NOVAFCSH
FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

Building Blocks:
A Translation Research Skills Initiative

Ana Sofia Marques Martins

Relatório de Estágio de Mestrado em Tradução

(Área de Especialização em Inglês)

Maio de 2023

Relatório de Estágio apresentado para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Tradução realizado sob a orientação científica do Prof. Doutor Marco Neves
e da Doutora Marina Sánchez Torrón.

Para o meu pai, obviamente.

(Mas só se estiver bom.)

Acknowledgments

Antes de mais, quero agradecer aos meus orientadores, Prof. Doutor Marco Neves e Doutora Marina Sánchez Torrón, por todo apoio ao longo deste processo.

Ao Professor Marco, pelos mil emails e mais 500 reuniões (com e sem COVID), por arranjar solução para todo o tipo de situações (um verdadeiro MacGyver) e por, de modo geral, ser uma excelente pessoa.

À Marina, por deixar que eu perguntasse qualquer tipo de coisa, por ser tão simpática e compreensiva, por estar disponível para ajudar em qualquer momento e por acreditar que consigo fazer muito mais do que eu acho que sou capaz.

À Unbabel, principalmente à Community Team, por me acolher e proporcionar esta oportunidade de estágio.

À Érica, Marina e Raul, companheiros de procrastinação (elas) e pressão (ele). O mestrado tinha sido muito menos divertido sem vocês. Às vezes os ombros dos gigantes são os dos amigos e nós somos um grupo de animais empilhados dentro de uma gabardina.

Às minhas amigas, só para não parecer mal. Apesar de serem todas umas chatas, apoiam-me nos momentos mais difíceis. Além disso, acham que tenho capacidades para todo o tipo de carreiras: tradutora, empresária e até comediante ("És tão engraçada. Devias ir para o circo."). Infelizmente, eu até gosto delas.

Aos meus avós, por fazerem o trabalho de quatro e estarem sempre a ligar mesmo quando não é para ir lá a casa mudar de HDMI. À minha avó em especial que só quer que eu acabe de escrever isto para vermos uns filmes (Está quase!).

À minha mãe, que me deixou ir para Lisboa no pior momento das nossas vidas; por tentar não reclamar muito comigo apesar de sermos as duas muito teimosas e por fazer sempre o melhor que pode. Só falta mesmo acertar nos números do Euromilhões.

Ao meu pai, que só gostava de ler legendas e manuais de instruções e, portanto, nunca ia ler isto. As pessoas só sabem falar sobre a tua paciência, apesar de teres sido e seres muito mais que isso. Vou cobrar-te este tempo e, sim, quero um ovo estrelado.

BUILDING BLOCKS:

A TRANSLATION RESEARCH SKILLS INITIATIVE

Ana Sofia Marques Martins

Abstract

Post-editing has always been a necessary component of Machine Translation. Nevertheless, when it comes to ensuring the quality of highly specialized and localized texts, this single step is not enough, and that is where revision comes into play. Despite this, many individuals hired to perform these tasks often lack experience or education in translation, which may result in some underdeveloped skills.

This report aims to examine why trial test content reviews failed to meet the quality expectations established by Unbabel and develop solutions for these problems in the context of the integration of Lingo24, a highly specialized translation-oriented language service provider. To achieve this goal, we took multiple approaches throughout this internship.

Firstly, we examined the MQM scores of previous test trial batches of Lingo24 content and how these compared to the minimum acceptable scores defined by the company for these texts. Then, after confirming the substandard quality of these trials, we analyzed 9952 Translation History Lookup searches to identify possible common issues across reviews. Next, with the information obtained from this analysis, we created a survey that targeted Senior Editors (i.e., reviewers) to gain first-hand empirical data on their most significant difficulties, work patterns, and preferences. Finally, after studying the survey results, we created materials to help them improve their instrumental competence in the tool they use the most: Google Search.

KEYWORDS: machine translation; post-editing; machine translation revision; instrumental competence; instrumental competence training; translation research skills

BUILDING BLOCKS:

A TRANSLATION RESEARCH SKILLS INITIATIVE

Ana Sofia Marques Martins

Resumo

A pós-edição sempre foi uma componente necessária da tradução automática. No entanto, no que toca a garantir a qualidade de textos altamente especializados e localizados, só este passo não chega e é aí que a revisão entra em jogo. Apesar disto, vários indivíduos contratados para realizar estas tarefas têm falta de experiência ou formação em tradução, o que pode resultar no subdesenvolvimento de algumas competências.

O presente relatório visa analisar as razões pelas quais a revisão de conteúdo para avaliação interna não foi de encontro às expectativas de qualidade estabelecidas pela Unbabel e desenvolver soluções para estes problemas no contexto da integração da Lingo24, uma fornecedora de serviços linguísticos e tradução altamente especializada. Para atingir este objetivo, utilizamos diversas abordagens ao longo deste estágio.

Em primeiro lugar, examinámos valores de MQM de conteúdo da Lingo24 para avaliação interna e como estes se equiparam aos valores mínimos aceitáveis definidos pela empresa. Mais tarde, após confirmámos a qualidade inferior deste conteúdo, partimos para a análise de 9952 pesquisas do Translation History Lookup de forma a identificar os problemas comuns a todas as revisões. Em seguida, com a informação obtida através desta investigação, criámos um questionário com foco nos Senior Editors (ou seja, revisores) de forma a obter dados empíricos sobre as suas maiores dificuldades, ritmo de trabalho e preferências. Finalmente, após o estudo dos resultados do questionário, desenvolvemos materiais para ajudar a melhorar a sua competência instrumental na ferramenta que mais utilizam: o Google Search.

PALAVRAS-CHAVE: tradução automática; pós-edição, revisão de tradução automática; competência instrumental; treino da competência instrumental; capacidades de pesquisa tradutória

Table of Contents

Acknowledgments

Abstract

Resumo

Table of Contents

List of Acronyms

Introduction **1**

1. Host Characterization **3**

1.1. Unbabel's Translation Workflow 4

1.1.1. Pre-Lingo24 4

1.1.2. Post-Lingo24 6

1.2. Unbabel's Quality Processes 9

1.2.1. Automatic Quality Processes 9

1.2.2. Manual Quality Processes 10

1.2.2.1. Translation Errors Annotation 10

1.2.2.2. Editors' Evaluation 11

1.2.2.2.1. Post-editor 11

1.2.2.2.2. Senior Editor 13

1.3. Internship goals 13

2. State of the art **15**

2.1. Machine Translation through the Ages 15

2.2. Historical overview of post-editing and post-editors 18

2.3. Translation, post-editing, or revision? 21

2.4. Competence Models 23

2.4.1. Translation Competence Models 23

2.4.1.1. PACTE's Translation Competence Model 24

2.4.1.2. Göpferich's Translation Competence Model 25

2.4.2. Robert's Translation Revision Competence Model 26

2.4.3. Nitzke's Post-editing Competence Model 27

2.5. Instrumental competence and information literacy 28
3. Methodology 30
3.1. MQM score requirements 30
3.2. Translation History Lookup 32
3.3. Translation Research Skills Survey elaboration and participant selection 33
4. Results and discussion 35
4.1. MQM Dataset Analysis 35
4.2. THL Dataset Analysis 36
4.3. Translation Research Skills Survey findings 38
4.3.1. Participation 38
4.3.2. Demographics 39
4.3.3. Main field of study and previous experience vs. preferences 40
4.3.4. Translation History Lookup and other common issues 48
4.3.5. Professional and Creative texts 53
4.3.5.1. Professional Texts 53
4.3.5.2. Creative Texts 58
4.3.6. Key questionnaire results 62
4.3.7. Participant's Error Type Baseline per client 63
4.4. Research Skills Materials 64
4.4.1. Google Search 64
4.4.1.1. Materials structure 65
4.4.1.2. Written format (PDF) 66
4.4.1.3. Video format 67
4.4.1.4. Knowledge Base Article 70
4.4.2. Senior Editor’s feedback on Research Skills Materials 70
Contributions, Limitations, and Future Work 73
Bibliography 76
List of Figures 81
List of Tables 82
Annexes 83

List of Acronyms

AI	Artificial Intelligence
ALPAC	Automatic Language Processing Advisory Committee
CAT	Computer-Assisted Translation
CRM	Customer Relationship Management
FAQ	Frequently Asked Questions
GDPR	General Data Protection Regulation
HT	Human Translation
IBM	International Business Machines
LLM	Large Language Models
LP	Language Pair
LSP	Language Service Provider
MQM	Multidimensional Quality Metrics
MT	Machine Translation
NE	Named Entities
NMT	Neural Machine Translation
PE	Post-Editing
PEC	Post-Editing Competence
PEMT	Post-Edited Machine Translation
PII	Personal Identifiable Information
QE	Quality Estimation
SE	Senior Editor
SM	Shadow Mode
SMT	Statistical Machine Translation
SPANAM	Spanish-English Machine Translation
TAT	Turnaround Time
TC	Translation Competence
THL	Translation History Lookup
TM	Translation Memory
TP	Translation Pipeline
TRC	Translation Revision Competence

Introduction

This report was written as a requirement for the master's degree in Translation of NOVA FCSH in the context of an internship at Unbabel, a Portuguese startup that combines the power of machine translation (MT) with human post-editing (PE).

While we acquired extensive knowledge about translation-related technology throughout the master's program, the topic of MT was barely discussed in the classroom. Nevertheless, it has become increasingly evident that its uses (and misuses) are something that professional translators must face. Therefore, we believed that an internship at a world-renowned Portuguese MT-focused company would provide the best insight into this new translation reality, and that is why we chose Unbabel. This internship took place between September 2022 and March 2023, and I was accompanied on this journey by my colleague, Érica Lemos.

The need for PE was foreseen even before there were real uses of MT. Researchers always assumed that a machine would be unable to fully grasp the intricacies of a text and that someone would be required to "fix" it, that is, post-edit it. While there have been enormous improvements in the quality of MT output, in some cases, there is a need for a second step after the PE of MT: revision. While people often equate PE to revision, Unbabel considers it an extra, distinct element that guarantees the quality of the translation.

After the integration of Lingo24, a language service provider that dealt with many specialized and localized texts, this revision became imperative. Nonetheless, the company discovered during trial tests that the quality of the final product was not meeting the established threshold. Consequently, this research aims to identify the key issues behind this situation and develop strategies and materials to improve the quality of the work produced by editors.

With this goal in mind, we examined quality metrics to verify the existence of this issue, followed by an in-depth analysis of possible common issues and the gathering of structured empirical data. Afterward, taking into account all the previous results, we developed materials that fit the community's needs to improve the instrumental competence of not only Senior Editors (SE) but also post-editors. Moreover, we expanded and improved the resources available to these members of Unbabel's community.

Firstly, in **Section 1**,[1] we will present the host of this internship, Unbabel, along with its workflows and quality processes. Subsequently, in **Section 2**, we will provide a historical overview of MT and PE, followed by a reflection on the differences and similarities between the translation, PE, and revision processes, and, finally, an introduction to multiple competence models. Afterward, in **Section 3**, we will present the methodology used in the various steps of this research. Finally, in **Section 4**, we will discuss the results of every step taken, including the analysis of quality metrics and common issue datasets, as well as the data obtained from a SE survey and the creation of Research Skills materials. This section will be followed by the **conclusions** of this work, where we will turn our attention to our contributions, limitations, and possible future work.

[1] All the text in bold (e.g., Section and Annex) is a link to facilitate the navigation across this document.

1. Host Characterization

Unbabel is a Portuguese startup founded in 2013, whose primary goal is to create universal understanding through multilingual quality translation services by combining artificial intelligence (AI) with human-aided post-edition.

In this globalized era, there is a growing demand for quick and high-quality content translation in multiple languages. In fact, according to Unbabel's 2021 Global Multilingual CX Report[2], 71% of clients surveyed responded that it is vital that "a brand promotes and supports their products and services in their native language," which confirms Unbabel's belief that consumer experience is affected by language inequity (Unbabel, 2021). Furthermore, a 2020 study from CSA Research[3] further corroborates this belief, with 40% of respondents stating they would not buy a product not offered in their mother tongue (CSA Research, 2020).

Nevertheless, while machine translation (MT), by its very nature, can match this need for quicker translation services, it still falls short in some scenarios where there are high-quality requirements. With this in mind, Unbabel developed a hybrid model where the speed of MT works in unison with the excellence of human translation. This output is then evaluated by myriad quality processes that train and improve the MT and ultimately result in the best of both worlds.

Currently, Unbabel works with 69 language pairs[4] (LP) through a vast network of post-editors, reviewers, annotators, terminologists, and other specialists. In terms of content, until December 2021, the company focused on translating Customer Support-oriented content. However, with the acquisition of Lingo24, a tech-enabled Language Service Provider (LSP) with expertise in delivering multilingual localized and specialized materials (Unbabel, 2021), the plethora of subject matters, languages, and client roster increased exponentially. In fact, until its integration, Lingo24 worked with roughly 500 companies (such as Eventbrite, Virgin Pulse, and Anuvu) in more than 120 languages (Sharma, 2021).

In the upcoming sections of this chapter, we will examine in detail the inner workings of Unbabel. In **Section 1.1**, we will thoroughly analyze the previous customer

[2] Annual report based on the responses to Unbabel's Multilingual Customer Experience Survey.

[3] Formerly known as Common Sense Advisory.

[4] It is important to point out that, at Unbabel, language varieties are classified as distinct languages. For example, European Portuguese and Brazilian Portuguese are two "different" languages.

support-oriented and current post-Lingo24 integration Unbabel workflows. Subsequently, in **Section 1.2**, we will elaborate on the automatic and manual quality processes in place at the company.

1.1. Unbabel's Translation Workflow

Prior to the acquisition of Lingo24, Unbabel mainly focused on the translation of Customer Support content. Consequently, its business model was adapted to excel at translating such materials. The original workflow consisted of three distinct pipelines tailored to suit the different degrees of complexity present in this kind of content, namely: chat, "tickets,"[5] and Frequently Asked Questions (FAQs). Nonetheless, to truly understand the present workflow in use at Unbabel, it is first necessary to understand its past, in other words, its Customer Support-oriented pipelines. Since this workflow was no longer in place at the start of this internship, some of the pre-Lingo24 information (such as images) was obtained from previous Unbabel interns' reports (see: Gonçalves, 2021; Silva, 2022) in order to establish a comparison between the old and new models.

1.1.1. Pre-Lingo24

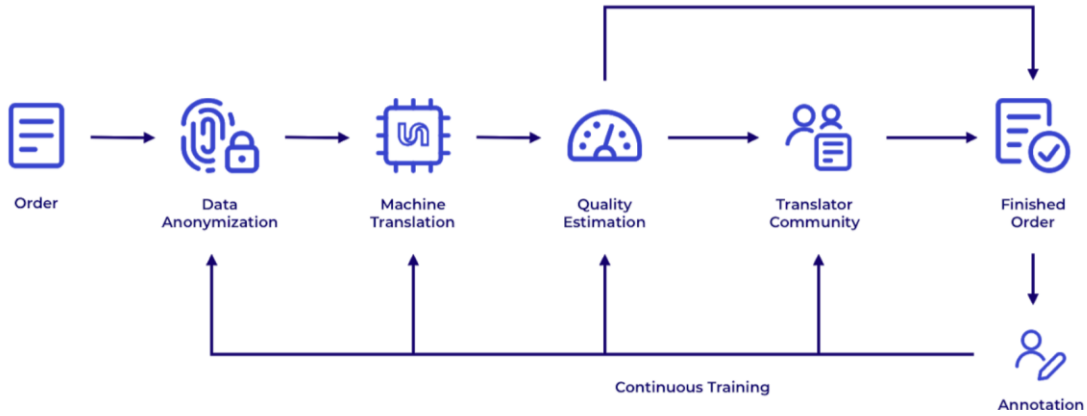


Figure 1- Pre-Lingo24 Integration General Pipeline (Gonçalves, 2021)

While there are some differences between the three original pipelines, Figure 1 illustrates the most important stages common to all of them, and we fully address any variations throughout the following explanation. First, the client submitted their order

[5] "Ticket" is the Customer Support jargon used for emails interchanged between customers and the service provider.

through their chosen Customer Relationship Management (CRM)[6] integrations offered by Unbabel, such as Salesforce, Zendesk, or Intercom.

This submission was followed by a vital step, especially when dealing with personal details such as the ones found in Customer Support, which is data anonymization. This stage is essential to ensure the privacy of Personal Identifiable Information (PII) in compliance with the General Data Protection Regulation (GDPR)[7]. This information is usually transmitted by what the industry designates as Named Entities (NE), which can include names, addresses, phone numbers, or even credit card details, and are automatically identified by the Named Entity Recognition (NER) system developed by Unbabel. After dividing the text into segmented sentences ("nuggets"), the PII was anonymized. Then, a posteriori to the complete translation process, the original data was reintegrated into the product delivered to the client. Since the process is entirely automated, this guaranteed that none of the members of the Unbabel community had access to any confidential information.

After the data anonymization step, the content was machine translated through Neural Machine Translation (NMT) (see **Section 2.1**). Subsequently, the translated nuggets went through Quality Estimation (QE), where the caliber of the translation was determined, as will be explained in **Section 1.2.1**. If the quality was deemed acceptable, the customer received the output directly, and this was designated a QE Skip. On the other hand, if the QE considered that the quality of the text was below the established threshold, it would be sent to a native-speaker-level editor of the language in question for PE. This is also one of the stages where the structure varies, as both the chat and FAQs pipelines did not possess a QE step.

Regarding the chat pipeline, since its instant messaging content required a quicker translation and not-so-strict quality requirements, where accuracy is prioritized over fluency or style, it was never sent to post-editors, making the QE unnecessary. In contrast, the FAQs pipeline did not have this step for the opposite reason. In addition, since the FAQs are usually longer texts, these were divided into smaller sections and allotted to

[6] According to Zendesk (2023), CRM is "a tool that helps companies track and drive revenue while maintaining and improving customer relationships. It enables teams to capture and manage customer interactions and sales activities across channels in one, centralized place.".

[7] According to the European Parliament this is a regulation implemented for the "protection of natural persons with regard to the processing of personal data and on the free movement of such data" (REGULATION (EU) 2016/679, 2016).

several post-editors. These texts were then, in an extra step, integrally reviewed by a Senior Editor (SE), as will be explained in further detail in **Section 1.2.2.2.2**.

In the following stage, seen in Figure 1, the completed translation was delivered to the customer. Nevertheless, the process did not end there. Professionals then annotated the finished product to once more assess the quality of the work delivered and retrain the MT.

It is also worth noting that numerous steps found in these pipelines are also present in Unbabel's current pipelines. In the following section, we will introduce Unbabel's recently developed framework and, to avoid repetition, only the most important or new stages in each corresponding pipeline will be thoroughly analyzed.

1.1.2. Post-Lingo24

The integration of Lingo24 led Unbabel to adopt its current Quality Framework, where quality evaluation consists of tiered quality ranks. This framework aims to meet a vast range of quality expectations across various clients. For example, the translation of chat content requires a faster turnaround time (TAT) but, simultaneously, a lower degree of quality. Contrastingly, content such as press releases, due to their non-perishable and publishable nature, demand more attention to detail since they might have a more direct severe impact on the client's reputation. Consequently, these intricacies require a bigger range of approaches than Unbabel had until then, and that is where the new Quality Framework comes in.

The Quality Framework is divided into six different levels, which range from 0 to 5 and have Multidimensional Quality Metrics (MQM) scores as a basis (Lommel, Burchardt, & Uszkoreit, 2014) (see **Section 1.2.2.1**). For example, Level 0, the first rank, corresponds to an MQM score of 75-85, and it applies to unedited MT output with no client-specific terminology, which might contain inaccuracies. Contrastingly, higher ranks, such as Level 4 and 5, have 94-98 and 98-100 MQM scores correspondingly, meaning that they are not only almost perfect scores but are also above what is considered the threshold for translation quality even as a human translator (Sánchez-Torrón & Koehn, 2016, p. 21).

This Quality Framework itself was “translated” into three pipelines: Realtime, Hybrid, and Specialist. Each of these is also divided into various subcategories and assigned a particular Quality Framework level[8].

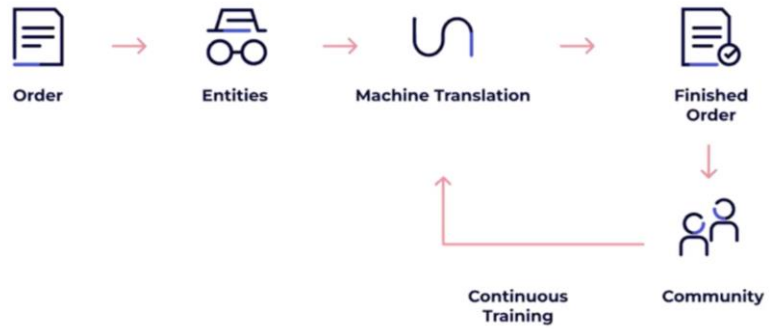


Figure 2 - "Realtime" Translation Pipeline (Internal resources)[9]

The first pipeline, titled Realtime (Figure 2), corresponds to Unbabel’s preceding Chat pipeline. The first two steps follow the same processes as “Order” and “Data Anonymization” found in Unbabel’s Customer Support-oriented pipeline and are present in every single new Quality Framework-based pipeline.

As detailed in the previous section, this structure does not only perform synchronous translations without any human input (before the delivery to the client) but also without any QE step. Nonetheless, Unbabel sends the translated content to its community, post hoc delivery, for annotation or PE for MT retraining purposes. Since this pipeline does not go through any human intervention or QE stage due to the prioritization of TAT, all these MT models use only post-edited data in an effort to maintain quality. As a result, this pipeline corresponds to the first two levels (1 and 2) of the new Quality Framework, with the lowest required MQM scores.

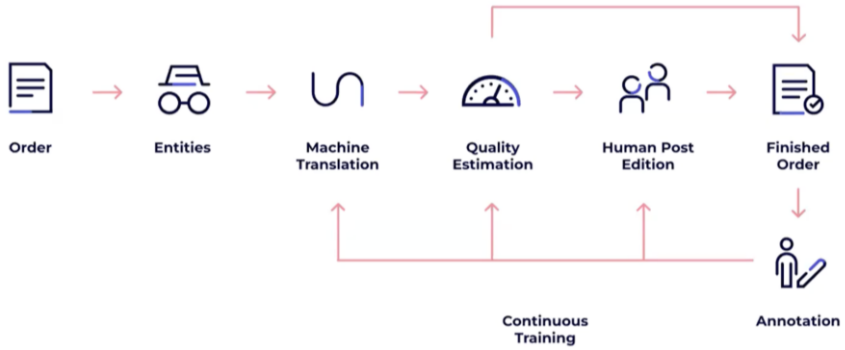


Figure 3 - "Hybrid" Translation Pipeline (Internal resources)

[8] The entirety of the Quality Framework will not be detailed in this work due to confidentiality.
[9] This and the following pipeline images were obtained from internal resources.

The second pipeline, named Hybrid (Figure 3), corresponds to Unbabel's previous "Tickets" pipeline. As aforementioned in **Section 1.1.1**, this structure features both the QE step and the human post-edition; the QE process triages the need for a human editor. Additionally, this new pipeline divides itself into three different Post-Edited Machine Translation (PEMT) subcategories, corresponding to increasing levels of QE calibration. For example, while the first subcategory does perform asynchronous translations with a degree of human PE since it requires a lower threshold of QE (in comparison to the two other subcategories), it features a higher level of automation, meaning that fewer orders go to post-editors for proofreading within the translation pipeline (TP) itself. Comparatively, the highest PEMT subcategory, which features more complex texts, is always sent to a post-editor. In addition, as in all the pipelines, after the order is complete, the data is always annotated for quality reporting and retraining of MT models. Regarding the Quality Framework, this pipeline and its subcategories range from level 1 to 3, with 94 being the highest MQM score.

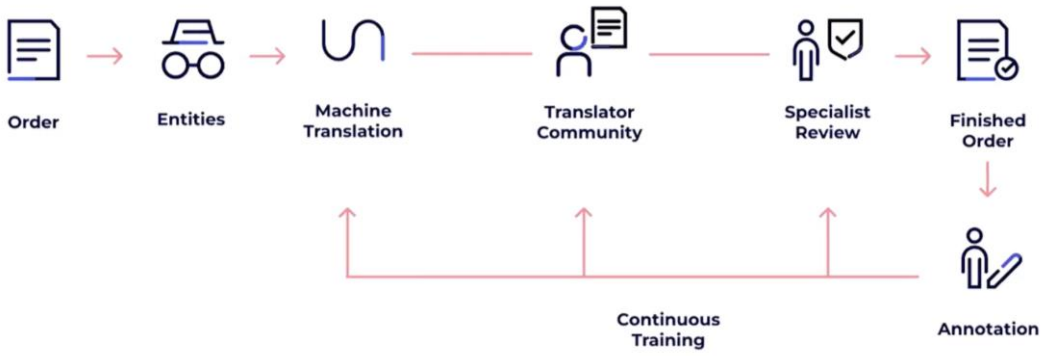


Figure 4 - "Specialist" Translation Pipeline (Internal resources)

Finally, the last pipeline, titled Specialist (Figure 4), not only shares similarities with Unbabel's previous FAQs pipeline but also aims to incorporate highly specialized and creative texts such as the ones brought by the integration of Lingo24. Therefore, this pipeline also features asynchronous translation but with a greater degree of human intervention and domain expertise. Similarly to the Realtime pipeline, this structure does not include a QE stage. Contrastingly, as was addressed in the previous section, this is due to the complexity of the texts, which would deem this step redundant since every order needs to go through a human post-editor. Nevertheless, contrary to Unbabel's older pipeline, the content is not broken down into parts and sent to multiple post-editors. Instead, the whole text is ideally sent to a community member with domain knowledge in the specific industry (e.g., medicine, engineering, …) or in a particular translation

discipline (e.g., marketing-oriented). This is followed by a review done by a SE to ensure a level of fidelity to the brand's tone of voice. Finally, the content goes through the usual annotation process for MT model retraining.

