

Research article

UDC 81'42 + 004.85

DOI: <https://doi.org/10.18721/JHSS.16410>

EDN: <https://elibrary/LYKUVR>



ENHANCING LLM INTERPRETATION OF APPRAISALS IN SPANISH DIGITAL DISCOURSE

S. Sikorskii , M.L. Carrio-Pastor 

Polytechnic University of Valencia, Valencia, Spain

 ssikors@upv.es

Abstract. This article examines the challenges Large Language Models (LLMs) face in interpreting evaluative language in digital discourse. These models often distort the semantics of evaluative expressions, hindering their accurate linguistic interpretation. The aim of the study is to determine whether integrating specialised knowledge improves a model's ability to correctly identify and classify evaluative meanings. Methodologically, the work draws on Martin and White's Appraisal Theory and includes an experimental evaluation of GPT-4 on a stratified corpus of posts from a social network. The analysis is conducted in two conditions – with and without external contextual knowledge – and the results are compared to expert annotations in terms of precision and recall. The findings demonstrate a substantial improvement in automatic classification: the accuracy of identifying evaluative categories increases, the range of detectable appraisal elements expands, and new patterns of meaning variation emerge. The conclusions emphasise that enriching LLMs with structured knowledge enhances the reliability of evaluative language analysis and provides deeper insight into how such meanings function in digital discourse. The proposed approach opens new avenues for improving automated methods for analysing evaluative meanings in linguistic research.

Keywords: appraisal analysis, context integration, Large Language Models, evaluative language, social media discourse.

Acknowledgements: Project CIAICO/2023/187, funded by the Conselleria d'Innovació, Universitats, Ciència i Societat Digital.

Citation: Sikorskii S., Carrio-Pastor M.L., Enhancing LLM Interpretation of Appraisals in Spanish Digital Discourse, Terra Linguistica, 16 (4) (2025) 159–184. DOI: 10.18721/JHSS.16410



АНАЛИЗ ОЦЕНОЧНОЙ ЛЕКСИКИ В ИСПАНОЯЗЫЧНОМ ЦИФРОВОМ ДИСКУРСЕ: ПРЕОДОЛЕНИЕ ОГРАНИЧЕНИЙ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

С. Сикорский , М.Л. Каррио-Пастор 

Политехнический университет Валенсии, Валенсия, Испания

 ssikors@upv.es

Аннотация. В статье рассматривается проблема понимания оценочной лексики большими языковыми моделями (БЯМ) в цифровом дискурсе. Отмечается, что языковые модели нередко искажают семантику оценочных высказываний, что затрудняет их корректную лингвистическую интерпретацию. Цель исследования — определить, улучшает ли интеграция специализированного знания способность модели точно идентифицировать и классифицировать оценочные значения. Методологически работа опирается на теорию оценки Мартина и Уайта и включает экспериментальную проверку модели GPT-4 на стратифицированном корпусе сообщений в социальной сети. Анализ проводится в двух режимах — с использованием внешнего знания и без него — с последующим сопоставлением результатов с экспертной разметкой по показателям точности и полноты. Результаты демонстрируют существенное улучшение качества автоматической классификации: повышается точность определения оценочных категорий, расширяется спектр обнаруживаемых элементов, выявляются новые паттерны смысловой вариативности. В выводах подчеркивается, что обогащение БЯМ структурированным знанием повышает надежность анализа оценочной лексики и позволяет глубже понять механизмы ее функционирования в цифровом дискурсе. Предложенный подход открывает перспективы для совершенствования автоматизированных методов анализа оценочных значений в лингвистических исследованиях.

Ключевые слова: анализ оценочной речи, интеграция контекста, большие языковые модели, оценочная речь, дискурс в социальных сетях.

Финансирование: Project CIAICO/2023/187, funded by the Conselleria d'Innovació, Universitats, Ciència i Societat Digital.

Для цитирования: Сикорский С., Каррио-Пастор М.Л. Анализ оценочной лексики в испаноязычном цифровом дискурсе: преодоление ограничений больших языковых моделей // Terra Linguistica. 2025. Т. 16. № 4. С. 159–184. DOI: 10.18721/JHSS.16410

Introduction

Large language models (LLMs) have revolutionized natural language processing through exposure to vast text corpora. LLMs develop in-context learning capabilities that enable them to generate contextually relevant responses without explicit task-specific training [1, 2]. However, these models exhibit persistent cultural biases in analysing evaluative language, often defaulting to anglophone normative frames that overlook how context shapes non-English attitudes [3, 4]. This limitation is particularly acute for non-English political and crisis discourse, where evaluative meaning is deeply entwined with local institutional norms, cultural values, and historical narratives.

Spanish, the world's second most widely spoken native language (L1), constitutes a major linguistic domain for digital political communication across Europe and Latin America, making it a critical test case for cross-cultural appraisal interpretation. Far less is known about how LLMs handle appraisal in major non-English digital spheres, despite the centrality of Spanish in global online crisis communication.



This study addresses these limitations through Martin and White's Appraisal Theory [5], rooted in Systemic Functional Linguistics. The framework distinguishes three domains of evaluative meaning – Attitude, Engagement, and Graduation – and is especially sensitive to Judgment subcategories such as propriety (ethical norms) and tenacity (persistence under adversity). These categories are crucial in crisis discourse, where institutional behaviour is evaluated against culturally embedded expectations [6, 7].

Context profoundly mediates evaluative interpretation, transforming neutral expressions into loaded assessments based on shared cultural frameworks [8, 9]. For instance, "traditional" may invoke positive propriety in conservative Spanish political talk but negative judgment in progressive critiques, highlighting polarity shifts via contextual interplay [10, 11]. Empirical studies demonstrate that evaluation processing changes when contextual cues are manipulated [12], and that culturally variant texts cannot be reliably interpreted through universalist or English-centric frameworks [13–15].

Recent assessments of LLMs indicate similar issues: cultural misalignment in non-English settings [3], inconsistent value alignment across languages [16], and insufficient sensitivity to culturally specific evaluative cues. Although multilingual strategies such as fine-tuning or prompt engineering [17–19] improve surface performance, they do not address the deeper problem that evaluative meaning is constituted – not merely modified – by cultural knowledge, especially in politically charged Spanish digital discourse [4].

To address this gap, the present study focuses on Spanish-language crisis communication during the 2024 DANA Valencia floods, using posts from X.com, Spain's most active platform for real-time political commentary and institutional accountability. X provides dense evaluative microtexts ideal for appraisal analysis; its use here is strictly academic¹.

To clarify the aim of this study, we formulate the following overarching research question: How does structured, multidimensional contextual modelling influence LLM-based appraisal interpretation in Spanish digital crisis discourse?

To address this, we investigate three focused research questions:

1. How does multidimensional cultural context integration via graph databases enhance LLM performance in Spanish appraisal analysis, particularly Judgment detection?
2. What systematic evaluation patterns in Spanish crisis discourse require cultural context for accurate interpretation, challenging English-centric frameworks?
3. Does structured context injection yield superior cross-cultural accuracy compared to baseline approaches?

Using 200 stratified Spanish posts about the 2024 DANA crisis, we compare GPT-4's performance with and without context injection. By integrating historical, social, political, cultural, and digital contextual metadata into a graph database, we show that cultural knowledge is constitutive of evaluative meaning, offering a scalable method for culturally responsive appraisal analysis in Spanish-speaking digital contexts.

Theoretical Background

Context Types and their Influence on Appraisal

From the diverse contextual factors discussed in the literature (e.g., [20–22]), five overarching dimensions of context are critical for cross-cultural understanding: Historical, Social, Cultural, Political, and (digital-specific) Thread Context. This selection is based on their demonstrated influence on evaluative meaning, their recurrent salience in social media interactions, and their potential for operationalization through identifiable markers (Figure 1).

Figure 1 not only lists these context types and their marker families (e.g., temporal reference, group norms, ideological stance), but also illustrates the conceptual influence relations between them such as historical context 'informing' cultural values or political structures 'guiding' social norms.

¹ In the Russian Federation, X (formerly Twitter) is blocked.

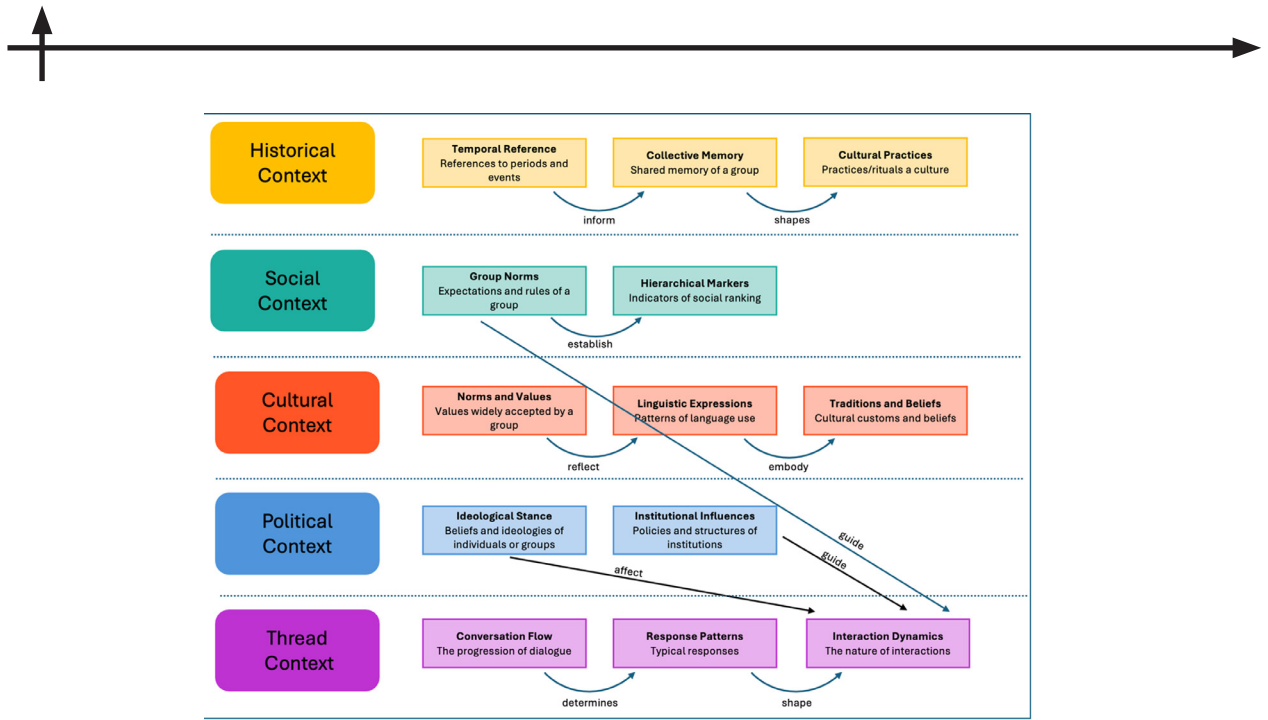


Fig. 1. Contextual integration for LLM-based appraisal analysis

These arrows represent theoretical pathways of contextual influence in discourse, rather than procedural steps in the computational method, and serve as a conceptual scaffold for the later modelling of contextual retrieval.

