

Estado de la publicación: El preprint ha sido publicado como artículo en una revista
DOI del artículo publicado: <https://doi.org/10.37135/chk.002.21.01>

ACERCAMIENTO A UN PERFIL LINGÜÍSTICO DE POTENCIALES AGRESORES SEXUALES INFANTILES: ANÁLISIS DE INTERVENCIONES TEXTUALES ONLINE

Luis Enrique Colmenares-Guillén, Yosbeli Delgado González

<https://doi.org/10.1590/SciELOPreprints.6563>

Enviado en: 2023-08-04

Postado en: 2023-08-04 (versión 1)

(AAAA-MM-DD)

ACERCAMIENTO A UN PERFIL LINGÜÍSTICO DE POTENCIALES AGRESORES SEXUALES INFANTILES: ANÁLISIS DE INTERVENCIONES TEXTUALES ONLINE
APPROACHING A LINGUISTIC PROFILE OF POTENTIAL CHILD SEXUAL AGGRESSORS: ANALYSIS OF ONLINE TEXTUAL INTERVENTIONS

Luis Enrique Colmenares-Guillén¹, ORCID: <https://orcid.org/0000-0002-9921-8813>

Yosbeli Delgado González², ORCID: <https://orcid.org/0000-0002-0877-564X>

¹Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Coodinador del laboratorio de Análisis Forense Digital. Puebla, México, email: enrique.colmenares@correo.buap.mx

²Benemérita Universidad Autónoma de Puebla, Facultad de Filosofía y Letras, Lic. En Lingüística y Literatura hispánica. Puebla, México, email: delgado.gonzalez.yosbeli@gmail.com

RESUMEN

Cada hablante posee características lingüísticas particulares condicionadas por las circunstancias de contacto. En la lingüística forense, lingüística aplicada y lingüística de corpus se considera que una regularidad puede ser reconocida como un *modus operandi* específico y este suele corresponder a comunidades que comparten un mismo fin, con sujetos igualmente precisos, mediante determinadas estrategias lingüísticas. En este artículo se realiza un acercamiento a un perfil lingüístico, correspondiente a potenciales agresores sexuales de infantes, con el fin de obtener un comportamiento lingüístico constante que posteriormente permita reconocerlo en conversaciones *online*. La metodología empleada en este estudio es empírica, basada en el análisis de lenguaje natural que propone la lingüística de corpus de Rayson. Mediante ella, se analiza el comportamiento lingüístico en la comunicación textual vía *online* extraído de un corpus (con un procesamiento previo), constituido de cinco conversaciones, en las cuales se estableció la frecuencia léxica para generar diecisiete módulos temáticos. Finalmente, se encontró regularidad en las primeras siete intervenciones, lo que señala un comportamiento lingüístico específico: un perfil lingüístico que podría guiar a un reconocimiento oportuno de las identidades *online* que representan un peligro para las infancias.

PALABRAS CLAVE: Identidad, comportamiento lingüístico, estructura sintáctica, frecuencia léxica.

ABSTRACT

Each speaker has linguistic characteristics conditioned by the circumstances of contact. In forensic linguistics, applied linguistics, and corpus linguistics it is considered that regularity can be recognized as a specific modus operandi, and this usually corresponds to communities that share the same purpose, with equally precise subjects, and by means of specific linguistic strategies. In this article, an approach to a linguistic profile corresponding to potential sexual offenders of infants is carried out to obtain a constant linguistic behavior that subsequently allows its recognition in online conversations. The methodology used in this study is empirical and based on the natural language analysis proposed by Rayson's corpus linguistics. Through it, the linguistic behavior in online textual communication extracted from a corpus (with prior processing) is analyzed. It consists of five conversations, in which the lexical frequency was established to generate seventeen thematic modules. Finally, regularity was found in the first seven interventions, which points to a specific linguistic behavior: a linguistic profile that could lead to a timely recognition of online identities that represent a danger to children.

KEYWORDS: *Identity, language behavior, syntactic structure, lexical frequency.*

Recibido: (22/03/2023)

Aceptado: (26/07/2023)

INTRODUCCIÓN

En la actualidad, gracias a internet la comunicación en tiempo real con personas no identificadas resulta posible. El desarrollo de identidades ocultas (Bauman, 2000) detrás de cada intercambio comunicativo es posible a través de redes sociales; no obstante, tiene importantes consecuencias debido a que —en ocasiones— puede llegar a facilitar conductas criminales (Perverted-Justice, 2008/2019). El presente estudio se ubicó en los delitos cibernéticos de carácter sexual que, pese a que continúan en un largo proceso de legislación, se comportan como frecuentes.

Por otro lado, la infancia con acceso a las redes sociales en línea (*online*) se conforma como uno de los sectores más vulnerables a delitos como abuso/explotación sexual, trata, trabajos forzados y rapto (Interpol, 2023), debido a la facilidad de prestadores para ocultar sus identidades y al uso descuidado del internet por parte de la comunidad infantil, lo que propicia una situación de constante riesgo.

Dentro de este contexto, el ambiente tecnológico crea un marco de convivencia, un ambiente para los usuarios *online* que atienden a características textuales específicas —que comparten como comunidad— pero distintivas —de otras comunidades—, por esto, se considera plausible la creación de posibles y futuras taxonomías textuales que compongan un perfil —donde las piezas léxicas funcionen como señales que den pie a la configuración de módulos temáticos para el posterior armado de un perfil lingüístico— cada vez más acotado, que represente una identidad grupal *online* sistemática.

