

Subtitling Emotion Words Between English and Arabic: A Corpus-Based Study of Collocational Patterns

Sara A. AlShubaily

Department of English, College of Language Sciences, King Saud University, Riyadh, Saudi Arabia

Abstract—This study compares verb collocations with emotion words related to *happiness* and *anger* in English and Arabic subtitles using Sketch Engine's corpus and Word Sketch functionality. The research employs a mixed-methods approach, combining quantitative analysis of LogDice statistics and qualitative examination of subtitle samples. The study aims to identify differences in verb collocations between the two languages, explore patterns and translation strategies in subtitling, and evaluate Sketch Engine's effectiveness in facilitating statistical analysis for translation research. The study reveals a prevalent use of literal translation in subtitling, which often results in unnatural expressions in the target language, mainly when dealing with emotion-related collocations that differ between English and Arabic. The research underscores the significance of corpus-based methodologies in translation studies and demonstrates Sketch Engine's efficacy in identifying linguistic patterns and collocations across different contexts. This study contributes valuable insights for enhancing subtitle translation efficiency and cultural appropriateness while deepening the understanding of linguistic and cultural nuances in subtitling emotional expressions.

Index Terms—subtitling, collocations, corpus-based methods, audiovisual translation

I. INTRODUCTION

The translation of emotion words and their collocations presents a significant challenge in audiovisual translation, particularly in subtitling, where linguistic precision and cultural sensitivity are crucial. Emotions are deeply embedded in language, and how they are expressed through collocations, such as verb-emotion word pairings, can vary considerably between languages. English and Arabic, for instance, exhibit distinct lexical and grammatical structures that influence how emotions are conveyed, making direct translation difficult. Subtitlers must navigate these differences while maintaining the original emotional impact and ensuring that subtitles remain concise and readable. Given the constraints of subtitle space and timing, understanding how verb-emotion word collocations are adapted across languages is essential for producing accurate and natural subtitles.

This research seeks to conduct a comparative analysis of emotion words and verb collocations in source and target languages, explicitly focusing on their application in subtitling. Emotion words are lexical items that explicitly convey emotional states, feelings, or affective responses, such as *happiness*, *anger*, *fear*, and *sadness*. These words can function as nouns (e.g., *joy*, *rage*), adjectives (e.g., *happy*, *furious*), or verbs (e.g., *love*, *hate*), and cultural and linguistic contexts often influence their meanings. In translation studies, emotion words pose challenges due to potential shifts in intensity, connotation, and collocational patterns between languages (Bednarek, 2008). Subtitling plays a crucial role in audiovisual translation, and understanding the nuances of verb-emotion word collocations can significantly improve the quality and accuracy of subtitle translations. The present study utilized Sketch Engine's Word Sketch functionality to extract and analyze collocations from a parallel corpus of English and Arabic subtitles, mainly focusing on the emotions of *happiness* and *anger*. By examining these collocations in both languages, the researcher aims to identify patterns, differences, and translation strategies that may effectively convey emotional content. This research intends to provide valuable insights that can contribute to the enhancement of subtitle translation, improving both its efficiency and cultural appropriateness while offering a better understanding of the linguistic and cultural nuances involved in subtitling emotional expressions.

The data selection for this study, which analyzes subtitling emotional word collocations between English and Arabic, was carefully designed to ensure a comprehensive examination of emotional expression across languages. A parallel corpus of English-Arabic subtitles from popular TV shows and movies was chosen, as it provides a diverse range of emotion-related contexts and colloquial language use. High-frequency emotion words, identified using Sketch Engine's Word Sketch function, were selected for detailed analysis, focusing on commonly used emotional expressions to provide a solid foundation for exploring translation patterns. Subtitles from various genres, such as drama and history, were included to capture a broad spectrum of emotional contexts and intensities, ensuring the inclusion of both subtle and intense emotional expressions.

For this study, the selection of two emotion words—*happiness* for positive emotions and *anger* for negative emotions—was driven by several considerations. Existing research on these emotions in linguistics and psychology provides a strong theoretical foundation for the study (Bednarek, 2008; Al-Kharabsheh & Yassin, 2017; Thawabteh, 2017; Bogucki, 2020),

allowing for comparisons with established findings and contributing to a deeper understanding of subtitling strategies for translating emotional language from English to Arabic. Based on these considerations, this study aims to answer the following research questions: (1) How do verb collocations with emotion words related to *happiness* and *anger* differ between English and Arabic, as identified through LogDice statistics? (2) What patterns and translation strategies are employed in subtitling verb collocations with *happiness* and *anger*, based on the qualitative analysis of subtitle samples from the corpus? (3) How effectively is Sketch Engine facilitating statistical analysis for translation research, particularly in identifying linguistic patterns and collocations across different contexts?

II. LITERATURE REVIEW

A. Collocations

Collocations are essential in language and translation, or frequently occurring word combinations with specific meanings. They enrich expression and enhance linguistic fluency, making them crucial for accurate translation, especially between structurally and culturally distinct languages like English and Arabic (Ibrahim, 2019). Since collocations often carry idiomatic meanings, direct translations can result in unnatural or incorrect phrasing. Understanding these patterns helps translators preserve the text's stylistic and semantic integrity, ensuring effective communication. This is particularly relevant in subtitling, where linguistic accuracy must be balanced with readability and time constraints.

Given the role of collocations in shaping meaning and fluency, their accurate translation is particularly crucial in subtitling, where naturalness and readability significantly impact viewer comprehension and experience (Thawabteh, 2017). Research on subtitling between English and Arabic is essential due to the linguistic and cultural differences between the two languages. Subtitling requires accurate translation and preservation of meaning, style, and emotional impact within space and time constraints (Gamal, 2014). These challenges are further amplified by structural differences, idiomatic expressions, and culture-specific references, which may not have direct equivalents. Effective subtitling ensures that audiovisual content remains accessible and engaging for diverse audiences.

B. Collocation Translation Between English and Arabic

Shraideh and Mahadin (2015) studied the challenges and strategies in translating English-Arabic collocations, focusing on students at the University of Jordan and Yarmouk University. Their research revealed no significant difference in strategies between the two universities, although elaboration and explication were underused. Yarmouk University students used collocations more frequently, while MA students encountered fewer. These findings stress the importance of accurately translating collocations, especially in subtitling, to preserve meaning and cultural nuances. Strategies like elaboration and explication help convey cultural relevance and context effectively.

Obeidat and Mahadi (2019) explored the translation of Arabic religious and cultural collocations into English, using domestication and foreignization strategies. The study found domestication was more common, as translators often adapted collocations to the target culture, while foreignization retained original cultural meanings. This research highlights the challenge of preserving cultural nuances when translating Arabic collocations into English, emphasizing the need for balanced cultural sensitivity and meaning, crucial in subtitling where linguistic and cultural fidelity is key.