Consequently, due to the intricacy of such content, this pipeline corresponds to the highest two levels within the Quality Framework (4 and 5), which divide themselves into two subcategories, “Professional” and “Creative,” correspondingly. Since the bulk of this project focuses on improving the skills of SEs involved in these review tasks, we will further address these categories in later sections.

In conclusion, the creation of these pipelines allows Unbabel to deliver a reliable set of translation combinations optimized to the speed, quality, and cost required. Moreover, and perhaps more important from a customer standpoint, this structure allows the customers themselves to define the best plan for their needs.

1.2. Unbabel’s Quality Processes

Unbabel employs a variety of automatic and manual quality processes to guarantee the quality of its translations. As briefly mentioned in previous sections, these quality procedures can range from the assessment of the caliber of the finished product through the use of QE proprietary software, the annotation of multiple issues, and the continuous evaluation of its community.

1.2.1. Automatic Quality Processes

As a service provider whose primary goal is to deliver the best translations as quickly as possible, Unbabel must have automatic tools to speed up the process and maintain the quality standard. One of these processes is Quality Estimation (QE), which aims to “evaluate a translation system’s quality without access to reference translations” (Martins, et al., 2017, p. 205). That is, QE allows the company to determine how good a translation is through an entirely automated process.

Unbabel does this through the use of a proprietary system named OpenKiwi, which, similarly to previous more complex systems, attributes a score to the translation post-MT step and automatically decides if PE is necessary or not (Kepler, Trénous, Treviso, Vera, & Martins, 2019, p. 1). This system, as the authors point out, has the potential of “informing an end user about the reliability of automatically translated content,” “deciding if a translation is ready for publishing or if it requires post-editing,”

and “highlighting the words that need to be post-edited” to both facilitate and speed up the PE process (Kepler, Trénous, Treviso, Vera, & Martins, 2019, p. 1).

1.2.2. Manual Quality Processes

1.2.2.1. Translation Errors Annotation

Annotation was defined as the “assignment of a category to a segment of the corpus,” which is often derived from a contained, that is, a limited set of options (Lüdeling & Hirschmann, 2015, p. 136). In essence, annotation allows for the assessment of the quality of a certain text by categorizing the errors and converting them into a metric, which helps improve future work.

The current error classification system in use at Unbabel is the Multidimensional Quality Metrics (MQM), created by Arle Lommel in the context of the QTLaunch-Pad project, described as a flexible framework for “declaring translation quality evaluation metrics” (Lommel, Burchardt, & Uszkoreit, 2014, p. 456). Its creation aimed to avoid the pitfalls of generalization encountered across previous models and hoped to develop a system that could change to fit a specific purpose (Lommel, Burchardt, & Uszkoreit, 2014, p. 458). In addition, establishing a standard of evaluation would also help to avoid subjectivity inherent to the evaluation process (Lommel, Burchardt, & Uszkoreit, 2014, p. 457).

Overall, the core element of MQM is the hierarchization of issue types through the creation of guidelines to cover different levels of granularity and allow each company to decide how fine-grained those must be. In general, MQM consists of a typology with multiple parent categories and children categories, as well as different levels of severity (which indicate how damaging a specific error can be to the target text) and a scoring mechanism to assign a numeric value representing quality.

While more are available in the original MQM structure, the typology in use at Unbabel focuses on eight main categories: Accuracy (how closely the target text reflects the message within the source text), Style (various stylistic problems, such as the level of register used), Linguistic Conventions (linguistic or grammatical errors), Locale Conventions (locale-specific content or format requirements, such as measurements), Terminology, Audience Appropriateness (how a text fits a certain audience or culture), Design and Markup (only concerned with the design aspect, such as emojis or HTML), Custom (for errors that do not fit other categories). Moreover, as mentioned above, there

is also a level of severity attributed to each issue type: Neutral (only for errors in the source text), Minor (do not alter the meaning of the text), Major (may alter the meaning of the text), and Critical (alter the meaning of the text to a dangerous degree).[10]

Unbabel's in-house system, Annotation Tool, is used during this annotation step. The annotations are directly added to the text by native-level speaker annotators by selecting the "wrong" text and choosing an error category. Nevertheless, to further guarantee the quality of the translations, the annotations themselves are also evaluated. According to Lüdeling and Hirschmann (2015, p. 148), there are two ways of evaluation: through the use of a gold standard (corpora of already well-annotated texts) or through inter-annotator agreement (IAA) (that is, by comparing annotations of the same text across annotators from the same LP and calculating their level of agreement). Besides ascertaining the veracity of the quality of the translations, this system of evaluation also allows us to gauge what exactly annotators most struggle with and to improve the comprehensibility and validity of the annotation guidelines.

1.2.2.2. Editors' Evaluation

Unbabel ensures the quality of its translations through multiple steps, one of which is the continuous evaluation of its editors. The work of these individuals has an immediate and, in some cases, long-lasting impact on numerous fronts. For example, through reviews, they have a visible and instantaneous effect on the quality of the texts delivered to customers. Simultaneously, since Unbabel also uses their work to retrain MT models, any error on their part may decrease the accuracy and quality of such tools. As a result, recurrent editor evaluations are a crucial part of maintaining the company's translation quality.

1.2.2.2.1. Post-editor

When an individual tries to join the platform, they will be first asked to select their LPs. Afterward, if there is a vacancy for that particular LP, they can continue filling out their profile. If not, their data will be stored in the event of a future spot. Upon admission, these individuals will be questioned about their linguistic background, what motivates them to work for Unbabel, as well as other personal information.

[10] It is also worth noting that, similarly to the TPs, the integration of Lingo24 also resulted in some modifications to the previous MQM Custom support-oriented guidelines.

Subsequently, in order to become an editor, they must complete three steps: an interface tour; a multiple-choice test to prove their language understanding and remove any robots or people using artificial translation tools; and a skill test, that is, performing five tasks within Unbabel's online PE and revision platform, Polyglot[11], in accordance with the specified language guidelines. The latter test will be sent to Unbabel's qualified evaluators (professional translators, linguists, and terminologists), who will determine if the person should deserve a promotion to the position of "paid editor." Given that both tests only give applicants one attempt, they are warned to be well-prepared since failing will result in an LP block (Unbabel, n.d. a).

Figure 5 - Editor's evaluation rating system (Silva, 2022)

This evaluation consists of the assessment of multiple key categories (detailed in **Section 1.2.2.1**), and each task within the skill test is assigned a rating ranging from 1 (bad quality) to 5 (excellent quality), as seen in Figure 5 (Unbabel, n.d. b). If the average score is above the minimum quality threshold,[12] they are allowed access to paid assignments; if not, they remain unpaid editors. Furthermore, a higher rating per LP guarantees that these paid post-editors are the first to have access to specific tasks. However, the time span between their access and the tasks becoming available to the remaining contributors may be less than a few seconds. It is also important to point out that PE tasks have a certain time allotted to them, which Unbabel calculates by taking into consideration the number of words and complexity of the text.

In addition, as aforementioned, these evaluations are continuous, which means that even paid editors will go through frequent tests. The evaluated assignments (roughly five) are randomly selected once every three months or after a certain number of

[11] This proprietary software works similarly to a CAT tool, and possesses features such as glossaries, formatting tags, and many more.
[12] Unbabel's Community Team adjusts this minimum quality threshold per LP. For example, if there are too many editors for a certain LP, Unbabel can afford to be more selective and raise the minimum threshold.

completed tasks. If the average score is below the established minimum quality threshold, this can result in a demotion back to unpaid editor. Concurrently, other factors may lead to a promotion to Senior Editor.

1.2.2.2.2. Senior Editor

Unlike a post-editor, a Senior Editor does not only have access to PE tasks but also reviews. Reviews are lengthy texts previously distributed into smaller parts (such as FAQs) and corrected by multiple editors. Alternatively, these can also be more complex texts (seen in the Specialist pipeline in **Section 1.1.2**) directly sent to a single editor with domain expertise. Due to their inherent length, SEs are also given longer to work on reviews than on PE tasks.

For a post-editor to become a SE, they must match a certain set of requirements. While no previous experience or specific academic background is obligatory, editors must deliver excellent translations across various types of content, maintaining a minimum average score of 4.0. They also must go through at least two paid evaluations, which correspond to a minimum of six months of consistent work with Unbabel, given that post-editor evaluations are customarily conducted every three months.

However, post-editors are not automatically promoted to SEs as happens in the process from unpaid editor to paid editor. In this case, only an Unbabel Community Manager can promote an editor, and sometimes it can take months for a particular language pair SE vacancy to be available (Unbabel, n.d. c).

1.3. Internship goals

As we have seen so far, the integration of Lingo24 brought forth new types of texts, which pose unique and intricate terminological and localization challenges. However, due to the novelty of this type of content within the company, Unbabel identified that during Shadow Mode (SM), that is, trials of translated content that would not be delivered to the client but would only serve to measure the caliber of the translations and permit Unbabel to act accordingly, that these texts were not meeting the quality standards expected for Lingo24 content.

With this in mind, the main objective of this work and internship would be to identify the key challenges associated with reviewing these texts and develop solutions to improve the quality of the work that the company ultimately submits to the customer. Nonetheless, since this previous first overall assessment that the quality of the translations

did not meet the quality requirements was made prior to the beginning of this internship, it was necessary to further analyze all possibilities.

In **Section 3**, we will detail all the methodological procedures employed during this research, followed by an analysis and discussion of its findings (**Section 4**).

2. State of the art

In this chapter, we will present a historical overview of MT and the role of PE and post-editors in this process (**Sections 2.1** and **2.2**). Afterward, in **Section 2.3**, we will compare translation, post-editing, and revision concepts. Finally, in **Section 2.4**, we will analyze multiple competence models and the development of certain competences within this context (**Section 2.5**).

2.1. Machine Translation through the Ages

Machine translation (MT) refers to the use of machines, namely computers, to translate a text into another language without the need for human intervention. In fact, Kenny (2019, p. 428) describes it as "the automatic translation of text from one human language into another." Meanwhile, Carmo (2017, p. 78) considers that MT is "the most common designation for any method or system that aims at producing a version of an ST [source text] in a different language by applying automated methods.". Nevertheless, these are just two recent examples of what MT is considered to be.

If we were to uncover the roots of the concept of MT, some agree that it was introduced in the 17th century, when René Descartes proposed the creation of a universal language using symbols (Qun & Xiaojun, 2014, p. 105). Meanwhile, in the 1945 novelette "First Contact," Murray Leinster idealized the creation of a mechanical translator that allows communication with aliens (Traductanet, 2019), thus going beyond the concept of MT proposed by Kenny (2019).

Nonetheless, one can trace the first instances of actual mechanical translation (as MT was initially known) back to two French and Russian patents issued in 1933 by Georges Artsrouni and Petr Trojanskji, respectively. The first patent aimed for a more generic approach to a mechanical dictionary, whereas the second intended to interpret grammatical rules (Hutchins, 2003, p. 1).

Over a decade later, Warren Weaver's "Translation Memorandum" (1949) stimulated MT research by suggesting that machines are the solution to rising translation problems (Kenny, 2019, p. 430). After this, MT research was catapulted to new heights, with the first MT conference happening in 1952 (Hutchins, 2003, p. 2). Another important event was the first demonstration of an MT system in 1954 by International Business Machines (IBM) in collaboration with Georgetown University. In this experiment, they translated "49 Russian sentences into English, using a very restricted vocabulary of 250 words and just 6 grammar rules" (Hutchins, 2003, p. 2), which resulted, according to

Wiggins (2017, as cited in Kenny, 2019, p. 431), the long-running joke about fully automated MT being available "in five years."

Nevertheless, despite the increasing popularity of this field during the 1950s, with the emergence of multiple MT research groups around the globe, and the creation of three distinct approaches to MT (direct translation, interlingua, and transfer approach) (Hutchins, 2003, p. 3), it was clear that fully automated MT would not be ready in five years. In fact, in a survey conducted by linguist Bar-Hillel (1960), he stated that MT could only produce low-quality output, which would have to be "fixed" by post-editors (Kenny, 2019, p. 431).

However, the final nail in the coffin of MT research was not this survey but the Automatic Language Processing Advisory Committee (ALPAC) 1966 report titled "Language and Machines - Computers in Translation and Linguistics," developed at the request of the main sponsors of MT research in the USA (ALPAC, 1966). In this report, ALPAC stated that "there is no immediate or predictable prospect of useful machine translation" (ALPAC, 1966, p. 32). Moreover, it was considered more costly than human translation (HT) due largely to the lack of quality and subsequent need for PE. Nevertheless, Hutchins (2003, p. 6) remarks that ALPAC was overly focused on high-quality texts and ignored areas and clients who would accept lower-quality texts that merely conveyed essential information.

Contrastingly, while this report hindered the development of MT studies in the USA, the same did not happen abroad. For instance, in 1976, the French-English version of the SYSTRAN system that the US Air Force (Russian-English) used until 1970 was purchased by the European Commission and used to translate numerous European languages (Hutchins, 2003, p. 7). This system was in place for over three decades until its discontinuation in 2010 (Kenny, 2019, p. 432). In another continent, the Traduction Automatique de l'Université de Montréal (TAUM) created the METEO (or Météo) system in 1977, which served to translate weather forecasts across Canada. According to Kenny (2019), it was "one of the most successful MT implementations of the twentieth century" in terms of "longevity" [1977 to 2002], "productivity" [45 thousand words per day], and "quality" [around 90%] (Kenny, 2019, p. 432). In addition, in the 1980s, other systems such as Logos and METAL (German-English) also appeared. However, according to Hutchins (2003, p. 7), these were general-purpose applications, while others were designed with a specific subject matter in mind. Such is the case of the Spanish-

English Machine Translation (SPANAM) system developed by the Pan American Health Organization (PAHO) in 1980 to translate medical texts into both languages (Vasconcellos & Leon, 1985, p. 122).

As the decade drew to a close, contrary to the linguistic trend until then (rule-based machine translation), the first real-world application of a corpus-based approach to MT appeared in 1989, with the development of what would become "statistical machine translation" (SMT) by IBM through the Candide project (Hutchins, 2003, pp. 11-12). This project used the Hansard reports of the Canadian parliament to create a corpus of bilingual texts, which would then serve as the basis to calculate the statistical probabilities of any word being the "correct" translation in a particular sentence. Kenny (2019, p. 433) points out that while this approach was first "met with incredulity," it was then applauded at the turn of the century with the growth of the amount of the bitext accessible with the advent of the Internet. Simultaneously, while MT was becoming popular among the masses, most professionals disliked it and preferred other aids, such as translation memories (TM). However, this was a double-edged sword because while they did, in fact, improve their overall productivity; these TMs also served as a way to produce large amounts of bitext, which would, in turn, promote the advancement of SMT (Kenny, 2019, p. 434).

The hegemony of SMT prevailed until around two decades later when Neural Machine Translation (NMT) appeared in 2015. Just one year later, Google announced they had shifted from a Phrase-based Statistical Machine Translation (PBSMT) model to an NMT one, named Google Neural Machine Translation (Wu, Schuster, Chen, Le, & Norouzi, 2016). According to Koehn (2020, p. 31), while similar to SMT in the way that its training consists of using massive amounts of parallel corpora, this MT approach does this by replicating the workings of a brain in the sense that it consists of "artificial neurons" which activate other neurons depending on the context and processes. However, since these need to follow a step-based logical architecture, these artificial neurons and their processes are, in fact, less intricate than an actual human brain. Notwithstanding, Forcada (2017, p. 295) states that this type of MT results "resemble a text completion device." Consequently, despite the lack of precision, this NMT output compensates in fluency, resulting in a greater challenge to discover any errors (Forcada, 2017, p. 300).

Nevertheless, with the appearance of NMT, the rate of progress and research in the field of MT increased significantly. For example, Koehn (2020, p. 40) points out that in 2015 there was only one NMT system at the Conference of Machine Translation

(WMT)[13], while in the following year, there was a winning NMT system in every LP. Finally, in 2017, almost all the submissions were comprised of NMT systems. Therefore, at this point, NMT may still be considered a state-of-the-art MT system. However, with the appearance of generative AI (such as ChatGPT[14]), that may not be the case for much longer.

2.2. Historical overview of post-editing and post-editors

From the inception of MT, researchers were aware of how it would require, to some extent, pre-editing and post-editing (García, 2012, p. 294). In fact, the director of the Mechanical Translation journal, Victor Yngve, pointed out that while MT "may never be able to produce a perfect translation" (Yngve, 1954, as cited in García, 2012, p. 294), its inadequacies would increase the need for human translators. Additionally, years earlier, Yngve had, in The Machine and The Man, detailed one of the first profiles for what would be the ideal post-editor: "skilled in the output language but who may be entirely ignorant of the input language" (Yngve, 1954, as cited in García, 2012, p. 299), even though he recognized bilingualism as a benefit he did not see it a must. Moreover, they should be an "expert in the particular field of knowledge," and it alludes to the fact that the level of productivity of the PE process was in itself dependent on the quality of the MT output and its purpose (García, 2012, pp. 299-300).

Simultaneously, in the 1950s, the demand for or even the existence of post-editors could only be found in MT research fields, not in real practical work. The first documented instance of this occurring is in a Russian-English project by RAND corporation, where they expressed the need for a post-editor who "must know both the English grammar and the subject matter of the article," with the knowledge of Russian required only of the linguistics involved in the pre-editing of the text, and the professionals who would review it after PE (Edmundson & Hays, 1958, as cited in García, 2012, p. 294). This would result in the publication of a "Manual for Postediting Russian Text" in November of 1961, which was canceled shortly after (García, 2012, p. 294). Nonetheless, it is interesting to observe that here we see the first documented instance of a PE situation as the one predicted by Trojanskji almost three decades earlier in his patent,

[13] The acronym initially stood for "Workshop on Statistical Machine Translation", which was changed to in 2016, after the appearance of NMT in the previous year. See: https://www.statmt.org/wmt16/
[14] See: https://openai.com/blog/chatgpt .

where he anticipated the need for a third entity, bilingual, which would then review the text integrally (an idea he later discarded in his own undertaking) (Kenny, 2019, p. 430).

A few years later, the 1966 ALPAC report also drastically impacted the fate of PE. At that juncture, only two projects were dealing with PE, namely, the IBM Mark II and Phase II systems by the US Air Force's Foreign Technology Division (FTD) in 1964[15] and the Euratom established in 1963 in Italy (Hutchins, 1996, p. 12). Regarding the first case, ALPAC considered it a waste of funds and time since the endeavor could have been performed in a more expeditious manner by the Joint Publications Research Service (JPRS), that is, by HT. As a matter of fact, the report further elaborated that MT could not serve its final purpose without the help of PE and that this "joint effort" was not only more costly and more arduous but also produced texts of lower quality (ALPAC, 1966, pp. 43-44).

Nevertheless, while repudiating the idea of MT and its necessary PE, the report saw benefit in what they would call "machine-aided human translation" (García, 2012, p. 296). Additionally, as also pointed out by García (2012), in one of the appendices of the report (Appendix 12), they verified that the use of electronic tools such as glossaries could "reduce errors by 50% and increase productivity by over 50%" (García, 2012, p. 296).

As a result, in the aftermath of the ALPAC report, PE also lost the spotlight it had achieved during all those years of MT research. Nonetheless, as previously mentioned, as much as this report was detrimental to the research done in the USA[16], the same was not seen outside the country.

In the years that followed, another technological development occurred. Translation was a time-consuming process at the time of the ALPAC report, with translators handwriting their translations, correcting them, and finally typing them. Consequently, the average typing speed of a translator was well below what we currently consider the norm (García, 2012, p. 296). Even though many translators used dictaphones to speed up the procedure, it was still a lengthy process. However, in the 1980s, computers gradually supplanted dictation and typewriting to the point that "inputting data was no longer a skill on its own" (García, 2012, p. 297). Of course, a similar effect was seen in PE, with Kenny (2019, pp. 432-433) stating that: "In the early days, post-editors simply

[15] It employed 43 people including post-editors and translated about 100,000 words per day.

[16] For example, the Association for Machine Translation and Computational Linguistics (AMTCL) removed MT from its name and became ACL in 1968.

identified and corrected faulty MT by hand, sometimes working with pen and paper printouts; by the 1980s word-processing programs were being used for this purpose".

Simultaneously, as MT output improved, two new types of PE appeared: rapid post-editing (in Europe, also known as "light post-editing," in the USA) and full post-editing (García, 2012, p. 297). In more detail, Jeffrey Allen (2003, p. 302) describes light PE as something utilized for "assimilation," that is, created so the gist of it can be understood, meaning that texts of low quality are accepted due to their perishable nature. Contrastingly, full PE is associated by Allen (2003, p. 303) with the concept of "dissemination," meaning that a higher level of quality, akin to HT, is expected since these texts will most likely be published (Carmo, 2017, p. 154).

Concurrently, after decades of dormancy, MT research began to flourish once more, but this time in more corporate domains rather than academic ones since there was a growing awareness of the commercial viability of MT and PE. In addition, in the early stages of computers, post-editors were expected to master additional skills beyond the ones they already possessed, such as "key proficiency" or "cursor positioning" (García, 2012, p. 297). At the same time, other tools and functions were introduced, such as the "find and replace" feature or the use of macros, which is described by García (2012, p. 297) as the starting point of "automated post-editing."

Soon after, the speed of technological improvements increased significantly across various fields, primarily due to the introduction of emails and the Internet in the following years. These two innovations, when combined, enabled communication between the many participants of the MT process, specifically quick interaction between customers, the provider of the MT service, and, of course, the post-editors. As a result, in this new era, PE "became a more economically viable proposition" than it had been until then (García, 2012, p. 298). In fact, Lueke (1995, as cited in García, 2012, p. 298) stated that implementing the METAL system at SAP would increase the number of both post-editors and words translated by a wide margin in just three years. Similarly, in the European Commission, the number of pages translated increased from 30 thousand to 180 thousand in just five years (1990-95) with the use of SYSTRAN (García, 2012, p. 298).

In 1996, with the increasing demand for translated content, PE began to be done by freelancers (using diskettes) rather than by in-house personnel (García, 2012, p. 298). While the client feedback was mostly positive, Senez (1998, as cited in García, 2012, p. 298) noted that only urgent materials would apply these translation methods. The

customer would have several options, such as "full quality translation, rapid post-editing, or oral or written summaries, with the completed work bearing the heading 'Rapidly revised machine translation'" (García, 2012, p. 298). This approach was perhaps one of the precursors for models currently seen on the market, such as Unbabel's (described in detail in **Section 1.1.2**), where the client chooses which kind of translation better fits their needs. Nevertheless, it is also worth noting that most of these projects would function without direct input from post-editors. However, there were some exceptions, such as the previously mentioned SPANAM project, which had both engineers and in-house post-editors working together and proposing improvements. To some extent, Unbabel also does this through its teams of engineers and in-house linguists.

Nonetheless, despite their spotlight, MT and PE were usually disregarded by experienced translators, and most people preferred HT over PE. Furthermore, with the introduction of computer-assisted translation (CAT) tools and TMs in the 1990s, translators were thought to be prolific enough to outperform MT in both quality and speed. However, this all changed with the arrival of "Free Online Machine Translation" at the turn of the millennium. Since there were now tools such as "Google Translate," which made "translation" available to everyone, and the output was good enough that people could disregard HT and now obtain this "service" at zero cost.

Notwithstanding, and as García (2012, p. 305) concludes, even after decades of studying MT and PE, the *wh*-questions remain: *when* and *what* should be post-edited; *how* the texts should be post-edited in relation to their purpose; and *who* should be the post-editor. While these are pertinent questions not only to PE but to translation studies as a whole, in this report, we will focus on the *who*, specifically, on what kind of competences should individuals have to excel at this type of work.

2.3. Translation, post-editing, or revision?

In the preceding section, we discussed how PE was always, to some extent, a component of the MT process. While we also indicated that in specific corporate workflows (such as Unbabel's), the PE process precedes revision, PE has often been deemed a synonym for revision, with MT filling the role of HT. As a result, before analyzing the competences needed to be a professional in this area, understanding the relationship between these three processes is beneficial to our study.

While there are many definitions for what translation *is*, this research is more concerned with what the translation process *entails* and how it differs or intertwines with the rest (revision and PE). According to Carmo & Moorkens (2020, p. 36), translation "is a process of creation, and it is by writing that the translation is created," in contrast to revision which people usually view as more of an act of reading than writing (Mossop, 2020, as cited in Carmo & Moorkens, 2020, p. 116). Nonetheless, if this is the case, what implications does this have for PE?

Carmo & Moorkens (2020) go against the notion that PE is closely related to revision and explain that many translators are resistant when they encounter what they were made to believe was revision but was PE. When faced with MT output, these individuals can innately gauge that the process will be more time-consuming than a revision task, which in itself demonstrates that PE must possess distinct characteristics than revision (Carmo & Moorkens, 2020, p. 40). This, in turn, leads to the question of whether MT is equivalent to HT in the sense of a "finalized translation." The answer depends on many factors (e.g., quality of the MT models, the LP [the amount of bitext], …), but generally, this is not the case. Additionally, Carmo & Moorkens (2020) point out that due to multiple restrictions (such as time), it is doubtful that even PE will produce a finalized text. They also add that in current corporate workflows, PE is followed by a revision process "because that is the only way to guarantee fitness for purpose" (Carmo & Moorkens, 2020, p. 41), as it happens in Unbabel's more specialized texts.

