

# UNA REPRESENTACIÓN BASADA EN CARACTERÍSTICAS LATENTES PARA LA IDENTIFICACIÓN DE LA PERSONALIDAD

TESIS DOCTORAL  
para obtener el grado de  
Doctora en Ciencias Naturales e Ingeniería

Por:

ADRIANA GABRIELA RAMÍREZ DE LA ROSA

Asesores:

Héctor Jiménez Salazar, director  
Esaú Villatoro Tello, co-director

Jurado:

Dr. Iván Vladimir Meza Ruiz, IIMAS, UNAM  
Dra. Rosa María Ortega Mendoza, UAEH  
Dra. Ángeles Belem Priego Sánchez, UAM Azcapotzalco  
Dr. Jorge Cervantes Ojeda, UAM Cuajimalpa

Ciudad de México, 2024



Para Sara y José,  
Sara Victoria y Brissa,  
Yubia, mi mayor soporte,  
y Omar.





# Agradecimientos

Una investigación de doctorado es una tarea compleja en la cual intervienen un sin fin de personas. Agradezco a cada una de ellas.

Agradezco a mis asesores, al doctor Héctor Jiménez por sus consejos y guías y al doctor Esaú Villatoro por trabajar conmigo en este proyecto de investigación.

Agradezco a mi comité evaluador que leyeron, comentaron y sugirieron aspectos para mejorar el trabajo de investigación. Muchas gracias por sus comentarios alentadores a las doctoras Rosa María Ortega y Ángeles Priego, y a los doctores Iván Meza y Jorge Cervantes.

Muchas gracias a todas las personas que intervinieron en la creación del corpus RxPI. A todas las personas que aportaron sus ensayos para ser analizados; a los estudiantes que recorrieron la Torre 3 de la UAM Cuajimalpa para aplicar el instrumento de recolección; y a las personas encargadas de transcribir manualmente los ensayos, muchas gracias por la dedicación al proyecto y por realizar un buen trabajo.

A las psicólogas que me orientaron y guiaron en diferentes etapas de la investigación: Cecilia Cisneros durante la creación del corpus; Jaime Rojas y la doctora Verónica Reyes en la etapa de análisis de datos y durante el desarrollo del método propuesto, sus aportaciones fueron extremadamente valiosas para el desarrollo de este proyecto.

Agradezco a Alfredo Mateos y Carlos Jaimez por el apoyo que siempre me brindaron para que pudiera trabajar en este proyecto de investigación. A Joel Rivero, además, por su constante apoyo y aliento. A todos los colegas del Departamento de Tecnologías de la Información porque de una u otra forma me motivaron a ser una mejor científica.

Agradezco mucho a los estudiantes que me brindaron la oportunidad de acompañarlos en diferentes proyectos de clases o de investigación. Especialmente a Alejandro Rosales que me introdujo al concepto de la disponibilidad léxica. A Tonantzin Siurob, Ángeles López y Érika

## **Agradecimientos**

---

Rosas por inspirarme con su entusiasmo por aprender y cuestionar.

Agradezco todas las facilidades que me brindó la Universidad Autónoma Metropolitana Unidad Cuajimalpa durante la realización de esta tesis.

Gracias a quienes con sus comentarios, inadvertidamente me motivaron a no abandonar el proyecto, se me vienen a la mente Manuel Montes, Luis Villaseñor y Grigori Sidorov, pero hay muchas personas más.

Gracias a mi familia. A Sara Victoria por su amor, a Brissa por su apoyo incondicional, a Sara y José por apoyarme durante 40 años, a Yubia por su compañía, aliento en todos los momentos y por siempre creer que puedo hacer cosas que ni yo creo que puedo hacer, y a Omar por todo el apoyo en todos los pasos de este camino.

A los creadores del arte que en diferentes formas me acompañaron durante estos años. Gracias a James S. A. Corey por crear *The Expanse*. Gracias a Kim N., Kim S., Min Y., Jung H., Park J., Kim T. y Jeon J. por la música y la inspiración.

*Oaxaca, 9 de julio de 2024*

G. Ram.



## Resumen

El conjunto de patrones de comportamiento individual que incluye emociones y pensamientos define la personalidad. Estudiar e identificar la personalidad de un individuo puede ayudar en la detección de aspectos que impactan su salud física y mental. Sin embargo, identificar la personalidad es una tarea costosa, pues aunque existen instrumentos para la autoevaluación, generarlos y analizarlos requiere conocimientos especializados y recursos. Los cuestionarios son solo uno de varios instrumentos que los psicólogos utilizan al identificar y evaluar la personalidad de sus pacientes. En este contexto, la utilización de sistemas automáticos para apoyar las actividades de especialistas en salud mental son cada vez más necesarios. Estos métodos automáticos pueden construirse a partir de analizar cómo una persona realiza actividades tales como el comportamiento ante diferentes situaciones, la forma de expresarse, el uso que da a sus redes sociales, entre otras actividades cotidianas. De acuerdo a la literatura en psicología, el lenguaje es un buen indicador de la personalidad pues a través de él se expresa la forma de pensar y sentir. Así, esta tesis propone que a través del uso de una representación de texto basada en el análisis del léxico disponible extraído de muestras de lenguaje, y bajo un esquema de aprendizaje supervisado, es posible identificar rasgos de la personalidad del modelo *Big-Five*. El método genera una representación interpretable por especialistas, permitiéndoles entender mejor los resultados automáticos que el método obtiene. La evaluación experimental apoya la hipótesis propuesta, obteniendo resultados de clasificación superiores a los métodos de referencia que hacen uso de un esquema tradicional de Bolsa de Palabras y otro que emplea representaciones recientes basadas en modelos *transformers*. Adicionalmente, se evaluó el método propuesto en dos tareas de salud mental: depresión y demencia, obteniendo resultados alentadores que sustentan la utilidad del uso de la disponibilidad léxica en la representación del método automático propuesto.







## Abstract

Personality refers to the set of distinctive behavioral patterns in an individual, including emotions and thoughts. Studying and identifying an individual's personality can help in the detection of aspects that impact their physical and mental health. However, identifying personality is a costly task; although there are instruments for self-assessment, generating and analyzing them requires specialized knowledge and resources. Questionnaires are only one of several instruments that psychologists use when identifying and assessing the personality of their patients. In this context, the use of automatic systems to support the activities of mental health specialists is becoming increasingly necessary. These automatic methods can analyze how a person performs activities such as their behavior in different situations, the way they express themselves, their social networks' use, among other daily activities. According to psychology's literature, language is a good indicator of personality since it expresses the way of thinking and feeling. Thus, this thesis proposes that through the use of a text representation based on the analysis of the available lexicon extracted from language samples, and under a supervised learning scheme, it is possible to identify personality traits of the Big-Five model. The method generates a representation that can be interpreted by specialists, allowing them to better understand the results obtained by the automatic method. The experimental evaluation supports the proposed hypothesis; the proposed method obtains superior performance than the reference methods (such as using as representation that uses a traditional Bag of Words scheme and another one that employs recent representations based on transformer models). Additionally, the method was evaluated in two mental health tasks: depression and dementia, obtaining encouraging results that support the usefulness of the use of lexical availability for these tasks.





# Índice general

<b>Agradecimientos</b>	<b>I</b>
<b>Resumen</b>	<b>III</b>
<b>Abstract</b>	<b>V</b>
<b>I Introducción</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
1.1. Hipótesis . . . . .	4
1.2. Objetivos . . . . .	5
1.3. Contribuciones . . . . .	5
1.4. Estructura del documento . . . . .	6
<b>II Revisión de antecedentes teóricos</b>	<b>7</b>
<b>2. Lexicón mental y disponibilidad léxica</b>	<b>9</b>
2.1. El lexicón mental . . . . .	9
2.2. La disponibilidad léxica . . . . .	11
2.2.1. Teoría de disponibilidad léxica: contexto histórico . . . . .	12
2.2.2. Usos de la disponibilidad léxica . . . . .	12
2.3. Análisis final . . . . .	14
<b>3. Clasificación automática de textos</b>	<b>15</b>
3.1. Representación . . . . .	16
3.2. Métricas de desempeño . . . . .	18
3.3. Calidad de los datos . . . . .	19
<b>4. Personalidad y su identificación automática</b>	<b>21</b>

## Índice general

---

4.1. El modelo <i>Big Five</i> de personalidad . . . . .	21
4.1.1. Un modelo generalizable . . . . .	22
4.2. Identificación automática de la personalidad . . . . .	23
4.2.1. Representaciones basadas en atributos lingüísticos . . . . .	23
4.2.2. Representaciones basadas en aprendizaje profundo . . . . .	27
4.2.3. Análisis de las representaciones usadas para la identificación de rasgos de personalidad . . . . .	30
<b>III Contribuciones</b>	<b>31</b>
<b>5. Colecciones de datos</b>	<b>33</b>
5.1. RxPI—un recurso para el análisis de la personalidad . . . . .	33
5.1.1. Motivación . . . . .	34
5.1.2. Composición . . . . .	35
5.1.3. Proceso de recolección . . . . .	42
5.1.4. Preprocesamiento/limpieza/etiquetado . . . . .	44
5.1.5. Usos . . . . .	45
5.1.6. Distribución . . . . .	46
5.1.7. Mantenimiento . . . . .	47
5.2. Otras colecciones utilizadas . . . . .	48
5.2.1. Conjunto de datos Auto–descripciones . . . . .	48
5.2.2. Conjunto de datos Essays . . . . .	49
<b>6. Método propuesto</b>	<b>51</b>
6.1. Adaptación de la prueba de disponibilidad léxica . . . . .	52
6.1.1. Preprocesamiento de textos . . . . .	52
6.1.2. Cálculo de la disponibilidad léxica con repeticiones . . . . .	53
6.1.3. Disponibilidad léxica en diferentes colecciones . . . . .	55
6.2. Representación de texto basada en disponibilidad léxica . . . . .	58
6.2.1. Representación básica: DL–base . . . . .	59
6.2.2. Representación basada en el valor de disponibilidad léxica: DL–va . . . . .	60
6.3. Autoselección de categorías . . . . .	61
6.3.1. Método de autoselección de categorías relevantes . . . . .	62
6.3.2. Diccionario de Categorías . . . . .	64
6.4. Esquema general de las contribuciones . . . . .	67
<b>7. Identificación automática de la personalidad: experimentación y análisis</b>	<b>69</b>
7.1. Configuración experimental . . . . .	70
7.2. Experimento 1: Disponibilidad léxica para una representación textual . . . . .	71