Alsaadi and Rabab'ah (2022) examined Arabic-to-English subtitling strategies for collocations during the World Government Summit, identifying challenges like cultural specificity and idiomatic expressions. Strategies like equivalence, calque, and paraphrasing were used, highlighting the importance of cultural awareness. The study stressed the need for further research and curriculum improvements in translation training.

C. Emotion Translation Between English and Arabic

Kayyal and Russell (2012) examined the equivalence of emotion words between English and Arabic, testing 12 English emotion labels and their Arabic counterparts. They found that only one label (happy) passed both equivalence tests, revealing that language differences were more significant than cultural ones. Their study challenges the assumption of equivalence in standard translation methods and underscores the need for nuanced approaches to ensure accurate emotional expression in subtitling.

Salameh et al. (2015) studied Arabic sentiment lexicons and the accuracy of translating emotion-related words between English and Arabic. They found that 88.3% of the words retained their original sentiment, while 11.7% had errors. Errors were categorized as mistranslations, translations into related but distinct words, and differences in usage. Dominant senses and cultural differences in sentiment perception influenced the latter. These findings highlight the complexities of translating emotions, especially in subtitling, where linguistic and cultural sensitivity are essential for preserving the intended emotional impact.

Mansour (2016) highlighted the importance of corpora in translation studies, particularly for analyzing emotional adjectives in English and Arabic. Large corpora, like the British National Corpus and Al-Hayat, provide authentic data that enables a deeper understanding of language usage and social significance. Using corpora helps identify regularities in language behavior, crucial for accurately translating emotional expressions.

D. Subtitling Challenges

Several studies have explored challenges in Arabic-English subtitling, particularly in handling euphemisms, taboo words, and colloquialisms. Thawabteh (2012) examined the translation of euphemistic and dysphemistic expressions in Arabic-English subtitling, highlighting the difficulties these expressions pose for the target audience and the strategies used to address them. Sahari (2023) analyzed euphemization strategies for private body part-related taboo words in Arabic subtitles of Hollywood films, finding that the vast majority were either euphemized or omitted due to cultural sensitivities. In a subsequent study, Sahari (2024) investigated the subtitling strategies used for the f-word in Hollywood films, revealing that translation choices were influenced by both linguistic constraints in Modern Standard Arabic and the multimodal nature of audiovisual content. Al-Kharabsheh and Yassin (2017) explored the subtitling of Arabic colloquialisms in *The Dupes*, identifying issues such as mis subtitling, omission, and meaning loss. Their study emphasized the importance of cultural context and suggested alternative strategies to improve subtitle accuracy. These studies collectively highlight the complexities of Arabic-English subtitling and the necessity of culturally and contextually aware translation strategies.

Bogucki (2020) applied Relevance Theory to analyze decision-making in subtitling films and series, arguing that the technical, linguistic, and translational constraints of subtitling lead to condensed target texts. By examining English-Polish Netflix subtitles, he demonstrated how Relevance Theory helps subtitlers select contextually relevant elements to enhance coherence and audience understanding, thereby offering a theoretical framework for subtitling within translation studies. Similarly, Rayyash and Haider (2022) investigated translation strategies for English swear words in Netflix subtitles, revealing that omission was the most frequent strategy, particularly in drama, and highlighting how such choices affect meaning and character portrayal. In a related vein, Alowaidi (2024) adopted a corpus-based approach to explore subtitling strategies for culture-specific items in films, finding that while English films favored foreignization, Arabic films often employed domestication, which tended to reduce the visibility of local cultural elements in the subtitles. This research resonates with Saudi Arabia's Vision 2030, underlining the vital role of translation in preserving and promoting cultural identity.

E. Computational Approaches to Translation Studies

Computational approaches to translation studies have increasingly utilized corpus-based methodologies to analyze linguistic patterns, particularly collocations. Dayter (2019) examined collocations in non-interpreted and simultaneously interpreted English, comparing window-based and syntax-based methods while addressing the challenges of syntactic parsing in spoken corpora. Using the SIREN corpus, the study found that simultaneous interpreting enhances oral features while reducing literate ones, supporting the equalizing universal hypothesis. The research highlighted differences in collocability between interpreted and non-interpreted English, emphasizing the need for further qualitative analysis. The SIREN corpus, comprising comparable and parallel texts of spontaneous speeches and their interpretations between Russian and English, provided insights into how interpretation affects language patterns, with comparisons to the ukWaC corpus helping measure shifts in orality and literateness. Additionally, statistical measures such as Mutual Information (MI), T-score, and Log-likelihood have been applied in collocation research to address data sparseness (Bernardini, 2022). Labrador (2024) further contributed to computational translation studies with a child-oriented corpus-based study on semantically related verbs using Sketch Engine, demonstrating the tool's application in analyzing verb collocations. These studies underscore the significance of corpus-based methodologies in translation research, highlighting theoretical implications and practical applications in computational linguistics.

Building on the existing literature, this study addresses a gap in research on the subtitling of emotion words between English and Arabic, particularly regarding collocational patterns. While previous studies have explored general translation strategies and challenges in audiovisual translation, limited attention has been given to how emotion words and their verb collocations are adapted in subtitling between linguistically and culturally distant languages. This study employs quantitative and qualitative analyses to examine these patterns, leveraging corpus technologies and Sketch Engine to extract statistical insights. The research aims to provide empirical findings that improve subtitling quality by systematically analyzing collocational shifts and their impact on meaning transfer. The study's focus on English-Arabic subtitling offers valuable insights for audiovisual translation in contexts where linguistic and cultural disparities present unique challenges, ultimately enhancing translation strategies and refining the accuracy and acceptability of subtitles in cross-cultural settings.

III. METHOD

This study employs a mixed-methods approach, integrating quantitative and qualitative analyses to examine verb collocations with emotion words in English and Arabic subtitles. The methodology consists of two main stages: corpus-based statistical analysis and qualitative examination of subtitle samples.

A. The Data

This study investigates the collocations of emotion words in English and Arabic subtitles, focusing on two key emotions: *anger* and *happiness*. The research utilizes Sketch Engine's subtitle corpus to analyze the collocational patterns of these emotion words, selected due to their fundamental role in human experience and their prevalence in dramatic and historical narratives. These emotions are frequently portrayed in movies and television series, making them ideal candidates for

analysis in the selected corpus. Their universal nature allows for meaningful cross-linguistic comparison, as *happiness* and *anger* transcend cultural boundaries. Additionally, both emotions exhibit linguistic richness, encompassing a wide range of associated vocabulary and expressions, providing a robust collocational analysis dataset. They also play a crucial role in narrative structure, significantly influencing plot development and character interactions. Furthermore, these emotion words often appear in metaphorical expressions, offering insights into the conceptual frameworks of emotions in both languages. By analyzing these aspects, the study provides a quantitative foundation for understanding collocational behavior and a qualitative perspective on how these collocations are treated in subtitling, ultimately enabling a comprehensive analysis of how emotion words are conveyed across languages in subtitle translation.

