SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 3902-3912
Open Access. Online & Print

DOI: 10.5281/zenodo. 121126288

A CORPUS-DRIVEN CONTRASTIVE ANALYSIS OF DIGITAL DISCOURSE MARKERS IN IRAQI ARABIC AND AMERICAN ENGLISH ONLINE COMMUNICATION: A REVIEW

Mohammadreza Pahlavannezhad[1]*, Ali Aswad Dookhi[2]

[1] *Department of Linguistics, Faculty of Letters and Humanities, Ferdowsi University of Mashhad. Mashhad, Iran.*
[2] *General Directorate of Education in Al-Muthanna Governorate, Ministry of Education, Iraq.*

Received: 01/12/2025
Accepted: 02/01/2026

Corresponding author: Mohammadreza Pahlavannezhad (pahlavan@um.ac.ir)

ABSTRACT

This paper reports on a corpus-driven contrastive analysis of digital discourse markers (DDMs) in Iraqi Arabic and American English online communication. By compiling, two corpora of circa 500,000 words each from public Facebook and X (formerly twitter) posts (2020-2025), the study examines the frequencies, forms, positional distributions, and functions of DDMs. The findings show that the overall frequency of DDMs was statistically significantly higher in American English, at 2.76 per 1,000 words, than in Iraqi Arabic, at 2.49. Iraqi Arabic favored elaborative markers like (wa) 'and' and contrastive markers like bas 'but', in initial and medial positions mainly, in establishing relational harmony and opposition management as part of a collectivist mentality. American English, on the other hand, favored inferential markers, for instance, 'so,' and hedging phrases, for example, 'you know,' whose usages are predominantly initial, as part of logical sequencing and politeness due to an individualistic orientation. Both languages used digital variants such as emojis and abbreviations, but with different functional emphases. These results indicate dynamism in the use of DDM and how its use intersects with both language structure, cultural norms (collectivism vs. individualism) and technological affordances. This research contributes to computer-mediated communication theory and is relevant to cross-cultural pragmatics, digital interface design, and language pedagogy.

KEYWORDS: discourse markers, digital communication, contrastive analysis, corpus linguistics, pragmatics.

1. INTRODUCTION

Communication has entered new territory with a multitude of digital media through which people communicate with one another rapidly and constantly (for instance, texting, e-mailing or instant messaging), cutting short the human interaction aspect and allowing for signification at an unprecedented speed and frequency on a global scale. This change has not only affected the means of information exchange, but it has also influenced the linguistic framework on which this exchange is built: Thus to arise peculiarities in online discourse itself (Crystal 2001). The spread of new media technologies, including social media and digital communication services has changed the way individuals have conversations (Bamman et al., 2014), and this trend shows no signs of abating; in confronting these issues scholars are increasingly concerned with how language changes to accommodate/conform to these circumstances as it becomes more evident that speakers have a tendency to form new innovations that are hybrids between informal and formal elements within digital environments (Herring, 2013). These adaptations are reflected in the use of language elements that assist in creating coherence and keeping readers interested, as to make sense and to have an effect on target receivers among brevity and attention-shifting required by online communication (Khudair, 2025).

Among these linguistic elements, discourse markers stand out as essential tools for organizing and signaling relationships within spoken and written communication. Discourse markers refer to words or phrases used in guiding the development of discourse without adding meaning to its propositional content; they are employed in a number of functions such as, among others, indicating transitions, expressing attitudes, and handling interpersonal relationships. Defined as words or phrases that guide the flow of discourse without adding to its propositional content, discourse markers serve such multiple functions as indicating transitions, expressing attitudes, and managing interpersonal dynamics (Schiffrin, 1987). In traditional contexts, such as face-to-face conversations or formal writing, these markers have been intensively studied for their role in enhancing textual cohesion and listener comprehension (Fraser, 1999). The digital context, however, brings new dimensions to their usage, whereby the constraints of character limits and real-time responses more often than not yield, instead, abbreviated or hybridized forms that strongly favor efficiency overelaboration (Androutsopoulos, 2011). For example, in online contexts, discourse markers may well evolve into emojis, abbreviations, or repeated punctuation to carry nuance, almost as spoken intonation does, in a way that covers the gulf between oral and written modes (Darics, 2013).

Research into the use of discourse markers within digital environments, sometimes referred to as digital discourse markers, demonstrates how such elements adapt to the multimodal nature of online communication by providing both visual and textual signals to keep the conversation flowing. Digital discourse markers signal not only logical relations, such as addition or contrast, but also social functions, such as rapport-building or mitigating possible misunderstandings in asynchronous messages, according to Georgakopoulou (2011). This adaptability becomes very active in cross-cultural contexts where users with diverse linguistic backgrounds employ discourse markers to navigate both cultural norms and technological affordances (Tagg, 2012). Comparative analyses further highlighted differences in their deployment, showing that while the discourse function of some markers is universal, frequency and form are far from being similar across languages due to underlying sociocultural influences (Crible, 2017).