This idea of PE being closer to the translation process than revision is supported by Sanchez-Gijón (2019, as cited in Carmo & Moorkens, 2020, p. 41), which sees both procedures as similar due to the use of CAT tools and the existence of fuzzy matches. Carmo & Moorkens' (2020, pp. 43-44) reinforce this concept by explaining that when professionals use CAT tools and encounter a segment with less than 75% match, it is considered something that needs to be translated and not simply edited. Consequently, since MT output can vary from 0 to 100% in quality, anything requiring above 25% of editing during a PE task should be, in fact, considered a translation action. The authors even go further by demonstrating how the four editing actions (deletion, insertion, movement, and replacement), seen in TS of Toury and Nida, are also found in PE (Carmo & Moorkens, 2020, p. 49).

In conclusion, no matter to what degree they are related to one another, we can say with certainty that translation, revision, and PE all possess some similarities. That

raises the question: would it be the same for the competences needed to perform these tasks?

2.4. Competence Models

As discussed above, there are some differences and similarities across the various processes that constitute MT, so they may require equally distinct (or similar) competences. As a result, since the bulk of this project is concerned with training a particular set of competences within the PE and revision scope, it seems pertinent to briefly examine relevant models.

Firstly, it is important to define what "competence" is in this context. The European Master's in Translation (EMT) group describes it as "the proven ability to use knowledge, skills, and personal, social and/or methodological abilities, in work or study situations and in professional and personal development" (EMT, 2022, p. 3). Consequently, we can surmise that a competence model is a structure where the necessary "qualities" of, in this case, a professional translator, post-editor, or reviewer are detailed.

2.4.1. Translation Competence Models

It is evident that because translation was the first of these processes, it was also the first to be studied by different researchers since the 1990s, who developed several translation competence (TC) models.

According to Robert, Remael, & Ureel (2017, p. 1), scholars have, through the years, evolved from the view that TC was the "summation of linguistic competence" to what today is a "multicomponential competence," which means that TC encompasses several competences. Nevertheless, since researchers disagree on which components are correct, many models were created. These TC models would eventually serve as the foundation for the development of a translation revision competence (TRC) model and a post-editing competence (PEC) model, which we will discuss in the following subsections. However, to fully comprehend these models, it is first necessary to briefly examine some TC models. Due to how extensive each one of these models is, we elaborated a table with a tentative correspondence structure across models and definitions for each competence (**Annex A**).

2.4.1.1. PACTE's Translation Competence Model

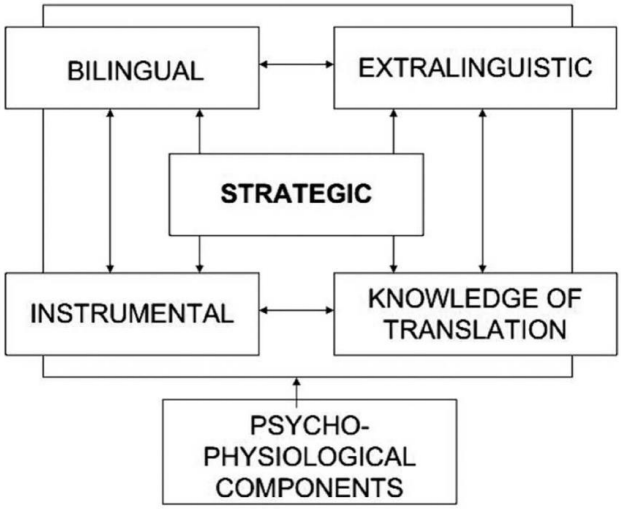


Figure 6 - PACTE's Translation Competence Model

The most widely known TC model is the PACTE model, developed in 1998 and then remodeled in 2003 after its 2000 pilot. In this remodeling, they concluded that TC is different from what the PACTE group classifies as "bilingual competence," that is, declarative knowledge (which all bilinguals speakers possess), and that it is "expert knowledge in which procedural knowledge is predominant" (PACTE, 2003, p. 59). Moreover, they also stated that this TC consists of multiple sub-competences (bilingual, extra-linguistic, instrumental, and knowledge about translation), which are all interconnected, with "strategic sub-competence" being the binding element between them (descriptions in **Annex A**). Moreover, they also add what they name "psycho-physiological components", that is, particular personal features (described by the PACTE group as: cognitive, attitudinal, and others) (PACTE, 2003, p. 59). Nonetheless, according to Pym (2013, p. 489), these last components "remain poorly categorized" in this model and the following ones.

In their study, PACTE focused on procedural knowledge competences (strategic, instrumental, and knowledge about translation) and concluded that when one competence is lacking, individuals use another to compensate for the former (specifically, strategic).

It is also important to consider that Kelly (2005, p. 33) advised that another sub-competence is essential for translators, namely, "interpersonal competence," that is, being able to collaborate with every party involved (clients, revisors, project managers, …). This concept will also make an appearance in other models.

2.4.1.2. Göpferich's Translation Competence Model

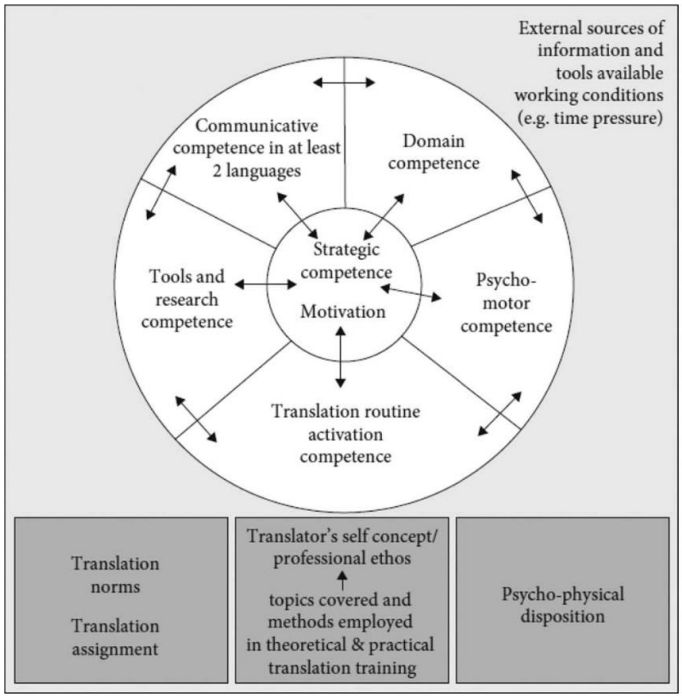


Figure 7 - Göpferich's Translation Competence Model

This model was developed years later by Göpferich in 2008, using the PACTE model as its basis. This, of course, resulted in multiple similarities between the two, such as equivalence between some of the sub-competences (in more detail in **Annex A**).

Nevertheless, this TC model also introduces another competence, namely, "psychomotor competence," that is, being able to read and write using electronic tools, which, in turn, saves time for actual translation problem-solving (Göpferich, 2009, pp. 21-22). Besides this, the author also discusses how translation norms and assignments, the translator's self-concept and professional ethos, and the translator's psycho-physical disposition, motivation, and access to external tools and working conditions may impact the entire translation process (Göpferich, 2009, p. 22). Nevertheless, unlike the previous model, this TC model was never tested.

2.4.2. Robert's Translation Revision Competence Model

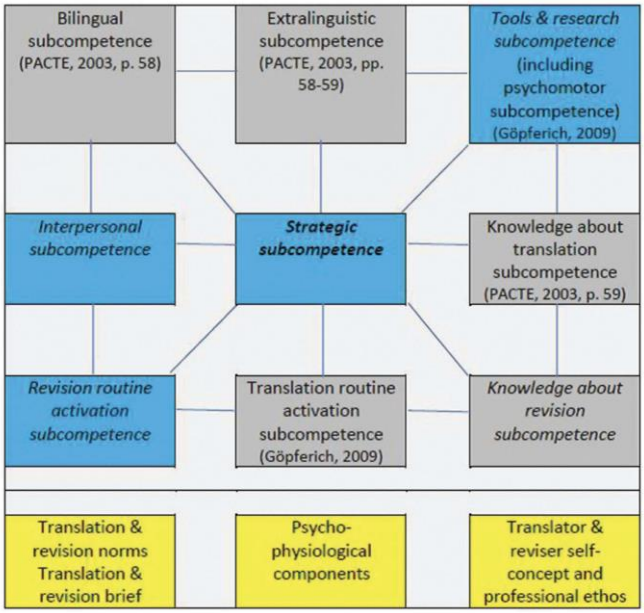


Figure 8 - Robert's Translation Revision Competence Model

Following the previous section, we now have the knowledge needed to approach the translation revision competence (TRC) model developed by Robert, Remael, & Ureel in 2017. According to Robert, Remael, & Ureel, while there were many studies concerning revision, there was not, until then, a competence model. The authors agree that the concept of revision in TS should "apply only to the revision of a translation by a reviser, who is someone other than the translator, carried out before it is delivered to the client" (Robert, Remael, & Ureel, 2017, p. 3). They solidify this idea by referring to the European Standard EN 15 038, which was replaced by ISO 17100 a posteriori to the publication of this TRC model. Nevertheless, ISO 17100 recognizes revision as a "bilingual examination of target language consent against source language content for its suitability for the agreed purpose," and it states that professionals should either have "experience in translation or the revision of the subject matter they are responsible for reviewing" (ISO 17100, 2015).

The TRC model itself consists of competences found in previous models (PACTE, Kelly, and Göpferich) (see **Annex A**). In terms of new sub-competences unique to revision, there is a "revision routine activation sub-competence" derived from Göpferich's TC version, but this time to fit revision (Robert, Remael, & Ureel, 2017, pp. 13-14). Moreover, all the other components in yellow in Figure 8 were also adapted from Göpferich's TC model.

2.4.3. Nitzke's Post-editing Competence Model

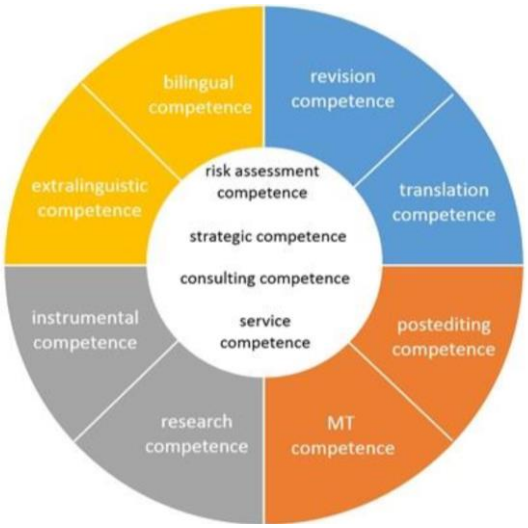


Figure 9 - Nitzke's Post-editing Competence Model

This post-editing competence (PEC) model presented in 2019 by Nitzke, Hansen-Schirra, & Canfora aims to offer a response to the competences explicitly needed for PE, and it draws inspiration from both the PACTE and the Robert models.

While a significant part of this proposal focuses on the economic risk factors of PE, the model itself comprises four core competences, namely: risk assessment, strategic, consulting, and service competence, and eight sub-competences: bilingual, extra-linguistic, instrumental, research, revision, translation, MT and PE sub competences (Nitzke, Hansen-Schirra, & Canfora, 2019, pp. 248-250). As evident from previous sections, many of these competences overlap with the ones seen in other models. Nevertheless, this model adds two new sub-competences: the MT and the PE sub-competences. In addition, it is worth noting that "instrumental" competence divides itself into two (instrumental and research competence). Moreover, the authors point out that, like in previous models, the PE task is also affected by "surrounding factors," such as: "psycho-physiological components," "the post-editor's self-perception," "the PE brief including guidelines for the PE task," "an affinity towards technology and computers," the last of which they consider essential since it is inseparable from PE (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 251).

Still, since there is such a considerable overlap between this model and the previous models, the authors conclude that there is no need for individual training of post-editors and that instead, these extra sub-competences should be integrated into translation training curricula (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 252).

2.5. Instrumental competence and information literacy

As **Annex A** shows, a few sub-competences stretch across every model discussed, namely (using PACTE nomenclature): bilingual, extralinguistic, instrumental, knowledge of translation, and strategic. If we agree with PACTE's reasoning that the first two are declarative knowledge that everyone possesses, then, when it comes to training, the focus should be on improving the remaining three sub-competences. Nevertheless, since it is the focus of this project, we shall focus on the improvement of "instrumental competence."

First, it is necessary to define this competence in more detail. Robert's (2017, p. 14) model describes it as an all-encompassing category of knowing how to simultaneously use electronic resources (like CAT tools), where to look for relevant information, and psycho-motor abilities. Contrastingly, Nitzke's (2019, p. 249) model divides it, to our understanding, into two competences: "instrumental" and "research." While "instrumental" corresponds to how to use the tools, "research" is preoccupied with how to research efficiently.

Nonetheless, this concept has been studied and categorized in TS and other areas as "information literacy." For example, the Association of College and Research Libraries (ACRL) defines it as "the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning" (ACRL, 2016, p. 8).

However, as far as we know, only one competence model focuses on developing this "information literacy," and that is the INFOLITRANS model developed by Molina & Sales (2008). While the authors consider that information literacy is unable to replace what they describe as "translation competence" (and we understand it as signifying the remaining TC sub-competences), they still agree that the development of this skill gives the translator a professional advantage (Molina & Sales, 2008, p. 433).

Throughout their proposal, the authors describe and suggest multiple classroom-oriented activities on how to develop each type of what they divide into knowledge, competences, and skills. In particular, we would like to point out what the authors name "digital-informational skills," that is, "skills that translators must be able to apply fluently and autonomously when it comes to handling and using information in their work; special

emphasis should be placed on everything concerning organization, processing, searching for, retrieving and evaluating information" (Molina & Sales, 2008, p. 427). One of these skills is partially described as the ability of "carrying out efficient searches in databases and selecting suitable query terms (controlled vocabulary, use of Boolean operators and designing search strategies)" (Molina & Sales, 2008, p. 428), which will be the focus of our work.

In this vein, there have also been many studies about research using web resources among translators (see: Enríquez Raido, 2011; Paradowska, 2015; Mohammed & Al-Sowaidi, 2023). In one of these, Enríquez Raido (2011) studies the web search behaviors of six participants (all of them with some degree of knowledge and/or experience in translation). Curiously, regarding the use of techniques to improve search engine queries, one of the participants mentioned that they only learned about certain features (in this particular case, search operators) after contacting their university's librarian within the scope of their Ph.D. research (Enríquez Raido, 2011, p. 277). Interestingly, NOVA University of Lisbon offers a course with similar content for doctoral students and investigators[17]. Nevertheless, neither the participant's inquiry nor the course by NOVA has as a primary goal translation training. Moreover, the rest of the participants, although mostly translation students, had a very shallow knowledge of these tools, except for one individual who was a translation professor (Enríquez Raido, 2011, p. 275).

Nevertheless, these studies are often preoccupied with training translation students or professional translators and use these individuals as study subjects. However, how could this "instrumental competence" or "information literacy" be improved when the subjects are not necessarily "translators" or even hope to be? Despite taking into consideration these models, the approach taken in this project had to suit a professional corporate context where the "students" may have no knowledge in the field of translation; and are not "forced" to interact with the materials as would happen in lectures and may possess different learning preferences.

[17] See: https://www.unl.pt/en/courses/study/escola-doutoral/information-literacy-course

3. Methodology

This chapter will give an in-depth description of all the methodological steps taken to develop our project during this internship. First, in **Section 3.1**, we will thoroughly discuss the MQM score requirements for Lingo24 texts, followed by an introduction to Translation History Lookup and how its analysis would further supplement our research (**Section 3.2**). Finally, we will present the elaboration of a research skills-oriented survey, for the purpose of obtaining empirical data, in **Section 3.3**.

3.1. MQM score requirements

As previously mentioned in **Section 1.1.2**, all Lingo24 clients fall under the scope of levels 4 (Excellent) and 5 (Best) of the Quality Framework in use at Unbabel. These quality levels are described internally in the following manner:

- Excellent (Level 4)
  - High level of quality
  - MQM between 94 and 98
  - The translation is fluent and accurately conveys the intended meaning. Minor style or fluency issues may still occur, but communication is effective and successful.
- Best (Level 5)
  - Premium quality
  - MQM equal or above 98
  - The translation is exceptional and displays outstanding fluency. Any potential errors are negligible and have minimal impact on the overall meaning and success of the communication.

Furthermore, these clients can be divided into two distinct pipeline subcategories, namely, “Professional Translation” and “Creative Translation,” with predetermined MQM score requirements based on both the tier of LP in use and the translation pipeline itself.

In terms of LP tiers, there are three of them, and they are determined based on the historical performance of and maturity of these language communities within Unbabel. This means that the higher the tier, the bigger the volume of work for a given language and the longer the company has worked with it. For example, one of the LPs with the biggest volume of work at Unbabel is English-German, which means it is a Tier 1 LP. In

contrast, an English-Thai LP has a lower content volume and is, therefore, a Tier 2 LP. As a result, it is also possible to say that the higher the tier, the higher the need for quality translations.

In terms of the "Professional Translation" pipeline, Unbabel expects that Tier 1 LPs will consist of 80% Best (Level 5) and 20% Excellent (Level 4) translations, while Tier 2 and 3 LPs should consist of 70% Best and 30% Excellent. The "Creative Translation" pipeline must be 95% of Best and 5% Excellent for Tier 1 LPs and 90% Best and 10% Excellent for the other two.

The combination of these two factors results in a baseline provided by Unbabel that determines the quality expectations, in other words, the MQM scores for each pipeline subcategory and LP tier (Table 1).

Translation Pipeline/ Language Pair Tiers	**Tier 1**	**Tier 2**	**Tier 3**
Professional Translation	80% Best 20 % Excellent	70% Best 30% Excellent	70% Best 30% Excellent
Creative Translation	95% Best 5% Excellent	90% Best 10% Excellent	90% Best 10% Excellent

Table 1 - Predetermined Pipeline/LP Quality Framework Percentages

In order to determine the average minimum threshold of the MQM score for each pipeline, in accordance with each LP tier, it was necessary to use the following formula[18]:

PIPELINE= (98*BEST_PERCENTAGE + (94*EXCELLENT_PERCENTAGE))

For instance, if we were to determine the average minimum quality expectations compliant MQM score for a Tier 1 LP in a Creative Translation pipeline, we should perform the following calculation: (98*0.8 + (94*0.2)), which would result in a 97.2 MQM score.

Translation Pipeline/ Language Pair Tiers	**Tier 1**	**Tier 2**	**Tier 3**
Professional Translation	97.2	96.8	96.8
Creative Translation	97.8	97.6	97.6

Table 2 - Predetermined Pipeline/LP MQM scores

[18] This is a simplified formula used to analyze the quality of content independently and not as a part of distribution.

As shown in Table 2, it was possible to conclude that the highest minimum average MQM score across all tiers and pipeline subcategories is 97.8, while the lowest is 96.8.

After we finished this preliminary step, a sizable amount of the SM test data accessible regarding Lingo24 clients had to be selected and analyzed. For this purpose, we chose four clients and assessed the MQM scores of the batches[19] completed over the course of a month. This analysis and its results will be detailed in **Section 4.1**.

3.2. Translation History Lookup

While the MQM score requirements addressed in the previous section can demonstrate whether the produced content is meeting those standards, they do not provide detailed insights into SEs' difficulties. Consequently, we decided that we should also analyze Translation History Lookup (THL) searches.

According to Unbabel's Knowledge Base article "What is a translation history lookup and how do I use it?", this is a tool that allows individuals working on the Polyglot interface to "look up individual source terms or sets of source terms in any client's database of previously-approved translations" (Unbabel, n.d. d). Generally speaking, it works the same as a concordance search done in a CAT tool. That is, the SE can select a certain portion of the text, and any matches (with varying degrees of accuracy) to previously approved and delivered translations will appear on the right side of the screen.

While it is unfeasible to obtain information about what each individual searches for on their personal computer, or the numerous tools they use outside the Polyglot interface, it is possible to observe some of their actions within the platform. In essence, when a user utilizes the THL feature to select and look up a word or phrase, this search query is stored within Unbabel's database and can be retrieved for analysis if necessary.

Through the analysis of this data, we could potentially shed some light on the difficulties encountered by SEs which, in turn, would help us develop the next step of our research. The findings of this particular analysis will be discussed in **Section 4.2**.

[19] A "batch" is a set of "jobs", also known as "tasks", which can be a group of segments or a complete text, as previously explained. In this particular case, these Lingo24 batches always signify complete texts.

3.3. Translation Research Skills Survey elaboration and participant selection

While, throughout this project, we have tried to obtain as much information as possible through MQM and THL dataset analyzes, we thought it was still important to collect first-hand empirical data about the invisible issues hindering SEs' ability to deliver quality work. We decided this should be done through a survey of the SE community since this would be the best way to "collect[ing] structured data on a large scale" (Saldanha & O'Brien, 2014, p. 152).

According to Saldanha & O'Brien (2014, p. 153), the most challenging part of designing a survey is defining "how will this question help answer my research question?". As a result, the first thing to do is to understand the goal of the research in itself. There were three main goals throughout this survey, which were: to verify if the previous MQM scores and THL analysis were correct, to understand the habits of research and the preferences of the SEs, and to obtain new resources in order to improve both the ones available in the Language Guidelines and in the client instructions.

While we could infer much from the data collected, it was necessary to develop questions that would allow us to get more insight into the number of issues or even other struggles the SEs faced. Moreover, since one of the goals was to obtain more resources, we quickly understood that the strategy adopted should be of mixed methods typologies; namely, concurrent strategies, that is, "data collection using both quantitative and qualitative approaches simultaneously, for example when a questionnaire contains both closed and open-ended questions" (Saldanha & O'Brien, 2014, p. 201). However, since open-ended questions would also be in the survey, this would increase its completion time. Consequently, it was necessary to create a structured questionnaire that would only show participants questions relevant to their own experiences.

This structure was designed by mainly establishing once again the division observed between "Professional" and "Creative" clients. The participants would be asked about which clients they had worked with and get a set of questions depending on which group they were assigned to. Other factors that reduced (or increased) the number of questions were if the subject had worked with Lingo24 before, if their LP was a part of the European Union or not (to obtain more information about terminological banks for other languages and varieties), and when selecting that they did "advanced search" in a particular multiple-choice question they would get a follow-up question.

This resulted in a 40-question survey done with Typeform[20] as is customary for Unbabel, in which 19 questions were mandatory and were concerned with mostly background information or issues related to both types of text. The survey was composed of yes or no questions, multiple choice questions (with an option where participants could write their answers if they wanted to), Likert-type questions, and, of course, there were also open-ended questions. Throughout the survey, we avoided jargon, explained any abbreviations or acronyms, and the language used was as simple as possible to avoid misunderstandings because most of the respondents were not native speakers of English.

However, even after all these steps, the approximate time needed to complete this questionnaire was around 25 minutes when answering all questions. As a result, having in mind the question "(...) is this a reasonable demand of my target population's time?" (Saldanha & O'Brien, 2014, p. 154), the company decided that the participants should get a symbolic monetary incentive to complete the survey.

In terms of the selection of participants, these were the individuals whom we identified as the top active SEs (total number of words reviewed) during the previous two weeks to the release of this questionnaire. When more than five SEs were available per LP, only the top five were selected. However, whenever this was impossible, all the top SEs from these other LPs were chosen. This resulted in a total of 70 individuals across 18 different LPs (Table 3). In addition, around 16% (11) of these selected SEs were originally from Lingo24, while 84% (59) had been with Unbabel since the start.

Five SEs per LP		Less than five SEs per LP
Chinese (Simplified)	Polish	Chinese (Traditional) - 1 editor
Dutch	Portuguese (BR)	Hindi - 2 editors
French	Russian	Portuguese (EU) - 2 editors
German	Spanish	Thai - 2 editors
Japanese	Spanish (LatAm)	Turkish - 2 editors
Korean	Swedish	Vietnamese - 1 editor

Table 3 - Number of SEs selected per LP

After the participant screening, a communication was sent to the selected individuals detailing the purpose of this survey and internship and how they would receive monetary compensation for their participation. The conclusions drawn from the analysis of this survey data will be presented in **Section 4.3**.

[20] See https://www.typeform.com/ .

4. Results and discussion

In this chapter, we will present and discuss the results obtained through the employment of the methodological approaches detailed in the previous section. Firstly, in **Section 4.1**, we will examine the chosen MQM dataset, followed by the analysis of the THL dataset (**Section 4.2**). Subsequently, in **Section 4.3**, we will present and interpret the results from the Translation Research Skills Survey. Finally, we will describe in-depth the creation of the Research Skills Materials and the corresponding feedback (**Section 4.4**).

4.1. MQM Dataset Analysis

As stated in **Section 3.1**, one of the first steps taken to validate the theory that SM content was not meeting quality requirements was the analysis of the MQM scores across four different Lingo24 clients over the course of one month (October 9th to November 9th of 2022).

To maintain the anonymity of the clients in question, we will substitute their names with pseudonyms and will refer to them as "Client A," "Client B," "Client C" and "Client D." It is important to observe that only one of the three clients, namely, "Client D" has been categorized as possessing texts which correspond to the Creative Translation pipeline, whereas the other two are a part of the Professional Translation pipeline category. This differentiation is crucial since it influences the minimum average MQM threshold.

Translation Pipeline	**Professional Translation**			**Creative Translation**
Clients	**Client A**	**Client B**	**Client C**	**Client D**
Batches	95.04	99.62	78.49	91.44
	73.25	92.21	91.45	95.77
	90.10	94.34	-	98.87
Average	**84.37**	**94.72**	**79.20**	**96.59**

Table 4 – Shadow Mode MQM scores from four Lingo24 clients

As depicted in Table 4, it was possible to obtain information from a total of 11 batches, with three of them analyzed for each client (except Client C, with only two batches). Among the Professional Translation pipeline subcategory clients, only one batch (Client B - 99.62) out of eight achieved the minimum MQM score for any of the

three LP tiers. In addition, another batch meeting the minimum MQM value belonged to Client D, with a score of 98.87, which also exceeded the minimum score for all three LP tiers.

