



Universidad Internacional de La Rioja

Facultad de Ciencias de la Salud

Máster Universitario en Investigación en Psicología
**Procesamiento del Lenguaje Natural (PLN) y
Salud Mental: Afinando un Modelo BERT
para Evaluar Ansiedad**

Trabajo de fin de estudio presentado por: Javier Vidal Marciel

Tipo de trabajo: Trabajo de Fin de Máster

Director: Jose Antonio García Arroyo

Fecha: 19 de julio de 2023

Resumen

El lenguaje es una importante vía de acceso a la mente humana, un cometido asumido teóricamente en disciplina de la psicolingüística. El desarrollo de los métodos computacionales de inteligencia artificial, específicamente el Procesamiento del Lenguaje Natural, ha permitido que surjan métodos objetivos para modelar el habla y la escritura. Las aplicaciones al campo psicológico pueden ser importantes tanto para la teoría como para la práctica clínica. Los sistemas más avanzados se basan en la tecnología de Transformadores (BERT o GPT). En este estudio se afina un modelo BERT en español para la discriminación según el nivel de ansiedad de transcripciones de entrevistas, en una tarea de clasificación supervisada, utilizando el cuestionario STAI como referencia. En un corpus generado por 11 mujeres de 20 a 24 años, se alcanzó un 60% de precisión para diferenciar dos mitades de la muestra. Pese a ser baja la precisión, se muestra el alto potencial de impacto de este tipo de herramientas para el diagnóstico automático en salud mental.

Abstract

Language is an important access tool into the human mind, an objective which has been theoretically grouped under psycholinguistics. The development of computational methods around artificial intelligence, specifically Natural Language Processing, has allowed the development of objective methods to model speech and writing. The applications towards the psychological field could be important to both theory and clinical practice. The most advanced NLP systems are built in the Transformers architecture (BERT or GPT). In this study, a BERT Spanish model is fine-tuned to differentiate between interview transcriptions according to anxiety levels, in a supervised classification task, using the STAI measure as reference. In a corpus generated by 11 women between 20 and 24 years-old, a 60% accuracy was reached when distinguishing two groups. Although a low accuracy, it is shown the high impact potential of these tools for automated diagnosis in mental health.

Índice de contenidos

| | |
|---|----|
| 1. Introducción..... | 1 |
| 1.1 Psicolingüística y Procesamiento del Lenguaje Natural..... | 1 |
| 1.1.1 División teórica y nacimiento de la IA (años 30 y 40)..... | 3 |
| 1.1.2 Primeros auspicios de la Revolución Cognitivo-Computacional (años 50 y 60)..... | 4 |
| 1.1.3 Nacimiento de la Revolución Cognitiva (años 70)..... | 7 |
| 1.1.4 Comercialización de la Computación y Revitalización del PLN (años 80 y 90)..... | 9 |
| 1.2 PLN y Salud Mental: Un estrechamiento del cisma cualitativo-cuantitativo..... | 14 |
| 1.3 Justificación del presente estudio..... | 18 |
| 2. Métodos..... | 20 |
| 2.1 Muestra..... | 20 |
| 2.2 Instrumentos..... | 21 |
| 2.3 Transformadores..... | 22 |
| 2.3.1 Arquitectura del modelo..... | 23 |
| 2.3.2 Modelos utilizados..... | 30 |
| 3. Resultados..... | 31 |
| 3.1 BETO..... | 32 |
| 3.2 ALBETO..... | 33 |
| 3.3 DistilBETO..... | 34 |
| 4. Discusión..... | 37 |
| Referencias bibliográficas..... | 42 |

Índice de figuras

| | |
|---|----|
| 1. Figura 1: Implementación del proceso de diagnóstico automático..... | 23 |
| 2. Figura 2: Estructura del codificador..... | 24 |
| 3. Figura 3: Ejemplo ilustrativo de la topografía creada en los Word Embeddings (WE)..... | 25 |
| 4. Figura 4: Arquitectura de cada capa del codificador..... | 26 |
| 5. Figura 5: Red neuronal artificial del modelo construido..... | 28 |
| 6. Figura 6: Arquitectura general del modelo de clasificación binaria en transformadores..... | 29 |
| 7. Figura 7: Curva ROC de BETO..... | 32 |
| 8. Figura 8: Curva ROC de AIBETO..... | 33 |
| 9. Figura 9: Curva ROC de DistilBETO..... | 34 |
| 10. Figura 10: Importancia atribuida por BETO en el Ejemplo 1..... | 35 |
| 11. Figura 11: Importancia atribuida por BETO en el Ejemplo 2..... | 36 |

Índice de tablas

| | |
|--|----|
| 1. Tabla 1: Matriz de confusión de BETO..... | 32 |
| 2. Tabla 2: Matriz de confusión de ALBETO..... | 33 |
| 3. Tabla 3: Matriz de confusión de DistilBETO..... | 34 |
| 4. Tabla 4: Comparación de los tres modelos de BERT..... | 35 |

1. Introducción

1.1 Psicolingüística y Procesamiento del Lenguaje Natural

El lenguaje, en su definición más básica, es un sistema de reglas lógicas concretas que permite la producción y entendimiento de objetos de significado de carácter infinito (Fromkin et al., 2003). De una u otra forma, domina la vida social y cognitiva del ser humano (Harley, 2014, Pennebaker et al., 2003). Su desarrollo ha estado siempre íntimamente relacionado con el de la disciplina psicológica, tanto en su ámbito teórico como en la práctica clínica.

Los cuestionamientos sobre el pensamiento y el lenguaje, las raíces de la psicología y la lingüística, pueden siempre remontarse a las primeras ideas filosóficas sobre el ser humano y su funcionamiento (Rieber y Vetter, 2013). Pero por mantener la concisión con la cuestión actual del trabajo, puede ser más sensato comenzar por las teorías que se han originado desde de la Revolución Científica. El estudio moderno del lenguaje se inicia durante la segunda mitad del siglo XVIII en Europa continental, principalmente en Francia y los Estados Alemanes. Surgiendo como muchas otras ciencias, en el epicentro de la transición de la Ilustración al Romanticismo (Levelt, 2013). La lingüística al inicio se trató como un asunto de la antropología filosófica, enfocada en arrojar luz sobre sus orígenes. Y en el cruce entre la empresa por explicar el mundo y colonizarlo, Europa tuvo durante el siglo XVIII y XIX acceso a una miríada de pueblos y culturas que habían permanecido en relativa impermeabilidad. Esta dinámica favoreció la posibilidad de estudiar múltiples lenguas, con recorridos hasta cierto punto independientes. Dando origen a una nutrida serie de trabajos sobre el origen histórico y antropológico de las culturas y sus lenguas (Burnett, 1773/2018; Herder, 1772/1966; Rousseau, 1781/1966). A la vez biólogos como J. Lamarck (1744-1829) desarrollaron tesis sobre la selección natural, que más tarde aparecerían en Darwin (1859), e hicieron del lenguaje un interés de la etnología evolutiva (Greenberg, 1948). Bajo la hipótesis de que diferentes lenguas pueden estar explicadas por distintas formas de pensar, emergieron las teorías alrededor del Darwinismo Social, que utilizaron el lenguaje para explicar cómo el avance de unas culturas sobre otras se enlazaba con la precariedad del pensamiento de los pueblos menos industrializados (Haller, 1971, Rogers, 1972). En parte sirvió para racionalizar la colonización europea del mundo, y fue evidentemente refutado por la biología contemporánea. Pero el caso es interesante, pues aporta un temprano ejemplo sobre cómo esta proto-ciencia lingüística se utilizó

para realizar inferencias psicológicas, acerca de la inteligencia. La lingüística como disciplina académica propia, escindida de la filosofía antropológica tiene su inicio a finales del siglo XVIII (Levelt, 2013).

Durante el siglo XIX, autores como J.F. Herbart (1776-1841), H. Steinhal (1823-1899) o R.C. Rask (1787-1832), continuaron preguntándose por las relaciones entre el lenguaje, el pensamiento y la cultura (Duignan, 2009; Kim, 2015; Levelt, 2013), coincidiendo con el nacimiento de la psicología experimental. En este aspecto, se rescata la importancia de Wilhem Von Humboldt (1776-1835), quien formula la hipótesis de la concepción de mundo, o *darstellung*. Según la cual, el lenguaje determina y es absolutamente necesario para la existencia de pensamiento (Galán, 1994; Levelt, 2013; Mills, 2000). También Wittgenstein (1922/1998), afirmó: “Los límites de mi lenguaje son los límites de mi mundo”. Esta postura se formulará más adelante como el determinismo lingüístico.

Avanzado el siglo XIX, el mundo científico vivió una fuerte expansión y diversificación en múltiples disciplinas (Delmas et al., 2010). Por primera vez, la lingüística se dividió en dos ramas:

1. El estudio formal y lógico de las dimensiones del lenguaje (gramática, morfología, semántica o léxica), vinculado al movimiento positivista y neopositivista. Muchos de los autores del notorio Círculo de Viena emplearon el análisis lingüístico de las hipótesis, enunciados y proposiciones para llevar a cabo su empresa epistémica. Unificar la ciencia bajo una frontera concisa entre la verdad analítica y sintética y el conocimiento pseudocientífico (Da Cunha, 2013; Perlovsky, 2009).
2. Otro abordaje enlazado con el estudio del ser humano, caracterizado por métodos más cualitativos y bautizado como psicolingüística (MacWhinney, 2015). El mismo W. Wundt ha sido considerado por algunos (Blumenthal, 1987) como fundador de esta disciplina ecléctica. También L. Vigotsky (1896-1934) por otros (Omoyeva, 2021), gracias a sus teorías sobre la adquisición del lenguaje en la infancia. Estas nociones pronto fueron llevadas a la práctica de las primeras terapias psicodinámicas, como fue el caso de K.G. Jung (1871-1961) y su método de la asociación libre (Jung, 1910).

Aunque psicología y lenguaje quedaron inevitablemente unidos desde finales del siglo XIX y principios del XX, el término es originalmente acuñado por Osgood et al. en 1954. Las teorías que asumían la influencia del lenguaje sobre el pensamiento se agruparon en dos líneas. El relativismo

lingüístico, que lo concibe como un soporte optimizador de la percepción y la cognición, y en el que destaca la hipótesis de Sapir-Whorf (Ellingsworth, 1992, Lucy, 2001). Y el determinismo lingüístico, una propuesta más extrema, defiende que no podemos conocer aquello que no podemos formular verbalmente. Los hablantes de cada idioma tendrían por lo tanto accesos perceptuales diferenciados a la realidad (Zahedi, 2008). Aquellos entendimientos que colocaron el lenguaje como el medio necesario para el pensamiento, fueron más propensos a compartir que la capacidad lingüística es innata y exclusiva al ser humano. Otros de carácter conductual, rechazarían más adelante estas ideas.

Hallamos entonces, una disciplina que nace enormemente dividida, entre el enfoque matemático-formal y el abordaje psicosocial del lenguaje. Dos planteamientos con métodos y filosofías del conocimiento opuestas, en un momento de fuerte discusión acerca de la verdad absoluta y el camino epistémico hacia ella. Durante el siglo XX, encontramos cuatro grandes etapas en lo que concierne al campo psicolingüístico:

1.1.1 División teórica y nacimiento de la IA (años 30 y 40)

Un primer momento marcado por una profunda escisión teórica y metodológica, y a la vez, caracterizado por un cometido común: el estudio del lenguaje como vía para entender la mente. Por un lado, el psicoanálisis mantuvo una fuerte hegemonía como paradigma psicológico del ser humano. Entendiendo el lenguaje como un código que contiene información encriptada del inconsciente, y por lo tanto debe ser descifrado por métodos interpretativos e introspectivos. En gran medida, este modelo fue un notorio precursor de que la terapia se configurase como un intercambio verbal, en el que el profesional accede a la vida interior de su paciente a gracias al diálogo clínico. Sin embargo, aquellos lingüistas formales puros, descendientes de un método logicista y neopositivista, encontraron escasa satisfacción con una metodología que se les hacía oscurantista y acientífica. Un notorio caso fue el de Allport (1942), que estimó el lenguaje como una herramienta predilecta de acceso al funcionamiento mental, pero se lamentaba por la precariedad metodológica del campo en estos años. Síntoma de esta insatisfacción, sería la llegada de otro paradigma con tesis radicalmente opuestas, como fue el conductismo clásico y operante.

De manera paralela, y en un área con otra pretensión, se estuvieron realizando avances en computación que cambiarían la psicolingüística; y realmente todas las ciencias. Se inició una

convergencia entre la biología (Didier y Bigand, 2011), la matemática (Schmidhuber, 2022) y la lógica epistemológica (Burnham, 1888) donde nació la Inteligencia Artificial (IA). Que, en su formulación más básica, es un planteamiento para modelar la mente de manera formal y computacional. Definiéndola mediante algoritmos y estructuras lógicas, que sirven de base a sistemas artificiales de procesamiento de la información. Los primeros conocimientos sobre el cerebro y la neurona, por S. Ramón y Cajal (1852-1934) y C. Golgi (1843-1926) (De Carlos y Borrell, 2007), inspiraron la concepción de las redes neuronales. La noción de que la mente se sostenía sobre un sistema de elementos eléctricos, interconectados a través de *inputs* y *outputs* dicotómicos, encajó perfectamente con la teoría de sistemas del momento. Pronto se enlazó esta visión del cerebro con el funcionamiento de la mente (Wang y Raj, 2017). Un modelo fundamental fue el formulado por Hebb (1949), considerado por algunos (Didier y Bigand, 2011) como padre de las redes neuronales. Hipotetizó la *Regla del Aprendizaje Hebbiana*: cuando la neurona A, sinapta persistentemente sobre la neurona B, ocurre un proceso de crecimiento metabólico que refuerza la eficiencia y predisposición para que sea A, la neurona que frecuentemente sinapta sobre B. Describió el proceso de aprendizaje y condicionamiento a nivel neuronal (Wang y Raj, 2017).

Más tarde, surgieron los primeros sistemas eléctricos capaces de modelar, de manera primitiva una red neuronal. La primera idea que se asocia a esta cuestión es la de Lenz-Ising (Brush, 1967; Niss 2004). Aunque los autores lo desarrollaron en un inicio para representar fenómenos electromagnéticos, más adelante se utilizó para describir formalmente un primer boceto de IA. En 1943, McCulloch y Pitts diseñaron la *Neurona McCulloch-Pitts* (MCP) con circuitos eléctricos.

Sin embargo, el autor que más se asocia con el nacimiento de la IA es A. Turing (1912-1954), que durante la II Guerra Mundial puso en práctica uno de los acercamientos iniciales al ordenador actual a través de *"The Bombe"*, con el objetivo de descifrar la máquina nazi *"Enigma"*. Turing (1950) plasmó sus avances en una icónica publicación y planteó la pregunta clave: *"¿Pueden las máquinas pensar?"*. Propuso su famoso método para la comprobación de este interrogante, el *Test de Turing*.

1.1.2 Primeros auspicios de la Revolución Cognitivo-Computacional (años 50 y 60)

Estas fueron décadas decisivas para definir la forma en la que actualmente entendemos la psicolingüística. Ocurrieron una serie de discusiones fundamentales y avances tecnológicos que facilitaron los primeros rastros de un marco común, solventando las diferencias que existieron

durante los años 30 y 40. Permitió que aquellos psicólogos que buscaban en el lenguaje un medio cualitativamente rico, pero científico de acceso al pensamiento, y los lingüistas lógicos que demandaban un análisis del contenido riguroso se encontrasen. La fusión entre psicolingüística y procesamiento computacional del lenguaje natural quedó fundada.

Primero, tuvieron lugar en estos años intensos choques acerca de la existencia del lenguaje en el ser humano, y los procesos por los que se adquiere universalmente (Cuoto, 2002). Discusiones que, en gran medida, recuerdan a choques cíclicos e irresueltos entre el empirismo y racionalismo. Durante los años 50, el conductismo reemplazó al psicoanálisis como teoría. Propuso una solución sumamente científica y neopositivista que, sin embargo, no ofrecía un mejor techo bajo el que agrupar a psicólogos y lingüistas. La negligencia de los contenidos mentales que no pueden observarse y medirse empíricamente, impedía el auténtico uso del lenguaje como medio de acceso al pensamiento (Fuchs y Milar, 2003). Siendo precisamente este, el cometido de la psicolingüística. La hipótesis conductista del lenguaje icónicamente se plasmó en la obra "*Verbal Behaviour*" de Skinner (1957). Tomando ideas que rescata de J. Locke (1632-1704), como el asociacionismo o la *tabula rasa*, defendió que la adquisición del lenguaje obedece un proceso de condicionamiento operante. La principal oposición fue ofrecida por Chomsky (1959), que al respecto argumentó que el lenguaje es un sistema demasiado complejo para plasmarse en términos tan reduccionistas. Llegó a proponer la existencia de estructuras cerebrales innatas (Dispositivo de Adquisición del Lenguaje, DAL) exclusivamente dedicadas a procesarlo y generarlo (Primero, 2008; Zahedi, 2008).

