Google Translate in Foreign Language Learning: A Systematic Review

Vahid Reza Mirzaeian[*1], Katayoun Oskoui [2]

[1] *Assistant Professor, Department of English, Faculty of Literature,*
Alzahra University, Tehran, Iran
[2] *MA, Department of English, Faculty of Literature,*
Alzahra University, Tehran, Iran

Abstract: Thanks to the significant achievements in Artificial Intelligence (AI), Machine Translation (MT), in general, and Google Translate (GT), in particular, have been extensively used in all facets of life, including language learning. However, faced with a plethora of research evidence on GT's educational contributions, erroneous translations create disparity regarding its use in language learning. To address this lacuna, the present study systematically reviewed 10 databases, namely, Web of Science, Scopus, ERIC, ScienceDirect, Taylor & Francis Online, Wiley Online Library, SAGE Journals, Springer Link, Springer Open, and DOAJ. Additionally, it hand searched the reference lists of 44 studies selected to be included in the synthesis from database search along with references cited in three previous systematic reviews on similar topics to capture a comprehensive view of the literature related to the use of GT in language learning between 2010-2021. It reviewed 50 studies witnessing a rise in the number of studies in this area. Studies reported that although significant improvements in the quality of GT led to pedagogical gains and more tendency to implement it in language learning, instructors still distrust it. Accordingly, this research provides pedagogical implications and suggests avenues for future research on the use of GT in language learning.

Keywords: Google Translate, Foreign language acquisition, Systematic review, CALL, Machine Translation.

Introduction

Emerging technological advances in the last decade have demonstrated a marked influence on the way language learning is viewed and practiced (Clifford, Merschel, & Joan, 2013; Knowles, 2022). Research findings in Computer Assisted Language Learning (CALL) have also documented the rewarding outcomes of technology employment in language pedagogy (Lin, 2021; Payant & Zuniga, 2022; Teng, 2022). As one category of CALL tools, Machine Translation (MT), most noticeably Google Translate (GT), is available to learners via smartphones, and computers (Darancik, 2016). The feasibility of MT as a CALL tool in language education has been apparent since the 1980s (Garcia & Pena, 2011).

Contrary to the long-lasting debates regarding the role of translation in language learning, mostly associated with the outdated 'Grammar Translation' method of instruction (Chang & Yamada, 2021), Fountain and Fountain (2009) argued that although translation must be considered a distinct skill, stakeholders must take into account the exceptional part it plays in language learning in its own right. Harmoniously, Darancik (2016) viewed translation, the 5th main language skill, as the catalyst for language pedagogy. Correspondingly, GT not primarily geared for educational purposes (Tsai, 2020), transformed the language learning landscape (Chung, 2020).

In 2006, Google launched an online translation engine to freely translate texts, speeches, websites, real-time videos, and images (Chon & Shin, 2020). Substantial leaps due to technological developments led to the creation of a new version called "Google Neural Machine Translation (GNMT)" in 2016 (Tsai, 2020) whose accuracy surpassed the previous algorithm. GNMT supports more than 103 languages (Le & Schuster, 2016). Figure 1 presents the rise in the number of GT-supported languages.
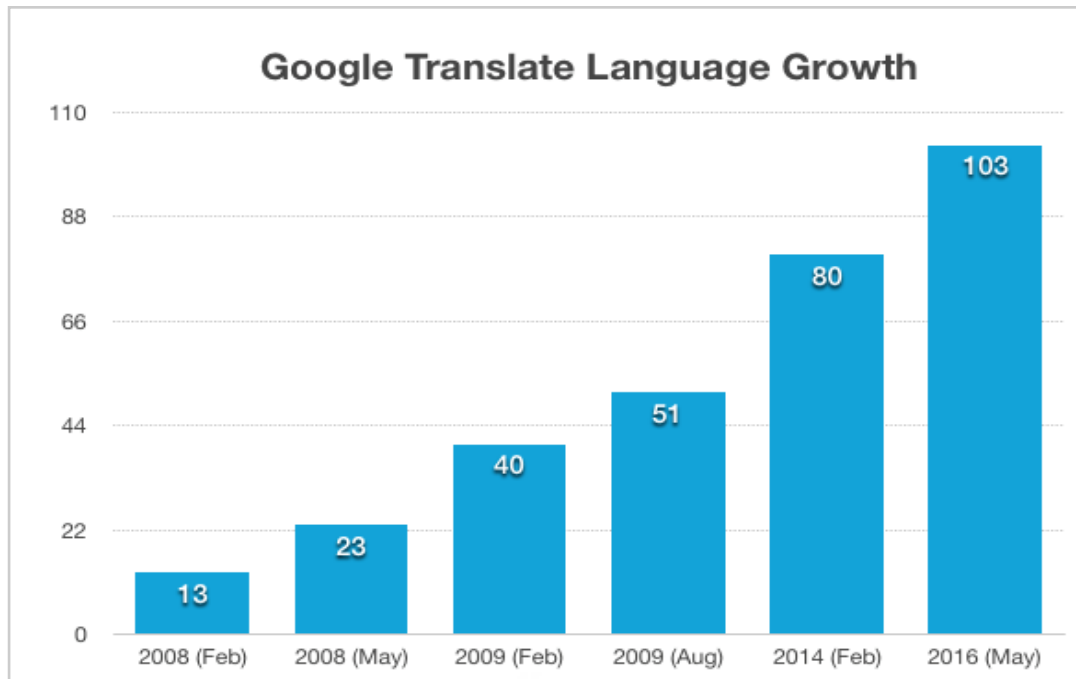
Figure 1. *GT Language Growth (Source: https://globalbydesign.com)*

Despite the mixed opinions about implementing GT in language learning (Stapleton, 2021), it has received a warm welcome and scaffolds learning experiences (Tourmen & Hoffmann, 2022). Given the growing number of studies on the use of GT in language learning, coupled with the unresolved disparities regarding its use, it is about time to conduct a systematic review to shed light on the status quo of its use in language learning. To elucidate this issue and assist in making sound pedagogical decisions, systematically reviewing research findings associated with the use of GT in language learning is essential.

To this end, the aim of the present study was to thoroughly explore a body of empirical research on the use of GT in language learning. The interest of this review article was mainly confined to studies investigating the use of GT in language acquisition and probing the effects of other MT platforms was beyond the scope of this systematic review. The considerable significance of this study is that although there are a couple of recent reviews concerning CALL (Fathali & Emadi, 2021), MT (Jolley & Maimone, 2022; Lee, 2021), and ELT (Khany & Kamalvand, 2022), to the best of the researchers' knowledge, this is the first systematic review conducted on the use of GT in language learning. By reviewing the current state of the literature on this newly emerged line of research, we wish to contribute to the existing literature and pave the way for future research in order to bridge the gaps and

provide insights into the use of GT in language learning for researchers and suggest practical pedagogical implications.

Review of the Literature

AI developments have equipped 21st-century classrooms with avant-garde programs including MT (Cancino & Panes, 2021). Research findings confirm the prevalence of GT in language learning (Stapleton, 2021; van Lieshout & Cardoso, 2022). The seed of GT was planted in 2004, but it was not until 2006 when Google launched its Statistical Machine Translation (SMT) algorithm. Due to emerging AI technologies such as deep learning and Natural Language Processing (NLP), Google introduced a radical departure, GNMT, in 2016. Although it employed both machine learning and human contributions to present more accurate translations (Stapleton, 2021), its outputs were not still exempt from flaws.

The early footstep of GT in language education was traced in the early 2010s when two researchers (Garcia & Pena, 2011) engaged Spanish learners in pre- and post-editing tasks to examine whether the GT-backended Tradukka interface helped them improve L2 writing. The results showed that beginners communicated better and their writing improved. Parallel with the above-mentioned quality improvements, researchers in different time spans reported multiple levels of accuracy for GT drafts, in turn influencing the way GT use in language learning was evaluated by stakeholders. The earlier studies reported more errors, and most of the practices were limited to post-editing (Case, 2015). However, GNMT compensated for most of the limitations concerning accuracy (Briggs, 2018).

A plethora of studies investigated the controversial use of GT in language learning. The holistic scenario was that regardless of the failures of GT coupled with existing preventive policies, learners worldwide consulted it due to its multilingualism, availability, immediacy, and cost (Maghsoudi & Mirzaeian, 2020; Murtisari et al., 2019). This issue raised two lines of debate within the language learning profession.

On the one hand, some scholars (Musk, 2014) held negative views of employing GT. According to Darancik (2016), this deceiving translation program harbored feelings of indolence in learners. Besides, she mentioned that gradual enrichment of GT leading to dissimilar outcomes upon the repeated requests of translating sentences would subsequently make students bewildered and unmotivated. Concerning ethicality, the research identified opposite trends. While some users took GT use as cheating, others viewed it as moral (Groves & Mundt, 2021). On the other hand, empirical studies reported positive effects of GT on writing improvement (Lee, 2019; Lee, 2020; Lin, 2021; Ryu et al., 2022), vocabulary

building (Chandra & Yuyun, 2018; Cohen & Wang, 2019; Cornell, Dean, & Tomaš, 2016; Ting & Tan, 2021), speaking (Klekovkina & Denié-Higney, 2022), and reading comprehension (Mirzaeian, 2020). As a result, the current situation forced us to wonder if it was possible for GT to serve as a pedagogical tool to facilitate language learning in the new millennia.

