ISSN 1799-2591 (print); ISSN 2053-0692 (online)
Theory and Practice in Language Studies, Vol. 16, No. 6, pp. 1906-1915, June 2026
DOI: https://doi.org/10.17507/tpls.1606.11

Evaluative Voice and Interpersonal Shifts in ChatGPT-Translated Feedback for EFL Writing

Asma Alshehri
King Saud University, Saudi Arabia

Abstract—**This study investigates how ChatGPT's machine translation from English to Arabic affects the realization of interpersonal meaning and learner engagement in academic feedback for EFL learners. The main aim is to compare evaluative features in original and machine-translated feedback, addressing a significant gap in understanding the implications of AI-assisted translation for EFL writing pedagogy. Adopting qualitative comparative analysis, ChatGPT-generated English feedback and its Arabic translations were examined through the lenses of Appraisal Theory by Martin and White (2005) and Hyland's (2006) Reader Engagement Model. The findings reveal that in the Arabic translations the authorial voice shifts to be more distant and directive, often resulting in more instructional but less empathetic feedback. The study concludes that while ChatGPT feedback has potential, human oversight is essential to maintain interpersonal depth, learner engagement, supportive and personalized characteristics crucial for effective feedback in EFL contexts. Future research is recommended to explore EFL learner perceptions of translated feedback and to develop strategies for achieving the full pedagogical potential of AI-assisted feedback translation.**

Index Terms—**ChatGPT feedback, English-Arabic translation, interpersonal meaning, learner engagement**

I. Introduction

The integration of generative artificial intelligence, such as ChatGPT, into translation and language instruction has recently been discussed in education. ChatGPT is increasingly used to generate formative feedback in EFL academic writing contexts. Effective feedback positively impacts EFL learners' cognitive, affective, and behavioral engagement (Hyland & Jiang, 2016). The affective dimension is essential to language learning since the students' emotional reactions and attitudinal responses to the feedback they receive are of great importance (Carless, 2022; Hwang, 2025; Sun, 2021).

Due to recent developments and the integration of AI in writing pedagogy, some studies explored the effectiveness of AI tools in providing feedback (Alshehri, 2025; Rad et al., 2024), reporting promising findings. Other studies compared traditional teacher feedback to AI-generated feedback. They found that teacher feedback cannot be replaced by AI-generated feedback, but supplemented by it (Awidi, 2024; Dai et al., 2024; Sardinha, 2024; Steiss et al., 2024). Additionally, they argued that teacher individualized praise, empathy and expressions of encouragement remain unique qualities and are not yet fully available in Large Language Models (LLMs). Therefore, questions arise about the interpersonal elements and the preservation of dialogic tone in feedback translated by ChatGPT into Arabic compared to that originally generated in English. Understanding the evaluative linguistic choices and the emotional framing of AI translated feedback into Arabic is essential for making pedagogical decisions in EFL contexts.

Despite growing interest in AI-generated feedback and the role of AI in translation, little is known about how the dialogic nature and evaluative features of feedback are transformed and maintained when ChatGPT-generated English feedback is translated into Arabic. To this end, the present study adopts a dual-framework approach, it draws on the Appraisal Theory Framework by Martin and White (2005) and Hyland's (2006) Reader Engagement Framework to systematically analyze shifts in interpersonal meaning and learner engagement positioning strategies in feedback generated in English and then translated into Arabic by ChatGPT for intermediate proficiency level EFL learners. This study sets to explore the extent to which ChatGPT can generate feedback with the same qualities of authorial stance and functions of supportive, and empathetic tone, in both original (i.e., English) translated (i.e., Arabic) ChatGPT-generated feedback, foregrounding the implications for Arabic-speaking learners in EFL academic contexts. Therefore, the current study aims to address this gap by answering the following research questions:

1. How does ChatGPT's machine translation from English to Arabic affect the realization of appraisal and engagement resources in academic feedback?
2. What challenges and patterns emerge in maintaining authorial tone and learner positioning in AI translated feedback?

II. Literature Review

ChatGPT-generated feedback, translation systems, and the use of Appraisal Theory in translation evaluation, have recently become important topics in linguistic research. The following review synthesizes findings from key studies on these topics, aligned with the objectives and questions of the present study aiming to explore interpersonal meaning, evaluative resources, and learner positioning in ChatGPT-generated feedback in English and Arabic for EFL learners.

A. ChatGPT-Generated Feedback Studies

Research comparing the efficacy of ChatGPT-generated feedback to traditional feedback in writing instruction has produced slightly similar results. For instance, Banihashem et al. (2024) compared automated feedback to peer feedback in argumentative writing. They found that ChatGPT provided descriptive comments, while peer feedback excelled at identifying writing problems. Alsofyani and Barzanji (2024) found that ChatGPT's feedback was as effective as teacher feedback in improving learner writing. In another study by Cao and Zhong (2023), they evaluated ChatGPT's performance as a feedback provider in Chinese and English translation tasks. They found that teacher feedback outperformed ChatGPT-generated feedback. However, ChatGPT excelled at enhancing lexical choice and cohesion. Also, Lin and Crosthwaite (2024) compared feedback in academic writing courses by ChatGPT to human teachers' feedback. The findings revealed that unlike teacher feedback, ChatGPT failed to provide individualized learner responses based on their proficiency levels. In addition, Steiss et al. (2024) noted that while teachers gave higher quality feedback in most domains, the consistency and quick accessibility of ChatGPT suggest a valuable supporting role of AI in academic writing assessment. Zou et al. (2025) explored learner perceptions of feedback generated by ChatGPT. They found that teacher praise, encouragement and empathy motivate learners and highly impact their engagement with feedback. In addition, Awidi (2024) compared teachers to ChatGPT in providing personalized feedback on reflective texts. The findings revealed that human oversight is essential when ChatGPT feedback is integrated into language learning educational contexts. In summary, the studies in the literature highlight the complementary potential of ChatGPT to teacher- led feedback in writing instruction.