Nevertheless, the average MQM score for each client fell below the defined quality expectations threshold, with Client D being the closest to meeting the desired score. Notably, as mentioned above, Client C had one less batch analyzed, which could have biased the results, resulting in the lowest overall score.

In conclusion, this analysis corroborated the belief that the content reviews submitted by Lingo24 clients were not meeting the quality standards established by Unbabel. As can be observed, merely two batches are over the minimum MQM threshold defined for both the Professional and Creative Translation pipelines, and the average total MQM score for each client is consistently below the cutoff required for these assignments.

Nonetheless, while it was possible to prove the existence of issues through this analysis, it did not necessarily give sufficient insight into which factors may be troubling SEs and causing a decrease in the quality of the final product. For this purpose, it was necessary to further investigate the problem, and we did this through the analysis of multiple THL searches.

4.2. THL Dataset Analysis

In this study, we collected a dataset of nearly ten thousand (precisely 9952) THL search queries from a period of one month, spanning multiple batches, four Lingo24 clients, 18 LPs, and over 100 SEs.

For the sake of this research, these clients were also divided into two groups that correspond to the pipeline subcategories previously mentioned: “Professional,” meaning that these clients usually have more specialized texts which use more technical jargon and terminology, and “Creative,” namely, clients whose texts may contain characteristics such as markers of informality, wordplay or slogans, and generally require more attention to the tone of voice or brand identity than necessarily terminological research.

Due to the enormity of search data and the lack of context regarding each query, it was necessary to make some inferences by identifying similar patterns. Consequently, the following analysis can be considered an approximation for every THL search.

Clients/ Common Searches	Professional clients	Creative clients
Acronyms	✓	✓
Countries/Cities/…	-	✓
Job titles	-	✓
Keywords and adjectives	✓	✓
Products or brands	✓	✓
Terminology	✓	✓
Units of measurement	✓	-
Wordplay or slogans	-	✓

Table 5 - THL most common searches within the dataset

In general, Table 5 illustrates a collection of categories we determined that are searched by SEs across these two sorts of clients.

In terms of searches common to both categories of customers, one of the most prominent is the use of THL to search for acronyms (e.g., "ADAS" [advanced driver assistance systems] for Professional clients, and "NFT" [non-fungible token] for Creative clients). Nevertheless, we identified more occurrences among texts from Professional clients, with over 100 searches for acronyms (this number refers to searches where only the acronym is subjected to the query and excludes any longer segments [groups of words or phrases] where they might feature).

Another common search is for keywords or certain adjectives, such as "powered," "premium," "advanced," "intelligent," "unique," and "high-end," among many others. While they tend to be more prevalent in Creative texts, they also occur in Professional reviews, particularly when describing specific products or services.

As a result, the following category identified in both circumstances is the search for a specific product or brand. In the case of Professional texts, this is especially visible in the search for particular items, which may contain a unique code or structure that complicates the translation process. However, when it comes to brands, there appears to be an equal number of searches across both types of text. For confidentiality reasons, none of these queries can be divulged in this work.

Next, we find the most prevalent category of query in this dataset, which is terminology. These terminological searches cover various fields depending on the client

and the subject matter. For example, in terms of Professional clients, the emphasis is usually on electrical, engineering, or even automotive jargon (e.g., "adjustable [bolt, clamping, output, …]", "cable clamp," "cam locks," …). In contrast, when it comes to Creative material, the type of terms may vary depending on the theme of a particular text (e.g., medicine ["sirolimus-eluting balloon"], biology ["polyplastics"], …).

Nevertheless, due to the different nature of these clients and texts, analyzed queries show different search category patterns. For example, when it comes to Creative texts, we identified that SEs often search for countries, cities, or regions (e.g., Vietnam, Shanghai, Guangdong, …), job titles (e.g., "Chief Product Officer," "Execution Manager," "Chief Technology Officer", …), and wordplay or slogans (no examples due to client confidentiality). Simultaneously, in Professional texts, we can find a high number of searches for all types of units of measurement (such as Celsius, Fahrenheit, inches, centimeters, microhenry, watts, volts, and many others).

While this, of course, does not show us the full picture of the research done while reviewing a text, it served as the basis for the next step taken during this research: the creation of a "Translation Research Skills" questionnaire.

4.3. Translation Research Skills Survey findings

4.3.1. Participation

As was addressed in the previous section, the selected participants included 70 individuals across 18 LPs. The results showed that 64% of the intended audience completed the survey, resulting in 45 respondents (Figure 10).

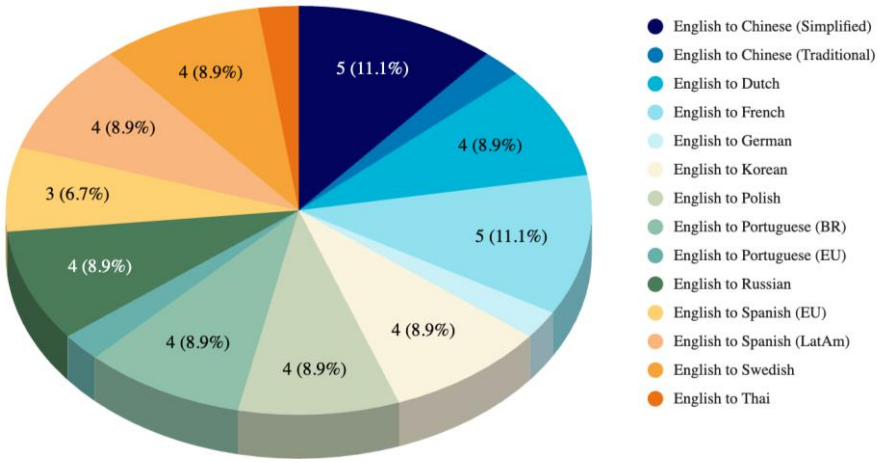


Figure 10 - Participation across LPs

The only languages that managed to garner full participation were Chinese (Simplified), Chinese (Traditional), and French, with a total of 11 SEs. Conversely, there were no participants from the following languages: Japanese, Hindi, Turkish, and Vietnamese (a total of 10 SEs), which reduced the number of languages surveyed to 14.

4.3.2. Demographics

The first three questions in this survey (see **Annex B**) were concerned with determining the demographic characteristics of the respondents, namely, the age range, gender, and level of education of each individual.

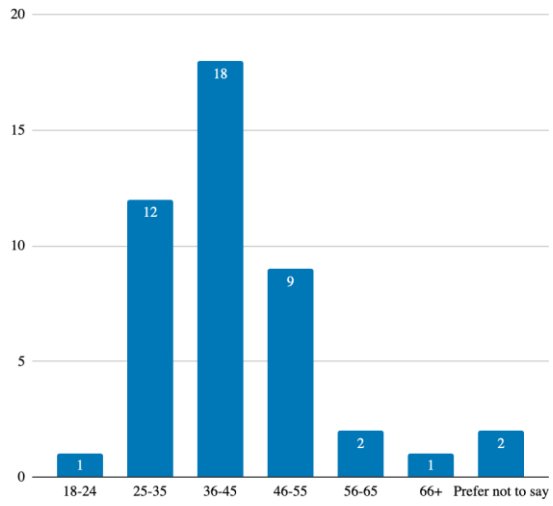


Figure 11 - Age range of participants

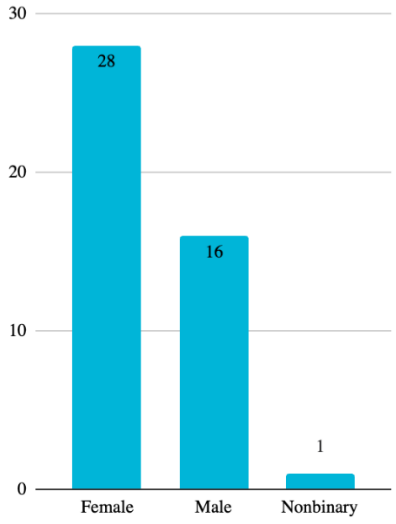


Figure 12 - Gender of participant

In terms of age range, most participants were between the ages of 25 to 35 (12 individuals - 26.7%) and 36 to 45 (18 individuals - 40%). Since there is only one participant between the ages of 18 and 24, the results indicate that the bulk of participants in this sample are young to middle-aged adults (Figure 11).

In terms of gender (Figure 12), over half of the participants in this sample identify as female (28 individuals - 62.2%), with the rest identifying as male (16 individuals - 35.6%) and nonbinary (one individual - 2.2%).

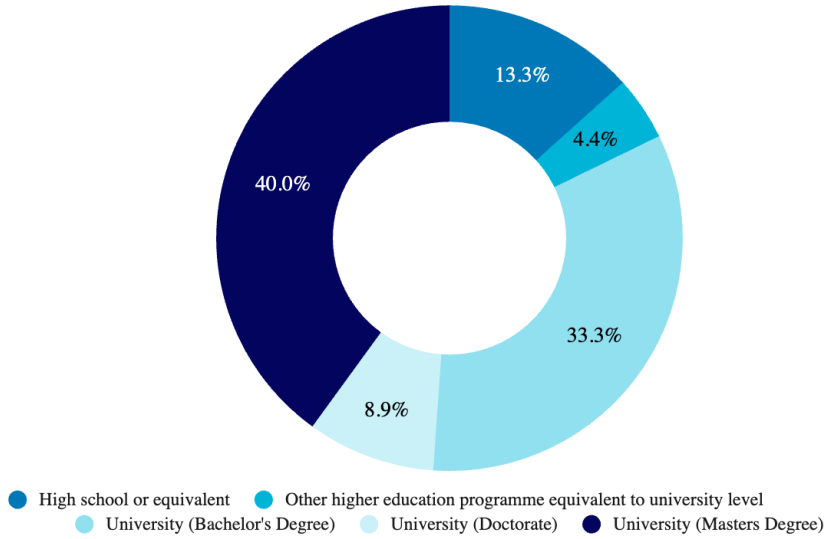


Figure 13 - Education level of participants

The information gathered about the educational background ("What is your level of education?" [**Annex B** – Q3])[21] of the 45 participants demonstrates a relatively high level of education. With 18 (40%) individuals possessing a master's degree, 15 (33.3%) holding a bachelor's degree, four (8.9%) having a doctorate, and the other two (4.5%) having finished "other higher education program equivalent," it is clear that a sizable majority of participants hold some type of university degree, as shown on Figure 13. Contrastingly, just six (13.3%) of the participants in this sample have completed high school or an equivalent program.

4.3.3. Main field of study and previous experience vs. preferences

After gathering the demographic-related information, one of the earliest questions the participants encountered was: "What is your favorite content to translate/review?" (**Annex B** – Q7). Since this was also the first open-ended question, it led to many different responses, both in content and length. Therefore, following Saldanha and O'Brien's (2014, p. 189) reasoning that "categories are not predetermined but derived from the data," a thematic analysis was chosen to classify the various statements. These authors describe this method as "a process of working with raw data to identify and interpret key ideas or themes" (Saldanha & O'Brien, 2014, pp. 189-190), meaning that the researcher

[21] For the sake of clarity, every following question will be presented in a similar format. However, due to question length, only the most crucial parts will be shown. For the full text, please refer to the corresponding question number in **Annex B**.

should develop an index of themes (or codes) through the analysis of all the information. While this coding template can be created prior to data analysis, it should be flexible enough to avoid incorrect inferences and biases. Moreover, due to the multifaceted and somewhat subjective nature of open-ended questions, more than one theme can be attributed to each response.

In this particular case, the coding template was developed after a first examination of the data, which led to the creation of four parent codes and six daughter codes (Figure 14). These themes encompass the type of content chosen; the level of familiarity with that type of text; if a specific client is mentioned; and if there is some exception to the content they enjoy reviewing.

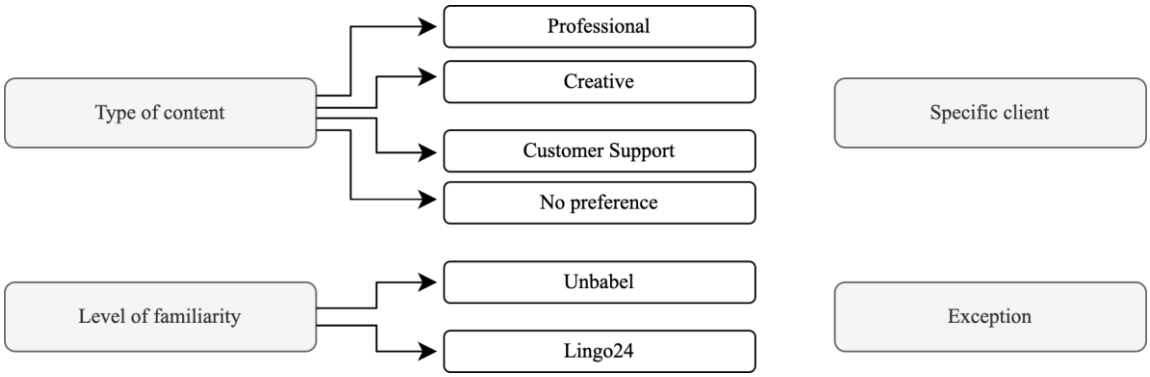


Figure 14 - Coding template

Regarding the type of content, this parent category was divided into four daughter codes:

- Professional
    - Areas where a great degree of terminology is used and in-depth knowledge may be required (i.e., technical, electrical, automotive, technological, engineering, medicine, law, …).
    - May lack context (e.g., lists of very specific electrical or automotive products);
    - The texts are more objective and are written in a formal or neutral tone.
    - Publishable content which may have an impact and real repercussions.
- Creative
    - Areas where there is not necessarily a need for in-depth knowledge (i.e., press releases, various articles, travel and marketing content, movie

reviews, ...). However, some press releases may contain specific subject matter terminology (e.g., crop names, wind turbine elements, …);
    - Usually, longer texts contain more context.
    - It may be informal depending on the style of the client and use less specialized terms.
    - Require more "creativity" in the sense of adapting the style of the client to another language and culture.
    - Publishable content, usually on the website of the company, may have an impact on its image.
- Customer Support
    - Depending on the subject, it may require specialized terms, but that is not always true.
    - Usually follows a particular template that facilitates the review.
    - It may be shorter and not necessarily publishable, which means it will not impact the brand (e.g., customer support emails).
- No Preference
    - When the participant explicitly states that they have no preference (including saying they enjoy reviewing any kind of content).

The level of familiarity consists of two subcategories: "Unbabel" and "Lingo24". This code addresses whether the participant has only worked with Unbabel or was also a SE for Lingo24.

Finally, the last two themes are "Specific Client" and "Exception." As the name suggests, "Specific Client" is used whenever a participant mentions a particular client (in a negative or positive light). Regarding "Exception," this code refers to any negative commentary added by the SE, such as a type of text they dislike reviewing.

After categorizing and dividing all the responses into this index of codes, it was possible to discover some patterns and develop a few theories. The first hypothesis was that there is a correspondence between the main field of study of the participant and the content they enjoy translating. To validate this theory, it was first important to observe the data concerning their main field of study ("What is your main field of study?" [**Annex B** – Q4]) and then compare it to the different subcategories of the type of content chosen by each individual.

Through this analysis, it was possible to identify a wide array of subject matters, namely, sixteen different disciplines, ranging from Sciences to Arts (Figure 15). The most common fields of study among the participants were "Business and management" and "Computer science," both with a percentage of 13.3% (6 participants each), followed by "Modern languages & linguistics" and "Natural sciences (Physics, Chemistry, Biology, and Earth Sciences)," tied with a percentage of 8.9% (4 participants each). Within the scope of our research, it is relevant to point out that only three (6.7%) participants hold a Translation Studies degree.

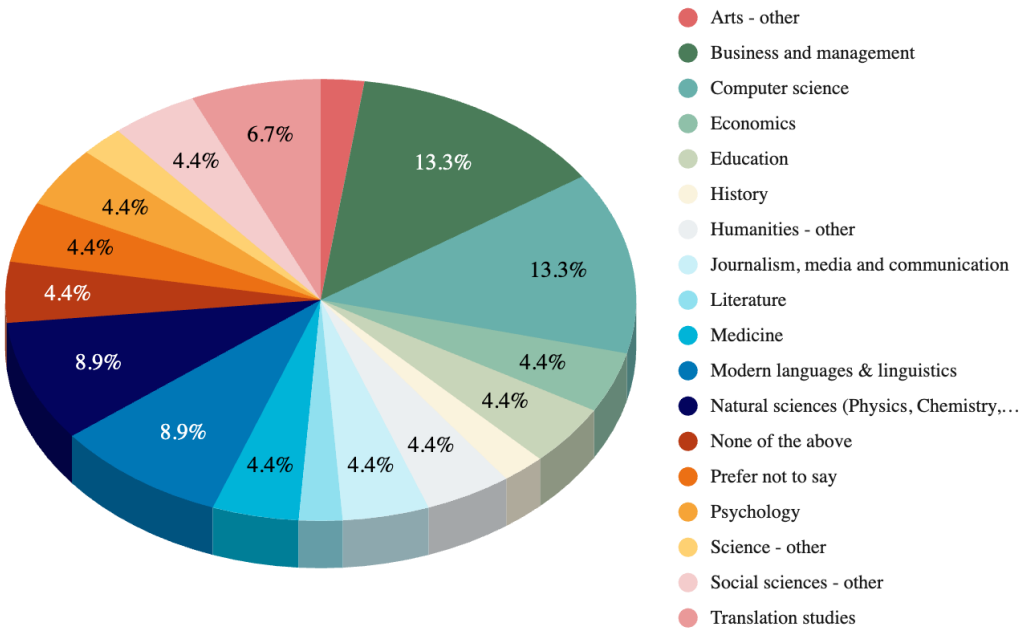


Figure 15 - Fields of study of the participants

In relation to the type of content, out of a total of 45 participants, 35.6% (16 individuals) mentioned they enjoy translating/reviewing "Professional" texts, while 71.1% (32 individuals) prefer "Creative" texts (Figure 16). Moreover, 11.1% (5 individuals) mentioned they enjoy reviewing Customer Support content, while 13.3% (6 individuals) stated they have no preference. Additionally, since "Creative" texts are the predominant interest, we can find 13 responses where individuals describe an interest in both "Professional" and "Creative" texts. In contrast, only one individual mentions "Customer Support" and "Creative" content simultaneously.

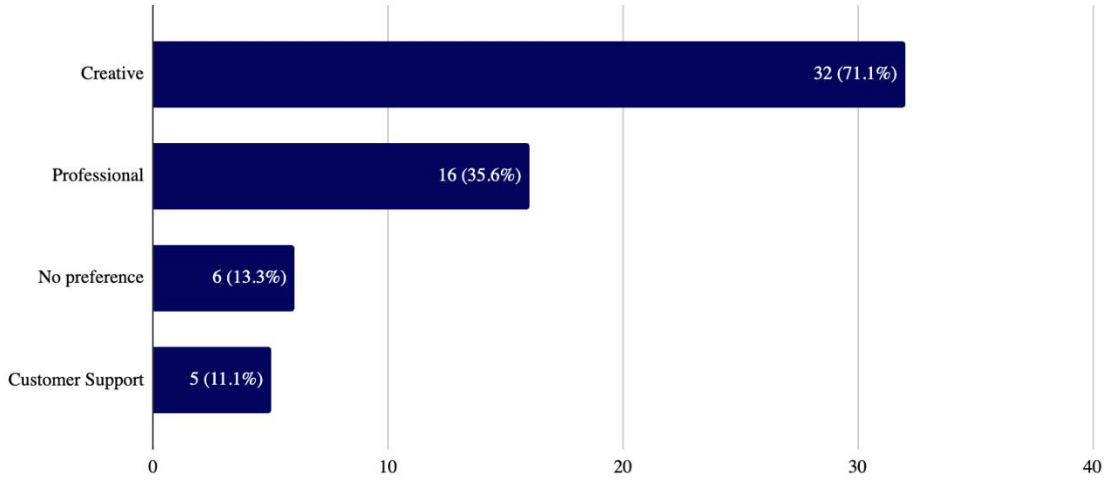


Figure 16 - Types of content the participants are interested in

Subsequently, we compared the type of content categories to the main fields of study of each SE. It is important to note that the six individuals (13.3%) who did not identify a specific kind of content were not included in this since their preferences may include types of text outside of the research topics of this study.

As previously mentioned, 16 of these participants were attributed the "Professional" type of content theme based on their responses across nine distinct disciplines (excluding the "None of the above" and "Prefer not to say" respondents). Within these SEs (Figure 17), seven (43.8%) had science-oriented degrees (Natural Sciences, Computer Science, and other Sciences), three (18.8%) had business/economics-related degrees (Business and management and Economics), other three (18.8%) studied humanities-related subjects (Humanities, Social Sciences, and Translation Studies). At the same time, the final three did not provide any concrete information (Prefer not to say or None of the above).

However, just three (18.8%) of these 16 individuals only mentioned "Professional" content. Notwithstanding, two of those three have science-oriented degrees (namely, "Natural Sciences" and "Science - other"). In conclusion, when it comes to "Professional" texts, this seems to be preferred by people with scientific-related backgrounds.

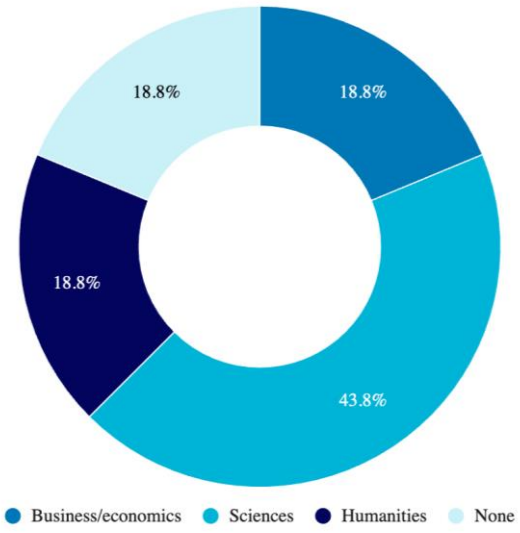


Figure 17 - Aggregated fields of study (Professional)

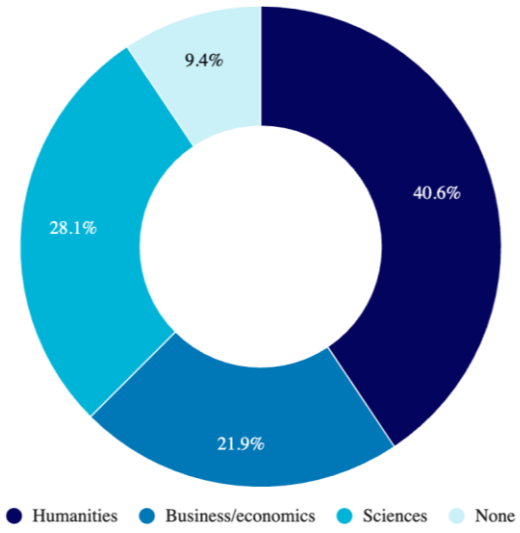


Figure 18 - Aggregated fields of study (Creative)

In terms of a preference for "Creative" texts, since there are more individuals (a total of 32), there is also a bigger range of fields of study, that is, 14 different disciplines.

However, if we aggregate the subjects as before, there seems to be a prevalence of individuals whose main field of study is humanities-oriented (Figure 18). In more detail, 13 (40.6%) participants had humanities-related subjects (Arts, Education, Humanities, Journalism media and communication, Literature, Modern languages & linguistics, Psychology, Social Sciences, and Translation Studies), nine (28.1%) studied science-oriented subjects (Computer Science, Medicine and Natural Sciences), seven (21.9%) had business/economics-related degrees (Business and management and Economics), and, finally, the last three (9.4%) did not provide any information ("None of the above" or "Prefer not to say").

Although the correspondence between the preference for "Professional" texts and science-focused areas of study is stronger (3.2% higher) than the one found in this case (Figure 18), there still exists a discernible majority of humanities-related degree holders who share a preference for "Creative" texts. As a result, it is once again evident that there

is a slight correspondence between the primary topic of study of a particular individual and the kind of content they enjoy reviewing or translating.

During the analysis of this data, we formulated a second hypothesis that there is also a relation between favorite content to translate/review and the previous experience of each individual, namely, "Unbabel" (customer support-oriented content) or "Lingo24" (multiple types of text such as the ones defined as "Professional" and "Creative" in this particular study). To verify this hypothesis, it was necessary to compare the "type of content" categories again, but in this case, to the data obtained concerning their previous work with Unbabel and/or Lingo24.

In a previous question, we invited participants to respond to the following: "What best describes your situation?" (**Annex B** – Q5).

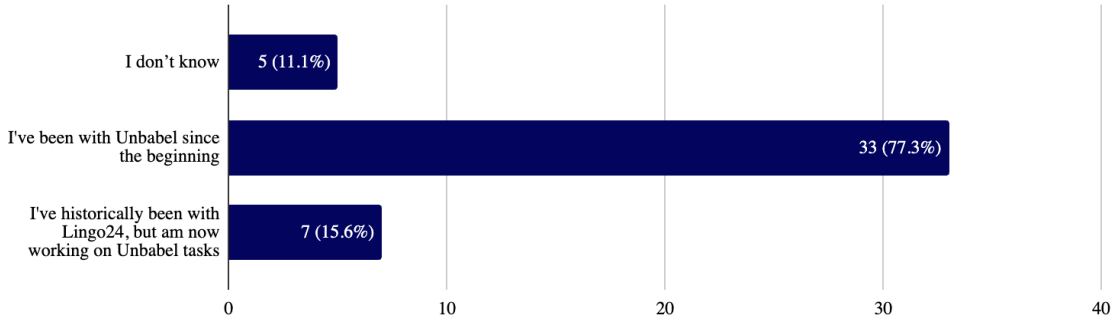


Figure 19 - Employment history (Unbabel and/or Lingo24)

As depicted in Figure 19, most participants (77.3% - 33 individuals) reported that they had been working for Unbabel since the beginning, whereas a smaller percentage (15.6% - 7 individuals) had previously worked for Lingo24. Notably, 11.1% (5 individuals) claimed to be unaware of their employment history. However, after examination of previous data obtained from the selection of participants (see **Section 3.3**), it was possible to determine that those five people had only ever worked for Unbabel, thereby raising the percentage of Unbabel-only SEs in this sample to 84.4% or a total of 38 individuals. With this data, it was possible to attribute them the "level of familiarity" categories, namely, "Unbabel" and "Lingo24".

