Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Digital English as Pragmatic Repertoire: A CorpusPragmatic Study of Multimodal Change and EFL Pedagogy

Abdelaziz Mohammed

amimohammed@bu.edu.sa

Al-Baha University

Research Article

Keywords: digital pragmatics, multimodal discourse, social media, digital literacies, EFL pedagogy

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10036473/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Digital platforms are now central to everyday communication, yet their impact on the pragmatic organization of English and on English as a foreign language (EFL) teaching remains poorly specified. This mixed-methods, corpus-pragmatic study investigates how social media communication reshapes English at the lexical, syntactic, semantic, typographic, multimodal, and discourse-pragmatic levels, and explores how these changes can inform EFL pedagogy. A multi-platform corpus of 70,000 English-dominant posts from seven platforms (X, Instagram, TikTok, Reddit, WhatsApp, Discord, and Telegram; January–December 2025) was compiled and annotated for lexical innovation, semantic shift, syntactic variation, typographic grammar, multimodal tokens, abbreviations, and pragmatic features. Complementary data were gathered from 50 expert interviews and questionnaires completed by 160 stakeholders, including teachers, curriculum designers, content creators, community moderators, and EFL learners. Abbreviations and clippings emerged as the most frequent phenomena (43.1% of coded instances), followed by multimodal tokens (33.0%) and lexical innovations (29.6%), while syntactic change was relatively rare (4.4%). These patterns indicate that digital communication redistributes pragmatic work onto lexical, orthographic, typographic, and visual resources while leaving core clause structure comparatively stable. Participants reported both opportunities (greater expressivity, engagement, and authenticity) and challenges (register confusion, exclusion, cultural appropriation, and exposure to algorithmically standardized language). The study concludes that digital English functions as a multimodal pragmatic repertoire and argues for integrating digital registers, multimodal literacy, and pragmatic awareness into EFL curricula and teacher education.

INTRODUCTION

Human communication has always been shaped by the tools available, but twenty-first-century digital platforms have accelerated and complicated this relationship in unprecedented ways [1]. Social media, messaging apps, and online communities now mediate a large proportion of everyday interaction, creating environments in which written and visual resources are deployed with speech-like immediacy and reach. Scholars have argued that this shift gives rise to new forms of “digital literacy” in which understanding and producing messages requires navigating a dense field of signs, symbols, images, and algorithms [2–4]. In such environments, mastery of Standard English, while still necessary, is no longer sufficient: users must interpret and control a wide range of pragmatic and multimodal cues.

From a corpus-pragmatic perspective, these developments invite systematic investigation of how pragmatic functions such as stance-taking, face-work, alignment, and irony are realized in large collections of naturally occurring digital discourse [5, 6]. Rather than treating abbreviations, emojis, or typographic play as marginal “extras,” corpus pragmatics conceptualizes them as recurrent, annotatable resources that perform interactional work and can be quantified, compared across platforms, and interpreted qualitatively. This study adopts such a corpus-pragmatic stance, using an annotated multi-platform corpus to examine how digital English constitutes a multimodal pragmatic repertoire and how this repertoire can be brought into EFL pedagogy.

Computer-Mediated Communication (CMC) research highlights how platform affordances and social dynamics reshape language norms. The internet has enabled “networked individualism” [5], where individuals craft highly personalized linguistic identities across loosely connected audiences. This environment often privileges communicative efficiency and affective expressivity over traditional prescriptivism. Online writing has become more conversational, blurring the boundary between speech and text, while simultaneously fostering a “typographic tone of voice” in which capitalization, punctuation, and spacing act as resources for expressing stance and emotion [7].

At the same time, digital discourse is increasingly multimodal. Emojis, GIFs, stickers, and visual memes co-occur with text and can be central to meaning. Engagement-based research on Twitter/X, for example, shows that posts containing emojis receive significantly more interaction, suggesting that such elements are not mere decoration but part of the communicative core [8]. Recent work conceptualizes emojis as illocutionary force indicating devices (IFIDs) that modulate tone, politeness, and alignment in compressed messages [9, 10]. Academics further argue that sequences of emojis exhibit a constrained “grammar of visual language” [11]. In parallel, studies of AI-mediated communication and predictive text suggest emerging forms of algorithmic mediation, where language models and recommendation systems shape which forms of English are more visible or normative [7, 12].

Despite growing interest in these developments, the implications for English linguistics and EFL pedagogy remain underspecified. Much public discourse continues to frame digital English as a threat to “proper” language, while educational responses oscillate between banning digital features (e.g., abbreviations, emojis) and uncritical celebration. There is a need for empirically grounded, theoretically informed work that treats digital English as a pragmatic and multimodal repertoire, examines where and how change occurs, and explores how this repertoire can be used responsibly in language education.

Corpus Pragmatic Perspective

Corpus pragmatics integrates methods from corpus linguistics with concepts and questions from pragmatics and discourse analysis [13]. Rather than focusing solely on word frequencies, corpus-pragmatic studies investigate how pragmatic functions such as stance-taking, politeness, alignment, and humor are realized in recurrent lexico-grammatical and discourse patterns in large datasets. This typically involves combining quantitative distributional analysis with close qualitative examination of concordance lines and text segments. In digital contexts, corpus pragmatics has been used, for example, to trace how users index affiliation and disaffiliation through quotation and irony on Twitter, or how politeness strategies vary across online communities.

