

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

**FACULTAD DE MEDICINA y HOSPITAL UNIVERSITARIO DR. JOSÉ
ELEUTERIO GONZÁLEZ**



“PROCESAMIENTO DE LENGUAJE NATURAL EN PSIQUIATRÍA”

Por

DR. CARLOS ANTONIO BERNAL ROCHA

**Como requisito parcial para obtener el Grado de:
ESPECIALISTA EN PSIQUIATRÍA GENERAL**

MARZO 2024

"PROCESAMIENTO DE LENGUAJE NATURAL EN PSIQUIATRÍA"

Aprobación de la tesis:



DR. ANTONIO COSTILLA ESQUIVEL
Director de la tesis



DR. ALFREDO BERNARDO CUÉLLAR BARBOZA
Coordinador de Enseñanza del Departamento de Psiquiatría



DR. MED. ERASMO SAUCEDO URIBE
Coordinador de Investigación del Departamento de Psiquiatría



DR. STEFAN MAURICIO FERNÁNDEZ ZAMBRANO
Jefe de Servicio o Departamento de Psiquiatría



DR. MED. FELIPE ARTURO MORALES MARTÍNEZ
Subdirector de Estudios de Posgrado

DEDICATORIA Y/O AGRADECIMIENTOS

Dedico este proyecto a mi esposa y a mi familia por su presencia y apoyo constante e incondicional. A mi padrino, que en paz descanse, quién no tuvo la oportunidad de ver la finalización de este proyecto y quién aún después de su partida me sigue enseñando sobre la vida.

Agradezco a mis maestros por sus enseñanzas durante mi formación académica, profesional y personal. Agradezco a mi director de tesis, Dr. Antonio Costilla Esquivel, por su paciencia, su tiempo y el apoyo que me brindó para culminar este paso en mi carrera profesional.

Tabla de contenido

1. RESUMEN	7
1.1 PROPÓSITO DEL ESTUDIO Y MÉTODO DEL ESTUDIO	7
1.2 CONTRIBUCIONES Y CONCLUSIONES.....	7
2. INTRODUCCIÓN	9
2.1 PLANTEAMIENTO DEL PROBLEMA	9
2.2 JUSTIFICACIÓN DEL PROYECTO	10
2.3 ANTECEDENTES.....	11
2.3.1 <i>La psicosis</i>	11
2.3.2 <i>Psicosis en el DSM-5</i>	12
2.3.3 <i>Análisis del lenguaje natural</i>	12
2.3.3.1 <i>Análisis del lenguaje natural en Psiquiatría</i>	13
2.4 ¿CÓMO RECARAR MUESTRAS DE LENGUAJE NATURAL?	16
3. HIPÓTESIS	18
4. OBJETIVO	18
4.1 OBJETIVO GENERAL	18
4.2 OBJETIVOS ESPECÍFICOS.....	18
5. MATERIAL Y MÉTODOS	19
5.1 DISEÑO METODOLÓGICO DEL ESTUDIO.....	19
5.2 CARACTERÍSTICAS DE LA POBLACIÓN DE ESTUDIO	19
5.3 CRITERIOS DE INCLUSIÓN DEL GRUPO CASOS.....	19
5.4 CRITERIOS DE INCLUSIÓN DEL GRUPO CONTROLES	19
5.5 CRITERIOS DE EXCLUSIÓN DE AMBOS GRUPOS	20
5.6 CRITERIOS DE ELIMINACIÓN	20
5.7 MÉTODOS DE RECLUTAMIENTO	20
5.6 PREPARACIÓN DE LOS DATOS PARA MODELACIÓN	21
5.7 MODELACIÓN.....	21
5.7.1 <i>Estandarización</i>	21
5.7.2 <i>Representaciones del texto</i>	21
5.7.3 <i>Entrenamiento de los modelos</i>	22
5.7.4 <i>Métricas de los resultados</i>	22
5.8 ANÁLISIS ESTADÍSTICO.....	22
5.8.1 <i>Tamaño de la muestra y plan de análisis estadístico</i>	22
5.8.1.1 <i>Tamaño de la muestra</i>	22
5.9 CONSIDERACIONES ÉTICAS.....	23
6. RESULTADOS	25
6.1 DESCRIPCIÓN DE LOS PACIENTES. CARACTERÍSTICAS DEMOGRÁFICAS.....	25
6.2 DESCRIPCIÓN DE LOS AUDIOS OBTENIDOS	28
6.2.1 <i>Sumatoria total de los audios</i>	28
6.2.2 <i>Número de preguntas contestadas por grupo</i>	28
6.2.3 <i>Palabras por respuesta y su promedio</i>	30
6.2.4 <i>Tópicos de los textos</i>	32
6.2.5 <i>Modelo de clasificación</i>	33
7. DISCUSIÓN	34

7.1 LIMITACIONES.....	37
7.2 DIRECCIONES FUTURAS.....	37
8. CONCLUSIÓN.....	38
9. ANEXOS.....	39
10. BIBLIOGRAFÍA.....	77
11. RESUMEN AUTOBIOGRÁFICO.....	80

Lista de Abreviaturas

PLN: Procesamiento de lenguaje natural

IA: Inteligencia Artificial

APA: American Psychiatry Association

DSM-V: Diagnostic and Statistical Manual of Mental Disorders

CIMAT: Centro de Investigación en Matemáticas A.C.

Capítulo I

1. RESUMEN

1.1 Propósito del estudio y método del estudio

El estudio tuvo como objetivo explorar la capacidad de un modelo de procesamiento de lenguaje natural para diferenciar, entre personas hablantes del español, las sanas de las que portaban un diagnóstico de problemas de salud mental, en población mexicana, en el Departamento de Psiquiatría del Hospital Universitario “Dr. José Eleuterio González”, con participación del Centro de Investigación en Matemáticas (CIMAT A.C.).

La población del estudio estuvo constituida por dos grupos de adultos entre los 18 a 64 años: 1) Personas en seguimiento en el Departamento de Psiquiatría y con un diagnóstico dentro del espectro de las psicosis; y 2) Personas sanas sin diagnóstico psiquiátrico. Los participantes recibieron una explicación detallada del objetivo del estudio y los procedimientos a realizar, y se les resolvieron las dudas presentadas, obteniéndose después el consentimiento informado por escrito o digital de aceptación para colaborar en el estudio. Los participantes grabaron audios de sus respuestas a al menos cuatro preguntas de un cuestionario de siete preguntas posibles. Estas grabaciones de audio se almacenaron en un repositorio al que sólo los integrantes del equipo médico tenían acceso. Posteriormente, los audios se recuperaron, transcribieron y se prepararon para su modelación mediante un proceso llevado a cabo por el CIMAT.

Se encontró un modelo de procesamiento de lenguaje natural que obtuvo una precisión aceptable (72%) para el reconocimiento de pacientes con diagnóstico dentro del espectro de la psicosis, en comparación con pacientes sanos.

1.2 Contribuciones y conclusiones.

Hasta nuestro conocimiento, este es el primer estudio realizado en población mexicana hispanohablante en el que se encontró que un modelo de procesamiento de lenguaje natural tuviera una precisión aceptable para el reconocimiento de personas que portan un diagnóstico dentro del espectro de la psicosis y aquéllas que son sanas.

Este estudio representa un avance en la investigación sobre el procesamiento del lenguaje natural aplicado a la psiquiatría en la población de habla hispana. Los resultados obtenidos sugieren que los modelos de procesamiento de lenguaje natural pueden ser utilizados como herramientas que ayuden a corroborar una sospecha diagnóstica, y para la identificación de patrones dentro del contenido verbal de las personas que portan un diagnóstico dentro del espectro de las psicosis.

Es importante destacar, sin embargo, que se requieren más investigaciones con poblaciones similares y con mayores tamaños de muestra para validar estos hallazgos, siempre considerando la privacidad y la ética en el manejo de los datos sensibles.

En conclusión, este estudio marca un precedente en la investigación del uso de modelos de procesamiento de lenguaje natural en la psiquiatría en la población mexicana e hispanoparlante, que podrían utilizarse en el futuro para el diagnóstico y seguimiento de trastornos psiquiátricos en esta población.

Capítulo II

2. INTRODUCCIÓN

2.1 Planteamiento del problema

El lenguaje es considerado una fuente de información indispensable para la comunicación de ideas y pensamientos, lo cual lo vuelve un aspecto fundamental de la conducta social humana. Al ser el lenguaje una expresión del pensamiento, y al proveer al ser humano una ventana hacia la mente, este se ha vuelto una fuente de información esencial para las ciencias que estudian la salud mental. El habla y el lenguaje proveen diferentes datos del pensamiento humano, dentro de los cuales se ha podido estudiar la semántica, la coherencia, la sintáctica, la complejidad y el contenido emocional.

Dentro del área de la salud mental, se ha considerado que el lenguaje y el habla son una fuente primaria de datos para la anamnesis, diagnóstico y vigilancia en respuesta al tratamiento de diferentes trastornos mentales, ya que a partir de ellos se brinda información sobre la organización y el contenido del pensamiento. Al ser estos elementos importantes dentro de la evaluación clínica dentro del área de Psiquiatría, los cambios en el lenguaje y en el habla (o alteraciones en la comunicación verbal) han sido objeto de estudio para los diferentes desórdenes existentes. Sin embargo, gran parte del lenguaje ha sido difícil de analizar y cuantificar debido a su amplia gama de aplicaciones y matices; sin embargo, con el advenimiento de las diferentes ciencias computacionales, ha sido posible el estudio del lenguaje producido por los individuos de manera cuantificable y objetiva.

La Inteligencia Artificial ha sido definida como un campo de la ciencia y la ingeniería que se encarga del entendimiento computacional de lo que es comúnmente llamado conducta inteligente y con la creación de artefactos que exhiban tal comportamiento (Shapiro, S. 1992). Dentro del campo de estudio de la inteligencia artificial existen diversas ramas que se han centrado en la aplicación de técnicas computacionales para el análisis y síntesis del lenguaje natural y del habla, a esta pertenece la rama conocida como procesamiento de lenguaje natural (PLN).

La aplicación de la inteligencia artificial en el área de la salud mental, como en la Neurología o Psiquiatría, ha sido principalmente investigada en los trastornos mentales que presentan síntomas psicóticos.

Los trastornos psicóticos son definidos como anormalidades en uno o más de los siguientes cinco dominios: delirios, alucinaciones, pensamiento desorganizado (habla), desorganización grosera o conducta motora anormal (incluida catatonía) y síntomas negativos (DSM-5). Dentro del espectro de los trastornos psicóticos, especialmente en la esquizofrenia, existe una reducción de la complejidad del lenguaje, por lo que de manera clínica se manifiesta en sintomatología como un lenguaje concreto y una disminución en el contenido del mismo; de la misma manera han descrito la disminución de la coherencia semántica, el flujo del significado del lenguaje y una tangencialidad en habla.

La utilización del procesamiento del lenguaje natural para la investigación se ha centrado principalmente en los síntomas psicóticos, ya que en estos la estructura del lenguaje y la coherencia del discurso y la complejidad del mismo se ven alterados, lo cual ha ayudado al estudio del lenguaje producido por un individuo de manera objetiva y cuantificable. Dentro de las investigaciones realizadas, se han estudiado métodos automatizados que utilizan el PLN que operacionalicen el examen del estado mental con respecto al lenguaje y a la conducta.

El uso de procedimientos automatizados de análisis del lenguaje natural ha mostrado su capacidad para diferenciar entre personas sanas y personas que reciben un diagnóstico de algunos trastornos mentales, como los del espectro de las psicosis, y los trastornos del estado de ánimo en otros idiomas. Las características definitorias de los agrupamientos logrados por estos modelos tienen potencial para convertirse en biomarcadores no invasivos de grupos de pacientes con síntomas psiquiátricos.

2.2 Justificación del proyecto

Hasta el momento de la realización del estudio y hasta el conocimiento de los autores, no se disponían de estudios que explorarán el procesamiento de lenguaje natural aplicado a los trastornos psicóticos en población mexicana y con idioma español.

Es importante destacar que, hay una brecha en la investigación para evaluar estas técnicas en hablantes del idioma español en países como México. Esta falta de información inhibe el desarrollo de herramientas y modelos más precisos para la detección y seguimiento de trastornos psiquiátricos en poblaciones de

habla hispana. Por lo tanto, el presente estudio pretende aportar a la literatura para el desarrollo de herramientas precisas para la detección temprana de trastornos psiquiátricos en la población mexicana.

El propósito de este estudio fue la exploración de la capacidad de un modelo de procesamiento de lenguaje natural para diferenciar entre un grupo de personas sanas y aquellas que presentaban un diagnóstico de problema de salud mental dentro del espectro de la psicosis, para lo cual fueron estudiados dos grupos con las características antes mencionadas. También fueron determinadas la sensibilidad y especificidad del modelo utilizado para distinguir el lenguaje entre los grupos.

2.3 Antecedentes

2.3.1 La psicosis

La Asociación Psiquiátrica Americana (American Psychiatric Association, APA) define a la psicosis como un estado caracterizado por la pérdida de contacto con la realidad, debido a una disrupción en la forma en que el cerebro procesa información. Cuando alguien experimenta un episodio psicótico, los pensamientos de la persona y sus percepciones se encuentran alterados; por esta razón los individuos pueden tener dificultad para entender qué es real y qué no lo es (Torres 2020; Schrimpf, Aggarwal, and Lauriello 2018; Bürgy 2008).

El Manual Diagnóstico y Estadístico de la Asociación Psiquiátrica Americana (Diagnostic and Statistical Manual of Mental Disorders DSM-5), conceptualiza la psicosis como anormalidades en uno o más de cinco dominios sintomáticos: delirios, alucinaciones, pensamiento desorganizado (trastornos formales del pensamiento), conducta motora anormal o desorganizada (incluyendo catatonia) y síntomas negativos (abulia, alogia, apatía, embotamiento afectivo)(Bürgy 2008b; American Psychiatric Association (APA), 2014).

Los síntomas psicóticos se manifiestan en los trastornos del espectro esquizofrénico definidos por el DSM-5: Esquizofrenia, Trastorno esquizofreniforme, Trastorno esquizotípico, Trastorno delirante, Trastorno psicótico breve, y Trastorno esquizotípico de la personalidad, además de los Trastornos psicóticos inducidos por sustancias o debidos a enfermedades médicas no psiquiátricas. Por otra parte, no existen rasgos psicóticos patognomónicos para cada diagnóstico, y muchos otros trastornos psiquiátricos

como los Trastornos neurocognitivos, los Trastornos del desarrollo, el Trastorno bipolar I y el Trastorno de depresión mayor pueden acompañarse de síntomas psicóticos.