B. Evaluative Meaning in Translation Research

Appraisal Theory by Martin and White (2005) offers a powerful lens for exploring evaluative meaning across source and target texts. Appraisal theory is concerned with how writers take a stance, voice their opinions, and position their readers. It provides writers with linguistic resources that allow them to be critical, to reject, to accept, to appreciate, and to challenge other positions. Appraisal theory has three subsystems: (1) attitude that refers to the resources used to positively or negatively evaluate phenomena; (2) engagement attends to the linguistics resources used to take a stance towards the positions and values in the text with consideration of their readers; (3) graduation is concerned with how writers decrease or increase the interpersonal impact of their feelings, judgements and assessments (i.e., toning-down or toning-up). In the context of academic feedback, teachers engage with EFL learners based on previous knowledge of their learners. As a result, the teacher's knowledge of the learners, shapes feedback and defines the rhetorical strategies employed. Martin and White (2005) argue that appraisal theory not only explores the writer's values and attitudes but also highlights the relations of rapport and solidarity between the teachers/ ChatGPT (i.e., producer of the text) and EFL learners (i.e., the receivers). Qian (2017) applied appraisal theory framework to bilingual public notices in English and Chinese and found that translations often showed reduced attitudinal resources and less dialogic engagement. Findings also revealed some shifts in sentence structure from declarative to imperative structures. Such functional adaptation suggests that some loss of interpersonal meaning in translation is expected. In addition, Li et al. (2025) drew on the appraisal framework to investigate how undergraduate students translated evaluative meanings in news articles from English to Chinese. The findings showed frequent misrepresentations and omissions of source text interpersonal meaning, such as shifting positive attitudes to be more neutral or negative. They argued that integrating appraisal-based analysis into translator training may raise awareness of evaluative expression and authorial stance. Moreover, Tajvidi and Arjani (2017) reviewed appraisal-driven translation studies across various genres and languages. They highlighted key trends such as tendencies for translated texts to exhibit frequent shifts in attitude, engagement and graduation. Typical trends included reduction or loss of attitudinal markers, adaptation of authorial stance, and alteration of reader positioning and rhetorical purpose in target texts. These tendencies can be due to processing constraints, target culture conventions, and institutional or personal ideological agendas. Abbamonte and Cavaliere (2006) found that translated news texts became less emotional and more distant. Munday (2009) reported changes in writer-reader relationships leading to less authorial presence in the translated political texts. Rosa (2013) found that translated literary texts exhibited less narrative power and more solidarity with readers. Zhang (2013) reported increased intensifications of negative judgments in the translated texts of news headlines. Al-Shunnag (2014) found that in translated news texts, authorial stance was realized more through value-laden lexis and frequent omissions of intensification resources. In summary, the existing literature suggests that some translations intensify attitudinal markers, while others downplay or omit evaluative and dialogic features, often shifting the overall interpersonal tenor of the original text.

C. ChatGPT Translation for English to Arabic Texts

Recent studies in the literature explored the quality of LLM translation for English-Arabic academic texts. Mohsen (2024) compared ChatGPT-4 to Google Translate and found that ChatGPT-4 outperformed traditional machine translation of academic texts in terms of accuracy and clarity. Similarly, Alkhawaja (2024) reported similar findings when he compared ChatGPT-generated translation to other machine translations. However, Khondaker et al. (2023) argued that ChatGPT and Google fail to translate cultural and aesthetic aspects of texts, such as proverbs. These findings may suggest challenges in preserving affective and evaluative features central to interpersonal and dialogic tone in ChatGPT feedback translation.

The present literature review highlights that ChatGPT provides effective feedback and improved translation performance compared to traditional systems in academic contexts. However, limitations persist in evaluative-based

research in AI-generated translation. The field calls for further investigation of evaluative shifts in translated texts produced by AI systems like ChatGPT, which are increasingly integrated into academic and educational contexts.

III. Methodology

A. Research Design

This study employs a qualitative comparative analysis to investigate the transformation of evaluative and interpersonal meanings in AI-generated academic feedback. The focus is on feedback produced by ChatGPT in English and its translation into Arabic. The analysis draws on Martin and White's (2005) Appraisal Theory and Hyland's (2006) Categories of Reader Engagement to capture shifts in interpersonal meaning and learner positioning within feedback for B1-level EFL Arabic-speaking learners.

B. Data

For the present study, twenty academic texts were selected. The texts were in English and written by B1-level Arabic-speaking EFL learners. For each text, feedback was generated using ChatGPT-4 in English, then subsequently translated into Arabic by ChatGPT's own translation function. The resulting corpus consists of twenty pairs of English and Arabic feedback texts, enabling direct comparative analysis of original and translated feedback at both the utterance and discourse levels.