Therefore, under a contrastive umbrella, Arabic and English, belonging to distinct families, Semitic and Indo-European, respectively, exhibit different syntactic and pragmatic conventions when it comes to the usage of specific markers, as pointed out by Holes (2004). While Arabic discourse markers often prey upon an elaborate stock of connective particles that stress relational harmony, their English counterparts, predicated on explicitness and logical sequencing, as remarked by Al-Khawaldeh et al. (2014), are an instance of linguistic polarization. It is further enlarged in the dialectal varieties like Iraqi Arabic with colloquial integrations which come with the peculiar history and social orientation of the region in contrast with the more standardized forms common in American-English online discourse, as described by Sultan (2011). The situation becomes more precarious in online communication since Iraqi Arabic users may further integrate code-switching with English loanwords or modernist digital slang, whereas American-English speakers frequently use informal markers to achieve immediacy and friendliness, as discussed by Silvano et al. (2022).

A corpus-based approach makes more empirical the examination of these markers in an us since where WhatsApp, twitter and Facebook are the arenas everywhere interaction takes place. By constructing vast databases of authentic user-generated text, such methodologies are able surface patterns that traditional introspection may not be sensitive to and offer a solid platform for investigating how the use of discourse markers is accommodated in dynamic, user-paced contexts (Biber et al., 1999). Accordingly,

not only does such an approach emphasize the lexical equivalent words between Iraqi Arabic and American English, but it also provides a clear observation of functional shifts in words for emphasis in emotional content and hedging expressions that are used across these two languages (Fu et al., 2024). In this vein, the current study goes a step further by conducting a corpus-driven contrastive analysis of digital discourse markers of Iraqi Arabic and American English in online communication to investigate their forms, functions and frequencies to expose cross-linguistic generalities and cultural subtleties.

2. LITERATURE REVIEW

The study of discourse markers is a major lens for investigating how speakers and writers construct coherent and meaningful interactions. Discourse markers were understood as a category of linguistic items which organize discourse, signal relationships between segments, and facilitate pragmatic inferences without altering propositional content, thus crucially guiding interpreters through the logical, interpersonal, and textual layers of communication (Fraser 1999; Schiffrin 1987). While foundational research established their functions in traditional, face-to-face contexts, the shift to digital mediums has occasioned the need to rethink their form and function.

The digital context creates new constraints and affordances that reshape discourse marker usage. Character limits, real-time responses, and asynchronous exchange favor economy, thus typically yielding elided and hybridized forms (Androutsopoulos 2011). Discourse markers are extended beyond textual words by emojis, abbreviations, and punctuation clusters, which, in these environments, carry paralinguistic nuance across the oral-written divide (Darics 2013). These digital discourse markers adapt to the semiotic mode of online communication, signaling not only logical relations such as addition and contrast but also social functions including rapport-enhancement and mitigation of misunderstanding (Georgakopoulou 2011). This adaptiveness tends to be highly active in cross-cultural contexts where users navigate diverse cultural norms and technological platforms (Tagg 2012).

Comparative analyses have, therefore, pointed out that while the basic discourse functions of markers are perhaps universal, their frequency, form, and pragmatic force are determined by underlying sociocultural influences (Crible 2017). This is very well seen in the comparison of Arabic and English. According to Holes (2004), these languages, representing the Semitic and Indo-European language families, respectively, are embedded in very different syntactic and pragmatic conventions. Arabic discourse markers often access an elaborated repertory of connective particles highlighting relational harmony and collectivist values. On the other hand, discourse markers in English are often based on explicitness and logical succession, appealing to a more individualistic orientation (Al-Khawaldeh et al. 2014).

This polarization becomes even stronger in dialectal varieties. Iraqi Arabic, due to its colloquial integrations and particular sociohistorical orientation, constitutes a very different case from the highly standardized varieties that are common in American English (Sultan 2011). Iraqi Arabic internet users often code-switch, combining English loanwords and digital slang with local dialec resource that indicates negotiating identity and modernity (Silvano et al. 2022). On the other hand, American English users commonly rely on informal markers for the purpose of immediacy and friendliness within a largely monolingual digital ecology.

Corpus-driven approaches have proveda to be an essential tool for the empirical investigation of these subtleties. Enabling such computational work is the fact that by collecting large amount of real-world language use data, researchers are able to extract patterns that traditional introspection can often miss as a strong base for understanding the function of DMs in interactional practices (Biber et al., 1999). The DM-only literature has broadened the range of analysis across genre, language, and pedagogical environments. The challenges encountered when annotating and analizing DMs across languages led to standardized frameworks, as presented by Silvano et al. (2022) proposed an ISO-based annotation scheme across a multilingual corpus. This proposal demonstrated that while certain core operations, such as "expansion," are universal, how they are implemented varies greatly across languages and hence schemes should be flexible enough to accommodate cross-linguistic differences.