Amongst the seven people who had previously worked for Lingo24, three individuals (42.9%) mentioned they enjoyed translating/reviewing both "Professional" and "Creative" texts. In contrast, the same number of participants (42.9%) indicated a preference for translating/reviewing only "Professional" texts, with simply one individual (14.3%) singling out "Creative" texts as their favorite content. It is also worth noting that

none of these seven participants expressed enjoyment in translating or reviewing customer support-oriented content, and all of them mentioned liking at least one of the other two types of text (“Professional” or “Creative”).

Contrastingly, among the 38 individuals who had exclusively worked with Unbabel, only five (13.2%) expressed a preference for translating and reviewing customer-oriented content (with one mentioning “Creative” content as well). Meanwhile, of these 38 members of the Unbabel community, 28 (73.7%) reported enjoying reviewing/translating “Creative" content. Additionally, 10 (26.3%) participants also expressed interest in translating and reviewing “Professional” texts, with only six (15.8%) participants indicating no preference for any particular type of content. Interestingly, all 10 participants who said they enjoyed “Professional” content also said the same about “Creative” content.

As a result, it is possible to conclude that the preference for certain types of content is not only dependent on the level of familiarity of each SE but also the complexity of each text. As can be observed, participants from Lingo24 clearly showed a preference for the content they were already familiar with (which often requires more in-depth knowledge). In contrast, Unbabel-only participants did not only show interest in customer support-oriented content, which was the content they were most familiar with, but the majority also claimed to enjoy translating/reviewing mostly content classified as “Creative.” For example, several responses cited a preference for working on movie reviews, travel-related materials, or marketing content. It is also important to point out that five Unbabel-only respondents who mentioned they enjoy reviewing “Professional” also hold degrees in science-oriented subjects. Other two of those ten did not disclose their educational background (“Prefer not to say” or “None of the above”). Consequently, it seems that as long as the new content does not require an extreme level of in-depth knowledge, most SEs seem receptive and interested in reviewing it.

This assertion is further supported by the “Specific client” and “Exception” categories. That is, among the five responses under the “Specific client” category, four participants named clients which fall under the “Creative” category, while only one individual expressed a preference for tasks from a particular “Professional” client. Furthermore, another one of these five participants cited the exact previous “Professional” client as one they did not find enjoyable to work with, thereby also falling under the “Exception” category. Additionally, two other respondents (out of three under the

"Exception" category) expressed a dislike for reviewing texts containing highly technical terminology. Incidentally, these two individuals are Unbabel-only personnel.

4.3.4. Translation History Lookup and other common issues

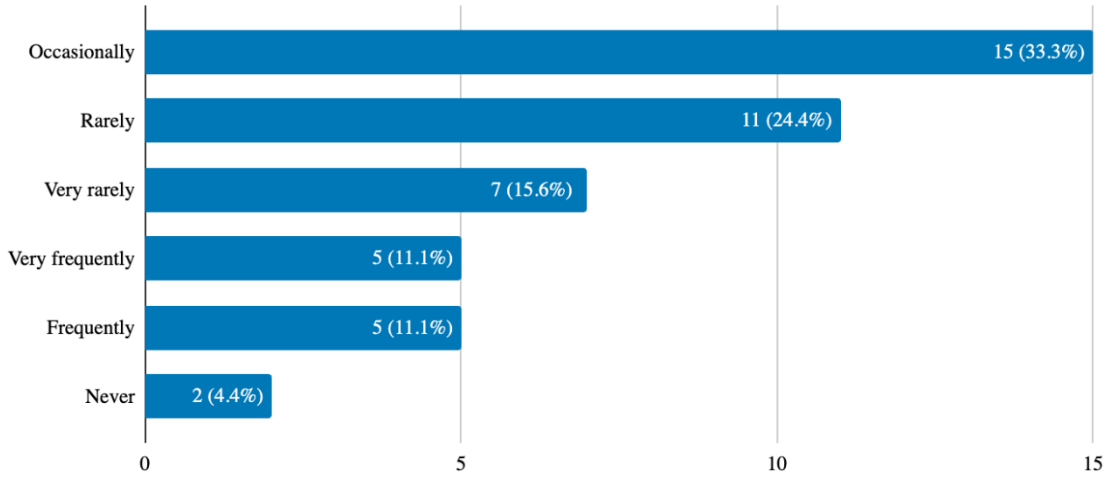


Figure 20 - Use of Translation History Lookup

As aforementioned in **Section 4.2**, the analysis of THL searches was one of the core elements for identifying potential issues and challenges SEs may encounter during the revision process. However, as has been discussed before, it is crucial to bear in mind that this is just one of the resources these individuals may utilize. Nevertheless, it was important to understand how often and for what purposes SEs use this tool and if these correspond or differ from the preliminary analysis of THL searches.

As a result, the first question ("How often do you use the Translation History Lookup in Polyglot (Editing Interface)?" [**Annex B** – Q12]) in relation to THL was concerned with the frequency with which individuals utilize this tool. This Likert scale question, as shown in Figure 20, presented six response options ranging from "very rarely" to "very frequently" and asked SEs to identify how often they use this feature. The most common response was "occasionally," which was chosen by 33.3% (15 individuals) of respondents, followed by "rarely," which amounted to 24.4% of the total responses (11 individuals), and then "very rarely," by 15.5% of participants (7 individuals). This suggests that while most individuals do not use this feature as frequently as desired, certain SEs still find it useful enough to use occasionally.

Furthermore, only 4.4% of participants selected the "never" option (2 individuals), meaning that most respondents had at least tried using the THL feature once. Additionally, 11.1% of the individuals chose the options "frequently" or "very frequently" (a total of 10 individuals), which indicates that some SEs find the THL tool helpful and use it regularly.

In conclusion, it is hard to pinpoint the precise reasons for the underuse of this tool, and several possible factors may be involved. These can include restrictions on how well it performs in specific situations (e.g., different types of texts across distinct LPs), issues related to its usability (e.g., when a SE uses this feature on a mobile device unpredictable technical difficulty may arise), or even simply personal preferences that favor other tools for research.

In a follow-up question ("You usually use the Translation History Lookup to search for…" [**Annex B** – Q13]), we asked SEs about the purposes for which they utilize this tool. This was a multiple-choice question with alternatives that included several of the potential issues listed in **Section 4.2** but also enabled the participant to enter their response (as every following multiple-choice question present in this questionnaire).

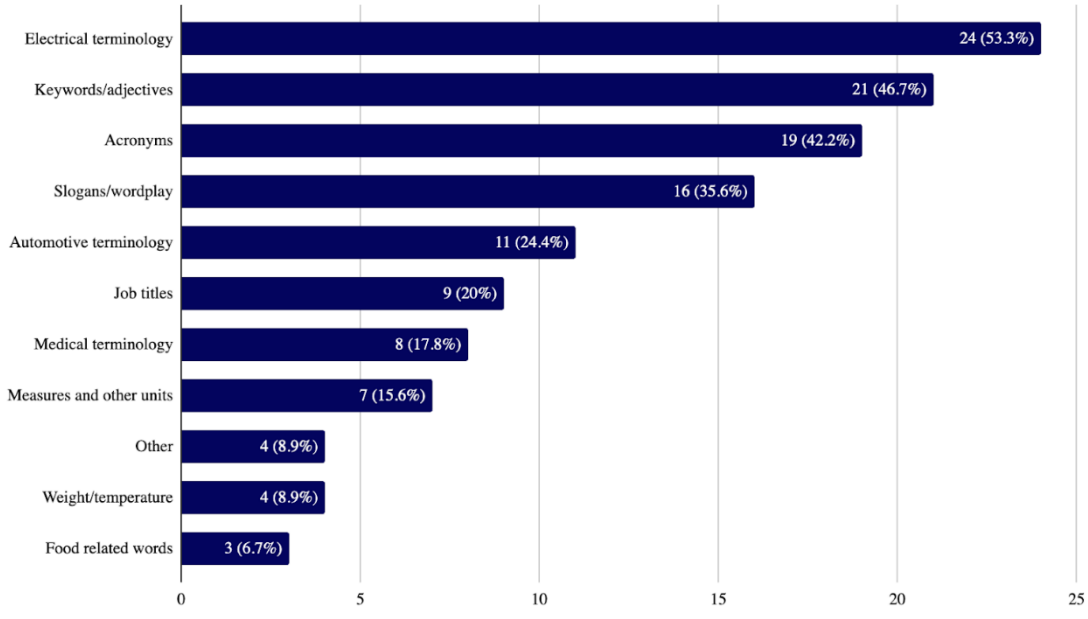


Figure 21 - What SEs use Translation History Lookup for

As displayed in Figure 21, and as was identified in the examination of the THL searches dataset, terminology appears to be one of the most searched-for subjects. However, in order to gain more in-depth insights, the category "terminology" was

separated into various subcategories. Nevertheless, even subdivided into subcategories, terminology scores the highest, with “electrical terminology” being selected by 24 (53.3%) individuals. Moreover, if we combine all the present types of terminology (namely, “automotive,” “electrical,” and “medical”), the score rises to a total of 43 responses. The following most popular topic searched with this tool is “keywords/adjectives" (46.7% - 21 respondents), once again confirming the previous findings. Then, we have “acronyms” (42.2% - 19 respondents), which we also identified as one of the most frequent searches. Lastly, “wordplay and slogans" is another query with more than 15 replies (35.6% - 16 responses). Furthermore, four of the respondents preferred to enter their own responses (“Other”), with two of those responding: “Anything” and “Everything.” Interestingly, in the preceding question, those two same respondents claimed to use the THL tool very frequently.

In conclusion, despite some variations, it was possible to find a correspondence between the initial analysis of searches and the THL usage purposes data provided by the SEs who participated in this survey.

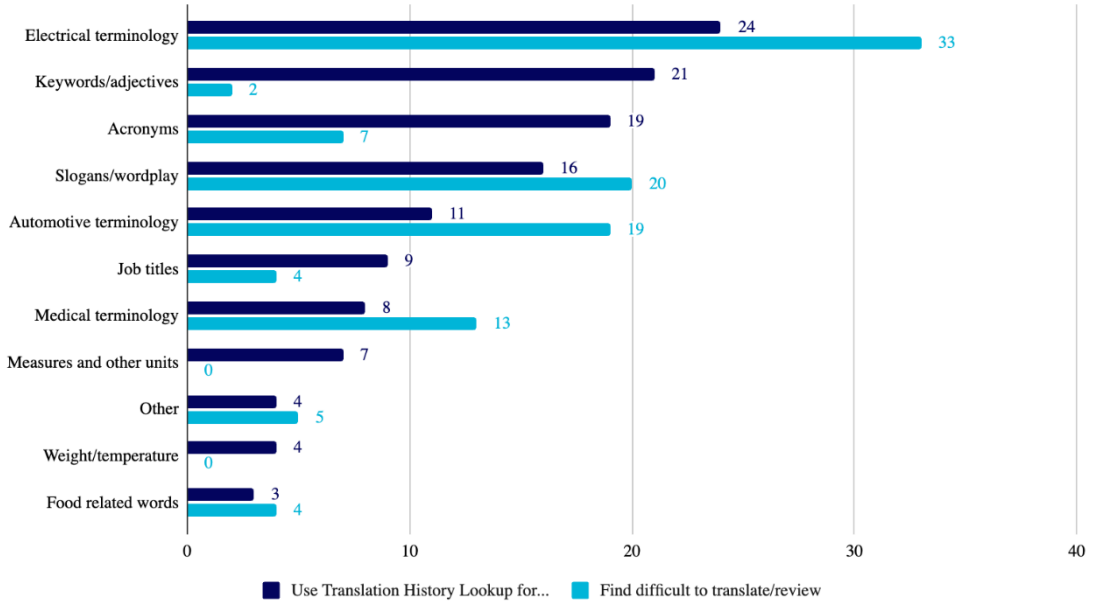


Figure 22 - Comparison between THL searches and what SEs find difficult to translate/review

In the same vein, one of the last questions (“Considering the work you’ve completed so far, what were the hardest things for you to translate/review?” [**Annex B** – Q38]) of this survey was concerned with understanding what SEs felt were the hardest things to translate/review. Similarly to the THL-related questions, this was also a multiple-choice question featuring the same options as the previous question analyzed.

As it is possible to observe in the graph above (Figure 22), terminology is still the subject these SEs struggle the most with. In fact, if we group all the different terminology options, we will get 65 answers. Additionally, the second most frequent issue seems to be related to slogans or wordplay, with 20 respondents. Therefore, it was possible to conclude that the two main issues are related to two different types of clients, demonstrating that they both entail extensive research that may require different approaches.

In addition, comparing this data to the previous ("What do you use THL for…" [**Annex B** – Q13]) data, it is possible to see that there is always a significant difference between what they seem to use THL for and what they identify as being difficult to review. For example, 21 respondents claim to use THL to search for keywords and adjectives, but only two find these difficult to translate. Opposingly, even though it is still the subject participants most use THL for, it is possible to see that, for example, when it comes to "electrical terminology," out of 45 participants, 33 find it challenging to translate. However, only 24 actually used THL when doing this research. In conclusion, it is possible to surmise that THL does better in some specific searches than others, which may reflect why these SEs find some subjects so hard to translate/review since they must do external research.

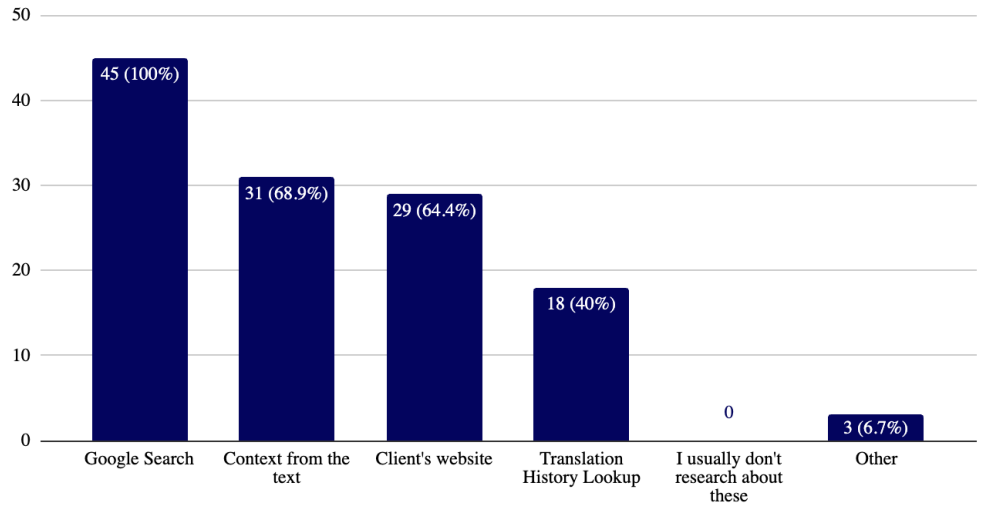


Figure 23 - How SEs verify what acronyms/initialisms stand for

The next multiple-choice question ("How do you verify what an acronym or initialism stands for?" [**Annex B** – Q14]) that every survey participant encountered inquired them about "acronyms" and "initialisms," and how they verified what they stand for (Figure 23). Overall, "Google Search" was the most popular response (with all 45 participants selecting it), followed by "Context from the text" (68.9% - 31 individuals),

and finally, the client's website, with 29 (64.4%) respondents. Although 19 respondents chose THL in the previous question, only 18 selected it again in this one. Moreover, the three (6.7%) individuals who opted to type out the answer referenced alternative search engines (namely, Yahoo and Baidu[22]). Therefore, it is reasonable to infer that the tool most used by these SEs to search for acronyms is Google Search.

The final qualitative question (also multiple-choice) ("How do you verify the validity and accuracy of a term?" [**Annex B** – Q17]) that every participant answered was in relation to one of the most prevalent searches identified in the first analysis of THL searches and later supported by data supplied by SEs themselves, that is, terminology (Figure 24). When asked how they verify the validity and accuracy of a term, 39 (86.7%) out of 45 respondents selected "Advanced Google Search," followed by "Client's website" (75.6% - 34 responses), "Trusted linguistic sources" (64.4% - 29 responses), and lastly "Translation History Lookup," with 27 (60%) responses. In conclusion, the values were evenly spread overall, with Google Search once again being the top choice. In addition, two of the six (13.3%) individuals who decided to input their responses also mentioned Google Search and another search engine (Naver). Finally, two other SEs responded with "My 15 years' experience" and "My own knowledge :P."

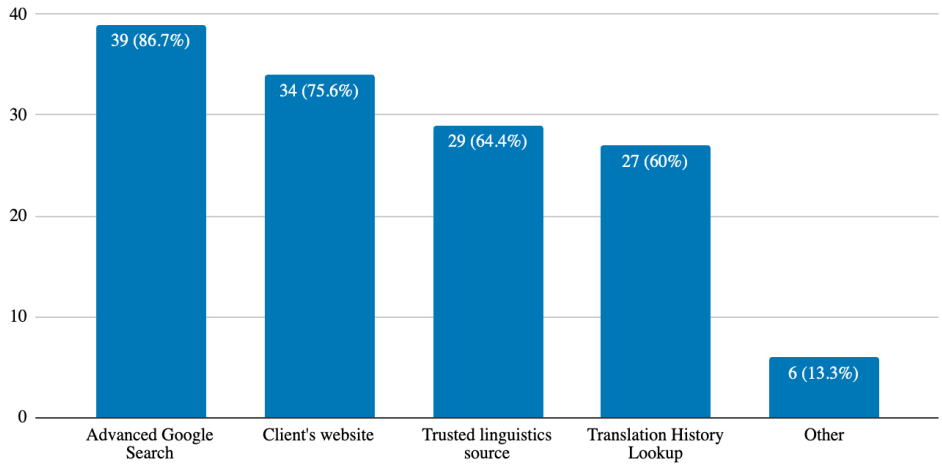


Figure 24 - How SEs verify the validity and accuracy of a term

[22] Baidu is a Chinese search engine (see: https://www.baidu.com/).

4.3.5. Professional and Creative texts

In an initial question ("While working at Lingo24/Unbabel, you've probably translated/reviewed texts for one (or more!) clients such as…" [**Annex B** – Q8]), we asked participants which Lingo24 clients they had previously worked with. This question served as the basis for the division of the questionnaire into two sections, one pertaining to "Professional" clients and the other to "Creative" clients. Since this was a multiple-selection question, it resulted in 42 individuals selecting Professional and 41 selecting Creative clients. Due to the fact that a significant number of participants worked with both types of clients, it is possible to find a large number of replies in both instances, which we will further examine in this section.

As pointed out in **Section 3.3**, we created this separation as a way to reduce the size of the questionnaire and ensure a higher number of participants. As a result, even though we found some categories (such as "products and brands") on both types in the prior THL search query analysis, it was necessary to attribute them to the type of text in which they were more prevalent. Therefore, as detailed in **Annex C**, the individuals in the "Professional" texts category encountered seven mandatory and three answer-dependent questions. In contrast, the individuals who were a part of the "Creative" texts category encountered one less mandatory question and two answer-dependent questions.

Furthermore, since the goal of this questionnaire was to acquire insight not only into the research habits and challenges of SEs but also to understand a few of their translation processes and obtain more resources, some of the questions (namely, the ones in red in **Annex C**) are not as pertinent to this particular analysis. However, we used those internally to improve the number and quality of resources.

4.3.5.1. Professional Texts

In terms of Professional texts, the first qualitative multiple-choice question ("How do you deal with uncommon and difficult IT, electrical, automotive related terms, for which there may already be a translation somewhere?" [**Annex B** – Q20]) addressed a situation in which there is no evident or standardized translation for a term (in the particular fields of IT, electrical or automotive engineering, …), with the goal of understanding how SEs deal with this terminology (Figure 25). The top response, with a wide margin, was that they would pick an existing translation after doing an advanced search (with a total of 22 responses, that is, 52.4%). The remaining choices received a

significantly smaller number of responses: “Keep the original followed by an explanation between parentheses” (16.7% - 7 responses), “I skip the task when there are too many of these” (14.3% - 6 responses), “Create your own translation by analyzing other similar (and already translated) terms” (11.9% - 5 responses), and two respondents (4.8%) stated that they would not translate these terms.

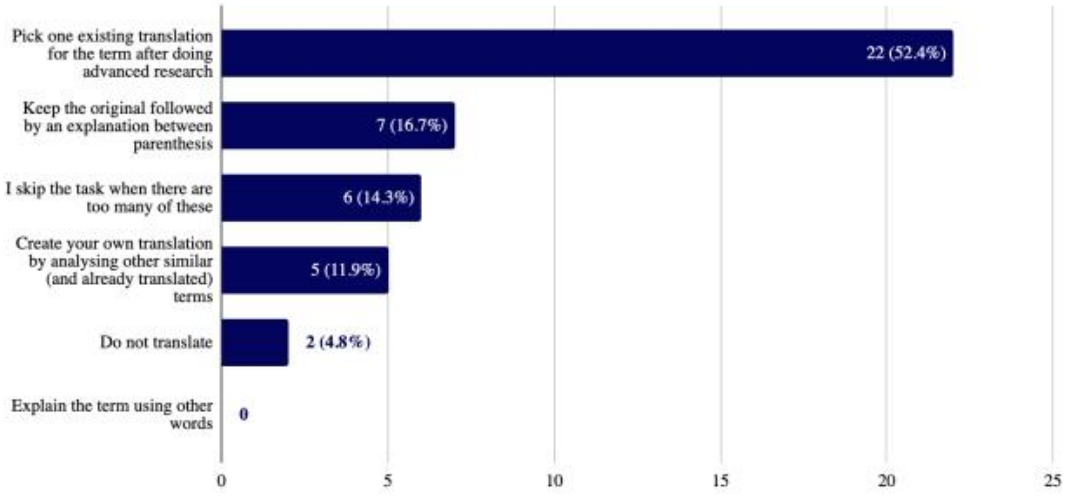


Figure 25 - How SEs deal with "Professional" terminology

Following this initial question, individuals who chose the “Advanced Search” option received an additional question requesting further information about their process (“How do you verify if the existing translation you’ve chosen is the correct one in your target language?” [**Annex B** – Q21]). That is, we asked them how they ensured that the translation they pick is the correct one in their target language. Since this was an open-ended question, it led to diverse answers. Nevertheless, after analyzing them and other related questions, another index of themes was created. Since these particular questions were always concerned with how SEs did research, it was possible to identify some patterns in the replies, namely, the use of “Google Search,” “Other search engines and/or advanced search,” the “Client’s website,” and “Other” (that is, various distinct approaches). We will use this coding template in the additional questions concerning the advanced search process.

In this case (Figure 26), it was evident that five (22.7%) of the total 22 responses expressly cited Google Search, while another six (27.3%) mentioned different search engines or simply advanced search. Furthermore, three (13.6%) of the 22 individuals who participated mentioned they compare the term to the ones present on the website of the client, and if they cannot locate the precise term, they will substitute it for something

lexically similar. Lastly, the remaining respondents (36.4%) suggested a variety of approaches, including examining various articles, utilizing paper dictionaries, or simply "knowing" if the term they have chosen is correct.

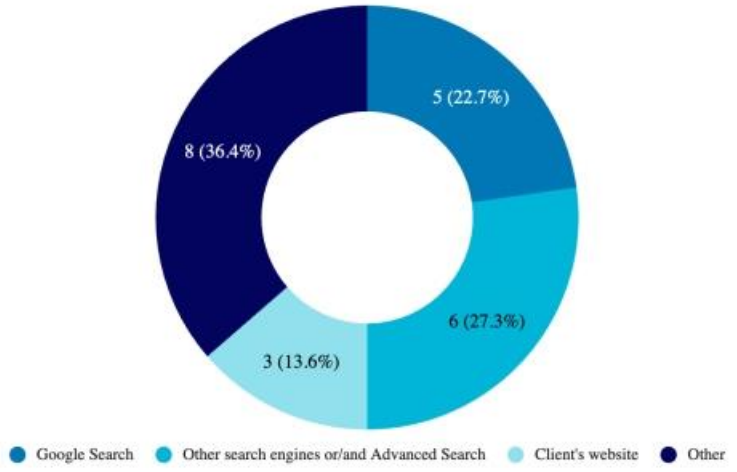


Figure 26 - Methods used to do "Professional"-related terminology advanced search

The following question ("How do you find out if the word(s) refer to a product or a brand?" [**Annex B** – Q22]) attempted to address another of the previously discovered most searched categories among technical texts, in this case, certain products or brands, and how SEs detect that a given word corresponds to the name of a product or brand.

According to the results seen in Figure 27, in terms of unknown product and brand names, "Client's website" was the most often selected option for this question, chosen by 33 of 42 participants (78.6%), followed by "Context from the text" with 30 responses (71.4%), "Google Search" with 29 responses (69%), and lastly, "Translation History Lookup" with 20 responses (47.6%). Additionally, two individuals (4.8%) opted to write their own answers, one mentioning another search engine (Baidu) and the other their professional experience. This demonstrates how SEs are focused on delivering in-brand translations, as it is imperative for these kinds of text.

