UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Computación

Modelos computacionales para la caracterización de estados mentales alterados

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Ciencias de la Computación

Facundo Carrillo

Directores de tesis: Dr. Diego Fernández Slezak

Dr. Mariano Sigman

Consejero de estudios: Dr. Agustín Gravano

Buenos Aires, 2019

Modelos computacionales para la caracterizacion de estados mentales alterados

Hoy en día, la mayoría de las áreas profesionales usan las Ciencias de la Computación (CS) como uno de sus fundamentales componentes. En muchas actividades, la penetración de tecnologías heredadas de las CS son imprescindibles, no solo para generar alta competencia, sino porque proponen avances disruptivos e imposibles de alcanzar sin estas. La inteligencia artificial probablemente sea el área de CS que mayor avanzó en la última década, dentro de la inteligencia artificial los sistemas que usan recursos de Procesamiento del Lenguaje Natural (NLP) son incontables. El abanico donde las herramientas de NLP cumplen un rol protagónico es muy amplio. En particular, son fuertemente usadas para cuantificar diversos atributos, no sólo sobre qué está hablando un mensaje sino sobre la intención y el estado del sujeto que lo produce.

En este trabajo construimos y usamos herramientas de NLP para caracterizar los estados mentales alterados a través del discurso como privilegiada ventana a la mente. A partir de este marco de trabajo, aplicamos distintas técnicas y definimos algoritmos que nos permiten modelar las propiedades particulares de cada alteración. Para validar el marco aplicamos esta estrategia en diferentes casos.

En el capítulo de estados mentales alterados por patologías, estudiamos diferentes casos. En primer lugar encontramos que el discurso de pacientes con esquizofrenia se ve alterado y que esta alteración es capturada por nuestro algoritmo de coherencia. Nuestro algoritmo mide cómo se vinculan semánticamente frases sucesivas. En el caso de los sujetos patológicos, observamos que estos tienen un nivel de coherencia medio inferior al del grupo control y a su vez presentan disrupciones más pronunciadas. Como continuación de este caso, estudiamos sujetos prodrómicos de psicosis por esquizofrenia, es decir sujetos que aún no presentan síntomas pero pertenecen a un grupo de riesgo de generarla. En este estudio vemos que analizando los niveles de coherencia podemos, aplicando un algoritmo de aprendizaje automático en un

contexto de validación cruzada, distinguir cuáles sujetos en el futuro desarrollarán psicosis por esquizofrenia y cuáles no. Otro caso de estudio es el de sujetos bipolares. En este caso vemos como la valencia del lenguaje, en términos de positividad y negatividad del uso de palabras, diferencia los sujetos patológicos. Usado la misma estrategia de aprendizaje supervisado, encontramos que podemos clasificar con buen nivel de performance entre los dos grupos. Por último, estudiamos el uso de herramientas de NLP y de aprendizaje automático de una perspectiva ortogonal a la anterior. Creamos un modelo que nos permite seleccionar qué sujetos son aptos para un tratamiento psicofarmacológico con psilocibina para depresión crónica y cuáles no. De esta manera mejoramos los valores de sensibilidad del tratamiento creando un test de susceptibilidad basado en herramientas de NLP disminuyendo la masa de sujetos intervenidos con resistencia al tratamiento.

Luego de estudiar los estados mentales alterados por patologías, trabajamos en dos casos de estudio de mentes alteradas por intoxicaciones farmacológicas. En este capitulo, medimos los efectos de un experimento con MDMA y con LSD donde caracterizamos los efectos en el lenguaje a partir de modelos computacionales.

Como último escenario, estudiamos como cambios endocrinos impactan en el lenguaje, modelando cuantitativamente con herramientas de NLP, cómo afectan estos cambios en el lenguaje.

Teniendo en cuenta los casos de estudios, esta tesis pone evidencia sobre cómo los métodos computacionales de NLP sumados a técnicas de aprendizaje automático son una herramienta útil y novedosa para estudiar el lenguaje. Particularmente en un escenario donde este se ve afectado producto de las alteraciones que subyacen a su productor: la mente.

PALABRAS CLAVES: estados mentales alterados, aprendizaje automatico, procesamiento del lenguaje natural, coherencia discursiva, inteligencia artificial

Computational models for the characterization of altered mental states

Nowadays, most professional areas use Computer Science (CS) as fundamental components. In many activities, the penetration of technologies inherited from CS are essential, not only to generate high competition but the CS allow to achieve disruptive advances. Artificial intelligence is probably the CS area that has developed more than any in the last decade, within artificial intelligence systems that use Natural Language Processing (NLP) resources are countless. The applications areas where the NLP tools play a leading role is very broad. In particular, they are strongly used to quantify different speech characteristics, not only what a message is talking about, but also the intention and state of the subject that produces it.

In this work, we build and use NLP tools to characterize altered mental state, using the speech as window to the mind. We tested this approach in different scenarios.

In the chapter altered mental state by pathologies, we study three different cases. First, we measured the incoherence level in schizophrenia patients and we built a model that allow to sort subjects between control subjects and patients. In the second case, we tested the coherence method in high risk patients of psychosis by schizophrenia. In this case we found that we could sort between those subjects that are going to convert to schizophrenic before they do it. Then we analyze subjects with bipolar disorder and we built a model to classify between patients or control subject. As last case of this chapter, we built a system that it measures the probability that a particular subject has to respond to a psychopharmacology treatment.

In the chapter, altered mental state by drug intoxications, we tested our approach in two different experiment. The first one, we replicated the results in a previous study with MDMA in a new cohort. Then we study in a small experiment the effect of the LSD in speech.

In the chapter, altered mental state by endocrine system. We quantified the effect of some hormones in two scenarios: the menstrual cycle and in pregnancy.

In this thesis, we developed new NLP tools and we combined them with machine learning methods to build models that characterize the alterations of mental states in different scenarios.

KEY WORDS: altered mental states, machine learning, natural language processing, discursive coherence, artificial intelligence

Agradecimientos

Al empezar a escribir esta página, empecé, borré y empecé de nuevo varias veces. Me permití un espacio para pensar por qué me siento tan agradecido y sin duda no es por la providencia ni el azar, sino por ustedes, aquellos que intentaré nombrar extensivamente a continuación y aquellos que sin duda olvidaré: perdón a estos últimos y gracias a todos.

El comienzo de casi cualquier proceso largo suele ser difuso, mi trabajo de los últimos años, plasmado en una experiencia de doctorado es uno de estos ejemplos. Como cualquier niño la curiosidad estuvo ahí latente, pero no fue gracias a nadie más que a mi familia llevarla a buen puerto. Pienso en la cantidad de tiempo que ellos estuvieron para mi, contestando mis preguntas no solo con respuesta sino con más preguntas, una actitud que agradezco mucho. Se me vienen a la mente charlas y conjeturas con mi madre, mis momentos con mi abuelo Kike compartiéndome sus ideas y dudas sobre el universo, o mis experimentos con mi abuela Marta, entre tantos otros. Gracias a toda mi familia por estar conmigo y fomentarme así.

Tengo el recuerdo en las épocas del colegio secundario al tener mis primeras experiencias programando, la sensación de poder manejar a mi computadora, imprimirle un comportamiento, fue un momento deslumbrante para mi. Recuerdo a Brian Curcio enseñándome Visual Basic.

Luego del colegio, me inscribí en Ingeniería en Informática en la UBA, esencialmente porque era la única opción que conocía, en el CBC tuve la suerte de conocer a Nicolás Echebarrena quien estaba anotado para la carrera de física en Exactas. Nico, aunque probablemente él no lo sepa, es un excelente divulgador científico que me enseñó un montón de física y de ciencias naturales en general, su entusiasmo contagioso hizo fundamentalmente que me cambiaria a Exactas.

Empecé mi carrera en computación y tuve la gran suerte de conocer a Rodrigo Castaño y a Federico Pousa con quien cursamos luego toda la carera, ambos dos fueron el mejor grupo que pude tener, unos genios de otro planeta ambos, súper

generosos y solidarios.

En ese momento disfrutada cursar computación, pero empezaba a sospechar que las ciencias naturales eran lo que me interesaba y tuve la suerte de tener en casa a una gran divulgadora científica. Con Lu, cursamos toda la carrera juntos, y todos los días ella me contaba todo lo divertido que aprendía cursando biología y se bancaba mis infinitas hipótesis donde simplificaba todo y creía revolucionar el mundo cada noche, pero lejos de callarme ella me entusiasmaba, no tengo duda que yo seguí la carrera científica por su fomento.

Por la misma época, descubrí la colección ciencia que ladra y el programa Proyecto G que hacia Diego Golombek, su programa de tv para mi fue muy importante, me conectaba con la ciencia y experimentos de una manera muy divertida. Contemporáneamente escuché a Alejandro Dolina hablar sobre el cuento de Borges *El idioma analítico de John Wilkins* y entendí la oportunidad que tenia la computación para modelar el pensamiento como un fenómeno natural.

Con curiosidad empecé a pensar mucho al respecto y tuve la tremenda suerte de toparme con una charla de Mariano Sigman. La charla fue genial, Mariano no solo les ponía las palabras exactas a las dudas que yo tenia, sino que presentaba experimentos y resultados de un montón de ellas. En ese momento descubrí que esto que me interesaba se llamaba neurociencias cognitivas.

Finalizando mi carrera de grado, me acerqué a Diego Slezak para hacer mi tesis. Desde ese momento trabajamos en los inicios de esta línea de investigación de una manera genial. Diego me ofreció un ambiente donde me pude desarrollar y él catalizaba mi entusiasmo y me ayudaba a conducirlo a algo concreto. Junto a Diego que fue mi director con Mariano, sumamos al trabajo diario al ídolo de Guillermo Cecchi, quien fue siempre tremendamente generoso con su tiempo dedicado y trabajar con él fue siempre un placer. A su vez, Mariano me abrió las puertas de su laboratorio en el departamento de física donde conocí un montón de gente súper interesante que me enseñó mucho y diferentes personas fuera del labo tambien interesadas en las

neurociencias cognitivas como Pablo Zivic.

Durante mi doctorado estuve financiado mayormente por una beca de Conicet, esto no hubiera sido posible sin el compromiso de dirigentes políticas que creyeran genuinamente que la ciencia debía ser motor fundamental del desarrollo del país. A su vez, a parte de este financiamiento, tuve la suerte de obtener dos financiamientos externos, el último fue una fellowship de Facebook que sin duda lo conseguí gracias al tremendamente generoso Greg Diuk que siempre, en diferentes instancias de mi doctorado, ayudó desinteresadamente de alguna forma.

Pienso que la industrialización de la ciencia nos facilita perder la brújula de la inquietud científica y caer en la burocracia del sistema de publicaciones. Para aislarme de esto, encontré un grupo de personas con quien discutir de ciencia y filosofía de la ciencia de una manera que apreció mucho: Alejo, Euge, Leti, Mec, Riera, Babino, Rieznik, Facu entre otros.

No puedo dejar de agradecerles a mis tres hijas con quien difruto aprendiendo infinito de como experimentan su entorno y son siempre un remedio perfecto en los dias donde los experimentos no dan. Por último y como parte más fundamental de toda esta experiencia, quiero agradecerle a Lucila Gallino. Nada de esto hubiera sido posible sin su incansable apoyo y amor. Gracias por hacer ciencia conmigo, discutir, enseñarme tanto y experimentar juntos en la crianza de nuestras hijas.

A las tres partes de mi que no soy yo,

Guada, Valen y Milu

Índice general

1. INTRODUCCIÓN

La mente, la relación con la materia y sus conflictos justificaron incontables teorías a lo largo de la historia de la cultura [1]. Sin embargo, en las ciencias naturales, la batalla entre materialistas e idealistas terminó hace tiempo y se arribó a un amplio consenso sobre practicas correctas de ciencia basadas en la observación, experimentación y replicación [2, 3] . Eludiendo entonces, las infinitas preguntas filosóficas que surgen a partir del concepto de la mente, en esta tesis abordamos el estudio de los estados mentales desde una perspectiva científica, pragmática usando modelos computacionales.

Tras superar históricamente los esquemas descriptivos dualistas donde la separación de mente y cerebro aun estaba en duda [1], el colectivo de las ciencias naturales ha avanzando, en los últimos años, en el conocimiento de este binomio complejo autoorganizado [4, 5]. Sin embargo, coexisten dos modos de estudios contrapuestos que no logran converger en un modelo integrativo de la mente.

Desde una perspectiva *bottom up* se intenta contribuir al entendimiento del cerebro/mente desde propiedades de las neuronas, como por ejemplo la dinámica de los canales iónicos [6], fundamentales para el funcionamiento de las células del cerebro, pasando por escalas físicas mas grandes donde se modelan las membranas de las células [7], experimentos donde células enteras son modeladas [8] hasta modelos de cómo fluye la información en redes complejas [9, 10]. Esta perspectiva, contribuye enormemente y es pieza fundamental de la neurociencias sin embargo aun se encuentra muy lejos de poder modelar fenómenos grandes de la órbita comportamental o cultural de los sujetos.

En el otro extremo, se encuentra una estrategia, que podríamos definir como *top down*, donde otros jugadores avanzan experimentalmente con metodologías contra-

puestas y con nuevas preguntas. La psicología experimental se ubica de este lado, donde se estudian preguntas de comportamiento y procesos cognitivos de más alto nivel ([11–13] entre otros). Si bien existen modelos validados y usados de esta perspectiva, tampoco es completa y no alcanza para describir un modelo integral que le permita ir desde fenómenos comportamentales hasta procesos de activación neuronal.

Estos dos perspectivas incluso se involucran en los mismos casos de estudios. Si un sujeto toma una droga en particular, por ejemplo MDMA, podemos entender que sucede farmacológicamente [14], se puede modelar bien la relación de afinidad con determinados receptores [15] , incluso modelar la distribución de receptores según áreas del cerebro [16], la biodisponibilidad de la molécula etc. A su vez, podemos describir con cierta precisión los cambios psicológicos del sujeto, los cambios comportamentales, el estado de ánimo, etc [17–20]. Sabemos que esos cambios mentales se producen debido a que el cerebro se encuentra en un estado distinto por la interacción farmacológica, pero no podemos describir ni predecir con un solo modelo fenómenos de la órbita neuronal y el cambio del estado mental o de conciencia alterada.

En general, exceptuando algunos procesos donde se pueden unir estos dos mundos, por ejemplo, la tarea de reconocer una cara es medida en la activación de una sola neurona [21], la neurociencia aun esta lejos de tener un modelo integrativo.

Una diferencia sustancial de las dos estrategias es que la segunda, al menos una gran parte de esta, usa como metodología de estudio herramientas menos objetivas que la otra, pues la psicología cognitiva, como disciplina fundamental, se acerca a la mente mayormente a través de sus producciones subjetivas y no se su sustento cerebral, como si lo hacen otras disciplinas como la biología molecular, electrofisiología, etc. Por ejemplo, diversos protocolos psicológicos usan como dato las respuestas de los pacientes a percepciones introspectivas. En algunos casos, incluso el experimentador tras registrar la respuesta del paciente, debe cuantificarla en una escala a

partir de ciertas reglas, aportando nuevamente su subjetividad. Esta característica no convierte a este enfoque en ciencia menos útil o valida por el contrario, solo exigen en la creación y diseño de protocolos que permiten tomar muestras validando hipótesis complejas, que luego resulten reproducibles.

Sin embargo, aceptando la tesis de Alan Turing [22], podríamos modelar la mente como un proceso computable independientemente de su intangibilidad y la de sus producciones. Teniendo en cuenta esto, la perspectiva aportada desde las ciencias de la computación a la neurociencia cognitiva puede ser y progresivamente lo es, enorme, pues aporta métodos objetivos y cuantitativos complementando un área con una gran porción de análisis cualitativo. Esta perspectiva es usada a lo largo de esta tesis.

Los modelos computacionales pueden tomar distintas producciones de la mente como entrada pues no existe una única producción a modelar sino diversas. Cada una permite caracterizar la mente desde ángulos distintos y a si ha avanzado el estudio de los mismos, algunas de las producciones estudiadas ampliamente por la neurociencias son: la toma de decisiones [23–25], cómo aprenden y se enseñan los sujetos [26–32], cómo se forman estrategias [33, 34], cómo almacenan y evocan memoria [35, 36], cómo se integran diferentes productores de información construyendo la conciencia y aspectos de la misma [37–40].

De todas las producciones posibles, en esta tesis usamos como ventana a la mente el **lenguaje** - el producto exclusivo de la mente humana (donde exclusivo nos referimos a la capacidad de un lenguaje complejo [41]). Para estudiar el lenguaje de los sujetos creamos y aprovechamos condiciones de alteraciones mentales para poder entender los mecanismos y relaciones entre cerebro/mente y lenguaje siempre modelando la fenomenología a partir de modelos de inteligencia artificial. Es decir, realizamos experimentos donde controlamos la alteración mental de los sujetos participantes y obtenemos producciones de texto de los mismos como respuesta a alguna consigna particular (por ejemplo, contar un sueño o relatar un recuerdo

cercano). Luego basándonos en la descripción de la fenomenología creamos modelos basados en procesamiento del lenguaje natural y *machine learning* que nos permiten caracterizar automáticamente la mente de los sujetos.

En las ultima década, los modelos de inteligencia artificial han crecido sin lugar a duda. Diversos factores contribuyeron a esto, sobre todo por uno fundamental, la masiva disponibilidad de datos [42]. Gracias a la penetración tecnológica de computadoras y teléfonos móviles con conexión a Internet, la población todo los días suma más productores de información que se registran en distintos bancos de datos, muchos de esos públicos como algunas redes sociales. Según la Unión Internacional de Telecomunicaciones [1] estima que 3200 millones de personas estaban conectadas a Internet en 2015. A su vez, los distintos proveedores de servicios lanzan políticas donde permiten a los usuarios, no solo guardar información y compartirla sino también *anotarla* o etiquetarla. Esto les permitió, a los grandes dueños de información, crear repositorios de datos etiquetados con categorías, valoraciones etc, a partir del uso ocioso de la gente en la red, inaugurando el concepto de *Human Computation* [43–48] (por ejemplo, cuando subimos una foto y etiquetamos una cara, cuando jugamos a un juego y describimos en texto una imagen, cuando proponemos una mejora en una traducción de un servicio de traducción automática, etc). Gracias a este fenómeno de masificación de acceso, aumento en la capacidad y el cambio de arquitecturas de computo [49], y avances en teoría de modelos de aprendizaje automático, un sin fin de tareas que hace algunas décadas eran imposibles de abordar por una computadora, hoy en día son resueltas por algoritmos con niveles comparables, y en algunos casos mejores, que los seres humanos [50–53].

Teniendo en cuenta estas dos características actuales, la falta de técnicas cuantitativas computables sobre el lenguaje en ciertos ámbitos de las neurociencias cognitivas, y a su vez, el avance desde la ciencias de la computación en inteligencia

[1] http://www.internetlivestats.com/internet-users/

artificial con amplia disponibilidad de datos masivos, proponemos el siguiente objetivo: **Modelar, con herramientas de procesamiento del lenguaje natural y aprendizaje automático, los estados mentales a partir de transcripciones del discurso de sujetos con mentes alteradas y así mejorar la caracterización mental de los sujetos con herramientas más objetivas, cuantitativas y computables**. Para satisfacer el objetivo general planteamos 3 objetivos particulares. Objetivo 1: Estudiar los casos donde la alteración se producen por patologías psiquiátricas. Logrando modelar con herramientas de procesamiento del lenguaje natural e inteligencia artificial diferentes propiedades del lenguaje. Objetivo 2: Estudiar los cambios en lenguaje en situaciones donde la alteración mental se produce por una interacción farmacológica conocida y controlada. Objetivo 3: Estudiar casos donde diversos cambios endocrinos de situaciones naturales generan alteraciones en el lenguaje medibles con modelos computacionales.

Estructura de la Tesis

La tesis esta organizada en 6 capítulos. El primero, la **Introducción**, donde se plantea el marco de trabajo y desarrollo de los trabajos luego expuestos.

El capítulo 2, **Métodos**, donde presentamos los diferentes métodos desarrollados y usadas en la tesis. Primero describimos el algoritmo de coherencia desarrollado para modelar alteraciones del discurso en pacientes con esquizofrenia (ver 2.1). Luego presentamos un nuevo algoritmo de similitud semántica de alta frecuencia y su uso como modelo de *word embedding* (ver 2.2). A continuación detallamos los distintos modelos de análisis de sentimiento usados a lo largo de la tesis (ver 2.3) y una breve descripción de métodos de análisis de gramáticas (ver 2.4). Posteriormente presentamos un modelo basado en grafos del discurso que desarrollamos para medir particularidades de sujetos con trastorno bipolar (ver 2.5). Para cerrar el capitulo describimos los conceptos básicos usados de *aprendizaje supervisado* (ver 2.6) y algunas herramientas estadísticas de corrección de múltiples comparaciones usadas

(ver 2.7).

El capítulo 3,**Estados mentales alterados por patológicas**, donde estudiamos distintos casos de estudios de sujetos con patologías psiquiátricas. Primero presentamos los resultados de estudiar pacientes esquizofrénicos (ver 3.1) y presentamos resultados de clasificación automática como asistencia al diagnóstico, luego estudiamos una población muy particular de sujetos de alto riesgo de desarrollar psicosis con una metodología similar al caso anterior (ver 3.2). A continuación, trabajamos con sujetos con trastorno bipolar (ver 3.3) donde también reportamos resultados de clasificación automática. Para cerrar el capitulo proponemos un modelo para predecir la susceptibilidad a un tratamiento psicofarmacológico antes del mismo para pacientes con depresión (ver 3.4).

El capítulo 4, **Estados mentales alterados por intoxicaciones farmacológicas**, donde estudiamos dos casos de estudios de alteraciones en el discurso en sujetos bajo efecto de dos drogas psicoactivas: MDMA y LSD. Para el caso de MDMA, replicamos con una segunda cohorte el experimento presentado en [54] estudiando los efectos de intercambiar los diferentes modelos entrenados en cada grupo de sujetos y analizamos cuan reproducible son los resultados previos (ver 4.1). Luego estudiamos los efectos en la emotividad en sujetos bajo el efecto de LSD (ver 4.2).

El capítulo 5, **Estados mentales alterados por cambios endocrinos**, donde estudiamos como cambios endocrinos producen alteraciones cognitivas usando el lenguaje como proxy. En este capítulo estudiamos dos casos. El primero, usamos el ciclo menstrual como caso de estudio en una población de usuarios de una red social donde analizamos como el uso del lenguaje es afectado por los cambios en la dinámica de concentración de las diferentes hormonas (ver 5.1). Luego, el segundo caso de estudio, estudiamos en las mismas condiciones anteriores, misma red social, los cambios en el lenguaje pero estaba vez en sujetos embarazadas, donde entendemos como esta condición y la dinámica hormonal produce cambios medibles en el lenguaje (ver 5.2).

El capítulo 6, donde discutimos los resultados reportados en la tesis analizando las limitaciones y extensiones de lo encontrado. También presentamos las direcciones de trabajo futuro de la linea de investigación propuesta en esta tesis.

2. MÉTODOS

A lo largo de esta tesis usamos distintos métodos y algoritmos, en este capitulo se definen varios de ellos y se detallan particularidades de las implementación pertinentes.

2.1. Algoritmo de coherencia

La coherencia es un concepto lingüístico definido no formalmente, por esto cuenta con distintas versiones. Por ejemplo, la Real Academia Español la define como: "Conexión, relación o unión de unas cosas con otras. Estado de un sistema lingüístico o de un texto cuando sus componentes aparecen en conjuntos solidarios" [1], el diccionario de Cambridge lo define como "the situation when the parts of something fit together in a natural or reasonable way" [2].

Estas vagas e intuitivas nociones son usadas por la psiquiatría para describir los pensamiento de sujetos psicóticos. Inspirados en esto, con una motivación exclusivamente pragmática, nos propusimos crear, mediante un simple algoritmo, una sesgada pero computable definición de coherencia. Si bien nuestro algoritmo resultó novedoso, profesionales de nuestra disciplina han aportado distintas interpretaciones del concepto de coherencia y por lo tanto han implementado distintas mediciones. Algunos trabajos destacados fueron desarrollados por Elvevag et al. en [55], y especialmente por Danielle McNamara en su libro *Coh-metrix* [56] y en otros trabajos [57, 58].

Nuestro algoritmo tiene como objetivo medir la coherencia semántica de un texto usando las oraciones como unidades, buscamos de esta manera que *encajen jun-*

[1] http://dle.rae.es/?w=coherencia

[2] http://dictionary.cambridge.org/dictionary/english/coherence

tas de una manera razonable como se define en el diccionario de Cambridge. Para lograr dicho propósito necesitamos previamente contar con alguna representación estructurada de la semántica de una frase. Existen diferentes maneras de encontrar representaciones estructuradas del lenguaje natural. En nuestro caso, como decidimos estudiar las relaciones semánticas usamos una representación vectorial (o *word embedding* en ingles) del léxico.

Las representaciones vectoriales son modelos del lenguaje usados en procesamiento del lenguaje natural (NLP) donde las frases del vocabulario son asignadas a vectores de números reales. Por lo general, dichos modelos son generados con algoritmos, que a partir de un corpus, generan una relación entre palabras y vectores. Estos métodos usan distintos familias de algoritmos para lograr el *embedding*. Algunos métodos usan la co-ocurrencia de aparición de palabras en documentos y luego usan métodos de reducción de dimensionalidad. Por ejemplo con *Latent Semantic Analysis* (LSA) [59] se puede conseguir un *embedding* de palabras a partir de una matriz de co-ocurrencia de palabras en documentos y posteriormente una reducción de dimensionalidad usando *Singular Value Descomposition*. Otros métodos se basan en modelar con redes neuronales la probabilidad de aparición de una palabra en función a su contexto, una implementación posible de esta política es el algoritmo *Word2Vec* [60] en sus distintas versiones. En resumen todos los modelos de *word embedding* se encargan de aproximar la función de similitud semántica en un corpus mediante el posicionamiento de palabras en un espacio de $\mathbb{R}^n$, donde n es un hiper parámetro del modelo a decidir, teniendo en cuenta que distintos valores condicionan al modelo de diferentes formás (un estudio de los efectos del tamaño de la dimensionalidad se puede ver en [61]). En la sección 2.2 presentamos un algoritmo de representación semántica desarrollado por nosotros.

Los modelos de *word embedding* tienen distintas propiedades relacionadas con la posición especial de los vectores. La característica más relevante que deben cumplir es que dos palabras semánticamente relacionadas deben tener una representación

vectorial, que bajo cierta métrica, estos dos vectores estén *cerca*. Por ejemplo, LSA usa la distancia coseno, es decir el angulo entre los dos vectores, como medida de cercanía de palabras. Esto quiere decir que dos palabras semánticamente relacionadas, como *perro* y *gato*, les corresponden dos vectores con distancia coseno pequeña, por otro lado, dos palabras no relacionadas semánticamente como *gato* y *helicóptero* deben tener una mayor distancia coseno. Por supuesto todos estos métodos son procedimientos que intentan modelar el espacio semántico del lenguaje natural, por lo que no existe una única forma de medir validar un modelo, ni tampoco existen garantías o cotas de los errores que estos modelos pudieran estar generando. Esta características no impiden que sean usados para modelar las relaciones semánticas del lenguaje en diferentes ámbitos. A demas de la propiedad de cercanía semántica y cercanía vectorial, estos modelos a veces tienen otras propiedades: por ejemplo el método Word2Vec ostenta poder capturar propiedades que permiten hacer *aritmética semántica*, en [62] presentan el ejemplo que si se toman los vectores de *rey* y se le restan el vector de *hombre* y se le suma el vector *mujer* el vector resultante es muy cercano al vector que representa el concepto *reina*.

Nuestro algoritmo de coherencia usa un modelo de *word embedding* ya entrenado. Para evaluar la coherencia de un texto, realiza los siguientes pasos. El texto es separado en frases. Por cada frase, se reemplazan las palabras por los vectores del *word embedding* elegido y se promedian. De esta manera, por cada frase se obtiene un vector que representa a la oración original. Habiendo hecho este cambio, un texto que consistía en una lista de frases, ahora es representado por una lista de vectores. A partir de esta lista de vectores, calculamos dos series distintas, que llamamos: coherencia de primer orden (COH_1) y coherencia de segundo orden (COH_2). La COH_1 corresponde a la serie generada a partir de medir distancia coseno entre vectores contiguos. La COH_2 representa a la distancia coseno entre vectores que tienen otro vector en medio, es decir:

Si un texto T es representado por la siguiente lista de vectores:

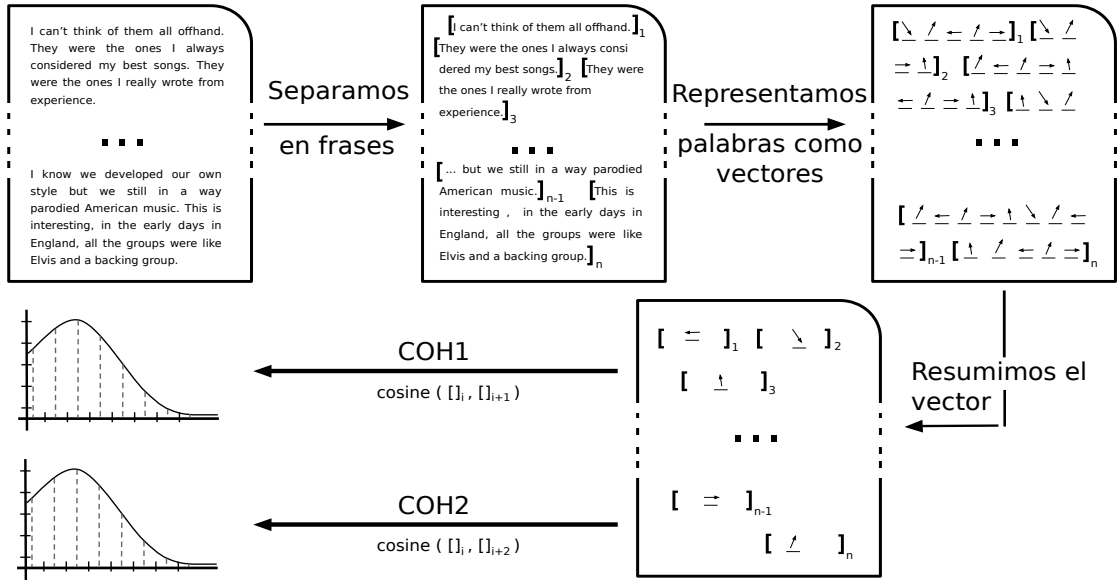

Fig. 2.1: Resumen de Algoritmo de Coherencia

$$T =< Vector\_Promedio\_Oracion_1, ..., Vector\_Promedio\_Oracion_n > \quad (2.1)$$

Definimos:

$$COH_1 =< cos(T_1, T_2), ..., cos(T_i, T_{i+1}), ..., cos(T_{n-1}, Tn) > \quad (2.2)$$

y

$$COH_2 =< cos(T_1, T_3), ..., cos(T_i, T_{i+2}), ..., cos(T_{n-2}, Tn) > \quad (2.3)$$

Teniendo definidas estas dos distribuciones, caracterizamos la coherencia de un texto tomando medidas estadísticas de cada serie, como por ejemplo: media, mediana, mínimo, máximos, desvío estándar y percentiles. La Figura 2.1 (similar a la presentada en [63]) resumen el método de coherencia desde el texto hasta la creación de las dos distribuciones de coherencia.

