

Quality Assessment in Interlingual Subtitling: A Systematic Review

Altaf Fakih

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia

Mozhgan Ghassemiazghandi*

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia

Abdul-Hafeed Fakih

Department of English/Translation, Najran University, Najran, Saudi Arabia

Masood Khoshsaligheh

Department of English, Faculty of Letters and Humanities, Ferdowsi University of Mashhad, Mashhad, Iran;
School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia

Abstract—This study is a systematic review of the literature on quality assessment in interlingual subtitling, a critical aspect of audiovisual translation, and media accessibility. Subtitling plays a vital role in bridging linguistic and cultural gaps; nevertheless, defining and assessing its quality optimally remains a complex and multifaceted task that is influenced and hindered by diverse stakeholders and parameters. While existing narrative reviews offer valuable insights, they often lack the systematic rigor required to comprehensively answer focused research questions. This systematic review aims to critically evaluate the existing body of literature, synthesizing relevant studies to uncover key concepts, trends, issues, and methodological approaches, as well as to identify research gaps and inform future investigations in this domain. Employing the PRISMA protocol, this review analyzed 42 research studies selected from five major databases and supplemented by reference citation searches. Thematic synthesis using ATLAS.ti software identified key trends, methodologies, challenges, and gaps in subtitling quality assessment. Findings reveal the diverse and intricate nature of subtitling quality evaluation, emphasizing the need for standardized frameworks, context-sensitive, and technology-driven approaches. This review, one of the first of its kind, provides a synthesized understanding of subtitling quality assessment and offers actionable recommendations for advancing research in AVT, Translation Studies (TS), and Media Localization.

Index Terms—audiovisual translation, human subtitling, quality assessment, interlingual translation

I. INTRODUCTION

Owing to the growing demand for accessibility and globalization, subtitling has become a powerful tool for accessing information, breaking language barriers, and providing rich viewing experiences by enabling viewers to access content in languages that lack proficiency (Gupta et al., 2019). Media accessibility (MA) involves providing access to content and services through media solutions for individuals with sensory impairments or linguistic barriers (Greco, 2019). Subtitling serves as a key tool for MA by facilitating communication across languages, cultures, and sign systems, while compensating for inherent communication losses. This rising demand for multilingual content, of which a significant part is audiovisual in nature, raises important questions for the subtitling industry (Bywood et al., 2017), which has become a vital mode in audiovisual communication and has recently become the focus of heated debate (Agulló, 2020). It was argued that quality in subtitling is challenging to define and assess owing to its complexity, which spans both process and product. Additionally, its evaluation lacks fixed parameters and varies based on the perspectives of the different stakeholders involved in subtitle production and reception (Szarkowska et al., 2021).

With this significant role of subtitling in audiovisual production and MA, more scholarly efforts should emphasize subtitling evaluation and its quality assessment, either as a process or product. Quality assessment is one of the most contentious and recurring topics in translation studies (TS) (Hu, 2021), as it contributes to the development and progress of the field (Karakanta, 2022a). The study of translation quality and its assessment holds significant value as it offers insights into the influence of various factors on the translation process, outcomes, and audience reception. This understanding enables the refinement of techniques, training, and tools to align better with quality standards (Saldanha & O'Brien, 2013). There is no lack of narrative reviews (e.g., Cintas & Remael, 2020; Kuo, 2020) that address important aspects of subtitling quality assessment (SQA) in terms of parameters, TQA models and procedures, challenges, and issues. Such efforts represent a solid contribution to our understanding of quality assessment during

* Corresponding Author.

subtitling. However, the selection of studies in narrative reviews is “left to the tacit good sense of the expert reviewer” (Norris & Ortega, 2007, p. 807). They can be subjective regarding the number and selection of previously published studies included, and they lack systematicity in examining whether these studies have been analyzed in detail (Macaro, 2019). In addition, narrative reviews only summarize the results of individual studies with no aim to answer a predefined research question.

Hence, this investigation was motivated by the lack of a systematic review that seeks to answer a focused question within an extensive and structured selection of relevant studies to develop a synthetic understanding of quality in subtitling and how it has been assessed. Given this gap in the literature, this study aimed to conduct a systematic review of interlingual SQA using comprehensive and robust procedures to select studies on which the findings will be drawn. The goal of this systematic review is to provide a comprehensive assessment of the existing body of knowledge and synthesize an inclusive selection of scholarly work in order to gather relevant information, concepts, and evidence, to explore the important trends, discussions, issues, methodologies, and research gaps, and to inform applicable future research. To our best knowledge, at the current time, this is the first systematic literature review addressing the TQA of interlingual subtitling; it will contribute to TS in general and Audiovisual Translation (AVT) field in particular, by providing deeper insights on how TQA of interlingual subtitling has been addressed so far and what can be done next. In progressing the review, the researchers will be guided by the following research question: How has the TQA of interlingual subtitling been addressed so far?

This paper is divided into four sections. The first section introduces the background and rationale of the study and research question. The second section describes the methodology used to identify and select relevant and representative publications for review. The third section presents the results, while the fourth section synthesizes the findings and highlights the research gaps through a systematic analysis of selected studies relevant to SQA. Finally, the last section presents conclusions and provides recommendations for future research.

II. METHOD

This systematic review aims to carry out a comprehensive assessment of a selection of previous relevant scholarly works using rigorous procedures throughout the process that mainly aim to reduce researchers' bias and prior assumptions. This research was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021). PRISMA Statement is a protocol to conduct systematic reviews, consisting of a 27-item checklist and a four-phase flow diagram. It was initially developed in the medical field by a group of 29 scholars to enhance the transparency and accuracy of the literature reviews. However, the protocol has been used for conducting systematic reviews in several disciplines beyond the medical fields, that is, AVT (Wu & Chen, 2021; Yonamine, 2022) and TS (Liu et al., 2022; Chen et al., 2023). PRISMA was chosen over other existing protocols because of its wide recognition of comprehensiveness and ability to increase consistency in conducting systematic reviews (Pahlevan-Sharif, 2019). The reviewers meticulously followed the PRISMA checklist guidelines. However, some modifications were made to ensure that the procedures were within the scope of the present study. For instance, items related to the risk of bias and combining results of meta-analysis studies (items 10b–15 and 18–22) were omitted as they were related to meta-analyses, while the present study used a qualitative approach to analyze data. In addition, item (24) related to registration and protocol were overlooked, as the review paper was registered.

A. Literature Search Strategy

For this study, a comprehensive literature search was undertaken to identify relevant papers in a structured and transparent manner, making it possible to replicate the method in the future. The study selection process was guided by the main steps in the PRISMA Flow Diagram: identification, screening, and eligibility (Figure 2).