In the present study, a corpus-pragmatic approach is adopted in two ways. First, categories such as ABB, MM, TYPO, and PRAG are explicitly defined as pragmatic resources that can be searched, counted, and compared across platforms. Second, the quantitative patterns arising from this annotation are interpreted qualitatively in terms of the interactional work they perform—for instance, how emojis function as illocutionary force indicating devices, how abbreviations serve as markers of in-group

identity, or how typographic features construct a “tone of voice” that mitigates or aggravates face-threatening acts. This dual perspective allows the analysis to move beyond listing forms towards explaining how digital English constitutes a multimodal pragmatic repertoire.

This study responds to three gaps in the literature. The first concerns the scope of linguistic change in digital environments. Existing discussions frequently focus on individual phenomena—such as slang, abbreviations, or emojis—without systematically comparing their relative prevalence or examining how they interact across multiple platforms. There is a need for large-scale corpus work that treats digital English as involving change in several components at once: lexical, semantic, syntactic, typographic, multimodal, and discourse-pragmatic.

Corpus-based work on digital language has tended to focus either on lexical innovation or on isolated multimodal features such as emojis. Much less work has simultaneously compared multiple pragmatic resources within the same annotated corpus or examined how these resources cluster at the post level. A corpus-pragmatic design that treats abbreviations, emojis, typographic cues, and pragmatic markers as jointly available resources is therefore needed to understand digital repertoires holistically.

The second gap involves theory. While key notions from pragmatics and discourse analysis are often invoked, many empirical studies of digital English do not use an explicit pragmatic framework to guide category construction or interpretation. This can lead to a focus on surface forms rather than on the underlying communicative actions they enable. A theoretically explicit approach is particularly important when considering how digital language practices redistribute work among lexical items, orthography, visual tokens, and interactional routines, and how these distributions vary by platform.

Within corpus pragmatics, category design is not a purely technical matter but a theoretical one: the chosen categories embody particular assumptions about what counts as pragmatic work and how it should be distributed across levels of analysis. Without an explicit framework, there is a risk that annotation collapses heterogeneous phenomena or overlooks subtle but recurrent patterns of stance and alignment.

The third gap is pedagogical. Work on digital literacies has highlighted the importance of preparing learners to participate in online environments, and some classroom studies show that tasks based on Twitter, Instagram, or other platforms can enhance motivation, vocabulary, and engagement. Nevertheless, formal curricula and assessment systems often remain centered on monomodal, decontextualized texts, and many teachers report uncertainty about whether, and how, digital registers should be addressed. The pragmatic complexities of digital English—especially around politeness, alignment, humor, and offense—are rarely treated systematically in EFL materials. There is a need for research that bridges corpus-based descriptions of digital pragmatics with concrete, empirically grounded proposals for EFL pedagogy.

From a corpus-pragmatic teaching perspective, what is missing is classroom material that makes these patterns visible to learners: concordances of emojis used to soften disagreement, sets of abbreviations

in workplace WhatsApp exchanges, or collections of typographic strategies used to signal irony. Such resources could support explicit instruction in digital pragmatics, but they require robust descriptive work on authentic corpora.

In response to these gaps, the present study adopts a dual theoretical framework that combines digital pragmatics and digital literacies. Digital pragmatics is used to conceptualize features such as abbreviations, emojis, hashtags, typographic patterns, and discourse markers as resources for accomplishing illocutionary acts, managing face, and negotiating stance in interaction. This perspective foregrounds how users adjust their language to audience, platform, and context, and how pragmatic work is distributed across words, symbols, images, and orthographic strategies. Digital literacies provide a complementary pedagogical lens, emphasizing that learners require not only technical ability to use digital tools but also critical awareness of how meaning is made across modes and how power and inequality are reproduced through language choices online.

Within this framework, the researcher treats features such as abbreviations, emojis, typographic conventions, and discourse markers as corpus-searchable pragmatic resources and interprets their distribution and co-occurrence patterns using corpus-pragmatic methods.

Within this framework, the study focuses on three research questions. The first asks how digital communication influences linguistic change in English, understood broadly to include changes in lexical forms, meanings, structures, typographic conventions, and multimodal practices. The second asks which component—lexical, syntactic, semantic, typographic, multimodal, or other—this influence is most pronounced in. The third asks how the multimodal features of social media platforms, particularly X, can be incorporated into EFL teaching to enhance learners' pragmatic skills, digital literacy, and communicative effectiveness, rather than undermining formal competence.

The study contributes to knowledge in three main ways. It provides one of the largest annotated corpora of English digital discourse to date, covering seven major platforms and distinguishing systematically among lexical, semantic, syntactic, typographic, multimodal, and pragmatic features. It develops and applies a transparent coding scheme grounded in pragmatic and multimodal theory, enabling fine-grained analysis of how specific resources are used across platforms. It integrates corpus findings with perspectives from experts, teachers, content creators, moderators, and learners, thereby linking descriptions of digital English to the concerns of pedagogy and practice.

The remainder of the article is organized as follows. The next section outlines the research methods, including corpus construction, annotation, and stakeholder data collection. Results are then presented and discussed together, focusing first on cross-platform patterns of digital pragmatics and then on questionnaire and interview data. The final section offers conclusions, including implications for EFL teaching and suggestions for future research on digital pragmatics and pedagogy.

DATA AND METHODS

The study employed a mixed-methods design to combine large-scale corpus analysis with multi-stakeholder perspectives (see Table 1). A quantitative strand constructed and annotated a multi-platform corpus of English digital discourse. A qualitative strand collected data through expert interviews and questionnaires administered to teachers, content creators, moderators, and learners (see Table 2; Appendices D and E). Both strands addressed the same research questions and were integrated at the interpretation stage. Quantitative analyses are corpus-based and corpus-pragmatic in that they focus on the distribution of pragmatic and multimodal features across platforms and relate these patterns to qualitative interpretations of their discourse functions.

