
UNA APROXIMACIÓN PARA LA DETECCIÓN DE CONTENIDO SEXUAL EN CONVERSACIONES DIGITALES

*An approach for the detection of sexual content in digital
conversations*

*Uma abordagem para detecção de conteúdo sexual em conversas
digitais*

Luis Enrique Colmenares-Guillén¹ , José Luis Jiménez-Aguilera¹ 

¹ Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla. Puebla -México. Correo: enrique.colmenares@correo.buap.mx, joseluisjimaqui@gmail.com

Fecha de recepción: 16 de marzo de 2023.

Fecha de aceptación: 17 de abril de 2023.

RESUMEN

INTRODUCCIÓN. El internet se ha hecho más accesible utilizando las TICs (Tecnologías de la información y Comunicación), ha crecido el uso de las redes sociales, y ha generado la exposición de menores a delitos como ciberacoso, sexting, pederastia, entre otros. **OBJETIVO.** Realizar un análisis de conversaciones con contenido pederasta y de esta forma establecer una estructura prototípica utilizada por los delincuentes. **MÉTODO.** Se utilizaron las conversaciones de la página Perverted-Justice, un corrector gramatical como LanguageTool y un paquete de herramientas lingüísticas y de extracción textual como Linguakit. Estas conversaciones podrán ser etiquetadas, categorizadas, y procesadas en diferentes etapas para obtener un corpus. **RESULTADOS.** Se encontraron verbos, nombres, adjetivos y adverbios más utilizados por los delincuentes. Además de un esquema de intervenciones, que proporciona un perfil preliminar del victimario. **DISCUSIÓN Y CONCLUSIONES.** El análisis de conversaciones en el idioma Inglés por pedófilos y pederastas en internet muestra el léxico que refleja los nombres y sustantivos que se utilizan para la obtención de información como fotografías y tipo de ropa. Además, los adjetivos se emplean para halagar a la víctima y así, ganar su confianza, obteniendo más fácil la información. La importancia de realizar un corpus en español ayudaría a impedir este tipo de delitos como la pornografía infantil y pederastia en México, propiciando la detención de estos delincuentes.

Palabras claves: Pederastia, pedofilia, abuso sexual, sexting.



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.
Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



 Compartir

ABSTRACT

INTRODUCTION. The internet has become more accessible using ICTs (information and communication technologies), the use of social networks has grown and has generated the exposure of minors to crimes such as cyberbullying, sexting, pederasty, among others. **OBJECTIVE.** Perform an analysis of conversations with pederast content and establish a prototypical structure used by criminals **METHOD.** Using the conversations obtained from Perverted-Justice, a grammar checker known as LanguageTool and a package of linguistic and text extraction tools such as Linguakit. These conversations could be tagged, categorized, and processed in different stages to obtain a corpus. **RESULTS.** Verbs, nouns, adjectives and adverbs most used by criminals were found. In addition, a scheme of interventions that provides a preliminary profile of the victimizer. **DISCUSSION AND CONCLUSIONS.** The analysis of conversations in the English language by pedophiles and pederasts on the Internet shows the lexicon that reflects the names and nouns that are used to obtain information such as photographs and type of clothing. In addition, adjectives are used to flatter the victim and thus gain their trust, obtaining the information easier. The importance of creating a corpus in Spanish would help prevent this type of crime, such as child pornography and pederasty in Mexico, leading to the arrest of these criminals.

Keywords: Pederasty, pedophilia, sexual abuse, sexting.

RESUMO

INTRODUÇÃO. A internet tornou-se mais acessível com recurso às TIC (Tecnologias de Informação e Comunicação), cresceu a utilização das redes sociais e gerou a exposição de menores a crimes como cyberbullying, sexting, pederastia, entre outros. **OBJECTIVO** Faça uma análise de conversas com conteúdo pederasta e estabeleça uma estrutura protótipica utilizada por criminosos. **MÉTODO.** Foram usadas conversas da página Perverted-Justice, um verificador gramatical como o LanguageTool e um pacote de ferramentas linguísticas e de extração de texto como o Linguakit. Essas conversas podem ser marcadas, categorizadas e processadas em diferentes etapas para obter um corpus. **RESULTADOS.** Foram encontrados os verbos, substantivos, adjetivos e advérbios mais utilizados pelos criminosos. Além disso, um esquema de intervenções, que fornece um perfil preliminar do vitimizador. **DISCUSSÃO E CONCLUSÕES.** A análise de conversas em língua inglesa por pedófilos e pederastas na Internet mostra o léxico que reflete os nomes e substantivos que são usados para obter informações como fotografias e tipo de roupa. Além disso, adjetivos são usados para bajular a vítima e assim ganhar sua confiança, obtendo as informações mais facilmente. A importância de criar um corpus em espanhol ajudaria a prevenir esse tipo de crime, como pornografia infantil e pederastia no México, levando à prisão desses criminosos.

Palavras-chave: pederastia, pedofilia, abuso sexual, sexting.

INTRODUCCIÓN

A lo largo de los años el acceso a Internet a través de las tecnologías de la información y la comunicación (TIC's), brinda un sinnúmero de datos que están a la disposición en cuestión de segundos, aunado a esto, las redes sociales se han convertido en uno de los medios de comunicación principales a nivel mundial; además, permiten compartir datos personales, actividades que se realizan día a día, gustos, intereses, entre otros [1]. De esta forma, la interacción entre usuarios en algunas redes sociales se da mediante conversaciones utilizando un chat, que se caracterizan por ser



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



de forma espontánea e instantánea y que puede darse entre usuarios de diferentes partes del mundo por medio de plataformas digitales [2].

Por otra parte, en la actualidad cualquier persona puede tener una red social con solo tener un dispositivo con acceso a internet [3], y aunque algunas de estas redes sociales permiten su uso solo a personas mayores o un rango de años establecido, existe la falta de interés por parte de usuarios y administradores de estas mismas páginas referente a la regulación de los usuarios y de la información que proporcionan, derivando en que los menores de edad tengan acceso a estas plataformas, exponiéndose a delitos como la pornografía infantil, el sexting [4], el ciberacoso [5], la trata de personas, la pederastía, y demás delitos que se han adaptado a los medios digitales [6].

El objetivo de este trabajo es realizar un análisis de conversaciones con contenido pedófilo/pederasta y de esta forma establecer una estructura prototípica de sus participaciones, por lo que es importante preguntarse ¿cuál es el léxico utilizado por un pederasta en conversaciones de chats con contenido sexual? y ¿cuál es la estructura prototípica que se presenta en estas conversaciones?

Así los objetivos específicos que se abordarán en este trabajo serán: 1. Analizar conversaciones con contenido sexual utilizando software para el preprocesamiento del lenguaje natural; 2. Determinar el léxico más utilizado en conversaciones con contenido sexual; y 3. Determinar cuál es la estructura prototípica que se presenta en estas conversaciones.