An extensive literature search was conducted using five electronic databases: Scopus, Web of Science (hereafter WoS), Bibliography of Interpreting and Translation – BITRA, EBSCOhost, and Semantic Scholar, for the identification phase. After conducting a pilot test, the researchers selected Scopus, WoS, BITRA, and EBSCOhost as the main databases because of their academic relevance to works on TS and AVT and the advanced search engine that enables the use of Boolean operators and search functions. Semantic Scholar, which is classified as a grey literature database, was added to the group later as supplementary to ensure wide-range coverage and a sufficient number of reviewed studies. In addition, reference citations were added as a source of data collection after the reviewers found closely related articles during the data-extraction phase.

The researchers sought synonyms, related terms, and variations to enrich keywords for better literature search results. The authors pilot-tested a combination of related terms before the final selection to ensure the best retrieval results; based on this process, the search terms similar to “Subtitl* and translation quality assessment” devised for the identification phase were checked in the selected databases, as shown in Figure 1. The combinations of these keywords were processed using search functions, such as field code, phrase searching, and truncation (e.g., asterisk (*)), which were used to truncate words and capture variations around a word stem, enabling spelling variations to be found whenever possible. Boolean operators (AND, OR, AND/OR), parentheses, and quotation marks were also used to insert keywords into the chosen databases, thereby enhancing literature retrieval.

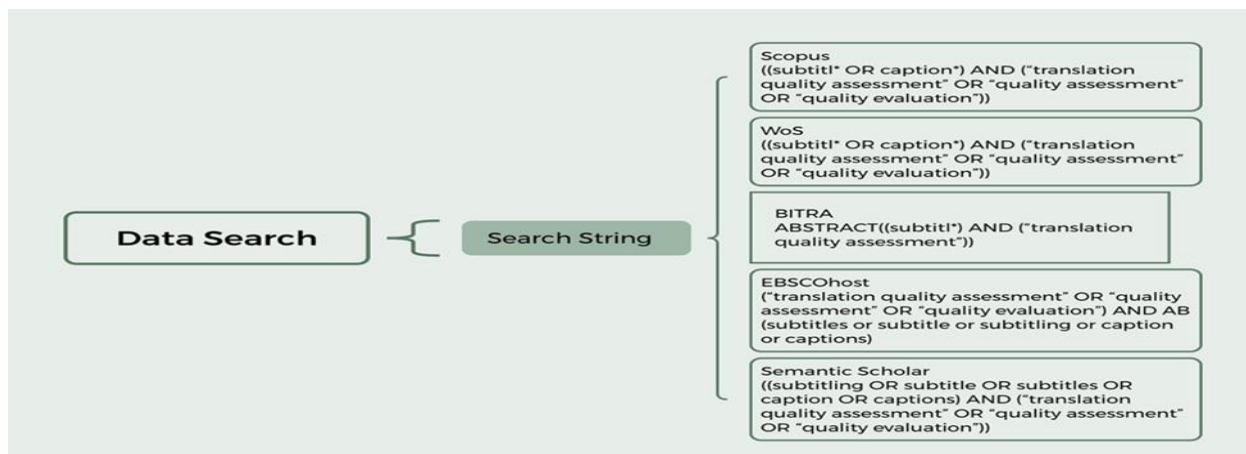


Figure 1. Databases and Search Strings Utilized for Literature Search

B. Eligibility Criteria

To comprehensively gather research on SQA, it is essential to establish precise inclusion and exclusion criteria, which serve as requisite requirements for conducting a non-biased systematic review (Pahlevan-Sharif, 2019).

The study was deemed eligible if: (1) the study reported data on the focus of the study, that is, the TQA of interlingual subtitling; otherwise, studies involving other modes of AVT, such as intralingual subtitling, dubbing, and voice-over, were excluded. (2) The study was written in English; only documents written in English were included, given that English has been the dominant language in international publications over the past decades (Hamel, 2007), particularly in translation and interpreting studies (Yan et al., 2018); (3) the study was published as a journal article, conference proceeding, book, or book chapter; and (4) the study was fully accessible.

On the other hand, studies were excluded if they met one or more of the following exclusion criteria: (1) The study was part of an unpublished study, a review or commentary, or any other write up that fall outside the inclusion criteria; (2) The study was irrelevant to the topic of this systematic literature review. In other words, if the study involved TQA of intralingual subtitling, live subtitling, surtitling, and dubbing, among others, it was exempted. No time restrictions were imposed to ensure that all relevant data were collected.

C. Data Identification and Screening

The identification phase started with searches of the five databases conducted on December 18, 2023, returning 526 results. Another round of literature search was carried out on October 30, 2024, to search for relevant documents that were not yet retrieved. Three additional documents were detected in the EBSCOhost database. This resulted in a total of 529 participants. The number of results per database is illustrated in Flow Diagram 1. The researchers inserted the retrieved records into a semi-automated tool, called “Rayyan” for duplication detection and screening phase. The automated tool detected 125 duplicate records. However, after human revision, three records were retrieved, as it was found that they were not duplicates, whereas one record was detected and excluded because it was duplicated. Thus, 71 duplicate records were removed. Finally, the identification phase detected 458 distinct studies that were eligible for the screening phase.

The reviewers then independently screened the titles, abstracts, and keywords of the records. Papers that did not meet the inclusion criteria were discarded. Of the 458 studies, 386 were excluded based on the title/abstract/keyword analysis, either because the investigations focused on MT-generated subtitling, live subtitling, speech translation, interpretation, closed captioning image/video quality assessment, engineering, etc., or were written in languages other than English (three studies published in Czech, Chinese, and Spanish). The full texts of the remaining papers (i.e., 72 studies) were carefully reviewed. In this phase, any disagreements among reviewers were addressed through discussion and resolved by reaching consensus.

D. Data Extraction

The full texts of the selected articles were processed in three stages: eligibility assessment, data extraction, and data analysis. First, the articles were stored as CSV Excel files containing the metadata, that is, bibliography, abstract, affiliations, keywords, etc., for eligibility assessment; out of 72 selected articles, 34 were excluded as they did not meet the eligibility criteria, as follows: 28 papers in which SQA was not the main focus; 4 papers were not fully accessible; and 3 papers were reviewed. Additionally, during the full-text review phase, the researchers identified 5 articles that were closely related to the topic of this systematic review and met the eligibility criteria. Thus, these articles were included in the review.

