Hermenēus
Revista de traducción e interpretación

EDICIONES Universidad Valladolid

The impact of translator training, experience and extrinsic information on translation evaluation: An eye-tracking study

Impacto de la información extrínseca en la evaluación de una traducción de acuerdo con la formación y experiencia traductoras: estudio de seguimiento ocular

MONTSERRAT BERMÚDEZ BAUSELA
Universidad Nacional de Educación a Distancia (UNED). Facultad de Filología. Departamento de Filologías Extranjeras y sus Lingüísticas. C/ P.º Senda del Rey, 7, 6.ª planta, despacho 600. 28040 Madrid. España.
Dirección de correo electrónico: mbermudez@flog.uned.es
ORCID: https://orcid.org/0000-0002-9578-1318

TABEA DE WILLE
University of Limerick. Department of Computer Science & Information Systems. Computer Science Building, Plassey Park Rd, Sreelane, Limerick. Irlanda.
Dirección de correo electrónico: tabea.dewille@ul.ie
ORCID: https://orcid.org/0000-0003-0248-9228

Recibido/Received: 22/1/2024. Aceptado/Accepted: 29/1/2025.
Cómo citar / How to cite: Bermúdez Bausela, Montserrat y Tabea de Wille, «The impact of translator training, experience and extrinsic information on translation evaluation: An eye-tracking study», *Hermēneus. Revista de Traducción e Interpretación*, 28 (2026): pp. 151-179.
DOI: https://doi.org/10.24197/f16fex28

Artículo de acceso abierto distribuido bajo una Licencia Creative Commons Atribución 4.0 Internacional (CC-BY 4.0). / Open access article under a Creative Commons Attribution 4.0 International License (CC-BY 4.0).

Abstract: In this paper, we describe the results of an eye-tracking study where we have observed the impact of translator training and experience as well as extrinsic information on the translation evaluation process. We presented three different groups of participants (novice, trained and experienced) with English source texts and Spanish translations and asked them to evaluate the translations. We recorded the cognitive effort spent on different information provided through eye-tracking technology. We seek to observe differences in evaluating behaviour and reliance on extrinsic information in these groups of participants with differing levels of expertise and relate measurement knowledge and quality uncertainty. Overall, our results indicate that translator's training and experience tends to add value to the evaluation process by making the assessors less

reliant on extrinsic information, even if it initially raises their measurement uncertainty, which is then overcome by greater effort spent on the intrinsic information.
Keywords: Perceived translation quality; translation training; eye-tracking; cognitive effort; extrinsic information.
Resumen: En este artículo se describen los resultados de un estudio centrado en la relación entre la formación y experiencia traductoras, y la información extrínseca en el proceso de evaluación de una traducción, siguiendo la metodología de seguimiento ocular. Para ello, se mostraron unos textos originales en inglés y diferentes traducciones al español de cada texto a tres grupos de participantes (principiantes, con formación y con experiencia) y se les solicitó que señalaran la que les parecía de mayor calidad. Mediante la grabación del esfuerzo cognitivo que nos permite esta metodología, se han detectado diferencias relevantes en el comportamiento evaluador y en la dependencia de la información extrínseca. Los resultados indican que la formación y experiencia traductoras añaden valor al proceso de evaluación, ya que contribuyen a disminuir la dependencia en la información extrínseca, aunque inicialmente se eleve el nivel de incertidumbre, que después se compensa gracias al mayor esfuerzo empleado en la información intrínseca.
Palabras clave: Percepción de la calidad de una traducción; formación en traducción; seguimiento ocular; esfuerzo cognitivo; factores extrínsecos.
Summary: Introduction; 1. Background, 1. 1. Previous work, 1. 2. (Translator's) expertise, 1. 3. Observing cognitive processes through eye tracking; 2. Methodology and data collection, 2. 1. Overview of setup, 2. 2. Participants, 2. 3. Environment setup, 2. 4. Setup of the stimulus, 2. 5. Instructions to participants; 3. Data cleaning and analysis, 3. 1. Data reduction and AOIs, 3. 2. Exclusion criteria, 3. 3. Analysis; 4. Results, 4. 1. Comparison with previous study, 4. 2. Did participants look at all intrinsic and extrinsic information?, 4.2.1. General observations, 4.2.2. Relationship to chosen scenarios, 4.3. How much cognitive effort did participants expend on different types of information provided?, 4.3.1. Fixation Durations, 4.3.2. Average time viewed, 4.3.3. Fixation duration intrinsic vs extrinsic information, 4.4. Revisits; Conclusions; References.
Sumario: Introducción; 1. Marco teórico, 1. 1. Revisión de la literatura, 1. 2. Competencia traductora, 1. 3. Observación de los procesos cognitivos mediante seguimiento ocular; 2. Metodología y recogida de datos, 2. 1. Descripción general del diseño experimental, 2. 2. Participantes, 2. 3. Configuración del entorno, 2. 4. Configuración del estímulo, 2. 5. Instrucciones para los participantes; 3. Limpieza y análisis de datos, 3. 1. Reducción de datos y AOI (Áreas de Interés), 3. 2. Criterios de exclusión, 3. 3. Análisis de datos; 4. Resultados, 4. 1. Comparación con estudio previo, 4. 2. ¿Prestaron atención los participantes a toda la información intrínseca y extrínseca?, 4.2.1. Observaciones generales, 4.2.2. Relación con los escenarios seleccionados, 4.3. Esfuerzo cognitivo dedicado a los diferentes tipos de información proporcionada, 4.3.1. Duración de las fijaciones, 4.3.2. Tiempo medio de permanencia, 4.3.3. Duración de las fijaciones: información intrínseca versus información extrínseca, 4.4. Segundas visitas; Conclusiones; Referencias bibliográficas.

INTRODUCTION

A fundamental challenge in the localisation and translation industry is the information asymmetry between buyer and seller, and the detrimental effects that quality uncertainty can have on this relationship. An example

can be found in prices decreasing to the point at which low quality products drive good quality products out of the market (Akerlof, 1978).

To manage quality uncertainty and information asymmetry, extrinsic cues commonly support the quality evaluation process, by giving a buyer clues as to whether a product is of high quality. This is especially the case if the customer is not in a position to evaluate the intrinsic attributes of a product sufficiently (Golder et al., 2012). In translation, a common extrinsic cue might be that of the price where a buyer might assume that an expensive translation is of higher quality than a cheaper one, possibly due to extra time spent, additional quality checks or a higher level of expertise, although the link between the external cue of 'price' and observable intrinsic quality attributes is generally weak and unreliable.

The present research examines the impact of extrinsic (rather than intrinsic) quality cues on perceived quality in translated texts. We focus on the professional status of the translator, such as whether they are novice, trained or expert, as extrinsic cues.