Taking these issues as points of departure, it appeared to be essential to provide an overview of research studies on the use of GT in language instruction to illuminate the current state of affairs and provide sound pedagogical implications (Jin & Deifell, 2013). As Cancino and Panes (2021) noted, GT was one of the most employed MTs across all languages, mostly for reading and writing purposes (Stapleton, 2021). A number of studies have concertedly presented GT as a viable resource for reading and writing promotion alike (Murphy, 2020).

According to Grabe and Zhang (2013), reading comprehension is the key to better execute reading and writing tasks. Recently, Mirzaeian (2020) explicated the comprehensibility and practicality of MT outputs among Iranian EFL university students. The remarkable educational implication of this study was that despite the long-term uncertainties regarding the use of MT in higher education, learners were advised to conveniently consult GT. This study aligned with that of Karnal and Pereira (2015) who documented the effectiveness of GT on reading comprehension. Although applied linguistics research verified the reciprocal impact of reading and writing as key linguistic elements (Dülger, 2021), given the pivotal role of writing in communication (Murphy, 2020), the complex relationship between GT and writing has received great attention (Chen, 2020; Lee & Briggs, 2021; Lin & Morrison, 2021).

Juxtaposing the GT drafts with Human Translations, Tsai (2019, 2020) backed up the idea that GT supported the writing process by producing higher-quality texts. Although at times researchers claimed that not only AI-based GT was acceptable enough to be used as an educational technology (Mirzaeian, 2020), also the quality of its hypertexts was comparable to that of human translations (Murphy, 2020), human texts were found to be more accurate (Egamberdievna & Daminovna, 2020). Hence, some scholars believe that students should not place blind faith in MT outputs and need to identify linguistic errors (White & Heidrich, 2013). Correspondingly, research on the limitations of GT has brought shortcomings to the forefront. Register problems (Maghsoudi & Mirzaeian, 2020) and grammatical inaccuracies (Chen et al., 2019) are among the examples of the reported downsides of this service.

In order to eliminate the existing problems, it is suggested that users manage the revision process of texts via two alternative approaches, namely, pre- and post-editing. The majority of the studies about post-editing drafts have been conducted before the launch of GNMT in 2016. Obviously, today the outputs contain fewer errors. Therefore, the research results have to be interpreted with greater care (Maghsoudi & Mirzaeian, 2020). Although error correction is a major side of the text revision process, the findings confirm that individual learner differences, including language proficiency, affect the extent to which learners are able to improve text quality (Xu, 2022). Chung (2020) highlighted the influence of L2 command and the ability to post-edit GT drafts and concluded that the number and level of corrections are influenced by the participants' competence.

In reviewing the literature, it has commonly been affirmed the efficiency of GT for proficient language learners (Clifford et al., 2013). In contrast to these findings, in a study on beginner Spanish learners, Garcia and Pena (2011) reported that GT helped students to improve their writing. What is worthwhile in our review, however, is that according to a Google report in 2010, language learners used GT to look up words, learn to write and speak, etc. (Garcia & Pena, 2011). In fact, the overall attitude of software users, namely, language learners reflects the practicality of MT; not the software itself (Levy, 2009).

Evaluating GT use from learners' and instructors' viewpoints alike has been the focus of some studies (Briggs, 2018; Case, 2015). The results of the surveys indicated that both drawbacks (Mundt & Groves, 2016) and benefits (Chon, Shin, & Kim, 2021) of GT form the two sides of the same coin factor to discrepancies among users (Briggs, 2018). It is crucial to note that integrating GT in language learning is a learner-centered approach (Tsai, 2019), and as Fredholm (2015) believed GT use per se does not give rise to linguistic improvements but learners' command of language along with their competence in technology use contribute to software efficiency. In view of the fact that free MT technologies have been present since the 1990s (Groves & Mundt, 2021) and are already used in language learning (Chon & Shin, 2020), it is imperative for educators not to take GT for granted and embrace it in real-world language learning territory (Groves & Mundt, 2015).

The foremost motivating factor of this systematic review, therefore, is that while there have been systematic reviews on CALL (Kalyaniwala & Ciekanski, 2021; Klimanova, 2021), MALL (Alzubi, 2021; Hou & Aryadoust, 2021), and MT (Lee, 2021), no study systematically targeted the use of GT in foreign language learning. Moreover, this field of study is still in its infancy and there are definitely several questions regarding the position of MT in foreign language learning that still have remained unanswered. To address this

research gap, we wish to contribute to the existing literature by enhancing our understanding of this under-researched area, providing constructive educational implications, and suggesting prospective research opportunities. As a result, the following six research questions were posed in this study:

1. What is the publication trend on the use of GT in language learning between 2010-2021?
2. What are the characteristics of the reviewed studies?
3. What are the purposes of using GT in language education?
4. What are the main research outcomes of the reviewed studies?
5. What are the educational implications of the reviewed studies?
6. What are the limitations of the reviewed studies?

Research Methodology

The present study focused on the use of GT in foreign language learning. To address research questions, this study employed a systematic review search strategy based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Page et al., 2021), and the PRISMA explanation and elaboration (Page et al., 2021). Following a protocol, a systematic review thoroughly selects and synthesizes accessible studies relevant to target review questions through transparent and replicable processes (Macaro, Handley, & Walter, 2012). This unbiased method provides an all-inclusive and clear account of a topic to end users (Diacopoulos & Crompton, 2020). Other systematic reviews on CALL also employed this approach (Cinquin, Guitton, & Sauzeon, 2019; Hou & Aryadoust, 2021). In addition, the instructions of well-known reference books on systematic reviews were consulted (Gough, Oliver, & Thomas, 2017; Littell, Corcoran, & Pillai, 2008) and a number of recent systematic reviews published in distinguished CALL journals such as Computers & Education, System, and Computer Assisted Language Learning (CALL) were also taken into consideration (Booton, Hodgkiss, & Murphy, 2021; Hou & Aryadoust, 2021).

Search Strategy

To reduce research bias, the search strings and inclusion/exclusion criteria for further evaluation and selection of the papers were specified. Moreover, in order to cover the required breadth of data to ensure reliability, we executed both an electronic database search and an additional hand

search of the reference lists of all the studies selected to be included in the in-depth review from THE database search, plus references cited in a number of previous systematic reviews on similar topics (Abdel-Reheem, 2020; Bowker, 2020; Lee, 2021) to cover further viable empirical research evidence on the investigation of GT use in foreign language learning.

To place a limit on the number of results and due to the authors' English language proficiency, the articles written in English were found eligible. There are definitely other valuable studies in languages other than English that have been overlooked. The twelve-year time span (i.e., 2010-2021) was specified because the database search result showed that it was during the early years of the 2010s that the use of GT in language learning was first investigated (Garcia & Pena, 2011). Besides, in order to keep track of the observable gradual increase in the number of studies in the literature and stay updated, reviewing this timeframe seemed mandatory. Moreover, considering the fact that most systematic reviews exclude non-empirical records to conduct a synthesis of investigations (McKinley & Rose, 2019), the present research also excluded non-empirical papers such as books and blogs. As mentioned earlier, the interest of this systematic review was merely confined to the studies investigating the use of GT in language education, and probing the effects of other educational technologies and/or other MT platforms was beyond the scope of this study. Moreover, the studies mainly concerned with the application of GT in translation and translation education were not included in the review. As Lee (2021) mentioned, while such investigations focused on other research objectives, their findings did not lead to implications for foreign language education.

Following multiple previous CALL systematic reviews, articles were retrieved from an electronic search of 10 major academic databases, namely Educational Resources Information Center (ERIC), Web of Science (WOS), Scopus, Science Direct, Springer Link, Springer Open, Taylor & Francis Online, SAGE Journals, Wiley Online Library, and DOAJ.

The multidisciplinary essence of the topic led the way for choosing and reviewing these databases. The searched terms used for this review were divided into two groups. The first cluster of terms related to MT included 'Google Translate' and 'Machine Translation'. The other collection of words about language learning involved 'language teaching' and 'language learning'. These keywords were selected based on both CALL literature and ERIC Thesaurus, and the Boolean operators would let the database search for any probable result by combining the search phrases of both aforementioned groups. The following search strings were used to perform the search:

("Google Translate" OR "machine translation") AND (" language teaching" OR " language learning").

Inclusion/Exclusion Criteria

We optimized and consented to the following inclusion and exclusion criteria (Table 1) to identify the most relevant research papers and ensure quality standards.