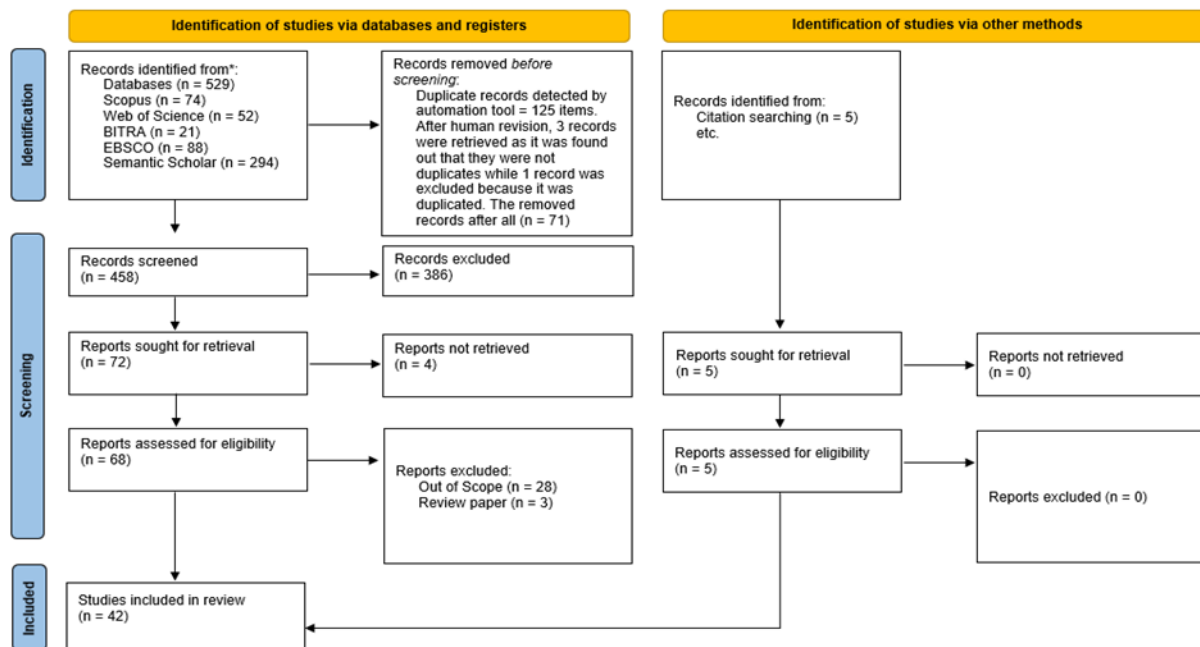


Figure 2. Selection of the Publications for the Review – PRISMA Flow-Chart

A total of 42 articles were imported into ATLAS.ti – Qualitative Data Analysis Software – for data extraction and synthesis. The data extraction process was guided by the primary research question, and qualitative data synthesis (thematic synthesis) was applied to analyze the extracted data. The two researchers independently extracted the data by following the standardized protocol and procedure for data extraction using software that enabled the identification and classification of the repeated codes in the selected studies into themes.

III. RESULTS

The included articles were thematically analyzed to identify and code patterns within existing studies by examining any similarities or connections to present how the topic has been addressed, thereby answering the research question by performing qualitative synthesis. At this stage, the reviewers thoroughly read the included articles, coded the repeated patterns, and grouped the patterns into four themes and subthemes.

The first theme of the analysis aimed to form a general overview of the studies, looking into the year of publication and analysis themes. The other three themes analyzed were: Factors of subtitling quality; SQA included four subthemes: SQA methods, language and genre, quality assessor, and measuring scales; and Models/frameworks for SQA.

A. Theme 1: Overview of Studies

From a diachronic perspective, the 42 studies selected for this review were published from 2008 to 2023, covering a 15-year period. A surge in publications was observed from 2016 onward. It is clearly a watershed year for the three-decade period because the number of publications from 2016 to 2023 accounts for 74% of the total sample (31 out of 42). It is fair to say that the past decade has witnessed increased scholarly interest in the quality and assessment of interlingual subtitling.

After a thorough examination, the focal studies were grouped according to their orientation of investigation (Table 1). Before presenting the results, the themes outlined in Table 1 are clarified. Product-oriented studies focus on investigating the quality of subtitling from the subtitled product’s perspective, meaning the assessments conducted mainly focused on examining the subtitles of the filmed content along with the dialogue, images, sounds, technical constraints, and so on. By the term “Practitioner”, we refer to the producers who involves in the process of subtitling production, i.e., subtitlers, simulators, proofreaders, editors, subtitling companies, etc. Viewer-oriented studies discuss the quality of subtitling from the perspective of the recipient of the subtitles, while academic-oriented studies investigate subtitling quality from pedagogical and didactic perspectives. Based on the results, the studies were categorized into two groups: unidirectional studies, focusing on one aspect in their investigation, and multidirectional studies, addressing two or more aspects. Most studies (29 of 42) were unidirectional and distributed as follows: product-oriented (25 cases), practitioner-oriented (seven cases), viewer-oriented (two cases), and academic-oriented (one case). Conversely, seven studies adopted a multi-directional approach indicated with an asterisk (*) in Table 1.

TABLE 1
ORIENTATION OF INVESTIGATION

Orientation type	Studies
Product oriented	Taivalkoski-Shilov (2008)*; Pedersen (2008); Bogucki (2009); Bittner (2011); Gamal (2013)*; Nurhidaya (2013); Horbačauskienė (2016); Alaa Eddin and Khuddro (2016)*; Jin (2017)*; Furgani (2019)*; Hidayati (2019); Pedersen (2017); Pedersen (2019); Furgani (2019); Abdelaal (2019); Ninsih (2020); Hudi et al. (2020); Kuo (2020); Kuswardani I and Septiani (2020); Hu (2021); Abdelaal and Al Sarhani (2021); Putri (2022); Batmanathan et al. (2022); Basriana et al. (2022); Bogucki (2022); Alaa and Al Sawi (2023); Mehdizadkhani (2023)*; Mounadil (2023); Sanatifar and Ghamsarian (2023); Gupta and Sharma (2023); Gil (2023).
Practitioner oriented	Taivalkoski-Shilov (2008); Munday (2012); Gamal (2013); Imre (2015); Kuo (2014); Alaa Eddin and Khuddro (2016); Robert and Remael (2016); Künzli (2020); Jin (2017); Szarkowska et al. (2020)*; Oziemblewska and Szarkowska (2020); Artegiani (2021); Mehdizadkhani (2023).
Viewer oriented	Künzli and Ehrensberger-Dow (2011); Jin (2017); Furgani (2019); Deckert (2021); Mehdizadkhani (2023).
Academic oriented	Gamal (2013); Martins and Ferreira (2019); Szarkowska et al. (2020).

B. Theme 2: Subtitling Quality Factors

Quality factors can be defined as the characteristics of a product or process that affect its ability to meet clients' and viewers' expectations. Identifying and managing both quality parameters and factors are essential for ensuring quality subtitling. Of the focal studies, six studies investigated the factors that led to the deterioration of quality in interlingual subtitling in different contexts, as shown in Table 2.

