Machine translation: Turkish–English bilingual speakers' accuracy detection of evidentiality and preference of MT

Sümeyra Tosun[1*]

Abstract

Machine translation (MT) is the automated process of translating text between different languages, encompassing a wide range of language pairs. This study focuses on non-professional bilingual speakers of Turkish and English, aiming to assess their ability to discern accuracy in machine translations and their preferences regarding MT. A particular emphasis is placed on the linguistically subtle yet semantically meaningful concept of evidentiality. In this experimental investigation, 36 Turkish–English bilinguals, comprising both early and late bilinguals, were presented with simple declarative sentences. These sentences varied in their evidential meaning, distinguishing between firsthand and non-firsthand evidence. The participants were then provided with MT of these sentences in both translation directions (Turkish to English and English to Turkish) and asked to identify the accuracy of these translations. Additionally, participants were queried about their preference for MT in four crucial domains: medical, legal, academic, and daily contexts. The findings of this study indicated that late bilinguals exhibited a superior ability to detect translation accuracy, particularly in the case of firsthand evidence translations, compared to their early bilingual counterparts. Concerning the preference for MT, age of acquisition and the accuracy detection of non-firsthand sentence translations emerged as significant predictors.

Introduction

Machine translation (MT) is the process of using automated software to translate text from one language to another, catering to a wide array of language pairs. While MT has been in existence since the 1950s, it experienced a significant transformation in the past two decades, shifting from rule-based MT to statistical MT and more recently to neural machine translation (NMT). This evolution, especially the advent of deep learning, has dramatically enhanced the quality of MT (Burchardt et al, 2017; Melby, 2020; Popović, 2017; Turovsky et al., 2022). Some scholars argue that recent improvements have elevated

MT quality to a level comparable to human translation (HT) for specific texts and language pairs (Hassan et al., 2018; Perrault et al., 2019). Nevertheless, there remains a lack of consensus regarding the validity of this assertion and the criteria used to evaluate MT quality (Pym, 2020; Toral et al., 2019). Even so, it is widely acknowledged that MT technology has made remarkable progress recently although these advancements vary across language pairs, with popularity influencing the level of investment in data acquisition (Perrault et al., 2019).

As neural MT continues to advance, it is likely that MT will become an even more prevalent technology, necessitating engagement from professional translators. This progress places MT at the forefront of Artificial Intelligence-Mediated Communication (AI-MC), characterized as "interpersonal communication that is not merely transmitted by technology but is modified, augmented, or generated by a computational agent to

*Correspondence:
Sümeyra Tosun
stosun@mec.cuny.edu
[1] Department of Psychology, Medgar Evers College, CUNY, New York, NY, USA

achieve communication goals" (Hancock et al., 2020, p. 90). Recent research on interactions with MT has shown that its usage does not hinder the translation process or necessarily impact the final quality of the translated text (e.g., Gaspari et al., 2014; Koponen, 2012; Moorkens, et al., 2015; Teixeira, 2014). However, professional translators still exhibit substantial resistance to adopting MT as an aid and notice more errors in MT. This study aims to explore whether the preferences of non-professional Turkish–English bilingual speakers for MT and their ability to detect translation errors are associated.

Modern technologies are increasingly deployed to surmount language barriers, extending their reach beyond personal use into critical domains like healthcare, legal proceedings, and law enforcement. The accessibility and societal influence of machine translation (MT) are currently under scrutiny, with a focus on ensuring the inclusive participation of diverse social groups in communication processes (Vieira et al., 2021). The proliferation of freely accessible MT systems online has democratized language translation, bridging gaps in both developed and developing nations. Google Translate, for instance, processes an astounding 150 billion words daily, highlighting the far-reaching societal implications of MT beyond the translation community (Davenport, 2018). However, there is a recognized gap in public awareness regarding the capabilities and quality of machine translation (Kasperé et al., 2021).

While MT is rapidly dismantling language barriers and closing in on human-level translation accuracy and efficiency, human intervention remains crucial to mitigate the potential negative consequences of its use in society (Hoi, 2020). Effective communication processes supported by MT can attain high quality, provided that participants are cognizant of the technology's limitations (Yasuoka & Bjorn, 2011). Furthermore, research has shown that MT can play a pivotal role in reducing the exclusion of ethnic minorities across various fields (Taylor et al., 2015).

Recent research on MT

Extensive research has investigated the quality of machine translation and the utilization of post-editing techniques (Nurminen & Koponen, 2020; Ortega et al., 2019; Rossi & Carré, 2022; Ueffing, 2018; Vardaro et al., 2019, among others). Various studies employing diverse research methodologies have recognized the advantages of post-editing machine translation across different language pairs (Carl et al., 2011, 2015; Moorkens, 2018; Stasimioti & Sosoni, 2021). Additionally, investigations have addressed the acceptability of machine-translated content (Castilho, 2016; Castilho & O'Brien, 2017; Rivera-Trigueros, 2021; Taivalkoski-Shilov et al., 2022). However,

these studies have primarily focused on the attitudes and perceptions of translation students, novice translators, professional translators, and post-editors, likely due to the convenience of access to respondents and research design (Moorkens et al., 2015; Rossi & Chevrot, 2019; Ferreiraa et al., 2021). Surprisingly, the acceptability of machine-translated content among non-professional users remains relatively understudied.

Prior research in machine translation has predominantly focused on professional translators' viewpoints, including their perception of MT as a potential industry threat (Vieira, 2020), the factors influencing their adoption or rejection of MT (Cadwell et al., 2018), and the impact of MT advancements on their practices, self-perception, and professional standing (Läubli & Orrego-Carmona, 2017; Sakamoto, 2019). These studies have consistently highlighted translators' concerns and their generally lower regard for MT regarding quality and utility. However, it is worth noting that some of this resistance diminishes when translators are involved in the development and implementation of MT software (Rossi & Chevrot, 2019). As MT increasingly permeates non-translator usage in various social communication contexts, focusing solely on professional translators' attitudes limits our understanding of the broader societal impacts of this technology. Therefore, it is essential to consider both translator and non-translator populations when exploring public perceptions of MT, particularly from ordinary users' perspectives, given the various purposes for which they casually employ machine translation in their daily lives (Kasperé et al., 2021).

The studies, particularly on professional translators, demonstrate that machine-translated text typically demands a higher cognitive load, to varying degrees, compared to human-translated or post-edited content. Earlier user-centered research analyzed raw machine translation, and identified reduced usability of machine-translated instructions compared to post-edited output (Carl et al., 2011; Castilho, 2016; Castilho et al., 2014; Doherty & O'Brien, 2014; Doherty, *2016*; Daems et al., 2017; Ferreiraa et al., 2021; Guerberof Arenas et al., 2021; Hu et al., 2020; Jakobsen & Jensen, 2008; Kasperé et al., 2023; Moorkens, 2018; Stasimioti & Sosoni, 2021; Vardaro et al., 2019).