B. Procedures

This study uses a mixed-methods approach to analyze verb collocations with emotion words in English-Arabic subtitling, focusing on *happiness* and *anger*. The first stage involved quantitative analysis using Sketch Engine's Word Sketch functionality to extract verb collocations with emotion words (سعادة and غضب) from an English parallel subtitles corpus. The extracted data was analyzed using LogDice scores to measure collocational strength, and comparisons between English and Arabic were made to identify similarities and differences in emotional expression. Statistical tests validated the reliability of the collocation patterns.

The second stage adopted a qualitative approach, building on the quantitative findings. A sample of 50 subtitles containing the identified verb collocations was extracted, and two translators rated their acceptability. An interrater test measured agreement between the raters to ensure consistency. The study then analyzed the translation of these collocations, exploring subtitling strategies and any shifts in meaning, intensity, or structure. The qualitative analysis highlighted recurring patterns in translation and provided insights into common strategies used in subtitle translation. Combining both methods, the study offers a comprehensive understanding of how verb collocations with emotion words are translated in subtitles, contributing to the broader field of audiovisual translation studies.

IV. FINDINGS

A. Quantitative Analysis of Verb Collocations With Emotion Words in Subtitling

The analysis began with comparing LogDice scores between English and Arabic verbs to understand how these languages differ in their verbal expression of emotions and how this is reflected in subtitling. LogDice is a statistical measure used to evaluate the collocation strength between two words, with higher values indicating stronger associations.

TABLE 1
LOGDICE STATISTICS FOR HAPPINESS

<i>Happiness</i> frequency in the parallel subtitles corpus			
English verb	LogDice	Arabic Verb	LogDice
Pursue	4.79	تكنم	6.67
Define	4.42	تدوم	5.68
Deserve	4.28	يعيش	5.62
Sacrifice	4.01	يشعر	5.56
Buy	3.70	تندفق	5.28
Find	3.30	يجد	1.18

As shown in the table above, in English, the verb 'pursue' shows the highest LogDice value (4.79), reflecting a strong and common phrase like 'pursuing happiness.' Other strong collocations include 'define' (4.42), 'deserve' (4.28), and 'sacrifice' (4.01), which suggest frequent references to 'defining, deserving,' and the sacrifices associated with *happiness*. The verb 'buy' (3.70) indicates a moderate, yet still notable, association, often appearing in expressions such as 'buy happiness.' 'Find' (3.30) has the weakest LogDice value, suggesting that while the phrase 'finding happiness' is common, it is not as strongly associated as other verbs.

The table presents LogDice values for verb collocations with the word سعادة in Arabic subtitles, highlighting the strength of each verb's association with happiness. The verb 'تكنم' (6.67) has the strongest relationship, suggesting happiness resides in specific conditions. 'تدوم' (5.68) reflects lasting happiness, while 'يعيش' (5.62) indicates experiencing happiness. 'يشعر' (5.56) shows that feeling happiness is common, and 'تندفق' (5.28) suggests happiness flows. The weakest association is with 'يجد' (1.18), indicating that 'finding happiness' is less frequently expressed in Arabic subtitles than other emotional expressions.

These findings highlight significant linguistic and cultural implications regarding how *happiness* is conceptualized in English and in Arabic. In English, the verbs with higher LogDice values (e.g., 'pursue,' 'define,' 'deserve') reflect active, goal-oriented, or defining aspects of happiness. In English, *happiness* is often treated as something that can be pursued, defined, or deserved, reflecting individual agency and conceptualization of happiness.

In Arabic, 'تكنم' (lie in) suggests that *happiness* is seen as inherent to specific conditions, while 'تدوم' (to last) emphasizes enduring happiness. The frequent use of 'يعيش' (to live) highlights happiness as an active, ongoing experience.

‘تندفق’ (to flow) indicates a view of happiness as organic and abundant. In contrast, the weak association of ‘يجد’ (to find) implies that Arabic speakers less often see *happiness* as something to be sought or found, unlike the English concept of ‘finding happiness.’ These patterns reveal cultural attitudes toward emotional well-being and fulfillment in Arabic discourse.

TABLE 2
LOGDICE STATISTICS FOR ANGER

Anger frequency in the parallel subtitles corpus			
English verb	LogDice	Arabic Verb	LogDice
Misdirect	7.35	صب / يصب	9.20
Vent	6.95	أثار / يثير	7.99
Express	6.09	ينفس / نفس	7.37
Disperse	5.96	يفرغ	7.01
Repress	5.68	أتفهم	6.53
Direct	5.53	يتحكم	6.13

Table 2 presents LogDice values for verb collocations with *anger* in English subtitles, showing the strength of each verb's association with the emotion. The verb ‘misdirect’ (7.35) has the strongest tie, suggesting anger is often redirected. ‘vent’ (6.95) follows, indicating the common expression ‘venting anger,’ or releasing it. ‘express’ (6.09) reflects the frequent phrase ‘expressing anger,’ meaning to outwardly show it. ‘disperse’ (5.96) suggests anger can lose intensity, while ‘repress’ (5.68) indicates suppressing anger. ‘direct’ (5.53) implies anger is often aimed at someone or something.

The table shows LogDice values for verb collocations with ‘غضب’ in Arabic subtitles, highlighting the strength of each verb's connection to anger. The verb ‘صب’ (9.20) has the highest value, suggesting that ‘pouring’ or ‘spilling’ anger is a powerful expression. ‘أثار’ (7.99) follows, indicating that anger is often provoked or stirred up. ‘ينفس’ (7.37) reflects the act of ‘venting’ anger, similar to English usage. ‘يفرغ’ (7.01) suggests anger can be ‘emptied’ or dissipated. ‘أتفهم’ (6.53) indicates ‘understanding’ anger, while ‘يتحكم’ (6.13) reflects the act of ‘controlling’ anger.

English verb collocations with anger, such as ‘misdirect,’ ‘vent,’ ‘express,’ and ‘repress,’ emphasize managing, releasing, or misdirecting anger. These verbs suggest that anger is often viewed as something to control or direct in a specific way. In contrast, Arabic verb collocations like ‘صب’ (pouring), ‘أثار’ (provoking), and ‘ينفس’ (venting) emphasize the release, provocation, and full expression of anger, indicating a cultural tendency to view anger as an emotion to be strongly expressed or triggered. While Arabic also recognizes controlling or emptying anger through verbs like ‘يفرغ’ (emptying) and ‘يتحكم’ (controlling), the focus remains on the intensity and expression of anger. The differences between English and Arabic reveal contrasting cultural perspectives: English prioritizes regulating and managing *anger*, while Arabic emphasizes its direct, intense expression, providing insights into the emotional norms of both cultures.