Genre-specific analyses further illuminate the functional adaptation of markers. Thus, in media discourse, Al-Khawaldeh et al. (2014) identified, within Arabic sports news, 73 different types of DMs, dominated by the elaborative marker wa (and), which maintains the flow of text. Correspondingly, Fu et al. (2024) noted that during mediatized political interviews, professional role and cultural background influence pragmatic marker selection, with non-native interviewers using "you know" more to build rapport. Similarly, contrastive rhetoric between English and Arabic shows distinct preferences: Sultan (2011) indicated that Arabic academic writing relies more on interactive markers that provide textual coherence, while English uses more engagement markers. Complementing these are

the granular analyses, such as Alsager et al. (2020), who demonstrated that Arabic lakin is mainly a correction marker with medial positional restrictions, while its English counterpart but is much more versatile and frequently used for addition. Empirical research in EFL contexts points to some specific difficulties for Arab learners. Research has shown L1 transfer may lead to an over-reliance on simple markers such as "and" and "but," while a range of other English DMs are underused (Al-Yaari et al., 2013). This "overuse or underuse" issue, based on a general lack of awareness of the variability and cultural appropriacy of DMs, is mitigated by explicit Discourse Marker Instruction, which significantly enhances learners' fluency and coherence, as Hasoon has demonstrated for Iraq (2024). However, significant gaps persist in the literature. Although Arabic and English discourse markers are well-represented in more traditional genres such as media and academic writing, there is still a dearth of research targeting specifically digital contexts, especially for dialectal varieties like Iraqi Arabic. There is a visible lack of contrastive, corpus-driven investigations into how cultural nuance and technological affordances shape the use of digital discourse markers on platforms like social media and messaging apps (Fu et al., 2024; Georgakopoulou, 2011). Moreover, very few explore either multimodal accommodation or functional shifts during real-time digital interactions, and thus there is still a gap in the current understanding regarding cross-linguistic patterns within the setting of globalized online communication. To fill these gaps, this study poses the following research questions:

1. What are the frequencies, forms, and positional distributions of digital discourse markers in Iraqi Arabic and American English online communication?
2. How do the functions of these discourse markers differ between the two languages in digital contexts?
3. What cultural and linguistic factors influence the use of digital discourse markers in Iraqi Arabic versus American English online interactions?

3. METHODOLOGY

This paper therefore uses a corpus-driven contrastive analysis of digital discourse markers in Iraqi Arabic and American English online communication, considering the methodology by Biber et al. (1998), which relies on empirical data obtained from naturally occurring texts to identify patterns without prior hypotheses. Compared to corpus-based methods that test hypotheses, corpus-driven methods let researchers inductively discover linguistic features directly from the data, and for this reason they are particularly appropriate for researching emergent phenomena in digital environments (Tognini-Bonelli, 2001). The design enables one to make systematic comparisons between frequencies, forms, functions, and cultural influences across the two languages and could answer the research questions by quantitative and qualitative perspectives (McEnery & Hardie, 2012).

3.1. Research Design

This research design is mainly quantitative, supplemented with qualitative interpretation, following established practices in discourse analysis, where corpus data provide measurable insights into marker usage, while contextual examples illuminate functional nuances (Baker, 2010). A contrastive framework is followed to highlight the differences and similarities across languages, underpinned by principles of contrastive linguistics, centered on systematic comparisons of equivalent elements across different languages (James, 1980).

To ensure robustness, the study incorporates elements of mixed-methods research: it integrates frequency counts with functional categorization, as recommended for corpus-driven studies into discourse markers (Crible, 2018). Ethical considerations, such as data anonymization and compliance with the terms of service of the platforms where data will be sourced from, guide the process and adhere to the guidelines on digital linguistic research (Page et al., 2014).

3.2. Corpus Compilation

Two parallel corpora were compiled to represent authentic online communication in Iraqi Arabic and American English, comprising approximately 500,000 words each, to ensure statistical reliability (Sinclair, 2005). Data for the Iraqi Arabic corpus are from public posts and comments from platforms such as Facebook and X, targeting Iraqis via location tags, dialectical features, and self-described profiles; it follows, in general, dialect-specific corpus building in Arabic studies (Al-Khawaldeh et al., 2014). Data collection for the American English corpus was sourced from corresponding sources, targeting users in the United States via geolocation and only standard American English variants, excluding non-native influences of any sort, in order to maintain comparability (Herring, 2013).

Other selection criteria included posts ranging from 2020 to 2025, capturing the most modern digital trends; informal interactions featuring topics such as daily life, politics, and entertainment were targeted to create natural discourse (Georgakopoulou, 2011). Data collection was carried out by platform APIs and web scraping tools such as Python's Beautiful Soup

library, which allows random samples to be taken from threads with high levels of engagement, thereby reducing bias (Levshina, 2015). Advertisements, spam, and non-textual items were excluded, thereby providing cleaned datasets that were then balanced in terms of genre and length.

3.3. Data Collection

Data collection involved several steps to collect representative samples. The initial step searches through X using advanced query options to collect up to 10,000 posts per platform, following semantic and keyword search approaches for online corpora (Silvano et al., 2022). Public Facebook groups and pages dealing with Iraqi communities, for example, "Iraq News" or dialect-focused forums, and American groups were sourced manually with the support of tools like ExportComments, which extracts comments in an ethical manner (Fu et al., 2024).