En el Estado de Puebla, el artículo 217 del Código Penal del Estado Libre y Soberano del Estado de Puebla [7] dicta:

Comete el delito de pornografía de menores e incapaces, quien con relación a una persona menor de dieciocho años de edad o que carezca de la capacidad de comprender el significado de los hechos o de quien por la razón que fuere no pudiere oponer resistencia, realice alguna de las siguientes conductas:

II. Fije, imprima, video grabe, audio grabe, fotografíe, filme o describa actos de exhibicionismo corporal o lascivos o sexuales, reales o simulados.

Se sancionará con prisión de ocho a catorce años y multa de cien a mil doscientos días de salario, según lo establece el artículo 221.

Dado que, en la actualidad, este tipo de delitos se ha hecho presente en medios de comunicación masiva como lo es el internet, la presente investigación se enfoca en la realización de un análisis de conversaciones con contenido sexual originarias del idioma inglés y disponibles en el sitio Perverted-Justice [8].

Debido al uso restringido de datos oficiales para la extracción de discursos escritos digitales con contenido sexual en español, esta investigación da apertura a la elaboración de corpus y análisis en lengua española. Así mismo, se busca la realización de investigaciones lingüísticas que se interesen en objetos de estudio provenientes de



medios digitales y relacionados a delitos cibernéticos [9], así como fomentar investigaciones de carácter interdisciplinario con ciencias: criminología, lingüística, derecho, psicología y computación, entre otras.

MÉTODO

Fuentes

Los textos que se analizaron para esta investigación son mensajes generados en los distintos sitios web de interacción social, así como redes sociales. La extracción de estas conversaciones fue obtenida de la página PervertedJustice la cual, se encarga de la recolección de chats con contenido pedófilo/pederasta.

La selección de las conversaciones son un total de veinticinco que pertenecen a hablantes que viven en diferentes estados de los EEUU, de sexo masculino y de edades diferentes, como se muestra en la Figura 1. Las conversaciones fueron recolectadas entre los años 2006 a 2015.



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.
Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



NÚMERO DE CONVERSACIONES	ESTADO	NOMBRE
1	California	Kent Douglas Grant
2	California	Aminder Singh
3	Michigan	Eric M. Skulsky
4	Cincinnati, Ohio	Jeff
5	New Jersey	Jeremy Keister
6	Grand Rapids, Michigan	Jarrad Lawrence Box
7	Bogalusa, Louisiana	Wendell W Green
8	Louisville, Kentucky	Damon
9	Florence, Mississippi	Johnny Chain
10	Fargo, North Dakota	Dustin
11	Michigan	Dennis Frick Jr.
12	California	Matt. 48
13	California	Gregory John Karnos, 49
14	Portsmouth, Virginia	Kenneth Hanna, 33
15	Hannibal, Missouri	Jeff W Crenshaw, 30
16	Suffolk County, New York	Michael J. Coffey, 54
17	Lawton/Ft. Sill, Oklahoma	John, 27
18	Brooksville, Florida	Thomas, 31
19	Dallas, Texas	José, 21
20	Houston, Texas	Randy, 50
21	Texarkana, Texas	Santigo Gaston, 34
22	Bogalusa, Louisiana	Wendell W Green 25
23	Metairie, Louisiana	Michael L. Chatellier
24	Metairie, Louisiana	Michael L. Chatellier
25	Tucson, Arizona	William

Figura 1. Conversaciones de pederastas.

Instrumentos

Se utilizaron dos herramientas de software, la primera para la corrección ortográfica de las intervenciones del hablante (pedófilo/pederasta) y otra para el conteo de frecuencia léxica.

El primer software que se utilizó fue LanguageTool [10], un corrector gramatical y de estilo para diferentes idiomas como inglés, español, alemán, entre otros. También proporciona una explicación del por qué, se consideran incorrectos los errores gramaticales que detecta.

El segundo software que se utilizó fue Linguikit [11], un paquete de herramientas lingüísticas y de extracción textual para explorar, analizar y obtener una mejor información de los textos y documentos escritos. Lo que ofrece esta paquetería es: análisis completo, resumidor, corrector, traductor, conjugador verbal, frecuencia de palabras, palabra clave en contexto, etiquetador morfosintáctico, analizador sintáctico, analizador de sentimiento, identificador de idioma, extractor de palabra clave, extractor multipalabra, reconocedor de identidades, supercorrector, extractor de tripletas y extractor de conceptos. De estas herramientas de análisis, la que se utilizó principalmente para este estudio fue la de frecuencia de palabras.

Técnicas de análisis de datos

Esta investigación se realizó, utilizando un software, un proceso de etiquetación de las categorías gramaticales encontradas en las conversaciones: verbos, nombres, adjetivos, artículos, adverbios, preposiciones, conjunciones y pronombres.

De la misma manera, para el análisis de frecuencias de palabras de las intervenciones de los hablantes se consideraron las frecuencias nominales, verbales, adjetivales y adverbiales.

Y para el análisis de los turnos en la conversación se optó por la construcción del esquema de los turnos del primer hablante pederasta/pedófilo dada su condición de victimario. La construcción de este esquema fue a partir del propuesto por Stephen C. Levinson [12], desde la teoría del análisis de la conversación, específicamente, la construcción de los pares adyacentes [13], los cuales mantienen de forma activa la conversación gracias a su fluidez [14].

Procedimiento

En la investigación se considera en una primera etapa, en el nivel léxico de las participaciones de los hablantes pedófilos/pederastas [15], y se afirman: “la referencia fundamental del texto se construye a través de elementos léxicos”. Estos introducen los temas. Para mantener la referencia iniciada hay diversos mecanismos a su servicio, todos ellos fundamentan la aparición de las palabras más frecuentes dentro de estas conversaciones para identificar los principales temas.

Por consiguiente, los resultados acerca de la frecuencia son clave para el análisis, como lo establece la información de frecuencia: La base en la tecnología con



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



respecto al lenguaje que se usa hoy en día, son muchos sistemas informáticos, que pueden entender el lenguaje y que dependen de la información de la frecuencia [16].

En la segunda etapa de la investigación se realiza un análisis de las intervenciones dentro de los chats, considerados como pederastas/pedófilos, lo cual se considera el orden canónico de los pares adyacentes. De acuerdo con la conversación se caracteriza por la alternancia de turnos [12]: un participante, A, habla, se detiene; otro, B, inicia, habla, detiene; de este modo se obtiene una repartición de los turnos de la charla A-B-A-B entre los dos participantes.

Los pares de adyacencia son secuencias de dos enunciados:

- Adyacentes.
- Producidos por hablantes diferentes.
- Ordenados por una primera parte y una segunda parte.

Clasificados de manera, que de una determinada primera parte requiere una determinada segunda (o un conjunto de segundas partes).

La figura 2, presenta una caracterización estructural de las primeras partes y de las segundas partes de la conversación, considerando que las primeras partes dan inicio a los turnos de habla y las segundas partes la continuación de la secuencia que pueden ser preferida (aceptación/respuesta esperada) o despreferida (rechazo/respuesta no esperada).