2.3.2 Psicosis en el DSM-5

En el DSM-5, que utiliza el numeral arábigo para incorporar más fácilmente revisiones menores mediante un punto y numerales crecientes, se conceptualizó a los trastornos psicóticos en una forma diferente, en parte como producto de la modelación con técnicas estadísticas no precisamente novedosas, pero que no habían sido aplicadas previamente, como Análisis de Factores y Teoría de respuesta a ítems. Así, se generó un grupo llamado Espectro de la esquizofrenia y otros trastornos psicóticos, que incluye al Trastorno esquizotípico de la personalidad, el Trastorno delirante, el Trastorno psicótico breve, el Trastorno esquizofreniforme, la Esquizofrenia, el Trastorno esquizoafectivo, el Trastorno psicótico inducido por sustancias o medicamentos, el Trastorno psicótico debido a otra afección médica, y criterios para el especificador de Catatonía y las categorías residuales. Otros trastornos psiquiátricos que pueden acompañarse de síntomas psicóticos, como el Trastorno bipolar I, el Trastorno de depresión mayor, el Trastorno neurocognitivo mayor y leve, y algunos trastornos de la personalidad diferentes al esquizotípico, entre otros, pero la sintomatología psicótica es considerada como fuera del grupo de síntomas nucleares definitorios del trastorno.

2.3.3 Análisis del lenguaje natural

El procesamiento de lenguaje natural es una técnica que surge a partir de la década de 1950 de la intersección de las ciencias computacionales y lingüísticas, y que se refiere a la capacidad de las computadoras para analizar y comprender de manera autónoma el lenguaje humano, tanto hablado como escrito. Esta técnica involucra diferentes algoritmos y modelos computacionales que se utilizan para procesar y transformar el lenguaje en una forma interpretable por las máquinas (Canuma Prince 2019; Grosz 1982). El procesamiento de lenguaje natural funciona a través de diferentes niveles de análisis lingüístico, que va

desde la identificación de palabras y frases que conforman el lenguaje hasta la interpretación semántica y pragmática del mismo.

Es importante resaltar que el procesamiento de lenguaje natural es una técnica en continua evolución y que presenta importantes desafíos en cuanto a la complejidad del lenguaje y a la variación de los patrones lingüísticos. Sin embargo, su potencial para mejorar la eficiencia de múltiples procesos y ofrecer una gama de soluciones innovadoras en diferentes áreas es indiscutible.

2.3.3.1 Análisis del lenguaje natural en Psiquiatría

Aplicado a la psiquiatría, el análisis del lenguaje natural consiste en el uso de modelos y algoritmos que analizan el contenido verbal de un individuo con el objetivo de detectar patrones de pensamiento y comportamiento relacionados con trastornos psiquiátricos.

En 1979, Andreasen realizó estudios que arrojaron información acerca del discurso de los pacientes que sufrían esquizofrenia y cómo en éste se reducía la complejidad sintáctica, y a su vez se evidenció que típicamente se operacionalizaban oraciones cortas y esto era más evidente en las narrativas abiertas y no utilizando preguntas cerradas, lo cual le llevó a investigar la relación entre la longitud de las oraciones y la predicción de psicosis. La propuesta fue entonces de utilizar entrevistas abiertas para que de esta manera se pudiera obtener el discurso natural, con ello realizar un análisis y finalmente ser capaz de medir la coherencia semántica de longitud de la oración y así medir mejor la captura de la complejidad sintáctica (Cheryl M. Corcoran et al. 2018).

A finales de la década de 1980 e inicios de 1990, existió una revolución en esta ciencia con la introducción y la unión de los métodos en inteligencia artificial que consistió en la creación de algoritmos para el procesamiento del lenguaje natural por medio del *machine learning* (Canuma Prince 2019; Grosz 1982). Durante esta época, Hand en 1985, Servan-Schreiber en 1986 y Hoffman en 1987, comenzaron a preguntarse sobre los potenciales beneficios en la práctica clínica de la tecnología basada en inteligencia artificial y en el campo de la Psiquiatría, ya que para este tiempo no estaba bien definido su papel dentro del área clínica. A partir de 1992, Garfield, et al. realizaron una revisión de las diferentes escuelas que utilizaban la inteligencia artificial, en la que destacaron la importancia de su

uso para la evaluación de la psicopatología en trastornos psiquiátricos. Concluyeron que estas técnicas pueden utilizarse como herramienta dentro del campo de la entrevista clínica, y que además sirven como herramienta para la investigación en el comportamiento verbal de la patología (Garfield, D., C. Rapp, 1992).

La esquizofrenia se considera la patología prototipo para poder describir una estructura anormal en el lenguaje, y además se ha descrito la posibilidad de que cambios en el lenguaje sutiles pudieran encontrarse en estados prepsicóticos.

Para un mejor entendimiento de la estructura de las perturbaciones en el lenguaje en estados psicóticos, se realizó una revisión para describir los diferentes métodos utilizados con un enfoque hacia el análisis del procesamiento del lenguaje natural en discursos transitorios. Se ha descrito que, de forma común, la discapacidad del lenguaje en la esquizofrenia se produce a nivel de la semántica, en la coherencia y en la cohesión, los cuales mantienen un flujo de significados y tienen consistencia de referencias a través de cláusulas y oraciones (Cheryl Mary Corcoran and Cecchi 2020). Se ha demostrado en revisiones de la literatura que los individuos con psicosis tienden a realizar comúnmente pausas y tener variaciones de lentitud en el ritmo del lenguaje, además de una reducción en la intensidad en el habla comparando con sujetos sanos, lo cual mostró que estos elementos se pueden utilizar para la producción de información (Cheryl M. Corcoran et al. 2020).

Estas características antes mencionadas son las características acústicas del lenguaje, las cuales mediante la identificación temprana permiten realizar intervenciones de tratamiento oportuna, así como brindar una pista acerca de la etiología de la psicosis.

El desorden formal del pensamiento se definió como un síntoma común de la psicosis y se propuso que pudiera ser considerado como un marcador de severidad de la enfermedad psicótica. Se propuso que una evaluación del desorden formal del pensamiento pudiera ayudar a clarificar los diagnósticos y pudiera ayudar a predecir los pronósticos (Roche et al. 2015).

En el 2018, se realizó un metaanálisis y una revisión sistemática, donde se evaluaron 22 artículos relevantes al campo del procesamiento de lenguaje natural, de los cuales 18 de ellos mostraron como resultado que las características de la técnica de modelos de espacio semántico funcionaban

como herramienta diagnóstica y pronóstica en cuanto a trastornos psiquiátricos como neurológicos (de Boer et al. 2018).

El valor potencial del procesamiento de lenguaje natural aunado a la técnica de inteligencia artificial de Machine Learning se ha propuesto para ser considerados como biomarcadores lingüísticos significativos. Existen además descubrimientos clave en las alteraciones de la producción del lenguaje en un episodio de psicosis, donde los métodos de medición automática del discurso arrojan como resultados anomalías en la coherencia, la sintáctica, la complejidad, pobreza en el contenido, coherencia referencial y lenguaje metafórico (Cheryl M. Corcoran et al. 2020).

En Mayo del 2021, Tang, et al. exploraron métodos que caracterizaban los cambios en el lenguaje de los trastornos del espectro de la esquizofrenia, y comparaban estos resultados con controles sanos utilizando algoritmos y representaciones bidireccionales de encoder de Transformers (BERT), donde a través de la partición del lenguaje se encontró que los trastornos del espectro de la esquizofrenia utilizaban más pronombres singulares y en primera persona y menos adverbios, adjetivos y determinadores, así mismo se encontró una tangencialidad y una distancia en el ensamblaje de oraciones. Los resultados de esta investigación demostraban que los modelos utilizados tenían una gran capacidad para discriminar a individuos sanos de aquellos que padecían un trastorno del espectro de la esquizofrenia (Tang et al. 2021).

El crecimiento de técnicas de procesamiento de lenguaje natural en la medicina se ha utilizado para ayudar al clínico a realizar diagnósticos, tamizajes clínicos, detección de interacciones medicamentosas y predicción de efectos adversos. Específicamente en el campo de la Psiquiatría, existe limitada información acerca de su uso y aplicaciones en la amplia gama de trastornos existentes en la rama psiquiátrica, siendo el mayor énfasis en los trastornos psicóticos (Cheryl Mary Corcoran and Cecchi 2020). Además, se ha propuesto el aplicar estos métodos de procesamiento de lenguaje natural en diferentes poblaciones de pacientes para una mejor prueba de reproducibilidad, además de aplicarlo en otros idiomas diferentes al inglés, e identificar la fuente de variabilidad (Cheryl M. Corcoran et al. 2018) .

Por lo anterior, el procesamiento del lenguaje natural ha demostrado ser una técnica prometedora para el apoyo diagnóstico y el seguimiento de los trastornos

psiquiátricos. Además, el análisis puede ofrecer beneficios de importancia ya que pueden recolectar y analizar grandes cantidades de datos de manera rápida y económica, lo que puede traducirse en un diagnóstico más preciso y un tratamiento dirigido.

2.4 ¿Cómo recabar muestras de lenguaje natural?

A partir de las características observadas en el discurso de una persona, se han propuesto a lo largo del tiempo diferentes tareas o paradigmas para obtener un discurso representativo por parte de pacientes, para su evaluación y análisis posterior.

La primera aplicación directa de técnicas de procesamiento del lenguaje natural a un discurso fue realizada por Elvegag en el 2007, a pacientes con diagnóstico de Esquizofrenia, en donde la tarea a realizar consistía en aplicar una entrevista semi-estructurada de 45 minutos de duración, en la cual se agregaban preguntas abiertas conocidas como "open-ended questions", que incluían preguntas sobre su sintomatología, eventos recientes y opiniones personales. Dentro de estas últimas existían preguntas con temas estándar y preguntas con temas abstractos, como: "¿Por qué algunas personas creen en Dios? o ¿Qué significa el libre albedrío? o ¿Qué tendría que hacer una persona para fumar un cigarrillo?". De manera interesante, la producción oral de los sujetos fue suficiente para encontrar alteraciones en el discurso (Elvevåg et al. 2007.). Otra de las tareas utilizadas como fuente para obtención de lenguaje fue a través de la elaboración de diferentes tipos de discursos, fuese narrativo (contando una historia) o expuesto (describiendo conceptos abstractos) , así como la utilización de preguntas de prueba y relatos como "¿Qué es lo que habían hecho el día de hoy? o ¿Qué es lo que te gusta hacer?" lo cual producía respuestas heterogéneas y eran sujetas a análisis (Elvevåg et al. 2010).

Para el 2015, Bedi G. utilizó entrevistas abiertas de aproximadamente una hora, así como descripciones personales de cambios observados durante el tiempo, y el impacto de los mismos, así como una opinión personal sobre si era útil o no el cambio, tras la obtención del lenguaje. Este quedo sujeto a técnicas de procesamiento de lenguaje natural donde, a través de la medición de análisis

latente, se obtuvieron resultados favorables para la predicción de psicosis en personas con alto riesgo de psicosis (Bedi et al. 2015).

Posteriormente, diversos estudios han utilizado tareas como dar respuesta a entrevistas estructuradas o semiestructuradas (Cheryl M. Corcoran et al. 2018; Tang et al. 2021; Elvevåg et al. 2007; Bedi et al. 2015; Rezaii, Walker, and Wolff 2019), realizar una descripción de uno mismo, contar relatos positivos o neutrales en su vida (Tang et al. 2021), contar historias y pedirle a un paciente que hiciera lo mismo utilizando material que recordara relacionado a la historia, así como dar respuesta a preguntas relacionadas con las historias (Corcoran et al. 2018; Elvevåg et al. 2010; Rosenstein et al. 2014).

Capítulo III

3. HIPÓTESIS

La modelación del lenguaje natural de las personas hablantes de español sanas y aquellas con diagnóstico de problemas de salud mental permite diferenciar entre los grupos.

Capítulo IV

4. OBJETIVO

4.1 Objetivo General

El objetivo del presente trabajo de investigación es explorar la capacidad de modelos de lenguaje natural para reconocer personas con un diagnóstico de psicosis.

4.2 Objetivos específicos

Establecer cuáles son las variables cuantitativas en el lenguaje que nos permitan realizar una distinción del lenguaje psicótico del que no lo es.

Capítulo V

5. MATERIAL Y MÉTODOS

5.1 Diseño metodológico del estudio

Se trató de un estudio de casos y controles de tipo transversal, comparativo, prospectivo, no ciego. Se realizó el reclutamiento de los sujetos del mes de febrero del 2022 a agosto del 2023.

5.2 Características de la población de estudio

- Se evaluaron dos tipos de poblaciones en adultez temprana y adultez media (entre 18 y 65 años), hablantes del idioma español.
 - o Una población quienes llevaran su atención por un diagnóstico dentro del espectro de la psicosis.
 - o Una población sin evidencia de problemas de salud mental, ni diagnósticos psiquiátricos previos.

5.3 Criterios de inclusión del grupo casos

- Adultos entre los 18 y los 65 años.
- Que solicitaran atención por síntomas psicóticos o con un diagnóstico dentro del espectro de la psicosis en la Consulta Externa o que hayan participado en un protocolo previo sobre psicosis en el Departamento de Psiquiatría del Hospital Universitario “Dr. José Eleuterio González”.
- Que acepten participar en el protocolo de investigación.
- Que hablen el idioma español.

5.4 Criterios de inclusión del grupo controles

- Adultos entre los 18 y los 65 años.
- Sujetos sanos sin diagnósticos previos de patología psiquiátrica.
- Que acepten participar en el protocolo de investigación.
- Que hablen el idioma español.

5.5 Criterios de exclusión de ambos grupos

- Menores de 18 años o mayores de 65 años.
- Rehusar participar en el protocolo de investigación.
- Pacientes no hablantes del idioma español.
- Con diagnóstico previo de discapacidad intelectual, trastorno de lenguaje, trastorno fonológico, afasia, mutismo o cualquier problema que impida la expresión verbal.

5.6 Criterios de eliminación

- Sujetos que soliciten su retiro del protocolo de investigación.
- Sujetos cuyos archivos de audio sean ininteligibles.

5.7 Métodos de reclutamiento

Se realizó el reclutamiento de los sujetos del mes de febrero del 2022 a agosto del 2023.