TABLE 2
THE MAIN FINDINGS OF THE REVIEWED STUDIES ADDRESSING QUALITY FACTORS IN SUBTITLING

Item	Context	Study Approach	Data Collection Instrument	Findings
Gamal (2013)	Egypt	Qualitative	Multimodal corpus analysis and interviews	<ul style="list-style-type: none"> ▪ Subtitlers work on a freelance basis ▪ Lack of quality control ▪ Subtitling is considered as a form of written translation ▪ AVT students and trainees are not given proper training in subtitling.
Kuo (2014)	Europe and China	Quantitative	Online survey	<ul style="list-style-type: none"> ▪ Poor remuneration ▪ Tight assignment deadlines ▪ Low quality of support materials ▪ Inadequate use of subtitling programs ▪ Inadequate quality control procedures ▪ Invisibility of subtitlers ▪ Recruitment of inexperienced subtitlers
Robert and Remael (2016)	International	Quantitative	Online survey	<ul style="list-style-type: none"> ▪ Falling subtitling fees ▪ Turning into translation technology solutions and templates ▪ Improper quality control
Szarkowska et al. (2020)	International	Qualitative	Qualitative analysis of online open-ended survey, eye-tracking tests, and semi-structured interviews	<ul style="list-style-type: none"> ▪ Falling subtitling fees ▪ Rise in subtitle display rates and less condensation ▪ Widespread use of templates ▪ Lack of quality control procedures ▪ Lack of proper recognition of subtitling as a profession ▪ Influx of inexperienced people into the profession ▪ Inaccurate transfer of content ▪ Issues in technical synchronization ▪ Visibility of the subtitles.
Artegiani (2021)	UK and Europe	Qualitative and Ethnographic	Observation and semi-structured interviews	<ul style="list-style-type: none"> ▪ Prioritizing client's satisfaction over the faithfulness of message conveyance ▪ Restrictions or impositions of technology use
Oziemblewska and Szarkowska (2022)	International	Mixed method	Online Survey	<ul style="list-style-type: none"> ▪ Low quality subtitling templates

C. Theme 3: SQA

The third theme discusses different aspects (see Table 3) of studies that conducted quality assessment in interlingual subtitling. It includes four subthemes: assessment instrument, quality evaluator, language and genre, and measurement scale.

TABLE 3
OVERVIEW OF THE SELECTED STUDIES CONDUCTING SQA

Item	Assessment Instrument	Evaluator	Language Pair	Genre	Measuring Scale
Taivalkoski-Shilov (2008)	Toury's Framework + corpus stylistic analysis	Researcher	English to Finnish, French, and Russian	Rap lyrics	Accuracy and acceptability
Bogucki (2009)	Error Analysis	Researcher	English to Polish	Fantasy action movie – Fansub	Subtitled's competence
Bittner (2011)	Comparative Discourse Analysis	Researcher	English to German	Crime and thriller movie	Accuracy
Künzli and Ehrensberger-Dow (2011)	Eye tracking + questionnaire	Viewers	French to German	Drama Movie – Fansub vs. Prosub	Audience's reception capacity, response, and satisfaction
Gamal (2013)	Corpus Multimodal Analysis + Interviews	Researcher	Egyptian Arabic to English	Classic films	Linguistic and semiotic aspects
Nurhidaya (2013)	Nababan's model	Researcher	English to Indonesian	Adventure/Fantasy movie	Accuracy, acceptability, readability
Artegiani and Kapsaskis (2014)	Comparative Analysis	Researchers + One of the drama series authors	English to French, Greek and Spanish	Drama series	Text reduction, narrative consistency, intertextuality, technical constraints
Imre (2015)	Discourse Analysis	Researcher	English to Romanian	Drama Series	Subtitled's competence
Alaa Eddin and Khuddro (2016)	Error Analysis	2 Researchers	English to Arabic	Animation movie	Linguistic and technical aspects
HorbaÅauski enÅ (2016)	Document Analysis using Pedersen's CSI taxonomy	Researcher	English to Lithuanian	Reality show	CSI transference
Jin (2017)	Corpus + interviews + questionnaire	Researcher + viewers + scholars	English to Chinese	Animation	Linguistic and cultural aspects
Abdelaal (2019)	Document analysis + FAR model	Researcher	English to Arabic	Sex comedy film series	Functionality, readability, acceptability
Furgani (2019)	Open-end questionnaires + House's TQA model	Researcher + viewers	English to Arabic	Not-mentioned	Accuracy
Hidayati (2019)	Document analysis + questionnaire using Nababan model	Viewers	English to Indonesian	Animation movie	Readability
Martins and Ferreira (2019)	FAR model, Kuo's guidelines (2014), Robert and Remael's model (2016)	Researchers	Multilingual	Multigenre	Technical constraints, stylistic, typographical and translation considerations.
Pedersen (2019)	FAR model	Researcher	English to Swedish	Multigenre – Fansub vs. Prosub	Functionality, acceptability, readability, creativity
Hudi et al. (2020)	Documentation + questionnaires and Nababan model	Expert (English lecturers and EFL students)	English to Indonesian	Comedy drama movie	Accuracy, acceptability, readability
Kuo (2020)	Ramos Pinto framework	Researcher	Singlish to Mandarin and English	Comedy drama movie series	Textual, diegetic, and socio-cultural dimensions
Kuswardani and Septiani (2020)	Nababan's model	2 expert evaluators	English to Indonesian	drama series	Clarity, accuracy and naturalness
Ninsih (2020)	Document analysis + Nababan's model	Researcher	English to Indonesian	Animation movie	Accuracy
Abdelaal and Al Sarhani (2021)	Document analysis + FAR model	Researchers	English to Arabic	Thriller crime movie	Functionality, acceptability, readability
Deckert (2021)	Audiovisual stimuli +	Viewers	Danish to Polish	Thriller drama film	Cognitive processing, reception cognitive load, enjoyment,

	questionnaire				comprehension, transportation
Basriana, et al. (2022)	Nababan's model	Raters (practitioners/ academics and informants)	Indonesian to English	YouTube - news	Communicativeness, acceptability, readability
Batmanathan et al. (2022)	Document analysis + Nababan's model	Researchers + peer validators	Malay to English	Animation movie	Accuracy
Putri (2022)	Document analysis + focus group discussion using Nababan's model	3 raters including the researcher	Indonesian to English	Folklore Horror Video game	Accuracy, acceptability, readability
Alaa and Al Sawi (2023)	Document analysis + FAR model	3 raters including the researchers	Egyptian Arabic to English	Comedy drama movie	Functionality, acceptability, readability
Gil (2023)	FAR model	Researcher	Spanish to English	Multigenre	Functionality, acceptability, readability
Mehdizadkha ni (2023)	FAR model + Interviews + Surveys	Researcher, Subtitlers, viewers	English to Iranian	Multigenre Fansub vs. Prosub	Functionality, acceptability, readability, reception
Mounadil (2023)	Corpus Analysis + FAR model	Researcher	English to Arabic	Sitcom	Functionality, acceptability, readability
Sanatifar and Ghamsarian (2023)	FAR model	Researchers	Persian to English	Multigenre	Functionality, acceptability, readability