Similarly, studies on non-professional users, that have examined the acceptability of machine-translated text, revealed that readers spend more time and cognitive effort on machine-translation errors compared to correct segments of text (Colman et al., 2021; Kasperavičienė et al., 2020). The presence of errors in machine-translated segments increased the cognitive processing demands. Hence, post-editing machine-generated translations remains crucial to ensure precise language translations

(Gaspari et al., 2014; Macías et al., 2020; Taylor et al., 2015).

Other studies have investigated how non-professionals or individuals with low proficiency perceive the acceptability of machine-translated texts in various languages. The individuals utilize machine translation for various purposes, often without a full understanding of how it works or the quality it provides. For instance, in a survey of 400 participants, the acceptability of machine-translated text from English to Lithuanian was influenced by factors like age and education. Less educated and older participants were more inclined to view machine translation as reliable and satisfactory (Kasperé et al., 2021). However, when the situation is ethically charged (e.g., in legal settings) both translators and non-translators exhibit a negative bias toward MT (Asscher & Glikson, 2023).

In a study by Rossetti et al. (2020), 61 participants assessed the "impact of machine translation and post-editing awareness" on comprehension and trust when reading crisis messages in English and Italian. The results showed no significant differences in comprehension and trust between raw machine-translated and post-edited text. However, participants with limited English proficiency had a more favorable view of raw machine-translated text in terms of comprehension and trust. Another study involving translation agencies, professional translators, and clients/users of professional translation investigated user awareness of machine translation through surveys (García, 2010). The study focused on the acceptability and evaluation of machine translation from Chinese into English. The findings indicated that less than 5% of professional translators considered the quality of machine translation to be very high. Translation agencies shared a similar view with the translators. Among clients/users of professional translation (about 30%) who were aware of and requested machine translation, there was an intermediate or positive assessment of its quality.

For closely related languages in MT, a lexical analysis aided by translation rules often suffices, omitting the need for extensive semantic analysis. Rule sets for translation are more manageable in close language pairs compared to unrelated ones, simplifying the rule coding process (Altintas & Cicekli, 2022). NMT systems utilize typological similarities among languages to create clusters (Tan et al., 2019). Typologically distinct languages require more effort in NMT and might reveal fewer satisfying results. Thus, the current study, following an offline methodology, focused on a typologically distinct language pair, namely Turkish and English, and their non-professional bilingual speakers. It aimed to determine their ability to detect accuracy in machine translation and their preference for MT in need of a translator, particularly concerning the linguistically subtle yet semantically significant concept of evidentiality.

Evidentiality

Evidentiality, a linguistic concept, relates to how a language conveys the source of knowledge regarding an event described (Aikhenvald, 2004). It allows speakers to indicate whether they personally witnessed an event or acquired information about it from others. In certain languages like Turkish, evidentiality is grammatically marked, requiring speakers to specify their information source when describing past events. In Turkish, firsthand experiences are marked with "-di," while non-firsthand information, such as hearsay or inference, is marked with "-miş." In contrast, languages like English offer optional means to express evidential-type meanings, relying on lexical or constructional choices rather than grammatical markers. This flexibility allows English speakers to decide whether to incorporate evidentiality in their descriptions, using words like "*apparently*" or "*seemed like*" to indicate their source of information. While evidentiality is not grammatically obligatory in English, it can still be expressed. However, due to its optional nature, it is less likely to be consistently and frequently expressed compared to Turkish, where it is grammatically mandatory. Translating from English to Turkish is relatively straightforward because there is a single marker for different sources of non-firsthand knowledge. However, challenges arise when translating from Turkish, which has one marker for non-firsthand information, to English, which offers multiple options (Filipović, 2017).

Linguistic and psycholinguistic research highlights that apart from indicating the source of knowledge, evidentiality communicates the epistemic value of information (Aikhenvald, 2004; Aksu-Koç, 2016; Arslan, 2020; Plungian, 2001; Tosun & Vaid, 2018; Willett, 1988). A proposition expressed with a firsthand source signifies higher confidence in its occurrence (higher epistemic value), while non-firsthand statements suggest less certainty, casting doubt on the proposition's actual occurrence (lower epistemic value).

Monolingual Turkish speakers typically acquire evidentiality in early childhood (Aksu-Koç et al., 2009). Research by Aksu-Koç and colleagues demonstrates that Turkish-speaking children start using evidential markers between 18 months and 3 years of age. However, some studies have indicated that comprehending sentences with evidential marking may be challenging until around the age of 3 (Öztürk & Papafragou, 2016). Bilingual speakers' acquisition of evidentiality demonstrated differences from monolingual speakers. It is noteworthy that some bilingual speakers may not fully acquire evidentiality. Heritage speakers of Turkish, who

are fluent in Turkish but learned it as a heritage language, have been found to have lower sensitivity to evidentiality markers compared to native Turkish speakers (Arslan, 2020; Arslan & Bastiaanse, 2020; Arslan et al., 2017; Schmid & Karayayla, 2020). Similarly, second-language learners of Turkish struggle with the various uses of non-firsthand evidential markers (Kaya-Soykan et al., 2023). Studies involving English speakers learning Japanese, a language that grammatically encodes evidentiality, have also shown that learners use evidential markers less frequently and correctly compared to native speakers (Ishida, 2006; Narita, 2011).

A recent study investigating the typological differences in evidentiality examined Turkish–English bilinguals in a bilingual setting (Tosun & Filipović, 2022). This study, similar to previous research, found that the AoA influenced bilingual speakers' performance in a bilingual context. Early bilinguals were typically heritage speakers of Turkish and their parents were born in Turkey and residing in the USA or the UK. Late bilinguals were immigrants from Turkey currently living in the USA or the UK. Participants were presented with sentences in both English and Turkish containing information marked as firsthand or non-firsthand. They were tasked with translating these sentences between the two languages and making judgments about the likelihood of the described events actually occurring. The results revealed that both groups performed similarly when translating firsthand-marked sentences. However, late bilingual speakers noticed non-firsthand sources when translating from English to Turkish, while early bilingual speakers tended to ignore information about non-firsthand sources when translating from Turkish to English. Additionally, late bilinguals were more likely to consider events described in firsthand-marked sentences as having actually occurred, compared to those described in non-firsthand sentences. These findings highlight the typological differences between English and Turkish and the challenges faced by early bilinguals in accurately discerning information sources. The study prompts questions about how accurately they can detect machine translation errors and how likely they are to prefer machine translation.

Present study
This study delved into the accuracy of Turkish–English bilingual speakers in detecting machine translation (MT) errors in translations of past events. The central focus was on evidentiality, a key linguistic property, given its distinctive structures in these two languages. The primary aim of the study was to explore how proficient MT users were in recognizing errors within MT outputs (rather than investigating the quality of MT outputs) and how this proficiency was associated with their use and

preference for MT tools. The research aimed to address several key questions:

1. How proficiently can bilingual speakers identify inaccuracies of evidentiality in MT of past events?
2. Does their ability to discern differences in evidentiality in translations predict their preference for MT tools in need of a translator?
3. Do specific contexts, where MT is applied, such as medical, legal, academic, or daily communication settings, influence their MT preference judgments?
4. Does the AoA of bilingual individuals impact their preference for MT and their accuracy in detecting translation errors?