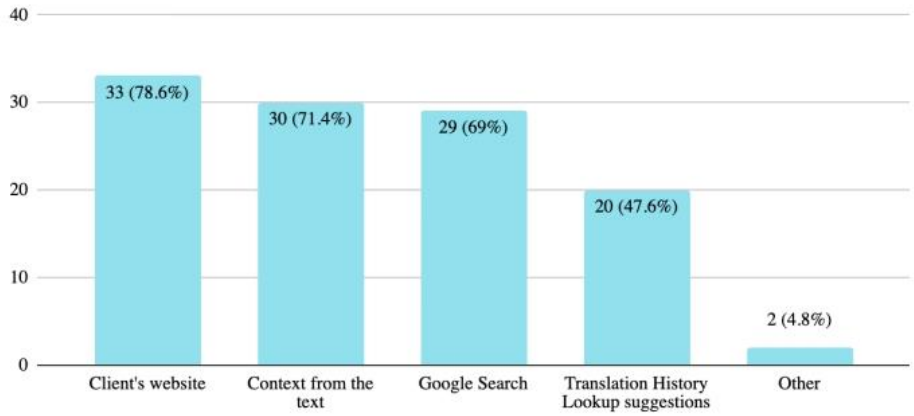


Figure 27 - How SEs deal with products/brands names

The following questions were connected queries that addressed how SEs would deal with specifically unknown products and brand names.

In terms of product names ("How do you deal with unknown product names?" [**Annex B** – Q23]), respondents said they do research outside of the Polyglot interface (81% - 34 respondents), followed by 27 individuals (64.3%) mentioning they use THL search, and lastly, some saying they leave them in English by default (33.3% - 14 respondents). Moreover, only one of the participants (2.4%) preferred to submit their answer, where they mentioned searching for context inside the website of the client (Figure 28).

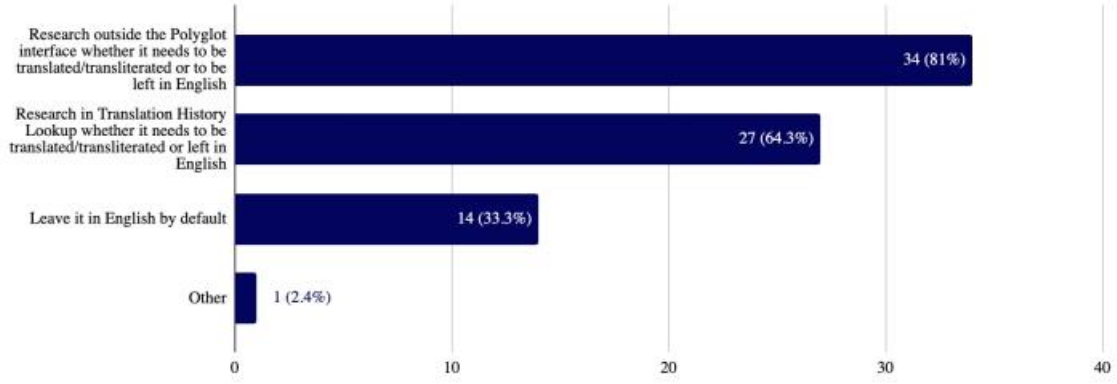


Figure 28 - What SEs do when faced with unknown product names

Regarding brand names ("How do you deal with unknown brand names?" [**Annex B** – Q25]), we gave individuals the same options present in the former question. This resulted in the majority stating that they either would leave it in English by default (52.4% - 22 respondents) or search for them using THL (also 52.4% - 22 responses), and finally,

researching outside of the Polyglot interface appears in third place (38.1% - 16 respondents) (Figure 29).

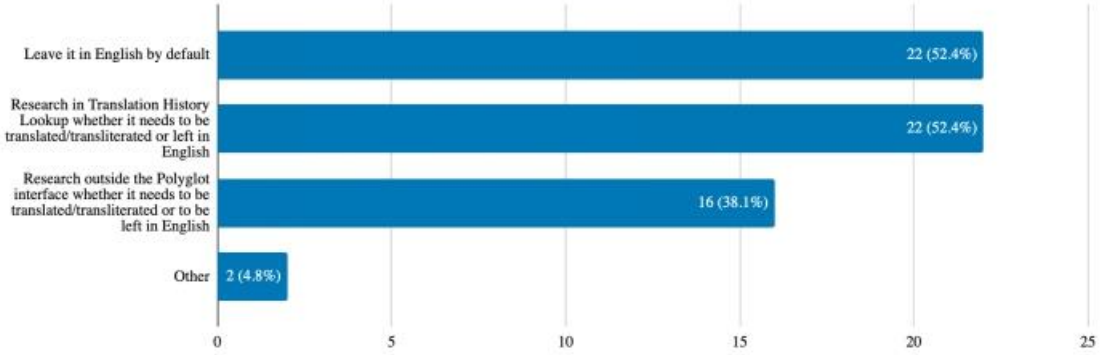


Figure 29 - What SEs do when faced with unknown brand names

Only two of these individuals (4.8%) elected to input their answers, which mentioned understanding it from the context provided by the text alone or their years of experience.

While we can sometimes find this type of content in the client-given glossaries (integrated into Polyglot), it is vital that SEs still research to double-check or even point out something they think should be in a glossary.

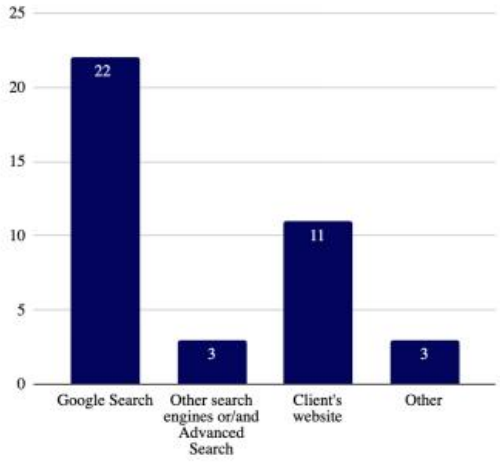


Figure 30 - Methods used for product-related advanced search

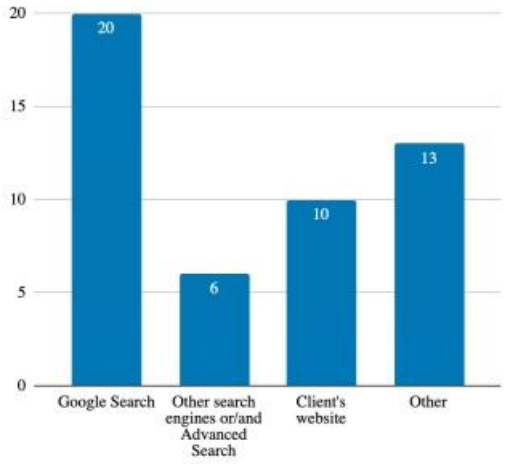


Figure 31 - Methods used for brand-related advanced search

Similarly to prior questions, when individuals said they conducted external research, they were assigned a follow-up question about this process. In terms of product names ("You've said you research unknown product names outside the Polyglot interface. How do you do this research?" [**Annex B** – Q24]), 22 (64.7%) respondents mention using Google Search, while three (8.8%) reference other search engines or advanced search,

leading up to a total of 25 responses. Furthermore, 11 (32.4%) of these 34 respondents also indicate the client's website. Finally, just three (8.8%) mentioned other methods (Figure 30).

Regarding brand names ("You've said you research unknown brand names outside the Polyglot interface. How do you do this research?" [**Annex B** – Q26]), due to an issue in the survey structure, even though some respondents did not mention doing research outside of Polyglot, they were still redirected to this question, which may skew some of the findings. Nonetheless, as we can see in Figure 31, 20 (47.6%) of the 42 respondents acknowledged mainly using Google Search, with the other six (14.3%) mentioning other search engines or advanced search. Moreover, ten (23.8%) reported searching on the client's website, and 13 (31%) reported various other methods.

4.3.5.2. Creative Texts

The first multiple-choice question ("How do you deal with uncommon and difficult medical terms, for which there may already be a translation somewhere?" [**Annex B** – Q30]) aimed at "Creative" texts asked SEs how they deal with medical terminology (e.g., press releases), which is not as well known or standardized as others (Figure 32). The top choice, with 21 responses out of 41 (51.2%), was that they picked an existing translation for the term after conducting extensive research. This was followed by individuals mentioning they skipped the task when there were too many medical terms, and the third top answer was that they keep the original and then provide an explanation within parentheses, both tied with six responses each (14.6% each). Next, the most selected option is the creation of a new translation by analyzing other comparable (and previously translated) terms (12.2% - 5 responses), followed by individuals stating that they do not translate these terms (4.9% - 2 responses). Finally, only one respondent (2.4%) indicated that they describe the term using other words.

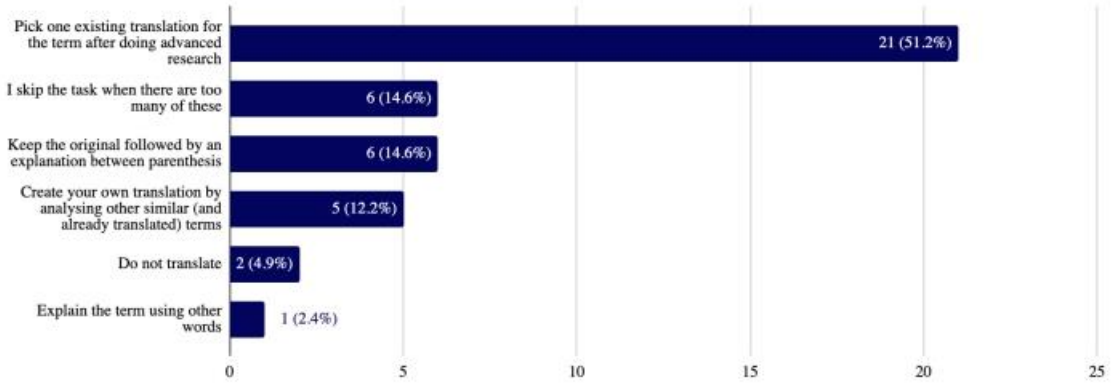


Figure 32 - How SEs deal with "Creative" terminology

Once again, the 22 respondents who selected “Advanced Search” in the previous question were redirected to a follow-up question (“How do you verify if the existing translation you’ve chosen is the correct one in your target language?” [**Annex B** – Q31]), to which six (27.3%) of them responded that they used Google Search (Figure 33). In addition, the other nine (40.9%) respondents mentioned using other search engines (resulting in 15 responses). Lastly, six (27.3%) individuals cited additional methods, with only one (4.5%) saying they would search on the client’s website.

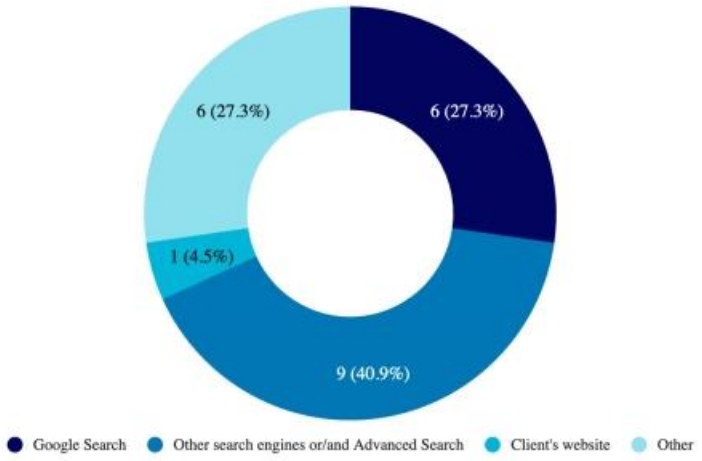


Figure 33 - Methods used to do "Creative"-related terminology advanced search

In the first examination of THL search queries, another category assigned to “Creative” texts included searches for certain job titles. As a result, SEs were questioned (“How do you research the standard job title(s) used in your target language?” [**Annex B** – Q33]) on how they would look for job titles in their target language. In this particular instance (Figure 34), the techniques varied somewhat, but a total of 14 responses indicated utilizing Google Search (8 respondents – 19.5%) or other search engines (7 respondents – 17.1%) to carry out their inquiries. Only two of these responses (4.9%) expressly

mentioned the website of the client, while 25 responses (61%) mentioned various methods. Interestingly, one of the participants mentioned using LinkedIn as a tool to find or sometimes verify job titles in different languages.

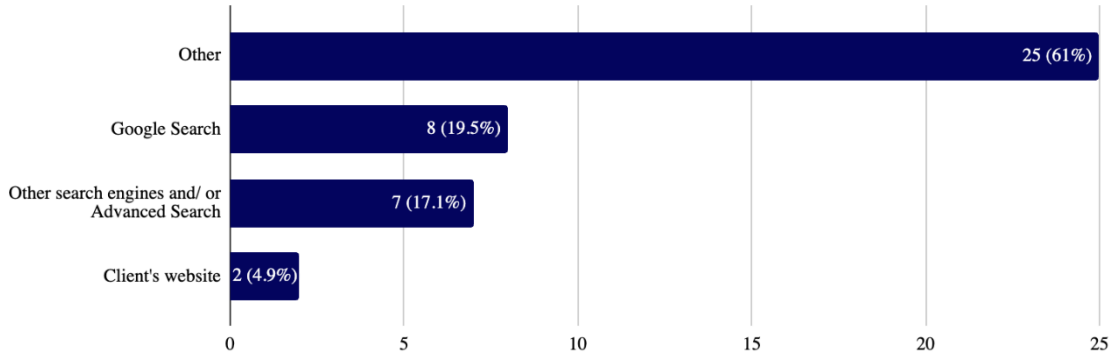


Figure 34 - How SEs look for job titles in their target language

The next multiple-choice question ("Since these clients also have more informal texts, you can sometimes find slogans and wordplay. How do you edit these?" [**Annex B** – Q34]) covered another of the identified categories, namely the usage of slogans and wordplay in "Creative" texts and how SEs would translate them (Figure 35). The most common response (37 out of 41, that is, 90.2%) was that they translated the terms after conducting research, followed by the use of THL suggestions (51.2% - 21 responses), and finally, preserving the original text (26.8% - 11 responses). Only four of the 41 respondents (9.8%) chose to input their answer, which ranged from basing their choices on experience and searching for "slang/idioms typical for a given environment/recipient, taking into account their emotional overtones, even if their literal translation is completely different."

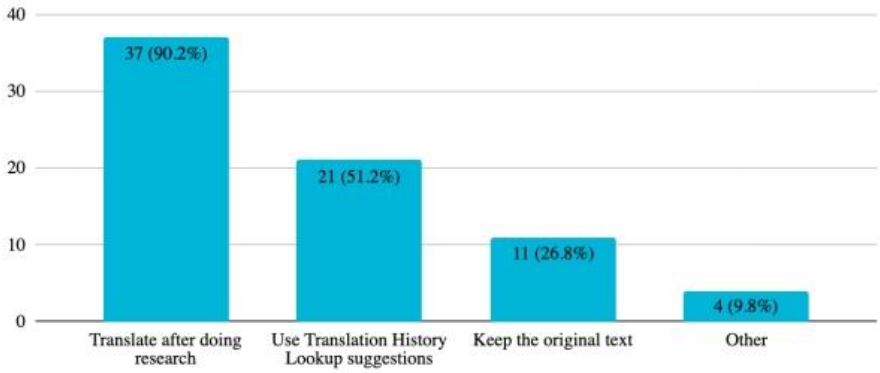


Figure 35 - How SEs deal with wordplay and slogans

As in earlier questions, this was followed by a supplementary question ("You've said you translate these slogans and wordplay after doing research. How do you do this

research?" [**Annex B** – Q35]) for those who reported they performed research before translating, asking them to elaborate on their process (Figure 36). In this case, 12 (32.4%) identified Google Search as their primary tool of research, while seven (18.9%) mentioned other search engines (Naver, Daum, or Nate[23]) or advanced search (resulting in a total number of 18 search engine-related responses). Furthermore, nine (24.3%) individuals indicated visiting the client's website to verify the usage and correctness of this proposed translation, while 11 (29.7%) mentioned other methods.

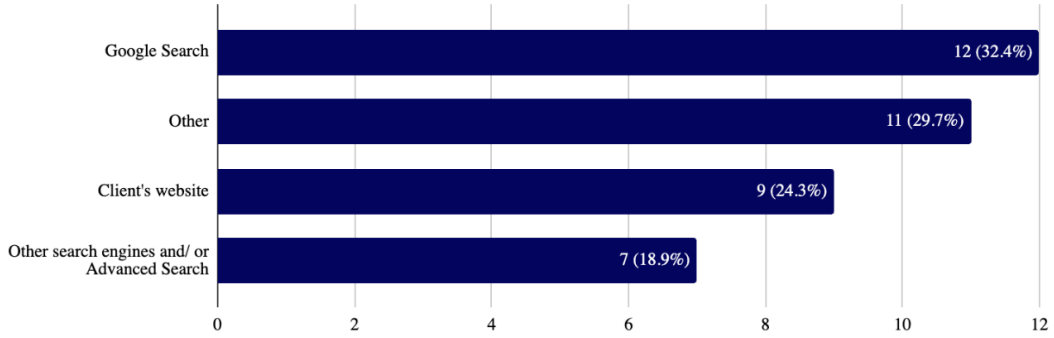


Figure 36 - Methods used by SEs to do wordplay or slogan-related advanced search

The final question ("How do you verify if the keywords and adjectives you've found during your research fit the client's style?" [**Annex B** – Q37]) focused on the usage of specific keywords or adjectives, such as the ones exemplified in the THL search analysis (see **Section 4.2**), and how SEs confirm if the ones they choose fit the style of the client (Figure 37). Out of 41 responses, the client's website was mentioned as the primary reference for analysis by the vast majority (53.7% - 22 responses), as predicted. Other individuals, however, suggested that they mainly used Google Search and other search engines (a total of six responses).

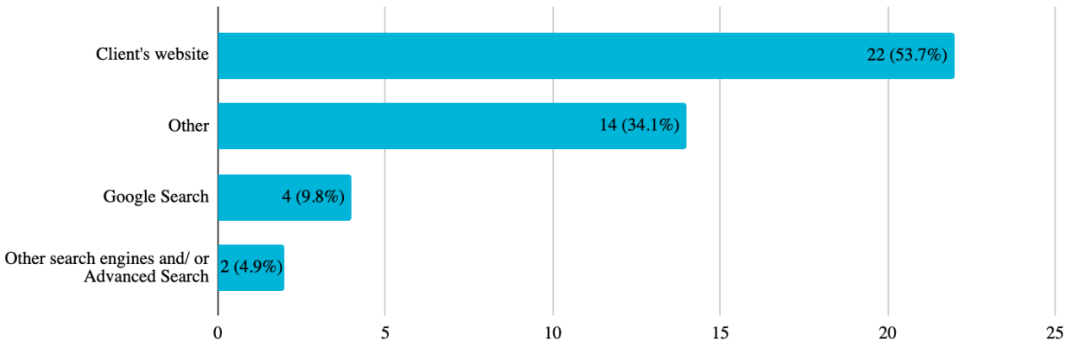


Figure 37 - Methods used by SEs to do a keyword-related advanced search

[23] These are three South-Korean search engines (see: https://www.naver.com/ ; https://www.daum.net/ ; https://www.nate.com/).

4.3.6. Key questionnaire results

In conclusion, while the answers vary per the topic, the type of issues, and the challenges they might raise, we can observe that when advanced research was required, the most frequently used tool and practice across any type of translation or review problem these participants encountered was Google Search. In fact, in Table 6, it is possible to observe that across ten multiple-choice and open-ended questions in both text categories, Google Search is the most common answer (a total of five questions where it is the top answer).

#	Question	Top Response	%	Type of question
14	While you're reviewing, you will sometimes find acronyms and initialisms. How do you verify what an acronym or initialism stands for?	Google Search	100%	Multiple-choice
17	Different types of texts have different terminology. How do you verify the validity and accuracy of a term?	Google Search	86.7%	Multiple-choice
21	How do you verify if the existing translation you've chosen is the correct one in your target language? (Electrical and automotive terminology)	Other	36.4%	Open-ended
22	Some texts mention certain products and brands. How do you find out if the word(s) refers to a product or a brand?	Client's website	78.6%	Multiple-choice
24	You've said you research unknown product names outside the Polyglot interface. How do you do this research?	Google Search	64.7%	Open-ended
26	You've said you research unknown brand names outside the Polyglot interface. How do you do this research?	Google Search	47.6%	Open-ended
31	How do you verify if the existing translation you've chosen is the correct one in your target language? (Medical terminology)	Other Search Engines or Advanced Search	40.9%	Open-ended
33	How do you research the standard job title(s) used in your target language?	Other	61%	Open-ended
35	You've said you translate these slogans and wordplay after doing research. How do you do this research?	Google search	32.4%	Open-ended
37	How do you verify if the keywords and adjectives you've found during your research fit the client's style?	Client's website	53.7%	Open-ended

Table 6 - Predominance of Google Search as a solution

In addition, whenever it is not the top answer, it is a part of the top three, and if SEs do not use Google Search, they seem to use other search engines as their primary source of knowledge (such as Baidu, Daum, Naver, Nate, Yahoo, …).

Nonetheless, while required to do so in multiple questions, most respondents do not explain their process in detail and only name these search engines.

4.3.7. Participant's Error Type Baseline per client

Client A		Client B		Client C		Client D	
Error	**%**	**Error**	**%**	**Error**	**%**	**Error**	**%**
Addition	22.6%	**Mistranslation**	**31.1%**	Date/Time Format	45.7%	**Mistranslation**	**17.4%**
Measurement Format	7.5%	Unnatural Flow	18.9%	Unnatural Flow	13.1%	Wrong Named Entity	14.9%
Word Order	7.1%	Whitespace	11.5%	Punctuation	4.1%	Currency Format	10.8%
Mistranslation	**6.6%**	Omission	10.4%	Lacks Creativity	4%	Inconsistency	9.4%
Whitespace	6.4%	Capitalization	7.3%	**Mistranslation**	**4%**	Unnatural Flow	9.3%

Table 7 - Participant's Error Baseline per Client

The table above (Table 7) shows an error type[24] baseline for the same dates as the MQM scores mentioned before (see **Section 4.1**), but this time only concerning the participants of the "Translation Research Skills Survey." Considering that this baseline presents error types and not only numbers, it provides another layer of information into quality that numbers alone cannot achieve. While "Mistranslation" is an error found in every one of the clients and can be attributed to terminology which is not a part of the glossaries each client has, it is not the only error or even the top overall error. In fact, we can find many types of errors across different clients.

Consequently, these findings raised some questions. If they all do extensive research and seem to do everything "right," why are there still so many errors? One

[24] All the error types found in Table 7 are a part of Unbabel's Annotation Typology. However, since this taxonomy is not accessible to external entities, please refer to MQM's definitions (see: https://themqm.info/typology/) (Unbabel's "Unnatural Flow" corresponds to MQM's "Unidiomatic Style"). The only two error types (featured in Table 7) not present in the original framework are "Lacks creativity" (when a text is not creative, flexible, or engaging enough from a linguistic point of view) and "Wrong Name Entity" (used to mark problems concerning NEs, such as orthographic mistakes).

answer to this may be that studies have shown that the Hawthorne effect[25] usually affects the results of questionnaires. Nevertheless, should we focus on what they find most difficult? However, does focusing on terminology only solve part of the problem? Because as was shown, there are still many issues in reviews, such as texts from Creative clients, which require much research but are not necessarily related to terminology.

How could we create materials useful to all SEs and not just a part of them? And even, perhaps, not only SEs but also editors and annotators. Moreover, which of the analyzed translation, revision, and PE competences would be the best and most feasible to develop?

Therefore, and since it was necessary to be realistic about the amount of time needed to do this work and how this "teaching" would not be done in the usual classroom environment, it was decided the focus should be on developing their "instrumental sub-competence." That is, help them use the tool they most mention in the most efficient way possible: Google Search.

4.4. Research Skills Materials

4.4.1. Google Search

The creation of these materials began with an attempt to comprehend how Google Search works and all the tools it provides. To execute this task, we conducted a first search, during which we examined several Google Support and Help pages in order to gain a better understanding of these features[26].

Nonetheless, it immediately became apparent that Google does not adequately advertise or explain all of its tools or functionalities and how we can utilize them. This fact might be one of the reasons why SEs (and many translation students, as seen in **Section 2.5**) seem unaware of them.

As a result, we had to broaden the investigation to include other websites and articles. These included websites from multiple institutions (such as universities) (UTSA,

[25] This effect is described by the Collins English Dictionary as the "improvement in the performance of employees, students, etc, brought about by making changes in working methods, resulting from research into means of improving performance." (Collins English Dictionary, n.d.)
[26] See: Do an Advanced Search on Google and Refine web searches.

n.d.), blogs about search engine optimization (SEO) strategies[27] (Harnish, 2023), and even articles about using Google more efficiently (Roy, 2022).

Through this research, a common thread was discovered and investigated. For example, most websites highlighted the usage of "search operators" (or Boolean operators as referred to by (Molina & Sales, 2008, p. 428)) to improve the use of Google Search. To put it briefly, search operators are keywords or commands that allow the user to limit (or occasionally expand) the scope of their search. For instance, using quotation marks around a group of words can restrict the search to only verbatim content instead of several websites containing all the terms or even just one of them. However, the number of search operators on every website varied greatly, which meant that perhaps some were more useful than others. Moreover, after a brief investigation, it was evident that some were deprecated due to a new Google Search feature or for no apparent reason. Consequently, it was essential to evaluate the reliability of each one by doing multiple tests.

After each test, the ones that appeared to operate as expected of them and simultaneously made the most sense in terms of what is necessary in the context of PE or revision research skills were compiled into a list of fourteen search operators.