2.2. Twitter Semantic Similarity y su transformación como modelo de *word embedding*

Cada modelo de *word embedding* tiene distintas propiedades en función a diversos factores como el corpus de entrenamiento, la definición de distancia o similitud, etc. Otra propiedad importante en varios de estos modelos, por ejemplo LSA, es que una vez que se calcula la representación semántica (y en el caso de este método es computacionalmente costoso hacerlo), esta no puede ser modificada. Es decir, si se quieren agregar documentos nuevos al corpus el cálculo de la proyección al espacio de R^n debe ser nuevamente calculado. Si bien este método es útil y está validado de manera exitosa para diferentes tareas, no resulta el mejor método para estudiar preguntas donde es necesario cambiar frecuentemente el corpus. Por eso, teniendo en cuenta esta propiedad e inspirados en el algoritmo Google Semantic Distance [64] creamos un nuevo método que llamamos *Twitter Semantic Similarity* (TSS).

TSS es un método de representación de relaciones semánticas que usa Twitter como corpus de entrenamiento. Este método aporta, diferenciándose de los métodos anteriormente presentados, la posibilidad de modelar las relaciones semánticas que subyacen a toda la red social *Twitter* en un escenario de dinámica de alta frecuencia. Esto quiere decir, que se puede computar las relaciones semánticas entre conjuntos de palabras a lo largo de una nueva dimensión: el tiempo. Con TSS armamos una herramienta que nos permitió estudiar la dinámica de cambios de las relaciones semánticas a lo largo del tiempo con precisión tan alta como 1 modelo por segundo.

Esta característica es oportuna en este escenario. *Twitter* es una de las red sociales de Internet más grandes del mundo. En esta red los usuarios pueden compartir mensajes de hasta 140 caracteres (llamados *tweets*). Es tan grande la cantidad de usuarios, que esta red tiene récords de aproximadamente 150 mil tweets por segun-

do [3]. Dado este volumen de información y las restricciones que su API REST [4] provee resultaria inviable o al menos muy dificil usar un esquema parecido a LSA o Word2vec donde el corpus es bajado y reducido su dimensionalidad. Por eso, ideamos una estrategia similar a la de Google Semantic Distance.

Nuestro algoritmo se basa en mirar la *co-ocurrencia* de aparicion de palabras en los *tweets* suponiendo que dos palabras semánticamente relacionadas tendrían una co-ocurrencia mayor que dos palabras no relacionadas semánticamente. El problema de este abordaje es que exige medir la aparición de diferentes palabras en una cantidad de tweets. Como mencionamos, la cantidad tweets puede ser tan grande que haga que esta tarea sea inviable, es por eso que propusimos usar la *velocidad* de producción de tweets como estimador de la cantidad de mensajes en una ventana corta de tiempo.

Por ejemplo, dada la palabra *perro*, le pedimos a la API REST que obtenga una lista con los últimos 200 tweets en la que esta palabra aparece. Cada mensaje cuenta con la fecha y hora a la cual fue producido, esta serie derivada de 200 tiempos la llamamos t. Generalizando, para la palabra w le pedimos a la API REST los últimos n tweets y definimos:

$$velocidad\_produccion\_media(w) = \frac{\sum_{i=1}^{n-1}(t_{i+1} - t_i)}{n-1} \tag{2.4}$$

$$freq(w) = (velocidad\_produccion\_media(w))^{-1} \tag{2.5}$$

Teniendo definida la estimación de frecuencia de aparicion de una palabra, definimos TSS como:

$$TSS(w1, w2) = (\frac{freq(w1 + w2)}{max(freq(w1), freq(w2))})^{\alpha} \tag{2.6}$$

[3] https://blog.twitter.com/engineering/en_us/a/2013/new-tweets-per-second-record-and-how.html

[4] https://dev.twitter.com/rest/public

Es decir, para calcular la similitud entre la palabra $w1$ y $w2$ calculamos primero la frecuencia de producción de cada una por separado, luego la frecuencia de producción del *string* que tiene $w1$ y $w2$. A continuación normalizamos este último valor por la máxima frecuencia de cada palabra por separado. Vale mencionar que cuando calculamos la frecuencia del *string* que tiene $w1$ y $w2$ la API REST retorna los mensajes que contienen ambas palabras en cualquier orden. Luego elevamos todo por un factor α para obtener una mejor escala de la medida, en los ejemplos siguientes fijamos $alpha = \frac{1}{4}$. En el caso donde no ocurran tweets con $w1$ y $w2$ definimos $TSS = 0$. Esta definición hace que TSS alcance el valor de 1 para palabras totalmente asociadas y 0 para cualquier par de palabras no semánticamente relacionadas.

Habiendo definido esta nueva medida lo primero que probamos fue su relación con metodologías y corpus ya fuertemente validadas. Para eso armamos a partir de los 1500 sustantivos más usados en inglés, 100 mil pares de palabras al azar y medimos TSS, similitud usando distancia coseno en LSA entrenado con TASA y algunas medidas de similitud de Wordnet. Los resultados muestran que existe una fuerte correlación lineal entre TSS y las demás medidas. Para TSS vs LSA obtuvimos un $\rho = 0{,}2199$ con $p-valor < 10^{-280}$. Para Wordnet, usamos distintas medidas pero los resultados fueron similares: el camino más corto usando vínculos de hipernimia e hiponimia ($\rho = 0{,}298, p-valor < 10^{-275}$), para la *theoretic definition similarity* [65] obtuvimos $\rho = 0{,}15, p - valor < 10^{-67}$ y para la similitud basado en contenido de Resnik [66] $\rho = 0{,}157, p - valor < 10^{-74}$. Estas comparaciones siguieren que TSS es una medida que captura, de una manera similar, la función de cercanía semántica que los otros métodos modelan.

Tras haber validado que TSS responde para palabras comunes de una manera similar a otros métodos decidimos probar cuán bien podíamos embeber nuestra medida en un espacio vectorial. Para eso usamos *Multidimencional Scalling* (MDS) [67]. MDS consiste en hacer transformaciones no lineales a partir de una matriz de distancias o similitudes intentando que la distancia (por ejemplo euclidiana), entre

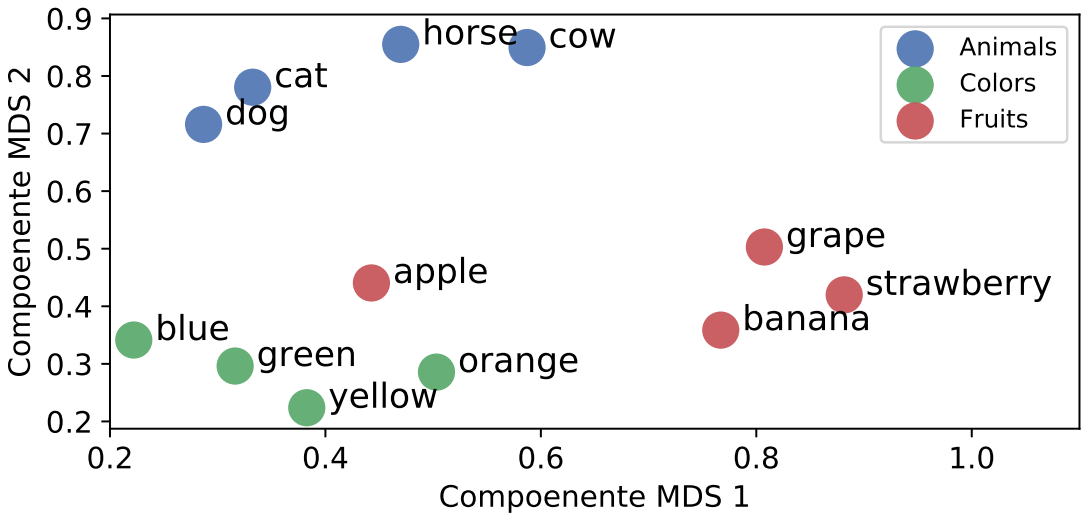

Fig. 2.2: Embedding en dos dimensiones hecho con MDS a partir de la matriz de similitud entre las palabras calculada con *Twitter Semantic Similarity*.

todos los puntos en el espacio embebido, se preserve relativamente lo más posible en función a la matriz de distancia original. La Figura 2.2 muestra, a modo de ejemplo, como se proyectan, usando MDS, un conjunto de palabras a partir de la matriz de similitud de TSS en dos dimensiones. Las palabras comunes fueron elegidas en tres categorías manualmente. En principio se ve que las tres categorías están bien separadas, a excepción de una sola palabras *apple*. Esto probablemente se deba a que esta palabra es altamente polisémica, más en el contexto de Twitter donde la acepción referida a la marca tecnológica tiene mucho peso. A su vez, también se nota como dentro de los grupos hay cierto orden, por ejemplo para el grupo de animales se ve que por un lado se encuentran los animales domésticos (*dog* y *cat*), y por otro lado, los dos animales de granja (*horse* y *cow*). La matriz de similitud calculada por TSS para este conjunto de palabras varia a lo largo del día, sin embargo para este conjunto, las distintas variaciones conservan las propiedades descriptas pues son conceptos fuertemente estacionarios.

Tras estudiar estas cotidianas palabras, decidimos medir la similitud entre los nombres de los países. En este caso, tomamos la medición de similitud por varios días dos veces por día y luego usamos la matriz de similitud promedio. Si bien tomamos esta estrategia para tener una muestra representativa de varios días, cuando miramos cada matriz de similitud particular, esta presenta la similitud entre palabras similar a la de la matriz promedio. Este ejemplo determinó que la medida TSS estaba capturando distintas propiedades semánticas. La Figura 2.3 muestra el *embedding* en dos dimensiones hecho con MDS a partir de la matriz de similitud entre los nombres de los países. Lo primero que muestra esta figura es que se respetó la organización geográfica de los países. Se conserva la organización por continente (los colores de los puntos representan la agrupación por continente agregados manualmente). A su vez, al mirar la figura encontramos que también existía una organización secundaria en el plano. Los países, dentro de cada continente se organizaban en función a su riqueza. Para medir este efecto medimos la relación entre el producto bruto interno de cada países y la distancia hacia el centro de masa de los 5 países más ricos y encontramos una correlación negativa ($rho = -0{,}58, p - valor < 10^{-4}$), a su vez graficamos el tamaño de cada punto en relación al producto bruto interno. Cabe destacar que la correlación con el PBI probablemente esta capturando algún tipo de sub-relación semántica mas compleja y que este índice solo sea un *proxy* de esta propiedad oculta.

Habiendo validado que la nueva medida se comportaba según lo esperado en un escenario estático, tanto por comparación con procedimientos estándares como en los experimentos de proyección de palabras, decidimos estudiar el efecto en un escenario dinámico. Para esto elegimos una fecha particular donde ocurrió un evento que considerábamos que reconstituiría las relaciones semánticas entre países. Tomamos el evento del sorteo de grupos para el mundial de fútbol 2014. En este evento la FIFA transmitía en vivo para todo el mundo un sorteo donde distintos participantes sacaban pelotas de un bolillero y asi constituían los diferentes 8 grupos del mundial.

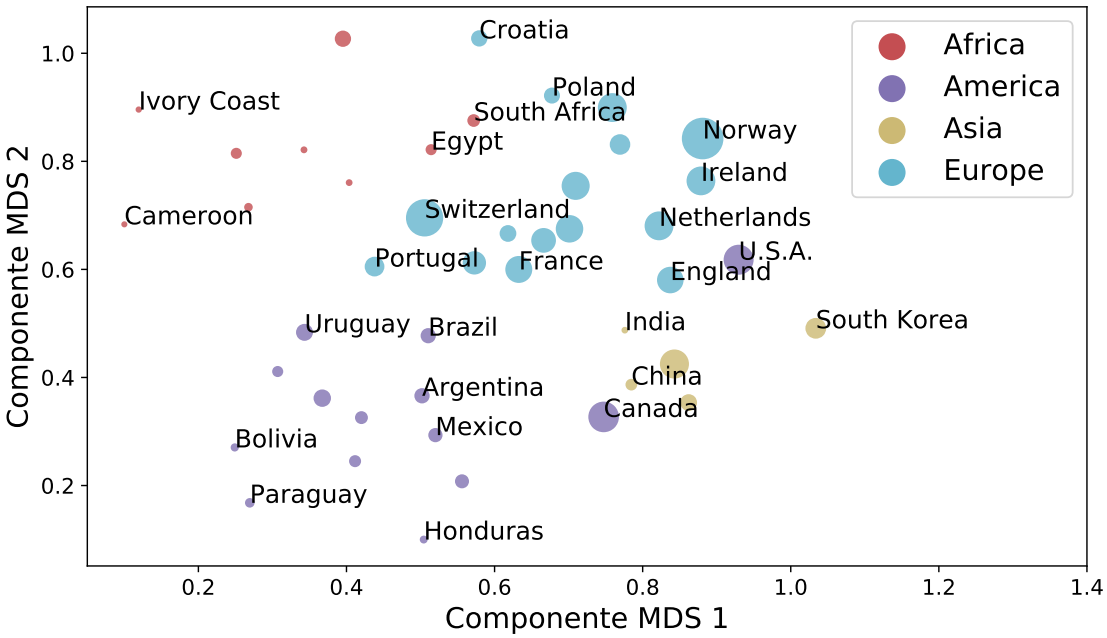

Fig. 2.3: Embedding en dos dimensiones hecho con MDS a partir de la matriz de similitud entre los nombres de los países calculado con TSS. Los colores de los puntos corresponden al continente al cual los países pertenecen. El tamaño de cada punto se relaciona con el valor del producto bruto interno de cada país.

Cada uno reúne 4 países que debían competir entre si todos contra todos. Este evento ocurrió el 4 de Diciembre del 2013.

Para estudiar la dinámica de cambios semánticos elegimos 64 países, 32 países clasificados al mundial (QT), aquellos que serian sorteados, y 32 que no (NQT). Luego con estos 64 países medimos la TSS entre todos ellos cada 2 horas desde el 3 de Diciembre al 14. A su vez, durante el sorteo, tomamos mediciones cada 2 minutos. La Figura 2.4 muestra las matrices de similitud para 4 momentos particulares: 1) 3 días antes del sorteo (D-3); 2) Justo antes del sorteo (D-1), exactamente dos horas antes del sorteo; 3) Justo después del sorteo (D+1), exactamente una hora después del mismo; y 4) Una semana después del sorteo (D+7). En principio se ve diferencia significativa (t-test $p - valor < 10^{-5}$ entre los valores de similitud) entre las TSS para el momento D-3 entre QT y NQT, (promedio y desvió estándar $TSS_{D-3}(QT) = 0{,}11 \pm 0{,}0089, TSS_{D-3}(NQT) = 0{,}07 \pm 0{,}0085$). Esto era de esperar pues, tres día antes del sorteo, las personas especulaban sobre como se agruparían los países ya clasificados. Este efecto se magnifica para los sub-siguientes momentos.

La figura muestra un segundo efecto, además de acercarse los paises de QT tras el sorteo, estos se organizan fuertemente tal como quedaron agrupados por el sorteo de grupos. Es decir, este evento hace que la TSS entre los nombres de los países cambie radicalmente la organización basada en la disposición geográfica a una nueva disposición meramente cultural producida por el evento particular del sorteo.

Como la organización por grupos del mundial fue tan fuerte cuantificamos este efecto con un esquema de clasificación. Dado un país elegíamos los 3 países más cercanos en la función de similitud definida por TSS y considerábamos que la clasificación era buena si coincidía con la del mundial. Por ejemplo, si de los tres países más cercanos, dos correspondían al mismo grupo definido por el sorteo, obteníamos una clasificación de 0.66. La Figura 2.5 muestra la *performance* de este clasificador medido como la tasa media de cantidad de países bien clasificados. Para cuantificar cuál era el efecto de elegir al azar para tener noción de cuan bien estaba andando el

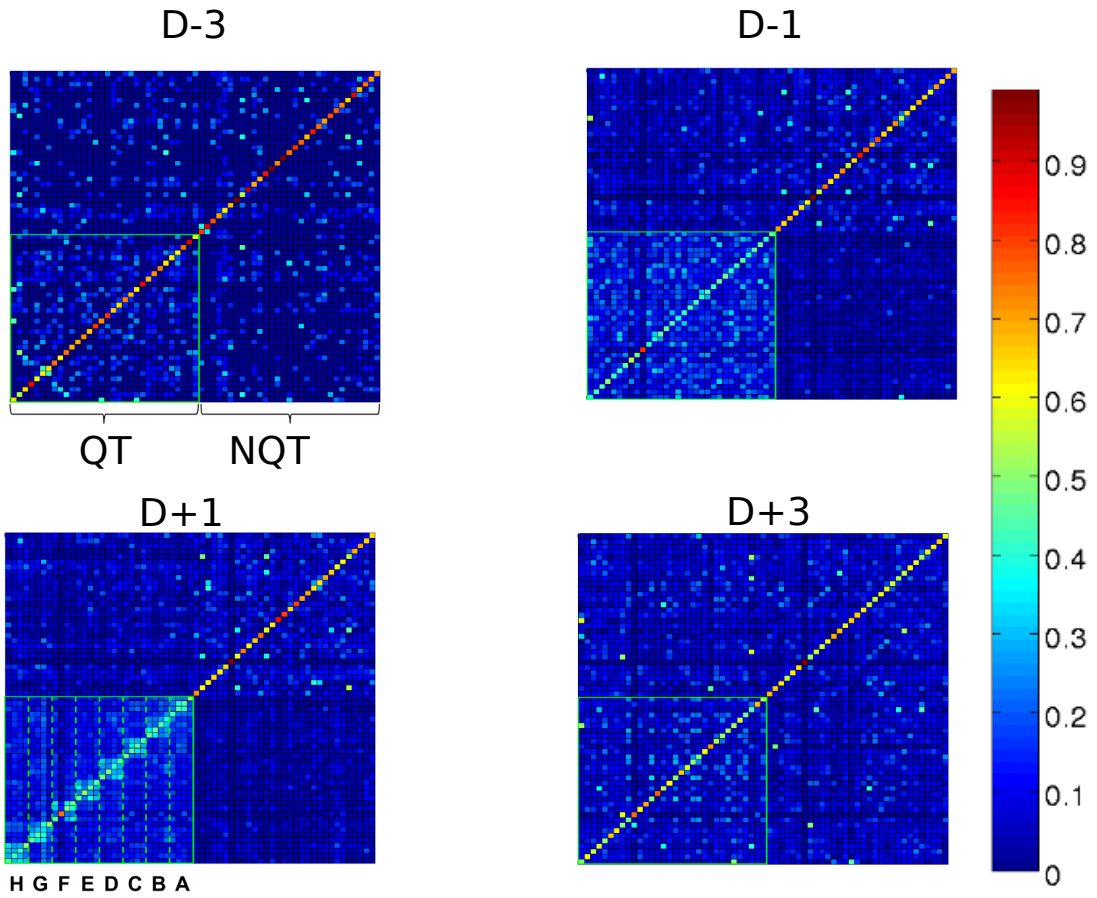

Fig. 2.4: Reorganización de la red semántica causada por el sorteo de grupos del Mundial. Cada matriz de similitud ordena primero los 32 países en clasificados (QT) y luego los no clasificados (NQT). A su vez, dentro de los de QT se ordenaron por como quedaron los grupos. Se muestran las 4 instancias de tiempo: 1) 3 días antes del sorteo (D-3); 2) Dos horas antes del sorteo (D-1); 3) Dos horas después del sorteo (D+1); 4) Una semana después del sorteo (D+7). Cada linea fue normalizada por norma 1 para que se visualice mejor. Valores cercanos a 1 corresponden a alta similitud semántica.

clasificador hicimos el mismo experimento permutando las matrices muchas veces. De esta manera calculamos empíricamente el valor del azar para esta medida, que se ubicó en $0{,}1471 \pm 0{,}0004$. Teniendo noción de cuánto clasifica un modelo totalmente azaroso, se ve en la figura que antes del sorteo la red semántica no aportaba información de como quedarían los grupos del mundial. Luego, a medida que el sorteo iba develando esta información, usando el método que dijimos y midiendo la matriz de similitud cada 2 minutos vemos que la *performance* aumenta rápidamente a valores cercanos 0.8 y luego vuelve a caer. Si bien la caída es rápida, es decir, un día después del sorteo la *performance* desciende de 0.8 a 0.25 el nivel se conserva incluso una semana después por encima del valor inicial e incluso un poco mayor al valor del azar. Es decir, el evento modificó la red semántica subyacente de Twitter capturada por TSS por un largo periodo de tiempo. Esta misma evaluación hecha en Marzo 2015, tres meses después, determinó que la red semántica de los países volvió a niveles muy anteriores al sorteo donde se refleja nuevamente la semántica geográfica mencionada.

Como resumen de este experimento, comprendemos que la red semántica se vio fuertemente afectada por un evento saliente, incluso respondiendo con cambios de alta frecuencia (minutos), sin embargo también presentó una dinámica de decaimiento rápida que se vio estabilizada a niveles basales luego de tres meses.

TSS resultó una herramienta muy útil para estudiar la dinámica de cambios en la red semántica. A su vez combinándola con sistemas de embedding como MDS o con algún otro tipo de proyección (SVD, PCA, ICA, etc) podemos establecer un mapa semántico promedio estático y ajeno a eventos salientes. Esto permitiría usar TSS como cualquier otra metodología de *word embedding* , incluso podríamos usarla como input de otros algoritmos como el de coherencia presentado en la sección 2.1.

Los resultados expuestos en esta sección fueron publicados en [68].

Carrillo, F., Cecchi, G. A., Sigman, M., & Slezak, D. F. (2015). Fast distributed dynamics of semantic networks via social media. Computational intelligence and

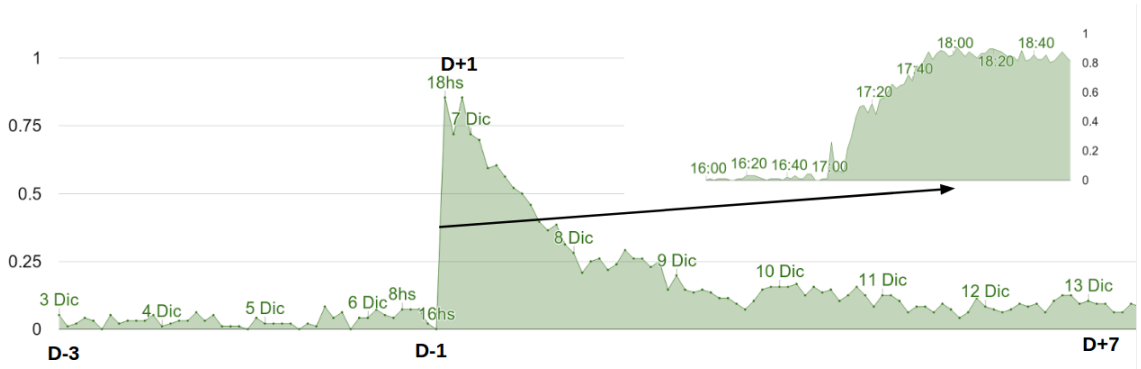

Fig. 2.5: Predicción de clasificación de grupos del mundial medida como la tasa de cantidad de países bien clasificados.

neuroscience, 2015, 50.

2.3. Análisis de sentimiento

El análisis de sentimiento es un área de aplicación de NLP que se encarga de evaluar un mensaje extrayendo valoraciones de alguna característica altamente subjetiva, generalmente referida a la positividad-negatividad pero también sobre otras dimensiones como grado de ironía, sarcasmo, nivel de felicidad del mensaje, nivel de metáforas, entre otros. Para resolver esta tarea se emplean diversas técnicas. Desde métodos muy simples, como a partir de diccionarios de palabras con su valoración etiquetada manualmente, calcular la frecuencia de apariación de las mismas en un mensaje, hasta técnicas de aprendizaje automático más complejas y modernas. En [69] los autores hacen una revisión del estado e historia de esta metodología hasta el 2008. A partir de ese momento, los sistemás de análisis de sentimiento fueron tornando más hacia el uso de técnicas modernas mayormente basadas en redes neuronales [70–72]. La ventaja de las metodologías más modernas incluyen que pueden entender contextos de sentimiento complejos como tonos irónicos o conjunciones lógicas de negación complejas [73,74], sin embargo la interpretación de sentiminiento

no es una tarea resuelta en las ciencias de la computación.

En esta tesis, usamos la metodología más simple, tomamos diccionarios de palabras anotados con su valoración y medimos la frecuencia de aparición de las palabras en los mensajes. De este modo, usamos distintas medidas definidas de la siguiente forma:

- Tasa de Positividad: tasa de palabras del mensaje que se encuentran en la lista de palabras positivas
- Tasa de Negatividad: tasa de palabras del mensaje que se encuentran en la lista de palabras negativas
- Intensidad: tasa de palabras del mensaje que se encuentran en la lista de palabras positivas o en la lista de palabras negativas
- Neutralidad: 1 - Intensidad

A lo largo de la tesis usamos cuatro diccionarios distintos:

- *Dictionary of Affective Language* (DAL) presentado en los trabajos [75–78]. Este corpus tiene más de 8 mil palabras en inglés evaluadas en 3 categorías. Para definir la lista de palabras positivas tomamos el 20 % de palabras más grande para la categoría *pleasantness*, análogamente pero con el 20 % más chico para la lista de palabras negativas.
- Spanish Dal: Es la versión de DAL pero en español, presentada en [79].
- Warriner en español: Tomamos el corpus traducido con más de 9 mil palabras. Para definir la lista de positividad y negatividad tomamos las palabras en el 20 % superior e inferior de la valoración de *valence* [80]. Disponibles online [5].

[5] `http://danigayo.info/PFCblog/index.php?entry=entry130117-183114`

- SentiWordNet3.0: Este corpus define más de 117 mil palabras con tres valuaciones: positividad, negatividad y objetividad. Para definir la lista de palabras positivas de este corpus se tomaron aquellas que tengan valuación mayor a 0 en su categoría positiva y tengan valuación igual a 0 en negativa. Para la lista de palabras negativas procedimos de manera análoga. Wordnet3.0 es presentado en el siguiente trabajo [81].

Si bien, la metodología usada captura de forma simple la valoración de sentimiento es importante destacar que no capturar fenómenos del lenguaje como negaciones. Esto hace que este tipo de análisis sea pobre para un sistema donde se quiere interpretar la intención del sujeto al producir un mensaje. Sin embargo en los casos de uso en esta tesis, usamos los modelos de análisis de sentimiento como herramientas para capturar que tipo de palabras eligen los sujetos, si son positivas o negativas y no nos focalizamos en modelar la intención por detrás del mensaje.

2.4. Estudio de distribución de uso de gramáticas del lenguaje natural

La abstracción automática de la gramática (*Part of Speech*) usada en lenguaje natural esta bien resuelta e implementada de diferentes formás. La tarea consiste en dado una oración etiquetar cada palabra con el símbolo gramatical que le corresponde. Diferentes *parsers* usan distintas categorías gramaticales pero esencialmente las etiquetas que usan son: sustantivos, pronombres, adjetivos, verbos, adverbios, preposiciones, conjunciones, interjecciones y determinantes. Los modelos computacionales que resuelven esta tarea son diversos, algunos de ellos son: cadenas de markov [82], redes neuronales [83], etc. Muchos de estos métodos están basados en ejemplos etiquetados manualmente por lo que cambiar de lenguaje natural (español, ingles, etc) implica conseguir nuevos datos etiquetados. A modo de ejemplo, usando la implementación de NLTK [84] la siguiente frase de “The house is red” es etique-

tada: [('The', 'DT'), ('house', 'NN'), ('is', 'VBZ'), ('red', 'JJ')]. DT corresponde a determinantes, NN a sustantivo singular, VBZ verbos en tercera persona singular y JJ a adjetivos.

En el contexto de esta tesis, encontramos que estudiar la frecuencia de uso de cada etiqueta gramatical en un discurso aporta información útil sobre qué tipo de gramáticas explota la mente del sujeto que las produce.

2.5. Estudio estructural del lenguaje en grafos del discurso

Si bien el estudio estadístico de los símbolos de gramáticas del lenguaje releva información de que tipo de hilo discursivo lleva adelante un sujeto al producir un discurso, esta metodología no es total en la tarea de capturar las gramáticas de la mente. Para complementarlo modelamos el lenguaje como un grafo donde estudiamos distintas características del discurso. Esta noción es introducida en [85] donde los autores parten un texto en lemas y construyen un grafo donde los nodos son lemás y conectan los lemás con aristas dirigidas cuando un lema es precedido por otro en un discurso. En [86] extendemos esta idea automatizandola y generando por cada discurso tres grafos distintos donde en cada uno complejizamos el preprocesamiento que le damos al texto para luego armar el grafo que se induce de este. Los tres grafos que armamos son:

- Naive graph: Este grafo es inducido a partir del texto sin ninguna transformación. Cada palabra corresponde a un nodo y tendemos una arista entre dos nodos cuando una palabra precede a otra en el texto.

- *Stem* graph: Este grafo es inducido a partir del texto posterior a aplicar la acción de *Stem* a cada palabra. El proceso de *stemming* consiste en eliminar los afijos morfológicos de las palabras. Por ejemplo, aplicándolo a la palabra *loving* obtenemos *love* al igual que si lo aplicamos a: *loves, loved, love*. Esta acción genera que el grafo pierda eventualmente algunos nodos pues distintas

palabras se colapsan en nodos existentes. Para realizar el proceso de *stemming* usamos la implementación de NLTK [87, 88]

- POS graph: Este grafo es inducido a partir del texto posterior a aplicar el etiquetado gramatical. En este caso el texto se convierte en una secuencia de etiquetas gramaticales (NN, WDT, JJ, etc) quedando un grafo que caracteriza las gramáticas usadas de una manera distinta al estudio de frecuencias.

La Figura 2.6 ejemplifica como a partir de un texto se generan los tres grafos y sus tres transformaciones. La inspiración original de modelar el discurso como un grafo fue entender que los sujetos en estado maníaco recurrían fuertemente a repetirse en ciertas dos o tres palabras. Tal vez un estudio de la frecuencia de n-gramás podría haber capturado fenómenos similares a los de los grafos del discurso.

Teniendo los grafos obtenemos las clásicas medidas usadas de la literatura de teoría de grafos [89].