(a). Normality Test

Table 3 below displays the normality of data distribution, which was assessed for *happiness* and *anger* frequency across English and Arabic verbs using both Kolmogorov-Smirnov (K-S) and Shapiro-Wilk tests. Both tests indicated deviation from normality for *happiness* frequency in English verbs (K-S: statistic = 0.144, df = 6, p = 0.029; Shapiro-Wilk: statistic = 0.989, df = 6, p = 0.047). The Arabic verb frequency showed a more substantial departure from normal distribution (K-S: statistic = 0.391, df = 6, p = 0.005; Shapiro-Wilk: statistic = 0.715, df = 6, p = 0.009). Similarly, *anger* frequency measurements demonstrated non-normal distributions in both languages. For English verbs, the tests revealed significant deviations (K-S: statistic = 0.259, df = 6, p = 0.009; Shapiro-Wilk: statistic = 0.892, df = 6, p = 0.039), while Arabic verbs also showed non-normal distribution patterns (K-S: statistic = 0.167, df = 6, p = 0.037; Shapiro-Wilk: statistic = 0.956, df = 6, p = 0.025). Given that all p-values were less than 0.05, these results indicate that the data did not have normal distribution across both languages, suggesting the need for non-parametric statistical approaches in data analysis as the Mann-Whitney U test, the alternative test for independent sample T-Test (Table 4, Figure 1 and Figure 2).

TABLE 3
NORMALITY TEST RESULTS

LogDice	Language	Kolmogorov-Smirnovb			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
<i>Happiness</i> frequency	English verb	.144	6	.029*	.989	6	.047*
	Arabic verb	.391	6	.005**	.715	6	.009**
<i>Anger</i> frequency	English verb	.259	6	.009**	.892	6	.039*
	Arabic verb	.167	6	.037*	.956	6	.025*

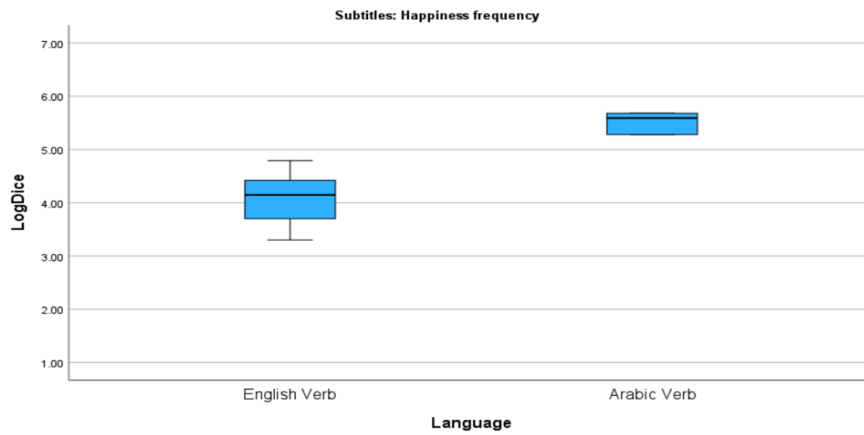


Figure 1. Happiness Frequency

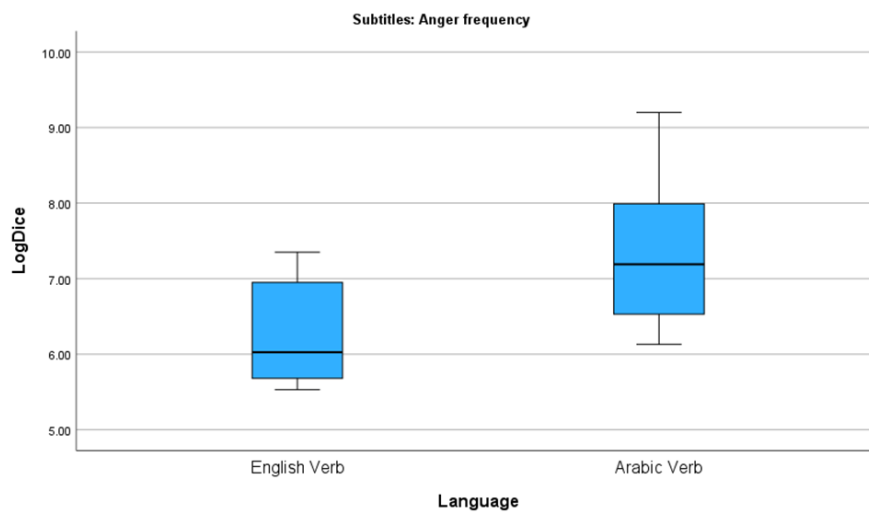


Figure 2. Anger Frequency

(b). Comparison Tests

The descriptive analysis of *happiness* frequency LogDice scores revealed distinct patterns between English and Arabic verbs. English verbs demonstrated a mean LogDice score of (4.08 ± 0.53) , while Arabic verbs showed a higher mean score of (5.00 ± 1.93) . This indicates more significant variability in the Arabic verb expressions of *happiness* than English verbs (Table 4).

TABLE 4
DESCRIPTIVE ANALYSIS OF HAPPINESS FREQUENCY

	Language	N	Mean	Std. Deviation	Std. Error Mean
LogDice	English Verb	6	4.0833	.53242	.21736
	Arabic Verb	6	4.9983	1.92977	.78783

The non-parametric analysis using the Mann-Whitney U test showed meaningful differences in rank distributions between the languages. English verbs had a lower mean rank of 4.50 (sum of ranks = 27.00) compared to Arabic verbs with a mean rank of 8.50 (sum of ranks = 51.00), suggesting systematically higher LogDice scores for Arabic *happiness* expressions (Table 5).

TABLE 5
RANK DISTRIBUTION BETWEEN LANGUAGES FOR HAPPINESS

	Language	N	Mean Rank	Sum of Ranks
LogDice	English Verb	6	4.50	27.00
	Arabic Verb	6	8.50	51.00
	Total	12		

The inferential statistics confirmed a statistically significant difference between the two languages (Mann-Whitney U = 6.000, Z = -1.922, p = 0.047). These results suggest that Arabic verbs have significantly higher LogDice *happiness* frequency scores than English verbs (Table 6).