An important aspect here is that translation is a relatively weakly protected and institutionalised profession, where anyone who wishes to do so could theoretically call themselves a 'professional translator', regardless of whether they are sufficiently skilled, trained and experienced to provide adequate results (Chesterman, 2001). It is unclear whether a 'professional' translator for example needs to have training or formal qualifications or whether it is sufficient to merely earn a living by doing translations (Paloposki, 2016). This suggests that the importance of different labels that mark the professional status of a translator are subject to interpretation and may or may not be a reliable indicator of the quality that can be expected and hence serve as a suitable extrinsic cue.

With this discussion, we do not intend to imply that there is no difference in the quality that a 'professional/trained/experienced' translator will be able to provide, compared to a 'non-professional/not trained/not experienced' one. Instead, it is our aim to provide insights into how information on the professional status of a translator impacts the way their work is perceived, and to what extent evaluators are then influenced by this information depending on their own professional background. Given these two variables under consideration, we have investigated two related aspects of translator expertise: 1. the attributes that would enable a translator to evaluate translations on their own merit, rather than being overly influenced by extrinsic information; And 2. the signals sent by extrinsic information on the translator's professional status.

1. BACKGROUND

1. 1. Previous work

We have been working for several years on how quality is perceived, and the results of this research have been published fragmentedly depending on the focus of the data collection and analysis. Like this, in De Wille and Bermúdez-Bausela (2020), we focused on translation editing, more particularly, how extrinsic information on the translators' professional status influences the edits introduced by reviewers to translations in proofreading; while in Bermúdez-Bausela and De Wille (2023), we measured the cognitive effort invested by students and professionals in the processing of lexical features applying an eye-tracking methodology. The main aim of the overall study responds to the need to see how quality is perceived among different groups of participants (novice, trained and expert) given some extrinsic information, and how this may influence more or less their choice of translated version and the number and nature of edits introduced to the translation of their choice.

For the purposes of this paper, we would like to refer more in-depth to De Wille and Bermúdez-Bausela (2018), where we examined the effects of the extrinsic cues of the translator's professional background on the way participants with different levels of measurement knowledge perceived translation quality. We divided participants into novice, trained and expert evaluators according to anticipated measurement knowledge to observe how strongly those groups were influenced by extrinsic cues relative to other participant groups.

Measurement knowledge refers to the individual's ability to assess attribute performance with minimal bias or variance compared to more objective measures (Golder et al., 2012). While in real-life scenarios, measurement motivation (the desire to assess attribute performance, depending on level of interest or time to evaluate a product) comes into play (Golder et al., 2012), we did not evaluate this aspect as study participation was voluntary and we therefore assumed a more or less equal level of measurement motivation among participants.

We found that quality evaluators with greater measurement knowledge were influenced less strongly by extrinsic cues than those with less measurement knowledge and that the extrinsic cue of professional status had a general strong influence on participants to differing degrees.

Expert evaluators trended towards selecting expert translation and showed a high level of agreement within the group on which translation was the best. They generally relied less on extrinsic cues than the other participant groups.

Trained (but inexperienced) evaluators were situated in the mid-range in terms of reliance on extrinsic cues. They selected the expert translator less frequently, and they had the smallest increase when extrinsic cues and intrinsic attributes matched. This suggests that while their translation studies had equipped them with a greater ability to evaluate quality, compared to novice participants, there was a greater element of disagreement. This may indicate that certainty in evaluation is linked to experience levels.

Novice evaluators were heavily influenced by extrinsic cues. Furthermore, levels of agreement were lowest for this group.

1. 2. (Translator's) expertise

As we investigate how the level of evaluators' expertise impacts their perception of quality, it is relevant to consider what translation expertise consists of and how we might be able to observe it through eye tracking. We have grouped our participants into 'novice', 'trained' and 'expert' based on their progression through a translation programme and the years spent working in a translation-related field after graduation. The question as to what distinguishing factors we can expect between these groups is examined in the following sections.

The concept of 'Translation Competence' is frequently discussed in literature, for example, Beeby et al. (2000) reference several authors in their review of the literature. Furthermore, several models of competence have been proposed, such as PACTE (2000, 2003) and Kelly (2002, 2007).

However, those perspectives on translation competence primarily focus on the visible output of competence rather than how processes might differ between levels of competence. A more relevant perspective for the purpose of our study is that of the different types of knowledge and how they relate to different stages of expertise. Beeby et al. (2000) outline declarative vs operative knowledge, where declarative knowledge is easy to verbalise, explicitly defined and processing essentially controlled. Operative knowledge, however, is difficult to verbalise, can be partially processed, is acquired gradually and through practice and processing is essentially automatic. Both may coexist or declarative knowledge may

give way to operative knowledge in the process of a novice becoming an expert (Beeby et al., 2000). This development from the novice to expert stage is, according to Dreyfus (2004), marked by a range of emotions in the intermediate stages, especially that of the competent learner (which is the third stage out of five). While the novice learner is provided with rules and examples, the "learning can be carried out in a detached, analytic frame of mind as the student follows instructions and is given examples" (Dreyfus, 2004, p. 177). However, with increased experience, the learner can become overwhelmed and "performance becomes nerve-wracking and exhausting" (Dreyfus, 2004, p. 178) as they are able to recognize an increasing number of potentially relevant elements and procedures but are lacking a sense of what is important for the situation they are in. While the proficient performer sees what needs to be done and then decides how to do it, the expert not only perceives what needs to be achieved but also immediately how to achieve it, due to a "vast repertoire of situational discriminations" (Dreyfus, 2004, p. 180).

Differences in how groups with different levels of expertise manage translation production have been noted by Dragsted (2010), Jakobsen and Hvelplund (2008), Carl and Kay (2011), Sharmin et al. (2008), and Göpferich and Jääskeläinen (2009). In relation to identifying errors, Schaeffer et al. (2019) observed in their eye-tracking study that "translation students correct significantly fewer errors as compared to professional translators" (Schaeffer et al. 2019, p. 599) and that "students take longer to correct errors as compared to professional translators" (Schaeffer et al., 2019, p. 600). They also conclude that professional translators are predictably more efficient in error recognition and correction than students, due to being more strategic in their reading behaviour than students.

In the process of translation training, assessment abilities will have been acquired in the middle and the last stages, thanks to experience and having absorbed all the competences in translation. Chesterman (2016) is of the opinion that contrastive analysis becomes relevant at this stage as it helps students take an analytical view on the source text and the target text. Having cultivated the required skills, their judgements will be more reliable. In this light, we argue that conscious engagement translates into cognitive effort: the more conscious engagement, the more the cognitive effort. This cognitive effort can then be visualised and measured using eye-tracking methodology.

1. 3. Observing cognitive processes through eye tracking

While our previous work provided important insights into the end-result of quality evaluation, it did not answer the question of how different participant groups conducted the quality evaluation, or how the extrinsic information provided impacted this process. In the present study, we used eye-tracking technology to observe whether extrinsic cues impacted evaluation processes in novice, trained and expert evaluators.