Given the escalating use of MT in our interconnected world, gaining insights into bilingual speakers' attitudes toward MT holds substantial significance, particularly in contexts where discourse plays a pivotal role in decision-making processes.

Method
Participants
A total of 36[1] participants (all female[2]) were recruited in the USA and South Africa. AoA threshold was established at 12 years old. Participants who acquired their second language (English) after the age of 12 were categorized as late bilinguals. The late L2 English bilingual group consisted of 15 participants, with a mean age of 33 ($SD = 10.38$). Their mother tongue was Turkish, and it was acquired in Turkey. They immigrated to the US or South Africa for educational or economic purposes, where they acquired their L2 English. The L2 English was rated as worse than their Turkish by 34% and the remaining stated that their English was as good as their Turkish in this group. They graded their Turkish proficiency a 7 and English proficiency a 6 out of a 7-point scale. They indicated their use of Turkish as 60% and English as 40% daily ($SD = 6.12$). The early bilingual group consisted of 21 participants with a mean age of 17.57 ($SD = 3.63$). They were generally heritage learners who learned and spoke Turkish at home and learned and spoke English at school

---

or outside of home. All early bilingual participants stated that their English was better than their Turkish. Out of a 7-point scale, they judged their Turkish proficiency as 5 (including reading and writing) and English proficiency as 7. Early bilinguals also indicated that approximately 78% of the time, they used English daily (*SD* = 8.56).

Materials and measures

For the translation accuracy detection phase, three variables were manipulated within subjects: Translation direction (English to Turkish vs Turkish to English), Source of information (Firsthand vs Non-firsthand), and Translation accuracy (Correct vs Incorrect). Thus, there were 8 total conditions. A total of 80 sentences were divided evenly between these 8 conditions, which made 10 sentences per condition.

Half of the sentences were presented as the source language was Turkish and the target language was English, while the other half were given the other way around. For half of the Turkish sentences (20 sentences), the past tense suffix used was the firsthand form (*-di: Suna eski kocasını affetti*) and for the remainder (20 sentences) the non-firsthand past tense suffix was used (*-miş: Suna eski kocasını affetmiş*). For half of each form, the English translations (10 sentences) were presented correctly (with a phrase added to indicate the source of information for non-firsthand sentences such as *apparently*, *it seemed like*) and the other half translations were presented incorrectly (10 sentences). Like Turkish to English stimuli, half of the sentences (20 sentences) of English to Turkish translations were introduced in firsthand form (e.g., *Sue forgave her ex-husband*). The other half (20 sentences) was in non-firsthand form and contained one of the following expressions that signal evidential meanings (and that are most often used as translation equivalents for the Turkish non-firsthand evidential): it appeared, it seemed, must have, it looked like, and apparently (e.g., *It seemed Sue forgave her ex-husband)*. Half of the English-to-Turkish translation sentences were presented in the correct translation format (where the source of information was translated accurately) and the other half were presented in an incorrect translation format. Stimuli were presented in the blocks of the direction of translation and the direction of translation blocks were counterbalanced. Per language, firsthand and non-firsthand sentences were presented in a fixed random order. Also, for each language, the particular sentences chosen to be in firsthand versus non-firsthand form and correct versus incorrect translation form were counterbalanced across participants.

The sentences were simple declarative, transitive sentences, each containing a verb in the past tense. All sentences were roughly of similar length and were adapted from stimuli created by Tosun and Filipović ([2022](#)). While the original sentences were translated using Google Translate, the author later manipulated them to create the above-mentioned conditions. This involved adding or removing evidential expressions in English and modifying the Turkish suffixes. Expressions such as "it seemed like," "apparently," "must have," and "it looked like" were used for non-firsthand sentences. These phrases were positioned either at the start of the sentence or just before the verb, depending on their natural usage in English.

For this section, participants' *hit rates* (total number of detections of translation accuracy correctly) and *false alarms* (judging the incorrect translations as correct) for each source were computed. The false alarm term was entered into the accuracy detection computation to achieve a more balanced and fine-tuned evaluation of detection performance. Finally, translation accuracy detection was calculated as the difference between hits and false alarms. Thus, the dependent variable presented the pure ability to accurately discriminate between correct detection and noise while minimizing false identifications. This increased the sensitivity and discriminability of the correct detection measure.

A language background questionnaire was utilized to detect the bilingual participants' language history, their AoA, and other relevant metadata such as frequency of daily use of each language and proficiency self-assessment. MT use, trust, and preference were measured by employing a questionnaire. In this questionnaire, participants were asked about their usage patterns of MT tools (How frequently do you use the machine translation tools on a daily basis? Please use the slider to adjust your frequency use from 0 to 100), and the level of trust they place in these tools (How much do you trust the accuracy of the translation that is produced by the machine translation tools? Please use the slider to adjust your trust rate from 0 to 100). Subsequently, they were asked about their likelihood of preferring MT in various settings (How much would you prefer to be translated by MT in need of a translator in the following settings? Please use the slider to adjust your preference level from 0 to 100), including medical, legal, academic, and daily contexts.

Procedure, design, and data analysis

The experiment was conducted through Qualtrics. Participants received the language background questionnaire first followed by the MT use and preference questionnaire. Then they were instructed that they would see some sentences followed by their equivalent translation completed by a machine translation tool. The sentences were presented side by side and without time limitation. They were asked to read each sentence and

Table 1 Summary of accuracy detection of MT across AoA, source of information and translation direction

	Firsthand				Non-firsthand			
	Late		Early		Late		Early	
	T to E	E to T	T to E	E to T	T to E	E to T	T to E	E to T
Mean	0.95	0.55	0.46	0.18	−0.22	0.40	0.00	0.24
SE	0.08	0.10	0.07	0.08	0.15	0.09	0.13	0.07

its translation carefully, and then make a judgment about whether it was translated accurately. The sentences (both the original sentence and the translation) were available on the screen while participants were making their judgments and indicating their responses. They were additionally asked to indicate the kind of inaccuracy if they thought it was not accurately translated. The order of the translation direction (Turkish to English and English to Turkish) was counterbalanced. The experiment took approximately forty-five minutes to complete, and participants received compensation when they submitted their completed forms.

Translation accuracy detection
This portion of the study had a 2 (AoA: Early vs. Late L2 English) × 2 (Direction: Turkish to English (T to E) vs. English to Turkish (E to T)) × 2 (Source: Firsthand vs. Non-firsthand) mixed design. Group was the between-subjects variable, and other variables were manipulated as within subjects. The dependent variable was participants' translation accuracy detection, which was calculated as the difference between hits and false alarms. A 2 × 2 × 2 repeated measures ANOVA was conducted and the Bonferroni method was used for post-hoc comparisons.