Texts were then normalized by applying libraries such as PyArabic to remove diacritics and perform dialect detection (Zerrouki, 2012), among other tasks. This ultimately produced 1,000 texts per language, and tokenization into sentences was done with the support of NLTK in the case of English and CAMeL Tools in the case of Arabic for comparability purposes (Obeid et al., 2019). Before continuing with further steps, user information was anonymized, and review by an institutional ethics board was sought for privacy protection.

3.4. Data Analysis

Analysis was conducted in three stages to investigate the research questions: identification, quantification, and interpretation. Digital discourse markers were first identified through a hybrid process combining automated keyword searches with manual annotation. Based on Fraser's seminal work (1999) on elaborative, contrastive, inferential, and temporal, an initial list of markers was generated from previous studies (e.g., "wa" and "bas" in Arabic; "and" and "but" in English), and then extended inductively to account for digital varieties like emojis (e.g. for mitigation) and abbreviations (e.g., "lol" for stance) (Darics, 2013).

Tools like AntConc (Anthony, 2022) helped to conduct concordance searches for frequency and collocation patterns, whereas positional distribution was analyzed by Python scripts with spaCy, considering initial, medial, and final positions (Honnibal & Montani, 2017). Quantitative analysis included normalization based on frequencies (per 1,000 words) by means of chi-square tests for the cross-linguistic comparisons, following contrastive corpus studies (Alsager et al., 2020). Differences in function were analyzed by manually coding 20% of instances using two coders who reached inter-rater reliability based on Cohen's kappa ($\kappa > 0.80$) (McHugh, 2012). Qualitative analyses studied contextual examples to identify cultural influences such as in-group collectivism versus individualism, considering thematic coding supported by NVivo software (QSR International, 2020). This approach allows for a wide-ranging and reproducible investigation that begins to fill in many of the gaps in existing digital discourse research, as it combines corpus-driven methods with contrastive perspectives (Sultan, 2011; Fu et al., 2024).

4. RESULTS

This section discusses the outcomes of the corpus-driven contrastive analysis of digital discourse markers in Iraqi Arabic and American English online communication. The results are categorized according to the three research questions, informed by data from the respective corpora compiled of around 500,000 words each from public posts on X (formerly Twitter) and Facebook across 2020–2025. Quantitative data encompass frequency counts per 1,000 words, forms, and positional distribution, using AntConc for concordancing (Anthony, 2022) and Python scripts for statistical testing (Honnibal & Montani, 2017) (chi-square tests, with $p < 0.05$ as indicative of statistical significance). Qualitative interpretations incorporate functional categorization along dimensions identified by Fraser's (1999) taxonomy, viz. elaborative, contrastive, inferential, and temporal, with contextual extracts from the corpora. Given that a corpus-driven approach is essentially inductive, extended forms of digital variants such as emojis, for instance used for mitigation, and abbreviations, for example, "lol" for stance, were included.

4.1. Results for Research Question 1

The first research question aimed to identify the frequencies, forms, and positional distributions of digital discourse markers in Iraqi Arabic and American English online communication. The analysis identified 1,245 discourse markers in the Iraqi Arabic corpus and 1,378 in the American English corpus, with normalized frequencies of 2.49 per 1,000 words for Iraqi Arabic and 2.76 per 1,000 words for American English. This slight difference was statistically significant ($\chi^2 = 12.34$, $p < 0.01$), suggesting higher overall use in American English digital interactions, probably due to the emphasis the platform puts on brevity and expressivity.

4.2. Frequencies and Forms

Table 1 summarizes the top 10 most frequent discourse markers in each language, including traditional textual forms and digital variants. In Iraqi Arabic, elaborative markers such as (wa, meaning "and") dominated, reflecting a preference for connective harmony in informal exchanges. Contrastive markers like (bas, meaning "but") were common; these were often used for opposition in emotional or argumentative posts. Inferential (yaani, meaning "meaning" or "so") appeared frequently for clarification. Digital forms included repeated punctuation, e.g.for emphasis, and emojis, e.g., for hedging. In American English, inferential markers such as "so" and hedging phrases like "you know" were common, as were contrastive "but" and elaborative "like" (as a filler). Digital variants like "lol" and emojis, e.g. performed similar mitigative functions.

Table 1: Top 10 Frequent Discourse Markers

Rank	Iraqi Arabic DM	Form/Type	Raw Count	Normalized Frequency (per 1,000 words)	American English DM	Form/Type	Raw Count	Normalized Frequency (per 1,000 words)
1	(wa)	Elaborative (conjunction)	458	0.92	so	Inferential (adverb)	512	1.02
2	(bas)	Contrastive (particle)	312	0.62	but	Contrastive (conjunction)	398	0.80
3	(yaani)	Inferential (particle)	245	0.49	you know	Hedging (phrase)	289	0.58
4	!!!	Digital emphasis	112	0.22	like	Elaborative/filler (adverb)	201	0.40
5	(emoji)	Mitigative (visual)	98	0.20	lol	Mitigative (abbreviation)	156	0.31
6	(la)	Contrastive (negative particle)	87	0.17	and	Elaborative (conjunction)	132	0.26
7	(ay)	Confirmatory (particle)	65	0.13	well	Hedging (adverb)	89	0.18
8	(hay)	Demonstrative (particle)	52	0.10	anyway	Topic shift (adverb)	67	0.13
9	(ami)	Interpersonal (vocative)	41	0.08	yeah	Agreement (interjection)	54	0.11
10	(hahaha)	Mitigative (laughter)	35	0.07	😊 (emoji)	Mitigative (visual)	48	0.10

Thus, Iraqi Arabic forms were mostly particles and conjunctions from dialectal vocabulary, while 68% were textual and 32% digital/multimodal. For American English, the division was similar, 65% textual and 35% digital, but there were more phrasal forms, such as "you know," suggestive of a hedging style.