- **Nodes:** La cantidad de nodos del grafo.
- **Edges:** La cantidad de aristas del grafo.
- **PE:** La suma entre la cantidad de ejes paralelos entre todo par de nodos.
- **LCC:** Una componente conexa de un grafo es un subconjunto de nodos tal que para todo par de nodos en él existe un camino. También debe cumplir que no puede existir otro nodo del grafo que esté conectado a algún nodo de la componente. LCC es la cantidad de nodos en la máxima componente conexa. Para computar esta medida, computamos todas las componentes conexas y tomamos la de mayor cantidad de nodos.
- **LSC:** Una componente fuertemente conexa en un grafo dirigido es un subconjunto de nodos tal que para todo par de nodos a y b en él existe un camino

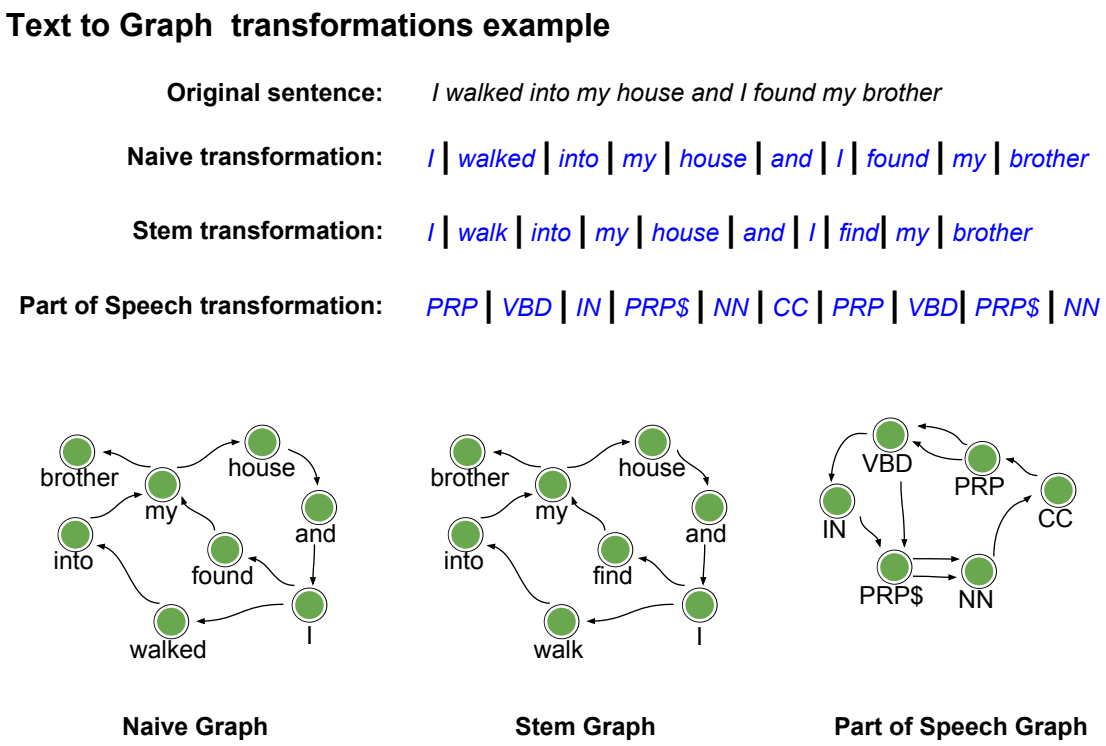

Fig. 2.6: Ejemplo de texto de paciente transformado en las tres versiones de grafos distintos.

entre a y b y entre b y a. LSC es la cantidad de nodos en la máxima componente fuertemente conexa. Para computar esta medida, computamos todas las componentes fuertemente conexas y tomamos la de mayor cantidad de nodos. Para calcular las componentes fuertemente conexas usamos el algoritmo de Tarjan [90].

- **ATD:** El grado de un nodo de un grafo es la cantidad de aristas que entran o salen de él. ATD es el promedio total de los grados para todos los nodos.
- **L1:** Cantidad de ciclos de longitud de un nodo. Para computar esta medida, tomamos la traza de la matriz de adyacencia del grafo.
- **L2:** Cantidad de ciclos de longitud de dos nodos. Para computar esta medida, tomamos la traza de la matriz de adyacencia del grafo elevada al cuadrado.
- **L3:** Cantidad de ciclos de longitud de tres nodos. Para computar esta medida, tomamos la traza de la matriz de adyacencia del grafo elevada al cubo.

2.6. Aprendizaje supervisado

El aprendizaje supervisado es un subarea del de aprendizaje automático que se encarga de modelar funciones existentes a partir de exponerse a *ejemplos* de la misma. Es decir se toman sucesivos valores del dominio de las funciones y sus respectivos valores del codominio para intentar inferir y modelar las funciones. El objetivo final que tiene esta área es poder modelar las funciones, de modo tal que generalizan a valores no vistos, es decir, que el modelo capture como se comparta la función para valores del dominio no usados en el entrenamiento (la etapa donde se le presentan los ejemplos).

Por ejemplo, si quisiéramos modelar la función que define de la temperatura en Buenos Aires podríamos tomar como datos de entrenamiento una lista con < (fecha, hora) , temperatura >, donde (fecha, hora) son los valores del dominio y

temperatura los valores del codominio que queremos predecir. Teniendo estos datos, deberíamos usar un algoritmo de entrenamiento que arme un modelo que permita luego para un nuevo valor de (fecha, hora) no visto previamente, poder devolver un valor de temperatura. Este ejemplo, corresponde a un modelo de regresión pues, la predicción ocurre sobre una variable continua.

Otro ejemplo, tomando del libro *Machine Learning* de Tom Mitchell [91] consiste en modelar si un sujeto jugará o no al tenis. Para ajustar el modelo, el algoritmo cuenta con evidencia de que días jugo y que días no, y de cada día tiene información del tiempo (humedad, viento, etc). En este caso, la función que se modela es una función que tiene como dominio el espacio de propiedades del tiempo/clima y como codominio dos valores posibles, *juega* o *no juega*. El aprendizaje automático nombra a este tipo de escenarios como modelos de clasificación pues la variable a predecir no es de tipo numérica sino de tipo categórica.

A continuación se mencionan las distintas particularidades de los modelos usados durante la tesis.

2.6.1. Modelos de clasificación y regresión

Árboles de decisión

Los arboles de decisión son métodos para modelar, típicamente, funciones de valores discretos. La función es modelada mediante una estructura de árbol, donde cada nodo representa una propiedad o dimensión del dominio y las aristas que salen de ese nodo representan distintos valores que puede tomar esta propiedad. Esto último pueden ser valores categóricos (por ejemplo para la propiedad *cielo* podría tomar como posibles valores: despejado o nublado) o también pueden ser rangos de variables continuas (por ejemplo para la propiedad *velocidad del viento*, podría separarse en *velocidad del viento*$>$ 20Km/h y *velocidad del viento* $\leq$ 20Km/h. Los nodos hoja del árbol corresponden a valores posibles de la función aprendida (para

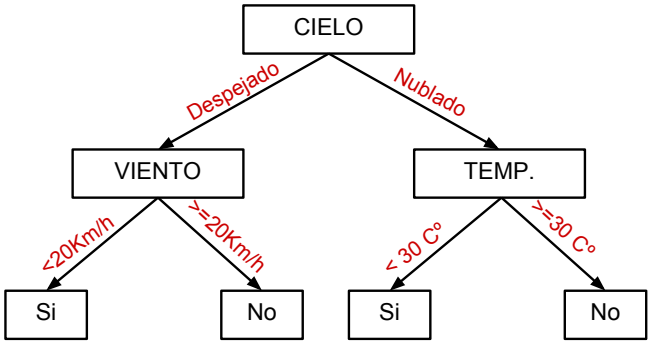

Fig. 2.7: Ejemplo de árbol de decisión ya entrenado tomado del libro de Mitchell [91].

el ejemplo del tenis, juega o no juega).

Las decisiones que expresan cómo se arma el modelo son parte del *sesgo inductivo* que define cada implementación de este tipo de modelos. Existen muchas implementaciones y políticas para definir el sesgo inductivo, cada una con distintas propiedades. El libro de Mitchell [91] presenta un panorama amplio, aunque no actualizado, de las distintas propiedades y sesgos inductivos. En definitiva, el sesgo inductivo va definir, entre otras cosas, las limitaciones con cual se abstrae y modela lo observado.

La Figura 2.7 ejemplifica un árbol de decisión ya entrenado a partir de unos supuestos datos de entrenamiento tomando el ejemplo presentado por Mitchell. En este ejemplo se ve que el algoritmo de entrenamiento a partir de los datos infirió que, por ejemplo, si esta *despejado* el cielo y si el viento es menor a 20Km/h entonces el sujeto juega, y a su vez, si la propiedad cielo es *nublado* y la temperatura es mayor a 30 Cº entonces el sujeto no juega.

Random Forest

Cuando se combinan diferentes algoritmos o modelos de aprendizaje automático se genera un nuevo modelo de *ensamble*. Los modelos de ensamble tienen la motivación de capturar los beneficios de cada submodelo que usan. Random Forest (RF)

es un modelo de ensamble que se construye a partir de varios arboles de decisión. La predicción del modelo entrenado corresponde en una ponderación de las diferentes respuestas de los diferentes arboles de decisión que lo constituyen [92]. RF tiene la motivación de componer diferentes modelos con bajo sesgo pero mucha varianza de modo que cuando se promedien por propiedades estadísticas se reduzca la varianza del modelo final conservando un bajo sesgo [93].

RF, en su etapa de entrenamiento, entrena los diferentes arboles de decisión teniendo dos particularidades. Cada árbol es entrenado con un subset de atributos y un subset de muestras de entrenamiento. Esto hace que el modelo tenga diferentes hiperparámetros a decidir como: cantidad de arboles de decisión y sus propiedades (como podría ser poda, criterio de orden de atributos, etc), cuántos atributos y cómo los elige para cada árbol, cuantas muestras se usan para entrenar cada árbol y si es o no con repetición, entre otros. Dependiendo si la tarea es de clasificación o de regresión también se debe decidir la política de ponderación de los resultados de los arboles para la constitución del resultado final.

En esta tesis usamos la implementación de Scikit-learn [94] con los parámetros por defecto, a menos que se explicite lo contrario [6].

Gaussian Naive bayes

Los clasificadores Naive Bayes son un conjunto de clasificadores probabilísticos basados en el Teorema de Bayes con la asunción ingenua de que existe independencia entre los atributos. El Teorema de Bayes establece la siguiente ecuación:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.7}$$

Donde A y B son eventos, $P(A)$ y $P(B)$ son las probabilidades de observar cada uno (con $P(B)$ mayor a 0), $P(A|B)$ es la probabilidad condicional de observar el

[6] http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

evento A dado la certeza del evento B y viceversa para $P(B|A)$. A partir de esta igualdad se puede modelar de manera simple un problema de clasificación.

Supongamos el mismo ejemplo de jugar al Tenis de Tom Mitchell [91], queremos saber cuál *clase* tiene mayor probabilidad. Dado que observe los atributos del día $A1$, ¿cuál es mayor $P(juega|A1)$ o $P(nojuega|A1)$? Para resolver esto deberíamos conocer la función de probabilidad condicional dado $A1$ que no conocemos, pero usando el teorema podemos hacer:

$$P(juega|A1) = \frac{P(A1|juega)P(juega)}{P(A1)} \tag{2.8}$$

y

$$P(nojuega|A1) = \frac{P(A1|nojuega)P(nojuega)}{P(A1)} \tag{2.9}$$

Como lo que queremos hacer es poder elegir entre las dos *hipótesis*, es decir entre *juega* o *nojuega* dado la observación $A1$, lo que queremos como output posible de nuestro modelo es h_{map} osea la hipótesis de máxima a posteriori

$$h_{map} = argmax_{h \in (juega, nojuega)} \frac{P(A1|h)P(h)}{P(A1)} \tag{2.10}$$

Como estamos buscando la h que maximice podemos sacar el denominador pues es igual para todas

$$h_{map} = argmax_{h \in \{juega, nojuega\}} P(A1|h)P(h) \tag{2.11}$$

Como $A1 = a1, a2, a3, ...$, donde en el caso del tenis $a1$ podría referirse a *despejado* para el atributo *cielo*, $a2$ "*>20Km/h*" para el atributo *viento*. Podemos tener en cuenta la asunción ingenua de suponer independencia en los atributos. Haciendo esto podemos escribir

$$h_{map} = argmax_{h \in \{juega, nojuega\}} P(a1 \cap a2 \cap ...|h)P(h) \tag{2.12}$$

como:

$$h_{map} = argmax_{h \in \{juega, nojuega\}}(\prod_{i=1} P(a_i|h))P(h) \tag{2.13}$$

Luego, podemos tomar dos asunciones más, a partir de la evidencia de entrenamiento, podemos computar $P(a_i|h)$ como la frecuencia de muestras de entrenamiento que la clase *target* corresponde es igual a h y definir $P(h)$ como la frecuencia total en la muestra de entrenamiento, la inversa o cualquier consideración válida. Habiendo hecho estas asunciones usando el teorema de bayes podemos, para atributos categóricos, elegir cuál es la clase más probable en función a unos datos de *entrenamiento*.

El problema surge con esta visión, en la manera en que computamos $P(a_i|h)$, si el tipo del atributo a_i es numérico. Para estos casos en [95] se introduce una extensión a este modelo que consiste básicamente en suponer una función de distribución subyacente a cada atributo. En el caso de Gausian Naive Bayes se supone que los atributos siguen una distribución normal, por lo que cuando se necesita estimar $P(a_i|h)$ se toman las muestras del subset de entrenamiento que tengan como valor de *target* la hipótesis h y se ajusta a una normal usando el promedio y el desvió calculados empíricamente. Con esta función de probabilidad ajustada, se calcula fácilmente $P(a_i|h)$ con la función de densidad.

KNeighbors

El clasificador KNeighbors es una metodología de clasificación basada en instancias [91]. La etapa de entrenamiento en este tipo de algoritmos consiste simplemente en almacenar los datos. Para evaluar una muestra no etiquetada el algoritmo toma los K puntos más *cercanos* a la nueva muestra y a partir de estos construye el resultado con alguna política, por ejemplo la clase con mayor frecuencia. Este clasificador tiene esencialmente dos parámetros, el primero es la cantidad de *vecinos* y el segundo la función de *cercanía*. En esta tesis usamos los valores por default del mismo de la implementación de scikit-learn [94], 5 vecinos y distancia euclidiana.

Regresión Logística

La regresión logística es un método que se suele usar para hacer clasificación [91] binaria. El entrenamiento consiste en ajustar una función logística encontrando los coeficientes β_i

$$f(t) = \frac{e^t}{1 + e^t} \tag{2.14}$$

con

$$t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \tag{2.15}$$

donde X es el vector de atributos.

Teniendo los valores de β calculado, se usa la función para determinar si una entrada pertenece a un valor categórico en función al valor numérico que la función produce y a un umbral de clasificación. Para encontrar los valores se usan diferentes algoritmos. En esta tesis usamos los valores por defecto del mismo de la implementación de scikit-learn [94].

2.6.2. Validación cruzada

La validación cruzada o *cross-validation* es una técnica utilizada para tener una valoración de la *performance* de un modelo en un contexto distinto al de los datos de entrenamiento, es decir evaluar cuán bien se comportaría un modelo frente a nuevos datos. Es una practica que consiste en tomar la muestra que se tiene y partirlas en n *folds* o particiones. Cada *fold* se espera que tenga la misma cantidad de muestras y que la distribución de cada *fold* con respecto a la variable de predicción sea similar en cada uno. Con esta partición se hace lo siguiente: por cada *fold*, se toma uno como test y los $n - 1$ restante como *folds* de entrenamiento. Luego, se entrena el modelo con los *folds* de entrenamiento y se testea con el de test. Haciendo esto n veces, donde cada *fold* toma el rol de test alguna vez, se reporta la media y alguna medida

de dispersión para los valores de performance sobre los test. Esta estrategia permite tener una noción de como un modelo de aprendizaje automático se comportará con datos no observados previamente, es decir, cuan bien generalizaría a datos no vistos. Cabe destacar que la validación cruzada es una estrategia usada para mitigar el sobreajuste y tener una noción *más real* de como generaliza el modelo, sin embargo esta estrategia no garantiza ninguna cota para el error en datos no vistos ni da certeza de la capacidad de generalización del modelo.

2.6.3. Medidas de performance

A lo largo de esta tesis hablamos de diferentes medidas de *performance*. Es necesario hablar de diferentes medidas de performance debido a que cada una de ellas pone el foco en alguna propiedad en particular o también se comporta con diferente susceptibilidad a distintas propiedades de lo que se esta evaluando. A continuación definimos las usadas durante la tesis.

Matriz de Confusión

La matriz de confusión la usamos en el contexto del análisis de un clasificador. Es una matriz cuadrada de tamaño n, donde n corresponde a la cantidad de clases. Por ejemplo, en una clasificación binaria ($n = 2$) donde se tienen dos clases: **A** y **B** la matriz de confusión se interpreta de la siguiente manera (Tabla 2.1). Cada fila de la matriz de confusión (mc) representa las instancias correspondientes a la clase real y la columna representa como el clasificador las predijo. Teniendo en cuenta esto, los elementos de la diagonal($mc[i, i]$) corresponden a cantidad de aciertos para la clase i, es decir cuantos de la clase i fueron predichos como clase i. Los valores de $mc[i, j]$, con $j \neq i$ corresponden a las instancias de la clase i predichas como de clase j osea los errores del clasificador. La matriz de confusión es una simple pero gran herramienta para entender donde están los errores de un modelo en un experimento. Claramente no es lo mismo matrices donde los errores están bien balanceados comparado con

Clasificados como			
A	**B**		
Verdaderos A	Falsos B	**A**	Clase Real
Falsos A	Verdaderos B	**B**	

Tab. 2.1: Ejemplo de matriz de confusión.

aquellas matrices donde los errores se encuentran en un extremo de la matriz, pues este caso es el de un modelo que insiste sobre un tipo de error que sobre otro. En la bibliografía donde se define una clase como *positiva* y otra como *negativa* los valores de $m[i, j]$ con $j \neq i$ son descriptos como falsos positivos o falsos negativos según la asignación de clase *positiva*.

Accuracy

En el contexto de clasificación binaria, es decir de dos clases la *accuracy* es la tasa de muestras bien clasificadas. Es decir la traza de la matriz de confusión dividido la cantidad de muestras.

$$accuracy = \frac{\sum_{i=0}^{n-1} mc[i, i]}{\sum_{i=0}^{n-1} (\sum_{j=0}^{n-1} mc[i, j])} \tag{2.16}$$

Los valores cercanos a 1 son las clasificaciones perfectas cuando aquellos cercanos a 0 son las malas clasificaciones.

F-score

Esta medida en realidad es una medida paramétrica llamada F_β score que permite reducir la información de la matriz de confusión en un número ponderando la información de esta de diferentes maneras. Esto responde a la necesidad de poder comparar matrices de confusión con alguna métrica de orden. Es decir, define un orden total entre matrices de confusión, útil para los casos donde las diferentes funciones objetivos de los algoritmos quieren tomar alguna decisión sobre que dirección

optimizar.

Para el caso de clasificaciones binarias, eligiendo una clase como *positiva* y la otra como *negativa* se define F_β como:

$$\frac{(1+\beta^2) \times true\ positives}{(1+\beta^2) \times true\ positives + \beta^2 \times true\ positives + false\ positives} \tag{2.17}$$

Los valores cercanos a 1 son las clasificaciones perfectas cuando aquellos cercanos a 0 son las malas clasificaciones.

Curva ROC y el área bajo la curva

La curva ROC se construye sobre un plano donde las coordenadas en el eje X corresponden a la tasa de Falsos positivos y las del eje Y corresponden a la tasa de Verdaderos Positivos, todos los puntos se construyen a partir de *mover* algún valor que funciona como umbral en el clasificador para binarizar la respuesta a una clase. Es decir, si un clasificador en vez de responder clase A o B responde la probabilidad de pertenecer a la clase A, para construir los distintos puntos de la curva ROC, se mueve el valor por el cual se define si la probabilidad es suficiente o no para que la predicción binaria sea finalmente clase A o B. El área bajo la curva ROC es una medida que cuantifica el área que hay debajo de esta curva. En el contexto de esta tesis, esta medida es usada pues tiene una propiedad interesante. En el caso de las clasificaciones binarias donde la cantidad de muestras por clase esta desbalanceada, medidas como *accuracy* desorientan con facilidad la evaluación de cuan bien funciona un modelo y cómo se compara este con tirar una moneda o el modelo *chance*. Sin embargo, el área de la curva ROC tiene como propiedad, que el valor para el modelo que tira una moneda es 0.5 independientemente al desbalance de elementos de la muestra. Esta medida permite tener una noción más intuitiva de cuan bien funciona un modelo. Un estudio y explicación de los alcances se reportó en [96]. En el contexto de esta tesis usamos la implementación de Scikit-learn [94] que permite aproximar

la curva roc a partir de la matriz de confusión.

2.7. Validación por múltiples comparaciones

En los experimentos donde testeamos diversas hipótesis a la vez sobre un mismo fenómeno, por ejemplo entender si existe diferencia estadística entre distintas propiedades del lenguaje para dos grupos de sujetos, debemos ser más rigurosos para interpretar la validez estadística de un resultado. Pues, si por ejemplo mido n comparaciones a la vez y no corrijo mi nivel de aceptación de un test de hipótesis voy a estar encontrando resultados espurios. Para resolver el problema de múltiples comparaciones, en esta tesis usamos la corrección por Bonferoni.

El método de corrección por Bonferoni consiste en modificar el valor de aceptación estadístico, es decir el valor por el cual consideramos aceptable la probabilidad de que un resultado no sea producto del azar (usualmente0,05 o 0,01) por la cantidad de múltiples comparaciones. Es decir, si hacemos 40 comparaciones múltiples y tenemos como valor de aceptación 0,05, luego de la corrección, el nuevo valor de aceptación para el test de hipótesis sobre las 40 hipótesis sera: $\frac{0{,}05}{40} = 0{,}00125$. Una explicación de este método y un detalle de cuándo es corrector usarlo de encuentra en [97]. Otra manera análoga de ver esto, consiste en dejar el umbral de significancia estable, es decir, por ejemplo 0,05 y multiplicar los $p - values$ por la cantidad de múltiples comparaciones.

3. ESTADOS MENTALES ALTERADOS POR PATOLÓGICAS

El concepto de salud mutó a lo largo del tiempo, hoy en día no existe un consenso en la definición sino múltiples declaraciones. La Organización Mundial de la Salud [1] la define como: "La salud es un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades.". Es evidente que este principio constituyente tan abstracto no alcanza para definir estados patológicos, transtornos y demas entidades nosologicas, por lo que cada área de la medicina, tomando alguna definición como guiá, ensaya diferentes descripciones propias respecto a diversos conceptos de la salud, como por ejemplo las patologias y sus síntomas. A su vez, a medida que nueva evidencia es recolectada, se cambia la taxonomía descriptiva de las enfermedades, transtornos, síndromes, síntomas, etc.

El siguiente caso ejemplifica como un transtorno, claro en la descripción de su origen, genera dificultades a la hora de encuadrar en cualquier definición de salud. La talasemia es un tipo de anemia caracterizada por una producción anormal de la hemoglobina muy presente en la población Argentina [98, 99]. Este *transtorno* se produce a partir de una mutación en alguno de los genes que codifica para una subunidad de la hemoglobina. Existen diferentes tipos de talasemias en función a cual gen y en que parte se ubica la mutación. Los síntomas de la talasemia [2] van desde ninguno, pasando por anemias severas, problemas de huesos, hasta malformaciones fetales que imposibilitan la vida. Sin embargo, los portadores de ciertas talasemia son casi inmunes a la malaria, debido al mecanismo de parasitación del parásito de la malaria. Esta característica sitúa en una posición ventajosa evolutivamente

[1] http://www.who.int/about/mission/es/

[2] https://www.nhlbi.nih.gov/health/health-topics/topics/thalassemia/signs

a los portadores de esta mutaciones en zonas donde los focos de malaria fueron considerables. El conflicto de definir el concepto de salud y las patologías por fuera de su contexto evolutivo se puede apreciar en muchos ejemplos diferentes donde la complejidad aumenta a medida que se desconocen factores referidos a la patología en si y a su contexto.

Si en casos como la talasemia, donde se conoce exactamente las mutaciones asociadas con el transtorno y a su vez se conoce la interacción especifica con el parásito de la malaria resulta dificil encajar el cuadro con cualquier definición de salud, las enfermedades psiquiátricas y neuronales constituyen un reto aun mayor.

La Organización Mundial de la Salud [3] , define la saludo mental como "Mental health is defined as a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community.". Nuevamente este concepto es definido de una forma imprecisa como el concepto de salud. Tal vez la dificultad en encontrar definiciones precisas responda a la complejidad del concepto y sea ineludible partir de una definición amplia y vaga. Entendiendo esta característica, en esta tesis tomamos como puntos axiomáticos las caracterizaciones de las patologías definidas en el *Manual diagnóstico y estadístico de los trastornos mentales* (DSM-5) [100] y los diferentes procedimientos y herramientas de diagnostico establecidos en la clínica psiquiátrica. En diferentes casos usaremos los diagnósticos como valoraciones a predecir en el contexto de aprendizaje supervisado. A su vez, tuvimos en cuenta en todos los casos posibles, la bibliografía psiquiátrica y las definiciones del colectivo sintomatológico de cada patología para orientarnos en la extracción de propiedades del lenguaje relevantes pues entendemos que capitalizar esta información es sumamente importante en nuestro aporte.

En este capitulo presentamos 4 casos de estudio donde la mente se los sujetos

[3] http://www.who.int/features/factfiles/mental_health/en/

se ve alterada por patologías psiquiátricas y a partir de modelos de procesamiento del lenguaje natural conseguimos cuantificar diferentes fenómenos de la mente y el lenguaje. El primer caso hacemos un análisis algorítmico de coherencia para el diagnostico automático de pacientes esquizofrénicos (Sección 3.1). En el segundo caso, analizamos propiedades del lenguaje para pacientes prodrómicos (Sección 3.2). En el tercer caso, estudiamos pacientes con trastorno bipolar (Sección 3.3) y por último estudiamos la predicción de respuesta a tratamiento piscofarmacologico en pacientes con depresión resistente al tratamiento (Sección 3.4).

3.1. Esquizofrenia

El *Manual diagnóstico y estadístico de los trastornos mentales* (DSM) es un manual psiquiátrico que clasifica los trastornos mentales y propone descripciones de cada cuadrado con el fin de definir un estándar para el diagnostico, estudio y tratamiento de trastornos mentales.

Cuando el DSM-5 [100] define las características claves de la esquizofrenia y otros desordenes psicóticos establece la *desorganización del pensamiento* como una de las más relevantes:

> *Disorganized thinking (formal thought disorder) is typically inferred from the individual's speech. The individual may switch from one topic to another (derailment or loose associations). Answers to questions may be obliquely related or completely unrelated (tangentiality). Rarely, speech may be so severely disorganized that it is nearly incomprehensible and resembles receptive aphasia in its linguistic disorganization (incoherence or "word salad").*

Teniendo en cuenta esta caracterización que define al discurso de pacientes esquizofrénicos, diseñamos un algoritmo que modela la desorganización del discurso a

través de la coherencia del mismo (ver Sección 2.1). Para evaluar si nuestro algoritmo capturaba bien los accidentes de desorganización del discurso lo aplicamos al estudio de un grupo de 20 sujetos esquizofrénicos.

Los 20 sujetos fueron diagnosticados usando los estándares del DSM mediante el rating SCID [101], los mismos son pacientes del Hospital Onofre Lopes dependiendo de la Universidad Federal de Rio Grande del Norte (UFRN), Brasil y del Hospital Machado, Natal, Brasil. A los participantes se les entregó la siguiente tarea para que respondan en Portugués: *Por favor reporte el recuerdo de un sueño reciente.* Luego, el discurso fue grabado y transcripto por un experimentador independiente al experimento que no conoce el mismo ni la condición del paciente a modo de evitar sesgos en la transcripción. Los sujetos firmaron un consentimiento para el estudio que fue aprobado por el comité de ética de UFRN. A su vez, contamos con 20 sujetos saludables como control que respondieron a la misma consigna.

El algoritmo de coherencia que diseñamos, como establecimos en la descripción usa un modelo de *word embedding* ya entrenado. Entrenar un modelo así es una tarea compleja si se quiere contar con un modelo útil. Como no contábamos con un modelo en portugués validado decidimos usar un modelo extremadamente usado en inglés y traducir los textos de los pacientes automáticamente al inglés usando Google Translate [4]. Esta transformación automática podría romper las características que nuestro algoritmo de coherencia captura sin embargo decidimos empezar con esta estrategia y estudiar los resultados. En caso que no encontráramos resultados esto podría ser responsabilidad de este paso y hubiera correspondido entrenar o conseguir un modelo de *word embedding* en portugués.

A partir del LSA entrenado y los textos en inglés calculamos las dos distribuciones de coherencia descriptas anteriormente COH_1 y COH_2 para cada texto. Luego para representar a un sujeto tomamos las siguientes medidas estadísticas para cada serie: media, mediana, desvió estándar, mínimo y máximo. A su vez, también armamos la

[4] https://translate.google.com/

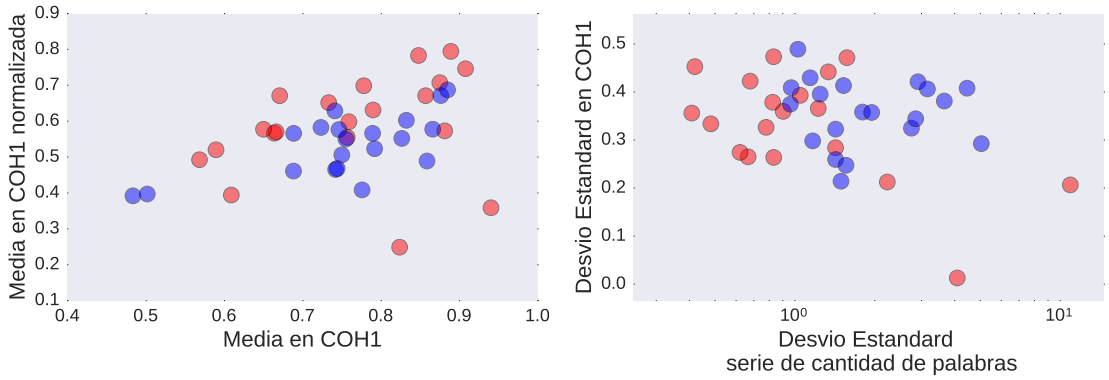

Fig. 3.1: Ejemplo de posicionamiento espacial de dos sujetos esquizofrénicos y control con 4 features.

serie de cantidad de palabras por frase y tomamos las mismas 5 medidas estadísticas. Notamos que había cierta variabilidad en la cantidad de palabras de las frases y esto de alguna manera iba a interceder en las series de coherencia (pues cuando tomamos el vector promedio para resumir una frase, no es lo mismo tomar un vector promedio a partir de una lista de 2 vectores que de una de 15 vectores). Por este motivo, decidimos agregar una normalización a COH_1 y COH_2 , esta consistió en dividir cada elemento de las series por la cantidad de palabras en los dos vectores. De esta manera cada sujeto es representado por 25 *features*.

Como primer paso calculamos la diferencia estadística entre los dos grupos (esquizofrénicos y control) de las *features* mediante Student Test. La Tabla 3.1 muestra los p-valores. Los resultados muestran que no hay diferencia estadísticas para los grupos luego de las correcciones por múltiples comparaciones usando corrección por Bonferoni (ver Sección 2.7).

Si bien esta primera medida no resultó significativa, los datos parecían mostrar cierta estructura observándolos por grupos. Por ejemplo, si mirábamos como se distribuyen en dos dimensiones los sujetos para 4 *features* podíamos notar que hay cierto orden entre ellos (ver la Figura 3.1).

Feature	p valor
COH_1 Mediana Normalizado	0.026
COH_2 Promedio Normalizado	0.061
COH_2 min	0.103
COH_2 min Normalizado	0.119
COH_1 Promedio Normalizado	0.128
Palabras por frase max	0.172
COH_2 std	0.192
COH_2 Mediana Normalizado	0.194
COH_1 min	0.273
COH_1 max	0.310
COH_2 max	0.323
Palabras por frase min	0.324
Palabras por frase std	0.351
COH_1 min Normalizado	0.355
COH_2 max Normalizado	0.439
COH_1 max Normalizado	0.452
COH_1 std	0.520
COH_2 Mediana	0.583
Palabras por frase Mediana	0.589
COH_1 Mediana	0.595
COH_2 Promedio	0.727
Palabras por frase Promedio	0.782
COH_1 Promedio	0.786
COH_2 std Normalizado	0.916
COH_1 std Normalizado	0.941

Tab. 3.1: Medidas de coherencia con sus p-valores y correcciones por múltiple comparación para el test de hipótesis Student Test entre población esquizofrénica y Control

Teniendo en cuenta que vemos cierta estructura no reflejada por los test de hipótesis consideramos abordar el problema a partir de entrenar algoritmos de aprendizaje supervisado (ver Sección 2.6).