MT preference
In this analysis, the possible predictors of MT preference in need of a translator were questioned. A series of multiple regression analyses were employed in which the MT preference was the dependent variable and MT use frequency, trust in MT, translation accuracy detection scores of non-firsthand sources[3] (both T to E and E to T), and AoA were entered as the predictors. No interaction terms were entered into the model. The analysis was conducted separately for the preference score of each setting (Medical, Legal, Academic, and Daily).

---

[3] The firsthand sources were not entered to the model as predictors due to collinearity issues. Because non-firsthand source is the critical distinction between Turkish and English typology, only these scores were entered to the model.

Results
Before presenting further analysis the early and late bilingual participants' frequency of MT use and trust in MT were analyzed. An independent sample t-test was utilized to examine the difference between the two bilingual groups. The results revealed that late bilinguals ($M = 26.8$, $SD = 15.9$) used MT more frequently than early bilinguals ($M = 16.4$, $SD = 10.9$, t (45) = 2.64, $p = 0.01$, Cohen's d = 0.78). However, their trust in MT did not reveal a significant difference ($M_{Late} = 66.3$, $SD_{Late} = 28.6$; $M_{Early} = 63.2$, $SD_{Early} = 20.3$, t (45) = 0.43, $p = 0.67$, Cohen's d = 0.13).

Translation accuracy detection
The summary of descriptive statistics is presented in Table 1. The results revealed a significant main effect of Source of information [F (1, 34) = 61.64, $p < 0.001$, $\eta_p^2 = 0.65$]. Firsthand source's translation accuracy was detected better than non-firsthand source (Mean Difference (MD) = 0.43). Source by AoA interaction was significant [F (1, 34) = 18.16, $p < 0.001$, $\eta_p^2 = 0.35$]. The late bilingual group was significantly better in the detection of firsthand translation accuracy than early bilinguals ($MD = 0.43$, $p < 0.001$) although, the difference disappeared in non-firsthand translations ($MD = -0.03$, $p = 1$). Within late bilinguals, firsthand translation accuracy was better detected than non-firsthand translation ($MD = 0.66$, $p < 0.001$). The same effect appeared in the early bilingual group, although the difference was relatively smaller ($MD = 0.20$, $p = 0.05$).

Direction by Source interaction was also significant [F (1, 34) = 47.6, $p < 0.001$, $\eta_p^2 = 0.58$]. T to E firsthand translations were detected significantly better than non-firsthand translations in the same direction ($MD = 0.81$, $p < 0.001$) although, the difference disappeared in E to T translations ($MD = 0.05$, $p = 1$). Within firsthand translations, T to E translation accuracy was better detected than E to T translation ($MD = 0.33$, $p < 0.001$). The reverse effect was found in non-firsthand sentences, where T to E translations were less accurately detected than E to T translations ($MD = -0.43$, $p < 0.001$).

Finally, the three-way interaction (see Fig. 1), Source by AoA by Direction was significant [F (1, 34) = 5.19, $p = 0.03$, $\eta_p^2 = 0.13$]. When translations were from T
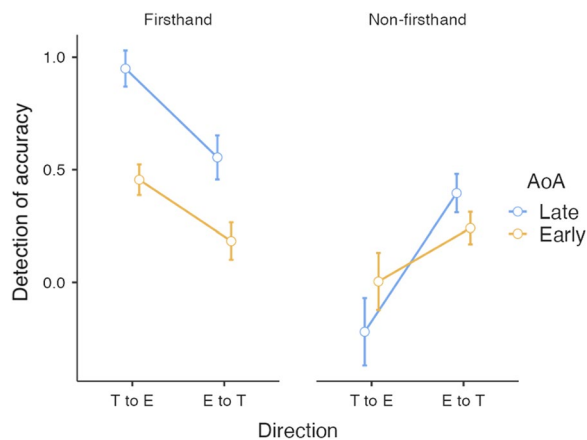
Fig. 1 This figure depicts the three-way interaction of Direction by Source by AoA

to E, late bilinguals detected firsthand sentences more accurately than early bilinguals ($MD = 0.49$, $p = 0.001$), although, the difference disappeared when translations were from E to T ($MD = 0.37$, $p = 0.18$). The direction effect also disappeared for early and late bilinguals in non-firsthand translations (T to E: $MD = -0.22$, $p = 1$; E to T: $MD = 0.15$, $p = 1$). Within late bilinguals, T to E firsthand translations were more accurately detected than T to E non-firsthand translations ($MD = 1.17$, $p < 0.001$) and E to T firsthand translations ($MD = 0.39$, $p = 0.012$). Non-firsthand translation accuracy demonstrated the otherwise, T to E translations were less accurately detected by late bilinguals than E to T translations ($MD = 0.61$, $p < 0.001$).

Within early bilinguals, T to E firsthand translations were detected more accurately than non-firsthand translations ($MD = 0.45$, $p = 0.05$), although they were equally better detected as E to T firsthand translations ($MD = 0.27$, $p = 0.09$). Further there was no difference between non-firsthand sentences of both directions ($MD = -0.23$, $p = 0.1$) for early bilinguals.

MT preference

The effect of settings

MT preference of four settings was compared by utilizing a repeated measure ANOVA as entering AoA as a between-subject factor. The results revealed a significant Settings main effect [F (3, 126) = 4.09, $p = 0.008$, $\eta_p^2 = 0.09$] and a significant Settings by AoA interaction [F (3, 126) = 6.03, $p < 0.001$, $\eta_p^2 = 0.13$]. Late bilingual speakers preferred to be translated by MT tools in all four settings roughly the same. Early bilingual speakers, on the other hand, demonstrated differences in some settings, in which, they preferred to be translated by MT tools in academic settings more likely than medical ($MD = 21.5$, $p = 0.06$) legal ($MD = 22.54$, $p = 0.03$) and daily ($MD = 36.92$, $p < 0.001$) settings (see Table 2).

Predictors of MT preference

The summary of the results is presented in Table 3. Four multiple regression analyses were conducted to examine the predictors of MT preferences in Medical, Legal, Academic, and Daily settings. The two non-firsthand source accuracy detection scores (T to E and E to T), MT use frequency, trust in MT, and AoA were entered as predictors.

Medical settings The results indicated that there was a collective significant effect among all five predictors (F (5, 30) = 11.4, $p < 0.001$, $R^2 = 0.65$). The individual predictors were examined further and indicated that T to E non-firsthand ($B = 1.62$, t (30) = 7.06, $p < 0.001$), E to T non-firsthand ($B = -1.05$, t (30) = 5.39, $p < 0.001$), and AoA ($B = -1.39$, t (30) = 5.15, $p < 0.001$) were significant predictors. Participants' accuracy detections of T to E non-firsthand translations had positively predicted their preference for MT in medical settings. The more accurately they could detect the translations from T to E the more likely they preferred MT. However, their accuracy detections of E to T non-firsthand translations were negatively related to their preference. The more accurately they could detect the translations of non-firsthand sentences the less likely they preferred MT in medical settings. AoA was another significant predictor revealing that late bilinguals were more likely to prefer MT than early bilinguals. Finally, the MT use frequency and trust in MT were not significant predictors in the model.

