowledge than experience since the inexperienced 3rd/4th year students performed nearly the same as the trained + experienced linguist group, suggesting similar reliance on extrinsic cues despite being less experienced.

For groups with higher measurement knowledge, extrinsic cues that confirm the perceived intrinsic attributes appeared to strengthen the perception of quality, while extrinsic cues appeared to get ignored more frequently if they didn't match with perceived intrinsic attributes.

While the findings on the role of price information were not conclusive, based on how strongly the participants agreed with each other, it appears that price less strongly influences perception of quality than training, and that participants ignored the extrinsic cue more frequently and considered intrinsic cues more strongly when the extrinsic cue was cost of translation rather than the translator's professional status.

Finally, buyers with low measurement motivation (or in this case being put in a situation that mimics low measurement motivation by being provided only reviewer comments) appear to be equally influenced by price of translation as an extrinsic cue independent of measurement knowledge. Here, the two respondent groups for this research were equally positive towards volunteer translation, indifferent towards mid-range cost translation and negative towards high-cost translation.

10.1. Original Contributions of the Research

While different definitions of quality and aspects influencing quality have been discussed in the literature, to the best of this researcher's knowledge, the present thesis presents the broadest investigation of definitions of quality and factors impacting quality from the practitioner's perspective in the context of crowdsourced localisation. It also provides an extensive discussion of these perspectives and puts them in the context of existing theory on skill acquisition and types of knowledge. It thereby provides insights into the relationship between the individual's professional experiences and their concept of quality.
In addition, this thesis includes previously unreported details on the Trommons community, as well as quality control and management mechanisms used in relation to Trommons. It also provides deeper insights into theories and attitudes around crowdsourced localisation from the perspective of different groups in localisation.

Further, this research provides insights into differences in measurement knowledge that can be observed in different groups in localisation. While this concept has previously been discussed in the literature based on individual experiences, this thesis further supports observations already made on the topic. In addition, this research highlights strategies used to address information asymmetry by different groups in the localisation process, the impact of training on the ability to express quality concepts as well as the process-based quality bias observed in research participants.

Finally, this thesis is to the best of this researcher's knowledge the first research project that investigates the impact the extrinsic cues voluntary/crowdsourced, price and professional status have on perceived quality and demonstrates a correlation between measurement knowledge and reliance on extrinsic cues, as well as differences in the relative impact of cues depending on levels of interpretation uncertainty.

10.2. Limitations and Future Research

While this thesis provides a range of original contributions to research, a set of limitations also need to be considered.

The researcher acknowledges that the findings are limited not only by the sampling approach of the different studies, but also by the focus on the Trommons community. As a result, these findings need to be considered as exploratory for future research rather than definitive and generalisable to the wider localisation industry as well as other crowdsourcing initiatives. By taking a broad approach and including different groups and goals, individual findings can inform future research design, for example by further comparing the impact of price vs. professional status within the same groups.

In addition, while this research took a broad approach to selecting respondent groups, it is not completely symmetrical in that it favours the perspective of those that are involved in crowdsourced or volunteer localisation. As a result, client-side for-profit buyers
with lower measurement knowledge as well as linguists without involvement in crowdsourcing or volunteering are missing from this research, which has potential to bias the results. This could be addressed by repeating individual studies with those absent groups and comparing outcomes with those presented in this thesis.

As previously mentioned, in the setup of experiment 3, professional buyers were essentially prevented from utilising strategies they might have in a real-world setting to address information asymmetry. This limitation suggests that further research is necessary to investigate the impact of extrinsic cues on perceived quality with higher ecological validity.

In addition, future research could investigate whether performance and impact of extrinsic cues on perceived quality is altered when criteria for evaluation are provided to different groups, thereby decreasing variability due to individual preferences or the assumed goals of the translations.

Finally, this current research does not uncover whether individuals take note of intrinsic attributes and then ignore them in favour of extrinsic cues, or whether they skipped reading the translations at least in parts and only focused on some aspects or on the extrinsic cues. This uncertainty is supported by the fact that in all experiments, several participants ignored the instructions in the test question asking them to select a specific option, but instead made their own selection. This suggests that at least some participants were not reading all the texts in the experiments. Future investigation could therefore be aimed at uncovering the extent to which intrinsic attributes are ignored vs. being not perceived.

10.3. Implications of the Research

In relation to the key findings presented in this chapter, the following implications have been identified:

**Key finding 1**

- ‘Expert buyers’ had greater measurement knowledge and detailed strategies for lowering quality uncertainty available than ‘non-expert buyers’.
• Implication: Expert buyers offer not only the benefit of lower workload to non-expert buyers, but also add value to the process through those strategies.

**Key finding 2**

• Training was shown to have increased the ability to express quality concepts.
• Implication: Training adds value to the localisation process by potentially improving the quality of localisation review.

**Key finding 3**

• There was wide agreement on some aspects of quality. In addition, buyers and linguists recognised that there are different types of quality.
• Implication: The purpose of quality evaluation needs to be clarified, distinguishing between approaches to satisfy business goals and those to achieve quality improvements.

**Key finding 4**

• Process-based quality bias could be observed in research participants, where importance was given to factors impacting quality in those parts the individual added to the localisation process.
• Implication: The individual perspective of groups needs to be considered in discussions around quality.

**Key finding 5**

• While for other extrinsic cues greater agreement on how to interpret those cues in relation to expected quality could be observed, the extrinsic cue ‘voluntary/crowdsourced localisation’ introduced uncertainty over where to place the translation on the ‘expected quality continuum’.
• Implication: Perception of such volunteer/crowdsourcing initiatives could benefit from greater classification and communication of the benefits and
limitations of such initiatives.

**Key finding 6**
- Despite uncertainty over where to place volunteer/crowdsourced localisation on the expected quality continuum, such initiatives were generally viewed in a positive light.
- Implications: The benefits of such initiatives might be seen to outweigh potential quality issues; quality might not be the most important factor.

**Key findings 7 and 8**
- Participants with high measurement knowledge were less strongly influenced by extrinsic cues than those with lower measurement knowledge for extrinsic cues that were easy to place on the ‘expected quality continuum’.
- If the extrinsic cue was harder to place on the ‘expected quality continuum’, interpretation uncertainty led to split results where some respondents were more heavily influenced than others.
- Implication: By lowering interpretation uncertainty of extrinsic cues, it might be possible to improve perceived quality also in those that negatively interpreted the cue.

**Key finding 9**
- Measurement knowledge was only an important factor in how strongly participants were influenced by extrinsic cues when sufficiently high amounts and quality of intrinsic attributes were available.
- Implication: Professional buyers’ strategies to address information asymmetry appear to be an important factor in their ability to add value to the localisation process chain.
Bibliography


social development, and well-being.’, *The American psychologist*, 55, 68–78.


Williams, Malcolm (2004) Translation quality assessment: An argumentation-centred approach,
University of Ottawa Press, Ottawa.


Appendices

Appendix A
This appendix contains a list of publications and presentations where parts of this research have previously been made public.

Title: "Multi-language communities, technology and perceived quality"
Authors: Tabea De Wille, Dr. Chris Exton, Reinhard Schäler
Venue: 7th International Conference on Communities and Technologies in Limerick, Ireland
Date presented: June 30th, 2015
Publication: In Gabriela Avram, Fiorella de Cindio, Volkmar Pipek (Eds.), International Reports on Socio-Informatics (IRSI), Proceedings of the Work-In-Progress Track of the 7th International Conference on Communities and Technologies (Vol. 12, Iss. 1, pp. 25-32)

Title: "Before Knowledge Management: Quality expectations in Volunteer Translation"
Authors: Tabea De Wille, Dr. Chris Exton, Reinhard Schäler
Venue: 16th European Conference on Knowledge Management at the University of Udine, Italy
Date presented: September 4th, 2015
Publication: In Maurizzio Massaro, Andrea Garlatti (Eds.), Proceedings of the 16th European Conference on Knowledge Management ECKM 2015, pp. 920-928.

Title: "Not-For-Profit Crowdsourced-Based Translation Technology"
Authors: Tabea De Wille, Dr. Asanka Wasala, Dr. Chris Exton, Reinhard Schäler
Venue: Conference for the 60th Anniversary of META – 1955-2015. Translation's horizons: back to the future at the University of Montreal, Canada
Date presented: August 20th, 2015

Title: "Technology in a multi-language community: opportunities and challenges"
Authors: Tabea De Wille, Dr. Chris Exton, Reinhard Schäler
Venue: XIII EXATEC Europe Summit 2015 at Dublin City University, Dublin, Ireland
Date presented: September 26th, 2015

Title: "Quality: Perspectives and Strategies for Localisation"
Authors: Tabea De Wille, Dr. Chris Exton, Reinhard Schäler
Venue: “Translating Europe” Workshops organised by University of Gdansk, Center for Translation Studies and Directorate-General for Translation, EU
Date presented: December 8th, 2016
Appendix B

This appendix contains the text used as basis for the experiments. It was translated into Spanish and German for experiments 1 and 2, and translations reviewed for experiment 3.

Succeeding in the MSc

Exciting times! You have been accepted to the MSc Multilingual Computing and Localisation and are rearing to go. Maybe you are fresh out of your undergraduate degree or maybe it’s been a few years since you have been involved in formalised learning. Whichever it is, here are some ideas that could help you not only cope with the workload, but also get the most out of the programme.

Know what you want to get out of it

Speaking of getting the most out of it, what is it you want to achieve by undertaking this programme? Maybe your focus is on deepening your knowledge in a specific area, getting credentials in order to get promoted into a specific role, switching careers, making contacts in the industry or getting into academia. All those and many others are good reasons, and knowing what your goals are helps with focusing on what is important.

Establish priorities in what to study and read

This is a demanding programme and you will receive a lot of material and opportunities to learn. It can be tempting to try and do everything perfectly and thereby get behind on the material. It is generally a good idea to look at everything available to you and consider your goals, as well as what you need to do in order to receive your degree. Check the programme outline, module outlines and when in doubt ask your lecturers or course director. Use this to decide which content you should go through right now and which can wait until a future date.

Do a little every day

To stay on top of things and also build good habits for writing your dissertation towards the end, it is ideal to do at least a little work for the programme every day. Maybe you can read
a page or two during your commute, mentally outline an essay in the shower or have a lecture running while doing the washing up? This won’t replace focused longer periods of work, but it helps with remaining immersed in the topics and also slowly chips away at the overall pile of tasks.

**Make use of having access to other students**

One of the great aspects of this programme is the wide diversity of students participating on all levels: different cultures; experience levels; specialities etc. Try and make use of that by socialising, but also by asking questions and opinions, providing your own thoughts, participating in discussions and being helpful.

**Make use of resources available to you like the library and magazine subscriptions**

While you will receive material to work through, don’t forget all the other resources available to you. The UL library to which you have access via your student ID has a good number of e-books you can borrow remotely, there are industry magazines and academic journals, websites, industry and standardization bodies with publications, blogs and so on. This will be especially important when it comes to preparing presentations, essays, reports and your dissertation.

**Be in charge of your learning experience**

During the programme you will likely encounter areas you find more difficult to understand or areas you want to know more about. Cultivating a curious mindset and trying to explore topics on your own is a step towards making the most of what you learn. The internet is full of possibilities, as are libraries and other practitioners in the industry. If you do come across something interesting, don’t forget to share!

**Show up**

In an online programme like this it is easy to disappear and fall behind. To avoid this, make use of the range of options to stay in touch with lecturers and other students, to engage with the material and to get help as well as share your own findings and ideas. From emails, forums, webinars, individual sessions or face to face meetings – there are many
possibilities. This doesn’t have to be a big deal, often showing up to the webinars and checking for new content will keep you ticking along nicely. Should you fall behind on content or miss deadlines, it is important that you reach out to the lecturer(s) as quickly as possible, or if you run into bigger issues, to the course director.

Share your strengths and weaknesses

As mentioned before, this is a diverse programme and sometimes it can seem a bit daunting if others seem to have a better grasp of a specific area. However, while you might be aware of your weaknesses, don’t forget your strengths. Inevitably, you will have a perspective on the topics that is unique to you and worthwhile sharing with other students and lecturers, and even just by asking questions you will open up discussions, make others aware of a topic they hadn’t even considered or show the lecturers which areas need to be covered again to clarify misunderstandings.

Think ahead to the exams but don’t panic

In some modules you will need to take a written exam at UL. While there is no need to panic, it is good to keep that exam in mind while working through the material during the semester. Will you be able to study based on the notes you are generating? Are you able to share and exchange notes with other students? Even if you only make sure you have all the resources in place when it’s time to start revising, that is a great advantage.