Los algoritmos de aprendizaje supervisado fueron entrenados con las 25 features de coherencia descriptos en un esquema de *cross-validation* de 10 *folds* para obtener una mejor noción del resultado (ver Sección 2.6.2). Probamos inicialmente un subconjunto de modelos y encontramos que la máxima *accuracy* (ver Sección 2.6.3) corresponde al modelo RandomForestClassifier (ver Sección 2.6.1) con 20 estimadores, alcanzando un valor de $0{,}85 \pm 0{,}052$ (media y error estandar).

Para testear la validez de nuestros resultados hicimos el mismo experimento probando los mismos modelos pero esta vez, permutamos los datos del vector que describe la clase a predecir. De esta manera podemos medir cuál es el resultado basal de los modelos si los datos conservan la distribución pero se cambian al azar la asignación de sujetos con su clase. Habiendo repetido la permutación 100 veces, este experimento nos mostró que los resultados de *accuracy* para cada modelo corresponden al valor del azar dada la distribución de clases (el %50 de las muestras corresponden a cada clase).

La Tabla 3.2 resume los resultados para todas las configuraciones probadas y los valores medios del experimento de permutación de clase.

Si bien la *accuracy* fue alta, esta información no bastó para entender los errores que cometió el modelo. Para eso miramos la matriz de confusión de los resultados. La Tabla 3.3 detalla esta información donde se ve que los errores fueron cometidos uniformemente entre clases. Es decir, tres sujetos que eran esquizofrénicos fueron predichos como control y también tres sujetos control fueron predichos como esquizofrénicos. Es interesante entender que dependiendo el caso de uso los errores podrían ser ponderados distintos. En este caso cuando entrenamos el modelo penalizamos de la misma manera estos dos errores, pero en un contexto donde quisiéramos usar esta herramienta como metodología de detección automática podríamos tomar

Clasificador	*accuracy*	*accuracy* en permutaciones
RandomForestClassifier	0.85 ± 0.052	0.487 ± 0.094
DecisionTreeClassifier	0.750 ± 0.05	0.499 ± 0.092
KNeighborsClassifier	0.725 ± 0.0740	0.515 ± 0.088
LogisticRegression	0.675 ± 0.061	0.517 ± 0.097
GaussianNB	0.5 ± 0.050	0.514 ± 0.077

Tab. 3.2: Resultados de distintos modelos de clasificación para la tarea de clasificar sujetos Esquizofrénicos vs sujetos Control en un esquema de validación cruzada de 10 *folds*. La columna *accuracy* corresponde a la media y el error estandar de cada modelo con los datos reales. La columna *accuracy* en permutaciones corresponde a la media y el error estandar del experimento de permutación de clases.

la decisión arbitraria de ponderar más un error sobre otro, en este caso no tomamos esta decisión pero es fácil adaptar el modelo para que intente minimizar no la *accuracy* sino en su lugar la sensibilidad o la especificidad del modelo o también ponderar con algún coeficiente cada clase usando F_β score (ver Sección 2.6.3).

Como mencionamos, estudiar la coherencia en el contexto de pacientes esquizofrénicos fue motivado por la descripción del DSM-5. Pero a su vez, en [85] los investigadores aplicaron los métodos de grafos del discurso (ver Sección 2.5) en otra cohorte de pacientes con esquizofrenia. Sumando los *features* de grafos a nuestra clasificación conseguimos aumentar levemente los valores de performance (no significativamente distintos si se toma en cuenta el desvió estándar de las soluciones presentadas en 3.2). Entender por qué tener caracterizaciones supuestamente independientes como son los grafos del discurso y la medida de coherencia no contribuyen a mejorar la *performance* del modelo debe ser un siguiente paso. Resultados preliminares de este análisis se pueden ver en [102].

Los datos y resultados de en esta sección fueron publicados en [86].

Carrillo, F., Mota, N., Copelli, M., Ribeiro, S., Sigman, M., Cecchi, G., & Slezak,

Clasificados como			
Esquizofrénicos	**Control**		
17	3	**Esquizofrénicos**	Clase Real
3	17	**Control**	

Tab. 3.3: Matriz de confusión de clasificación esquizofrénicos vs control con Random Forest.

D. F. (2014, December). Automated Speech Analysis for Psychosis Evaluation. In International Workshop on Machine Learning and Interpretation in Neuroimaging (pp. 31-39). Springer International Publishing.

3.2. Prodrómicos

La falta de marcadores biológicos para patologías como la esquizofrenia implica que el diagnostico se subjetivice, más que en otras áreas de la medicina, a la interpretación del profesional médico y a la propia valoración del paciente en algunos casos. Por ejemplo, usando la herramienta estándares del DSM para el diagnostico, el rating SCID [101] los pacientes tienen que contestar la siguiente pregunta:

> *During (TIME PERIOD FOR EPISODE) were you feeling so good or hyper that other people thought you were not your normal self or you were so hyper that you got into trouble? (Did anyone say you were manic? Was that more than just feeling good?)*

Si bien no alcanza una sola respuesta para conformar el diagnostico, los psiquiatras se basan en este tipo de evaluaciones para decidir los tratamientos a seguir. Dada las características extremadamente subjetivas de las herramientas y pese al gran esfuerzo de los médicos, el inicio de una patología no siempre queda claro. Por este motivo nos preguntamos si nuestras herramientas podrían prever o adelantar el reconocimiento del inicio de alguna de las patologías (estado prodrómico).

Para responder esta pregunta diseñamos un experimento donde 34 participantes sanos de alto riesgo de psicosis se sometieron a una entrevista. Los sujetos tenían al momento de la primera entrevista entre 14 a 27 años con un nivel de inglés fluido (solo 3 no tenían a este idioma como lenguaje natal). Los participantes fueron considerados de alto riesgo usando la herramienta *Structured Interview for Prodromal Syndromes/Scale of Prodromal Symptoms* (SIPS/SOPS) [103]. SIPS/SOPS es una herramienta para evaluar síntomas positivos y negativos (taxonomía de sintomatología usada en esquizofrenia [104]), desorganización del pensamiento y síntomas en general. Para incluir a los sujetos dentro del grupo de riesgo estos tenían que cumplir con algún criterio de los siguientes:

1. Síndrome de síntomas positivos atenuados

2. Riesgo genético (parientes de grado uno con episodios psicóticos o diagnosticados con esquizofrenia)

3. Breve síntomas de psicosis

Cada uno de estos criterios esta definido en SIPS/SOPS y se cumple en función a un puntaje alcanzado por las respuestas a preguntas particulares. En definitiva los criterios caracterizan a sujetos que tienen riesgo alto de devenir en episodios psicóticos pero aun no han experimentado uno y no son diagnosticados con ninguna patología psiquiátrica con desorden psicótico.

Teniendo bien definidos los sujetos experimentales se entrevisto a cada uno. La entrevista consistió en la siguiente consigna: *"como se sienten, que cambios han experimentado recientemente y cuales fueron los impactos de estos cambios, y también que reporten su expectativa personal futura"*. Las entrevistas, que duraron al rededor de una hora, fueron conducidas por un experto en entrevistas cualitativas en *New York State Psychiatric Institute en Columbia* . Un investigador independiente al experimento transcribió del audio las entrevistas. El protocolo diseñado y usado

en este experimento fue aprobado por la Junta de Revisión Institucional donde el experimento fue conducido. A su vez, un consentimiento escrito fue firmado por cada sujeto y por los padres en caso de que estos fueran menores de edad. Todos los experimentos

Posterior a la toma de las entrevistas, los sujetos son revaluados cada 3 meses sobre su condición psíquica para un eventual diagnostico. Luego de 2 años y medio, 5 de los 34 sujetos desarrollaron esquizofrenia diagnosticada con las mismas herramientas de diagnostico que, en un principio y en cada evaluación intermedia, no los habían diagnosticados como tales.

Dada esta configuración experimental, donde contamos con 34 sujetos sanos pertenecientes a grupos de alto riesgo con entrevistas en momentos donde no eran patológicos y luego de 2.5 años, 5 de ellos se diagnosticaron esquizofrénicos, refinamos nuestra hipótesis inicial. Ahora, nos preguntamos si había información en el discurso producido en las entrevistas que identifique a esos 5 sujetos y los distinga. En el caso de encontrar información, podríamos explicar que los métodos clásicos psiquiátricos no cuentan con tal grado de susceptibilidad suficiente para identificar la patología en un momento tan insipiente.

Teniendo en cuenta que podía existir la posibilidad de que los síntomas estuvieran presentes en aquellos 5 sujetos pero simplemente las herramientas clásicas de la psiquiatría aun no pudieran detectar, nos propusimos buscar aquellos síntomas que sabíamos que habíamos medido con éxito en sujetos esquizofrénicos (ver Sección 3.1). Por ello decidimos volver a aplicar el algoritmo de coherencia (definido en 2.1) y estudiar si había alguna diferencia entre los 29 sujetos no convertidos (CHR-) y los que se convirtieron en sujetos esquizofrénicos (CHR+).

Previo a aplicar el algoritmo de coherencia a los datos reales estudiamos más las propiedades de este algoritmo. Nuestra intención era evaluar cuán sensible es el algoritmo a cambios en el discurso, de esta manera generar intuición sobre que

tipo de efecto podíamos reconocer y cuáles no. Para eso tomamos distintas obras clásicas de la literatura en ingles (10 obras de Clásicas de Proyecto Gutenberg al azar [5]) y evaluamos como cambiaban las medidas de coherencia en función a cuanto alterábamos el texto original. Definimos una tasa de cambio como la tasa de frases que cambiamos de posición original en el texto. Por ejemplo un texto con tasa de cambio de 0.1 significa que 10 % de las frases fueron cambiadas de lugar en el texto a posiciones nuevas. Usamos la misma configuración para el algoritmo de coherencia implementada en el caso de sujetos esquizofrénicos (LSA con Tasa entrenada con 300 dimensiones como *word embedding*, ver Sección 3.1).

Los 10 textos tomados fueron:

- Alices Adventures in Wonderland por Lewis Carroll
- A Study in Scarlet por Arthur Conan Doyle
- A Tale of Two Cities por Charles Dickens
- Metamorphosis por Franz Kafka
- Moby Dick; Or, The Whale por Herman Melville
- Peter Pan por J. M. Barrie
- Pride and Prejudice por Jane Austen
- The Adventures of Sherlock Holmes por Arthur Conan Doyle
- The Adventures of Tom Sawyer por Mark Twain
- The Count of Monte Cristo por Alexandre Dumas

Para cada texto tomamos las primeras 4 mil frases y las convertimos en vectores de 300 dimensiones. Luego por cada libro, corrimos distintas configuraciones de tasa de cambios, tomamos: [0 0,16 0,32 0,48 0,64 0,8], para cada tasa de

[5] https://www.gutenberg.org/

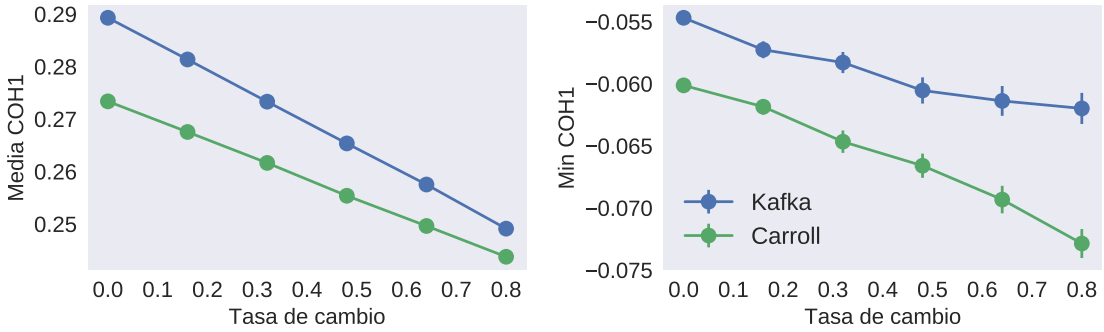

Fig. 3.2: Ejemplo de dinámica de decaimiento para distintos niveles de tasa de cambios para experimento de coherencia en obras clásicas

cambio corrimos 1000 muestras (por ejemplo, si tomamos la tasa de cambio 0,16 que correspondería a 16 % de cambio, armamos 1000 muestras distintas que contuvieron cada una 16 % de cambios). La Figura 3.2 muestra a modo de ejemplo la dinámica de la coherencia media y mínima de dos libros para los distintos valores de tasa de cambio mencionados. Para cuantificar este efecto medimos la correlación de Pearson tomando todos los valores de la serie de COH_1 media para todos los libros, es decir miramos si existe una correlación entre la tasa de cambio y el valor medio de COH_1 . Los datos muestran una fuerte correlación negativa ($rho = -0{,}9918, p-valor = 2{,}7305e-10$) como era de esperar, esto quiere decir que a medida que mas mezclamos el texto en comparación al orden original, la coherencia media del mismo baja. El mismo efecto se preserva al observar la correlación pero esta vez usando el mínimo de COH_1 en vez de la media como medida de caracterización de la serie ($rho = -0{,}9674, p-valor = 2{,}7111e-7$). Este resultado también fue el esperado pues suguiere que al cambiar el orden al azar bajamos el mínimo de la serie de coherencia. Pues dos frases que tienen poco que ver semánticamente se sitúan contiguas en el texto.

Tras haber validado la herramienta de coherencia en modificaciones de textos clásicos, corrimos el algoritmo de coherencia para los datos de las entrevistas reales.

Las entrevistas cuentan con una distribución distinta a los ejemplos de los libros pues son mas cortas, tienen en promedio 5161 palabras con un desvió estándar de 2846, a su vez cuentan con, en promedio, 265±118 frases. También computamos la distribución de símbolos gramaticales (ver Sección 2.4) y medidas de cantidad de palabras por frase. Tomamos la decisión de agregar estas dos mediciones pues la primera captura información sintáctica ortogonal a la que registra el algoritmo de coherencia. La información sintáctica provee noción del uso de gramáticas que los sujetos usan inconscientemente, por ejemplo podría ser un discurso con un uso exhaustivo de adjetivos caracterizando un discurso muy descriptivo o tal vez uno mas procedural, viéndose esto último en el exceso del uso de etiquetas gramaticales referidas a por ejemplo, verbos en infinitivo. Con respecto a la segunda, la cantidad de palabras por frase, esta serie aporta información sobre propiedades no intensivas del lenguaje (como si lo es la frecuencia de etiquetas gramaticales o la coherencia media) que caracterizan propiedades tan simple como la verborragia del sujeto.

Habiendo calculado todas estas medidas por sujeto, dado el gran desbalance entre las dos clases (CHR+ 5 sujetos, CHR- 29 sujetos) y a su vez que la cantidad de sujetos no era demasiado grande, hacer comparaciones estadísticas entre las distintas medidas perdía sentido. Por lo que decidimos abocarnos en la tarea de clasificación entre clases.

El primer experimento que hicimos fue intentar calcular usando el clasificador *Random Forest* con 10 arboles (ver Sección 2.6.1) cual era la *performance* de la clasificación con todas las features. A su vez modificamos la función de costo para que los errores no pesen lo mismo sino que estén en función a la inversa de la frecuencia de las clases. Esto genera que cometer un error en la clasificación de un CHR+ aporte mucho mas que un error en CHR- (exactamente $\frac{29}{4} = 7{,}25$ veces mas). Decidimos usar un esquema de *cross-validation* de 5 folds, de esta manera nos aseguramos que en todos los folds siempre tenemos una muestra de la clase CHR+. Usando todos las medidas y corriendo el algoritmo 100 veces (pues *RandomForest*

cuenta con una cuota de azar al muestrear los atributos y las muestras) obtuvimos una *accuracy* (ver Sección 2.6.3) de 0.8497±0.013 (media y error estándar). Si bien esta valor parece grande y significativo es difícil de evaluarlo pues el valor de tirar una moneda en vez de usar un clasificador, en este caso no es 50 %, debido a que las clases no están balanceadas. Para responder esto, hicimos el experimento de permutar las clases de los sujetos, hacer *cross-validation* y reportar la media. Esto nos aporta una noción empírica de cual es el valor de *chance*, es decir, cual es el valor que tomaría un clasificador si respetamos las distribuciones de los atributos de entrada pero cambiamos al azar intencionalmente la relación que este clasificador estaría modelando. Este experimento reporta, corrido 100 veces, una *accuracy* de 0.8457±0.021 (media y error estándar), es decir el resultado que parecía bueno en realidad no esta logrando predecir de manera correcta. Una mejor manera de evaluar un algoritmo cuando esta desbalanceada las clases, en la tarea de clasificación, es usar *rocauc* dado que su valor teórico de *chance* es siempre 0,5 (ver Sección 2.6.3).

Tras haber obtenido un valor de predicción igual al de no tener información decidimos reducir la cantidad de atributos que usamos para la clasificación para entender mejor el funcionamiento de los experimentos pues inicialmente habíamos testeado con todas los atributos de coherencia, distribución de símbolos de gramáticas y medidas estadísticas sobre palabras por frase. Todos estos atributos totalizaban 122, teniendo 34 sujetos decidimos ir a un esquema donde la cantidad de atributos sea menor a la cantidad de sujetos. Por esto, probamos elegir un atributo de cada categoría (coherencia, gramática y extensión) y tratar de clasificar con 3 atributos. Tras experimentar un poco de esta manera llegamos a una configuración de atributos que produjeron una buen resultado. Estos atributos fueron:

- Mínimo de la serie COH_1 normalizado (como se normalizado en el caso del experimento en sujetos esquizofrénicos, ver Sección 3.1).

- La frecuencia del símbolo gramatical referido a los determinantes (WDT),

que incluyen: *that, what, whatever, which, and whichever*, este último símbolo gramatical fue elegido por que entendemos que captura el concepto de escasez de discurso que constituye un síntoma negativo en la esquizofrenia, concluimos esto luego de ver que es fuertemente usado como conector para armar frases mas grandes y complejas.

- Tamaño de la frase mas larga

Para obtener estos tres atributos, armamos mil conjuntos distintos de tres atributos de las tres categorías y reportamos el mejor.

Estos tres atributos, representando las categorías resultaron buenos para la clasificación automática dando un resultado de *rocauc* 0.8286±0.0057 (media y error estándar de correr 100 veces el experimento). Si bien el valor teórico de *chance* usando *rocauc* es de un medio, lo comprobamos empíricamente. Para eso permutamos las clases 100 veces y obtuvimos un valor medio de 0.5046 con un error estándar de 0.0169. A su vez, repetimos el experimento usando la misma política de búsqueda de los tres atributos habiendo fijado una permutación. Es decir, dada una permutación de las clases, buscamos los 3 atributos que dieran la mejor clasificación. Repitiendo este experimento control 100 veces tampoco encontramos resultados que difieren significativamente del valor de chance. Este último experimento control nos da información para creer que no estamos haciendo un sobre ajuste indebido al probar tantas configuraciones distintas de atributos para el experimento real.

Si bien presentamos tres atributos, uno por categoría, si los cambiamos respetando las categorías, en muchos casos obtenemos valores de *rocauc* similares (por ejemplo cambiando el mínimo de COH_1 por la media, obtenemos una *rocauc* media de 0.8601 con un error estándar de 0.004). Esto probablemente se deba a la alta correlación que hay entre distintos grupos de atributos dentro de cada una de las tres categorías. La Figura 3.3 muestra la correlación entre los atributos separados por los tres grupos (coherencia, gramáticas y serie de cantidad de palabras). Teniendo

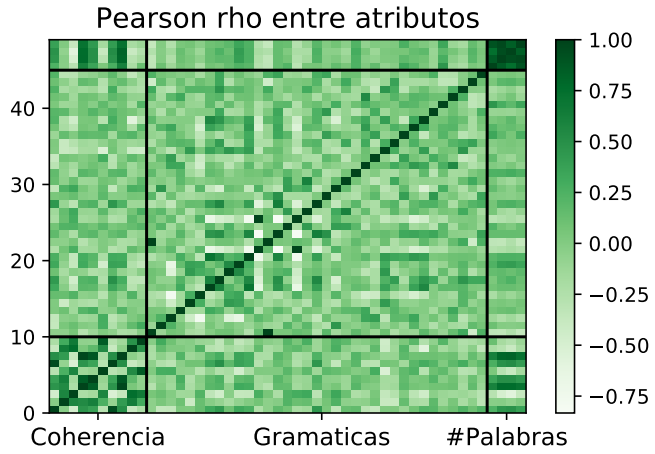

Fig. 3.3: Matriz de *rho* de correlación lineal entre los atributos para sujetos prodrómicos.

en cuenta la corrección por múltiple comparaciones, los p-valores, para el cluster de Coherencia el 15 % de los atributos correlacionan significativamente entre si, mientras que ese valor baja a 1 % para el cluster de gramáticas, para el caso de la serie de cantidad de palabras todos los atributos correlacionan significativamente entre si. Las comparaciones entre los atributos de distintos clusters, para las tres comparaciones (coherencia vs gramáticas, coherencia vs serie de cantidad de palabras y gramáticas vs serie de cantidad de palabras) da 0 % de comparaciones significativas. Esta característica justifica que tenga sentido empírico la partición que propusimos inspirada en tres tipos de atributos distintos. Tal vez hubiera correspondido usar alguna estrategia de reducción de dimensionalidad o cambio de base que promueva una nueva base para el espacio con alguna propiedad particular de independencia, ortogonalidad o anticorrelación entre los vectores de la base (como por ejemplo con SVD o ICA [105])

Los resultados presentados muestran fuertemente que había información que indicaba que el lenguaje de los 5 sujetos CHR+ era distinto al de los sujetos CHR-. Si bien es difícil interpretar la diferencia creemos fuertemente que está presente, pues de otro modo no podríamos haber encontrado una clasificación exitosa teniendo en cuenta los métodos control probados para mitigar el efecto del sobre ajuste. Tenien-

do en cuenta que estamos midiendo características del lenguaje relacionadas con el cambio en la coherencia del discurso y que ya vimos que para sujetos con patologías con trastornos psicóticos esta propiedad ya se veía afectada, consideramos que estos resultados pueden ser interesantes para avanzar a re-definir una metodología de diagnostico apoyada en este tipo de herramientas. A su vez creemos que el siguiente paso es ampliar la muestra para poder empezar a reconstruir una noción mas real sobre el cambio de esta propiedad tan particular del lenguaje. Los resultados de esta prueba de concepto fueron publicados en [63].

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., ... & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophrenia, 1, 15030.

Como continuación de este trabajo, realizamos el mismo experimento en una cohorte nueva de pacientes pudiendo replicar exitosamente los resultados. Dicho trabajo fue publicado en [106].

Corcoran, Cheryl M., Facundo Carrillo, Diego Fernández-Slezak, Gillinder Bedi, Casimir Klim, Daniel C. Javitt, Carrie E. Bearden, and Guillermo A. Cecchi. "Prediction of psychosis across protocols and risk cohorts using automated language analysis."World Psychiatry 17, no. 1 (2018): 67-75.

3.3. Trastorno bipolar

El trastorno bipolar es una patología psiquiátrica definida en el DSM-5 [100] conocida antiguamente como trastorno maníaco-depresivo. Los sujetos que la padecen transitan entre dos estados fuertemente antagónicos: un estado depresivo y un estado maníaco. Si bien hay varios tipos, en todos los casos, los pacientes transitan de un estado a otro y lo pueden hacer con diferentes velocidades. A su vez, cada estado puede estar asociado con distintos niveles de expresión. Esta enfermedad es padecida por una alta parte de la sociedad, según el trabajo realizado en [107] mas

del 2 % de la población mundial sufre algún tipo de trastorno bipolar. Esta enfermedad tiene consecuencias muy graves en lo que lo padecen. Si bien no hay un fuerte consenso, las tasas de mortalidad reportadas en algunos estudios son superiores al 15 % [108, 109] y los intentos de suicidio mayores al 30 % según el DSM-5 [100]. Las causas de la enfermedad aun son desconocidas, sin embargo existe información que siguiere una importante relación genética [110] sumado a distintos factores ambientales que promueven la aparición.

Históricamente el diagnostico de esta patología fue complejo. Originalmente eran consideradas dos enfermedades distintas, luego fue reconocida como un cuadrado donde ocurren dos estados fuertemente distintos pero relacionados entre si. Actualmente, también surgen algunas hipótesis que sugieren un continuo entre la esquizofrenia y el trastorno bipolar [111]. Si bien esto aun es solo una hipótesis, ayudaría a entender la complejidad de la identificación de este tipo de enfermedades.

El diagnostico del trastorno bipolar es difícil pues este cuadro engloba diferentes estados por lo que la evaluación tiene que ser hecha con información histórica del paciente que no siempre es fácil de conseguir. A su vez, para conseguir el diagnostico, el DSM-5 describe los distintos estadios (episodio maníaco, episodio hipomaníaco, episodio depresivo severo) y las propiedades que el sujeto expresa en cada oportunidad. Por ejemplo para el estado maníaco tienen que estar presentes tres o más de los siguientes síntomas:

1. Inflated self-esteem or grandiosity.
2. Decreased need for sleep (e.g., feels rested after only 3 hours of sleep).
3. More talkative than usual or pressure to keep talking.
4. Flight of ideas or subjective experience that thoughts are racing.
5. Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli), as reported or observed.

6. Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation (i.e., purposeless non-goal-directed activity).

7. Excessive involvement in activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments).

Es decir, para reconocer uno de los estado de la patología se deben cumplir tres de una serie de criterios altamente subjetivos. Teniendo en cuenta las características de esta patología decidimos estudiar el discurso de 20 pacientes con trastorno bipolar. Aplicamos el mismo protocolo del experimento de pacientes con esquizofrenia (ver Sección 3.1), es decir los 20 sujetos fueron diagnosticados usando los estándares del DSM mediante el rating SCID [101]. Los mismos son pacientes del Hospital Onofre Lopes dependiendo de la Universidad Federal de Rio Grande del Norte (UFRN), Brasil y del Hospital Machado, Natal, Brasil. A los participantes se les entregó la siguiente tarea para que respondan en Portugués: *Por favor reporte el recuerdo de un sueño reciente*, el discurso es grabado y transcripto por un experimentador independiente al experimento que no conoce el mismo ni la condición del paciente a modo de evitar sesgos en la transcripción. Como fue el caso de los sujetos esquizofrénicos de la Sección 3.1, las entrevistas fueron traducidas al inglés usando Google Translate). Los sujetos firmaron un consentimiento para el estudio que fue aprobado por el comité de ética de UFRN. A su vez, contamos con 20 sujetos saludables como control que respondieron a la misma entrevista.

Teniendo los discursos de los 20 sujetos patológicos y los 20 control decidimos estudiar la *intensidad emocional* (IE). La motivación de avanzar en esta dirección surgió de la interpretación de que los sujetos patológicos tuvieran mayor rango dinámico de variabilidad emocional y esto podríamos eventualmente notarlo mirando el nivel medio de la intensidad emocional como también la dispersión de las emociones por sujeto. Esta apreciación esta sustentada por trabajos de la psiquiatría clásica donde

se analiza los cambios emocionales de los pacientes [112–114]. También la literatura cuenta con estudios donde se siguiere que, usando resonancia magnética funcional, existe una modulación anormal entre las regiones prefrontal y límbica en los cerebros de pacientes con trastorno bipolar y se entiende que esto contribuye probablemente a una pobre regulación emocional y otros síntomas del humor [115].

Para cuantificar la IE usamos el algoritmo definido en la Sección 2.3. La idea del algoritmo consiste en definir la intensidad emocional de una frase como la tasa de cantidad de palabras con valoración emocional, es decir la tasa de palabras de una frase que pertenecen a la lista de palabras positivas o la lista de palabras negativas. Por ejemplo la frase "Es un hermoso día pero me siento triste" tiene solo dos palabras que tienen algún tipo de valencia emocional (según la lista de palabras positivas y negativas) que son: *hermoso* y *triste*, en este ejemplo la intensidad emocional es de $\frac{2}{8} = 0{,}25$. Para este estudio, usamos el *Dictionary of Affective Language* en inglés (ver la Sección 2.3).

Como prueba inicial para entender cuan bien cuantificaba un texto este algoritmo decidimos comparar esta medida en dos tipos de textos que creíamos que deberían tener intensidad emocional fuertemente distintas. Para eso armamos dos corpus: Artículos de Wikipedia y Selección de poemas. Para el corpus de artículos de Wikipedia tomamos los primeros 100 artículos al azar [6]. Para la selección de poemas, usamos los 75 mejores poemas definidos por *Best-Poems* [7]. Aplicando IE sobre los documentos de los dos corpus vimos una diferencia significativa (t-test, $p - valor < 10^{-46}$). Los artículos de Wikipedia presentaron una IE media y desvió estándar de $0{,}0394 \pm 0{,}0207$, para los poemas, estos presentaron una media y desvió de $0{,}1017 \pm 0{,}0363$. Tal como esperábamos, el corpus de artículos de una enciclopedia cuenta con un valor medio de emocionalidad mucho más bajo que el de poemas. Tras analizar este resultado consideramos que correspondía cuantificar el efecto del

[6] https://en.wikipedia.org/wiki/Special:Random

[7] http://100.best-poems.net/

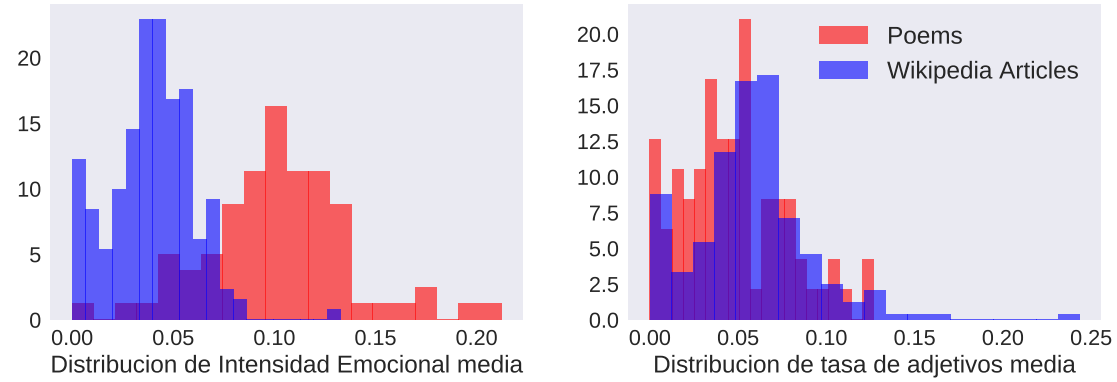

Fig. 3.4: La figura de la izquierda muestra las dos distribuciones del valor de intensidad emocional para los dos curpus distintos (poemas y artículos de Wikipedia). Se aprecia, como era de esperar que las dos distribuciones sean altamente disjuntas (t-test $p-valor = 0{,}00793$). La figura de la derecha muestra la misma compración pero no para el valor de intensidad emocionalid sino para la tasa de adjetivos en los textos. En este caso se ve que las dos distribuciones no son distintas.

género literario en las gramáticas usadas. Es decir, tal vez el algoritmo lo que este capturando sea que en textos poéticos el uso de adjetivos es mayor que en textos enciclopédicos y como probablemente haya una relación entre las palabras adjetivos (en comparación con palabras con otro rol gramatical) y la probabilidad de estar en una lista de emocionalidad decidimos cuantificar la tasa de adjetivos en los dos corpus. Comparando las distribuciones de tasa de adjetivos en los dos corpus vemos que no hay diferencia significativa para las dos poblaciones (artículos enciclopédicos IE media y desvió $0{,}0574 \pm 0{,}034$, para textos poéticos $0{,}0473 \pm 0{,}0300$ (t-test p-valor$> 0{,}6$). Para computar la tasa de adjetivos calculamos los símbolos de *part of speech* usando el *parser* de NLTK. La Figura 3.4 muestra las distribuciones de las dos comparaciones.