Legal settings The results demonstrated a collective significant effect among all five predictors (F (5, 30) = 13, $p < 0.001$, $R^2 = 0.68$). The individual predictors were examined further and indicated that T to E

Table 2 Summary of MT preference of bilingual speakers by AoA and settings

	Medical		Legal		Academic		Daily	
	Late	**Early**	**Late**	**Early**	**Late**	**Early**	**Late**	**Early**
Mean	52.1	42.8	60.3	41.8	49	64.3	51.5	27.4
SE	7.21	6.58	6.43	5.87	4.91	4.48	6.58	6.01

Table 3 Summary of the regression analysis: possible predictors of MT preference in various settings

Model summary	R^2	F	p	ß	p
Medical	0.65	11.40	< 0.001		
Frequency of use				− 0.19	0.12
Trust to AI				0.26	0.09
T to E Non-Firsthand				1.62	< 0.001
E to T Non-Firsthand				− 1.05	< 0.001
AoA				− 1.39	< 0.001
Legal	0.68	13.00	< 0.001		
Frequency of use				− 0.23	0.05
Trust to AI				0.29	0.05
T to E Non-Firsthand				1.53	< 0.001
E to T Non-Firsthand				− 1.07	< 0.001
AoA				− 1.80	< 0.001
Academic	0.50	5.94	< 0.001		
Frequency of use				0.19	0.19
Trust to AI				− 0.37	0.05
T to E Non-Firsthand				0.59	0.04
E to T Non-Firsthand				− 1.07	< 0.001
AoA				− 0.45	0.18
Daily	0.75	18.00	< 0.001		
Frequency of use				0.30	0.01
Trust to AI				− 0.71	< 0.001
T to E Non-Firsthand				− 0.35	0.08
E to T Non-Firsthand				0.16	0.34
AoA				− 0.99	< 0.001

non-firsthand ($B = 1.53$, t (30) = 6.99, $p < 0.001$), E to T non-firsthand ($B = − 1.07$, t (30) = 5.7, $p < 0.001$), MT use frequency ($B = − 0.23$, t (30) = 2.08, $p = 0.05$), trust in MT ($B = 0.29$, t (30) = 2.01, $p = 0.05$) and AoA ($B = − 1.8$, t (30) = 6.99, $p < 0.001$) were significant predictors. Participants' accuracy detections of non-firsthand sentences in both directions were significant predictors although in opposite directions. T to E non-firsthand accuracy detection had positively predicted their preference for MT in legal settings. The more accurately they could detect the translations from T to E the more likely they preferred MT. However, their accuracy detections of E to T non-firsthand translations were negatively related to their preference. The more accurately they could detect the translations of non-firsthand sentences the less likely they preferred MT in legal settings. The frequency of MT use was negatively related to preference for MT in legal settings where the more frequently participants use MT the less likely they preferred MT as their translator. As expected, trust in MT is a positively related predictor, the more participants trust in MT the more likely they preferred MT. Finally, the AoA effect was similar to the medical

setting in which late bilinguals were more likely to prefer MT than early bilinguals.

Academic settings The results showed a collective significant effect among all five predictors (F (5, 30) = 5.94, $p < 0.001$, $R^2 = 0.5$). Similar to the other two settings, T to E non-firsthand accuracy detection ($B = 0.59$, t (30) = 2.13, $p = 0.04$), E to T non-firsthand accuracy ($B = − 1.07$, t (30) = 4.54, $p < 0.001$) were revealed as significant predictors. Additionally, trust in MT ($B = − 0.36$, t (30) = 2.03, $p = 0.05$) was another significant predictor. As participants' T to E non-firsthand accuracy detection increased their preference of MT increased. As opposed to T to E, as their E to T non-firsthand accuracy detection increased their preference for MT decreased. Further, as participants indicated less trust in MT, they more likely preferred MT. Differently than the other settings AoA was not a significant predictor of MT preference in academic settings along with the frequency of use.

Daily settings The results demonstrated a collective significant effect among all five predictors (F (5, 30) = 18, $p < 0.001$, $R^2 = 0.75$). The individual predictors were examined further. Differently than the other settings, the translation accuracy detection of non-firsthand sources in both directions were not significant predictors of participants' preference for MT in daily settings. The frequency of MT use ($B = 0.3$, t (30) = 3.06, $p = 0.005$), trust in MT ($B = − 0.71$, t (30) = 5.58, $p < 0.001$) and AoA ($B = − 0.99$, t (30) = 4.33, $p < 0.001$) were significant predictors. The more frequently participants used MT the more likely they preferred to use MT in daily settings. However, the more they trusted in MT the less likely they preferred it in daily settings. The AoA effect was similar to the other setting in which late bilinguals were more likely to prefer MT than early bilinguals.

Discussion
This study focused on assessing the MT accuracy detection capabilities of Turkish–English bilingual speakers, particularly in the context of translating past events, and examined their MT preferences. The choice of past event translations was driven by a typological difference between Turkish and English, specifically related to evidentiality—a structural variation in indicating the source of information about past occurrences. Despite its subtle structural nature, this distinction had significant semantic implications (e.g., Aikhenvald, 2004; Aksu-Koç, 2016; Arslan, 2020; Plungian, 2001; Tosun & Vaid, 2018, in press; Willett, 1988). Given the substantial rise in MT usage and the importance of accurately conveying the source of past events, the ability to detect the accuracy of evidentiality translations in MT became crucial.

The study's findings revealed that both early and late bilinguals exhibited better error detection in firsthand

sentences compared to non-firsthand sentences, but this difference was noticeable only when translations were from Turkish to English. In this context, late bilinguals were particularly adept at detecting machine translations. However, when the translation direction was from English to Turkish, the distinction between firsthand and non-firsthand sources vanished. Although firsthand sentences were somewhat more accurately detected than non-firsthand sentences, this difference did not reach statistical significance. These findings align with previous research, demonstrating that firsthand sources tend to be better remembered (Tosun et al., 2013) and translated more accurately (Tosun & Filipović, 2022).