C. Analytical Frameworks

Appraisal Theory (Martin & White, 2005)

All feedback texts in English and Arabic were coded for Attitude (Affect, Judgment, Appreciation), Engagement (contractive/expansive), and Graduation (force/focus). Appraisal's three subsystems provide writers with the means to value, to accept, to reject, and to challenge other points of view. The first subsystem is attitude, and it refers to the resources used by writers to make negative or positive evaluations. Attitude includes three categories; affect, judgement and appreciation. Affect deals with how writers express their feelings (e.g., feeling sad, bored, happy, worried, or interested), judgment attends to the moral judgments of human behavior (e.g., admiring, criticizing, or praising) and appreciation refers to the evaluation of phenomena but not people's behaviors. Second, engagement refers to the linguistic resources that writers use to take a stance towards the propositions in the text and with consideration of their reader audience. Third, graduation deals with how writers increase or decrease (i.e., toning-down or toning-up) the interpersonal impact of their feelings, judgements and assessments. The Appraisal Theory framework was selected because it reveals how feedback conveys approval, criticism, encouragement, and alignment which are all crucial factors in shaping student motivation. The three subsystems allow for a detailed examination of how evaluative meanings are constructed in academic feedback and how these meanings may be preserved, transformed or lost through ChatGPT translation. Therefore, this framework directly addresses the study's focus on evaluative and dialogic meanings across languages.

Hyland's (2006) Categories of Reader Engagement

This framework was used to locate and identify instances of reader engagement in the feedback generated by ChatGPT in English and in the translated version into Arabic. The instances were categorized according to Hyland's model, including reader mentions, questions, appeals to shared knowledge, directives, and personal asides. Hyland's engagement model specifically focuses on how writers interact with and involve their audience readers through strategic linguistics choices. Therefore, this framework complements appraisal theory in capturing evaluative stance and dialogic and participatory qualities of feedback. Its categories identify key linguistic elements that play an essential role in establishing rapport, guiding learner action, and encouraging active engagement with feedback.

Together, the two frameworks offer a robust tool for revealing evaluative stance and interactive engagement qualities of feedback before and after translation by ChatGPT. Their combined use accounts for what meaning is conveyed (appraisal theory) and how students are drawn into a dialogic and interactive process (Hyland's engagement model). This provides a holistic overview of how interpersonal meaning in feedback is effectively shaped, reshaped or transformed by AI across languages in EFL contexts.

D. Data Analysis

First, all feedback texts in English and Arabic were automatically coded for appraisal using the Appraisal Theory Tagger Tool for coding. The Appraisal Theory Tagger is free and automatically annotates texts based on the three subsystems of appraisal (Alshehri et al., 2025). Then, the two sets of texts (i.e., English and Arabic) were manually coded for reader engagement resources based on Hyland's five categories of reader engagement. Second, the coded English and Arabic feedback were compared to identify patterns of loss, transformation, or amplification of evaluative and reader engagement meanings during translation. Finally, recurring patterns and translation-specific challenges were identified and classified. This allowed the researcher to address how evaluative and dialogic meanings are affected and to identify implications for EFL learner perceptions and pedagogical effectiveness. For coding reliability purposes, the data was coded twice. First, it was imported into the Appraisal Theory Tagger Tool for coding and manually for Hyland's engagement model. Then, the researcher recoded 20% of the data to guarantee consistency. For ethical considerations,

informed consent was ensured for the use of anonymized EFL learner texts. The study also observes appropriate ethical guidelines for the use of AI-generated data.

IV. Results

The present study aimed to explore the extent to which ChatGPT can generate feedback with the same qualities of authorial stance and functions of supportive, and empathetic tone, in both feedback produced by ChatGPT in English and in Arabic (i.e., translated by ChatGPT). Therefore, the study had two objectives. The first was to explore how ChatGPT's machine translation from English to Arabic affects the realization of appraisal and engagement resources in academic feedback. The second was to identify challenges that emerge in maintaining the authorial tone and learner positioning strategies when feedback is translated by AI into Arabic.

To achieve the research objectives and answer the research questions, evaluative resources and learner engagement strategies were explored and compared in the two sets of texts. The three subsystems of appraisal were explored, and the findings are illustrated in Figure 1.

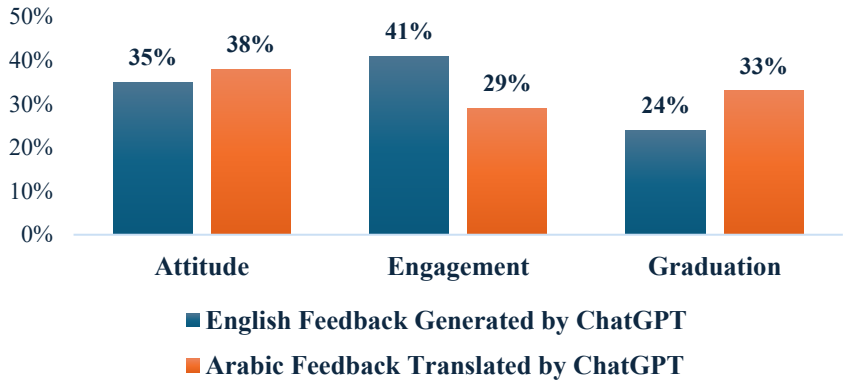

Figure 1. Distribution of Appraisal Systems in the Data

The findings illustrated in Figure 1 demonstrate some variation in the distribution of appraisal systems across the two datasets. The figure shows that the resources of engagement were the most frequent in the English feedback accounting for 41%, while the most frequent appraisal system in the Arabic Feedback was attitude accounting for 38%. The least frequent appraisal system in both sets of feedback was graduation accounting for 24% in the English feedback and for 33% in the Arabic feedback. In the feedback in Arabic, the prevalence of attitude resources indicates an emphasis on expressing feelings, judgments, and appreciations of the students writing in the texts. Graduation resources were the least frequently employed in both sets of data indicating similar tendencies in scaling or intensifying the evaluative meanings.