Think ahead to the dissertation but don’t panic

Similarly to the exams, the dissertation will need to be completed in order to get the MSc degree. While there is no need to worry too much about the dissertation early in the programme, it is a good idea to have a look out for potential topics as you go along. Modules or topics you find particularly interesting or a question you had that hasn’t been answered yet are good directions to look into further. You might also want to give some thought to the questions of how you will organise yourself, where to find material, and when and how you will be able to do your research and writing.

Try to apply what you learn
Remember that what you are learning is relevant to your day-to-day work and your past experiences. By connecting what you learn with your own professional life, you might notice that you understand things a little better, or do things a little differently. This sense of achievement in itself can be immensely satisfying.

**Enjoy being part of this**

Maybe your main motivation for studying in the MSc programme isn’t having fun, but that doesn’t mean it can’t be enjoyable. In a way, you are embarking on an intensive immersion in the world of localisation, together with other people with similar interests. While it is important to have goals and be focused, don’t forget to smell the roses on the way.
Appendix C

This appendix contains the participant information sheet and consent form used for the interview part of this research. The same information, although in a web based format, was provided to questionnaire and experiment participants.

INFORMATION SHEET

Dear Participant,

My name is Tabea De Wille and I am currently undertaking PhD research at the University of Limerick under the supervision of Dr. Chris Exton and Reinhard Schäler. The title of my proposed research is Quality Management in Volunteer Translation. The purpose of this project is to gain insights into the way participants involved in volunteer, non-market translations view quality matters and how quality is defined and managed within volunteer translation/localisation projects.

Participants will join either an online group discussion of approximately 60 minutes, or an individual interview of approximately 20 minutes. The discussion will be recorded for the purpose of transcription. No names or details that could lead to identification of individuals will be published.
Participants can withdraw from the interviews or group discussions at any point and request that their contributions are not used.

If you have further questions regarding this research please feel free to get in touch with either myself or my supervisor using the email addresses listed below.
If you have concerns about this study and wish to contact someone independent, you may contact: The Chair, Faculty of Science & Engineering Research Ethics Committee, University of Limerick, Limerick. Tel: 061 202802

Yours sincerely,

Tabea De Wille
Tabea.dewille@ul.ie
+353 (0)61 234797

Dr. Chris Exton
Chris.exton@ul.ie
+353 (0)61 202705
CONSENT FORM

Consent Section:
I, the undersigned, declare that I am willing to take part in research for the project entitled “Quality Management in Volunteer Translation”.

- I declare that I have been fully briefed on the nature of this study and my role in it and have been given the opportunity to ask questions before agreeing to participate.
- The nature of my participation has been explained to me and I have full knowledge of how the information collected will be used.
- I am also aware that my participation in this study may be recorded (video/audio) and I agree to this. However, should I feel uncomfortable at any time I can request that the recording equipment be switched off. I am entitled to copies of transcripts of my recordings on request, the recordings will be destroyed once the study is completed.
- I fully understand that there is no obligation on me to participate in this study
- I fully understand that I am free to withdraw my participation at any time without having to explain or give a reason
- I am also entitled to full confidentiality in terms of my participation and personal details

____________________________________         __________________________
Signature of participant                      Date
Appendix D
This appendix contains additional parts of the redacted transcripts of the interviews conducted as part of this research.

Defining and evaluating quality
Non-profit clients

"I mean like maybe the transmission is not perfect but it is ... I think the core of Rosetta is granting access to information and an exchange of expertise between organisations to expand the work that some are doing, that can be accessible and then exchange around the work. So maybe for very official or more commercial or market oriented projects the requirements are different. For us our main project it's actually a guide tool for activists so the important thing is that people get access to the core ideas and to the core tools and instruments we are trying to expand to other countries." [OrgClient 2]

"For what is the readability like when you're reading a document. I know sometimes it is translated, of course you want to do it word for word but sometimes it seems like it was a little - you know - the grammar is a little off like they're trying to be precise with the word choice but it doesn't read well" [OrgClient 2]

“For us our main project it's actually a guide tool for activists so the important thing is that people get access to the core ideas and to the core tools and instruments we are trying to expand to other countries.” [OrgClient 2]

“For what is the readability like when you're reading a document. I know sometimes it is translated, of course you want to do it word for word but sometimes it seems like it was a little - you know - the grammar is a little off like they're trying to be precise with the word choice but it doesn't read well” [OrgClient 2]

Professional buyers

"I mean there's the purely sort of linguistic quality per se so the style and how the person writes and stuff but then there's the more sort of technical or typically terminology and
things like that. So, we know not making mistakes in variables and all that stuff so there's those two sort of (...) basically there's a technical side to quality and there's a purely linguistic side. (...) So the technical stuff is basically the stuff that a non-native person can check. So, it's technically going to be - you know - with TMs now it's very easy for us to check on - you know - terminology and see if the same line has been translated - basically consistency so if the same terms have been translated same way throughout the file. It's gonna be about the - you know - double spaces, wrong capitalisation compared to the source text, weird punctuation, numbers not matching, people making mistakes in variables and things like that. That basically a non-native person can check. And we have, we have a series of tools that allow us to run through files and those would pop up quite quickly. And so, everything else is what we can't sort of judge in house so it's going to be about style and all that." [ProBuyer 1]

"There's a lot of - we test translators to start with. I mean we have a database with translators that have been tested on different subjects. And then we also have a network, we know from other translators that this translator has already for example translated this game or that game and he's a good gaming translator so we can have him on board. And then we give him a few thousand words and then we check the results. And you know this confirms what we've heard. And so, then we build a network of translators. We have game translators who are gamers and make a living from translating games and they just do this every day mostly. So, there's testing and then there's we check the work. Not everything of course, not every word that he's translated but we check as much as we can." [ProBuyer 2]

"The technical side is quite easy for us to check. For the purely linguistic side it's gonna be about how of course so that person did well or not on the test. It’s also going to be about feedback that we see so we release games and we always check for feedback in all the languages even those that we don't speak in house. We ask the vendors or partners to send us feedback from - you know - internet sites and the like." [ProBuyer 1]

"I mean German we can absorb a bit of translation in house but when - you know - during the peak season typically during summer and stuff, they won't be able to translate in house so but what we always do in house to the extent which we can is proofread. So basically, all
the linguistic coordinators in house do the proofreading, which sort of I also very much like the idea because it's ... I mean we have a control over quality and it's also the [company name] quality because it's our people that have the final word and that's what we're selling sort of." [ProBuyer 1]

Linguists

"Well, translation is very subjective - you know - so what I think is a good rendition, another translator might think it's not the best, but I think there's always room for disagreement on subtleties and things that are a matter of preference.

But I think on being accurate and reflecting the original meaning and following the rules of the target language, I think everybody would agree on that." [Translator 1]

"I suppose translation is ... there's a lot of it that's creative and if it wasn't subjective in a way then it wouldn't be and then machines would be able to do all the work. It takes a human mind to interpret and when you have to interpret there's always room for more than one interpretation." [Translator 1]

"If it's an agency the agency should have their own proof-readers and their proof-readers would kind of have the last word and but they should always encourage conversation with the translator too with the proof-reader. So, they can maybe come to an agreement. But if the client is a direct client then it's trickier because the client doesn't know anything about the language service market and doesn't have any linguistic skills to begin with. So that would be harder to resolve." [Translator 1]

"But with agencies I have and I've always come to ... I've had proof-readers proofread my work and say I don't like the way you used this word. I always had arguments to back up my choice but a professional translator would always know what is a matter of preference and what is not." [Translator 1]
Factors impacting quality

Professional buyers

"I think there is a link even though there should not be one. From experience and by discussing this with freelance translators, I know that a better paid project will be allocated more time, through research and additional revision steps." [ProBuyer 3]

"From what I have seen, a freelancer will treat a well-paid project more carefully so they do not lose a client that will keep on bringing good revenue." [ProBuyer 3]

"What I also noticed is that some of the best translators, with very good work ethics usually have reasonable rates. Not too high, not too low, they respect their work but also the client and they rarely abuse in terms of time spent on revision tasks, etc." [ProBuyer 3]

Linguists

"And when you're, if you're translating for two cent a word then neither the client can expect a very good quality translation and the translator probably doesn't try to provide amazing quality either because we're aware that the price is very low." [Translator 1]

"Well I think if you pay little money, usually people don't have the time if they're trying to make a living, they'll have to translate a lot of information so they won't put that much work (...) so yeah you could expect a difference." [Translator 4]

"And then they deal with clients that may not have very high expectations as well. (...) And may not know whether the translation is good or bad. They just want, they see translation as just another box to tick, this is done and they just want it done and they may not have a very good understanding of language and how it affects their image and that things should be written properly. And as long as it's Spanish they don't care if there are mistakes or not and I think that's the bottom line." [Translator 1]
Volunteer translation and crowdsourcing

Non-profit clients

“And I think it's something that they understand better than the rest of us; how important is to have grant access to people in other parts of the world.” [OrgClient 2]

“It was definitely more personal when Rosetta was emailing because you thought okay like there's somebody connecting you” [OrgClient 1]

“Whereas the platform is kind of like you just kind of upload the project, hope somebody picks it up” [OrgClient 1]

"honestly, I thought that the translation quality of the volunteers as long as they were fluent in both languages was quite good, was very good" [OrgClient 1]

"I mean it was rare that we actually had to go back to Trommons to get something re-translate, but you know obviously, it's unknown because you don't know who's picking up the project" [OrgClient 1]

"And sometimes they were helpful and able to do that. other times if we didn't want to go through that whole process or were in more of a rush, we would find other like student interpreters who are fluent in Spanish and English, have them read the article and fix those translations."[OrgClient 1]

"They wouldn't ... it's a project that is founded in trust. Trust that organisations are going to use it only because they don't have the resources and translators are in it because they care. So maybe it's not the most based in facts assumption but they are in it because they want to give their time." [OrgClient 2]

"It was definitely more personal when Rosetta was emailing because you thought okay like there's somebody connecting you" [OrgClient 1]

"Whereas the platform is kind of like you just kind of upload the project, hope somebody picks it up" [OrgClient 1]
Professional buyers

"Okay so they did I don't know, maybe ten years ago or something they did a little episode series called 'Stand Alone Complex' or something like that. And it it's pretty technical so it's Japanese, right? At the time, I was in Japan and it's pretty technical vocab you know it's all high-tech shit with you know cyborgs and what not. And the subtitles were done by fans and they were jewels. I mean they were English subtitles and it was ... I was very very much impressed. I mean I've lost my Japanese skills bit now but at the time I could understand maybe 40-50% and I would sort of - you know - get the meaning in my head but it was really well written and really well researched and stuff. And I saw at the end it was fansubbed. So, and the first thing I thought to myself was: I'm really not sure a professional - you know - company would be able to produce such high-quality stuff because in the end, I mean in the in the end it's a job also for people - you know - and you're not always in the best dispositions or you're not always passionate about the stuff you're working on. I think that does come into play at some point." [ProBuyer 1]

"Well I mean UI is UI. (...) UI doesn't really leave that much space for - you know - to be creative and all that so I guess it depends on the nature of the text you're translating. In the case of Facebook I can only imagine, I don't know why people would actually do that. I mean have difficulty understanding why Facebook wouldn't pay for a service like that or pay by some sort of in-game points thing that people that do the translations ... but maybe they do it for fame or for be able to say yeah, I translated the 'like button' or something like that. I don't know." [ProBuyer 1]

"You may ... sometimes you have a lot of typos, grammar errors ... but in general you can feel that the people who have translated do know the subject so they are more or less subject experts but they are not linguistic experts so they tend to leave some obvious typos or they forget to use the spell-checkers. They forget to spell-check their work. They forget to read or have it proofread. They don't ... when you're not really a professional translator some habits you need to have to perform a good translation. So, my idea is that crowdsourcing can be very interesting to get translations that do mean something but contain quite a few errors." [ProBuyer 2]
"Well then it really depends on the subject. I think if you're a volunteer translator, if you volunteer to translate something you should still consider the subject. I mean, why translate something you don't know anything about and volunteer to translate that and on top of that you're not even a ... you've never translated before. And you don't know it's - you know - to translate thousand words it can be quite time consuming so it really depends on if you just want to translate a few sentences here and there for let's say a good cause you can of course do that." [ProBuyer 2]

Linguists

"But I think the volunteers ... I think they would do to the best of their ability and I think they try hard. When I do translation ... to make sure you are using the right vocabulary and check on the internet, check it on websites and make sure that is the right term. I still think volunteer volunteers will do a good job. I think it will still be of a high quality." [Translator 3]

"I think there would be some people who would try really hard even if they do it for free. But then even people who are paid, it doesn't mean they are trying really hard at translating, so I think depends on the person." [Translator 3]

"I would expect that I would give the best I can. If I cannot translate I could maybe say I'm starting or the people at let's say the Rosetta foundation, when they receive a translated piece that's not very good, well of course give it to somebody else to proofread it to make sure that it's the best - you know - it can reach a client. I think that it doesn't matter that it's ... not that it doesn't matter but it being a volunteer does not mean that you don't give your best. (...) It could be very experienced translators that are volunteering so that it's not paid for doesn't mean it's not linked to quality for me" [Translator 2] (talking about a hypothetical volunteer)