El presente trabajo busca construir un acercamiento a un perfil lingüístico —mediante un corpus— donde se reconozcan patrones lingüísticos (Sacks, 2010; Schegloff, 2012; Sacks et al., 2015), recurrentes (MacMartin & Wood, 2007) y detectables (Rashid et al., 2013). Este objetivo se integra a la construcción de una posible identidad lingüística *online* (Chiang & Grant, 2019); en este caso, describe a una comunidad lingüística.

En otras palabras, el presente análisis ejecuta un primer acercamiento —en el idioma español— a la frecuencia de piezas léxicas; en este caso particular, se refiere a la identificación gramatical de nominales, verbos y adjetivos, de carácter sexual o sexualizado (en el documento se entiende por sexualización todo uso del lenguaje con connotación, significado y sentido sexual, a toda construcción sintáctica o pieza léxica que sea propia del campo de la sexualidad (órganos sexuales) y sus alcances (relaciones o contactos sexuales)).

Asimismo, se consideran los cambios de significado o doble sentido que se apoyen de símiles, atenuaciones o ejemplificaciones, que respondan o referencien al carácter sexual, teniendo como meta su reconocimiento a nivel léxico, y de agrupación a nivel pragmático por medio de las intervenciones y la clasificación en módulos por temas que, de acuerdo con la función comunicativa comprendida, arrojen como resultado la generación de un perfil lingüístico.

METODOLOGÍA

El presente artículo de investigación utilizó métodos empíricos que parten esencialmente de la lingüística forense y de corpus que plantea la evaluación de un perfil lingüístico que implica distintos niveles. La lingüística forense, define Esteva (2016), se erige como un tema amplio que requiere mucho por granjear en la lengua española para generar un estudio puntual:

En último término, cabría incluir (...) los estudios sobre dialectología, sociolingüística, semántica, pragmática y psicolingüística forenses enfocados a la elaboración de perfiles

lingüísticos, los cuales servirán de ayuda en la investigación policial previa y en la posterior judicial para la preparación de la vista oral, puesto que podrían identificar a los posibles autores de textos orales y escritos. (p. 15)

Debido a que se busca “lograr una identificación de hablantes; el análisis o la atribución de autoría de textos escritos” (Esteva, 2016, p. 15) estriba en distintos niveles y precisiones de la lengua y, a su vez, en los hablantes específicos. Este primer acercamiento busca encontrar ejes temáticos en el *modus operandi* de lengua escrita *online* de los posibles agresores como comunidad lingüística.

IDENTIDAD *ONLINE* Y PERFIL LINGÜÍSTICO

La representación de un perfil lingüístico, sus ejes y posibilidades como estudio interdisciplinario requieren fundamentos que lo soporten y que planteen la identidad lingüística y la identidad *online* o *performance* (Bauman, 2000) como un concepto básico.

Así, para el asentamiento de esta base, la psicolingüística y la lingüística aplicada forense suelen colaborar, aunque también se utiliza la sociolingüística (condicionamientos por edad, género, lugar de origen, entre otros) y la pragmática (acotamientos de circunstancias específicas del momento del habla), especialmente mediante el uso de investigaciones de género, edad (Argamon et al., 2007) y de jerarquía de los hablantes aunada a su proyección en los diferentes tipos de discursos (Ehrlich, 2007). Cada disciplina delimita y matiza la acotación en la expresión textual.

Para este propósito, deviene necesario partir de un amplio espectro de estudios: el análisis del discurso por Rojas & Suárez (2008), la lingüística forense por Turell (2010) y de corpus de Rayson (2015), Stefanowitsch (2020), Cheng (2009) como ramas generales por medio de una indagación especializada.

En primer lugar, desde el análisis del discurso existe una serie de estructuras sintácticas e interpretaciones dentro de un marco de alternativas y ambigüedades recurrentes en contextos específicos (Sacks et al., 2015).

Las técnicas lingüísticas utilizadas pueden ser descritas o identificadas mediante la ocurrencia de unidades —sintácticas, léxicas o de estrategias discursivas— en la interacción —sin hablar de estilística— regidas por motivaciones.

Asimismo, la conversación contiene reglas generales de intercambio comunicativo (saludo, invitaciones, proposiciones, despedida, entre otros; expresado como turnos bajo sistemas operantes con distribuciones, secuencias y transferencias evidentes), con unidades como las secuencias lingüísticas que pueden describirse, organizarse y clasificarse para dar pie a modelos que permiten adecuaciones para afiliaciones de comunidades (Tannen et al., 2015). De esta manera se construye secuencias desde los turnos con una guía elegida (Sacks, 2010), en este caso, léxica. Es preciso exponer un marco del estudio pragmático para la creación del perfil de frecuencia de módulos, en virtud de que se extraen los turnos de habla del posible agresor y, posteriormente, solo las intervenciones explícitamente sexuales que conduzcan a identificar un perfil lingüístico específico.

De este modo, resulta imprescindible reconocer las guías esenciales del análisis del discurso (Calsamiglia & Tusón, 2001; Jaworski & Coupland, 2014; Schiffrin, 2003; Van Dijk, 1997). El intercambio comunicativo, investigado por Sacks et al. (2015) en la interacción de turnos; en la conducta comunicativa (Goffman, 1955); en contextos específicos y generales (Sacks, 2010); y en la organización de las secuencias en la interacción comunicativa (Schegloff, 2012).