Each subsystem was explored in detail. First, the subsystem of attitude was observed in both sets of feedback as presented in Table 1.

Table 1
Distribution of Attitude Resources in the Data

	Affect		Judgement		Appreciation		Total	
	No	%	No	%	No	%	No	%
English Feedback Generated by ChatGPT	1	0.5%	25	14%	157	85.5%	**183**	**100%**
Arabic Feedback Translated by ChatGPT	0	0%	15	11%	119	89%	**134**	

The findings in Table 1, show similar distribution of affect, judgment, and appreciation resources when learner feedback is generated in English and then translated into Arabic by ChatGPT. In the original English feedback, appreciation overwhelmingly dominates (85.5%) of the resources, with judgment accounting for 14% and affect being infrequent at only 0.5%. After translation into Arabic, appreciation increases further to 89%, while judgment decreases to 11% and affect drops to 0%, indicating a complete absence of authorial feelings or emotional language. These shifts may suggest that translating feedback into Arabic by ChatGPT highlights descriptive and evaluative language about qualities and processes (i.e., appreciation) but diminishes explicit evaluations of behavior (i.e., judgments) and authorial emotional engagement (i.e., affect). The results suggest a tendency for AI translated feedback to become more impersonal and objective. Example excerpts illustrating attitudinal resources:

	Feedback in English	Feedback translated into Arabic
1.	The text has a **clear** [Attitude; appreciation; positive] introduction that sets the stage for the main points.	لدي النص مقدمة **واضحة** [Attitude; appreciation; positive] تضع الأساس للنقاط الرئيسية.
2.	Overall, this is a **solid** effort [Attitude; judgement; positive].	بشكل عام، هذا جهد **جيد** [Attitude; judgement; positive]
3.	The phrase "and another way to make your advertisement effective' is used repeatedly, which makes the text **feel redundant** [attitude; affect; negative].	عبارة "and another way to make your advertisement effective" تُستخدم بشكل متكرر، مما قد يجعل النص يشعر بالتكرار.
4.	Certain phrases are unclear or **awkwardly** [Attitude; judgement; negative (criticize)] constructed.	بعض العبارات غير واضحة أو مبنية بشكل **غير مناسب** [attitude; appreciation; negative]

In examples 1 and 2, the attitudinal expressions were preserved in Arabic translations. However, in example 3, the authorial affect was mistranslated and consequently omitted. In example 4, the sentence contained criticism and judgment, but during translation into Arabic, the judgmental aspect was canceled and was shifted to an evaluation of objects instead of judgment of human behavior. Second, the polarity of the attitudinal resources was explored in both sets of feedback as presented in Table 2.

TABLE 2
POLARITY IN ATTITUDE RESOURCES

	Positive Attitude		Negative Attitude		Total	
	No	**%**	**No**	**%**	**No**	**%**
English Feedback Generated by ChatGPT	123	67%	60	33%	**183**	**100%**
Arabic Feedback Translated by ChatGPT	82	61%	52	39%	**134**	

The findings in Table 2 reveal similar frequencies in the polarity of attitude resources when feedback is generated in English and subsequently translated into Arabic by ChatGPT. In the English feedback, most of the attitude resources are positively encoded accounting for 67%, while negatively encoded attitudes account for 33% of the instances. When the feedback was translated into Arabic, the frequency of positive attitudes decreased to 61%, whereas negative attitudes increased to 39% of the total. These results indicate that the translation process led to a slight shift toward more negative evaluative language in the Arabic feedback compared to the original English. Examples from the data are as follows:

	Feedback in English	Feedback translated into Arabic
5.	Some sentences are quite **long** [Attitude; appreciation; negative] and can be broken down for clarity.	بعض الجمل **طويلة** [Attitude; appreciation; negative] جدًا ويمكن تقسيمها لزيادة الوضوح.
6.	The text contains punctuation **errors** [Attitude; appreciation; negative], such as the **incorrect** [Attitude; appreciation; negative] use of commas and missing conjunctions.	النص يحتوي على **أخطاء** [Attitude; appreciation; negative] في الترقيم، مثل الاستخدام **غير الصحيح** [Attitude; appreciation; negative] للفواصل وغياب أدوات الربط.
7.	Some phrases are **unclear** [Attitude; appreciation; negative] or **poorly** [Attitude; appreciation; negative] structured.	بعض العبارات **غير واضحة** [Attitude; appreciation; negative] أو **غير منظمة** [Attitude; appreciation; negative] جيدًا.

As indicated in examples 5-7, there were both positive and negative evaluative expressions, mostly in the form of appreciation. Next, the system of engagement provides a systematic analysis of how writers can linguistically voice their views and how they position their audience readers. Engagement in the context of this study refers to the linguistics resources employed by the authorial voice in both sets of data to engage with the EFL learners in the feedback. According to Martin and White (2005) there are two types of engagement, contractive and expansive. When an utterance is dialogic and allows for alternative voices and positions, it is considered expansive, but when it restricts or challenges alternative positions, it is contractive. The dialogic strategies in both sets of feedback are presented in Table 3.

