ANÁLISIS DEL DISCURSO PARA EL DESARROLLO DE UN MODELO DE DETECCIÓN DE CIBERGROOMING EN ROBLOX

Ana Paola Castañón Marroquín, Brenda Ailed Rodríguez Colis, Andrea Bazán Durán, Luis Enrique Colmenares-Guillen

https://doi.org/10.1590/SciELOPreprints.16320

Envíado en: 2026-05-28
Postado en: 2026-05-28 (versión 1)
(AAAA-MM-DD)

Research Article

DISCOURSE ANALYSIS FOR THE DEVELOPMENT OF A CYBERGROOMING DETECTION MODEL ON ROBLOX

ANÁLISIS DEL DISCURSO PARA EL DESARROLLO DE UN MODELO DE DETECCIÓN DE CIBERGROOMING EN ROBLOX

Ana Paola Castañón Marroquín[1], ORCID: https://orcid.org/0009-0007-3411-0418
Brenda Ailed Rodríguez Colis [2], ORCID: https://orcid.org/0009-0002-2694-6542
Andrea Bazán Durán [3], ORCID: https://orcid.org/0009-0001-7722-1878
Luis Enrique Colmenares-Guillen[4], ORCID: https://orcid.org/0000-0002-9921-8813

[1]Benemérita Universidad Autónoma de Puebla, Facultad de Filosofía y Letras, Licenciatura en Lingüística y Literatura Hispánica, Puebla, México, email: ana.castanonm@alumno.buap.mx
[2]Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ingeniería en Ciencias de la Computación, Puebla, México, email: brenda.rodriguezco@alumno.buap.mx
[3]Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ingeniería en Ciencias de la Computación, Puebla, México, email: andrea.bazand@alumno.buap.mx
[4]Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Coordinador del Laboratorio de Análisis Forense Digital, Puebla, México, email: enrique.colmenares@correo.buap.mx

ABSTRACT
Cybergrooming represents a growing threat on online gaming platforms such as Roblox, where anonymity and frequent interaction among child users create conditions conducive to child abuse and sexual harassment. The objective of the research that led to this article was to identify linguistic patterns in the discourse of groomers in Spanish-speaking Roblox communities and incorporate them into a computational model for the automatic detection of this cybercrime through text. To this end, a mixed-methods approach was developed, integrating Corpus-Assisted Discourse Studies with the CRISP-DM data mining methodology. A specialized corpus of 25 conversations was compiled and processed, then subjected to detailed analysis. As a main result, a pattern of discursive organization consisting of a sequence of seven conversational modules with specific predictive value and a set of 21 functional lexicogrammatical patterns with 224 associated collocations were identified, described, and subsequently incorporated into a text classification model capable of distinguishing grooming conversations with 93.33% accuracy. In this way, the study demonstrated the efficacy of discourse analysis as a basis for the development of systems for the automatic detection of cybercrimes against minors.
KEYWORDS: Discourse analysis, computer crime, child abuse, videogame, computational linguistics.

RESUMEN
El cibergrooming representa una amenaza creciente en plataformas de juego en línea como Roblox, donde el anonimato y la frecuente interacción de usuarios infantiles generan condiciones propicias para el acoso sexual y abuso de menores. El objetivo de la investigación que originó este artículo fue identificar patrones lingüísticos en el discurso de groomers de comunidades hispanohablantes de Roblox e incorporarlos a un modelo computacional para la detección automática de este delito informático a través del texto. Para ello se construyó una metodología de enfoque mixto que integró los Estudios del Discurso Asistidos por Corpus y la metodología de minería de datos CRISP-DM. Se compiló y procesó un corpus especializado de 25 conversaciones

que fue sometido a un análisis pormenorizado. Como resultado principal, se delimitó y describió un patrón de organización discursiva constituido por una secuencia de siete módulos conversacionales con un valor predictivo determinado y 21 patrones léxico-gramaticales funcionales con 224 colocaciones asociadas, elementos integrados a un modelo de clasificación textual capaz de distinguir conversaciones de grooming con un 93.33% de exactitud. De esta forma, el estudio demostró la efectividad del análisis discursivo como base para el desarrollo de sistemas de detección automática de delitos informáticos contra menores.
PALABRAS CLAVE: *Análisis del discurso, delito informático, abuso de menores, videojuego, lingüística computacional.*

Recibido: (07/03/2026)
Aceptado: (27/05/20026)

INTRODUCTION

The prolonged use of internet-enabled electronic devices among children and adolescents is a growing phenomenon that has transformed the social dynamics and access to information. In this regard, Red Grooming Latam (2024) reported that Latin American minors spend an average of four hours a day online, split between forums, messaging applications, social media, and videogames, which are of particular interest to this study.

Online gaming platforms have established a strong presence as significant spaces for social interaction within youth culture (Morreale & Rosa, 2024), a phenomenon related both to the integration of chat and voice channels into their interfaces (Excelin et al., 2024), and to the formation of user communities on external applications (Pyslar, 2025). A paradigmatic example is Roblox, a platform that enables the creation of gaming experiences shared by millions of users (Kou et al., 2025; Red Grooming Latam, 2024).

However, interacting with strangers in virtual environments has led to the emergence of harmful behaviors and cybercrimes (Rozgonjuk et al., 2023), such as cybergrooming, a practice in which an adult contacts minors through digital platforms with the aim of involving them in sexual activity or obtaining sexual content (Lorenzo-Dus, 2023; Ortiz, 2024). In particular, cases of grooming associated with the use of Roblox have been reported and have led to legal complaints (Carville & D'Anastasio, 2024; Dorsey, 2024).

As a form of child abuse, online grooming has been extensively studied from a behavioral science perspective, although research adopting a discursive approach is scarce (Lorenzo-Dus, 2023). Nevertheless, the practice of grooming materializes by means of a series of communicative exchanges between adults and minors.

This study performed the first analysis of the conversational discourse of groomers in Spanish-speaking Roblox virtual communities. To this end, in line with previous research (Broome et al., 2025; Joleby et al., 2021; Lorenzo-Dus et al., 2020; Lorenzo-Dus & Izura, 2017; Pienczykowski & Madella, 2026), the procedures of Corpus-Assisted Discourse Studies, as defined in the work of Gillings et al. (2024), were adopted. The aim was to identify linguistic patterns at the level of conversational organization and lexico-grammatical resources with functional value (Halliday & Matthiessen, 2004; Vaamode & González, 2008) associated with pragmatic-discursive strategies.