- En el reclutamiento del grupo de casos se recopilaron los datos de pacientes que tuvieran diagnóstico dentro del espectro de la psicosis. Se les localizó vía telefónica para invitarlos a participar en el presente estudio, acordando una cita para realizar la firma del consentimiento informado por escrito. Posteriormente se realizó el llenado de la encuesta en línea correspondiente al grupo de casos. A continuación se recabó una grabación de audio a sus respuestas de al menos a cuatro preguntas abiertas de siete posibles. La grabación se realizó en un dispositivo electrónico con acceso a Internet y un micrófono.
- En el reclutamiento del grupo de controles se hizo una invitación al estudio en las redes sociales del departamento e investigadores, donde se invitó a personas que no tenían ningún diagnóstico psiquiátrico a participar. A continuación, aquellos que aceptaron la participación realizaron el llenado de la encuesta en línea correspondiente al grupo de controles. Posteriormente se recabó una grabación de audio a sus respuestas de al menos a cuatro

preguntas abiertas de siete posibles. La grabación se realizó en un dispositivo electrónico con acceso a Internet y un micrófono.

- Las preguntas posibles a responder fueron:
 - o ¿Cuáles fueron sus actividades el día de ayer?
 - o ¿Qué es para usted un buen día?
 - o ¿A qué se debe que algunas personas creen en Dios y otras no?
 - o ¿Cuál es la actividad que más le gusta hacer, y porqué es su favorita?
 - o ¿Podría describir cómo se lleva con los miembros de su familia?
 - o ¿Podría describir uno o más recuerdos bonitos de su infancia?
 - o ¿Cómo es la colonia o barrio en el que vive?

5.6 Preparación de los datos para modelación

Independientemente de la cohorte, los archivos con la grabación fueron recuperados y transcritos automáticamente con la aplicación Whisper de Open AI, que fue instalada en un servidor desconectado de internet para la transcripción, empleando el nivel de calidad denominado “Large” (disponible en <https://openai.com/index/whisper>).

5.7 Modelación

La modelación de los datos contenidos en la base anonimizada fue realizada por el Centro de Investigación en Matemáticas, A.C., el cuál es un centro público de investigación de la red CONACYT. En la modelación, se realizó la caracterización de la producción oral de los sujetos del estudio. Las fases de la modelación abarcan las siguientes:

5.7.1 Estandarización

En este paso se verificó la integridad de los archivos, y se realizó la normalización del texto donde se incluyó la transformación a letras minúsculas, la eliminación o transformación de caracteres especiales y palabras funciones.

5.7.2 Representaciones del texto

Se obtuvieron las representaciones vectoriales de texto adecuadas para las tareas de aprendizaje, tanto supervisado como no supervisado. Las representaciones que se exploraron fueron aquellas basadas en características léxicas, semánticas y en embeddings contextuales usando diferentes propuestas como word2vec, fastText, ELMO, VERT, entre otros los cuáles están pre-entrenados en corpus en español.

5.7.3 Entrenamiento de los modelos

Se implementaron diferentes metodologías de inteligencia artificial para obtener modelos de predicción adecuados, partiendo de las representaciones de texto obtenidas. Estos métodos están basados en modelos profundos del tipo encoder-decoder con mecanismos de atención, y usando arquitecturas de redes neuronales adecuadas para datos secuenciales, tales como LSTM, GRU, capas convolucionales o Transformers, así como modelos early y late fusion para incorporar características léxicas o definidas por expertos.

5.7.4 Métricas de los resultados

Para cada uno de los modelos se reportó la tabla de contingencia o matriz de confusión y las siguientes métricas derivadas.

- 1) Precisión o valor predictivo positivo (precisión)
- 2) Exhaustividad o tasa de verdaderos positivos (recall)
- 3) Exactitud (accuracy)
- 4) F1 ponderado (Weighted F1)
- 5) Matthews Correlation Coefficient (MCC)
- 6) MCC multiclase

5.8 Análisis estadístico

5.8.1 Tamaño de la muestra y plan de análisis estadístico

5.8.1.1 Tamaño de la muestra

Este estudio ha sido de prueba de concepto, el cuál consistió en una modelación a través de las técnicas que han sido descritos en el apartado de Materiales y

Métodos. No existe una prueba de hipótesis; por lo tanto, no hay, hasta nuestro conocimiento, fórmulas para calcular el tamaño de muestra o la potencia estadística, y la pregunta de investigación carece por el momento de un tamaño del efecto.

Por otro lado, la literatura nos muestra que en estudios realizados en hablantes de Inglés, se han encontrado modelos con muestras de treinta pacientes (Corcoran et al. 2018; Tang et al. 2021; Bedi et al. 2015; Haas et al. 2020; Sommer et al. 2004) y precisiones aceptables.

De acuerdo al estudio de Bujang y colaboradores, asumiendo que la prevalencia de algún trastorno dentro del espectro de la esquizofrenia entre los sujetos del grupo de casos sea de 40%, se necesitaron treinta y un sujetos con el diagnóstico y setenta y ocho como tamaño de muestra total para diferenciar entre una sensibilidad del 0.70 (nula) y una de 0.90 (alterna) con una potencia estadística mayor o igual a 0.80, y un valor de "p" para declarar significancia de 0.05 o menor (Bujang and Adnan 2016).

En base a lo mencionado previamente, se propuso como tamaño de muestra mínimo de 30 sujetos por cohorte.

5.9 Consideraciones éticas

El presente estudio implicó como procedimiento la realización de grabaciones de audio, las cuales se almacenaron en una base de datos de identificación.

La base de datos de identificación consistió exclusivamente en el número de folio dentro del estudio, y sólo fue manejada por el equipo médico de la investigación. Esta base de datos incluyó el folio, género, edad, medicamentos al momento de la entrevista, diagnóstico, medicamentos prescritos, y el archivo de texto anonimizados. Sólo esta última fue compartida con el equipo de modelación del Centro de Investigación en Matemáticas A.C. (CIMAT).

La base de datos de audio permaneció separada y fue sometida a los procedimientos de modelación, que incluyeron estandarización, representación del texto, entrenamiento de modelos, y evaluación de métricas de resultados.

El equipo de modelación de CIMAT otorgó su acuerdo de participación, en el cual se incluye la obligación de asesorar al equipo médico en la anonimización automatizada de los textos, tratar los textos anonimizados como información

confidencial, y no compartir los archivos de texto con ninguna institución o investigador.

No se incluyó material biológico no infeccioso o terapia génica/material genético. El estudio fue financiado por recursos económicos propios del departamento o servicio, sin generar un costo para el participante. Se consideró un estudio de riesgo mínimo de acuerdo con el Artículo 17 del Reglamento de la Ley General de Salud en Materia de Investigación para la Salud. Los participantes tuvieron la posibilidad de beneficiarse directamente por su participación en el estudio al recibir psicoeducación sobre su patología, así como la obtención de información sobre los diversos protocolos que se realizaban dentro del Departamento de Psiquiatría.

Las medidas adicionales de protección para los sujetos vulnerables incluyeron la recolección de consentimiento informado por parte del sujeto que participó en la investigación. No se requirió una vigilancia posterior al término de la participación en el estudio. En caso de rechazar la participación, no se generó ninguna consecuencia que afectada o repercutiera en el tratamiento o seguimiento correspondiente al padecimiento de base.

Capítulo VI

6. RESULTADOS

6.1 Descripción de los pacientes. Características demográficas.

La invitación se realizó a 221 sujetos quienes conformaban el grupo de casos y el grupo de controles. Dentro del grupo de casos correspondiente a los sujetos que solicitaron atención por síntomas psicóticos, aquellos con un diagnóstico dentro del espectro de la psicosis, o quienes hayan participado en un protocolo previo sobre psicosis, se le realizó la invitación a 104 pacientes, de los cuáles 49 rechazaron participar en el estudio sin especificar el motivo, 55 aceptaron participar, de estos, 32 acudieron para realizar las grabaciones correspondientes, los 23 restantes se desconoce la causa del rechazo a la grabación. Respecto al grupo de los controles correspondientes a sujetos sin diagnóstico psiquiátrico previo la participación se recolectó una muestra de 117 sujetos, de los cuales 44 fueron excluidos por presentar un diagnóstico o tratamiento psiquiátrico previo, 2 fueron excluidos por rechazar la invitación a participar y 71 de ellos fueron incluidos al estudio y realizaron las grabaciones correspondientes.

La muestra final fue de 103 participantes (N= 103) (Grupo control: 32 participantes; Grupo control: 71 participantes).

Correspondiente al grupo de casos: el 13% (N= 13) de los participantes fueron del género femenino y 59% (N= 19) fueron del género masculino (Ver gráfico 1), la edad varió entre los 20 y 62 años, media= 35.59. No se recabó información demográfica para este grupo de población.

Correspondiente al grupo de controles: el 55.6% (N= 40) de los participantes fueron del género femenino y 44.4% (N= 31) fueron del género masculino (Ver gráfico 2), la edad varió entre los 18 y 64 años, media= 31.15.

En cuanto a la población estudiada del grupo de control se recabó información sobre si la persona era de nacionalidad mexicana, a lo cual solo el 2.8% (N= 2) respondió "no" y el 97.2% (N= 69) respondió "sí". Quienes no tenían una nacionalidad mexicana, respondieron que su nacionalidad era colombiana. Dentro de quienes tenían una nacionalidad mexicana, se recabó la información de cuál era el estado de la República Mexicana donde había vivido más tiempo la persona, obteniendo lo siguiente: la mayor parte de la población estudiada era

de Nuevo León correspondiente al 56.5% (N= 39), seguido de Veracruz con un 7.2% (N= 5), Coahuila con 4.3% (N= 3), Tamaulipas con 4.3% (N= 3), Estado de México con 4.3% (N= 3), , Chihuahua con 2.9% (N= 2), Michoacán con 2.9% (N= 2), Sinaloa con 1.4% (N= 1), Ciudad de México con 1.4% (N= 1), Jalisco con 1.4% (N= 1), Guanajuato con 1.4% (N= 1), Zacatecas con 1.4% (N= 1), Quintana Roo con 1.4% (N= 1), y Chiapas con 1.4% (N= 1). Un porcentaje de 7.2% (N= 5) de los participantes no contestaron sobre cuál estado de la República ha sido donde han vivido por más tiempo (Ver gráficos 3 y 4). Se estudió también el máximo grado de estudios de las personas que participaron en el estudio y realizaron las grabaciones correspondientes obteniendo lo siguiente: el 59.2% (N= 42) de los participantes tenían carrera universitaria, el 22.5% (N= 16) tenía como máximo nivel de estudios un posgrado, el 14.1% (N= 10) estudió hasta la preparatoria, el 2.8% (N=2) estudió hasta la secundaria, y el 1.4% (N= 1) estudió hasta la primaria.

Gráfico 1.- Recuento de género del grupo casos

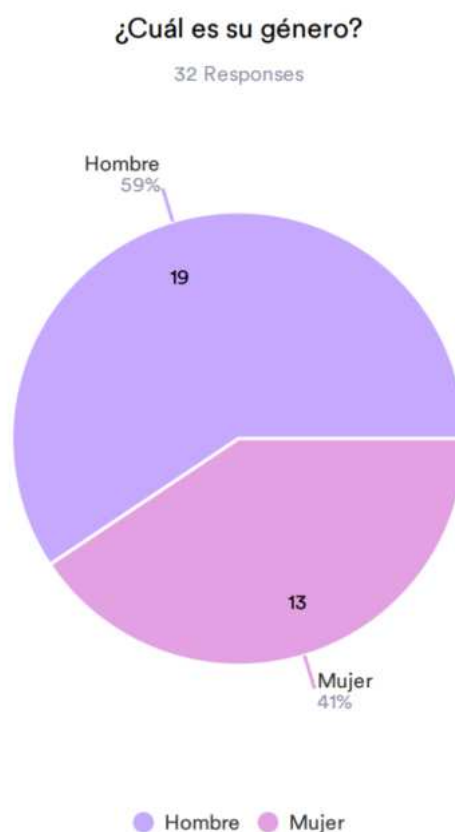


Gráfico 2.- Recuento sobre el género del grupo control

Recuento de ¿Cuál es su género?

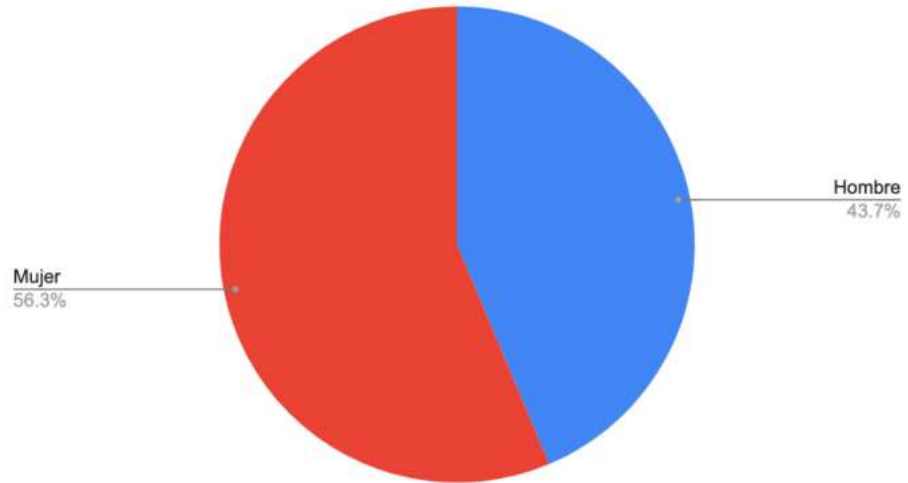


Gráfico 3.- Recuento sobre la nacionalidad del grupo control

Recuento de ¿Es usted mexicana o mexicano?