"Organisation like The Rosetta Foundation or Translators Without Borders, because there these organisations are created by linguists or translators, so people in that field so they know what they're doing. But an NGO doesn't necessarily know the first thing about
translation or about languages or they may not be aware of the subtleties or how the quality might affect the end product.” [Translator 1]

"Overall I was surprised that it was quite poor. I think there was only one where I was impressed. Maybe a couple that were okay and maybe there were a few that quite bad, where you could see that the translator didn't understand the text." [Translator 4]

"And it's really a shame, it's really a shame because I don't know what motivates people to volunteer with a skill that they don't have, and I wish that the NGOs or the organisations would have a better way of screening their translations because there's loads of people willing to translate for free for a good cause and a lot of professionals and a lot of translators who need the experience, who are already translators, but they might be students who are just finished and they would do a fantastic job, but they just don't have the experience yet and need to get their foot in the market. So that's always something that I find concerning. That the quality is in general very poor.” [Translator 1]

"From what I saw I'd say most of them are amateurs but there were a few that were good so it's hard to tell. Some that could be professional, yeah” [Translator 4]

"I guess they want the experience and they want to learn French so they can start getting paid for it, I think maybe some people volunteer too soon, they should get maybe to a certain level before they take a type of task” [Translator 4]

**Trommons: Managing and producing quality**

"I guess what I was trying to point out is that in practice there is a lot more intervention in practice than was planned at the beginning.” [TRF Staff 1]

"So, this in itself is almost sort of, sometimes like a quality control mechanism in itself because she will contact people who she knows they're reliable and who she knows will - you know - do good work.” [TRF Staff 1]

"I mean they can choose the tasks that they want to work on and then we trust them and that they are going to select the task that they feel that they are good enough to work on. And
then in some cases, like for example there's an NGO that works with elderly people and they always ask French into Spanish and I know there is a Spanish volunteer that really likes that field and his father is working in the elderly sector. Her expertise is the elderly sector and she is always willing to work on these projects so we select for example, in this case we ask directly the girl to work on the projects and there's a lot of other cases like that, that we know that they have some special connection with the field so we ask directly these volunteers. But in general, it's up to them to choose any project they want to work on."

[TRF Staff 2]

"I would say that in that case that normally like NGOs wants to spread a message, so I would say that you can read the content, it's easily understandable, then of course some would be like the deadline involved as well in that cases because some of them made it quite quick and they are long documents. So yeah, I would say that they are aware that the quality, if they need - you know - in a short deadline, some of them might be aware that the quality might degrade." [TRF Staff 2]

"I would say it's mostly good. Sometimes excellent, and sometimes awful." [TRF Staff 1]

"So, I would say overall it's fine. You know the texts are you know, they make sense, there might be the occasional - you know - weakness in them, they may be not the best sort of ... I'd say in terms of correctness they're fine, mostly good enough. So, I'd say it's a rare occasion that we get a text that either linguistically or - you know - in some other way is not acceptable. But I would say that for most translations, you could probably improve them a little bit. So, in terms of fluency or - you know - just the way maybe the way they're written, they would sound better if - you know - you put a little bit more work into them but I would say the majority of them are fine." [TRF Staff 1]

"To be honest, I mean, sometimes there is some bad feedbacks, but in general there's not a lot of bad feedbacks from the proof-readers or from the organisation. Actually, all the bad feedbacks are from the volunteers who proofread some document because the organisations never complain" [TRF Staff 2]
"This is something that bothers me quite a lot is that I feel that because we are doing them a favour, they're slow to give us feedback. On what they get. Or they're just not bothered. I don't know, I'm not entirely sure what the reason is. But I would welcome a lot more feedback from them. I would welcome feedback on every single translation. I will welcome overall feedback once a year, I would very much welcome more complaints. We don't, we hardly ever get complaints. And I think the reason they don't complain is they feel they can't because they haven't paid for it. Whereas our idea was that - you know - that there would be sort of a feedback loop. Feedback from proof-readers, feedback from organisations and - you know - that the volunteers would then take on board, and - you know - not make the same mistake again the next time. And the proof-readers are - you know - more often the majority of proof-readers are doing their job. In that regard. The majority of organisations are not. They do not give feedback, I can't give you another answer other than they don't tell us. So, either this means that they're happy and that's why they don't tell us anything, because they're fine with what they get. And they do, I mean when we do ask them, most of the time we get: yes, everything's fine, everything's great, we're very happy. But my sense is that there must be problems, at least some of the time that they don't tell us about." [TRF Staff 1]

"I guess that before you register and you try it, it's a little bit weird that suddenly some volunteers that are gonna translate for you in some NGO that is based in Ireland. You may not be based here and it's a little bit, when you work face to face it's easier to feel more reliable and that, but then when it's all remotely, it's a little bit more (laughs) I don't know I'd say it's a little bit difficult to trust a different style you just come across with The Rosetta Foundation, but then just - you know - I'm every day in contact with the translators and I know a lot of them, you kind of get a personal contact and relation, and I know that most of them work as a professional translators. (...) But I can see why the organisation that doesn't know all of that can just be a little bit like feeling that it's little bit strange of the person. And it's remotely and I think that when it's remotely it's always like, feel a little more untrustful." [TRF Staff 2]
"Well actually we have some of them that you see, at first, they just create short projects that doesn't seems that urgent or not important for them and then when they see that it works fine, I mean it's quick and they get good results, they start creating more and more projects. Sometimes they had said to first, but really is this for free and how can I trust them." [TRF Staff 2]
Appendix E

This appendix contains questionnaire flows and questions used in the survey part of this research. Questionnaires 1 and 2 only include the parts relevant for the topic of this research.

Questionnaire 1

- Gender:
  - Female
  - Male
- Age:
  - 17 or younger
  - 18-24
  - 25-34
  - 35-49
  - 50-64
  - 65 or older
- Native Language:
- Secondary Language(s):
- Level of education:
  - No studies
  - Primary education
  - Secondary education
  - Third level
  - Short-cycle tertiary education
  - Bachelor or equivalent
  - Master or equivalent
  - Doctorate or equivalent.
- Have you studied translation or a related discipline?
  - No.
  - Yes.
- How did you hear about The Rosetta Foundation?
  - Friends/family
  - Internet search
  - School/college
  - Social media
  - Volunteer centre
  - Work/colleagues
  - Other (please specify)
- When did you register on Trommons?
○ Less than 3 months ago
○ Between 3 and 6 months ago
○ Between 6 months and a year ago
○ More than 1 year ago

● Are you following Trommons or The Rosetta Foundation on any of the following social media channels?
  ○ Facebook
  ○ Google +
  ○ LinkedIn
  ○ Twitter
  ○ You Tube
  ○ None of the above

● Are you donating to any non-profit project?
  ○ Yes
  ○ No

● What would encourage you to contribute to the funding of Trommons?
  ○ Support the cause
  ○ The credibility and transparency of the organisation
  ○ The effectiveness or impact of the project
  ○ Tax advantage
  ○ Public recognition

● Have you ever claimed a task on Trommons?
  ○ Yes
  ○ No

● Why have you not yet claimed a task on Trommons?
  ○ I realised it was not what I had expected
  ○ Lack of interest
  ○ Lack of time
  ○ Technical problems
  ○ There is no project available for my language skills
  ○ Other (please specify)

● Have you ever claimed but not completed a task on Trommons?
  ○ No, never
  ○ Yes, once
  ○ Yes, more than once
● If you have ever claimed and not completed a task on Trommons, what prevented you from completing the task?

● How frequently do you volunteer on Trommons?
  ○ Once a week
  ○ 2 to 3 times a month
  ○ Once a month
  ○ Less than once a month
  ○ Other (please specify)

● How likely are you to recommend The Rosetta Foundation to others as a place to volunteer?
  ○ Extremely likely
  ○ Very likely
  ○ Moderately likely
  ○ Slightly likely
  ○ Not at all likely

● How much do you agree with the following statements about volunteering?
  ○ I feel it is important to help others
  ○ I like to learn through hands-on experience
  ○ I feel better about myself when I help others
  ○ I want to gain experience to get a job
  ○ I like to be part of a group that shares an interest in community service
  ○ It is a way to focus on something other than my own life
  ○ The reason behind my volunteering varies according to the type of task
  ○ I always pick my tasks for the same reasons
  ○ It is important to me to receive feedback on my work
  ○ I regularly see tasks that fulfil my most important reason for volunteering
<table>
<thead>
<tr>
<th>How easy do you find Trommons to work with?</th>
<th>Very easy</th>
<th>Easy</th>
<th>Average</th>
<th>Difficult</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding the information I need on Trommons is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navigating my way around Trommons is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viewing tasks on Trommons is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claiming tasks on Trommons is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using the profile section is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filling in details on my profile page is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How would you like to be recognised and rewarded publicly for your work on Trommons?</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displaying rewards on my profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Displaying rewards from Trommons on other platforms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicating my activities on Trommons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalising my profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How much do you agree with the following statements? A quality service...</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>… must fulfil predefined requirements.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… must be free of any errors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… does not have to be perfect, but it must serve its purpose.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… exceeds the organisation’s expectations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… is available at the given deadline.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Which of the following criteria are the most important for you when it comes to translation quality? Please rank them (1 = most important, 4 = least important).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The translation sounds like a text that was originally written in the target language.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The translation precisely renders the meaning of the original text.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The translation fulfils real-world requirements (e.g. legally, culturally etc.).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The translation fulfils the design requirements (formatting, tags etc.).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Have you encountered any problems in the past in relation to quality on Trommons?
  - Yes
  - No

<table>
<thead>
<tr>
<th>Problems encountered on Trommons</th>
<th>Very often</th>
<th>Often</th>
<th>Sometimes</th>
<th>Rarely</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical problems (e.g. file format, embedded images).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic problems (e.g. terminology, grammar, style, register) with original (source language) file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic problems (e.g. terminology, grammar, style, register) with translated (target language) file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems with linguistic resources (e.g. glossary, style guide)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems contacting the non-profit organisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems contacting The Rosetta Foundation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- How did you solve the problem(s) you have encountered?
  - I fixed it myself.
  - I ignored it.
  - I contacted The Rosetta Foundation for help.
  - I contacted other translators/friends/volunteers for help.
  - I contacted the non-profit organisation for help.
  - I gave up and returned the task on Trommons.
○ I used the Trommons Forum to ask a question / discuss the issue with other volunteers and the organisation that posted the task.
○ Other (please specify)

● What could have helped you to resolve the issue more easily?

● Overall, how satisfied are you with the support Trommons currently offers when it comes to translation quality?
○ Very satisfied
○ Satisfied
○ Neither satisfied nor dissatisfied
○ Dissatisfied
○ Very dissatisfied

● Assuming a hypothetical scenario without resource restrictions, in which Trommons became a “quality paradise” for translators and proofreaders, which of the following features would you find most helpful? Please pick THREE items from the list below and give reasons for your choice.

○ Tutor-system for inexperienced translators
○ A project-specific communication platform for organisations, translators and proofreaders
○ Linguistic resources such as translation memories, glossaries, and style guides.
○ Option to participate in groups for specific domains (e.g. legal, medical)
○ Option to participate in groups for specific languages or language pairs
○ Language leads with proven track record as mentors and sounding boards
○ Open voting system for proposed translation segments
○ Task restrictions depending on level of professional experience/quality feedback received on Trommons (e.g. proofreading only for volunteers with several years professional experience)
○ Translated material must be published with by-line (names of translator and proofreader)
○ Being pre-selected / tested by specific organisation(s)
○ Badges that recognise experience and domain expertise
○ Please give reasons for your choice.

● Do you use translation tools for tasks you have downloaded from Trommons?
○ Yes
○ No

● If you use translation tools... a. List your favourite tools (max. of 5) b. What are the top three issues you would like to see resolved in relation to the tools you are using?
● If you do not use translation tools... Please tell us why you do not use translation tools.
  ○ Not aware of any
  ○ Too expensive
  ○ Not suitable for tasks on Trommons

● Would you like to have access to translation tools on Trommons? Please explain your answer.
  ○ Yes
  ○ No

● Would you like to have access to a mobile version of Trommons? Please explain your answer.
  ○ Yes
  ○ No

● Would a mobile version of a translation tool be useful? Please explain your answer.
  ○ Yes
  ○ No

● Is it important for you to have access to an organisation’s activity metrics (eg. project history, impact history, past reviews/comments, etc.)? Please explain your answer.
  ○ Yes
  ○ No

● Please choose the THREE most important reasons why you select a task.
  ○ Interest in volunteering in the organisation’s activity domain
  ○ Approval of the organisation’s affiliations (religious, cultural, political, financial etc.)
  ○ Impact of the specific project
  ○ Reputation of the organisation
  ○ Source and/or target language fluency
  ○ Previous professional or volunteer experience in the domain
  ○ Interest in improving language skills
  ○ Previous experience in translation (specific language pair)
  ○ Interest in gaining translation experience (any)
  ○ Interest in gaining translation experience (domain-specific)
  ○ Size of project
  ○ Deadlines
  ○ Acknowledgement of your contribution/credit/certification
  ○ Other (please specify)
● Please explain why these criteria are important to you.