In this theoretical framework, different types of context contribute distinctly to the interpretation of appraisal:

Historical Context. Historical context provides a temporal dimension that anchors evaluations within broader narratives. As van Dijk [21] notes, historical context helps comprehend complex phenomena by providing a background against which communication occurs. This context shapes how judgments evaluate behaviour against past events and how affect connects to collective experiences. Historical references in digital discourse often function as evaluative shortcuts, activating shared interpretative frameworks. For instance, references to "La Transición" ("*Spain's democratic transition*") in Spanish political discourse evoke specific normative expectations that shape how political actors are judged.

Social Context. Social context encompasses relationships, roles, and social structures that frame evaluations. KhosraviNik [22] emphasizes how social factors influence the production and interpretation of discourse, from immediate interpersonal dynamics to broader institutional structures. These social dimensions shape affective expressions, condition judgment criteria, and influence appreciation standards. In Spanish digital discourse, social markers like formal address terms ("*usted*") or informal language signal particular evaluative frames that influence how attitude is expressed and interpreted.

Political Context. Political context introduces ideological frameworks and power dynamics that influence evaluative positioning. Van Dijk [21] shows how political contexts shape how speakers position themselves and others, often through strategic language use. These contexts determine which behaviours warrant positive or negative judgment and which values receive appreciation. Political markers in Spanish digital discourse (e.g., "*progresista*" vs. "*conservador*") activate distinct evaluative frameworks that influence how policies, leaders, and social phenomena are assessed.

Cultural Context. Cultural context provides shared knowledge, values, and practices that ground evaluative meaning. Manca [23] demonstrates how cultural context intertwines with social and historical contexts to determine what is considered appropriate or inappropriate in communication. Cultural norms shape how emotions are expressed, which behaviours receive positive judgment, and what



qualities merit appreciation. Spanish digital discourse often exhibits more emotionally expressive language than English equivalents, reflecting cultural differences in affective communication norms.

Digital Discourse Context. Digital discourse introduces unique contextual elements through conversation structure, platform conventions, and multimodal resources. Thread context, as examined by Diethei et al. [20], significantly impacts how evaluations evolve across conversational exchanges. Digital markers like hashtags function as contextual frames that influence evaluative interpretation, while emojis and other multimodal elements modify evaluative meaning. These digital discourse features create complex contextual configurations that shape how appraisals are expressed and understood [24–27].

These types interact dynamically, per van Dijk [21], to shape interpretative environments essential for appraisal.

Contextual Interaction Mechanisms

Building on Kecskes' (2023) Dynamic Model of Meaning, we delineate three mechanisms for evaluative shaping in digital discourse, drawing from socio-cognitive analogs (e.g., dialogic syntax, pragmatic negotiation): a) Contextual Resonance: Activates frameworks beyond literals, with historical/cultural macros conditioning others [28, 29]. In Spanish crises, "El 15-M" resonates political Judgment (tenacity as collective persistence), amplifying propriety via shared echoes [30]. b) Contextual Shifting: Stable terms shift valences via configurations [31], mediated socio-politically in dynamic digital spaces [25, 32]. "Radical" signals positive tenacity in progressive Judgment but negative propriety in conservative. c) Contextual Evolution: Threads reconfigure frames diachronically [33, 34], evolving Judgment (e.g., positive tenacity shifting to ironic propriety in accumulating replies).

These imply LLM needs: cultural access for resonance [35], dynamic parsing for shifting [36], memory for evolution [37].

Spanish Digital Discourse: Contextual Characteristics

Spanish digital discourse, prevalent in global online environments [38], serves as a case study for contextual influences on appraisal. Spanish social media exhibits heightened context dependence, with increased engagement markers compared to English equivalents [39], reflecting cultural reliance on shared knowledge. For instance, idiomatic expressions like "Esto es más largo que un día sin pan"² ("This is longer than a day without bread") invoke cultural context to express negative appreciation of duration appearing in 14 instances (7%) This idiom frequently co-occurs with political dissatisfaction, e.g. "Esto es más largo que un día sin pan con el gobierno" ("This is longer than a day without bread with this government"), where temporal appraisal blends with negative propriety.

Spanish digital discourse often mirrors spoken conventions, deviating from written norms. Lage and Diego [40], identified oralized features in chats, extended in recent crisis analyses [41] to social media, where explicit Affect predominates [42]. In the DANA corpus, this includes expressions like "¡Vaya tela con este gobierno!" ("What a mess with this government!"), which conveys negative propriety and appears in 35 instances (18% of Judgments).

Similarly, emotional reactions include:

- "Da miedo ver como gestionan esto."
"It's scary to see how they're handling this."
→ Negative Affect: insecurity.
- "Letizia con cara de 'menudo palo estar aquí'... vergüenza."
"Letizia with a face of 'what a pain to be here... shameful.'"
→ Negative Judgment: propriety.

² Examples in this section are drawn from the DANA Valencia crisis corpus used in this study; the corpus construction and sampling procedure are described in Section Method.



- "Autocontrol nivel DIOSA"

"Self-control at GODDESS level"

→ Positive Appreciation: valuation + hyperbolic Graduation.

- "borbones corruptos haciendo el paripé."

"corrupt Bourbons putting on an act."

→ Negative Judgment: veracity + propriety.

Cultural contexts frame appraisals collectivistically, emphasizing community welfare and shared responsibility [43]. Qualitative evidence from protest calls [41], identity construction [44], and indigenous online forums [45] highlights first-person plural forms "Hemos conseguido algo importante" (*"We have achieved something important"*) graduating positive appreciation via collective achievement. In the DANA corpus, this manifests in 38% of Judgments, such as "estamos juntos en esto" (*"We're in this together"*) invoking tenacity as communal persistence during floods.

- "Aquí nadie se rinde"

"Nobody gives up here"

→ Collective tenacity, aligning with crisis rhetoric [7].

- "Gracias por ir y validar el dolor del pueblo."

"Thank you for going and validating the people's pain."

→ Positive Judgment: propriety.

- "Gracias por ir y validar el dolor del pueblo."

"Thank you for going and validating the people's pain."

→ Positive Judgment: propriety.

Noteworthy are culturally specific appraisals of bravery and honour:

- "El Rey le ha echado pelotas."

"The King showed guts."

→ Positive Judgment: tenacity + capacity.

- "Se la ha bancado con honor y probidad."

"He endured it with honour and integrity."

→ Positive Judgment: propriety.

These examples demonstrate how Spanish digital discourse intertwines institutional ethics, emotional response, and communal identity. They highlight the need for context-sensitive appraisal analysis that accounts for Spanish-specific evaluative configurations, which resist universal categories and depend on institutional, cultural, and collective baselines.

Method

Corpus and Operationalization

To operationalize the framework, we fused computational context injection with LLM analysis for empirical validation.

The corpus comprises 200 X.com posts, stratified-random sampled from #DANAValencia and #Valencialnundaciones queries via X's advanced search (October 29–November 15, 2024, aligning with flood peak and royal visits). From an initial pool of ~3,000 posts, we balanced supportive ($n = 100$, keywords like "apoyo" "support", "solidaridad" "solidarity") and critical ($n = 100$, keywords like "vergüenza" "shame/disgrace", "fracaso" "failure") replies, excluding bots/reposts via metadata filters. This ensured representation of evaluative language polarity in crisis discourse.

Tweets vary considerably in length: the shortest contain 2–4 tokens, while extended argumentative posts exceed 70 tokens, with emojis treated as separate tokens following multimodal corpus conventions. The mean tweet length is approximately 22–27 tokens, with a median of 18–20 tokens. This distribution captures the natural variation in crisis-related evaluative discourse and is appropriate for the study's aim of assessing a pilot appraisal annotation workflow.



A single expert (Spanish linguist, Universitat Politècnica de València) annotated using Martin and White framework [5], operationalizing Judgment subcategories. Propriety (ethical norms) was frequently realized through lexical items such as "vergüenza" (*"shame; moral disgrace"*) in 45% of negative instances. Tenacity (persistence, resilience) was commonly expressed through idioms such as "dio la cara" (*"he/she stepped forward; faced the situation"*) in supportive replies. Given the modest corpus size ($n = 200$), a single annotator suffices for initial validation, aligning with pilot practices in discourse analysis where multi-annotator ICR is typically applied at larger scales [6, 46]. For reliability, self-consistency checks re-annotated a 20% random subset (40 posts, per standard thresholds for intra-rater agreement in small corpora – McHugh [47] recommends at least 30 comparisons for precise estimates to balance effort and detect category drift, yielding 92% agreement. Disagreements (e.g., 3 propriety-tenacity overlaps) were resolved via framework reference, ensuring stability without inflating resources.

Drawing on mechanisms introduced in Section Theoretical Background, we implemented a two-stage appraisal procedure:

Stage 1: Transparent Scaffolding – Explicitly apply categories to discover evaluations, documenting Spanish resistances (e.g., collective propriety exceeding individual-centric English frames).

Stage 2: Context-Dependent Pattern Recognition – Identifying structures shaped by Spanish discourse conventions, such as "quedar bien/mal" (*"to look good/bad"*, social-image management modulating propriety), "vergüenza ajena" (*"second-hand embarrassment"*, a collective judgment device), and references to "La Transición" (*"Spain's democratic transition"*) anchoring tenacity and legitimacy evaluations.

This operationalization integrates lexical, cultural, and institutional cues to ensure compatibility between the annotation scheme and Spanish-specific evaluative practices.

The Design of the Proposal: Contextual Markers and Patterns

The operationalization of our multi-contextual framework requires systematic identification of linguistic and semiotic markers that activate specific interpretative frameworks within each contextual dimension. Drawing on Van Dijk's [21] socio-cognitive approach to context, KhosraviNik's [22] social media discourse framework, and Martin and White's appraisal categories [5], we propose a taxonomic classification of contextual markers that function as computational cues for activating contextual knowledge essential for culturally-informed evaluative interpretation.

Our marker classification addresses a fundamental challenge in cross-cultural appraisal analysis: the systematic detection of context-dependent evaluative triggers that operate beyond surface linguistic features. Unlike approaches that treat context as supplementary background information, our framework positions contextual markers as constitutive elements that actively shape – rather than merely influence-evaluative meaning formation.

The systematic taxonomy encompasses five contextual dimensions, each containing multiple marker types with distinct identificational and functional properties. This classification enables both theoretical analysis and computational implementation by providing explicit criteria for marker detection alongside their specific roles in activating evaluative frameworks. Table 1 presents the complete taxonomic structure.

Historical and cultural markers function as macro-contextual frames that condition the interpretation of political and social markers. For instance, references to "La Transición" (*"Spain's democratic transition"*) establish historical frameworks that modify how contemporary political terminology is evaluated.

Digital discourse markers, particularly emojis and platform conventions, systematically modify the evaluative intensity established by textual markers. The combination of political hashtags with specific emojis creates evaluative configurations unpredictable from individual marker analysis.