(a). *Assessment Instrument*

The results revealed that a variety of methods have been used in the focal studies to assess the quality of interlingual subtitling (see Table 3). The data indicated a distinct preference for using models and frameworks as the primary method for assessing the quality of interlingual subtitling. This approach, adopted by 48% of the studies, provides a systematic way to evaluate subtitling, ensuring that the assessment is aligned with the established standards and criteria. The second most common approach, employed by 35% of the studies, is a MIXED-METHOD strategy approach, in which two or three methods are mixed and utilized for the quality assessment of subtitling. For instance, eye tracking was employed along with a questionnaire (Künzli & Ehrensberger-Dow, 2011) and audiovisual stimuli with written questionnaire-elicited input including participant self-reports (Deckert, 2021) in studies that attempted to examine the reception and viewing experiences of AVT product recipients. Furthermore, corpus-based approaches have also been utilized in SQA, supported by error analysis (Mounadil, 2023), multimodal analysis (Gamal, 2013), and stylistic comparative analysis (Taivalkoski-Shilov, 2008). Another group of instruments, including discourse/comparative analysis (10%) and error analysis (7%), were also used individually yet less frequently in the focal studies.

Notably, only nine studies applied the FAR model, one of the few specifically designed and tailored for assessing the quality of interlingual subtitling. The FAR model (Pedersen, 2017) is an error-based framework that focuses on examining three core aspects within subtitled products: functionality, acceptability, and readability. Moreover, eight studies conducted assessments using a model for assessing the quality of monomodal written translations developed by Nababan et al. (2008, 2012). This model uses a holistic approach to assess quality using three measuring scales: accuracy, acceptability, and readability. It was noted that all the studies utilizing Nababan's TQA model (2008, 2012) were either English-Indonesian or English-Malay language pairs, as the model was originally developed in Indonesian. Nevertheless, a number of studies have borrowed theoretical frameworks from neighboring disciplines. For example, Gamal (2013) adopted a Multimodal Analysis framework consisting of polysemiotic text analysis (Delabastita, 1989), film analysis and description (Thibault, 2000; Taylor, 1993), and multimodal pragmatic analysis of film discourse (Mubenga, 2009). In addition, Eddin and Khuddro (2016) conducted an error analysis based on practical functional approach theories accumulated over three decades. Similarly, Taivalkoski-Shilov (2008) utilized Toury's framework (1995) to conduct an indirect comparison measuring stylistic shifts made by subtitlers. Kuo (2020) utilized Ramos Pinto's analytical framework (2018) for examining non-standard language varieties in subtitling.

Furthermore, several focal studies have investigated textual categorizations and strategy typologies applied in subtitling audiovisual texts using document analysis to assess the quality of the texts. Nurhidaya (2013), Hudi et al. (2020), and Putri (2022), for instance, adopted Molina and Albir's (2002) translation techniques typology for identifying the techniques applied by the subtitlers. In addition, Ninsih (2020) adopted the written-text translation techniques developed by Baker (2001) to examine the quality of subtitling idioms in an animation movie. Strategy typologies established for subtitling, on the other hand, have been utilized by other scholars. Jin (2017), Hidayati (2019), Kuo (2020) and Batmanathan et al. (2022) used the subtitling strategies typology proposed by Gottlieb (1992). In addition, Pedersen's (2005, 2011) subtitling strategy typology has been applied in a number of studies (HorbaÅauskienÅ, 2016; Abdelaal, 2019; Abdelaal & Al Sarhani, 2021; Mounadil, 2023).

(b). *Quality Evaluator*

This subtheme presents the roles of individuals involved in quality assessment activities across focal studies. Most of the assessments (58%) were conducted by the researchers themselves. Additionally, 14% of the studies utilized inter-rater agreement, and another 14% involved participants (i.e., viewers and readers) to assess the quality of the subtitles displayed. Notably, only four studies utilized a combination of different assessors. For instance, Furgani (2019) supplemented viewer data with a quality assessment carried out by the researcher, whereas Basriana et al. (2022) engaged raters, either practitioners or academics with a background in translation, to evaluate the communicativeness and acceptability of the subtitles. Simultaneously, informants without translation backgrounds were asked to assess the readability of the same subtitles.

(c). *Language and Genre*

According to the Publication Office of the UN, subtitling is the most commonly used language transfer practice in Europe, encompassing 28 countries, including Bulgaria, Denmark, Finland, the Netherlands, Norway, Portugal, Sweden, Switzerland, Turkey, and the United Kingdom (Angrisani et al., 2011). However, this review revealed that English-Indonesian and English-Arabic language pairs were dominant, with 26% each in the SQAs conducted, as illustrated in Table 3. As a hegemonic or dominant language in the international audiovisual industry, English was present as one of the language pairs (either the source or target language) in most of the investigations (26 cases). Furthermore, only three studies made an exception to evaluating the standard form of language and investigated non-standard language varieties, for example, Egyptian Arabic in classic films (Gamal, 2013; Alaa & Al Sawi, 2023) and Singlish, that is, a variety of English spoken in Singapore, incorporating elements of Chinese and Malay (Kuo, 2020), in comedy drama movie series.

To explore whether there was a certain tendency of quality assessment toward a particular genre, titles and corpora were classified following IMDb's genre typology (IMDb, 2023), an authoritative source for media content classification. The results showed a wide variety of genres involved in the investigations, with an almost balanced ratio. Five studies drew the orientation assessment of a compilation of excerpts from multiple titles with different genres, they were labelled "multigenre", it was not possible to analyze the correlation of each genre with the overall SQA trend represented in the study. Further, subtitling production when it comes to the producer is classified into professional (also prosubbing, which is officially produced by professional subtitlers and agencies) and non-professional, including crowdsourcing, amateur, or fansubbing made by fans and non-specialists who do not have academic or professional backgrounds on subtitling or translation. Only five studies discussed non-professional subtitling activities either to assess the quality of the subtitled products in comparison with subtitles produced by professionals (Künzli & Ehrensberger-Dow, 2011; Pedersen, 2019; Mehdizadkhani, 2023), or to evaluate the amateur subtitlers' competence, as in Bogucki (2009) and Munday (2011).