● If an organisation could handpick volunteers for their next project, what do you think it would value most? Please choose THREE of the following options.
  ○ Linguistic expertise (source and/or target language fluency)
  ○ Translation experience
  ○ Domain expertise
  ○ Volunteer's areas of interest matching the organisation’s domain of activity
  ○ Previous voluntary work (any)
  ○ Previous work for their organisation
  ○ Presence on Trommons
  ○ Consistent online professional behaviour
  ○ Volunteer availability
  ○ Technical skills
  ○ History of meeting time and scope requirements from previous projects
  ○ Other (please specify)

● How likely are you to recommend Trommons to others as a place to volunteer?
  ○ Extremely likely
  ○ Very likely
  ○ Moderately likely
  ○ Slightly likely
  ○ Not at all likely

● Please share any other ideas/suggestions related to Trommons.

Questionnaire 2

● Number of staff
  ○ a. 0-10
  ○ b. 11-50
  ○ c. 51-100
  ○ d. 100+

● Which domain is covered by your organisation?
  ○ Culture and Recreation
  ○ Education and Research
  ○ Health
  ○ Social Services
  ○ Environment
  ○ Development and Housing
  ○ Law, Advocacy and Politics
  ○ Philanthropy/Volunteerism
○ International
○ Religion
○ Professional Association/Union
○ Other (please specify)

• Do you work with volunteers?
  • Yes
  • No

• If yes, how many?
  o a. 0-10
  o b. 11-50
  o c. 51-100
  o d. 100+

• How did you hear about Trommons?
  o Internet search
  o Social media
  o Through a network
  o Through other organisations that were already registered
  o Volunteer centre
  o Word of mouth
  o Other (please specify)

• Is your organisation following Trommons or The Rosetta Foundation on any of the following social media channels?
  o Facebook
  o Google +
  o LinkedIn
  o Twitter
  o YouTube
  o None of the above

• Have you created a project on Trommons?
  o No
  o Yes (Please go directly to question 8)

• Why have you never created a project?
  o We realised it was not what we had expected
  o Lack of time
  o Technical problems
  o We have not had any project for translation
  o Other (please specify)
- Why does your organisation work with volunteer translators?

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The quality is better than machine translation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The volunteers cover a wider range of languages than machine translation available to us.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is a good way to involve others in our community work.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To save money for our organisation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is important for us to get our message out quickly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We want to cover a wider range of languages than the ones we currently cover using paid translators or agencies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is no other way for us to translate our content.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is no other way for us to translate content from clients / other organisations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- What is it that you are looking for in a translation? Please rank the following criteria (1 = most important, 3 = least important).
  - Accuracy (the translation has to be accurate)
  - Speed (we need translations done quickly)
  - Cost (we have a low budget)
• How much do you agree with the following statements? A quality service...

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>… must fulfil predefined requirements.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… must be free of any errors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… does not have to be perfect, but it must serve its purpose.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… exceeds our expectations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>… is available at the given deadline.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Which of the following criteria are the most important for you when it comes to translation quality? Please rank them (1 = most important, 4 = least important).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>The translation sounds like a text that was originally written in the target language.</td>
<td></td>
</tr>
<tr>
<td>The translation precisely renders the meaning of the original text.</td>
<td></td>
</tr>
<tr>
<td>The translation fulfils real-world requirements (e.g. legally, culturally etc.).</td>
<td></td>
</tr>
<tr>
<td>The translation fulfils the design requirements (formatting, tags etc.).</td>
<td></td>
</tr>
</tbody>
</table>

• Have you encountered any problems in the past when it comes to quality on Trommons?
  o Yes
  o No (Please go to question 17)

• Which problems have you encountered on Trommons in relation to quality?
<table>
<thead>
<tr>
<th>Category</th>
<th>Very often</th>
<th>Often</th>
<th>Sometimes</th>
<th>Rarely</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical problems (e.g. file format, embedded images)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic problems (e.g. terminology, grammar, style, register) with <em>original (source language)</em> file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic problems (e.g. terminology, grammar, style, register) with <em>translated (target language)</em> file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems with <em>linguistic resources</em> (e.g. glossary, style guides)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems contacting the volunteer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems contacting The Rosetta Foundation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- How did you solve the problem(s) you have encountered? (multiple selections possible)
  
  o We fixed it ourselves.
  o We ignored it.
  o We contacted the volunteer.
  o We contacted our own experts to rectify the issue.
  o We contacted The Rosetta Foundation for help.
  o We posted the task on Trommons again.
  o We used the Trommons Forum to ask a question / discuss the issue with volunteers.
  o Other (please specify)
• What could have helped you to resolve the issue more easily? Comment:

• Do you review quality after the translation/proofreading tasks on Trommons have been completed?
  - Yes
  - No

• If yes, how? Comment:

• Overall, how satisfied are you with the support Trommons currently offers when it comes to translation quality?

<table>
<thead>
<tr>
<th>Very satisfied</th>
<th>Satisfied</th>
<th>Neither satisfied nor dissatisfied</th>
<th>Dissatisfied</th>
<th>Very dissatisfied</th>
</tr>
</thead>
</table>

• Assuming a hypothetical scenario without resource restrictions, in which Trommons became a “quality paradise” for organisations, which of the following features would you find most helpful? Please pick three items from the list below and give reasons for your choice.

<table>
<thead>
<tr>
<th>Features</th>
<th>My three picks</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>A project-specific communication platform for organisations, translators and proofreaders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Option to select volunteers from groups for specific domains (e.g. legal, medical)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Option to select volunteers from groups for specific languages or language pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Option to send tasks to preferred (by you) volunteers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Option to send tasks to pre-selected (by us) volunteers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open voting system for proposed translation segments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Task restrictions depending on level of professional experience/quality feedback received on *Trommons* (e.g. proofreading only for volunteers with several years of professional experience)

Option to define a “quality priority” for each task (e.g. speed, design etc.)

Possibility of excluding volunteers if previous work was not satisfactory

Possibility of rewarding volunteers directly for exceptional work

Badges that recognise experience and domain expertise

- Please share any other ideas/suggestions concerning translation quality on *Trommons.*
  
  **Comment:**

- Is it important for you to have access to user activity metrics (e.g. task claim/completion history, tag subscription, past reviews, consistency, submission delays) of volunteer translators?
  - Yes
  - No

- Please give reasons for your choice.
  
  **Comment:**

- Please pick the three most important reasons why you would select a volunteer.

| Linguistic expertise (source and/or target language fluency) | |
| Translation experience | |
| Domain expertise | |
| Volunteer's areas of interest matching our domain of activity | |
| Previous voluntary work (any) | |
| Previous work for our organisation | |
Presence on Trommons
Consistent online professional behaviour
Volunteer availability
Technical skills
History of meeting time and scope requirements from previous projects
Other (please specify)

- Please explain why these criteria are important to you.
Comment:

- Why do you think volunteer translators claim a specific task? (Please pick three)

| Interest in volunteering in the organisation’s activity domain |
| Approval of the organisation’s affiliations (religious, cultural, political, financial etc.) |
| Impact of the specific project |
| Reputation of your organisation |
| Source and/or target language fluency |
| Previous professional or volunteer experience in the domain |
| Interest in improving language skills |
| Previous experience in translation (specific language pair) |
| Interest in gaining translation experience (any) |
| Interest in gaining translation experience (domain-specific) |
| Size of project |
| Deadlines |
| Acknowledgement of their contribution/credit/certification |
• How likely are you to recommend Trommons to other organisations as a place to have their projects translated?

<table>
<thead>
<tr>
<th>Extremely likely</th>
<th>Very likely</th>
<th>Moderately likely</th>
<th>Slightly likely</th>
<th>Not at all likely</th>
</tr>
</thead>
</table>

• Please share any other ideas/suggestions related to Trommons. Comment:

Questionnaire 3
Thank you for participating in this survey. We define translation as the linguistic and cultural adaptation of text-based content in any media format digital or otherwise. Please answer questions in English if you are able to, if necessary then Spanish is also okay.

1. Which of the following categories * describes you best?
   • Student, University of Limerick
   • Student, Dublin City University
   • Student, Universidad Alfonso X el Sabio
   • Lecturer, University of Limerick
   • Lecturer, Dublin City University
   • Lecturer, Universidad Alfonso X el Sabio
   • Other (please specify)

2. Are you trained in reviewing texts and assessing quality? Please check * all that apply.
   • Yes, at an academic institution
   • Yes, at a non-academic institution
   • Yes, at a for-profit company
   • Yes, at a non-profit organisation
   • No, I have not received translation training
   • Other (please specify)

3. Do you have professional translation * experience?
• Yes, Fewer than 12 months
• Yes, 1 up to (not including) 5 years
• Yes, 5 up to (not including) 10 years
• Yes, 10 up to (not including) 15 years
• Yes, 15 up to (not including) 20 years
• Yes, 20 up to (not including) 25 years
• Yes, 25 up to (not including) 30 years
• Yes, 30+ years
• No, I have no professional translation experience
• Other (please specify)

4. Do you consider yourself a native speaker of Spanish?

• Yes
• No
• Other (please specify)

5. In your own words, how would you describe translation quality?

6. Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?

• Yes, in most cases
• Yes, in some cases
• Only sometimes
• No, not usually
• I don’t know/skip
• Other (please specify)

7. In your own words, can you describe what defines a professional translator?

8. Do you have any additional comments on anything relating to this questionnaire?

Questionnaires 4 and 5: Shared demographic section
Thank you for participating in this survey. In the context of this questionnaire, we may use translation and localisation interchangeably and define "localisation" as the linguistic, cultural and
technical adaptation of digital content. We define "translation" as the linguistic and cultural adaptation of text-based content, regardless of digital or not. "Volunteering", "crowdsourcing" and "community/social localisation" refer to services performed without payment, independent of the content that is translated/localised or the processes involved. Whenever we ask you to rate or rank factors, please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different. A final comment before we get started: This survey only covers some aspects of quality and doesn't consider all possible factors that can impact quality. It makes no claims to consider quality as a whole, but only the aspects mentioned. If you feel unable to answer a question due to this, please feel free to include a comment.

1. Are you currently a student at school/* college/university?
   - Yes
   - No

2. At which level are you attending or enrolled in school/college/university?
   - Secondary student
   - Tertiary undergraduate student
   - Tertiary postgraduate student
   - Mature student studying a field I have previously worked in
   - Other (please specify)

3. Is the programme you are currently enrolled in translation or localisation * related?
   - Yes
   - No

4. What is the highest level of education you have completed? Please select the option closest to the equivalent in your country.
   - Less than high school
   - High school diploma
   - Associate, two-year, junior college degree
   - Bachelor’s degree
   - Master’s degree
   - Doctorate
   - Professional (MD, JD, DDS, etc.)
   - Do not wish to answer
   - Other (please specify)

5. Is the highest level of education you have completed so far translation/localisation related?

526
6. You have indicated that your highest level of education is not in translation or localisation. Are you otherwise trained or educated in the field? Please check all that apply.

- Yes, at an academic institution
- Yes, at a non-academic institution
- Yes, at a for-profit organisation
- Yes, at a non-profit organisation
- No, I have not received translation or localisation training
- Other (please specify)

7. Independent of your highest level of education, you have indicated that you are trained or educated in a translation or localisation related field. Please specify which. Select all that apply.

- Linguistics/translation
- Computer science/programming related with focus on localisation (e.g. localisation engineer, internationalisation ...)
- Project management or business
- Marketing
- Content creation
- Testing
- Other (please specify)

8. Do you have experience working or volunteering in a localisation related discipline? Please select all that apply.

- Yes, volunteer (translation related) at Trommons/The Rosetta Foundation
- Yes, volunteer (translation related) at other organisations
- Yes, linguist (translator, proofreader/reviewer, terminologist etc.) registered on proz.com
- Yes, linguist (translator, proofreader/reviewer, terminologist etc.) NOT registered on proz.com
- Yes, project management
- Yes, business/management
• Yes, marketing
• Yes, testing
• Yes, programming/engineering
• Yes, content creation
• No
• Yes, Other (please specify)

9. Which role have you primarily filled in the localisation/translation field?

• Buyer* of localisation/translation services (paid-for)
• Buyer of localisation/translation services (volunteer)
• Reviewer**/proofreader/tester of localisation/translation services (paid-for)
• Reviewer/proofer/tester of localisation/translation services (volunteer)
• Translator (paid-for)
• Translator (volunteer)
• Technical work in localisation/translation (paid-for)
• Technical work in localisation/translation (volunteer)
• Academic researcher in localisation/translation
• Student in localisation or translation area
• I have no work or study related ties to localisation or translation
• Other (please specify)

* By "buyer" we mean the person who decides on or recommends to decision makers whether a translation service will be requested and later its product accepted.