Tras comprobar, mediante el experimento control sobre corpus conocidos, que el algoritmo IE captura información referida a la valencia de un texto, nos abocamos a

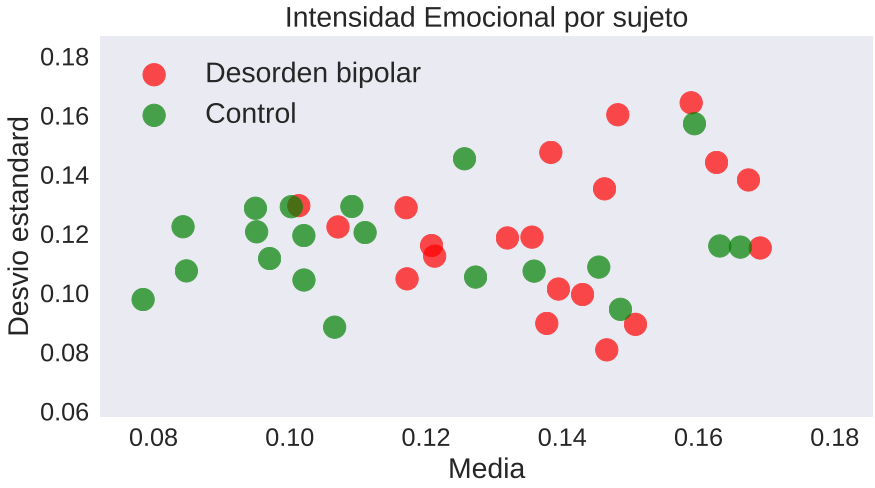

Fig. 3.5: Proyección en dos dimensiones (media y desvió estandar de cada sujeto según la clase a la que pertenecen (control o patológico).

la pregunta principal de esta sección. Para eso, analizamos los textos de pacientes con desorden bipolar y los sujetos control. Computamos IE en cada frase de cada texto, luego resumimos cada texto de cada paciente con la media y el desvió estándar. La Figura 3.5 muestra la proyección de cada sujeto. Se puede ver en esta que existe una cierta separación entre los grupos, se nota que la media de IE para los sujetos control es típicamente más baja que para los sujetos con desorden bipolar (comparando las dos poblaciones encontramos diferencias significativas, t-test $p - valor = 0{,}00793$). La IE media y desvió para el grupo control fue $0{,}1168 \pm 0{,}0277$ mientras que para el grupo patológico fue $0{,}1380 \pm 0{,}0193$.

Luego de encontrar diferencia estadística nos preguntamos si podíamos entrenar un algoritmo para clasificar automáticamente si un texto fue producido por un sujeto patológicos o uno control. Para eso tomamos como vector de *features* o dominio, como representante de cada sujeto, al vector compuesto por la media y el desvió estándar de la serie de IE de cada frase, tal como habíamos reportado anteriormente en la Figura 3.5. Usando un esquema de validación cruzada de 10 *folds* obtenemos, usando regresión logística (ver Sección 2.6.1), una performance de 75 % cuando el

Tab. 3.4: Matriz de confusión en clasificación automática con regresión logistica en esquema de 10 *folds cross-validation.*

	Predicción		
	Control	Patológico	
Control	12	8	**Clase Real**
Patológico	2	18	

valor de un modelo sin información (o modelo chance) se ubica en 50 %. La Tabla 3.4 muestra la matriz de confusión donde se ve que el modelo comete más errores de falsos positivos, es decir 8 sujetos control los clasifica como patológicos.

En este caso de estudios pudimos, mediante una técnica simple, medir parcialmente una característica, sugeridas por la psiquiatría clásica y la neurociencia, particular que definen a los sujetos con desorden bipolar: los niveles emocionales del discurso. Como se presentó, la intención de avanzar en esta dirección fue motivada por información previa. Sin embargo, también probamos que sucede con los niveles de coherencia de los sujetos patológicos y no encontramos diferencias con los sujetos control. Creemos que es importante avanzar en estos casos de estudio con hipótesis previas, dado que la relación del tamaño de la muestra y la cantidad de características del discurso a medir son muy diferentes pudiendo encontrar resultados productos del azar si se explora sin reparo y no se tiene control de este fenómeno.

Los resultados de esta sección fueron presentados en *5th NIPS Workshop on Machine Learning and Interpretation in Neuroimaging: Beyond the Scanner* y estan disponibles en [116].

3.4. Predicción de susceptibilidad de tratamiento farmacológico en depresión

En las secciones anteriores aplicamos herramientas de procesamiento del lenguaje natural y aprendizaje automático para asistir el diagnostico psiquiátrico de manera automática y objetiva. Las evidencias provistas por la experimentación con estas nuevas tecnologías abre la puerta a diferentes usos de las mismas. En esta sección exploramos un caso de uso ortogonal a lo presentado anteriormente. Diseñamos un simple modelo que funciona como un test de susceptibilidad para un tratamiento piscofarmacologico para pacientes con depresión resistente al tratamiento (TRD).

En [117, 118], Carhart-Harris et al. proponen un tratamiento piscofarmacologico usando psilocibina como agente psicodélico para tratar a pacientes con TRD. La psilocibina es un agonista de los receptores de serotonina disponible en la naturaleza en diferentes hongos usados por diversas culturas a lo largo del tiempo. Sin embargo la aplicación para tratamientos contra la depresión es reciente. Esta nueva función es posible gracias a que a diferencia de muchos fármacos antidepresivos clásicos, que funcionan como inhibidores de la recaptación de serotonina, la psilocibina funciona directamente como agonista de los receptores de serotonina.

Usando esta droga junto a una terapia psicológica los investigadores en [117] consiguen que respondan al tratamiento un 41 % de la muestra de 17 sujetos resistentes a tratamientos clásicos. Para considerar que un sujeto se encuentra en remision de la depresion o que el tratamiento fue exitoso, lo hacen en función a la reducción a menos de la mitad del valor *baseline* del test *Quick Inventory of Depressive Symptoms*(QIDS). Es decir, si los sujetos a partir de la semana 5 del tratamiento (el tiempo considerado correcto para evaluar la remisión por parte de los investigadores), consiguen un QIDS menor o igual a la mitad del valor con el que comenzaron el tratamiento, se los considera que respondieron al mismo.

En cualquiera área de la medicina y en particular en la psiquiatría las ventajas

de un diagnostico correcto son obvias pues un buen diagnostico eleva las chances de que el tratamiento asociado funcione mejor [119]. Por este motivo nos preguntamos si podríamos con información previa de los sujetos predecir cual será la respuesta al tratamiento piscofarmacológico con psilocibina. Para esto, tomamos entrevistas anteriores al tratamiento y estudiamos, como en el caso de los pacientes bipolares (ver Sección 3.3) la emocionalidad de los sujetos, pues nuevamente en pacientes depresivos los síntomas se evidencian en el contenido emocional de las expresiones que usan.

Contábamos con el registro de un *autobiographical memory test* (AMT) en inglés. Los AMT son entrevistas donde el investigador propone una palabra y el sujeto experimental debe contestar con el relato de la evocación de un recuerdo asociado. En esta versión contamos con 12 palabras por sujeto. La respuesta de los sujetos contiene en promedio (y desvió estándar) $73{,}93 \pm 47$ palabras. Para cada sujeto, calculamos la tasa de palabras positivas y tasa de palabras negativas (ver Sección 2.3 usando SentiWordNet [81] por cada una de las 12 respuestas y luego tomamos el promedio.

Antes de evaluar si los niveles de emotividad calculados eran informativos para predecir la respuesta al tratamiento. Decidimos probar esto entre sujetos con TRD y sujetos control saludables (18). Si bien existen diversos modelos y resultados en la bibliografía que abordan el diagnostico de depresión a partir del discurso [120–123] la motivación de esta experimentación no era focalizarse en este problema sino usarlo como un escenario para entender más la aplicación a sujetos con TRD. Esta comparación arrojó una diferencia significativa entre la población con TRD y los sujetos control en la tasa de palabras positivas (control media y desvió $0{,}0532 \pm 0{,}013$, TRD media y desvió $0{,}0384 \pm 0{,}01$, t-test $p - valor = 0{,}0011$). Para el caso de tasa de palabras negativas no se encontró diferencia estadística. Tomando los dos valores para representar a un sujeto, es decir el vector conformado por la tasa de palabras positivas y la tasa de palabras negativas, usando Gaussian Naive Bayes (ver Sección

Tab. 3.5: Matriz de confusión en clasificación automática de respuesta al tratamiento con Gaussian Naive Bayes en esquema de 10 *folds cross-validation*.

	Predicción		
	Control	TRD	
Control	8	2	**Clase Real**
TRD	1	6	

2.6.1) en un esquema de validación cruzada (ver Sección 2.6.2) de 10 *folds* obtuvimos 82,85 % de *accuracy* (precision = 0,82, recall = 0,82, sensitivity = 0,82, specificity = 0,83). Corrimos un experimento control donde al azar mezclamos la clase *target* y encontramos el resultado esperado, *accuracy* de 50 %. En esta evaluación usamos solo propiedades intensivas del discurso pues la longitud de la respuesta suele ser distinta. Los pacientes con depresión son menos verborrágicos que los control saludables y esto no nos hubiera permitido hacer una justa comparación del método de emotividad.

Habiendo constatado experimentalmente que la simple tasa de palabras positivas y negativas funciona para separar entre sujetos TRD y control, decidimos estudiar si esta información es útil para predecir cuales sujetos respondieron y cuales no al tratamiento. En este caso no hubo diferencia significativa ni en el uso de palabras positivas y palabras negativas para los sujetos que respondieron (7) comparándolos con los sujetos que no respondieron al tratamiento (10). Sin embargo, usando el mismo esquema de validación cruzada y el mismo clasificador obtuvimos un *accuracy* de 85 % (precisión de 75 %) en la predicción de cuales sujetos responderían y cuales no. La Tabla 3.5 muestra la matriz de confusión donde se aprecia más información de la clasificación automática.

Teniendo en cuenta esta prueba de concepto. Armar un modelo de procesamiento del lenguaje natural y aprendizaje automático que permita definir que sujetos responderán y cuales no a un tratamiento psicofarmacológico puede ser usado como un test previo a la aplicación de la intervención a modo de poder descartar sujetos

que la probabilidad de respuesta sea baja. Esto permite ahorrarle tiempo a sujetos buscando un esquema de tratamiento más adecuado para este. En función a este idea, en la Figura 3.6 esquematizamos el cambio del protocolo médico, de modo que previo al tratamiento, se ejecute una instancia de test de susceptibilidad.

Los resultados de este trabajo fueron publicados en [124]

Carrillo, Facundo, Mariano Sigman, Diego Fernández Slezak, Philip Ashton, Lily Fitzgerald, Jack Stroud, David J. Nutt, and Robin L. Carhart-Harris. "Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression."Journal of affective disorders 230 (2018): 84-86.

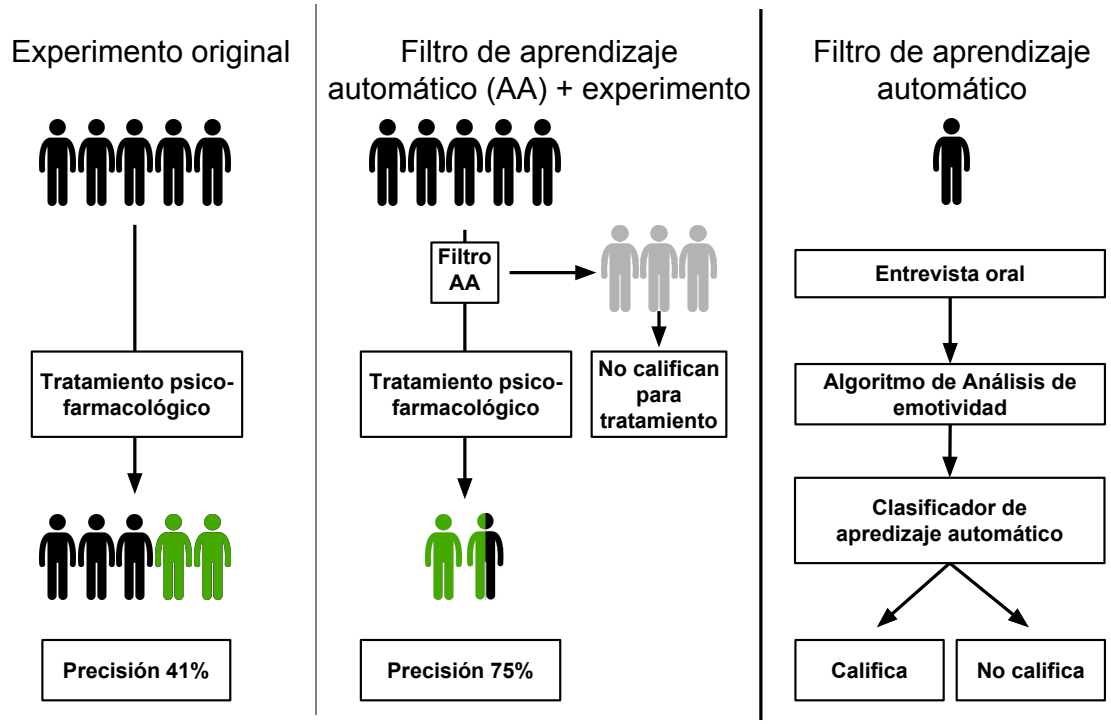

Fig. 3.6: La figura describe la diferencia del protocolo experimental clínico anterior y la nueva versión con el módulo de aprendizaje automático agregado. La primer columna muestra el protocolo previo a la intervención computacional donde se ve que los sujetos pasan por el nuevo tratamiento psicofarmacológico alcanzando una precisión de 41 %. La segunda columna muestra el nuevo protocolo, donde previo al tratamiento se evalúa con el programa (caja de Filtro AA) y este decide cuáles sujetos son aptos para el tratamiento y cuáles no. De esta manera, los sujetos que acceden al tratamiento consiguen alcanzan una precisión de 75 % en la remisión de la depresión. La tercer columna describe el filtro. Donde se ve que, primero, el sujeto produce la entrevista, a continuación el algoritmo de valuación de emocionalidad cuantifica el discurso y luego este es usado como input del clasificador ya entrenado para decidir si califica o no para recibir el tratamiento.

4. ESTADOS MENTALES ALTERADOS POR INTOXICACIONES FARMACOLÓGICAS

Las células del sistema nervioso tienen diferentes mecanismos para interactuar entre ellas. Las farmacología de las moléculas involucradas y sus receptores se estudia en diferentes aspectos. Una propiedad importante de los receptores de las neuronas es la función de afinidad que tienen con diversas moléculas, tanto de origen endógeno como exógenas. Algunas de las moléculas que interactúan con los receptores lo hacen como agonistas mientras otras como antagonistas, inclusive esta diferencia no es binaria pues existen antagonistas parciales como también ocurren fenómenos de selectividad funcional donde la activación de un receptor incluso desencadena diferentes respuestas.

El párrafo anterior, no hace mas que argumentar la complejidad enorme en la dinámica a nivel molecular sobre la interacciones moleculares. Aún así, sin entender plenamente la farmacología o la cascada de consecuencias que se tienen en la interación, el ser humano experimenta continuamente con fármacos que le producen efectos cognitivos diversos. Inclusive, muchas de estas experimentaciones tienen un fin recreativo o cultural y no estrictamente médico o necesario para la supervivencia.

En este capitulo presentamos dos casos de estudios. El primero, un experimento sobre intoxicaciones farmacológicas producto de MDMA y el segundo intoxicaciones por LSD.

4.1. Intoxicación por MDMA

El MDMA es una droga psicoactiva que produce, en el que lo consume, efectos en el estado de ánimo y en diferentes capacidades sociales. La farmacología del MDMA es conocida. Este actúa bloqueando diferentes transportadores, particularmente los

de serotonina, lo que genera que haya un exceso de este neurotransmisor disponible. A esto último le adjudican la percepción de cambios emocionales.

En [54] estudiamos el efecto del MDMA en el discurso en sujetos bajo el efecto de diferentes dosis de MDMA. El diseño experimental consistía en llevar a 13 sujetos al hospital 4 veces (separadas por una semana). Cada día se le daba una pastilla que podía contener placebo, MDMA dosis baja, MDMA dosis alta y metanfetaminas, los sujetos no sabían que pastilla tomaban cada día. Luego de tomar la pastilla, los sujetos relataban un recuerdo (en promedio relatos de 780 palabras). Con ese recuerdo trascripto usamos Latent Semantic Analysis, como Word Embedding (ver Sección 2.2), para medir la similitud semántica de los discursos de cada sujeto/condición con un grupo de conceptos elegidos. La lista de palabras que usamos para comparar surgió de la descripción de síntomas cognitivos del uso de MDMA del articulo de Wikipedia del mismo. Los resultados de ese experimentos fueron que las diversas palabras seleccionadas tenían una diferente similitud en función a la condición entre sujetos. La Figura 4.1 muestra los ejemplos de palabras que tomamos y la diferencia entre condiciones para ese experimento.

En ese experimento encontramos que existían diferencias significativas respecto a como los sujetos cambiaban sus relatos bajo el efecto de MDMA hacia direcciones semánticas propias de la descripción subjetiva de la sintomatología. Con esa información armamos un clasificador que nos permitió distinguir en un esquema de validación cruzada para cada sujeto en que momento había tomado el placebo y en que momento había tomado MDMA con una *accuracy* de 88 %.

En el 2016, conjunto a los mismos investigadores lanzamos la continuación del experimento pero con una muestra de sujetos más grande. En este caso tuvimos una muestras de 44 sujetos en la misma condición experimental que el caso anterior, pero solo placebo y dosis alta de MDMA (1.5 mg). Sin embargo, para estos sujetos las respuestas dadas son considerablemente de menor tamaño (300 palabras en promedio vs 780 en el experimento anterior), por lo que nos interesa saber si en una muestra

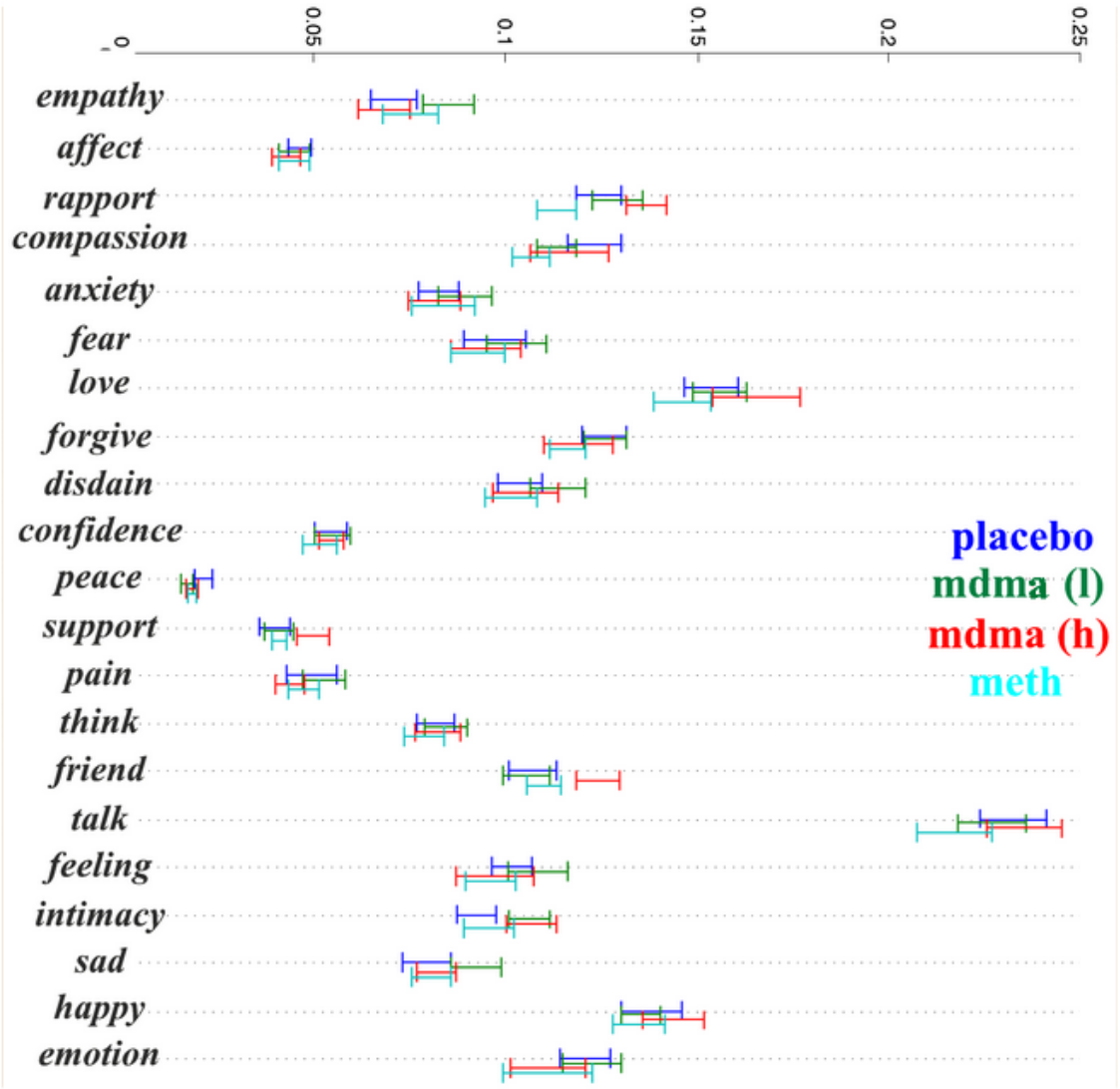

Fig. 4.1: Similitud entre palabras elegidas del articulo de Wikipedia a los discursos de sujetos agrupados por condición (placebo, MDMA(l), MDMA(h) y Metanfetamina). Cada vector representa la media y el estándar error de cada grupo. Imagen de tesis de mi tesis de Licenciatura. [125]

con un tamaño menor se conserva el efecto encontrado en la anterior cohorte y si acaso podemos entrenar modelos con una y predecir en la otra de manera exitosa. Proponer un caso de *repetibilidad* es importante para este tipo de experimentos dado la diferencia en la dimensionalidad del espacio de cambios semánticos estudiado en comparación con la cantidad de sujetos experimentales.

Para respetar el diseño experimental del trabajo anterior creamos dos vectores de *features* por sujetos. El correspondiente a la condición placebo (PBO) y el correspondiente a la condición de toma de MDMA. Cada elemento del vector correspondió a la similitud semántica usando LSA entrenado con TASA (como en el experimento original) entre una palabra elegida y el discurso en la condición correspondiente. Para resumir la similitud semántica entre un texto y una palabra, medimos la palabra contra todas las palabras del discurso y tomamos la mediana de esta distribución. La serie de palabras que usamos fueron: *compassion, rapport, intimicy, empathy, friend, nerve, connect, meaning*. A su vez, como en el experimento original, sumamos como nueva *feature*,la cantidad de palabras.

Teniendo esta representación de dos vectores por sujeto, para poder repetir el esquema de clasificación del trabajo anterior creamos dos *samples* por sujeto. Una correspondiente a tomar el vector de MDMA-PBO y la otra PBO-MDMA. Con esta representación por cada sujeto tenemos dos muestras, una por cada resta. Luego, implementamos un esquema de clasificación donde entrenamos con todos los sujetos menos uno y testeamos sobre este sujeto (con sus dos muestras), esta estrategia es conocida como *leave-one-out cross-validation*. A su vez, repetimos esto para cada sujeto y reportamos la media y el error estándar. Esta estrategia fue tomada en el experimento anterior porque la muestra era considerablemente chica (13 sujetos) por lo que no podíamos hacer *cross-validation* y caer en una situación muy distinta con información relevante en cada fold. Es importante destacar que un sujeto queda plenamente o en el conjunto de entrenamiento o en el de test pero nunca una muestra en cada uno.

Con esta estrategia definida, heredada del trabajo anterior, probamos las siguientes configuraciones experimentales:

1. Usamos exclusivamente la cohorte nueva (44 sujetos) con un esquema de *leave-one-out cross-validation*

2. Entrenamos el modelo con la cohorte nueva (44 sujetos) y testeamos en la cohorte del primer experimento (13 sujetos)

Para todos los casos probamos entrenar con dos clasificadores distintos para compatibilizar el experimento anterior, Support Vector Machine [126] y K-vecinos más cercanos (ver Sección 2.6.1).

Los resultados obtenidos fueron: para el caso 1, usando el clasificador K-vecinos más cercanos ($k = 3$) obtuvimos *accuracy* promedio de de 0,775±0,008, usando SVM con kernel RBF obtuvimos un accuracy de $0{,}8125 \pm 0{,}0055$. Para el caso 2, usando K-vecinos $k = 3$ obtuvimos *accuracy* de $0{,}84 \pm 0{,}043$, usando SVM $0{,}62 \pm 0{,}021$.

Estos resultados muestran dos propiedades interesantes del método y el efecto poblacional. Primero encontramos que pudimos replicar los resultados del experimento en una población mayor, pero a su vez, el tamaño de la muestra por sujeto es menor. En principio creímos que esto seria un problema dado la propiedad intrínseca de un sistema ruidoso como es el semántico en un ambiente inclusive de análisis con muestras chicas. En segundo lugar, encontramos que las dos poblaciones comparten entre si las propiedades de deformación del discurso, pues con el experimento 2, entrenando en el dataset con más sujetos pudimos obtener una buena *performance* de clasificación testeando sobre un dataset independiente. Esto constituye y le otorga mayor confianza y robustez al método y al efecto cognitivo poblacional del MDMA.

Analizando los cambios estadísticamente con un test de hipótesis apareado por sujeto, para las 9 palabras, no encontramos diferencias significativas luego de hacer correcciones por múltiples comparaciones. Probablemente el efecto encontrando por

los clasificadores se pueda medir estadísticamente en algún tipo de interacción de las variables no presente en la comparación unidimensional.

Como dijimos, intentamos reproducir las condiciones del experimento anterior, sin embargo este experimento con un tamaño de muestra más grande permitiría estudiar diferentes propiedades del lenguaje e inclusive modificar el diseño experimental de validación. Si bien no nos extendimos en esto, medimos el efecto de agrandar el conjunto de test pasando de un esquema de *leave-one-out cross-validation* a uno de *leave-n-out cross-validation*, es decir estudiamos que sucedía si no testeabamos solo en un sujeto sino en varios a la vez. Este experimento lo realizamos para el caso 1, es decir donde testeamos sobre la nueva cohorte de 44 sujetos para SVM. Los resultados de este experimento mostraron que la *accuracy* que empieza con un valor promedio de 0,8125, si dejamos dos sujetos cambia a 0,77, si dejamos de test 5 sujetos pasa a 0,75 y este programación continua disminuyendo hasta 20 sujetos donde se mantiene en 0,66, todavía aun por arriba del nivel del modelo azar. Esta última evidencia contribuye a sostener la validez del resultado obtenido.

Los resultados presentados en esta sección fueron publicados en [54].

Bedi, G., Cecchi, G.A., Slezak, D.F., Carrillo, F., Sigman, M. and De Wit, H., 2014. A window into the intoxicated mind? Speech as an index of psychoactive drug effects. Neuropsychopharmacology, 39(10), p.2340.

4.2. Intoxicación por LSD

LSD es una droga psicoactiva que afecta al sujeto de diferente maneras. Desde cambios fisiológicos hasta cambios psicológicos, de performance cognitiva, de percepción, cambio de ánimo, entre otros. El estado psicodélico al que el sujeto llega mediante la ingesta de dosis correctas incluyen delirios, alucinaciones y otros efectos buscados por el consumo recreativo[1]. En la década de los 70' se exploro el uso del

[1] https://elgatoylacaja.com.ar/sobredrogas/psicodelicos/

LSD en terapias psicofarmacológicos [127] y situaciones como el alcoholismo [128]. Debido a políticas prohibicionistas el uso en la mayoría de los países se vió cancelado, incluso para investigaciones médicas. Recientemente diversos estudios con drogas psicodélicas en general están mostrando cómo estas pueden servir en terapias psicofarmacológicas por lo que su uso terapéutico es discutido nuevamente por la comunidad científica. Por esto mismo, recientes trabajos están aportando las primeras evidencias en un nuevo nivel, un ejemplo de esto es presentado en [129] donde los resultados muestran correlatos en la actividad neuronal producto del LSD abriendo un nuevo espacio de estudio por fuera del psicológico subjetivo y el meramente farmacológico.

Para estudiar el efecto del LSD en el lenguaje desde una perspectiva computacional contamos con los datos de un experimento realizado por el grupo de Robin Carhart-Harris de Imperial College London. En el experimento, los investigadores, entre otras condiciones no incluidas en nuestro experimento, someten a cada sujetos (17) a dos condiciones: placebo y LSD, ambas condiciones son ignoradas por los sujetos. En cada condición se les practica un estudio médico no invasivo y luego se registra una conversación respecto a cómo se sintieron durante el estudio. Para nuestro experimento obviamos las intervenciones orales de los investigadores y solo nos quedamos con lo dicho por lo sujetos. En [130] se encuentra la información completa del experimento completo e información demográfica de los sujetos.

En este experimento no contamos con una hipótesis previa a medir sobre los efectos del LSD en el lenguaje, a diferencia del caso de estudio de pacientes con esquizofrenia donde modelamos la incoherencia. Debido a esta falta, decidimos estudiar los efectos en la emocionalidad ya que es una de las características del lenguajes más simples y menos multidimencionales para explorar. Para eso, medimos la tasa de palabras positivas, negativas y neutrales, como fue descripto en 2.3, usando SentiWordNet3.0 como corpus de entrenamiento. La Tabla 4.1 muestra los resultados de comparar estas 4 características. Para comparar entre sujetos normalizamos

Tab. 4.1: Características de emocionalidad. Se reporta la media y el desvió estándar de el cociente entre el valor para cada medida de LSD sobre Placebo para todos los sujetos. La normalización se debe a que cada sujeto puede tener un nivel basal distinto y la condición placebo normalizado esto. El p-valor reportado corresponde a el resultado de aplicar *Student Test* apareado sobre la muestra. Haciendo corrección por múltiples comparaciones la significancia es valida para: positivo y neutral.

Neutral	**Positivo**	**Negativo**	**LSD / Placebo**
0,9757	1,0159	1,9285	**media**
±0,0073	±0,1118	±0,2098	**error estandar**
0,0051	0,9056	0,0009	**p-valor Student Test**

cada característica por sujeto resumiendo la dos condiciones en un valor tomando el cociente entre el valor en el estado LSD sobre el valor para el estado Placebo. Esta normalización es necesaria para poder comparar en efecto que produce el LSD en términos de la diferencia con el estado basal del sujeto. Por eso, por ejemplo, el valor de *Positivo* es 0,9757, debido a que corresponde a tomar la media entre todos los cocientes de LSD/Placebo para cada sujeto. La Figura 4.2 muestra la proyección de los sujetos en las dos condiciones para las dos características con diferencia significativa.