TABLE 6
RESULT OF THE NON-PARAMETRIC ANALYSIS OF ANGER FREQUENCY

	LogDice
Mann-Whitney U	6.000
Z	-1.922
Asymp. Sig. (2-tailed)	.047

The descriptive analysis of *Anger* frequency LogDice scores revealed distinct patterns between English and Arabic verbs. English verbs demonstrated a mean LogDice score of (6.26 ± 0.73) , while Arabic verbs showed a higher mean score of (7.37 ± 1.11) . This indicates more significant variability in the Arabic verb expressions of anger than English verbs (Table 7).

TABLE 7
DESCRIPTIVE ANALYSIS OF ANGER FREQUENCY

	Language	N	Mean	Std. Deviation	Std. Error Mean
LogDice	English Verb	6	6.2600	.72834	.29734
	Arabic Verb	6	7.3717	1.10509	.45115

The non-parametric analysis using the Mann-Whitney U test showed meaningful differences in rank distributions between the languages. English verbs had a lower mean rank of 4.33 (sum of ranks = 26.00) compared to Arabic verbs with a mean rank of 8.67 (sum of ranks = 52.00), suggesting systematically higher LogDice scores for Arabic *anger* expressions (Table 8).

TABLE 8
RANK DISTRIBUTION BETWEEN THE LANGUAGES FOR ANGER EXPRESSIONS

	Language	N	Mean Rank	Sum of Ranks
LogDice	English Verb	6	4.33	26.00
	Arabic Verb	6	8.67	52.00
	Total	12		

The inferential statistics confirmed a statistically significant difference between the two languages (Mann-Whitney U = 5.000, Z = -2.082, p = 0.037). These findings demonstrate that Arabic verbs consistently exhibit significantly higher LogDice scores for *anger* frequency than English verbs, suggesting a more varied and intense lexical expression of *anger* in Arabic verbal forms (Table 9, Figure 3).

TABLE 9
RESULT OF THE NON-PARAMETRIC ANALYSIS, THE MANN-WHITNEY U TEST FOR ANGER EXPRESSIONS

	LogDice
Mann-Whitney U	5.000
Z	-2.082
Asymp. Sig. (2-tailed)	.037

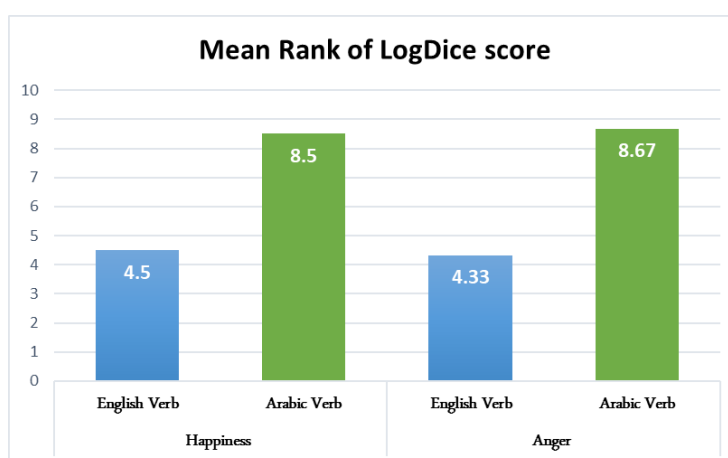


Figure 3. Mean Rank of LogDice Score

In the study's second phase, a qualitative approach was employed to analyze subtitling strategies for translating verb collocations identified in the quantitative findings. To enhance reliability, an interrater statistical test was conducted before data analysis. A sample of subtitles was extracted from the corpus for qualitative assessment, where two translators rated the acceptability of the translations.

(c). *Interrater Reliability Tests*

To evaluate the reliability of translators' assessment of subtitling emotion word verb collocations, the researcher extracted 50 random samples from the subtitles corpus and asked two experts to rate them for acceptability. Inter-rater reliability measures the degree of agreement between two or more raters. A high inter-rater reliability indicates strong agreement, whereas a low value suggests weak agreement. The frequency analysis revealed that, out of 50 total ratings, there were 26 agreements (52%) between the raters, while 24 cases (48%) showed disagreement. This relatively balanced distribution warranted further statistical investigation (Table 10).

TABLE 10
FREQUENCY OF AGREEMENT BETWEEN RATERS

Agreements		n	%
	1	26	52.0
0	24	48.0	
Total	50	100.0	

Pearson's correlation test assessed inter-rater reliability, assuming a continuous rating scale. The correlation strength was classified as follows: 0.20–0.39 (low), 0.40–0.69 (moderate), 0.70–0.89 (high), and 0.90 and above (very high) (Cohen, 1982). As shown in Table 11, the correlation between the two raters was $r = 0.733$, $p < 0.01$, indicating a high level of inter-rater reliability and substantial agreement between the raters.

TABLE 11
PEARSON CORRELATION RESULTS FOR THE INTER-RATER RELIABILITY

		Rater1	Rater2
Rater1	Pearson Correlation	1	.733**
	Sig. (2-tailed)		<.001
	N	50	50
Rater2	Pearson Correlation	.733**	1
	Sig. (2-tailed)	<.001	
	N	50	50

However, Cohen's Kappa coefficient was also calculated to account for chance agreement. The cross-tabulation of ratings showed observed consistency, particularly in extreme ratings. The Kappa coefficient was 0.381 ($p < 0.001$), indicating moderate agreement. These findings highlight the importance of employing multiple methods to assess inter-rater reliability, as different statistical approaches may reveal different aspects of agreement (Table 12).

TABLE 12
MEASURE OF AGREEMENT (KAPPA) FOR INTER-RATER RELIABILITY

		Rater2					Total
		1	2	3	4	5	
Rater1	1	2	0	0	0	0	2
	2	4	5	0	0	0	9
	3	2	3	3	0	0	8
	4	0	2	1	4	1	8
	5	0	4	4	3	12	23
Total		8	14	8	7	13	50
Measure of Agreement (Kappa) = .381, p-value <.001							

These results suggest that while Pearson's correlation indicates a high level of reliability, Cohen's Kappa reveals a more moderate level of agreement. This underscores the necessity of using complementary statistical methods to obtain a more nuanced understanding of inter-rater reliability in analyzing emotion word-verb collocations in English-Arabic subtitles.

While the statistical measures provide insight into the degree of agreement between raters, they do not explain the underlying reasons for agreement or disagreement. The following section presents a qualitative analysis of 10 selected samples from the data used in the inter-rater reliability test to better understand the rated samples. These samples illustrate the subtitling strategies employed and provide insights into the linguistic and contextual factors influencing the subtitlers' decisions.

B. Qualitative Analysis of Verb Collocations With Emotion Words in Subtitling

This section presents a qualitative analysis of verb collocations with *happiness* and *anger* in English-to-Arabic subtitling. A sample of 10 subtitles was extracted from the study's corpus for in-depth qualitative analysis.