Por otro lado, ciertas investigaciones muestran ejes de análisis para la identificación de personas e identidades digitales en crímenes cibernéticos (Rashid et al., 2013), donde se realizan comparación de estilos, análisis categorial y de discurso, focalización en los patrones de interacción en tiempo real con categorizaciones de campos semánticos, la huella del estilo de lenguaje (*emojis*, puntuación y vocabulario), análisis del lenguaje natural y estructural de las identidades engañosas, además de las diferentes estrategias lingüísticas en las formas de control

de los tópicos (Chiang, 2019; Chiang & Grant, 2019; Tannen, 2016; Tusón, 2016; Brookes & McEnery, 2020).

Estos elementos contribuyen a la construcción de mecanismos que concretan la atribución de una identidad textual o lingüística que, inicialmente, suele generarse con base en descripciones de un estilo lingüístico con frecuencia de patrones en distintos niveles.

El gran abanico de posibilidades del lenguaje hace difícil comprometerse con patrones rígidos de expresión textual; sin embargo, pese a ser conscientes de la libertad infinita de expresión que se tiene, es un hecho que los humanos tienden a un comportamiento repetitivo (Grant & MacLeod, 2016).

Por su parte, la lingüística computacional trabaja con piezas léxicas, estructuras recurrentes, detecciones de perfiles o de atribuciones de autoría (Sierra et al., 2020) proyectados en la sistematización mediante técnicas y métodos computacionales, huellas dirigidas a una atribución/identificación de autoría.

A su vez, la lingüística de corpus realiza análisis masivos —perfil de frecuencia, concordancia, colocación, grupos o paquetes léxicos en distintos niveles de la lengua— (Rayson et al., 2016), señala frecuencias; además, describe el comportamiento de las piezas léxicas en diferentes niveles lingüísticos en función de encontrar patrones y repeticiones de datos.

Finalmente, desde la lingüística forense se plantean varios métodos: la acústica forense; la comparación forense de textos escritos e informe pericial; la elaboración de una identidad que atiende estudios de conducta con atención en las recurrencias; la manipulación del lenguaje (MacMartin & Wood, 2007; O’Keeffe & Breen, 2007); de interpretación (Greenlee, 2012); la polisemia y ambigüedad (Ng, 2013); la velocidad de las respuestas (Jenkins & Dando, 2012); comparación de similitudes, de patrones de preferencia desde el lexicón de referencia, enfocando el género, edad, la longitud de los textos, los tiempos de respuestas o los temas de la conversación (Tomblin, 2012); hasta las frecuencias de las distintas funciones, distribuciones o colocaciones de las palabras; el léxico, la puntuación, análisis de contenido, y de niveles de sofisticación sintáctica (Grant & MacLeod, 2016).

Se proyecta así un tema amplio que requiere una delimitación de diversas disciplinas que susciten la elaboración de perfiles lingüísticos orales o escritos. Debido a lo complejo del asunto, este acercamiento se centra en encontrar ejes temáticos en el *modus operandi* de la lengua escrita *online* de los posibles agresores que forman una comunidad lingüística.

EL CORPUS Y EL CONTEXTO OPERACIONAL

El corpus corresponde a conversaciones escritas que, en este caso, se encuentran en el ciberespacio, mediadas por un mecanismo tecnológico, por lo que no existe muestra de vocalizaciones, tonos ni acentos; sin embargo, posee otros aspectos externos a la lengua como *emojis*, *stickers* y *gifts*, pertenecientes a la plataforma cibernética, que de igual manera matizan los sentidos en la comunicación.

Se entiende que no se trata de una comunicación real (se refiere a que no existe contacto directo, por lo que ciertos componentes pueden ser ocultos o fingidos como la edad, el carácter e incluso el género), sino de una vía *online* —mayormente— textual; esto brinda la oportunidad de apuntar al desarrollo y automatización de técnicas y métodos para la detección de una conducta particular (Grant & MacLeod, 2012; Jenkins & Dando, 2012; Egbert et al., 2022; MacLeod & Wright, 2020).

Las estudiantes de Criminología que trabajaron en el proyecto: “Creación de una metodología para el análisis forense, utilizando el texto digital”, colaboraron en el laboratorio de análisis forense, realizando la recolección del corpus durante un periodo de seis meses. Los aspectos metodológicos y teóricos específicos corresponden a una investigación puntual (Colmenares-Guillén et al., 2020) constituida por cinco conversaciones en español obtenidas del perfil de una red social *Facebook*, elaborado a partir de la identidad fingida de una niña aproximadamente de 14 años de edad.

Para la obtención del corpus se realizó previamente una revisión teórica documental desde la perspectiva socio/jurídica y criminológica. Sin embargo, también requirió un estudio experimental auxiliar con bases de la ciencia computacional y de la lingüística, con el objetivo de comprender la diferencia de los soportes empleados para la comunicación y sus matices que caracterizan la obtención de los datos.

Debido a que son conversaciones *online*, se debe precisar: 1. Las características del corpus y 2. Los roles de los usuarios: aquel que recabó información y quien la proporcionó.

Las investigadoras son acreedoras de una identidad falsa con objetivos de investigación (Criminal Justice Inspectorates, 2014); por ello, el rol que personifican requiere preparación, tales como, nociones de identidades criminales y de identidades vulnerables, entre otros aspectos.

Sin embargo, referimos a esta identidad de infante fingido/identidad fingida (Bauman, 2000), desde la idea de un perfil falso construido en una red social virtual para la específica adquisición del corpus —suponiendo la caracterización de una niña— que ayudó a recabar la presente información a partir de su constitución artificial o fingida.