In the last few years, eye tracking has gained support in the field of Translation Studies as a research methodology, particularly in a diverse range of process-oriented translation studies. Examples that are especially relevant to our research include Doherty et al. (2010), who use eye tracking as a Machine Translation evaluation technique with positive results; Pavlović and Hvelplund (2009), who investigate directionality in translation processes, conducting an experiment that involves student and professional subjects. Jakobsen and Hvelplund (2008) carried out an experiment using eye-tracking methodology in which novice and professional translators are presented with translation-oriented tasks, and the cognitive effort employed by each group of participants in terms of gaze duration is measured. One of their findings suggests that professional translators dedicate more time to the translation process and its revision than to the source text. In contrast, students tend to invest more time in grasping the intricacies of the source text. In a study conducted by Dragsted and Carl (2013), where they explored common features among students and professional translators, they also studied individual behavioural traits and potential categorizations based on these characteristics.

Myriad studies have been carried out along these lines and this section does not seek to be a comprehensive collection of them. For a deeper review of eye-tracking methodology applied to process-oriented translation research, we recommend the overviews carried out by Walker (2021) and Hvelplund (2017).

2. METHODOLOGY AND DATA COLLECTION

Eye tracking allows the researcher to measure the cognitive load (or effort) invested by the translator in the process of translating through data points such as the amount of time spent on a particular word or a whole Area of Interest (AOI), the shifts and sequences of eye movements from

one AOI to another, the revisits to an AOI, the first fixation points in the processing of a text and other related eye movements. Eye tracking helps to investigate different aspects of cognitive processing, in our case, determining the cognitive effort spent by our participants in the established AOIs (source text fragments and their corresponding translated versions) and the influence that extrinsic information (which we also defined as AOIs) might exert on the process.

Deriving cognitive load from eye movement and fixations is based on "the eye-mind assumption", which states that "the eye remains fixated on a word as long as the word is being processed. So, the time it takes to process a newly fixated word is directly indicated by the gaze duration" (Just and Carpenter, 1980, p. 330). That is, there is a correlation between what the eye is looking at and the amount of cognitive effort devoted to its processing (Schäffner and Shuttleworth, 2013).

Eye tracking has enhanced our understanding of how a particular individual approaches and processes a specific translation task as it allows us to measure eye movements and fixation points. However, this methodology is not without its challenges and obstacles, which we discuss in the following sections.

2. 1. Overview of setup

Participants were assigned to three groups based on their level of training and experience. Each participant was presented with the same MS PowerPoint slides on a screen. The slides comprised the source text (ST), three translations or target texts (TT) and, in certain instances, information (correct or incorrect) on who had done the translation. We refer to this information as 'extrinsic'. Participants were asked to select which of the three translations they thought was best and state their preference verbally. One of the researchers in the room documented their preference and their eye movements were recorded using eye-tracking technology.

2. 2. Participants

Students and lecturers at the University Alfonso X El Sabio in Madrid were invited to participate in this study. The reason behind the choice of institution is that one of the researchers of this study was a lecturer at the time in this University and both the circumstances and the context were the optimal ones to conduct the study here.

We would like to point out that even though this group of participants was different from the group of participants that took part in the first study conducted on perceived quality and translation editing in this same University, there was some overlap due to the fact that some participants were, during the first experiment, at the beginning of their degree, and in this second experiment, they were finishing their degree. Also, among the group of expert evaluators there were some lecturers that had conducted both experiments. However, due to the long period between the studies and the difference in tasks participants were asked to perform, we do not believe that factors such as fatigue and lack of motivation impacted the results.

Participants volunteered their time and did not receive compensation or rewards for it. To initiate the data collection process, there was a pilot phase involving two participants. The data collected in this phase were excluded from the analysis. This allowed the researchers to acquaint themselves with the setup and to resolve issues related to the physical space and the instructions for participants.

The twenty-four participants included in data collection and analysis were grouped into three categories, based on their level of training received at that point and their working experience (Table 1).

Participant grouping	**Number**	**Percentage**
Novice evaluators (1^{st} /2^{nd} year students)	12 (5 first year, 7 second year)	50%
Trained evaluators (3^{rd} /4^{th} year students)	6 (4 third year, 2 fourth year)	25%
Expert evaluators (lecturers)	6	25%

Table 1. Participant groups.

Students in years one and two of their studies were considered novice evaluators as their programme did not cover translation or evaluation of translation in its first year. This meant that at the time of data collection, participants would have not yet received extensive training in translation. The only subject related to translation training as part of their first-year

curriculum is English for future translation trainees. In their second year, students do start having some translation-related courses.

The trained evaluators were third and fourth-year students in the Translation and Interpreting degree programme. During these academic years, students would have already received in-depth training in translation and covered a range of relevant subjects but would not have had a lot of real-life translation experience.

It is worthy of note that in their third year, students have a specific revision course that provides the necessary skills to conduct a linguistic and textual analysis applied to text revision as an integral part of the translation process. Students focus on the difference between revision and correction, types of errors in translation, the revision process and industry quality standards.

The expert evaluators were lecturers at the University Alfonso X El Sabio in Madrid. Regarding their formal training, 60% of them have an undergraduate degree and/or a doctorate in English Philology (and one participant in German), while the remaining 40% hold an undergraduate degree and/or a doctorate in Translation and Interpreting Studies.

In Spain, before the 90s, a degree in Translation and Interpreting was not very common, and translators from that period usually hold degrees in Philology and have afterwards acquired specific competences in Translation through a master or doctoral studies, as well as hands-on experience. Regarding our group of experts, 80% of them have more than twenty years of experience teaching translation at the university level as well as over twenty years working as professional translators.

2. 3. Environment setup

Participants carried out the task individually with only the two researchers present in the room with them. They were asked to sit at a desk previously set up with a computer monitor elevated so that they could comfortably read the text on the screen. The room was well lit, with a combination of natural daylight and artificial light. One of the researchers gave instructions and assistance to the participants, while the other operated the eye-tracking software and monitored the data collection. Although using an eye-tracking bar rather than head-mounted eye-tracking hardware may lower accuracy of the measurements (O'Brien, 2009), we opted for an eye-tracking bar placed below the screen that the participants

were viewing to improve ecological validity as this technology is less intrusive for participants. We used eye-tracking technology by Gazepoint, along with its associated software for data recording and initial analysis. We used the built-in calibration procedure for the system and confirmed successful calibration with a test procedure on our recording screen.

However, in certain instances, calibration proved challenging despite several attempts, leading to the exclusion of those participants from the study. The information on the number of participants provided in Table 1 already excludes them.

2. 4. Setup of the stimulus

For this study, we used a selection of texts created and translated for previous research described in De Wille and Bermudez Bausela (2018) and De Wille (2017).

The specific source text used for our study was written by one of this paper's authors and proofread by a native speaker with training and experience in editing texts. Among its features we can point out that it had not been previously translated; it included thirteen paragraphs consisting of two to three sentences each, with individual headings (between two and thirteen words each) trying to avoid issues with missing references, context or consistency that may arise in longer paragraphs, and it dealt with a general subject matter that did not require high levels of specialisation.