Table 1
Qualitative participants by role and region (N = 160)

Role	**N**	**%**	**Major Regions Represented**
Academic researcher	40	25	UK; US; China; Brazil
EFL teacher (K–12)	40	25	Spain; Indonesia; Mexico
EFL teacher (Higher ed)	24	15	Saudi Arabia; Sudan; Yemen; Egypt
Curriculum designer/materials writer	12	7.5	Global
Content creator/influencer	20	12.5	US; Nigeria; Philippines
Community moderator/platform practitioner	12	7.5	Global
EFL learners (varied proficiencies)	12	7.5	Multiple L1s
Total	160	100	

Table 2
Participant qualifications and practitioner status (N = 160)

Category	**N**	**%**
PhD / equivalent (academia)	40	25
MA / MEd / CELTA-qualified teachers	56	35
In-service teachers (no advanced degree)	36	22.5
Industry professionals (platform/edtech)	12	7.5
Active content creators/moderators	16	10
Total	160	100

Table 2 shows that 25% of participants hold doctoral degrees and 35% have advanced teaching qualifications, while the remainder are in-service teachers, industry professionals, and active content creators or moderators. This mix ensures that both academic and practitioner perspectives are represented (see Appendix E, Tables E1–E4).

Corpus construction and sampling

A two-stage process was used to build the corpus. In the first stage, a pilot dataset of 10,000 English-dominant tweets from X was compiled using the official application programming interface. Boolean queries targeting common digital discourse markers were used to retrieve posts likely to contain abbreviations, slang, emojis, and other features of interest. This pilot corpus was used to refine the coding scheme and to test annotation procedures (see Appendix C).

In the second stage, the corpus was expanded to include seven platforms: X, Instagram, TikTok, Reddit, WhatsApp, Discord, and Telegram. For each platform, 10,000 public posts were collected between January and December 2025, yielding a corpus of 70,000 posts. Where possible, platform-specific interfaces and export tools were used; in all cases, only public or ethically accessible group content was included, and private messages were excluded unless explicit permission was granted. Language identification tools were applied to ensure that English was the dominant language, while allowing for some code-mixing in multilingual contexts.

To mitigate temporal and topical skew, posts were sampled using monthly quotas for each platform, with approximately equal numbers drawn from each month of 2025. Where APIs allowed, random samples were taken from public timelines; where that was not feasible (e.g., for some WhatsApp exports and Discord servers), posts were sampled from public or consenting channels using time-bounded windows and, where appropriate, topic or hashtag filters. As a result, the corpus reflects a mixture of "ambient" everyday talk, topical conversations around hashtags or events, and more institutional discourse such as organizational announcements and news.

Engagement-based bias is a known issue in social-media corpora. To address this, posts were not exclusively sampled by engagement; rather, high-engagement posts were included only as a supplementary stratum to ensure that highly visible pragmatic practices (such as viral memes or widely shared crisis messages) were represented. In the analysis, engagement was included as a covariate, and both unweighted and engagement-weighted rates were reported to make this potential bias visible. While the corpus cannot claim to be fully representative of any platform, these procedures support meaningful comparative analysis of major platform-level tendencies.

Pre-processing and validation

Pre-processing was conducted using established natural language processing tools. Stop words and URLs were removed, while linguistically significant elements such as hashtags, mentions, emojis, and orthographic cues were retained. A Unicode-aware tokenizer ensured that emojis and clippings were treated as distinct tokens. Language identification and token normalization procedures were validated on a stratified sample of 2,100 posts (300 per platform), which were manually annotated by three coders. Accuracy scores exceeded 90% for language identification and 92% for abbreviation expansion and emoji mapping, with ambiguous cases resolved through adjudication. The emoji sense list and

abbreviation mappings used in this process are summarized in Appendices B and A, respectively, with full mappings documented in the Appendix C folder.

Error analysis indicated that most misclassifications occurred in very short posts (one or two tokens) and in heavily code-mixed messages. Such cases were also pragmatically complex, often consisting of reaction-only posts (e.g., a single emoji or interjection), where multimodal and contextual cues carry the bulk of the illocutionary force. For this reason, low-confidence posts were manually inspected and either corrected or excluded from quantitative analyses, and sensitivity checks were run with and without these items.

Coding scheme and annotation

To operationalize the theoretical framework, each post was treated as a unit of analysis and coded for up to seven categories: lexical innovation, semantic shift, syntactic or grammatical change, typographic grammar, multimodal tokens, abbreviations or clippings, and pragmatic or discourse features. These categories were defined in an annotation manual that specified inclusion criteria, subtypes, and examples (see Appendix C). For instance, lexical innovation covered new or repurposed lexemes, often associated with particular communities; semantic shift involved established words used with new meanings; typographic grammar captured meaningful patterns of capitalization, spacing and punctuation; multimodal tokens included emojis, GIFs, stickers and reaction images used as integral parts of meaning; pragmatic features included vocatives, disclaimers, tone tags and other devices that manage interpersonal relations and stance.

Each category was explicitly linked to pragmatic functions. Lexical innovations were treated as potential markers of identity and alignment with particular communities; semantic shifts were seen as sites where shared pragmatic inferences have been conventionalized; syntactic variants were examined for their role in constructing informality or emphasis; typographic grammar was interpreted as a written analogue of prosody; multimodal tokens were analyzed as IFIDs, evaluative devices, or stance markers; abbreviations were examined both as efficiency devices and as in-group signals; and pragmatic features (PRAG) captured explicit face-work, mitigation, and stance-taking constructions. This functional mapping informed both the annotation and the subsequent corpus-pragmatic interpretations (see Appendix C.6).