Pares adyacentes		
1 ^a parte	2 ^a parte	
petición	preferida aceptación	no preferida rechazo
ofrecimiento/invitación	aceptación	rechazo
valoración	acuerdo	desacuerdo
pregunta	respuesta esperada	respuesta inesperada o no respuesta
acusación	negación	admisión

Figura 2. Pares adyacentes de la conversación [12].

Aunado a lo anterior, se realizó una aproximación de la elaboración del corpus y del análisis de conversaciones con contenido pedófilo. Un corpus se puede definir como una colección de documentos (textos o grabaciones) que ha sido recopilada con fines específicos [17], y que se considera representativa de la lengua que se analiza (es decir que ofrece ejemplos significativos de la población de interés) y además es balanceada, lo cual quiere decir que los criterios de selección del corpus se mantengan a lo largo del



mismo, y que sea posible acceder a una porción conservando sus características [18]. El tamaño de la muestra para su elaboración no tiene límites, basta con que garantice representatividad del objeto de estudio.

Para esta elaboración se hizo una serie de pasos que se mencionan a continuación:

En primer lugar, después de la extracción de las intervenciones del hablante pedófilo/pederasta y debido a que son conversaciones de chat, se eliminaron datos como la hora, el username y la fecha. En la figura 3, se muestra una parte de la conversación 11, para el análisis.

working_loving_goodman (10/14/08 9:44:39 PM): hi asl	hi asl
working_loving_goodman (10/14/08 9:45:23 PM): im 33 mi single and looking are you	im 33 mi single and looking are you
working_loving_goodman (10/14/08 9:45:43 PM): you have pic	you have pic

Figura 3. Ejemplo de una parte de la Conversación 11.

En segundo lugar, fue la corrección ortográfica de las intervenciones de los pedófilos/pederastas, las variantes léxicas, acrónimos, las contracciones de algunas palabras como verbos y pronombres, igualmente con la corrección de sintagmas modales en las preguntas, ver figura 4, las correcciones fueron realizadas a través del software LanguageTool.

El interés de las correcciones fue para proceder a los siguientes pasos del análisis como la etiquetación y contabilidad de palabras. Para la ubicación y corrección de todos los elementos, se estableció una nomenclatura para facilitar la ubicación dentro de los turnos de habla: rojo (ortografía), amarillo (abreviaturas, contracciones y acrónimos), azul (variantes léxicas), rosa (rastreo de indicios de texto sexual).

u mean this Friday	you mean this Friday
Thks	thanks
u have sex in ur life	you have sex in your life
u can do sax	you can do sex
lol	laughing out loud

Figura 4. Ejemplo de una parte de la conversación 4.

RESULTADOS

A continuación, se exponen los resultados del análisis léxico de las veinticinco conversaciones que se consideraron como corpus. Se muestran los diez términos más frecuentes que se presentan en cada conversación.

Tabla General Léxica/ Verbal



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.
Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



En las figuras 5, 6, 7, 8 y 9 se exponen de manera general la lista de verbos o lexical, de acuerdo a la categorización de Linguikit, se consideró los primeros diez términos por ser de mayor frecuencia en las participaciones de los pedófilos/pederastas. Algunos de los verbos más utilizados son: be, do y have. Asimismo, dentro de estos primeros se abordaron primordialmente los verbos con connotaciones sexuales como suck, lick, fuck, entre otros. También otros verbos representativos para peticiones y proposiciones sexuales dentro de las conversaciones: wear, see y meet.

1		2		3		4		5	
Apariciones	Verbo								
78	be	31	be	27	be	28	be	72	be
28	do	19	do	16	do	27	do	42	do
22	have	12	mean	11	have	7	get	40	have
12	get	11	have	10	like	6	come	12	want
10	come	10	call	8	go	5	eat	10	know
8	like	8	want	6	want	4	suck	8	say
7	nees	7	know	5	meet	4	have	7	come
6	pick	7	go	5	see	4	know	6	call
6	know	7	set	4	wear	3	see	6	guess
6	go	5	see	4	ask	3	stay	6	tel

Figura 5. Resultado de las Conversaciones 1-5 (Verbos).

6		7		8		9		10	
Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo
58	be	15	have	45	do	38	be	50	be
39	do	15	be	26	be	16	get	38	do
24	want	11	do	11	have	14	do	14	come
22	get	10	get	7	want	13	see	12	see
20	have	7	let	6	like	12	come	12	go
20	go	5	go	5	see	10	wan	11	have
10	come	5	know	5	get	9	know	10	get
8	think	4	wear	5	know	9	want	9	live
7	see	4	promise	4	come	8	have	8	lick
7	lick	3	meet	4	need	6	hang	7	want

Figura 6. Resultado de las Conversaciones 6-10 (Verbos).

11		12		13		14		15	
Apariciones	Verbo								
21	be	63	be	31	be	23	be	28	be
12	do	24	do	15	have	14	do	22	do
11	have	18	have	13	mean	11	have	7	get
9	like	7	come	12	call	10	like	6	come
8	go	6	go	12	do	8	go	5	eat
6	sit	6	get	10	want	7	sit	4	suck
6	want	6	think	9	get	6	want	4	have
5	know	5	like	8	know	5	meet	4	know
5	meet	5	see	8	come	5	see	3	see
		5	want	7	tell	4	wear	3	stay

Figura 7. Resultado de las Conversaciones 11-15 (Verbos).

16		17		18		19		20	
Apariciones	Verbo								
36	yo	84	you	12	do	28	be	82	be
17	I	36	I	10	want	22	do	65	do
15	do	29	your	9	get	10	want	42	want
12	ok	21	be	6	sec	9	got	37	make
10	be	21	to	5	meet	7	get	30	see
10	what	19	and	5	see	8	know	30	get
9	to	15	what	4	know	7	and	29	like
8	me	12	it	3	it	7	your	27	rub
8	like	12	do	3	stay	6	think	22	wish
7	and	11	have	3	go	5	like	20	feel

Figura 8. Resultado de las conversaciones 16-20 (Verbos).

21		22		23		24		25	
Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo	Apariciones	Verbo
47	be	42	be	24	have	62	be	98	be
73	do	30	have	4	have	15	get	46	do
52	have	26	get	4	do	14	do	36	have
43	want	22	do	3	love	10	love	22	want
39	know	14	get	3	rub	10	see	20	meet
37	get	12	wear	3	laugh	8	rub	19	go
34	think	12	go	3	bet	8	look	18	cool
20	go	10	try	3	think	7	try	17	get
15	look	10	know	3	play	6	tell	16	see
14	do not	10	promise	3	lift	6	want	15	see

Figura 9. Resultado de las conversaciones 21-25 (Verbos).



Tabla General de Nombres o Sustantivos

En las Figuras 10, 11 y 12 se extrajeron los primeros 10 nombres o sustantivos de cada conversación con alta frecuencia y con connotaciones sexuales que son: sex, cock, picture. Algunos otros como company, photo y address, también se abordaron por las intenciones de índole sexual que estas conllevan dentro de las conversaciones por parte del pedófilo/pederasta. Se muestran resultados de los nombres o sustantivos de las conversaciones del 1 al 10 y del 21 al 25.