Finally, a computational model was developed using the CRISP-DM methodology, in which the detected patterns were organized into structures suitable for use in automated text analysis (Manning et al., 2009): a sequence of conversational modules with a specific predictability percentage and collocations or collocation pairs (Daohuan & Xuri, 2023; Schweinberger, 2024) associated with the patterns of each module.

METHODOLOGY

This study adopted a mixed-methods approach (Molina et al., 2024) that combined two distinct methodologies: Corpus-Assisted Discourse Studies or CADS for discursive analysis, and the CRISP-DM data analysis methodology for the development of the computational model. The result of this integration is presented in Figure 1.

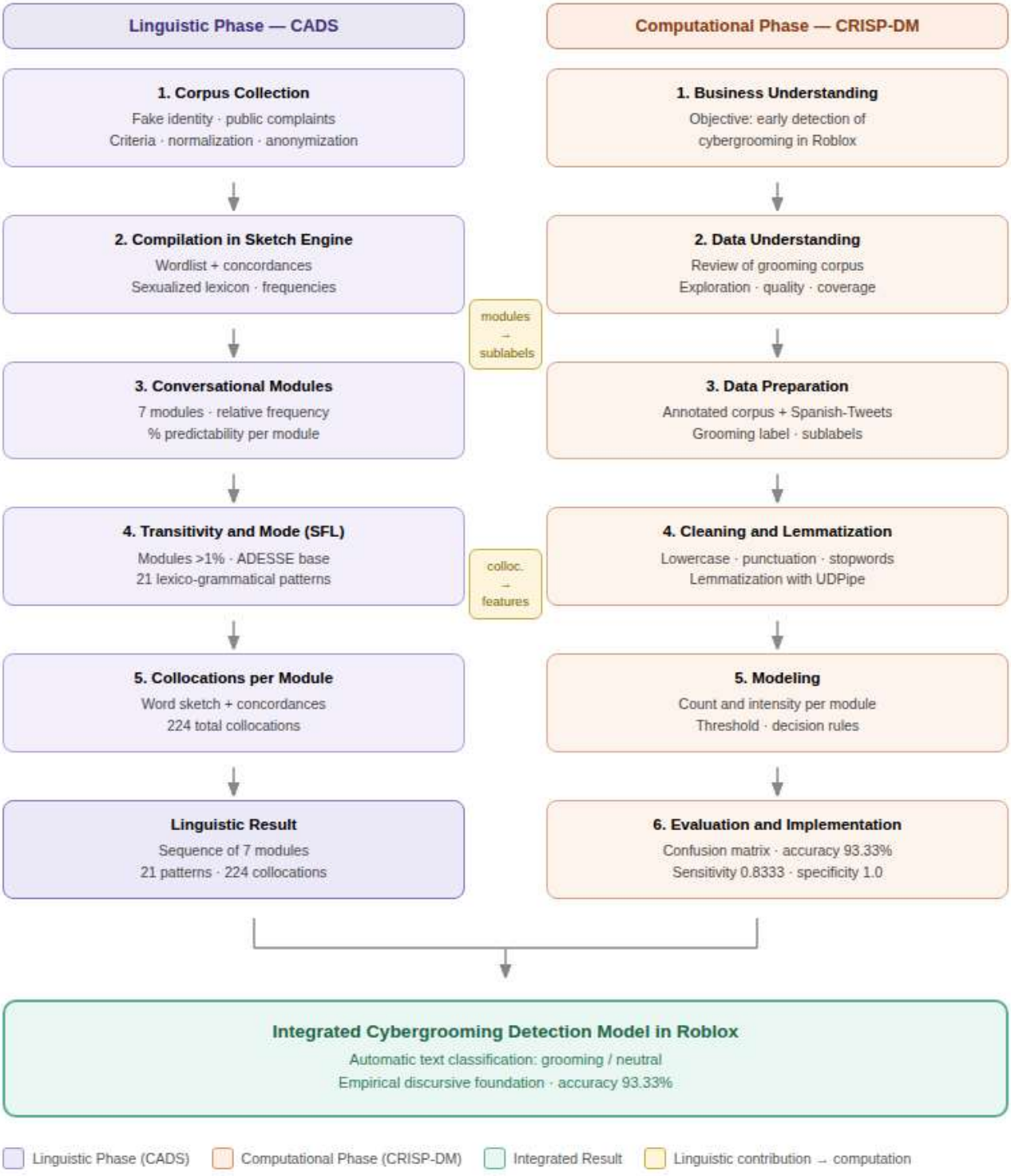

Figure 1: Diagram of the integrated CADS ad CRISP-DM methodology and its implementation in the study

CADS is a mixed-methods paradigm that employs corpus linguistics tools in discourse analysis, namely the study of language in use (Gillings et al., 2024; Partington et al., 2013). In the first phase, this methodology requires the identification of high-frequency linguistic patterns, and in the second phase, the qualitative interpretation of discourse characteristics (Gillings et al., 2024).

In conceptualizing cybergrooming as a distinct discursive practice, two theoretical postulates were adopted. First, the definition proposed by Van Dijk (1997) of discourse as a complex and bounded set of linguistic constructions with effects on context, conceived as "the structure of all properties of the social situation that are relevant for the production or the reception of discourse" (p. 19). Second, the characterization offered by Lorenzo-Dus (2023) of digital grooming as a form of manipulative discourse with specific resources and strategies.

For pattern identification, Sketch Engine was employed, a software program specialized in the management and analysis of corpus or large collections of natural language data (Lutzuky & Kehoe, 2022), which enabled the generation of statistically representative results (Isti'anah et al., 2023) and the determination of their relative frequency.

In this study, a linguistic pattern as a recurring choice of form and meaning (Hunston, 2010), was examined at two levels: as a unit that reflects a regular discursive organization, and as a set of valency structures and preferred lexical collocations associated with a particular meaning or implicature (Hunston, 2025).

Regarding the first level, drawing on the concept of sequential organization proposed by Schegloff (2007) and considering that discourse may be analyzed "in terms of a number of typical formal categories and their specific order and function" (Van Dijk, 1997, p. 37), it was posited that a series of communicative acts or interventions —referred to here as conversational modules— bearing a specific discursive or thematic function, constitute a pattern.

With respect to the second level, categories from Systemic Functional Linguistics were employed to detail the valency structure that configures the pattern, following the semantic role annotation methods for patterns proposed by Hunston (2025). This theoretical framework is conceived as a tool for the socio-semiotic study of language (Halliday & Matthiessen, 2004), thereby rendering it compatible with concepts from discourse analysis and pragmatics drawn upon in the second phase of CADS.