Table 1. Taxonomic structure of context dimensions

Dimension	Marker Type	Definition	Spanish Examples	English Translation	Identification Method	Appraisal Function
Historical	Temporal References	Direct lexical references to specific historical periods, events, or chronological markers that activate collective memory frameworks	"durante el franquismo," "desde la Transición," "en el 23-F"	"during Franco's regime," "since the Transition," "during the 23-F coup attempt"	Named entity recognition + historical knowledge base	Activates evaluative frameworks grounded in shared historical experience
	Collective Memory	Terms invoking shared cultural-historical experiences that carry implicit evaluative loading through collective remembrance	"los años de plomo," "la movida madrileña," "la crisis del 2008"	"the years of lead," "the Madrid scene," "the 2008 crisis"	Cultural lexicon matching + semantic field analysis	Establishes evaluative baseline through historical precedent
	Historical Analogies	Comparative constructions linking current situations to historical events with embedded evaluative implications	"como en el 23-F," "peor que en los 80," "igual que en la guerra"	"like during 23-F," "worse than in the 80s," "just like in the war"	Comparative syntax detection + historical entity linking	Transfers historical evaluative frameworks to contemporary assessment
Social	Formality Indicators	Address forms, register markers, and linguistic choices signaling social hierarchies and relationship dynamics	"usted" vs. "tú," "señor ministro," "oiga"	formal "you" vs. informal "you," "Mr. Minister," "listen"	Pragmatic register analysis + addressee detection	Modulates evaluative intensity based on social positioning
	Group Identity Signals	Linguistic markers establishing ingroup/out-group boundaries that frame evaluative standards and loyalties	"los nuestros," "esta gente," "la gente como nosotros"	"our people," "those people," "people like us"	Pronominal reference analysis + social deixis detection	Activates groupbased evaluative criteria and solidarity frameworks
	Solidarity Markers	Expressions of collective unity and shared identity that intensify evaluative meaning through group cohesion	"estamos juntos en esto," "todos a una," "somos familia"	"we're in this together," "all as one," "we are family"	Collective pronoun analysis + unity semantic field detection	Amplifies evaluative force through collective emotional investment
Political	Ideological Terminology	Lexical items associated with specific political positions that activate partisan evaluative frameworks	"neoliberal," "progresista," "conservador," "de izquierdas"	"neoliberal," "progressive," "conservative," "leftwing"	Political lexicon classification + ideological stance detection	Frames evaluative standards according to political worldview
	Institutional References	Mentions of political institutions, offices, or constitutional entities that activate institutional evaluation frameworks	"el Congreso," "la Moncloa," "el Constitucional," "las Cortes"	"Congress," "the Presidency," "the Constitutional Court," "Parliament"	Named entity recognition (political institutions) + constitutional knowledge base	Activates institutional propriety and constitutional legitimacy frameworks
	Political Hashtags	Platform-specific topic markers signaling political framing and mobilization contexts	"#EleccionesYa," "#CambioClimaticoYa" "#NoALaGuerra"	"#ElectionsNow," "#ClimateActionNow," "#NoToWar"	Hashtag extraction + political topic classification	Establishes political mobilization context for evaluative interpretation
Cultural	Cultural References	Allusions to shared cultural knowledge, traditions, or common expressions that modify evaluative interpretation	"ser más listo que el hambre," "como Pedro por su casa," "hacer el indio"	"to be smarter than hunger," "like Peter in his house," "to act the fool"	Cultural idiom detection + metaphorical expression analysis	Activates culture-specific evaluative schemas and value systems
	Value-laden Terms	Lexical items explicitly encoding cultural values and moral frameworks central to Spanish cultural identity	"respeto," "dignidad," "honor," "vergüenza," "orgullo"	"respect," "dignity," "honour," "shame," "pride"	Value semantic field analysis + cultural axiological mapping	Directly instantiates cultural value systems in evaluative assessment
	Regional Expressions	Dialectal and regional linguistic markers signaling geographic cultural positioning and local identity	"molar," "guay," "chachi," "flipar," "estar piripi"	"to be cool," "awesome," "great," "to freak out," "to be tipsy"	Regional dialect detection + geographic linguistic variation analysis	Signals regional cultural affiliation affecting evaluative authenticity
Digital Discourse	Thread Positioning Signals	Metalinguistic markers indicating structural position within digital conversation and discourse organization	"abro hilo," "como decía antes," "siguiendo el hilo," "cierro hilo"	"opening thread," "as I was saying," "following the thread," "closing thread"	Discourse marker detection + conversation structure analysis	Establishes evaluative context within ongoing digital dialogue
	Multimodal Elements	Non-textual semiotic resources (emojis, images, GIFs) that modify, intensify, or recontextualize textual evaluative meaning	😂, 😏, 🤔, 🙄, 🤪	laughing emoji, eye-roll emoji, clapping emoji, flexing emoji, clown emoji	Computer vision + emoji sentiment analysis + multimodal integration	Modulates evaluative polarity and intensity through visual-textual interaction
	Platform-specific Conventions	Format elements and interactive features specific to particular platforms that frame evaluative meaning	"RT @usuario," "via @," "#TT," "@menciones"	"retweet @user," "via @," "trending topic," "@mentions"	Platform API analysis + social media convention detection	Activates platform-specific evaluative norms and interaction patterns



Marker combinations across dimensions create emergent evaluative meanings through what we term "contextual resonance." The phrase "como en la Transición" ("*as in Spain's democratic transition*") (historical) combined with "los progresistas" ("*the progressives*") (political) and 🤖 (digital) generates a complex evaluative framework that requires simultaneous access to multiple contextual knowledge bases.

Illustrative Inventory of Appraisal and Cultural Markers

Interpreting evaluative meaning in Spanish crisis discourse requires identifying the linguistic cues that typically signal Attitude (Affect, Judgment, Appreciation) and culturally embedded evaluative triggers. The following inventory presents representative examples of these markers as they appeared in the DANA corpus. It is not intended as an exhaustive lexicon, but rather as a demonstration of the types of expressions that guide appraisal interpretation within the framework of Martin and White [5].

Affect expressions in the corpus were often direct emotional reactions to institutional actions, crisis events, or political actors. Table 2 presents typical markers.

Table 2. Illustrative affect markers

Subtype	Spanish Examples	English Rendering	Appraisal Function
Dis/satisfaction	"¡Vaya tela!", "qué desastre," "así no avanzamos"	"What a mess!", "what a disaster," "we can't progress like this"	Signals frustration, displeasure, or negative emotional stance
In/security	"da miedo," "intranquilidad," "es para preocuparse"	"It's scary," "unease," "something to worry about"	Encodes emotional instability, fear, or perceived threat
Happiness/Relief	"menos mal," "alivio," "gracias por estar"	"good thing," "relief," "thanks for being here"	Expresses reassurance, gratitude, or positive emotional evaluation

These expressions frequently appeared alongside emojis (e.g., 😡, 😞, 🙏), which intensified or modified the affective force.

Judgment was the most frequent Attitude type in the dataset, encompassing evaluations of moral conduct, institutional responsibility, competence, and tenacity. Table 3 presents representative markers.

Table 3. Illustrative judgement markers

Subcategory	Spanish Examples	English Rendering	Appraisal Function
Propriety	"vergüenza," "ridículo," "poca dignidad," "no dio la cara", "ausente"	"shame," "ridiculous," "little dignity," "did not show up," "absent"	Evaluates ethical norms, moral obligations, and breaches of institutional duty
Capacity	"echó pelotas," "dio la cara," "aguantó," "estuvo ahí"	"showed guts," "faced the situation," "held on," "was there"	Assesses persistence and resilience, often tied to civic responsibility
Tenacity	"incapaz," "fracasó," "no supo gestionar," "ineficiencia"	"incapable," "failed," "did not know how to manage," "inefficiency"	Evaluates competence and effectiveness in crisis management
Veracity	"mentira," "manipulación," "cuento," "falso"	"lie," "manipulation," "story," "false"	Judges honesty, credibility, and truthfulness

Implicit Judgment frequently emerged through understatement, irony, or omission. For instance, "no fue invitado" ("*wasn't invited*") functioning as an indirect sanction of political propriety.

Appreciation markers evaluated actions, crisis responses, symbolic gestures, and the perceived quality of institutional performance (Table 4).

Spanish crisis discourse often activates culturally shared evaluative frameworks. Several recurring triggers influenced appraisal interpretation (Table 5).



Table 4. Illustrative appreciation markers

Subtype	Spanish Examples	English Rendering	Function
Reaction	"gesto bonito," "detalle," "triste escena"	"nice gesture," "detail," "sad scene"	Aesthetic or emotional response to events and symbolic acts
Valuation	"acto importante," "papel clave," "momentos complicados"	"important act," "key role," "difficult moments"	Evaluates significance or social value of actions
Composition	"bien organizado," "mal planteado," "caos total"	"well organized," "poorly planned," "total chaos"	Evaluates coherence, planning, and structural quality

Table 5. Cultural and contextual triggers relevant to appraisal

Cultural Trigger	Example Expression	English Rendering	Cultural Significance	Evaluative Effect
Honour and dignity norms	"vergüenza ajena," "sin dignidad"	"vicarious shame," "without dignity"	Social expectations of institutional comportment	Strengthens negative propriety judgments
Collective identity	"estamos juntos en esto," "todos a una"	"we're in this together," "all as one"	Community-oriented ethos and shared responsibility	Amplifies positive/negative judgments via collective framing
Symbolic monarchy roles	"los Reyes se quedaron," "estuvieron presentes"	"the King and Queen stayed," "they were present"	Expectations tied to Spain's ceremonial monarchy	Reinterprets presence/absence as propriety or tenacity
Crisis duty expectations	"dar la cara," "estar al pie del cañón"	"face the situation," "be on the front line"	Norms of visibility and commitment in emergencies	Activates tenacity and propriety
Strategic exclusion	"no fue invitado," "ausente"	"was not invited," "absent"	Institutional condemnation expressed through omission	Functions as implicit negative propriety

These triggers regularly co-occurred with political identifiers and historical references, shaping evaluative polarity through contextual resonance.

Digital discourse features contributed additional layers of evaluative meaning. Frequent modifiers included:

- Emojis: 🙄 (sarcasm or dismissal), 👏 (praise or ironic applause), 🤡 (ridicule), 💪 (strength/tenacity);
- Hashtags: #vergüenza, #inundaciones, #DANAValencia (framing devices signalling stance);
- Thread markers: "abro hilo," "como decía antes" (structuring evaluative sequences);
- Image/GIF references: used to amplify or invert textual appraisal.

These multimodal elements often intensified Judgment readings or shifted polarity through sarcasm.

Validation of the Proposal

To operationalize our multi-contextual framework and demonstrate its systematic effects on appraisal interpretation, we developed a computational approach integrating graph database contextual modelling with LLM analysis. This implementation bridges theoretical linguistic frameworks with practical Natural Language Processing (NLP) applications, providing both empirical validation and a replicable methodology for context-aware appraisal analysis.

Our validation approach centres on systematic context injection using graph database architecture to model the complex relationships between contextual dimensions and evaluative interpretation. We constructed a comprehensive knowledge graph encoding the five contextual dimensions as interconnected nodes and relationships, enabling structured representation of the contextual complexity inherent in Spanish digital discourse.