Table 1. *Inclusion and Exclusion Criteria*

Inclusion criteria
- Empirical   - English   - Published between 2010-2021   - Studies focusing on the use of Google Translate in foreign language teaching and learning
Exclusion criteria
- Non-empirical   - Non-English   - Studies on the application of educational technologies and/or machine translation platforms other than Google Translate   - Studies on the application of Google Translate in Translation and translation education

Study Selection

Throughout this procedure, we kept logs to record search dates, database names, the type of filters applied to each database where available, and the number of records per search. Agreements on unclear points or probable conflicts were achieved through online meetings and discussions. The initial database and hand searches yielded 2783 results. The researchers gathered 1492 records from an electronic database search between 9-16 October 2021. Table 2 provides a summary of the number of research papers retrieved from each database.

Table 2. *Distribution of Identified Records in Databases*

Database	Frequency
Scopus	496
Taylor & Francis Online	378
Science Direct	211
Wiley Online Library	184
ERIC	56
Web of Science	56
SAGE Journals	39

Springer Link	34
DOAJ	27
Springer Open	11
Total	1492

To organize the collected data and remove duplicates, we exported the results to the Endnote web reference manager. Through the deduplication processes, 559 duplicate records were excluded (497 records were identified by Endnote web software, and 62 records were found by manual checking). Further refinements resulted in relevant articles based on the inclusion/exclusion criteria to answer the research questions. Following these steps, only 44 research articles fulfilled the inclusion criteria and were included in the qualitative synthesis. The authors independently reviewed the papers against the inclusion and exclusion criteria, resulting in an inter-reviewer agreement of 96%. Disagreements were further discussed and 100% agreement was achieved.

In addition, one researcher hand searched the reference lists of 44 eligible studies along with references cited in three previous systematic reviews on similar topics (Abdel-Reheem, 2020; Bowker, 2020; Lee, 2021), between 17-22 October 2021. Collectively, 1291 research papers were identified from these processes. In this regard, 1159 studies were identified from 44 eligible studies and 132 records from previous systematic reviews. The Microsoft Excel worksheet was used to compile the returned results and run the deduplication process. Of these documents, through the first step of screening, 746 papers were found ineligible. The remaining 545 items entered the second or final step of screening and 6 papers were included in the qualitative synthesis. Finally, from the initial 2783 returned results, 50 articles were deemed suitable to enter into the in-depth qualitative synthesis to answer the research questions. The study selection process is detailed in Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021) (Figure 2).
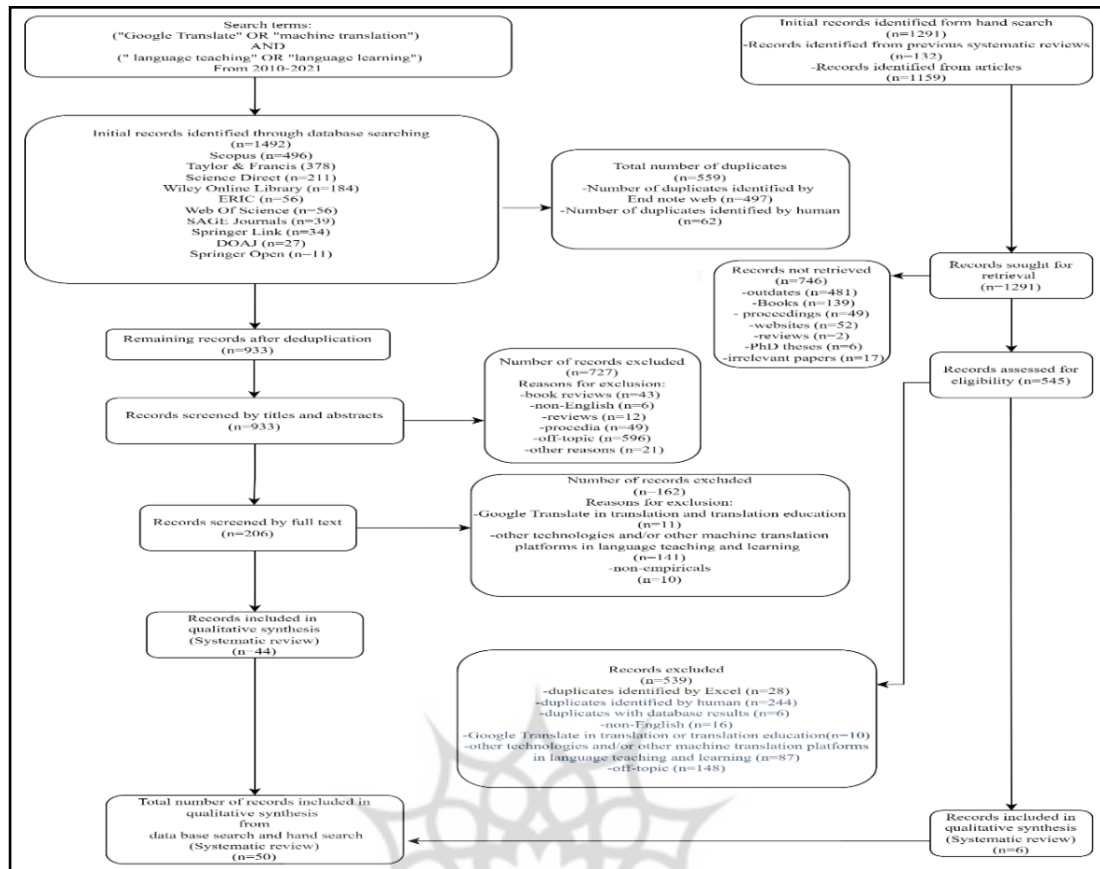
Figure 2. *PRISMA Flowchart*

Quality Appraisal

In this review, 50 studies were selected to answer the research questions. The main author and an expert researcher independently performed the critical appraisal of the quality of each individual paper. To guarantee the quality of the selected papers, the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018) was adopted. Based on the suggestion of the new version of MMAT, we did not use the metrics to present the results in the form of uninformative categories or single numbers.

Data Coding

Having identified the final 50 research papers, the next step was to qualitatively analyze the content of the papers to answer the research questions. Firstly, we entered the collected data into the Microsoft Excel program. Then, we conferred on the coding process and selected six main variables for analysis based on the research questions, and a primary screening of the papers included in the review: 1) publication trend, 2) characteristics of the reviewed studies

(i.e., the educational settings, and the research methods), 3) research purposes, 4) main research outcomes, 5) educational implications, and 6) limitations.

Results

In this section, we provide the answers to the research questions representing the results of this systematic review.

What is the publication trend on the use of GT in language learning between 2010-2021?

A total of 50 research papers were published in 34 journals (Figure 3); CALL journal published the majority of the articles (N = 6), System, English for Specific Purposes, and ReCALL each published three articles, followed by The JALT CALL Journal, The Journal of Asia TEFL, English Teaching & Learning, CALL-EJ, and Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas each published two articles. Other journals each hosted one article (Fig. 3).
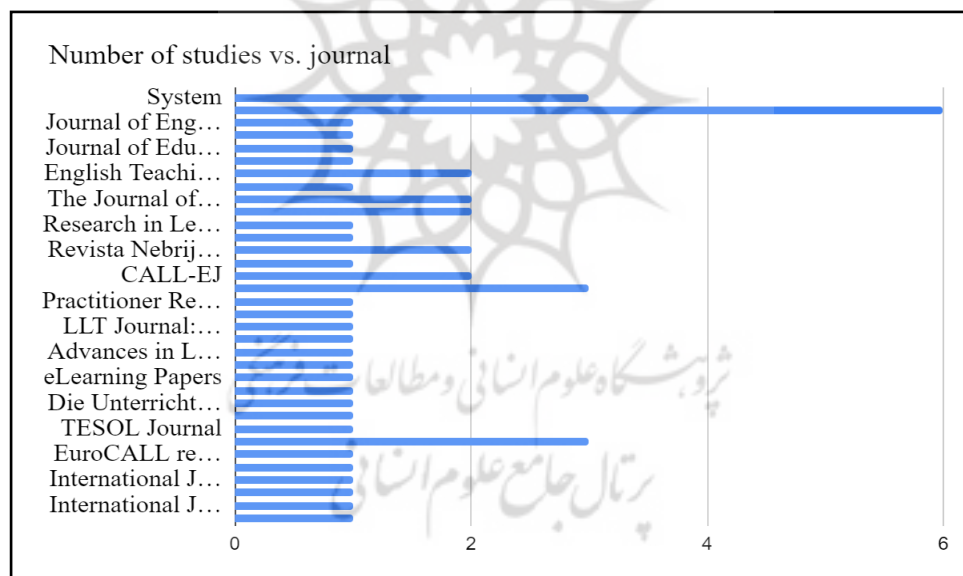
Figure 3. *The Title of the Journals in Which the Reviewed Papers Were Published*