Within this theory, the unit of analysis is the clause, a construction with a conjugated verb in which three metafunctions are realized at the lexico-grammatical level (Halliday & Matthiessen, 2004). In particular, the experiential component of the ideational metafunction was employed within the transitivity subsystem, where six types of verbal processes and their participants are identified, as defined in Table 1 based on Halliday and Matthiessen (2004) and the ADESSE database project reviewed by Vaamode and González (2008).

Table 1: Types of verbal processes and their participants in the systemic-functional approach

Verbal process type	Description	Participants
Material	An entity capable of physical action causes a visible process.	Actor Goal Circumstance Beneficiary Scope
Mental	An entity endowed with psychic life experiences a state, change of state, or inner perceptive, sensitive, and/or cognitive activity.	Experiencer Stimulus
Relational	An entity is related to another entity, property, function, or any type of fact that allows the entity to be characterized.	Entity Attribute
Verbal	An entity endowed with communicative capacity transfers information through any semiotic system to another entity.	Sayer Receiver Verbiage
Existential	An entity, property, or process is present or absent, exists or lacks existence.	Being Initiator Circumstance
Behavioral	A living being, prototypically human, experiences or performs a process that manifests externally and is related to the psycho-biological and/or interpersonal or social existence of the being.	Behaver Behavior

Furthermore, the interpersonal metafunction was addressed through the mode subsystem, which distinguishes three discursive realizations: declarative, imperative, and interrogative (Halliday & Matthiessen, 2004). In addition to collocations, high-frequency lexical pairs that occur within the same construction (Daohuan & Xuri, 2023; Schweinberger, 2024), these elements completed the structure of the patterns.

The corpus analyzed was composed of 25 conversations collected by means of a convenience sampling technique (Molina et al., 2024), confirming that the conversations complied the following selection criteria: relation to Roblox, explicit request for sexual content, and participation of a fictional or real minor.

Corpus compilation involved the systematic search for public reports on social media documenting, by means of conversations, cases of child abuse in Roblox communities, as well as the creation of fictional minor profiles to interact with potential groomers in public and private social media groups related to the videogame, registering the conversations that met the established criteria.

The conversations were transcribed, anonymized by removing direct identifiers of the interlocutors, and normalized through orthographic correction, translation of anglicisms, and completion of contextually relevant apocopated forms. Subsequently, the complete corpus was compiled in Sketch Engine to facilitate both the qualitative analysis aimed at delimiting the conversational modules and the identification of linguistic patterns by means of the specific tools of the software in the next stage.

Conversational modules, clauses, the predictability percentage —a value established for the automatic detection of each module, where a zero percentage confirms that constitutes a neutral scheme not directly associated with cybergrooming —, relevant conversational modules —those exceeding 1% predictability—, and the interpretation of discourse through the identification of pragmatic-discursive strategies (Calsamiglia & Tuson, 2012), were delimited and developed qualitatively. In contrast, the calculation of the relative frequencies of these elements, as well as the keyword list automatically generated by means of Sketch Engine's Wordlist, Word Sketch, and

Concordances tools —consulted for the delimitation of modules, patterns, and collocations — correspond to the quantitative dimension of the approach.

For the development of the computational model, the CRISP-DM data mining methodology was adopted. According to Mariscal et al. (2010), this methodology consists of six interrelated phases applied in the research as detailed in Figure 1.

In the data preparation phase of CRISP-DM, a dataset was constructed. This is a collection of data with observations in a structured format that facilitates storage, retrieval, and automatic analysis (Badman & Kosinski, 2024). This collection was created by integrating a sample of 50 observations from Spanish-Tweets (Pérez et al., 2022), a large, open-source, public dataset hosted on Hugging Face by the Pysentimiento Project (McDonough, 2023) which compiles Spanish-language tweets characterized by their neutrality and diversity of regional and thematic variants. Finally, the dataset was supplemented with the conversational corpus.

Two fields were also incorporated for each unit of analysis in grooming conversations: an identifier and the associated text. This structure, visualized in Table 2, retained the original text sequence, and facilitated the subsequent stages of data cleaning, computational analysis, and modeling.

Table 2: Structure of the conversational units processed for the cybergrooming dataset

Field	Description
id_conversation	Unique identifier for the conversation.
text	Complete transcript of the message, already processed, from the sender.

Regarding the dataset cleaning process, this included converting text to lowercase, removing punctuation marks and grammatical categories without lexical value, as well as lemmatization, a widely used text preprocessing technique in natural language processing —an area of computational linguistics— that performs automatic morphological analysis to convert a word to the base form or lemma, reducing textual complexity (Manning et al., 2009). Lemmatization was performed using the *udpipe* library (Straka & Straková, 2020), which removes prefixes and suffixes from words.

The following techniques and metrics were applied during the model evaluation and implementation phases: confusion matrix, to examine correct and incorrect classifications by class; accuracy, as an overall measure of performance; recall and specificity, to evaluate performance differences between positive and negative classes (IBM, 2025; Steward, 2023; Ting, 2010); as well as the visualization of a classification test.

Combining the conversational corpus with the sample obtained from social media creates a dataset containing different types of language and contexts, which could introduce biases in the evaluation of the model. Therefore, this validation strategy was established to measure the performance of the classifier and the ability to adapt to the differences between the two data sources within the test set.

RESULTS AND DISCUSSION

A sequence of seven conversational modules was identified in the conversations of the corpus, described and exemplified in Table 3, which also specifies their relative frequency and predictability percentage. This sequence is presented as a pattern of discursive organization characteristic of digital groomers in Spanish-speaking Roblox communities. Moreover, three to five patterns were obtained for the relevant modules, contained in Table 4, which also presents a sample of 10 of the 224 total collocations associated with the modules.