4.3. Positional Distributions

Table 2 shows positional preferences. Iraqi Arabic markers thus tended to occur in medial positions (52%), as facilitating embedded clarifications in flowing dialectal sentences, where the initial positions (38%) are used for turn-taking in replies, while final positions were least used at 10%, much like emotional closures. American English preferred initial positions for topic shifts at 45%, medial for hedging at 40%, and final for emphasis at 15%. The differences are significant ($\chi^2 = 18.67$, $p < 0.001$), pointing to syntactic divergences: Arabic's VSO structure embeds markers medially, while the SVO of English allows initial signaling.

Table 1: Positional Preferences

Position	Iraqi Arabic (% of Total DMs)	Example Extract	American English (% of Total DMs)	Example Extract
Initial	38%	(But imagine the year...) – Starting opposition in a political thread.	45%	"But you know they are getting huge kickbacks..." – Initiating contrast in a discussion.
Medial	52%	(...meaning they came out (in slippers)) – Embedded inference in casual commentary.	40%	"...they are getting huge kickbacks for doing these procedures but you know..." – Hedging mid-sentence.
Final	10%	(...and I check again 100 times) – Closing elaboration in a personal anecdote.	15%	"...she just sits there judging my indecision lol" – Mitigative closure in a pet naming post.

These patterns suggest Iraqi Arabic uses DMs for relational continuity within sentences, while American English employs them for explicit structuring at boundaries.

4.4. Results for Research Question 2

The second research question was an intent to unseal how the functions of these discourse markers differ between the two languages in digital contexts.

Functions were coded for 20% of instances with inter-rater reliability (κ = 0.85) (McHugh, 2012), which indicated distinct profiles. Iraqi Arabic favored elaborative (45%) and contrastive (30%) functions in developing harmony and opposition in social debates, and inferential (20%) for explanations, while temporal functions were rarely (5%) used. On the other hand, American English preferred inferential (35%) and hedging (25%) to clarify and be polite, while contrastive (20%), elaborative (15%), and temporal (5%) supported argumentative flow.

4.5. Functional Differences

Table 3 highlights significant functions with examples from both varieties. As shown below, differences were significant ($\chi^2 = 22.45$, $p < 0.001$), with Iraqi Arabic's elaborative use fostering collectivist rapport, e.g., "و" linking ideas in family-oriented posts, versus American English's inferential/hedging for individualistic nuance, e.g., "you know" assuming shared knowledge in opinions.

Table 3: Functional Differences

Function	Iraqi Arabic (%)	Example Extract & Interpretation	American English (%)	Example Extract & Interpretation
Elaborative	45%	(Iraq's spring is short but sweet, flowers and greenery) – Adds details to evoke shared cultural nostalgia.	15%	"You've put so little time... but other gods are absolutely your problem" – Expands argument logically.
Contrastive	30%	(Any one who comes doesn't help or harm, meaning harder than the fool) – Opposes expectations in political critique.	20%	"I'm actually a rape survivor... but to protect women from men" – Contrasts personal experience with broader advocacy.
Inferential	20%	(Meaning it's been years and they haven't los hope) – Draws conclusions from ongoing even	35%	"So here's Rosie hers first TOT" – Infers a decision based on prior context.
Hedging/ Mitigative	5%	(Hahaha meaning who loves Zainab) – Softens humor in social commentary.	25%	"You guys can't even defend... Thats why." – Mitigates criticism with "you know"-like assumption.
Temporal	5%	Rare; e.g. (After some time) – Sequences narratives.	5%	Rare; e.g., "then they start asking" – Marks progression in explanations.

Interpretations show that Iraqi Arabic DMs express the importance of social cohesion in emotional digital spaces, while American English uses them for pragmatic efficiency in debates.

4.6. Results for Research Question 3

The third research question was intended to explore the cultural and linguistic factors that influence the use of digital discourse markers in Iraqi Arabic versus American English online interactions. Qualitative thematic analysis revealed cultural influences rooted in the respective linguistic orientation toward collectivism versus individualism, and linguistic factors such as dialectal flexibility versus standardization. For instance, in Iraqi Arabic, DMs reflected Mesopotamian emphasis on relational harmony (Holes, 2004), whereby was building community in posts like (And the Euphrates river flows in China), satirically linking local identity to global absurdity. Factors included code-switching, such as mixes of Arabic and English, which amplified emotional stance amid sociopolitical tensions.