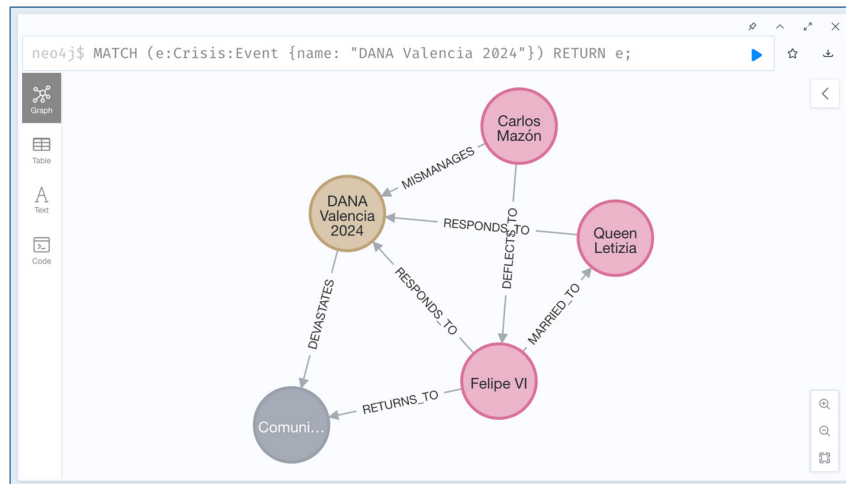


Fig. 2. Fragment of the contextual graph database for the DANA Valencia 2024 crisis

The graph database implementation modelled contextual entities as typed nodes: historical events (Paiporta incident, La Transición), political figures (Sánchez, Felipe VI), cultural concepts (institutional propriety, solidarity norms), social relationships (formality markers), and digital discourse structures (thread positioning, platform conventions). Relationships between nodes captured the theoretical interactions proposed in our framework, with weighted edges representing contextual influence strength based on discourse analysis patterns.

To illustrate the structure of the contextual modelling, Figure 2 displays a fragment of the Neo4j graph centred on the DANA Valencia 2024 Crisis-Event node. The subgraph shows its immediate connections to key contextual entities, including royal figures (Felipe VI, Queen Letizia), political actors (Carlos Mazón), and the affected regional entity. Figure 2 visualises these entities as typed nodes and the relationships between them through labelled edges, providing a concrete example of how heterogeneous contextual elements are represented within the database.

This fragment illustrates how the knowledge graph encodes contextually meaningful relationships that the system later retrieves during appraisal interpretation. Crisis-related behaviours (e.g., RESPONDS_TO, MISMANAGES, DEFLECTS_TO) function as structurally defined links that formalise institutional roles, crisis dynamics, and social expectations. Although only a small portion of the full database is shown, the example demonstrates how contextual associations are modelled and how these connections support targeted, non-redundant context injection during LLM analysis.

The operational value of this graph structure lies in its capacity to support selective, entity-driven context retrieval. When an entity such as Felipe VI or the DANA event is detected in a reply, the system traverses only the locally relevant neighbourhood of connected nodes, extracting relational patterns (e.g., crisis behaviours, institutional roles, territorial impact) as structured contextual metadata. This ensures that the LLM receives precise, discourse-relevant information rather than an undifferentiated dump of background knowledge. By constraining retrieval to the subgraph surrounding the referenced entities, the system avoids context pollution while preserving culturally and institutionally salient cues essential for accurate appraisal interpretation.

For the DANA crisis corpus, the knowledge graph incorporated 847 contextual nodes across all five dimensions, with 2,341 weighted relationships encoding Spanish-specific cultural knowledge, historical precedents, political dynamics, social hierarchies, and digital discourse conventions. This structured representation enabled systematic context injection rather than ad hoc contextual consideration.



```
{
  "contextual_metadata": {
    "political_actor": "King of Spain (Direct engagement)",
    "historical_precedent": "Remained when PM evacuated",
    "cultural_expectation": "Head of State, Captain General Armed Forces",
    "crisis_context": "DANA Valencia 2024 (223 casualties)",
    "symbolic_significance": "High",
    "institutional_contrast": "Higher than any Spanish politician"
  }
}
```

Fig. 3. Sample context extraction for DANA crisis corpus

The knowledge graph operationalizes theoretical relationships through typed nodes and labelled edges with contextual properties. The system performs selective context injection based on entity recognition, retrieving relevant contextual dimensions only when entities are identified in the target posts. Figure 3 illustrates the structured output generated after graph traversal, showing only the contextual slice relevant to the entity 'Felipe VI'. This figure does not represent the entire graph, but the result of multi-dimensional context filtering performed prior to LLM injection.

We employed GPT-4 as the base LLM for appraisal analysis, selected for its superior instruction-following capabilities, robust multilingual performance, and demonstrated capacity for complex contextual reasoning tasks. GPT-4 was selected for contextual framework testing because comparative studies demonstrate that major LLMs exhibit similar cross-cultural limitations, making model choice less critical than framework validation. Cao et al. [48] found that ChatGPT "exhibits a strong alignment with American culture, but adapts less effectively to other cultural contexts," with English prompts systematically reducing cultural variance – a pattern of Western bias documented across state-of-the-art models [4]. While domain-specific performance varies across GPT-4, Claude, and Gemini [16, 49], these differences appear task-specific rather than indicative of fundamental advantages for cross-cultural understanding. We employed GPT-4 as the base LLM for appraisal analysis, selected for its superior instruction-following capabilities, robust multilingual performance, and demonstrated capacity for complex contextual reasoning tasks³.

The computational validation implemented a controlled comparison between context-free and context-enhanced analysis through systematic context injection protocols.

Phase 1: Context-Free Analysis. GPT-4 analysed 200 sampled Spanish social media posts in isolation (without the context), applying Martin and White's appraisal categories [5] through crafted prompts requesting categorization of evaluative expressions as Affect, Judgment, or Appreciation with positive/negative polarity determination (see Figure 4). This phase established baseline LLM performance on Spanish evaluative discourse without contextual enhancement.

Figure 4 corresponds to Phase 1 of the evaluation procedure, showing the baseline prompt used for context-free appraisal analysis before applying contextual metadata.

Phase 2: Context-Enhanced Analysis. For each reply, the system queried the knowledge graph to retrieve relevant contextual metadata across all five dimensions. Graph traversal algorithms identified contextual nodes connected to entities mentioned in the target reply, generating structured contextual summaries including historical background, political relationships, cultural significance, social dynamics, and digital discourse positioning. This contextual metadata was systematically injected via API into GPT-4 prompts through analysis templates (see Figure 5).

³ Although the prompts used to guide the model were written in English, all linguistic material analysed by the LLM remained entirely in Spanish, and GPT-4 was instructed to produce all appraisal classifications exclusively in Spanish. English was selected as the prompt language for methodological reasons: recent evaluations of multilingual LLMs show that models often exhibit higher instruction-following stability and lower variance when task instructions are provided in English, even when processing non-English content [50, 51]. This tendency is not universal – performance varies by language and task – but English prompts frequently minimise misalignment effects caused by instruction parsing rather than linguistic content. We therefore used English solely at the level of metainstructions to ensure consistency and replicability, while keeping the analysed texts and the output language fully in Spanish.



```
You are an expert in Spanish discourse analysis using Appraisal Theory [5].

## POST TO ANALYSE:
"{reply_text}"

## TASK:
Analyse this Spanish social media reply for evaluative language using Martin and White's framework.

## ANALYSIS REQUIREMENTS:

### APPRAISAL ANALYSIS
Identify and categorize evaluative language:
- ATTITUDE: Judgment (Capacity, Tenacity, Veracity, Propriety),
  Affect (Satisfaction, Security, Happiness),
  Appreciation (Reaction, Composition, Valuation)
- ENGAGEMENT: Monogloss vs Heterogloss, Attribution patterns
- GRADUATION: Force and Focus

### OUTPUT FORMAT:
{
  "appraisal_instances": [
    {
      "text_span": "exact evaluative phrase",
      "category": "Judgment:Capacity|Affect:Security|etc",
      "polarity": "positive|negative|neutral"
    }
  ]
}
```

Fig. 4. Prompt for context-free analysis

Figure 5 corresponds to Phase 2 of the procedure, where database-derived contextual metadata is injected into the prompt to enable context-informed appraisal analysis.

To establish the baseline, a Spanish linguistics expert (Universitat Politècnica de València, specializing in digital discourse analysis) independently annotated all 200 posts using Martin and White's framework. The expert coded all instances for attitude category, polarity, and cultural specificity, providing baseline accuracy measurements for computational analysis comparison.

We compared three conditions: GPT-4 without context, GPT-4 with graph database context injection, and expert ground truth. Primary metrics included accuracy of computational conditions against expert annotations, precision and recall for appraisal detection, McNemar's test for statistical significance, and Cohen's d for effect size calculation. Secondary analysis involved expert validation of Spanish-specific patterns and correlation analysis between cultural marker presence and accuracy improvements.

Results

Our analysis demonstrates systematic improvements in LLM appraisal performance through cultural context integration (Table 6). The expert annotation served as the baseline for evaluating computational outputs across conditions.

Context integration significantly improved GPT-4's accuracy from 67.3% to 84.7% against expert annotations (McNemar's test: $p < 0.001$). Table 6 details detection counts and accuracy rates across conditions: context-free analysis, context-enhanced analysis, and expert baseline. The "Detection Increase" column quantifies additional instances surfaced by context (e.g., 111.1% more Judgments as well as 110.8% more total appraisals), while accuracy percentages match to expert codings. To illustrate qualitative improvements, several appraisals that the baseline model either missed or misclassified were detected only after contextual injection. For example:

- "Estar 20 minutos tras el barro y piedras, oye."
"Standing there 20 minutes under mud and stones, listen..."



```

You are analysing Spanish royal crisis discourse using Appraisal Theory with specific focus on evaluation patterns
during the DANA Valencia 2024 crisis.

[FORMATTED DATABASE CONTEXT INCLUDING:]
## KEY FIGURES
**Felipe VI**: Role: King of Spain, Leadership Style: Direct engagement,
Crisis Style: Constitutional, Approval Rating: 5.7,
Crisis Management: Remained when PM evacuated

## LOCATIONS
**Paiporta**: DANA Deaths: 62, DANA Impact: Catastrophic,
Public Hostility: High, Royal Visit Date: 2024-11-03

## EVENTS
**DANA Valencia 2024**: Date: 2024-10-29, Casualties: 223,
Political Impact: High, Symbolic Significance: High

## POST TO ANALYSE:
"{reply_text}"

## ANALYSIS REQUIREMENTS:

### APPRAISAL ANALYSIS
Identify and categorize evaluative language using Martin and White's framework:
- ATTITUDE: Judgment (Capacity, Tenacity, Veracity, Propriety),
  Affect (Satisfaction, Security, Happiness),
  Appreciation (Reaction, Composition, Valuation)
- ENGAGEMENT: Monogloss vs Heterogloss, Attribution patterns, Dialogic positioning
- GRADUATION: Force (intensification, quantification), Focus (sharpen, soften)

### CONTEXTUAL INTEGRATION
Using database context: How does this reply relate to specific events, persons, or locations? How do appraisal
patterns connect to broader crisis narrative? What institutional relationships are being evaluated?

OUTPUT FORMAT:
{
  "appraisal_instances": [
    {
      "text_span": "exact evaluative phrase",
      "category": "Judgment:Capacity|Affect:Security/etc",
      "polarity": "positive|negative|neutral",
      "contextual_significance": "how database context enhances interpretation"
    }
  ],
  "institutional_positioning": "relationship to royal/political authority",
  "crisis_narrative_connection": "how this fits broader DANA crisis story"
}

```

Fig. 5. Prompt for context enhanced analysis

→ Correctly identified as Positive Judgment: Tenacity, not mere observation.