Table 3: Conversational modules of cybergrooming on Roblox

	Conversational module	Description	Examples (No. of conversation + Text)	RF	Predictability percentage
1	Salutation	The groomer initiates contact with the minor through an adjectivized salutation, a salutation with politeness formulae, or a personal introduction.	20. Hola bonita 25. Hola cómo estás qué haces? 1. Hola soy el administrador	76%	0%
2	Personal information request	The groomer elicits personal information from the minor, such as name, age, daily habits, or family background.	1. Me dices tu edad y nombre para anotarte 14. Y eres hija única o tienes más hermanos? 8. Tienes horario? / es decir te dicen tus padres a qué hora dejar el cel y eso	76%	17%
3	Robux or reward offer	The groomer offers the minor some type of reward related to Roblox.	10. Mira estoy buscando a una chica para consentir con robux 18. Oye y no te gustaría tener robux? 4. Estás interesado en los 200 robux	96%	22%
4	Sexual content request	The groomer solicits photographs, videos, or video calls of a sexual nature from the minor.	3. Te daré los 2000 a cambio de videos / videos de ti, sin un poco de ropa 5. Ok te propongo algo si en videollamada haces todo lo que te diga te daré varios robux 7. Manda foto / ahora la vagina porfis	100%	28%
5	Resistance management	The groomer produces statements aimed at persuading the minor to agree to provide sexual content or more content than already provided, such as new offers or mitigated reformulations of previous constructions.	14. Te puedo seguir dando robux muy seguido 4. Acá tengo pruebas / si no te preocupes de eso 7. Porfis / no te dé pena mostrármelo confía / no seas malita	92%	21%
6	Compliments	The groomer produces positive evaluative statements that appeal to the minor's physical traits, personality, or actions.	20. Qué bonita estás amor 18. Asi, asi, qué guapa estás, muy hermosa 1. Muy bien / me caes bien	40%	12%
7	Closing	The groomer says goodbye, thanks the child or threatens them.	4. Bueno amigo gracias por confiar 20. Pinche putita te voy a buscar y te voy a matar ya sé dónde vives	32%	0%

The first module corresponds to the salutation, presented in 76% of conversations. This speech act is also referred to as an opening frame sequence, which includes interrogative formulas or exclamatory expressions that serve a basic interpersonal function: opening and acknowledging the existing relationship between the two dialogue participants (González-Sanz, 2024). Within the corpus, this module evidenced the definition of the interactional relationship through the calibration of parameters such as social position and degree of intimacy (Calsamiglia & Tuson, 2012; González-Sanz, 2024), as corroborated by the use of vocatives or adjectivized appellatives as rapport-building strategies.

Additionally, in some cases the salutation was accompanied by personal introductions aimed at projecting an apparently legitimate identity of the perpetrator, appealing to the strategic configuration of a face or image in the terms of Goffman (Walsh, 2022; Calsamiglia & Tuson,

2012). Together, these resources foster the development of a trust bond with the minor and initiate manipulation (Sorlin, 2017) in the access phase of the model of Lorenzo-Dus (2023).

Table 4: Exemplification of linguistic patterns and collocations from the cybergrooming corpus on Roblox

Module	Pattern	Example (No. of conversation + Text)	Sample of collocations
Personal information request	A	1. Me dices tu edad y nombre para anotarte	decir edad decir nombre tener año estar solo querer conocer tener hermano habitación solo qué edad cuál nombre tu familia
	B	10. Qué edad tienes? / eres atrevida?	
	C	1. Con quién duermes	
	D	20. Y qué te gusta jugar?	
	E	1. Yo soy de estados unidos y tú	
Robux or reward offer	A	20. Te gustaría ganarte robux	dar robux querer robux buscar robux buscar chico interesar robux robux video robux foto reto robux robux cambio cuántos robux
	B	4. Estás interesado en los 200 robux?	
	C	10. Estoy regalando tarjetas de robux	
	D	24. Yo te doy 200 robux a tu cuenta	
Sexual content request	A	12. Toma una foto abriendo mucho la boca	mandar foto mandar video pedir foto hacer videollamada enseñar tu tu cuerpo tu parte primero luego sin ropa foto intima
	B	23. Mándame unos videos con ropa interior	
	C	3. Te daré los 2000 a cambio de videos/ videos de ti, sin un poco de ropa	
	D	18. Me gustaría que me mandaras una foto tuya pero como si te fueras a ir a dormir	
	E	2. El primer desafío será mandar una foto tuya	
Resistance management	A	18. Y por 500 no te animarías?	ser normal ser fácil dar más confiar mi no preocupar más robux solo foto entre nosotros robux cuenta tu decisión
	B	10. No me digas eso	
	C	12. Algo simple/ es fácil	
	D	D. 18. Además, te daría los robux ahorita mismo	
Compliments	A	20. Qué bonita estás amor	ser guapo estar hermoso estar guapo ver hermoso caer bien excelente bien muy hermoso muy bonito niño hermoso mi amor
	B	8. Tu personalidad es inocente en parte y madura/ aggg qué linda	
	C	1. Excelente muy bien/ me caes bien	

The second conversational module presented a frequency of 76% and was designated as the personal information request. In line with the two subsequent modules, it was named in terms of the illocutionary force of a speech act with a greater or lesser degree of indirectness that threatens

the groomer's face (Calsamiglia & Tuson, 2012; Walsh, 2022); in this case, a request whose primary function is the solicitation of personal data from the victim. Five patterns were identified:

- Pattern A, with a relative frequency of 3%, in which the groomer employs a verbal process that positions the minor as the sayer of personal data or information, with the groomer as the receiver.
- Pattern B, with a frequency of 68%, configures the minors as entities that assign themselves attributes or personal traits, both psychological and physical, by means of a relational process.
- Pattern C, which inquires into the habits and family context of the minor by means of a behavioral process, related to risk assessment for isolation (Lorenzo-Dus et al., 2020), and with a frequency of 3%.
- Pattern D, with a frequency of 12%, explores the preferences of the minor as a stimulus regularly experienced through a mental process of sensation or perception.
- Pattern E, identified in constructions with a greater degree of indirectness, whereby the groomer discloses potentially false personal information (Pienczykowski & Madella, 2026) by means of a relational process, which presented a frequency of 6%.

The most frequent mode of discursive realization was the interrogative, with 82%, as visualized in Figure 2 alongside the relative frequency data of the three modes analyzed in relevant modules.