In American English, hedging out of politeness was driven by individualism (Tagg, 2012), as in "but you know they are getting huge kickbacks", assuming shared skepticism of critiques. Linguistic standardization favored explicit inferences, whose directness was tempered in diverse interactions through such digital forms as "lol".

Overall, Iraqi cultural collectivism favored elaborative/contrastive DMs in solidarity, while American individualism highlighted the use of inferential/hedging for autonomy. Iraqi Arabic dialectalism allowed expressive hybrids, whereas structured varieties are characteristic of English. All these aspects point to the adaptation to digital affordances, which platforms amplify.

5. DISCUSSION

5.1. Research Question 1: Frequencies, Forms, and Positional Distributions

While the normalized frequency is higher in American English than in Iraqi Arabic, 2.76 and 2.49 per 1,000 words, respectively, the differentiated forms and positional preferences point to cross-linguistic variation as a function of communicative demands within online platforms. In Iraqi Arabic, the prevalence of elaborative forms such as "و" (wa) and medial positioning supports Coherence Theory in its assumptions that discourse markers create textual coherence through connective ties, especially in languages that rely more on relational flow (Halliday & Hasan, 1976). The medial embedding thus enables embeddings for clarifications within dialectal sentences, reduces cognitive effort, and enhances relevance according to Relevance Theory by optimizing relevance in fragmented digital messages (Sperber & Wilson, 1986). For example, the extract

reflects how the inferential markers like (yaani) are medially placed in a sentence to infer meaning mid-sentence, in tune with the harmony-oriented character of Arabic discourse.

American English, on the other hand, leans more toward inferential forms such as "so" and initial positioning, reinforcing pragmatic approaches, which see markers as means of explicit turn management and logical sequencing (Schiffrin, 1987). In the extract, "But you know they are getting huge kickbacks...," there are clear initial contrastive and hedging markers signaling topic shifts, which is propitious to efficiency in asynchronous interactions within the frames of Computer-Mediated Communication Theory, where compensation for nonverbal cues is needed in digital contexts (Walther, 1996). Digital forms include emojis and abbreviations such as "lol" in both languages, but higher multimodal proportions in American English (35% vs. 32%) relate to the adaptations to platform affordances, where ambiguity is reduced by visual cues (Darics, 2013).

These patterns echo previous studies. Al-Khawaldeh et al. (2014) highlighted the high frequencies of elaborative "wa" in Arabic sport texts, with initial positioning dominant but now in this digital context tends to medial for relational continuity, extending their findings onto online brevity. Similarly, Alsager et al. (2020) noted medial "lakin" in Arabic newspapers, which is paralleled in the medial bias here, while Fu et al. (2024) reported medial "you know" in English interviews, in line with American English medial hedging, in contrast to the initial preference in digital data. Recent work on non-native Arab English speakers reflects a similar pattern of frequency disparities, with DMs such as "you know" put to hedging use in speech, and suggesting that in digital contexts these are amplified in writing (Al-Harahsheh, 2025). Sultan (2011) also notes higher frequencies of interactive markers in Arabic for textual coherence, to which the present medial forms would attest, in digital harmony-building. However, the altogether lower frequencies in Iraqi Arabic may be a result of dialectal conciseness, as with Jordanian Arabic DM "ʔinno" tending towards embedded positions (Al-Mansoori, 2025). The suggestion that Iraqis prefer medial flexibility is an extension of Silvano et al. (2022), where multilingual corpora revealed positional variation in translations, into digital realms.

5.2. Research Question 2: Functions of Discourse Markers

This functional divergence, Iraqi Arabic's focus on elaborative (45%) and contrastive (30%) roles, while American English focuses on inferential (35%) and hedging (25%), sheds light on how digital discourse markers serve pragmatic and social purposes in cross-cultural online interactions. Iraqi Arabic elaborative functions, such as adding cultural details, (Iraq's spring is short but sweet, flowers and greenery), support the Relevance Theory by assuming a shared contextual knowledge to evoke nostalgia and minimizing processing effort in collectivist exchanges (Sperber & Wilson, 1986). Contrastive usages for opposition, like in tune with pragmatic theories that frame markers as stance managers within emotional critiques (Schiffrin, 1987), which would allow for easy rapport in sociopolitically charged digital threads.

Conclusiveness of American English, such as in "So here's Rosie hers first TOT" drawing conclusions and hedging in "...they are getting huge kickbacks for doing these procedures but you know...," indicates logical sequencings (Halliday & Hasan, 1976) extended to digital argumentation via Coherence Theory. Hedging, "...but you know...," mitigates face-threats, according to Computer-Mediated Communication Theory, promoting hyperpersonal connections in individualistic debate (Walther, 1996). The infrequence of temporal functions in both, 5%, indicates synchronic relations have priority over sequencing in the digital platform, unlike traditional discourse.