Este resultando muestra que existe una diferencia significativa en el cambio de las emociones producto del efecto de LSD en una muestra pequeña de 17 sujetos observable con un algoritmo trivial. La siguiente pregunta que nos hicimos, fue si podíamos entrenar un clasificador y medir si esta información nos servía para clasificar automáticamente las condiciones por sujeto. Para eso, usamos la misma metodología que la del caso de MDMA, en la cual armamos dos muestras por sujeto, correspondientes a LSD-Placebo y Placebo-LSD y usamos la estrategia de *leave-one-out cross-validation* (ver Sección 4.1). Haciendo esto, y usando un clasificador de

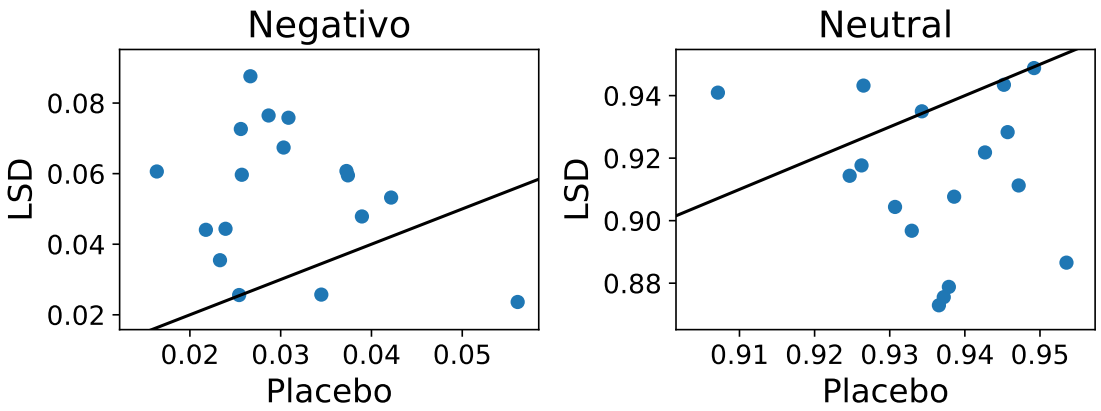

Fig. 4.2: Negatividad y Neutralidad de los sujetos proyectado para las dos condiciones. La Linea negra marca la relación x=y. La Figura muestra claramente como para el caso de la característica Negatividad los sujetos, en la condición LSD están por encima de la recta negra lo que significa que crecen en el uso de expresiones negativas. En el caso de Neutral, los sujetos están mayormente por debajo de la recta negra lo que significa que en LSD disminuyen el uso de palabras neutrales.

Vecinos más cercanos con $k = 3$ obtuvimos un *accuracy* de $0{,}8823 \pm {,}0781$. Es decir, tomando solo propiedades intensivas del lenguaje referidas a la emocionalidad, de la forma mas simple, es decir tomar la tasa de uso, podemos clasificar muy bien la condición en que se encuentran los sujetos. A su vez, da la diferencia tan marcada en la Neutralidad por la Figura 4.2 nos preguntamos si hacia falta recurrir a normalizar por sujeto para poder clasificar. Para eso armamos un dataset donde agregamos por cada sujeto dos muestras, una para la condición LSD y otra para Placebo tomando las 3 *features* definidas anteriormente. Haciendo esto y usando un Árbol de Decisión encontramos que tomando solo Neutralidad podíamos clasificar con 0.75 de *accuracy* en un esquema de validación cruzada de 10 *folds*. Es decir, la normalización por sujeto ayuda para detectar y predecir la condición pero no es necesaria para tener una buena clasificación.

En este análisis mostramos que el LSD repercute poblacionalmente en la emocionalidad, al menos, para este experimento donde los sujetos charlar con los investigadores. Por supuesto este caso es chico para generalizar el efecto del LSD como disparador del uso de palabras negativas. A su vez, creemos que para el investigador resulto evidente la condición del sujeto por lo que este puede haber sesgado las preguntas y respuestas que conducen a que el sujeto experimental se comporte de una manera distinta. Por lo que creemos que correspondería repetir el experimento pero en una entrevista más estandarizada donde el sujeto responda siempre a las mismas preguntas y no dependan las preguntas del experimentador a las respuestas dadas por este. Sin embargo contando con estos datos analizar este tipo de propiedades resulta interesante.

5. ESTADOS MENTALES ALTERADOS POR CAMBIOS ENDOCRINOS

El sistema endocrino usa sustancias llamadas hormonas como mensajeros en un complejo sistema de señalización entre células . Los hormonas son usadas para coordinar diferentes mecanismos de la fisiología de los seres vivos. Alguno de las funciones que coordina son: el crecimiento, el metabolismo, funciones inmunológicas, funciones cardíacas, regulación del estrés, regulación del sueño, entre otras. Estas hormonas tienen blancos en distintas partes del cuerpo, en especial en el cerebro donde la interacción de este con el sistema endocrino afecta directamente en sus producciones y la mente.

En este capitulo exploramos dos casos de estudio relacionados entre si. 1) Alteraciones mentales producto del ciclo menstrual (Sección 5.1, 2) Alteraciones mentales producto del embarazo (Sección 5.2).

5.1. Ciclo Menstrual

El ciclo menstrual (CM) ocurre en las mujeres en edad fértil típicamente cada $28,1 \pm 3{,}95$ días [131] como resultado de una coordinación compleja de hormonas, diferentes órganos interactuando y sistemas de *feedback* [132]. Desde una perspectiva evolutiva el CM es el encargado de terminar de madurar las gametas y preparar el cuerpo de la mujer para un posible embarazo. Para coordinar estas funciones usa esencialmente 4 hormonas (FSH, LH, estrógenos, progesterona) producidas por básicamente el eje hipotálamo-hipófisis-ovario. La Figure 5.1 detalla la dinámica poblacional de estas hormonas.

El CM afecta diferentes funciones fisiológicas, incluyendo regulaciones del sistema inmunológico [134], tasa metabólica [135], cómo responde el cuerpo al ejercicio [136,

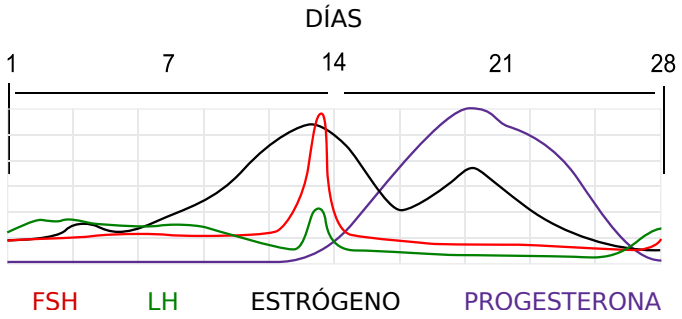

Fig. 5.1: Dinámica de concentración de hormonas durante el ciclo menstrual. Datos tomados de [132], recursos gráfico basada de [133].

137], etc. También regula funciones cognitivas, de comportamiento y emocionales, como la memoria [138, 139] , toma de decisiones [140, 141], percepción dolor [142–144], preferencia y frecuencia sexual [145–148] y estado del ánimo [149–152]. Sin embargo, no existe una caracterización sobre los efectos en el lenguaje. Es por eso nos preguntamos si existe alguna relación entre el CM y cambios del lenguaje.

Como mencionamos anteriormente, la evidencia en la bibliografía reporta cambios en el estado de ánimo de los sujetos producidos por ciclo menstrual o por las hormonas que lo coordinan. Estos reportes usan test estandarizados o apreciaciones subjetivas para interpretar el estado de ánimo de los sujetos y no métodos más objetivos que operen sobre el lenguaje. Considerando las innumerables propiedades del lenguaje a estudiar decidimos comenzar por aquella que sugiere una continuación de las observaciones de más alto nivel ya reportadas. Por eso estudiamos la emocionalidad como primer propiedad a relevar del lenguaje. Teniendo definida la propiedad a estudiar nos preguntamos de que modo opera el CM en el lenguaje y si esta diferencia podíamos verla en sujetos mujeres en comparación con los sujetos hombres.

Para estudiar esta pregunta armamos un conjunto de datos que nos permitiera estudiar a lo largo del tiempo los efectos en el lenguaje producto del ciclo menstrual, para eso usamos Twitter. Recolectamos todos los mensajes de los 418 sujetos

(184 mujeres y 234 hombres) entre 18 a 40 años (en promedio 26,8 años con un desvío estándar de 5,9 años). Twitter es una red social particular, sus mensajes se caracterizan por esta limitados a 140 caracteres, esto probablemente condicione a un uso a su vez particular en comparación con otras redes sociales. Muchos de los usuarios de Twitter son extremamente constantes en la generación de tweets (los mensajes de la plataforma). En nuestra muestra obtuvimos una producción promedio de 7,31 tweets por día con un desvío estándar de 2,22 por sujeto. El conjunto de datos que armamos es comparable e incluso mayor con otros trabajos donde se analizan fenómenos del lenguaje usando Twitter, u otras redes sociales, como fuente de datos [120, 153, 154].

Habiendo armado un corpus para estudiar como el CM opera sobre el lenguaje computamos las medidas relacionadas con la valoración de sentimientos (ver Sección 2.3) entrenando los modelos con Spanish DAL [79]. Por lo que para cada tweet computamos la tasa de palabras positivas, negativas y la intensidad (ver Figura 5.2 A). Luego para cada sujeto armamos dos series temporales que describen como la dinámica en los puntajes de emociones fluctúa a lo largo del tiempo. La primera de las series usa la media para obtener un valor diario por sujeto, es decir esta serie temporal representa la media agrupada por día de la propiedad del lenguaje para cada día para cada sujeto. La segunda serie usa el máximo como función para resumir un día. Decidimos usar el máximo pues consideramos que este captura el mensaje más intenso, positivo o negativo según que propiedad del lenguaje estuviéramos viendo. La Figura 5.2 B ejemplifica cómo a partir de una serie de tweets de un sujeto se forman las dos series temporales. Un análisis estadístico sobre la magnitud del uso de las distintas emociones no muestra diferencia significativa en el uso por parte de un grupo u otro (sujetos masculinos vs sujetos femeninos), esto significa que no habría diferencia en cuanto al uso de emotividad según el sexo estudiándolo de manera estacionaria.

En materia de diferencias observables como fenómenos dinámicos, el ciclo mens-

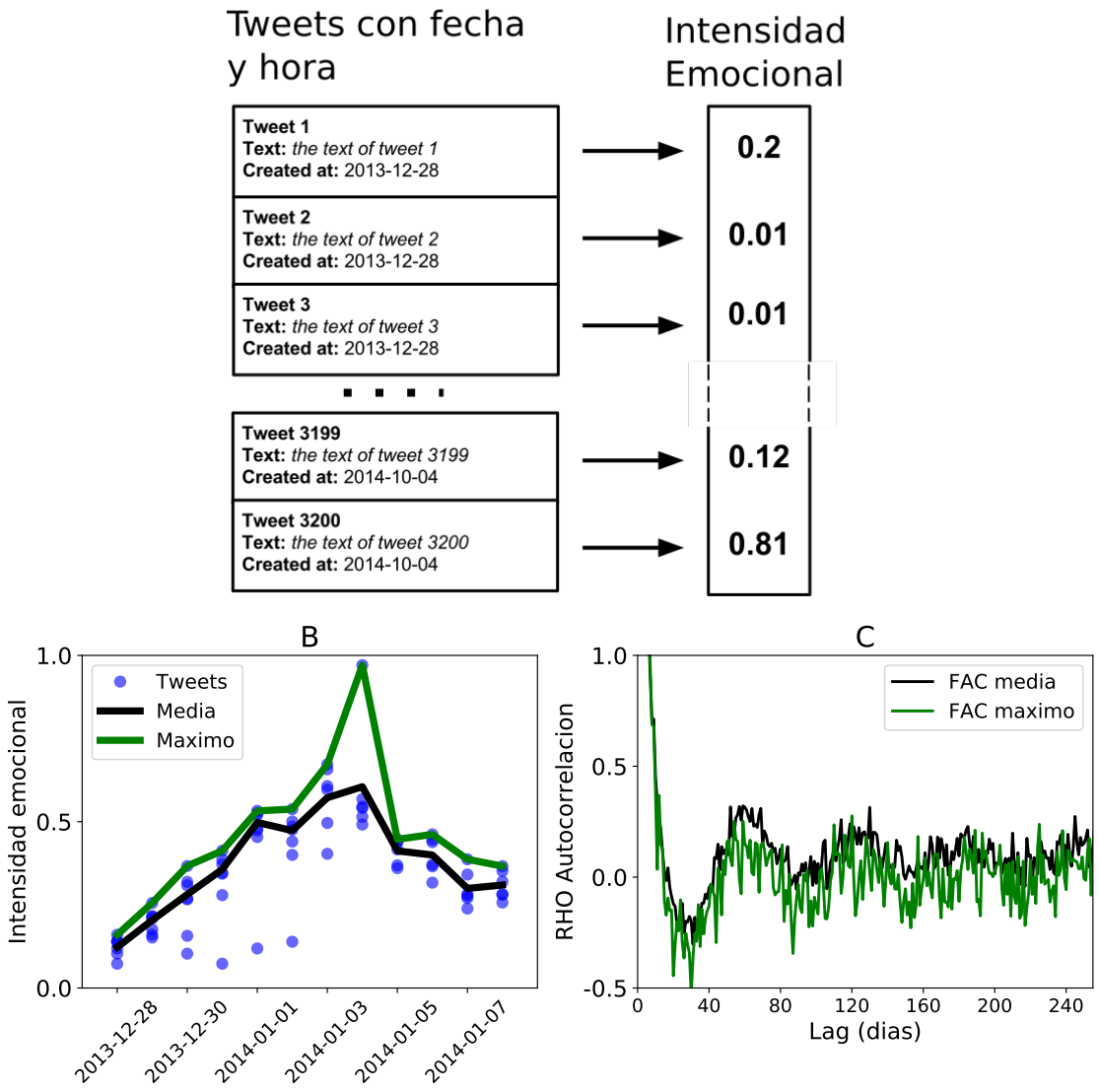

Fig. 5.2: Diseño experimental. La Figura muestra un ejemplo del proceso que implementamos por sujeto. La primer parte (A) describe el proceso de bajar los Tweets y procesarlos, donde obtenemos los últimos 3200 tweets por sujeto y calculamos la intensidad emocional para cada mensaje. La segunda parte (B) corresponde a resumir los días con más de un mensaje a un solo valor en dos versiones. La versión que usa el máximo valor y otra que corresponde a usar la media. De este modo para este paso, cada sujeto tiene dos series temporales proveniente de la intensidad emocional. La tercer parte corresponde en calcular la función de autocorrelación para las dos series anteriores para distintos valores de *lag*.

trual podría estar afectando de diferente forma el lenguaje en función a en que momento del ciclo menstrual cada mujer se encontrase. Sin embargo no contábamos con la fase del ciclo de cada sujeto, es decir, dado un día del calendario no sabíamos en que día del ciclo menstrual propio de cada mujer se encontraba. Por lo que decidimos estudiarlo desde otra perspectiva. Si el ciclo menstrual estuviera operando sobre el lenguaje, y este tiene en promedio 28 días, deberíamos ver que el grupo de sujetos femeninos tuviera, cada uno en particular, un sincronismo consigo mismo más fuerte y distinto que el grupo de sujetos masculinos.

Para responder esta pregunta, usamos la función de autocorrelación (FAC). La FAC es la correlación entre una señal consigo mismo pero con su versión *movida* por distintos valores llamados *lag*. Es decir definimos:

$$FAC(i) = correlacion(s[i:], s) \tag{5.1}$$

donde $s[i:]$ significa descartar los primeros i valores de la serie o mover la serie para el valor de $lag{=}i$. Esta herramienta permite estudiar patrones repetitivos dentro de una señal, pues si existe un patrón de longitud n que se repite, $FAC(n)$ cuantifica esto. Con esta herramienta, tomamos las dos series temporales, la que usa la media para resumir los días y la del máximo. Por lo que cada sujeto es representado ahora con dos funciones de autocorrelación, una que llamamos FAC media y la otra FAC max, la Figura 5.2 C completa el diseño experimental. En este experimento siempre como medida de correlación dentro de FAC usamos la correlación de Pearson.

A diferencia de las series temporales, donde no conocíamos la fase del CM, las dos funciones de autocorrelación (FAC media y FAC max) son agrupables. Esto nos permite estudiar el efecto poblacionalmente. La Figura 5.3 muestra las dos versiones de funciones de autocorrelación originadas a partir de la serie de intensidad emocional, agrupadas por sexo reportando el promedio y el error estándar como medida de de dispersión.

Las *FACs* muestran dos patrones claros. El primero es un fuerte decaimiento,

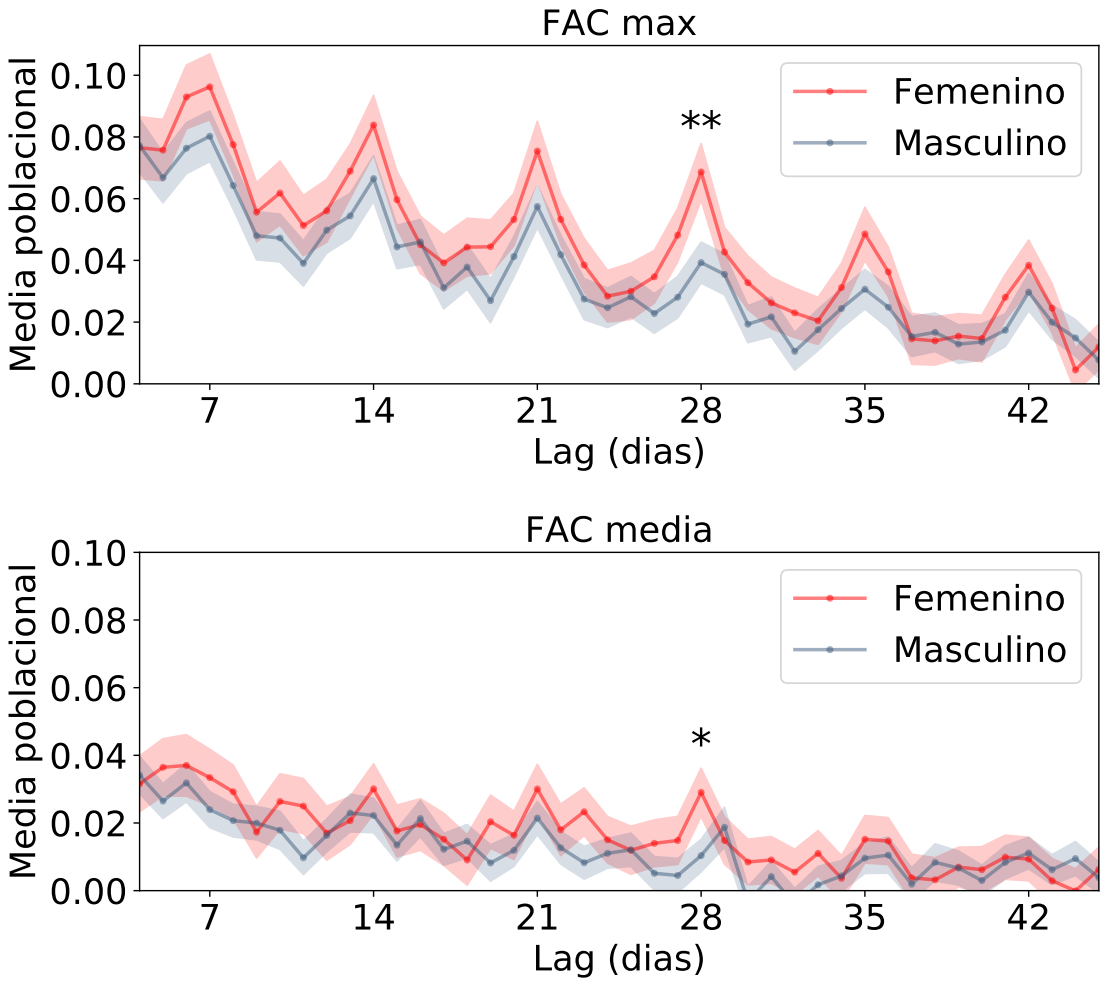

Fig. 5.3: Media y error estándar de la función de autocorrelación para las series de Intensidad Emocional agrupadas por genero. La serie roja representa a los sujetos femeninos y la azul a los masculinos. La parte A muestra las FAC derivadas de la Intensidad Emocional usando máximo como función de resumen diario. El doble asterisco marca solo el punto para $lag = 28$, único punto donde los dos grupos son diferentes significativamente ($pval = 0{,}0085$). El Panel B muestra que FAC media. El asterisco simple marca solo el punto para $lag = 28$, donde los dos grupos son significativamente diferentes ($pval = 0{,}0260$).

esto simplemente indica que a medida que el tiempo pasa, la similitud emocional entre dos días disminuye. El segundo efecto muestra que existen picos semanales sobre la señal principal, lo que indica que los niveles de emoción tienden a tener patrones similares en los mismos días de la semana (es decir, los martes son emocionalmente similares a los martes como los sábados a los sábados). Más allá de estos dos efectos, nuestra hipótesis principal era que los sujetos femeninos deberían presentar, comparado a los sujetos masculinos, un comportamiento diferente al rededor de 28 días (el promedio poblacional del ciclo menstrual). Las *FACs* de sujetos femeninos se encontró ligeramente por arriba a la de los sujetos masculinos para todo valor del dominio. Sin embargo, comparándolo estadísticamente encontramos que el único punto significativamente distinto entre los dos grupos es para el *lag* de 28 días, para las dos versiones de *FACs*. Para FAC max $p - valor = 0{,}0085$, para FAC media $p - valor = 0{,}0260$. Esto quiere decir, que el único lag donde los dos grupos son estadísticamente diferentes entre si es para 28 días. Es decir, el único patrón temporal diferente en la emocionalidad entre grupos es para 28 días, vale aclarar que, como se menciono, para los sujetos femeninos este patrón es más fuerte, es decir la autocorrelación es significativamente mayor que para los sujetos masculinos. La hipótesis que mejor explica este fenómeno es la intervención del ciclo menstrual en procesos cognitivos como el estado de ánimo repercutiendo en el lenguaje.

En el resultado expuesto, consideramos que no era necesario hacer correcciones por múltiples comparaciones pues la hipótesis planteada era estudiar si la diferencia a los 28 días era significativa. Tras contestar esto y cuantificar con un test de hipótesis la probabilidad de un resultado espurio, medimos, por completitud, si el mismo efecto se ve presente para ciclos de otro tamaño. Si el experimento hubiera sido buscar para qué ciclos la diferencia fuera significativa, entonces si correspondería hacer una corrección por múltiples comparaciones pues, por azar con un nivel de aceptación de 0,05 algún resultado en 20 comparaciones hubiera pasado probablemente el umbral de aceptación.

Decidimos usar la intensidad emocional en principio porque consideramos que no teníamos evidencia suficiente para suponer que alguna de las dos direcciones posibles de la emocionalidad, según el modelo que usamos, tuviera un cambio frente a la otra. Sin embargo, habiendo mostrado que la intensidad cambió corresponde de manera exploratoria describir los cambios en la positividad y la negatividad pues podría ser que la intensidad se viera afectada por el cambio en una de sus componentes. Un análisis análogo al ejecutado con las *FACs* provenientes de la serie de intensidad emocional muestran que no hay diferencia significativa si se toman las dos series por separado. Es decir, que el cambio en la intensidad se expresa como un cambio integral en la emocionalidad y no en una dirección particular de esta.

Este experimento, presentó evidencia sobre patrones que sincronizan la emocionalidad en sujetos femeninos que ocurren cada 28 días no presentes, o presentes mas atenuados en la población masculina. A su vez, mostró que en el rango analizado este ciclo fue el único que presentó diferencias entre las dos poblaciones. La explicación más simple a esto es debido a la interacción entre las hormonas que coordinan el ciclo menstrual y ciertas partes del cerebro ligadas. Para poder argumentar más esta hipótesis correspondería realizar más experimentos que aborden la misma pregunta desde otras perspectivas.

A su vez, tomamos la dirección del lenguaje más simple y relacionada con los cambios de estado de ánimo. Esta decisión fue determinada debido a que el corpus que teníamos no permitía hacer un análisis de otro tipo de propiedades del lenguaje que requieren más texto (como coherencia del discurso).

Desde nuestro campo de trabajos, con motivo de presentar más evidencia sobre como las hormonas pueden estar afectando el lenguaje, una de las más privilegiadas funciones cognitivas, decidimos estudiar otro *caso natural* y disponible en Twitter, donde estas hormonas cambian en su dinámica de concentración. Para eso en la Sección 5.2 estudiamos los efectos en el lenguaje en mujeres durante el embarazo.

Los resultados presentados en esta sección fueron publicados en [155].

Gallino, L., Carrillo, F. and Cecchi, G.A., 2019. Differential 28-days cyclic modulation of affective intensity in female and male participants via social media. Frontiers in integrative neuroscience, 13.

5.2. Embarazo

El embarazo de las mujeres se estudia desde casi todas las áreas de la ciencia. La biológica contribuye con conocimientos básicos pero también pragmáticos, desde un nivel molecular estudiando las interacciones farmacológica hasta el estudio de regulaciones endocrinas complejas [156–159]. Pero no solo los estudios se basan en la *madre*, lógicamente también se estudia el sujeto gestante. En la bibliografía abunda la evidencia que muestra como el macro-ambiente durante la gestación impacta directamente en el futuro del infante, no solo en la interacciones química/farmacológica [160–164] sino también en aspectos psicológicos y sociales donde se lleva adelante el embarazo [165–170]. En relación a este último tipo de factor, distintas perspectivas neurocientíficas estudiaron la relación entre los estados mentales de la madre y los problemas gestacionales, mayormente centrándose en el estrés, depresión, ansiedad y como todos estos fenómenos interactúan con el desarrollo del infante [171, 172].

Más allá de los estudios relacionados con patologías o problemas durante el embarazo, algunos trabajos [173–176] estudiaron aspectos psicológicos de la madre en casos sanos usando metodología cualitativa sobre auto-reportes de sentimiento o mecanismos de formularios estandarizados [177]. Mayormente en estos trabajos los sujetos reportan su estado emocional en diferentes momentos del embarazo. Por ejemplo en [178] los autores describen los cambios de afectividad durante el embarazo usando test estandarizados psicológicos. Los trabajos mencionados tienen una fuerte limitación, los experimentos deben ser llevados a cabo en laboratorios o consultorios médicos pues o bien las entrevistas son ejecutadas personalmente o son

requeridos practicas de laboratorio para tomar muestra de marcadores fisiológicos. En contraste a esto, en este experimento, proponemos realizar un experimento que tenga un muestreo casi continuo de alta frecuencia sobre la emocionalidad de los sujetos durante el embarazo. Para lograr esto sacamos el experimento del laboratorio usando los mensajes de Twitter de mujeres embarazadas como *proxy* al estado emocional.

Este experimento nos permite seguir explorando la relación entre la concentración hormonal y los estados de ánimos reflejados en el lenguaje como continuación del experimento de ciclo menstrual. Puntualmente planteamos dos hipótesis de trabajo: 1) Los cambios hormonales durante el embarazo producen efectos emocionales consistentes en la población y pueden medirse a través del lenguaje. 2) Existe una correlación entre alguna medida del lenguaje y la dinámica hormonal.

Para este experimento conseguimos una muestra de 75 mujeres argentinas de lenguaje español nativo mayores de 18 años que reportaron que durante todo el embarazo no presentaron complicaciones y llegaron con su embarazo a termino sin ninguna patología asociada. A su vez conseguimos un grupo control de mujeres no embarazadas y otro de hombres. Para cada sujeto descargamos la máxima cantidad de *tweets* disponibles (3200 por limitaciones de la API REST). Luego de bajar los mensajes calculamos tres valoraciones: la positividad, la negatividad y la neutralidad a cada mensaje y armamos 3 series por sujeto tomando el puntaje promedio por día (como en el caso del ciclo menstrual, ver Sección 2.3 y Sección 5.1). Como teníamos la fecha del parto indexamos los mensajes teniendo ese día como día 0 por lo que los días anteriores al parto corresponden a números negativos y los posteriores a positivos. Para poder comparar entre sujetos normalizamos las tres series dividiendo los valores de cada serie por la media de la misma. Esta nacionalización *pierde* la valoración absoluta pero para este experimento eso no es un problema pues nos interesaba medir los cambios intra-sujeto, por lo que la magnitud podía ser descartada. La

nacionalización usada es la típica en el estudio de emociones en Twitter [153].

Para estudiar la primera hipótesis, decidimos separar los mensajes de cada sujeto en 4 etapas:

- *antes* : Los mensajes anteriores al estado *embarazo*. Para representar este intervalo tomamos los tweets entre 57 semanas antes del parto hasta 42 semanas antes del parto.

- Primer trimestre (*1T*): Los mensajes correspondientes al primer trimestre de embarazo. Para representar este intervalo tomamos los tweets entre 42 semanas antes del parto hasta 27 semanas antes del parto.

- Segundo trimestre (*2T*): Los mensajes correspondientes al segundo trimestre de embarazo. Para representar este intervalo tomamos los tweets entre 27 semanas antes del parto hasta 13 semanas antes del parto.

- Tercer trimestre (*3T*): Los mensajes correspondientes al tercer trimestre de embarazo. Para representar este intervalo tomamos los tweets entre 13 semanas antes del parto hasta un día antes a la fecha de parto inclusive.

- *luego* : Los mensajes posteriores al parto, por limitaciones de la muestra los mensajes posteriores al parto van desde el día del parto hasta 4 semanas después de este.

Luego de tener esta partición descartamos aquellos sujetos que tuvieran menos de 200 tweets por intervalo (quedándonos con 59 sujetos). Con esta nueva muestra estudiamos para las tres medidas su correlación con etapas o periodos. La Figura 5.4 muestra, para cada medida, la media y el error estándar por etapa. La dinámica de las tres medidas emocionales fue distinta.

La serie de valores positivos presentó un comportamiento de decaimiento desde el periodo *antes* hasta el tercer trimestre (*3T*) (correlación de Pearson negativa

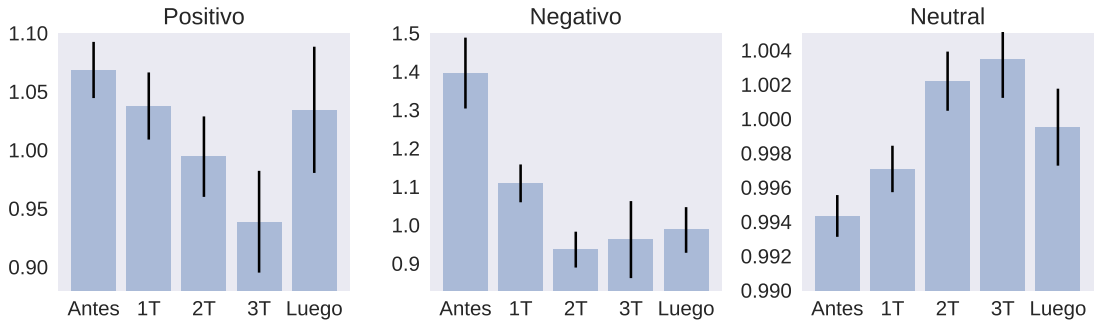

Fig. 5.4: Medias y errores estándar de medida emocional poblacional agrupadas por trimestres. Se aprecia las diferentes dinámicas de decaimiento. La serie *positivo* presenta un comportamiento de decaimiento desde el periodo *antes* hasta el $3T$. Luego del parto, la emocionalidad positiva crece a un nuevo estado indistinguible con el periodo *antes* . La serie *negativo* presenta un decaimiento desde el periodo *antes* hasta al último trimestre. En este caso la comparación entre estado *antes* y *luego* son significativamente diferentes (paired ttest $pval = 0{,}0031$). La serie *neutralidad* presentó, como esperábamos un comportamiento de crecimiento desde el periodo *antes* hasta el último trimestre $3T$. Comparando los estados *antes* y *luego* no encontramos diferencia significativa

$\rho : -0{,}1838$, $pval = 0{,}0048$). Aunque luego del parto, la emocionalidad positiva crece a un nuevo estado indistinguible con el periodo *antes* (paired ttest $pval = 0{,}6212$).

La serie de valores negativos presentó un fuerte decaimiento desde el periodo *antes* hasta el último trimestre *3T* (correlación de Pearson negativa $\rho : -0{,}2663$, $pval = 3{,}8044 \times 10^{-5}$). Para esta serie, incluso luego del parto, el valor de emocionalidad negativa permanece en un valor bajo estadísticamente diferente al estadio anterior al embarazo (*antes*) (paired ttest $pval = 0{,}0031$).