An annotation team of twelve coders was trained on the manual and applied the scheme to a pilot set of 7,000 posts, evenly distributed across platforms. Intercoder agreement was assessed using Krippendorff’s alpha, which reached satisfactory levels overall and by category. Disagreements were resolved by a senior coder, and the manual was refined accordingly.

Table 3
Annotation Team and Reliability Pilot

Item	Value/Result
Annotation team size	12 annotators
Pilot sample size	7,000 posts (1,000 per platform × 7 platforms)
Inter-coder reliability (Krippendorff's alpha) overall	0.82
Alpha by category – Lexical	0.85
Alpha by category – Semantic	0.81
Alpha by category – Syntactic	0.78
Alpha by category – Typographic	0.80
Alpha by category – Multimodal	0.86
Alpha by category – Abbrev/Acronym	0.88
Alpha by category – Pragmatic	0.79
Reconciliation protocol	Double annotation + adjudication by senior coder

Table 3 summarizes the reliability statistics from the pilot annotation. Overall alpha was 0.82, with particularly high agreement for abbreviations and multimodal tokens, indicating that the coding scheme was applied consistently across annotators. The full 70,000-post corpus was then annotated, with occasional double coding for quality control. (see Appendix C.7).

Data Collection Tools

The qualitative strand consisted of semi-structured interviews (see Appendix G) with 50 academic experts in linguistics/applied linguistics and digital media/communication (see Table 4), and questionnaires administered to 160 participants representing different roles: academic researchers, EFL teachers in primary, secondary, and higher education, curriculum designers, content creators, community moderators, and EFL learners. Participants were purposively sampled through institutional networks and professional contacts. Interviews were conducted online, recorded with consent, and transcribed verbatim. Questionnaires combined Likert-scale items with open-ended questions about perceptions of digital English, classroom practices, challenges, and needs (see Appendix D and E).

Table 4
Demographic distribution of the academic expert interview participants (N = 50)

Linguistics experts (n = 31)	Experience			Ranks			Affiliation	
	5–10 years	10–15 years	> 15 years	Assistant Professor	Associate Professor	Professor	Al-Baha University	Other
	11	**15**	**5**	**14**	**10**	**7**	**20**	**11**
Digital media experts (n = 19)	5	9	5	6	7	6	6	13

The 50 academic experts are all PhD holders and include slightly more linguistics specialists (n = 31) than digital media specialists (n = 19), with the majority across both groups having 10–15 years of experience. Rank distributions are relatively balanced, but assistant professors are somewhat more represented in linguistics, while associate and full professors are proportionally stronger in digital media. Al-Baha University accounts for just over half of the linguistics experts and about one-third of the digital media experts, with the remainder affiliated with other universities, ensuring a mix of local and external perspectives.

Table 5
Participant Experience, Teaching Context, and Platform Familiarity (N = 160)

Experience	Count	Teaching/Practice Context	Average Platforms Used
< 5 years	24	Secondary classrooms / private schools	3.8
5–10 years	56	University; language institutes	4.2
> 10 years	80	National curriculum; teacher training	3.7
Total/Mean	160		3.9

As summarized in Table 5, half of the participants have more than 10 years of experience, and, on average, respondents report using about 4 platforms regularly, indicating both substantial professional experience and broad digital engagement.

Data analysis

Corpus data were analyzed using descriptive statistics to calculate per-platform and overall frequencies and rates for each category. Cross-platform comparisons were made to identify patterns, such as which features were particularly associated with certain platforms. Questionnaire responses were summarized using means, standard deviations, and percentage agreement, and open-ended responses were analyzed thematically. Interview transcripts were coded for recurring pragmatic themes (e.g., stance, politeness,

alignment, offense, register awareness). Corpus findings and qualitative themes were then integrated in a corpus-pragmatic interpretation that links distributional patterns of ABB, MM, TYPO, and PRAG to their functions in interaction and to participants' perceptions (see Appendix E, Tables E1–E4).

In addition to descriptive statistics, selected inferential tests were conducted. Chi-square tests compared the prevalence of ABB, MM, and SYN across platforms, and mixed-effects logistic regressions were used to model the probability of each feature as a function of platform, post length, and engagement, with random intercepts for user ID where available. For count data such as emoji frequencies, negative-binomial models with offsets for post length were estimated. Time-trend analyses were conducted on quarterly aggregates (see Appendix F, Table F1) to assess whether ABB usage exhibited significant upward or downward trends across platforms.

RESULTS AND DISCUSSION

The Results and Discussion are presented together to facilitate interpretation. The first part examines the distribution of coded categories across platforms and addresses the first two research questions about how and where digital communication influences English. The second part considers stakeholder perspectives and discusses implications for EFL pedagogy in light of the theoretical framework.

Cross-platform patterns of digital pragmatics

Table 5
Feature distribution by platform (N = 10,000 posts per platform)

Platform	ABB count	ABB/100 posts	MM count	MM/100 posts	SYN count	SYN/100 posts
X	6,200	62.0	2,400	24.0	420	4.2
Instagram	4,300	43.0	5,000	50.0	330	3.3
TikTok	3,800	38.0	4,900	49.0	410	4.1
Reddit	2,400	24.0	1,200	12.0	540	5.4
WhatsApp	5,600	56.0	3,700	37.0	600	6.0
Discord	3,900	39.0	3,100	31.0	450	4.5
Telegram	4,000	40.0	2,800	28.0	320	3.2
Total	30,200	43.1	23,100	33.0	3,070	4.4

Note. ABB = abbreviations, acronyms, and clippings; MM = multimodal tokens; SYN = syntactic or grammatical change. Rates are per 100 posts. Categories are non-exclusive, as posts may contain multiple features.