1		2		3		4		5	
Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre
11	girl	10	home	8	cock	3	today	10	sex
5	night	9	cam	7	pic	3	face	6	tomorrow
5	one	7	time	6	addres	3	room	6	girl
4	sister	5	hello	5	panty	2	hello	4	guy
3	way	4	bye	4	mom	2	hour	4	wok
3	person	3	age	4	one	2	cum	3	boyfriend
3	day	3	lol	4	time	2	yeah	3	problem
3	weekend	3	sex	3	pussy	2	leg	3	dad
3	company	3	friend	3	weekend	1	weather	3	month
3	photo	3	plane	3	color	1	cool	3	home

Figura 10. Resultados conversaciones 1-5 (Sustantivos).

6		7		8		9		10	
Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre
20	lol	5	sex	5	nothing	12	yeah	8	yeah
10	ass	5	picture	5	door	5	lol	6	sweetie
8	yeah	4	mom	5	picture	4	one	6	pussy
7	minute	4	weekend	3	way	4	weekend	5	house
5	dude	4	camera	3	stuff	4	got to	4	time
5	guy	3	hey	3	guy	4	girl	4	door
5	mine	3	panty	3	yeah	3	hey	4	store
5	cam	2	thong	3	girl	3	mom	3	place
4	pic	2	phone	2	address	3	going to	3	one
4	condom	2	promise	2	nice	3	band	3	sex

Figura 11. Resultados conversaciones 6-10 (Sustantivos).

21		22		23		24		25	
Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre	Apariciones	Nombre
37	friend	10	mom	5	pie	10	breast	18	girl
32	i	10	sex	3	shirt	7	guy	11	baby
32	boy	10	pic	2	age	6	cock	11	pic
21	time	8	weekend	2	self	4	tee	10	cum
17	stuff	8	cam	2	profile	4	boyfriend	8	cam
16	house	8	panty	2	guy	3	size	8	one
14	you	6	phone	2	nipple	3	mouth	5	sex
14	laptop	6	hey	2	year	3	mall	5	fun
12	girl	5	thong	2	kiss	3	pussy	5	wife
11	room	4	finger	2	breast	2	girlfriend	5	park

Figura 12. Resultados conversaciones 21-25 (Sustantivos).

Tabla General de Adjetivos

En las Figuras 13, 14, 15, 16 y 17, corresponden a los adjetivos, se extrajeron los primeros 10 términos con mayor iteración de cada conversación con connotaciones sexuales como: sexy, naked y horny. Se evoca que los adjetivos califican a los nombres agregando una cualidad, algunos adjetivos considerados para la atracción, seducción u otros aspectos dentro de las conversaciones son: pretty, little y sweet.

1		2		3		4		5	
Apariciones	Adjetivos								
6	sexy	27	ok	5	horny	12	ok	68	ok
6	good	8	okay	4	ok	9	naked	7	only
4	pretty	7	good	4	single	3	pussy	6	good
4	ok	4	numb	3	I	1	bald	4	big
4	late	3	sorry	2	bald	1	cold	4	nice
3	other	3	sure	2	later	1	ready	3	old
2	sorry	3	I	2	other	1	tight	2	oh
2	sure	2	only	2	wido	1	other	2	sure
2	sweet	2	nice	2	old	1	last	2	little
2	short	2	shy	2	real	1	dick	2	legal

Figura 13. Resultados de las conversaciones 1-5 (Adjetivos).

6		7		8		9		10	
Apariciones	Adjetivos								
9	ok	25	ok	8	ok	7	ok	17	ok
6	little	7	pussy	3	cool	3	ready	5	cool
4	asshole	3	dick	3	other	3	next	3	hot
4	sorry	2	top	3	little	3	nice	2	naked
4	okay	2	naked	3	single	3	new	2	oh
3	only	1	nude	2	ready	2	other	2	pussy
3	cute	1	serious	2	only	2	oh	2	sexy
3	good	1	wrong	2	good	2	wan	2	sure
3	gay	1	other	2	wrong	2	good	2	comfortable
2	dick	1	young	2	I	2	old	2	later

Figura 14. Resultados de las conversaciones 6-10 (Adjetivos).



11		12		13		14		15	
Apariciones	Adjetivos								
6	horny	6	sexy	50	ok	5	horny	12	ok
6	good	4	pretty	17	ur	4	single	9	naked
4	pretty	4	good	7	good	4	ok	3	pussy
4	ok	4	ok	4	sorry	3	I	1	bald
4	late	4	late	4	sure	2	other	1	cold
3	other	3	old	3	only	2	real	1	ready
2	sorry	2	sorry	3	numb	2	bald	1	tight
2	sure	2	other	3	I	2	later	1	other
2	sweet	2	sweet	2	dick	2	wide	1	last
2	short	2	little	2	full	1	ready	1	red

Figura 15. Resultados de las conversaciones 11-15 (Adjetivos).

16		17		18		19		20	
Apariciones	Adjetivos								
12	ok	5	horny	34	ok	13	sexy	20	good
3	cool	4	ok	12	naked	10	pretty	12	ok
3	sexy	4	single	11	pussy	9	good	10	hard
3	sure	2	bald	10	bald	2	later	9	sweet
2	sorry	2	later	9	cold	2	other	6	mad
2	good	2	other	1	other	2	wide	5	sorry
2	old	2	wide	1	busy	2	full	5	next
2	green	2	hot	1	avy	2	long	5	same
2	other	2	real	1	big	1	ready	5	little
2	busy	1	ready	1	same	1	tight	5	pussy

Figura 16. Resultados de las conversaciones 16-20 (Adjetivos).

21		22		23		24		25	
Apariciones	Adjetivos								
29	good	62	ok	6	cool	17	ok	49	ok
26	cool	14	pussy	6	ok	9	old	26	cool
12	ols	6	dick	5	old	8	sexy	11	good
7	sure	4	top	4	horny	5	real	11	awesome
6	pussy	4	cool	3	cute	5	cool	9	hot
5	happy	4	naked	3	young	4	naked	9	sexy
5	new	2	nite	2	short	3	cute	5	quick
5	bad	2	nude	1	true	3	other	5	yummy
5	weird	2	serious	1	uneasty	3	small	5	true
5	big	2	other	1	borwn	3	honest	5	horny

Figura 17. Resultados de las conversaciones 21-25 (Adjetivos).

Tabla General de Adverbios

Por último, en las Figuras 18, 19, 20, 21 y 22 que corresponden a los adverbios y que son los que se presentan con mayor frecuencia en las conversaciones y por esta razón este tipo de palabras tendrán una connotación significativa dentro del contexto de los enunciados. Algunos que se encontraron son: always, tonight y just, entre otros, que además que poseen contenido sexual. Estos representan algunos puntos como el encuentro entre el pederasta y el menor de edad.