La serie de neutralidad presentó, como esperábamos por como esta definida, un comportamiento de crecimiento desde el periodo *antes* hasta el último trimestre (*3T*) (correlación de Pearson positiva $\rho : 0{,}2676$, $pval = 3{,}4947 \times 10^{-5}$). Comparando los estados *antes* y *luego* no encontramos diferencia significativa (paired ttest $pval = 0{,}066$).

Haciendo el mismo análisis pero no separando los mensajes en periodos sino que tomando el día como *offset* con relación al parto, las correlaciones se mantienen.

La Figura 5.5 resumen los cambios en comparación con el estadio *antes* . Observamos que durante el embarazo el uso de palabras neutrales aumenta, es decir, la muestra de sujetos mujer se mueve a un estado emocional menos emotivo o menos intenso. Pero luego del parto las emociones crecen nuevamente pero esta vez con un balance distinto entre emociones positivas y negativas. Las mujeres luego del parto se vuelvan a un uso del lenguaje mas positivo. Para esta comparación agrupamos los estadios de los tres trimestres en uno llamado *durante*. Para la positividad, el estadio *durante* presenta una caída de un 11 % respecto al estadio *antes* ($p-valor = 0{,}007$), y sin cambio estadísticos para el estadio *luego* ($p - valor > 0{,}6$). Para la serie de negatividad, para los dos casos (*durante* y *luego*) estos presentaron una caída significativa ($p - valor < 0{,}0031$ y $p - valor < 0{,}0029$ respectivamente). Para la emocionalidad neutral el estadio *durante* presento un incremento ($p - valor < 0{,}00074$) y en comparación con el estadio *luego* no presentó ninguna diferencia significativa $p - valor < 0{,}081$.

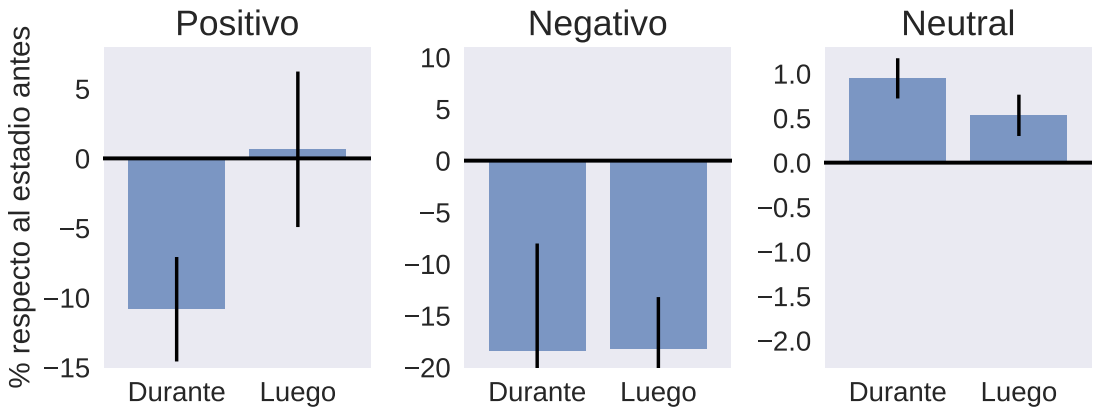

Fig. 5.5: Media y error estándar de las diferentes medidas emocionales como porcentaje respecto al estado basal de los estados *durante* y *luego* .

Si bien todas las medidas fueron significativas, repetimos todas las comparaciones con la población control. Para esto armamos diferentes grupos de 75 sujetos no embarazados (de cuentas de Twitter de un experimento anterior) probando diferentes combinaciones de sexo y edad. En los mil grupos distintos de 75 sujetos que armamos al azar (simulando los distintos periodos conservando la distribución de los sujetos embarazados) no conseguimos reproducir ninguna correlación significativa. Esto sostiene más el valor de las correlaciones anteriores en la población no control.

El resultado anterior manifiesta que existe un efecto en el lenguaje producto del embarazo. Relacionar este efecto a una hormona o varias puede ser difícil pues la construcción emocional del sujeto durante el embarazo es compleja. Sin embargo, nos propusimos estudiar también si hay alguna relación entre la dinámica de concentración en sangre de las hormonas relacionadas en el embarazo y alguna medida del lenguaje.

En el experimento del ciclo menstrual encontramos una relación entre la intensidad del lenguaje (es decir, 1 - neutralidad) y el ciclo. Teniendo como mejor modelo explicativo que la regulación hormonal puede estar implicando los cambios en el len-

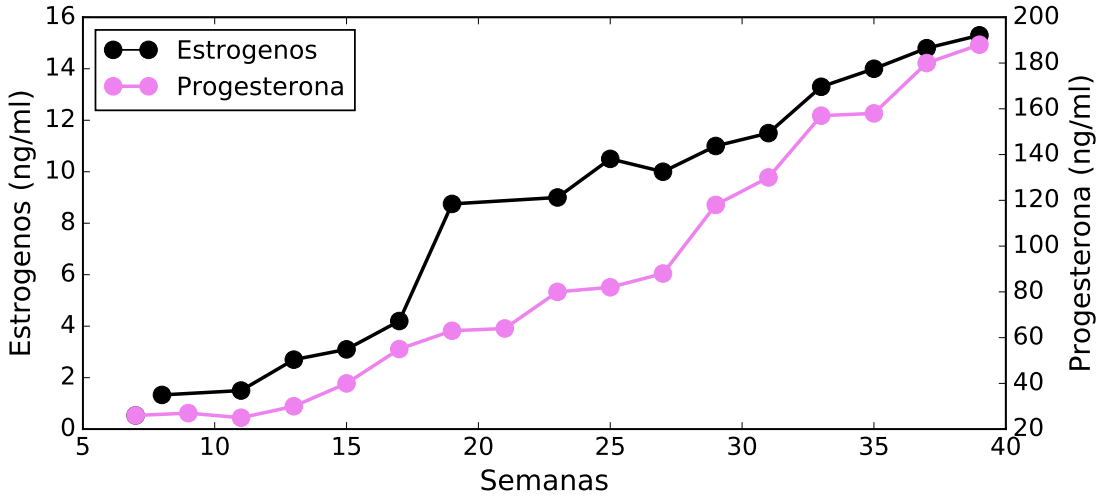

Fig. 5.6: Curvas poblacional media de concentración de estrógenos y progesterona [158]

guaje y en el estado de ánimo. A su vez, esencialmente los hormonas involucradas en el ciclo menstrual y durante el embarazo son progesterona y estrógenos, nos proponemos ahora medir si existe una relación entre la neutralidad (o 1 - intensidad) y alguna de la dinámica de concentración de moléculas en sangre para estas hormonas. El estudio de efectos cognitivos producido por la interacción farmacológica de estos dos grupos de hormonas fue estudiado ampliamente pero nunca desde el lenguaje y menos en un escenario de alta frecuencia de muestreo, sin embargo existen excelentes reportes sobre estado emocional y comportamiento sexual entre otros [179–185].

Para medir este efecto tomamos los datos de [158] donde los investigadores reportan la curva promedio de niveles normales de estrógenos y progesteronas para embarazos llevados a cabo exitosamente sin patologías ni en el feto ni en la madre, la Figura 5.6 muestra la curva tomando los valores del trabajo mencionado. En principio, la información de esta tabla muestra una alta correlación en la concentración de las dos hormonas en sangre durante el transcurso del embarazo (ver [158]) debido a esta relación la comparación la hicimos con la curva de estrógenos. En [158] los investigadores reportan niveles de concentración de estrógenos cada 2 semanas durante

el embarazo. Para poder comparar estos datos con los nuestros lo que hicimos fue: Para cada sujeto normalizamos como ya mencionamos dividiendo por la media por sujeto, esto nos permite obviar la magnitud y el rango por sujeto y poder agrupar. Luego, tomamos para cada sujeto todos los tweets producidos en las dos semanas del intervalo definido por [158]. Por lo que cada sujeto tiene reportado para cada dos semanas una valoración media de *neutralidad*. Luego, construimos una serie poblacional tomando la media entre sujetos para cada intervalo de dos semanas. Con esta nueva serie, que llamamos neutralidad poblacional cada dos semanas, calculamos la correlación de *pearson* entre esta y la serie de concentración de estrógenos en sangre de [158]. Los resultados muestran que existe una correlación positiva y significativamente aceptable ($\rho = -0{,}46, pval < 0{,}04672$). Es decir, los niveles de concentración de las dos hormonas correlacionan positivamente con la neutralidad del lenguaje. A medida que el embarazo avanza, y la concentración de las hormonas también, el lenguaje se vuelve más neutral. Haciendo un análisis análogo con las series de positividad y negatividad no encontramos correlación significativa.

La Figura 5.7 muestra las dos series graficadas, asumiendo crecimiento lineal en la concentración de hormonas entre los puntos de dos semanas reportados y aplicando un suavizado de ventana de 60 días para atrás para la serie de neutralidad. Donde se aprecia la correlación medida anteriormente.

En la bibliografía encontramos suficiente para sostener que existe una interacción de estrógenos, progesterona y diferentes mecanismo cognitivos. Sin embargo, el uso del lenguaje en un escenario de alta frecuencia no había sido abordado. Este acercamiento propone una ventana de posibilidades para formular nuevas preguntas. En los dos experimentos que hicimos intentamos abordar las dimensiones del lenguaje mas *transparentes* al estado de ánimo de los sujetos. Sin embargo quedan pendiente diferentes ejes del lenguaje por los que avanzar.

En ambos casos, resultó difícil evaluar y ligar efectivamente el efecto cognitivo

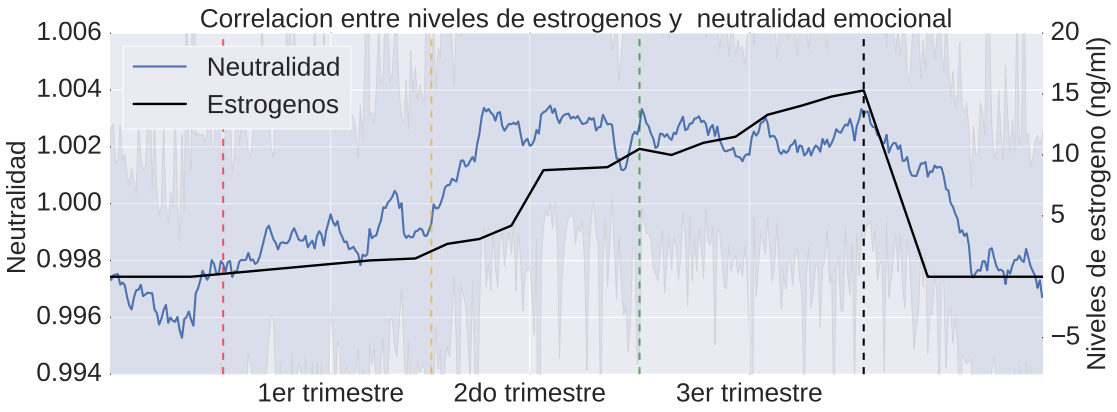

Fig. 5.7: Dinámica de la serie de neutralidad media con error estándar suavizada con ventana de 60 días para atrás y de la serie de niveles de estrógenos poblacional reportada en [158].

asociado a la interacción farmacológica debido fuertemente a que no propusimos un diseño experimental intervencionista sino que aprovechamos la situación natural de dos estados particulares de la mujer donde la dinámica de concentración de estas dos hormonas es radicalmente distinta. Hubiera sido extremadamente interesante contar con la fase del ciclo menstrual en que las mujeres se encontraban para el experimento de la Sección 5.1 pues nos hubiera permitido no solo estudiar que existe un efecto sino también si el efecto opera en función a cual hormona o si acaso surge de la relación de ambas. También, dada la diferencia en la magnitud de concentración en los dos experimentos (el ciclo menstrual opera con un orden de concentración hormonal menor que durante el embarazo) nos hubiera permitido evaluar la sensibilidad de la interacción farmacológica. Otra situación de experimento no intervencionista que sería interesante estudiar es el caso de la etapa menopáusica pues la dinámica de concentración hormonal en estos casos vuelve a ser diferente a los casos estudios. Inclusive existen tratamientos para mujeres en esa etapa que incluyen la intervención del paciente suministrando estas hormonas modificando la dinámica basal. Intentamos armar un conjunto de datos de esta población pero resulto muy difícil pues la

población etaria en Twitter esta sesgada hacia gente mas joven y a esta complejidad también se agrega la intrínseca alta diferencia que hay en el inicio de esta etapa en la población.

Teniendo en cuenta que en los dos experimentos vemos que la neutralidad (o la intensidad) del lenguaje se ve afectada tal vez podamos pensar que estas dos hormonas o en la regulación de las mismas podrían participar en un tratamiento psicofarmacológico para aplacamiento de rango dinámico de emociones. Si bien la interacción farmacológica de las hormonas es muy alta en diferentes partes del cuerpo los efectos adversos podrían ser comparados con los de otros psicofármacos ampliamente usados.

6. DISCUSIÓN

El lenguaje es estudiado de diferentes perspectivas por distintas disciplinas científicas (lingüística, antropología, psicología, etc). Creemos que hacerlo desde una perspectiva algorítmica y cuantitativa, contribuye a mejorar el entendimientos de mecanismos que explican la relación entre el cerebro y la mente. Con este fin, en esta tesis, ahondamos distintos casos de estudios de sujetos con estados mentales alterados. En todos los casos intentamos usar modelos computacionales simples y sobre todo, guiarnos por hipótesis previas de la distintas disciplinas (por ejemplo, el caso de estudio de pacientes esquizofrénicos donde modelamos la coherencia tras relevar la caracterización del cuadro en bibliografía psiquátrica). A nuestro entender, esto resulta en una estrategia acertada pues, en el contexto donde los modelos computacionales pueden ser tan complejos (hecho indiscutiblemente útil en ciertas tareas pragmáticas), los modelos simples proponen no solo resultados útiles sino descripciones entendibles de los fenómenos observados. Estas descripciones entendibles, promueven luego más preguntas y permiten sostener un dialogo entre los expertos del dominio de la fenomenología y aquellos que contribuyen desde la computación. Creemos fuertemente que este dialogo es necesario para avanzar con preguntas claras y no caer en un escenario de producción de ciencia estéril.

Teniendo en cuenta nuestro marco de trabajo a continuación presentamos una discusión de cada capitulo en particular donde planteamos las limitaciones de la experimentación y las direcciones para avanzar en el futuro.

En el Capítulo 3, *Estados mentales alterados por patológicas* estudiamos tres casos donde los sujetos presentaban estados mentales alteradas debido a patologías psiquiátricas. En el caso de estudio de pacientes esquizofrénicos 3.1 recurrimos a la bibliografía psiquiátrica para entender los fenotipos del lenguaje típicos de los

pacientes con esta patologías. Tras esta revisión, encontramos que la coherencia discursiva era una buena dimensión para estudiar y caracterizar a los sujetos, pues el DSM-V [100] resultaba elocuente al describir la deformación del lenguaje pero a su vez resultaba vaga, lo que permitía modelar y experimentar con métodos para capturar esta propiedad. El algoritmo que diseñamos para capturar esto (presentado en 2.1) fue testeado inicialmente en un entorno controlado donde estudiando como funcionaban las distintas mediciones en un experimento en el cual intercambiábamos el orden en en las frases. Este experimento presentó el comportamiento esperado en las distintas series de coherencia. Tras validar el algoritmo, estudiamos si las mediciones de coherencia en los sujetos esquizofrénicos era distinta en comparación con los sujetos control. Las primeras comparaciones con test estadísticos, no arrojaron diferencias significativas entre los grupos. Sin embargo al inspeccionar la disposición espacial de estas propiedades entre los grupos notamos que había información que separaba a los grupos. Probablemente habría que haber estudiado algún test estadístico en la interacción de mas de una variable de coherencia para encontrar diferencias significativas. En vez de avanzar en esta dirección, usamos una estrategia de aprendizaje supervisado en la cual, usando validación cruzada, estudiamos cuan bien podíamos aprender los patrones subyacentes a la coherencia que definían la fenomenología patológica del las características de un discurso control. Este experimento arrojó muy buenos resultados (*accuracy* mayor a 0.85), esto significó que la información de coherencia como propiedad del lenguaje era capturable por un mecanismo de inferencia de patrones, conformando un modelo el cual serviría para predecir nuevas muestras, no usadas en el entrenamiento, con una alta chance de hacerlo bien. Si bien el tamaño de la muestra no era grande, creemos que los mecanismos de mitigación de sobre-ajuste nos permiten entender que nuestro modelo puede generalizar a casos nuevos. Es importante destacar, como limitación del resultado, que harían falta experimentos donde se use otro disparador discursivo para entender si el efecto de alteración de coherencia discursiva solo se observa bajo este

disparador (relatar un sueño reciente) o, de lo contrario, es un fenómeno capturable en el discurso independientemente del disparador. Estos experimentos nos permitirían entender mejor si, acaso el relato de un recuerdo reciente, o la descripción de un ser querido generan un desarrollo del lenguaje donde podemos medir alteraciones de la coherencia o solo es exclusivo del disparador usado. Luego, entender en función a las limitaciones por qué los disparadores eventualmente condicionan la alteración del lenguaje o por el contrario, cuán universal es el cambio en el lenguaje independientemente al disparador.

También surgen dudas respecto a la longitud necesaria del discurso o inclusive si el método es susceptible a textos referidos a un tema particular. Es decir, para entender las limitaciones del modelo corresponde hacer mas experimentos. A su vez, creemos que, si bien nuestro algoritmo de coherencia captura información relevante, el éxito del mismo no es invariante al disparador del discurso y sin duda corresponde avanzar en esta dirección.

Tras estudiar la muestra de pacientes ya diagnosticados como esquizofrénicos, nos avocamos al estudio de sujetos de alto riesgo de conversión a la enfermedad. En la Sección 3.2 presentamos los resultados de estudiar como nuestro algoritmo de coherencia se comportaba en una muestra de sujetos sanos clasificados de alto riesgo. En el análisis de este caso de estudio descubrimos que aquellos sujetos que desarrollaron psicosis ya tenían un nivel de coherencia alterado en el momento en que la clínica médica consideraba que eran sujetos sanos. Este resultado propone que esta metodología podría ser útil no solo para replicar de una manera razonable los resultados de la clínica sino para aumentar la capacidad de diagnostico precoz. En un esquema de aprendizaje automático encontramos que podíamos armar un modelo que predecía bien con niveles de *rocauc* mayores a 0.84. La fuerte limitación de este experimento fue el tamaño de la muestra, donde contábamos con 34 sujetos que solo 5 convirtieron. El problema de este tipo de experimentos es la dificultad por parte de los médicos en llevar a adelante un protocolo donde el seguimiento de

los pacientes es a largo plazo. Como trabajo futuro en esta misma dirección estamos realizando nuevas mediciones comparables en una nueva cohorte estudiando cuán generalizable es el modelo ya entrenado en la nueva muestra. A su vez, estamos diseñando nuevos protocolos basados en hipótesis de la fenomenología para intentar capturar nuevas deformaciones del lenguaje.

Luego de trabajar con estos dos casos nos centramos en estudiar pacientes con trastorno bipolar (Sección 3.3). Inspirados nuevamente por la descripción bibliográfica de los fenotipos mentales encontramos que estudiar la emocionalidad podía ser un buen vehículo para avanzar. Teniendo en cuenta esto, medimos intensidad emocional del discurso cuantificando la tasa de palabras emotivas en el discurso de los sujetos (es decir, aquellas que fueran positivas o negativas). Haciéndolo descubrimos que la media y el desvió estándar de la intensidad de los sujetos por frases aportaba información relevante para caracterizar a los sujetos. Usando un test de hipótesis identificamos que la intensidad en los dos grupos era distinta, mayor para el grupo patológico. También, recurrimos nuevamente a un método de aprendizaje automático con validación cruzada. Con este prueba pudimos clasificar con 75 % de *accuracy* simplemente con los dos atributos mencionados anteriormente. Este experimento describe que mirar en un estado estacional a los pacientes en cuanto a su intensidad emocional ya aporta para caracterizarlos. Dada la dinámica de transición de estados emocionales de la patología lo interesante seria poder estudiar los cambios y no una solo visión estacionaria. Armando un modelo con esta característica creemos que podría estudiarse la patología de una manera mas a fin a la descripción de la fenomenología. Dada la complejidad de tener una muestra así queda como trabajo futuro armar un buen conjunto de datos para este tipo de análisis.

En los tres casos de estudios anteriores, intentamos mediante metodología computacional, predecir el diagnóstico del sujeto. Sin embargo, modelar los estados mentales a partir del lenguaje podrían describir automáticamente distintas propiedades del sujeto más allá de la condición de diagnostico. Otras propiedades podrían ser, por

ejemplo, intentar regresionar diferentes escalas de la literatura clásica psiquiátrica. En la Sección 3.4 estudiamos otra propiedad diferente. Usamos modelos computacionales para predecir el éxito de un tratamiento psicofarmacológico antes de aplicarlo en los sujetos. Las ventajas de contar con un test de susceptibilidad son relevantes pues acortan el tiempo en que los médicos dan con el tratamiento correcto para un determinado sujeto. En este caso de estudio, medimos usando la emocionalidad nuevamente, si los sujetos con depresión resistente a tratamientos clásicos responderían positivamente al nuevo tratamiento o no. Para eso armamos un modelo que usando la positividad y negatividad y el clasificador Gaussian Navie Bayes, logramos predecir con 75 % de precisión aquellos sujetos que responderían al tratamiento (con 85 % de *accuracy*). Si bien la cantidad de sujetos fue pequeña y el tratamiento solo uno en particular, creemos que esta perspectiva de estudio, es decir intentar predecir otra propiedad del sujeto es novedosa pues en la bibliografía se suele poner foco en la predicción del diagnostico.

En el Capítulo 4, *Estados mentales alterados por intoxicaciones farmacológicas* estudiamos dos casos de alteraciones mentales producto de interacciones farmacológicas controladas. En principio, en la Sección 4.1, nos dedicamos a replicar los resultados de un experimento previo donde encontrábamos como la ingesta de MDMA produce alteraciones homogéneas en el discurso. En este nuevo experimento, tomamos una nueva cohorte de sujetos y repetimos la experimentación. Los resultados mostraron no solo que pudimos replicar de una manera similar los resultados previos, sino que lo hicimos en un contexto experimental levemente diferente. Los discursos, en este nueva configuración, eran notablemente mas cortos. Esta característica resulta relevante en este tipo de análisis, pues la metodología de abstracción semántica suele tener una sensibilidad diferente en función a la cantidad y la calidad de texto. Inicialmente hipotetisamos que deberíamos cambiar el tipo de análisis pero los resultados mostraron que esta cantidad de palabras fue suficiente. Tras

comprobar que podíamos replicar los resultados en la muestra levemente diferente, corroboramos que también podíamos crear un modelo con la nueva muestra y predecir en la anterior. Este experimento mostró exitosamente que el modelo fue más robusto de lo esperado. Pues creíamos que no conseguiríamos generalizar a los nuevos datos. Sin embargo, los valores alcanzados en el experimento donde integramos las dos cohortes fueron tan buenos como los resultados de los modelos intra-cohorte.

En la Sección 4.2 estudiamos superficialmente los efectos del LSD en la emotividad de los sujetos. En este simple análisis encontramos un resultado muy claro que describió que el discurso bajo el efecto de LSD se movió hacia una emocionalidad negativa. Sin embargo, este resultado no puede interpretarse como que el LSD promueve sentimiento negativos pues se encuentra muy contextualizado a la situación de los sujetos. El disparador discursivo fue hablar sobre la experiencia de los sujetos durante la realización de un estudio médico. Probablemente bajo otro disparador el LSD opere de otra manera y lo que estuviéramos viendo como efecto en realidad fuera otro, tal vez el fármaco potencia los sentimientos en diferentes direcciones, operando en la amplitud y no en la valencia. Esta, es solo una hipótesis posible, para resolver esto, como trabajo futuro , esperamos seguir colaborando con los investigadores clínicos buscando más situaciones donde esta droga promueva relatos con diferentes disparadores, como por ejemplo, relatar un sueño reciente o un recuerdo feliz.

En el Capítulo 5, *Estados mentales alterados por cambios endocrinos* nos preguntamos si podíamos encontrar cambios en la mente de los sujetos, mediante el lenguaje, producto de interacciones hormonales. Los experimentos para contestar esta pregunta podrían ser diversos. En particular, por una cuestión de practicidad decidimos no realizar un protocolo intervencionista en sujetos para evaluar los cambios, sino que aprovechamos dos situaciones naturales de la especie, donde la descripción de la dinámica de los hormonas esta ampliamente reportada. A su vez, tomamos

lenguaje de redes sociales, de modo que la medición experimental no operase sobre los sujetos sesgándolos de ninguna manera.

En primer lugar, tomamos como escenario el Ciclo Menstrual (Sección 5.1). La dinámica de concentración de hormonas en sangre del ciclo menstrual en humanos es conocida y bien descripta. A su vez, la bibliografía cuenta con mucha evidencia presentada sobre la interacción de las hormonas involucradas en el CM y los diversos sistemas, en particular como interactúa con diferentes mecanismos cognitivos. Sin embargo, en la revisión bibliografía no encontramos ningún trabajo sistemático que use el lenguaje como *proxy* a la mente de la manera en que nosotros lo presentamos. Es decir, en un escenario de muestreo de alta frecuencia (varios mensajes diarios). En nuestro experimentamos, con las limitaciones presentadas en el trabajo donde no contábamos con la fase del ciclo por sujeto, encontramos que la emocionalidad de los sujetos femeninos es diferente y más intensa a la de los sujetos masculinos cuando se estudia el comportamiento de ciclos emocionales de 28 días. La explicación mas plausible que encontramos, es que el CM estaría operando sobre la emocionalidad de los sujetos femeninos. Nuestro experimento cuenta con una gran limitación, no estamos midiendo cómo opera el CM sobre el lenguaje, sino solo que lo modifica. Eso se debe particularmente a dos factores. Primero, no tenemos información real de concentración hormonal en sangre de los sujetos, solo suponemos que los sujetos cuentan con una dinámica poblacional. Esto no es una gran suposición pues nuestro resultado también describe une efecto sobre la población. Sin embargo el problema de esto, es que nos perdemos los sujetos que tienen ciclos no medios, es decir aquellas sujetos mujeres que tienen ciclos distintos de 28 días o que la concentración hormonal no es exactamente la promedio. Tampoco, contamos con la información sobre interacciones farmacológicas, como por ejemplo el uso de un método anticonceptivo hormonal. Este último factor podría ser importante pues los métodos anticonceptivos hormonales orales operan, la mayoría, sobre las mismas hormonas que regulan el ciclo menstrual cambiando la dinámica. Teniendo en

cuenta esto, podríamos estar mezclando dos sub-poblaciones, aquellas mujeres que usan anticonceptivos hormonales orales y aquellas que no. Sin embargo, el resultado observado debería ser similar, pues estos métodos también operan sobre un ciclo de 28 días. A su vez, lo hacen bajo las mismas hormonas, por lo que mediante nuestra metodología donde usamos la función de autocorrelación, probablemente también veríamos el efecto similar en ambas sub-poblaciones. Planteadas estas dos limitaciones entendemos que un experimento que permita presentar conclusiones mas fuertes es necesario. En este experimento deberíamos poder tomar mediciones de la concentración hormonal, controlar si usan o no métodos anticonceptivos hormonales y por supuesto tener la fase del ciclo menstrual. De esta manera podríamos entender mejor el efecto hormonal en el lenguaje.

Como no contamos con un experimento mas controlado, decidimos seguir estudiando el efecto de las hormonas en otro escenario conocido natural. Por eso, en la Sección 5.2, estudiamos el efecto en el lenguaje en mujeres embarazadas. Este caso de estudio nos permitió continuar el anterior pues, si bien el embarazo es un estado complejo desde la fisiología hasta la construcción mental del sujeto gestador, en la regulación hormonal del mismo operan esencialmente las mismas hormonas (y otras más). En la revisión bibliográfica encontramos que no habían trabajos donde el lenguaje fuera estudiando sistemáticamente en un escenario de muestro de alta frecuencia como en el caso del ciclo menstrual. Por eso, decidimos replicar las condiciones experimentales del ciclo menstrual y usamos una red social para estudiar el lenguaje de los sujetos. A su vez, como continuación al trabajo anterior, decidimos usar la emocionalidad del lenguaje como característica del lenguaje a medir, pues si entendíamos que las hormonas modulaban esta propiedad, también deberíamos poder medirlo en este nuevo escenario. Para eso, medimos la emocionalidad, antes, durante y después del embrazo en mujeres y encontramos que existía una correlación fuerte entre el tiempo de gestación y la valoración de las propiedades del lenguaje usadas (positivadad, negatividad y neutralidad). Esta nueva información describe

un estado emocional claro poblacionalmente. Luego de estudiar la relación entre las propiedades del lenguaje y el tiempo gestacional, nos preguntamos si, suponiendo una curva de concentración hormonal poblacional media, podíamos encontrar alguna relación directa entre las propiedades discursivas y la concentración hormonal. Tomando la asunción fisiológica poblacional, encontramos que a mayor concentración de dos hormonas aumentaba la neutralidad del lenguaje. Esto es interesante, pues en el ciclo menstrual donde también operan estas dos (a concentraciones muy distintas), el lenguaje era modificado en la misma propiedad (en el caso del CM hablamos de intensidad, en este caso de neutralidad que fue definida como 1 - intensidad). Esta evidencia resulto tremendamente alentadora pues encontramos en dos experimentos separados un efecto cognitivo del lenguaje similar. Desafortunadamente la dinámica de estas dos hormonas durante el embarazo es muy similar, y correlacionan entre si, por lo que pudimos explicar el efecto en la neutralidad como función de una de ellas, o incluso entender si el efecto ocurre en la interacción. Para entender esto intentamos estudiar otro escenario donde estas curvas nuevamente se separan en su dinámica (más allá del ciclo menstrual), y es en el caso de las mujeres que ya no son fértiles. Sin embargo, intentamos armar una muestra en la red social de esta población pero no pudimos. Teniendo en cuenta el resultado encontrado creemos que el experimento pendiente del CM va a aportar información relevante y una descripción mas precisa sobre como operan las hormonas en el lenguaje.

Habiendo discutido e interpretado los resultados de los casos de estudio, creemos que durante la tesis presentamos suficiente evidencia relevante y novedosa para cumplir con el objetivo general y los objetivos particulares establecidos en la introducción. A su vez, planteamos las direcciones que nos parecen correctas para continuar, definiendo así, una linea de investigación interdisciplinaria pero con una fuerte presencia de modelos de inteligencia artificial. Creemos que es necesario avanzar y explorar esta linea de investigación pues el campo medico en el futuro se nutrirá

de diferentes herramientas de procesamiento del lenguaje natural e inteligencia artificial y necesitamos instanciar todas estos métodos y herramientas a las distintas coyunturas e idiosincrasias de nuestro país para sostener lo que creemos que en un futuro serán nuevos estándares de diagnostico médico.

A continuación listamos las publicaciones que realizamos en el contexto de esta tesis.

- Gallino, L., Carrillo, F. and Cecchi, G.A., 2019. Differential 28-days cyclic modulation of affective intensity in female and male participants via social media. Frontiers in integrative neuroscience, 13.

- Carrillo, F., Sigman, M., Slezak, D.F., Ashton, P., Fitzgerald, L., Stroud, J., Nutt, D.J. and Carhart-Harris, R.L., 2018. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. Journal of affective disorders, 230, pp.84-86.

- Corcoran, C.M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D.C., Bearden, C.E. and Cecchi, G.A., 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry, 17(1), pp.67-75.

- Carrillo, F., 2017. Computational characterization of mental states: A natural language processing approach. In Proceedings of ACL 2017, Student Research Workshop (pp. 1-3).