The text was then independently translated by six Spanish translation students and lecturers (two from each group). The groups were:

- Novice translators: first-year translation students who had not yet received training in translation.
- Trained translators: fourth-year translation students who had received in-depth training in translation but had not yet gained real-**life translation experience.**
- **Expert** translators: lecturers at the University Alfonso X El Sabio in Madrid. Both experts had many (25 and 15 respectively) years of teaching experience in the translation degree programme and were trained and experienced in judging language quality.

All translators were told that their translations would be used for a study, thought the specific nature of the study was not disclosed. The translations were screened (but not edited) by the researchers to ensure that the translations were mostly complete and did not include additional text such as translator comments. For this eye-tracking study, individual sentences were laid out on MS PowerPoint slides. Text on the individual slides was limited to one single ST sentence per slide and three translations. Each translation had been done by a different translator as described above. Text was laid out in large font sizes (between twenty-seven and forty-four px) to increase the accuracy of the eye-tracking data. The initial slides did not include extrinsic information on the translator. They were followed by slides that included the correct information on the translator's professional status. We refer to this extrinsic information as GoodInfo henceforth. GoodInfo was positioned above the target texts and included the phrases 'Novice Translator', 'Trained Translator' and 'Trained + Experienced Translator' without further explanation on those phrases (Figure 1). The third set of slides also included extrinsic information, but the extrinsic information provided did not match the professional status of the translator who had done the translation. Thus, a novice translator's TT might be labelled as 'Trained Translator' and that of the expert translator as 'Novice Translator', etc. BadInfo was included to observe whether participants' behaviour was primarily influenced by the extrinsic information or the intrinsic quality of the translations.

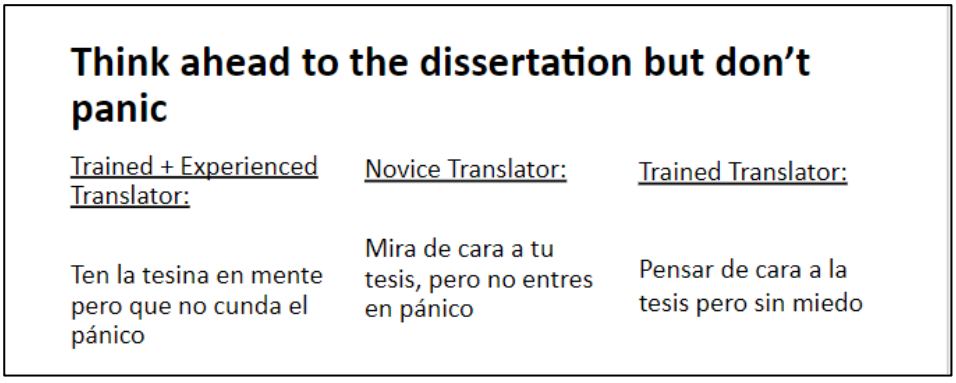


Figure 1: Sample of slides shown to participants.

2. 5. Instructions to participants

Participants were asked to choose which of the TT they thought was best out of the three options. Participants stated their selection verbally, the researchers recorded their choices and participants were then shown the next slide. Participants did not receive instructions on how to evaluate the translations or whether to consider the extrinsic information provided or not.

3. DATA CLEANING AND ANALYSIS

3. 1. Data reduction and AOIs

A major challenge when analysing eye-tracking data is the large volume of data points and potential approaches to analysis which need to be related to the overall research goals and evaluated for validity. With this in mind, we limited our analysis to two slides per type of scenario (NoInfo, GoodInfo, BadInfo). We then manually defined Areas of Interest (AOIs) for each slide. AOIs were placed as rectangular fields covering the text itself and a border around the text, reaching no further than the border of any adjacent AOI. Each type of text (ST, TT1, TT2, TT3, Extrinsic 1, Extrinsic 2, Extrinsic 3) received its own AOI field for each slide.

3. 2. Exclusion criteria

One participant initially wore contact lenses when attending the study, leading to problems calibrating the device. Once the participant replaced their contact lenses with spectacles, calibration and tracking were possible without further issues.

Other issues we observed involved one participant with a migraine headache, which led to the device only being able to calibrate one eye correctly, and one participant with very long eyelashes, which caused similar issues. Whenever calibration was not possible, the participants were excluded from the study.

During data analysis, we observed that one participant had looked away from the screen to an excessive degree for some of the slides, which could be seen in the video recording of their eyes and in the absence of fixation points for large durations of the recording. We therefore excluded the data collected for those slides from analysis for this case.

3. 3. Analysis

Based on previously defined AOIs, we used Gazepoint software to generate data from the recordings that indicated what AOIs individual participants fixated on, for how long and for how often. We then collapsed this data into participant and AOI categories to be able to evaluate general trends, which we describe in the following section.

4. RESULTS

4. 1. Comparison with previous study

We had previously conducted a similar study (without the eye-tracking component), using a similar setup and participant groups (De Wille and Bermúdez-Bausela, 2018; De Wille, 2017). The following compares the results of this current study with those from the previous one. Our expectation was that the results of this study would match those results. Analysis steps included were:

Frequency analysis for selection of each translation by participant group, cross-tabulated with information on the given translators' status (NoInfo, GoodInfo, BadInfo).

Figure 2 shows the mean values for selection per translator and participant group. Figure 3 shows the same data for our previous study (De Wille and Bermúdez-Bausela, (2018).

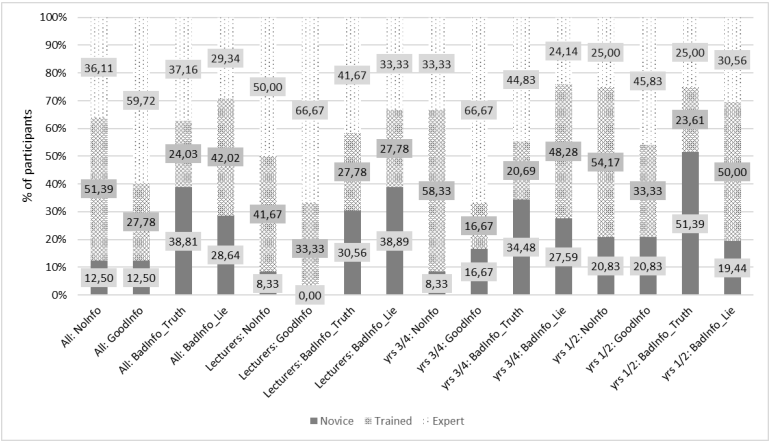


Figure 2. Results of current study, mean values for selection per translator and participant group.