Siguiendo diálogo recíproco entre la teoría del lenguaje y el desarrollo computacional de la IA, durante los años 50 emergieron modelos más sofisticados (Newell et al., 1958; Rosenblatt, 1958). Este momento tras la II Guerra Mundial, en el epicentro del apogeo tecnológico y científico de los Estados Unidos, estuvo caracterizado por un gran optimismo. Se vaticinó que, en pocos años, existirían máquinas indistinguibles de la inteligencia humana. En 1956, M. Minsky y J. McCarthy, acuñan el término "inteligencia artificial" en una conferencia en Darmouth College (EE. UU.) (Hahlein y Kaplan, 2019). Entre los científicos que asistieron se encontraba el fundador de la teoría de la información (Shannon, 1948) o N. Rochester, que diseñaría el primer ordenador científico (IBM 701). Este impulso y energía se mantuvo durante los años 60, dando lugar a los primeros sistemas de Procesamiento del Lenguaje Natural (PLN), basados en sistemas de *Information Retrieval* (IR) (Manning et al., 2009). Destacan ELIZA, el primer programa capaz de simular el

lenguaje humano (Weizenbaum, 1966), y SHRDLU, un programa para interactuar con un ecosistema virtual a través de frases (Winograd, 1971). También el Experimento de Georgetown-IBM, un programa capaz de traducir sesenta frases del ruso al inglés (Hutchins, 2004). Durante estos años la IA estuvo basada en los llamados *Enfoques Basados en Reglas* o *Sistemas Expertos*, que empleaban estructuras muy robustas y con poco espacio para el aprendizaje. Dependían de una enorme cantidad de reglas preprogramadas por expertos en el área al que pretendiesen ser aplicados. Por ejemplo, en los árboles de decisión (Morgan y Sonquist, 1963), la eficacia dependía totalmente de la selección y adecuada implementación de variables predictoras, que capturasen los patrones y pudiesen por lo tanto realizar extrapolaciones. Rosenblatt (1962), publicó un importante libro (*“Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms”*) que sella el nacimiento del *Deep Learning* (DL). Aunque Ivachenko y Lapa (1967) fueron los primeros en implementarlo.

En el contexto del PLN, los sistemas computacionales únicamente eran capaces de manejar información gramática sin significado ni contexto (Nadkarni et al., 2011). Para ello, se desarrollaron los *Lenguaje Libres de Contexto*, como UNIX (Kleene, 1956). Éstos probaron ser inadecuados para trabajar con lenguaje natural humano directamente, que es especialmente simbólico, significativo y contextual. Los primeros lenguajes de programación, FORTRAN (Backus et al., 1954) y LISP (McCarthy, 1960), son coetáneos. Un caso icónico es el *“General Inquirer”*, desarrollado por Stone et al., en 1961 sobre un IBM 7090 en la Universidad de Harvard (Stone et al., 1963). Trajo al mundo el análisis estadístico de contenido lingüístico. Era capaz de construir frecuencias relativas de ciertas palabras en un texto, previamente programadas en su librería interna y correlacionar tales métricas con variables psicológicas (Psathas, 1969).

Todos estos avances en IA, PLN y sistemas informacionales; otorgaron a la psicolingüística un nuevo paradigma para analizar el lenguaje e inferir sobre el pensamiento. El descubrimiento de técnicas que hacían posible extraer propiedades del texto y el habla, para traducirlas en modelos cuantitativos, supuso un primer acercamiento a un estudio objetivo y operativo en psicolingüística (Jelinek y Mercer, 1980; Kennedy et al., 2022, Winograd, 1972).

Sin embargo, en el cambio a los años 70, se experimentó un declive en el optimismo y la energía que el campo de la IA acarreaba desde los 50. Varias razones justifican este repentino giro de los acontecimientos:

1. La falta de poder computacional, demostrada por Minsky y Papert en 1969, el coste de mantenimiento de los primeros ordenadores y la falta de aplicabilidad inicial a problemas prácticos, fueron factores determinantes. En los años 70, llegó además una grave crisis económica global, generada por la especulación reivindicativa del petróleo por parte de la OPEP (Organización de Países Exportadores de Petróleo). Todavía sin la llegada masificada de tecnologías de la información que años después ocurriría, el volumen de datos que se capturaba de la realidad socioeconómica era muy escaso. Contando con que la IA y los sistemas de computación nutren su utilidad de información masiva, las ventajas que proponían no eran rentables o atractivas para las empresas del momento. Pues debía superar unos altísimos costes de mantenimiento y un personal altamente cualificado, que entonces era escaso.
2. El ordenador era una máquina bastante poco común en el mundo académico, y especialmente en las ciencias sociales. De nuevo, la falta de sistemas que viablemente pudiera poseer u operar una persona o un departamento hacía muy difícil que su uso se extendiera. Además, las lenguas de programación sólo habían sido concebidas 15 años antes, y muy pocas personas fuera de contextos matemáticos y de ingeniería sabían manejarlas. Incluso, hoy en día se sigue hablando de una brecha tecnológica generacional. El optimismo respecto a la computación y a la IA quedó más bien encapsulado en pequeños núcleos académicos, y no se produjo la adopción vaticinada de manera tan precipitada.

Todo ello, contribuyó a un fuerte declive durante la década de los 70 en Occidente, titulado “el invierno de la IA” (Haenlein y Kaplan, 2019; Kaul et al., 2020). Un caso atípico fue el de Japón, que durante estos años dominó el mercado electrónico y digital, entrando en la siguiente etapa (años 80-90) con mayor inercia industrial.

1.1.3 Revolución Cognitiva (años 70)

Durante los años 70, si bien la IA y la computación frenaron su expansión, la psicología vivió una fuerte revolución y expansión sobre sus dos pilares:

1. Psicología básica o académica: tuvo lugar la expansión de la Psicología Cognitiva, como nuevo paradigma teórico y científico. Aunque muchas de sus tesis se habían ido desarrollando en los

50, fue durante estos años los que realmente empaparon la disciplina. Halló un campo psicológico en gran parte insatisfecho con el conductismo radical y reduccionista, que había dominado la academia desde mediados de los 50 (Fuchs y Milar, 2003). Fue un cambio que se gestó de ideas que no eran nuevas. En Europa, la Gestalt ya había desarrollado sus tesis sobre la percepción (Köhler, 1929; Koffka, 1935/2013; Wertheimer, 1923) y Piaget (1929), tenía desde hace tiempo parte de sus trascendentales teorías sobre el desarrollo cognitivo. La cibernética (Wiener, 1965) y la teoría de la información (Attneave, 1959; Hick, 1952), que posteriormente rescataría la sistémica, ganaron popularidad años después de su concepción original. Éstas últimas, influenciadas por la revolución tecnológica y teórica en computación comentada antes, extendieron una nueva concepción del ser humano. Fuertemente caracterizada por analogías entre la mente y los ordenadores, y por nociones del cerebro como un sistema de procesamiento de la información (Fuchs y Milar, 2003).

2. Psicología aplicada o clínica: el modelo terapéutico vivió una profunda remodelación. Principalmente en lo referente a la figura del psicólogo y los objetivos de la intervención clínica. Se motivó una transición del modelo clínico desde un planteamiento mecanicista y patológico a uno humanista y más generalizado. Como autores principales encontramos a Rogers (1942) o Maslow (1968), aunque hasta mediados de los 70 no fueron consistentemente reconocidos.

El cognitivismo se expandió gracias a su éxito científico para diseñar y probar la existencia de procesos mentales superiores. Para hacer emerger una aproximación al ser humano objetiva y neopositivista, sin caer necesariamente en el reduccionismo. Un equilibrio más sostenible entre lo cualitativo y lo cuantitativo. La mayoría de los científicos comportamentales se integraron en el nuevo campo, añadiendo los procesos mentales como mediadores entre los estímulos y la conducta. Dando lugar, a este tándem tan extendido en el actual panorama científico y terapéutico, el Cognitivo-Conductual. Durante principios de los años 70, motivados por las ideas de Skinner que rechazan la cualidad innata del lenguaje y su exclusividad humana, un gran volumen de autores intentó implementar capacidades lingüísticas en animales (Kellogg, 1968; Gardner y Gardner, 1969; Premack, 1971; Hayes y Nissen, 1971; Miles, 1978; Patterson, 1978). Esta empresa acabó colapsando por la mediatización y comercialización de sus progresos, que llevó a una grave pérdida del rigor (Pepperberg, 2017). En parte, siendo un acto de claudicación simbólico de las tesis de Skinner sobre la adquisición del lenguaje.

1.1.4 Comercialización de la Computación y Revitalización del PLN (años 80 y 90)

Finalmente, el desarrollo tecnológico trajo al mundo los ordenadores personales (*Personal Computers*, PC). IBM introdujo en el mercado de 1975 el IBM 5100 por 1,500\$, Apple el modelo Apple II en 1977 por 1,300\$ y el Macintosh en 1984, costando 2,500\$. Por primera vez desde la invención de la computación electrónica, la posesión masificada de estas máquinas era factible. Además, se introdujeron actualizaciones en el diseño de usuario (los sistemas operativos), que hacían posible poder manejar un ordenador sin saber lenguajes de programación. Si a mediados de siglo mantener un terabyte de datos costaba 9.2 billones de dólares, en 2022 costaba 14,3\$ (McCallum, 2022). Todo ello, ha posibilitado la exponencial y vertiginosa invasión de los ordenadores en prácticamente todos los ámbitos de la vida en pocas décadas.

Al mismo tiempo, la Psicología Cognitiva se hallaba en pleno auge (Miller, 2003). Se rescató un mentalismo renovado (Spino y Loewen, 2018), que secundaba las hipótesis de Chomsky sobre la cualidad innata del lenguaje. Por otro lado, los antiguos conductistas acogieron las tesis del referencialismo (Echeverri, 2014), que propuso un marco teórico para la adquisición del lenguaje donde los conceptos de mente se catalogan a través del aprendizaje de palabras. Favoreciendo la idea de la adquisición del lenguaje mediante experiencias, más cercano a lo defendido por el condicionamiento operante. Con el tiempo, la evidencia científica ha favorecido más a las tesis chomskianas y mentalistas (Anastasiy y Cyprien, 2021; Fitch, 2019).

Los mayores avances se vivieron al converger el cognitivismo con nuevas técnicas de neuroimagen, potenciadas por el persistente progreso tecnológico, especialmente la imagen por resonancia magnética funcional (Ogawa et al., 1992). Las más importantes teorías de la mente fueron el conexionismo (Campanario, 2004) y el modularismo (Mesulam et al., 2021), que otorgaron aún más reconocimiento científico a la psicología a través del establecimiento de correlatos anatómicos cerebrales. Durante los años 80 y 90, nacieron la neurociencia y la neuropsicología, a través de autores como Fodor (1983), Geschwind (1985) interesado en la dislexia y otras patologías neurológicas o Rumelhart et al. (1986) que formularon el modelo del *Procesamiento Distribuido en Paralelo* (PDP).

La IA, pese a seguir viviendo su invierno académico, siguió progresando en la retaguardia. Especialmente en Japón con figuras como Amari (1972), que creó la red neuronal recurrente (RNN,

en inglés) o Fukushima (1980) y Zhang (1990) que diseñaron la primera red neuronal convolucional (CNN). En Estados Unidos, se produce mucha bibliografía entorno a la visión computarizada (Le Cun y Bengio, 1995) y el *Deep Learning* (Le Cun et al., 2015). Muchos de estos autores, permanecieron en la sombra académica durante los 90 y están emergiendo hoy en día, cuando realmente se presencia la expansión práctica de la IA.

El Procesamiento del Lenguaje Natural (PLN) se asienta como un espacio teórico con más reconocimiento y utilidad para la psicolingüística durante estos años. En gran medida, debido al desarrollo que se ha venido narrando en ciencia cognitiva, computación e IA. En los años 80 y 90 encontramos tres giros metodológicos fundamentales (Nadkarni et al., 2011):

1. La psicolingüística adoptó un espíritu absolutamente abierto al tratamiento matemático de la información de gran volumen. Aquello que no había sido posible durante los años 40 y 50 por la precariedad tecnológica y la ausencia de información electrónica, se materializó durante los 80 y 90 a través de técnicas probabilísticas y de algoritmos de *Machine Learning* (ML) (Kennedy et al., 2022). Surgió el concepto del *Knowledge Discovery in Databases* (KDD) y la metodología del *Data Mining* (DM). Estas perspectivas, que tuvieron un contundente impacto en todos los sectores académicos e industriales, representan una nueva forma de obtener información de grandes cantidades de datos de baja escala (Fayyad et al., 1996; Maimon y Rokach, 2019). De hecho, el número de información que se recopila de nuestra realidad física y social no ha dejado de crecer. Hasta el punto de que ha dado lugar a nuevas metodologías de alto impacto (*Big Data*) y ha permitido que cuatro de las cinco empresas con mayor volumen de mercado en 2023 basen sus modelos de negocio total o parcialmente en el manejo y tratamiento de la información (Apple, Microsoft, Google y Amazon). Esta aproximación fuertemente apoyada sobre la extracción de conocimiento a través de los datos (*Data Driven Approach*), favoreció que los modelos estadísticos del lenguaje fuesen mucho más sofisticados, convirtiéndose en el estándar metodológico de la psicolingüística analítica (Han et al., 2012).
2. El desarrollo del ML y los modelos estadísticos probabilísticos, agrupados alrededor de la teoría del Aprendizaje Explicativo (Explained-based Learning, Minton et al., 1989) resolvieron varios de los problemas que habían encontrado los ya mencionados Sistemas Basado en Reglas (General Inquirer, ELIZA, SHRDLU, Georgetown-IBM...). Entre otras cosas posibilitaron que se reemplazasen las Gramáticas Libres de Contexto (GLCs), y se lograra capturar la dimensión léxica del lenguaje. Durante estos años se diseñan los "*Hidden Markov Models*" (Huang et al.,

1990), los “*Conditional Random Fields*” (Lafferty et al., 2001) o los “*Naive Bayes Classifiers*” (McCallum y Nigam, 2001). A grandes rasgos, hicieron posible construir modelos del lenguaje mucho más cercanos al que los seres humanos empleamos de manera natural. Por lo tanto, tareas clásicas en PLN como clasificar textos o predecir frases a partir de un input, se convirtieron en procesos mucho más válidos y aplicables a problemas reales.

3. Por último, se hizo necesaria la construcción de grandes bases de textos, llamadas *corpus*. El volumen de datos requerido para entrenar los engranajes de estas nuevas técnicas de PLN, que aún estaban poco optimizadas y eran “menos inteligentes”, se convirtió en un activo científico de suma importancia. Además, agrupar cantidades masivas de texto se hizo más productivo si éste contenía categorías y marcadores que lo hicieran distintivo (*annotated corpora*), pues optimizaba el ajuste de algoritmos para encontrar patrones y predecir con mayor eficiencia y agudeza. Este conjunto de nuevas aproximaciones encuentra una productiva materialización en la metodología del “*Corpus Linguistics*”, que básicamente es una herramienta del *Data Mining*. Glynn y Fischer (2010) o Gries (2017) aportaron nutridas evidencias a esta rama de la psicolingüística estadística, que sigue aun generando conocimiento (Girard et al., 2022). Su cometido general, es extraer características de un texto, como su longitud o la cantidad de verbos de una particular naturaleza, para generar correlaciones con rasgos detectados en el emisor.

Encontramos las primeras aplicaciones del PLN a la psicología en estas décadas de los 80 y 90 (Pauselli et al., 2018). La invención del LIWC (*Linguistic Inquiry and Word Count*) por Pennebaker y Francis en 1999, el primer software accesible y manejable por profesionales en ciencias sociales, favoreció mucho esta interacción (Boyd y Schwartz, 2021). Por ejemplo, Manschreck et al., (1981) o Morice e Ingram (1982) estudiaron el rango de repetición y frecuencia de errores verbales en pacientes con esquizofrenia. Otros estudios posteriores de PLN en psicosis se pueden encontrar en Fraser et al., 1986 o Hoffman y Sledge, 1988.

Durante las dos últimas décadas del siglo XXI, el crecimiento y diversificación de la Inteligencia Artificial y el PLN ha llegado a tal punto, que escapa el alcance de este estudio describir de manera consistente y detallada el avance en todos sus subcampos. Sí podemos describir cuatro tendencias globales que facilitan comprender el estado en el que se encuentra la cuestión (Lauriola et al., 2022; Nadkarni et al., 2011; Sun et al., 2021). También permiten entender, cómo en cuestión de tres

décadas, las técnicas de IA hayan absorbido el PLN por completo, desplazando otros modelos anteriores.

1. La transición de los Sistemas Basados en Reglas a los Modelos Estadísticos de IA: como se ha mencionado ya, la década de los 90 fue testigo de la desaparición de los modelos rígidos, con variables independientes preestablecidas y exhaustivos programas que regulaban cualquier acción. Fueron reemplazados por sistemas estadísticos probabilísticos, fundamentalmente potenciados por métodos de ML, capaces de actualizar sus parámetros autónomamente. La transición ha hecho posible modelar los aspectos del lenguaje más subjetivos (semántica y contexto), más allá de la gramática, la morfología o la sintaxis.
2. Introducción de los *Word Embeddings* (WE): En 2013, cuatro científicos de Google (Mikolov et al., 2013) crean *Word2Vec*, y en 2014 otro modelo similar, *GloVe*, se diseña en Stanford (Pennington et al., 2014). Ambos, son métodos para traducir palabras a vectores manteniendo información contextual ("*word embeddings*"). De manera intuitiva, una red neuronal es entrenada en un gran *corpus* de palabras y atiende al orden que éstas siguen. Es decir, aprende a establecer qué tipo de palabras van encadenadas y construye probabilidades de que, por ejemplo, la palabra "*plátano*" sea seguida por "*amarillo*". De esta manera, construye vectores en base a la distancia topográfica que existe entre ambas palabras, determinada por la probabilidad de ser encontradas en la misma secuencia. Ello permite capturar parte de la información semántica de la palabra y también hace posible manipular sus representaciones de forma matemática. Por ejemplo, restarle al vector "*Rey*" el vector "*Hombre*", suele resultar en un vector parecido al de "*Reina*". Tradicionalmente, la sintaxis, la gramática y la morfología fue lo primero que se consiguió modelar en PLN, gracias a su naturaleza especialmente formal. En cambio, la semántica y el contexto ha sido siempre un arduo problema, y fue de hecho en él que los Sistemas Basados en Reglas fracasaron. Nuestro lenguaje humano, es bastante sofisticadas (aunque varía según el idioma) y sigue unas reglas semánticas altamente complicadas. Más aún, los elementos que rodean a una palabra pueden determinar lo que significa más que la misma morfología de la palabra. Por ejemplo, "Pablo es una torre comparado con sus amigos", supone un grave problema de semántica para un modelo de PLN. Esencialmente porque "*torre*", no se relaciona completamente con el sustantivo con el que explícitamente está asociado (ie. Una edificación vertical cilíndrica con una base por lo general más reducida espacialmente que su altura). En esta frase, "*torre*", es un sustantivo cuyo significado se desgrana para simplemente emplear uno de los adjetivos (altura) que suelen

atribuírsele a este tipo de edificios. En el ejemplo, vemos como las palabras que envuelven a “torre” definen más su significado que aquel que pudiéramos encontrar en el diccionario. La gramática y el vocabulario que se programa en un modelo de PLN tradicional jamás podría capturar este tipo de significado, incluyendo cada palabra que usamos en una librería programada, porque la cantidad de sentidos con el que puede manejarse un sustantivo es sencillamente infinita. Por eso los modelos probabilísticos, que no atienden a la palabra si no a las concatenaciones entre ellas, han tenido superiores éxitos para capturar la semántica del lenguaje natural. Es una solución que asume que modelar el contexto, permite cierto grado de comprensión semántica. Otro problema, más bien derivado del contexto y la pragmática lingüística, podría ser la ironía. Donde una frase formalmente coherente, donde cada palabra adquiere significados perfectamente concretos, tiene una intención hacia el sujeto opuesta a la que aparentemente encierra. Estos problemas de la comprensión no han sido del todo resueltos, la IA no puede entender la ironía tal y como la concebimos nosotros, pero es muy certera en capturar la estructura probabilística que estas problemáticas construcciones tienen. Con el suficiente entrenamiento, se ha logrado que las anomalías de este tipo perjudiquen lo menos posible a la precisión en la tarea en cuestión.