Consistently, the study also highlighted the influence of AoA. Late bilingual speakers exhibited a greater awareness of the evidentiality distinction compared to their early bilingual counterparts (Arslan et al., 2015, 2017; Arslan & Bastiaanse, 2020; Karayayla, 2020; Schmid & Karayayla, 2020; Tosun et al., 2013; Tosun & Filipović, 2022). Most MT research has traditionally focused on examining proficiency levels rather than AoA. These studies have generally revealed that translators exhibit higher cognitive loads and lower acceptability compared to less proficient bilinguals (Carl et al., 2011; Castilho, 2016; Doherty, 2016; Daems et al., 2017; Ferreiraa et al., 2021; Guerberof Arenas et al., 2021; Hu et al., 2020; Kasperé et al., 2023; Moorkens, 2018; Stasimioti & Sosoni, 2021; Vardaro et al., 2019). In the current study, it was observed that late bilingual speakers exhibited a superior ability to detect errors in translations compared to their early bilingual counterparts. The late bilinguals in the sample also reported a higher overall proficiency level than the early bilinguals. These findings align with previous research on MT, underscoring the influence of proficiency levels and the consistency of these patterns in our study. It is important to highlight that the AoA, age of participants, and the language proficiency of the sample demonstrated a large overlap (where the late bilinguals were older and more proficient than the early bilinguals). This made it challenging to isolate the sole impact of AoA. Previous studies on age and MT usage (e.g., Kasperé et al., 2021) indicate that older individuals generally utilize MT less often compared to younger ones. However, intriguingly, in this study, late bilinguals (the older group) displayed more frequent MT usage than early bilinguals (the younger group), contrary to prior age-related findings. Future studies should prioritize investigating the distinct effects of AoA, age of participants, and language proficiency in machine translation (MT).

Conversely, the findings regarding the direction of translation yielded results that differed from prior research. In the present study, late bilinguals exhibited a greater proficiency in detecting the accuracy of Turkish-to-English translations compared to English-to-Turkish translations. This contrasted with the outcomes of Tosun and Filipović's study (2022), where bilingual speakers displayed more accurate translations when translating English sentences into Turkish. The inconsistency in results could be attributed to procedural differences between the two studies. In the current study, participants were tasked with determining whether the source sentence had been accurately translated into the target language. When the target language was English, participants encountered additional phrases like "it seemed" or "apparently," influencing their decisions. These evidential phrases made the sources of information more conspicuous in the Turkish-to-English direction. In contrast, Tosun and Filipović's experiment required participants to produce the correct markers when translating between the two languages. The production of such phrases is relatively more challenging in English because the language does not mandate speakers to indicate the source of information. Consequently, their study found lower accuracy in translating from Turkish to English. Furthermore, the translation literature has explored the role of direction as a factor affecting translation accuracy (Ferreira & Schwieter, 2017; García et al., 2014). It was concluded that the advantage of forward (L1 to L2) or backward (L2 to L1) translations did not consistently yield robust results and was influenced by various factors, including experience and proficiency.

In terms of MT preference, the study revealed several noteworthy findings. The average accuracy in detecting MT errors related to evidentiality stood at 32%. In terms of preference for MT, the figures were as follows: 46% in medical settings, 53% in legal, 56% in academic, and 38% in daily settings. Participants reported using MT tools with a frequency of 21% and expressed a trust level of 65% in MT. The results demonstrated that, except in daily settings, the accuracy detection of non-firsthand sources significantly predicted preferences for MT. Intriguingly, the direction of translation (from Turkish to English or vice versa) had contrasting effects. Participants who more accurately detected non-firsthand translations from English to Turkish were less likely to prefer MT, while those who more accurately detected such translations from Turkish to English were more inclined to opt for MT. Similarly, in academic and daily contexts, those who reported lower trust in MT were more likely to prefer MT. Additionally, AoA emerged as a significant predictor, except in academic settings, with late bilingual speakers showing a greater preference for MT, even though they exhibited better accuracy in error detection compared to

early bilinguals. These findings were consistent with prior research (e.g., García, 2010; Kasperé et al., 2021, 2023).

Summing up these findings, it becomes evident that despite the ability of bilinguals to spot errors in machine translations and their lack of trust in MT products, they still preferred MT. In response to an open-ended question regarding their preference for MT, participants commonly cited ease of access, cost-effectiveness, and convenience as the primary reasons. As Kasperé et al. () discussed, the bilingual speakers in this study indicated a preference for MT even when translations were only partially accurate because they valued the ability to convey their message quickly and conveniently. However, it is worth noting that trust in MT had a positive correlation with MT preference in legal settings, whereas trust in MT did not significantly predict MT preference in medical contexts. This discrepancy might be attributed to the critical importance of clear communication in legal settings, leading participants to be more cautious in their choice of MT (e.g., Asscher & Glikson, 2023). For future research, it is recommended to delve into the actual use of MT in vital contexts such as medical or legal, assessing both the accuracy of the translations and user satisfaction.

In summary, the study highlighted the influence of AoA on accuracy detection, particularly in aspects where a linguistic property is present grammatically in one language and lexically in the other. Alongside AoA, the accuracy of detecting non-firsthand sources and reported trust in MT emerged as significant predictors of MT preference, although the outcomes varied depending on the translation context. Furthermore, participants offered insights into their preference for MT, emphasizing factors like accessibility, cost-effectiveness, and convenience, even if the translations were only partially correct. For future research, exploring the attitudes and accuracy detection abilities of bilingual speakers with diverse backgrounds, including brokers (who are individuals, often children or adolescents, who facilitate communication between their family members and dominant language speakers), professional translators, and foreign language users, could provide valuable insights. In conclusion, the study underscores the multifaceted factors influencing preferences for MT, highlighting the significance of AoA, accuracy detection, and trust in MT across different translation contexts.

Acknowledgements

Significance statement
This study delved into the attitudes of bilingual users from diverse backgrounds towards machine translation (MT) across medical, legal, academic, and daily contexts. It specifically focused on their proficiency in detecting the accuracy of MT output, particularly when translating between typologically distinct languages like Turkish to English, considering the linguistic concept of evidentiality. Evidentiality, denoting the source of information, presents subtle differences between languages, being optional in English and obligatory in Turkish grammar. Results unveiled that late bilinguals exhibited superior accuracy in detecting translation errors, especially in firsthand evidence translations, compared to early bilinguals. This sheds light on the nuanced variations in how bilinguals process linguistic information, showcasing the diversity within this population. Given the prevalence of bilingual speakers globally, these findings contribute crucial insights into a significant segment of the population. Concerning MT preference, age of acquisition (AoA) and accuracy detection of non-firsthand sentence translations emerged as significant predictors. Late bilinguals, despite their enhanced accuracy detection, displayed a preference for MT when needed. This finding holds relevance for professionals in legal and medical fields involved in multilingual interviews, offering valuable insights into effective communication with bilingual speakers. Moreover, understanding user trust and preferences in MT tools in identifying areas for improvement, providing valuable input for developers to enhance the efficiency of MT tools.

Author contributions
The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding

Availability of data and materials
The data supporting the findings of this study are available from the corresponding author on request.

Declarations

Ethics approval and consent to participate
"All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards." The institutional Review Board of the author's affiliated institution approved the procedure. Informed consent was obtained from all individual participants included in the study.

Consent for publication
The author, the undersigned, give my consent for the publication of identifiable details, which can include details within the text ("Material") to be published in the above Journal.

Competing interests
The author certifies that she has NO affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

References
Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford Univesity Press.
Aksu-Koç, A. (2016). The interface of evidentials and epistemics in Turkish. *Exploring the Turkish Linguistic Landscape: Essays in Honor of Eser Erguvanlı-Taylan, 175*, 143.
Aksu-Koç, A., Ögel-Balaban, H., & Alp, İE. (2009). Evidentials and source knowledge in Turkish'. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development* (pp. 13–28). Jossey-Bass.
Altintas, K., & Cicekli, I. (2022). A machine translation system between a pair of closely related languages. In *International symposium on computer and information sciences* (pp. 192–196). CRC Press.