7.2.1. Análisis sobre el uso de la disponibilidad léxica . . . . .	75
7.2.2. Dispersión temática . . . . .	76
7.3. Experimento 2: Autoselección de categorías para mejorar la representación textual	78
7.3.1. Autoselección de categorías predefinidas en LIWC . . . . .	79
7.3.2. Autoselección de categorías autogeneradas con LDA . . . . .	93
7.3.3. Comparación del uso de los diccionarios generados: manualmente y automáticamente . . . . .	97
<b>8. Identificación automática de Demencia y Depresión: experimentación y análisis</b>	<b>103</b>
8.1. Corpora para la identificación de <i>Demencia y Depresión</i> . . . . .	103
8.1.1. ADReSS: <i>Alzheimer's Dementia Recognition through Spontaneous Speech</i> .	104
8.1.2. <i>Distress Analysis Interview Corpus</i> . . . . .	104
8.2. Configuración experimental . . . . .	105
8.3. Experimento 1: Disponibilidad léxica para una representación textual . . . . .	107
8.4. Experimento 2: Autoselección de categorías para mejorar la representación . . .	109
8.4.1. Comparación entre DL–base y DL–va en tareas de identificación de trastornos psicológicos . . . . .	114
8.4.2. Resumen . . . . .	115
<b>IV Conclusiones generales</b>	<b>117</b>
<b>9. Conclusiones</b>	<b>119</b>
9.1. Perspectivas y trabajo futuro . . . . .	120
9.2. Consideraciones éticas . . . . .	121
9.3. Colecciones de datos para investigación en el ámbito de la salud . . . . .	123
9.4. Publicaciones científicas . . . . .	124
<b>A. Creación del conjunto de datos RxPI</b>	<b>125</b>
A.1. El instrumento de recolección . . . . .	125
A.2. Asignación de valores numéricos y clases nominales . . . . .	128
A.3. Transcripción manual . . . . .	130
A.3.1. Etiquetas para anotadores . . . . .	130
A.3.2. Preguntas frecuentes o FAQs . . . . .	133
A.3.3. Formulario para transcriptores . . . . .	135
<b>Bibliografía</b>	<b>141</b>



A horizontal decorative bar with a repeating pattern of overlapping purple circles and arcs, creating a textured, scale-like appearance.

# Parte I

## Introducción





# 1



## Introducción

La personalidad se refiere al conjunto de patrones de comportamiento —que incluyen emociones y pensamientos— distintivos en una persona. Estos patrones caracterizan la forma en que cada individuo se adapta a una situación particular (Mischel, 1993; Funder, 2001). Usualmente, los patrones de comportamiento que conforman la personalidad se mantienen estables a lo largo del tiempo y a través de diferentes situaciones.

Estudiar e identificar la personalidad de un individuo puede ayudar en la detección de aspectos como el nivel de bienestar de las personas. Por ejemplo, se ha encontrado que la personalidad tiene un impacto en la salud física y mental; así, conocer la personalidad ayuda en la prevención de conductas o enfermedades mentales, como depresión y ansiedad (Klein et al., 2011; Jourdy and Petot, 2017). Además, existe una relación de la personalidad con la identificación de trastornos de la personalidad detonados por algún evento inesperado (Ozer and Martínez, 2006).

Sin embargo, identificar la personalidad de una persona es una tarea costosa, pues aunque existen instrumentos como cuestionarios de personalidad para la autoevaluación, generar estos instrumentos requiere conocimientos especializados y tiempo. Contar con instrumentos de autoevaluación requiere primero ser creados por especialistas y luego validados para la comunidad a la que se deba aplicar. Por otro lado, los cuestionarios son solo una de varias herramientas que los psicólogos utilizan cuando van a identificar y evaluar la personalidad de sus pacientes. Y desafortunadamente, en la sociedad actual el sistema de salud mental se ve rebasado por la demanda. En este contexto, la utilización de sistemas automáticos pretende aliviar la sobrecarga de los especialistas para apoyar en la tarea de identificación de personalidad.

Tales métodos automáticos pueden construirse a partir de analizar actividades realizadas por un individuo (Matthews et al., 2009). Por ejemplo, la forma de presentarse ante otros (Batinca et al., 2011), su comportamiento en reuniones informales (Kalimeri, 2013), la forma de expresarse

## Capítulo 1. Introducción

---

verbal y de forma escrita (Mairesse et al., 2007; Pennebaker et al., 2001), el uso de sus redes sociales (Adali and Golbeck, 2012; Ortigosa et al., 2014), entre otras. Sin embargo, es deseable contar con sistemas confiables que requieran *poca* información en un ambiente poco o nada controlado.

Cuando una persona especialista analiza la personalidad a través de la observación directa, está reportando la personalidad percibida. De acuerdo al modelo cognitivo de Brunswik (*Brunswik Lens*) (Brunswik, 1956), la personalidad *real* (en contraposición a la personalidad *percibida*) de un sujeto puede ser capturada mediante pistas observables. Estas pistas observables, una vez capturadas pueden usarse en el análisis de la personalidad.

En este sentido, es importante determinar las pistas que se observarán de cada individuo para identificar su personalidad. De acuerdo con Keirse (1998), los tipos de personalidad de una persona probablemente se derivan de la interconexión de dos de las acciones más básicas de los humanos: cómo nos comunicamos y cómo utilizamos herramientas para lograr objetivos.

De manera general, al comunicarnos buscamos y recuperamos léxico almacenado en nuestra mente —lexicón mental—. De forma análoga, el modelo psicológico de rasgos de personalidad *Big-Five* se basa en la hipótesis léxica —que existe un léxico que es relevante para cada rasgo de personalidad del modelo—. Así, en esta tesis presento un método para la identificación automática de rasgos de la personalidad del modelo *Big-Five*. El método propuesto tiene como base la identificación del léxico disponible —del lexicón mental— de una comunidad de personas que comparten el mismo rasgo de personalidad. Luego, este léxico disponible es utilizado para generar una representación textual de muestras de una comunicación corta dentro de un modelo de clasificación automática.

Para evaluar el método propuesto se construyó un corpus específico para la tarea de identificación de la personalidad que se basa en la obtención de muestras de lenguaje; estas muestras comunican aspectos personales y son escritas en español mexicano. El método también fue evaluado con otros corpora tanto en español como en inglés. Los resultados obtenidos son comparables con métodos del estado del arte basados en redes artificiales profundas y representaciones que usan un enfoque de aprendizaje de características. Adicionalmente, el método propuesto revela aspectos del lexicón mental de la comunidad estudiada más allá de un estado psicológico particular. El análisis de aspectos del lexicón mental ha mostrado ser relevante para entender la memoria y el lenguaje y para ayudar en el diagnóstico de enfermedades neurológicas.

### 1.1. Hipótesis

La hipótesis de esta investigación establece que:

*A través del uso de una representación de texto basada en el análisis del lexicón disponible extraído de muestras de lenguaje, y bajo un esquema de aprendizaje supervisado, es posible identificar rasgos de la personalidad de una persona.*

### 1.2. Objetivos

*Proponer una metodología para la identificación automática de la personalidad en muestras de texto producidas por una persona basada en el análisis del lexicón disponible.*

Los objetivos específicos de esta investigación son:

1. Crear un corpus de textos cortos en español mexicano etiquetados con los rasgos de personalidad del modelo *Big-Five*.
2. Adaptar la prueba de disponibilidad léxica para ser usada en muestras de lenguaje producido en lugar de listas de palabras elicitadas.
3. Proponer una nueva representación de texto basada en la incorporación de información obtenida con la prueba de disponibilidad léxica adaptada.
4. Diseñar e implementar un método de clasificación automático de textos que incorpore la representación propuesta para identificar rasgos de personalidad.

### 1.3. Contribuciones

A continuación se listan las principales contribuciones de esta tesis doctoral:

1. Se diseñó, colectó y creó el corpus RxPI que contiene muestras de lenguaje en español mexicano de 1484 participantes etiquetadas con los rasgos de personalidad del modelo *Big-Five*.
2. Se adaptó la prueba de disponibilidad léxica utilizada en el área de Sociolingüística para obtener aspectos del lexicón mental a partir de muestras de lenguaje.
3. Se propusieron dos variantes de una representación basada en la información obtenida del léxico disponible.
4. Se diseñó un método automático para seleccionar atributos usados en la representación dentro de un esquema de clasificación automática de textos.

### 1.4. Estructura del documento

Esta tesis trata de contestar si la información del lexicón mental obtenida a través de la disponibilidad léxica ayuda a la identificación automática de rasgos de la personalidad en muestras cortas de textos. La experimentación realizada muestra evidencia de que la hipótesis es correcta. El resto del documento está organizado como sigue:

La Parte II muestra una revisión de antecedentes teóricos con el objetivo de proveer contexto y conceptos relevantes. Esta revisión de antecedentes teóricos contiene 3 capítulos. En el Capítulo 2 se introduce el *lexicón mental* y la *disponibilidad léxica*, conceptos utilizados en diferentes áreas adicionales a las ciencias de la computación. El Capítulo 3 presenta conceptos relacionados directamente con la clasificación automática de textos desde el punto de vista computacional. El Capítulo 4 hace un recorrido por el estado del arte sobre la identificación automática de rasgos de personalidad.