Se aclara que este modelo surge bajo una noción de mecanismo formal computacional que responde a posibilidades, probabilidades, convenciones y estrategias lingüísticas desde tendencias léxicas (Kaplan, 2004).

Por otro lado, la justificación para considerar a los usuarios —de quienes se extrajo la información— como posibles agresores sexuales se basa en:

1. Las peticiones de fotografías a la menor de edad, lo cual corresponde a una conducta delictiva (Artículo 202, 2018) en México y,
2. El acuerdo al que llegaban los usuarios respecto a las actividades sexuales que planeaban llevar a cabo, antes de solicitar la cita presencial.

Ahora, para los datos lingüísticos obtenidos y la cuantificación de frecuencias de las clases gramaticales —se recuperan los nominales, verbos y adjetivos que dirigen el primer acercamiento a los turnos de habla—, se retomó la metodología —empírica, basada en datos que explotan muestras de lenguaje natural— de la lingüística de corpus de Rayson (2015), que a su vez adapta la de Rayson (2008):

1. Pregunta: idear una pregunta de investigación.
2. Construcción: diseño y compilación del *corpus*.
3. Anotación: análisis manual o automático del corpus.
4. Recuperación: análisis cuantitativos y cualitativos del corpus.
5. Interpretación: interpretación manual de resultados.

El corpus presentó irregularidades ortográficas; debido a esto, se hizo un pre-procesamiento que consistió en la extracción de las intervenciones de habla del potencial agresor y de su limpieza —hora, fecha, *username*, de la corrección ortográfica, la eliminación de recursos externos a la lengua como fueron los *emojis*, además de las contracciones—. De este modo, fue posible contar y etiquetar las piezas léxicas.

BASES LINGÜÍSTICAS: PIEZAS LÉXICAS, INTERVENCIONES Y AGRUPACIÓN

Se reconoce la intervención de habla como la parte del turno de conversación en que el victimario escribe una línea de texto. Asimismo, se reconoce al turno conversacional o de habla como la unidad básica de la conversación. Desde un punto de vista formal, la conversación se caracteriza precisamente por la alternancia de varios turnos, es decir, por la sucesión de intervenciones a cargo de diferentes interlocutores (Centro Virtual Cervantes, 2008).

A partir de los criterios de turnos e intercambios, este perfil lingüístico tiene como base las intervenciones construidas con piezas léxicas sexuales o sexualizadas que dan pie a una intervención destacada. Aquellas que sean recurrentes, constituyen un módulo comunicativo que se clasificó de manera temática.

Además, se entiende aquí como un módulo temático a una intervención de carácter sexual o sexualizada que tenga una frecuencia y contenga características propias como piezas léxicas que la definan, desde la perspectiva de los usos marcados —o no comunes— de ciertas piezas léxicas, también pueden tener significados especiales en una comunidad lingüística propia que permitan realizar análisis acerca de usos y distinciones en contextos específicos (en lingüística forense refieren a las características analizables del lenguaje en contextos específicos como juicios legales; no obstante, los tintes marcados de manipulación y significados múltiples pueden analizarse en diferentes contextos), para la evaluación de un perfil lingüístico.

RESULTADOS Y DISCUSIÓN

Para la creación del perfil lingüístico, se identificaron y extrajeron las intervenciones del posible agresor, distinguiéndolas de la infanta fingida. A partir de estas se precisaron las piezas léxicas: nominales, verbos y adjetivos sexuales o sexualizados. Después se analizó la estructura sintáctica general de las intervenciones para agruparlas en módulos temáticos que se ordenaron al distinguir su frecuencia y con base en ella, se organizó un patrón lingüístico (May-Chahal et al., 2014) de intervenciones, del cual resulta un primer piloto experimental, empírico y descriptivo. El análisis constó de cinco pasos:

1. Pregunta o planteamiento: se refiere a la construcción de un perfil lingüístico de potenciales agresores sexuales de infantes en línea.
2. Construcción: consta de cinco conversaciones desde el perfil ficticio de una menor de edad.
3. Anotación: apunta al análisis manual del *corpus*. La contabilización de las piezas léxicas, comprendidas como palabras de cualquier categoría gramatical, presentó el siguiente número de intervenciones y de piezas léxicas como se muestra en la Tabla 1.

Tabla 1: Cuantificación general de intervenciones y piezas léxicas

CONVERSACIÓN	I	II	III	IV	V
Número de intervenciones	536	185	47	32	35
Número de piezas léxicas	2742	562	221	244	174

4. Recuperación. Análisis cuantitativos y cualitativos del corpus: léxico, clases de palabras; en este apartado se muestran los resultados de las tres clases de palabras seleccionadas — nominales, verbos y adjetivos— y su frecuencia en el corpus. En las Tablas 2, 3 y 4 de frecuencia léxica los números arábigos remiten al número de veces que las piezas aparecieron en las conversaciones.

NOMINALES

Los nominales más frecuentes en las intervenciones que se identificaron con significado contextual distintivo se muestran en la Tabla 2.

Tabla 2: Frecuencia de piezas léxicas: nominales

Conversación	I	II	III	IV	V
Frecuencia de nominales	28 Chiquita	28 Amor	8 Bebé	15 Bebé	7 Amor
	14 Niña	10 Bebé	2 Viejito	1 Ropa	3 Bebé
	6 Amiga	6 Chiquita	1 Chiquita	1 Amor	2 Novia

La tabla muestra que el nominal con mayor constancia en las conversaciones fue bebé (conversaciones II-III-IV-V).