- "Le creen el dolor a Letizia?"

"Do they believe Letizia's pain?"

→ Detected as Negative Judgment: Veracity, previously labelled as Affect.

These examples demonstrate how culturally anchored cues related to shame, honour, emotional authenticity, and symbolic presence require contextual metadata to be interpreted correctly.

Context proved particularly effective at detecting Appreciation evaluations (215.9% increase) and improving accuracy for negative polarity expressions (63.7% to 83.4%). All categories showed consistent improvement, with context-enhanced analysis achieving over 82% accuracy across all appraisal types, but invoked types lagged slightly (82.6%), suggesting limits in implicit polarity without fuller dialogic Engagement analysis.

A clear example of context-dependent invoked appraisal is the standalone acronym "VERDE". Without contextual metadata, the baseline model classified it as non-evaluative because it appears as a colour term. However, within Spanish discourse and monarchical support registers, the acronym is widely used to mean "Viva El Rey De España" ("Long live the King of Spain"). With context injection,



Table 6. Contextual enhancement with expert validation

	Metric	Annotation				Accuracy in %	
		Context-Free	Context-Enhanced	Expert Baseline	Detection Increase in %	Expert vs. Context-Free	Expert vs. Context-Enhanced
	Total Appraisals	416	877	642	110.8	67.3	84.7
Category	Affect	116	218	142	87.9	71.2	87.3
	Appreciation	44	139	116	215.9	69	85.3
	Judgment	244	515	384	111.1	64.8	83.9
Polarity	Negative	212	371	344	75	63.7	83.4
	Positive	196	457	298	133.2	72.1	86.2
Type	Inscribed	384	812	596	111.5	68.1	85.1
	Invoked	32	65	46	103.1	65.2	82.6

GPT-4 correctly identified it as Positive Appreciation: Valuation and Positive Judgment: Capacity/Propriety signalling symbolic support for the monarch. This demonstrates the necessity of cultural and intertextual context for detecting evaluations encoded in acronymic or formulaic expressions.

Table 7 reveals how context enhancement works beyond just improved accuracy. Beyond detection, context enabled qualitative shifts: surfacing missed appraisals in 116 posts (58%, e.g., invoked propriety in omission like "ausente" as deliberate exclusion) and reclassifying 144 instances (e.g., from Appreciation to Judgment in restraint descriptions). Expert validation confirmed 127 reclassifications (87.8%) as accurate, but 17 (12.2%) were overcorrections (e.g., context inflating tenacity in neutral persistence), highlighting potential for hallucinated cultural inferences.

Table 7. Systematic enhancement through contextual integration

Metric	Context-Free	Context-Enhanced	Expert validation	Change
Posts Analysed	200	200	200	N/A
Additional Detected Appraisals	0	116	N/A	58%
Reclassified Appraisals	164	144	127 confirmed	87.8% of shared

Analysis surfaced five context-dependent patterns shaping Judgment in the corpus (expert rejected Systemic Political Disengagement as non-specific crisis talk, reducing from six). Occurrences and percentages reflect proportion of 196 context-reliant instances; patterns fit Martin and White's framework but require Spanish cultural baselines for polarity/nuance (Table 8).

These patterns, validated in 85.3% of context-specific instances as requiring Spanish knowledge, manifest culturally shaped Judgment applications. Honour-Driven Valor (12% of instances) links personal courage expressions to civic duty expectations – evaluations requiring Spanish cultural knowledge about institutional responsibility. Institutional Moral Disgrace (15.5% of instances) employs shame based terms that reflect Spanish collectivist values about institutional behaviour.

Meaningful Exclusion (9% of instances) achieves criticism through strategic omission – stating someone "no fue invitado" ("wasn't invited") functions as institutional critique requiring understanding of Spanish protocol expectations. Ceremonial Monarchy as Symbolic Leadership (11% of instances) evaluates royal presence during crisis through constitutional monarchy frameworks unique to Spanish political culture.



Table 8. Spanish-specific appraisal patterns

Pattern Label	Appraisal type(s)	Example expressions	English rendering	Cultural significance	Occurrences	%
Honour-Driven Valor	Judgment: Tenacity	"echó pelotas," "dio la cara"	" <i>He showed guts,</i> " " <i>He faced the situation</i> "	Heroism grounded in civic responsibility	24	12
Institutional Moral Disgrace	Judgment: Propriety	"vergüenza," "pena," "ridículo"	" <i>Shame,</i> " " <i>disgrace,</i> " " <i>ridiculous</i> "	Moral judgment of institutional failure	31	15.5
Meaningful Exclusion	Judgment: Propriety	"no fue invitado," "ausente"	" <i>wasn't invited,</i> " " <i>absent</i> "	Critique by deliberate exclusion	18	9
Ceremonial Monarchy as Symbolic Leadership	Judgment: Propriety/ Capacity	"los Reyes se quedaron"	" <i>The Royals stayed</i> "	Elevating symbolic presence as duty	22	11
Institutionally-Aligned Emotional Display	Judgment: Propriety + Affect	"Letizia abraza a las víctimas"	" <i>Letizia embraces the victims</i> "	Emotionally authentic response aligned with institutional norms	42	21
Systemic Political Disengagement	Judgment: Capacity/ Propriety	"Sánchez ha huido," "fracasó"	" <i>Sánchez has fled,</i> " " <i>he failed</i> "	Abandonment of civic duty by political figures	16	8

Institutionally-Aligned Emotional Display (21% of instances – the most common pattern) combines emotional expression with institutional judgment in ways reflecting Spanish cultural norms about collective emotional expression. However, Systemic Political Disengagement (8% of instances) was identified by the expert as general crisis discourse rather than Spanish-specific.

Discussion

Our analysis of Spanish crisis discourse provides empirical evidence for the constitutive role of cultural context in cross-cultural appraisal analysis. The findings address the three research questions, highlighting the potential and limitations of structured contextual integration for enhancing LLM performance in non-English evaluative tasks.

Systematic Context Integration and LLM Performance Enhancement

The systematic integration of multi-dimensional cultural context demonstrates substantial effects on LLM appraisal analysis capabilities, with a 110.8% increase in detected instances (see Table 6) that extends beyond quantitative gains to qualitative interpretative depth. Context-enhanced analysis reveals evaluative complexity invisible to surface-level processing, particularly in the systematic emergence of judgment evaluations embedded within seemingly neutral descriptive language.

As we can see in Example 1, context-free analysis categorizes this as positive affect – emotional reassurance about royal presence. However, contextual integration activating knowledge of Prime Minister Sánchez's evacuation while the monarchy remained transforms the evaluative landscape entirely. The statement becomes complex judgment evaluation encompassing tenacity (persistence under adversity) and propriety (institutional duty fulfillment), demonstrating how cultural knowledge constitutes rather than supplements evaluative meaning.

The predominant transformation involves revelation of implicit judgment evaluations operating through Spanish cultural frameworks for institutional behaviour, collective responsibility, and symbolic leadership. These patterns suggest that Spanish evaluative discourse operates through culturally embedded mechanisms largely invisible to context-free computational analysis.

These findings build upon foundational work in evaluation studies that has long recognized context's crucial role in interpretative processes. Hunston and Thompson [8] established that evaluation



Example 1

Reply:	"Los Reyes se quedaron en la zona afectada mientras otros se marcharon."	"The Royals stayed in the affected area while others left."
Context-Free:	Affect: Security (positive emotional tone, interpreted as reassurance)	
Context-Enhanced:	<ul style="list-style-type: none">• Judgment: Tenacity – persistence in adversity• Judgment: Propriety – ethical commitment to presence• Affect: Admiration – invoked admiration for bravery	

serves multiple functions including expressing community value systems and constructing speaker-hearer relationships, while Martin and White [5] demonstrated how appraisal operates dialogically to position speakers relative to value systems and alternative viewpoints. Our analysis extends these insights by providing systematic evidence for how cultural context transforms evaluative interpretation in computational environments.

The enhancement in appraisal detection provides empirical evidence for Moroshkina's [9] theoretical claim that context can transform neutral language units into evaluative expressions, moving beyond her descriptive account to systematic documentation of transformative processes in Spanish digital discourse. Predominant shifts reveal implicit Judgments in neutral language, yet 12.2% over-corrections (see Table 7) underscore risks of contextual overdetermination, aligning with Tao et al. [4] Western misalignment but exposing LLM vulnerabilities in non-English polarity.

Context-Dependent Evaluation Patterns

Our analysis identifies systematic evaluation patterns that appear distinctively embedded within Spanish cultural frameworks. Honour-driven valor emerges through expressions like "echó pelotas" and "dio la cara" that link personal courage to institutional responsibility in culturally specific ways. These evaluations operate within Martin and White's Judgment: Tenacity category but require Spanish cultural knowledge about civic duty and collective responsibility for accurate interpretation.

Institutional moral disgrace operates through shame-based assessment that reflect Spanish collectivist frameworks emphasizing social harmony and institutional respect. As seen in Example 2, the statement "Sánchez no fue invitado a la ceremonia" demonstrates sophisticated evaluative mechanisms operating through strategic omission rather than explicit criticism. Surface neutrality masks judgment of political propriety: exclusion signals institutional disapproval without requiring explicit condemnation.

Example 2

Reply:	"Sánchez no fue invitado a la ceremonia."	"Sánchez was not invited to the ceremony."
Context-Free:	Factual, neutral report. No appraisal detected.	
Context-Enhanced:	<ul style="list-style-type: none">• Judgment: Propriety (negative) – implies moral impropriety, exclusion as institutional condemnation.• Attitude Modality: Possibly invoked through irony – rhetorical understatement masking criticism.	

Meaningful exclusion achieves evaluative force through cultural knowledge about Spanish institutional protocols and ceremonial significance. Without understanding these frameworks, exclusion statements remain informationally neutral. With contextual activation, they become sophisticated political evaluation achieving critical force through systematic understatement. This pattern occurs consistently across 18 instances in our subcorpus (see Table 8), suggesting systematic rather than isolated phenomena.

Example 3 reveals systematic appraisal reclassification when Spanish cultural contexts activate. Surface analysis categorizes this as appreciation of behavioural restraint. Contextual integration reveals judgment evaluation operating through Spanish frameworks for institutional authority, where



Example 3

Reply:	"Felipe estuvo correcto, sin grandes gestos."	" <i>Felipe was proper, without grand gestures.</i> "
Context-Free:	Appreciation: Valuation – interpreted as a neutral or positive aesthetic comment on restraint and composure.	
Context-Enhanced:	<ul style="list-style-type: none"> • Judgment: Propriety (positive) – behaviour viewed as ethically appropriate, reflecting public expectations of royal dignity. • Judgment: Capacity (optional reading) – possible interpretation of symbolic competence in representing national unity during crisis. • Attitude Modality: Possibly blended – culturally coded masculinity frames emotional restraint as moral strength, especially in patriarchal institutional roles. 	

emotional restraint signals leadership competence during collective trauma. This reclassification pattern – from Appreciation to Judgment – occurs in 87.8% of shared detections (see Table 7), indicating systematic interpretative transformation rather than random variation.