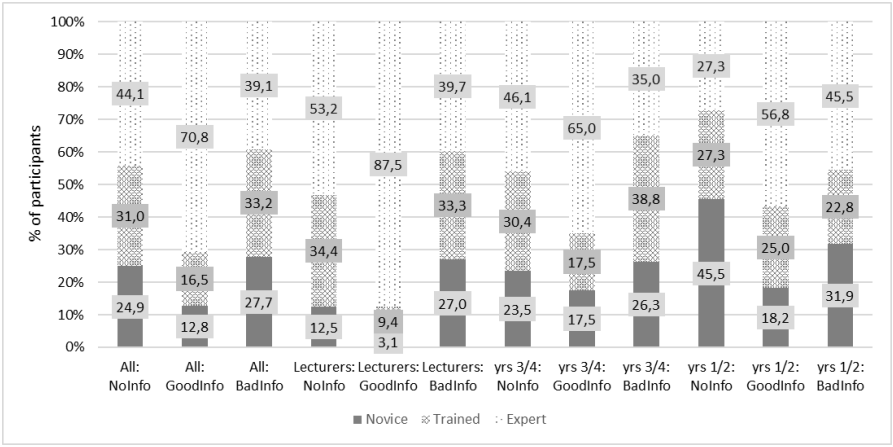


Figure 3. Results of previous study (De Wille and Bermudez Bausela, 2018), mean values for selection per translator and participant group.

Throughout the comparison of the current study with the previous ones, it became clear that this group of evaluators preferred the translations provided by the trained translators (in NoInfo, GoodInfo and BadInfo scenarios) more often than in the previous study, despite having re-used the translations from that study for the present one.

This was also evident in the NoInfo scenarios, where no information on the professional status of the translator was provided. Here, the expert translator was selected by 36.1% of the participants, with the translations provided by trained translators selected most frequently in general (51.39%). This is in contrast with the previous study where the expert had been selected more frequently than the other two groups (44.1%).

Across all participant groups, evaluators favoured either the trained translator's version (trained evaluators) or the expert translator's version (expert evaluators). The novice translator's versions were selected substantially less often than the other groups (12.5% across all participants) when no extrinsic information was provided.

Given correct information on the professional status of the translator (GoodInfo scenarios), the expert translation was selected more frequently (mean: 59.7%) and the novice translation least frequently (mean: 12.5%) across all participant groups, which matches the observations in our previous study.

The patterns for the individual groups were also similar to that of the previous study, with expert evaluators choosing expert translations more frequently in GoodInfo scenarios compared to NoInfo. However, the increase was not as pronounced, with 33.33% of participants still choosing the trained translator. Trained participants also chose the expert translator in GoodInfo scenarios more often than in NoInfo scenarios, as did Novice evaluators. However, Novice evaluators did so to a lesser degree than expert and trained evaluators.

When incorrect information on the professional status of the translator was provided (BadInfo scenarios), the mean expert selection dropped to 29.3% from 36.1% in NoInfo scenarios, which follows the trend from our previous study.

As with our previous study, expert and trained participants selected what was actually the expert translation (BadInfo_Truth) more often than what was ostensibly the expert translation (BadInfo_Lie). The only group that picked expert translations in BadInfo_Lie over BadInfo_Truth more often were the novice participants.

Furthermore, in the present study, the trained and novice evaluators picked the 'so-called' trained translation most frequently (BadInfo_Lie), while in the case of novice evaluators this resulted in them picking what was the novice translator's versions most frequently (BadInfo_Truth).

Overall, we observe a slightly stronger preference for the trained translator's version over that of the expert, in contrast to the previous study, while the novice translator's versions are still least frequently favoured, when no information on professional status is provided. When extrinsic information reflects who has done the translation, preference for expert translation increased, but did so least in the Novice participant group relative to trained and expert evaluators. When extrinsic information did not reflect reality, trained and expert evaluators seemed to be more often guided by intrinsic information, while novice evaluators were more strongly influenced by extrinsic information.

However, by merely observing selections of different participant groups we are unable to confirm whether those groups were indeed influenced by the extrinsic information, or whether the differences in selections are due to differences in preference related to the translations themselves. To gain a more holistic picture of participant behaviour, we first analyse whether different groups looked at all the intrinsic and extrinsic information provided.

4. 2. Did participants look at all intrinsic and extrinsic information?

4.2.1. General observations

We found that across all groups, participants looked at the source text in all cases, and in most cases all the target texts. Trained and expert evaluators ignored target texts more frequently than the novice evaluators, and they did so to a similar degree in GoodInfo and BadInfo scenarios. Extrinsic information was ignored most frequently in all participant groups, more often in BadInfo scenarios (nearly or more than half of participants) than GoodInfo ones (up to nearly a quarter of participants) (Tables 2 and 3).

In GoodInfo scenarios, expert participants were most likely to ignore extrinsic information provided, with 60% (3) of the experts ignoring AOIs Extrinsic 2 and Extrinsic 3 for scenario 3, and a third ignoring Extrinsic 2 in scenario 4. A smaller percentage of novice and trained evaluators ignored extrinsic information in scenarios 3 and 4.

In BadInfo scenarios, between a third and more than two-thirds of trained evaluators ignored extrinsic information. The number of expert and novice evaluators who ignored extrinsic information also increased, although not as much.

In summary, across all groups, evaluators generally viewed source text and the three target texts. Extrinsic information was ignored more frequently in BadInfo scenarios, and a higher percentage of trained and expert evaluators ignored extrinsic information more than novice evaluators did.

Slide	Group	ST	TT1	TT2	TT3	Extrinsic1	Extrinsic2	Extrinsic3
S3	**Novice**	0	0	0	0	0	0	8,3
S3	**Trained**	0	0	0	0	0	16,7	0
S3	**Expert**	0	0	20	20	0	60	60
S3	**All**	0	0	4,3	4,3	0	21,7	21,7
S4	**Novice**	0	0	0	8,3	8,3	25	16,7
S4	**Trained**	0	0	0	0	16,7	16,7	0
S4	**Expert**	0	0	0	0	16,7	33,3	16,7
S4	**All**	0	0	0	4,2	12,5	25	12,5

Table 2: Percentages of participants that did not look at a particular AOI, GoodInfo. Percentages are calculated within each participant group.

Note that S3 only included five experts as one participant's data was excluded due to being invalid. This means that for S3 the total number of participants was only twenty-three, whereas it was twenty-four for the other scenarios.

Slide	Group	ST	TT1	TT2	TT3	Extrinsic 1	Extrinsic 2	Extrinsic 3
S5	**Novice**	0	0	0	0	25	25	25
S5	**Trained**	0	0	16,7	0	33,3	66,7	83,3
S5	**Expert**	0	0	0	0	16,7	50	50
S5	**All**	0	0	4,2	0	25	41,7	45,8
S6	**Novice**	0	0	0	0	8,3	8,3	50
S6	**Trained**	0	16,7	0	0	33,3	50	66,7
S6	**Expert**	0	0	0	0	16,7	16,7	66,7
S6	**All**	0	4,2	0	0	16,7	20,8	58,3

Table 3. Percentages of participants that did not look at a particular AOI, BadInfo. Percentages are calculated within each participant group.

4.2.2. Relationship to chosen scenarios

Only one participant selected a translation they had not looked at. This was a fourth-year student in a BadInfo scenario. In all other instances, participants had looked at the translations they ended up selecting as the best translation.