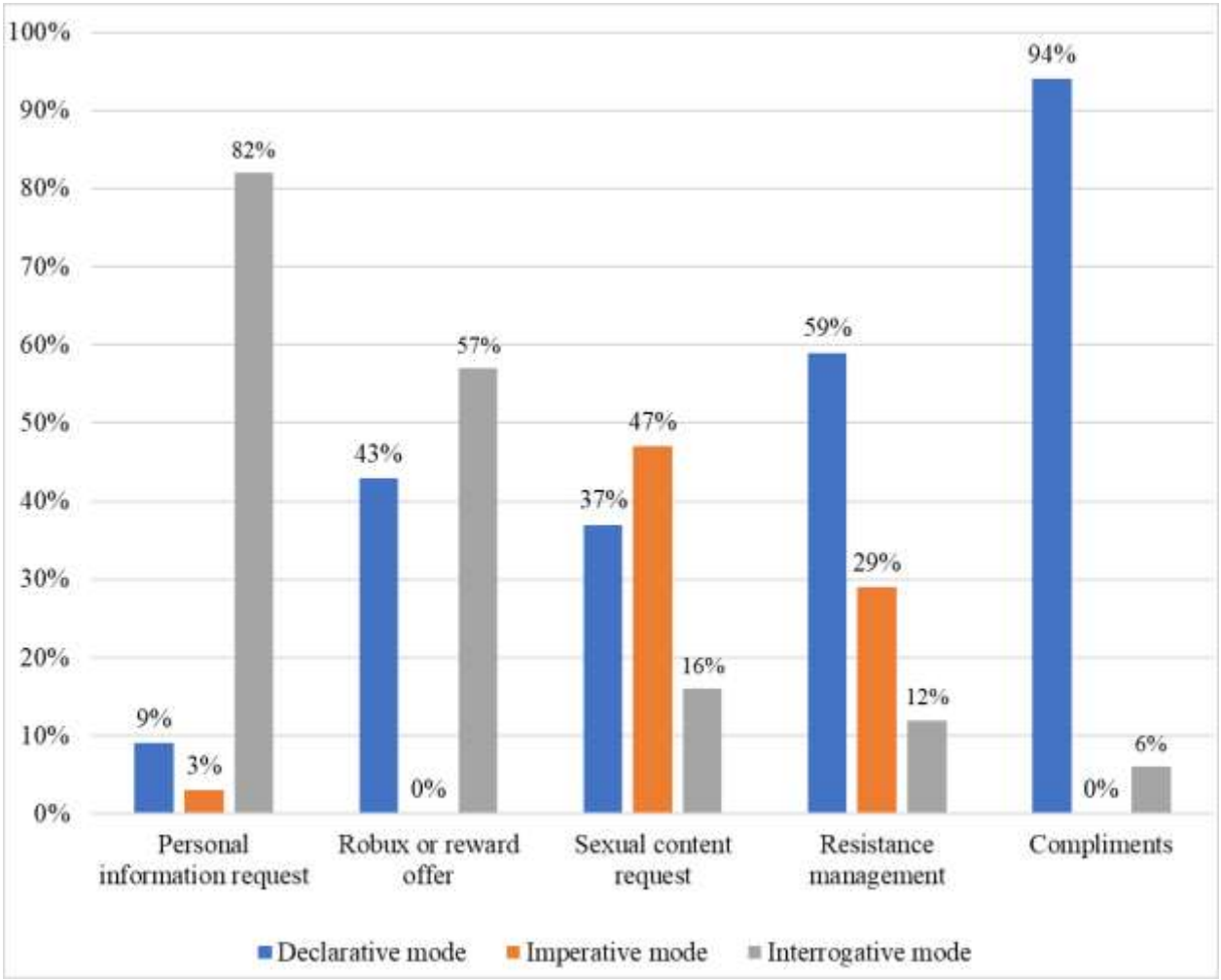

Figure 2: Graph of the frequency of the modes of discursive realization in relevant conversational modules of cybergrooming on Roblox

The third module, Robux —the virtual currency of the videogame— or rewards offer, presented a frequency of 96%. This constitutes the most relevant contribution of this study and particularizes the cybergrooming discourse on Roblox. Contrary to the implicit offer of affective goods —such

as validation and recognition— in this context, the incentives are material and related to the dynamics of the videogame. The offer of these rewards is "conditional and thereby put pressure on the child" (Joleby et al., 2021, p.10). Four patterns were identified:

- Pattern A, presented in 34% of the conversations, entails a mental process of sensation that configures the minor as the experiences of diverse, though generally positive, internal states in response to a possible reward.
- Pattern B, with a frequency of 11%, presents an attributive relational process that posits a positive state related to the possession of Robux or rewards.
- Pattern C, with a frequency of 17%, involves a behavioral process performed by the groomer and employed for indirect speech acts that trigger conversational implicatures or situated inferential processes that speakers activate to understand utterances (Calsamiglia & Tuson, 2012); in this case, to interpret the clause as an offer.
- Pattern D, presented in 29% of the conversations, consists of a relational process of transference with the groomer as the entity or possessor and the reward as the attribute.

The interrogative mode presented a relative frequency of 57% and the declarative mode of 43%, the latter corresponding to indirect offers.

The fourth conversational module, sexual content request, was present in 100% of the corpus and is the defining scheme of the discursive practice of cybergrooming.

Five patterns were identified in this module:

- Pattern A, identified in 40% of the conversations, consists of a material process of creation that positions the minor as the actor that produces sexual content with the groomer as the beneficiary.
- Pattern B, with a relative frequency of 21%, in which a relational process of transference or possession configures the minor as the possessor or entity capable of providing sexual content.
- Pattern C, with a frequency of 19%, also entails a relational process of transference or possession, but one in which the groomer is presented as the entity related to positive states derived from the possession of sexual content.
- Pattern D, with a frequency of 7%, constitutes a mental process of sensation in which the groomer is presented as the experiences of positive states derived from the possession of sexual content.
- Pattern E, with a frequency of 7%, comprises a relational process in which a challenge created by the groomer is defined as the action of sending sexual content in the form of an indirect request.

In this module of request, the imperative and declarative modes were employed with greater frequency. Furthermore, this was characterized by the inclusion of both explicit and vague sexual language: explicit terms coincide with the onset of the sexual gratification process (Joleby et al., 2021), while vague terms serve the function of "minimizing the offence (...) and desensitizing the child to sexual communications" (Pienczykowski & Madella, 2026, p. 11), thereby preserving the face of the groomer in face-threatening acts (Calsamiglia & Tusón, 2012).

The penultimate module, with a relative frequency of 92%, is that of resistance management. This module exhibited greater variability in clause types and constructions, as well as different speech acts and an overlap of cyber-grooming phases (Joleby et al., 2021; Lorenzo-Dus, 2023).

This module employs both overt persuasion and coercion strategies (Chiang & Grant, 2019), as well as politeness strategies to improve the image of the victims and attribute to them a sense of responsibility for their actions comparable to that of the aggressor, or what Pienczykowski & Madella (2026) refer to as "role and responsibility reversal" (p. 10).

Moreover, this was presented in the form of hypothetical statements, softened or approximate rephrasings, conditional statements, requests for specific information, or new offers, with the aim of managing or defusing the conflict that arises during the interaction in response to the resistance of the victim (Evans & Lorenzo-Dus, 2025). The four patterns identified were:

- Pattern A with a mental process of sensation or cognition that motivates the child and provides them with a false sense of security by appealing to positive internal states, with a frequency of 22%.
- Pattern B, with a 4% frequency, comprises a verbal process that inquires information regarding the sexual content and the refusal to comply with the request of this content from the minor.
- Pattern C, with a frequency of 22%, entails a relational process that assigns a positive evaluation with an attenuating effect to the action of sending content.
- Pattern D, with a frequency of 19%, presents a relational process of transference in which the groomer offers additional rewards.