(d). *Measuring Scales*

As shown in Table 3, there are a variety of measuring scales upon which the quality assessments have undergone, and in favor of clarity, the studies were categorized according to the model, theoretical framework, or any assessment tool used for the evaluation. First, the analysis of the studies utilized the FAR model (Abdelaal, 2019; Martins & Ferreira, 2019; Pedersen, 2019; Abdelaal & Al Sarhani, 2021; Alaa & Al Sawi, 2023; Gil, 2023; Mehdizadkhani, 2023; Mounadil, 2023; Sanatifar & Ghamsarian, 2023) concentrated on three measuring scales: functionality, acceptability, and readability. Studies adopting Nababan's TQA framework have mainly focused on three measuring instruments: accuracy, acceptability, and readability (Nurhidaya, 2013; Hudi et al., 2020; Basriana et al., 2022; Putri, 2022), while others have confined the assessment to only one aspect (Hidayati, 2019; Ninsih, 2020). Kuswardani and Septiani (2020), however, made some amendments to their model to serve their research objectives. From a practical functional perspective, Alaa et al. (2016) investigated the extent to which linguistic and technical elements of the source text and target text achieve cohesion, coherence, and acceptability. Taivalkoski-Shilov (2008) explored the stylistic shifts in rap lyrics, focusing on accuracy and natural language style as key assessment instruments.

Two studies (Gamal, 2013; Kuo, 2020) directly addressed the multimodal nature of subtitling and investigated the semiotic dimension, along with linguistic and cultural aspects, in their analysis. With regard to culture in subtitling, Bittner (2011) and HorbaÄauskienÄ (2016) addressed the transfer of language and culture between the original dialogue and target subtitles. Furthermore, assessment scales of reception studies (Künzli & Ehrensberger-Dow, 2011; Furgani, 2019; Deckert, 2021) mainly focused on concepts related to how the audience cognitively received and experienced the subtitled products, such as comprehension scores, satisfaction rates, cognitive load, and enjoyment.

All of the above-mentioned studies focused on examining the subtitles for assessing the quality of the product itself, except for Bogucki (2009) and Imre (2015), who evaluated the performance of the subtitler by analyzing the subtitles by whom they were produced.

D. *Theme 4: Frameworks and Models for SQA*

Remarkably, eight studies go even further beyond quality assessment to developing models or proposing theoretical frameworks tailored to investigate various aspects of interlingual subtitling and its quality. Pedersen (2008) proposed a framework integrating Speech Act theory and Skopos theory, emphasizing the prioritization of the speaker's illocutionary intent over surface structures to ensure fidelity to the original message within media-specific constraints. Munday (2012) employed the Appraisal Theory of Evaluation (Martin & White, 2005) to examine subtitler decision

making and its impact on interpersonal meaning, while Bogucki (2022) employed Relevance Theory to analyze the intersemiotic nature of subtitling through four semiotic channels. Hu (2021) combined House's (1977, 1997, 2015) TQA with Multimodal Discourse Analysis, using Systemic Functional Theory to account for audiovisual dimensions. Pedersen (2017) developed a viewer-centered FAR model based on the notion of a tacit 'contract of illusion' (Pedersen, 2007, pp. 46–47) between the subtitler and the viewer. Künzli (2020) introduced the CIA model, grounded in professional subtitlers' insights, which assesses correspondence, intelligibility, and authenticity to facilitate the viewing experience. Gupta and Sharma (2022) proposed a model for improving subtitle timing by identifying non-transcribed dialogue segments using language-agnostic methods.

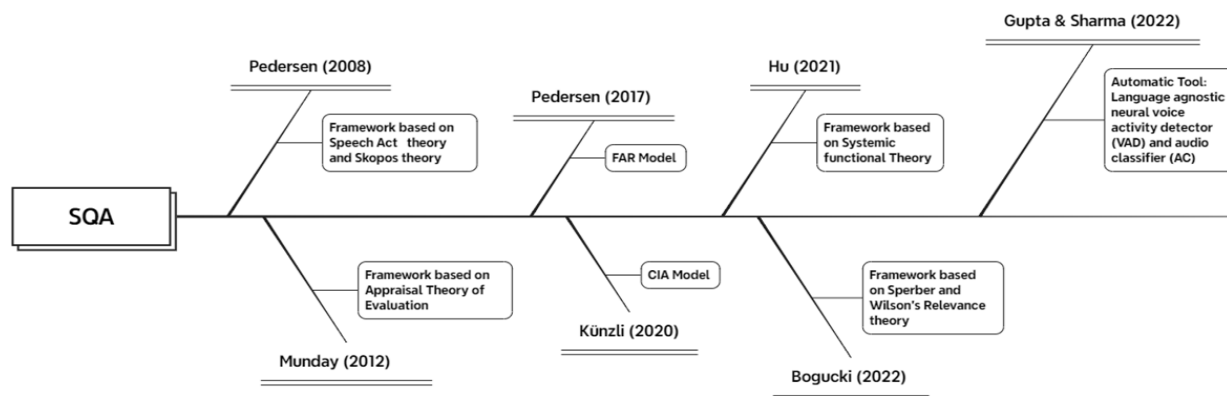


Figure 3. Frameworks and Models for Subtitling Quality Assessment

IV. DISCUSSION

A. Overview

The results revealed that although there is a notably growing interest in focusing on the topic under investigation, the publication trend remains limited compared to interrelated disciplines such as TQA. This advocates the claims of many scholars who have attempted to address this issue (Gamal, 2013; Kuo, 2014; Künzli, 2020; Hu, 2021). This may be attributed to the marginal role of subtitling in the film industry (Bittner, 2011; Kuo, 2014).

As for investigation orientation, Table 1 shows that product-oriented studies have attracted sustained scholarly attention, followed by investigations on subtitling practitioners. In contrast, the investigation of subtitling viewers and learners has received little attention. Künzli and Ehrensberger-Dow (2011) confirmed that few reception studies have been conducted in AVT, more specifically in subtitling which can be concluded that the situation has not changed much ever since. Translation has been conceptualized as a complex interplay of relationships among the target text, source text, their respective text types, context, and stakeholders involved in the translation process (Bittner, 2011; Szarkowska et al., 2020). This perspective frames subtitling quality as a dynamic construct shaped by various factors that interact during both production and reception stages. To address subtitling quality more comprehensively, assessments should consider both production and reception perspectives. However, existing research has predominantly focused on one dimension at a time. Only three studies (Jin, 2017; Szarkowska et al., 2020; Mehdizadkhani, 2023) have examined both subtitling production and reception, underscoring a significant gap in empirical research. This highlights the need for further investigation into subtitling quality that integrates the perspectives of all stakeholders involved.

B. Subtitling Quality Factors

To bridge the gap between theory and practice in subtitling, scholarly efforts have been made to investigate the quality factors related to the subtitling profession and industry. Based on the findings shown in Table 2, the factors contributing to the deterioration of subtitling quality are of two root-cause types, namely subtitler-related and profession-related (working conditions) factors.