As mentioned earlier, the first research paper in this area appeared in 2011 (Garcia & Pena, 2011). However, we witnessed a rise in the number of papers from 2019 to 2021. Lee (2021) congruently reported a slight growth in the number of articles between 2019 and 2020 on the effect of MT on language education. The results showed that of the 50 available articles, eleven studies were published in 2021 up to the time of our data collection in October 2021, ten in 2020, twelve in 2019, five each in 2018, and 2015, two studies each in

2017, 2016, and 2013. Finally, only one paper was published in 2011. No empirical studies were conducted in 2010, 2012, and 2014. The increasing count of the publications indicates that this line of research is in its infancy, but has gained considerable attention in recent years (Figure 4).
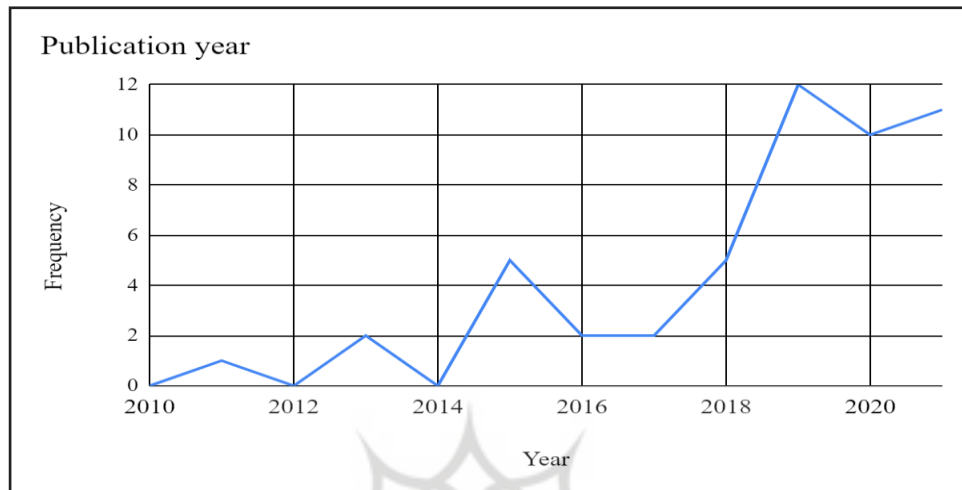
Publication year

Figure 4. *Publication Years*

Regarding the country of origin, Figure 5 shows that South Korea and The United States each possess the most research conducted in this area (N = 7), followed by Hong Kong (N = 5). Other countries carried out fewer studies. Luef, Ghebru, and Ilon's (2019) claim regarding South Korea confirms our findings. According to them, South Korea is one of the most technologically advanced countries, possessing the fastest and broadest internet connection, with a great tendency to incorporate mobile apps into daily activities including education. In addition, foreign language learning is an inevitable requisite in this country. Undoubtedly, as Korean language learners prefer to consult emerging technologies including GT, more scholarly publications will be cultivated.
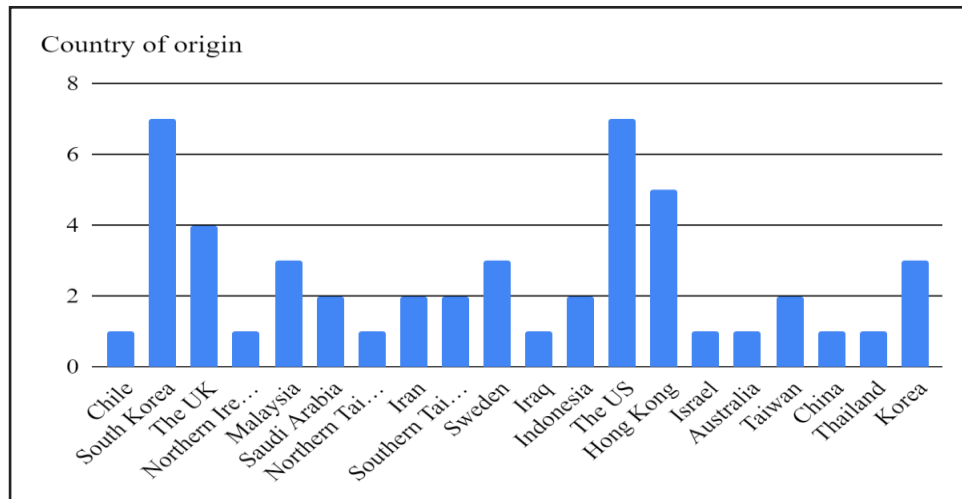
Figure 5. *Country of Origin*

What are the characteristics of the reviewed studies?

The characteristics of the reviewed studies the included educational setting and research method.

Educational Setting

As the demographic characteristics of the reviewed studies illustrate, the majority of studies (87.9%) were carried out in higher education where university students served as participants of the studies (Fig. 6). The associated literature provides multifold reasons for this selection. According to Groves and Mundt (2021), improvements in GT output quality has led to its ubiquity to the extent that academic journals such as Taylor & Francis have embedded GT widget to provide on-the-spot and aboriginal translation of scholarly content. They also remind us of the influential role of MT in English-speaking higher education courses that host a large number of multilingual students worldwide.
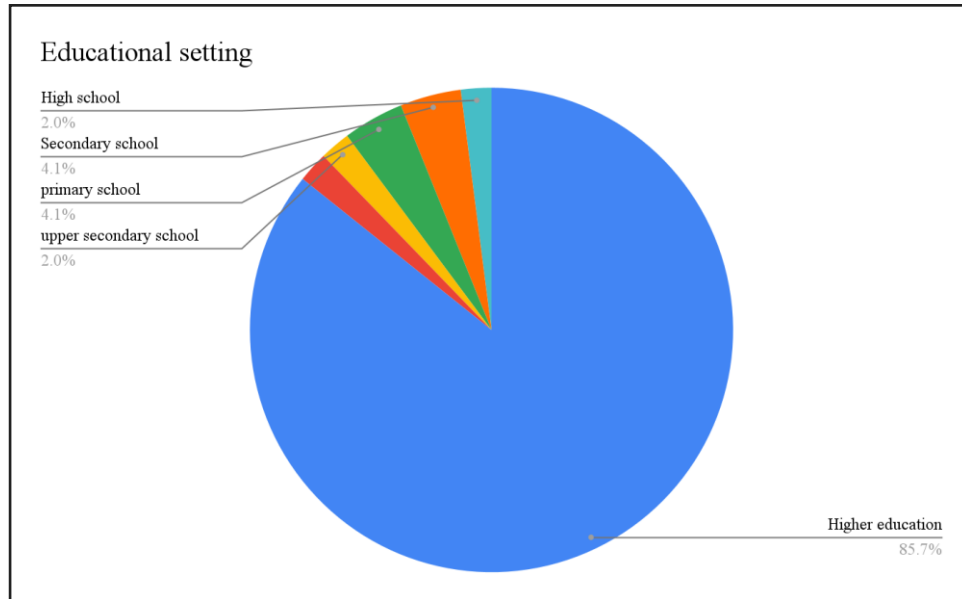
Figure 6. *Educational Setting*

Research Methodology
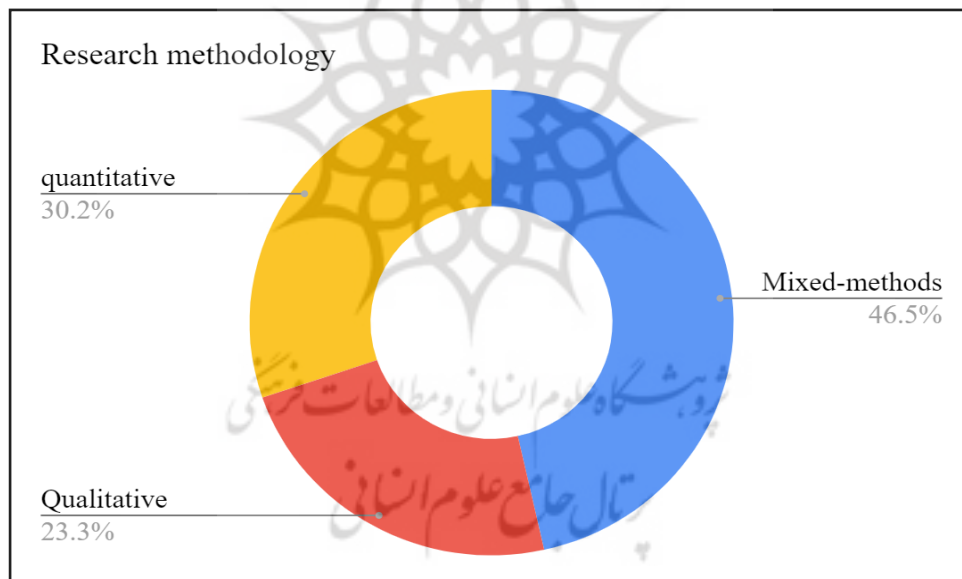
Figure *7. Research Methodology*

As presented in the 3D pie chart of Fig. 7, most of the studies used mixed-methods (46.5%), followed by a quantitative approach (30.2%), and finally qualitative studies (23.2%). Given that mixed-methods (Chung, 2020) provide both statistical analyses and more in-depth accounts of participants (Groves & Mundt, 2021), these findings seem logical. In some cases, the adopted research methodologies were not clearly stated (Nino, 2020). Due to the probable misidentification of methods, we found it sufficient to merely mention them.