The Spanish-specific patterns we identify provide empirical support for established observations about cultural differences in discourse strategies. Ivorra-Pérez and Giménez-Moreno [39] documented systematic differences between Spanish and English digital discourse, particularly regarding contextual dependence in communication strategies. Our documentation of honour-driven valor, institutional moral disgrace, and meaningful exclusion extends such comparative analysis into evaluative language territory, demonstrating how Spanish cultural frameworks systematically shape appraisal beyond surface linguistic features.

The emergence of strategic omission as evaluative practice aligns with Van Dijk's [21] socio-cognitive framework for understanding how political contexts shape discourse positioning strategies. Our analysis provides specific mechanisms through which Spanish political culture achieves evaluative force through systematic understatement, offering empirical grounding for theoretical frameworks about context-dependent meaning while identifying culturally specific manifestations requiring specialized analytical approaches.

Structured Context Injection and Computational Enhancement

The graph database approach demonstrates significant potential for improving cross-cultural appraisal analysis accuracy through structured context injection. The systematic context injection protocol reveals interpretative transformations invisible to ad hoc contextual approaches. Historical events like the DANA crisis activate complex relationship networks connecting institutional figures, geographical locations, cultural values, and political dynamics. The graph structure captures these multidimensional relationships representing contextual influence strength, enabling dynamic context retrieval rather than static information provision. This operationalizes KhosraviNik [22] social-computational integration, validating Zhang [11] dependency arguments.

While our computational validation demonstrates clear performance improvements, future research would benefit from additional validation approaches. The systematic reclassification of 87.8% of shared detections provides strong evidence for contextual enhancement effects, though future multi-coder ICR needed for pattern consistency.

Methodological Limitations and Future Directions

Several methodological constraints limit the generalizability of this study. The corpus comprises 200 posts from a single crisis event, which restricts the scope of claims that can be made about broader Spanish evaluative discourse. The dataset was intentionally bounded to enable controlled comparison between context-free and context-augmented conditions, but larger and multi-event corpora will be required for more robust generalization. Token-length distribution, thematic scope, and sampling window have been specified to clarify the characteristics of the dataset.

A further limitation concerns the use of a single expert annotator. Although multi-annotator designs are the preferred standard for assessing inter-coder reliability, single-coder annotation is sometimes



employed in exploratory and pilot-stage linguistic studies, where the primary goal is methodological development and internal consistency [46, 52, 53]. Nonetheless, this design necessarily limits conclusions about annotation stability across coders. Future work will therefore adopt multi-annotator protocols, including double-blind coding, adjudication procedures, and inter-coder reliability metrics such as Krippendorff's α , to strengthen the robustness of the appraisal annotations.

In addition, although the computational validation demonstrates consistent improvements under context-augmented conditions, larger datasets and independent expert evaluation will be needed to fully assess interpretive robustness. Extending the analysis to multiple crisis contexts and to Spanish varieties beyond the Peninsular domain will further test the generality of the proposed framework. Integrating the contextual-graph approach with established Spanish discourse-analytic models may also help consolidate theoretical alignment.

Author Contribution and Study Significance

This study makes three key contributions. First, it provides empirical evidence that culturally structured context substantially enhances LLM-based appraisal interpretation in Spanish crisis discourse, improving accuracy from 67.3% to 84.7%. Second, it identifies five culturally embedded evaluation patterns – honour-driven tenacity, institutional moral disgrace, emotional authenticity, collectivist propriety, and symbolic valuation – that shape Judgment in Spanish digital communication and are not detectable through English-centric or context-free approaches. Third, it proposes a reproducible graph-based context-injection workflow that integrates cultural, institutional, and historical meta-data into LLM processing. These contributions advance both computational appraisal analysis and cross-cultural LLM evaluation.

Conclusions

This study advances the linguistic description of evaluative meaning in Peninsular Spanish crisis discourse and demonstrates how appraisal resources behave in digital environments. The analysis identifies five discourse-level appraisal patterns – institutional moral sanction, honour-linked tenacity, emotional authenticity judgments, collectivist responsibility, and symbolic valuation – that structure Judgment and Graduation in Spanish crisis communication. These findings refine the understanding of how evaluative stance is encoded in Spanish, showing that key Judgment subtypes (especially Propriety, Capacity, and Tenacity) are activated through interactional cues, rhetorical framing, and discourse sequencing rather than through lexical items alone. In doing so, the study contributes new empirical evidence to Appraisal Theory and offers a more detailed linguistic characterisation of evaluation in crisis-related communication.

In direct response to the research questions, the findings demonstrate that incorporating discourse-level contextual information substantially improves the linguistic interpretation of appraisal in Spanish. Regarding RQ1, integrating structured contextual cues enhanced the accuracy of Judgment detection, particularly Propriety and Capacity, by enabling the model to recognise evaluative configurations grounded in discourse organisation rather than lexicon-only patterns. For RQ2, the results show that the appraisal patterns identified here (e.g., institutional propriety, moral accountability, leadership legitimacy, crisis-response expectations) rely on pragmatic and interactional cues that must be made explicit for correct polarity and target assignment. For RQ3, the context-augmented model consistently outperformed the baseline, producing more stable polarity assignments and richer linguistic coherence across posts, indicating that context-sensitive modelling supports more precise, linguistically faithful appraisal annotation in Spanish.

Methodologically, the study demonstrates that context-informed modelling has clear benefits for linguistic analysis: accuracy improved from 67.3% to 84.7%, and implicit Judgments, often missed in surface-level appraisal extraction, were more reliably detected. This workflow illustrates how computational tools can meaningfully support theory-driven linguistic analysis when designed to preserve linguistic nuance rather than prioritise general-purpose predictions.



Future work should expand the corpus size and include multi-coder reliability assessment to test the stability of the appraisal patterns across Spanish varieties. Further comparative studies across Romance and Slavic languages may clarify to what extent the evaluation patterns observed here are language-specific or cross-linguistically recurrent. Overall, the study contributes both new empirical insights into Spanish evaluative discourse and a replicable approach to integrating Appraisal Theory with computational linguistic methods.

REFERENCES

- [1] **Bigelow E.J., Lubana E.S., Dick R.P., Tanaka H., Ullman T.D.**, In-Context Learning Dynamics with Random Binary Sequences, arXiv:2310.17639v3, 2023. DOI: 10.48550/ARXIV.2310.17639
- [2] **Wibisono K.C., Wang, Y.**, From Unstructured Data to In-Context Learning: Exploring What Tasks Can Be Learned and When, arXiv:2406.00131v2, 2024. DOI: 10.48550/ARXIV.2406.00131
- [3] **Naous T., Xu W.**, On The Origin of Cultural Biases in Language Models: From Pre-training Data to Linguistic Phenomena, arXiv:2501.04662v1, 2025. DOI: 10.48550/ARXIV.2501.04662
- [4] **Tao Y., Viberg O., Baker R.S., Kizilcec R.F.**, Cultural bias and cultural alignment of large language models, PNAS Nexus, 3 (9) (2024) pgae346. DOI: 10.1093/pnasnexus/pgae346
- [5] **Martin J.R., White P.R.R.**, The Language of Evaluation: Appraisal in English, Palgrave Macmillan, New York, 2005. DOI: 10.1057/9780230511910
- [6] **Bednarek M.**, Evaluation in Media Discourse: Analysis of a Newspaper Corpus, Continuum, London, 2006.
- [7] **Puspita D., Pranoto B.E.**, The attitude of Japanese newspapers in narrating disaster events: Appraisal in critical discourse study, Studies in English Language and Education, 8 (2) (2021) 796–817. DOI: 10.24815/siele.v8i2.18368
- [8] Evaluation in Text: Authorial Stance and the Construction of Discourse, ed. by S. Hunston, G. Thompson, Oxford University Press, Oxford, 2000. DOI: 10.1093/oso/9780198238546.001.0001
- [9] **Moroshkina H.**, Specificities of the Mutual Influence of Context and Assessment in French Evaluative Utterances, Advanced Education, 12 (2019) 244–248. DOI: 10.20535/2410-8286.148506
- [10] **Ingendahl M., Woitzel J., Prophet N., Wänke M., Alves H.**, From Deviant Likes to Reversed Effects: Re-Investigating the Contribution of Unaware Evaluative Conditioning to Attitude Formation, Collabra: Psychology, 9 (1) (2023) 87462. DOI: 10.1525/collabra.87462
- [11] **Zhang, D.**, Evaluation in Context, Australian Journal of Linguistics, 37 (1) (2017) 76–80. DOI: 10.1080/07268602.2015.1091282
- [12] **Moran T., Nudler Y., Bar-Anan Y.**, Evaluative Conditioning: Past, Present, and Future, Annual Review of Psychology, 74 (2023) 245–269. DOI: 10.1146/annurev-psych-032420-031815
- [13] **Kareem R., Farhan H.**, The Language of Evaluation in Jose Saramago’s Blindness via Appraisal Theory, International Linguistics Research, 5 (1) (2022) 25–36. DOI: 10.30560/ilr.v5n1p25
- [14] **Lim W., Wu Q.**, Vague language and context dependence, Frontiers in Behavioral Economics, 2 (2023) 1014233. DOI: 10.3389/frbhe.2023.1014233
- [15] **Tilakaratna N., Mahboob A.**, Appraisal in the time of conflict, Linguistics and the Human Sciences, 8 (1) (2013) 63–90. DOI: 10.1558/lhs.v8i1.63
- [16] **Liu H., Cao Y., Wu X., Qiu C., Gu J., Liu M., Hershcovich D.**, Towards realistic evaluation of cultural value alignment in large language models: Diversity enhancement for survey response simulation, Information Processing & Management, 62 (4) (2025) 104099. DOI: 10.1016/j.ipm.2025.104099
- [17] **Gupta V., Chowdhury S.P., Zouhar V., Rooein D., Sachan M.**, Multilingual Performance Biases of Large Language Models in Education, arXiv:2504.17720v2, 2025. DOI: 10.48550/arXiv.2504.17720
- [18] **Hengle A., Bajpai P., Dan S., Chakraborty T.**, Can LLMs reason over extended multilingual contexts? Towards long-context evaluation beyond retrieval and haystacks, arXiv:2504.12845v1, 2025. DOI: 10.48550/arXiv.2504.12845
- [19] **Huang Z., Zhu W., Cheng G., Li L., Yuan F.**, MindMerger: Efficient Boosting LLM Reasoning in non-English Languages, arXiv:2405.17386v1, 2024. DOI: 10.48550/arXiv.2405.17386
- [20] **Diethei D., Colley A., Wienert, J., Schoning J.**, Different Length, Different Needs: Qualitative Analysis of Threads in Online Health Communities, 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, 2022, pp. 348–356. DOI: 10.1109/ICHI54592.2022.00056