VERBOS

Los verbos se muestran en infinitivo, considerando dentro de ellos las diferentes conjugaciones que tuvieron en las conversaciones. Es importante señalar que, aislados como aparecerán pueden no evidenciar las cargas sexuales; esto se debe a que algunos actúan como cópulas verbales. Por lo que resulta provechoso estudiar los casos de los modificadores, pues crean los contextos y demuestran las cargas semánticas que guían la conversación y su tema.

No obstante, en este acercamiento han sido identificados y se evalúa su importancia. Los verbos más frecuentes en las intervenciones que se identificaron con connotaciones con significado contextual distintivo se muestran en la Tabla 3:

Tabla 3: Frecuencia de piezas léxicas: verbos

Conversación	I	II	III	IV	V
Frecuencia de verbos	10 Gustar	7 Gustar	2 Mandar	4 Regalar	2 Gustar
	12 Querer	5 Extrañar	1 Gustar	4 Enviar	2 Pasar
	9 Tener	3 Chatear	1 Encantar	2 Gustar	1 Mandar

Un aspecto a resaltar es la utilización de tres elecciones léxicas verbales —verbos directivos o de petición— para solicitar una fotografía: mandar, enviar, regalar; lo que apunta a la evidencia de la variación léxica en las peticiones. Asimismo, el verbo con mayor constancia en las conversaciones fue gustar (presente en todas), un verbo psicológico o de estado mental.

ADJETIVOS

Los adjetivos más frecuentes en las intervenciones que se identificaron con connotaciones con significado contextual distintivo se muestran en la Tabla 4.

Los adjetivos que tienen solo un número de aparición se muestran en el orden en que se presentaron en la conversación, o debido a que dentro de la misma se encuentra la raíz de la palabra con modificadores. El adjetivo con mayor constancia en las conversaciones fue hermosa (conversaciones I-II-IV-V).

Tabla 4: Frecuencia de piezas léxicas: adjetivos

Conversación	I	II	III	IV	V
Frecuencia de adjetivos	24 Hermosa	6 Hermosa	1 Sexy	4 Hermosa	1 Hermosa
	3 Sexy	1 Chiquita	1 Rica	1 Ricas	1 Bonita
	4 Bonita	1 Buena	1 Bebé		1 Chica

5. Interpretación de los módulos: Se identificaron 17 módulos temáticos por medio de su función comunicativa, en la Tabla 5 se enumeran y muestran ejemplos para comprender mejor la referencia de la categorización.

El número romano de la columna ejemplo da la referencia al número de la conversación; mientras que los números arábigos de la misma columna aluden al número de turno enlazado al corpus del que se extrajeron y conciernen a la organización específica del análisis generado a partir de la clasificación de turnos.

De este modo, de acuerdo con el patrón de intervenciones, el resultado más representativo consistió en la frecuencia de las primeras siete intervenciones de las conversaciones, con respecto a la reiteración de los módulos:

- En la primera intervención existe la regularidad del módulo I. Saludo adjetivado.
- En la segunda intervención es constante una petición de información (conversación I-IV-V), específicamente de la localización de la infanta.
- En la tercera intervención, como petición de información, se localizan preguntas acerca de la edad (conversación III-V) o el reconocimiento de la brecha de edades (conversación III).
- La cuarta intervención presenta dos módulos temáticos en el mismo porcentaje: el reconocimiento de la brecha de edad (conversación I-V) y halagos físicos de la infanta (conversación III-IV).
- En la quinta intervención hay mayor aparición del módulo de petición de información, específicamente de fotografías (conversación III-IV-V).
- En la sexta intervención se presentan dos módulos temáticos: 1. petición de información (conversación I-II-IV) donde se solicita vestimenta, si tiene alguna pareja, vida cotidiana, y 2. halagos físicos (conversación III-V).
- En la séptima intervención también hay mayor aparición del módulo temático de petición de información (conversación I-II-III-V): insistencia de información, compañía, vida cotidiana y fotografías.

Tabla 5: Módulos temáticos

Módulo temático	Ejemplo
I. Saludo: adjetivado o con agradecimiento.	III. 1. Hola, cosita rica, ¿cómo estás, sofí? IV. 1. Hola, gracias por aceptarme
II. Petición de información: localización	II. 9. ¿De dónde eres? II. 11. ¿Dónde vives?
III. Petición de información: edad	V. 5. Oye, eres muy chica ¿verdad?
IV. Reconocimiento: brecha de edad	III. 9. Estás bien chiquita, bebé. III. 10. Yo ya estoy viejito jejejeje pero sabroso jejejeje soy ruco chavo I. 10. Creo que eres una niña para mí
V. Petición de información: fotografías	I. 24. Regálame una foto, te veo 138. A ver 140. Tus pechos II. 79. Mándame una foto donde se vea tu carita
VI. Halagos: físico	II. 19. Has de estar súper hermosas III. 30. Qué guapa
VII. Petición de información: indagación de pareja	I. 15. Qué afortunado tu novio. 16. O esposo, lo que tengas II. 81. ¿Tienes novio?
VIII. Petición de información: insistencia de respuestas	I. 327. ¿Por qué ya no respondes? 328. Por favor, dime algo
IX. Preguntas y proposiciones hipotéticas	I. 493. Mi niña, me da miedo ¿y si me enamoro de ti? 506. Imaginate qué afortunado sería que te fijaras en mí 282. Me gustaría acariciarte, besarte. 188. ¿Te gustaría tocarla, acariciarla?
X. Petición de información: vestimenta	I. 301. ¿Cómo te veras ahorita? 33. ¿Cómo te ves ahorita?
XI. Proposición directa	I. 283. Vente conmigo 284. Cásate conmigo
XII. Proposición indirecta	I. 55. ¿Te gustaría que te invitara a salir? 60. A mí me gustaría saber qué se siente tener una chica como tú, de tu edad 61. Hacerla mía
XIII. Petición de información: vida íntima	I. 74. ¿A poco nunca lo has hecho? 75. ¿Y te gustaría?
XIV. Petición de información: compañía	I. 367. Pero ¿a poco no te regañan tus papás si te ven? II. 26. ¿No te regañan por chatear conmigo?
XV. Petición de información: vida cotidiana	I. 403. Vivo solo en mi casa. 404. Sí, hoy las vi, las llevé a comer y estuve un rato con ellas 405. Menos con mi ex, con ella ya nada.
XVI. Expresión de deseos	I. 311. Quiero verte desnuda. 316. Quería yo verte desnuda, aunque sea de la mitad para abajo, aunque tu carita no se te vea.
XVII. Ofrecimiento	IV. 31. Si necesitas recargas a tu número, me dices 32. Amor, si necesitas plata para un celular, dime que yo te ayudo, si quieres II. 91. Te mando dinero