La Parte III de este documento está enfocada en las contribuciones de la tesis y contiene 4 capítulos. En el Capítulo 5 se describe a detalle el corpus *RxPI* creado en este trabajo de investigación; así también se describen dos corpora más utilizados en la experimentación. En el Capítulo 6 se muestra el método propuesto que incorpora la adaptación de la prueba de disponibilidad léxica; la representación propuesta basada en la disponibilidad léxica; y el método automático diseñado para autoseleccionar los mejores atributos para la representación. El Capítulo 7 describe la experimentación realizada para evaluar la hipótesis planteada en la identificación de rasgos de la personalidad. A continuación, en el Capítulo 8 se presentan experimentos realizados para evaluar el desempeño del método propuesto en la identificación automática de demencia y depresión.

Finalmente, la Parte IV contiene las conclusiones de la tesis. El Capítulo 9 presenta las conclusiones generales de este trabajo, perspectivas y trabajo futuro; así como una discusión sobre consideraciones éticas sobre la utilización de métodos automáticos para identificar aspectos de salud mental.

## Parte II

### Revisión de antecedentes teóricos



# 2



## Lexicón mental y disponibilidad léxica

Este capítulo introduce el lexicón mental como modelo para representar aspectos de la mente humana, describe las pruebas y tareas desarrolladas para observar estructuras en el lexicón mental, y presenta la relación entre el lexicón mental y el léxico disponible. Finalmente, el capítulo introduce la prueba de disponibilidad léxica que permite extraer y analizar el léxico disponible de una comunidad de hablantes con características particulares.

### 2.1. El lexicón mental

Cuando nos comunicamos a través del lenguaje, buscamos y recuperamos palabras almacenadas en nuestra mente. Este conjunto de palabras, o vocabulario, es lo que los investigadores han llamado *lexicón mental* (Allport and Funnell, 1981). Este concepto fue mencionado por primera vez en 1961 en la tesis doctoral de Ann Triesman aunque con el nombre de *diccionario mental* (Coltheart et al., 2001).

El *lexicón mental* puede definirse como un modelo representado por una estructura altamente interconectada de diversas dimensiones que organiza el vocabulario en la memoria a largo plazo. Algunas dimensiones de este modelo corresponden a la sintaxis, partes de la oración (o *part of speech*), pronunciación o fonología, significado afectivo, entre otras propiedades y características de las palabras almacenadas (Tamariz-Martel Mirêlis, 2005; Rensbergen et al., 2015). La importancia del lexicón mental radica en que al entender su estructura, así como los procesos que se llevan a cabo en él, es posible comprender aspectos del procesamiento del lenguaje en los humanos; por ejemplo, determinar la fuerza en la relación entre pares de palabras—aspecto importante para entender la memoria y *priming* (cómo y en qué medida el contexto previo influencia lo que viene después) (De Deyne and Storms, 2015; Stella, 2020)—.

## Capítulo 2. Lexicón mental y disponibilidad léxica

---

Con instrumentos actuales no es posible observar o medir directamente las conexiones en la mente humana con respecto a la organización de nuestro vocabulario (Aitchison, 2003); por lo tanto, se han desarrollado herramientas que permiten construir partes del *lexicón mental*. Una de las herramientas ampliamente usadas en las ciencias cognitivas para construir el lexicón mental es la asociación de palabras (Deyne et al., 2003).

La prueba de *asociación de palabras* se usa para elicitarse *respuestas* que primero vengan a la mente de los participantes dado un *detonante*. Normalmente existe un conjunto de detonantes que son usados para generar diversas respuestas y generalmente tanto las respuestas como los detonantes son palabras (De Deyne and Storms, 2015). Al analizar las *respuestas* obtenidas de un conjunto de participantes, se puede formar una estructura semántica en forma de grafo  $G = (V, E)$  donde los vértices  $V$  son la unión de las palabras detonantes y las respuestas; y los arcos  $E$  se forman por la relación de una respuesta y el detonante usado. Por lo que utilizar un número grande de detonantes es importante para generar una estructura compleja.

Una vez que se conoce el lexicón mental, puede ser usado para evaluar el funcionamiento verbal como la capacidad de organización temporal y secuencia de la información. Específicamente, existen estudios neurológicos que utilizan un tipo de pruebas de asociación de palabras simplificada denominadas *pruebas de fluidez verbal*. El estudio de esta capacidad de organización ha ayudado a diagnosticar trastornos o enfermedades neurológicas y psiquiátricas (Allen et al., 1993; Cerhan et al., 2002; Andreou and Trott, 2013). En las *pruebas de fluidez verbal* quienes participan responden a solo *una* palabra detonante, de aquí que es una simplificación de las pruebas de asociación de palabras.

Las *pruebas de fluidez verbal* pueden llevarse a cabo mediante dos tareas conocidas como: i) fluidez semántica, donde el detonante y las respuestas son palabras, y ii) fluidez fonológica, donde el detonante es una letra y la respuesta es una palabra que empiece con la letra dada. Estas tareas típicamente se elicitarse en un tiempo límite, normalmente un minuto y de forma hablada o escrita. Las tareas consisten en mencionar o escribir todas las palabras que vienen a la mente de una categoría semántica (e.g., animales) o que empiecen con una letra (e.g., que empiecen con la letra P) respectivamente. Al finalizar la actividad, cada participante obtiene un valor que corresponde al número de palabras únicas listadas, donde entre mayor sea este valor, mayor fluidez verbal.

Existe un tipo de fluidez semántica que tiene sus bases en la teoría de disponibilidad léxica usada principalmente por lingüistas, llamada prueba de disponibilidad léxica. La siguiente sección explica el objetivo de la prueba de disponibilidad léxica y las diferencias con las pruebas de fluidez semántica; también presenta la razón de utilizarla como principal herramienta para obtener aspectos del lexicón mental en el contexto de esta tesis.



### 2.2. La disponibilidad léxica

La *prueba de disponibilidad léxica* tiene como objetivo encontrar un conjunto de palabras que una comunidad dada —conjunto de personas que comparten alguna característica de interés— tiene en su lexicón mental y que está condicionado por una intención comunicativa. Una *intención comunicativa* o intención de comunicación se define como un mensaje que tiene como objetivo transmitir información a otros a través del lenguaje escrito o hablado.

La asociación de palabras, descrita en la sección anterior, es una generalización de las pruebas de fluidez semántica; y en particular es una generalización de la prueba de disponibilidad léxica. Mientras que la asociación de palabras es una de las pruebas más utilizadas por la comunidad científica cuando se busca entender el lexicón mental, tienen diferencias importantes con la prueba de disponibilidad léxica en cuanto al propósito de la prueba, el foco y el orden de las respuestas elecitadas:

- **Propósito.** Mientras que la asociación de palabras tiene como finalidad encontrar estructuras basadas en conexiones entre las palabras detonantes usadas, donde para obtener estructuras más nítidas se requiere la participación de más personas y un conjunto grande de palabras detonantes; las pruebas de disponibilidad léxica busca determinar un valor de disponibilidad o accesibilidad de las palabras respuestas.
- **Foco.** En la asociación de palabras el foco es la construcción de las estructuras que se pueden formar de pares detonante–respuesta. Pero en la disponibilidad léxica, el foco está en el detonante; pues se busca encontrar respuestas relacionadas con estos detonantes que usualmente son elegidos de forma cuidadosa para obtener respuestas significativas.
- **Orden.** El orden de respuestas no es importante en la asociación de palabras pues mientras existe una respuesta a un detonante, existirá un vértice que conecte a ambas palabras. En contraste, en la disponibilidad léxica, el orden en que se recuperan de la memoria las respuestas impactará el análisis posterior. Entre más pronto vengan a la mente más disponibles están las palabras.

Dadas las características mencionadas, la hipótesis de esta tesis establece que es posible utilizar una adaptación de la prueba de disponibilidad léxica a una intención comunicativa para establecer la disponibilidad de palabras usadas por una comunidad —con características particulares, por ejemplo su personalidad—. El léxico disponible luego permitirá contrastar entre comunidades de interés; por ejemplo, una comunidad de personas introvertidas de una comunidad de personas extrovertidas.

### 2.2.1. Teoría de disponibilidad léxica: contexto histórico

Hacia 1954, en Francia, existió la necesidad de enseñar la lengua francesa a quienes vivían en territorios franceses (*Union Française*) y por tanto, la necesidad de crear recursos educativos de enseñanza del francés como segunda lengua (Gougenheim, 1954). En el desarrollo de este recurso educativo, los investigadores notaron que incluir palabras basándose únicamente en la frecuencia de uso en el lenguaje francés produciría un vocabulario no necesariamente útil para la comunicación del día a día. Por tanto, en su trabajo pionero, Gougenheim et al. (1964) desarrollaron la idea de crear un diccionario de palabras basado en una combinación entre disponibilidad y frecuencia.

De acuerdo con López Morales (2014), René Michéa —coautor en (Gougenheim et al., 1964)— fue quien hizo la distinción entre lo que es frecuente y lo que es disponible, como propiedades de las palabras. Así, la disponibilidad léxica es la combinación de propiedades que hacen útil a un vocabulario para una situación comunicativa dada. Por lo tanto, la idea propuesta por Gougenheim et al. (1964) fue la de seleccionar palabras no solamente por su frecuencia en textos formales en la lengua francesa —que era el método tradicional usado para la construcción de diccionarios hasta entonces— sino también por su disponibilidad en hablantes nativos (lengua madre o L1) en categorías semánticas específicas (o detonantes).

A partir de aquí, el objetivo se centró en encontrar aquellas palabras accesibles. De acuerdo con Dimitrijević (1981) “una palabra, independiente de su frecuencia y otras características léxicas, es accesible si en una situación y contexto particular esta palabra viene a la mente fácilmente y sin necesidad de buscarla”. Para esto los investigadores empezaron a desarrollar una tarea que les permitiera generar un léxico disponible dada una temática, o categoría semántica. Al final, la idea fue usar la tarea de fluidez semántica otorgando al detonante el nombre de *centro de interés*.

### 2.2.2. Usos de la disponibilidad léxica

A partir de este momento, los centros de interés de las pruebas de disponibilidad léxica toman su puesto como principal foco de estudio. Dada su génesis, el primer uso que se le dio a la disponibilidad léxica fue el de ayudar en la generación de material didáctico para aprender un segundo lenguaje (L2) en otras lenguas diferentes al francés. Y más tarde para comparar entre comunidades de estudiantes bilingües en diferentes regiones del mundo.