Arslan, S. (2020). When the owner of information is unsure: Epistemic uncertainty influences evidentiality processing in Turkish. *Lingua, 247*, 102989.

Arslan, S., & Bastiaanse, R. (2020). First language exposure predicts attrition patterns in Turkish heritage speakers' use of grammatical evidentiality. In F. Bayram (Ed.), *Studies in Turkish as a heritage language* (pp. 105–126). John Benjamins.

Arslan, S., Bastiaanse, R., & Felser, C. (2015). Looking at the evidence in visual world: Eye-movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality. *Frontiers in Psychology, 6*, 1387. https://doi.org/10.3389/fpsyg.2015.01387

Arslan, S., De Kok, D., & Bastiaanse, R. (2017). Processing grammatical evidentiality and time reference in Turkish heritage and monolingual speakers. *Bilingualism: Language and Cognition, 20*(3), 457–472.

Asscher, O., & Glikson, E. (2023). Human evaluations of machine translation in an ethically charged situation. *New Media & Society, 25*(5), 1087–1107.

Burchardt, A., Macketanz, V., Dehdari, J., Heigold, G., Jan-Thorsten, P., & Williams, P. (2017). A linguistic evaluation of rule-based, phrase-based, and neural MT engines. *The Prague Bulletin of Mathematical Linguistics, 108*(1), 159.

Cadwell, P., O'Brien, S., & Teixeira, C. S. C. (2018). Resistance and accommodation: Factors for the (non-) adoption of machine translation among professional translators. *Perspectives: Studies in Translatology, 26*(3), 301–321.

Carl, M., Dragsted, B., Elming, J., Hardt, D., & Lykke Jakobsen, A. (2011). The process of post-editing: A pilot study. *Copenhagen Studies in Language* (Frederiksberg), 131–142.

Carl, M., Gutermuth, S., & Hansen-Schirra, S. (2015). Post-editing machine translation. *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting, 115*, 145.

Castilho, S. (2016). *Measuring acceptability of machine translated enterprise content [Unpublished doctoral dissertation]*. Dublin City University.

Castilho, S., & O'Brien, S. (2017). Acceptability of machine-translated content: A multi-language evaluation by translators and end-users. *Linguistica Antverpiensia, New Series–Themes in Translation Studies, 16*.

Castilho, S., O'Brien, S., Alves, F., & O'Brien, M. (2014). Does post-editing increase usability? A study with Brazilian Portuguese as target language. In *Proceedings of the 17th annual conference of the European association for machine translation* (pp. 183–190). European Association for Machine Translation.

Colman, T., Fonteyne, M., Daems, J., & Macken, L. (2021). It's all in the eyes: An eye tracking experiment to assess the readability of machine translated literature. In *31st Meeting of computational linguistics in The Netherlands (CLIN 31)*.

Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology, 8*, 01282. https://doi.org/10.3389/fpsyg.2017.01282

Davenport, C. (2018). Google Translate processes 143 billion words every day. *Android Police, 9*.

Doherty, S. (2016). Translations| the impact of translation technologies on the process and product of translation. *International Journal of Communication, 10*, 23.

Doherty, S., & O'Brien, S. (2014). Assessing the usability of raw machine translated output: A user-centered study using eye tracking. *International Journal of Human Computer Interactions, 30*, 40–51. https://doi.org/10.1080/10447318.2013.802199

Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. (2009). Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior Research Methods, 41*, 1149–1160.

Ferreiraa, A., Griesa, S. T., & Schwieterc, J. W. (2021). Assessing indicators of cognitive effort in professional translators: A study on language dominance and directionality. *Translation, Interpreting, Cognition, 115*.

Ferreira, A., & Schwieter, J. W. (2017). Directionality in translation. In A. Ferreira & J. W. Schewieter (Eds.), *The handbook of translation and cognition* (pp. 90–105). Wiley-Blackwell.

Filipović, L. (2017). Applied language typology: Applying typological insights in practice. *Languages in Contrast, 17*(2), 255–278.

García, l. (2010). Is machine translation ready yet? *Target, 22*, 7–21.

García, A. M., Ibáñez, A., Huepe, D., Houck, A. L., Michon, M., Lezama, C. G., & Chadha, S. (2014). Word reading and translation in bilinguals: The impact of formal and informal translation expertise. *Frontiers in Psychology, 5*, 112660. https://doi.org/10.3389/fpsyg.2014.01302

Gaspari, F., Toral, A., Kumar Naskar, S., Groves, D., & Way, A. (2014). Perception vs reality: Measuring machine translation post-editing productivity. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of the 11th conference of the association for machine translation in the Americas: Workshop on post-editing technology and practice (WPTP3)* (pp. 60–72). AMTA.

Guerberof Arenas, A., Moorkens, J., & O'Brien, S. (2021). The impact of translation modality on user experience: An eye-tracking study of the Microsoft word user interface. *Machine Translation, 35*, 205–237. https://doi.org/10.1007/s10590-021-09267-z

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication, 25*(1), 89–100.

Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., … & Zhou, M. (2018). Achieving human parity on automatic Chinese to English news translation. arXiv:1803.05567.

Hoi, H. T. (2020). Machine translation and its impact in our modern society. *International Journal of Scientific and Technology Research., 9*, 1918–1921.

Hu, K., O'Brien, S., & Kenny, D. (2020). A reception study of machine translated subtitles for MOOCs. *Perspectives, 28*, 521–538. https://doi.org/10.1080/0907676X.2019.1595069

Ishida, K. (2006). How can you be so certain? The use of hearsay evidentials by English speaking learners of Japanese. *Journal of Pragmatics, 38*(8), 1281–1304.

Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. *Copenhagen Studies of Language, 36*, 103–124.

Karayayla, T. (2020). Effects of first language attrition on heritage language input and ultimate attainment: two generations of Turkish immigrants in the UK. In B. Brehmer, J. Treffers-Daller, & D. Berndt (Eds.), *Lost in Transmission: The role of attrition and input in heritage language development.: John Benjamins*

Kasperavičienė, R., Motiejūnienė, J., & Patašienė, I. (2020). Quality assessment of machine translation output. *Texto Livre, 13*, 271–285. https://doi.org/10.35699/1983-3652.2020.24399

Kasperė, R., Horbačauskienė, J., Motiejūnienė, J., Liubinienė, V., Patašienė, I., & Patašius, M. (2021). Towards sustainable use of machine translation: Usability and perceived quality from the end-user perspective. *Sustainability, 13*(23), 13430.

Kasperė, R., Motiejūnienė, J., Patašienė, I., Patašius, M., & Horbačauskienė, J. (2023). Is machine translation a dim technology for its users? An eye tracking study. *Frontiers in Psychology, 14*, 1076379.