A recurring issue in the industry is the lack of proper quality control procedures, which are often confused with quality management and assurance. Robert and Remael (2016) clarified these concepts, positioning quality control as a subset of quality assurance, which itself is part of quality management. Despite the importance of these procedures, Gamal (2013), Kuo (2014), and Robert and Remael (2016) highlight that clients often fail to enforce them, focusing instead on style and technical specifications at the expense of linguistic accuracy. This is exemplified by the absence of clear project briefings, guidelines, or adequate resources, as well as a lack of emphasis on scripts and audiovisual materials. Robert and Remael (2016) observed that while subtitlers prioritize linguistic and technical aspects equally, clients often prioritize technical parameters. Such practices are reflected in industry guidelines such as Netflix's style

guide. However, both linguistic and technical qualities are crucial for a smooth viewer experience. A grammatically correct yet poorly timed subtitle, or vice versa, can disrupt viewers' immersion and enjoyment.

The subtitling industry's structural dynamics further exacerbate quality concerns. Artegianni (2021) argued that AVT practices increasingly prioritize profit and client satisfaction over the original communicative function of films. Another factor for quality deterioration is falling subtitling fees due to economic pressure and high competition (Kuo, 2014; Robert & Remael, 2016; Szarkowska et al., 2020), leading to reduced motivation among subtitlers to ensure accuracy and thoroughness. Furthermore, agencies often hire inexperienced subtitlers to cut costs and undermine professional standards (Bittner, 2011; Gamal, 2013). Imre (2015) emphasized that subtitlers' competence is pivotal for quality, advocating skill-based approaches to address these challenges.

Globalization has intensified the demand for multilingual subtitling, resulting in widespread use of templates or spotting lists. Although templates can improve efficiency, their poor quality often hinders subtitling outcomes (Oziemblewska & Szarkowska, 2022). Although Artegianni and Kapsaskis (2014) suggested that templates do not compromise quality, they acknowledged that rigid and normative formats often limit the subtitlers' creativity and standards. Tight deadlines further strain subtitlers, pushing the industry toward technological solutions such as machine translation (MT). Artegianni and Kapsaskis (2014) warned that while MT improves efficiency, it risks reducing subtitlers' skills and professional identity, with post-editing becoming a standard practice.

Overall, consistent quality factors across different regions suggest the potential of unified guidelines for best practices. Szarkowska et al. (2020) advocated for greater collaboration among stakeholders, including academia, practitioners, clients, and the public, to enhance awareness of subtitling quality and improve the working conditions of subtitlers. Future research should validate these findings across diverse settings to establish more comprehensive quality standards.

C. SQA

Most focal studies (48%) conducted SQA using models and frameworks, emphasizing the pivotal role of structured methodologies in the field. These approaches offer systematic evaluations that are aligned with established standards and reduce subjectivity. However, such models face criticism for their reductionist tendencies as they attempt to encapsulate the multifaceted nature of subtitling, such as categorizing and annotating errors (Saldanha & O'Brien, 2013). Given the unpredictable and context-sensitive nature of quality, expecting a single QA model, especially one not specifically designed for subtitling, such as Nababan's TQA model, to be universally valid, is unreasonable. Effective assessments must account for the context, purpose, and end-user needs of the translation product, necessitating models that balance flexibility with precision (Lauscher, 2000).

Recent trends in research methodology highlight the growing consensus on methodological triangulation, where product-focused evaluations are complemented by studies on reception (Jin, 2017), production (Gamal, 2013), or both (Mehdizadkhani, 2023). This approach acknowledges the complexity of subtitling quality assessment. Despite this progress, many assessments remain highly subjective as they rely on one-sided judgments by researchers. While complete objectivity is unattainable owing to the inherent nature of translation (Bittner, 2011), subjectivity can be mitigated through methods such as inter-rater agreement. Consensus among multiple annotators reduces bias and distortion, resulting in more reliable and comparable outcomes (Saldanha & O'Brien, 2013). As House (2005) argued, future research on TQA must move beyond subjective, assertive judgments. Developing intersubjectively verifiable evaluative criteria based on large-scale observational and empirical studies is essential for advancing the field.

D. Models and Frameworks for SQA

Based on the results, we can conclude that the suggested theoretical frameworks for assessing quality in interlingual subtitling (Pedersen, 2008; Munday, 2012; Hu, 2021; Bogucki, 2022) suffer from fuzzy definitions of their core concepts and still lack distinct guidelines and measuring instruments for assessments to be ready for application in research and practice. In addition, they need to be tested for validation through observational data and empirical studies that involve a large number of samples with varying language pairs and text genres.

The FAR (Pedersen, 2017) and CIA models (Künzli, 2020) provide more structured frameworks for assessing subtitle quality, with well-defined scales and parameters designed to enhance subtitle reception and foster a flow experience. However, the CIA model requires further validation, including assessments by professional subtitlers employing a think-aloud protocol, and the development of weighting systems for its parameters (Künzli, 2020). The FAR model, on the other hand, has been the most widely adopted for interlingual subtitling assessment, utilized in 21% of relevant studies. Its popularity is attributed to its adaptability, which allows localization through integration with in-house guidelines, best practices, or national subtitling norms. The analytical approach can be considered a weakness of the model because it does not provide any bonus points for creative translation solutions. However, the numerically measured characteristic adds value to the model, as it makes it less subjective and suitable for didactic purposes (Künzli, 2020).

From the results of this theme, we conclude that the SQA area still needs well-established models and frameworks that evaluate subtitling from different angles and varying settings. QA models need to extend beyond basic frameworks to consider a broader array of influencing factors, such as multimodality, subtitler competence and ideology, perception of quality in the current technological turn, global and local context, genre, purpose, viewers reception, and so on,

which would clearly be highly complex, yet not impossible. As with all assessment methods and themes, whether they are product-, process-, or viewer-oriented, researchers need to seek transparency, giving necessary details on the QA model and methods of evaluator recruitment and training in order to facilitate replicability and reinforce reliability of results. Furthermore, with the ever-increasing demand for content subtitling, automatic evaluation metrics [(Gupta & Sharma, 2022) is one example] can provide a great contribution to the field as they compensate for cost, time, and subjectivity issues involved in the human assessment of translation quality. Therefore, we advocate increased scholarly efforts to develop robust automatic evaluation metrics.