Por ejemplo, Dimitrijević (1981) realizó un estudio para comparar la disponibilidad léxica entre estudiantes monolingües y bilingües utilizando como hipótesis que la disponibilidad puede estar determinada por diversos factores como género, edad, intereses profesionales, además del contexto social y cultural. En este trabajo se analizaron dos ciudades de Europa con un número reducido de participantes: 228 estudiantes de Serbia y 185 de Escocia. Y particularmente

en dos comunidades de estudiantes de Serbia: monolingües y bilingües. Entre sus resultados, Dimitrijević (1981) encontró que basado en el número de unidades léxicas producidas, las mujeres producen más palabras que los hombres, pero los hombres producen más palabras distintas que las mujeres, independientemente de su lengua materna. El nivel socioeconómico, la profesión y nivel de estudios de los padres no tuvo efecto en el número de palabras producidas pero sí en la disponibilidad léxica, particularmente en los centros de interés *Política*, *Alimentos y bebidas* y *Profesiones*; lo que indica que aunque las unidades léxicas sean las mismas, el valor de disponibilidad es relevante al estudiar comunidades.

Posteriormente, el Proyecto Panhispánico de Disponibilidad Léxica (López Morales, 2014) definió 16 centros de interés para ser utilizados por la comunidad científica con el objetivo de permitir comparaciones entre estudios. Algunos de estos centros de interés vienen de la propuesta original de Gougenheim et al. (1964) y son: partes del cuerpo humano, la ropa, partes de la casa (sin muebles), alimentos y bebidas, objetos colocados en la mesa para la comida, la cocina y sus utensilios, la escuela: muebles y materiales, iluminación, calefacción y medios de airear un recinto, la ciudad, el campo, medios de transporte, trabajos del campo y del jardín, los animales, juegos y distracciones, y profesiones y oficios. El objetivo de este proyecto internacional, iniciado en 1990, fue el de construir diccionarios de disponibilidad léxica de los centros de interés preestablecidos, para diferentes variantes del español homogeneizando la metodología que permitiera hacer comparaciones lingüísticas, etnográficas y culturales entre diversas comunidades (Grupo de Investigación DispoLex, 2023).

La investigación de la disponibilidad léxica ha estado presente en otras áreas del conocimiento además de la lingüística, entre ellas en pedagogía, sociología y psicología. En pedagogía, se ha utilizado la disponibilidad léxica para identificar problemas en los procesos de aprendizaje de estudiantes de diferentes áreas. Por ejemplo Salcedo et al. (2016) y Rojas et al. (2016) comparan el desempeño de aprendizaje en estudiantes de matemáticas en el sur de Chile utilizando centros de interés específicos a cada subdisciplina de las matemáticas estudiadas, como *sistemas numéricos*, *cálculo*, *geometría*, etc.

En sociología, diversos estudios han usado la prueba de disponibilidad léxica para entender las diferencias léxicas entre variables demográficas como el género. Por ejemplo, Hernández Muñoz (2010) usa la disponibilidad léxica para explorar cómo las mujeres y los hombres cambian su producción léxica en seis centros de interés entre dos canales: escrito y hablado. Interesantemente, este trabajo utiliza 5 centros de interés de la investigación tradicional —de los 16 centros usados por la comunidad científica— adicional a un centro de interés abstracto, que la autora llama *inteligencia*. En el centro de interés abstracto, *inteligencia*, la autora encontró que el grupo de mujeres participantes utilizan términos considerados positivos con mayor disponibilidad que términos considerados negativos; en contraste con el grupo de hombres participantes, quienes tienen más accesibilidad a términos considerados negativos o peyorativos. Este resultado puede

## Capítulo 2. Lexicón mental y disponibilidad léxica

---

indicar cómo el contexto social (particularmente el condicionamiento cultural en Latinoamérica) permea el lexicón mental y en consecuencia afecta la disponibilidad léxica de estas dos poblaciones. Esta conclusión es similar a la que Ríos González (2008) obtuvo en su estudio sobre el centro del interés *tópicos de conversación*. Al examinar el léxico disponible, la autora encontró que las mujeres expresan términos estereotípicos de lo que se espera que una mujer en países latinoamericanos utilice en público.

Otros trabajos que comparan la disponibilidad léxica entre mujeres y hombres son los de Lázaro and Rubino (2020) y Pérez Durán (2020) quienes utilizan dos centros de interés nuevos: *homosexualidad e insulto*, respectivamente. Mientras que Lázaro and Rubino (2020) comparan entre género y edad con el objetivo de describir posturas sociales–culturales sobre inclusión y tolerancia en estudiantes de bachillerato de dos escuelas (pública y privada) de la Ciudad de Culiacán, México; Pérez Durán (2020) analiza el vocabulario de estudiantes de secundaria en la ciudad de San Luis Potosí, México para determinar qué vocabulario se utiliza como insultos.

Todos los trabajos mencionados proponen un cambio en el estudio de la disponibilidad léxica, particularmente en utilizar otros centros de interés a los propuestos inicialmente que se adapten al interés particular de cada investigación. Entre los centros de interés propuestos, en su mayoría abstractos son: *inteligencia* (incorporado por Hernández Muñoz (2010)), *temas de conversación* (usado por Ríos González (2008)), *insulto* (usado por Pérez Durán (2020)) y *homosexualidad* (incorporado por Lázaro and Rubino (2020)). Además, de acuerdo con los protocolos de las investigaciones, particularmente en sociología y psicología, se ha ido cambiando la prueba de disponibilidad léxica de la elicitación a la producción espontánea, como es el caso del trabajo de Pérez Durán (2020).

### 2.3. Análisis final

De los trabajos descritos anteriormente, sabemos que las pruebas de disponibilidad léxica (y otras pruebas de asociación) son útiles para estudiar comunidades, ya que el lenguaje es un producto de nuestro contexto social. Además, podemos establecer un enlace entre el léxico disponible y aspectos del lexicón mental; en consecuencia, podemos intentar comprender ciertos aspectos de la mente humana. Esta tesis propone crear un modelo que incorpore la obtención del léxico disponible a partir de una producción del habla escrita o transcrita, y luego utilizar el léxico disponible de dos comunidades con personalidad diferente para identificar automáticamente un rasgo de esta característica psicológica. La hipótesis principal de esta tesis, establece que analizar el léxico disponible entre dos comunidades permite distinguir rasgos de la personalidad de quien produce lenguaje.

# 3



## Clasificación automática de textos

La identificación automática de la personalidad como se aborda en esta tesis es un problema que desde la inteligencia artificial cae dentro del aprendizaje supervisado. La principal característica del aprendizaje supervisado es que necesita un conjunto de ejemplos del cual *aprender*.

Formalmente, el **aprendizaje supervisado** se define como (Russell and Norvig, 2009): Dado un *conjunto de entrenamiento* de  $N$  pares de ejemplos, con la forma  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ , donde cada  $y_j$  es generada por una función desconocida  $f(x_j) = y_j$ , se construye una función  $h$  que aproxime la función real  $f$ . La función  $h$  se evalúa con un conjunto diferente de pares entrada–salida, si la función  $h$  predice de forma correcta el valor de  $y$  para los nuevos ejemplos, entonces  $h$  generaliza bien a  $f$ .

Dependiendo de la forma de  $y$  en el conjunto de entrenamiento, existen dos tareas dentro del aprendizaje supervisado: clasificación y regresión. Por un lado, si  $y$  es un conjunto finito de etiquetas o clases es un problema de clasificación; por otro lado, si  $y$  es un valor numérico entonces es un problema de regresión. En el problema de la identificación automática de personalidad se han realizado ambos tipos de tareas. Bajo el modelo psicológico *Big-Five* (explicado en el Capítulo 4), la personalidad está representada por 5 rasgos de personalidad con dos polos opuestos de acuerdo con un valor numérico asociado en cada rasgo.

Así, si se identifica el polo de cada rasgo de personalidad —por ejemplo si es alto, medio o bajo— se lleva a cabo una tarea de clasificación, donde las clases predefinidas son  $y = \{\text{alto}, \text{medio}, \text{bajo}\}$ . Si se predice el valor que se obtiene a través de instrumentos estándares como cuestionarios de personalidad —dependiendo del instrumento de personalidad, el valor a predecir puede estar en el conjunto de números reales en el rango  $y \in [1, 7]$  o  $y \in [1, 5]$ — entonces es una tarea de regresión.

Para el problema de la identificación de personalidad mediante instrumentos psicológicos

estandarizados, los valores numéricos obtenidos por estos cuestionarios son interpretados dentro de clases categóricas dependiendo de la población específica a la cual se adapta el instrumento y de forma independiente para cada rasgo de personalidad. Estas clases predefinidas son las que utilizaré en esta tesis para abordar el problema de personalidad como una tarea de clasificación; específicamente, como una tarea de clasificación binaria. Es decir, los valores de  $y$  son fijos y predefinidos como  $y = \{alto, bajo\}$ . En los siguientes párrafos me concentraré en presentar conceptos relacionados con la clasificación, particularmente con la clasificación automática de textos.

### 3.1. Representación

El otro componente esencial del *conjunto de entrenamiento* es la forma que toman las entradas  $x$ . Estos elementos  $x$  son llamados **ejemplos o instancias** —que en nuestro caso serán textos— y se codifican mediante un modelo de representación de datos. En clasificación automática de textos, el modelo ampliamente utilizado es conocido como el modelo de representación vectorial, donde cada instancia en el conjunto de entrenamiento es representada de la forma  $(\vec{x}, y)$ , donde  $\vec{x} \in \mathbb{R}^n$  es un vector de tamaño  $n$  y  $y$  es la etiqueta de la clase de la instancia representada. Existen dos enfoques principales para formar este vector  $\vec{x}$  para cada instancia: 1) ingeniería de atributos y 2) aprendizaje de atributos.