El orden de las intervenciones agrupadas en módulos temáticos se presenta en la siguiente Tabla 6.

Tabla 6: Perfilación de módulos temáticos

No. de intervención	CONVERSACIÓN I	CONVERSACIÓN II	CONVERSACIÓN III	CONVERSACIÓN IV	CONVERSACIÓN V
1	I. Saludo: adjetivado o con agradecimiento.	I. Saludo: adjetivado o con agradecimiento.	I. Saludo: adjetivado o con agradecimiento.	I. Saludo: adjetivado o con agradecimiento.	I. Saludo: adjetivado o con agradecimiento.
2	II. Petición de información: localización	VIII. Petición de información: insistencia de respuestas	III. Petición de información: edad	II. Petición de información: localización	II. Petición de información: localización
3	III. Petición de información: edad	II. Petición de información: localización	IV. Reconocimiento: brecha de edad	V. Petición de información: fotografías	III. Petición de información: edad
4	IV. Reconocimiento: brecha de edad	V. Petición de información: fotografías	VI. Halagos: físico	VI. Halagos: físico	IV. Reconocimiento: brecha de edad
5	VI. Halagos: físico	VI. Halagos: físico	V. Petición de información: fotografías	V. Petición de información: fotografías	V. Petición de información: fotografías
6	VII. Petición de información: indagación de pareja	XV. Petición de información: vida cotidiana	VI. Halagos: físico	X. Petición de información: vestimenta	VI. Halagos: físico
7	VIII. Petición de información: insistencia de respuestas	XIV. Petición de información: compañía	V. Petición de información: fotografías	VI. Halagos: físico	XV. Petición de información: vida cotidiana

Con base en ello, el resultado del patrón de intervenciones por módulos temáticos es el siguiente *modus operandi*: I. Saludo: adjetivado o con agradecimiento; II. Petición de información: localización; III. Petición de información: edad; IV. Halagos: físico; V. Petición de información: fotografías; VI y VII también petición de información: indagación de pareja, vestimenta, insistencia de respuestas y vida cotidiana.

CONCLUSIONES

El presente artículo generó un acercamiento a un perfil lingüístico de posibles agresores *online* como comunidad lingüística, mediante la creación de un corpus en español resultado de los módulos temáticos por su intención comunicativa en las cinco conversaciones analizadas. Este perfil lingüístico precisa información léxica en tres categorías gramaticales extraídas de las intervenciones textuales *online* entre una infante ficticia/fingida y un agresor sexual. Lo anterior permitió observar un patrón en las piezas léxicas, dirigiendo a un armado de módulos temáticos, lo que posibilita identificar posibles agresores sexuales en otros escenarios reales.

Asimismo, de acuerdo con el reconocimiento de los módulos con cargas sexuales se identificó una frecuencia de las piezas léxicas que genera un habla específica de los victimarios hacia las víctimas: el nominal más frecuente fue bebé, el verbo con mayor aparición fue gustar y la recurrente aparición del adjetivo hermosa.

Ahora bien, en cuanto a los módulos temáticos, se identificó que el *modus operandi* se genera por una secuencia de intervenciones en un orden particular: un saludo específico, dos peticiones de información, la primera de localización y una posterior de edad; de manera consecutiva se presentan halagos físicos, que a su vez anteceden a la petición de fotografías.

Con ello, se logra construir un perfil frecuente en la dinámica de comunicación —se reconoce una conducta lingüística— de agresores sexuales con sus posibles víctimas. En vista de lo abordado, se permite una posible intervención por medio de aspectos lingüísticos y alienta a considerar nuevas formas preventivas y de acción que versen propiamente en ellos.

Los hallazgos laterales al propósito regente del estudio, pero innegablemente evidentes, parten del notable cambio de la interacción, de acuerdo al uso de los tiempos verbales —el destacable uso del subjuntivo y las estructuras copulativas, que a su vez abordan la semántica de las piezas léxicas en contexto y sus interacciones—.

Finalmente, a nivel sintáctico, existe la posibilidad de que las palabras se sexualicen dentro del contexto de contacto léxico; su recurrencia en el corpus denota aspectos sintácticos marcados en el *modus operandi*, y esto abre la posibilidad a futuras investigaciones en otros niveles de la lengua que sumen al trabajo de la lingüística aplicada, a fin de consolidar herramientas que sean de utilidad para la sociedad.