Kaya-Soykan, D., Antonova-Unlu, E., & Sagin-Simsek, C. (2023). The production and perception of Turkish evidentiality markers by Turkish–German returnees. *Applied Linguistics Review, 14*(2), 251–270.

Koponen, M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the 7th workshop on statistical machine translation* (pp. 181–190). Association for Computational Linguistics.

Läubli, S. & Orrego-Carmona, D. (2017). When Google Translate is better than some human colleagues, those people are no longer colleagues. In *Proceedings of the 39th conference translating and the computer, one birdcage walk* (pp. 59–69). Editions Tradulex.

Macías, L. P., Ramos, M. D. M. S., & Rico, C. (2020). Study on the usefulness of machine translation in the migratory context: Analysis of translators' perceptions. *Open Linguistics, 6*(1), 68–76.

Melby, A. K. (2020). Future of machine translation: Musings on Weaver's memo. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 419–436). Routledge.

Moorkens, J. (2018). Chapter eye-tracking as a measure of cognitive effort for post-editing of machine translation. *Eye Tracking and Multidisciplinary Studies on Translation*, 55–69.

Moorkens, J., O'Brien, S., da Silva, I. A. L., de Lima Fonseca, N. B., & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation, 29*(3–4), 267–284. https://doi.org/10.1007/s10590-015-9175-2

Narita, R. (2011). The effects of pragmatic consciousness-raising activity on the development of pragmatic awareness and use of hearsay evidential markers for learners of Japanese as a foreign language. *Journal of Pragmatics, 44*(1), 1–29.

Nurminen, M., & Koponen, M. (2020). Machine translation and fair access to information. *Translation Spaces, 9*, 150–169. https://doi.org/10.1075/ts.00025.nur

Ortega, J., Sánchez-Martínez, F., Turchi, M., & Negri, M. (2019). Improving translations by combining fuzzy-match repair with automatic post-editing. In *Proceedings of machine translation summit XVII: Research track* (pp. 256–266). European Association for Machine Translation.

Öztürk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*, 199–230.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., & Grosz, B. (2019). *The AI index 2019 annual report.* Technical report, Stanford: AI Index Steering Committee, Human-Centered AI Institute, Stanford University.

Plungian, V. A. (2001). The place of evidentiality within the universal grammatical space. *Journal of Pragmatics, 33*(3), 349–357.

Popović, M. (2017). chrF++: words helping character n-grams. In *Proceedings of the second conference on machine translation* (pp. 612–618).

Pym, A. (2020). Quality. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 437–452). Routledge.

Rivera-Trigueros, I. (2021). Machine translation systems and quality assessment: A systematic review. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-021-09537-5

Rossetti, A., O'Brien, S., & Cadwell, P. (2020). Comprehension and trust in crises: investigating the impact of machine translation and post-editing. In *Proceedings of the 22nd annual conference of the European Association for machine translation* (pp. 9–18). European Association for Machine Translation.

Rossi, C., & Carré, A. (2022). How to choose a suitable neural machine translation solution: Evaluation of MT quality. Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence. https://doi.org/10.5281/zenodo.6759978

Rossi, C., & Chevrot, J.-P. (2019). Uses and perceptions of machine translation at the European Commission. *Journal of Specialized Translation, 31*, 177–200.

Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts. *Journal of Specialised Translation, 31*, 201–216.

Schmid, M. S., & Karayayla, T. (2020). The roles of age, attitude, and use in first language development and attrition of Turkish-English bilinguals. *Language Learning, 70*(S1), 54–84.

Stasimioti, M., & Sosoni, V. (2021). Chapter 5: Investigating post-editing: A mixed-methods study with experienced and novice translators in the English-Greek language pair. In Tra&Co (Ed.), Translation, *Interpreting, cognition: The way out of the box* (pp. 79–104). Language Science Press.

Taivalkoski-Shilov, K., Toral, A., Hadley, J. L., & Teixeira, C. S. C. (2022). *Using technologies for creative-text translation. Routledge Advances in Translation and Interpreting Studies*. Routledge.

Tan, X., Chen, J., He, D., Xia, Y., Qin, T., & Liu, T. Y. (2019). Multilingual neural machine translation with language clustering. arXiv:1908.09324.

Taylor, R. M., Crichton, N., Moult, B., & Gibson, F. (2015). A prospective observational study of machine translation software to overcome the challenge of including ethnic diversity in healthcare research. *Nursing Open, 2*, 14–23. https://doi.org/10.1002/nop2.13

Teixeira, C. S. C. (2014). Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. In S. O'Brien, M. Simard, and L. Specia (Eds.), *Proceedings of the 11th conference of the association for machine translation in the Americas: Workshop on post-editing technology and practice (WPTP3)* (pp. 450–459). AMTA.

Toral, A, Castilho, S, & Hu, K. (2019). Attaining the unattainable? Reassessing claims of human parity in neural machine translation. In *Proceedings of the 3rd conference on machine translation: Research Papers* (pp. 113–123). Association for Computational Linguistics.

Tosun, S., & Filipović, L. (2022). Lost in translation, apparently: Bilingual language processing of evidentiality in a Turkish–English translation and judgment task. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728922000141

Tosun, S., & Vaid, J. (in press). Does firsthand vs. non-firsthand evidential framing affect a narrative's construal? A preliminary study of Turkish and English users. *Cognitive Linguistic Studies.*

Tosun, S., & Vaid, J. (2018). Activation of source and stance in interpreting evidential and modal expressions in Turkish and English. *Dialogue and Discourse, 9*(1), 128–162.

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence influence source memory? A Turkish/English investigation. *Journal of Memory and Language, 69*(2), 121–134.

Turovsky, Y., Wolf, D., Meshcheryakov, R., & Iskhakova, A. (2022). Dynamics of frequency characteristics of visually evoked potentials of electroencephalography during the work with brain-computer interfaces. In *International conference on speech and computer* (pp. 676–687). Springer.

Ueffing, N. (2018). Automatic post-editing and machine translation quality estimation at eBay. In *Proceedings of the AMTA 2018 workshop on translation quality estimation and automatic post-editing* (pp. 1–34). Association for Machine Translation in the Americas.

Vardaro, J., Schaeffer, M., & Hansen-Schirra, S. (2019). Translation quality and error recognition in professional neural machine translation post-editing. *Informatics, 6*, 41. https://doi.org/10.3390/informatics6030041

Vieira, L. N. (2020). Machine translation in the news. *Translation Spaces, 9*, 98–122. https://doi.org/10.1075/ts.00023.nun

Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases. *Information. Communication & Society, 24*, 1515–1532. https://doi.org/10.1080/1369118X.2020.1776370

Willett, T. (1988). A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language, 12*(1), 51–97.

Yasuoka, M., & Bjorn, P. (2011). Machine translation effect on communication: What makes it difficult to communicate through machine translation? In *2011 Second international conference on culture and computing*. IEEE.

Publisher's Note