La **ingeniería de atributos** ha sido el enfoque tradicional en el que manualmente se decide el conjunto de atributos que formarán las dimensiones del vector, así como el esquema de pesado que cada elemento en este vector tendrá. En la forma más simple, las dimensiones del vector son palabras que aparecen en las instancias del conjunto de entrenamiento y el valor de elemento (palabra) en el vector puede asignarse con un valor booleano dependiendo de la aparición o no de la palabra en la instancia a representar; o también, con la frecuencia en que cada palabra aparece en esa instancia. A este esquema de representación se le ha llamado Bolsa de Palabras o BoW (*Bag of Words*) y los vectores correspondientes se conocen como *vectores de documento*. El tipo de palabras que serán utilizadas en la representación son seleccionadas dependiendo del problema y en el escenario ideal, con conocimiento *a priori* del mismo.

Una de las desventajas de la ingeniería de atributos es que la dimensionalidad del vector  $\vec{x}$  puede llegar a ser muy grande, ya que cada palabra presente en las instancias se trata de forma independiente de otras, aun cuando compartan el mismo contexto. Por ejemplo, una instancia que contenga solo la palabra *perro* y otra que contenga solo la palabra *gato*, tendrán vectores ortogonales aun cuando tienen contextos similares —ambos son nombres de animales domésticos—. Para mitigar esta desventaja se han propuesto representaciones contextuales, desde la utilización de  $n$ -gramas de palabras (como uno de los primeros enfoques que intenta incluir algo de contexto en la representación) hasta representaciones distribucionales. De esta idea de representaciones

distribucionales, se propone el segundo enfoque: aprendizaje de atributos.

El **aprendizaje de atributos** actualmente tiene como objetivo aprender una representación con base en el descubrimiento de patrones escondidos en los datos de entrada, que los modelos de aprendizaje desarrollados pretende desenredar (Bengio et al., 2013). Para aprender estas representaciones se pueden utilizar una variedad de redes neuronales. Al final, estas representaciones aprendidas capturan el contexto y características embebidas en las instancias de entrada, y se condensan en capas ocultas de redes que tienen como objetivo una tarea de aprendizaje supervisado.

Entre las representaciones obtenidas de esta forma, una de las más usadas y preentrenadas es word2vec (Mikolov et al., 2013). Este tipo de representaciones asignan vectores de baja dimensionalidad —pues están limitados por el número de neuronas en las capas intermedias— usualmente 100 o 300 dimensiones. De modo que los vectores de palabras semánticamente cercanas, por ejemplo *perro* y *gato*, serán cercanos en el espacio vectorial mientras existan este contexto en el conjunto de entrenamiento. Adicionalmente, con el avance de las arquitecturas profundas de redes neuronales, es posible aprender representaciones de oraciones e incluso de instancias completas.

Hay dos principales ventajas del aprendizaje de atributos, la primera es que la representación obtenida es generalizable, pues expresan aspectos inherentes al mundo cuando el conjunto de entrenamiento contiene instancias de este y no está enfocado en ninguna tarea en particular. Esta capacidad de generalización se ejemplifica muy bien con las representaciones word2vec (Mikolov et al., 2013) que se construyen con un conjunto de millones de páginas Web de uso general, como la Wikipedia, para aprender los vectores de cada término y que son utilizadas en una variedad de tareas de procesamiento de lenguaje natural (PLN) con resultados positivos.

Sin embargo, esta generalización no es necesariamente adecuada en todos los casos. Por ejemplo, el Capítulo 4 muestra sistemas de clasificación automática de personalidad que utilizan los dos enfoques de representación; mientras que los métodos que utilizan aprendizaje de atributos muestran ligeras mejoras en la clasificación de los rasgos de personalidad, esta mejora es mínima en comparación con la utilización del enfoque basado en ingeniería de atributos. Lo que lleva a plantearse la posibilidad de que a diferencia de otras tareas de PLN, donde utilizar aprendizaje de atributos muestra una evidente mejora en los resultados de clasificación, para identificación de personalidad y otras tareas asociadas a la salud mental los resultados no son contundentes.

La segunda ventaja del aprendizaje de atributos es que no es necesario un conocimiento *a priori* de la tarea de clasificación que se desea automatizar, pues este recurso requiere de mucha mano de obra y de ingenio humano. En palabras de Bengio et al. (2013) esta limitante impide que se construyan aplicaciones novedosas de forma rápida. Sin embargo, no todos los problemas

necesitan ser resueltos rápidamente; además, en el caso de la identificación de personalidad, ya existen recursos construidos por expertos que han mostrado ser útiles para tareas de clasificación automática.

### 3.2. Métricas de desempeño

Una vez que se entrenó un modelo de clasificación —se encontró la función  $h$ —, debe evaluarse para determinar qué tanto se acerca a la función real  $f$ . Esta validación se realiza en un *conjunto de evaluación* que debe ser diferente al conjunto de instancias de entrenamiento. Las métricas ampliamente utilizadas en clasificación automática son exactitud, precisión, recuerdo y  $f$ -score (Manning and Schütze, 1999).

En un problema de clasificación binario, con clases positiva y negativa; existen 4 valores que son utilizados para definir las métricas de evaluación:

- $tp$ : número de instancias que son clasificadas correctamente como positivas.
- $tn$ : número de instancias que son clasificadas correctamente como negativas.
- $fp$ : número de instancias que son clasificadas como positivas, pero que son realmente negativas. O número de instancias incorrectamente clasificadas como positivas.
- $fn$ : número de instancias que son clasificadas incorrectamente como negativas.

Así, la *precisión* es el porcentaje de predicciones correctas en la clase positiva sobre el número total de predicciones en la clase positiva.; i. e. el ratio del total de aciertos de la clase positiva y el número de instancias que el sistema clasificó como positivas (Eq. 3.1).

$$precisión = \frac{tp}{tp + fp} \quad (3.1)$$

El *recuerdo* es el porcentaje de predicciones correctas de la clase positiva sobre el número total de casos de la clase positiva.; i. e. el ratio entre el total de aciertos de la clase positiva y el número total de instancias positivas que hay en el conjunto de evaluación (Eq. 3.2).

$$recuerdo = \frac{tp}{tp + fn} \quad (3.2)$$

Una métrica que ha sido utilizada como una combinación de *precisión* y *recuerdo* es la



medida F (f-score) que se define en la Eq. 3.3.

$$f - score = \frac{2 * precisión * recuerdo}{(precisión + recuerdo)} \quad (3.3)$$

Finalmente, la exactitud es el porcentaje de instancias de ambas clases que son correctamente clasificadas (Eq. 3.4).

$$exactitud = \frac{tp + tn}{tp + fp + fn + tn} \quad (3.4)$$

Existen otras métricas de evaluación que son utilizadas cuando se realiza regresión (estas métricas se utilizan por algunos de los trabajos presentados en el Capítulo 4) donde se calcula el error del modelo al predecir el valor asociado con cada instancia. Sea  $\hat{y}$  el valor que el modelo predijo debe tomar  $y$ ; los errores principales son el error absoluto  $L_1(y, \hat{y}) = |y - \hat{y}|$  y el error cuadrado  $L_2(y, \hat{y}) = (y - \hat{y})^2$ . De aquí se definen las métricas MAE (o *mean absolute error*) que calcula el promedio de  $L_1$  para cada instancia en el conjunto de evaluación. Y por otro lado, la métrica MSE (o *mean square error*) que calcula el promedio del error  $L_2$  de todas las instancias del conjunto de evaluación.

### 3.3. Calidad de los datos

El conjunto de entrenamiento es un componente importante en la tarea de clasificación de textos; y en general, en el aprendizaje automático. A partir de los ejemplos en el conjunto de entrenamiento es que los sistemas de aprendizaje supervisado *aprenden*; por lo tanto, considerar las características de este conjunto de entrenamiento es importante para obtener sistemas automáticos eficaces. La primera característica es el número de ejemplos; pues intuitivamente, entre mayor sea el número de ejemplos disponibles hay más datos para aprender un modelo automático. Históricamente, esta ha sido la característica a la que más atención se ha puesto en el área de procesamiento de lenguaje natural (PLN). Sin embargo, obtener de forma ética un conjunto de ejemplos suficientemente grande para que los sistemas actuales aprendan es una tarea difícil, especialmente en problemas de clasificación automática dentro del área de la salud.

Ante la falta de un conjunto grande de ejemplos que formen parte del conjunto de entrenamiento, la segunda característica se vuelve más importante: la calidad de los datos. Mientras que determinar cuándo un conjunto de datos es grande depende de un valor numérico, determinar la calidad de un conjunto de datos es complejo. Por ejemplo, en la revisión de la literatura realizada por Priestley et al. (2023) la calidad de los datos tiene que ver tanto con las diferentes etapas del desarrollo de sistemas de aprendizaje automático como con 4 dimensiones de la calidad de los datos: intrínseca, contextual, representacional y de accesibilidad.

### Capítulo 3. Clasificación automática de textos

---

Las etapas que más influyen la calidad de los datos en el desarrollo de sistemas automáticos son Casos de uso de datos y diseño; y la etapa de Recolección. En estas etapas es importante considerar a humanos expertos tanto para determinar las necesidades de datos como en la etiquetación de los mismos en la etapa de recolección. Además de la documentación tanto en la etapa de planeación como en la recolección y publicación.

Así, un punto importante en la calidad de los datos se ve reflejado en un enfoque intervencionista, como lo explican Priestley et al. (2023) y Jo and Gebru (2020). El enfoque intervencionista contrasta con el enfoque de mínima supervisión de los métodos de recolección de datos que se realizan vía *web-crawling* o *crowdworking* utilizados tradicionalmente para obtener grandes cantidades de datos de forma rápida —atendiendo a la primera característica sobre el número de ejemplos mencionada antes—. El problema de este último enfoque radica en que no se evalúa el origen, motivación, plataforma o potencial impacto de los datos recolectados. En la actualidad, algunas áreas de PLN se están moviendo hacia la recolección de datos específicos para tareas particulares, dando mayor importancia a la alta calidad en los datos recolectados que al número de instancias recolectadas (Hutchinson et al., 2021).