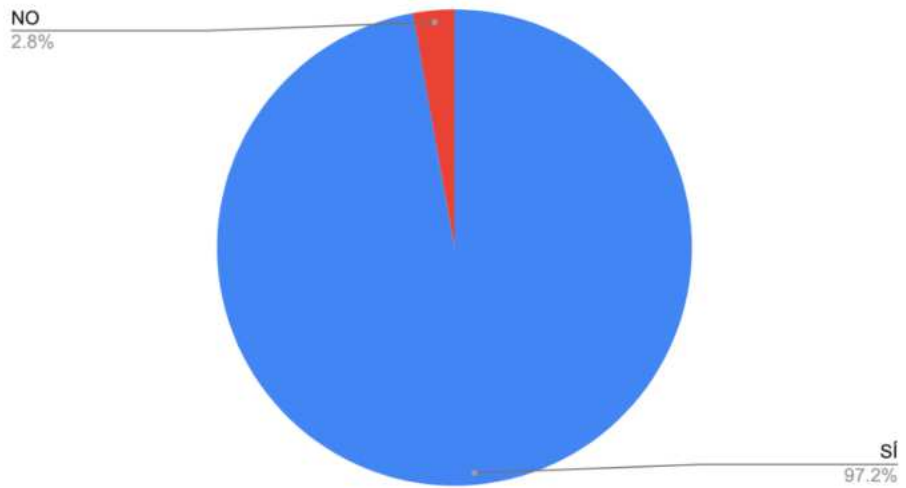
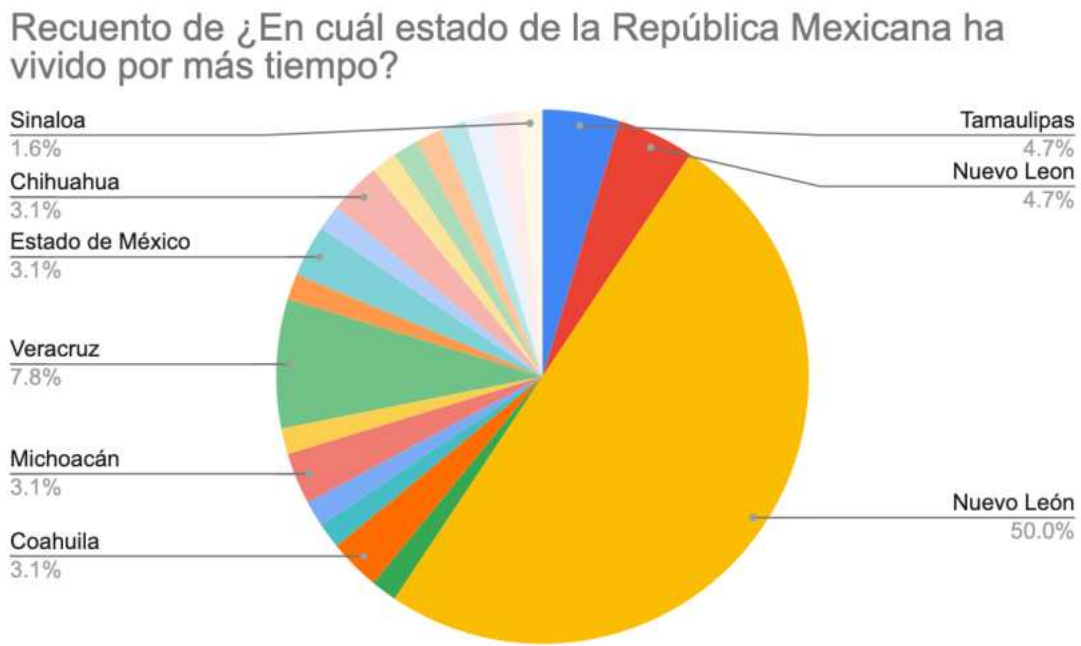


Gráfico 4.- Recuento sobre el estado de la República en el que los sujetos del grupo control han vivido por más tiempo



6.2 Descripción de los audios obtenidos

6.2.1 Sumatoria total de los audios

La sumatoria total de los audios recabados para el grupo de casos y el grupo control fueron las siguientes:

- Grupo de casos: 201 minutos y 26 segundos.
- Grupo de control: 436 minutos y 47 segundos

6.2.2 Número de preguntas contestadas por grupo.

Se realizó una medición de cuantos participantes habían contestado las diferentes preguntas planteadas dentro del grupo de casos y el grupo de control, donde se obtuvieron los siguientes resultados (Ver Tabla 1)

- Pregunta 1 “¿Cuáles fueron sus actividades el día de ayer?”.
 - Respecto al grupo de casos, un total de 30 participantes decidió contestar la pregunta. A comparación del grupo control, un total de 73 participantes decidió contestar esta pregunta.
- Pregunta 2 “¿Qué es para usted un buen día?”

- En el grupo de casos, 27 personas respondieron a la pregunta, mientras que en el grupo de control, 70 personas lo hicieron.
- Pregunta 3 “¿A qué se debe que algunas personas crean en Dios y otras no?”
 - Un total de 26 participantes del grupo de casos decidió contestar la pregunta, en comparación con los 60 del grupo control que también respondieron.
- Pregunta 4 “¿Cuál es la actividad que más le gusta hacer, y por qué es su favorita?”
 - Para el grupo de casos, 31 personas eligieron responder la pregunta, comparado con 65 personas del grupo control que hicieron lo mismo.
- Pregunta 5 “¿Podría describir cómo se lleva con los miembros de su familia?”
 - En el grupo de casos, 27 individuos contestaron la pregunta, mientras que en el grupo control fueron 56 los que la respondieron.
- Pregunta 6 “¿Podría describir uno o más recuerdos bonitos de su infancia?”
 - El grupo de casos tuvo 27 respuestas a la pregunta, en comparación con las 54 respuestas del grupo control.
- Pregunta 7 “ ¿Cómo es la colonia o barrio en el que vive?”
 - Dentro del grupo de casos, 29 participantes optaron por responder la pregunta, en contraste con los 54 participantes del grupo control que también lo hicieron.

Tabla 1: Número de participantes que contestaron las preguntas.

Pregunta	Grupo de casos	Grupo de control
1	30	73
2	27	70
3	26	60
4	31	65
5	27	56
6	27	54
7	29	49

6.2.3 Palabras por respuesta y su promedio

Al realizar la modelación de los datos obtenidos, se recopiló un total de 73,675 palabras, de las cuales 22,020 correspondieron al grupo de casos y 51,655 correspondieron al grupo de control. Se obtuvo un vocabulario de 8,925 palabras totales, de las cuales 4,088 correspondieron al grupo de casos y 6,838 correspondieron al grupo de control. Obteniendo una riqueza léxica total del 12.1%, un 18.5% y un 13.2% de riqueza léxica para el grupo de casos y el grupo de control respectivamente (Ver Tabla 2).

Tabla 2: Número de palabras totales, vocabulario y riqueza léxica.

		Grupo de casos	Grupo de control
Palabras totales	73675	22020	51655
Vocabulario	8925	4088	6838
Riqueza léxica	0.121	0.1856494096	0.132378279

A continuación se muestran los resultados en cuanto al análisis de palabras promedio y su desviación dependiendo de la pregunta contestada para ambos grupos (Ver tablas 3 y 4).

Tabla 3: Promedio de palabras por pregunta del grupo de casos.

Pregunta	Promedio	Desviación	Max	Min
1	166.6	230.9313895	1176	3
2	92.37037037	102.4720913	460	3
3	131.8461538	126.6335392	636	3
4	68.4516129	55.10391281	209	3
5	123.7407407	123.1017624	424	3
6	103.5555556	110.6942179	433	3
7	104.7586207	98.564404	402	3
Todas	112.7766497	135.0335204	1117	3

Tabla 4: Promedio de palabras por pregunta grupo de control.

Pregunta	Promedio	Desviación	Max	Min
1	105.8767123	139.187559	1074	3
2	97.77142857	92.26269505	547	3
3	147.2333333	127.6211015	823	3
4	98.38461538	80.69461755	481	3
5	147.5714286	148.0438952	692	3
6	149.1851852	130.3363276	585	3
7	121.6326531	108.0309279	553	3
Todas	121.971897	121.8340683	1074	3

Realizando una comparación del promedio de palabras empleadas por pregunta en los distintos grupos, podemos observar lo siguiente:

- Pregunta 1: En el grupo de casos se obtuvo un promedio de 166 palabras en comparación con el grupo de control donde se obtuvo un promedio de 105 palabras.
- Pregunta 2: En el grupo de casos, el promedio de palabras fue de 92, en contraste con el grupo control, donde el promedio fue de 97 palabras.
- Pregunta 3: El promedio de palabras en el grupo de casos fue de 131, comparado con el promedio de 147 palabras en el grupo control.
- Pregunta 4: Se obtuvo un promedio de 68 palabras en el grupo de casos, en comparación con el promedio de 98 palabras en el grupo control.
- Pregunta 5: El grupo de casos registró un promedio de 123 palabras, frente a las 147 palabras promedio del grupo control.

- Pregunta 6: En el grupo de casos se observó un promedio de 103 palabras, comparado con las 149 palabras promedio del grupo control.
- Pregunta 7: El grupo de casos tuvo un promedio de 104 palabras, a diferencia del grupo control que tuvo un promedio de 121 palabras.

6.2.4 Tópicos de los textos

Dentro del análisis de las respuestas y las palabras empleadas por los participantes del grupo de casos y control, se encontraron diferentes repeticiones de palabras que permitieron el análisis de la pobreza del lenguaje entre cada cohorte. (Ver Tabla 5 y 6).

En las tablas 5 y 6 podemos ver que existe una diferencia en la cantidad de veces que una palabra es repetida en las diversas preguntas que se emplearon, donde se puede concluir que existe una menor cantidad de palabras repetidas en el grupo de casos a comparación del grupo control.

Tabla 5: Tópicos utilizados en la cohorte 2

Tópico 1
bien', 'día', 'ahí', 'casa', 'bueno',
Tópico 2
dios', 'creen', 'creo', 'personas', 'gente', 'cosas',

Tabla 6: Tópicos utilizados en la cohorte 3

Tópico 1
día', 'buen', 'hacer', 'salir', 'trabajo', 'tiempo', 'poder', 'casa', 'comer'
después', 'ayer', 'casa', 'día', 'dd', 'puse', 'hice', 'tarde', 'trabajo', 'mañana'
gusta', 'actividad', 'hacer', 'ver', 'disfruto', 'favorita', 'cosas', 'creo', 'música', 'poder'
Tópico 2
bien', 'llevo', 'familia', 'relación', 'mamá', 'papá', 'buena', 'hermanos'
recuerdo', 'recuerdos', 'infancia', 'bonitos', 'bonito', 'jugar', 'mamá'
dios', 'creo', 'creer', 'personas', 'religión', 'creen', 'existe',
colonia', 'vecinos', 'vivo', 'aquí', 'barrio', 'tranquila', 'gente',

6.2.5 Modelo de clasificación

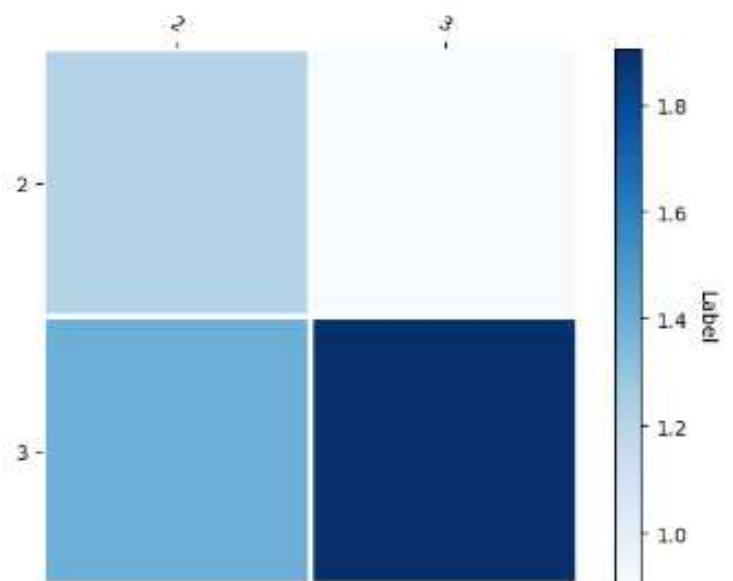
Se utilizó un modelo basado en transformador BETO-cased (basado en BERT) el cual encontró que el modelo de clasificación obtuvo un valor de F de 0.66, lo cual indicaría que el modelo estudiado tiene una precisión del 66% para el reconocimiento de personas que portan un diagnóstico psiquiátrico de aquellos que no entre las diferentes cohortes.

Se observó una exactitud del modelo del 75.2%, así como una exhaustividad de 0.65 (65%) y una precisión de 0.72 (72%).

Tabla 7: Modelo de clasificación

Gráfico 1: Modelo de clasificación BETO

Model: Beto-cased Optimizer: Adam Learning rate: 2e - 5 Steps: 1e - 8 Epochs: 50	
Accuracy	75.2
Recall	0.65
Precision	0.72
F-Measure Clase 2	0.48
F-Measure Clase 3	0.88
F-Measure	0.66



Capítulo VII

7. DISCUSIÓN

El Departamento de Psiquiatría pertenece al Hospital Universitario Dr. José Eleuterio González, el cual es un hospital de tercer nivel de atención en la zona metropolitana de Monterrey, Nuevo León, México. A pesar de la alta prevalencia de pacientes en el espectro de la psicosis atendidos en nuestra institución, consideramos que el tamaño de la muestra fue limitado.

Con base en las características clínicas de la población estudiada en el grupo de casos, teorizamos que la población de pacientes estudiada y que acude a nuestra institución tiene características que complican el abordaje diagnóstico y su estudio en momentos tempranos de la patología, lo anterior debido a la naturaleza de la sintomatología psicótica existente en algunos de ellos, lo cual influencia a ciertos individuos a optar por abandonar su seguimiento terapéutico. De la misma manera, notamos que esto afectaba en algunos sujetos a optar por no participar o hacerlo de manera parcial dentro de nuestro estudio, dificultando así su reclutamiento. Dicho lo anterior, consideramos importante el impacto de la red de apoyo de los sujetos estudiados para su afluencia y constancia en su seguimiento farmacológico y terapéutico, así como en su involucro en los diferentes estudios de investigación dentro de nuestra institución. Dentro de los motivos de ello, consideramos que puede influir de manera importante el brindar psicoeducación por parte del médico tratante, así como psicoeducación por parte de la institución para ayudar a contener y disminuir angustia en este tipo de población. Esta distinción podría servir para contrastar la influencia del diagnóstico previo en las respuestas de los participantes en el estudio.

El procesamiento de lenguaje natural, al ser una herramienta que se ha utilizado como apoyo diagnóstico en diferentes especialidades de la medicina, el auge que ha tenido dentro de las ciencias que estudian la salud mental como la Psiquiatría en el estudio de los trastornos psicóticos es de especial interés. Dicho esto, creemos que el apoyo de un modelo basado en el procesamiento de lenguaje natural es de ayuda para la identificación temprana de la población con alto riesgo de padecer un trastorno psicótico. En este estudio, la diversidad de perfiles permitió analizar las diferencias y similitudes en las respuestas a las preguntas planteadas durante el estudio en los diferentes grupos.