The declarative mode was the most common in the resistance management module, accounting for 59% of the clauses in the corpus.

The final relevant conversational module identified, with a relative frequency of 40%, is that of compliments. The function of this module is to build trust by explicitly praising the appearance, abilities, possessions, or personality traits of the victim (Lorenzo-Dus & Izura, 2017).

The degree of directness of this module changed depending on the profile of the aggressor, whether these were hypersexual, with more direct compliments, or intimacy-focused and adaptive, with more indirect compliments (Broome et al., 2025). Three patterns were identified:

- Pattern A, with a frequency of 56%, constitutes a relational process in which positive attributes are assigned to the minor or to some of their physical or psychological traits in the form of a direct compliment which emphasizes maturity or docility to equalize the power relationship.
- Pattern B, with a frequency of 13%, involves a relational process that positively defines a personal trait of the minor in the form of an indirect compliment.
- Pattern C presents a mental process of sensation in which the groomer openly declares their reaction to the victim or to the content as a stimulus.

The most common discursive realization mode in the compliment module was the declarative, which appeared in 94% of the clauses.

Finally, the defined sequence concludes with a closing module that appears in 32% of the conversations. In this section, the groomer engages in actions such as expressing gratitude, making threats, and saying goodbye; or, once the conversation has ended —whether voluntarily or involuntarily— the groomer ceases to interact, thereby violating the cooperative principle proposed by Grice (Calsamiglia & Tuson, 2012)

On the other hand, a computational model was developed capable of classifying conversational text as either cybergrooming or neutral, as well as identifying the relevant thematic module in each instance of grooming text by detecting collocations of the predefined linguistic patterns.

For the performance of the model, cross-validation was employed (Allgaier & Pryss, 2024), which consisted of partitioning the data into k distinct groups where $k - 1$ constituted the training sets. The remaining group contained the data used to evaluate the accuracy; this process was repeated k times until each group had been used as a validation set.

The automatic analysis performed by the developed model was based on the detection of collocations associated with each conversational module. Each time a collocation was identified, the corresponding count was incremented; subsequently, the intensity was calculated by dividing the total number of collocations found by the absolute frequency of the collocation pairs within the module.

Based on these intensities, a prediction function was proposed, illustrated in Figure 3. This function generated a risk score by combining the intensities through a weighted scoring scheme, following the principle described by Manning et al. (2009) for classification models based on linguistic features.

With the score calculated for each conversation, an optimal threshold was established (Lyu & Ishwaran, 2023), the point at which specificity and sensitivity are optimized for detection and classification in each case.

Additionally, based on the patterns identified by means of the linguistic analysis, a conditional decision was implemented that classified a conversation as positive when it included the sexual content request module or at least three relevant conversational modules. These criteria were determined empirically during the model evaluation process.
Overall, a conversation was classified as grooming if the score was equal to or higher than the optimal threshold, or if the conditional described above was met.

$$Pred = (P \geq threshold) \vee (condicional = 1)$$

Figure 3: Prediction function for the cybergrooming detection model on Roblox

Regarding the first method for evaluating the performance of the model, the confusion matrix (Ting, 2010) determines the total number of correct and incorrect predictions. Figure 4 presents the confusion matrix corresponding to the test set and indicates zero false positives, which means that no neutral conversations were classified as grooming.

Confusion Matrix and Statistics

Prediction	Reference 0	1
0	9	1
1	0	5

Figure 4: Confusion matrix of the test set in the cybergrooming detection model on Roblox

Subsequently, accuracy was calculated, which indicates how frequently the model makes a correct prediction (IBM, 2025). The model achieved an accuracy of 0.9333, meaning that the system correctly classified 93.33% of all evaluated conversations. Similarly, as shown in Figure 5, other evaluation metrics were calculated: sensitivity or recall, which measures the correct detection of positive cases, and specificity, which measures the detection of negative cases (Steward, 2023).

Kappa	:	0.8571
Mcnemar's Test P-Value	:	1.000000
Sensitivity	:	0.8333
Specificity	:	1.0000
Pos Pred Value	:	1.0000
Neg Pred Value	:	0.9000
Prevalence	:	0.4000
Detection Rate	:	0.3333
Detection Prevalence	:	0.3333
Balanced Accuracy	:	0.9167

Figure 5: Results of the evaluation metrics of the cybergrooming detection model on Roblox

A sensitivity of 0.8333 indicates that the model correctly identified 83.33% of grooming cases, although it implies that the remaining 16.67% went undetected. A specificity of 1.0 indicates that the model correctly classified all cases that were not grooming, without generating any false positives. Finally, this system was tested with new conversations. The results were organized in an output table that includes the estimated probability interpreted as the risk level, the number of modules identified, and the final binary classification with grooming=yes and neutral=no.

Table 5: Visualization of a prediction test of the cybergrooming detection model on Roblox

id_conversation	Grooming	Grooming_percentage	n_modules	modules_detected
NUEVO_CONV_CA_15	No	9.8	0	
TWEET_30	No	9.8	0	
TWEET_3	No	9.8	0	
TWEET_13	No	9.8	0	
NUEVO_CONV_CA_6	No	9.8	0	
GROOM_CONV_CG_12	Yes	65.3	1	resistance_management
NUEVO_CONV_CA_8	No	9.8	0	
GROOM_CONV_CG_11	No	66.7	2	sexual_content_request, personal_information_request
NUEVO_prueba2	No	94.1	2	sexual_content_request, personal_information_request
TWEET_38	No	9.8	0	
GROOM_CONV_CG_10	Yes	99.7	4	compliments, resistance_management, sexual_content_request, personal_information_request
NUEVO_prueba3	No	87.1	2	compliments, sexual_content_request
TWEET_5	No	9.8	0	
NUEVO_prueba1	No	9.8	0	
TWEET_15	No	9.8	0	
GROOM_CONV_CG_8	Yes	98.5	3	compliments, sexual_content_request, personal_information_request
TWEET_29	No	9.8	0	
TWEET_28	No	9.8	0	
TWEET_31	No	9.8	0	
GROOM_CONV_CG_6	Yes	67.2	2	compliments, personal_information_request

Ultimately, the results in Table 5 reveal a directly proportional relationship in which the estimated probability of grooming increases as the density of identified modules augments, suggesting that the accumulation and aggregation of linguistic indicators positively influence the statistical reliability of the classification. This trend is evident in the detected modules column of Table 5, where the absence of these modules results in a negative prediction; however, this behavior is conditioned by the established decision threshold. This threshold regulates the sensitivity to ambiguous cases, acting as a methodological filter designed to reduce the false positive rate. Accordingly, instances that record high probabilities but are ultimately categorized as negative (No) reflect a precise adjustment in the decision criteria of the model.