** By "reviewer" we mean the person who assesses translation quality and makes recommendations regarding improvements and potentially the purchasing decision.

Questionnaire 4

64. For how long have you been involved in localisation/translation?

• Fewer than 12 months
• 1 to 5 years
• 6 to 10 years
• 11 to 15 years
• 16 to 20 years
• 21 to 25 years
• 26 to 30 years
• 31+ years
• Other (please specify)
65. What is your general view on volunteer translations (also called crowdsourcing, community translation or social translation)?
   - I think it is overall positive
   - I think it is overall negative
   - No opinion/neutral

66. Would you like to add your thoughts on the previous question?

67. How would you describe the average translation volunteer?

68. Why do you think people provide translations voluntarily, without payment? Please pick the main 3 motivations. I think volunteer translators ...
   - ... want to do good and help other people
   - ... care for a specific cause
   - ... want to get experience with translation
   - ... enjoy translation as a hobby
   - ... aren't very good translators and need to improve
   - ... are on a break and bored
   - ... want to learn about new things
   - ... want to be affiliated with a non-profit organisation for career purposes
   - ... want to contribute to a creative effort they are passionate about
   - Other (please specify)

69. In your own words, how do you define translation/localisation quality?

Reminder: Whenever we ask you to rate or rank factors, please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different.

This survey only covers some aspects of quality and doesn't consider all possible factors that can impact quality. It makes no claims to consider quality as a whole, but only the aspects mentioned. If you feel unable to answer a question due to this, please feel free to include a comment.

70. Below are some criteria commonly mentioned in relation to translation and localisation quality. Please rank them by importance with 1 being the most important. Please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user.

You can sort them by dragging and dropping the individual options into the preferred order.
   - Spelling, Grammar, Punctuation (there are no linguistic errors)
   - Accuracy, Omissions, Additions (the meaning is represented correctly)
   - Terminology (correct Terminology)
   - Consistency (the same terms, style etc. are used throughout the translation)
   - Style, Register (the translation sounds good/appropriate)
   - Adherence with customer specifications
   - Formatting
   - Technical requirements
   - Factors not listed here
71. Which factors do you think impact translation/localisation quality most? Pick 3
- Translator/reviewer training or education
- Translator/reviewer experience
- Source text quality
- Source text type/topic
- Translator/reviewer level of motivation
- Translator/reviewer level of familiarity with the subject matter
- Information/guidelines provided
- Translator/reviewer high or low payment
- Technology and tools used (CAT tools, glossaries etc.)
- Translator/reviewer works into their native language
- Time spent on translation
- The quality checks performed after translation/review
- Internationalisation (technical)
- Other (please specify)

72. In your opinion, how important are the following when it comes to the quality of a translation/localisation one can expect to receive?

- High impact on quality
- Some impact on quality
- No or nearly no impact on quality
- Payment/rates
- Formal training
- Experience
- Other (please specify here)

73. Would you like to add your thoughts to the previous question?

74. How do you evaluate quality if you are not provided with customer requirements? Select * all that apply.
- I use criteria developed by an individual or organisation other than myself
- I use criteria developed by an individual or organisation other than myself. I have adapted these criteria to suit my needs
- I use my own assessment criteria
- Other (please specify)

Would you like to comment?

75. Do you think that the majority of reviewers would agree with your change requests if you were to do a full review on a translation? Please assume that both you and the other reviewer are given the same translation.
- Yes, for the majority of criteria
- Yes, but only for a few criteria
- No, not for most criteria
- No, I think it would be mostly a matter of individual preference
• I don’t know what others are looking for in a translation/Can’t say

76. Please assume you have been asked to review a translation. Rank the following by quality you’d expect the translator to deliver, with 1 being the highest. Please assume that in all cases, the translator knows the content equally well.
You can sort them by dragging and dropping the individual options into the preferred order.
• Unpaid volunteer
• Paid 5 cents per word
• Paid 10 cents per word
• Paid 15 cents per word

77. Please assume you have been asked to review a translation. Rank the following by quality you’d expect the translator to deliver, with 1 being the highest. Please assume that in all cases, the translator knows the content equally well.
You can sort them by dragging and dropping the individual options into the preferred order.
• Volunteer translator, 5 years experience
• Volunteer translator, 3 months experience
• Paid 5 cents per word, 5 years experience
• Paid 5 cents per word, 3 months experience
• Paid 15 cents per word, 5 years experience
• Paid 15 cents per word, 3 months experience

78. Please assume you have been asked to review a translation. Rank the following by quality you’d expect the translator to deliver, with 1 being the highest. Please assume that in all cases, the translator knows the content equally well.
You can sort them by dragging and dropping the individual options into the preferred order.
• Volunteer translator, WITH formal training
• Volunteer translator, NO formal training
• Paid 5 cents per word, WITH formal training
• Paid 5 cents per word, NO formal training
• Paid 15 cents per word, WITH formal training
• Paid 15 cents per word, NO formal training

79. Would you like to comment?

80. Do you think you would be able to tell from looking at a translation whether it was expensive or cheap?
• Yes, always
• Yes, in most cases
• No, not in most cases
• No, Never

Would you like to comment?

81. Below are some reasons why translators work at much lower rates than others within the same
language pair. Please rank them with 1 being the most common. You can sort them by dragging and dropping the individual options into the preferred order.

The translator ...
- ... is not well trained/qualified
- ... is new to the field and wants to gain experience
- ... works in a country where cost of living is low
- ... works in a field or language pair where rates are usually low
- ... doesn’t do good work and already lost clients
- ... urgently needs money and is working extra hours
- ... plans to work quickly without much care to justify the low rate
- ... doesn’t have good tools so doesn’t have to cover an investment
- ... is naive and doesn’t understand that the rates are too low

82. Would you like to comment?
83. Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?
- Yes, in most cases
- Yes, in some cases
- No, not usually
- I don’t know/skip

Would you like to comment?
84. In your own words, can you describe what defines a localisation or translation professional?
85. We are nearly done with this part! Do you have any additional comments on anything relating to this questionnaire?
86. Do you speak either German or Spanish? If you speak both, please select the language where you have higher proficiency.
- I speak German
- I speak Spanish
- I speak neither German nor Spanish

The following sections were also asked from those that had selected ‘Spanish’ knowledge. ‘German’ was replaced with ‘Spanish’ in this case.

87. Please select your level of knowledge for German
- I do not speak German
- Beginner (A)
- Intermediate (B)
- Proficient (C)
- Native speaker
88. If you provide paid-for services in German, what kind of work does it include? Please select all that apply.
- I translate into German
- I translate from German
- I review translations into German
- I review translations from German
- I do not provide paid-for services in German
- Other (please specify)
89. If you provide volunteer services in German, what kind of work does it include? Please select all that apply.
- I translate into German
- I translate from German
- I review translations into German
- I review translations from German
- I do not provide volunteer services in German
- Other (please specify)

Questionnaire 5

10. Are you trained in reviewing texts and assessing quality? Please check all that apply.
- Yes, at an academic institution
- Yes, at a non-academic institution
- Yes, at a for-profit company
- Yes, at a non-profit organisation
- No, I have not received review/translation training
- Other (please specify)

11. For how long have you been involved in localisation/translation?
- Fewer than 12 months
- 1 to 5 years
- 6 to 10 years
- 11 to 15 years
- 16 to 20 years
- 21 to 25 years
- 26 to 30 years
- 31+ years
- Other (please specify)

12. What is your general view on volunteer translations (also called crowdsourcing, community translation or social translation)?
- I think it is overall positive
- I think it is overall negative
- No opinion/neutral
13. Would you like to add your thoughts on the previous question?
14. How would you describe the average translation volunteer?
15. Why do you think people provide translations voluntarily, without payment? Please pick the main 3 motivations. I think volunteer translators …
   - … want to do good and help other people
   - … care for a specific cause
   - … want to get experience with translation
   - … enjoy translation as a hobby
   - … aren’t very good translators and need to improve
   - … are on a break and bored
   - … want to learn about new things
   - … want to be affiliated with a non-profit organisation for career purposes
   - … want to contribute to a creative effort they are passionate about
   - Other (please specify)

16. In your own words, how do you define translation/localisation quality?

Reminder: Whenever we ask you to rate or rank factors, please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different. This survey only covers some aspects of quality and doesn’t consider all possible factors that can impact quality. It makes no claims to consider quality as a whole, but only the aspects mentioned. If you feel unable to answer a question due to this, please feel free to include a comment.

17. Below are some criteria commonly mentioned in relation to translation and localisation quality. Please rank them by importance with 1 being the most important. Please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user.

You can sort them by dragging and dropping the individual options into the preferred order.
   - Spelling, Grammar, Punctuation (there are no linguistic errors)
   - Accuracy, Omissions, Additions (the meaning is represented correctly)
   - Terminology (correct Terminology)
   - Consistency (the same terms, style etc. are used throughout the translation)
   - Style, Register (the translation sounds good/appropriate)
   - Adherence with customer specifications
   - Formatting
   - Technical requirements
   - Factors not listed here

18. When you receive a translation, how do you decide if you are comfortable using that translation for your project assuming you don’t speak the language in question?
19. Which factors do you think impact translation/localisation quality most? Pick 3
   - Translator/reviewer training or education

534
• Translator/reviewer experience
• Source text quality
• Source text type or topic
• Translator/reviewer level of motivation
• Translator/reviewer level of familiarity with the subject matter
• Information/guidelines provided
• Translator/reviewer high or low payment
• Technology and tools used (CAT tools, glossaries etc.)
• Translator/reviewer works into their native language
• Time spent on translation
• The quality checks performed after translation/review
• Internationalisation (technical)
• Other (please specify)

20. In your opinion, how important are the following when it comes to the quality of a translation/localisation one can expect to receive?

   o High impact on quality
   o Some impact on quality
   o No or nearly no impact on quality

• Payment/rates
• Formal training
• Experience
• Other (please specify here)

21. Would you like to add your thoughts on the previous question?

22. Please rank the following by quality you’d expect a translator to deliver, with 1 being the highest.
   Please assume that in all cases, the translator knows the content equally well.
   You can sort them by dragging and dropping the individual options into the preferred order.

   • Unpaid volunteer
   • Paid 5 cents per word
   • Paid 10 cents per word
   • Paid 15 cents per word

23. Please rank the following by quality you’d expect a translator to deliver, with 1 being the highest.
   Please assume that in all cases, the translator knows the content equally well.
   You can sort them by dragging and dropping the individual options into the preferred order.

   • Volunteer translator, 5 years experience
   • Volunteer translator, 3 months experience
   • Paid 5 cents per word, 5 years experience
   • Paid 5 cents per word, 3 months experience
   • Paid 15 cents per word, 5 years experience
   • Paid 15 cents per word, 3 months experience
24. Please rank the following by quality you’d expect a translator to deliver, with 1 being the highest.
Please assume that in all cases, the translator knows the content equally well.
You can sort them by dragging and dropping the individual options into the preferred order.

- Volunteer translator, WITH formal training
- Volunteer translator, NO formal training
- Paid 5 cents per word, WITH formal training
- Paid 5 cents per word, NO formal training
- Paid 15 cents per word, WITH formal training
- Paid 15 cents per word, NO formal training

25. Would you like to comment?

26. Do you think you would be able to tell from the comments of a reviewer whether the translation was expensive or cheap?

- Yes, always
- Yes, in most cases
- No, not in most cases
- No, Never

Would you like to comment?

27. Below are some reasons why translators work at much lower rates than others within the same language pair. Please rank them to indicate how common you think they are, with 1 being the most common. The translator ...
You can sort them by dragging and dropping the individual options into the preferred order.

- ... is not well trained/qualified
- ... is new to the field and wants to gain experience
- ... works in a country where cost of living is low
- ... works in a field or language pair where rates are usually low
- ... doesn’t do good work and already lost clients
- ... urgently needs money and is working extra hours
- ... plans to work quickly without much care to justify the low rate
- ... doesn’t have good tools so doesn't have to cover an investment
- ... is naive and doesn’t understand that the rates are too low

28. Would you like to comment?

29. Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?

- Yes, in most cases
- Yes, in some cases
- No, not usually
- No, never
- I don’t know/skip
30. In your own words, can you describe what defines a localisation or translation professional?