In 19 cases (out of 144, 13.4%), participants chose a scenario without looking at the extrinsic information. In most cases (15) this occurred in BadInfo scenarios, with only in 4 in GoodInfo scenarios.

Seven of those cases were novice evaluators, with two participants making a choice without having looked at extrinsic information twice, and three participants once each.

Eight of the cases were trained evaluators who made a choice without having looked at extrinsic information, which one participant did three times, two participants twice and one participant only once.

Among the experts, in four cases a choice was made without having looked at extrinsic information. One expert did so three times and 1 only once.

It is important to remember that there were twice as many novice participants as trained or expert ones. Adjusting for this, expert and novice evaluators made a choice without looking at extrinsic information to a similar extent, while trained evaluators were twice as likely to make a choice without having viewed the extrinsic information.

4.3. How much cognitive effort did participants expend on different types of information provided?

As discussed previously, we can measure cognitive effort by measuring the time participants fixated on a particular piece of information. With this in mind, we present fixation durations for intrinsic and extrinsic information, and the relationship between fixation duration and scenarios chosen in the following sections.

4.3.1. Fixation Durations

Looking at fixation durations overall, we expected expert evaluators to spend the least amount of cognitive effort, followed by novice evaluators. We also expected trained evaluators to have the longest fixation durations in line with Dreyfus (2004).

4.3.2. Average time viewed

Looking at average times across all participant groups, average duration decreased linearly from NoInfo slides to BadInfo slides. This might appear surprising as GoodInfo and BadInfo slides contained more information to evaluate due to the extrinsic information being included. However, NoInfo slides were shown before GoodInfo slides, which were followed by BadInfo ones, which meant that fatigue, dropping interest levels and/or greater familiarity with the setup likely played a role in average time spent. We therefore evaluate individual groups in relation to one other for each of the types of scenarios and do not consider the time spent for the different types of scenarios.

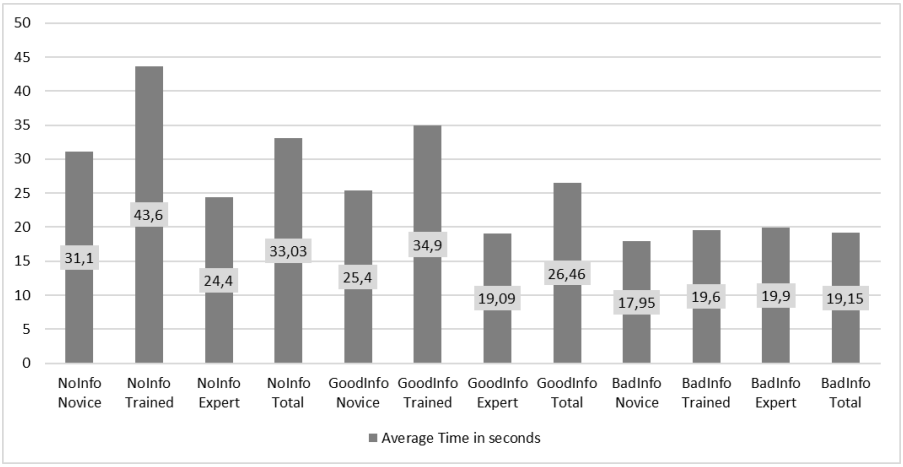


Figure 4. Average time viewed in seconds per respondent group, broken down by NoInfo, GoodInfo and BadInfo scenario.

In NoInfo and GoodInfo scenarios, expert evaluators spent the least amount of time assessing each slide, as expected, and a near equal amount of time as trained evaluators in BadInfo scenarios. Trained evaluators spent more time in evaluating NoInfo and GoodInfo scenarios than any other group, and novice evaluators ranked in between expert and trained evaluators. In BadInfo scenarios, the average time spent is similar for all participant groups with negligible differences.

The same pattern regarding duration spent on evaluations for each participant group can be observed when factoring in the differences in word length for each scenario, by considering average time spent per word. We also observed that when considering the per word duration, there is a slight increase in time spent for each individual group and across all groups between NoInfo and GoodInfo scenarios and only a decrease in time for BadInfo scenarios.

4.3.3. Fixation duration intrinsic vs extrinsic information

The following section breaks down the fixation durations to observe on what type of information (source text, target texts or extrinsic information) the evaluators spent most cognitive effort. Our expectation was that experts would fixate less on the AOI containing the extrinsic information than trained evaluators, and trained evaluators would do so

less than novice evaluators. This would indicate a reverse correlation between level of expertise and dependence on extrinsic information for quality assessment.

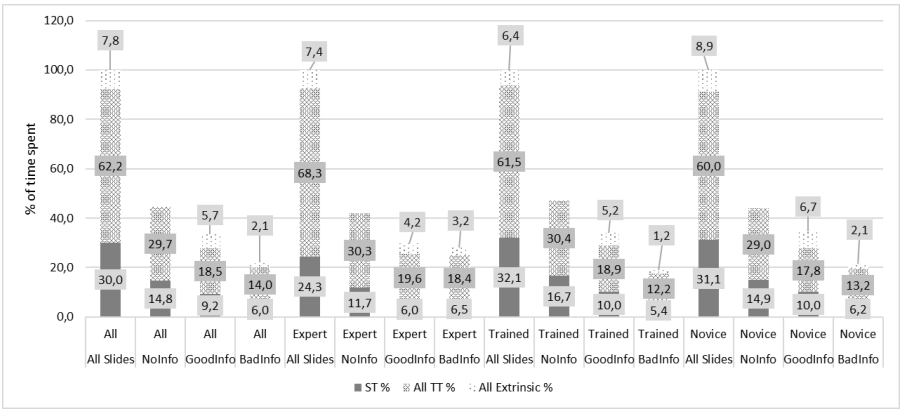


Figure 5. Percentage of time spent on each AOI relative to overall time spent, per group.

As trained evaluators spent more time on all AOIs than any other group except in one instance (extrinsic information in BadInfo scenarios), we related the time spent on each AOI to the overall time spent on evaluating for each of the groups.

Trained evaluators spent the least amount of time on extrinsic information in total, followed by experts and then novice evaluators. In GoodInfo scenarios, the progression is linear with experts at the bottom, followed by trained and then novice evaluators. However, in BadInfo scenarios, trained evaluators spent the least amount of time on extrinsic information, followed by novice and then expert evaluators at the top.

4.4. Revisits

We use the number of revisits to measure uncertainty management, where high numbers of revisits relative to the revisits of other participant groups indicate higher levels of uncertainty.

Generally, we note that participants revisited extrinsic information least frequently, followed by the ST and most frequently the combined TTs. However, each group revisited ST more frequently than each individual TT in all instances with some exceptions in the expert group

(Figure 6). In scenario 1, experts revisited TT2 more frequently than the ST. In scenario 3, they revisited TT1 and Extrinsic 1 more frequently than the ST. In scenario 4, they revisited TT2 more frequently than ST and TT3 nearly as frequently as ST (Figure 7). Altogether, experts revisited the TTs least frequently, followed by novice evaluators in the middle, and trained evaluators with the highest number of revisits (Figure 6).