CONCLUSIONS

This study performed the first discourse analysis of groomers in Spanish-speaking Roblox communities by identifying a series of specific linguistic patterns. After analyzing a corpus of 25 conversations, a pattern of discursive organization was delimited in the form of a sequence of seven conversational modules with specific relative frequencies and predictability percentages that reflect a consistent strategic progression: salutation, personal information request, robux or reward offer, sexual content request, resistance management, compliments, and closing.

The robux or rewards offer module was particularly significant, presented in 96% of the conversations analyzed. This demonstrates that the material incentives inherent to the videogame replace the affective goods offer, reported in previous research on digital grooming discourse. Similarly, the module of sexual content request was identified in 100% of the conversations and recognized as the defining criterion of cybergrooming discourse on Roblox.

With reference to the lexico-grammatical patterns, 21 specific patterns were identified, with 224 associated collocations, which collectively demonstrated their effectiveness for integration into automated detection systems. An analysis based on concepts from Systemic Functional Linguistics and the identified pragmatic-discursive strategies revealed that the groomer systematically employs constructions involving verbal, relational, mental, material, and behavioral processes, as well as declarative and interrogative modes of expression, most frequently to manipulate minors and obtain sexual content while projecting an image of legitimacy and closeness.

Furthermore, the computational model developed demonstrated to be an effective tool for the automatic detection of cyber-grooming, due to the integration of contributions from discourse analysis as the basis for development and the data processing techniques employed. This is reflected in the evaluation of performance metrics, which indicates a correct classification with 93.33% accuracy.

DECLARATION OF CONFLICTS OF INTEREST: The authors state that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS STATEMENT AND ACKNOWLEDGMENTS: The contribution of each author is listed using the CRediT Taxonomy below:

- Ana Paola Castañón Marroquín: Lead autor, Conceptualization, Formal Analysis, Research, Methodology, Validation, Visualization, Writing-revision, and editing.
- Brenda Ailed Rodríguez Colis: Conceptualization, Formal Analysis, Research, Methodology, Validation, Visualization, Writing-original draft, Writing-revision, and editing.
- Andrea Bazán Durán: Conceptualization, Formal Analysis, Research, Methodology, Project Management, Resources, Validation, Visualization, Writing-original draft, Writing-revision,
- Luis Enrique Colmenares-Guillen: Conceptualization, Research, Methodology, Project Management, Resources, Supervision, Validation, Writing-original draft, Writing-revision, and editing.

The authors would like to thank the Benemérita Universidad Autónoma de Puebla for its support, and in particular to students Karla Hernández and Pedro Vera of the Bachelor's Program in Forensic Science, and Jesús Semita and María Sarmiento of the Bachelor's Program in Criminology, who participated in the collection of the corpus over an eight-month period at the Digital Forensic Analysis Laboratory of the School of Computer Science at the Benemérita Universidad Autónoma de Puebla, although they are not responsible for the content of this article.

DECLARACIÓN DE DISPONIBILIDAD DE DATOS: The authors declare that the data used in this research are available and freely accessible for analysis by interested parties, in the repository: https://doi.org/10.5281/zenodo.19241233

BIBLIOGRAPHIC REFERENCES

Allgaier, J., & Pryss, R. (2024). Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Machine Learning and Knowledge Extraction*, *6*(2), 1378–1388. https://doi.org/10.3390/make6020065

Badman, A., & Kosinski, M. (2024). *¿Qué es un conjunto de datos?* IBM. https://www.ibm.com/mx-es/think/topics/dataset

Broome, L., Izura, C., & Davies, J. (2025). An investigation of the linguistic and deceptive characteristics of online grooming types. *Journal of Sexual Aggression*, *31*(3), 378–395. https://doi.org/10.1080/13552600.2023.2300470

Calsamiglia, H., & Tuson, A. (2012). *Las cosas del decir*. Ariel.

Carville, O., & D'Anastasio, C. (2024, July 22). *Roblox Predator Problem Potentially Exposes Kids to Pedophiles*. Bloomberg. https://www.bloomberg.com/features/2024-roblox-pedophile-problem/

Chiang, E., & Grant, T. (2019). Deceptive Identity Performance: Offender Moves and Multiple Identities in Online Child Abuse Conversations. *Applied Linguistics*, *40*(4), 675–698. https://doi.org/10.1093/applin/amy007

Daohuan, L., & Xuri, T. (2023). The Contextualized Representation of Collocation. En M. Sun, B. Qin, X. Qiu, J. Jiang, & X. Han (Eds.) *The 22nd Chinese National Conference on Computational Linguistics* (1st ed., pp. 836–846). ACL Anthology. https://aclanthology.org/2023.ccl-1.71/

Dorsey, E. (2024). *Problems at Roblox (RBLX) #4*. The Bear Cave. https://thebearcave.substack.com/p/problems-at-roblox-rblx-4

Evans, C., & Lorenzo-Dus, N. (2025). A corpus-assisted discourse analysis of children's and groomers' talk in online grooming interactions. *Applied Corpus Linguistics*, *5*(3), 100147. https://doi.org/10.1016/j.acorp.2025.100147

Excelin, G., Dinansyah, F., & Anugrah, C. (2024). Communication Patterns in the Use of Communication Features in Online Games Case Study: Valorant. *Jurnal Vokasi Indonesia*, *11*(2), 45–56. https://doi.org/10.7454/jvi.v11i2.1203

Gillings, M., Learmonth, M., & Mautner, G. (2024). Taking the Road Less Travelled: How Corpus-Assisted Discourse Studies Can Enrich Qualitative Explorations of Large Textual Datasets. *British Journal of Management*, *35*(4), 1667–1679. https://doi.org/10.1111/1467-8551.12816

González-Sanz, M. (2024). Análisis pragmalingüístico de las secuencias de apertura en conversación de WhatsApp. *Rilce Revista de Filología Hispánica*, *40*(3), 995–1023. https://doi.org/10.15581/008.40.3.995-1023

Halliday, M., & Matthiessen, C. (2004). *An Introduction to Functional Grammar*. Routledge.