From a corpus-pragmatic perspective, Table 5 shows that abbreviations, acronyms, and clippings are by far the most frequent phenomena, occurring in 43.1 posts per 100 across the corpus, with particularly high rates on X and WhatsApp, consistent with the confessionalization of online writing described by [2, 3] and the concentration of change in non-syntactic layers as noted [14]. A ranked list of the most frequent internet slangs and abbreviations, together with glosses and primary category labels, is provided in Appendix A. Examples include conventional abbreviations such as “LOL”, “BRB”, and “IDK”, as well as numeronyms and creative clippings. This confirms that brevity and efficiency are pervasive pressures in digital interaction [15, 16]. A supplementary quarterly breakdown of ABB rates by platform (Q1–Q4 2025) is provided in Appendix E and Appendix F, Table F1, which suggests modest but consistent upward trends on several platforms, particularly X and TikTok. However, abbreviations also function pragmatically as markers of alignment and in-group membership and can create barriers for those unfamiliar with them. Teachers and learners reported both benefits and difficulties, noting that many learners understand full forms but struggle to interpret less transparent items such as “IYKYK” or “POV”, a pattern that echoes the observation in [10] that heavily abbreviated forms can deepen digital divides when users are unfamiliar with niche acronyms.

(1) X: “brb, meeting rn, ttyl” (be right back, right now, talk to you later) — here ABB compresses turn-taking, signaling temporary withdrawal while preserving interpersonal rapport.

(2) WhatsApp: “pls send the rpt by EOD thx” — “EOD” functions as a workplace shorthand that indexes efficiency and shared professional norms; “thx” softens the imperative.

(3) Discord: “gg, that was wild fr” — “gg” (“good game”) and “fr” (“for real”) both signal alignment and positive evaluation within a gaming community.

In such cases, abbreviations do not merely save keystrokes but convey stance and in-group membership.

Do likewise (briefly) for LEX (slang items as stance), MM (emoji strings as mitigation or humor), TYPO (all caps, ellipsis), PRAG (e.g., “no offense but…”, “/s”), and SYN (e.g., ellipsis “because science”).

Lexical innovations account for 29.6% of coded instances and are especially frequent on TikTok and Discord [17, 18]. These include slang items and neologisms such as “bussin”, “rizz”, and “sus”, many of which originate in specific communities or varieties, including African American English, gaming culture, and fandoms. The rapid circulation of such items across platforms illustrates how digital environments facilitate bottom-up lexical change [6]. At the same time, expert and teacher comments highlight concerns about the volatility of this slang, which can quickly become outdated, and about the unacknowledged appropriation of forms from minoritized communities into global English. From a pragmatic perspective, lexical innovations serve as indexical markers of identity, solidarity, and stance, but their meanings and sociocultural implications are not always transparent to EFL learners. Appendix A flags items that are dual-coded as LEX or SEM in addition to ABB, illustrating these overlaps.

Multimodal tokens, comprising emojis, GIFs, and stickers, appear in 33.0 percent of coded instances, with the highest densities on Instagram and TikTok and substantial usage on X, WhatsApp, and Discord [8]. These visual forms routinely perform pragmatic work: softening imperatives, intensifying evaluations, conveying sarcasm, signaling shared amusement, or positioning the writer in relation to others [9, 10]. The corpus contains many examples where emojis replace or supplement words, effectively functioning as lexical or grammatical elements. Sequences of emojis sometimes form narrative-like strings, consistent with the idea of a visual grammar [11]. Appendix B lists the most frequent emojis in the pilot subset and notes their dominant pragmatic senses, which informed the MM role attribute in the annotation scheme (Appendix C.3.5). Stakeholders uniformly stressed the importance of these resources for expressing tone, but learners also reported confusion about context-specific meanings. For example, the skull emoji is interpreted by some users as indicating extreme laughter, but its literal association with death can puzzle others. In corpus-pragmatic terms, these multimodal tokens function as recurrent, patternable resources that serve roles comparable to those of discourse markers and IFIDs in spoken interaction.

Semantic shifts are common in the corpus and are particularly evident on platforms where professional and informal registers intersect, such as Reddit and WhatsApp. Words such as “ping”, “bandwidth,” and “ghost” are used with extended meanings related to messaging, cognitive capacity, and disengagement. Pragmatic use of intensifiers such as “literally” in non-literal ways illustrates semantic bleaching. These shifts illustrate how meanings are co-constructed through recurrent usage and shared contexts, but they can also lead to misunderstandings when interlocutors interpret words in more traditional senses.

Typographic grammar and pragmatic markers are pervasive across platforms. By treating these typographic patterns as a dedicated annotation category (TYPO), the study makes them available for corpus-pragmatic analysis, showing how users systematically exploit written form to perform pragmatic work. Users rely on capitalization, punctuation, and spacing to convey a typographic tone of voice [7, 19]. All-caps sequences often signal shouting or excitement; multiple exclamation marks or question marks convey strong emotion or incredulity; a deliberate lack of capitalization can index informality or irony. Pragmatic markers such as “no offense but”, “girl…” and explicit tone tags like “/s” and “/j” are widely used to manage potential face threats and to signal joking or sincerity. These resources support nuanced pragmatic work in short texts, but they require shared knowledge of conventions and can be difficult for learners to decode. The PRAG category in the annotation manual (Appendix C.3.7) captures such mitigation and stance devices, while TYPO attributes (Appendix C.3.4) record their pragmatic functions.