Example 1:

Source Text: *It hasn't dulled their anger.*

Target Text: ولكنها لم تخدر غضبهم

In English, the verb 'dull' in this context means to reduce the intensity of anger, a commonly used collocation. However, in the Arabic translation, the verb يُخَدِّر (to numb or anesthetize) is used, which conveys a state of relaxation or sedation, often leading to sleep. In this example, the translation strategy used is literal translation, which has resulted in an unnatural

collocation in Arabic, as *anger* does not typically collocate with numbing. A more idiomatic rendering would be to use *يهدئ غضبهم* (to calm their anger), which better captures the intended meaning in Arabic.

Example 2:

Source Text: *Good, embrace your anger.*

Target Text: جيد، يعتنق الغضب

In English, 'embrace' anger conveys the idea of acknowledging, accepting, and managing the emotion, often in a therapeutic or self-reflective context. However, the Arabic translation uses *يعتق*, which typically means (to adopt or strongly believe in an ideology or intangible concept). In this example, the translation strategy used is literal translation, which has resulted in an unnatural collocation in Arabic, as *anger* is not something one 'adopts' in Arabic. A more accurate translation would involve a verb such as *يقبل* (to accept) or *يحتوي* (to contain), which better reflects the intended meaning in the source text.

Example 3:

Source Text: *Tell your mother you conquered your anger.*

Target Text: قل لأمك أنك تغلبت على غضبك

This example illustrates a successful rendition of the verb collocation. The English phrase 'conquer your anger' conveys overcoming anger through strength and self-control. The Arabic translation accurately captures this meaning using the verb *تغلب*, which denotes (triumph over a challenge or difficulty). This choice effectively preserves the original phrase's semantic and emotional nuances. The translation aligns well with natural Arabic usage, making the subtitle accurate and acceptable in conveying the intended message.

Example 4:

Source text: *Someone who takes his anger out on men.*

Target Text: شخص يأخذ غضبه خارجاً على الرجال

In English, the phrase 'to take one's anger out on someone' refers to expressing frustration or anger by directing it toward another person, sometimes aggressively. However, the Arabic translation uses a literal, word-for-word rendering (*يأخذ غضبه خارجاً*), which is not a natural or meaningful collocation in Arabic. This results in an awkward and unclear expression that fails to convey the intended meaning. A more idiomatic translation would be *ينفجر غضباً* (to explode in anger) or *يثور على الرجال* (to revolt against men), both of which align with natural Arabic usage and accurately reflect the sense of emotional outburst in the source text.

Example 5:

Source Text: *Maybe put to rest some anger and resentment we have for each other*

Target Text: ربما لإراحة بعض الغضب و الاستياء تجاه بعضنا البعض

The phrase 'إراحة الغضب' is a literal translation that does not align with natural Arabic usage. According to the quantitative analysis of the subtitles corpus, the most frequent verbs collocating with *anger* in Arabic include *express* (*يعبر عن*), *release* (*يفرح*), and *control* (*يسيطر على*). These verbs reflect common conceptualizations of anger in Arabic, where it is typically something to be expressed, contained, or released rather than put to rest. A more accurate and idiomatic rendition in this context would be 'نهدئ الغضب' (to calm the anger), as it better conveys the intended meaning while maintaining natural Arabic collocational patterns. This example highlights the importance of selecting verbs that align with established collocational norms in Arabic to ensure both accuracy and acceptability in subtitling.

Example 6:

Source Text: *So, we will swallow our anger.*

Target Text: إننا سوف نبتلع غضبنا

In this example, the translation of 'swallow our anger' results in an unnatural collocation in Arabic. The verb 'نبتلع' (to swallow) is a direct translation from English, but it does not naturally collocate with *anger* in Arabic. According to the quantitative analysis of the subtitles corpus, anger in Arabic is more commonly associated with verbs such as *suppress* (*تكبت*), *control* (*تضبط*), or *restrain* (*تتمالك*). These verbs better reflect how the concept of anger is metaphorically framed in Arabic. A more natural and acceptable translation in this case would be 'إننا سوف نكبت غضبنا' (So, we will suppress our anger), as it aligns with conventional Arabic usage while preserving the original text's intended meaning. This example underscores the necessity of adapting idiomatic expressions rather than relying on literal translation to maintain linguistic fluency and cultural appropriateness in subtitling.

Example 7:

Source Text: *And may nothing ever happen to mar your happiness.*

Target Text: .. شيئاً يفسد سعادتك.

In the example, the translation of 'mar your happiness' results in a subtle yet important shift in meaning due to collocational differences between English and Arabic. The verb 'mar' in English conveys the idea of slightly damaging or spoiling something, often in a way that diminishes its perfection but does not completely ruin it. However, the Arabic translation 'يفسد سعادتك' (to spoil/corrupt your happiness) carries a stronger negative connotation, implying a more severe and permanent form of damage. A more natural and idiomatic translation would be: 'و ألا يحدث شيء يعكر صفو سعادتك' (And may nothing ever happen to disturb the purity of your happiness). This alternative preserves the nuance of the original phrase while aligning with conventional Arabic collocational patterns. This example highlights the importance of

selecting verbs that accurately reflect the intensity and connotation of emotions in translation, ensuring both linguistic and cultural appropriateness in subtitling.

Example 8:

Source Text: *You can't obtain happiness by relying on the likes of red water or a philosopher's stone!*

Target Text: أنت لا تستطيع كسب السعادة إذا كنت تحاول الاعتماد على الماء الأحمر أو حجر الفلسفة

The expression 'obtain happiness' presents a challenge when translated into Arabic, as the verb 'كسب' (to earn/gain) does not naturally collocate with 'happiness' in Arabic. In English, 'obtain' can be used with both tangible and intangible concepts. However, in Arabic, happiness is more commonly associated with verbs like 'تحقيق السعادة' (achieve happiness), 'الوصول إلى السعادة' (reach happiness), or 'إيجاد السعادة' (find happiness). These verbs are more idiomatic and align with conventional Arabic collocational patterns. Therefore, a more natural translation would replace 'كسب السعادة' with one of these alternatives, preserving the intended meaning while ensuring the expression sounds natural in Arabic. This example highlights the importance of selecting verbs that align with established collocational norms in the target language, ensuring both accuracy and fluency in subtitling.