Siendo nuestra hipótesis: “La modelación del lenguaje natural de las personas hablantes de español sanas y aquellas con diagnóstico de problemas de salud mental permite diferenciar entre los grupos”, empleando un modelo de procesamiento de lenguaje natural, se encontró una precisión del mismo de un 72% para la diferenciación entre el grupo de casos y el grupo control.

En cuanto a la población estudiada, las comparaciones entre el grupo de casos y el grupo control referente al género, observamos que el grupo de casos tiene una mayor representación masculina en comparación al grupo de control, por lo cual, nuestros hallazgos son consistentes con lo reportado en la literatura. Asimismo, esta disparidad de género entre los grupos podría influir en la percepción de la salud mental y en la disposición de los participantes a compartir sus experiencias durante las grabaciones.

En cuanto a la población estudiada en nuestro estudio, en la sección 6.1 *Descripción de los pacientes*, se destaca la información sobre la zona de residencia y la nacionalidad de los participantes del grupo control. La mayor parte de los participantes del grupo control residen en una amplia variedad de estados de residencia de la República Mexicana. Estos datos podrían estar relacionados con factores socioeconómicos y culturales, los cuales influyen en la salud mental de la población estudiada.

El nivel educativo de los participantes del grupo control se observó un nivel educativo de estudios universitarios en la mayoría de los participantes (59.2%, N= 42). Este dato podría influir en la participación del estudio, así como en la interpretación de las preguntas y en la profundidad de las respuestas proporcionadas durante el estudio.

Relativo al análisis de las grabaciones de los audios obtenidos, se destaca que el grupo de control tuvo una mayor duración de grabación en comparación con el grupo de casos, lo que podría indicar una mayor participación e involucramiento en las preguntas planteadas. Esta mayor interacción podría revelar patrones de pensamiento más complejos o una mayor disposición a expresar pensamientos y emociones a las preguntas planteadas, lo que podría ser relevante para entender mejor las preocupaciones y pensamientos de los participantes.

En cuanto al análisis de las palabras y la riqueza léxica, los resultados obtenidos en el grupo de casos presentaron un total de 22,020 palabras con un vocabulario (es decir, palabras diferentes entre sí) de 4,088 palabras, a comparación del grupo control que presentó un total de 51,655 palabras con un vocabulario de 6,838 palabras. Con lo anterior, se obtuvo un diferencial de riqueza léxica total del 12.1%, una riqueza léxica del grupo de casos fue de 18.5% y para el grupo control se obtuvo una riqueza léxica de 13.2%. Con respecto al análisis de las palabras utilizadas y la riqueza léxica entre los grupos, los datos muestran que este rubro en el grupo de casos es menor que en el grupo de control. Esta diferencia podría reflejar diferentes niveles de expresión verbal y una mayor complejidad en la comunicación entre los grupos estudiados, asimismo podrían señalar diferentes aspectos cognitivos relevantes de los diferentes grupos.

En lo que respecta al análisis de los tópicos de los textos y las repeticiones de palabras, al analizar el apartado de 6.2.4 *Tópicos del lenguaje* (ver Tablas 5 y 6), observamos que existe una diferencia en la frecuencia y variedad de palabras que se utilizan en los diferentes grupos, donde se puede observar que existe un mayor número de repeticiones de palabras en las variadas respuestas dentro del grupo de casos, lo cual concuerda con la pobreza del vocabulario. Por el contrario, en el grupo de control existe una amplia variedad de palabras. Esta variación en las palabras utilizadas podría indicar diferentes enfoques cognitivos y en la expresión verbal de los participantes. Esto a su vez puede revelar patrones de pensamiento, interés y enfoque en los diferentes grupos estudiados. En resumen, el estudio realizado proporciona una fuente de información que permite explorar diversos aspectos relacionados con la salud mental, la comunicación y la respuesta a estímulos en pacientes con y sin un diagnóstico psiquiátrico. Creemos que el integrar estos hallazgos podría enriquecer la comprensión del comportamiento y la percepción e interacción con el entorno de la población estudiada en relación con el proceso de la evaluación clínica en quienes se sospecha un trastorno psicótico.

Para las futuras investigaciones en este ámbito de la inteligencia artificial, recomendamos que el aumento del tamaño de muestra con respecto al presentado en este estudio podría mejorar la precisión y exactitud del modelo en la diferenciación del discurso entre individuos con y sin trastornos psiquiátricos. Además, la inclusión de un mayor número de participantes provenientes de

diversas regiones geográficas de México y Latinoamérica, hispanohablantes, podría enriquecer el léxico y el vocabulario del modelo, contribuyendo así a una mayor precisión del mismo.

7.1 Limitaciones

Una de las principales limitaciones de nuestra investigación fue el reducido tamaño de la muestra, atribuido a la falta de información sobre pruebas de hipótesis al inicio del estudio. Por consiguiente, se recomienda llevar a cabo más estudios con poblaciones similares para confirmar estos resultados. Además, otra limitación fue la homogeneidad en los diagnósticos del grupo de casos y la uniformidad geográfica de la muestra, los cuales representaron una restricción significativa, por lo que se necesitan realizar investigaciones con poblaciones más diversas para enriquecer la variedad léxica y el vocabulario, contribuyendo así a una mayor precisión en el modelo utilizado.

7.2 Direcciones futuras

Dado el creciente interés en la investigación sobre el uso del procesamiento del lenguaje natural en la rama de la Psiquiatría, especialmente enfocada en los trastornos del espectro de la psicosis, se considera factible emplear estas herramientas ,basadas en inteligencia artificial, para el cribado o la diferenciación complementaria de los discursos psicóticos. Esto permitiría una intervención terapéutica más oportuna y eficaz, lo cual podría mejorar la calidad de vida de los pacientes. A pesar de los desafíos en el desarrollo de modelos de procesamiento de lenguaje natural para la identificación temprana de los trastornos psicóticos, confiamos en la utilidad del modelo presentado en esta tesis y su contribución al avance en este campo en constante evolución.

Capítulo VIII

8. CONCLUSIÓN

El modelo de clasificación desarrollado en este estudio logró una precisión del 66% en la distinción de individuos con diagnóstico dentro del espectro de la psicosis en comparación con aquellos sin dicho diagnóstico. Este hallazgo sugiere la posibilidad de entrenar modelos de procesamiento de lenguaje natural para diferenciar sujetos con diversos diagnósticos psiquiátricos. Por ende, recomendamos que las investigaciones futuras en el ámbito del procesamiento del lenguaje natural en psiquiatría incluyan una mayor diversidad en individuos con diagnósticos psiquiátricos, abarcando espectros como los trastornos afectivos, trastornos de ansiedad, suicidio, trastornos del sueño, trastorno por consumo de sustancias, entre otros.




Consideramos que el estudio de diversas técnicas de inteligencia artificial, en particular el procesamiento de lenguaje natural, podría ser de gran utilidad para los especialistas al enfrentar dudas diagnósticas entre patologías que comparten alteraciones del pensamiento, contribuyendo así a una atención terapéutica más temprana y eficaz.

Finalmente, consideramos relevante que la utilización de inteligencia artificial en la práctica clínica de la psiquiatría se emplee principalmente como una herramienta de respaldo para corroborar sospechas diagnósticas. Este uso debe enfocarse en proporcionar un diagnóstico descriptivo y en la identificación de patrones en el análisis del contenido verbal, sin remplazar el abordaje clínico integral del profesional.

Capítulo IX

9. ANEXOS

Consentimiento informado de cohorte 2 (Grupo de casos)

 UANL <small>UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN</small>	 DEPARTAMENTO DE PSQUIATRÍA <small>HOSPITAL UNIVERSITARIO</small>	 Facultad de Medicina y Hospital Universitario
--	--	--

FORMATO DE CONSENTIMIENTO INFORMADO COHORTE 2

Título del Estudio	Procesamiento de lenguaje natural en Psiquiatría
Nombre del Investigador Principal	Dr. Antonio Costilla Esquivel
Servicio / Departamento	Psiquiatría
Teléfono de Contacto	8116913795
Persona de Contacto	Carlos Antonio Bernal Rocha
Versión de Documento	Versión 5.0
Fecha de Documento	Julio 2022

Usted ha sido invitado(a) a participar en un estudio de investigación. Este documento contiene información importante acerca del propósito del estudio, lo que Usted hará si decide participar, y la forma en que nos gustaría utilizar su información personal y la de su salud. Puede contener palabras que Usted no entienda. Por favor solicite a su médico o al personal del estudio que le explique cualquier palabra o información que no le quede clara.

1.-¿CUÁL ES EL PROPÓSITO DEL ESTUDIO?

El propósito de este estudio es averiguar si hay diferencias entre la forma de hablar de personas con y sin problemas de salud mental que se puedan distinguir automáticamente. Se le pide participar porque antes ha participado en otros estudios o ha recibido atención en el Departamento de Psiquiatría del Hospital Universitario "Dr. José Eleuterio González" de la UANL. La investigación en la que Usted participará es importante porque podría apoyar a las y los médicos en la detección y tratamiento de problemas de salud mental.

2.-¿CUÁL SERÁ LA DURACIÓN DEL ESTUDIO Y CUÁNTOS PARTICIPANTES HABRÁ EN ESTE ESTUDIO?

Aceptaremos personas que quieran participar en el estudio durante 9 meses aproximadamente. Su participación individual consistirá en responder un cuestionario en línea en una sola ocasión. Se incluirá un mínimo de 90 personas en total. Si Usted acepta participar, el grupo que le correspondería es de 30 personas como mínimo.

3.-¿CUÁLES SON LOS REQUISITOS QUE SE TOMARÁN EN CUENTA PARA MI PARTICIPACIÓN?

Los criterios de inclusión y de exclusión son los siguientes:
Inclusión (debe cumplir todos los siguientes):

- Personas entre los 18 y 65 años de edad.
- Que hayan participado previamente en estudios de investigación o que reciban atención para problemas de salud mental en el Departamento de Psiquiatría del Hospital "Dr. José Eleuterio González" de la UANL.
- Que hablen español.

Página 1 de 6
Formato de Consentimiento Informado Cohorte 2 v5.0



Exclusión (basta cumplir alguno de los siguientes para ser excluido):

- Personas menores de 18 años o mayores de 65 años.
- Personas que se rehúsen a participar en el protocolo de investigación.
- Personas que no hablen español
- Personas con problemas que les impidan expresarse hablando.

4.-¿CUÁL ES EL TRATAMIENTO DEL ESTUDIO?

Si Usted decide participar en este estudio de investigación no recibirá ningún tratamiento, sólo se le pedirá que responda un cuestionario en línea, con algunas respuestas escritas, y otras grabadas en audio.

5.-¿CUÁLES SON LOS PROCEDIMIENTOS QUE SE ME REALIZARÁN?

Se le invitará a firmar este Formato de Consentimiento Informado, luego de que le sea leído, que se satisfagan todas sus dudas, y ante la presencia de dos testigos.

Se le pedirá que conteste el cuestionario en línea.

6.-¿QUÉ VA A HACER SI USTED DECIDE PARTICIPAR EN ESTE ESTUDIO?

Si Usted da su consentimiento para participar, se le pedirá responder un cuestionario que tiene dos tipos de preguntas. Algunas están relacionadas con datos generales: edad, género, medicamentos que se emplean al momento de la entrevista, e intensidad de sus síntomas o molestias valuados en una escala del 0 al 10, correspondiendo "0" a nada de sintomatología a "10" sintomatología incapacitante. Las cuatro preguntas finales son abiertas, y se le pedirá que las responda lo más ampliamente posible, grabando su respuesta en el mismo cuestionario en línea. Si alguna pregunta le parece difícil o prefiere no contestarla, podrá solicitar que se la asigne otra diferente. No hay respuestas correctas o incorrectas; lo único que le pedimos es que las conteste tan ampliamente como sea posible.

7.-¿CUÁLES SON LOS POSIBLES RIESGOS O MOLESTIAS?

No anticipamos que sufra riesgos o molestias, excepto por la inversión de tiempo para responder el cuestionario. Este tiempo para contestar el cuestionario está estimado entre quince y veinte minutos.

8.-¿CUÁLES SON LOS POSIBLES BENEFICIOS PARA USTED O PARA OTROS?

Aunque no es probable que Usted tenga un beneficio directo por participar en este estudio de investigación, su participación nos ayudará a verificar si analizar automáticamente la manera en que las personas hablan puede ayudar a distinguir las que necesitan atención de una o un psiquiatra, por tener un posible problema de salud mental.

La participación en este estudio puede ayudar a las y los médicos a comprender mejor las manifestaciones de los problemas de salud mental, y a medir el efecto de los tratamientos para atenderlos.

9.-¿QUÉ OTROS PROCEDIMIENTOS O TRATAMIENTOS PODRÍAN ESTAR DISPONIBLES PARA USTED?

El estudio no involucra ningún tratamiento. Usted no tiene que participar en este estudio de investigación si no lo desea.

10.-¿SU PARTICIPACIÓN EN ESTE ESTUDIO LE GENERARÁ ALGÚN COSTO?

No habrá costos para Usted por participar en este estudio.

11.-¿SE LE PROPORCIONARÁ ALGUNA COMPENSACIÓN ECONÓMICA PARA GASTOS DE TRANSPORTACIÓN?

A Usted se le proporcionará una compensación para gastos de transporte de \$200.00 (doscientos pesos 00/100 MN) por una ocasión en el día que acuda al Departamento de Psiquiatría del Hospital Universitario "Dr. José Eleuterio González", a firmar el consentimiento informado por escrito.

12.-¿RECIBIRÁ ALGÚN PAGO POR SU PARTICIPACIÓN EN ESTE ESTUDIO?

Usted no recibirá ningún pago por la participación en este estudio.

13.-¿SE ALMACENARÁN MUESTRAS DE SANGRE O TEJIDOS PARA FUTURAS INVESTIGACIONES?

No. Sus respuestas al cuestionario serán almacenadas y podrían ser usadas en futuras investigaciones, pero no se guardarán datos que permitan identificarle como su nombre o dirección.

14.-¿QUÉ DEBE HACER SI LE PASA ALGO COMO RESULTADO DE PARTICIPAR EN ESTE ESTUDIO?

Si Usted sufre una lesión o enfermedad durante su participación en el estudio, debe buscar tratamiento a través de su médico de cabecera o centro de atención médica de elección y debe informárselo inmediatamente al médico del estudio.