Hunston, S. (2010). How can a corpus be used to explore patterns? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (1st ed., pp. 152–166). Routledge. https://doi.org/10.4324/978003856949

Hunston, S. (2025). *Pattern, Construction, System*. Cambridge University Press. https://doi.org/10.1017/9781009629065

IBM. (2025). *Métricas de evaluación de precisión*. IBM. https://www.ibm.com/docs/es/ws-and-kc?topic=metrics-accuracy

Isti'anah, A., Febrina, R., Suhandano, S., Winarti, D., Sutrisno, A., & Jumanto, J. (2023). Big Data, Computer, and Technology in Language Studies: The Potentials of Sketch Engine in Indonesia's Research. *2023 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 46–51. https://doi.org/10.1109/iSemantic59612.2023.10295357

Joleby, M., Lunde, C., Landström, S., & Jonsson, L. (2021). Offender strategies for engaging children in online sexual activity. *Child Abuse & Neglect*, *120*, 105214. https://doi.org/10.1016/j.chiabu.2021.105214

Kou, Y., Hernández, R., & Gui, X. (2025). "The System is Made to Inherently Push Child Gambling in my Opinion": Child Safety, Monetization, and Moderation on Roblox. In N. Yamashita, V. Evers, K. Yatani, X. Ding, B. Lee, M. Chetty, & P. Toups-Dugas (Eds.) *CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (1st ed., pp. 1-18). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713170

Lorenzo-Dus, N. (2023). *Digital Grooming Discourses of Manipulation in Cyber-crime*. Oxford University Press. https://doi.org/10.1093/oso/9780190845193.001.0001

Lorenzo-Dus, N., & Izura, C. (2017). "cause ur special": Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics*, *112*, 68–82. https://doi.org/10.1016/j.pragma.2017.01.004

Lorenzo-Dus, N., Kinzel, A., & Di Cristofaro, M. (2020). The communicative modus operandi of online child sexual groomers: Recurring patterns in their language use. *Journal of Pragmatics*, *155*, 15–27. https://doi.org/10.1016/j.pragma.2019.09.010

Lutzky, U., & Kehoe, A. (2022). Using corpus linguistics to study online data. In C. Vasquez (Ed.), *Research methods for digital discourse analysis* (2nd ed., pp. 219–236). Bloomsbury Academic.

Lyu, J., & Ishwaran, H. (2023). Commentary: To classify means to choose a threshold. *Journal of Thoracic and Cardiovascular Surgery*, *165*(4), 1443–1445. https://doi.org/10.1016/j.jtcvs.2021.08.009

Manning, C., Raghavan P., & Schütze H. (2009). *Introduction to Information Retrieval*. Cambridge University Press. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, *25*(2), 137–166. https://doi.org/10.1017/S0269888910000032

McDonough, M. (2023). *Cross-validation*. Encyclopedia Britannica. https://www.britannica.com/technology/cross-validation-computer-science

Molina, J., Fàbregues, S., & Escalante, E. (2024). *Métodos mixtos de investigación*. Ediciones Pirámide.

Morreale, D., & Rosa, A. (2024). Media Technologies and Epistemologies: The Platforming of Everything| Roblox and the Pervasiveness of Play: What Game-Making Communities Can Teach Us About Participatory Practices in Affinity Spaces. *International Journal of Communication*, *18*, 4281–4299. https://ijoc.org/index.php/ijoc/article/view/21902

Ortiz, N. (2024). *El grooming como delito trasnacional: Avances y desafíos de la legislación y la cooperación entre Colombia, Argentina y México* [Master's Thesis, Universidad Externado de Colombia]. https://bdigital.uexternado.edu.co/handle/001/26312

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins Publishing Company.

Pérez, J. M., Furman, D. A., Alonso Alemany, L., & Luque, F. M. (2022). RoBERTuito: A pre-trained language model for social media text in Spanish. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference (1st ed., pp. 7235–7243. https://aclanthology.org/2022.lrec-1.785

Pienczykowski, B., & Madella, P. (2026). 'I was so Afraid All You Wanted from Me Was Sex': A Corpus-Assisted Study in the Pragmatics of Manipulation in Online Child Sexual Groomers' Discourse. *Corpus Pragmatics*, *10*, article number 11. https://doi.org/10.1007/s41701-025-00218-0

Pyslar, M. (2025). The Evolution of Community Management in Gaming: From Forums to AI-driven Engagement. *Актуальные Исследования*, *9*(244), 69–73. https://doi.org/10.5281/zenodo.14968205

Red Grooming Latam. (2024). *Los riesgos de niñas, niños y adolescentes en internet* [[PDF file]. https://www.groominglatam.org/wp-content/uploads/2025/02/INFORME-2024-FLYER-INFORMATIVO.pdf

Rozgonjuk, D., Schivinski, B., Pontes, H. M., & Montag, C. (2023). Problematic Online Behaviors Among Gamers: the Links Between Problematic Gaming, Gambling, Shopping, Pornography Use, and Social Networking. *International Journal of Mental Health and Addiction*, *21*(1), 240–257. https://doi.org/10.1007/s11469-021-00590-3

Schegloff, E. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge University Press.

Schweinberger, M. (2024). *Analyzing Collocations and N-grams in R*. Language Technology and Data Analysis Laboratory (LADAL). https://slcladal.github.io/coll.html

Sorlin, S. (2017). The pragmatics of manipulation: Exploiting im/politeness theories. *Journal of Pragmatics*, *121*, 132–146. https://doi.org/10.1016/j.pragma.2017.10.002

Steward, K. (2023). *Sensitivity vs specificity: Test accuracy explained*. Technology Networks. https://www.technologynetworks.com/analysis/articles/sensitivity-vs-specificity-318222

Straka, M., & Straková, J. (2020). UDPipe at EvaLatin 2020: Contextualized Embeddings and Treebank Embeddings. En R. Sprugnoli & M. Passarotti (Eds.), *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages* (1st ed., pp. 124–129). European Language Resources Association (ELRA). https://aclanthology.org/2020.lt4hala-1.20/

Ting, K. M. (2010). Confusion Matrix. In G. I. Sammut Claude & Webb (Eds.), *Encyclopedia of Machine Learning* (1st ed., p. 209). Springer. https://doi.org/10.1007/978-0-387-30164-8_157

Vaamode, G., & González, F. (2008). Clasificación verbal, etiquetación semántica e información lexicográfica en el proyecto ADESSE. En M. Verdejo & A. Serrano (Eds.), *Acceso y visibilidad de la información multilingüe en la red: El rol de la semántica* (1st ed., pp. 225–236). UNED.

Van Dijk, T. (1997). *Discourse as Structure and Process*. Sage Publications.

Walsh, M. J. (2022). About ‘face’’: Reconsidering Goffman’s theory of face-work for digital culture.’ In M. Jacobsen & G. Smith (Eds.), *The Routledge International Handbook of Goffman Studies* (1st ed., pp. 207–218). Routledge. https://doi.org/10.4324/9781003160861-20

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia Creative Commons CC-BY.
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.