TABLE 3
DISTRIBUTION OF ENGAGEMENT RESOURCES

			English Feedback Generated by ChatGPT		Arabic Feedback Translated by ChatGPT	
			No	**%**	**No**	**%**
Contractive	Disclaim	Deny	3	1%	0	0%
		Counter	36	17%	12	11%
	Proclaim	Concur	0	0%	0	0%
		Pronounce	5	2%	1	1%
		Endorse	38	18%	6	6%
Total			**82**	**38%**	**19**	**18%**
Expansive	Entertain		125	58%	86	82%
	Attribute	Acknowledge	10	4%	0	0%
		Distance	0	0%	0	0%
Total			**135**	**62%**	**86**	**82%**

As presented in Table 3, clear differences in the use of engagement resources between English feedback produced by ChatGPT and its Arabic translation were found. In the English feedback, expansive engagement resources were dominant, accounting for 62%, while contractive resources accounted for 38%. However, after translation into Arabic, the frequency of expansive resources increased to 82%, and contractive resources decreased markedly to only 18%. This shift suggests

that the authorial tone in translated Arabic feedback is more dialogic and opens up space allowing for alternative viewpoints. However, contractive engagement strategies that limit alternatives and assert stronger authorial tone are reduced. In summary, the contractive engagement strategies were higher in the English feedback than in Arabic and expansive engagement strategies increased after translating feedback into Arabic. Examples (8) and (10) are illustrations of expansive engagement where the authorial voice allows for dialogic openness, while in example (9) the authorial voice contracts the dialogic space.

	Feedback in English	Feedback translated into Arabic
8.	Some sentences are quite long and **can** [Engagement; expand; entertain] be broken down for clarity.	**يمكن** [Engagement; expand; entertain] تقسيمها لزيادة الوضوح.
9.	The essay **demonstrates** [Engagement; contract; proclaim; endorse] a good understanding of the topic.	**يُظهر** [Engagement; contract; proclaim; endorse] المقال فهماً جيدًا للموضوع.
10.	You've **clearly** [Engagement; contract; proclaim; pronounce] put thought into identifying several key factors that contribute to effective advertising.	**ومن الواضح** [Engagement; contract; proclaim; pronounce] أنك فكّرت جيدًا في تحديد عدة عوامل رئيسية تساهم في الإعلان الفعّال.

Graduation is a subsystem of appraisal, and it refers to how the authorial voice grades feelings, judgements and evaluations. It is how meanings are either toned down or toned up. For example, the word mistake can be toned down by saying a small mistake or it can be toned up by saying a huge mistake. Martin and White (2005) noted that force refers to the grading of meaning from low to high intensity, while focus refers to the sharpening (e.g., the sentence structure was really weak) or softening meanings (e.g., the sentence structure was sort of weak). The graduation resources employed in both sets of data is illustrated in Table 4.

TABLE 4
DISTRIBUTION OF GRADUATION RESOURCES

	Force				Focus				Total	
	Upscale		Downscale		Sharpen		Soften			
	No	%	No	%	No	%	No	%	No	%
English Feedback Generated by ChatGPT	96	74%	33	26%	0	0%	0	0%	**129**	**100**
Arabic Feedback Translated by ChatGPT	77	65%	34	29%	7	6%	0	0%	**118**	**100**

The findings in Table 4 reveal some differences in the use of graduation resources between English feedback generated by ChatGPT and its Arabic translation. In the English feedback, graduation is realized through force exclusively with no presence of focus resources (sharpen or soften). Resources of upscaling (intensifying meanings) were more dominant at 74%, and downscaling (softening meanings) at 26%. When the feedback was translated into Arabic, force resources remained prevalent but decreased slightly. The upscaling of meanings decreased to 65% and downscaling increased to 29%. Notably, instances of focus-sharpen emerged in the Arabic translation accounting for 6%, whereas focus-soften remained unused. The following are excerpts from the data:

	Feedback in English	Feedback translated into Arabic
11.	There are **several** [Graduation; force; upscale] grammatical errors that can distract the reader.	هناك **العديد** [Graduation; force; upscale] من الأخطاء النحوية التي قد تشتت انتباه القارئ.
12.	With **a few** [Graduation; force; downscale] adjustments for clarity, punctuation, and sentence structure, your insights will come across even **more** [Graduation; force; upscale] effectively.	مع **بعض** [Graduation; force; downscale] التعديلات من حيث الوضوح، الترقيم، وتركيب الجمل، ستظهر رؤاك بشكل **أكثر** [Graduation; force; upscale] فعالية.
13.	**Some** [Graduation; force; downscale] sentences are **a bit** [Graduation; force; downscale] fragmented.	**بعض** [Graduation; force; downscale] الجمل يُمكن أن تكون مجزأة **قليلًا** [Graduation; force; downscale]
14.	No instances of sharpen in the English data.	**بعض** [Graduation; force; downscale] الجمل معقدة جدًا [Graduation;] أو مجزأة. [focus; sharpen

To analyze how feedback generated by ChatGPT in English and its Arabic translation engages learners, Hyland's engagement framework was applied to both datasets. According to Hyland (2006), instances of reader engagement are categorized into five types: reader mentions, questions, appeals to shared knowledge, directives, and personal asides. This framework focuses on the strategic linguistic choices that the authorial voice employes to interact with and involve its learners, highlighting dialogic and participatory qualities in the feedback. In the present study, Hyland's model provides insights into how linguistic elements, such as learner direct address, instructional directives can establish rapport and encourage learner engagement with feedback. The distribution of Hyland's engagement features in the English and Arabic datasets is presented in Table 5, enabling a comparative overview of how ChatGPT's reader engagement strategies shift from English to Arabic.