- [21] **Van Dijk T.A.**, Discourse and Context: A Sociocognitive Approach, Cambridge University Press, Cambridge, 2008.
- [22] Social Media and Society: Integrating the digital with the social in digital discourse, ed. by M. KhosraviNik, John Benjamins Publishing Company, 2023. DOI: 10.1075/dapsac.100
- [23] **Manca E.**, Context and Language, Università del Salento, 2012.
- [24] **Felbo B., Mislove A., Søgaard A., Rahwan I., Lehmann S.**, Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017, pp. 1615–1625. DOI: 10.18653/v1/D17-1169
- [25] **Sykes J.M.**, Emergent Digital Discourses: What Can We Learn From Hashtags and Digital Games to Expand Learners' Second Language Repertoire?, Annual Review of Applied Linguistics, 39 (2019) 128–145. DOI: 10.1017/S0267190519000138
- [26] **Zappavigna M., Logi L.**, Emoji and Social Media Paralanguage, Cambridge University Press, Cambridge, 2024. DOI: 10.1017/9781009179829
- [27] **Hadour T.**, The pragmatics of hashtags in French tweets, Internet Pragmatics, 8 (1) (2025) 86–112. DOI: 10.1075/ip.00117.had
- [28] **Du Bois J.W.**, Towards a dialogic syntax, Cognitive Linguistics, 25 (3) (2014) 359–410. DOI: 10.1515/cog-2014-0024
- [29] **Pöldvere N., Johansson V., Paradis C.**, Resonance in dialogue: The interplay between intersubjective motivations and cognitive facilitation, Language and Cognition, 13 (4) (2021) 643–669. DOI: 10.1017/langcog.2021.16
- [30] **Kamsinah N.N., Nuraziza A.**, Pragmatic Analysis in Digital Communication: A Case Study of Language Use on Social Media, International Journal of Economics, Commerce, and Management, 1 (4) (2024) 375–383. DOI: 10.62951/ijecm.v1i4.259
- [31] **Eragamreddy Dr.N.**, Exploring Pragmatics: Uncovering the Layers of Language and Meaning, International Journal of Current Science Research and Review, 7 (3) (2024) 1886–1895. DOI: 10.47191/ijcsrr/V7-i3-50
- [32] **Eke C., Asak M., Adeyem O.**, Pragmatics and Communication in the Digital Age: A Focus on the Nigerian Experience, Advance Journal of Linguistics and Mass Communication, 9 (2) (2025) 34–52. Available at: <https://aspjournals.org/Journals/index.php/Ajlmc/article/view/1057> (accessed 08.12.2025).
- [33] **Herring S.C.**, The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis, Analyzing Digital Discourse, ed. by P. Bou-Franch, P. Garcés-Conejos Blitvich, Palgrave Macmillan, Cham, 2019, 25–67. DOI: 10.1007/978-3-319-92663-6_2
- [34] **Kopf S.**, Unravelling social media critical discourse studies (SM-CDS) – four approaches to studying social media through the critical lens, Critical Discourse Studies, (2025) 1–18. DOI: 10.1080/17405904.2025.2463622
- [35] **Bisk Y., Holtzman A., Thomason J., Andreas J., Bengio Y., Chai J., Lapata M., Lazari-dou A., May J., Nisnevich A., Pinto N., Turian J.**, Experience Grounds Language, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, pp. 8718–8735. DOI: 10.18653/v1/2020.emnlp-main.703
- [36] **Sap M., Gabriel S., Qin L., Jurafsky D., Smith N.A., Choi Y.**, Social Bias Frames: Reasoning about Social and Power Implications of Language, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020, pp. 5477–5490. DOI: 10.18653/v1/2020.acl-main.486
- [37] **Asher N., Lascarides A.**, Logics of Conversation, Cambridge University Press, Cambridge, 2005.
- [38] **Aquino Y., Bento P., Buzelin A., Dayrell L., Malaquias, S., Santana C., Estanislau V., Dutenhofner P., Evangelista G.H.G., Porfírio L.G., Grossi C.S., Rigueira P.B., Almeida V., Pappa G.L., Meira W.**, Discord Unveiled: A Comprehensive Dataset of Public Communication (2015–2024) // arXiv:2502.00627v1. 2025. DOI: 10.48550/arXiv.2502.00627
- [39] **Ivorra-Pérez F.M., Giménez-Moreno R G.-M.**, The level of context dependence of engagement markers in Peninsular Spanish and US business websites, Revista de Lenguas Para Fines Específicos, 24.2 (2018) 38–53. DOI: 10.20420/rlfe.2018.364
- [40] **Lage A.H., Diego Á.R.**, La oralización de textos digitales: Usos no normativos en conversaciones instantáneas por escrito, Caracteres: Estudios Culturales y Críticos de La Esfera Digital, 2 (2) (2013) 92–108.



- [41] **Alarcón Silva M., Cárdenas Neira C.**, Convocatoria de protesta a través de Instagram, un análisis socio cognitivo de estrategias discursivas en el contexto del movimiento social en Chile (2019–2020), *Revista Latina de Comunicación Social*, 79 (2021) 127–149. DOI: 10.4185/RLCS-2021-1524
- [42] **Deas N., Turcan E., Mejia I.E.P., McKeown K.**, MASIVE: Open-Ended Affective State Identification in English and Spanish, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, FL, 2024, pp. 20467–20485. DOI: 10.18653/v1/2024.emnlp-main.1139
- [43] **Millán-Franco M., Domínguez De La Rosa L., Gómez-Jacinto L., Hombrados-Mendieta M.I., García-Cid A.**, Análisis discursivo sobre el sentido de comunidad de los latinoamericanos en Málaga, *Migraciones internacionales*, 12, (2021) 5. DOI: 10.33679/rmi.v1i1.2137
- [44] **Águila Díaz J.**, Los efectos perniciosos de una identidad colectiva excluyente. El caso del rap español, *Encrucijadas. Revista Crítica de Ciencias Sociales*, 23 (3) (2024) v2301. Available at: <https://recyt.fecyt.es/index.php/encrucijadas/article/view/96747> (accessed 08.12.2025).
- [45] **Falconi Calvachi F.R., Velásquez Cajas Á.P., Lema Flores K.M., Lomas Chacón P.E.**, Análisis crítico del discurso intercultural en Facebook. Caso ‘Conaie Comunicación’, *Culturas Revista de Gestión Cultural*, 8 (2) (2021) 101–122. DOI: 10.4995/cs.2021.16529
- [46] **Krippendorff K.**, *Content Analysis: An Introduction to Its Methodology*, SAGE Publications, Inc., Thousand Oaks, CA, London, New Delhi, 2019. DOI: 10.4135/9781071878781
- [47] **McHugh M.L.**, Interrater reliability: The kappa statistic, *Biochemia Medica*, 22 (2012) 276–282. DOI: 10.11613/bm.2012.031
- [48] **Cao Y., Zhou L., Lee S., Cabello L., Chen M., Hershovich D.**, Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study, *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, ed. by S. Dev, V. Prabhakaran, D. I. Adelani, D. Hovy, L. Benotti, Association for Computational Linguistics, Dubrovnik, 2023, pp. 53–67. DOI: 10.18653/v1/2023.c3nlp-1.7
- [49] **Ashrafimoghari M., Doroudi S., Tabatabaei S.M.S., Valipour M.**, Comparative analysis of GPT-4, Gemini AI, and Claude: Strengths and weaknesses in content generation. SSRN, 2024. DOI: 10.2139/ssrn.4681307
- [50] **Moroni L., Aula-Blasco J., Conia S., Baucells I., Perez N., Suárez S.P., Sallés A., Ostendorff M., Falcão J., Son G., Gonzalez-Agirre A., Navigli R., Villegas M.**, Multi-LMentry: Can Multilingual LLMs Solve Elementary Tasks Across Languages?, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Suzhou, 2025, pp. 34114–34145. DOI: 10.18653/v1/2025.emnlp-main.1731
- [51] **Schut L., Gal Y., Farquhar S.**, Do Multilingual LLMs Think In English?, *arXiv:2502.15603v1*, 2025. DOI: 10.48550/arXiv.2502.15603
- [52] **Anglin K., Boguslav A., Hall T.**, Improving the Science of Annotation for Natural Language Processing: The Use of the Single-Case Study for Piloting Annotation Projects, *Journal of Data Science*, 20 (3) (2022) 339–357. DOI: 10.6339/22-JDS1054
- [53] **Artstein R., Poesio M.**, Inter-Coder Agreement for Computational Linguistics, *Computational Linguistics*, 34 (4) (2008) 4. DOI: 10.1162/coli.07-034-R2

СПИСОК ИСТОЧНИКОВ

1. **Bigelow E.J., Lubana E.S., Dick R.P., Tanaka H., Ullman T.D.** In-Context Learning Dynamics with Random Binary Sequences // *arXiv:2310.17639v3*. 2023. DOI: 10.48550/ARXIV.2310.17639
2. **Wibisono K.C., Wang Y.** From Unstructured Data to In-Context Learning: Exploring What Tasks Can Be Learned and When // *arXiv:2406.00131v2*. 2024. DOI: 10.48550/ARXIV.2406.00131
3. **Naous T., Xu W.** On The Origin of Cultural Biases in Language Models: From Pre-training Data to Linguistic Phenomena // *arXiv:2501.04662v1*. 2025. DOI: 10.48550/ARXIV.2501.04662
4. **Tao Y., Viberg O., Baker R.S., Kizilcec R.F.** Cultural bias and cultural alignment of large language models // *PNAS Nexus*. 2024. Vol. 3, Iss. 9. Art. no. pgae346. DOI: 10.1093/pnasnexus/pgae346
5. **Martin J.R., White P.R.R.** *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan, 2005. 278 p. DOI: 10.1057/9780230511910
6. **Bednarek M.** *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. London: Continuum, 2006. 253 p.