These findings extend earlier results. Fraser (1999) had characterized English markers as inferential for procedural guidance, following the American pattern, while Al-Khawaldeh et al. (2014) identified elaborative dominance in Arabic texts, to which the present study extends digital elaborations for harmony. Fu et al. (2024) found "you know" for hedging in interviews, paralleling digital uses here, but with higher inferential roles in American online data. In Arabic-English contrasts, Sultan (2011) noted Arabic's interactive functions for coherence, akin to elaborative here, versus English's interactional hedging. Recent studies on Arab EFL writing show similar functional shifts, with DMs strengthening persuasion through inferential means in English but elaborative in Arabic-influenced texts (Alotaibi, 2025). Emotional markers in Lebanese social media during 2024 hostilities reveal contrastive functions for identity expression, mirroring Iraqi oppositional uses (Al-Azzawi, 2025). This corroborates Silvano et al. (2022), where multilingual DMs mapped cause-expansion relations, but digital contexts amplify hedging in English for politeness.

5.3. Research Question 3: Cultural and Linguistic Influences

Pragmatic stance management, according to Schiffrin (1987), influences cultural factors such as

Iraqi Arabic for elaborative markers in driving collectivism, for instance, linking local identity satirically: (And the Euphrates river flows in China), and American English for hedging in driving autonomy, "but you know they are getting huge kickbacks", assuming shared skepticism. The linguistic dialectalism of Iraqi Arabic accommodates expressive code-switching in line with Relevance Theory's principled optimization in variable contexts (Sperber & Wilson, 1986), and the English standardization lends itself more to explicit inferences according to Coherence Theory (Halliday & Hasan, 1976). Digital affordances amplify these through Computer-Mediated Communication Theory by means of which markers compensate for cues in global interactions (Walther, 1996).

These influences echo earlier work. Holes (2004) identified Arabic markers for relational harmony, extended here to Iraqi digital collectivism, contrasting Tagg's (2012) English hedging for inclusivity. Sultan (2011) noted metadiscourse differences, with Arabic interactive for engagement, paralleling collectivist functions. Recent analyses of Jordanian "Xalasʕ" show pragmatic mitigation in spoken Arabic, similar to Iraqi emotional stance (Al-Qadi, 2025). In digital emotions, Lebanese markers reflect identity amid conflict, aligning with Iraqi sociopolitical uses (Al-Azzawi, 2025). Cross-linguistic metadiscourse in Qatari texts shows Arabic glosses for explicitness but digital data here highlight collectivist elaboration (Al-Mutairi, 2025). This extends Fu et al. (2024) and Silvano et al. (2022), where functional shifts in multilingual corpora reflect cultural adaptations, to dialectal online influences.

6. CONCLUSION

This corpus-based contrastive study has revealed the nuanced pragmatic roles of the DDSs as discourse makers in Iraqi Arabic and American English distributed discourses, highlighting the interaction between language patterns, cultural practices, and technological transfers. The research reveals that while both English and Chinese draw on discourse markers to achieve coherence and interactiveness in the online environment, they have different frequency, form, position distribution, function and influence. Iraqi Arabic uses elaborative and contrastive mediative forms in order to preserve relationship symmetry, indicative of a collectivist orientation in informal talk. American English, in contrast, makes its preference-wide for inferential and hedging markers clear from the placement of these in the initial position within sentences, indicating a wide concern with explicitness and individual nuance (among speakers) concerning argumentation or opinionated speech. Online varieties Online, both the English (/aps/) and the British (p/s/h/a) are frequent but serve as linguistic fillers for softening more so in American than British English. In general, these findings point to the way in which online environments exaggerate language-specific strategies and that discourse markers mutate under the influence of multimodality and descript postage constraints to effect communication across cultural borders.

The theoretical, the practical and the pedagogical axes of this research have implications that are discussed. Theoretically, the research further strengthens discourse markers as being dynamic entities that do change and adapt in digital environments and adds to models of computer-mediated communication by demonstrating how they can make up for the lack of nonverbal signals through hybrid forms. Such crosslinguistic contrasts may fine-tune perspectives on the analysis ofprocedural meaning in online discourse, particularly in terms of relational vs logical priority in Semitic layout relative to that found in (e.g.) most Indo-European layouts. Concretely, this may contribute to the design of digital communication tools such as chat algorithms or translation programs by drawing attention to how culture-specific marker usage can serve to minimize misunderstanding on a global scale. For instance, dialect-aware social network features would be particularly helpful for users from regions like Iraq, where the embedded Meta data have an emotional connect. Pedagogically, the results suggest that this language of the internet should be part of Arabic and English language instruction in these online environments. It could be, for instance, of functional oppositions at the level of a curriculum module design to promote pragmatic competence and hence afford communication among peoples across different cultures in the digital era.

REFERENCES

Al-Azzawi, H. (2023). Arabic discourse markers in Al-Jazeera.net news articles. Arabiyat: Journal of Arabic Education and Arabic Studies, 10(2), 1-20.

Al-Azzawi, H. (2025). Emotional markers in Lebanese social media during hostilities. Discourse, Context & Media, 55, 100-115.

Al-Harahsheh, M. A. (2025). Functions of discourse markers in nonnative English speech: The case of Arab English speakers. Journal of Pragmatics, 220, 45-60.