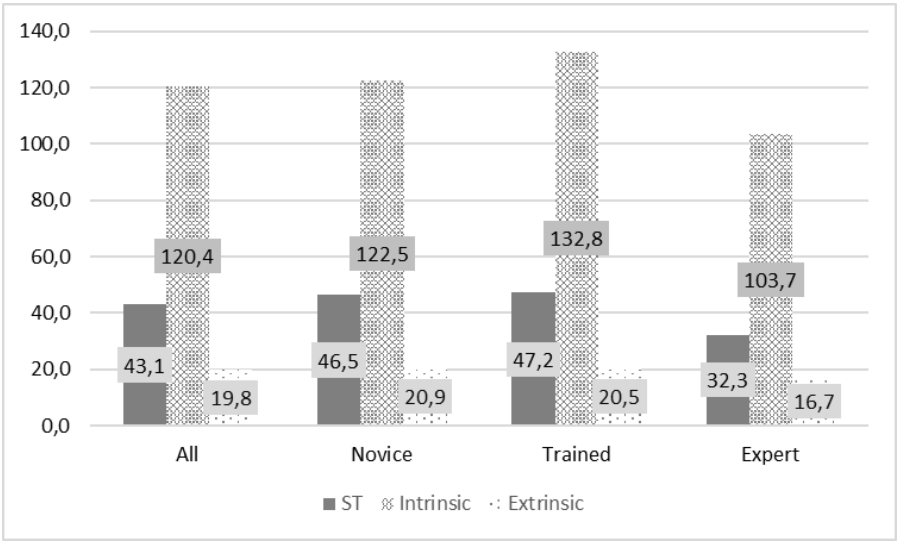


Figure 6. Total of revisits for types of AOI by participant group.

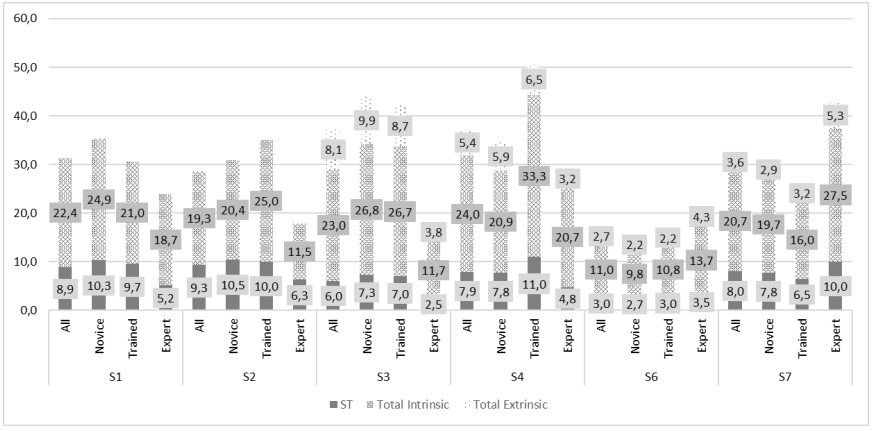


Figure 7: Revisits by type of AOI for each participant group, broken down by scenario.

CONCLUSIONS

In this study, we have investigated two related aspects of translator expertise: 1. the attributes that would enable a translator to evaluate translations on their own merit, rather than being overly influenced by extrinsic information; and 2. the signals sent by extrinsic information on the translator's professional status.

We have sought to observe differences in evaluating behaviour and reliance on extrinsic information in participant groups with differing levels of expertise by measuring cognitive effort through eye-tracking technology.

As we have previously discussed, uncertainty management has been suggested as a significant factor in the behaviour of evaluators, and how they will evaluate texts. In particular, we have imbalanced measurement knowledge and quality uncertainty suggested as a factor in literature that gets mitigated by evaluating extrinsic information. As a second aspect, Dreyfus (2004) outlined the stages of expertise and saw the intermediate stages of expertise as having to deal most with uncertainty as the learner knows that they do not have the knowledge and need to mitigate that.

Reliance on extrinsic aspects: Our study results indicate that participant groups differed in their reliance on extrinsic information depending on their levels of expertise and depending on whether the extrinsic information matched the intrinsic quality parameters of the translation. However, this observation is limited by the fact that there is not necessarily a direct correlation between the training/experience level of the translator and the quality of the translation itself.

When considering Dreyfus (2004), we assumed that our trained evaluators would have higher levels of uncertainty than both expert and novice evaluators. We expected novice evaluators to have higher levels of uncertainty than experts, but lower than trained ones. Interestingly, we did not observe trained evaluators managing this uncertainty through heavy reliance on extrinsic information. We found that of all the three participant groups, trained evaluators were most likely to ignore extrinsic information completely. They also spent the least amount of time on extrinsic information that they did look at, compared to the other two groups. Instead of heavier reliance on extrinsic information, we observed that trained evaluators spent the most amount of time on the evaluations and had the highest number of revisits of TTs relative to the other groups.

By contrast, novice evaluators were less likely than the other two groups to ignore extrinsic information entirely and made a choice without having looked at the respective extrinsic information to a similar extent as expert evaluators did. However, they also spent the least amount of time viewing extrinsic information compared to the other groups. Novice evaluators ignored TTs less frequently than the other groups and ranked between expert and trained evaluators in overall time spent evaluating the texts. Novice evaluators also ranked in the middle for revisits.

Expert evaluators were more likely to ignore extrinsic information than novice evaluators and did so to a similar extent as trained evaluators. Overall, they spent the least amount of time evaluating the scenarios, but they did rank between novice evaluators at the top and trained evaluators at the bottom when it came to the time spent on extrinsic information. Expert evaluators also had the lowest number of revisits among all three groups.

All together, these observations suggest that, as expected, based on Dreyfus (2004), there were high levels of uncertainty in the trained evaluator group, but they were not primarily mitigated through reliance on extrinsic information, but instead by taking longer to evaluate the translations and revisiting texts more frequently. Expert evaluators showed low levels of uncertainty, which could be seen in their overall lowest time spent on the evaluations and the lowest number of revisits. While they did consider extrinsic information more than trained evaluators did, they appeared to be less influenced by the extrinsic information, especially when the extrinsic information did not match their perception of intrinsic quality. Anecdotally, this was also confirmed by expert participants, who commented that they ignored the extrinsic information when they felt that it did not match their perception of the quality. All in all, this suggests that expert evaluators will consider all information points provided to them but will not be influenced by it as strongly if it is not coherent. We believe that this pattern of considering but then dismissing information provided is another indicator of low levels of uncertainty and suggests that a major difference between a trained evaluator and the expert evaluator is that the expert evaluator has, through experience, overcome uncertainty and gained confidence in their ability to evaluate translations. On the other hand, while a trained evaluator could be interpreted as showing more diligence through increased time spent on evaluations and number of revisits relative to the other groups, higher levels of uncertainty also make

them slower, albeit less prone to being influenced by extrinsic information than novice evaluators.