31. Do you have any additional comments on anything relating to this questionnaire?

**Questionnaire 6**

Thank you for participating in this survey. In the context of this questionnaire, we define "translation" as the linguistic and cultural adaptation of text-based content, regardless of digital or not. "Volunteering", "crowdsourcing" and "community/social translation" refer to services performed without payment, independent of the content that is translated or the processes involved. Whenever we ask you to rate or rank factors, please assume that the text is a description of the features and benefits of a popular software product like MS Office, aimed at the end-user. Please assume that all translators were given the same amount of time, tools, information etc. - only the variables mentioned are different.

A final comment before we get started: This survey only covers some aspects of quality and doesn't consider all possible factors that can impact quality. It makes no claims to consider quality as a whole, but only the aspects mentioned. If you feel unable to answer a question due to this, please feel free to include a comment.

1. What is your connection to The Rosetta Foundation/Trommons? Please select the option that matches you the closest.

   - I am the main contact person for translation projects
   - I am one of several contacts for translation projects
   - I am the main contact person but I don't manage translation projects
   - I have occasional contact, but not for translation projects.
   - I am a translator for The Rosetta Foundation/Trommons
   - I have no ties to The Rosetta Foundation/Trommons
   - I don't know what The Rosetta Foundation or Trommons are
   - Other (please specify)

2. Including your contact with The Rosetta Foundation/Trommons, have you spent time managing translations? If so, how long?

   - No, never
   - Only on occasion
   - Regularly, less than 12 months experience
   - 1 up to (not including) 5 years
   - 5 up to (not including) 10 years
   - 10 up to (not including) 15 years
   - 15 up to (not including) 20 years
   - 20 up to (not including) 25 years
   - 25 up to (not including) 30 years
• 30+ years
• Other (please specify)

3. Apart from your current position, do you have experience working in a translation related discipline? Please select all that apply.

• Yes, linguistics/translation
• Yes, computer science/programming related with focus on international software (e.g. localisation engineer, internationalisation ...)
• Yes, project management or business
• Yes, marketing
• Yes, content creation
• Yes, testing
• No, I do not have work experience in a translation related field
• Other (please specify)

4. Do you have training or education in a translation related field? Please select * all that apply

• Yes, linguistics/translation
• Yes, computer science/programming related with focus on international software (e.g. localisation engineer, internationalisation ...)
• Yes, project management or business
• Yes, marketing
• Yes, content creation
• Yes, testing
• No, I do not have training or education in a translation related field

5. Is this your main area of training or education?

• Yes
• No
• Other (please specify)

6. Are you trained in reviewing texts and assessing quality? Please check all that apply.

• Yes, at an academic institution
• Yes, at a non-academic institution
• Yes, at a for-profit company
• Yes, at a non-profit organisation
• No, I have not received review/translation training
• Other (please specify)
7. What is your general view on volunteer translations (also called crowdsourcing, community translation or social translation)?

- I think it is overall positive
- I think it is overall negative
- No opinion/neutral

8. Would you like to add your thoughts on the previous question?

9. How would you describe the average translation volunteer?

10. Why do you think people provide translations voluntarily, without payment? Please pick the main 3 motivations. I think volunteer translators ...

- ... want to do good and help other people
- ... care for a specific cause
- ... want to get experience with translation
- ... enjoy translation as a hobby
- ... aren’t very good translators and need to improve
- ... are on a break and bored
- ... want to learn about new things
- ... want to be affiliated with a non-profit organisation for career purposes
- ... want to contribute to a creative effort they are passionate about
- Other (please specify)

The following two questions are optional, if you find them too difficult to answer. However, please feel free to answer impulsively - the goal here is to get your personal thoughts, not a generally accepted or "correct" definition.

11. In your own words, how do you define translation quality?

12. If you find it difficult to define translation quality, can you tell us what factors are most important to you or your organisation when it comes to translations?

13. When you receive a translation, how do you decide if you are comfortable using that translation for your project assuming you don’t speak the language in question?

The following question is optional, if you find it too difficult to answer. However, please feel free to answer impulsively - the goal here is to get your personal thoughts, not a generally accepted or "correct" definition.
14. Which factors do you think impact translation quality most? Pick 3

- Translator/reviewer training or education
- Translator/reviewer experience
- Source text quality
- Source text type or topic
- Translator/reviewer level of motivation
- Translator/reviewer level of familiarity with the subject matter
- Information/guidelines provided
- Translator/reviewer high or low payment
- Technology and tools used (CAT tools, glossaries etc.)
- Translator/reviewer works into their native language
- Time spent on translation
- The quality checks performed after translation/review
- Internationalisation (technical)
- Other (please specify)

15. In your opinion, how important are the following when it comes to the quality of a translation one can expect to receive?

- High impact on quality
- Some impact on quality
- No or nearly no impact on quality
- Payment/rates
- Formal training
- Experience
- Other (please specify here)

16. Would you like to add your thoughts on the previous question?

17. Do you think the question whether someone is a professional translator or not is a deciding factor in the quality of a translation one can expect to receive?

- Yes, in most cases
- Yes, in some cases
- No, not usually
- No, never
- I don’t know/skip
18. In your own words, can you describe what defines a translation professional?