# 4



## Personalidad y su identificación automática

Este capítulo describe el modelo *Big-Five* para representar la personalidad y presenta investigaciones relevantes que reportan métodos automáticos para la identificación de rasgos de personalidad mediante el análisis de textos. Estas investigaciones se pueden dividir en dos principales grupos: i) las que utilizan ingeniería de atributos, que en su mayoría proponen el uso de atributos con significado lingüístico; y ii) las que utilizan representaciones basadas en el enfoque de aprendizaje de atributos.

### 4.1. El modelo *Big Five* de personalidad

La personalidad define a un conjunto de patrones de comportamiento individual que incluye emociones y pensamientos. Estos patrones, también caracterizan cómo cada individuo se adapta a ciertas situaciones de la vida y se ha encontrado que estos permanecen estables a través del tiempo y en diversas situaciones (Mischel, 1993; Funder, 2001).

La comunidad científica en el área de la Psicología afirma que los modelos psicológicos basados en rasgos de personalidad son eficientes para medir aspectos de la vida (Goldberg, 1993). Si bien existen diversos modelos basados en rasgos —como el MBIT (Myers and Myers, 1995)—, el modelo dominante es el conocido como Modelo de los Cinco Grandes (o *Big Five Model*) (McCrae and Costa Jr, 1997) que consiste en cinco rasgos que se nombran por alguno de sus polos (Costa Jr and McCrae, 2008) y son:

- Apertura a Nuevas Experiencias. En uno de sus polos, la persona tiende a mostrar creatividad, imaginación, rápida comprensión, y apreciación a ideas inusuales. En el polo opuesto, la persona tiende a ser conservadora, convencional y con falta de curiosidad y de imaginación, también se considera inflexible al cambio.

## Capítulo 4. Personalidad y su identificación automática

---

- Responsabilidad. Una persona responsable tiende a mostrar autodisciplina, diligencia y logros; estas personas planean para alcanzar sus objetivos, además de mostrar un alto nivel de organización y perseverancia en actividades orientadas a tareas. En el polo contrario, una persona no responsable exhibe pereza, espontaneidad y falta de fiabilidad.
- Extroversión. En un polo, la persona extrovertida muestra interés en otras personas y en eventos sociales, un deseo de comunicar sentimientos y emociones, exhibe energía y disfrute; además de ser asertiva y activa. Contrariamente, personas introvertidas prefieren estar solas y no muestran interés en entablar comunicación.
- Amabilidad. En uno de sus polos, esta persona tiende a ser comprensiva y cooperativa; demuestra compasión, simpatía, gentileza, confianza en otras personas y calidez. En el polo contrario, este rasgo hace referencia a la desconfianza y apatía, al antagonismo y agresividad, así como al deseo de no cooperar con otras personas.
- Estabilidad Emocional. Una persona emocionalmente estable muestra autocontrol de impulsos, tranquilidad, confianza, satisfacción, y calma; además de ser capaces de afrontar el estrés. En el polo opuesto, llamado usualmente Neuroticismo, una persona tiende a experimentar emociones no placenteras como enojo, ansiedad, depresión, miedo, preocupación y cambios de ánimo repentinos.

### 4.1.1. Un modelo generalizable

Desde 1884, Galton F. (1884) exploró la idea de que el carácter humano es medible y ha reconocido que las diferencias individuales más importantes en los humanos están codificadas en términos tan simples como palabras en uno o todos los lenguajes naturales. Este concepto es lo que luego se denominó *hipótesis léxica*.

Investigadores como Goldberg (1993) y Ashton and Lee (2005) han establecido que todos los rasgos importantes del carácter humano se manifiestan a través del lenguaje. Por consiguiente, dado que el modelo de personalidad del *Big-Five* está basado en la hipótesis léxica, en esta investigación considero que es posible encontrar un conjunto de términos descriptivos de los rasgos de la personalidad a partir del análisis del lenguaje.

Ciertamente, a través de la hipótesis léxica ya se ha propuesto un modelo general de personalidad, como el del *Big-Five* que considera una lista del léxico que codifica cada uno de los cinco rasgos. Mi interés no es encontrar un nuevo modelo de personalidad; sino, en términos generales obtener evidencia de que esta codificación del carácter humano (como la personalidad) puede verse reflejada en comunicaciones cortas. Así, como un producto adicional de la investigación, se genera una lista del léxico dependiente de la población de interés que, de manera acotada, caracterizan a cada polo de cada rasgo de personalidad del modelo *Big-Five*.

### 4.2. Identificación automática de la personalidad

La arquitectura general de métodos automáticos para identificar la personalidad tienen tres partes principales: el modelo de personalidad a utilizar, la muestra del lenguaje del individuo, y una representación de la muestra del lenguaje. El modelo de personalidad que se utiliza en esta investigación es el *Big-Five* (definido en la Sección 4.1); las muestras del lenguaje son comunicaciones cortas que se definen en el Capítulo 5; y la representación de tales muestras forman vectores numéricos bajo el enfoque de ingeniería de atributos, como se definió en el Capítulo 3.

El objetivo de esta arquitectura es generar una función que, como se explicó en el Capítulo 3, mapea vectores de muestras o instancias de lenguaje a un valor categórico o numérico. Si el valor es categórico se habla de clasificación mientras que si el valor es numérico se habla de una tarea de regresión. La función se evalúa con los valores de personalidad obtenidos mediante instrumentos psicológicos (cuestionarios utilizados en el área psicológica, diseñados y adaptados a la población objetivo, para la identificación de los rasgos de personalidad del *Big-Five*) y a estos valores se les conoce como *gold-standard*.

A continuación se describe el estado del arte para la identificación automática de personalidad agrupado de acuerdo al tipo de representación generada de las muestras del lenguaje: representaciones basadas en atributos o elementos lingüísticos (enfoque de ingeniería de atributos), y representaciones basadas en arquitecturas de aprendizaje profundo (enfoque de aprendizaje de atributos).

#### 4.2.1. Representaciones basadas en atributos lingüísticos

El lenguaje es un buen indicador de la personalidad pues a través de él se puede expresar la forma de pensar o sentir (Pennebaker, 2011). Es por esto que los primeros intentos de identificación automática de personalidad a través del lenguaje tiene raíces en encontrar correlaciones entre la personalidad y atributos lingüísticos previamente identificados en campos de la psicolingüística.

Uno de los primeros trabajos que aparecieron sobre el reconocimiento de la personalidad a través del texto se basó enteramente en la identificación de tipos de palabras utilizadas por los individuos. Por ejemplo, Argamon et al. (2005) realizaron experimentos con ensayos escritos por estudiantes (corpus referido como *Essays*<sup>1</sup>) para determinar la identificación de dos rasgos: *Extraversión* y *Neurotismo*; en este trabajo, categorizaron las palabras usadas en cuatro categorías: palabras de función (como artículos y preposiciones), palabras de cohesión (como pronombres), palabras de valoración (términos que evalúan el contexto como aquellas palabras usadas para

---

<sup>1</sup>La descripción del corpus *Essays* se presenta en la Sección 5.2.2 del Capítulo 5.

## Capítulo 4. Personalidad y su identificación automática

---

validar, desear, aprobar, etc.), y palabras de evaluación (términos que se usan para expresar la actitud del escritor sobre lo que escribe).

Por otro lado, Mairesse et al. (2007) propusieron identificar los cinco rasgos del *Big Five* utilizando 88 categorías de palabras de la herramienta LIWC (*Linguistic Inquiry and Word Count*) (Pennebaker et al., 2001) —como palabras de enojo, asuntos metafísicos, procesos sociales, entre otras—, más un conjunto de 14 atributos tomados de la base de datos psicolingüística MRC (Coltheart, 1981) que contiene estadísticas de más de 150,000 palabras —como frecuencia de uso, sílabas por palabra, o más complejas como edad de adquisición, familiaridad, entre otras—. Los autores exploraron dos enfoques para identificar la personalidad de los participantes: clasificación binaria (indicando, por cada rasgo, el polo al que cada persona pertenece) y regresión (prediciendo el valor numérico dada por la prueba de personalidad o *gold-standard*). En la fase de evaluación de la tarea de clasificación se obtuvieron resultados estadísticamente significativos en la detección automática de la personalidad que van de una exactitud de 52 % a 62 % en los cinco rasgos. Sin embargo, los resultados para la tarea de regresión, medido por el Error Absoluto Relativo —el radio entre el error absoluto del método base y el error absoluto del modelo a evaluar— no arrojaron resultados favorecedores, como muestra la Tabla 4.1.

En una dirección diferente al uso de vocabulario preidentificados por lingüistas o psicólogos, surgió la línea de investigación basada en vocabulario abierto. Esta línea se ha centrado en analizar la relación de unidades textuales, ya sea el conjunto de palabras independientes o las secuencias de estas (*n*-gramas de palabras), con la personalidad. El enfoque basado en vocabulario abierto se justifica dada la constante evolución del lenguaje; por lo tanto, no limita la representación de las muestras de lenguaje a palabras o categorías preestablecidas. Al mismo tiempo, este enfoque intenta incorporar, en cierta medida, la dependencia de las palabras utilizadas en el texto o el contexto en el que estas muestras de lenguaje se obtienen o producen.

Por ejemplo, Oberlander and Nowson (2006) evitan el uso de recursos lingüísticos y deciden hacer una representación de la muestra de lenguaje que obtienen mediante texto en forma de *n*-gramas de palabras de longitud  $n = 2$  y  $n = 3$ . A pesar de que el conjunto de individuos que formaron el conjunto experimental fue pequeño —de solamente 71 sujetos— la exactitud de su método es alta: entre 70 % y 90 % de exactitud, dependiendo del rasgo detectado.