In contrast, syntactic or grammatical change is relatively rare, accounting for only 4.4% of coded instances. Where it does occur, it often takes the form of ellipsis, subject drop, fragmentary “Because X” constructions, or creative part-of-speech shifts. These patterns suggest that, in this 2025 corpus, core clause structure appears comparatively stable and that most innovation occurs in the distribution of resources around that structure. From a theoretical perspective, this supports the view that digital English is not characterized by wholesale grammatical breakdown, but rather by a reallocation of

pragmatic work to lexicon, orthography, and multimodality [14]. Decision rules for coding patterns as SYN, including the requirement for recurrence across users and contexts, are detailed in Appendix C.3.3 and C.5.

Mitigation and politeness.

Across platforms, users often combine lexical, typographic, and multimodal resources to mitigate potentially face-threatening acts. Requests on WhatsApp, for example, frequently include ABB or clipped forms (“pls”, “thx”) together with emojis (e.g.) and hedging adverbials (“maybe”, “if you don’t mind”). On X and Instagram, criticism is often buffered by disclaimers such as “no offense but…” and softened through laughing or smiling emojis. These combinations function as complex pragmatic packages where no single element carries the full illocutionary load; corpora allow us to identify the recurrent co-occurrence of such patterns.

Irony and sarcasm.

Irony and sarcasm are often marked by specific lexical items (“sure, Jan”), typographic cues (scare quotes, capitalization), and emojis (e.g. ,), sometimes supplemented by explicit tone tags such as “/s”. On Reddit, sarcastic responses frequently use deadpan lexical content combined with typographic emphasis, whereas on Discord, users rely more heavily on emojis and GIFs. The corpus shows that explicit tone tags remain relatively rare compared to implicit cues, suggesting that readers must still infer ironic intent from distributional patterns of these pragmatic markers.

Affiliation and alignment.

Expressions of alignment, such as “this,” used as a standalone comment, or emphatic agreement via “same”, “so true”, and reaction emojis, are particularly visible on platforms like TikTok and Instagram, where commenting systems foreground such responses. The corpus reveals that simple tokens like “same” frequently co-occur with particular emojis (e.g. ,), forming multimodal sequences that index strong affective alignment.

Offense and conflict.

While the main focus of this study is not online aggression, some patterns related to disagreement and offense emerged. For instance, the use of all caps and repeated punctuation in conjunction with negative lexical items can function as escalation resources, whereas adding a laughing emoji or “jk” (“just kidding”) can serve as a retroactive softener. Content creators and moderators reported that misreading these combinations is a common source of conflict, especially for users less familiar with the community’s pragmatic norms.

Participants’ perspectives and pedagogical implications

The questionnaire and interview data from academic researchers, teachers, content creators, moderators, and learners provide insight into how digital English is experienced and evaluated in practice. Academic respondents overwhelmingly agreed that digital communication influences English at multiple levels beyond the lexicon and supported treating multimodal resources as part of digital grammar [1]. Many argued that a distinct digital English register can be described and that EFL learners need explicit instruction in its features, provided that this is done contrastively to preserve awareness of formal academic norms (see Appendix E, Table E1), a position consistent with calls in digital literacies and TESOL scholarship for explicit work on register and multimodality [4, 20, 21].

EFL teachers reported that their students regularly use abbreviations, emojis, and slang in and out of class and that social media use often supports vocabulary development, fluency, and confidence. At the same time, teachers observed that digital features sometimes appear in formal writing, such as essays and emails, where they are considered inappropriate. Levels of teacher confidence in explaining differences between digital and formal English varied, and many respondents expressed a need for training and materials to help them integrate digital registers into teaching while managing risks. There was strong interest in corpus-informed resources based on authentic posts from X and other platforms, particularly when accompanied by guidance on ethical use and register comparison (see Appendix E, Table E2), in line with findings that social-media-based tasks can enhance learners' engagement and performance when pedagogically framed [22–24].

Content creators and moderators (see Appendix E, Table E3) described unwritten norms governing language use in their communities and emphasized that misunderstandings frequently arise when digital cues are misread. They viewed emojis, GIFs, and abbreviations as essential tools for conveying tone and believed that EFL learners would benefit from exposure to community-specific practices, provided that ethical issues regarding anonymization and consent were respected. They also felt that current teaching materials rarely reflect the realities of social media English [5, 19].

Learners reported that they often use digital English in their online communication and believe that doing so helps their vocabulary, confidence, and reading skills. Many acknowledged that they sometimes find English on social media difficult to interpret because of slang, memes, and emojis, and that they occasionally carry digital features into formal writing unintentionally. They expressed a strong desire for classroom activities based on real social media content, explicit explanations of abbreviations, slang, and emojis, and guidance on when digital English is appropriate and when more formal styles are required (see Appendix E, Table E4), which underscores the need for explicit digital register awareness in EFL curricula [4].

Taken together, these perspectives suggest that digital English is already central to learners' linguistic lives and that both teachers and learners see value in treating it as a legitimate object of study. At the same time, they underscore the need for a critical, pragmatically informed approach to pedagogy. Integrating X's multimodal features into EFL teaching should not mean simply reproducing online

practices, but rather helping learners develop the ability to interpret and manage pragmatic meaning across contexts and registers.