What are the purposes of using GT in foreign language education?

Having analyzed the objectives of the selected papers, we classified research purposes into the following topics: 1) Investigating the effect of GT on different aspects of language acquisition (N = 34, 68%). The majority of articles were concerned with the effect of GT on L2 writing (N = 24, 48%). Note that some of them focus on more than one variable. 2) Investigating the effect of language learners' proficiency on post-editing GT outputs (N = 1, 2%). 3) Investigating GT end users' perceptions (N = 13, 26%). These articles discussed issues from language learners, instructors, and academic staff. Some of the studies pursued more than one of the identified objectives.

As mentioned above, L2 writing has attracted the most attention of researchers in the field of machine translation with about 48% of the reviewed articles. Most recently, Cancino and Panes (2021) evaluated the linguistic features of narrative written tasks. Besides, a number of scholars focused on the effect of GT on other features of foreign language learning. Maghsoudi and Mirzaeian (2020) inquired into the relation between GT use and L2 reading comprehension. Considering the importance of vocabulary learning (Nation, 2001), other publications advanced the application of GT in vocabulary acquisition (Chen, 2020; Fredholm, 2019; Kol, Schcolnik, & Spector-Cohen, 2018). Most recently, the practicality of GT and its text-to-speech synthesis (TTS) and automatic speech recognition (ASR) components for independent Dutch vocabulary learning and pronunciation practice have been investigated (Van Lieshout & Cardoso, 2022).

Regarding the highly disputable interaction between language proficiency and post-editing capabilities, only one study exclusively paid attention to this topic (Chung, 2020). Ultimately, given the disagreements among language learners and instructors arising from the prevalence of GT, multiple investigations sifted GT end users' perceptions to spell out the current status quo. Lately, Groves and Mundt (2021) interviewed the faculties of two UK universities and discussed some issues of concern such as MT policy-making and legitimacy in the academic context. Kelly and Hou (2021) interviewed teenage multilingual pupils and their teachers in Northern Island.

What are the main research outcomes of the reviewed studies?

We presented the main research outcomes in harmony with the above-mentioned research purposes. Based on our findings, language learners with any linguistic background use GT extensively for educational and daily activities despite its drawbacks and external

prohibitions. They expressed satisfaction with its multilingualism, convenience, and being free of charge. Generally, they found it a versatile tool for language learning. Although O'Neill (2019a) reported the frequent use of GT for completing assignments before the onset of Covid-19, research admitted more intensified use of GT for multiple educational purposes during the pandemic (Bailey & Lee, 2020; Bin Dahmash, 2020). GT has also been a productive resource for Independent Language Learning (ILL) (Lai & Zheng, 2018; Nino, 2015, 2020).

As mentioned above, most of the studies (N = 24, 48%) of the first group examined the effect of GT on L2 writing. Research confirmed that students perceived GT as a time-saving writing assistant to translate at word level and beyond, improving accuracy, fluency, and cohesion. Another outstanding characteristic of GT was that it provided individualized and on-the-spot feedback comparable to teachers' corrections or peer feedback. Definitely, it does not substitute instructors or the language learning process itself. Psychologically speaking, GT reduced writing anxiety and improved learners' self-efficacy, confidence, and motivation. In parallel, a number of studies reported the plausible effects of the use of GT in vocabulary learning (Chang & Yamada, 2021), reading comprehension (Maghsoudi & Mirzaeian, 2020), and language learning in general (Clifford et al., 2013).

According to Tsai (2019), text genre is one of the factors that influences the quality of GT drafts and resulting research outcomes. Research on the effect of text genre on language learners' use of GT indicated that while it made more fluent and longer argumentative texts containing more complex syntactic structures, it affected the accuracy and lexical enrichment of narrative texts (Chung & Ahn, 2021).

Another factor that highly affects the practicality of GT is language pairs (Nino, 2020). Tarsoly and Valijärvi (2019) claimed that Finnish and Hungarian learners preferred to use GT to 'get the job done' rather than learning a morphologically complex language. GT as a practical assistant has proven to be influential in other aspects of language acquisition as well. For instance, university students found GT beneficial for learning Bahasa Malaysia. They believed that this service was appropriate for vocabulary learning, writing, and reading (Bahri & Mahadi, 2016). In total, they found it necessary and beneficial for language learning.

The second group is made up of one work that especially highlighted the impact of learners' proficiency and their post-editing capabilities (Chung, 2020). In her work, Chung (2020) tracked this issue while investigating the relationship between learners' proficiency and their post-editing

capabilities along with the level of errors they were able to detect and correct. She observed that word-level errors were recognizable and editable for all learners. Consistent with another researcher (Briggs, 2018), she contended that as low-skilled learners lacked the required lexical knowledge to distinguish literal mistranslations, they tended to use GT drafts with minimum modifications. Furthermore, although low-skilled and intermediate learners attempted to differentiate phrase and clause-level errors, they preferred to deal with word-level errors.

In contrast, high-skilled learners as critical users of GT found additional errors even above the word level and favored full-sentence revisions due to their familiarity with nuanced lexical connotations. In a nutshell, the more proficiency, the better error detection, and correction. Nevertheless, according to our findings, researchers have not come up with an agreed-upon mutual relation between language proficiency and error management. Collectively, as Briggs (2018) stated, GNMT has satisfied learners at all linguistic levels.

Several studies identified lexical and grammatical errors influencing text quality and comprehension (Groves & Mundt, 2015; White & Heidrich, 2013). However, the highly improved GNMT has resolved most errors of the earlier versions (Briggs, 2018). Although its accuracy is comparable to that of human translations in most languages, it is still far from perfect and must be used cautiously. This service is still unable to provide accurate equivalents for jargon, idioms, double meanings, grammatical equivalents, or proper names. It also fails to recognize registers and lacks textual features such as coherence, cohesion, and co-reference.

Additionally, insensitivity to cultural and contextual connotations, as well as sociolinguistic, language variety, and genre-specific issues are problematic for GT. Therefore, it is not capable of substituting for a professional translator. However, findings confirmed that GT's speech-enabled language translation (SELT) and computer-aided-translation (CAT) features have facilitated two highly valuable educational objectives, namely cross-cultural understanding and intercultural sensitivity (Shadiev, Sun, & Huang, 2018).

The time of conducting studies, input (i.e. source text) quality, and language pair determine the number and type of reported output flaws (Habeeb & Muhammadb, 2020). Correspondingly, according to a number of scholars, as language learners were aware of the value of pre-editing practices for more accurate outputs, they attempted to use lexically and grammatically simpler and more refined inputs for smoother translation.

With regard to the last group of the studies in which language learners, instructors, and academic staff were interviewed to elicit their opinions about using GT for foreign language acquisition, most recently, academic staff at two UK universities claimed that given that no explicit policy has ever been enacted around this issue and available regulations precede the latest MT improvements, serious negotiations pertaining to legal cases including misconduct and policy-making along with graduates expected linguistic competence and their efficacy in professional contexts seem crucial.

Generally speaking, academic staff confirmed using GT for receptive skills such as reading and did not consider it as academic misconduct in terms of academic integrity, but were mostly worried about GT use for productive skills such as writing. Consequently, not only such a prescriptive dictum is unlikely to be widespread and fully implemented, but as well it may also threaten learners' honor and self-confidence in academic contexts (Groves & Mundt, 2021).

Last but not least, Bin Dahmash (2020) provided noteworthy accounts of Saudi Arabian university students' personal histories with GTA. Learners initially downloaded GT based on parents' or friends' advice or they might have found some of its features such as 'text image' translation interesting. As these learners thought that GTA acted as a free English teacher, they valued it as a necessary application on smartphones for conversations and language learning.

Overall, language learners, instructors, and faculty members were not in unanimous agreement concerning the part GT played in foreign language learning. Most learners believed that it was their legal, civil, and natural right to make informed choices of their own learning supports and that universities were responsible to set the scene for them. In agreement with these learners, several faculty members, as some researchers (Clifford et al., 2013) claimed, perceived GT as an appropriate support and language teacher assistant mostly for advanced language learners rather than a threat to the profession. Simultaneously, they were of the opinion that they should remind their learners of its shortcomings.

Unlike the proponents of GT, some instructors and faculties were opposed to using it (Organ, 2022). They viewed it as cheating and plagiarism. Moreover, they accused GT of making learners unmotivated and lazy due to over-reliance on it, leaving the outputs unedited and accepting them as they were and consequently leading to language avoidance.