- Mota, N.B., Carrillo, F., Slezak, D.F., Copelli, M. and Ribeiro, S., 2016, November. Characterization of the relationship between semantic and structural language features in psychiatric diagnosis. In 2016 50th Asilomar Conference on Signals, Systems and Computers (pp. 836-838). IEEE.

- Carrillo, F., Mota, N., Copelli, M., Ribeiro, S., Sigman, M., Cecchi, G. and

Slezak, D.F., 2016. Emotional intensity analysis in Bipolar subjects. arXiv preprint arXiv:1606.02231.

- Bedi, G., Carrillo, F., Cecchi, G.A., Slezak, D.F., Sigman, M., Mota, N.B., Ribeiro, S., Javitt, D.C., Copelli, M. and Corcoran, C.M., 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophrenia, 1, p.15030.

- Carrillo, F., Cecchi, G.A., Sigman, M. and Slezak, D.F., 2015. Fast distributed dynamics of semantic networks via social media. Computational intelligence and neuroscience, 2015, p.50.

- Bedi, G., Cecchi, G.A., Slezak, D.F., Carrillo, F., Sigman, M. and De Wit, H., 2014. A window into the intoxicated mind? Speech as an index of psychoactive drug effects. Neuropsychopharmacology, 39(10), p.2340.

- Carrillo, F., Mota, N., Copelli, M., Ribeiro, S., Sigman, M., Cecchi, G. and Slezak, D.F., 2013. Automated speech analysis for psychosis evaluation. In Machine Learning and Interpretation in Neuroimaging (pp. 31-39). Springer, Cham.

Bibliografía

[1] Steve Alsop. *Beyond Cartesian Dualism: Encountering Affect in the Teaching and Learning of Science.*, volume 29. Springer Science & Business Media, 2005.

[2] Círculo de Viena. La concepción científica del mundo: El círculo de viena. *REDES, Revista de Estudios sobre la Ciencia y la Tecnología*, 9:103–50, 1987.

[3] Otto Neurath. Wissenschaftliche weltauffassung: Der wiener kreis. In *Empiricism and sociology*, pages 299–318. Springer, 1973.

[4] Olaf Sporns. The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1):109–125, 2011.

[5] Gyorgy Buzsaki. *Rhythms of the Brain.* Oxford University Press, 2006.

[6] SUSUMU Hagiwara and Leukocytes Byerly. Calcium channel. *Annual review of neuroscience*, 4(1):69–125, 1981.

[7] Anthony N Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95(1):1–19, 2006.

[8] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[9] Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1):018102, 2005.

[10] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode

hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.

[11] Stanislas Dehaene. *The number sense: How the mind creates mathematics*. OUP USA, 2011.

[12] Stanislas Dehaene, Serge Bossini, and Pascal Giraux. The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3):371, 1993.

[13] Stanislas Dehaene and Lionel Naccache. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1):1–37, 2001.

[14] A Richard Green, Annis O Mechan, J Martin Elliott, Esther O'Shea, and M Isabel Colado. The pharmacology and clinical pharmacology of 3, 4-methylenedioxymethamphetamine (mdma,"ecstasy"). *Pharmacological reviews*, 55(3):463–508, 2003.

[15] Robert A Lyon, Richard A Glennon, and Milt Titeler. 3, 4-methylenedioxymethamphetamine (mdma): stereoselective interactions at brain 5-ht 1 and 5-ht 2 receptors. *Psychopharmacology*, 88(4):525–526, 1986.

[16] George Battaglia, Brian P Brooks, Chaiyaporn Kulsakdinun, and Errol B De Souza. Pharmacologic profile of mdma (3, 4-methylenedioxymethamphetamine) at various brain recognition sites. *European journal of pharmacology*, 149(1):159–163, 1988.

[17] AC Parrott, E Sisk, and JJD Turner. Psychobiological problems in heavy 'ecstasy'(mdma) polydrug users. *Drug and alcohol dependence*, 60(1):105–110, 2000.

[18] Michelle Wareing, John E Fisk, and Philip N Murphy. Working memory deficits in current and previous users of mdma ('ecstasy'). *British journal of psychology*, 91(2):181–188, 2000.

[19] Andy C Parrott. Human psychopharmacology of ecstasy (mdma): a review of 15 years of empirical research. *Human Psychopharmacology: Clinical and Experimental*, 16(8):557–577, 2001.

[20] Richard S Cohen. Subjective reports on the effects of the mdma ('ecstasy') experience in humans. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 19(7):1137–1145, 1995.

[21] R Quian Quiroga, Leila Reddy, Gabriel Kreiman, Christof Koch, and Itzhak Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, 2005.

[22] A.M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[23] Irving L Janis and Leon Mann. *Decision making: A psychological analysis of conflict, choice, and commitment.* free press, 1977.

[24] Max H Bazerman and Don A Moore. *Judgment in managerial decision making.* Wiley, 2008.

[25] Scott Plous. *The psychology of judgment and decision making.* Mcgraw-Hill Book Company, 1993.

[26] Eilon Vaadia. Cognitive neuroscience: Learning how the brain learns. *Nature*, 405(6786):523–525, 2000.

[27] Stanislas Dehaene and Jean-Pierre Changeux. Reward-dependent learning in neuronal networks for planning and decision making. *Progress in brain research*, 126:217–229, 2000.

[28] Lisa Holper, Andrea P Goldin, Diego E Shalóm, Antonio M Battro, Martin Wolf, and Mariano Sigman. The teaching and the learning brain: A cortical hemodynamic marker of teacher–student interactions in the socratic dialog. *International Journal of Educational Research*, 59:1–10, 2013.

[29] Matías Lopez-Rosenfeld, Andrea Paula Goldin, Sebastián Lipina, Mariano Sigman, and Diego Fernandez Slezak. Mate marote: A flexible automated framework for large-scale educational interventions. *Computers & Education*, 68:307–313, 2013.

[30] Antonio M Battro, Cecilia I Calero, Andrea P Goldin, Lisa Holper, Laura Pezzatti, Diego E Shalóm, and Mariano Sigman. The cognitive neuroscience of the teacher–student interaction. *Mind, Brain, and Education*, 7(3):177–181, 2013.

[31] CI Calero, A Zylberberg, J Ais, M Semelman, and M Sigman. Young children are natural pedagogues. *Cognitive Development*, 35:65–78, 2015.

[32] Sidney Strauss, Cecilia I Calero, and Mariano Sigman. Teaching, naturally. *Trends in neuroscience and education*, 3(2):38–43, 2014.

[33] Marc Jeannerod. *The cognitive neuroscience of action*, volume 1997. Blackwell Oxford, 1997.

[34] Lesley K Fellows. The cognitive neuroscience of human decision making: a review and conceptual framework. *Behavioral and cognitive neuroscience reviews*, 3(3):159–172, 2004.

[35] Eric W Stein. Organization memory: Review of concepts and recommendations for management. *International journal of information management*, 15(1):17–32, 1995.

[36] Lynn Hasher and Rose T Zacks. Working memory, comprehension, and aging: A review and a new view. *Psychology of learning and motivation*, 22:193–225, 1988.

[37] Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience*, 5(1):42, 2004.

[38] Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, pages 1632–1634, 1995.

[39] Giulio Tononi and Gerald M Edelman. Consciousness and complexity. *science*, 282(5395):1846–1851, 1998.

[40] Giulio Tononi, Olaf Sporns, and Gerald M Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, 1994.

[41] Marc D Hauser, Noam Chomsky, and W Tecumseh Fitch. The faculty of language: what is it, who has it, and how did it evolve? *science*, 298(5598):1569–1579, 2002.

[42] Big Data. for better or worse: 90 % of world's data generated over last two years. *SCIENCE DAILY, May*, 22, 2013.

[43] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004.

[44] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 294–311. Springer, 2003.

[45] Luis Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.

[46] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.

[47] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[48] Luis Von Ahn. Human computation. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 1–2. IEEE, 2008.

[49] Thomas N Theis and H-S Philip Wong. The end of moore's law: A new beginning for information technology. *Computing in Science & Engineering*, 19(2):41–50, 2017.

[50] Yunhe Pan. Heading toward artificial intelligence 2.0. *Engineering*, 2(4):409–413, 2016.

[51] Charles Beattie, Joel Z Leibo, Denis Teplyashin, Tom Ward, Marcus Wainwright, Heinrich Küttler, Andrew Lefrancq, Simon Green, Víctor Valdés, Amir Sadik, et al. Deepmind lab. *arXiv preprint arXiv:1612.03801*, 2016.

[52] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[53] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[54] Gillinder Bedi, Guillermo A Cecchi, Diego F Slezak, Facundo Carrillo, Mariano Sigman, and Harriet de Wit. A window into the intoxicated mind?

speech as an index of psychoactive drug effects. *Neuropsychopharmacology*, 39(10):2340–2348, 2014.

[55] Brita Elvevåg, Peter W Foltz, Daniel R Weinberger, and Terry E Goldberg. Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia research*, 93(1):304–316, 2007.

[56] Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press, 2014.

[57] Danielle S McNamara, Eileen Kintsch, Nancy Butler Songer, and Walter Kintsch. Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, 14(1):1–43, 1996.

[58] Arthur C Graesser, Danielle S McNamara, Max M Louwerse, and Zhiqiang Cai. Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2):193–202, 2004.

[59] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.

[60] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[61] Edgar Altszyler, Mariano Sigman, and Diego Fernández Slezak. Comparative study of lsa vs word2vec embeddings in small corpora: a case study in dreams database. *arXiv preprint arXiv:1610.01520*, 2016.

[62] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in

continuous space word representations. In *Hlt-naacl*, volume 13, pages 746–751, 2013.

[63] Gillinder Bedi, Facundo Carrillo, Guillermo A Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B Mota, Sidarta Ribeiro, Daniel C Javitt, Mauro Copelli, and Cheryl M Corcoran. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia*, 1:15030, 2015.

[64] Rudi L Cilibrasi and Paul MB Vitanyi. The google similarity distance. *IEEE Transactions on knowledge and data engineering*, 19(3), 2007.

[65] Dekang Lin et al. An information-theoretic definition of similarity. In *Icml*, number 1998 in 1, pages 296–304, 1998.

[66] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.

[67] Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[68] Facundo Carrillo, Guillermo A Cecchi, Mariano Sigman, and Diego Fernández Slezak. Fast distributed dynamics of semantic networks via social media. *Computational intelligence and neuroscience*, 2015:50, 2015.

[69] Bo Pang, Lillian Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.

[70] Cícero Nogueira Dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69–78, 2014.

[71] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 513–520, 2011.

[72] Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 142–150. Association for Computational Linguistics, 2011.

[73] Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, pages 60–68. Association for Computational Linguistics, 2010.

[74] Cristina Bosco, Viviana Patti, and Andrea Bolioli. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems*, 28(2):55–63, 2013.

[75] Cynthia M Whissell and Michael RJ Dewson. A dictionary of affect in language: Iii. analysis of two biblical and two secular passages. *Perceptual and Motor Skills*, 62(1):127–132, 1986.

[76] CM Whissell. The dictionary of affect in language. emotion: theory, research and experience 4. the measurement of emotions, 1989.

[77] Cynthia Whissell. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological reports*, 105(2):509–521, 2009.

[78] Kevin Sweeney and Cynthia Whissell. A dictionary of affect in language: I. establishment and preliminary validation. *Perceptual and motor skills*, 59(3):695–698, 1984.

[79] Matıas G Dell'Amerlina Rıos and Agustın Gravano. Spanish dal: A spanish dictionary of affect in language. *WASSA 2013*, page 21, 2013.

[80] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

[81] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[82] J. Kupiec. Robust part-of-speech tagging using a hidden markov model. *Computer Speech & Language*, 6(3):225–242, 1992.

[83] Helmut Schmid. Part-of-speech tagging with neural networks. In *Proceedings of the 15th conference on Computational linguistics-Volume 1*, pages 172–176. Association for Computational Linguistics, 1994.

[84] Steven Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006.

[85] N.B. Mota, N.A.P. Vasconcelos, N. Lemos, A.C. Pieretti, O. Kinouchi, G.A. Cecchi, M. Copelli, and S. Ribeiro. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PloS one*, 7(4):e34928, 2012.

[86] Facundo Carrillo, Natalia Mota, Mauro Copelli, Sidarta Ribeiro, Mariano Sigman, Guillermo Cecchi, and Diego Fernandez Slezak. Automated speech analysis for psychosis evaluation. In *International Workshop on Machine Learning and Interpretation in Neuroimaging*, pages 31–39. Springer International Publishing, 2014.

[87] E. Loper and S. Bird. NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching*

natural language processing and computational linguistics-Volume 1, pages 63–70. Association for Computational Linguistics, 2002.

[88] Martin F Porter. Snowball: A language for stemming algorithms, 2001.

[89] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2001.

[90] R.E. Tarjan. Applications of path compression on balanced trees. *Journal of the ACM*, 26(4):690–715, 1979.

[91] Tom M Mitchell. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877, 1997.

[92] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[93] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):25, 2007.

[94] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[95] George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

[96] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[97] Richard A Armstrong. When to use the bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508, 2014.

[98] Karen G Scheps, Liliana Francipane, Abigail Nash, Gloria E Cerrone, Silvia B Copelli, and Viviana Varela. Bases moleculares de alfa-talasemia en la argentina. *Medicina (Buenos Aires)*, 75(2):81–86, 2015.

[99] HH Kazazian Jr. The thalassemia syndromes: molecular basis and prenatal diagnosis in 1990. In *Seminars in hematology*, volume 27, pages 209–228, 1990.

[100] Fifth Edition, American Psychiatric Association, et al. *Diagnostic and statistical manual of mental disorders*. Washington, American Psychological Association, 1994.

[101] Michael B First, Robert L Spitzer, Miriam Gibbon, Janet BW Williams, et al. Structured clinical interview for dsm-iv axis i disorders. *New York: New York State Psychiatric Institute*, 1995.

[102] NB Mota, F Carrillo, DF Slezak, M Copelli, and S Ribeiro. Characterization of the relationship between semantic and structural language features in psychiatric diagnosis. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on*, pages 836–838. IEEE, 2016.

[103] Tandy J Miller, Thomas H McGlashan, Joanna L Rosen, Kristen Cadenhead, Joseph Ventura, William McFarlane, Diana O Perkins, Godfrey D Pearlson, and Scott W Woods. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophrenia bulletin*, 29(4):703, 2003.

[104] Stanley R Kay, Abraham Flszbein, and Lewis A Opfer. The positive and

negative syndrome scale (panss) for schizophrenia. *Schizophrenia bulletin*, 13(2):261, 1987.

[105] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[106] Cheryl M Corcoran, Facundo Carrillo, Diego Fernández-Slezak, Gillinder Bedi, Casimir Klim, Daniel C Javitt, Carrie E Bearden, and Guillermo A Cecchi. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, 17(1):67–75, 2018.

[107] Kathleen R Merikangas, Robert Jin, Jian-Ping He, Ronald C Kessler, Sing Lee, Nancy A Sampson, Maria Carmen Viana, Laura Helena Andrade, Chiyi Hu, Elie G Karam, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of general psychiatry*, 68(3):241–251, 2011.

[108] Anita Weeke and Michael Væth. Excess mortality of bipolar and unipolar manic-depressive patients. *Journal of affective disorders*, 11(3):227–234, 1986.

[109] Urban Ösby, Lena Brandt, Nestor Correia, Anders Ekbom, and Pär Sparén. Excess mortality in bipolar and unipolar disorder in sweden. *Archives of general psychiatry*, 58(9):844–850, 2001.

[110] Shaun M Purcell, Naomi R Wray, Jennifer L Stone, Peter M Visscher, Michael C O'donovan, Patrick F Sullivan, Pamela Sklar, Douglas M Ruderfer, Andrew McQuillin, Derek W Morris, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256):748–752, 2009.

[111] Hans-Jürgen Möller. Bipolar disorder and schizophrenia: distinct illnesses or a continuum? *The Journal of clinical psychiatry*, 64:23–7, 2002.

[112] SR Platman, R Plutchik, RR Fieve, and WG Lawlor. Emotion profiles associated with mania and depression. *Archives of general Psychiatry*, 20(2):210–214, 1969.

[113] Sheri L Johnson, June Gruber, and Lori R Eisner. Emotion and bipolar disorder. *APA*, 2007.

[114] June Gruber, Sheri L Johnson, Christopher Oveis, and Dacher Keltner. Risk for mania and positive emotional responding: too much of a good thing? *Emotion*, 8(1):23, 2008.

[115] Stephen M Strakowski, Caleb M Adler, Jorge Almeida, Lori L Altshuler, Hilary P Blumberg, Kiki D Chang, Melissa P DelBello, Sophia Frangou, Andrew McIntosh, Mary L Phillips, et al. The functional neuroanatomy of bipolar disorder: a consensus model. *Bipolar disorders*, 14(4):313–325, 2012.

[116] Facundo Carrillo, Natalia Mota, Mauro Copelli, Sidarta Ribeiro, Mariano Sigman, Guillermo Cecchi, and Diego Fernandez Slezak. Emotional intensity analysis in bipolar subjects. *arXiv preprint arXiv:1606.02231*, 2016.

[117] Robin L Carhart-Harris, Mark Bolstridge, James Rucker, Camilla MJ Day, David Erritzoe, Mendel Kaelen, Michael Bloomfield, James A Rickard, Ben Forbes, Amanda Feilding, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *The Lancet Psychiatry*, 3(7):619–627, 2016.

[118] Robin L Carhart-Harris and Guy M Goodwin. The therapeutic potential of psychedelic drugs: Past, present, and future. *Neuropsychopharmacology*, 2017.

[119] Nick Craddock and Laurence Mynors-Wallis. Psychiatric diagnosis: impersonal, imperfect and important, 2014.

[120] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. *ICWSM*, 13:1–10, 2013.

[121] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 47–56. ACM, 2013.

[122] Adam Mourad Chekroud, Ryan Joseph Zotti, Zarrar Shehzad, Ralitza Gueorguieva, Marcia K Johnson, Madhukar H Trivedi, Tyrone D Cannon, John Harrison Krystal, and Philip Robert Corlett. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, 3(3):243–250, 2016.

[123] John P Pestian, Pawel Matykiewicz, and Jacqueline Grupp-Phelan. Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 96–97. Association for Computational Linguistics, 2008.

[124] Facundo Carrillo, Mariano Sigman, Diego Fernández Slezak, Philip Ashton, Lily Fitzgerald, Jack Stroud, David J Nutt, and Robin L Carhart-Harris. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. *Journal of affective disorders*, 230:84–86, 2018.

[125] Facundo Carrillo. Caracterización automática del lenguaje natural en sujetos con alteraciones mentales. *Tesis de Licenciatura, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*, 2012.

[126] Bernhard Scholkopf, Kah-Kay Sung, Christopher JC Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio, and Vladimir Vapnik. Comparing support

vector machines with gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, 45(11):2758–2765, 1997.

[127] Charles Grob. *Psychiatric Research with Hallucinogens-what Have We Learned?* VWB, Verlag für Wissenschaft und Bildung, 1996.

[128] FS Abuzzahab and BJ Anderson. A review of lsd treatment in alcoholism. *International pharmacopsychiatry*, 6:223–235, 1971.

[129] Enzo Tagliazucchi, Leor Roseman, Mendel Kaelen, Csaba Orban, Suresh D Muthukumaraswamy, Kevin Murphy, Helmut Laufs, Robert Leech, John McGonigle, Nicolas Crossley, et al. Increased global functional connectivity correlates with lsd-induced ego dissolution. *Current Biology*, 26(8):1043–1050, 2016.

[130] Robin L Carhart-Harris, Suresh Muthukumaraswamy, Leor Roseman, Mendel Kaelen, Wouter Droog, Kevin Murphy, Enzo Tagliazucchi, Eduardo E Schenberg, Timothy Nest, Csaba Orban, et al. Neural correlates of the lsd experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences*, 113(17):4853–4858, 2016.

[131] Leonard Chiazze, Franklin T Brayer, John J Macisco, Margaret P Parker, and Benedict J Duffy. The length and variability of the human menstrual cycle. *Jama*, 203(6):377–380, 1968.

[132] ERNST Knobil. The neuroendocrine control of the menstrual cycle. *Recent progress in hormone research*, 36:53–88, 1980.

[133] Wikimedia Commons. File:menstrualcycle2 es.svg — wikimedia commons, the free media repository, 2016.

[134] Dee Unglaub Silverthorn, William C Ober, Claire W Garrison, Andrew C

Silverthorn, and Bruce R Johnson. *Human physiology: an integrated approach*. Pearson/Benjamin Cummings San Francisco, CA, USA:, 2009.

[135] L Dye and JE Blundell. Menstrual cycle and appetite control: implications for weight regulation. *Human reproduction*, 12(6):1142–1151, 1997.

[136] Xanne AK Janse de Jonge. Effects of the menstrual cycle on exercise performance. *Sports Medicine*, 33(11):833–851, 2003.

[137] JAMES M Pivarnik, CARLOS J Marichal, THOMAS Spillman, and JR Morrow. Menstrual cycle phase affects temperature regulation during endurance exercise. *Journal of Applied Physiology*, 72(2):543–548, 1992.

[138] Albert Postma, Joke Winkel, Adriaan Tuiten, and Jack van Honk. Sex differences and menstrual cycle effects in human spatial memory. *Psychoneuroendocrinology*, 24(2):175–192, 1999.

[139] Elizabeth Hampson. Variations in sex-related cognitive abilities across the menstrual cycle. *Brain and cognition*, 14(1):26–43, 1990.

[140] Tara J Chavanne and Gordon G Gallup. Variation in risk taking behavior among female college students as a function of the menstrual cycle. *Evolution and Human Behavior*, 19(1):27–32, 1998.

[141] Jeanne M Meadowcroft and Dolf Zillmann. Women's comedy preferences during the menstrual cycle. *Communication Research*, 14(2):204–218, 1987.

[142] Tykeysha Powell-Boone, Timothy J Ness, Ronda Cannon, L Keith Lloyd, Douglas A Weigent, and Roger B Fillingim. Menstrual cycle affects bladder pain sensation in subjects with interstitial cystitis. *The Journal of urology*, 174(5):1832–1836, 2005.

[143] LA Houghton, R Lea, N Jackson, and PJ Whorwell. The menstrual cycle affects rectal sensitivity in patients with irritable bowel syndrome but not healthy volunteers. *Gut*, 50(4):471–474, 2002.

[144] E Anne MacGregor and Allan Hackshaw. Prevalence of migraine on each day of the natural menstrual cycle. *Neurology*, 63(2):351–353, 2004.

[145] S Marie Harvey. Female sexual behavior: Fluctuations during the menstrual cycle. *Journal of psychosomatic research*, 31(1):101–110, 1987.

[146] Diana Sanders, Pamela Warner, Torbjorn Backstrom, and John Bancroft. Mood, sexuality, hormones and the menstrual cycle. i. changes in mood and physical state: description of subjects and method. *Psychosomatic Medicine*, 45(6):487–501, 1983.

[147] Torbjorn Backstrom, Dlana Sanders, Rosemary Leask, David Davidson, Pamela Warner, and John Bancroft. Mood, sexuality, hormones, and the menstrual cycle. ii. hormone levels and their relationship to the premenstrual syndrome. *Psychosomatic Medicine*, 45(6):503–507, 1983.

[148] John Bancroft, Diana Sanders, David Davidson, and Pamela Warner. Mood, sexuality, hormones, and the menstrual cycle. iii. sexuality and the role of androgens. *Psychosomatic Medicine*, 45(6):509–516, 1983.

[149] Rudolf H Moos, Bert S Kopell, Frederick T Melges, Irvin D Yalom, Donald T Lunde, Raymond B Clayton, and David A Hamburg. Fluctuations in symptoms and moods during the menstrual cycle. *Journal of Psychosomatic Research*, 13(1):37–44, 1969.

[150] Robert R May. Mood shifts and the menstrual cycle. *Journal of psychosomatic research*, 20(2):125–130, 1976.

[151] Haiyan Wu, Chunping Chen, Dazhi Cheng, Suyong Yang, Ruiwang Huang, Stephanie Cacioppo, and Yue-Jia Luo. The mediation effect of menstrual phase on negative emotion processing: Evidence from n2. *Social neuroscience*, 9(3):278–288, 2014.

[152] Sarah Romans, Rose Clarkson, Gillian Einstein, Michele Petrovic, and Donna Stewart. Mood and the menstrual cycle: a review of prospective data studies. *Gender medicine*, 9(5):361–384, 2012.

[153] Scott A Golder and Michael W Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

[154] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638. ACM, 2014.

[155] Lucila Gallino, Facundo Carrillo, and Guillermo A Cecchi. Differential 28-days cyclic modulation of affective intensity in female and male participants via social media. *Frontiers in integrative neuroscience*, 13, 2019.

[156] Frank E Hytten, Isabella Leitch, et al. The physiology of human pregnancy. *The physiology of human pregnancy.*, 1964.

[157] Mark McLean, Andrew Bisits, Joanne Davies, Russell Woods, Philip Lowry, and Roger Smith. A placental clock controlling the length of human pregnancy. *Nature medicine*, 1(5):460–463, 1995.

[158] Dan Tulchinsky, Calvin J Hobel, Elizabeth Yeager, and John R Marshall. Plasma estrone, estradiol, estriol, progesterone, and 17-hydroxyprogesterone

in human pregnancy: I. normal pregnancy. *American journal of obstetrics and gynecology*, 112(8):1095–1100, 1972.

[159] STEPHEN C Robson, STEWART Hunter, RICHARD J Boys, and WILLIAM Dunlop. Serial study of factors influencing changes in cardiac output during human pregnancy. *American Journal of Physiology-Heart and Circulatory Physiology*, 256(4):H1060–H1065, 1989.

[160] D Elkharrat, JC Raphael, JM Korach, MC Jars-Guincestre, Cl Chastang, C Harboun, and Ph Gajdos. Acute carbon monoxide intoxication and hyperbaric oxygen in pregnancy. *Intensive care medicine*, 17(5):289–292, 1991.

[161] Andrew Czeizel, István Szentesi, Ildikó Szekeres, George Molnár, Anna Glauber, and Péter Bucski. A study of adverse effects on the progeny after intoxication during pregnancy. *Archives of toxicology*, 62(1):1–7, 1988.

[162] Marian Semczuk and Anna Semczuk-Sikora. New data on toxic metal intoxication (cd, pb, and hg in particular) and mg status during pregnancy. *Medical Science Monitor*, 7(2):332–340, 2001.

[163] Frank D Gilliland, Yu-Fen Li, and John M Peters. Effects of maternal smoking during pregnancy and environmental tobacco smoke on asthma and wheezing in children. *American journal of respiratory and critical care medicine*, 163(2):429–436, 2001.

[164] Frank D Gilliland, Kiros Berhane, Rob McConnell, W James Gauderman, Hita Vora, Edward B Rappaport, Edward Avol, and John M Peters. Maternal smoking during pregnancy, environmental tobacco smoke exposure and childhood lung function. *Thorax*, 55(4):271–276, 2000.

[165] Katharina M Hillerer, Stefan O Reber, Inga D Neumann, and David A Slattery. Exposure to chronic pregnancy stress reverses peripartum-associated

adaptations: implications for postpartum anxiety and mood disorders. *Endocrinology*, 152(10):3930–3940, 2011.

[166] Xi-Kuan Chen, Shi Wu Wen, Nathalie Fleming, Kitaw Demissie, George G Rhoads, and Mark Walker. Teenage pregnancy and adverse birth outcomes: a large population based retrospective cohort study. *International journal of epidemiology*, 36(2):368–373, 2007.

[167] R Jay Turner, Carl F Grindstaff, and Norman Phillips. Social support and outcome in teenage pregnancy. *Journal of Health and Social Behavior*, pages 43–57, 1990.

[168] Bruce J Ellis, John E Bates, Kenneth A Dodge, David M Fergusson, L John Horwood, Gregory S Pettit, and Lianne Woodward. Does father absence place daughters at special risk for early sexual activity and teenage pregnancy? *Child development*, 74(3):801–821, 2003.

[169] Christine Dunkel Schetter. Psychological science on pregnancy: stress processes, biopsychosocial models, and emerging research issues. *Annual review of psychology*, 62:531–558, 2011.

[170] Janet A DiPietro, Melissa M Ghera, K Costigan, and M Hawkins. Measuring the ups and downs of pregnancy stress. *Journal of Psychosomatic Obstetrics & Gynecology*, 25(3-4):189–201, 2004.

[171] Heather A Bennett, Adrienne Einarson, Anna Taddio, Gideon Koren, and Thomas R Einarson. Prevalence of depression during pregnancy: systematic review. *Obstetrics & Gynecology*, 103(4):698–709, 2004.

[172] Kristin Bergman, Pampa Sarkar, THOMAS G O'CONNOR, Neena Modi, and Vivette Glover. Maternal stress during pregnancy predicts cognitive ability

and fearfulness in infancy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(11):1454–1463, 2007.

[173] Charles H Zeanah. *Handbook of infant mental health*. Guilford Press, 2009.

[174] Jeanne Brooks-Gunn and Frank F Furstenberg. The children of adolescent mothers: Physical, academic, and psychological outcomes. *Developmental review*, 6(3):224–251, 1986.

[175] Juanita H Williams. *Psychology of women: Behavior in a biosocial context* . WW Norton & Co, 1987.

[176] Dinora Pines. Pregnancy and motherhood: interaction between fantasy and reality. *British Journal of Medical Psychology*, 45(4):333–343, 1972.

[177] Felicia Otchet, Mark S Carey, and Lorraine Adam. General health and psychological symptom status in pregnancy and the puerperium: what is normal? *Obstetrics & Gynecology*, 94(6):935–941, 1999.

[178] J Galen Buckwalter, Frank Z Stanczyk, Carol A McCleary, Brendon W Bluestein, Deborah K Buckwalter, Katherine P Rankin, Lilly Chang, and T Murphy Goodwin. Pregnancy, the postpartum, and steroid hormones: effects on cognition and mood. *Psychoneuroendocrinology*, 24(1):69–84, 1999.

[179] Bruce S McEwen. The molecular and neuroanatomical basis for estrogen effects in the central nervous system. *The Journal of Clinical Endocrinology & Metabolism*, 84(6):1790–1797, 1999.

[180] Whitney Wharton, Carey E Gleason, Olson Sandra, Cynthia M Carlsson, and Sanjay Asthana. Neurobiological underpinnings of the estrogen-mood relationship. *Current psychiatry reviews*, 8(3):247–256, 2012.

[181] Barbara B Sherwin. The impact of different doses of estrogen and progestin on mood and sexual behavior in postmenopausal women*. *The Journal of Clinical Endocrinology & Metabolism*, 72(2):336–343, 1991.

[182] PN Wang, SQ Liao, RS Liu, CY Liu, HT Chao, SR Lu, HY Yu, SJ Wang, and Hsiu-Chih Liu. Effects of estrogen on cognition, mood, and cerebral blood flow in ad a controlled study. *Neurology*, 54(11):2061–2066, 2000.

[183] Hadine Joffe and Lee S Cohen. Estrogen, serotonin, and mood disturbance: where is the therapeutic bridge? *Biological Psychiatry*, 44(9):798–811, 1998.

[184] Stefan Hammarbäck, TorbjÖRn Bäckström, Juhani Hoist, Bo Schoultz, and Sven Lyrenäs. Cyclical mood changes as in the premenstrual tension syndrome during sequential estrogen-progestagen postmenopausal replacement therapy. *Acta obstetricia et gynecologica Scandinavica*, 64(5):393–397, 1985.

[185] Uriel Halbreich and Linda S Kahn. Role of estrogen in the aetiology and treatment of mood disorders. *CNS drugs*, 15(10):797–817, 2001.