Example 9:

Source Text: *This is happiness plotted against me.*

Target Text: هذا رسم السعادة عبر الزمن

In this example, the translation of 'happiness plotted against me' involves a significant shift in meaning due to differences in metaphorical expression and collocation between English and Arabic. The source text, 'This is happiness plotted against me', uses the verb 'plotted' metaphorically, suggesting that happiness is mapped or arranged in opposition to the speaker. In English, 'plotted' often implies a calculated or intentional arrangement, usually in the context of conflict or schemes, and in this case, it implies that happiness is working against the speaker. However, the Arabic translation, 'هذا رسم السعادة عبر الزمن' (This is the drawing of happiness over time), introduces a very different image. 'رسم' (drawing) and 'عبر الزمن' (over time) shift the focus from opposition to a neutral or even positive depiction of happiness as something evolving with time. This translation misses the adversarial nuance of the original expression. In Arabic, a better way to express the intended meaning would be 'إنها السعادة تبتعد عني' (happiness keeps drifting farther away from me). This alternative preserves the sense of happiness as something working against the speaker's favor. This example illustrates how important it is to select expressions that match the cultural and linguistic conventions of the target language while staying true to the metaphorical meaning in the source language. In this case, the revised translation better conveys the original intent, highlighting the nuances of happiness as a force that either works in opposition or is actively sought.

Example 10:

Source Text: *It's been a long time since this house has seen happiness and love.*

Target Text: لقد مر وقت طويل منذ أن شهد هذا البيت السعادة و الحب.

The translation of 'It's been a long time since this house has seen happiness and love' as 'لقد مر وقت طويل منذ أن شهد هذا البيت السعادة و الحب' is an acceptable and accurate rendering. Both the source and target texts maintain equivalent expressions, with the verb 'شهد' (to witness/see) in Arabic effectively capturing the meaning of the original English verb 'has seen.' The phrase 'long time' is also naturally translated as 'وقت طويل', and the emotional connotation of 'happiness and love' is preserved in both languages. This translation effectively conveys the same sentiment and emotional weight, ensuring clarity and cultural appropriateness in the Arabic rendition.

The analysis of the examples discussed highlights how the literal translation of emotional expressions often leads to unnatural collocations in the target language, which can distort the intended meaning. For instance, using 'إراحة الغضب' (resting anger) or 'نبتلع غضبنا' (swallowing anger) results in awkward expressions that do not align with natural Arabic usage. However, successful translations prioritize conveying the intended meaning over maintaining direct equivalence in collocates. In these cases, expressions like 'نهدي الغضب' (calming anger) or 'نكبت غضبنا' (suppressing anger) are more effective, even if they use different verbs from the original text. The overall meaning and emotional impact truly matter in subtitling, as other elements such as context, tone, and timing shape the audience's understanding. This underscores the importance of capturing the essence of the original message while adapting it to the target language's natural collocational patterns, ensuring both linguistic accuracy and cultural resonance.

V. DISCUSSION AND CONCLUSION

This study examined the subtitling of emotion words between English and Arabic through a corpus-based analysis of collocational patterns using Sketch Engine. Collocations often carry idiomatic meanings that do not always align directly across languages, making literal translations prone to unnatural or misleading phrasing. By analyzing these patterns, translators can better preserve the text's stylistic and semantic integrity, ensuring that emotional expressions are conveyed accurately. Because collocations play a crucial role in shaping meaning and fluency, their accurate translation is especially important in subtitling, where readability and naturalness directly impact viewer comprehension and engagement. As Thawabteh (2017) emphasizes, effective subtitling depends on maintaining linguistic authenticity while adapting content for the target audience. Subtitling requires accurate translation and preservation of meaning, style, and emotional impact within space and time constraints (Gamal, 2014).

The analysis of verb collocations with emotion words related to *happiness* and *anger* in English and Arabic, based on LogDice statistics, highlights distinct cultural conceptualizations of these emotions. In English, *happiness* is often framed

as an active pursuit, with verbs like 'pursue, define, and deserve' reflecting an individualistic and goal-oriented perspective, whereas Arabic collocations do not emphasize pursuit in the same way, suggesting a different cultural framing. Regarding *anger*, English collocations focus on regulating and managing the emotion, as seen in verbs like 'misdirect, vent, express, and repress,' while Arabic collocations emphasize intensity and direct expression, with verbs such as 'pouring' and 'provoking' highlighting their externalization. While both languages include terms for control, English favors moderation, whereas Arabic presents *anger* as a force to be fully expressed, underscoring the role of language in shaping emotional perception across cultures. These findings align with Kayyal and Russell (2012), who revealed that language differences were generally more significant than cultural differences, suggesting that emotion words are culturally specific and reflect varying categorizations of emotional experiences across languages, directly implying translating emotion words in subtitling. Additionally, Mansour (2016) highlights the importance of corpora in translation studies, particularly in analyzing emotional adjectives in English and Arabic. She emphasizes that large corpora reveal language regularities essential for accurately translating emotional expressions and preserving their intended meaning across cultures, further supporting the significance of corpus-based approaches in understanding emotional conceptualization in translation.

Analyzing the subtitle samples reveals that the literal translation of verb collocations related to *happiness* and *anger* often results in unnatural or awkward expressions in the target language. Directly translating collocations without considering their natural usage in Arabic can lead to distortions in meaning, as seen in examples like (resting anger) or (swallowing anger), which do not align with conventional Arabic expressions. While maintaining lexical equivalence, such translations fail to convey the intended emotional nuance effectively. Instead, more natural Arabic renderings, such as (calming anger) or (suppressing anger), better preserve the intended meaning while ensuring linguistic fluency. This demonstrates that effective subtitling goes beyond word-for-word translation, requiring careful adaptation to reflect both the target language's collocational norms and the emotional intent of the original text.

Additionally, the qualitative analysis highlights various translation strategies used to handle verb collocations in subtitling. In some cases, translators opt for modulation, shifting the perspective or emphasis of an expression to align with the target language's conventions. For example, instead of translating 'venting anger' literally, a more natural rendering like 'expressing anger' is used. When prioritizing meaning and emotional impact over strict lexical correspondence, subtitles can maintain authenticity while resonating with the target audience's linguistic and cultural expectations. Studies such as Rayyash and Haider (2022) and Al-Kharabsheh and Yassin (2017) highlight the complexities of Arabic-English subtitling. Those studies emphasize the importance of cultural context and suggest alternative strategies to improve subtitle accuracy, reinforcing the need for a nuanced, contextually aware approach to translating emotion words in subtitling.

The findings of this study indicate that Sketch Engine is an effective tool for facilitating statistical analysis in translation research, particularly in identifying linguistic patterns and collocations across different contexts. By leveraging corpus-based methodologies, Sketch Engine provides valuable statistical data that enhances the analysis of translation phenomena, thereby contributing to theoretical and practical advancements in subtitling studies (Labrador, 2024). These findings align with prior research in computational approaches to translation studies, which have increasingly adopted corpus-based methodologies to examine linguistic patterns, particularly collocations. Previous studies have highlighted significant differences in collectivity between translated and non-translated texts, reinforcing the importance of corpus-based approaches in translation research (Dayter, 2019). Statistical measures such as Mutual Information (MI), T-score, and Log-likelihood have been widely employed (Bernardini, 2022) in collocation research to address data sparseness, contributing to a more refined understanding of translation patterns. Additionally, research has demonstrated the effectiveness of Sketch Engine in computational translation studies (Labrador, 2024), particularly in analyzing verb collocations and large-scale linguistic data. These studies underscore the significance of corpus-based methodologies in translation research, particularly in subtitling.