Otros estudios que siguen el enfoque de vocabulario abierto son, por un lado, el estudio de Iacobelli et al. (2011) donde utilizan *n*-gramas de palabras para la representación de entradas de blog de 3000 participantes y que mediante un algoritmo de máquina de vectores de soporte (SVM) obtienen desempeños entre 70 % y 84 % de exactitud. Por otro lado, está el trabajo realizado por Iacobelli and Culotta (2013), donde usan una combinación de *n*-gramas de palabras con *n* de 1 a 3 para entrenar clasificadores con el algoritmo de Regresión Logística utilizando dos de los

## 4.2 Identificación automática de la personalidad

---

conjuntos de datos más usados hasta el momento: *Essays* y *myPersonality*<sup>2</sup>. En este segundo trabajo, los autores obtienen entre 53 % y 61 % de exactitud para el conjunto de datos de *Essays* y una exactitud entre 51 % y 66 % para *myPersonality*.

Las investigaciones anteriores establecieron las bases de la identificación automática de la personalidad en muestras de lenguaje, específicamente muestras de textos. Cabe mencionar que la mayoría de los trabajos en esta sección usan clasificación —particularmente la clasificación binaria— y como métrica de desempeño la exactitud. Sin embargo, la comparación directa de los resultados reportados por cada estudio no es posible debido a los diferentes corpora utilizados.

Hay tres tipos de fuentes de donde se obtienen los datos usados en las investigaciones previas: i) *Essays* que es un corpus de flujos de pensamientos que tenían los participantes por aproximadamente 20 minutos; ii) los dos corpora de Blogs (que son creados específicamente por cada investigación) son textos que tienen un propósito particular, ya sea temático o de estilo; y que por lo regular pasan por un proceso de edición. iii) Y, *myPersonality*, que es un corpus de mensajes o publicaciones de Facebook de más de 115 mil usuarios, con más de 1 millón de entradas o *posts*; este corpus al contener muestras de lenguaje publicadas en redes sociales, tienden a ser textos cortos, con jerga particular para este tipo de mensajes y suelen ser espontáneos. De esta revisión, una conclusión preliminar es que no existe una variedad de conjuntos de datos disponibles para entrenar métodos automáticos para la identificación de rasgos de personalidad.

Sin embargo y a pesar de lo heterogéneo de los estudios, los atributos usados en la representación de las muestras de lenguaje en cada uno son similares. Si comparamos los estudios que realizan clasificación y usan *Essays* podemos notar que los atributos lingüísticos de vocabulario cerrado tuvieron mejor desempeño que los atributos de vocabulario abierto. Adicionalmente, de las investigaciones que utilizan diccionarios como LIWC, se han encontrado correlaciones entre las categorías de palabras que contiene este diccionario con algunos rasgos de personalidad. Por ejemplo, dependiendo del origen de los textos (ensayos o conversaciones) diferentes categorías de LIWC son relevantes (Mairesse et al., 2007). En cuanto al vocabulario abierto, dependiendo del estudio se pudieron identificar  $n$ -gramas que surgieron y que no se encuentran en algún diccionario como LIWC que son importantes para algún rasgo.

---

<sup>2</sup>El corpus *myPersonality* fue utilizado inicialmente en el taller científico organizado por Celli et al. (2013).

**Tabla 4.1:** Resultados de la evaluación de los métodos en el estado del arte para la identificación automática de rasgos de personalidad. *R* y *C* en la columna *Tarea* representan Regresión y Clasificación, respectivamente. Bajo la columna *Métrica*, *Ex.* indica exactitud.

Referencia	Corpus	Representación-Modelo	Tarea	Métrica	Rasgo				
					Ext.	Ama.	Resp.	Ape.	
Mairesse et al. (2007)	Essays	LIWC+MRC-LR	R	RAE	0.99	0.98	0.98	0.96	0.93
Mairesse et al. (2007)	Essays	LIWC-SMO	C	Ex.	0.52	0.56	0.55	0.58	0.62
Oberlander and Nowson (2006)	Blogs	word n-grams-NB	C	Ex.	0.74	0.81	0.77	0.63	-
Iacobelli et al. (2011)	Blogs	word n-grams-SVM	C	Ex.	0.71	0.78	0.79	0.7	0.84
Iacobelli and Culotta (2013)	Essays	word n-grams-LR	C	Ex.	0.56	0.53	0.55	0.57	0.61
Iacobelli and Culotta (2013)	myPersonality	word n-grams-LR	C	Ex.	0.6	0.51	0.52	0.52	0.66
Leonardi et al. (2020)	myPersonality sm.	BERT-Stack	R	MSE	0.47	0.26	0.3	0.29	0.17
Xue et al. (2018)	myPersonality	Attention+CNN-SVR	R	MAE	0.47	0.38	0.42	0.49	0.35
Kampman et al. (2018)	YouTube	word2vec - CNN	R	MAE	0.11	0.09	0.12	0.11	0.11
Sun et al. (2018)	YouTube	GloVe-BiLSTM+CNN	C	macro-P	0.67	0.58	0.61	0.61	0.55
Kampman et al. (2018)	YouTube	word2vec - CNN	C	Ex.	0.88	0.9	0.87	0.88	0.88
Darlansyah et al. (2019)	myPersonality	GloVe - CNN+LSTM	C	Ex.	0.58	0.54	0.57	0.62	0.72
Majumder et al. (2017)	Essays	word2vec+CNN+Ling-MLP/SVM	C	Ex.	0.58	0.56	0.57	0.59	0.62
Sun et al. (2018)	Essays	pre-train GloVe-BiLSTM+CNN	C	macro-P	0.55	0.58	0.53	0.56	0.54
Jiang et al. (2020)	Essays	RoBERTa-FFN	C	Ex.	0.6	0.59	0.58	0.61	0.65
Kez et al. (2022)	Essays	LingEmbedding+BERT-MFFN	C	Ex.	0.63	0.6	0.61	0.6	0.71



### 4.2.2. Representaciones basadas en aprendizaje profundo

En años recientes, después del desarrollo de enfoques de representación de textos mediante aprendizaje de atributos con redes neuronales, la investigación se ha centrado en mejorar el desempeño de modelos automáticos de clasificación. De este conjunto de investigaciones se derivan dos líneas dependiendo de la fuente de datos utilizada. Por un lado están los trabajos que usan datos provenientes de redes sociales; por otro lado están los trabajos que han utilizado ensayos, particularmente el corpus *Essays*.

#### Métodos que estudian datos de redes sociales

Los enfoques de la primera línea de investigación se han centrado en utilizar textos provenientes de redes sociales para entrenar y evaluar sus modelos, esto se debe al fácil y libre acceso que se tiene sobre esos textos<sup>3</sup>. Lo que ha permitido obtener una gran cantidad de ejemplos que, además, son requeridos por modelos de aprendizaje profundo. Una segunda razón es que se asume que gracias al anonimato que proporcionan las redes sociales (particularmente Twitter<sup>4</sup> y Reddit<sup>5</sup>) los usuarios de ellas pueden expresarse tal cual son (Boyd and Pennebaker, 2015).

Entre las características principales de estas investigaciones es que usan textos en inglés y métricas de evaluación distintas entre un trabajo y otro, esto último agrega dificultad a la comparación directa entre diferentes modelos y propuestas. La desventaja de que solo utilizan corpora con textos en inglés no ayuda en la reutilización de modelos costosos de entrenar —modelos basados en aprendizaje profundo— para otros lenguajes, lo que complica la evaluación de estos métodos en corpora en diferentes idiomas.

Igual que con los métodos que utilizan representaciones basadas en recursos lingüísticos (ver Sección 4.2.1), la mayoría de los trabajos presentados aquí realizan la tarea de clasificación en lugar de la tarea de regresión. Entre la minoría que realizaron regresión están la investigación de Xue et al. (2018) quienes combinan *embeddings* pre-entrenados de word2vec (Mikolov et al., 2013) con estadísticas de algunas categorías de LIWC. Luego, los autores utilizan una red neuronal convolucional (CNN) con atención para representar oraciones; posteriormente usan otra CNN para codificar el documento. Tanto el entrenamiento como la evaluación se realizaron en el corpus de Facebook *myPersonality* y la medida de desempeño usada fue el Error Absoluto Medio (o MAE por sus siglas en inglés) para el que el valor de 0 es mejor. Su mejor resultado fue un valor de MAE de 0.35 y 0.38 para *Apertura a Nuevas Experiencias* y *Amabilidad*, respectivamente; por el contrario, el rasgo con el mayor error fue *Neuroticismo* con un MAE de 0.48.

---

<sup>3</sup>En la Sección 9.2 se mencionan las consideraciones éticas que deben tomarse en cuenta al utilizar datos de redes sociales.

<sup>4</sup>Actualmente llamada X <https://x.com/>.

<sup>5</sup><https://www.reddit.com/>.

## Capítulo 4. Personalidad y su identificación automática

---

Otro estudio que usa el corpus *myPersonality* es el realizado por Leonardi et al. (2020) quienes utilizan una arquitectura *transformer* para generar *embeddings* de oraciones o frases que usa como entrada para una red neuronal apilada. En este trabajo, los autores reportan como medida de desempeño el Error Cuadrado Medio (o MSE por sus siglas en inglés) donde 0 es el mejor valor esperado y 1 el peor. Su mejor resultado fue obtenido para el rasgo de *Apertura a Nuevas Experiencias* con un MSE de 0.17; el rasgo más difícil fue *Extroversión* con un valor de MSE de 0.47.

Por otro lado, Kampman et al. (2018) utilizan un corpus de YouTube con 8000 participantes que consiste en clips de 15 segundos. Aunque en el estudio completo realizan un método que combinan tres modalidades —audio, video, y texto— los resultados que discuto aquí son referentes a la modalidad de texto únicamente. En este trabajo, los autores realizan una tarea de regresión bajo una representación de *embeddings* pre-entrenados word2vec; luego, para generar el valor de personalidad de cada rasgo hacen uso de una CNN. El desempeño fue evaluado con el Error Absoluto Medio (MAE) donde obtuvieron errores entre 0.12 a 0.09, lo cual es relativamente bajo comparado con el error obtenido en el corpus de *myPersonality*.