Essentially, most of the scholars believed that instructors and faculties had to be aware that GT was an undeniable fact (Garcia & Pena, 2011). What they were supposed to do was

that instead of futile restrictions, as their learners asked (Alhaisoni & Alhaysony, 2017), they needed to work on empowering language learners for more productive use of this valuable service (Cancino & Panes, 2021; Nino, 2020). The answer to the next research question will shed light on this crucial issue.

What are the educational implications of the reviewed studies?

With regard to the ever-growing number of international students, incorporating GT as a source of reference for future generations into classrooms and enhancing learners' digital literacy seem crucial (Groves & Mundt, 2021). Research findings have proven that due to the importance of GT on the web (Garcia & Pena, 2011), it continues to improve (Ducar & Schocket, 2018). This time and energy-saving learning strategy lays the ground for more linguistic support for the act of language acquisition. Hence, it should not be taken for granted and be embedded into linguistic practices (Fredholm, 2019). Most importantly, educational institutions and instructors have to welcome this emergent technology (Groves & Mundt, 2021). What they all need to be fully aware of is that as GT empowers itself based on human-generated corpora, its excellent rendition production is guaranteed (Stapleton & Kin, 2019).

It seems that contrary to the common beliefs that viewed existing errors as the biggest threat to language learning, these high-quality outputs have been the source of concern (Case, 2015). According to some researchers (e.g. Maghsoudi & Mirzaeian, 2020), GT output quality and comprehensibility are comparable to that of human translations. Today, it is almost impossible to detect learners' GT use. Therefore, any preventive policy or objection is subject to total failure. Language learners' independent interactions with technology and authentic materials alter the essentials of language acquisition which in turn reforms the role of educational organizations to fulfill the requirements of their learners (Case, 2015). Moreover, using GT in language learning is not just simply a matter of consulting software for translations. In essence, as Vinall and Hellmich (2022) stated, machine translation puts forth contextual, cultural, and pragmatic issues in meaning construction which in turn affect further methodological and curricular policies and practices.

To manage this paradigm shift, it is imperative that initially instructors take into account the incremental advances in AI technologies and attain a thorough understanding of GT to bring it to their learners' service (Tsai, 2020). It is better to discuss issues related to

online technologies to identify learners' needs and objectives (Nino, 2020). They have to be mindful that language learners are dealing with GT (Fredholm, 2019). Instead of viewing it as taboo, educators should encourage the optimal use of this novelty (Groves & Mundt, 2015).

In order for more productive planning, instructors should consult learners to design more practical tasks to expose learners to natural language use. Moreover, discussing the pros and cons of this tool with learners, especially low-skilled ones, makes them aware of probable errors and teaches them more critical application of GT (O'Neill, 2019b; Wuttikrikunlaya, Singhasiri, & Keyuravong, 2018). In return, instead of relying on GT, language learners should improve their first and target language competence to be able to use this service more responsibly and productively. We hope that these implications help administrators, instructors, and learners to make more informed decisions and achieve greater educational objectives in the future.

What are the limitations of the reviewed studies?

The efficiency of GT for language learning is anticipated to be momentous over time. Hence, addressing the limitations of the reviewed studies is of great value. Consistently, we observed that about half of the studies (N = 23) did not detail their limitations. Of the remaining 27 articles, more than one limitation was reported in 20 articles, and the rest (N = 7) indicated only one. The identified limitations can be outlined as follows: a) small sample size and its consequent lack of generalizability of the findings are the most typical gaps reported by the reviewed studies (N = 11); b) the majority of the studies have been conducted in higher education and university students are commonly recruited as participants; c) the only data collection method employed to investigate whether GT acts as a shortcut to produce more improved texts or contributes to the acquisition of writing skills was a survey; d) most of the studies have looked into the effects of GT on writing skills and exploring the impact of this service on other linguistic skills is underestimated; and e) not much effort has been made to investigate the attitudes and perceptions of language instructors' and administrators' regarding the use of GT in language learning.

Discussion

The purpose of this research was to systematically review the publications on the use of GT in language learning. Through the succeeding section, we provide a synopsis of evidence of the reviewed studies.

Synopsis of Evidence

The obtained data demonstrate that there has been a slight increase in the number of studies between 2019 and 2020 on the use of GT in language learning. This trend is in line with Lee (2021), who reported a steady rise in the frequency of publications on the effect of MT on language education. A strength of the evidence for this systematic review is that almost 74% of the reviewed studies have been published in 22 reputable journals indexed by InCites Journal Citation Reports (JCR) via Web of Science or Scopus (see Appendix A). The evidence from the reviewed studies indicates that 85.7% (N = 42) of the studies were conducted in higher education. This aligns with views expressed by some researchers (Abdel-Reheem, 2020; Lee, 2021) who found that most investigations were organized in universities and involved university students as subjects.

Considering the essential role of English in international academic investigations and its use as the medium of instruction in tertiary education (Bowker, 2020), it is not surprising that English is the most frequent target language of the reviewed studies (N = 33). This finding mirrors Lee (2021), who reported English (N = 57) as the most focused L2. The most striking congruency with the previous systematic reviews is that most of the reviewed studies employed mixed-methods (46.5%) for data collection, followed by a quantitative approach (30.2%), and finally qualitative studies (23.3%). Accordingly, Lee (2021) realized that a mixed research method was the most preferred approach (N = 42), followed by quantitative (N = 25) and qualitative (N = 20) methods. In total, despite the fact that GT end users expressed mixed opinions and perceptions, research findings confirm the practicality of this service for language learning from linguistic, affective, and cognitive perspectives.

The improvement of language skills through the use of GT has been proven. It has been revealed that GT boosts learners' self-efficacy and confidence. Considering together, in agreement with the previous reviews, we found that given the pivotal part of GT in enhancing different aspects of language learning, much more attention has to be paid to the extensive introduction and informed application of this profitable CALL tool. This study refers to this

issue as '*GT literacy'*. Moreover, we suggest the concept of *GTALL* (GT Assisted Language Learning) for the use of GT in foreign language learning.

Conclusion

Based on the answers to the proposed research questions, we concluded that GT has proven to be useful for enhancing language learning. Learners with diverse linguistic competence are aware of the strengths and weaknesses of GT. Yet, they have accepted it as a valuable aid during their language-learning journey. Instructors and administrators, on the other hand, are more skeptical concerning this attitude. Some of them assume using GT as plagiarism or cheating. In addition, no official policy has even been enacted to act upon. The fact is that ongoing AI improvements have already turned GT into a reliable and inevitable entity. Therefore, we recommend rich introductory and training sessions to take more advantage of this reality in lieu of any availing restrictions.

What motivated us to carry out a systematic review was that a number of factors such as advances in GT performance, learners' proficiency level, language pairs, and text genre led to inconsistent results in sole investigations. This systematic review provided an accumulative and in-depth outlook of the current individual studies concerning areas such as research purposes, methodological approach, and outcomes. Moreover, it created an opportunity to pinpoint their directions for future research to reveal the real place and value of this novelty for language learning.

A summary of the probable suggestions for future investigations is presented below:

1) Probing into the uptakes of GT in diverse educational settings other than academia,

2) Progressively evaluating the quality of learners' GT-supported texts by undertaking case studies and longitudinal studies accompanied by data collection methods such as screen casts, think-aloud protocol, and stimulated recall interviews,

3) Paying more attention to the effects of GT on other skills including reading comprehension, and translation,

4) Inquiring about the worthwhile self-reports of the 21st-century language instructors and administrators about the direct and indirect effects of GT on language skills through interviews and more structured surveys, and

5) Conducting interviews and focus group discussions with larger populations of learners with different linguistic competence to clarify the rationale behind their reliance or better say over-reliance on GT in face of continuous improvements and ubiquity as

well as existing drawbacks that lead to the discovery of the probable evolutionary paths in MT use parallel to their linguistic growth.

We hope that the findings of this study add to the growing body of literature on this relatively recent avenue of research to enhance GT end users' understanding of the status quo to make more informed pedagogical decisions and integrate this novelty into classroom activities for more educational achievements.

Declaration of Conflicting Interests

The authors of this study do not have any conflicts of interest to declare.

References

Abdel-Reheem, E. (2020). A review of research into Google Apps in the process of English language learning and teaching. *Arab World English Journal (AWEJ)*, *11*(1), 399-418. https://dx.doi.org/10.24093/awej/vol11no1.27

Alzubi, A. A. F. (2021). The Role of Mobile Technologies in Impacting Learner Autonomy in an EFL Context: A Systematic Review. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *11*(3), 56-73. https://doi.org/10.4018/IJCALLT.2021070104

Booton, S. A., Hodgkiss, A., & Murphy, V. A. (2021). The impact of mobile application features on children's language and literacy learning: a systematic review. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2021.1930057

Bowker, L. (2020). Chinese speakers' use of machine translation as an aid for scholarly writing in English: a review of the literature and a report on a pilot workshop on machine translation literacy. *Asia Pacific Translation and Intercultural Studies*, *7*(3), 288-298. https://doi.org/10.1080/23306343.2020.1805843

Cinquin, P., Guitton, P., & Sauzeon, H. (2019). Online e-learning and cognitive disabilities: A systematic review. *Computers and Education*, *130*, 152-167. https://doi.org/10.1016/j.compedu.2018.12.004

Darancik, Y. (2016). The Effect of Data-Based Translation Program Used in Foreign Language Education on the Correct Use of Language. *Turkish Online Journal of Educational Technology-TOJET*, *15*(4), 88-106.