19. Do you have any additional comments on anything relating to this questionnaire?
### Appendix F

#### Experiment 1 scenarios

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Novice translator 1</td>
<td>none</td>
<td>Exit en el máster</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>none</td>
<td>Cómo tener éxito en el máster</td>
</tr>
<tr>
<td></td>
<td>Trained + experienced translator 1</td>
<td>none</td>
<td>Alcanzar el éxito con MSc</td>
</tr>
<tr>
<td>2</td>
<td>Novice translator 2</td>
<td>none</td>
<td>Hablando de aprovechar al máximo, ¿qué es lo que quieres conseguir al comenzar este programa? A lo mejor, quieres centrarte en profundizar tus conocimientos en un área específica, conseguir referencias para ascender a un puesto específico, cambiar tu carrera profesional, hacer contactos en la industria o entrar en la Escuela Mayor. Todas, son buenas razones, el saber cuáles son tus objetivos te ayudará a centrarte en lo que es realmente importante.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td>none</td>
<td>Hablando de sacar el máximo partido a esto, ¿qué es lo que quieres lograr mediante la realización de este programa? Quizás tu objetivo es aumentar tu conocimiento en un área determinada, obtener las credenciales para conseguir un ascenso en un papel específico, alternando carreras, haciendo contactos en el sector o adentrándote en el mundo académico. Todas estas y muchas otras son buenas razones, al igual que saber cuáles son tus objetivos porque ayuda a centrase en lo que es importante.</td>
</tr>
<tr>
<td></td>
<td>Trained + experienced translator 2</td>
<td>none</td>
<td>Cuando hablamos de aprovechar al máximo este máster, nos referimos a qué quieres lograr el estudiante. Quizás tu objetivo se centre en mejorar tus conocimientos en un área concreta, conseguir una acreditación necesaria para lograr un ascenso a un puesto concreto, reorientar tu carrera profesional, conocer a gente del sector o acceder al mundo académico. Cualquier razón es buena, y tener claro tus objetivos te ayudará a centrarte en qué es importante.</td>
</tr>
<tr>
<td>Scenario</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td>Novice translator 1</td>
<td>none</td>
<td>During the programme you will likely encounter areas you find more difficult to understand or areas you want to know more about. Cultivating a curious mindset and trying to explore topics on your own is a step towards making the most of what you learn. The internet is full of possibilities, as are libraries and other practitioners in the industry. If you do come across something interesting, don't forget to share!</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>none</td>
<td>Durante el programa probablemente encontrarás áreas que te resultarán más difíciles de entender, o áreas de las que quieres saber más. Con una actitud curiosa e intentando analizar temas por tu cuenta, es un paso hacia la mayoría de lo que aprendes. Internet está lleno de posibilidades, al igual que las bibliotecas u otros profesionales de la industria. Si encuentras algo interesante, no olvides compartirlo!</td>
</tr>
<tr>
<td></td>
<td>Trained + experienced translator 1</td>
<td>none</td>
<td>Durante el programa encontrarás áreas más difíciles de entender o que te gustaría conocer con más profundidad. Cultivar una actitud curiosa e intentar, de manera personal, explorar algunos temas son un paso para conseguir sacar el mayor beneficio a tu proceso de aprendizaje. Internet está lleno de posibilidades, así como las bibliotecas y otros profesionales de la industria. Si te topa con algo interesante, no olvides compartirlo!</td>
</tr>
<tr>
<td>4</td>
<td>Novice translator 2</td>
<td>none</td>
<td>Mantente en contacto</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td>none</td>
<td>Estar involucrado durante todo el programa</td>
</tr>
<tr>
<td></td>
<td>Trained + experienced translator 2</td>
<td>none</td>
<td>No faltes</td>
</tr>
<tr>
<td>Scenario</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>5</td>
<td>Novice translator 2</td>
<td>Novice translator</td>
<td>¡Un momento apasionante! Has sido admitido para el programa máster de ciencias plurilingüe de adaptación regional de las nuevas tecnologías y ya estás listo para comenzar. Puede que no salgas muy preparado de tu carrera o quizás te llevas un par de años sin tener un aprendizaje formalizado. Sea lo que sea, aquí tienes varias ideas que no solo te puedes ayudar a lidiar con el trabajo, sino que también, te van a ayudar a aprovechar el programa al máximo.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td>Trained translator</td>
<td>¡Qué emocionante! Has sido admitido en el Multilingual Computing and Localisation cuyas siglas son MSc (Máster en Computación Multilingüe y Localización) y te estás preparando para esto. Quizás acabas de terminar tu grado o quizás hayas pasado algunos años desde que estuviste en estudios oficiales. Sea lo que sea, aquí tienes algunas ideas que podrían ayudarte a hacer frente a la carga de trabajo, pero también sacar el máximo partido al programa.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>Trained translator</td>
<td>¡Llegó el momento! Has sido admitido en el MSc Multilingual Computing and Localisation (Máster en Tecnologías Multilingües y Localización) y estás calentando motores. Puede que acabas de terminar tus estudios de grado o quizás hayas pasado algunos años que no siguen una formación académica regulada. Sea cual sea tu caso, aquí tienes algunas pautas que quizás te resulten útiles, no solo para enfrentarte a los contenidos de este programa de máster, sino también para aprovecharlo al máximo.</td>
</tr>
<tr>
<td>6</td>
<td>Novice translator 1</td>
<td>Trained translator</td>
<td>Al igual que con los exámenes, la tesis tiene que estar completa para conseguir el máster. Al inicio del programa no tienes que preocuparte por la tesis, pero es una buena idea estar al tanto de los posibles temas. Las unidades o los temas que encuentres interesantes, o las preguntas que no hayas sido respondidas todavía, son buenas direcciones para investigar más a fondo. Probablemente tengas que reflexionar la manera de organizarla, cómo encontrar el material, y cuando y cómo podrías hacer tu investigación y tu escrito.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>Trained translator</td>
<td>Al igual que los exámenes, la tesis debe completarse para obtener el título del máster. Aunque no debes preocuparte demasiado sobre la tesis al inicio del programa, es una buena idea buscar temas potenciales mientras pasa el tiempo. Los módulos o temas que encuentres especialmente interesantes o una interrogante que tenga que no hayas sido previamente respondida, son buenos rumbo para empezar a indagar. Tal vez deba también pensar en la manera de organizarla, cómo conseguir el material, y cuándo y cómo podrías investigar y redactar.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td>Novice translator</td>
<td>Al igual que con los exámenes, necesitas realizar la tesis para conseguir un título de Máster en MSc. Aunque durante el comienzo del programa no te debe preocuparte demasiado la tesis (o trabajo final), es una buena idea que vayas buscando un posible tema. Una buena forma es ver que módulos o áreas encuentras particularmente interesantes o pensar en preguntas que no han sido todavía respondidas. Puede también que debas pensar la forma en la que vas a organizar, dónde vas a encontrar el material y cuándo y como vas a poder realizar tu investigación y redacción.</td>
</tr>
<tr>
<td>Scenario</td>
<td>Translator 1</td>
<td>Translator 2</td>
<td>Text</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>--------------</td>
<td>------</td>
</tr>
<tr>
<td>7</td>
<td>Novice</td>
<td>Novice</td>
<td>Establish priorities in what to study and read</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>Establecer prioridades en la lectura y el estudio</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Trained</td>
<td>Establecer prioridades en lo relativo a qué estudiar y leer</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>Establecer prioridades a la hora de studiar y leer</td>
</tr>
<tr>
<td>8</td>
<td>Novice</td>
<td>Trained</td>
<td>In an online programme like this it is easy to disappear and fall behind. To avoid this, make use of the range of options to stay in touch with lecturers and other students, to engage with the material and to get help as well as share your own findings and ideas. From</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>emails, forums, webinars, individual sessions or face to face meetings - there are many possibilities. For this question, please select the novice translator and move to the next page. Now the original text continues. Should you fall behind on content or miss deadlines, it is important that you reach out to the lecturer(s) as quickly as possible, or if you run into bigger issues, to the course director.</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Trained</td>
<td>En un programa online como este, es fácil desaparecer y quedarse atrás. Para evitar esto, utiliza la variedad de opciones para permanecer en contacto con los lectores y otros estudiantes para dedicarte al material y recibir ayuda, así como para compartir tus propios descubrimientos e ideas. Desde emails, foros, conferencias web, clases individuales o encuentros cara a cara, hay muchas posibilidades. Esto no tiene que ser un gran problema, a menudo aparecer en las conferencias web y compartir un nuevo contenido, mantenrás esperando educadamente. Si te atasas en cuanto a contenido o en las fechas límite, es importante que contactes con los profesores lo más rápido posible, o con el director si se trata de asuntos mayores.</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>Es fácil desaparecer y atrasarse en un programa en línea como este. Para evitar esto, utiliza la gama de opciones para estar en contacto con</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Trained</td>
<td>profesores y otros estudiantes, para dedicarte al material y obtener ayuda, así como compartir tus propios hallazgos e ideas. Hay muchas posibilidades: desde correos electrónicos, foros, webinars, sesiones individuales o reuniones cara a cara. Asistir con</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>frecuencia a los webinars y revisar nuevo contenido te ayudará a ir por buen camino y no tiene por qué convertirse en gran cosa. Es importante comunicárse todo antes posible con el profesor si se atasas con el contenido o no puede cumplir con las fechas de entrega, o</td>
</tr>
<tr>
<td></td>
<td>Trained</td>
<td>Novice</td>
<td>con el director del máster si surjan problemas mayores.</td>
</tr>
<tr>
<td></td>
<td>translator</td>
<td>translator</td>
<td>En un programa online como este es fácil desaparecer y retrasarse. Para evitar esto, utilizas todas las posibilidades para mantenerte en contacto con los profesores y estudiantes, involucrarte en el material y pedir ayuda a la vez que compartes tus hallazgos. Tienes la posibilidad de utilizar los correos electrónicos, pasando por los foros, webinars, sesiones individuales o clases presenciales. Esto no debe suponer un gran esfuerzo. Participar en los webinars y actualizar los contenidos son suficientes para mantener el ritmo. Si en el caso de que te retrasar con el contenido o no cumplir con las fechas de entrega es importante que te pongas en contacto con el/los profesor/es lo antes posible, y para asuntos más complicados, con el director del curso.</td>
</tr>
<tr>
<td>Scenario</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>9</td>
<td>Novice translator 2</td>
<td>Novice translator</td>
<td>To stay on top of things and also build good habits for writing your dissertation towards the end, it is ideal to do at least a little work for the programme every day. Maybe you can read a page or two during your commute, mentally outline an essay in the shower or have a lecture running while doing the washing up? This won't replace focused longer periods of work, but it helps with remaining immersed in the topics and also slowly chips away at the overall pile of tasks.</td>
</tr>
<tr>
<td></td>
<td>Novice translator 1</td>
<td>Trained translator</td>
<td>Para estar al día con todo y desarrollarse buenas costumbres para elaborar tu tesis de cara al futuro, sería conveniente que hicieras un poco cada día. Quizá puedas leer una página o dos de camino al trabajo, esquematizar mentalmente tu ensayo en la ducha o, simplemente, ensaya una pequeña conferencia mientras friegas los platos. Esto no significa que te vayas a quitar trabajo, pero te puede ayudar a seguir centrado en el tema y a quitarte paulatinamente la cantidad de tareas.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>Trained translator</td>
<td>Estar al tanto de las cosas y también de los buenos hábitos desarrollados para escribir tu tesis al final. Es ideal hacer al menos un poco de trabajo del programa cada día. Quizá puedas leer una o dos páginas durante tu desplazamiento, mentalmente hazte un resumen en la ducha o haced una lectura rápida, por ejemplo, mientras friegas. Esto no sustituirá a los periodos de trabajo más largos y concentrados, pero te ayudará con el resto de tareas y tareas gradualmente aparte de todo el montón de trabajos.</td>
</tr>
<tr>
<td>10</td>
<td>Novice translator 2</td>
<td>Trained translator</td>
<td>Be in charge of your learning experience</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>Trained translator</td>
<td>Estate al mando de tu experiencia de aprendizaje</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td>Novice translator</td>
<td>Ser responsable de tu experiencia de aprendizaje</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td>Trained translator</td>
<td>Gestiona tu propio aprendizaje</td>
</tr>
<tr>
<td>Scenario #</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>11</td>
<td>Novice translator 1</td>
<td>Make use of having access to other students</td>
<td>Hace uso del acceso a otros estudiantes</td>
</tr>
<tr>
<td></td>
<td>Novice translator 2</td>
<td></td>
<td>Aproveche el contacto con otros estudiantes</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td></td>
<td>Aprovecha la posibilidad de contactar con otros estudiantes</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Novice translator 1</td>
<td>Know what you want to get out of it</td>
<td>Aprende lo que quieres conseguir</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td></td>
<td>Conozca lo que quiere obtener de ello</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td></td>
<td>Saber los beneficios que quieres obtener</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Novice translator 1</td>
<td>One of the great aspects of this programme is the wide diversity of students participating on all levels: different cultures; experience levels; specialities etc. Make use of that by socialising, but also by asking questions and opinions, providing your own thoughts, participating in discussions and being helpful.</td>
<td>Uno de las grandes cosas que cuenta este programa, es la gran diversidad de estudiantes de diferentes culturas, niveles de experiencia, especialidades. Usa esta oportunidad, socialízate, pregunta y opina, para adquirir tus propias ideas, participa en debates e intenta ser de ayuda para los demás.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 2</td>
<td></td>
<td>Uno de los grandes aspectos de este programa es la amplia diversidad de estudiantes que participan en todos los niveles: diferentes culturas, niveles de experiencia, especialidades, etc. Intenta y haz uso de esto no solo por relacionarte con ellos, sino también por intercambiar preguntas y opiniones, aportando tus propias ideas, participando en debates y siendo provechoso.</td>
</tr>
<tr>
<td></td>
<td>Trained translator 1</td>
<td></td>
<td>Una de las mayores ventajas de este máster, a todos los niveles, es la gran variedad de estudiantes que lo cursan; encontrarás diferentes culturas, grados de experiencia, especialidades, etc. Intenta aprovecharlo relacionándote con tus compañeros, planteándoles preguntas, interesándote por sus opiniones, aportando tus ideas, participando en debates y ofreciendo tu ayuda.</td>
</tr>
</tbody>
</table>
### Experiment 2 scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Succeeding in the MSc</td>
<td>Erfolgreich mit dem Master of Science (M.Sc.)</td>
</tr>
<tr>
<td>0.10 Euro per word</td>
<td>none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05 Euro per word</td>
<td>none</td>
<td>Erfolg im MSc</td>
<td></td>
</tr>
<tr>
<td>0.15 Euro per word</td>
<td>none</td>
<td>Erfolgreich beim MSc</td>
<td></td>
</tr>
<tr>
<td>Unpaid Volunteer</td>
<td>none</td>
<td>Hilfreiche Tipps für den MSc</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>Speaking of getting the most out of it, what is it you want to achieve by undertaking this programme? Maybe your focus is on deepening your knowledge in a specific area, getting credentials in order to get promoted into a specific role, switching careers, making contacts in the industry or getting into academia. All those and many others are good reasons, and knowing what your goals are helps with focusing on what is important.</td>
<td></td>
</tr>
<tr>
<td>0.10 Euro per word</td>
<td>none</td>
<td>Wenn wir davon sprechen, den besten Nutzen daraus zu ziehen: „Was möchten Sie durch Ihre Teilnahme an dem Programm erreichen?“ Vielleicht konzentrieren Sie sich darauf, Ihr Wissen in einem bestimmten Bereich zu vertiefen, Sie erhalten Berechtigungsnachweise, um für eine spezifische Rolle befördert zu werden, Quereinstieg, Kontakte zur Branche zu knüpfen oder in die Welt der Wissenschaft einzudringen. All jene und viele andere, sind gute Gründe, und da Sie Ihre Ziele kennen, sind dies Hilfen, um sich auf das Wesentliche zu konzentrieren.</td>
<td></td>
</tr>
<tr>
<td>Unpaid Volunteer</td>
<td>none</td>
<td>Da wir gerade dabei sind: Was willst du mit diesem Studium erreichen? Vielleicht liegt dein Fokus darauf, dein Wissen in einem bestimmten Bereich zu vertiefen, akademischen Referenzen für eine Beförderung zu erhalten, den Beruf zu wechseln, Kontakte in der Branche zu knüpfen oder die akademische Welt kennenzulernen. Diese und andere gute Gründe und das Wissen, was deine Ziele sind, helfen dabei dich auf das Wesentliche zu konzentrieren.</td>
<td></td>
</tr>
<tr>
<td>0.05 Euro per word</td>
<td>none</td>
<td>Wenn wir über das „Meiste“ sprechen, was willst Du mit der Belegung des Studiengangs erreichen? Vielleicht liegt Dein Fokus auf die Vertiefung Deiner Kenntnisse in einem gewissen Bereich oder Du willst Deinen Abschluss machen, um in einer bestimmten beruflichen Laufbahn aufsteigen zu können, um den Arbeitsplatz zu wechseln, um Kontakte zur Industrie zu bekommen oder um in der Wissenschaft zu arbeiten. All diese und vieles Weitere sind gute Gründe und das Wissen, was Deine Ziele sind, hilft Dir, Dich darauf zu konzentrieren, was wichtig ist.</td>
<td></td>
</tr>
<tr>
<td>0.15 Euro per word</td>
<td>none</td>
<td>Wenn wir darüber sprechen, den größtmöglichen Nutzen aus dem Programm zu ziehen, müssen wir darüber nachdenken, was du mit der Teilnahme daran erreichen möchtest. Vielleicht ist dein Ziel die Vertiefung deines Wissens auf einem bestimmten Gebiet, das Erwerben weiterer Fähigkeiten, um eine bestimmte Position zu bekommen, ein Karriere wechsel, das knüpfen von Kontakten in der Branche oder das Vordringen in die akademische Welt. All diese und noch viele andere sind gute Gründe, und deine Ziele zu kennen hilft dir dabei, dich auf das zu konzentrieren, was wichtig ist.</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Translation</td>
<td>Information Provided</td>
<td>Text</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------</td>
<td>----------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>This is a demanding programme and you will receive a lot of material and opportunities to learn. It can be tempting to try and do everything perfectly and thereby get behind on the material. It is generally a good idea to look at everything available to you and consider your goals, as well as what you need to do in order to receive your degree. Check the programme outline, module outlines and when in doubt ask your lecturers or course director. Use this to decide which content you should go through right now and which can wait until a future date.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10 Euro per word</td>
<td>none</td>
<td></td>
<td>Dies ist ein herausforderndes Programm und Sie werden eine Menge Material und Gelegenheiten zum Lernen erhalten. Es kann verlockend sein, alles perfekt auszuprobieren und perfekt zu machen und somit bei der Bearbeitung des Materials in den Rückstand zu geraten. Im Allgemeinen ist es eine gute Idee, in allen Ihnen zur Verfügung stehenden Quellen nachzuschauen und dabei Ihre Ziele zu beachten sowie das was Sie tun müssen, um Ihren Materiabschluss zu erhalten. Überprüfen Sie die Programm- und Modulübersichten und falls Zweifel auftauchen sollten, fragen Sie Ihre Dozenten oder den Direktor des Studiengangs. Verwenden Sie dies, um zu entscheiden, welchen Inhalt Sie durcharbeiten sollten und welcher bis zu einem späteren Zeitpunkt warten kann.</td>
</tr>
<tr>
<td>Unpaid Volunteer</td>
<td>none</td>
<td></td>
<td>Da dieses Programm anspruchsvoll konzipiert wurde, würst du viele Materialien und Möglichkeiten zum Lernen erhalten. Dies kann einen dazu verleiten, alles perfekt machen zu wollen, was sich oft im Stoff zurückfallt. Im Allgemeinen sollte man alle verfügbaren Materialien sichten und dabei die eigenen Ziele berücksichtigen sowie herausfinden, was getan werden muss, um den Abschluss zu erwerben. Seh dir die Studiums- und Modulhandbücher an und frage im Zweifel deinen Dozenten oder deinen Kursleiter. Dies hilft dir zu entscheiden, welche Inhalte du sofort durchgehen solltest und welche erst zu einem späteren Zeitpunkt notwendig sind.</td>
</tr>
<tr>
<td>0.05 Euro per word</td>
<td>none</td>
<td></td>
<td>Dies ist ein anspruchsvolles Studiengang und Du wirst eine Menge an Materialien und Möglichkeiten zum Lernen bekommen. Man kann versucht sein, alles perfekt zu machen und dabei der Bearbeitung des Materials im Rückstand zu geraten. Es ist allgemein eine gute Idee, sich alles anzusehen, was es überhaupt für Dich gilt und dabei Deine Ziele zu betrachten und was Du dafür brauchst, um Deinen Abschluss zu bekommen. Sehe dir den Umfang des Programms an, den Umfang des Moduls und wenn Du Zweifel hast, frage Deinen Dozenten oder Kursleiter. Benutze diese Informationen, um zu entscheiden, welche Inhalte du jetzt durchgehen solltest und welche bis zu einem späteren Termin warten können.</td>
</tr>
<tr>
<td>0.15 Euro per word</td>
<td>none</td>
<td></td>
<td>Dieses Programm ist anspruchsvoll und du wirst viel Material und viele Möglichkeiten zum Lernen bekommen. Die Verlockung ist groß, zu versuchen alles perfekt zu machen und so bei der Bewältigung des Materials in Verruf zu geraten. Allgemein ist es eine gute Idee, dass du dir alles zur Verfügung stehende aneignst und deine Ziele ins Auge fasst, sowie anschließend festlegst, was du tun musst, um deinen Abschluss zu schaffen. Seh dir die Übersicht über das Programm und die Module an und wende dich bei Fragen an deine Dozenten oder an den Studiengangleiter. Verwende dies als Anhaltspunkt dafür, welche Inhalte du jetzt durcharbeiten solltest und welche noch etwas warten können.</td>
</tr>
<tr>
<td>Scenario</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>4</td>
<td>0.10 Euro per word</td>
<td>none</td>
<td>Do a little every day</td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>none</td>
<td>Taglich ein bisschen tun</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>none</td>
<td>Mach' jeden Tag etwas</td>
</tr>
<tr>
<td></td>
<td>Unpaid volunteer</td>
<td>none</td>
<td>Investiere jeden Tag etwas Zeit in das Studium</td>
</tr>
</tbody>
</table>