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



1		2		3		4		5	
Apariciones	Adverbios								
19	not	15	so	7	now	9	not	44	not
9	so	12	not	7	yes	6	just	17	so
7	here	10	yes	4	just	5	so	10	yes
6	really	6	now	4	not	4	then	8	just
6	yes	6	really	3	hard	3	right	6	ever
6	there	5	alone	3	there	3	oh	5	too
5	now	4	right	3	so	3	tonight	5	really
5	just	3	else	3	about	2	always	4	here
4	then	2	then	2	north	2	now	3	now
4	always	2	up	2	pretty	1	yeah	3	yet

Figura 18. Resultados conversaciones 1-5 (Adverbios).

6		7		8		9		10	
Apariciones	Adverbios								
25	not	11	not	14	not	13	well	24	not
7	then	6	here	12	so	12	just	11	just
7	just	6	now	10	just	9	not	11	there
6	so	5	yes	10	well	8	there	10	so
6	well	5	right	7	yes	5	here	9	then
4	here	4	down	3	too	5	really	9	back
4	ever	4	there	3	really	4	right	5	right
4	too	3	then	3	there	3	back	4	now
4	yes	3	just	3	right	2	then	4	yes
3	hard	3	so	2	actually	2	pretty	4	tonight

Figura 19. Resultados conversaciones 6-10 (Adverbios).

11		12		13		14		15	
Apariciones	Adverbios								
8	yes	14	so	7	yes	6	just	5	yes
5	now	13	yes	6	now	5	so	4	no
4	ust	11	not	4	just	4	then	3	not
4	here	9	now	3	hard	3	not	2	o
3	hard	7	right	3	there	3	right	2	then
3	so	6	really	3	so	3	oh	2	far
2	many	6	alone	2	north	3	tonight	2	back
2	north	3	else	2	not	2	always	1	sometime
2	not	2	o	2	right	2	now	1	just
		2	then	2	about	2	don't	1	here

Figura 20. Resultados conversaciones 11-15 (Adverbios).



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



16		17		18		19		20		
Apariciones	Adverbios									
8	yes	3	hard	hard	14	so	7	yes	22	just
7	now	3	so	so	13	yes	6	now	22	yes
4	just	2	many	many	11	not	4	just	20	not
4	there	2	north	north	9	now	3	hard	16	ever
3	hard	2	not	not	7	right	3	there	14	maybe
3	so	2	back	back	6	really	3	so	13	now
2	not	1	sometime	sometime	6	alone	2	north	11	so
2	north	1	just	just	3	else	2	not	10	here
2	not	1	here	here	2	o	2	right	9	don't
2	right	1	always	always	2	then	2	about	6	hard

Figura 21. Resultados conversaciones 16-20 (Adverbios).

21		22		23		24		25	
Apariciones	Adverbios								
37	well	22	not	9	too	22	yes	48	so
36	just	14	here	7	here	14	not	36	yes
27	not	14	now	6	just	13	here	23	now
25	do not	10	down	6	so	13	so	17	not
20	so	10	right	5	yes	10	just	7	then
13	yup	8	then	4	many	10	really	7	too
11	here	8	yes	3	now	8	now	7	also
10	yes	8	there	2	foward	7	too	6	maybe
10	now	7	just	2	then	6	hard	5	right
8	would	6	yeah	2	not	5	ever	5	still

Figura 22. Resultados conversaciones 21-25 (Adverbios).

Análisis del léxico dentro de la conversación

A partir de los datos arrojados por el programa y de las tablas de frecuencia léxica, se exponen algunas muestras de las conversaciones analizadas, se observó la frecuencia léxica en relación con su aparición en las intervenciones con connotaciones sexuales de los pedófilos/pederastas.

A continuación, se presentan dos ejemplos de las tablas léxicas de las Conversaciones 1 y 4, las cuales se muestran en las figuras 23 y 24.



PALABRAS							
VERBOS		NOMBRES		ADJETIVOS		ADVERBIOS	
Apariciones	Palabra	Apariciones	Palabra	Apariciones	Palabra	Apariciones	Palabra
78	be	11	girl	6	sexy	19	not
28	do	5	night	6	good	9	so
22	have	5	one	4	pretty	7	here
12	get	4	sister	4	ok	6	really
10	come	3	way	4	late	6	yes
8	lie	3	person	3	other	6	there
7	need	3	day	2	sorry	5	now
6	pick	3	weekend	2	sure	5	just
6	know	3	company	2	sweet	4	then
6	go	3	photo	2	short	4	always
6	think						

Figura 23. Tabla de frecuencia de palabras de la Conversación 1

PALABRAS							
VERBOS		NOMBRES		ADJETIVOS		ADVERBIOS	
Apariciones	Palabra	Apariciones	Palabra	Apariciones	Palabra	Apariciones	Palabra
28	be	3	today	12	ok	9	not
27	do	3	face	9	naked	6	just
7	get	3	room	9	pussy	5	so
6	come	2	hello	1	bald	4	then
5	eat	2	hour	1	cold	3	right
4	suck	2	cum	1	ready	3	oh
4	have	2	yeah	1	tight	3	tonight
4	know	2	leg	1	other	2	always
3	see	1	weather	1	last	2	now
3	stay	1	cool	1	dick	1	yeah
3	go						
3	like						

Figura 24. Tabla de frecuencia de palabras de la Conversación 4

Las figuras 23 y 24 de las Conversaciones 1 y 4 arrojan las palabras de mayor frecuencia (verbos, nombres, adjetivos y adverbios) dentro de la conversación, y que representan diferentes connotaciones sexuales y no sexuales que hay en ellas.

Se puede observar que la frecuencia léxica con estas connotaciones sexuales se presentan especialmente en la construcción de preguntas de índole sexual en los diversos temas que el victimario desea obtener información de la víctima.

Para las intervenciones con connotaciones sexuales se muestran las figuras 25 y 26 de las conversaciones 1 y 4.

39. That is cool so what will it be my little girl. Do you want to come <u>hang out</u> for the night? <u>Would you be okay</u> spending a fun sexy night with me?	PREGUNTA ÍNTIMA: PROPOSICIÓN SEXUAL
40. A sexy little skirt would be nice but plain old jean would look hot on you. Where would I pick you up at? It will take me about 20 to 30 minutes to get there. Here is my cell number	PREGUNTA ÍNTIMA: VESTIMENTA PREGUNTA ÍNTIMA: POSIBLE ENCUENTRO
41. May I ask do you like to kiss?	
42. Ok so girl when and where do I pick you up at????	PREGUNTAS HIPÓTETICA: SEXUAL
43. <u>Yes</u> I think so. Do you want to come <u>find out</u> ?	PREGUNTA ÍNTIMA: PROPOSICIÓN
44. Well girl can you come here <u>tonight</u> ?	ENCUENTRO
45. Ok all I am waiting for is for you to tell me how to	

Figura 25. Muestra de la conversación 1 de intervenciones con connotación sexual

10. naked	PREGUNTA ÍNTIMA: PROPOSICIÓN SEXUAL
11. suck or ride me	PREGUNTA ÍNTIMA: PROPOSICIÓN SEXUAL
12. would you ride or suck <u>me</u>	PREGUNTA ÍNTIMA: PROPOSICIÓN SEXUAL
13. you a virgin	PREGUNTA ÍNTIMA: VIRGINIDAD
14. you going to sleep naked	PREGUNTA ÍNTIMA: INFORMACIÓN SEXUAL
15. yes	
16. no	
17. <u>do you sleep naked</u>	

Figura 26. Muestra de la conversación 4 de intervenciones con connotación sexual.