Diacopoulos, M., & Crompton, H. (2020). A systematic review of mobile learning in social studies. *Computers and Education*, *154*(103911). https://doi.org/https://doi.org/10.1016/j.compedu.2020.103911

Ducar, C., & Schocket, D. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*(4), 779-795. https://doi.org/https://doi.org/10.1111/flan.12366

Dülger, O. (2021). The effect of out-of-class reading tasks on developing EFL writing skills. *European Journal of Foreign Language Teaching*, *5*(3). 87-106. https://doi.org/10.46827/ejfl.v5i3.3591

Egamberdievna, T., & Daminovna, D. (2020). Statistical Analysis in Performance of GT (Google Translate) Among Learners and Professor-Teachers & Comparative Analysis in Translation of GT and Human Mind. *Journal of Critical Reviews*, *7*(8), 2607-2619.

Fathali, S., & Emadi, A. (2021). CALL Research in Iran: An Integrative Review of the Studies between 2007 and 2019. *Computer Assisted Language Learning Electronic Journal*, *22*(3), 33-51.

Fountain, A., & Fountain, C. (2009). A New Look at Translation: Teaching tools for language and literature. *Empowerment through Collaboration: Dimension*, 1-15.

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*. London: Sage.

Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL, 4(1),* 9-24. https://doi.org/10.1002/tesj.65

Hong, Q., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., & Pluye, P. (2018). The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285-291. https://doi.org/10.3233/EFI-180221

Hou, Z., & Aryadoust, V. (2021). A review of the methodological quality of quantitative mobile-assisted language learning research. *System*, *100*, 102568. https://doi.org/10.1016/j.system.2021.102568

Jin, L. & Deifell, E. (2013). Foreign language learners' use and perception of online dictionaries: A survey study. *MERLOT Journal of Online Learning and Teaching*, *9*(4), 513-533.

Jolley, J. R., & Maimone, L. (2022). Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. *L2 Journal, 14(1)*. 26-44.

Klimanova, L. (2021). The evolution of identity research in CALL: From scripted chatrooms to engaged construction of the digital self. *Language Learning & Technology*, *25*(3), 186-204.

Kalyaniwala, C., & Ciekanski, M. (2021). Autonomy CALLing: A systematic review of 22 years of publications in learner autonomy and CALL. *Language Learning & Technology*, 25(3), 106–131. http://hdl.handle.net/10125/73452

Khany, R., & Kamalvand, A. (2022). 100 Years of Research on English Language Learning/Teaching Materials: A Systematic Literature Review. *The Teaching English as a Second Language Electronic Journal, 25(4),* 1-27.

Klekovkina, V., & Denié-Higney, L. (2022). Machine Translation: Friend or Foe in the Language Classroom?. *L2 Journal*, *14(1),* 105-135.

Klimanova, L. (2021). The evolution of identity research in CALL: From scripted chat rooms to engaged construction of the digital self. *Language Learning & Technology, 25(3),* 186–204. http://hdl.handle.net/10125/73455

Knowles, C. L. (2022). Using an ADAPT Approach to Integrate Google Translate into the Second Language Classroom. *L2 Journal*, *14*(1), 195-236.

Le, Q., & Schuster, M. (2016). *A neural network for machine translation, at production scale*. https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Lee, S. (2021). The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis. *Computer Assisted Language Learning*, 1-23. https://doi.org/10.1080/09588221.2021.1901745

Levy, M. (2009). Technologies in use for second language learning. *The modern language journal*, *93*, 769-782. https://doi.org/10.1111/j.1540-4781.2009.00972.x

Lin, P. (2021). Developing an intelligent tool for computer-assisted formulaic language learning from YouTube videos. *ReCALL First View*, 1–16. https://doi.org/10.1017/S0958344021000252

Littell, J., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.

Luef, E. M., Ghebru, B., & Ilon, L. (2019). Language Proficiency and Smartphone-aided Second Language Learning: A look at English, German, Swahili, Hausa and Zulu. *Electronic Journal of e-Learning*, *17*(1), pp25-37.

Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, *45*(1), 1-43. https://doi.org/10.1017/S0261444811000395

McKinley, J., & Rose, H. (2019). *The Routledge handbook of research methods in applied linguistics: Realities, dilemmas, and solutions*. New York: Routledge.

Mundt, K., & Groves, M. (2016). A double-edged sword: the merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education*, *6(4),* 387-401. https://doi.org/10.1080/21568235.2016.1172248

Musk, N. (2014). Avoiding the Target Language with the Help of Google: Managing Language Choices in Gathering Information for EFL Project Work. *TESOL Quarterly*, *48*(1), 110-135. https://doi.org/10.1002/tesq.102

Nation, I. (2001). *Learning vocabulary in another language*. Cambridge, UK: University Press.

Organ, A. (2022). Attitudes to the use of Google Translate for L2 production: analysis of chatroom discussions among UK secondary school students. *The Language Learning Journal*, 1-16. https://doi.org/10.1080/09571736.2021.2023896

Page, M., McKenzie, J., Bossuyt, P.M., Boutron, I, Hoffmann, T.C.... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*. http://dx.doi.org/10.1136/bmj.n71

Payant, C., & Zuniga, M. (2022). Learners' flow experience during peer revision in a virtual writing course during the global pandemic. *System*, 102715. https://doi.org/10.1016/j.system.2021.102715

Ryu, J., Kim, Y., Park, S., Eum, S., Chun, S., & Yang, S. (2022). Exploring Foreign Language Students' Perceptions of the Guided Use of Machine Translation (GUMT) Model for Korean Writing. *L2 Journal*, *14(1)*, 136-165.

Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, 102736. https://doi.org/10.1016/j.system.2022.102736

Tourmen, C., & Hoffmann, D. (2022). A "Hands-On" Approach to Raise Awareness of Technologies: A Pilot Class and its Lessons. *L2 Journal*, *14*(1), 237-257.

Van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning & Technology, 26(1)*, 1–19. http://hdl.handle.net/10125/73460

Vinall, K., & Hellmich, E. (2022). Machine Translation & Language Education: Implications for Theory, Research, & Practice. *L2 journal, 14(1),* 4-25.

Xu, J. (2022). Proficiency and the Use of Machine Translation: A Case Study of Four Japanese Learners. *L2 Journal*, *14*(1), 77-104.

References to the reviewed articles

Alhaisoni, E., & Alhaysony, M. (2017). An Investigation of Saudi EFL University Students' Attitudes towards the Use of Google Translate. *International Journal of English Language Education*, *5*(1), 72-82. https://doi.org/10.5296/ijele.v5i1.10696

Bahri, H., & Mahadi, T. (2016). Google Translate as a Supplementary Tool for Learning Malay: A Case Study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, *7*(3), 161-167. https://doi.org/10.7575/aiac.alls.v.7n.3p.161

Bailey, D. R., & Lee, A. R. (2020). Learning from experience in the midst of covid-19: Benefits, challenges, and strategies in online teaching. *Computer-Assisted Language Learning Electronic Journal*, *21*(2), 178-198.

Bin Dahmash, N. (2020). I Can't Live Without Google Translate: A Close Look at the Use of Google Translate App by Second Language Learners in Saudi Arabia. *Arab World English Journal*, *11*(3), 226-240. https://doi.org/10.24093/awej/vol11no3.14

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *JALT CALL Journal*, *14*(1), 3-24.

Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, *98*(102464). https://doi.org/10.1016/j.system.2021.102464

Case, M. (2015). Machine Translation and the Disruption of Foreign Language Learning Activities. *eLearning Papers, 45*, 4-16.

Chandra, S. O., Yuyun, I. (2018). The use of Google Translate in EFL essay writing. *LLT Journal, 21(2)*, 228-238. https://doi.org/10.24071/llt.2018.210212

Chang, C. C., & Yamada, M. (2021). Translation Tasks for Learning Collocations: Effects of Machine Translation plus Post-editing and Sight Translation. *English Teaching and Learning*, *45*(1), 27-44. https://doi.org/10.1007/s42321-020-00059-x

Chen, W. (2020). Using Google Translate in an authentic translation task: The process, refinement efforts, and students' perceptions. *Current Trends in Translation Teaching and Learning E*, *7*, 213-238.