TABLE 5
HYLAND'S CATEGORIES OF READER ENGAGEMENT IN THE DATA

	English Feedback Generated by ChatGPT		Arabic Feedback Translated by ChatGPT	
	No	%	No	%
Reader mentions	98	72%	108	57%
Questions	0	0%	0	0%
Appeals to shared knowledge	0	0%	0	0%
Directives	38	28%	80	43%
Personal asides	0	0%	0	0%
Total	**136**	**100%**	**188**	**100%**

The findings in Table 5 show interesting shifts in Hyland's reader engagement categories when learner feedback is generated in English by ChatGPT and then translated into Arabic. There were more instances of reader engagement in the Arabic data compared to the English dataset. In the English feedback, reader mentions (addressing the learner directly and the readers) were the most frequent accounting for 72%, while directives (explicit instructions or suggestions to the learner) accounted for 28%. However, in Arabic translations, the engagement instances were split between reader mentions and directives, accounting for 57% and 43%, respectively. Engagement categories such as questions, appeals to shared knowledge, and personal asides were entirely absent in both English and Arabic outputs. Examples from both sets of data are illustrated below: Examples 15 - 20 provide concrete illustrations of how Hyland's reader engagement categories particularly reader mentions and directives are realized in both English feedback and their Arabic translations.

	Feedback in English	Feedback translated into Arabic
15.	This helps structure **your** [Learner Mention] argument effectively.	**لقد أدرجت** [Learner Mention]مجموعة غنية من الأفكار الجيدة ومن الواضح **أنك** **فكّرت** [Learner Mention] جيدًا في تحديد عدة عوامل رئيسية تساهم في الإعلان الفعّال.
16.	By focusing on clarity, coherence, and precision in **your** [Learner Mention] writing, **you** [Learner Mention] can enhance **your** [Learner Mention] message even further.	لقد **أبرزت** [Learner Mention] بشكل فعّال عدة أساليب مهمة.
17.	**Ensure** [Directives] that **you** [Learner Mention] use the correct form of words consistently.	**تأكد** من استخدام الترقيم بشكل موحد عند سرد الأفكار
18.	**Keep working** [Directives] on **your** [Learner Mention] writing	**استمر** [Directives]في تطوير كتابتك، **وواصل** [Directives]مشاركة **رؤاك** القيمة! [Learner Mention]
19.	**Pay attention** [Directives] to plural forms and proper punctuation throughout the text to enhance professionalism.	**الأفضل أن تقول** [Directives]"إنشاء إعلانات فعّالة هو أحد الطرق للتميز."

Examples 15 and 16 demonstrate the use of reader mentions, where learners are directly addressed through personal pronouns such as "your" and "you" in English, and their marked equivalents in Arabic (e.g., " لقد أدرجت", "أنك فكّرت", "ابرزت""). The use of reader mentions helps establish a dialogic relationship with the learners and positions them as active participants. Examples 17, 18, and 19 illustrate how directives are used as explicit instructions and encouragement (e.g., "Ensure that you use...", "Keep working on your writing", "Pay attention to plural forms"). In the Arabic feedback, these are presented as imperatives such as "استمر", "تأكد", maintaining a direct and instructional tone. In both sets of feedback, directives served to guide learners toward improvement.

V. DISCUSSION

The first research objective was to explore how ChatGPT's machine translation from English to Arabic affects the realization of appraisal and learner engagement resources in academic feedback. The findings revealed that ChatGPT's machine translation from English to Arabic introduces shifts in appraisal and engagement in academic writing feedback. In both English and Arabic feedback, instances of appreciation overwhelmingly dominated the attitude subsystems. Affect and judgment occurred less frequently especially after ChatGPT translation into Arabic, where instances of affect disappeared and judgment decreased further. This aligns with findings reported by Qian (2017), Tajvidi and Arjani (2017), and Abbamonte and Cavaliere (2006), who found that translation often leads to decreased affect, resulting in a more distanced tone. Additionally, the present data revealed that the translated feedback expressed a more objective authorial tone. This finding mirrors Munday's (2009) observations, who found reduced authorial presence in translated political texts.

Polarity analysis revealed that the frequency of negative attitude tokens increased in Arabic feedback (from 33% to 39%), and positive evaluations slightly decreased (from 67% to 61%). These results indicate that the translation process led to a slight shift toward more negative evaluative language in the Arabic feedback compared to the original English. This aligns with Li et al. (2025) and Zhang (2013), who found that translation can increase the frequency of negativity polarity or shift source text evaluations to be more neutral or critical. The findings suggest implications for how translated feedback may be perceived by EFL learners. This is a potential concern for EFL contexts where praise and encouragement in writing feedback are pedagogically valued (Sun, 2021; Zou et al., 2025).

In English and Arabic feedback, engagement resources, central to establishing dialogic space, are similarly transformed. In English feedback, expansive markers were dominate, and the frequency of contractive instances (38%) reflected an authorial tone that is assertive and guiding. After translation, expansive resources increased to 82%, while contractive

markers decreased to 18%. This suggested that feedback in Arabic shifted to be perceived as more dialogic, less authoritative, and more open to interpretation. The results indicated that the Arabic translations produced by ChatGPT may present feedback in a more dialogically open manner compared to the original English versions, with a notable decrease in authoritative tone. Tajvidi and Arjani (2017) noted similar shifts in engagement markers for translated texts leading to increased dialogic space and decreased authorial alignment. However, Cao and Zhong (2023) raised concerns about decreased contractiveness. They noted that a less directive and less personalized feedback may not align with EFL learners' needs for individualized support. Regarding graduation, English feedback highly relied on force (upscaling and downscaling) with no use of focus resources. Arabic translations introduce some focus-sharpen markers (6%) of evaluative meaning with a slight decrease in instances of upscale. This suggests that the translation process led to more subtle evaluative scaling and a decrease in upscaling the force of evaluative meaning. This is consistent with Al-Shunnag's (2014) findings, who reported that translations may impact the persuasive power of feedback by decreasing the frequency of intensification resources.