En estas muestras de las conversaciones 1 y 4 con énfasis en las participaciones de los pederastas/pedófilos con connotaciones sexuales se observa que los turnos se presentan en forma de interrogantes, esto es, con el fin de obtener información personal íntima o condiciones íntimas de las víctimas, así como proposiciones de índole sexual y de un posible encuentro.

Construcción del esquema de intervenciones del victimario (pedófilo/pederasta)

Dentro de la conversación y dada la necesidad de la participación de dos hablantes se deja entre ver que ambos elementos son parte fundamental para el proceso de la comunicación. En un primer momento, se optó por analizar la intervención del primer hablante, que es considerado como pedófilo/pederasta en el momento y que, por su conocimiento, el segundo hablante es menor de edad y da seguimiento a proposiciones de índole sexual.

Dentro de las conversaciones se ha identificado, como se observa en la figura 27 que se alude a un esquema de intervenciones propuesto por Stephen Levinson, a la 1ra. parte de los pares adyacentes a lo largo de toda la conversación. La 2da, parte de los pares adyacentes ha recaído en su mayoría en el segundo hablante, mostrando que el inicio de los turnos se concentra en el primer hablante, el victimario. ostrando que el inicio de los turnos se concentra en el primer hablante, el victimario. Al utilizar como referencia los pares adyacentes, nos ayudan a ordenar, interpretar y explicar lo que esta pasando en la conversación [19]. Es por esto que se optó por desglosar el tipo de intervenciones del primer hablante (pedófilo/pederasta) dada la cantidad y el tipo de construcciones que se realizan. Lo cual fue observado en el análisis de las



intervenciones de pedófilos/pederastas en cinco conversaciones de las 25 conversaciones seleccionadas.

1ra. parte de pares adyacentes	Módulos encontrados en las conversaciones del corpus		Intención	Temas/Tipos
Petición	Proposiciones sexuales	Acercamiento sexual a la víctima		-Fotografías -Encuentro
Ofrecimiento/ invitación	Secuencia de la conversación	Mediación con el hablante		-Fotografías -Encuentro
Valoración	Expresión de juicios	Cortejo		-Halagos
	Preguntas de experiencia sexual	Petición de información		-Play Self -Virginidad -Sexuales
Pregunta	Preguntas personales	Petición de información		-Fotografías -Pareja -Vestimenta -Medidas -Condición íntima
	Preguntas hipotéticas	Proposiciones, descripciones		-Sexo -Phone Sex -Encuentro -Locación
Acusación			x	x

Figura 27. Construcción del esquema de las intervenciones del pedófilo/pederasta basado en la propuesta de Levinson.

Para Levinson, los pares adyacentes constituyen los turnos de habla que los hablantes utilizan con mayor cantidad en las conversaciones. Él propone cinco aspectos más frecuentes que están ligados a ambas intervenciones de los hablantes.

Se consideran de esta forma las peticiones, el ofrecimiento, la valoración, las preguntas y las acusaciones que giran en torno al tema del que están hablando o escribiendo el emisor y el destinatario.

A partir de este esquema se presentan dos ejemplos de análisis de las participaciones de pedófilos pederastas de las Conversaciones 1 y 4 en la figura 28 y 29 respectivamente, basado en la propuesta de Levinson.



1ra. parte de pares adyacentes	Módulos encontrados en las conversaciones del corpus	Intención	Temas/Tipos	Categorías gramaticales utilizadas	Muestra. Conversación 1
Petición	Proposiciones sexuales	Acercamiento sexual a la víctima	-Fotografías	Verbo: Have Sustantivo: Photo	17. Do you have another photo besides the one on your post?
			-Encuentro	Verbo: Pick	40. Where would I pick you up at?
Ofrecimiento/ invitación	Secuencia de la conversación	Mediación con el hablante	-Encuentro	Verbo: See Adjetivo: Other, here	32. Well the offer is there if you would like to see each other tonight i am here all alone just watching coocking show
Valoración	Expresión de juicios	Cortejo	-Halagos	Adjetivo: Pretty, sexy Sustantivo: Girl	18. You are very pretty sexy girl. I bet the boys are always chasing you.
Pregunta	Preguntas de experiencia sexual	Petición de información	-Play Self -Virginidad -Sexuales	Verbo: Like, be	60. Do you like to be licked?
	Preguntas personales	Petición de información	-Vestimenta	Adjetivo: Sexy, little	40. A sexy little skirt would be nice but plain old jean would look hot on you
	Preguntas hipotéticas	Proposiciones, descripciones	-Sexo	Sustantivo Night Adjetivo: Sexy	39. Would you be ok spending a fun sexy nighth with me?
Acusación	x	x	x	x	x

Figura 28. Construcción del esquema de las intervenciones del pedófilo/pederasta de la conversación 1.

Con la 1ra parte de pares adyacentes se puede observar que existe un uso reiterativo por parte del pedófilo/pederasta en usos de verbos nucleares para hacer peticiones como en la intervención 17 de la figura 28 o proposiciones para encontrarse como en la intervención 40 de la misma figura.

En la sección de la figura 28 que es el ofrecimiento o la invitación, el uso en conjunto de adjetivos, verbo y adverbios el pedófilo intenta convencer a la víctima, ofreciendo una invitación a cenar esa misma noche como se manifiesta en la intervención 32, está implícito en esta sección la proposición de un encuentro entre ellos.

Para la sección de valoración en la figura 28 del 1rer par adyacente se alude al cortejo a través de halagos que se encuentran dentro de la conversación y que son expresados por el pedófilo/pederasta en la intervención 18. En esta conversación el pedófilo/pederasta enfatiza la belleza del menor de edad.

En la sección de preguntas de la misma figura 28, el uso de verbos y adjetivos son representativos para la información sobre experiencia sexual que ha tenido el menor, expuesta en la intervención 60, seguido de las preguntas acerca de la vestimenta que portan o que pueden usar para su posible encuentro en la intervención 40, de la figura 28. Así como también las preguntas hipotéticas que son las más usuales dentro de la conversación, son formuladas para saber qué están dispuestos a hacer los menores de edad, declaradas en la intervención 39, de la figura 28. La sección de acusaciones de la 1er parte de pares adyacentes no se manifiesta en esta conversación.

Con los pares adyacentes de la figura 29 se puede notar que existe una cierta preferencia por parte del pedófilo/pederasta en usos de verbos esenciales para hacer



peticiones de proposición sexual, es evidente en la intervención 65, siendo inherente la proposición de encuentro en esta misma intervención.