3. La llegada y auge del Deep Learning: el salto de los algoritmos de ML a la creación de redes neuronales de varias capas transformó la IA en general, y con ello el PLN. La introducción de arquitecturas más refinadas como las *Long-Short-Term Memory Networks* (LSTM) o *Gated Recurrent Units* (GRUs), permite capturar patrones en los datos de manera mucho más sofisticada. Las LSTM (Hochreiter y Schmidhuber, 1997), por ejemplificarlas superficialmente, al fluir cíclicamente los datos del input por su estructura son capaces de retener selectivamente parte de la información de cada iteración (memoria a corto plazo), para actualizarla con los patrones de las siguientes secuencias (memoria a largo plazo). Esto permite que el aprendizaje sobre el conjunto de datos sobre los que se entrena sea un ejercicio mucho más complejo, y hace posible que se extraigan las variables independientes que dan forma a los datos, de una manera mucho más rica y potente. Hace factible, que el modelo sea mucho más “inteligente y autónomo”. Tanto, que en ocasiones se dificulta la comprensión de la lógica seguida en su proceso de aprendizaje, y provocando que llamemos a parte de su estructura “estados ocultos o escondidos” (*hidden states*).
4. Por último, y para fundamentar la siguiente sección del texto, se resalta la importancia que han tenido los modelos de Transformadores y el aprendizaje transferido. En este tipo de abordaje

(*transfer learning*), se entrena un modelo sobre un inmenso *corpus* genérico, para permitir que capture las probabilidades que le aseguran “capturar” la semántica y el contexto de un idioma (*pretrained model*). A posteriori, esta misma IA se puede re-entrenar en nuevas tareas de PLN, que requieren infinitamente menos datos y en las cuales se utiliza el aprendizaje acumulado en el modelo pre-entrenado y se incorpora la nueva información. Este desarrollo tecnológico en la computación lingüística ha supuesto un importantísimo cambio para la adopción y aplicabilidad de estas tecnologías. Hace unos pocos años, entrenar modelos del lenguaje desde el inicio hacía necesario contar con gigantes bases de datos y costosos equipos de ordenadores. Actualmente, cualquier persona con conexión a internet puede aplicar un modelo del lenguaje pre-entrenado en grandes laboratorios a su particular interés, necesitando bases de datos usuales para los que se emplean en cualquier tarea estadística tradicional. Este es el tipo de método que se emplea en este estudio.

Como se puede concluir, la inteligencia artificial ha permitido generar propuestas de PLN más inteligentes y, que, de manera más fiel, capturan el intrincado lenguaje humano. A continuación, se presentan algunas de las direcciones en las que los últimos avances del PLN se han aplicado en psicología.

1.2 PLN y Salud Mental: Un estrechamiento del cisma cualitativo-cuantitativo

Prácticamente desde que la lingüística y la psicología nacieron, pocos años tardaron en enlazarse y converger. El lenguaje es la principal vía humana que poseemos para exteriorizar o percibir contenido cognitivo en la interacción social. Sin embargo, la psicología tiene la particularidad de no poder acceder directamente a su objeto de estudio: la mente. (Tausczik y Pennebaker, 2010). Por ello, depende del lenguaje para inferir los fenómenos emocionales, memorísticos, perceptuales, conscientes e inconscientes que ocurren en el cerebro y que determinan el comportamiento. Ciertamente es que existen otros canales, como la observación de la conducta o la medición de efectos fisiológicos. Sin embargo, si bien esta aproximación se emplea ocasionalmente en psicología experimental, la rama clínica o aplicada de nuestro campo tiene el cometido adicional de pretender ocasionar cambios en la persona. Debido a esta doble orientación de la disciplina, de explicar y también de generar procesos, el lenguaje se asienta como la principal herramienta del psicólogo. De hecho, así se manifiesta en todos los paradigmas y teorías terapéuticos, al emplear la entrevista

clínica como base para la navegación y orientación clínica (Cain et al., 2016; Cooper et al., 2014; Shapless y Barber, 2009). Es algo que también se halla en la psicometría, la metodología que tradicionalmente más se ha usado para medir constructos en psicología. Los tests, esencialmente operativizan el lenguaje a través de frases preestablecidas (ítems) (Buchanan y Finch, 2005). Otras técnicas clásicas han empleado el lenguaje para operativizar el contenido mental fuera de la psicometría. El Test de Rorschach (1921), por ejemplo, se basa en la reacción verbalizada de los participantes a estímulos abstractos y ambiguos, y toda su sistematización se realiza sobre el texto generado.

Por lo tanto, la abstracción de los aspectos, contenidos y atributos del lenguaje a un formato estructurado, sistematizado y lógicamente homogéneo siempre ha sido el principal esfuerzo de la psicología por hacerse ciencia. Aunque este proceso acarrea profundas divisiones que siguen presentes. El principal problema, reside en la pérdida de información subjetiva e idiosincrásica que conlleva la conversión del testimonio a datos numéricos. Pero, al mismo tiempo, lo imposible que resulta comparar y medir las verbalizaciones naturales de dos sujetos diferentes de manera objetiva y científica. De hecho, las ciencias sociales se dividen en métodos cualitativos y cuantitativos por la existencia de este mismo problema (Ercikan y Roth, 2006; Griffin y Phoenix, 1994). Cualquier disciplina que tome al ser humano como objeto de estudio, se encuentra permanentemente en la encrucijada entre dos formas de aproximarse (American Psychological Association, 2023; Englander, 2016; Shaffer y Serlin, 2004).

1. Emplear un abordaje cualitativo, que permite capturar la identidad, el contexto y la expresión natural y subjetiva de la persona, pero que hace imposible tratar la información de manera científica.
2. Acercarse con un método cuantitativo y operativo, que traduce los fenómenos en piezas numéricas manipulables, comparables y extrapolables, pero que exigen una enorme reducción del testimonio o el autorreporte.

Por lo general, la academia psicológica ha encontrado fructífera la convivencia de ambos enfoques y más productivo aún, el uso complementario de los dos. Sin embargo, este es siempre un ejercicio que se realiza en tensión, pues la información de uno no es fácilmente traducible a la del otro. Este problema impacta claramente en la rama aplicada de la psicología, aquella que funciona en el ámbito clínico y terapéutico. Los métodos más utilizados para formular un caso en salud mental,

para elaborar un planteamiento sobre la etiología de cierta psicopatología, son esencialmente cualitativos (Calixto et al., 2022; Jackson et al., 2016). Pese a la existencia de herramientas psicométricas, normalmente se han diseñado para un uso complementario y no exclusivo en el proceso diagnóstico (Borsboom, 2008; American Educational Research Association et al., 2014). Éste se realiza a través de la entrevista clínica, que es la base para que el profesional pueda construir una idea sobre qué temáticas abordar en terapia y qué estrategias seguir (Haynes et al., 1995). Enfocada en la detección de síntomas en las verbalizaciones del paciente, tiene algunas dificultades asociadas:

1. Siempre hay cierto grado de subjetividad involucrado. Primero en la forma única en la que el paciente externaliza su vida mental interior. Segundo, en cómo el juicio clínico del profesional interpreta estos contenidos.
2. Es costosa en tiempo y dinero, pues el grado en el que el paciente comparte una narrativa del problema que lleva a terapia normalmente depende de dos factores: el vínculo y confianza con el profesional (Safran y Wallner, 1991; Simpson y Reid, 2014) y la propia consciencia del problema (Richards et al., 2010; Rickleman, 2009). Ambos son procesos que, por lo general, requieren de un número determinado de sesiones.

Para estos problemas, esencialmente originados por la complejidad del lenguaje y las relaciones humanas, el PLN podría aportar ciertas ventajas:

1. El procesamiento del lenguaje natural es automático. Una vez se programa el modelo no requiere de capital humano para funcionar. De esta forma, puede extraer conclusiones que dependen únicamente de la cantidad y variedad de texto con el que la IA haya sido entrenada.
2. El PLN modela el reporte del paciente de manera numérica y objetiva, haciendo posible tratar el lenguaje de manera puramente científica y matemática. Establecer comparaciones y medir variables con total exactitud, a diferencia de las conclusiones extraídas a través del juicio clínico.
3. Los tests psicométricos, tienen la desventaja de ser un ejercicio de prueba. Ello implica que el cansancio (Süss y Schmiedek, 2000), la motivación (Schleicher et al., 2010) y haber realizado el test más de una vez (Bartels et al., 2010) deterioren seriamente la validez de estas medidas. El PLN, no requiere otra actividad adicional en el paciente que la de hablar en la entrevista clínica, porque analiza pasivamente el lenguaje. Por ello, estos efectos no deterioran la capacidad predictiva y diagnóstica del PLN. Esto permite, por ejemplo, comparar cómo diferentes variables extraídas por la red neuronal se comportan a lo largo de un proceso terapéutico.

Durante los últimos cinco años, los intentos por construir este tipo de sistemas y aplicarlos al contexto clínico en psicología ha crecido exponencialmente. Actualmente, la literatura más reciente se puede dividir según dos criterios (Arseniev-Koehler et al., 2018; Calixto et al., 2022; Zhang et al., 2022).

Primero, el tipo de datos que se utiliza para entrenar y comprobar la precisión de la red neuronal. La mayoría de los estudios, se han enfocado en el análisis de publicaciones en redes sociales como Twitter, Reddit o Facebook (Bosco et al., 2023; Fine et al., 2020; Garg et al., 2023; Haque et al., 2022; Hawa et al., 2020; Hu et al., 2020). Este tipo de abordaje ofrece la facilidad que aporta la enorme cantidad de lenguaje que cada día se genera en estas plataformas de manera gratuita y de libre acceso. Sin embargo, existen dos problemas (Calixto et al., 2022; Conway y O'Connor, 2016).

1. Muchos de estos participantes no han sido diagnosticados previamente, en muchos casos son anónimos. Provocando que cualquier inferencia que se realice sobre ellos dependa exclusivamente de los patrones que la IA es capaz de detectar.
2. Frecuentemente, muchos de estos estudios han empleado un contenido público y de libre acceso, pero sin el consentimiento de los participantes. Este es un área gris dentro de la legalidad y de la ética en la investigación psicológica.

Pero también en el análisis de historiales médicos (Downs et al., 2017), notas tomadas por profesionales de la salud (Kshatriya et al., 2021; Tran y Kavuluru, 2017), grabaciones o transcripciones de entrevistas (Ringeval et al., 2017; Tlachac et al., 2020; Valstar et al., 2014; Voleti et al., 2019), agentes conversacionales telemáticos (Milne-Ives, 2020; Mouseh et al., 2019) o redacciones escritas (Rauchenberger et al., 2016).

Segundo, los estudios varían también según el espectro psicopatológico que buscan estudiar. Algunas de las temáticas principales son:

1. El Trastorno del Espectro Autista (TEA): Tanaka et al. (2014) estudiaron tanto las características verbales como las acústicas en niños japoneses de entre 10 y 12 años a través del software LIWC. Con una precisión del 70% pudieron distinguir las verbalizaciones narrativas de la muestra con TEA del control. Parish-Morris et al. (2018) también analizan señales verbales y hallan que la frecuencia de pausas prosódicas en el discurso se asocia con una menor sintomatología del TEA. MacFarlane et al. (2022) aplicaron un modelo de Machine Learning (Support Vector

Machine, SVM) para predecir el diagnóstico de TEA. Con una precisión del 80% lograron diferenciar la muestra experimental del control.

2. Dislexia: al ser un trastorno del lenguaje caracterizado por la mayor presencia de errores ortográficos en la escritura, ha sido un nicho muy productivo para la aplicación de estos modelos. Tolami et al. (2021) crearon una red neuronal capaz de predecir dislexia con un 95% de precisión en niños iraníes. También Richard y Serrurier (2020), han empleado algoritmos de machine learning para diferenciar la escritura de niños con dislexia en base a imágenes de su escritura a mano.
3. Esquizofrenia y espectro psicótico: los trastornos psicóticos se caracterizan por una fuerte desorganización del pensamiento, que consecuentemente se traduce en un lenguaje bastante característico (Covington et al., 2005). Diversos estudios desde el comienzo del PLN se han enfocado en este conjunto de trastornos, tanto por su severidad y prevalencia, como por su alta saliencia sintomatológica. Variables como la coherencia del discurso, la complejidad sintáctica, la riqueza del vocabulario o el uso de metáforas son algunos de los marcadores con los que se entrena a redes de PLN para esta tarea (Corcoran et al., 2020).

1.3 Justificación del presente estudio

En este estudio, se aplica un modelo BERT (*Bidirectional Encoder Representations from Transformers*) a una tarea muy común en el ámbito terapéutico de la salud mental, el diagnóstico de niveles altos de ansiedad. Teniendo en cuenta la literatura mencionada en el apartado anterior, una investigación de este tipo tiene sentido por las siguientes razones:

1. La presencia de modelos del lenguaje en español es muy inferior a la existente en inglés (Calixto et al, 2022; Zhang et al., 2022). Suponiendo un grave problema a la hora de poder aplicar en países hispanohablantes todos los avances tecnológicos alrededor del PLN en general. Aunque en algunos casos los modelos más avanzados (ie. RoBERTa) son capaces de extraer algo de información, la precisión en sus diferentes casos de uso desciende notablemente.
2. Como también se ha expuesto, la mayoría de las investigaciones aplicando modelos de PLN para el diagnóstico de la salud mental se realiza en redes sociales. Muchas veces implicando que los sujetos a los que se “diagnostica” nunca han sido contrastados por un profesional o alguna herramienta psicométrica validada. Además de los riesgos éticos que supone el manejo de estos datos. Por ello, este estudio se centra en un *corpus* de texto generado en entrevistas en directo,

con un instrumento ampliamente validado con el que comprobar la eficiencia de las predicciones y con sujetos identificados que han consentido legalmente su participación.

3. Por último, en este estudio se entrena una red neuronal basada en los Transformadores, la tecnología que actualmente consigue los mejores resultados en prácticamente todas las tareas de PLN (OpenAI, 2023; Vaca, 2022).

La pregunta de investigación que vertebra de este estudio es de carácter exploratorio, ¿pueden los modelos de PLN asistir en la tarea psicodiagnóstica? Existen dos objetivos principales:

1. Conocer qué precisión puede conseguirse en una tarea de clasificación supervisada, es decir, en la predicción de qué extractos de texto pertenecen a personas con mayores o menores niveles de ansiedad.
2. Realizar una labor de extracción de los patrones que la red ha empleado para realizar sus predicciones (*feature extraction*). Observar en qué tipo de palabras se ha enfocado más el modelo BERT para asociar una pieza lingüística natural con mayores niveles de ansiedad.
 1. La presencia de estados ansiosos, ya sean momentáneos o estén asociados con la personalidad, tiene efectos evidentes sobre el lenguaje. Generalmente, se asocian con un discurso más auto-referencial, representado por una mayor presencia de pronombres personales singulares (y no plurales) (Lyions et al., 2018). Rico en la expresividad emocional negativa (Rook et al., 2022) y también con un mayor uso de tiempos verbales futuros (Geromini y Woodruff-Borden, 2014). Poder comparar las características extraídas por el modelo BERT con estas conclusiones acerca del lenguaje en estado y rasgo ansioso es también un objetivo de este trabajo.

2. Métodos

2.1 Muestra

Para recoger texto en un formato natural, se realizaron 11 entrevistas telefónicas con una duración media de 15 minutos a mujeres de entre 20 y 24 años de Madrid. Todas ellas hablan castellano de manera nativa, siendo su idioma principal. Los criterios de selección se aplicaron por dos motivos:

1. Reducir al máximo posible las diferencias interindividuales entre los sujetos, para controlar posibles variables extrañas como la edad o el sexo que pudieran afectar al lenguaje. El objetivo, es que la presencia o ausencia de ansiedad sea el principal rasgo que diferencie el lenguaje entre las participantes.
2. La elección exclusivamente de mujeres se decidió siguiendo la hipótesis de que el sexo femenino expresa con mayor frecuencia e intensidad procesos psicológicos y emocionales. Una idea sobre la que existe bastante consenso en la literatura científica (Chaplin, 2015; Kring y Gordon, 1998; Newman et al., 2008). Por supuesto, existe un enorme debate acerca de los factores que hacen emerger estas diferencias. Con el objetivo de capturar niveles de ansiedad en el texto, la posibilidad de incluir ambos sexos funciona entonces como una variable extraña. Acotar la muestra a mujeres, provoca por lo tanto que el contenido del lenguaje registrado tenga mayores cargas emocionales, haciendo la tarea de clasificación más eficiente.