Hyland's engagement model analysis revealed that reader engagement increased in translation, but the balance altered. In English feedback, the frequency of reader mentions (72%) enhanced rapport and signaled individualized attention, as reported by teacher feedback studies (Zou et al., 2025; Carless, 2022). However, Arabic translations featured more directives (43%) leading to changes in the balance between directly addressing the learners and instructing them. This signals a tendency toward providing explicit instruction and guidance to learners in the Arabic translation of feedback. While increased directives may offer learners clear actionable steps, the decreased frequency of reader mentions (57%) may weaken the supportive and relational tone. This is parallel to findings reported by Abbamonte and Cavaliere (2006) and Rosa (2013) who cautioned that increased shifts to directives may negatively affect learner motivation. No new engagement strategies (questions, appeals to shared knowledge, asides) emerged after translation. This mirrors Steiss et al.'s (2024) concern that AI-generated feedback lacks the engaging diversity of teacher input.

The second research objective was to identify the challenges and patterns that emerge in maintaining authorial tone and learner positioning in AI translated feedback. Maintaining authorial tone and effective learner positioning in ChatGPT-translated feedback from English to Arabic, presents a few challenges. One is the increased reliance on appreciation (object-focused evaluation) and the decline in judgment and affect which may be interpreted by learners as impersonal. After translation, the feedback becomes more descriptive about processes and less evaluative of learners' performance, a pattern documented in several translation studies cited in the literature (Li et al., 2025; Tajvidi & Arjani, 2017; Munday, 2009). Another emerging pattern is the increase in negative evaluative language in Arabic outputs from 33% to 39%. This pattern may be challenging as the feedback may feel less supportive and risk learner discouragement, especially in contexts where positive reinforcement is highly recommended (Rad et al., 2024; Sun, 2021; Zou et al., 2025). Additionally, the absence of affective and empathetic tone in translation might lead to feedback sounding more detached and less warm. This aligns with findings by Qian (2017) and Abbamonte and Cavaliere (2006), suggesting that emotions are difficult to preserve in translation, resulting in a more detached authorial voice.

In terms of learner positioning, in Arabic feedback the use of directives rises from 28% to 43%, while reader mentions decrease from 72% to 57%. This demonstrates a clear shift in learner positioning in the feedback dialogue. The learners are repositioned as passive recipients of directive advice, rather than dialogically involving them in the process. This emerging pattern contradicts effective EFL feedback practices (Hyland & Jiang, 2016; Carless, 2022). A further emerging pattern is the expansion of dialogic openness in the translated feedback. In the Arabic set, expansive engagement is increased from 62% to 82% and contractiveness is decreased from 38% in English to 18% in Arabic which theoretically promotes learner agency and critical thinking but may also lead to ambiguity and loss of pedagogical clarity (Tajvidi & Arjani, 2017). Additionally, it may risk the loss of the authoritative guidance that learners with an intermediate proficiency level may need or require. Such trade-offs between learner agency, clarity, motivation, and rapport mirror concerns raised in the literature on both the translation of evaluative meaning (Tajvidi & Arjani, 2017; Munday, 2009) and AI-assisted feedback (Lin & Crosthwaite, 2024; Steiss et al., 2024).

In summary, ChatGPT's translation from English to Arabic alters both the distribution and function of appraisal and engagement resources. The emerging patterns show that ChatGPT-mediated translation tends to decrease authorial voice, interpersonal depth, relational warmth, and learner engagement. Although the feedback produced is more expansive and dialogically open, it may flatten interpersonal meaning and challenge the creation of personalized supportive effective feedback for EFL learners.

VI. Conclusion

The findings of this study offer important implications for EFL translation and writing instruction as AI-generated feedback becomes more common in language education. The results suggest that ChatGPT's translation of feedback from English to Arabic results in feedback that is less individualized. This shift may affect the quality of learner engagement, as empathy and individualized encouragement characteristic of effective teacher feedback decrease in machine translation. Such emerging patterns and changes should prompt educators to approach AI translations with caution in EFL feedback contexts, especially when it is intended to foster rapport and guide learners' affective responses. Educators can address these challenges by adopting a hybrid approach to feedback and combining AI-generated translations with teacher review to preserve supportive tone, encouragement, and personalized engagement. In addition, professional development and

EFL teacher and translator training programs should focus on continuous evaluations of LLM outputs. In addition to AI-generated or translated feedback, teachers' own context-sensitive comments remain crucial.

Despite this study's insights, it also presents some limitations. The sample size was relatively small and focused exclusively on B1-level Arabic-speaking EFL learners. This limits the generalizability of the findings to other L1 backgrounds and proficiency levels. Additionally, the exclusive use of ChatGPT-4 for translation may not reflect the full range of possible outcomes from other LLMs. Also, the study did not address the learners' perspectives of the translated feedback, which is critical for evaluating the real impact of these shifts on EFL learners.