1ra. parte de pares adyacentes	Módulos encontrados en las conversaciones del corpus	Intención	Temas/Tipos	Categorías gramaticales utilizadas	Muestra. Conversacion 4
Petición	Proposiciones sexuales	Acercamiento sexual a la víctima	-Encuentro	Verbo: Get Sustantivo: Room	65. so can't we meet somewhere and get a hotel room.
Ofrecimiento/ invitación	Secuencia de la conversación	Mediación con el hablante	-Encuentro	Adjetivo: Tonight, then	61. How about tonight then
Valoración	Expresión de juicios	Cortejo	-Halagos	x	x
	Preguntas de experiencia sexual	Petición de información	-Play Self -Virginidad -Sexuales	Verbo: Virgin	13. You a virgin
					31. Is your pussy tight
Pregunta	Preguntas personales	Petición de información	-Vestimenta -Condición íntima	Adjetivo: Pussy, tight,	48. what are your pijamas 49 what just a shirt
	Preguntas hipotéticas	Proposiciones, descripciones	-Sexo	Adjetivo: Naked Verbo: Suck Sustantivo: Dick	50. panties and bra
					7. Would you like to get me naked? 24. you can always suck my dick
Acusación	x	x	x	x	x

Figura 29. Construcción del esquema de las intervenciones del pedófilo/pederasta de la conversación 4.

En el siguiente par adyacente que es ofrecimiento o invitación de la figura 29 , el uso en conjunto de verbos y adverbios, el pedófilo intenta convencer a la víctima, ofreciendo una invitación a encontrarse esa misma noche como se muestra en la intervención 61, de la misma forma está inherente en este par la proposición de un encuentro entre ellos.

Para la valoración como siguiente par adyacente se refiere al cortejo a través de halagos, sin embargo, en esta conversación no está presente en la figura.

En las preguntas, el uso de verbos y adjetivos son significativos para las preguntas sobre información de experiencia sexual que ha tenido el menor en la intervención 13 siendo completamente explícito al preguntar sobre su virginidad, seguido de las preguntas acerca de su condición física de partes íntimas y la vestimenta que portan, se muestran en las intervenciones 13, 31, 48, 49 y 50 de la figura 29, así como las preguntas hipotéticas y las más frecuentes dentro de la conversación para saber que están dispuestos a hacer los menores de edad, presentes en la intervenciones 7 y 24 a través de verbos de connotación sexual.

DISCUSIÓN Y CONCLUSIONES

Las investigaciones sobre delitos cibernéticos desde la Lingüística computacional y la Lingüística formal han ido en aumento en países que han dado apertura a nuevas metodologías de investigación para resolver casos penales que



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



Compartir

requieren de mecanismos a otros niveles. En México, la Lingüística y las investigaciones que se desprenden de ella han tenido mayor auge en sectores como el estudio de las lenguas naturales. Respecto a esto, las investigaciones de carácter social, específicamente, en el sector jurídico y penal, han cobrado relevancia por la utilidad que conlleva la perspectiva del lenguaje en delitos donde existe evidencia de este tipo [20].

La realización del análisis de conversaciones escritas de pedófilos y pederastas en internet desde el idioma inglés muestra que el léxico de los hablantes refleja los tópicos ligados a las intenciones del emisor y del receptor de los mensajes [21]. La frecuencia de cada uno de los términos encontrados ha representado una conexión con el tipo de intervención donde aparece dentro de la conversación general, en donde el tipo de turno predominante han sido las preguntas de diversas índoles y cuyos temas se concentran en la obtención de información de distintos rubros de la víctima, esto junto con el engaño es solo una de las estrategias para incentivar a la víctima a caer en el juego del victimario [22].

Se pudo observar que la mayor información para detectar este tipo de conversaciones la dan los nombres o sustantivos que arrojaron a los instrumentos que el victimario usa para la obtención de información de la víctima como fotografías, direcciones y el tipo de ropa que utilizan [23]. De igual modo, los adjetivos son de uso frecuente por el victimario para halagar a la víctima y así ganar confianza para la obtención de los mismos datos mencionados antes. Por otro lado, la ambigüedad del uso de los adverbios por el victimario resulta difícil para la clasificación de estos y establecer en qué contextos son utilizados para la obtención de información de índole sexual.

Herramientas que pueden ser utilizadas para el análisis de conversaciones, y tener opciones en la personalización como spaCy [24] o VaderSentiment [25], además de facilitar la automatización de resultados.

Para finalizar, la importancia de hacer un corpus en español sobre este tipo de conversaciones ayudaría a diversos sectores para frenar, en primera instancia la pornografía infantil y trata de personas que predomina en México, y en una segunda instancia la detención de las personas que ejercen estos delitos, así como los delincuentes que están involucrados en el delito de la pederastia.

FUENTES DE FINANCIAMIENTO

Esta investigación fue financiada parcialmente por la Vicerrectoría de Investigación y posgrado de la Benemérita Universidad Autónoma de Puebla.

DECLARACIÓN DE CONFLICTO DE INTERÉS

No existe conflicto de interés entre los autores.



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.

Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



Compartir

APORTE DEL ARTÍCULO EN LA LÍNEA DE INVESTIGACIÓN

Esta investigación, es un trabajo original realizado por los autores, que ayudará a solucionar una problemática en redes sociales que ha puesto a los menores de edad en peligro por el uso de la tecnología.

DECLARACIÓN DE CONTRIBUCIÓN DE CADA AUTOR

El trabajo colaborativo para el desarrollo de este trabajo de investigación fue fundamental para la sinergia entre diferentes disciplinas. Los autores, Jose Luis y Luis Enrique trabajaron junto con las estudiantes de lingüística en el análisis de las conversaciones en el laboratorio de análisis forense digital. El autor Luis Enrique estuvo supervisando en todo momento el desarrollo de la investigación.

AGRADECIMIENTOS

El reconocimiento para las estudiantes Joselin Hernández Oseguera y Anyeli Vianey Tiempo Carbente de la Facultad de Filosofía y Letras de la licenciatura en lingüística y literatura hispánica de la Benemérita Universidad Autónoma de Puebla que trabajaron arduamente en el laboratorio de análisis forense digital de la Facultad de Ciencias de la Computación de la misma universidad.