DECLARACIÓN DE CONFLICTOS DE INTERESES: Los autores declaran no tener conflictos de intereses.

DECLARACIÓN DE CONTRIBUCIÓN DE LOS AUTORES: A continuación, se menciona la contribución de cada autor, en correspondencia con su participación, utilizando la Taxonomía Crédito:

- Luis Enrique Colmenares-Guillén: Autor principal, Conceptualización, Investigación, Metodología, Administración de proyectos, Recursos, Supervisión, Validación, Visualización, Redacción - borrador original, Redacción - revisión y edición
- Yosbeli Delgado Gonzalez: Conceptualización, Investigación, Análisis formal, Metodología, Redacción-revisión y edición.

Los autores agradecen el apoyo brindado por la Benemérita Universidad Autónoma de Puebla, y en particular a las estudiantes de la Facultad de Derecho y Ciencias Sociales de la Licenciatura en Criminología Ledezma Sánchez Alexa Fernanda, Rosas Martínez Lucero Ivone, Sánchez Juárez Elizabeth y Zago Galeazzi Marieta, que realizaron la recolección del corpus durante un periodo de seis meses en el Laboratorio de análisis forense digital, de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla.

REFERENCIAS BIBLIOGRÁFICAS

- Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the Blogosphere: Age, gender and the varieties of self-expression. *First Monday*, 12(9). <https://doi.org/10.5210/fm.v12i9.2003>
- Artículo 202. 1931, de 14 de agosto, de *Nuevo Código Publicado en el Diario Oficial de la Federación*, texto vigente. Estados Unidos Mexicanos, México. Últimas reformas publicadas DOF 21-06-2018.
- Bauman, R. (2000). Language, Identity, Performance. *Pragmatics*, 10(1), 1-5. <https://doi.org/10.1075/prag.10.1.01bau>
- Brookes, G. & McEnery, T. (2020). Corpus Linguistics. En S. Adolphs & D. Knight (Eds.). *The Routledge Handbook of English Language and Digital Humanities*. (pp. 378-404). Taylor & Francis.

- Calsamiglia, H., & Tusón, A. (2001). Las cosas del decir. En H. Calsamiglia & A. Tusón (Eds.), *Manual de análisis del discurso* (pp. 27-61). Ariel.
- Centro Virtual Cervantes. (2008). Diccionario de términos clave de Español Lengua Extranjera. *Conversación*.
https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/conversacion.htm
- Chiang, E. (2019). *Rhetorical moves and identity performance in online child abuse interactions* [Tesis de doctorado, Universidad de Aston]. Repositorio de la Universidad de Aston.
https://publications.aston.ac.uk/id/eprint/39062/1/Chiang_E_2018_Redacted.pdf
- 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>
- Cheng, W. (2009). Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. En J. Flowerdew & M. Mahlberg (Eds.), *Lexical Cohesion and Corpus Linguistics [Benjamins Current Topics]* (pp. 65-83). John Benjamins Publishing Company.
- Colmenares-Guillén, L. E., Hernández, J. L., & De Jesús, J. (2020). Análisis del Iter Criminis de la pornografía infantil en Facebook. *Revista Visión Criminológica-Criminalística*, 8(31), 9-15. <https://revista.cleu.edu.mx/new/descargas/2003/Articulo06.pdf>
- Criminal Justice Inspectorates. (2014). Inspecting policing in the public interest, Her Majesty's Inspectorate of Constabulary (HMIC). *An inspection of undercover policing in England and Wales*. <https://bit.ly/3nsACEi>
- Egbert, J., Biber, D., & Gray, B. (2022). *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness*. Cambridge University Press.
<https://www.cambridge.org/core/books/abs/designing-and-evaluating-language-corpora/references/FD9201C173C7BAD1561A29C976E0947E#>
- Esteva, M. (2016). *La prueba en el proceso penal: La relevancia de la Lingüística Forense en la prueba pericial* [Trabajo de pregrado, Universidad de Salamanca]. Repositorio de la Universidad de Salamanca.
https://gredos.usal.es/bitstream/handle/10366/131740/TFG_ESTEVA_BELOSO_Maria.pdf?sequence=1&isAllowed=y
- Ehrlich, S. (2007). Normative Discourses and Representations of Coerced Sex. En J. Cotterill (Ed.). En *The Language of Sexual Crime* (pp. 126-138). Editorial Palgrave Macmillan.
- Goffman, E. (1955). On Face-Work: An Analysis of Ritual Elements in Social Interaction. *Psychiatry*, 18(3), 213-231. <https://doi.org/10.1080/00332747.1955.11023008>
- Grant, T., & Macleod, N. (2012). Whose Tweet? Author analysis of micro-blogs and other short-form messages. En *Proceedings of the International Association of Forensic Linguists' Tenth biennial conference*, July 2011. Centre for Forensic Linguistics. Aston University, Birmingham, UK.
- Grant, T., & Macleod, N. (2016). Assuming Identities Online: Experimental Linguistics Applied to the Policing of Online Paedophile Activity. *Applied Linguistics*, 37(1), 50-70.
<https://doi.org/10.1093/applin/amv079>
- Greenlee, M. (2012). I object, or did it? En *Proceedings of the International Association of Forensic Linguists' Tenth biennial conference*, July 2011. Centre for Forensic Linguistics. Aston University, Birmingham, UK.
- Interpol. (2023). *Delitos contra menores*. <https://www.interpol.int/es/Delitos/Delitos-contra-menores>
- Jaworski, A., & Coupland, N. (2014). *The Discourse reader*. (3rd. ed.). Routledge Reprinted.
- Jenkins, M. & Dando, C. J. (2012). Computer-mediated investigative interviews: a potential screening tool for the detection of insider threat. En *Proceedings of the International Association of Forensic Linguists' Tenth biennial conference*, July 2011. Centre for Forensic Linguistics. Aston University, Birmingham, UK.