In general, while on a small scale and in need of further confirmation through future studies, our results indicate that translator training tends to add value to the evaluation process by making evaluators less reliant on extrinsic information, even if it initially raises their measurement uncertainty, which is then overcome by greater effort spent on the intrinsic information. Once a trained translator moves into the expert stage where training and experience are combined, measurement uncertainty is decreased again and evaluators gain confidence, as well as the ability to consider and dismiss information as appropriate or not. However, based on our results it is unclear whether this effect can only be observed in individuals with training and experience combined, or whether experience without formal training would be sufficient to achieve similar results.

References

Akerlof, George A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* 48 (3), 488-500. https://doi.org/10.2307/1879431.

Beeby, Allison, Ensinger, Doris, and Presas, Marisa (2000). Acquiring translation competence: Hypotheses and methodological problems of a research project. In Allison Beeby, Doris Ensingerand Marisa Presas (Eds.), *Investigating Translation* (pp. 99-106). John Benjamins.

Bermúdez-Bausela, Montserrat and De Wille, Tabea (2023). Estudio del esfuerzo cognitivo en el procesamiento de características léxicas en estudiantes y expertos utilizando una metodología de seguimiento de jos. *Hikma* 22 (1), 219-248. https://doi.org/10.21071/hikma.v22i1.15063

Carl, Michael and Kay, Martin (2011) Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta* 56 (4), 952-975. https://doi.org/10.7202/1011262ar

Chesterman, Andrew (2001). Proposal for a Hieronymic oath. *The Translator* 7 (2), 139-154. https://doi.org/10.1080/13556509.2001.10799097

Chesterman, Andrew (2016). *Memes of Translation: The Spread of Ideas in Translation Theory*. John Benjamins.

De Wille, Tabea (2017). *A Mixed Methods Investigation on Perceived Quality in Crowdsourced Localisation* [Unpublished doctoral dissertation]. University of Limerick. (Supervised by Chris Exton and Reinhard Schäler).

De Wille, Tabea and Bermúdez-Bausela, Montserrat (2018). Quality perceptions and professional status in translation. *Clina: Revista Interdisciplinaria de Traducción, Interpretación y Comunicación Intercultural*, 4 (2), 103-122. https://doi.org/10.14201/clina201842103122

De Wille, Tabea and Bermúdez-Bausela, Montserrat (2020). Quality perceptions and professional status: The impact of extrinsic information on translation editing. *Revista de Lenguas para Fines Específicos*, 26 (1), 84-107. https://doi.org/10.20420/RLFE.2 020.315

Doherty, Stephen, O'Brien, Sharon, and Carl, Michael (2010). Eye tracking as an MT evaluation technique. *Machine Translation*, 24, 1-13. https://doi.org/10.1007/s10590-010-9070-9

Dragsted, Barbara (2010). Coordination of reading and writing processes: An eye on uncharted territory. In Gregory Shreve and Erik Angelone (Eds.), *Translation and Cognition* (pp. 41-62). John Benjamins.

Dragsted, Barbara and Carl, Michael (2013). Towards a classification of translation styles based on eye-tracking and keylogging data. *Journal of Writing Research*, 5 (1), 133-158. https://doi.org/10.17239/jowr-2013.05.01.6

Dreyfus, Stuart E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology and Society*, 24 (3), 177-181. https://doi.org/10.1177/0270467604264992

Golder, Peter N., Mitra, Debanjan, and Moorman, Christine (2012). What is quality? An integrative framework of processes and states. *Journal of Marketing*, 76, 1-23.

Göpferich, Susanne and Jääskeläinen, Riitta (2009). Process research into the development of translation competence: Where are we, and where do we need to go? *Across Languages and Cultures*, 10(2), 169-191. https://doi.org/10.1556/Acr.10.2009.2.1

Hvelplund, Kristian T. (2017). Eye tracking in translation process research. In John W. Schwieter and Aline Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 248-264). John Wiley & Sons. https://doi.org/10.1002/9781119241485.ch14

Jakobsen, Arnt Lykke and Hvelplund, Kristian T. (2008). Eye movement behaviour across four different types of reading task. *Copenhagen Studies in Language*, 36, 103-124.

Just, Marcel A. and Carpenter, Patricia A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87 (4), 329-354. https://psycnet.apa.org/doi/10.1037/0033-295X.87.4.329

Kelly, Dorothy A. (2002). Un modelo de competencia traductora: bases para el diseño curricular. *Puentes*, 1, 9-20.

Kelly, Dorothy A. (2007). Translator competence contextualized, translator training in the framework of higher education reform: In search of alignment in curricular design. In Dorothy Kenny and Kyongjoo Ryou (Eds.), *Across Boundaries: International Perspectives on Translation Studies* (pp. 128-142). Cambridge Scholars Publishing.

O'Brien, Sharon (2009). Eye tracking in translation process research: methodological challenges and solutions. In Inger M. Mees, Fabio

Alves, and Susanne Göpferich, Susanne (Eds.), *Methodology, Technology and Innovation in Translation Process Research* (pp. 251-266). Samfundslitteratur.

PACTE (2000). Acquiring translation competence: Hypotheses and methodological problems in a research project. In Allison Beeby, Doris Ensingerand MarisaPresas (Eds.), *Investigating Translation* (pp. 99-106). John Benjamins.

PACTE (2003). Building a translation competence model. In Fabio Alves (Ed.), *Triangulating Translation: Perspectives in Process Oriented Research* (pp. 43-66). John Benjamins. https://doi.org/10.1075/btl.45.06pac

Paloposki, Outi (2016). Translating and translators before the professional project. *JoSTrans. The Journal of Specialised Translation*, 25, 15-32. https://doi.org/10.26034/cm.jostrans.2016.297

Pavlović, Natasa and Hvelplund Jensen, Kristian T. (2009). Eye tracking translation directionality. *Translation Research Projects*, 2, 93-109.

Schaeffer, Moritz, Nitzke, Jean, Tardel, Anke, Oster, Katharina, Gutermuth, Silke, and Hansen-Schirra, Silvia (2019). Eye-tracking revision processes of translation students and professional translators. *Perspectives*, 27 (4), 589-603. https://doi.org/10.1080/0907676X.2019.1597138

Schäffner, Christina and Shuttleworth, Mark (2013). Metaphor in translation: Possibilities for process research. *Target*, 25 (1), 93-106. https://doi.org/10.1075/target.25.1.08shu

Sharmin, Selina, Špakov, Oleg, Räihä, Kari-Jouko, and Jakobsen, Arnt Lykke (2008). Where on the screen do translation students look while translating, and for how long? In Susanne Göpferich,, Arnt Lykke Jakobsen, and Inger M. Mees (Eds.), *Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing* (pp. 31-51). Samfundslitteratur.

Walker, Callum (2021). *An Eye-Tracking Study of Equivalent Effect in Translation: The Reader Experience of Literary Style.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-55769-0