V. CONCLUSION

This systematic review demonstrated that while interlingual subtitling quality assessment has gained increasing attention in recent years, research in this area remains considerably fragmented. The analysis of the 42 studies revealed a diverse array of theoretical approaches, methodologies, and measurement scales. Product-oriented assessments, often guided by established or adapted TQA models, dominate the literature and underscore the tendency to focus on the subtitled text itself. At the same time, the relatively sparse investigations into subtitling practitioners and viewers confirm that multiple stakeholder perspectives, especially the experiential and pedagogical dimensions, remain insufficiently explored. Across different contexts, two broad categories of quality challenges frequently emerge: those arising from the profession itself (e.g., limited quality control, poor remuneration, reliance on inflexible templates), and those linked to subtitler competence (e.g., insufficient training, inexperience, or lack of linguistic and cultural awareness). A key takeaway from this review is the absence of a unified, context-sensitive framework that can adequately capture the nuances of substitution. Although promising models such as the FAR and CIA models have been introduced, they still lack extensive empirical validation and often do not fully account for multimodality, evolving technology, or shifts in viewer demand. Likewise, many industry-oriented studies highlight the growing reliance on post-editing and machine translation, developments that demand systematic investigation into how technological interventions shape both subtitler workflows and viewers' reception of subtitled content. Therefore, future research should strive for greater methodological triangulation, integrating production and reception data, and considering subtitler competencies, working conditions, and regional and cultural specificities. Ultimately, these findings underscore the need for continued collaborative efforts among academic researchers, industry practitioners, clients, and technology developers to establish robust shared metrics and protocols for high-quality subtitling. Advances in automatic evaluation methods may offer a promising path toward more objective and scalable assessments if they are developed in tandem with human-centered insights into the linguistic, cultural, and semiotic complexities inherent in audiovisual texts. By pursuing more holistic, multifaceted, and technologically informed approaches, stakeholders can better ensure that subtitling fulfills a crucial role in media accessibility and cross-cultural communication.

REFERENCES

- [1] Agulló, B. (2020). Technology for subtitling: a 360-degree turn. *Hermēneus. Revista de traducción e interpretación*, (22), 11-40. <https://doi.org/10.24197/her.22.2020.11-40>
- [2] Angrisani, S., Eugeni, C., Fontanel, H., Hamaoui, N., Modot, A., Safar, H., & Verstrepen, S. (2011). *Study on the use of subtitling. The potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages* (Report No. EACEA/2009/01). The European Commission, Directorate-General Education and Culture. Retrieved June 16, 2024, from http://publications.europa.eu/resource/cellar/e4d5cbf4-a839-4a8a-81d0-7b19a22cc5ce.0003.01/DOC_1.
- [3] Bywood, L., Georgakopoulou, P., & Etchegoyhen, T. (2017). Embracing the threat: machine translation as a solution for subtitling. *Perspectives*, 25(3), 492-508. <https://doi.org/10.1080/0907676x.2017.1291695>
- [4] Chen, S., Mansor, N. S., Mooisan, C., & Abdullah, M. A. R. (2023). Examining Cultural Words Translation in Tourism Texts: A Systematic Review. *Journal of Language Teaching and Research*, 14(2), 499-509. <https://doi.org/10.17507/jltr.1402.26>
- [5] Cintas, J. D., & Remael, A. (2020). *Subtitling: Concepts and practices*. Routledge. <https://doi.org/10.4324/9781315674278>
- [6] Duff, P. A., Norris, J. M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: Beyond art or science. *Tesol Quarterly*, 805-815. <https://doi.org/10.1002/j.1545-7249.2007.tb00105.x>.
- [7] Greco, G. M. (2019, July). Accessibility studies: Abuses, misuses and the method of poietic design. In *International conference on human-computer interaction* (pp. 15-27). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-30033-3_2
- [8] Gupta, P. Sharma, M., Pitale, K., & Kumar, K. (2019). *Problems with automating translation of movie/TVshow subtitles*. ArXiv. <http://dx.doi.org/10.48550/arXiv.1909.05362>.
- [9] Hamel, R. E. (2007). The dominance of English in the international scientific periodical literature and the future of language use in science. *Aila Review*, 20(1), 53-71. <http://dx.doi.org/10.1075/aila.20.06ham>
- [10] Karakanta, A. (2022b). *Automatic Subtitling a New Paradigm* [Doctoral dissertation]. Università Degli Studi Di Trento.
- [11] Kuo, S. Y. (2020). The Tangled Strings of Parameters and Assessment in Subtitling Quality: An Overview. *The Palgrave Handbook of Audiovisual Translation and Media Accessibility*, 437-458. https://doi.org/10.1007/978-3-030-42105-2_22
- [12] Lauscher, Susanne. (2000). 'Translation Quality Assessment: Where Can Theory and Practice Meet?'. *The Translator*, 6(2), 149-168. <https://doi.org/10.1080/13556509.2000.10799063>
- [13] Liu, Q. Y., Ang, L. H., Waheed, M., & Kasim, Z. M. (2022). Appraisal Theory in Translation Studies--A Systematic Literature Review. *Pertanika Journal of Social Sciences & Humanities*, 30(4). <http://dx.doi.org/10.47836/pjssh.30.4.07>

- [14] Macaro, E. (2019). Systematic reviews in applied linguistics. In *The Routledge handbook of research methods in applied linguistics* (pp. 230-239). Routledge.
- [15] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. <https://doi.org/10.1136/bmj.n71>
- [16] Pahlevan-Sharif, S., Mura, P., & Wijesinghe, S. N. (2019). A systematic review of systematic reviews in tourism. *Journal of Hospitality and Tourism Management*, 39, 158-165. <https://doi.org/10.1016/j.jhtm.2019.04.001>
- [17] Pedersen, Jan (2007). *Scandinavian Subtitles: A Comparative Study of Subtitling Norms in Sweden and Denmark with a Focus on Extralinguistic Cultural References* [Doctoral dissertation]. Stockholm University.
- [18] Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. Routledge. <https://doi.org/10.1080/0907676X.2015.1109906>
- [19] Wu, Z., & Chen, Z. (2021). A systematic review of experimental research in audiovisual translation 1992–2020. *Translation, Cognition & Behavior*, 4(2), 281-304. <https://doi.org/https://doi.org/10.1075/tcb.00056>
- [20] Yan, J. X., Pan, J., & Wang, H. (2018). *Research on Translator and Interpreter Training*. Springer Berlin Heidelberg. <http://dx.doi.org/10.1007/978-981-10-6958-1>
- [21] Yonamine, M. (2022). Domestication and foreignization in interlingual subtitling: A systematic review of contemporary research. *Translation & Interpreting*, 14(1), 198-213. <http://dx.doi.org/10.12807/ti.114201.2022.a11>

Altaf Fakh is a PhD candidate at Universiti Sains Malaysia. Her research areas of interest include machine translation, artificial intelligence, audiovisual translation, and translation quality assessment.

Mozhgan Ghassemiazghandi, PhD, is a Senior Lecturer at the School of Languages, Literacies, and Translation, Universiti Sains Malaysia. Her research focuses on translation technology, machine translation, and audiovisual translation.

Abdul-Hafeed Fakh, PhD, is Full Professor of Linguistics at the Department of English, Najran University, Saudi Arabia (and formerly at Ibb University, Yemen). He is a member of different editorial boards of indexed journals.

Masood Khoshsaligeh, PhD, researches the convergence of translation, multimedia, cultural adaptation, and technology.