REFERENCIAS

- [1] Marugán Solís F, Martín Critikián D. Redes sociales y Generación Z. Doxa Comunicación [Internet]. 1 de enero de 2023;(36):381-99. Disponible en: <https://revistascientificas.uspceu.com/doxacomunicacion/article/view/1707>
- [2] Trigueros Cervantes C, Rivera García E, , Torre Navarro E. D. El Chat como estrategia para fomentar el aprendizaje cooperativo. Una investigación en el Prácticum de Magisterio. Profesorado. Revista de Currículum y Formación de Profesorado [Internet]. 2011;15(1):195-210. Recuperado de: <https://www.redalyc.org/articulo.oa?id=56717469014>
- [3] Cánovas-Pelegrín R, Ballesta-Pagán FJ, Ibañez-López FJ. Percepción de los adolescentes sobre el consumo de redes sociales. RELATEC [Internet]. Zene.2023; 22(1):123-34. disponible en: <https://relatec.unex.es/article/view/4481>
- [4] Ballester-Arnal R., Castro-Calvo J., Gil-Llario M. D. & Giménez-García C. (2014) Relationship Status as an Influence on Cybersex Activity: Cybersex, Youth, and Steady Partner, *Journal of Sex & Marital Therapy*, 40:5, 444-456, DOI: 10.1080/0092623X.2013.772549
- [5] Ballester-Arnal R., Giménez-García C., Gil-Llario M.D. & Castro-Calvo J., Cybersex in the “Net generation”: Online sexual activities among Spanish adolescents, *Comput. Human Behav.*, vol.57, pp.261–266, abril de 2016. Disponible en:<https://doi.org/10.1016/j.chb.2015.12.036>



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.
Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



Compartir

- [6] ACCIÓN PARA PONER FIN A LA EXPLOTACIÓN Y EL ABUSO SEXUALES DE LAS NIÑAS, NIÑOS Y ADOLESCENTES [Internet]. Unicef.org. [citado el 1 de abril de 2023]. Disponible en: <https://www.unicef.org/media/102211/file/CSAE-SUMMARY-WEB-ES.pdf>
- [7] Código Penal del Estado Libre y Soberano de Puebla [Internet]. Gob.mx. Orden Jurídico Poblano; 1986 [citado el 1 de abril de 2023]. Disponible en: <https://ojp.puebla.gob.mx/legislacion-del-estado/item/375-codigo-penal-del-estado-libre-y-soberano-de-puebla>
- [8] Perverted-justice.com - The largest and best anti-predator organization online [Internet]. Perverted-justice.com. [citado el 1 de abril de 2023]. Disponible en: <http://www.perverted-justice.com/>.
- [9] Vera Morales, K. N. La violencia de género en línea: Guía de conceptos básicos, herramientas de seguridad digital y estrategias de respuesta. Secretaría General de la Organización de los Estados Americanos. 21 de octubre de 2021.
- [10] LanguageTool - Corrector ortográfico, gramatical y de estilo [Internet]. LanguageTool. [citado el 1 de abril de 2023]. Disponible en: <https://languagetool.org/es>.
- [11] Linguakit [Internet]. Linguakit.com. [citado el 1 de abril de 2023]. Disponible en: <https://linguakit.com/es/analisis-completo>
- [12] Levinson SC. Cambridge textbooks in linguistics: Pragmatics [Internet]. Cambridge, Inglaterra: Cambridge University Press; 2018. Disponible en: <http://dx.doi.org/10.1017/cbo9780511813313>.
- [13] Vázquez Carranza, A. La perspectiva lingüística del análisis conversacional. Propuestas metodológicas para el trabajo y la investigación lingüística: aplicaciones teóricas y descriptivas (pp.510-530) Publisher: Universidad de Colima.
- [14] Garayzábal Heinze E, Hidalgo de la Guía I. Detección de depredadores sexuales en los chats y la captación de menores: el papel de la lingüística forense. Tonos Digit [Internet]. 2020 [citado el 1 de abril de 2023]; Disponible en: <https://digitum.um.es/digitum/handle/10201/96042>.
- [15] Blancafort HC, Fort B, Valls Tuson J, Tuson Valls A. Las Cosas del Decir: Manual de Análisis del Discurso. Ariel; 2002.
- [16] McGlashan, M. Series Editing & Design. Corpus linguistics: Some key terms (The ESRC Centre for Corpus Approaches to Social Science (CASS: Briefings), Published 2013). <http://cass.lancs.ac.uk/wp-content/uploads/2013/12/CASS-Gloss-final1.pdf>

- [17] Parodi G. Lingüística de corpus: Una introducción al ámbito. *Revista de Lingüística Aplicada* [Internet]. 2008; 46(1). Disponible en: <http://dx.doi.org/10.4067/s0718-48832008000100006>
- [18] Manning C, Schuetze H. *Foundations of statistical natural language processing*. Londres, Inglaterra: MIT Press; 1999.
- [19] Henning S. La construcción de la imagen social en dos pares adyacentes: Opinión-acuerdo/desacuerdo y ofrecimiento-aceptación/rechazo : Un estudio de la conversación familiar sueca y española. Department of Romance Studies and Classics, Stockholm University; 2015
- [20] Delitos informáticos: Impunidad organizacional y su complejidad en el mundo de los negocios. *Rev Venez Gerenc* [Internet]. 2020; 89(89). Disponible en: <http://dx.doi.org/10.37960/revista.v25i89.31534>
- [21] DeHart D, Dwyer G, Seto MC, Moran R, Letourneau E, Schwarz-Watts D. Internet sexual solicitation of children: a proposed typology of offenders based on their chats, e-mails, and social network posts. *J Sex Aggress* [Internet]. 2017; 23(1):77–89. Disponible en: <http://dx.doi.org/10.1080/13552600.2016.1241309>
- [22] De Santisteban P, Almendros C, Gámez-Guadix M. Estrategias de persuasión percibidas por adolescentes en situaciones de seducción pederasta en Internet (online grooming) (pp. 243-262) [Internet]. Behavioralpsycho.com. [citado el 1 de abril de 2023]. Disponible en: https://behavioralpsycho.com/wp-content/uploads/2018/09/02.Santisteban_26-2r.pdf.
- [23] Whittle H, Hamilton-Giachritsis C, Beech A, Collings G. A review of online grooming: Characteristics and concerns. *Aggress Violent Behav* [Internet]. 2013;18(1):62–70. Disponible en: <http://dx.doi.org/10.1016/j.avb.2012.09.003>
- [24] SpaCy · industrial-strength Natural Language Processing in Python [Internet]. Spacy.io. [citado el 1 de abril de 2023]. Disponible en: <https://spacy.io>.
- [25] Welcome to VaderSentiment's documentation! — VaderSentiment 3.3.1 documentation [Internet]. Readthedocs.io. [citado el 1 de abril de 2023]. Disponible en: <https://vadersentiment.readthedocs.io/en/latest/index.html>.

NOTA BIOGRÁFICA



Luis Enrique Colmenares-Guillen. **ORCID iD**  <https://orcid.org/0000-0002-9921-8813>
Docente-investigador de la Facultad de Ciencias de la Computación, coordinador del Laboratorio de Análisis Forense Digital de la Benemérita Universidad Autónoma de Puebla. Ha trabajado en algoritmos de aprendizaje Automático y de procesamiento de lenguaje natural. Pertenece al Padrón de Investigadores y de consultores de la BUAP y al Sistema Nacional de Investigadores.



José Luis Jiménez Aguilera. **ORCID iD**  <https://orcid.org/0000-0001-7965-4207>
Tesis de la Benemérita Universidad Autónoma de Puebla del proyecto de tesis aprobado, “Detección de pederastia en conversaciones digitales con procesamiento de lenguaje natural”. Colaborador del Laboratorio de Análisis Forense Digital de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla.



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Colmenares-Guillén & Jiménez -Aguilera. Una aproximación para la detección de contenido sexual en conversaciones digitales.
Enero – Junio 2023

<https://doi.org/10.33210/ca.v12i1.410>



[Compartir](#)