7. **Puspita D., Pranoto B.E.** The attitude of Japanese newspapers in narrating disaster events: Appraisal in critical discourse study // *Studies in English Language and Education*. 2021. Vol. 8, No. 2. P. 796–817. DOI: 10.24815/siele.v8i2.18368
8. *Evaluation in Text: Authorial Stance and the Construction of Discourse* / ed. by S. Hunston, G. Thompson. Oxford: Oxford University Press, 2000. 238 p. DOI: 10.1093/oso/9780198238546.001.0001
9. **Moroshkina H.** Specificities of the Mutual Influence of Context and Assessment in French Evaluative Utterances // *Advanced Education*. 2019. Iss. 12. P. 244–248. DOI: 10.20535/2410-8286.148506
10. **Ingendahl M., Woitzel J., Propheter N., Wänke M., Alves H.** From Deviant Likes to Reversed Effects: Re-Investigating the Contribution of Unaware Evaluative Conditioning to Attitude Formation // *Collabra: Psychology*. 2023. Vol. 9, Iss. 1. Art. no. 87462. DOI: 10.1525/collabra.87462
11. **Zhang D.** Evaluation in Context // *Australian Journal of Linguistics*. 2017. Vol. 37, Iss. 1. P. 76–80. DOI: 10.1080/07268602.2015.1091282
12. **Moran T., Nudler Y., Bar-Anan Y.** Evaluative Conditioning: Past, Present, and Future // *Annual Review of Psychology*. 2023. Vol. 74. P. 245–269. DOI: 10.1146/annurev-psych-032420-031815
13. **Kareem R., Farhan H.** The Language of Evaluation in Jose Saramago's *Blindness* via Appraisal Theory // *International Linguistics Research*. 2022. Vol. 5, No. 1. P. 25–36. DOI: 10.30560/ilr.v5n1p25
14. **Lim W., Wu Q.** Vague language and context dependence // *Frontiers in Behavioral Economics*. 2023. Vol. 2. Art. no. 1014233. DOI: 10.3389/frbhe.2023.1014233
15. **Tilakaratna N., Mahboob A.** Appraisal in the time of conflict // *Linguistics and the Human Sciences*. 2013. Vol. 8, Iss. 1. P. 63–90. DOI: 10.1558/lhs.v8i1.63
16. **Liu H., Cao Y., Wu X., Qiu C., Gu J., Liu M., Hershcovich D.** Towards realistic evaluation of cultural value alignment in large language models: Diversity enhancement for survey response simulation // *Information Processing & Management*. 2025. Vol. 62, No. 4. Art. no. 104099. DOI: 10.1016/j.ipm.2025.104099
17. **Gupta V., Chowdhury S.P., Zouhar V., Rooein D., Sachan M.** Multilingual Performance Biases of Large Language Models in Education // *arXiv:2504.17720v2*. 2025. DOI: 10.48550/arXiv.2504.17720
18. **Hengle A., Bajpai P., Dan S., Chakraborty T.** Can LLMs reason over extended multilingual contexts? Towards long-context evaluation beyond retrieval and haystacks // *arXiv:2504.12845v1*. 2025. DOI: 10.48550/arXiv.2504.12845
19. **Huang Z., Zhu W., Cheng G., Li L., Yuan F.** MindMerger: Efficient Boosting LLM Reasoning in non-English Languages // *arXiv:2405.17386v1*. 2024. DOI: 10.48550/arXiv.2405.17386
20. **Diethel D., Colley A., Wienert J., Schoning J.** Different Length, Different Needs: Qualitative Analysis of Threads in Online Health Communities // 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). Rochester, MN, 2022. P. 348–356. DOI: 10.1109/ICHI54592.2022.00056
21. **Van Dijk T.A.** *Discourse and Context: A Sociocognitive Approach*. Cambridge: Cambridge University Press, 2008. 283 p.
22. *Social Media and Society: Integrating the digital with the social in digital discourse* // ed. by M. KhosraviNik. John Benjamins Publishing Company, 2023. 210 p. DOI: 10.1075/dapsac.100
23. **Manca E.** *Context and Language*. Università del Salento, 2012. 105 p.
24. **Felbo B., Mislove A., Søgaard A., Rahwan I., Lehmann S.** Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm // *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, 2017. P. 1615–1625. DOI: 10.18653/v1/D17-1169
25. **Sykes J.M.** Emergent Digital Discourses: What Can We Learn From Hashtags and Digital Games to Expand Learners' Second Language Repertoire? // *Annual Review of Applied Linguistics*. 2019. Vol. 39. P. 128–145. DOI: 10.1017/S0267190519000138
26. **Zappavigna M., Logi L.** *Emoji and Social Media Paralanguage*. Cambridge: Cambridge University Press, 2024. 252 p. DOI: 10.1017/9781009179829
27. **Hadour T.** The pragmatics of hashtags in French tweets // *Internet Pragmatics*. 2025. Vol. 8, Iss. 1. P. 86–112. DOI: 10.1075/ip.00117.had
28. **Du Bois J.W.** Towards a dialogic syntax // *Cognitive Linguistics*. 2014. Vol. 25, Iss. 3. P. 359–410. DOI: 10.1515/cog-2014-0024
29. **Pöldvere N., Johansson V., Paradis C.** Resonance in dialogue: The interplay between intersubjective motivations and cognitive facilitation // *Language and Cognition*. 2021. Vol. 13, Iss. 4. P. 643–669. DOI: 10.1017/langcog.2021.16



30. **Kamsinah N.N., Nuraziza A.** Pragmatic Analysis in Digital Communication: A Case Study of Language Use on Social Media // *International Journal of Economics, Commerce, and Management*. 2024. Vol. 1, Iss. 4. P. 375–383. DOI: 10.62951/ijecm.v1i4.259
31. **Eragamreddy Dr.N.** Exploring Pragmatics: Uncovering the Layers of Language and Meaning // *International Journal of Current Science Research and Review*. 2024. Vol.7, No. 3. P. 1886–1895. DOI: 10.47191/ijcsrr/V7-i3-50
32. **Eke C., Asak M., Adeyem O.** Pragmatics and Communication in the Digital Age: A Focus on the Nigerian Experience // *Advance Journal of Linguistics and Mass Communication*. 2025. Vol. 9, No. 2. P. 34–52. URL: <https://aspjournals.org/Journals/index.php/Ajlmc/article/view/1057> (дата обращения: 08.12.2025).
33. **Herring S.C.** The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis // *Analyzing Digital Discourse* / ed. by. P. Bou-Franch, P. Garcés-Conejos Blitvich. Cham: Palgrave Macmillan, 2019. P. 25–67. DOI: 10.1007/978-3-319-92663-6_2
34. **Kopf S.** Unravelling social media critical discourse studies (SM-CDS) – four approaches to studying social media through the critical lens // *Critical Discourse Studies*. 2025. P. 1–18. DOI: 10.1080/17405904.2025.2463622
35. **Bisk Y., Holtzman A., Thomason J., Andreas J., Bengio Y., Chai J., Lapata M., Lazaridou A., May J., Nisnevich A., Pinto N., Turian J.** Experience Grounds Language // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online, 2020. P. 8718–8735. DOI: 10.18653/v1/2020.emnlp-main.703
36. **Sap M., Gabriel S., Qin L., Jurafsky D., Smith N.A., Choi Y.** Social Bias Frames: Reasoning about Social and Power Implications of Language // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online, 2020. P. 5477–5490. DOI: 10.18653/v1/2020.acl-main.486
37. **Asher N., Lascarides A.** *Logics of Conversation*. Cambridge: Cambridge University Press, 2005. 526 p.
38. **Aquino Y., Bento P., Buzelin A., Dayrell L., Malaquias S., Santana C., Estanislau V., Dutenhefner P., Evangelista G.H.G., Porfirio L.G., Grossi C.S., Rigueira P.B., Almeida V., Pappa G.L., Meira W.** Discord Unveiled: A Comprehensive Dataset of Public Communication (2015–2024) // *arXiv:2502.00627v1*. 2025. DOI: 10.48550/arXiv.2502.00627
39. **Ivorra-Pérez F.M., Giménez-Moreno R.G.-M.** The level of context dependence of engagement markers in Peninsular Spanish and US business websites // *Revista de Lenguas Para Fines Específicos*. 2018. Vol. 24.2. P. 38–53. DOI: 10.20420/rlfe.2018.364
40. **Lage A.H., Diego Á.R.** La oralización de textos digitales: Usos no normativos en conversaciones instantáneas por escrito // *Caracteres: Estudios Culturales y Críticos de La Esfera Digital*. 2013. Vol. 2, Núm. 2. P. 92–108.
41. **Alarcón Silva M., Cárdenas Neira C.** Convocatoria de protesta a través de Instagram, un análisis socio cognitivo de estrategias discursivas en el contexto del movimiento social en Chile (2019–2020) // *Revista Latina de Comunicación Social*. 2021. Núm. 79. P. 127–149. DOI: 10.4185/RLCS-2021-1524
42. **Deas N., Turcan E., Mejia I.E.P., McKeown K.** MASIVE: Open-Ended Affective State Identification in English and Spanish // *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, FL, 2024, pp. 20467–20485. DOI: 10.18653/v1/2024.emnlp-main.1139
43. **Millán-Franco M., Domínguez De La Rosa L., Gómez-Jacinto L., Hombrados-Mendieta M.I., García-Cid A.** Análisis discursivo sobre el sentido de comunidad de los latinoamericanos en Málaga // *Migraciones internacionales*. 2021. Vol. 12. Art. núm. 5. DOI: 10.33679/rmi.v1i1.2137
44. **Águila Díaz J.** Los efectos perniciosos de una identidad colectiva excluyente. El caso del rap español. Encrucijadas // *Revista Crítica de Ciencias Sociales*. 2024. Vol. 23, Núm. 3. Art. núm. v2301. URL: <https://recyt.fecyt.es/index.php/encrucijadas/article/view/96747> (дата обращения: 08.12.2025).
45. **Falconi Calvachi F.R., Velásquez Cajas Á.P., Lema Flores K.M., Lomas Chacón P.E.** Análisis crítico del discurso intercultural en Facebook. Caso ‘Conaie Comunicación’ // *Culturas Revista de Gestión Cultural*. 2021. Vol. 8, Núm. 2. P. 101–122. DOI: 10.4995/cs.2021.16529
46. **Krippendorff K.** *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA; London; New Delhi: SAGE Publications, Inc., 2019. 422 p. DOI: 10.4135/9781071878781
47. **McHugh M.L.** Interrater reliability: The kappa statistic // *Biochemia Medica*. 2012. Vol. 22. P. 276–282. DOI: 10.11613/bm.2012.031



48. **Cao Y., Zhou L., Lee S., Cabello L., Chen M., Hershcovich D.** Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study // Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP) / ed. by S. Dev, V. Prabhakaran, D. I. Adelani, D. Hovy, L. Benotti, Dubrovnik: Association for Computational Linguistics, 2023. P. 53–67. DOI: 10.18653/v1/2023.c3nlp-1.7
49. **Ashrafimoghari M., Doroudi S., Tabatabaei S.M.S., Valipour M.** Comparative analysis of GPT-4, Gemini AI, and Claude: Strengths and weaknesses in content generation // SSRN, 2024. DOI: 10.2139/ssrn.4681307
50. **Moroni L., Aula-Blasco J., Conia S., Baucells I., Perez N., Suárez S.P., Sallés A., Ostendorff M., Falcão J., Son G., Gonzalez-Agirre A., Navigli R., Villegas M.** Multi-LMentry: Can Multilingual LLMs Solve Elementary Tasks Across Languages? // Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. Suzhou, 2025. P. 34114–34145. DOI: 10.18653/v1/2025.emnlp-main.1731
51. **Schut L., Gal Y., Farquhar S.** Do Multilingual LLMs Think In English? // arXiv:2502.15603v1. 2025. DOI: 10.48550/arXiv.2502.15603
52. **Anglin K., Boguslav A., Hall T.** Improving the Science of Annotation for Natural Language Processing: The Use of the Single-Case Study for Piloting Annotation Projects // Journal of Data Science. 2022. Vol. 20, Iss. 3. P. 339–357. DOI: 10.6339/22-JDS1054
53. **Artstein R., Poesio M.** Inter-Coder Agreement for Computational Linguistics // Computational Linguistics. 2008. Vol. 34, Iss. 4. Art. no. 4. DOI: 10.1162/coli.07-034-R2

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Sergei Sikorskii

Сикорский Сергей

E-mail: ssikors@upv.es

ORCID: <https://orcid.org/0009-0009-1874-4394>

Maria Luisa Carrio-Pastor

Каррио Пастор Мария Луиса

E-mail: lacarrio@upv.es

ORCID: <https://orcid.org/0000-0002-3040-5362>

Submitted: 24.07.2025; Approved: 30.11.2025; Accepted: 05.12.2025.

Поступила: 24.07.2025; Одобрена: 30.11.2025; Принята: 05.12.2025.