Los gastos que genere dicha lesión o enfermedad sólo le serán pagados si el médico del estudio ha decidido que la lesión / enfermedad está directamente relacionada con los procedimientos del estudio, y no es el resultado de una condición pre-existente de la progresión normal de su enfermedad, o porque no se han seguido las indicaciones que el médico de estudio ha recomendado.

15.-¿CUÁLES SON SUS DERECHOS COMO SUJETO DE INVESTIGACIÓN?

Si decide participar en este estudio, Usted tiene derecho a ser tratado con respeto, incluyendo la decisión de continuar o no su participación en el estudio. Usted es libre de terminar su participación en este estudio en cualquier momento.

16.- ¿PUEDE TERMINAR SU PARTICIPACIÓN EN CUALQUIER MOMENTO DEL ESTUDIO?

Su participación es estrictamente voluntaria. Si desea suspender su participación, puede hacerlo con libertad en cualquier momento. Si elige no participar o retirarse del estudio, su atención médica presente y/o futura no se verá afectada y no incurrirá en sanciones ni perderá los beneficios a los que usted tendría derecho de algún otro modo.

Su participación también podrá ser suspendida o terminada por el médico del estudio, sin su consentimiento, por cualquiera de las siguientes circunstancias:

- Que el estudio haya sido cancelado.
- Que el médico considere que es lo mejor para Usted.
- Que necesita algún procedimiento o medicamento que interfiere con esta investigación.
- Que no ha seguido las indicaciones del médico lo que pudiera traer como consecuencias problemas en su salud.

Si Usted decide retirarse de este estudio, deberá realizar lo siguiente:

- Notificar a su médico tratante del estudio
- Deberá de regresar todo el material que su médico le solicite.

Si su participación en el estudio se da por terminada, por cualquier razón, por su seguridad, el médico continuará con seguimientos clínicos. Además, su información médica recabada hasta ese momento podrá ser utilizada para fines de la investigación.

17.- ¿CÓMO SE PROTEGERÁ LA CONFIDENCIALIDAD DE SUS DATOS PERSONALES Y LA INFORMACIÓN DE SU EXPEDIENTE CLÍNICO?

Si acepta participar en la investigación, el médico del estudio recabará y registrará información personal confidencial acerca de su salud y de su tratamiento. Esta información no contendrá su nombre completo ni su domicilio, pero podrá contener otra información acerca de Usted, tal como iniciales y su fecha de nacimiento. Toda esta información tiene como finalidad garantizar la integridad científica de la investigación. Su nombre no será conocido fuera de la Institución al menos que lo requiera nuestra Ley.

Usted tiene el derecho de controlar el uso de sus datos personales de acuerdo a la Ley Federal de Protección de datos Personales en Posición de Particulares, así mismo de solicitar el acceso, corrección y oposición de su información personal. La solicitud será procesada de acuerdo a las regulaciones de protección de datos vigentes. Sin embargo, cierta información no podrá estar disponible hasta que el estudio sea completado, esto con la finalidad de proteger la integridad del Estudio.

La Facultad de Medicina y Hospital Universitario, así como el Investigador serán los responsables de salvaguardar la información de acuerdo con las regulaciones locales.

Usted tiene el derecho de solicitar por escrito al médico un resumen de su expediente clínico.

La información personal acerca de su salud y de su tratamiento del estudio podrá procesarse o transferirse a terceros en otros países para fines de investigación y de reportes de seguridad, incluyendo agencias reguladoras locales (Secretaría de Salud SSA a través de la COFEPRIS), así como al Comité de Ética en Investigación y al Comité de Investigación de nuestra Institución.

Para los propósitos de este estudio, autoridades sanitarias como la Secretaría de Salud y el Comité de Ética en Investigación y/o el Comité de Investigación de nuestra Institución, podrán inspeccionar su expediente clínico, incluso los datos que fueron recabados antes del inicio de su participación, los cuales pueden incluir su nombre, domicilio u otra información personal.

En caso necesario estas auditorías o inspecciones podrán hacer fotocopias de parte o de todo su expediente clínico. La razón de esto es asegurar que el estudio se está llevando a cabo apropiadamente con la finalidad de salvaguardar sus derechos como sujeto en investigación.

Los resultados de este estudio de investigación podrán presentarse en reuniones o en publicaciones.

La información recabada durante este estudio será recopilada en bases de datos del investigador, los cuales podrán ser usados en otros estudios en el futuro. Estos datos no incluirán información médica personal confidencial. Se mantendrá el anonimato.

Al firmar este documento, Usted autoriza el uso y revelaciones de la información acerca de su estado de salud y tratamiento identificado en esta forma de consentimiento. No perderá ninguno de sus derechos legales como sujeto de investigación. Si hay cambios en el uso de su información, su médico le informará.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



DEPARTAMENTO DE
PSIQUIATRÍA
HOSPITAL UNIVERSITARIO

Facultad de Medicina y
Hospital Universitario



18.- SI TIENE PREGUNTAS O INQUIETUDES ACERCA DE ESTE ESTUDIO DE INVESTIGACIÓN, ¿A QUIÉN PUEDE LLAMAR?

En caso de tener alguna pregunta relacionada a sus derechos como sujeto de investigación de la Facultad de Medicina y Hospital Universitario podrá contactar al **Dr. med. Oscar de la Garza Castro**, Presidente del Comité de Ética en Investigación de nuestra Institución o al **Lic Antonio Zapata de la Riva** en caso de tener dudas en relación a sus derechos como paciente.

Comité de Ética en Investigación del Hospital Universitario "Dr. José Eleuterio González".

Av. Francisco I. Madero y Av. Gonzalitos s/n
Col. Mitras Centro, Monterrey, Nuevo León México.
CP 64460

Teléfonos: 8183294050 ext. 2870 a 2874

Correo electrónico: investigacionclinica@meduanl.com

RESUMEN CONSENTIMIENTO

PARA LLENAR POR EL SUJETO DE INVESTIGACIÓN

- € Mi participación es completamente voluntaria.
- € Confirmando que he leído y entendido este documento y la información proporcionada del estudio.
- € Confirmando que se me ha explicado el estudio, que he tenido la oportunidad de hacer preguntas y que se me ha dado el tiempo suficiente para decidir sobre mi participación. Sé con quién debo comunicarme si tengo más preguntas.
- € Entiendo que las secciones de mis anotaciones médicas serán revisadas cuando sea pertinente por el Comité de Ética en Investigación o cualquier otra autoridad regulatoria para proteger mi participación en el estudio.
- € Acepto que mis datos personales se archiven bajo códigos que permitan mi identificación.
- € Acepto que mi médico general sea informado de mi participación en este estudio.
- € Acepto que la información acerca de este estudio y los resultados de cualquier examen o procedimiento pueden ser incluidos en mi expediente clínico.
- € Confirmando que se me ha entregado una copia de este documento de consentimiento firmado.

Nombre del Sujeto de Investigación _____

Firma _____

Fecha _____

Página 5 de 6

Formato de Consentimiento Informado Cohorte 2 v5.0

PRIMER TESTIGO

Nombre del Primer Testigo

Firma

Dirección

Fecha

Relación con el Sujeto de Investigación

SEGUNDO TESTIGO

Nombre del Segundo Testigo

Firma

Dirección

Fecha

Relación con el Sujeto de Investigación

PERSONA QUE OBTIENE CONSENTIMIENTO

He discutido lo anterior y he aclarado las dudas. A mi más leal saber y entender, el sujeto está proporcionando su consentimiento tanto voluntariamente como de una manera informada, y él/ella posee el derecho legal y la capacidad mental suficiente para otorgar este consentimiento.

Nombre de la Persona que obtiene el Consentimiento

Firma

Fecha

Consentimiento informado de cohorte 3 (Grupo de control)



Consentimiento informado en línea

Cohorte 3

Usted ha sido invitado(a) a participar en un estudio de investigación llamado "Procesamiento de lenguaje natural en Psiquiatría", que tiene como investigador principal al Dr. Antonio Costilla Esquivel, y como persona de contacto al Dr. Carlos Antonio Bernal Rocha, con quien se puede comunicar en el teléfono 8116913795, en el Departamento de Psiquiatría del Hospital Universitario "Dr. José Eleuterio González", ubicado en la Avenida Francisco I. Madero y Gonzalitos S/N, Col. Mitras Centro, C.P. 64460, en Monterrey, N.L., México.

El propósito de este estudio es determinar la utilidad del análisis automatizado del lenguaje para apoyar a los médicos en el diagnóstico de problemas de salud mental. La investigación es importante porque con los resultados obtenidos se espera crear información para una mejor atención de los padecimientos de salud mental.

Su participación no generará ningún costo para usted. Tampoco se le remunerará económicamente por ella.

Los requisitos necesarios para formar parte de esta cohorte de sujetos en el estudio son:

- No haber recibido algún diagnóstico psiquiátrico al momento del estudio.
- Tener entre 18 y 65 años de edad.
- Ser hablante del idioma español.

Le pedimos su permiso para grabar y transcribir sus respuestas a cuatro preguntas muy generales. Se recabarán así mismo algunos datos como su género, su edad y la ausencia de diagnósticos o tratamientos psiquiátricos. Lo que usted diga será transformado en texto y sometido a análisis automatizados. Estimamos que le tomará aproximadamente 12 minutos responder el cuestionario y grabar sus respuestas.

No guardaremos datos que permitan su identificación. Aún así, sus respuestas al cuestionario serán estrictamente confidenciales. La Facultad de Medicina y Hospital Universitario, así como el Investigador serán los responsables de salvaguardar la información de acuerdo con las leyes de nuestro país.

Su participación es completamente voluntaria y puede negarse a participar en el estudio sin hacerse acreedor a ningún tipo de sanción o problema. Su participación dentro de este estudio de investigación no conlleva costo alguno a los participantes.

Si desea acceder al Aviso de Privacidad del Departamento de Psiquiatría, puede consultarlo en el Departamento de Psiquiatría, Hospital Universitario, UANL.

En caso de tener alguna pregunta relacionada a sus derechos como sujeto de investigación de la Facultad de Medicina y Hospital Universitario podrá contactar al **Dr. Óscar de la Garza Castro**, Presidente del Comité de Ética en Investigación de nuestra Institución o al **Lic. Antonio Zapata de la Riva** en caso de tener dudas en relación a sus derechos como paciente.

Comité de Ética en Investigación del Hospital Universitario "Dr. José Eleuterio González".

Av. Francisco I. Madero y Av. Gonzalitos s/n
Col. Mitras Centro, Monterrey, Nuevo León México.

CP 64460

Teléfonos: 8183294050 ext. 2870 a 2874

Correo electrónico: investigacionclinica@meduanl.com

¿Está de acuerdo en participar en el estudio?

Sí.



No.



Consentimiento informado en línea

Usted ha sido invitado(a) a participar en un estudio de investigación llamado "Procesamiento de lenguaje natural en Psiquiatría", que tiene como investigador principal al Dr. Antonio Costilla Esquivel, y como persona de contacto al Dr. Carlos Antonio Bernal Rocha, con quien se puede comunicar en el teléfono 8116913795, en el Departamento de Psiquiatría del Hospital Universitario "Dr. José Eleuterio González", ubicado en la Avenida Francisco I. Madero y Gonzalitos S/N, Col. Mitras Centro, C.P. 64460, en Monterrey, N.L., México.

El propósito de este estudio es determinar la utilidad del análisis automatizado del lenguaje para apoyar a los médicos en el diagnóstico de problemas de salud mental. La investigación es importante porque con los resultados obtenidos se espera crear información para una mejor atención de los padecimientos de salud mental.

Su participación no generará ningún costo para usted. Tampoco se le remunerará económicamente por ella.

Los requisitos necesarios para formar parte de esta cohorte de sujetos en el estudio son:

- No haber recibido algún diagnóstico psiquiátrico al momento del estudio.
- Tener entre 18 y 65 años de edad.
- Ser hablante del idioma español.

Le pedimos su permiso para grabar y transcribir sus respuestas a cuatro preguntas muy generales. Se recabarán así mismo algunos datos como su género, su edad y la ausencia de diagnósticos o tratamientos psiquiátricos. Lo que usted diga será transformado en texto y sometido a análisis automatizados. Estimamos que le tomará aproximadamente 12 minutos responder el cuestionario y grabar sus respuestas.

No guardaremos datos que permitan su identificación. Aún así, sus respuestas al cuestionario serán estrictamente confidenciales. La Facultad de Medicina y Hospital Universitario, así como el Investigador serán los responsables de salvaguardar la información de acuerdo con las leyes de nuestro país.

Su participación es completamente voluntaria y puede negarse a participar en el estudio sin hacerse acreedor a ningún tipo de sanción o problema. Su participación dentro de este estudio de investigación no conlleva costo alguno a los participantes.

Si desea acceder al Aviso de Privacidad del Departamento de Psiquiatría, puede consultarlo en el Departamento de Psiquiatría, Hospital Universitario, UANL.

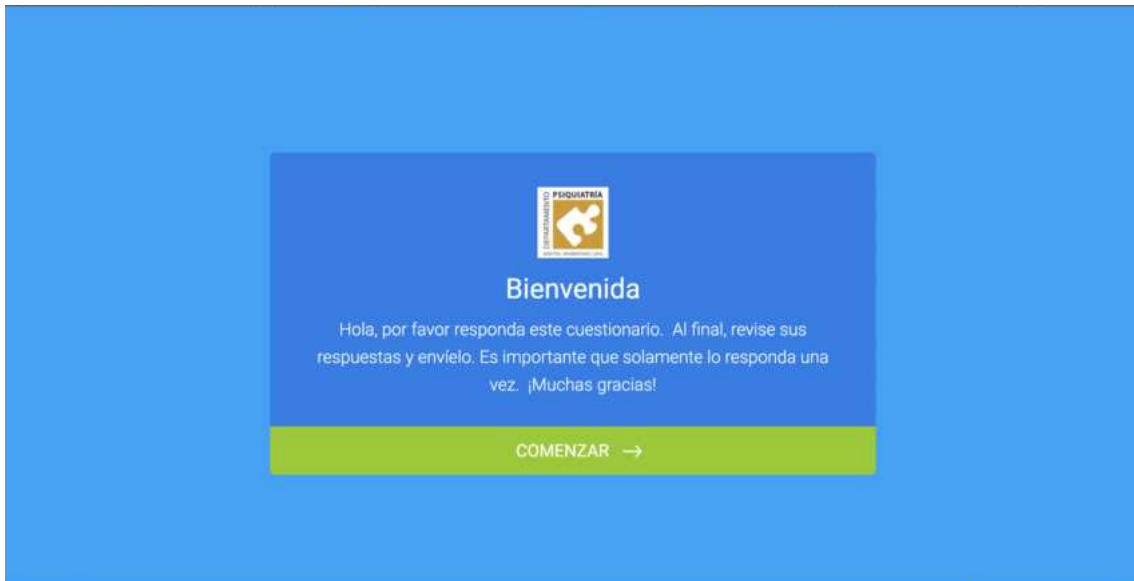
En caso de tener alguna pregunta relacionada a sus derechos como sujeto de investigación de la Facultad de Medicina y Hospital Universitario podrá contactar al Dr. Óscar de la Garza Castro, Presidente del Comité de Ética en Investigación de nuestra Institución o al Lic Antonio Zapata de la Riva en caso de tener dudas en relación a sus derechos como paciente.