One implication is that EFL classrooms can use authentic or adapted posts from X and comparable platforms to teach pragmatic competence through multimodal analysis. Learners can examine how combinations of text, emojis, and typographic choices shape perceived politeness, humor, and stance, and then practice rewriting posts for different audiences or purposes. Such tasks make visible the pragmatic work that digital features perform and can support learners in developing a flexible repertoire. Another implication is the importance of teaching digital register awareness, including discussions of the social meanings and ethical dimensions of forms such as AAVE-derived slang, as well as the risks of exclusion or offense when these are used without understanding.

Finally, the findings point to the need for teacher education that includes digital pragmatics and digital literacies. Teachers need opportunities to familiarize themselves with contemporary digital practices, interpret corpus-based evidence, and design tasks that leverage learners' existing digital engagement while fostering critical reflection and control [25, 26]. In this sense, the study supports calls for language education that prepares learners not only to use English accurately but also to navigate the complex pragmatic ecologies of digitally mediated communication.

CONCLUSIONS

This study examined how digital communication influences English and what these changes mean for the teaching of English as a foreign language. By combining a large, annotated corpus of posts from seven platforms with data from experts, teachers, content creators, moderators, and learners, it showed that digital English is best understood as a multimodal pragmatic repertoire rather than as a simple collection of slang terms or deviations from standard norms. Abbreviations, lexical innovations, emojis, typographic patterns, and pragmatic markers emerged as central resources for managing efficiency, stance, and interpersonal relations online, while core syntactic structures remained comparatively stable.

Regarding the first two research questions, the findings indicate that digital communication has its strongest impact on lexical, orthographic, typographic, and multimodal components, with abbreviations and multimodal tokens being particularly prominent. Semantic shifts and pragmatic routines also play a crucial role in organizing meaning, whereas syntactic change appears limited and largely confined to stylized ellipsis and fragmentary constructions. This pattern suggests that digital platforms redistribute pragmatic work across a wider set of semiotic resources without fundamentally destabilizing the grammatical backbone of English.

In relation to the third research question, the study showed that these developments can be productively incorporated into EFL pedagogy when approached through the lens of digital pragmatics and digital literacies. Stakeholders across roles recognized that digital English is already central to learners' communicative repertoires and that classroom engagement with authentic or adapted social-media discourse can support motivation, vocabulary growth, and confidence. At the same time, they

emphasized the need for explicit work on register, pragmatic appropriateness, and ethical issues, including the cultural origins of certain forms and the risks of misinterpretation or exclusion. The study therefore recommends that EFL curricula and teacher education explicitly incorporate digital registers, multimodal analysis, and pragmatic awareness as learning goals, and that corpus-based materials drawn from platforms such as X be used to design tasks which contrast digital and formal varieties and foster critical reflection.

The study has several limitations. The corpus, while large and multi-platform, represents a single year of data and focuses on English, so further research is needed to track diachronic change and to compare patterns across languages and varieties. The annotation scheme, though reliable, cannot capture all the nuances of pragmatic meaning, and some category boundaries remain approximate. In addition, the qualitative sample, although diverse, does not fully represent less connected or lower-resourced learner populations. Future work could combine corpus-based analysis with ethnographic and interactional approaches in specific communities, incorporate more systematic analysis of AI-mediated communication, and extend the pedagogical dimension through classroom intervention studies that evaluate the effects of digital-pragmatics-informed tasks on learners' performance.

Furthermore, although the annotation scheme captures a range of pragmatic phenomena, many aspects of digital interactional work remain beneath the surface. For example, irony, play, and affect are sometimes only recoverable through extended conversational context, shared history, or visual media that were not fully preserved in the corpus. Even with multimodal and typographic categories, some pragmatic meanings can therefore only be approximated. A more interactionally oriented corpus pragmatics, combining detailed thread-level reconstruction with sequential analysis, could shed further light on how pragmatic moves unfold over time. Future research might also compare different language pairs or varieties, investigate multilingual digital repertoires, and apply similar coding schemes to spoken corpora of digitally mediated talk (e.g., voice messages) to explore how online pragmatic norms bleed back into speech.

Overall, the study advances the understanding of how digital communication is reshaping English by showing that change is concentrated in a multimodal pragmatic layer that sits atop a relatively stable grammatical core. It also demonstrates that engaging with this layer is essential for contemporary EFL teaching, not only to reflect learners' lived linguistic practices but also to equip them to navigate complex digital environments in informed and responsible ways. More broadly, the study illustrates how corpus-pragmatic approaches can capture the multimodal repertoires of digital English and provide an empirical basis for updating descriptions of English usage and for designing pragmatics-oriented EFL pedagogy.

Funding Declaration

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Declaration

Not applicable.

Declarations

Author Contribution

The author is solely responsible for the conceptualization, study design, corpus construction, data collection, analysis, and interpretation, as well as for drafting, revising, and approving the final manuscript.

Acknowledgement

I am grateful to all participants, including teachers, learners, content creators, moderators, and academic experts, for generously sharing their time and insights. I also thank colleagues who provided feedback on the research design and interpretation of findings.