To stay on top of things and also build good habits for writing your dissertation towards the end, it is ideal to do at least a little work for the programme every day. Maybe you can read a page or two during your commute, mentally outline an essay in the shower or have a lecture running while doing the washing up? This won't replace focused longer periods of work, but it helps with remaining immersed in the topics and also slowly chips away at the overall pile of tasks.

0.10 Euro per word

0.10 Euro per word

Damit Sie ständig am Ball bleiben und für das Schreiben Ihrer Dissertation, am Ende des Studiums gute Gewohnheiten aufbauen. Es ist optimal, wenn Sie zumindest jeden Tag ein bisschen für das Programm tun. Vielleicht können Sie ein oder zwei Seiten auf der Fahrt zu Ihrem Arbeitsort lesen, in der Duuse im Gedanken einen Aufsatz entwerfen oder beim Laufen bez. beim Geschirrspülen einen Vortrag halten? Dadurch werden natürlich nicht längere Arbeitszeiträume ersetzt, aber es hilft Ihnen dabei, dass Sie bei den Themenbereichen auf dem Laufenden bleiben und ebenfalls dabei, dass der allgemeine Aufgabenbeicher langsam schrumpft.

Unpaid volunteer

0.05 Euro per word

0.15 Euro per word


0.15 Euro per word

0.15 Euro per word

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Translator</th>
<th>Fee provided</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td>0.10 Euro per word</td>
<td>Make use of having access to other students</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.10 Euro per word</td>
<td>Kontakt zu anderen Studenten pflegen und nutzen</td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutze den Kontakt zu anderen Kursteilnehmern</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutzen Sie die Kontakte zu anderen Studenten</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutze die Möglichkeit, Zugang zu anderen Studierenden zu haben</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.10 Euro per word</td>
<td>Share your strengths and weaknesses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05 Euro per word</td>
<td>Stärken und Schwächen mit anderen teilen</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Teile deine Stärken und Schwächen</td>
</tr>
<tr>
<td></td>
<td>0.10 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Teile deine Stärken und Schwächen</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.10 Euro per word</td>
<td>Make use of resources available to you like the library and magazine subscriptions</td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutze verfügbare Ressourcen wie die Bibliothek und Magazin-Abos</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutze die Dir zur Verfügung stehenden Ressourcen wie die Bibliothek und Zeitschriftenabos</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Nutze die dir zur Verfügung stehenden Ressourcen, wie z.B. die Bucherei und Zeitschriftenabonnements</td>
</tr>
</tbody>
</table>
### Scenario #

<table>
<thead>
<tr>
<th>5</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
</table>

In an online programme like this it is easy to disappear and fall behind. To avoid this, make use of the range of options to stay in touch with lecturers and other students, to engage with the material and to get help as well as share your own findings and ideas. From emails, forums, webinars, individual sessions or face to face meetings – there are many possibilities. This doesn’t have to be a big deal, often showing up to the webinars and checking for new content will keep you ticking along nicely. Should you fall behind on content or miss deadlines, it is important that you reach out to the lectures as quickly as possible, or if you run into bigger issues, to the course director.

| 0.10 Euro per word | Unpaid volunteer | In einem Onlineprogramm wie diesem ist es einfach zu verschwinden und zurückzubleiben. Um dieses zu vermeiden, nutzen Sie eine Palette an Optionen, damit Sie mit den Dozenten und anderen Studenten in Kontakt bleiben, sich mit dem Material beschäftigen und Hilfe bekommen, sowie Ihre eigenen Erkenntnisse und Ideen mit anderen zu teilen. Es gibt viele Möglichkeiten, einige dieser sind E-Mails, Foren, Webinare, Einzelsitzungen bis zu persönlichen Meetings. Es dürfte kein großes Problem darstellen, bei den Webinaren anwesend zu sein und nach neuen Inhalten zu suchen und somit gut vorankommen. Falls Sie bei den Inhalten zurückfallen oder den Abgabetermin verpassen, ist es wichtig, dass Sie dieses dem/den Dozenten schnellstmöglich mitteilen. Falls ernsthafte Probleme vorliegen, dann wenden Sie sich bitte an den Direktor des Studienganges. |


| 0.15 Euro per word | 0.15 Euro per word | Bei einem Online-Studienprogramm ist es einfach, zu verschwinden und zurückzubleiben. Um das zu vermeiden, nutzen die verschiedenen Möglichkeiten, um mit den Dozenten und anderen Studenten in Kontakt zu bleiben, sich mit dem Material zu beschäftigen und um Hilfe zu erhalten und Deine Anmerkungen und Ideen zu teilen. Nutze E-Mails, Foren, Webinare, Einzelsitzungen oder persönliche Besprechungen – es gibt viele Möglichkeiten. Es sollte kein großes Problem sein, oft an Webinaren teilzunehmen und nach neuen Inhalten Ausschau zu halten, wird Dich immer schon auf dem Laufenden halten. Solltest Du beim Inhalt im Rückstand geraten oder eine Frist versäumst, ist es wichtig, dass Du so schnell wie möglich Kontakt zu(m) (den) Dozenten oder – bei größeren Problemen – mit dem Kursleiter aufnimmt. |

<p>| 0.05 Euro per word | 0.05 Euro per word | Bei Online-Programmen wie diesem ist es leicht, zu verschwinden und ins Hinterhinein zu geraten. Um das zu vermeiden, nutzen die Vielzahl der Optionen, um mit Dozenten und anderen Studenten in Kontakt zu bleiben, sich mit dem Material zu beschäftigen sowie um Hilfe zu bitten und deine eigenen Erkenntnisse und Ideen zu teilen. Von E-Mails, Foren, Webinaren, Über Einzel-Sessions bis hin zu persönlichen Treffen – es gibt viele Möglichkeiten. Das muss keine große Sache sein, durch eine häufige Teilnahme an den Webinaren und das Suchen nach neuen Inhalten hältst du dich problemlos auf dem Laufenden. Solltest du im Hinterblick auf die Inhalte ins Hinterhinein geraten oder eine Deadline verpasst haben, ist es wichtig, dass du dem/den Dozenten so schnell wie möglich Bescheid gibst, oder dich, falls größere Probleme auftreten, mit dem Studiengangsdirektor in Verbindung setzt. |</p>
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
</table>
| 10       | Unpaid volunteer | 0.10 Euro per word | Unpaid volunteer
<pre><code>     |             | 0.10 Euro per word | While you will receive material to work through, don’t forget all the other resources available to you. For this question only, please select “Unpaid volunteer” for this option and move to the next page. Now the original text continues. This will be especially important when it comes to preparing presentations, essays, reports and your dissertation. |
</code></pre>
<p>| 11       | Be in charge of your learning experience | 0.10 Euro per word | Verantwortung für Ihre Lernfahrung übernehmen |
|          |           | 0.05 Euro per word | Übernimme die Verantwortung über deine Lernfahrung |
|          |           | 0.15 Euro per word | Steure Deine eigenen Lernfahrungen |
|          | Unpaid volunteer | 0.10 Euro per word | Nimm dein Lernleben selbst in die Hand |</p>
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Translator</th>
<th>Information provided</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.10 Euro per word</td>
<td>0.10 Euro per word</td>
<td>Think ahead to the exams but don’t panic</td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Früh genug an die Prüfungen denken, aber keine Panik!</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Denke früh genug an Prüfungen, aber gerate nicht in Panik</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Denke voraus an die Prüfung, aber verfalle nicht in Panik</td>
</tr>
<tr>
<td></td>
<td>Unpaid volunteer</td>
<td>0.15 Euro per word</td>
<td>Denke bereits jetzt an die Prüfungen, aber verfalle nicht in Panik</td>
</tr>
<tr>
<td>23</td>
<td>0.10 Euro per word</td>
<td>0.05 Euro per word</td>
<td>One of the great aspects of this programme is the wide diversity of students participating on all levels: different cultures; experience levels; specialities etc. Try and make use of that by socialising, but also by asking questions and opinions, providing your own thoughts, participating in discussions and being helpful.</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Einer der großartigsten Aspekte dieses Programms ist die große Vielfalt der Studenten, die auf allen Ebenen teilnehmen: unterschiedliche Kulturen, Erfahrungsniveaus, Spezialisierungen usw. Versuchen Sie es und pflegen Sie Kontakte, aber stellen Sie auch Fragen bzw. sagen Sie Ihre Meinung, bieten Sie Ihre Ideen und Einfälle an, nehmen Sie an Diskussionen teil und seien Sie hilfsbereit.</td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Ein toller Aspekt des Programms ist die Vielfalt der teilnehmenden Studenten auf allen Ebenen: verschiedene Kulturen, Erfahrungswerte, Fachgebiete etc. Versuche soziale Kontakte zu knüpfen, aber auch Fragen zu stellen und Meinungen nachzufragen, eigene Gedanken mitzuteilen, an Diskussionen teilzunehmen und anderen zu helfen.</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Einer der schönsten Seiten dieses Studiengangs ist die breite Vielfalt der Studenten, die daran auf den verschiedenen Ebenen teilnehmen: verschiedene Kulturen, Erfahrungsniveaus, Spezialisierungen usw. Geh hin und versuche Dich, mit ihnen anzufreunden, auch indem Du Fragen stellst, Meinungen äußertest, Deine eigenen Ansichten einbringst, an Diskussionen teilnimmst und mit Hilfe zur Seite stehst.</td>
</tr>
<tr>
<td></td>
<td>Unpaid volunteer</td>
<td>0.15 Euro per word</td>
<td>Einer der großartigsten Aspekte dieses Programms ist die breite Vielfalt an Studenten, die auf allen Ebenen daran teilnehmen: verschiedene Kulturen, Erfahrungsniveaus, Spezialisierungen usw. Versuche dir das zusätzliche zu machen, indem du dich mit anderen trifft, aber auch indem du Fragen stellst und Meinungen einholt, deine eigenen Gedanken einbringst, dich an Diskussionen beteiligst und hilfsbereit bist.</td>
</tr>
<tr>
<td>Scenario #</td>
<td>Translator</td>
<td>Information provided</td>
<td>Text</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>14</td>
<td>0.10 Euro per word</td>
<td>0.15 Euro per word</td>
<td><strong>Know what you want to get out of it</strong></td>
</tr>
<tr>
<td></td>
<td>0.05 Euro per word</td>
<td>0.10 Euro per word</td>
<td>Dem Nutzen genau kennen</td>
</tr>
<tr>
<td></td>
<td>0.15 Euro per word</td>
<td>Unpaid volunteer</td>
<td>Werde dir deiner Ziele bewusst</td>
</tr>
<tr>
<td></td>
<td>Unpaid volunteer</td>
<td>0.05 Euro per word</td>
<td>Mach Dir klar, was Du überhaupt willst, um dann auch zu erreichen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dein Ziel vor Augen haben</td>
</tr>
</tbody>
</table>