- Kaplan, N. (2004). Nuevos desarrollos en el estudio de la evaluación en el lenguaje: La Teoría de la Valoración. *Boletín de Lingüística*, (22), 52-78. <https://www.redalyc.org/articulo.oa?id=34702203>
- May-Chahal, C., Mason, C., Rashid, A., Walkerdine, J., Rayson, P., & Greenwood, P. (2014). Safeguarding cyborg childhoods: Incorporating the on/offline behaviour of children into everyday social work practices. *British journal of social work*, 44(3), 596-614. <https://doi.org/10.1093/bjsw/bcs121>
- MacMartin, C., & Wood, L. A. (2007). Sentencing Sexual Abuse Offenders: Sex Crimes and Social Justice. En J. Cotterill (Ed.), *The Language of Sexual Crime* (pp. 180-197). Editorial Palgrave Macmillan.
- MacLeod, N. & Wright, D. (2020). Forensic linguistics. En S. Adolphs & D. Knight (Eds.). *The Routledge Handbook of English Language and Digital Humanities* (pp. 360-377). Taylor & Francis.
- Ng, E. (2013). Garment, or upper-garment? A matter of interpretation? *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 26(3), 597-613. <https://doi.org/10.1007/s11196-012-9290-9>
- O'Keeffe, A., & Breen, M. (2007). At the Hands of the Brothers: a Corpus-based Lexico grammatical Analysis of Stance in Newspaper Reporting of Child Sexual Abuse Cases. En J. Cotterill (ed.). *The Language of Sexual Crime*, (pp. 217-236). Reino Unido: Editorial Palgrave Macmillan.
- Perverted-Justice. (22 Sep 2008/2019). *Ceasing decoy operations, Deck convicted again*. <http://www.perverted-justice.com/index.php>
- Rashid, A., Baron, A., Rayson, P., May-Chahal, C., Greenwood, P., & Walkerdine, J. (2013). Who am I? Analyzing digital personas in cybercrime investigations. *Computer*, 46(4), 54-61. <https://doi.org/10.1109/mc.2013.68>
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Rayson, P. (2015). Computational tools and methods for corpus compilation and analysis. En D. Biber, & R. Reppen (Eds.). *The Cambridge Handbook of English Corpus Linguistics* (pp. 32-49). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.003>
- Rayson, P., Mariani, J., Anderson-Cooper, B., Baron, A., Gullick, D., Moore, A., & Wattam, S. (2016). Towards Interactive Multidimensional Visualisations for Corpus Linguistics. *Journal for Language Technology and Computational Linguistics*, 31(1), 27-49. <https://doi.org/10.21248/jlcl.31.2016.200>
- Rojas, L. C., & Suárez, M. T. (2008). El lenguaje como instrumento de poder. *Cuadernos de Lingüística Hispánica*, (11), 49-66. <https://www.redalyc.org/articulo.oa?id=322227496005>
- Sacks, H. (2010). *Lectures on conversation*. Wiley-Blackwell Publishing.
- Sacks, H., Schegloff, E., & Jefferson, G. (2015). A simplest systematics for the organization of turn-taking for conversation (A. Korbut, Trad.). *Sotsiologicheskoe Obozrenie / Russian Sociological Review*, 14(1), 142-202. <https://doi.org/10.17323/1728-192x-2015-1-142-202>
- Schegloff, E. A. (2012). *Sequence Organization in Interaction: Volume I: A Primer in Conversation Analysis*. Cambridge University Press.
- Schiffrin, D. (2003). *Approaches to discourse* (2a. ed.). Wiley-Blackwell.
- Sierra, G., Hernández, T., Sánchez, B., & Ramos, S. (2020). *Atribución de autoría: ¡El Móndrigo! Bitácora del Consejo Nacional de Huelga*. <https://interactivo.eluniversal.com.mx/online/pdf-20/PDF-informe-atribucion-autoria-may2020.pdf>
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press. <https://zenodo.org/record/3735822>

- Tannen, D., Hamilton, E. H., & Schiffrin, D. (2015). *The handbook of discourse analysis* (2a. ed.). John Wiley & Sons.
- Tannen, D. (2016). La relatividad de las estrategias lingüísticas: repensando el poder y la solidaridad en el género y en la dominación. *La manzana de la discordia*, 3(2), 91-105. <https://doi.org/10.25100/lamanzanadeladiscordia.v3i2.1466>
- Tomblin, S. (2012). Investigating formulaic language as a marker of Authorship. En Proceedings of The International Association of Forensic Linguistics' Tenth Biennial Conference, July 2011. Centre for Forensic Linguistics. Aston University, Birmingham, UK.
- Turell, T., M. (2010). *Los retos de la lingüística forense en el siglo XXI*. Universidad de Alicante.
- Tusón, A. (2016). Lenguaje, interacción y diferencia sexual. *Enunciación*, 21(1), 138-151. <https://dialnet.unirioja.es/servlet/articulo?codigo=5623346>
- Van Dijk, T. A. (1997). *Discourse studies: Discourse as structure and process v. 1: A multidisciplinary introduction*. SAGE Publications Ltd.

Este preprint fue presentado bajo las siguientes condiciones:

- 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 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 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.