A partir de las 11 entrevistas se generaron 118 extractos, aislando cada una de las preguntas y dividiendo en dos aquellas cuya respuesta superase los 512 caracteres (BERT no asume más dentro de cada observación durante su entrenamiento). Los textos cuentan con una extensión media de 179 palabras y en total suman 12,024. Las entrevistas se grabaron en audio y se transcribieron para poder ser introducidas en el programa. La transcripción del audio se realizó manteniendo las palabras exactas originadas por las participantes. No se acomodó el texto para otorgarle la coherencia y cohesión que el lenguaje verbal comúnmente carece. Tampoco se incluyeron elementos prosódicos o pragmáticos (ie. "Eh", "Ah"). Teniendo en cuenta el funcionamiento de la red neuronal utilizada y la base de datos con la que ha sido pre-entrenada, estas palabras no son reconocidas por el modelo.

2.2 Instrumentos

La entrevista realizada a los sujetos contaba con diez preguntas genéricas, que buscan simular aquellas que podrían ocurrir en el contexto terapéutico.

1. ¿Cómo ves tu vida en este momento?
2. ¿Cómo te sientes al hablar de tu vida ahora mismo?
3. ¿Cómo ves tu futuro cercano?
4. ¿Cómo te sientes al hablar de tu futuro ahora mismo?
5. ¿Cómo ves tus relaciones cercanas actualmente?
6. ¿Cómo te sientes al hablar de tus relaciones cercanas?

Las preguntas 1, 3 y 5 abordan temas generales sobre cómo la participante se siente en el momento actual, respecto a asuntos a los que comúnmente se dedica pensamiento: el presente, el futuro y las relaciones cercanas. Las preguntas 2, 4 y 6 están diseñadas para ejercitar el metalenguaje, y suscitar contenido de carácter emocional.

7. ¿Qué cosas o situaciones te suelen preocupar?
8. ¿Qué cosas o situaciones te suelen irritar o enfadar?
9. ¿Qué cosas o situaciones te suelen entristecer?
10. ¿Qué cosas o situaciones hacen que te sientas tensa?

Las cuatro preguntas restantes tienen la pretensión de hacer emerger contenido que abarque diferentes dimensiones de los estados ansiosos o neuróticos: la preocupación, la irritación, la tristeza y la tensión. Con estas preguntas, se busca elicitarse un lenguaje que sea más diferenciado entre aquellas personas con niveles superiores e inferiores de ansiedad. Están especialmente dirigidas al segundo objetivo de la investigación, extraer características en el contenido lingüístico en los sujetos que puntúen más alto en ansiedad.

Para obtener una medida válida y sólida de la ansiedad, se administró a los sujetos el cuestionario STAI (*State-Trait Anxiety Inventory*, Spielberger et al., 1970). Esta herramienta, se construye sobre un abordaje de la ansiedad desde dos perspectivas:

1. La ansiedad como estado transitorio caracterizado por un sentimiento de tensión, aprensión y un aumento de la actividad del sistema nervioso autónomo.
2. La ansiedad como rasgo de personalidad, que contempla las diferencias individuales a largo plazo y de manera intercontextual.

Constituye una de las medidas para medir ansiedad más utilizadas en la disciplina psicológica, contando con un alfa de Cronbach total de 0.93 en su versión en español (Fonseca-Pedrero et al., 2012; Guillén-Riquelme y Buela-Casal, 2011). También ha demostrado tener una alta consistencia interna tanto en muestras clínicas como no clínicas en español, y una fiabilidad test-retest superior al 0.8 en un intervalo de dos semanas (Ortuño-Sierra et al., 2016).

Las medidas del STAI, permiten convertir la tarea de clasificación en un proceso de aprendizaje supervisado, puesto que el modelo BERT cuenta con esta información para entrenarse, especialmente útil dado el reducido tamaño de la muestra. Para realizar la tarea “diagnóstica”, se dividió el total de participantes en dos grupos. El punto de corte se hizo sobre la mediana de las puntuaciones globales del STAI. De esta forma, se creó una mitad “ansiosa” y otra “no ansiosa” asegurando, por lo tanto, que ambos grupos tuvieran el mismo número de casos. Esto es especialmente importante para evitar problemas de balance en el entrenamiento del modelo.

Finalmente, como modelo de PLN de Deep Learning se utilizó la familia de transformadores BERT en sus versiones en español. El objetivo de los modelos aplicados es predecir correctamente, si un determinado texto pertenece al grupo “ansioso” o al grupo “no ansioso”. Debido a su complejidad, se incluye un apartado exclusivo para describir este funcionamiento. Gracias a que estos modelos permiten ser pre-entrenados en enormes cantidades de texto, pueden ser luego aplicados a muestras más pequeñas como la de este estudio. Incorporando toda la información aprendida en su desarrollo original.

2.3 Transformadores

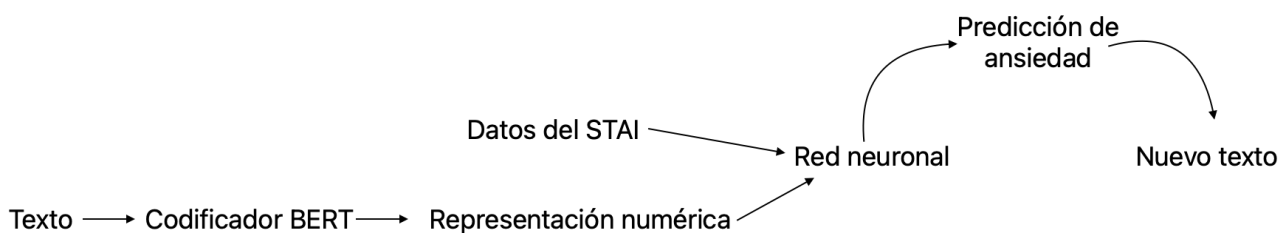
Los transformadores, fueron introducidos por Vaswani et al. en 2017, científicos computacionales de Google. La gran novedad que trajeron al campo del PLN fue una potente aproximación al problema de la semántica y el contexto. Ambas propiedades del lenguaje, como se ha descrito en la

introducción, han supuesto un importante problema histórico en la aplicación de modelos computacionales al análisis lingüístico. Existen dos importantes ramas en los modelos de transformadores:

1. Aquellos destinados a tareas de comprensión del lenguaje.
2. Aquellos cuyo objetivo es adicionalmente la generación de lenguaje.

En el caso de este experimento, únicamente se utiliza la tarea de comprensión. Es decir, la capacidad para codificar el lenguaje en un modelo numérico, a través de vectores. Esta representación numérica se asocia con las dos categorías posibles determinadas por el STAI (“grupo ansioso” y “grupo no ansioso”) a través de una red neuronal artificial (*artificial neural network, ANN*). La ANN utiliza una función de costo de entropía cruzada (*cross-entropy loss function*) para perfeccionar este proceso de aprendizaje durante las épocas (*epochs*) necesarias. La Figura 1 es una representación genérica del proceso:

Figura 1: Implementación del proceso de diagnóstico automático.



Fuente: Elaboración propia

Por lo tanto, aunque generalmente se describe un modelo de transformador con dos grandes secciones en su arquitectura (codificador y decodificador). En este caso únicamente se usa el codificador, es decir, la parte de “comprensión”.

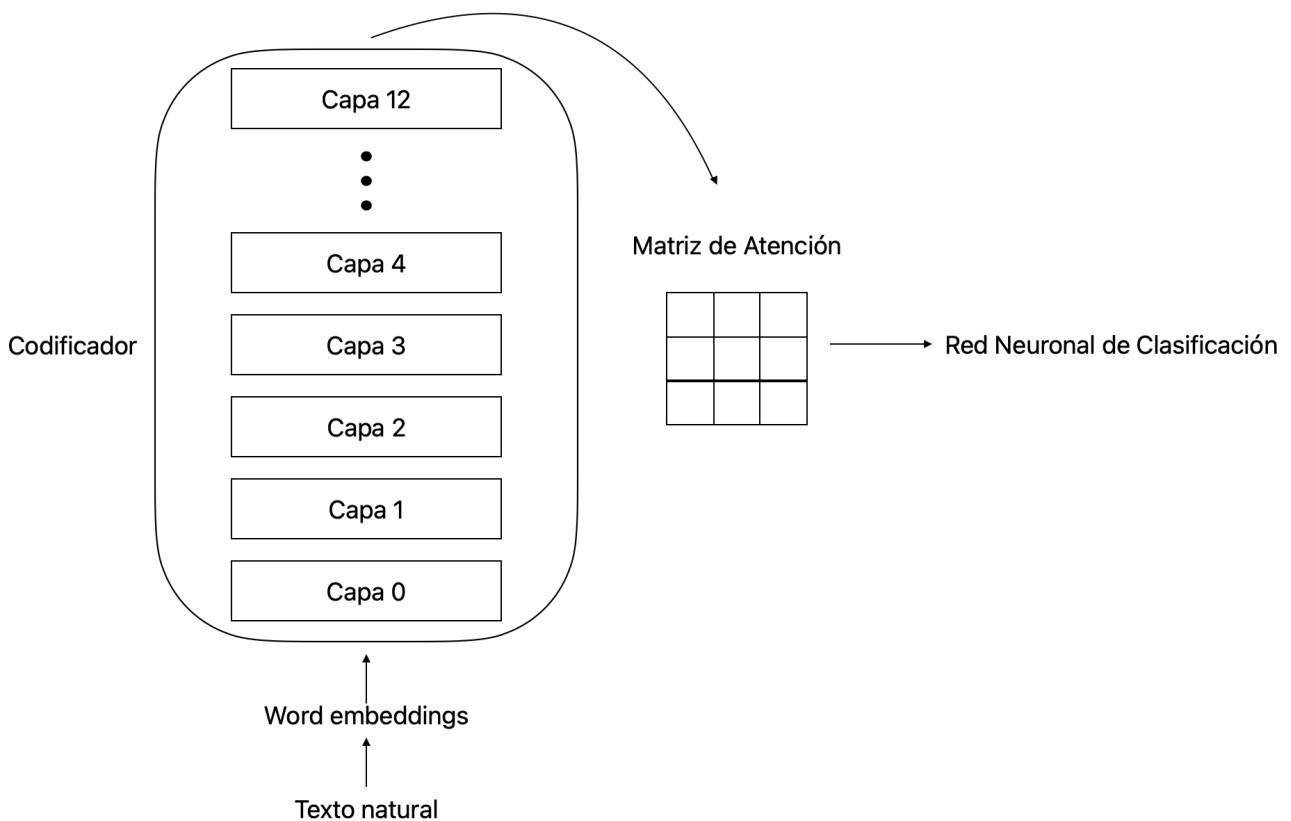
2.3.1 Arquitectura del modelo

En la red construida para este estudio, existen dos mecanismos principales:

1. Codificador BERT
2. Red neuronal artificial

El codificador tiene el objetivo de generar una representación numérica del texto que capture de manera óptima y compacta las propiedades del lenguaje de ejemplo. La arquitectura del codificador utilizado en este estudio se compone de 12 capas formadas por un mecanismo de propio-atención y una red neuronal en cada una de ellas. Estas capas son secuenciales, de forma que el output de cada una es procesado por la siguiente. La abstracción es acumulativa. Para apreciarla de forma más intuitiva se muestra la estructura en la Figura 2:

Figura 2: Estructura del codificador.

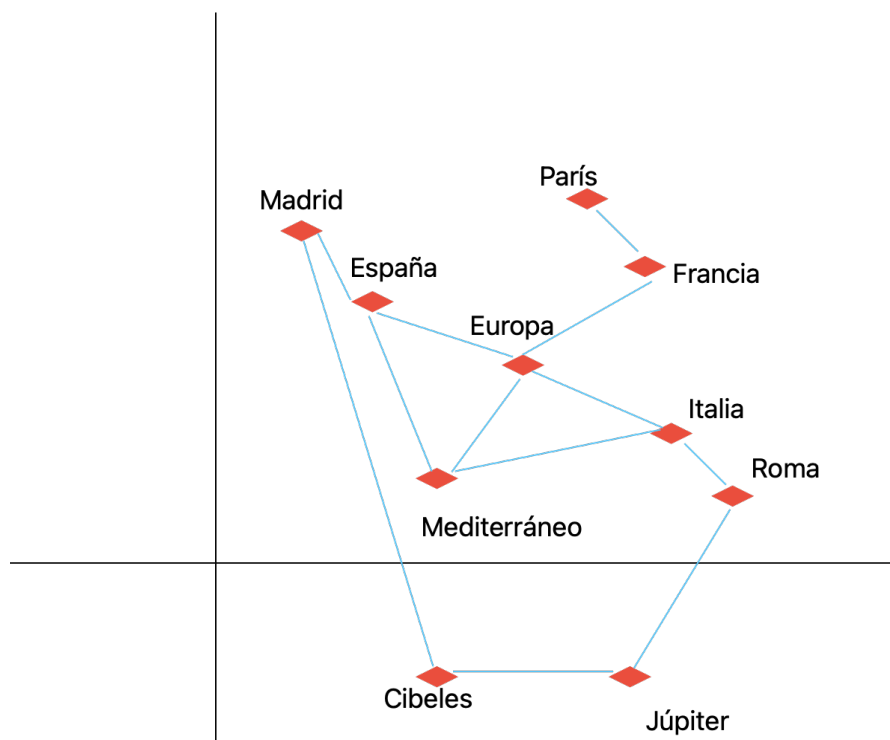


Fuente: Elaboración propia

Como se puede observar, antes de introducirse en el codificador, el texto natural se está traduciendo a “palabras incrustadas” (*word embeddings*, WE). Los WE son un tipo de representación de las palabras a través de vectores, otorgando a cada palabra un vector en función de su proximidad léxica. A través del entrenamiento del modelo en billones de palabras, se crea una topografía del vocabulario de una determinada lengua, un mapa parecido al que se incluye en la Figura 3. La “distancia” entre una palabra u otra se determina por la probabilidad de que aparezcan juntas en la misma secuencia, o en el mismo contexto. Una probabilidad estimada por la frecuencia

en la que dos palabras se encuentran en la misma frase dado un contexto. Por ejemplo, “Madrid” tiene mayor proximidad a la palabra “España” que a la palabra “China”, porque en los millones de textos en los que el algoritmo de WE ha sido entrenado, la frecuencia en la que “Madrid” y “España” aparecen juntas en el texto es mucho mayor. De esta forma, cada palabra es traducida a un vector que le otorga una posición dentro todo el vocabulario de una lengua. Esta es la forma en la que la arquitectura de los Transformadores resuelve el problema de la semántica. En este caso captura la semántica utilizando el campo léxico, de forma que realmente, el modelo no comprende el significado de una palabra. Únicamente conoce los otros conceptos con los que generalmente se asocia.

Figura 3: Ejemplo ilustrativo de la topografía creada en los Word Embeddings (WE)



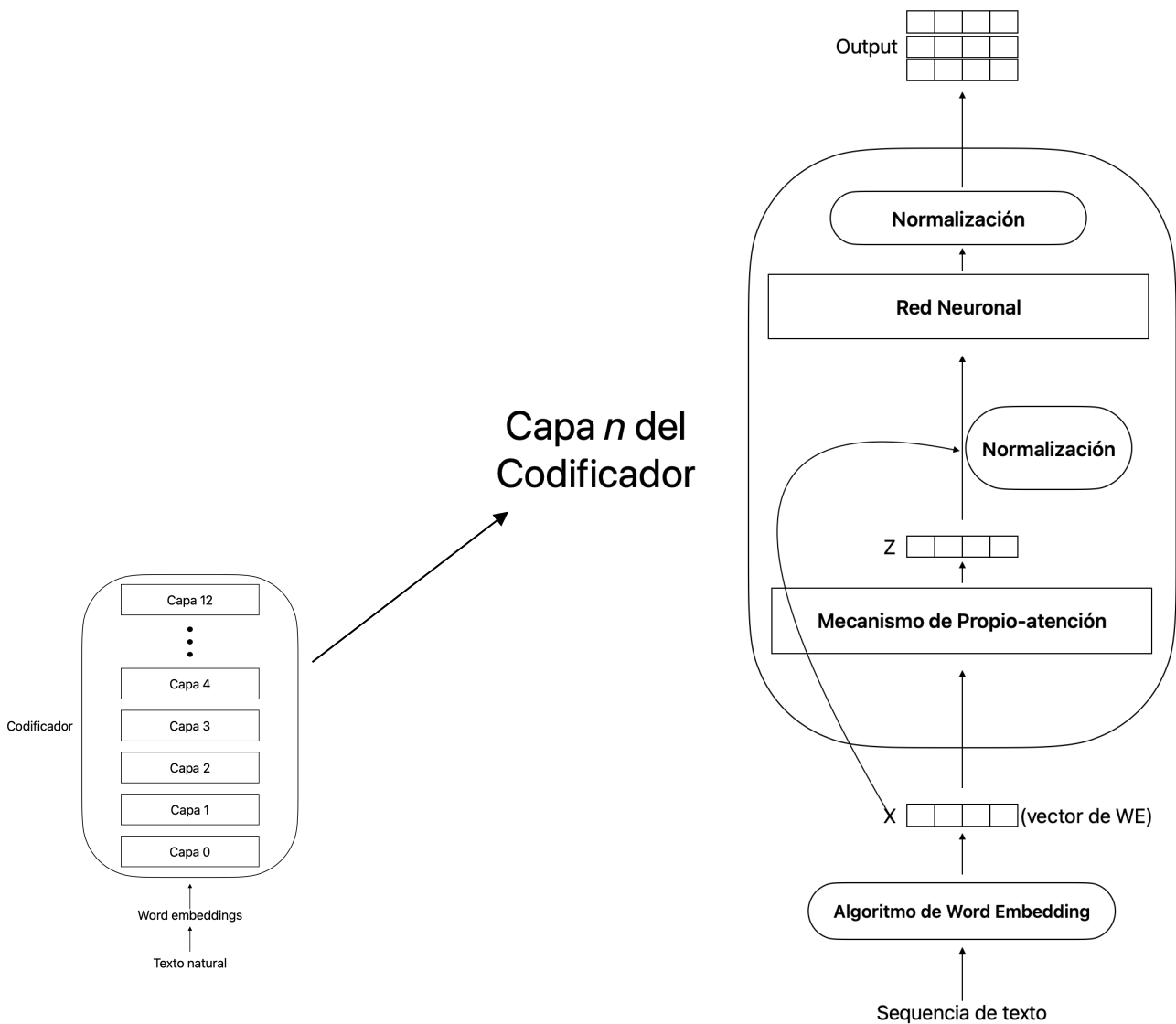
Fuente: Elaboración propia.