Nevertheless, these were not the only tools discovered during the initial search. Additional features, such as the "Advanced Search" page available within Google Search and the standard tools (or buttons) found under the search bar after performing a search, were also focus points. Both of these (and the options contained within them) were studied to determine how they could benefit various types of translation-related research. It is also worth pointing out that we carried out most of these tests using past THL searches to simulate an actual real-life review situation as closely as possible.

4.4.1.1. Materials structure

This research work culminated in a three-part structure that included what we labeled Basic Search Tools (the options that appear beneath the search bar), the previously mentioned Search Operators, and the Advanced Search page options. Ultimately, this resulted in the following layout: four Basic Search Tools, 14 Search Operators, and 11 Advanced Search options.

[27] These are techniques used to increase a website's traffic by increasing the odds of it appearing on the first results of search engines.

Moreover, we gave each tool a name to facilitate understanding and navigation across this structure. When names were already available, we used those, as seen in the case of both Basic Search Tools and Advanced Search (**Annex D**). However, in the case of Search Operators, websites commonly refer to them through their keyword, symbol, or function. Therefore, in that case, they were given titles descriptive of the function of the said operator (**Annex D**). Additionally, since some of the tools in each section (Basic Search Tools, Search Operators, or Advanced Search) have the same purpose as others found in other sections, we designed materials that would allow users to search in the most intuitive, faster, and efficient way possible.

The basic structure of each section consists of an introduction to a particular group of tools, a description of how to start using them, and then a comprehensive explanation of each tool. As previously mentioned, we developed this structure using examples from the THL searches since those were things we already knew SEs struggled with (primarily relying on the subjects they mentioned finding the most difficult to translate, such as terminology, as seen in **Section 4.3.4**). In addition, relevant client websites were utilized in examples to provide SEs with situations they most likely had encountered before.

4.4.1.2. Written format (PDF)

As a result of the investigation detailed in the previous section, we decided to present the information in the form of a 52-page PDF document. While it could have been possible to separate the three sections into distinct documents, we opted for grouping all the content into one file to avoid confusing SEs with various files.

Furthermore, given that certain portions of each section might occasionally overlap, it was considered preferable to maintain the materials in a single document. In addition to that, this format was initially a priority because numerous community members had previously requested PDF files for the Language and Annotation Guidelines available on Unbabel's website, so they could consult them offline and perhaps print them out.

Nonetheless, that is not currently feasible since those articles are continuously updated. The presence of PDF files may result in individuals downloading an older version and using it indefinitely, ignoring the changes constantly made to improve both the quality and understandability of the guidelines.

On the contrary, even though this project is susceptible to modifications due to Google's ever-changing features (e.g., deprecated search operators), we considered it something that did not have such a long-lasting impact. If SEs misused one of the search tools, it would not be as dangerous as, for example, if they wrongly employed an error category from the Annotation Guidelines. Should the research skills materials not be appropriately used, the repercussions, such as longer search time, would be smaller. Nevertheless, the PDF file features a disclaimer with the date when it was last updated and requests that users report any inaccuracies to Unbabel's Community Team.

It is important to point out that in the PDF, images outlining each procedure accompany the descriptions and examples. Moreover, these materials include several tables, one for each section, with a short description of each tool, examples, and links that will redirect the user to the full explanation of the tool or another tool that serves the same purpose (**Annex D**). If there were multiple tools with overlapping functionalities, we mainly approached them similarly. That is, using the same examples to maintain consistency across the entire document and avoid any confusion that various examples for the same feature might cause.

4.4.1.3. Video format

Due to the existence of different learning styles, this information was also converted into video format. We did this so that the content could reach a larger audience and appeal to a broader range of individuals. In fact, Trados Technology Insights 2023 reports that when asked about "what sources of learning are the most useful," all types of professionals (corporate, LSP, or freelance) selected videos as their preferred learning medium (RWS, 2023, p. 29). Interestingly, "articles on the internet" was the second most voted medium by all professionals (RWS, 2023, p. 29), once again validating our current work.

Initially, the plan was that the videos would be brief to involve the community as much as possible. In fact, studies show that any video over 6 minutes seems to reduce the median engagement of students (Brame, 2016, pp. 15:es6, 4). However, due to Zendesk constraints (the platform Unbabel uses to host their Knowledge Base), the Community Team decided that we should make three lengthier videos for each section. Consequently, the videos would follow the same structure as the PDF: Basic Search Tools, Search Operators, and Advanced Search.

After deciding on the structure, the first challenge was determining which program to utilize to create these videos. For that purpose, we contacted Unbabel's Creative Team and were made aware that Loom[28] is generally the most popular platform for hosting videos at Unbabel due to its simple interface. Moreover, it allows users to record themselves and their computer screen simultaneously, generates transcriptions, and is easier to attach to Zendesk articles. Unfortunately, since it is a program that only records what the user is doing, it does not allow for much customization regarding brand design. For example, while Loom allows adding a logo or trimming a video, we cannot change how the elements on the screen are presented after the recording is done. As a result, we did not choose Loom as the primary tool to produce the final videos. Nevertheless, since it contains interesting functionalities, such as highlighting mouse clicks, we selected it as the tool to record them.

In terms of video editing tools, after multiple trials, the program chosen for that purpose was CapCut[29], due to its features and the fact that it is free software. Following this decision, it was vital to understand how to convert the written material into engaging audiovisual content since numerous editing options were now available. Ultimately, we accomplished this by demonstrating how each search example would perform with or without the intended tool and how individuals can use it.

The overall structure of the videos consisted of an initial introduction to the subject (Google Search), another to the particular section (e.g., Basic Search Tools), and then a step-by-step analysis of each tool within that section. The description of each tool included an introductory segment to the tool, a practical example of a search made without it and its results, and a subsequent search and results obtained while using it. Furthermore, a review section was included after each feature for Search Operator-related tools, detailing the main concepts, common mistakes, and viable alternatives (if applicable) for achieving the same purpose.

Next, it was necessary to consider the visual aspect of the videos, as this is the most crucial feature given the medium in use. Therefore, we studied several videos from Unbabel's Knowledge Base and YouTube pages to comprehend the brand design's identity and evolution. Moreover, Unbabel's Brandpad[30] page was an essential piece in

[28] See https://www.loom.com/ .
[29] See https://www.capcut.com/ .
[30] See https://brandpad.io/unbabel-brand/ .

the development of this project. This platform contains and explains most of the details and resources regarding Unbabel's brand design, such as the official colors, illustrations, typeface, and even logos (Unbabel, n.d. e).

First and foremost, we decided that the videos should not feature a video recording of a speaker (as often happens in Unbabel's Loom resources) but that the tools themselves should be the main focus. As a result, these should primarily consist of browser recordings where we would demonstrate the different types of Google searches and tools. Therefore, the browser was filmed with Loom and later edited to fit the screen as a "floating window" in accordance with Unbabel's official colors.

In addition, for the sake of brand design consistency, some of the illustrations on the company's Brandpad page were selected and edited using the image editing software Procreate[31]. We also created other illustrations, so they would fit the desired concept (in this case, someone sitting at their computer and the image of a search bar to fit the theme of Google Search). After creating those graphics, we loaded them into CapCut and animated them (using the effects available in the software) into what would become the "Research Skills Materials" intro.

Moreover, even though we did not intend to include an actual person on screen, there was still a need to have someone narrating all the processes. In order to create this voiceover, we utilized Capcut's text-to-speech functionality since it was more convenient in terms of sound quality (due to the absence of professional equipment) and recording time. Additionally, the audio effects used were from CapCut, and the background music is a copyright-free song available at Bandcamp[32].

Finally, we also decided that the videos should be close-captioned to be as accessible as possible to all types of audiences (hard of hearing, non-native speakers, …). After testing other options (such as Subtitle Edit[33]), the videos were also subtitled through CapCut. Generally, subtitling software uses timecodes to match the subtitles to the video. However, since CapCut is not a subtitling-oriented program, the subtitles are instead directly connected to each video clip. Consequently, using CapCut simplified the process of altering the subtitles if we made any last-minute changes to the length or order of the videos.

[31] See https://procreate.com/ .
[32] See https://alexiaction.bandcamp.com/track/slideshow-corporate-inspiring .
[33] See https://www.nikse.dk/subtitleedit .

After completing this project phase, the following step entailed choosing the most appropriate platform for uploading the videos. As previously stated, Loom is the tool of choice for recording and hosting most of Unbabel's videos. Nonetheless, while it is possible to upload external videos to Loom, we chose YouTube as the ideal platform for this particular content. The reasons behind this decision are that YouTube is not only a platform that almost everyone can access but is also available on most devices. Moreover, because these videos have subtitles, YouTube's auto-translation feature allows users to automatically translate embedded subtitles into other languages, such as their native language if they feel more comfortable learning that way.

Additionally, timestamps were also included in each of the videos, allowing viewers to either be redirected to a specific part of the video from Unbabel articles or quickly browse and find the portions they are interested in, hopefully reducing the possible negative impact of their length.

In conclusion, this resulted in 45 minutes of video (distributed across three videos), each detailing every tool and information needed. Each video also references the other two videos and the "Research Skills Materials" playlist mentioned in the outro. Furthermore, in the description of the videos, there is also a brief introduction to the content, along with links to the "Research Skills Materials" playlist, the other videos, the Knowledge Base article, and the PDF file.

4.4.1.4. Knowledge Base Article

Finally, we integrated the written document and videos into a Knowledge Base article titled "How to research on Google more efficiently," where there is a succinct explanation of the materials (**Annex D**).

4.4.2. Senior Editor's feedback on Research Skills Materials

At the end of the Translation Research Skills survey, we asked participants if they had any interest in taking part in a research skills training program or having access to those materials. Out of 45 participants, 40 (88.9%) answered yes, and we requested they provide their email addresses so that the Unbabel Community Team could contact them when these became available.

Upon completing the materials, we reached out to the interested participants. We shared the Knowledge Base article with them, along with a short anonymous survey

(seven questions) regarding how they felt about the materials. Among 40 contacted SEs, seven replied to the survey, and one contacted us through email.

In terms of the survey itself (Table 8), the first question (**Annex E** – Q1) used a Likert scale (1 to 5) and asked participants how helpful they found the materials, which resulted in an average value of 4.3 (2 participants = 3 [28.6%]; 1 participant = 4 [14.3%]; 4 participants = 5 [57.1%]).

The next question (**Annex E** – Q2) was concerned with which format the SEs preferred. This resulted in a tie between PDF and video (3 participants - 42.9% each), with only one participant saying they had no preference (14.3%).

The following three yes-or-no questions (**Annex E** – Q3, Q4, and Q5) were concerned with: the clarity of the explanations; if the format was visually appealing; and if they had enough practical examples to help them learn. All seven participants replied yes to every question (100%).

	Q1	**Q2**	**Q3**	**Q4**	**Q5**
Participant 1	4	Videos	Yes	Yes	Yes
Participant 2	5	Videos	Yes	Yes	Yes
Participant 3	3	PDF	Yes	Yes	Yes
Participant 4	5	Videos	Yes	Yes	Yes
Participant 5	3	PDF	Yes	Yes	Yes
Participant 6	5	NP	Yes	Yes	Yes
Participant 7	5	PDF	Yes	Yes	Yes

Table 8 - Participants' answers (Questions 1 to 5)

Finally, the last two questions were both open-ended questions (**Annex E** – Q6 and Q7). The first inquired about how these materials helped them improve their work. The answers varied. One participant mentioned that since they had just read the materials, they did not have an actual impact yet, and they already knew most of what they called "Google voodoo." However, this was the only mildly negative response. The rest of the respondents mentioned that the materials helped with: working speed (by using the verbatim feature); narrowing the search to an exact match in word, country, and language simultaneously; to search for specific terminology; finding the names of items for various clients (even beyond the Lingo24 rooster).

The last question asked participants if they had any input on how to improve the learning materials or other things they would like to see. One of the participants pointed

out that since Google is the most used search engine, most SEs probably already knew about the features within the materials, and they would like to see similar materials for other websites. Another participant mentioned using Google Lens for a specific client and pointed out that it may be useful for other SEs. Moreover, another participant said we could present a condensed format version of the information in the editors' user dashboard, which Unbabel also uses for teaching within the platform. The rest of the participants expressed how much they liked the materials and would like to see more. They also requested that other Unbabel materials be available in PDF format and not only in articles. Finally, the SE who sent us an email stated that they thought the videos were "really helpful" and had "never tried most of those tips before."

In terms of other reception, the Knowledge Base article also has an option where viewers are asked, "Was this article helpful?" to which seven people responded "yes" at the time of this report. Finally, the videos themselves have stacked, at this point in time, over 270 views total views[34]. In conclusion, the response gathered to these materials was overwhelmingly positive.

[34] Since all the videos are unlisted on YouTube, these numbers are representative of people who accessed them through the article and are not skewed by "outside" views.

Contributions, Limitations, and Future Work

The main objective of this research was to understand why Lingo24 content reviews fell short of the quality expectations established by Unbabel and develop solutions to improve the work of SEs.

In order to achieve this goal, it was first necessary to verify that this perception was indeed accurate, and we did this by examining MQM scores across four different Lingo24 clients. The method employed for this analysis was the calculation of the average baseline MQM requirement for each TP and LP Tier, followed by an assessment of how the MQM scores of 11 batches compared to these. The data showed that merely two of those batches were above the minimum MQM criteria, demonstrating the need for this project.

Nevertheless, while it was a useful starting point, this data did not provide enough information about the root cause of these issues. Consequently, we decided to further investigate the potential problems by examining almost one thousand THL searches. From this examination, we managed to draw some conclusions. However, these queries were not enough to prove that these were the aspects hindering their work. Therefore, we determined that it was necessary to obtain empirical data from the source, that is, from the SEs themselves.

As a result, we created a 40-question-long survey using the THL dataset analysis as its basis. The aim of this questionnaire was threefold: to confirm if the identified issues in the THL dataset analysis were correct, to detect work patterns and preferences among SEs, and to obtain more resources to improve both the Language Guidelines and client instructions.

Through the survey results, we were able to confirm our previous THL searches hypothesis and gain more insights into the preferences of SEs (relation between type of content, previous experience, and field of study), their main difficulties (THL usage or overall struggles), the solutions used to solve translation/review problems (Google Search, mostly), and multiple resources (websites for specific subject matters).

After this phase, we analyzed MQM error data only pertaining to the participants of this survey and were able to observe that there was a high percentage of errors across the same clients. This, in turn, raised some questions: why do people who seem to research everything correctly make so many errors? While much can be inferred from

this, we concluded that perhaps these errors were due to inaccurate or inefficient searches. As a result, we decided to create Google Search-oriented materials since this was the tool most mentioned by SEs.

As detailed in **Section 4.4**, these materials were developed in a way that they only include up-to-date information, are accessible in terms of different preferences or any type of impairment, and are comprised of engaging content with relevant examples that address a business problem. Considering the context within which these materials insert themselves, this last aspect was one of the most important. While Unbabel continuously assesses the work quality of its community, most editors do not come from a Translation Studies-related background. In fact, only three of the survey respondents had a degree in this area. Consequently, it was necessary not to make assumptions about their knowledge level and to create materials as inclusive as possible. Furthermore, we hoped these materials could be helpful for not only SEs but also other members of the community, such as post-editors.

Unfortunately, due to the time constraints of this internship, there was not enough time to do a long-term follow-up of the impact of these materials (e.g., new MQM analysis) and the maintenance of this work. Nonetheless, upon the release of the materials, a short survey was sent out to interested SEs, to gauge its usefulness. The response was mostly positive, which, together with the number of views of the materials, indicates a lack of instrumental competence skills, which this project hoped to correct.

Moreover, throughout this internship, we also helped improve Language Guidelines resources (obtained from the survey and personal search, mostly terminology-focused); worked in the development of a template for client instructions, which aims to maintain the same structure of instructions across clients; and provided input in the design of other instructional videos developed by Unbabel.

It is also important to point out that, with the increasing popularity of large language models (LLMs), the training of instrumental competence may become even more nuanced. In addition to generative AI (e.g., ChatGPT or Bard[35]) already being used as the new Google (and making quite a few mistakes[36]), LLMs are improving at an impressive rate, and some are already being created with specific uses in mind, such as

[35] See: https://bard.google.com/ .
[36] Bard made a factual mistake in its first demo. As James Vincent states: "The mistake highlights the biggest problem of using AI chatbots to replace search engines — they make stuff up." (Vincent, 2023)

academic research[37]. As a result, it is not unthinkable that these tools will have a significant impact in the translation field, even beyond "translation" itself. Nonetheless, we believe that, since these tools are here to stay, we should investigate and learn to use them to our benefit instead of fighting against the current of what already is a tsunami.

[37] For example, "Elicit". See: https://elicit.org/

Bibliography[38]

ACRL. (2016). *Framework for Information Literacy for Higher Education.* Retrieved from https://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/framework1.pdf

Allen, J. H. (2003). Post-editing. In H. S. (Ed.), *Computers and Translation: A translator's guide* (pp. 297–317). Amsterdam/Philadelphia: John Benjamins. doi:https://doi.org/10.1075/btl.35.19all

ALPAC. (1966). *Language and machines: Computers in Translation and Linguistics.* Washington, D.C.: National Academy of Sciences National Research Council. Retrieved from https://nap.nationalacademies.org/resource/alpac_lm/ARC000005.pdf

Bar-Hillel, Y. (1960). The Present Status of Automatic Translation of Languages. In F. L. Alt., *Advances in Computers* (pp. 91-163). doi:https://doi.org/10.1016/S0065-2458(08)60607-5

Brame, C. J. (1 de December de 2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences EducationVol. 15, No. 4*, pp. 15:es6, 1 - 15:es6, 6.

Carmo, F. d. (2017). *Post-editing: a Theoretical and Practical Challenge for Translation.* Porto: Faculdade de Letras da Universidade do Porto.

Carmo, F. d., & Moorkens, J. (2020). Differentiating Editing, Post-Editing and Revision. In M. Koponen, B. Mossop, I. S. Robert, & G. S. (Eds.), *Translation Revision and Post-editing: Industry Practices and Cognitive Processes* (pp. 35-49). London: Routledge.

Collins English Dictionary. (n.d.). *Definition of 'Hawthorne effect'*. Retrieved March 26, 2023, from Collins English Dictionary: https://www.collinsdictionary.com/dictionary/english/hawthorne-effect

CSA Research. (2020, July 7). *Survey of 8709 Consumers in 29 Countries Finds that 76% Prefer Purchasing Products with Information in their Own Language.* Retrieved January 5, 2023, from CSA Research: https://csa-research.com/Blogs-Events/CSA-in-the-Media/Press-Releases/Consumers-Prefer-their-Own-Language

EMT. (2022). Competence Framework 2022. Retrieved from https://commission.europa.eu/system/files/2022-11/emt_competence_fwk_2022_en.pdf

[38] This bibliography follows APA 6th Edition rules.

Enríquez Raido, V. (2011). *Investigating the Web Search Behaviors of Translation Students: An Exploratory and Multiple-Case Study.* Spain: Universitat Ramon Llull.

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces 6:2*, 291-309.

García, I. (2012). A brief history of postediting and of research on postediting. *Revista Anglo Saxonica*, 291-310.

Gonçalves, M. (2021). *Analysis on the impact of the source text quality: Building a data-driven typology.* Lisboa: Faculdade de Letras da Universidade de Lisboa. Retrieved from http://hdl.handle.net/10451/51178

Google. (n.d.). *Do an Advanced Search on Google*. Retrieved January 1, 2023, from Google Search Help: https://support.google.com/websearch/answer/35890?hl=en&co=GENIE.Platform%3DDesktop#zippy=

Göpferich, S. (2009). Towards a model of translation competence and its acquisition. In S. Göpferich, A. L. Jakobsen, & I. M. Mees, *Behind the Mind: Methods, Models and Results in Translation Process Research* (pp. 11-37). Samfundslitteratur.

Harnish, B. (2023, February 12). *An SEO Guide to Google Advanced Search Operators*. Retrieved from Search Engine Journal: https://www.searchenginejournal.com/google-search-operators-commands/215331/

Hutchins, W. J. (1996, June). ALPAC: the (in)famous report. *MT News International, no. 14,*, pp. 9-12.

Hutchins, W. J. (2003). Machine translation: a concise history. pp. 1-21. Retrieved from https://xwiki.recursos.uoc.edu/wiki/mat00001ca/download/Research%20on%20Translation%20Technologies/Working%20with%20epub%20files%20using%20Python/WebHome/document3.pdf

ISO 17100. (2015, 05). *Translation services — Requirements for translation services*. Retrieved from https://www.iso.org/standard/59149.html

Kelly, D. (2005). *A Handbook for Translator Trainers.* St. Jerome Pub.

Kenny, D. (2019). Machine Translation. In P. W. J Piers Rawling, *Routledge Handbook of Translation and Philosophy* (pp. 428-445). Routledge.

Kepler, F., Trénous, J., Treviso, M., Vera, M., & Martins, A. F. (2019). OpenKiwi: An Open Source Framework for Quality Estimation. doi:https://doi.org/10.48550/arXiv.1902.08646

Koehn, P. (2020). *Neural Machine Translation.* Cambridge: Cambridge University Press. doi:10.1017/9781108608480

Lommel, A. R., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Revista Tradumàtica*, 455-463. doi:10.5565/rev/tradumatica.77

Lüdeling, A., & Hirschmann, H. (2015). Error annotation systems. In S. Granger, G. Gilquin, & F. M. (Eds.), *The Cambridge Handbook of Learner Corpus Research 1st Edition* (pp. 135-158). Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9781139649414.007

Martins, A. F., Junczys-Dowmunt, M., Kepler, F. N., Astudillo, R., Hokamp, C., & Grundkiewicz, R. (2017). Pushing the Limits of Translation Quality Estimation. *Transactions of the association for Computational Linguistics, 5*, 205-218. doi:10.1162/tacl_a_00056

Mohammed, T. A., & Al-Sowaidi, B. (2023). Enhancing Instrumental Competence in Translator Training in a Higher Education Context: A Task-Based Approach. *Theory and Practice in Language Studies 13(3)*, 555-566. doi:10.17507/tpls.1303.03

Molina, M. P., & Sales, D. (2008). INFOLITRANS: A model for the development of information competence for translators. *Journal of Documentation 64 (3)*, 413–437. doi:10.1108/00220410810867614

Nitzke, J., Hansen-Schirra, S., & Canfora, C. (2019). Risk management and post-editing competence. *The Journal of Specialised Translation*, 239-259.

PACTE. (2003). Building a Translation Competence Model. In F. Alves, *Triangulating Translation: Perspectives in process oriented research* (pp. 43-66). Amsterdam: John Benjamins.

Paradowska, U. (2015). Expert web searching skills for translators -a multiple-case study. In M. D. Paulina Pietrzak, *Constructing Translation Competence* (pp. 227-244). Peter Lang.

Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *META*, 487-503.

Qun, L., & Xiaojun, Z. (2014). Machine Translation: General. In C. S.-w. (Ed.), *Routledge Encyclopedia of Translation Technology 1st Edition* (pp. 105-119). London: Routledge.

REGULATION (EU) 2016/679. (2016, April 27). *On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Robert, I. S., Remael, A., & Ureel, J. J. (2017). Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, 1-19.