References

1. Salfin, S., Kurniadi, P., & Erwin, E. (2024). Language development in the digital age: a literature review on the influence of technology on human communication. *Sciences Du Nord Humanities and Social Sciences*, *1*(1), 1–7. https://doi.org/10.58812/0yhk5d80
2. Crystal, D. (2001). *Language and the internet*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164771
3. Crystal, D. (2011). *Internet linguistics: a student guide*. Routledge. https://doi.org/10.4324/9780203830901
4. Jones, R. H., & Hafner, C. A. (2021). *Understanding digital literacies: a practical introduction* (2nd ed.). Routledge. https://doi.org/10.4324/9781003177647
5. Zappavigna, M. (2022). Social media quotation practices and ambient affiliation: weaponising ironic quotation for humorous ridicule in political discourse. *Journal of Pragmatics*, *191*, 98–112. https://doi.org/10.1016/j.pragma.2021.12.003
6. Lyons, A., & Tagg, C. (2024). Post-digital connectivities: framing offline encounters in a digital prospection space. *Applied Linguistics*, *45*(5), 829–848. https://doi.org/10.1093/applin/amae008
7. McCulloch, G. (2023). *Because internet: understanding the new rules of language*. Updated edn. Riverhead Books, New York.
8. McShane, L., Pancer, E., Poole, M., & Deng, Q. (2021). Emoji, playfulness, and brand engagement on Twitter. *Journal of Interactive Marketing*, *53*, 96–110. https://doi.org/10.1016/j.intmar.2020.06.002
9. Uzan, S., Freud, D., & Elalouf, A. (2025). Optimizing chatbots to improve customer experience and satisfaction: research on personalization, empathy, and feedback analysis. *Applied Sciences*, *15*(17), 9439. https://doi.org/10.3390/app15179439

10. Al-Rawi, A. (2025). Emojified communication: how do news organizations use emojis on social media? *Telematics and Informatics Reports*, *19*, 100233. https://doi.org/10.1016/j.teler.2025.100233
11. Cohn, N., Engelen, J., & Schilperoord, J. (2019). The grammar of emoji? Constraints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications*, *4*, 33. https://doi.org/10.1186/s41235-019-0177-0
12. Martins, A., Londral, A., Nunes, I. L., & Lapão, L. V. (2024). Unlocking human-like conversations: scoping review of automation techniques for personalized healthcare interventions using conversational agents. *International Journal of Medical Informatics*, *185*, 105385. https://doi.org/10.1016/j.ijmedinf.2024.105385
13. Kaibel, G. (2026). Constructing expertise through questions?—Potentials of corpus pragmatic regression analyses. *Corpus Pragmatics*, *10*, 19. https://doi.org/10.1007/s41701-025-00206-4
14. Ovsiienko, L., Kulyk, O., Kardash, L., Makarchuk, O., & Boyko, S. (2025). Language shifts in the digital age: the transformation of lexical and grammatical structures under the influence of social networks. *Metaverse Basic and Applied Research*, *4*, 180. https://doi.org/10.56294/mr2025180
15. Joseph, A., & Devi, S. (2025). The evolution of English language in the digital era. *International Journal of Social Impact*, *10*(3), 62–72. https://doi.org/10.25215/2455/1003012
16. Ekayati, R., Sibarani, B., Ginting, S. A., Husein, R., & Amin, T. S. (2024). Digital dialects: the impact of social media on language evolution and emerging forms of communication. *International Journal of Educational Research Excellence*, *3*(2), 605–609. https://doi.org/10.55299/ijere.v3i2.986
17. Wan Mohd Muzani, A., & Mohammad Lotfie, M. (2024). Morphological neologisms: the emergence of social media slang on TikTok. *Sains Insani*, *9*(1), 103–113. https://doi.org/10.33102/sainsinsani.vol9no1.597
18. Liu, W., Zhao, X., Zhan, M., & Hernandez, S. (2024). Streaming disasters on TikTok: examining social-mediated crisis communication, public engagement, and emotional responses during the 2023 Maui wildfire. *Public Relations Review*, *50*(5), 102512. https://doi.org/10.1016/j.pubrev.2024.102512
19. Han, Y. (2024). The impact of digital media on language styles and communication methods based on text, image, and video forms. *Lecture Notes in Education Psychology and Public Media*, *40*(1), 211. https://doi.org/10.54254/2753-7048/40/20240754
20. Grewal, D., Herhausen, D., Ludwig, S., & Ordenes, F. V. (2022). The future of digital communication research: considering dynamics and multimodality. *Journal of Retailing*, *98*(2), 224–240. https://doi.org/10.1016/j.jretai.2021.01.007
21. Lee, T. K., & Li, W. (2025). Refashioning linguistic expertise: translanguaging TESOL in social media. *Applied Linguistics*. https://doi.org/10.1093/applin/amaf007
22. Chen, Y., & Hsu, L. (2022). Enhancing EFL learners' self-efficacy beliefs of learning English with emoji feedbacks in CALL: why and how. *Behavioral Sciences*, *12*(7), 227. https://doi.org/10.3390/bs12070227
23. Abdel Al Ibrahim, K. A. A., & Basim, N. (2024). Applying task-supported language learning via Instagram: effects on EFL learners' self-esteem, academic buoyancy, engagement, and

achievement. *Asian-Pacific Journal of Second and Foreign Language Education*. https://doi.org/10.1186/s40862-023-00223-3

24. Sancho Ortiz, A. E. (2024). Analyzing the functionality of Twitter for science dissemination in EFL teaching and learning. *Profile: Issues in Teachers' Professional Development*, *26*(1), 133–147. https://doi.org/10.15446/profile.v26n1.105685
25. Feng, L., & Sumettikoon, P. (2024). An empirical analysis of EFL teachers' digital literacy in Chinese higher education institutions. *International Journal of Educational Technology in Higher Education*, *21*, 42. https://doi.org/10.1186/s41239-024-00474-1
26. Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: a scoping literature review. *Computers and Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.pdf