El ejemplo de la Figura 3 es ilustrativo, facilita una intuición de cómo se construyen los vectores del WE. En la realidad, prácticamente todas las palabras de una lengua en la que el algoritmo de WE ha sido entrenado se distribuyen dentro de este plano. Una vez el vector se introduce en la primera capa ocurre el mecanismo de propio-atención (*self-attention*), que emite una matriz. Esta

información se normaliza y se introduce en una red neuronal que extrae más información de la relación entre las palabras de una secuencia.

El mecanismo de propio-atención es la gran aportación de los transformadores, pues está pensado para resolver el problema de la semántica por contexto. Dentro de cada capa, ocurre un proceso que de manera esquemática se representa en la Figura 4.

Figura 4: *Arquitectura de cada capa de codificador.*



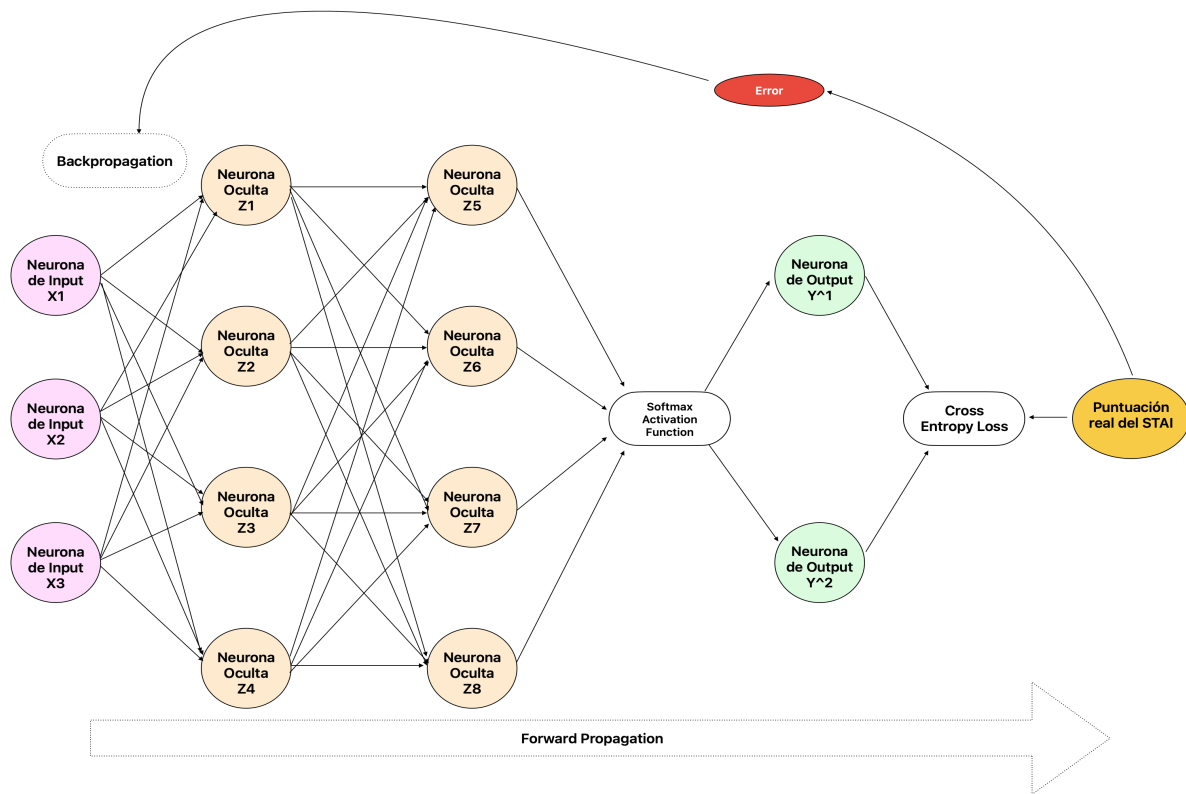
Fuente: Elaboración propia

Lo que ocurre dentro del mecanismo de atención, tiene una funcionalidad bastante compleja como para ser introducida con detalle aquí. Para ofrecer un acercamiento tentativo al mecanismo de propio-atención, se intuyen estas nociones:

1. El mecanismo de atención parte de la idea de capturar el significado de una palabra a través de la relación entre aquellas que la rodean y su propia función en una secuencia. Es decir, nace de un paradigma de que la semántica puede ser hasta cierto punto extraída del contexto léxico, aunque no se comprenda la unidad en sí.
2. BERT codifica cada Word Embedding, según su relevancia para sí y para los que le acompañan en una secuencia. Se calcula una puntuación (*attention score*) para cada palabra en la frase a través de tres vectores (*Query, Key y Value*), que guía al modelo en la importancia de cada unidad lingüística para el conjunto de la oración y viceversa.
3. Estos tres vectores funcionan como una matriz de pesos, que se inicializa en cada iteración y se actualiza tras la fase de *backpropagation*. Cuando finaliza la fase de entrenamiento, BERT ha aprendido qué puntuaciones de importancia para cada palabra en el contexto de su secuencia se asignan a cada grupo (ansioso y no ansioso).

El output, que produce una matriz con la información capturada por los mecanismos de propio-atención de todas las 12 capas, es finalmente introducido en una red neuronal artificial para la tarea de clasificación. En este punto, el modelo ya no está operando sobre texto si no sobre matrices de puntuaciones de importancia. En la Figura 5 se incluye un ejemplo ilustrativo de la estructura y funcionamiento de esta red:

Figura 5: Red neuronal artificial del modelo construido



Fuente: Elaboración propia

Las neuronas de input ($x_1, x_2, x_3... x_n$), reciben la matriz de vectores del codificador de BERT, que es una representación numérica del texto tras las doce capas de las que está formado. Las dos capas de neuronas ocultas ($z_1, z_2, z_3... z_n$) son transformaciones de los valores que contienen las neuronas de input. Estas transformaciones se realizan multiplicando los vectores por valores de 0 a 1, conocidos como pesos (*weights*). Los valores pasan de una capa a otra hasta llegar a una función de activación, que en el caso de este modelo es una función *softmax* para interpretarlas como probabilidades.

$$Softmax(x_i) = \frac{exp(x_i)}{\sum_i exp(x_j)}$$

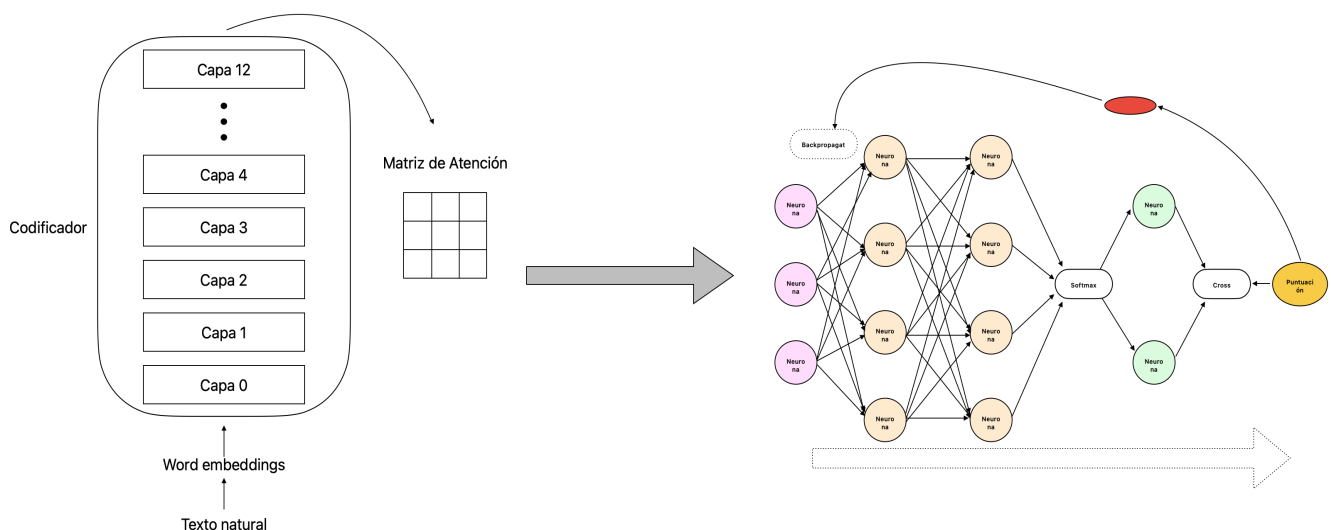
Esta función convierte un vector de números reales en una distribución, que determina la probabilidad que la red asigna a que cierta observación pertenezca al “grupo ansioso” (1) o al “grupo no ansioso” (0). Esta distribución se compara con el valor real de la observación (sección de texto convertida en matriz a través del codificador), es decir, la que es estimada a través del STAI. Dicha comparación se realiza a través de una función de costo de entropía cruzada (*cross entropy loss*):

$$Cross Entropy Loss = -(y \log(p) + (1 - y) \log(1 - p))$$

Esta primera parte del desarrollo de la red se conoce como *Forward Propagation* (Propagación hacia delante). La comparación entre la probabilidad propuesta por la red ($y^{\wedge}_1, y^{\wedge}_2$) y el valor real de una observación (y_1, y_2), genera un error (e). El tamaño de dicho error se utiliza para corregir los pesos que unen las neuronas de la red, proceso denominado *Backpropagation* (Propagación hacia atrás). Conforme pasa por la estructura todo el conjunto de observaciones, el tamaño del error se reduce progresivamente y la actualización de los pesos es cada vez menor. Este es el proceso mediante el cual se entrena una red neuronal. Los pesos se deben siempre modificar en una dirección que reduzca el tamaño del error computado por la función de *Cross Entropy*. Para que esto se produzca de manera eficiente y de manera exponencial entre las iteraciones, se aplica una técnica de optimización de descenso del gradiente a través del algoritmo *Adam* (*Adaptive Moment Estimation*), que actualmente es el más potente.

En resumen, el modelo de PLN con el que se realiza la clasificación entre textos del grupo ansioso y el grupo no ansioso tiene esta estructura y funcionalidad (Figura 6):

Figura 6: Arquitectura general del modelo de clasificación binaria en transformadores



Fuente: Elaboración propia

Dado el nivel de complejidad de este tipo de cuestiones, este estudio se centra en la aplicabilidad del modelo. Otros autores (Liu et al., 2017; Sewak et al., 2020) han realizado y continúan elaborando

trabajos que específicamente tratan esta arquitectura y las consecuencias de las diferentes alteraciones que se pueden implementar.

2.3.2 Modelos utilizados

Para este estudio se aplicaron tres modelos del conjunto de derivaciones del BERT original. Estas variaciones ocurren por la diferencia de tareas y de condiciones en los que se usan, tales como constricciones en la capacidad computacional, falta de memoria en los sistemas, el idioma específico o las características de los datos a procesar. Los tres siguen la misma arquitectura descrita durante esta sección, pero se diferencian en el número de neuronas input que se introducen en la red neuronal final. Es decir, en la complejidad de la matriz de atención con la que el codificador representa numéricamente el texto. Los modelos son:

1. BERT original: se utilizó la versión ligera (*base*) del modelo, debido a la limitada capacidad computacional con la que se contaba en el experimento. Este fue introducido por primera vez por Devlin et al. (2019). Cuenta con 110 millones de parámetros, 12 capas de codificación y es capaz de computar 768 características del lenguaje. El modelo completo (*BERT large*) cuenta con 340 millones de parámetros, 24 capas de codificación y trabaja con 1,024 características. Debido a las limitaciones financieras de este estudio, no fue posible implementar el modelo completo. De todas formas, cuando ambos sistemas se prueban en la Evaluación de Comprensión General del Lenguaje (GLUE, en inglés) no se encuentran diferencias excesivas. La evaluación GLUE (Wang et al., 2018a) es el estándar científico con el que se prueban estos modelos. La versión reducida de BERT (*base*) obtuvo una precisión media 79.6 sobre 100. El BERT completo obtiene un 82.1 (Devlin et al., 2019).
2. DistilBERT (Distilled BERT): se trata de un modelo más reducido en complejidad y por lo tanto en la demanda de energía y capacidad de computación. Fue introducido por Sanh et al. (2019), y ha obtenido una puntuación de 77.0 en el examen GLUE.
3. ALBERT (A Lite BERT): otro modelo más reducido introducido por Lan et al. (2020). La versión completa (*large*) de ALBERT cuenta con 18 millones de parámetros, 24 capas y opera con 128 características lingüísticas.

Debido a que estos modelos han sido pre-entrenados en inglés originalmente, se ha tenido que acceder a las versiones publicadas por el Departamento de Ciencias de la Computación de la Universidad de Chile. Para implementar el primero, BERT original, se utilizó BETO (Cañete et al.,

2020). Para ALBERT y DistilBERT se emplearon también versiones desarrolladas por los mismos autores (Cañete et al., 2023), denominadas ALBETO y DistilBETO. Los tres ejemplares han sido originalmente entrenados de manera no supervisada en el mismo *corpus*. Se trata de una base de datos denominado el *Spanish Unannotated Corpora*, compuesto por 3 billones de palabras. Los textos contienen todos los artículos en español de la Wikipedia, y otros muchos orígenes (artículos del Parlamento Europeo, de las Naciones Unidas o del índice JCR, entre muchos otros). Los modelos como BERT permiten que los aprendizajes obtenidos de este intenso proceso de pre-entrenamiento luego sean aprovechados en tareas como las que componen este trabajo. Este proceso de ajuste a la nueva información se conoce como *Fine Tuning* (afinación).

Para el entrenamiento posterior del modelo se seleccionó aleatoriamente el 80% de los textos que constituyen la muestra de este estudio, reservando el 20% para que la evaluación de la efectividad pudiera ser sobre material nuevo. Por lo tanto, las predicciones o “diagnóstico” del modelo se realizan sobre estos 24 textos aleatoriamente seleccionados, del total de los 118 generados para esta investigación.

Todos los modelos se han programado en Python 3, utilizando las librerías PyTorch (Pazke et al., 2019), pandas (McKinney, 2010), Scikit-learn (Pedregosa et al., 2011), NumPy (Harris et al., 2020) y HuggingFace Transformers (Wolf et al., 2019).

3. Resultados

Todos los modelos se aplicaron con las mismas condiciones, para asegurar que la comparación es homogénea.

- Tamaño de las tandas (*batchsize*) de entrenamiento = 5
- Tamaño de las tandas (*batchsize*) de evaluación = 5
- Épocas de entrenamiento (*epochs*) = 5

3.1 BETO

El modelo BETO programado cuenta con un codificador de 12 capas, un total de 768 neuronas en sus capas ocultas y 12 mecanismos de propio-atención (*attention heads*), funcionando en paralelo dentro de cada capa del codificador. Trabaja con un vocabulario compuesto por 31.002 palabras. Cada observación contiene un máximo de 512 caracteres. Se encontraron las siguientes métricas del desempeño en la tarea de clasificación binaria. La precisión, alcanzó el 0,625 (62.5% de los ejemplos se asignaron al grupo correcto), mientras que la puntuación F1 (que se obtiene a partir de la especificidad y la sensibilidad) fue de 0,594. Para observar con mayor claridad se muestra la matriz de confusión en la Tabla 1.

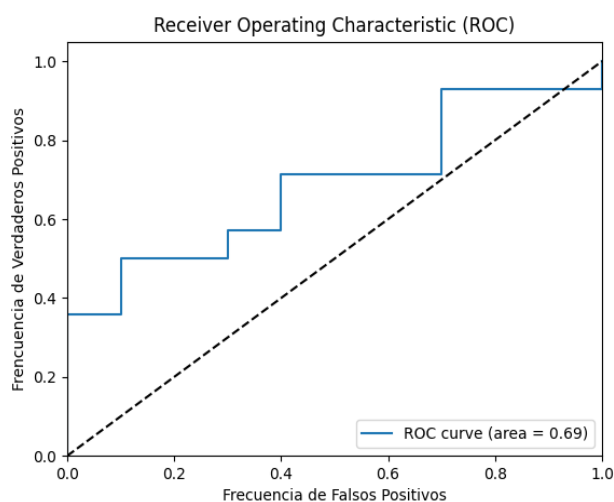
Tabla 1: Matriz de confusión de BETO.

| | | Diagnóstico real del STAI | |
|-------------------------------|------------------|---------------------------|------------------|
| | | Grupo ansioso | Grupo no ansioso |
| Diagnóstico asignado por BETO | Grupo ansioso | 10 | 0 |
| | Grupo no ansioso | 9 | 5 |

Por lo tanto, 9 de los 24 textos han sido erróneamente catalogados. De los cuales 9 son falsos negativos de ansiedad, sin que se halle ningún falso positivo. Supone pues una sensibilidad de 0,526 y una especificidad total de 1.

También se calculó el Área Bajo la Curva (*area under the curve, AUC*), con una puntuación de 0,685. La curva ROC (*Receiver Operating Characteristics*) se muestra en la Figura 7:

Figura 7: Curva ROC de BETO.