Comité de Ética en Investigación del Hospital Universitario "Dr. José Eleuterio González". Av. Francisco I. Madero y Av. Gonzalitos s/n Col. Mitras Centro, Monterrey, Nuevo León México, CP 64460

Teléfonos: 8183294050 ext. 2870 a 2874

Correo electrónico: investigacionclinica@meduanl.com

Formulario de grupo de casos



A welcome screen for a questionnaire. It features a blue background with a central white box containing a logo for 'PSIQUIATRÍA' and the text 'Bienvenida'. Below the text is a green button labeled 'COMENZAR' with a right-pointing arrow.

PSIQUIATRÍA

Bienvenida

Hola, por favor responda este cuestionario. Al final, revise sus respuestas y envíelo. Es importante que solamente lo responda una vez. ¡Muchas gracias!

COMENZAR →



A questionnaire form titled 'Procesamiento de lenguaje natural en Psiquiatría'. It features a blue background with a white form area. The form has an 'Email*' label, an input field with a placeholder 'ejemplo@ejemplo.com', and a green bar at the bottom with 'PREVIA' and 'SIGUIENTE' buttons. A red error message 'Este campo es obligatorio' is visible. At the bottom, there is a progress indicator and a '2 de 26' label.

Procesamiento de lenguaje natural en Psiquiatría

Email*

← PREVIA Este campo es obligatorio. SIGUIENTE →

2 de 26

¿Cuántos años tiene?*

← PREVIA

SIGUIENTE →



3 de 26



¿Cuál es su género?*

Mujer

Hombre

← PREVIA

SIGUIENTE →



4 de 26



¿Cuál es su máximo nivel de estudios?*

Primaria

Secundaria


Preparatoria

Carrera universitaria

Posgrado

← PREVIA SIGUIENTE →

5 de 26



En el momento de contestar este cuestionario, ¿está tomando algún medicamento?*

Sí








No

← PREVIA SIGUIENTE →


6 de 26



¿Qué medicamentos está tomando actualmente? Escriba el nombre, y si las recuerda, la dosis y la fecha en que empezó a tomarlo.

⌵ B I U       

← PREVIA SIGUIENTE →

7 de 26 

¿Qué tan intensas están sus molestias o dificultades psiquiátricas en este momento? (0 = nada intensas; 10 = extremadamente intensas)

← PREVIA SIGUIENTE →

8 de 26 

En las siguientes preguntas le pedimos que grabe una respuesta de audio tan amplia como le sea posible hasta completar 4 preguntas contestadas . Si la pregunta le parece difícil o poco clara puede omitirla. Si la acepta, oprima el botón de "record" y hable cerca del micrófono de su dispositivo; cuando haya terminado su respuesta oprima "stop". Puede revisar su grabación oprimiendo "play",y regrabarla si lo desea. ¿Ha comprendido las instrucciones?*

SÍ

NO

← PREVIA

SIGUIENTE →

9 de 26



¿Cuáles fueron sus actividades del día de ayer?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ

NO

← PREVIA

SIGUIENTE →

10 de 26



Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

11 de 26

¿Qué es para usted un buen día?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

← PREVIA SIGUIENTE →

12 de 26

Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

13 de 26

¿A qué se debe que algunas personas creen en Dios y otras no?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

← PREVIA SIGUIENTE →

14 de 26

Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

15 de 26

¿Cuál es la actividad que más le gusta hacer y porqué es su favorita?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

← PREVIA SIGUIENTE →

16 de 26

Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

17 de 26

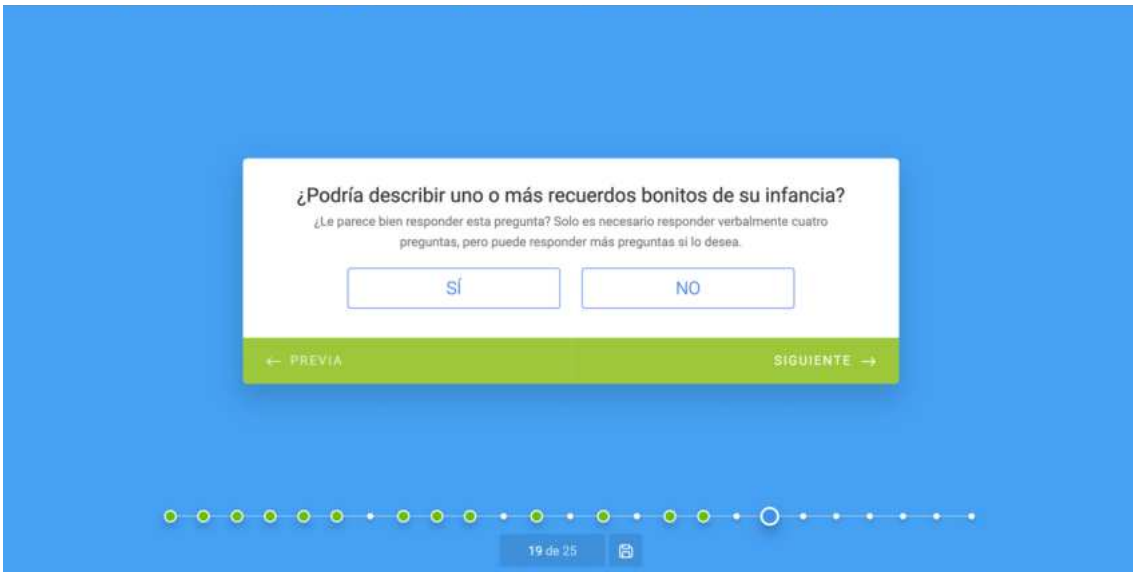
¿Podría describir cómo se lleva con los miembros de su familia?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

← PREVIA SIGUIENTE →

17 de 25



Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

21 de 26

¿Cómo es la colonia o barrio en el que vive?

¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

← PREVIA SIGUIENTE →

21 de 25

Grabadora de voz

Record 0:00 / 10:00

← PREVIA SIGUIENTE →

23 de 26

Le recordamos que debe guardar sus respuestas con el ícono que parece un disco, guardar en la parte inferior de la presentación para tablet y computadora de cada pregunta. Una vez guardado puede dar click en la "X" de la parte superior derecha, luego en volver, y finalizar con el formulario.

← PREVIA SIGUIENTE →

23 de 25

¡Gracias por participar!
Ha llegado al final de cuestionario

← PREVIA SIGUIENTE →

24 de 25



Estamos a sus órdenes para resolver sus dudas en el siguiente correo:
lenguajepsiq@gmail.com

← PREVIA REVISAR Y ENVIAR

25 de 25



Procesamiento de lenguaje natural en Psiquiatría – Cohorte 2

2. Email
c@gmail.com

3. ¿Cuántos años tiene?
28

4. ¿Cuál es su género?
Hombre

5. ¿Cuál es su máximo nivel de estudios?
Carrera universitaria

6. En el momento de contestar este cuestionario, ¿está tomando algún medicamento?
Sí

7. ¿Qué medicamentos está tomando actualmente? Escriba el nombre, y si las recuerda, la dosis y la fecha en que empezó a tomarlo.

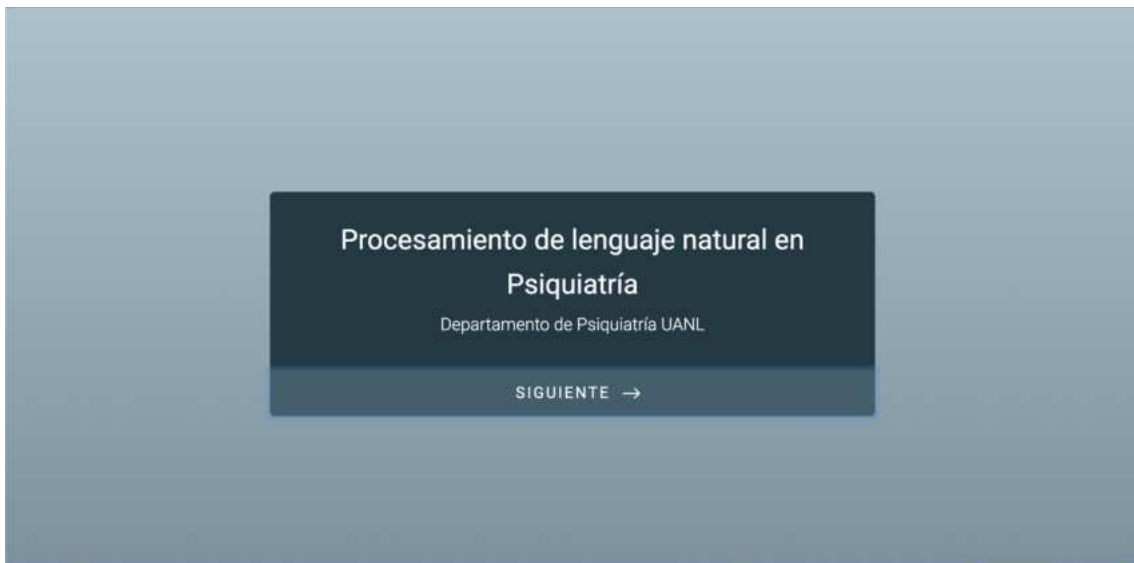
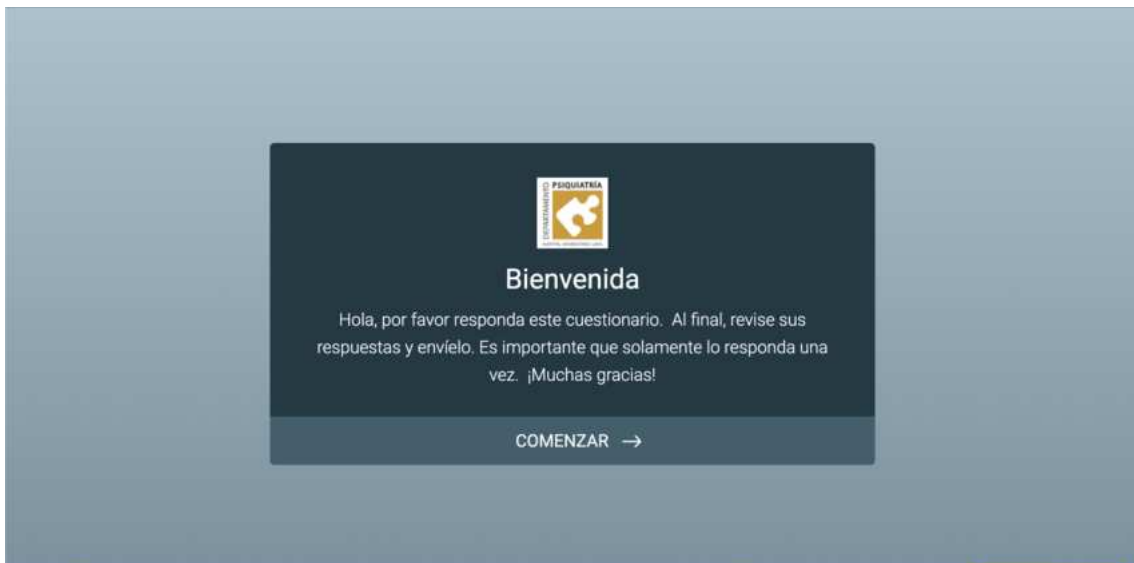
← IR AL FORMULARIO ENVIAR



Gracias!

¡Su envío ha sido recibido!

Formulario del grupo de control



Procesamiento de lenguaje natural en Psiquiatría

Consentimiento informado en línea

Usted ha sido invitado(a) a participar en un estudio de investigación llamado "Procesamiento de lenguaje natural en Psiquiatría", que tiene como investigador principal al Dr. Antonio Costilla Esquivel, y como persona de contacto al Dr. Carlos Antonio Bernal Rocha, con quien se puede comunicar en el teléfono 8116913795, en el Departamento de Psiquiatría del Hospital Universitario "Dr. José Eleuterio González", ubicado en la Avenida Francisco de M... B... C... P... 51468... M... M...

SIGUIENTE →

¿Esta de acuerdo en participar en el estudio?*

Si su respuesta es "No", por favor envíela oprimiendo el botón que aparecerá en la parte inferior del recuadro.

Sí No

← PREVIA

SIGUIENTE →



2 of 30




¿En qué tipo de dispositivo está respondiendo el cuestionario?*

Teléfono móvil Tableta o computadora

← PREVIA SIGUIENTE →

3 of 30



Por favor, lea con atención los siguientes enunciados, y capture, seleccione o grabe la respuesta a ellos.

← PREVIA SIGUIENTE →

4 of 30



Por favor capture su dirección de Email*

ejemplo@ejemplo.com

← PREVIA SIGUIENTE →


5 of 30



¿Cuántos años tiene?*

← PREVIA SIGUIENTE →

6 of 30



¿Cuál es su género?*

Mujer

Hombre

← PREVIA SIGUIENTE →

7 of 30



¿Es usted mexicana o mexicano?

← PREVIA SIGUIENTE →

8 of 30



¿En cuál estado de la República Mexicana ha vivido por más tiempo?

← PREVIA SIGUIENTE →

9 of 29

¿Cuál es su máximo nivel de estudios?*

Primaria

Secundaria

Preparatoria

Carrera universitaria

Posgrado

← PREVIA SIGUIENTE →

10 of 29

En el momento de contestar este cuestionario, ¿está tomando algún medicamento?*

Sí

No

← PREVIA SIGUIENTE →

11 of 29



¿Alguna vez ha recibido un diagnóstico psiquiátrico, o tratamiento con una o un psiquiatra?*

Sí NO

← PREVIA SIGUIENTE →

12 of 28



En las siguientes preguntas le pedimos que grabe una respuesta de audio (de 1 a 15 minutos), hasta completar 4 preguntas contestadas. Si la pregunta le parece difícil o poco clara puede omitirla. Si la acepta, oprima el botón de "record" y hable cerca del micrófono de su dispositivo; cuando haya terminado su respuesta, oprima "stop". Puede revisar su grabación oprimiendo "play", y regrabarla si lo desea. ¿Ha comprendido las instrucciones anteriores?