Roy, S. D. (2022, August 12). *How to Google like a Pro – 10 Tips for More Effective Googling*. Retrieved from freeCodeCamp: https://www.freecodecamp.org/news/how-to-google-like-a-pro-10-tips-for-effective-googling/

RWS. (2023). *Translation Technology Insights 2023.* Retrieved from https://www.trados.com/download/translation-technology-insights-2023-report/215294/

Saldanha, G., & O'Brien, S. (2014). *Research Methodologies in Translation Studies* (1st Edition ed.). London: Routledge. doi:https://doi.org/10.4324/9781315760100

Sánchez-Torrón, M., & Koehn, P. (2016). Machine Translation Quality and Post-Editor Productivity. *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA) Vol. 1: MT Researchers' Track* (pp. 16-26). Austin, Texas: AMTA. Retrieved from https://amtaweb.org/wp-content/uploads/2016/10/AMTA2016_Research_Proceedings_v7.pdf#page=22

Sharma, S. (2021, December 16). *Unbabel acquires Lingo24 to bolster AI-driven translations for enterprises.* Retrieved January 5, 2023, from VentureBeat: https://venturebeat.com/business/unbabel-acquires-lingo24-to-bolster-ai-driven-translations-for-enterprises/

Silva, B. (2022). *Translation error annotation : building an annotation module for east asian languages.* Lisboa: Faculdade de Letras da Universidade de Lisboa. Retrieved from http://hdl.handle.net/10451/56071

Traductanet. (2019, 8 19). *BREVE HISTÓRIA DA TRADUÇÃO AUTOMÁTICA I : Os primeiros anos.* Retrieved 2 5, 2023, from Traductanet: https://traductanet.com/pt-pt/ultimas-noticias/breve-historia-da-traducao-automatica-i-os-primeiros-anos/

Unbabel. (2021, December 16). *Fast-tracking Our LangOps Expansion: Unbabel Welcomes Lingo24 into the Fold*. Retrieved October 20, 2022, from Unbabel: https://resources.unbabel.com/blog/unbabel-acquires-lingo24

Unbabel. (2021, October 26). *Unbabel's 2021 Global Multilingual CX Survey Reveals 68% of Consumers Prefer to Speak with Brands in.* Retrieved January 5, 2023, from Resource Center: https://resources.unbabel.com/press-releases/unbabel-s-2021-global-multilingual-cx-survey-reveals-68-of-consumers-prefer-to-speak-with-brands-in-their-native-language

Unbabel. (n.d. a). *Getting started at Unbabel: a definitive guide*. Retrieved February 23, 2023, from Unbabel Support: https://help.unbabel.com/hc/en-us/articles/360047664013-Getting-started-at-Unbabel-a-definitive-guide

Unbabel. (n.d. b). *What are evaluators looking for? How do I succeed?* Retrieved March 3, 2023, from Unbabel Support: https://help.unbabel.com/hc/en-us/articles/360048020473-What-are-evaluators-looking-for-How-do-I-succeed-

Unbabel. (n.d. c). *I want to become a reviewer*. Retrieved December 20, 2022, from Unbabel Support: https://help.unbabel.com/hc/en-us/articles/360048026513-I-want-to-become-a-Senior-Editor

Unbabel. (n.d. d). *What is translation history lookup and how do I use it?* Retrieved January 5, 2023, from Unbabel Support: https://help.unbabel.com/hc/en-us/articles/9630139968791-What-is-translation-history-lookup-and-how-do-I-use-it-

Unbabel. (n.d. e). *Unbabel Brand*. Retrieved from Brandpad: https://brandpad.io/unbabel-brand/

UTSA. (n.d.). *Advanced Google Search*. Retrieved January 15, 2023, from UTSA Libraries: https://libguides.utsa.edu/advancedgoogle#s-lg-box-10834701

Vasconcellos, M., & Leon, M. (1985). Spanam and Engspan: Machine Translation at the Pan American Health Organization. *Computational Linguistics, 11(2-3),* 122–136. Retrieved from https://aclanthology.org/J85-2003.pdf

Vincent, J. (2023, February 2). *Google's AI chatbot Bard makes factual error in first demo*. Retrieved March 5, 2023, from The Verge: https://www.theverge.com/2023/2/8/23590864/google-ai-chatbot-bard-mistake-error-exoplanet-demo

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., & Norouzi, M. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

Zendesk. (2023, January 20). *What is CRM (Customer Relationship Management)?* Retrieved March 7, 2023, from Zendesk: https://www.zendesk.com/sell/crm/what-is-crm/

List of Figures

Figure 1- Pre-Lingo24 Integration General Pipeline (Gonçalves, 2021) 4
Figure 2 - "Realtime" Translation Pipeline (Internal resources) 7
Figure 3 - "Hybrid" Translation Pipeline (Internal resources) 7
Figure 4 - "Specialist" Translation Pipeline (Internal resources) 8
Figure 5 - Editor's evaluation rating system (Silva, 2022) 12
Figure 6 - PACTE's Translation Competence Model 24
Figure 7 - Göpferich's Translation Competence Model 25
Figure 8 - Robert's Translation Revision Competence Model 26
Figure 9 - Nitzke's Post-editing Competence Model 27
Figure 10 - Participation across LPs 38
Figure 11 - Age range of participants 39
Figure 12 - Gender of participant 39
Figure 13 - Education level of participants 40
Figure 14 - Coding template 41
Figure 15 - Fields of study of the participants 43
Figure 16 - Types of content the participants are interested in 44
Figure 17 - Aggregated fields of study (Professional) 45
Figure 18 - Aggregated fields of study (Creative) 45
Figure 19 - Employment history (Unbabel and/or Lingo24) 46
Figure 20 - Use of Translation History Lookup 48
Figure 21 - What SEs use Translation History Lookup for 49
Figure 22 - Comparison between THL searches and what SEs find difficult to translate/review 50
Figure 23 - How SEs verify what acronyms/initialisms stand for 51
Figure 24 - How SEs verify the validity and accuracy of a term 52
Figure 25 - How SEs deal with "Professional" terminology 54
Figure 26 - Methods used to do "Professional"-related terminology advanced search .. 55
Figure 27 - How SEs deal with products/brands names 56
Figure 28 - What SEs do when faced with unknown product names 56
Figure 29 - What SEs do when faced with unknown brand names 57
Figure 30 - Methods used for product-related advanced search 57
Figure 31 - Methods used for brand-related advanced search 57

Figure 32 - How SEs deal with "Creative" terminology 59
Figure 33 - Methods used to do "Creative"-related terminology advanced search 59
Figure 34 - How SEs look for job titles in their target language 60
Figure 35 - How SEs deal with wordplay and slogans 60
Figure 36 - Methods used by SEs to do wordplay or slogan-related advanced search .. 61
Figure 37 - Methods used by SEs to do a keyword-related advanced search 61

List of Tables

Table 1 - Predetermined Pipeline/LP Quality Framework Percentages 31
Table 2 - Predetermined Pipeline/LP MQM scores 31
Table 3 - Number of SEs selected per LP 34
Table 4 – Shadow Mode MQM scores from four Lingo24 clients 35
Table 5 - THL most common searches within the dataset 37
Table 6 - Predominance of Google Search as a solution 62
Table 7 - Participant's Error Baseline per Client 63
Table 8 - Participants' answers (Questions 1 to 5) 71

Annexes

Annex A – Competence Models

PACTE's TC	Göpferich's TC	Robert's TRC	Nitzke's PEC	Definition(s)
Bilingual	Communication competence in two languages	Bilingual	Bilingual	"Predominantly procedural knowledge needed to communicate in two languages. It includes the specific feature of interference control when alternating between the two languages. It is made up of pragmatic, socio-linguistic, textual, grammatical, and lexical knowledge in the two languages." (PACTE, 2003, p. 58)
Extralinguistic	Domain	Extralinguistic	Extralinguistic	"Predominantly declarative knowledge, both implicit and explicit, about the world in general and special areas. It includes: (1) bicultural knowledge (about the source and target cultures); (2) Building a translation competence model encyclopaedic knowledge (about the world in general); (3) subject knowledge (in special areas)." (PACTE, 2003, pp. 58-59)
Instrumental	Tools and research	Tools and research (+ Psycho-motor)	Instrumental	"Predominantly procedural knowledge related to the use of documentation sources and information and communication technologies applied to translation: dictionaries of all kinds, encyclopaedias, grammars, style books, parallel texts, electronic corpora, searchers, etc." (PACTE, 2003, p. 59) "Predominantly procedural knowledge related to the use of translation- and revision- specific conventional and electronic tools. Definition based on PACTE (2003) and Göpferich (2009)." (Robert, Remael, & Ureel, 2017, p. 14)
			Research	"A post-editor needs to know where and how to find information he or she does not know. Depending on the thematic field of the translation, specialised (online) dictionaries might be the first choice, whereas, for others, parallel corpora or thesauri might be a better option. Efficient research strategies positively influence the workflow time of a post-editing task. (...)" (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 249)
Knowledge of translation	Translation routine activation (+translation self-concept)	Knowledge of translation	Translation	"Predominantly declarative knowledge, both implicit and explicit, about what translation is and aspects of the profession. It includes: (1) knowledge about how translation functions: types of translation units, processes required, methods and procedures used (strategies and techniques), and types of problems; (2) knowledge related to professional translation practice: knowledge of the work market (different types of briefs, clients, and audiences, etc.)." (PACTE, 2003, p. 59)
		Translation routine activation		"This competence comprises the knowledge and the abilities to recall and apply certain – mostly language-pair-specific- (standard) transfer operations (or shifts) which frequently lead to acceptable target-language equivalents. In Hönig's terminology, this competence could be described as the ability to activate productive micro-strategies." (Göpferich, 2009, p. 21)
Strategic	Strategic (+ Motivation)	Strategic	Strategic	"Procedural knowledge to guarantee the efficiency of the translation process and solve the problems encountered. This is an essential sub-competence that affects all the others and causes inter-relations amongst them because it controls the translation process. Its functions are: (1) to plan the process and carry out the translation project (choice of the most adequate method); (2) to evaluate the process and the partial results obtained in relation to the final purpose; (3) to activate the different sub-competencies and compensate for deficiencies in them; (4) to identify translation problems and apply

				procedures to solve them." (PACTE, 2003, p. 59) "(...) How strictly translators adhere to employing this macro strategy depends on their strategic competence and their situation specific motivation, which may be both intrinsic (enjoying translating) or extrinsic (payment, fear of compensatory damages, etc.)" (Göpferich, 2009, p. 22)
		Interpersonal (Kelly, 2005)	Service	"Ability to work with other professionals in- volved in translation process (translators, revisers, documentary researchers, terminologists, project managers, layout specialists), and other actors (clients, initiators, authors, users, subject area experts). Teamwork. Negotiation skills. Leadership skills." (Kelly, 2005, p. 33) "Service competence means that the post-editor should be able to calculate prices competently, consciously, and transparently considering the quality of the MT output and the necessary PE effort (cf. translation competence model by EMT Expert Group (2009) (...) The post-editor should be able to match the needs of the customer with the set-up and conditions of the PE task as well as with the resources available to be able to make an appropriate offer that, for example, calculates a realistic time frame for the job." (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 248)
			Consulting	"Depending on the risk assessment and the strategic decisions, a post- editor has to inform the customer or project manager about potential risks as well as problem-solving strategies, respectively, i.e., the risk assessment should enable the post-editor to give advice on these questions. (...)" (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 248)
			Risk assessment	"As explained above, one of the most important competences a post-editor needs is the ability to assess the risk of the text to be translated. The proposed decision tree helps the customer or the post-editor to competently assess and document the risks around a specific source text as well as the corresponding translation workflow." (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 248)
		Revision routine activation	Revision	"Knowledge and abilities to recall and apply certain – mostly language-pair-specific – (standard) revision operations (detection strategies and immediate solutions) which frequently lead to acceptable target-language solutions. Definition based on Göpferich (2009)." (Robert, Remael, & Ureel, 2017, p. 14)
		Knowledge about revision		"Predominantly declarative knowledge, implicit and explicit, about what revision is and about aspects of the profession (...)" (Robert, Remael, & Ureel, 2017, p. 14)
	Psycho-motor			"These are the psychomotor abilities required for reading and writing (with electronic tools). The more developed these competences are, the less cognitive capacity is required, leaving more capacity for other cognitive tasks. (...)" (Göpferich, 2009, p. 21)
			Machine Translation	"A post-editor needs to know how an MT system works and which possible pitfalls it may generate. MT systems often generate different problems than human translators produce (...)" (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 250)
			Post-editing	"(...) errors triggered by neural MT are harder to identify since the MT output is more fluent and correct, which leads to the problem of overlooking mistakes which are not obvious (Toral et al. 2018 show that pauses become fewer but longer in PE NMT output). Therefore, the post-editor has to be trained in spotting exactly these more fine-grained problems." (Nitzke, Hansen-Schirra, & Canfora, 2019, p. 250)

Annex B – Translation Research Skills Survey

1. [Multiple Choice - One Option]
   First, how old are you?
   - 18-24
   - 25-36
   - 36-45
   - 46-55
   - 56-65
   - 66+
   - Prefer not to say
2. [Multiple Choice - One Option]
   How would you describe your gender?
   - Female
   - Male
   - Nonbinary
   - Prefer not to say
   - Other (allows input)
3. [Multiple Choice - One Option]
   What is your level of education?
   - High school or equivalent
   - University (Bachelor's Degree)
   - University (Masters Degree)
   - University (Doctorate)
   - Other higher education programme equivalent to university level
   - None of the above
   - Prefer not to say
4. [Multiple Choice - One Option]
   What is your main field of study?
   - Architecture
   - Art and history of art
   - Arts - other
   - Business and management
   - Classics (including ancient history and/or classical languages)
   - Computer science
   - Economics
   - Education
   - History
   - Humanities - other
   - Journalism, media and communication
   - Law
   - Literature
   - Mathematics and statistics
   - Medicine
   - Modern languages & linguistics
   - Natural sciences (Physics, Chemistry, Biology and Earth Sciences)
   - Performing arts (Music, dance, drama, theater)
   - Philosophy
   - Political science
   - Psychology
   - Religion and divinity
   - Science - other
   - Social sciences - other
   - Sociology
   - Translation studies
   - Vocational studies
   - None of the above
   - Prefer not to say
5. [Multiple Choice - One Option]
   What best describes your situation?
   - I've been with Unbabel since the beginning
   - I've historically been with Lingo24, but I am now working on Unbabel tasks
   - I don't know
6. [Multiple Choice - One Option]
   What is your main language pair?
   - English to Arabic
   - English to Bulgarian
   - English to Chinese (Simplified)
   - English to Chinese (Traditional)
   - English to Czech
   - English to Danish
   - English to Dutch
   - English to Finnish
   - English to French
   - English to German
   - English to Greek
   - English to Hebrew
   - English to Hindi
   - English to Hungarian
   - English to Indonesian
   - English to Japanese
   - English to Korean
   - English to Norwegian
   - English to Polish
   - English to Portuguese (BR)
   - English to Portuguese (EU)
   - English to Romanian
   - English to Russian
   - English to Spanish (EU)
   - English to Spanish (LatAm)

- English to Swedish
- English to Thai
- English to Turkish

7. [Open Ended]
What kind of content is your favorite to translate/review?

While working at **Lingo24/Unbabel** you've probably translated/reviewed texts for one (**or more!**) clients such as… (description of each client with examples)

8. [Multiple Selection]
- Client A
- Client B
- Client C
- Client D

9. [Open Ended]
[If the answer to question 4 was "I've historically been with Lingo24…"]
Estimated amount of time you've spent translating and/or reviewing texts for [client/s] while working only at Lingo24.

10. [Open Ended]
[If the answer to question 4 was "I've historically been with Lingo24…"]
Estimated amount of words you've spent translating and/or reviewing texts for [client/s] while working only at Lingo24.

11. [Open Ended]
[If the answer to question 4 was "I've historically been with Lingo24…"]
Any other information regarding your work with [client/s] while working only at Lingo24.

12. [Likert question]
How often do you use the Translation History Lookup in Polyglot (Editing Interface)?
Translation History Lookup is often known as a concordance search.
- Very frequently
- Frequently
- Occasionally
- Rarely
- Very rarely
- Never

13. [Multiple Selection]
You usually use the Translation History Lookup to search for…
Select all that apply.
- Slogans/wordplay
- Acronyms
- Weight/temperature
- Job titles
- Keywords/adjectives
- Medical terminology
- Electrical terminology
- Automotive terminology
- Measures and other units [inches, kiloohm,..]
- Food related words
- Other (allows input)

14. [Multiple Selection]
While you're reviewing, you will sometimes find **acronyms** and **initialisms.** How do you verify what an acronym or initialism stands for?
Acronym is when the group of letters is read as new word (e.g., **NATO**)
Initialism is when each letter is pronounced individually (e.g., **USA**)
- Context from the text
- Translation History Lookup
- Google Search
- Client's website
- I usually don't research about these
- Other (allows input)

15. [Multiple Choice - One Option]

If there are no client instructions, how will you deal with an **acronym** or **initialism**? You can only select **ONE** option.

- Do not translate
- Use the standard target language acronym/initialism
- Keep the original followed by an explanation or translation between parenthesis
- Spell it out
- Other (allows input)

16. [Open Ended]

Which trusted linguistic sources do you usually use to verify **acronyms** and **initialisms**? Please provide **concrete** examples (**NOT** Google search, Linguee, ...).

17. [Multiple Selection]

Different types of texts have different terminology. How do you verify the validity and accuracy of a **term**? A term is a word or group of words used to describe a specialized concept or thing within a branch of study. For example, "**aortic stenosis**" is a **medical term** and "**chassis**" is an **automotive term**.

- Client's website
- Translation History Lookup
- Advanced Google Search
- Trusted linguistic sources
- Other (allows input)

18. [Open Ended]

[If the LP chosen in Question 5 isn't included in IATE]

Which official linguistic banks (like IATE or Termium), that feature **your target language**, do you use?

[Professional Client only questions]

19. [Open Ended]

Besides the resources provided by Unbabel/Lingo24 (if any), what specialized (**IT, electrical, automotive, …**) terminology websites do you use? Please provide **concrete** examples (**NOT** Google search, Linguee, ...).

20. [Multiple Choice - One option]

Oftentimes there isn't an obvious or standardized translation for a term. How do you deal with **uncommon and difficult IT, electrical, automotive related terms**, for which there may already be a translation somewhere? You can only select **ONE** option.

- Do not translate
- Keep the original followed by an explanation between parenthesis
- Explain the term using other words
- Create your own translation by analyzing other similar (and already translated) terms
- Pick one existing translation for the term after doing advanced research
- I skip the task when there are too many of these

21. [Open Ended]

How do you verify if the **existing translation** you've chosen is the correct one in your target language? Please provide a **detailed** description of your process.

22. [Multiple Selection]

Some texts mention certain **products** and **brands.** How do you find out if the word(s) refer to a **product** or a **brand**? Select all that apply.

- Translation History Lookup suggestions
- Context from the text
- Client's website
- Google Search
- Other (allows input)

23. [Multiple Selection]

How do you deal with **unknown product names**? Select all that apply.

- Research in Translation History Lookup whether it needs to be translated/transliterated or left in English
- Leave it in English by default
- Research outside the Polyglot interface whether it needs to be translated/transliterated or to be left in English
- Other (allows input)

24. [If answer to the previous question was "Research outside the Polyglot Interface…"] [Open Ended]

You've said you research **unknown product names** outside the Polyglot interface. How do you do this research?

25. [Multiple Selection]

How do you deal with **unknown brand names**? Select all that apply.

- Research in Translation History Lookup whether it needs to be translated/transliterated or left in English
- Leave it in English by default
- Research outside the Polyglot interface whether it needs to be translated/transliterated or to be left in English
- Other (allows input)

26. [If answer to the previous question was "Research outside the Polyglot Interface…"] [Open Ended]

You've said you research **unknown brand names** outside the Polyglot interface. How do you do this research?

27. [Multiple Choice - One Option]

Many texts have different **units of measure** and **symbols**. If there are no client instructions regarding **weight**, **temperature**, **other measures** and **symbols**, how will you edit them? You can only select **ONE** option.

- Keep the original format and values (e.g., 38 °C → 38 °C; 6 ft → 6 ft)
- Convert format and values to the ones used by your target language (e.g., 38 °C → 100.4 °F; 6 ft → 183 cm)
- Other (allows input)

28. [Open Ended]

If you need to convert **weight**, **temperature** and **other measures** to the ones expected in your target language, what tools do you use? Please provide **concrete** examples (**NOT** Google search, Linguee, ...).

[Creative Client only questions]

29. [Open Ended]

Besides the resources provided by Unbabel/Lingo24 (if any), what specialized (**medical, culinary, marketing, …**) terminology websites do you use? Please provide **concrete** examples (**NOT** Google, Linguee, ...).

30. [Multiple Choice - One Option]

Medical texts are usually written in English. Consequently, most medical terminology is translated from English into other languages. However, some translations aren't as known or standardized as others. How do you deal with **uncommon and difficult medical terms**, for which there may already be a translation somewhere? You can only select **ONE** option.

- Do not translate
- Keep the original followed by an explanation between parenthesis
- Explain the term using other words
- Create your own translation by analyzing other similar (and already translated) terms
- Pick one existing translation for the term after doing advanced research
- I skip the task when there are too many of these

31. [Open Ended]

How do you verify if the **existing translation** you've chosen is the correct one in your target language? Please provide a **detailed** description of your process.

32. [Multiple Choice - One Option]

Occasionally you might come across **job titles**. If there are no client instructions, how will you deal with them? You can only select **ONE** option.

- Do not translate ("Chief Technology Officer" → "Chief Technology Officer")
- Keep the original job title followed by an explanation or literal translation between parenthesis
- Use the standard job title(s) in the target language ("Chairman" → "Presidente" [portuguese])
- Other (allows input)

33. [Open Ended]

How do you research the **standard job title(s)** used in your **target language**? Please provide a **detailed** description of your process and, if applicable, include **specific websites** you use (**NOT** Google, Linguee, ...).

34. [Multiple Selection]

Since these clients also have more informal texts, you can sometimes find **slogans** and **wordplay**. How do you edit these? Select all that apply.

- Use Translation History Lookup suggestions
- Keep the original text
- Translate after doing research
- Other (allows input)

35. [If answer to the previous question was "Translate after doing research"]
[Open Ended]

You've said you translate these **slogans** and **wordplay** after doing research. How do you do this research? Please provide a **detailed** description of your process.

36. [Open Ended]

Many of these texts can also contain certain **keywords** or **adjectives** which are currently trendy in certain areas (e.g., **"unleashing", "premium", "high-efficiency", ...**). Where do you look for the closest equivalent word(s) in your target language? Please provide **concrete** examples (**NOT** Google, Linguee, ...).

37. [Open Ended]

How do you verify if the **keywords** and **adjectives** you've found during your research fit the client's style? Please provide a **detailed** description of your process.

[All Participants]

38. [Multiple Selection]

Considering the work you've completed so far, what were the hardest things for you to translate/review?

- Slogans/wordplay
- Acronyms
- Weight/temperature
- Job titles
- Keywords/adjectives
- Medical terminology
- Electrical terminology
- Automotive terminology
- Measures and other units [inches, kiloohm,..]
- Food related words
- Other (allows input)

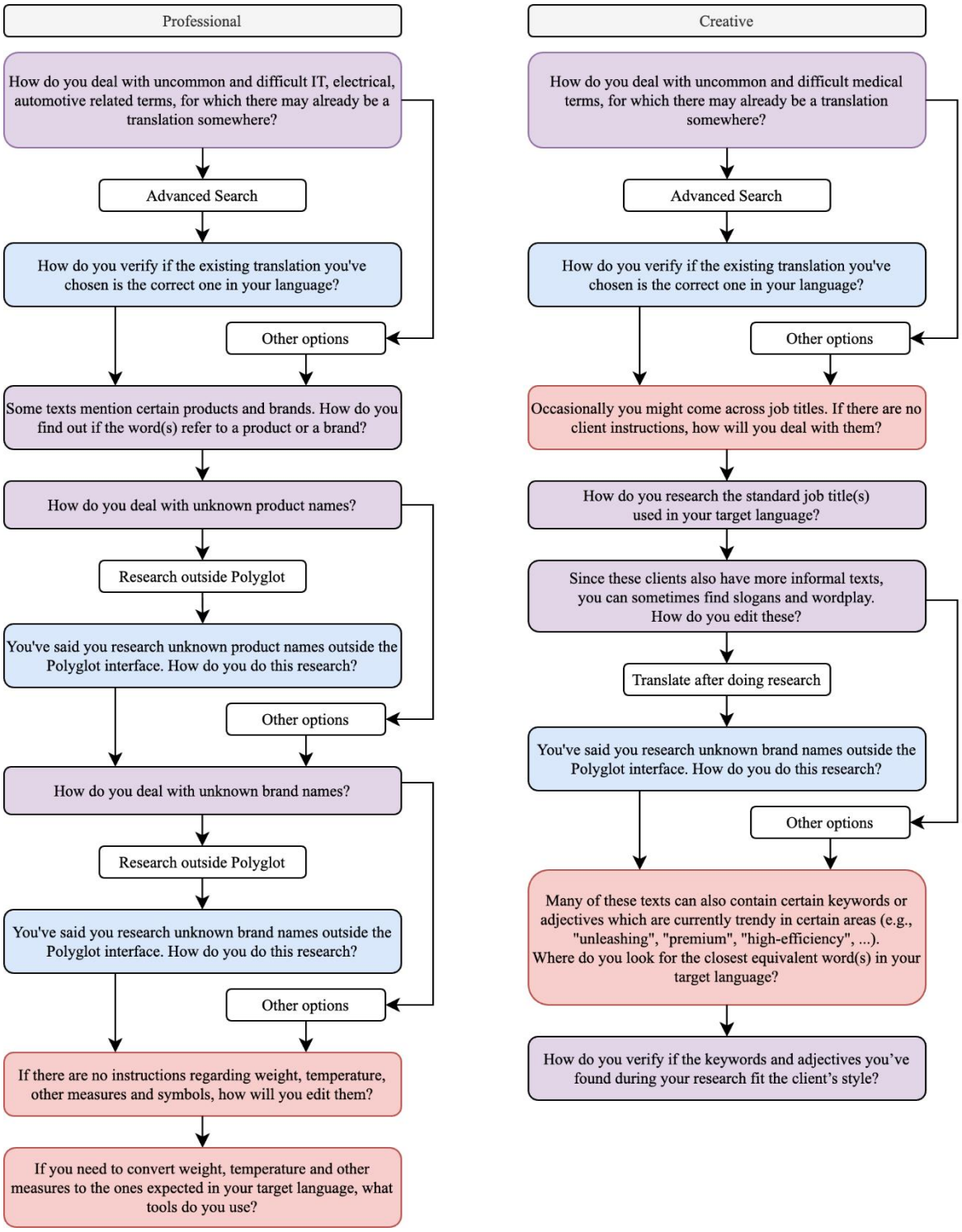
39. [Yes/No Question]

Would you be interested in participating in a **translation research skills training program**?

40. [Open Ended]

Please write your email so we can notify you when the **training program** is available.

Annex C – Questions per client type



Annex D – Research Skills Materials

Knowledge Base Article
Google Search PDF
Research Skills Playlist

Annex E – Research Skills Materials Feedback Survey

Research Skills Materials Feedback Survey

Access them **here**!

1. [Likert question]
   On a scale of 1-5, how helpful were the learning materials?
   (**1** = Not helpful at all, **5** = Extremely helpful)
2. [Multiple Choice - One Option]
   Did you find one format (**Videos** or **PDF**) more helpful than the other?
   If you didn't notice, the PDF is at the end of the article!
   - Videos
   - PDF
   - No preference
3. [Yes/No Question]
   Were the instructions in the learning materials clear and easy to follow?
4. [Yes/No Question]
   Were the learning materials presented in a format that was easy to understand and visually appealing?
5. [Yes/No Question]
   Did the learning materials provide enough practical examples to help you learn?
6. [Open Ended]
   How did these learning materials help you improve your work? Please **specify a situation** where that was the case.
7. [Open Ended]
   What suggestions do you have for improving the learning materials or other things you would like to see? This is the **last question**, please leave any suggestions you have here!