3.2 ALBETO

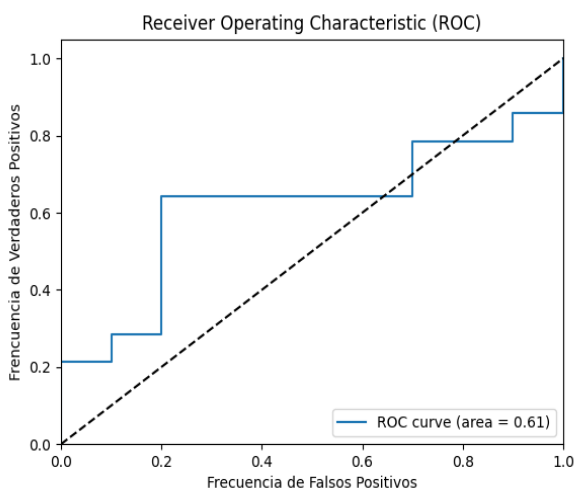
El modelo ALBETO está compuesto por 10 capas, cada una con 12 mecanismos de propio-atención y una profundidad neuronal vertical de 768 nodos ocultos. Cada observación puede contener un máximo de 512 caracteres. Tiene un vocabulario de 31.000 palabras. El modelo consiguió una precisión de 0,541 y un F1 igual a 0,504. La matriz de confusión en la Tabla 2:

Tabla 2: Matriz de confusión de ALBETO.

| | | Diagnóstico real del STAI | |
|---------------------------------|------------------|---------------------------|------------------|
| | | Grupo ansioso | Grupo no ansioso |
| Diagnóstico asignado por ALBETO | Grupo ansioso | 9 | 1 |
| | Grupo no ansioso | 10 | 4 |

11 de los 24 textos se asignaron al grupo equivocado, con 10 falsos negativos y 1 falso positivo. Obteniendo entonces una sensibilidad de 0,473 y una especificidad de 0,909. El tamaño del área bajo la curva es de 0,61, y la curva ROC se expone en la Figura 8:

Figura 8: Curva ROC de ALBETO



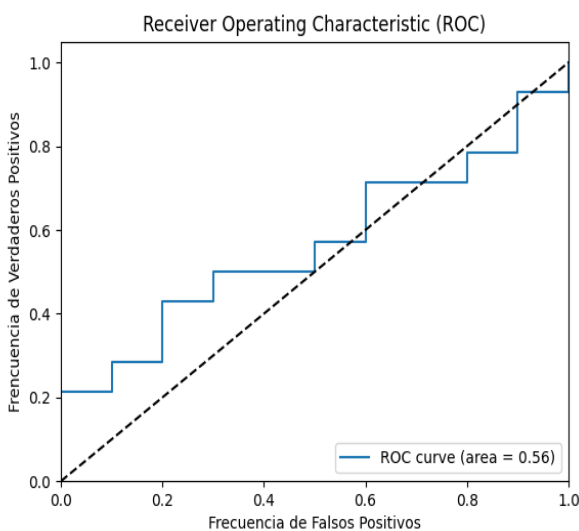
3.3 DistilBETO

En este caso se halla un modelo de 6 capas con 12 mecanismos de propio-atención, un número máximo de 512 por observación y 768 neuronas en cada capa oculta. El vocabulario que manipula es de 31.002 palabras. Es por lo tanto el más pequeño de los codificadores. DistilBETO consiguió una precisión de 0,583 en los 24 textos destinados a la evaluación, con una puntuación F1 de 0,54. La sensibilidad es de 0,5 y la especificidad de 0,8, como se puede apreciar en la matriz de confusión de la Tabla 3. El área bajo la curva es de 0,56.

Tabla 3: Matriz de confusión de DistilBETO

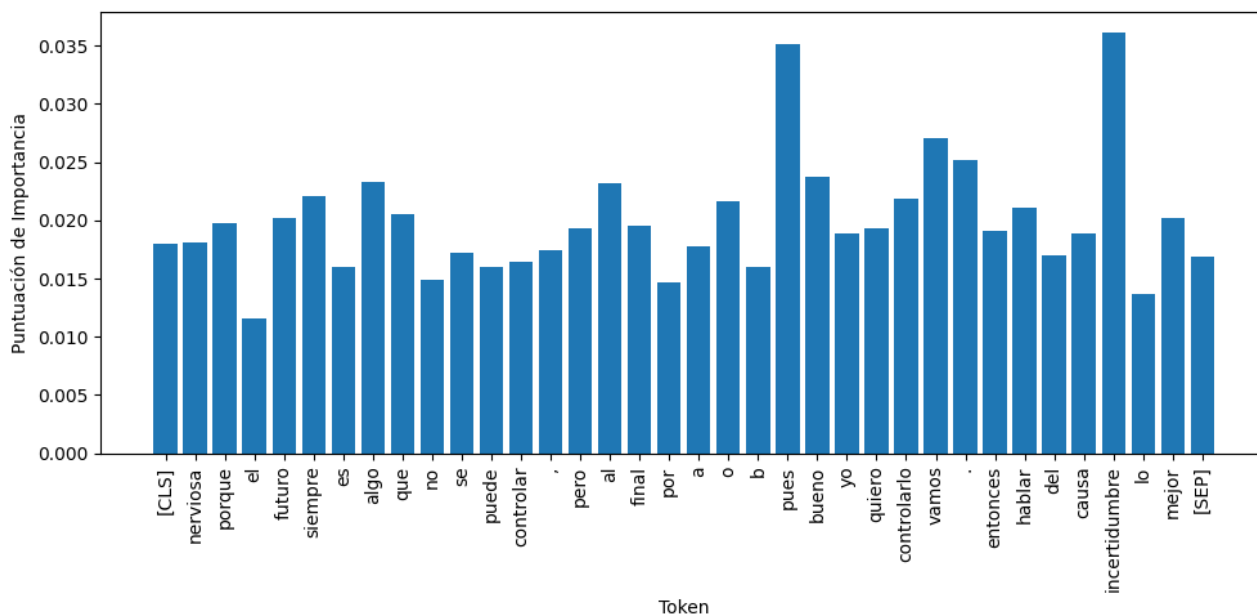
| | | Diagnóstico real del STAI | |
|-------------------------------------|------------------|---------------------------|------------------|
| | | Grupo ansioso | Grupo no ansioso |
| Diagnóstico asignado por DistilBETO | Grupo ansioso | 8 | 2 |
| | Grupo no ansioso | 8 | 6 |

Figura 9: Curva ROC de DistilBETO



En la Tabla 4, se muestra una comparación de la capacidad diagnóstica de los tres modelos:

Figura 11: Importancia atribuida por BETO en el Ejemplo 2



Por ilustrar los gráficos de las Figuras 9 y 10 con algo más de matiz y contexto, se resaltan aquellas palabras que superan el nivel de 0,018 en la puntuación de atención (*attention score*) del primer ejemplo y 0,02 en el segundo. Se recuerda que estas palabras son las que se han determinado de mayor valor discriminador para la secuencia a la que pertenecen. Aunque pudiera parecer poco lógico que palabras como “pues” tengan tanto peso, se entiende que el modelo no atiende a la naturaleza de la palabra, si no su importancia para la secuencia completa. La puntuación de atención obedece a la probabilidad asignada por BERT.

Ejemplo 1:

“¿Qué cosas o situaciones te hacen sentir tensa?”

Bueno, ir a hacer algo nuevo sola. O sea, algo que no he hecho sola antes me genera tensión. Quedar con ciertas personas. En situaciones sociales, el ir a una fiesta en la que no conoces a nadie me genera tensión también. Cuando procrastino demasiado y tengo que entregar alguna cosa en el último momento también me pongo tensa.”

Ejemplo 2:

“¿Cómo te sientes al estar hablando de tu futuro?

Nerviosa, porque el futuro **siempre es algo que** no se puede controlar. Pero **al final** por A **o** por B **pues bueno**, que yo siempre quiero **controlarlo vamos**, pero no se puede. Entonces, al final hablar del futuro es algo que causa **incertidumbre** a lo mejor.”

4. Discusión y conclusión

En este estudio se han afinado modelos de Procesamiento del Lenguaje Natural de la familia de Transformadores, con dos objetivos:

1. Explorar la capacidad diagnóstica de los modelos para evaluar ansiedad en una tarea de clasificación supervisada.
2. Extraer los resultados del mecanismo de propio-atención, para poder inferir variables o determinantes que expliquen cómo BERT realiza sus predicciones.

Por lo general, la capacidad de los tres modelos ha sido bastante baja. El modelo con mayor precisión es el más voluminoso, BETO. Puede tener sentido, ya que es capaz de codificar con mayor exhaustividad y por lo tanto es capaz de aprovechar mejor una reducida cantidad de datos. Es importante mencionar, que estos modelos pueden aplicarse bajo grandes variaciones de sus parámetros, cualquier replicación de los resultados obtenidos debe tener en cuenta las características pre-programadas en el sistema. Pese a los bajos resultados, el hecho de que se pueda diferenciar el 62% de una muestra lingüística correctamente habiendo empleado únicamente 11 sujetos, es una señal del considerable poder de estos modelos. También la agilidad y optimización en la implementación, que permiten realizar estudios como este sin requerir billones de palabras para su entrenamiento y a un insignificante coste computacional. Este estudio se ha realizado en un ordenador portátil de uso comercial estándar, las tareas más potentes no han llegado a superar los 8 GB de ocupación de la memoria RAM.

Otro aspecto que destacar es la propensión de los modelos afinados a infradiagnosticar a los sujetos como parte del grupo no ansioso cuando sí lo son. En los tres casos, el número de falsos negativos

ha sido considerablemente mayor que el de falsos positivos. Podría explicarse por una falta de balance en la muestra, es decir por la descompensación de las observaciones de un grupo sobre el otro. Sin embargo, este no es el caso, ya que ambos grupos estuvieron perfectamente equilibrados en la muestra. Otra razón puede ser el uso de la mediana como punto de corte y la división de los sujetos en sólo dos grupos. Dados los recursos y la muestra, no habría sido posible generar un modelo que distinguiese entre más de dos mitades, pues para cada subgrupo que se crea debe haber un suficiente número de ejemplos. Esta decisión metodológica tiene serias limitaciones, pues se ha tratado a la muestra de manera dicotómica. Cuando en la realidad encontrarse en el percentil 10 y el 45 tiene profundas diferencias, en el entrenamiento de estos modelos se han tratado ambos casos por igual. Es sensato suponer que el modelo no ha logrado comprender un lenguaje auténticamente discriminante del nivel de ansiedad, lo cual se evidencia en la precisión. Es necesario ampliar la muestra para poder partir el total de participantes en más estratos y contar en cada grupo con suficientes ejemplos de los que extraer un patrón coherente.

El segundo objetivo de esta investigación era el de poder desvelar algunos de los patrones del funcionamiento de BERT para compararlos con la literatura sobre el lenguaje en estados de ansiedad. Como métrica para seguir esta hipótesis exploratoria, se utilizó el grado de importancia o de atención que BERT ha asignado a cada palabra para discernir un texto de otro, como se ha representado en las Figuras 9 y 10 e ilustrado en los ejemplos escritos.

No se pueden obtener conclusiones claras, de nuevo probablemente el tamaño muestra y la partición en dos mitades poco naturales contribuya. Hay que recordar también, que la muestra (como se ha pretendido) es muy poco representativa de las mujeres. No sólo por el tamaño, también por la acotada edad o localización geográfica. En cierta medida, deberían analizarse miles de textos para poder extraer un patrón claro. Es verdad que BERT destaca muchos fragmentos como “incierto” o “genera tensión”, “tensa” o “procrastino demasiado”, que quizás se podrían relacionar con el malestar ansioso con más claridad. Pero también palabras que no tienen ningún vínculo aparentemente obvio con el discurso ansioso como “pues”, “siempre” o “fiesta”. BERT no sólo asigna atención a una palabra en sí misma, por el hecho de su frecuencia en los textos de un grupo u otro. Tampoco tiene en cuenta la naturaleza o el significado de la palabra. El mecanismo de propio-atención funciona puntuando en la medida que las palabras que envuelven a otra son importantes para su presencia en una frase. La llegada indirecta al significado, de la semántica a través del

contexto, es propensa a hallar estos resultados. Probablemente si se extendiese considerablemente el número de sujetos y se lograra partir la muestra de forma acorde a la distribución normal, las puntuaciones de ansiedad por palabra tendrían más armonía teórica.

No se encontró ningún artículo científico que trate medidas de ansiedad a través de transformadores en textos clínicos, por lo que es imposible comparar los resultados.

Por lo tanto, una fuerte limitación de este trabajo ha sido el volumen muestral. Habiéndose podido acceder a más sujetos se podría haber obtenido un conjunto de participantes que se acomodasen con más realismo a una distribución normal de ansiedad. Ello habría hecho posible utilizar la desviación típica o los centiles para seccionar los grupos y el trabajo de BERTO sería más acorde a la realidad. Además, la variabilidad del lenguaje en muestras reducidas es menor, y por lo tanto BERTO se fija en palabras menos representativas de la ansiedad. Es más probable que asigne una alta puntuación de atención a palabras cuya presencia se debe a la casualidad o a variables extrañas, y no a una frecuencia significativa en niveles ansiosos bajos o altos.

Otra limitación de este estudio es la falta de recursos computacionales. Los modelos más potentes bien no están accesibles de forma gratuita o son demasiado demandantes para ser programados en un ordenador genérico comercial. Contar con estructuras más robustas como RoBERTa (*Robustly optimized BERT approach*; Liu et al., 2019), o los últimos modelos en español como RigoBERTa. (Vaca et al., 2022) y MarIA (Guitérrez-Fandiño, 2022), podría haber mejorado los resultados. Hay razones para suponerlo, pues en la comparativa de este estudio ha sido el modelo más voluminoso el que mejor precisión ha podido obtener.

Por último, hay que contar con el material con el que estas IAs han sido entrenadas. A fecha de hoy, no existe un *annotated corpora* en español dedicado a la salud mental. O al menos no ha sido reportado en el espacio comunitario científico. Los más cercanos, son aquellos desarrollados para trabajar con informes médicos. Los que existen en otros idiomas para esquizofrenia o depresión tampoco son lo suficientemente voluminosos, y no son accesibles (Zhang et al., 2022). La mayoría de las bases de datos provienen de la Wikipedia o de comentarios sobre productos comerciales. Con la dificultad añadida, de que muchos de los *corpus* clínicos construidos no pueden ser publicados para no romper con la confidencialidad y la ética en investigación. Una común alternativa ha sido

dirigirse a las redes sociales, pero las implicaciones éticas y la baja validez que tienen estas bases de datos son muy cuestionables. Por lo tanto, hace falta generar modelos en español y pre-entrenados en notas de profesionales clínicos o transcripciones de entrevistas, entre otros.

Las implicaciones de este estudio son escasas si se coloca el foco sobre la precisión obtenida. Si se plantea la aproximación aquí aplicada como metodología de análisis, la importancia de esta área de trabajo es enorme. Existen tres razones por las que el diagnóstico automático puede ser sumamente disruptivo y constructivo para la psicología clínica y académica:

1. La optimización del proceso terapéutico basado en la evidencia. Gracias al diagnóstico automático e inteligente de este tipo, los profesionales podrían delegar las tareas cuantitativas de medición a modelos como este, para centrar más recursos en el vínculo terapéutico. En el caso de poder validarse un modelo real que asignase puntuaciones certeras y válidas, bastaría con grabar en audio una sesión de terapia para obtener medidas del constructo en cuestión. El modelo tomaría el conocimiento con el que ha sido entrenado y emitiría una predicción para el paciente. En estos casos, que el lenguaje sea el medio de acceso tiene dos ventajas:
 - a. Se eliminan los efectos de fatiga y aprendizaje asociados a la medición psicométrica. La evaluación podría hacerse cada sesión de terapia, pudiendo observar el progreso de cada paciente sin perder validez.
 - b. Se pueden capturar muchos más matices, al tratarse de lenguaje natural subjetivo, que aquellos extraídos de un test o escala. Mientras que la riqueza de la psicometría está en las inferencias estadísticas con cientos de participantes, este tipo de modelos permiten obtener conocimiento del discurso generado por un solo participante.
2. Autoevaluación por parte del paciente: cierto es que ningún modelo existente puede sustituir la interacción con profesional. Sin embargo, si se entrena un modelo de estas características para atender a múltiples dimensiones (*multi feature*), cualquier persona podría monitorizar su salud mental de manera independiente. A través de las conversaciones en redes sociales o WhatsApp, llamadas telefónicas o conversaciones con IAs virtuales, se podría construir un modelo de PLN que detecte síntomas subclínicos o prematuros en diferentes dimensiones psíquicas, como el afecto o la cognición. Esto facilitaría la

intervención preventiva de los servicios de salud mental, a los que se acudiría con una descripción etiológica muy completa. Las aplicaciones para la psicogerontología, en la detección precoz del

3. Ampliación de los modelos psicopatológicos: hasta el momento, la teoría psicopatológica y los grandes manuales se construyen sobre la observación profesional de síntomas reportados. Algunos de los problemas más difíciles de modelizar a través de este acercamiento son (Caplan y Cosgrove, 2004; Kinderman, 2014):
 - a. El inicio y desarrollo de cada patología en cada tipo de sujeto.
 - b. El carácter dimensional y no categórico de la sintomatología.
 - c. La comorbilidad entre áreas diagnósticas.
 - d. La expresión sintomática en la población normal, no clínica ni subclínica.

Elaborar sistemas que permitan construir representaciones formales del lenguaje natural en cualquier ámbito, de forma constante y automática, puede ser una herramienta para paliar estas barreras. Puede llevar a componer un abordaje de la enfermedad mental menos dependiente de la observación de la conducta verbal y motora en contextos clínicos.

En futuras investigación alrededor de esta cuestión, es necesario contar con muestras considerablemente mayores y con modelos pre-entrenados en textos cercanos a la salud mental. El desarrollo de corpus del lenguaje en español es necesario para que se puedan aplicar estas tecnologías en países hispanohablantes.