The findings of this study, in conjunction with previous research, highlight the advantages of employing computational tools like Sketch Engine to investigate translation phenomena systematically. Sketch Engine's ability to provide robust statistical insights not only supports quantitative analysis but also complements qualitative approaches, offering a comprehensive perspective on linguistic patterns in translation. As a result, Sketch Engine proves to be an indispensable resource for researchers examining collocational behavior across different languages and contexts, ultimately contributing to the broader field of computational linguistics and translation studies.

Further research on subtitling between English and Arabic can explore Bogucki's (2020) approach, which applies Relevance Theory to subtitling, guiding translators in selecting the most contextually relevant elements to maintain coherence and enhance audience comprehension. This perspective could provide valuable insights into how subtitlers prioritize meaning while balancing linguistic and cultural constraints. Additionally, future studies can utilize Sketch Engine to investigate other collocational or idiomatic expressions across various genres, further enriching our understanding of translation patterns and challenges in audiovisual contexts. Such research would contribute to developing more effective subtitling strategies, ensuring greater linguistic accuracy and cultural authenticity in cross-language media translation.

REFERENCES

- [1] Abu-Rayyash, H., & Haider, A. (2022). Construction of a Parallel Corpus of English-Arabic Movie Subtitles: A Genuine Source for Audiovisual Translators. *The International Journal of Humanities Education*, 21(1), 21–37. <https://doi.org/10.18848/2327-0063/CGP/v21i01/21-37>
- [2] Al-Kharabsheh, A., & Yassin, O. (2017). Translation of colloquialisms in the Arabic-into-English subtitled film, the dupes. *International Journal of Comparative Literature and Translation Studies*, 5(3), 18-27.
- [3] Alowedi, N. (2024). Bridging cultural divides: A comparative study of subtitling English and Arabic films in the Saudi context. *Proceedings of New Trends in Translation and Technology*, 17–33. 3–6 July, Varna, Bulgaria.
- [4] Alsaadi, A. A., & Rabab'ah, G. (2022). Arabic-English Subtitling of Collocations: The Case of the World Government Summit Held in the UAE. *International journal of linguistics*, 14(2), 168-192.
- [5] Bednarek, M. (2008). *Emotion talk across corpora*. Springer.
- [6] Bernardini, S. (2022). How to use corpora for translation. In *The Routledge Handbook of Corpus Linguistics* (pp. 485-498). Routledge.
- [7] Bogucki, Ł. (2020). *A relevance-theoretic approach to decision-making in subtitling*. Cham: Palgrave Macmillan.
- [8] Dayter, D. (2019). Collocations in non-interpreted and simultaneously interpreted English: a corpus study. In *New empirical perspectives on translation and interpreting* (pp. 67-91). Routledge.
- [9] Gamal, M. Y. (2014). Audiovisual translation in the Arab world (v 0.4): Mapping the field. *Arab media & society*, 19, 1-12.
- [10] Ibrahim, M. A. E. (2019). The Problems of Collocations in Translation. *International Journal of Linguistics, Literature and Translation*, 2(3), 42–53. <https://doi.org/10.32996/ijllt.2019.2.7.15>
- [11] Kayyal, M. H., & Russell, J. A. (2013). Language and emotion: certain English–Arabic translations are not equivalent. *Journal of Language and Social Psychology*, 32(3), 261-271.
- [12] Labrador, B. (2024). A child-oriented corpus-based study of semantically related verbs using sketch engine. *Focus on ELT Journal*, 6(2), 50-69.
- [13] Mansour, Salma. (2016). 'Appraisal Emotional Adjectives in English/Arabic Translation: A Corpus Linguistic Approach'. <http://www.lancaster.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/papers/Mansour.pdf>. Accessed 8 January 2025.
- [14] Obeidat, A. M., & Mahadi, T. S. T. (2019). The translation of Arabic religious cultural collocations in literary texts into English: An application of domestication and foreignization translation strategies. *International Journal*, 2(6), 155-165.
- [15] Sahari, Y. (2023). A corpus-based study of euphemising body parts in Arabic subtitles. *Humanities and Social Sciences Communications*, 10(1), 1-11.
- [16] Sahari, Y. (2024). Subtitling the f-word into Arabic in Hollywood films: a corpus-based study. *Humanities and Social Sciences Communications*, 11(1), 1-10.
- [17] Salameh, M., Mohammad, S., & Kiritchenko, S. (2015). Sentiment after translation: A case-study on Arabic social media posts. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 767-777).
- [18] Shraideh, K. W., & Mahadin, R. S. (2015). Difficulties and strategies in translating collocations in BBC political texts. *Arab World English Journal (AWEJ)*, 6(3), 320-356.
- [19] Thawabteh, M. A. (2012). The translatability of euphemism and dysphemism in Arabic-English subtitling. *Lexis. Journal in English Lexicology*, (7), 145-154.
- [20] Thawabteh, M. (2017). Subtitling on the Intersection of Theory and Practice: Pedagogical Research-Based Approach to Subtitler Training. *AWEJ for Translation & Literary Studies*, 1(1), 32-47.

Sara A. AlShubaily was born in Riyadh, Saudi Arabia. She received the B.A. degree in Translation Studies from King Saud University, Riyadh, Saudi Arabia, in 2001, and the M.A. degree in Applied Linguistics from the same institution in 2008. She earned the Ph.D. degree in Translation Studies from the University of Birmingham, Birmingham, United Kingdom, in 2019.

She has worked as an Assistant Professor at the College of Language Sciences, English Language Department, King Saud University, Riyadh, Saudi Arabia, where she teaches translation, linguistics, and corpus-based research. Her work has been published in peer-reviewed journals, including *Humanities and Social Sciences Communications* and *AWEJ for Translation & Literary Studies*. She is also the author of a doctoral dissertation titled *A corpus-based study of semantic shifts in the translation of self-help books from English into Arabic: a theoretical perspective based on relevance theory* (University of Birmingham, 2019). Her current research interests include audiovisual translation, corpus-based translation studies, semantic shifts in translation, and the pedagogical applications of corpora in translator training.

Ms. AlShubaily has participated in academic conferences and peer-review activities in the field of translation studies. She is committed to advancing research on translation practices between English and Arabic.