Al-Khawaldeh, A. A., Mat Awal, N., & Zainudin, I. S. (2014). A corpus-based description of discourse markers

in Arabic sport journalistic texts. Journal of Islamic and Human Advanced Research, 4(4), 200-215.
Al-Mutairi, F. (2025). Cross-linguistic metadiscourse in Qatari English texts. Journal of Second Language Writing, 67, 101-120.
Al-Qadi, N. (2025). Gender differences in using discourse markers in spoken Jordanian Arabic. Jordanian Educational Journal, 10(3), 342-365.
Alsager, H. N., Afzal, N., & Aldawood, A. A. (2020). Discourse markers in Arabic and English newspaper articles: The case of the Arabic lakin and its English equivalent but. Arab World English Journal, 11(1), 154-165. https://dx.doi.org/10.24093/awej/vol11no1.13
Alotaibi, K. (2025). Attitudinal resources in Arabic and English discourse: A comparative study. International Journal of Applied Linguistics, 35(1), 78-95.
Androutsopoulos, J. (2011). Language change and digital media: A review of conceptions and evidence. In K. Tore (Ed.), Standard languages and language standards in a changing Europe (pp. 145-159). Novus Press.
Anthony, L. (2022). AntConc (Version 4.0.6) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/
Baker, P. (2010). Sociolinguistics and corpus linguistics. Edinburgh University Press.
Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press.
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. Longman.
Crible, L. (2017). Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. International Journal of Corpus Linguistics, 22(2), 242-269.
Crible, L. (2018). Discourse markers and (dis)fluency: Forms and functions across languages and registers. John Benjamins Publishing Company.
Crystal, D. (2001). Language and the Internet. Cambridge University Press. https://doi.org/10.1017/CBO9781139164771
Darics, E. (2013). Non-verbal signalling in digital discourse: The case of letter repetition. Discourse, Context & Media, 2(3), 141-148. https://doi.org/10.1016/j.dcm.2013.07.002
Fischer, K. (2006). Approaches to discourse particles. Elsevier.
Fraser, B. (1999). What are discourse markers? Journal of Pragmatics, 31(7), 931-952. https://doi.org/10.1016/S0378-2166(98)00101-5
Fu, Y., Afzaal, M., & El-Dakhs, D. A. S. (2024). Investigating discourse markers "you know" and "I mean" in mediatized English political interviews: A corpus-based comparative study. Frontiers in Communication, 9, 1427062. https://doi.org/10.3389/fcomm.2024.1427062
Georgakopoulou, A. (2011). Small stories as a paradigm for narrative analysis in social media. In Proceedings of the 44th Hawai'i International Conference on System Sciences (HICSS-44).
Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. Longman.
Herring, S. C. (2013). Introduction to the pragmatics of computer-mediated communication. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), Pragmatics of computer-mediated communication (pp. 3-32). De Gruyter Mouton.
Holes, C. (2004). Modern Arabic: Structures, functions, and varieties. Georgetown University Press.
Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing [Unpublished manuscript].
James, C. (1980). Contrastive analysis. Longman.
Khudair, R. N. (2025). Corpus-based contrastive analysis of cohesion in academic English writing: The case of Arab and English research articles (Unpublished doctoral dissertation). Islamic Azad University, Isfahan (Khorasgan) Branch, International Faculty.
Levshina, N. (2015). How to do linguistics with R: Data exploration and statistical analysis. John Benjamins Publishing Company.
McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. Biochemia Medica, 22(3), 276-282. https://doi.org/10.11613/BM.2012.031
Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Al Afif, A., Malykh, O., Inoue, G., & Eryani, F. (2019). CAMeL tools: An open source Python toolkit for Arabic natural language processing. Proceedings of the 12th Language Resources and Evaluation Conference, 7022-7032.

Page, R., Barton, D., Unger, J. W., & Zappavigna, M. (2014). Researching language and social media: A student guide. Routledge.

QSR International Pty Ltd. (2020). NVivo (Version 12) [Computer software]. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Schiffrin, D. (1987). Discourse markers. Cambridge University Press.

Schourup, L. (1985). Common discourse particles in English conversation. Garland.

Silvano, P., Damova, M., Valūnaitė Oleškevičienė, G., Liebeskind, C., Chiarcos, C., Trajanov, D., Truică, C.-O., Apostol, E.-S., & Bączkowska, A. (2022). ISO-based annotated multilingual parallel corpus for discourse markers. Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), 2739-2749.

Sinclair, J. (2005). Corpus, concordance, collocation (Reprint ed.). Oxford University Press. (Original work published 1991)

Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Blackwell.

Sultan, A. H. J. (2011). A contrastive study of metadiscourse in English and Arabic linguistics research articles. Acta Linguistica, 5, 28-41.

Tagg, C. (2012). The discourse of text messaging: Analysis of SMS communication. Continuum.

Tognini-Bonelli, E. (2001). Corpus linguistics at work. John Benjamins Publishing Company.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Communication Research, 23(1), 3-43.

Zerrouki, T. (2012). PyArabic: An Arabic language library for Python [Software]. https://pypi.org/project/PyArabic/