In conclusion, while ChatGPT shows significant potential for increasing the efficiency of feedback in EFL writing instruction, educators must ensure that the interpersonal qualities essential for learner engagement are not lost in translation. Therefore, to achieve the full pedagogical benefits of AI-assisted feedback in EFL classrooms, human oversight remains essential.

REFERENCES

[1] Abbamonte, L., & Cavaliere, F. (2006). Lost in translation: The Italian rendering of UNICEF "The State of the World's Children 2004" report. In M. Gotti & S. Šarčević (Eds.), *Insights into specialized translation* (pp. 235-258). Berg: Peter Lang.

[2] Alkhawaja, L. (2024). Unveiling the New Frontier: ChatGPT-3 Powered Translation for Arabic-English Language Pairs. *Theory and Practice in Language Studies*, *14*(2), 347-357. https://doi.org/10.17507/tpls.1402.05

[3] Alshehri, M. (2025). Integrating Automated Writing Evaluation into Saudi EFL Students' Writing Practice. *Arab World English Journal*, *16*(1), 70-87. https://doi.org/10.24093/awej/vol16no1.5

[4] Alshehri, A., Alshehri, E., & Aljamal, N. (2025). *Appraisal Theory Tagger* [Computer software]. S. A. F. I. Property. Retrieved April 18, 2025, from https://appraisal-tagger.netlify.app/

[5] Al-Shunnag, M. (2014). *Stance in political discourse: Arabic translations of American newspaper opinion articles on the 'Arab Spring'* [Unpublished doctoral dissertation, University of Salford].

[6] Alsofyani, A. H., & Barzanji, A. M. (2024). The Effects of ChatGPT-Generated Feedback on Saudi EFL Learners' Writing Skills and Perception at the Tertiary Level: A Mixed-Methods Study. *Journal of Educational Computing Research*, *63*(2), 431-463. https://doi.org/10.1177/07356331241307297

[7] Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, *6*, 100226. https://doi.org/10.1016/j.caeai.2024.100226

[8] Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00455-4

[9] Cao, S., & Zhong, L. (2023). *Exploring the effectiveness of ChatGPT-based feedback compared with teacher feedback and self-feedback: Evidence from Chinese to English translation*. arXiv.org. https://doi.org/10.48550/ARXIV.2309.01645

[10] Carless, D. (2022). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, *23*(2), 143–153.

[11] Dai, W., Tsai, Y.-S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, *7*, 100299. https://doi.org/10.1016/j.caeai.2024.100299

[12] Hyland, K. (2006). Representing readers in writing: Student and expert practices. *Linguistics and Education*, *16*, 363–377.

[13] Hyland, K., & Jiang, F. (2016). "We must conclude that…": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, *24*, 29–42.

[14] Hwang, S. (2025). Unpacking the Impact of Writing Feedback Perception on Self-Regulated Writing Ability: The Role of Writing Self-Efficacy and Self-Regulated Learning Strategies. *Behavioral Sciences*, *15*(2), 100. https://doi.org/10.3390/bs15020100

[15] Khondaker, M. T. I., Waheed, A., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). *GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP.* arXiv.org. https://doi.org/10.48550/ARXIV.2305.14976

[16] Li, X., Meng, Y., & Jing, Y. (2025). An Appraisal-based Study on Divergence of Evaluative Meanings in Student Translations. *WORD*, *71*(2), 97-113. https://doi.org/10.1080/00437956.2025.2499341

[17] Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, *127*, 103529. https://doi.org/10.1016/j.system.2024.103529

[18] Martin, J., & White, P. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

[19] Mohsen, M. (2024). Artificial Intelligence in Academic Translation: A Comparative Study of Large Language Models and Google Translate. *Psycholinguistics*, *35*(2), 134-156. https://doi.org/10.31470/2309-1797-2024-35-2-134-156

[20] Munday, J. (2009). The concept of the interpersonal in translation. *SYNAPS*, *23*, 1527. Retrieved June 21, 2025, from http://hdl.handle.net/11250/2406113

[21] Qian, H. (2017). Investigating "Unfaithful" Translations via the Appraisal Theory: A Case Study of Public Notices. *Arab World English Journal for Translation & Literary Studies*, *1*(1). http://doi.org/10.24093/awejtls/vol1no1.13

[22] Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, *32*(9), 5020–5040.

[23] Rosa, A. A. (2013). The power of voice in translated fiction: Or, following a linguistic track in translation studies. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartłomiejczyk (Eds.), *Tracks and treks in translation studies*. Amsterdam: John Benjamins Publishing Company.

[24] Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

[25] Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, *91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894
[26] Sun, Y. (2021). The effect of teacher caring behavior and teacher praise on students' engagement in EFL classrooms. *Frontiers in psychology*, *12*, 746871. https://doi.org/10.3389/fpsyg.2021.746871
[27] Tajvidi, G.-R., & Arjani, S. H. (2017). Appraisal Theory in Translation Studies: An Introduction and Review of Studies of Evaluation in Translation. *Journal of Research in Applied Linguistics*, *8*(2), 3-30.
[28] Zhang, M. (2013). Stance and mediation in transediting news headlines as paratexts. *Perspectives*, *21*(3), 396-411.
[29] Zou, S., Guo, K., Wang, J., & Liu, Y. (2024). Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: a comparison study. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2024.2447279

Asma Alshehri is an assistant professor of applied linguistics. Currently, Dr. Asma is the vice chair of the English Language Skills department at the Common First Year Deanship at King Saud University, Riyadh, Saudi Arabia. Her research interests include academic writing, systemic functional linguistics, discourse analysis, and assessment.