Si he comprendido las instrucciones

No he comprendido las instrucciones

← PREVIA SIGUIENTE →

13 of 28

¿Cuáles fueron sus actividades del día de ayer?

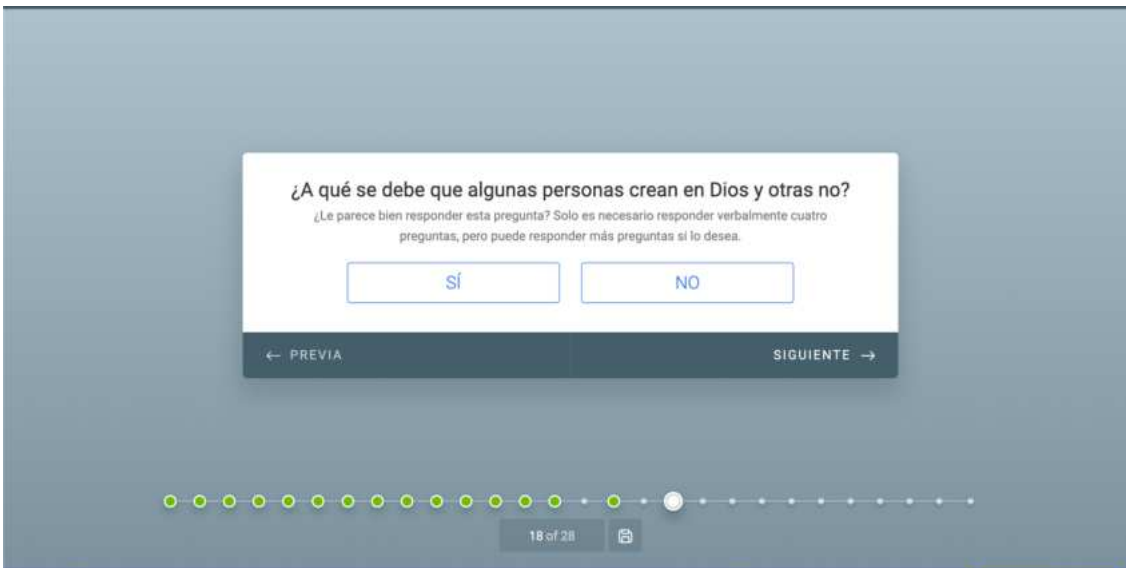
¿Le parece bien responder esta pregunta? Solo es necesario responder verbalmente cuatro preguntas, pero puede responder más preguntas si lo desea.

SÍ NO

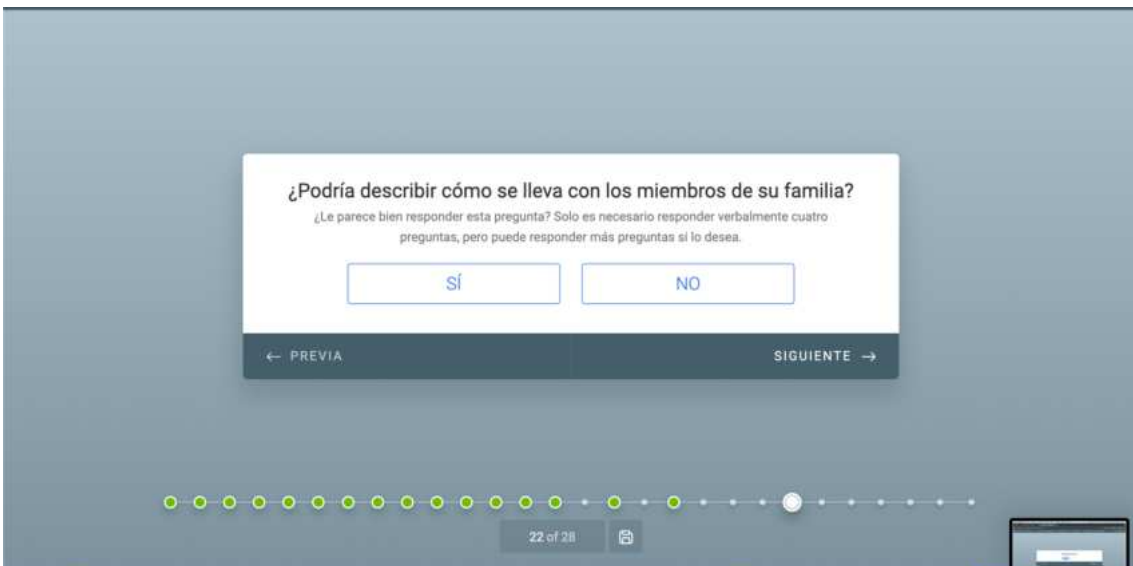
← PREVIA SIGUIENTE →

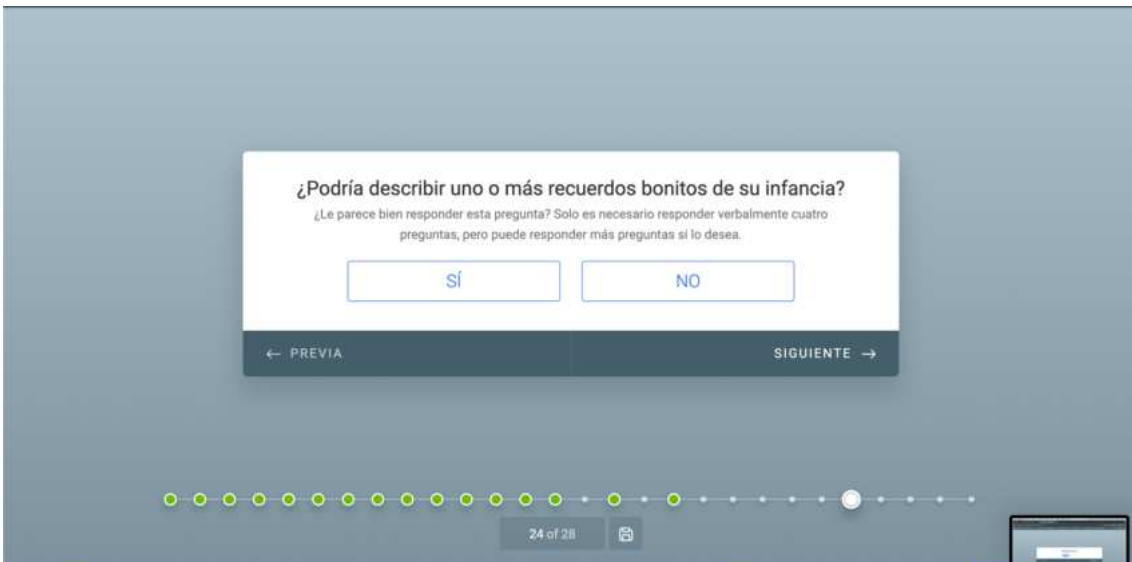
14 of 28















Capítulo X

10. BIBLIOGRAFÍA

- American Psychiatric Association (APA). n.d. “American Psychiatric Association. (2014c). DSM-5. Manual Diagnóstico y Estadístico de Los Trastornos Mentales (5.a Ed.). Editorial Médica Panamericana.”
- Bedi, G, F Carrillo, GA Cecchi, DF Slezak, M Sigman, and NB Mota. 2015. “Automated Analysis of Free Speech Predicts Psychosis Onset in High-Risk Youths.” *NPJ Schizophr.* 1 (1): 15030. <https://doi.org/10.1038/npjpsz.2015.30>.
- Bürgy, Martin. 2008a. “The Concept of Psychosis: Historical and Phenomenological Aspects.” *Schizophrenia Bulletin*. <https://doi.org/10.1093/schbul/sbm136>.
- . 2008b. “The Concept of Psychosis: Historical and Phenomenological Aspects.” *Schizophrenia Bulletin*. <https://doi.org/10.1093/schbul/sbm136>.
- Canuma Prince. 2019. “The Brief History of NLP. In This Article, We Are Going to Take A... | by Prince Canuma | DataDrivenInvestor.” August 29, 2019. <https://medium.datadriveninvestor.com/the-brief-history-of-nlp-c90f331b6ad7>.
- Corcoran, Cheryl M., Facundo Carrillo, Diego Fernández-Slezak, Gillinder Bedi, Casimir Klim, Daniel C. Javitt, Carrie E. Bearden, and Guillermo A. Cecchi. 2018. “Prediction of Psychosis across Protocols and Risk Cohorts Using Automated Language Analysis.” *World Psychiatry* 17 (1): 67–75. <https://doi.org/10.1002/wps.20491>.
- Elvevåg, Brita, Peter W. Foltz, Mark Rosenstein, and Lynn E. DeLisi. 2010. “An Automated Method to Analyze Language Use in Patients with Schizophrenia and Their First-Degree Relatives.” *Journal of Neurolinguistics* 23 (3): 270–84. <https://doi.org/10.1016/j.jneuroling.2009.05.002>.
- Elvevåg, Brita, Peter W Foltz, Daniel R Weinberger, and Terry E Goldberg. n.d. “Quantifying Incoherence in Speech: An Automated Methodology and Novel Application to Schizophrenia.”
- Grosz, Barbara J. 1982. “Natural Language Processing.” *Artificial Intelligence* 19 (2): 131–36. [https://doi.org/10.1016/0004-3702\(82\)90032-7](https://doi.org/10.1016/0004-3702(82)90032-7).
- Rezaii, Neguine, Elaine Walker, and Phillip Wolff. 2019. “A Machine Learning Approach to Predicting Psychosis Using Semantic Density and Latent Content Analysis.” *Npj Schizophrenia* 5 (1). <https://doi.org/10.1038/s41537-019-0077-9>.

- Rosenstein, Mark, Catherine Diaz-Asper, Peter W Foltz, and Brita Elvevåg. 2014. "A Computational Language Approach to Modeling Prose Recall in Schizophrenia." *Cortex* 55: 148–66. <https://doi.org/10.1016/j.cortex.2014.01.021>.
- Schrimpf, By Lindsey A, Arpit Aggarwal, and John Lauriello. 2018. "Psychosis," 845–60.
- Tang, Sunny X., Reno Kriz, Sunghye Cho, Suh Jung Park, Jenna Harowitz, Raquel E. Gur, Mahendra T. Bhati, Daniel H. Wolf, João Sedoc, and Mark Y. Liberman. 2021. "Natural Language Processing Methods Are Sensitive to Sub-Clinical Linguistic Differences in Schizophrenia Spectrum Disorders." *Npj Schizophrenia* 7 (1). <https://doi.org/10.1038/s41537-021-00154-3>.
- Torres, Felix. n.d. "What Is Schizophrenia?" Accessed September 5, 2021. <https://www.psychiatry.org/patients-families/schizophrenia/what-is-schizophrenia>.
- Bedi, G, F Carrillo, GA Cecchi, DF Slezak, M Sigman, and NB Mota. 2015. "Automated Analysis of Free Speech Predicts Psychosis Onset in High-Risk Youths." *NPJ Schizophr.* 1 (1): 15030. <https://doi.org/10.1038/npjSchz.2015.30>.
- Bujang, Mohamad Adam, and Tassha Hilda Adnan. 2016. "Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis." *Journal of Clinical and Diagnostic Research* 10 (10): YE01–6. <https://doi.org/10.7860/JCDR/2016/18129.8744>.
- Corcoran, Cheryl M., Facundo Carrillo, Diego Fernández-Slezak, Gillinder Bedi, Casimir Klim, Daniel C. Javitt, Carrie E. Bearden, and Guillermo A. Cecchi. 2018. "Prediction of Psychosis across Protocols and Risk Cohorts Using Automated Language Analysis." *World Psychiatry* 17 (1): 67–75. <https://doi.org/10.1002/wps.20491>.
- Haas, S. S., G. E. Doucet, S. Garg, S. N. Herrera, C. Sarac, Z. R. Bilgrami, R. B. Shaik, and C. M. Corcoran. 2020. "Linking Language Features to Clinical Symptoms and Multimodal Imaging in Individuals at Clinical High Risk for Psychosis." *European Psychiatry* 63 (1). <https://doi.org/10.1192/J.EURPSY.2020.73>.
- Shi, Leming, Gregory Campbell, Wendell D. Jones, Fabien Campagne, Zhining Wen, Stephen J. Walker, Zhenqiang Su, et al. 2010. "The MicroArray Quality Control (MAQC)-II Study of Common Practices for the Development and Validation of Microarray-Based Predictive Models." *Nature Biotechnology*

- 2010 28:8 28 (8): 827–38. <https://doi.org/10.1038/nbt.1665>.
- Sommer, IE, NF Ramsey, RC Mandl, CJ Van Oel, and RS Kahn. 2004. “Language Activation in Monozygotic Twins Discordant for Schizophrenia.” *Br J Psychiatry*. 184 (FEB.): 128–35. <https://doi.org/10.1192/bjp.184.2.128>.
- Tang, Sunny X., Reno Kriz, Sunghye Cho, Suh Jung Park, Jenna Harowitz, Raquel E. Gur, Mahendra T. Bhati, Daniel H. Wolf, João Sedoc, and Mark Y. Liberman. 2021. “Natural Language Processing Methods Are Sensitive to Sub-Clinical Linguistic Differences in Schizophrenia Spectrum Disorders.” *Npj Schizophrenia* 7 (1). <https://doi.org/10.1038/s41537-021-00154-3>.

Capítulo XI

11. RESUMEN AUTOBIOGRÁFICO

Carlos Antonio Bernal Rocha, nació en Monterrey, Nuevo León, México el 31 de Marzo de 1995.

Realizó el bachillerato en la Escuela Preparatoria No. 7 San Nicolas, Unidad Oriente del 2010 al 2012. Del 2012 al 2018 realizó sus estudios universitarios en Medicina general en la Facultad de Medicina de la Universidad Autónoma de Nuevo León, recibiendo su título en el 2019. En marzo del 2020 comenzó sus estudios de Posgrado en Psiquiatría General en el Hospital Universitario 'Dr. José Eleuterio González', para concluir en febrero del 2024. En el cuarto año de posgrado realizó una rotación la Clínica de Salud Mental Humind en San Pedro Garza García, Nuevo León.

En el 2023 recibió la Certificación por el Consejo Mexicano de Psiquiatría. La tesis desarrollada para obtener el grado de Psiquiatría General fue "Procesamiento de Lenguaje Natural en Psiquiatría".