Chen, F., Tsai, S. C., & Tsou, W. (2019). The application of translanguaging in an English for specific purposes writing course. *English Teaching & Learning*, *43*(1), 65-83. https://doi.org/10.1007/s42321-018-0018-0

Chon, Y., & Shin, D. (2020). Direct writing, translated writing, and machine-translated writing: A text level analysis with Coh-Metrix. *English Teaching*, *75*(1), 25-48.

Chon, Y., Shin, D., & Kim, G. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, *96*(102408). https://doi.org/10.1016/j.system.2020.102408

Chung, S. (2020). The Effect of L2 Proficiency on Post-editing Machine Translated Texts. *The Journal of Asia TEFL*, *17*(1), 182-193. http://dx.doi.org/10.18823/asiatefl.2020.17.1.11.182

Chung, E., & Ahn, S. (2021). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, 1-26. https://doi.org/10.1080/09588221.2020.1871029

Clifford, J., Merschel, L., & Joan, M. (2013). Surveying the Landscape: What is the Role of Machine Translation in Language Learning? *Atic*, *10*, 108-121. https://doi.org/10.7203/attic.10.2228

Cohen, A. D., & Wang, I. K. H. (2019). Fine-tuning word meanings through mobile app and online resources: A case study of strategy use by a hyper polyglot. *System*, *85*, 102106. https://doi.org/10.1016/j.system.2019.102106

Cornell, R., Dean, J., & Tomaš, Z. (2016). Up Close and Personal: A Case Study of Three University-Level Second Language Learners' Vocabulary Learning Experiences. *Tesol Journal*, *7*(4), 823-846. doi: 10.1002/tesj.247

Fredholm, K. (2015). Online translation use in Spanish as a foreign language essay writing: Effects on fluency, complexity and accuracy. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas, 9*(18), 7-24.

Fredholm, K. (2019). Effects of Google translate on lexical diversity: vocabulary development among learners of Spanish as a foreign language. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, *13*(26), 98-117.

Garcia, I., & Pena, M. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471-487. https://doi.org/10.1080/09588221.2011.582687

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, *37*, 112-121. https://doi.org/10.1016/j.esp.2014.09.001

Groves, M., & Mundt, K. (2021). A ghostwriter in the machine? Attitudes of academic staff towards machine translation use in internationalized Higher Education. *Journal of English for Academic Purposes*, *50*(100957). https://doi.org/10.1016/j.jeap.2021.100957

Habeeb, L. S., & Muhammedb, M. R. (2020). Investigating the Effectiveness of Google Translate Among Iraqi Students. *International Journal of English Linguistics, 12(12)*, 1-13.

Karnal, A. R., & Pereira, V. W. (2015). Reading strategies in a L2: A study on machine translation. *The Reading Matrix*.

Kelly, R., & Hou, H. (2021). Empowering learners of English as an additional language: translanguaging with machine translation. *Language and Education, 1(16)*. https://doi.org/10.1080/09500782.2021.1958834

Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in Academic Writing Courses? *The EUROCALL Review*, *26*(2), 50-57.

Lai, C., & Zheng, D. (2018). Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL*, *30(3)*, 299-318. https://doi.org/10.1017/S0958344017000258

Lee, S. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning, 33(3),* 157-175. https://doi.org/10.1080/09588221.2018.1553186

Lee, C. (2020). A study of adolescent English learners' cognitive engagement in writing while using an automated content feedback system. *Computer Assisted Language Learning, 33(1-2)*, 26-57. https://doi.org/10.1080/09588221.2018.1544152

Lee, S. (2021). An investigation of machine translation output quality and the influencing factors of source texts. *34*(1), 81-94. https://doi.org/10.1017/S0958344021000124

Lee, S., & Briggs, N. (2021). Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL, 33(1)*, 18–33. https://doi.org/10.1017/S0958344020000191

Lin, L. H., & Morrison, B. (2021). Challenges in academic writing: Perspectives of Engineering faculty and L2 postgraduate research students. *English for Specific Purposes*, *63*, 59-70. https://doi.org/10.1016/j.esp.2021.03.004

Maghsoudi & Mirzaeian (2020). Machine versus human translation outputs: Which one results in better reading comprehension among EFL learners? [Article]. *JALT CALL Journal*, *16*(2), 69-84. https://doi.org/10.29140/jaltcall.v16n2.342

Mirzaeian (2021). The effect of editing techniques on machine translation-informed academic foreign language writing. *The Euro CALL Review*, *29*(2), 33-43. https://doi.org/10.4995/eurocall.2021.13120

Murphy, D. (2020). Supporting Pre-Service English Teachers' Academic Reading and Writing with Online Machine Translation. *STEM*, *21*(2).

https://doi.org/ 10.16875/stem.2020.21.2.123 123

Murtisari, E. T., Widiningrum, R., Branata, J., & Susanto, R. D. (2019). Google Translate in Language Learning: Indonesian EFL Students' Attitudes. *Journal of Asia TEFL*, *16*(3), 978-986. https://doi.org/10.18823/asiatefl.2019.16.3.14.978

Niño, A. (2015). Language Learners Perceptions and Experiences on the Use of Mobile Applications for Independent Language Learning in Higher Education. *IAFOR Journal of Education*.

Nino, A. (2020). Exploring the use of online machine translation for independent language learning. *Research in Learning Technology*, *28*. https://doi.org/10.25304/rlt.v28.2402

O'Neill, E. (2019a). Online translator, dictionary, and search engine use among L2 students. *CALL-EJ: Computer-Assisted Language Learning–Electronic Journal*, *20*(1), 154-177.

O'Neill, E. M. (2019b). Training students to use online translators and dictionaries: The impact on second language writing scores. *International Journal of Research Studies in Language Learning*, *8(2)*, 47-65. http://dx.doi.org/10.5861/ijrsll.2019.4002

Shadiev, R., Sun, A., & Huang, Y. M. (2018). A study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. *British Journal of Educational Technology*, *50*(3), 1415-1433. http://dx.doi.org/10.1111/bjet.12648

Stapleton, P. (2021). Using Google Translate as a Tool to Improve L2 Writing: A Case Study of Primary-Level Writing in Hong Kong. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *11*(3), 92-98.

Stapleton, P., & Kin, B. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, *56*, 18-34. https://doi.org/10.1016/j.esp.2019.07.001

Tarsoly, E., & Valijärvi, R.-L. (2019). 'Language students as critical users of Google Translate': Pitfalls and Possibilities. *Practitioner Research in Higher Education*, *12*(1), 61-74.

Ting, F. K., & Tan, K. H. (2021). Enhancing English Language Vocabulary Learning among Indigenous Learners through Google Translate. *Journal of Education and e-Learning Research*, *8(2)*, 143-148. https://doi.org/10.20448/journal.509.2021.82.143.148

Tsai, S.-C. (2019). Using google translate in EFL drafts: a preliminary investigation. *Computer Assisted Language Learning*, *32(5-6),* 510-526. https://doi.org/10.1080/09588221.2018.1527361

Tsai, S.-C. (2020). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, 1-23. https://doi.org/10.1080/09588221.2020.1799412

White, K. D., & Heidrich, E. (2013). Our policies, their text: German language students' strategies with and beliefs about web-based machine translation. *Die Unterrichtspraxis/Teaching German*, *46(2)*, 230-250.

Wuttikrikunlaya, P., Singhasiri, W., & Keyuravong, S. (2018). The use of online tools in L2 writing: A study of Thai university students. *Journal of English Language Teaching and English Linguistics*, *30(1)*, 107-148.

Appendix A

Journal Indices and Number of Studies

Journal	WOS	IF (2020)	Scopus	Number of studies
Computer Assisted Language Learning	*	3.51	*	5
The Journal of Asia TEFL	*	0.4	*	2
The JALT CALL Journal	-	-	*	2
English for Specific Purposes	*	2	*	2
System	*	2	*	2
Journal of English for Academic Purposes	*	2.24	*	1
Language and Education	*	1.97	*	1
Journal of Education and e-Learning Research	-	-	*	1
Arab World English Journal (AWEJ)	*	0.51	-	1
English Teaching & Learning	*	0.61	*	1
English Teaching	-	-	-	1
Research in Learning Technology	*	0.97	*	1
STEM Journal	-	-	-	1
Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas	-	-	-	1
International Journal of Innovation, Creativity and Change	-	-	*	1

CALL-EJ	*	-	*	1
Practitioner Research In Higher Education	-	-	-	1
The EUROCALL Review	-	-	-	1
LLT Journal: A Journal on Language and Language Teaching	-	-	-	1
International Journal of English Language Education	-	-	-	1
Advances in Language and Literary Studies	-	-	-	1
The Reading Matrix: An International Online Journal	-	-	-	1
eLearning Papers	-	-	-	1
@ tic. revista d'innovació educativa	-	-	-	1
Die Unterrichtspraxis/ Teaching German	*	-	-	1

Note: InCites Journal Citation Reports (JCR) via Web of Science (https://mjl.clarivate.com/search-results) and Scopus (https://www.scopus.com).