Los esfuerzos por quitarle el estigma a la asistencia a terapia psicológica son un avance, está incrementando la demanda y el interés político. Pero este nuevo mercado coloca una presión no conocida sobre un sistema que no fue concebido para dar una cobertura nacional. Por ello, las estrategias por desarrollar herramientas preventivas, que optimicen el sistema de salud serán necesarias en esta transición. La automatización del diagnóstico, la implementación de herramientas para el cuidado propio o una reformulación de los modelos psiquiátricos psicopatológicos, son propuestas en esta línea. La tecnología ya las hace posibles y sigue un rápido avance, el trabajo debe ir dirigido a su integración en la psicología y al desarrollo de propuestas de aplicación real.

Referencias bibliográficas

- Allport, G. W. (1942). The use of personal documents in psychological science. *Social Science Research Council Bulletin*, 49, xix-210.
- Amari, S. (1972). Learning Patterns and Pattern Sequences y Self-Organising Patterns of Threshold Elements. *IEEE Transactions of Computers*, 21(11), 1197-1206. <https://doi.org/10.1109/T-C.1972.223477>.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychological Association. (2023). Qualitative research. En G. R. VandenBos (Ed.) *APA Dictionary of Psychology*. <https://dictionary.apa.org/qualitative-research>.
- American Psychological Association. (2023). Quantitative research. En G. R. VandenBos (Ed.) *APA Dictionary of Psychology*. <https://dictionary.apa.org/quantitative-research>.
- Anastasi, U., y Cyprien, T. (2021). Theories underpinning language acquisition/learning: behaviourism, mentalism and cognitivism. *International Journal of Contemporary Applied Researches*, 8(4), 1-15.
- Arseniev-Koehler, A., Mozgai, S., y Scherer, S. (2018). What type of happiness are you looking for? A closer look at detecting mental health from language En K. Loveys, K. Niederhoffer, E. Prud'hommeaux, R. Resnik y P. Resnik (Eds.) *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 1-12). Association for Computational Linguistics.
- Attneave, F. (1959). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods and Results*. Holt, Rinehart & Winston.
- Backus, J. W., Herrick, H., y Ziller, I. (1954). Preliminary Report: Specifications for the IBM Mathematical FORMula TRANSlating System, FORTRAN. *IBM Corporation*, 1-29.

- Bartels, C., Męgrzyn, M., Wiedl, A., Ackermann, V., y Ehrenreich, H. (2010). Practice effects in health adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, 16(11). doi: 10.1186/1471-2202-11-118.
- Blumenthal, A. L. (1978). The emergence of psycholinguistics. *Synthese*, 72, 313-323.
- Borsboom, D. (2008). Psychometric Perspectives on Diagnostic Systems. *Journal of Clinical Psychology*, 64(9), 1089-1108.
- Bosco, C., Patti, V., Frenda, S., Cignarella, A. T., Paciello, M., y D'Errico, F. (2023). Detecting racial stereotypes: An Italian social media corpus where psychology meets NLP. *Information Processing & Management*, 60(1), 1-18. <https://doi.org/10.1016/j.ipm.2022.103118>.
- Boyd, R. L., y Schwartz, H. A. (2021) . Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present and Future States of the Field. *Journal of Language and Social Psychology*, 40 (1), 21-41.
- Brush, S. G. (1967). History of the Lenz-Ising Model. *Reviews of Modern Physics*, 39(4), 883-893. [10.1103/RevModPhys.39.883](https://doi.org/10.1103/RevModPhys.39.883).
- Buchanan, R. D., y Finch, S. J. (2005). History of Psychometrics. En B. Everitt y B. Howell (Eds.) *Encyclopedia of Statistics in Behavioural Science* (vol. 2). John Wiley and Sons, Ltd. <https://doi.org/10.1002/0470013192.bsa282>.
- Burnett, J. (2018). *Of the Origin and Progress of Language*. (2ª Ed. Vol. 1). Gale ECCO. (Originalmente publicado en 1773).
- Burnham, W. H. (1888). Memory, Historically and Experimentally Considered. *The American Journal of Psychology*, 2(1), 39-90. <https://doi.org/10.2307/1411406>.
- Cain, D. J., Keenan, K., y Rubin, S. (2016). *Humanistic Psychotherapies: Handbook of Research and Practice*. American Psychological Association.
- Campanario, J. M. (2004). El Enfoque Conexionista en Psicología Cognitiva y algunas aplicaciones sencillas en Didáctica de las Ciencias. *Enseñanzas de las Ciencias*, 22(1), 93-104.

- Cañete, J., Chaperon, G., Fuentes, R., Ho, J., Kang, H., y Pérez, J. (2020, 26 de abril). *Spanish Pre-Trained BERT Model and Evaluation Data* [Presentación de artículo]. Practical ML for Developing Countries Workshop, Addis Ababa, Etiopía.
- Cañete, J., Donoso, S., Bravo-Marquez, F., Carvallo, A., y Araujo, V. (2023). ALBETO and DistilBETO: Lightweight Spanish Language Models. *arXiv*. <https://arxiv.org/pdf/2204.09145.pdf>.
- Caplan, P. J., y Cosgrove, L. (2004). *Bias in Psychiatric Diagnosis*. Jason Aronson, Inc.
- Chaplin, T. M. (2015). Gender and Emotion Expression: A Developmental Contextual Perspective. *Emotion Review*, 7(1), 14-21. doi: 10.1177/1754073914544408.
- Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, 35(1), 26-58.
- Conway, M., y O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. *Current Opinion in Psychology*, 9, 77-82. <https://doi.org/10.1016/j.copsy.2016.01.004>.
- Corcoran, C. M., Mittal, V. A., Bearden, C. E., Gur, R. E., Hitczenko, K., Bilgrami, Z., Savic, A., Cecchi, G. A., y Wolff, P. (2020). Language as a biomarker for psychosis: A natural language processing approach. *Schizophrenia Research*, 226, 158-166. <https://doi.org/10.1016/j.schres.2020.04.032>.
- Covington, M. A., Congzhou, H., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S., Semple, J., y Brown, J. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research*, 77(1), 85-89. <https://doi.org/10.1016/j.schres.2005.01.016>.
- Cuoto, D. (2022). Can There be Thought Without Words? Donald Davidson on Language and Animal Minds. *Top*, 41, 587-598. <https://doi.org/10.1007/s11245-022-09801-5>.
- Da Cunha, I. F. (2013). The Utopia of Unified Science: The Political Struggle of Otto Neurath and the Vienna Circle. *Principia*, 17(2), 319-329.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. John Murray.

- De Carlos, J. A., y Borrell, J. (2007). A historical reflection on the contributions of Cajal and Golgi to the foundations of neuroscience. *Brain Research Reviews* 55(1), 8-16. doi:10.1016/j.brainresrev.2007.03.010.
- Delmas, C., Vandame, C., y Andréolle, D. S. (2010). Introduction. En C. Delmas, C. Vandame y D. S. Andréolle (Eds.), *Science and Empire in the Nineteenth Century: A Journey of Imperial Conquest and Scientific Progress* (pp. vii-xvii), Cambridge Scholars.
- Devlin, J., Chang, M., Lee, K., y Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*. <https://arxiv.org/pdf/1810.04805.pdf>.
- Didier, J., y Bigand, E. (2011). *Rethinking physical and rehabilitation medicine: New technologies induce new learning strategies*. Springer Science & Business Media.
- Downs, J., Velupillai, S., George, G., Holden, R., Kikoler, M., Dean, H., Fernandes, A., y Dutta, R. (2017). Detection of Suicidality in Adolescents with Autism Spectrum Disorders: Developing a Natural Language Processing Approach for Use in Electronic Health Records. *AMIA Annual Symposium Proceedings Archive*, 2017, 641-649. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977628/>.
- Duignan, B. (2009). Rasmus Rask. En *Encyclopædia Britannica*. <https://www.britannica.com/biography/Rasmus-Rask>.
- Echeverri, S. (2014). Explaining Reference: A Plea for Semantic Psychologism. En J. Dutant, D. Fassio y A. Meylan (Eds.), *Liber Amicorum Pascal Engel* (pp. 550-579). Université de Genève.
- Ellingsworth, H. W. (1992). The Shadow of Benjamin Lee Whorf: Continuing Issues in Linguistic Relativism. *Intercultural Communication Studies*, 2(2), 43-59.
- Englander, M. (2016). The phenomenological method in qualitative psychology and psychiatry: Empirical Study. *International Journal of Qualitative Studies in Health and Well-being*, 11(30682), 1-11. <https://dx.doi.org/10.3402/qhw.v11.30682>.
- Ercikan, K., y Roth, W. (2006). What Good Is Polarising Research Into Qualitative and Quantitative? *Educational Researcher*, 35(5), 14-23.

- Fayyad, U., Piatetsky-Shapiro, G., y Smyth P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3).
- Fine, A. B., Crutchley, P., Blase, J., Carroll, J., y Coppersmith, G. (2020). Assessing population-level symptoms of anxiety, depression, suicide risk in real time using NLP applied to social media data. En D. Bamman, D. Hovy, D. Jurgens, B. O'Connor y S. Volkova (Eds.) *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science* (pp. 50-54). Association for Computational Linguistics.
- Fitch, W. T. (2019). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Philosophical Transactions of the Royal Society B*, 375(20190046), 1-9. <https://doi.org/10.1098/rstb.2019.0046>.
- Fodor, J. (1983). *The Modularity of the Mind*. MIT Press.
- Fraser, W. I., King, K. M., Thomas, P., y Kendall, R. E. (1986). The diagnosis of schizophrenia by language analysis. *British Journal of Psychiatry*, 148, 275-278.
- Fromkin, V., Rodman, R., y Hyams, N. (2003). *An Introduction to Language*. Wadsworth.
- Fuchs, A. H., y Milar, K. S. (2003). Psychology as a Science. En D. K. Freedheim y I. B. Weiner (Eds.) *Handbook of Psychology: Volume I, History of Psychology* (pp. 1-26). John Wiley & Sons, Inc.
- Fukushima, K. (1980). Neurocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36, 193-202.
- Galán, C. (1994). La teoría lingüística en Wilhem von Humboldt. *Anuario de Estudios Filológicos*, 17, 165-185. <https://hdl.handle.net/10662/2679>.
- Gardner, B. T., y Gardner, R. A. (1998). Development of phrases in the utterances of children and cross-fostered chimpanzees. *Human Evolution*, 13, 161-188.
- Garg, M., Saxena, C., Naseem, U., y Dorr, B. J. (2023). NLP as a Lens for Causal Analysis and Perception Mining to Infer Mental Health on Social Media. *arXiv*. <https://arxiv.org/abs/2301.11004>.

- Geromini, E. M. C., y Woodruff-Borden, J. (2014). The language of worry: Examining linguistic elements of worry models. *Cognition and Emotion*, 29(2), 1-9. doi: 10.1080/02699931.2014.917071.
- Girard, J. M., Vail, A. K., Liebenthal, E., Brown, K., Kilciksiz, C. M., Pennant, L., Liebson, E., Ögnur, D., Morency, L., y Baker, J. T. (2022). Computational analysis of spoken language in acute psychosis and mania. *Schizophrenia Research*, 245, 97-117.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. MIT Press.
- Greenberg, J. H. (1948). Linguistics and Ethnology. *Southwestern Journal of Anthropology*, 4(2), 140-147. <https://www.jstor.org/stable/3628708>.
- Gries, S. T. (2017). *Ten Lectures on Quantitative Approaches in Cognitive Linguistics*. Brill.
- Gries, S. T., y Divjak, D. S. (2010). Quantitative approaches in usage-based Cognitive Semantics: Myths, erroneous assumptions, and a proposal. En D. Glynn y K. Fischer (Eds.), *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches* (pp. 333-354). De Gruyter Mouton.
- Griffin, C., y Phoenix, A. (1994). The Relationship Between Qualitative and Quantitative Research: Lessons from Feminist Psychology. *Journal of Community & Applied Social Psychology*, 4, 287-298.
- Guillén-Riquelmen, A., y Buela-Casal, G. (2011). Actualización psicométrica y funcionamiento diferencial de los ítems en el State Trait Anxiety Inventory (STAI). *Psicothema*, 23(3), 510-520.
- Hahlein, M., y Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present and Future of Artificial Intelligence. *California Management Review*, 61(4), 5-14.
- Haller, J. S. (1971). Race and the Concept of Progress in the Nineteenth Century American Ethnology. *American Anthropologist*, 73(3), 710-724. <https://www.jstor.org/stable/671764>.
- Han, J., Kamber, M., y Pei, J. (2012). Introduction. En J. Han, M., Kamber y J. Pei (Eds.) *Data Mining: Concepts and Techniques* (3ª Ed., pp. 1-32). Elsevier.

- Haque, R., Islam, N., Islam, M., y Ahsan, M. (2022). A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. *Technologies*, 10(57), 2-15. doi: 10.3390/technologies10030057.
- Harley, T. A. (2014). *The Psychology of Language*. Psychology Press. <https://doi.org/10.4324/9781315859019>.
- Harris, C. R., Millman, K. J., Van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357-362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hayes, K. J., y Nissen, C. H. (1971). Higher mental functions of a home-raised chimpanzee. En A. M. Schrier y F. Stollnitz (Eds.), *Behavior of non-human primates* (vol. 4, pp.57-115), Academic Press. (Originalmente publicado en 1956).
- Hawa, S., Akella, S., Kaushik, S., Joshi, V., y Kalbande, D. (2020). Analysis of Therapy Transcripts using Natural Language Processing. *International Journal of Engineering and Advanced Technology*, 9(6), 489-494.
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. John Wiley and Sons, Inc.
- Herder, J. G. (1966). *Essay on the Origin of Language* (A. Goode, Trad.). Frederick Ungar Publishing Co. (Originalmente publicado en 1772).
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1), 11-26. <http://dx.doi.org/10.1080/17470215208416600>.
- Hochreiter, S., y Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- Hoffman, R. E., y Sledge, W. (1988). An analysis of grammatical deviance occurring in spontaneous schizophrenic speech. *Journal of Neurolinguistics*, 3, 89-101.

- Hu, Y., Huang, H., Anfan, C., y Mao, X (2020). Weibo-COV: A Large-Scale COVID-19 Social Media Dataset from Weibo. En K. Verspoor, K. Bretonnel, M. Conway, B. De Bruijn, M. Dredze, R. Mihalcea, y B. Wallace (Eds) *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*. Association for Computational Linguistics.
- Huang, X., Akiri, Y., Jack, M. (1990). *Hidden Markov Models for Speech Recognition*. Columbia University Press.
- Hutchins, W. J. (2004). The Georgetown-IBM Experiment Demonstrated in January 1954. En R. E. Frederking y K. B. Taylor (Eds), *Machine Translation: From Real Users to Research* (pp. 102-114. Springer-Verlag.
- Ivakhnenko, A. G., y Lapa, V. G. (1967). *Cybernetics and Forecasting Techniques (Modern Analytic and Computational Methods in Science and Mathematics, vol. 8)*. Elsevier America.
- Jung, C. G. (1910). The Association Method. *The American Journal of Psychology*, 21(2), 219-269. <https://doi.org/10.2307/1413002>.
- Kshatriya, B. S. A., Sagheb, E., Wi, C., Yoon, J. Y., Seol, H. Y., Juhn, Y., y Sohn, S. (2021). Identification of asthma control factor in clinical notes using a hybrid deep learning model. *BMC Medical Informatics and Decisions Making*, 21(272).
- Kaul, V., Enslin, S., y Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*, 92(4), 807-812. <https://doi.org/10.1016/j.gie.2020.06.040>.
- Kellogg, W. N. (1968). Communication and language in home-raised chimpanzee. *Science*, 162, 423-438.
- Kennedy, B, Ashokkumar, A., Boyd, R. L., y Dehghani, M. (2022). Text Analysis for Psychology. En M. Dheghani y R. L. Boyd (Eds.) *Handbook of Language Analysis in Psychology* (pp. 3-60). Guildford Publications.
- Kim, A. (2015). Johann Friedrich Herbart. En *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2015/entries/johann-herbart/#Aca>.

- Kinderman, P. (2014). *A Prescription for Psychiatry: Why We Need a Whole New Approach to Mental Health and Wellbeing*. Palgrave Macmillan.
- Kleene, S. C. (1951). Representations of Events in Nerve Nets and Finite Automata. *The RAND Corporation*.
- Köffka, K. (2013). *Principles of Gestalt Psychology*. Routledge. [Originalmente publicado en 1935].
<https://doi.org/10.4324/9781315009292>.
- Köhler, W. (1929). *Gestalt psychology*. Liveright.
- Kring, A. M., y Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of Personality and Social Psychology*, 74(3), 686-703. doi: 10.1037//0022-3514.74.3.686.
- Lafferty, J., McCallum, A., y Pereira, F. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. En C. E. Brodley y A. P. Danyluk (Eds.) *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 282-289). Morgan Kaufmann Publishers, Inc.
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., y Soricut, R. (2020, 26 de abril al 1 de mayo). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations [Presentación de artículo]. The Eighth International Conference on Learning Representations, Addis Ababa, Etiopía.
- Lauriola, I., Lavelli, A., y Aiolli, F. (2022). An Introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, 470, 443-456.
- Levelt, W. J. M. (2013). *A History of Psycholinguistics: The Pre-Chomskyan Era*. Oxford University Press.
- Lilly, J. C. (1967). Dolphin's vocal mimicry as a unique ability and a step toward understanding. En K. Salzinger y S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications* (pp. 21-27). Academic Press.

- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., y Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*. <https://arxiv.org/abs/1907.11692>.
- Liu, W., Wand, Z., Liu, X., Zeng, N., Liu, Y., y Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11-26. <https://doi.org/10.1016/j.neucom.2016.12.038>.
- Le Cun, Y., y Bengio Y. (1995). Pattern Recognition and Neural Networks. En M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*. MIT Press.
- Le Cun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lucy, J. A. (2001). Sapir-Whorf Hypothesis. En N. J. Smelser y P. B. Bates (Eds.) *International Encyclopedia of Social and Behavioural Science* (1ª ed., pp. 13486-13490). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/03042-4>.
- Lyons, M., Aksayli, N. D., Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87, 207-211. <http://dx.doi.org/10.1016/j.chb.2018.05.035>.
- MacCarthy, J. (1960). Recursive Function of Symbolic Expressions and their Computation by Machine, Part I. *Communications of the Association for Computer Machinery*, 3(4), 184-195. <https://doi.org/10.1145/367177.367199>.
- MacCarthy, J. (1978). History of LISP. En R. L. Wexelbat (Ed.), *History of programming languages* (pp. 173-185). Association of Computing Machinery. <https://doi.org/10.1145/800025>.
- MacWhinney, B. (2015). Psycholinguistics: Overview. En J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioural Sciences* (2ª ed., vol. 19, pp. 353-358). Elsevier.
- Maimon, O., y Rokach, L. (2019). Introduction. En O. Maimon y L Rokach (Eds.) *Data Mining and Discovery Knowledge Handbook* (2ª Ed., pp. 1-18). Springer.
- Manning, C. D., Raghavan, P., y Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press.

- Manschreck, T. C., Maher, B. A. , y Ader, D. N. (1981). Formal thought disorder, the type-token ratio and disturbed voluntary motor movement in schizophrenia. *The British Journal of Psychiatry*, 139, 7-15.
- Maslow, A. H. (1968). *Toward a psychology of being*. Van Nostrand.
- Gutiérrez-Fandiño, A., Armengol-Estapié, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Pio, C., Armentano-Oller, C., Rodríguez-Penagos, C., González-Aguirre, A., y Villegas, M. (2022). MarIA: Spanish Language Models. *Procesamiento del Lenguaje Natural*, 68, 39-60.
- McCallum, A., y Nigam, K. (2001). A Comparison of Events Models or Naive Bayes Text Classification. En M. Shami (Ed.) *Learning for Text Categorization: Papers from the 1998 AAAI Workshop* (pp. 41-48). AAAI Press.
- McCallum, J. C. (2023, 22 mayo). *Historical Memory Prices 1957+*. <https://jcmmit.net/memoryprice.htm>.
- McKinney, W. (2010). Data Structure for Statistical Computing in Python. En S. Van der Walt, J. Millman (Eds.) *Proceedings of the 9th Python in Science Conference* (pp. 56-61). SciPy 2010.
- Mesulam, M. M., Conventry, C. A., Rader, B. M., Kuang, A., Sridhar, J., Martersteck, A., Zhang, H., Thompson, C. K., Weintraub, S., y Rogalski, E. J. (2021). Modularity and granularity across the language network: A primary progressive aphasia perspective. *Cortex*, 141, 482-496.
- Meyer, D. E., y Schvaneveldt, R. W. (1971). Facilitation in Recognising Pairs of Words: Evidence of a Dependence between Retrieval Operations. *Journal of Experimental Psychology*, 90(2), 227-234. <https://doi.org/10.1037/h0031564>.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144. [https://doi.org/10.1016/S1364-6613\(03\)00029-9](https://doi.org/10.1016/S1364-6613(03)00029-9).
- Miles, H. L. (1978). Language acquisition in apes and children. En F. C. C. Peng (Ed.), *Sign language and language acquisition in man and ape* (pp. 108-120). Westview Press.

- Mills, J. (2000, noviembre). *Linguistic Relativity and Linguistic Determinism: Idiom in 20th Century Cornish*. [Presentación de artículo]. New Directions in Celtic Studies Conference, Newquay, Reino Unido.
- Milne-Ives, M., De Cock, C., Lim, E., Shedadeh, M. H., De Pennington, Mole, G., Normando, E., y Meinert, E. (2020). The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *Journal of Medicine and Internet Research*, 22(10), e20346. doi: 10.2196/20346.
- Milokov, T., Chen, K., Corrado, G., y Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv*, 1301.3781. <https://arxiv.org/abs/1301.3781>.
- Minsky, M., y Papert, S. (1969). *Perceptrons*. MIT Press.
- Minton, S., Carbonell, J. G., Knoblock, C. A., Kuokka, D., Etzioni, O., y Gil, Y. (1989). Explanation-Based Learning: A Problem Solving Perspective. *Artificial Intelligence*, 40, 63-118. doi: 10.1016/0004-3702(89)90047-7.
- Morice, R. D., e Ingram, J. C. L. (1982). Language analysis in schizophrenia: diagnostic implications. *Australian & New Zealand Journal of Psychiatry*, 16, 11-21.
- Morgan, J. N., y Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(302), 415-434. <http://www.jstor.org/stable/2283276>.
- Mouseh, M., Schneider, J., Ahmad, K., Alam, T., Al-Thani, D., Siddig, M. A., Fernández-Luque, L., Qaraqe, M., Alfuquha, A., y Saxena, S. (2019). An Evolutionary Bootstrapping Development Approach for Mental Health Conversational Agent. En A. Hasman, M. S. Househ, J. Mantas, A. Kolohathi, A., J. Liaskos y P. Gallos (Eds.) *Health Informatics Vision: From Data via Information to Knowledge* (pp. 228-231). IOS Press.
- Nadkarni, P. M., Ohno-Machado, L., y Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18, 544-551.

- Newell, A., Shaw, J. C., y Simon, H. A. (1959). Report on a General Problem-Solving Program. *The RAND Corporation*.
- Newel, A., y Simon, H. A. (1956). The logic theory machine: A complex information processing system. *IRE Transactions on Information Theory*, 2(3), 61-79.
- Newman, M. L., Groom, C. J., Handelman, L. D., y Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211-236. <https://doi.org/10.1080/01638530802073712>.
- Niss, M. (2004). History of the Lenz-Ising Model 1920-1950: From Ferromagnetic to Cooperative Phenomena. *Archive for History of Exact Sciences*, 59, 267-318. <https://link.springer.com/article/10.1007/s00407-004-0088-3>.
- Ogawa, S., Tank D. W., Menon, R., Ellermann, J. K., Seong-Gi, K., Merkle, H., y Ugurbil, K. (1992). Intrinsic signal changes accompanying sensor stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Science*, 89, 5951-5955.
- Omonvoyeva, K. S. (2021). On the Emergence of Psycholinguistics, Its Current Significance and Benefits for Future Generations. *International Journal of Academical Pedagogical Research*, 5(5), 145-146.
- OpenAI (2023). GPT-4 Technical Report. *arXiv*. <https://arxiv.org/pdf/2303.08774.pdf>.
- Ortuño-Sierra, J., García-Velasco, L., Inchausti, F., Debanné, M., y Fonseca-Perero, E.(2016). New approaches on the study of the psychometric properties of the STAI. *Actas Españolas de Psiquiatría*, 44(3), 83-92.
- Osgood, C. E., Sebeok, T. A. (Eds), Gardner, J. W., Carrol, J. B., Newmark, L. D., Ervin, S. M., Saporta, S., Greenberg, J. H., Walker, D. E., Jenkins, J. J., Kellogg, W., y Lounsbury, F. G. (1945). Psycholinguistics: a survey of theory and research problems. *The Journal of Abnormal and Social Psychology*, 49(4, Pt. 2), i-204. <https://doi.org/10-1037/h0063655>.
- Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., Donaher, J., Ferguson, E., Pandey, J., y Schultz, R. T. (2017). Linguistic camouflage in girls with autism

spectrum disorder. *Molecular Autism*, 8(48), 1-12. <https://doi.org/10.1186/s13229-017-0164-6>.

Patterson, F. G. (1978). Linguistic capabilities of a lowland gorilla. En F. C. C. Peng (Ed.), *Sign language and language acquisition in man and ape* (pp. 161-201). Westview Press.

Pazke, A., Gross, S., Massa, F., Lerer, A., Bradhury, J., Chanan, G, Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., De Vito, Z., Raison, M., Tejani, A., Chilamkutyhy, S., Steiner, B., Fang, L., Bai, J., y Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. En H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. A. Fox y R. Garnett (Eds.) *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*. NeurIPS 2019.

Pegredosa, F., Varoquaux, G., Gramfort, A., Michael, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, V., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, E. (2011). Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Pennebaker, J. W., y Francis, M. E. (1999). *Linguistic Inquiry and Word Count*. Lawrence Erlbaum Associates, Inc.

Pennebaker, J. W., Mehl, M. R., y Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54, 547-577.

Pennington, J., Socher, R., y Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. En A. Mosschitti, B. Pang y W. Daelemans (Eds.) *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 1532-1543). Association for Computational Linguistics.

Pepperberg, I. M. (2017). Animal language studies: What happened? *Psychonomic Bulletin & Review*, 24, 181-185. <https://doi.org/10.1007/s11245-022-09801-5>.

- Perlovsky, L., y Sakai, K. L. (2014). Language and Cognition. *Frontiers in Behavioural Neuroscience*, 8(426), 1-2. <https://doi.org/10.3389/fnbeh.2014.00436>.
- Piaget, J. (1929). *The child's conception of the world*. Kegan Paul, Trench & Trubner.
- Premack, D. (1976). *Intelligence in ape and man*. Earlbaum.
- Primerio, G. G. (2008). Actualidad de la polémica Chomsky-Skinner. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 10(2), 263-269.
- Psathas, G. (1969). The General Inquirer: Useful or Not? *Computers and the Humanities*, 3(3), 163-174. <https://www.jstor.org/stable/30203989>.
- Rauschenbenberger, M., Rello, L., Füsichel, S., Thomaschewski, J. (2016) A Language Resource of German Errors Written by Children with Dyslexia. En N. Calzolari, K. Choukri, T. Declerck, S., Goggi, M. Grobelnik, B., Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odjik y S. Piperidis (Eds.) *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 83-87). European Language Resources Association (ELRA).
- Richard, G., y Serrurier, M. (2020). Dyslexia and Dysgraphia prediction: A new machine learning approach. *arXiv*. <https://arxiv.org/abs/2005.06401>.
- Richards, K. C., Campenni, C. E., y Muse-Burke, J. L. (2010). Self-care and Well-being in Mental Health Professionals: The Mediating Effects of Self-awareness and Mindfulness. *Journal of Mental Health Counseling*, 32(3), 247-264.
- Rickleman, B. L. (2009). Anosognosia in Individuals with Schizophrenia: Toward Recovery of Insight. *Issues on Mental Health Nursing*, 25(3), 227-242.
- Rieber, R. W., y Vetter, H. (2013). Theoretical and Historical Roots of Psycholinguistic Research. En R. W. Rieber (Ed.) *Psychology of Language and Thought: Essays on the Theory and History of Psycholinguistics* (pp. 3-49). Springer.
- Ringeval, F., Schuller, B., Valstar, M., Gratch, J., Cowie, R., Scherer, S., Mozgai, S., Cummins, N., Schmitt, M., y Pantic, M. (2017). Real-life Depression, and Affect Recognition Workshop and Challenge. En *AVEC'17: Proceedings of the 7th Annual Workshop in Audio/Visual Emotion*

Challenge (pp. 3-9). Association for Computer Machinery.
<https://doi.org/10.1145/3133944.3133953>.

- Rogers, C. R. (1942). *Counseling and Psychotherapy: Newer Concepts in Practice*. Houghton Mifflin.
- Rogers, J. A. (1972). Darwinism and Social Darwinism. *Journal of the History of Ideas*, 33(2), 265-280.
<http://www.jstor.org/stable/2708873>.
- Rook, L., Mazza, M. C., Lefter, I., y Brazier, F. (2022). Toward Linguistic Recognition of Generalized Anxiety Disorder. *Frontiers in Digital Health*, 4(779039). doi: 10.3389/fdgth.2022.779039.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organisation in the brain. *Psychological Review*, 65(6), 386-408.
<https://doi.org/10.1037/h0042519>.
- Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books.
- Rorschach, H. (1921). *Psychodiagnostik*. Hans Huber.
- Rousseau, J. (1966). *Essay on the Origins of Languages which treats of Melody and Musical Imitation* (J. H. Moran, Trad.). Frederick Ungar Publishing Co. (Originalmente publicado en 1781).
- Rumelhart, D. E., Hinton, G. E., y McClelland, J. L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition, Vol. 1: Foundations*. MIT Press.
- Safran, J. D., y Wallner, L. K. (1991). The Relative Predictive Validity of Two Therapeutic Alliance Measures in Cognitive Therapy. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(2), 188-195.
- MacFarlane, H., Salem, A. C., Chen, L., Asgari, M., y Fombonne, E. (2022). Combining voice and language features improves automated autism detection. *Autism Research*, 15, 1288-1300.
doi: 10.1002/aur.2733.
- Sanh, V., Debut, L., Chaumond, J., y Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv*. <https://arxiv.org/pdf/1910.01108.pdf>.

- Schleicher, D. J., Van Iddenkinge, C. H., Morgeson, F. P., y Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology*, *95*(4), 603-617. <https://doi.org/10.1037/a0018920>.
- Schmidhuber, J. (2022). Annotated History of Modern AI and Deep Learning. *Technical Report IDSIA*. <https://people.idsia.ch/~juergen/deep-learning-history.html>.
- Sewak, M., Sahay, S. K., y Rathore, H. (2020). An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Neuroscience*, *17*(1), 182-188.
- Shaffer, D. W., y Serlin, R. C. (2004). What Good are Statistics that Don't Generalise? *Educational Researcher* *33*(9), 14-25.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Sharpless, B. A., Barber, J. (2009). Psychodynamic Therapy. En R. E. Ingram (Ed.) *The International Encyclopedia of Depression*, (pp.460-465). Springer Publishing Co.
- Simpson, S. G., y Reid, C. L. (2014). Therapeutic alliance in videoconferencing psychotherapy: A review. *Australian Journal of Rural Health*, *22*, 280-299.
- Skinner, B. F. (1957). *Verbal Behavior*. Harvard University Press.
- Spielberger, C. D., Gorsuch, R. L., y Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.
- Stone, P. J., y Hund, E. B. (1963). A Computer Approach to Content Analysis: Studies Using the General Inquirer System. *Proceedings of the May 21-23rd spring joint computer conference, Estados Unidos*, 241-256.
- Sun, T., Liu, X., Qiu, X., y Huang, X. (2022). Paradigm Shift in Natural Language Processing. *Machine Intelligence Research*, *19*(3), 169-183.

- Tanaka, H., Sakti S., Neubig, G., Toda, T., y Nakamura, S. (2014). Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorder in Children's Narrative. En P. Resnik, R. Resnik y M. Mitchell (Eds.) *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 88-96). Association for Computational Linguistics. <https://aclanthology.org/W14-3211>.
- Tausczik, Y. R., y Pennebaker, J. W. (2010). The Psychological Meaning of Words: LLIWC and Computerised Test Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24-54.
- Tlaschac, M. L., Sargent, A., Toto, E., Paffenroth, R., y Rundensteiner, E. (2020, 14-17 de diciembre). *Topological Data Analysis to Engineer Features from Audio Signals for Depression Detection* [Presentación de artículo]. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). Miami, EdysEUU. doi: 10.1109/ICMLA51294.2020.00056.
- Tolami, F. A., Khorasani, M., Kahani, M., Yazdi, S. A. A., Ghalenoiei, M. A. (2021, 28-29 de octubre). *An intelligent linguistic error detection approach to automated diagnosis of Dyslexia disorder in Persian speaking children* [Presentación de artículo]. 2021 11th IEEE International Conference on Computer Engineering and Knowledge (ICCKE). Mashhad, Irán. <https://ieeexplore.ieee.org/abstract/document/9721446/authors>.
- Tran, T., y Kavuluru, R. (2017). Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. *Journal of Biomedical Informatics*, 75, S138-S148. <http://dx.doi.org/10.1016/j.jbi.2017.06.010>.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, 59(236), 433-460.
- Vaca, A., García G., Montoro, H., Aldama, N., Samy, D., Betancur, D., Moreno, A., Guerrero, M., y Barbero, A. (2022) RigoBERTa: State-of-the-Art Language Model For Spanish. *arXiv*. <https://arxiv.org/abs/2205.10233v2>.
- Vaswani, A., Shazeer, N., Parmar, N., Uzokoreit, J., Jones, L., Gómez, A. N., y Kaiser, L. (2017). Attention Is All You Need. En U. Von Luxburg, I. Guyon, y S. Bengio (Eds.) *NIPS'17: Proceedings*

of the 31st International Conference on Neural Information Processing Systems. Curran Associates, Inc.

- Valstar, M., Schüller, B., Smith, K., Almaev, T., Eyben, F., Krajewski, J., Cowie, R., y Pantic, M. (2014). Dimensional Affect and Depression Recognition Challenge. En *AVEC'14 Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (pp. 3-10). Association for Computing Machinery. <https://doi.org/10.1145/2661806.2661807>.
- Voleti, R., Liss, J. M., y Berisha, V. (2019). A Review of Automated Speech and Language Features for Assessment of Cognitive and Thought Disorders. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 282-298. doi: 10.1109/jstsp.2019.2952087.
- Wang, H., y Raj, B. (2017). On the origin of deep learning. *arXiv*.
- Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. <https://doi.org/10.1145/365153.365168>.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, 4, 301-350. <https://doi.org/10.1007/BF00410640>.
- Wiener, N. (1965). *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press.
- Winograd, T. (1971). *Procedures as a representation for data in a computer program for understanding natural language*. MIT Press.
- Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1), 191. [https://doi.org/10.1016/0010-0285\(72\)90002-3](https://doi.org/10.1016/0010-0285(72)90002-3).
- Wittgenstein, L. (1998). *Tractatus Logico-Philosophicus* (471ª Ed.). Dover Publications. [Originalmente publicado en 1922].
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Von Platen, P., Ma, C., Jernite, Y., Plu, J., Canwen, X., Le

Scao, T., Gugger, S., ... Rush, A. M. (2020). HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv*. <https://arxiv.org/abs/1910.03771>.

Zahedi, K. (2008). Determinist Inquiries: Debates on the Foundation of Language. *International Journal of Criminology and Sociological Theory*, 1(1), 26-50.

Zhang, W., Itoh, K., Tanida, J., y Ichioka, Y. (1990). Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics*, 29(32), 4790-4797.