En la tarea de clasificación Sun et al. (2018) también utilizan datos de YouTube bajo una representación de *embeddings* pre-entrenados GloVe en una red bidireccional de memoria a largo y corto plazo (BiLSTM) concatenada a una arquitectura CNN. En este caso, los autores reportaron como medida de desempeño la macro-precisión, donde 1 es mejor. Los mejores resultados fueron obtenidos para el rasgo de *Extroversión* con una macro-precisión de 0.67 y el peor rasgo a clasificar fue en rasgo de *Apertura a Nuevas Experiencias* con una macro-precisión de 0.55.

Mientras tanto, Kampman et al. (2018), también en una tarea de clasificación con datos de YouTube, utilizan *embeddings* pre-entrenados word2vec en una CNN. Estos autores reportan el desempeño en términos de la exactitud con valores entre 0.88 y 0.90. Cabe mencionar que de estos dos últimos trabajos, aunque ambos usan datos de YouTube, los datos no son los mismos.

Un último ejemplo del uso de redes sociales para identificar rasgos de personalidad es el trabajo realizado por Darliansyah et al. (2019), quienes utilizan el corpus *myPersonality*. En esta investigación, los autores usan *embeddings* pre-entrenados GloVe como entrada en una CNN más una red neuronal recurrente LSTM. El desempeño también es medido con exactitud donde obtienen resultados de 0.72 para *Apertura a Nuevas Experiencias* y 0.54 para el rasgo *Amabilidad*.

Para este grupo de investigaciones, es difícil realizar comparaciones directas entre una propuesta y otra en función de las métricas de desempeño que presentan. Pero si comparamos entre dos estudios que realizan la tarea de clasificación, que usan *myPersonality*, y que reportan exactitud —entre el estudio realizado por Iacobelli and Culotta (2013) (método basado en vocabulario abierto) y el realizado por Darliansyah et al. (2019)— es posible notar que el método propuesto por Darliansyah et al. (2019) tiene mejor exactitud para 4 de los 5 rasgos (el rasgo en el que un

## 4.2 Identificación automática de la personalidad

---

enfoque más simple es mejor fue *Extroversión*). Mientras que las diferencias entre los otros cuatro rasgos varían entre 0.03 y 0.10. Estos resultados permiten inferir que utilizar representaciones más complejas —como los *embeddings* en lugar de *n*-gramas de palabras— cuando no se usan recursos lingüísticos predefinidos, ayuda a mejorar el desempeño de clasificación, aunque no por mucho y no en todos los rasgos de manera contundente.

### Métodos que estudian datos de ensayos

El segundo grupo de investigaciones utiliza datos de ensayos para entrenar modelos para la identificación de rasgos de personalidad. Majumder et al. (2017) fueron de los primeros en proponer el uso de arquitecturas de redes neuronales para identificación de rasgos de personalidad. Su red consta de una representación a base de *embeddings* pre-entrenados word2vec y una CNN para generar la representación de las muestras de lenguaje (ensayos). A este vector agregado de los textos concatenaron un vector que contiene los atributos propuestos por Mairesse et al. (2007). Con esta propuesta los autores logran un desempeño basado en exactitud de entre 0.56 (obtenido para *Amabilidad*) y 0.62 (obtenido para el rasgo de *Apertura a Nuevas Experiencias*). A pesar de la arquitectura mucho más robusta, Majumder et al. (2017) consiguieron mejorar el desempeño de solo 3 de los 5 rasgos entre 0.02 y el 0.06 en el valor de exactitud, en relación a la propuesta de Mairesse et al. (2007).

Así mismo, Sun et al. (2018) con el mismo método previamente descrito en la Sección 4.2.2 —*embeddings* pre-entrenados GloVe y la concatenación de dos arquitecturas de redes neuronales: BiLSMT más CNN— pero con el corpus de *Essays*, reportan un desempeño basado en macro-precisión con resultados entre 0.53 y 0.58.

En años más recientes, se han explorado representaciones basadas en *transformers* en lugar de *embeddings* pre-entrenados de palabras como word2vec y GloVe. Por ejemplo, en el trabajo de Jiang et al. (2020) experimentaron con los modelos de BERT y RoBERTa para representar cada ensayo. Reportando una exactitud de 0.65 para *Apertura a Nuevas Experiencias* —su mejor resultado— y una exactitud de 0.58 para el rasgo de *Responsabilidad*.

Finalmente, Kerz et al. (2022) presentan una representación combinada entre *embeddings* de oraciones hechas con BERT más una construcción de vectores basados en atributos lingüísticos. La clasificación la llevaron a cabo con una red neuronal multicapa con retroalimentación hacia adelante (o *multi-layer feed-forward neural network*). En este trabajo, los autores reportan el mejor desempeño en el rasgo de *Apertura a Nuevas Experiencias* (con una exactitud de 0.71) y el peor desempeño para los rasgos de *Amabilidad* y *Neuroticismo* (con 0.60 de exactitud).

### **4.2.3. Análisis de las representaciones usadas para la identificación de rasgos de personalidad**

Este capítulo presentó trabajos que basan sus representaciones en *embeddings* o vectores generados con redes neuronales profundas (ya sea pre-entrenados o mediante el uso de *transformers*); métodos que también hacen uso de arquitecturas complejas basadas en redes neuronales para la identificación de la personalidad a partir de textos.

De los métodos comparables —los que usan el corpus *Essays* y que reportan resultados basados en exactitud— el mejor método es el propuesto por Kerz et al. (2022) y seguido por el propuesto por Jiang et al. (2020); ambos utilizan un método de clasificación basado en aprendizaje profundo y representaciones basada en *transformers*. La diferencia más notable entre estos dos métodos es el uso de *embeddings* de atributos lingüísticos por el primero. De esta forma, es posible intuir que la información lingüística incluida en la representación sigue siendo relevante para identificar personalidad en textos. Esto es más evidente al observar la información presentada en la Tabla 4.1 si comparamos el mejor método de identificación de personalidad presentada en este capítulo (último renglón en la tabla) contra el método más simple propuesto en 2007 por Mairesse et al. (2007) (segundo renglón en la tabla). La ganancia del método de Kerz et al. (2022) sobre el de Mairesse et al. (2007) es de 0.11 para el rasgo de *Extroversión* y solo aumenta el 0.02 de exactitud para el rasgo de *Neuroticismo*.

De lo anterior, la tarea de identificación de personalidad es una tarea complicada y lejos de estar resuelta. Los métodos que hacen uso de tecnologías recientes y complejas, y mejoras que se han dado dentro del área de procesamiento de lenguaje natural en la construcción de modelos grandes y robustos —útiles en otras tareas de PLN—, no tienen el mismo impacto en el contexto de identificar la personalidad mediante textos; esto se puede deber a dos factores.

Por un lado, los algoritmos basados en redes neuronales profundas requieren de cantidades muy grandes de información; para el caso de la identificación de personalidad y otros problemas dentro de la psicología humana conseguir datos es una actividad compleja, que no solo tiene implicaciones técnicas sino éticas. Si bien se han hecho intentos con datos de redes sociales, conseguir el *gold-standard* confiable es un problema documentado (Chancellor and De Choudhury, 2020; Harrigan et al., 2021). Por otro lado, la tarea que se trata de resolver, identificación de rasgos de personalidad, es compleja *per se*. Psicólogos especialistas utilizan varias sesiones, diversos instrumentos de evaluación y muchos años de entrenamiento para realizarla de forma adecuada.

A pesar de lo anterior, esta revisión del estado del arte arroja dos conclusiones: primero, aunque el desempeño de los métodos revisados no es generalmente alto, se ha demostrado que es posible encontrar información relevante en los textos que ayudan a observar indicios de la personalidad.

## **4.2 Identificación automática de la personalidad**

---

Segundo, la información lingüística generada o compilada por lingüistas o psicólogos es relevante para este problema.



A horizontal decorative bar with a repeating purple pattern of overlapping, stylized wave or shell shapes. The text 'Parte III' is centered within this bar in a white, serif font.

**Parte III**

**Contribuciones**





# 5

## Colecciones de datos

En capítulos previos menciono la existencia de diversos conjuntos de datos utilizados para el análisis automático de rasgos de personalidad a partir de muestras del lenguaje, específicamente a partir de texto o audio transcrito. Aunque existen recursos de diferentes orígenes, como redes sociales, entrevistas semi-estructuradas, ensayos, autopresentaciones, etc., la mayoría de estos recursos son en inglés provenientes en su mayoría de habitantes de Estados Unidos o Europa. Esto es una limitación al estudiar personalidad en otras poblaciones y en otros lenguajes. Por ello, con el objetivo de contar con un recurso específico para el estudio y análisis de la personalidad en textos en español y particularmente para la población de México, en esta investigación se creó el conjunto de datos *RxPI*.

**Este capítulo describe a detalle la creación del conjunto de datos RxPI, como la primera contribución de este trabajo de investigación.** También describe dos conjuntos de datos adicionales que se utilizan en la fase de experimentación de esta tesis. Un conjunto de datos en español: *Auto-descripciones*. El segundo conjunto de datos es disponible en inglés: *Essays*. Estos tres conjuntos de datos contienen muestras de lenguaje asociadas a la personalidad de quien las produjo.

### 5.1. RxPI—un recurso para el análisis de la personalidad

RxPI (viene de las siglas en inglés *Resource for Personality Identification* y puede leerse como /ˈɪkspɛɪ/) es un conjunto de ensayos escritos en español mexicano construido para el análisis automático de la personalidad. RxPI está etiquetado con un valor numérico de personalidad y clases nominales para los rasgos del modelo de personalidad *Big-Five*.

RxPI fue concebido para el estudio de rasgos de personalidad, por lo tanto, para su obtención

## Capítulo 5. Colecciones de datos

---

se usó una instrucción inspirada en la entrevista de historia de vida (o *life story interview*) (Atkinson, 2007). Este tipo de instrucciones ha sido utilizada para el análisis de otras condiciones relacionadas con la salud mental, como en el estudio de Tackman et al. (2019) sobre depresión.

Las siguientes secciones presentan la motivación, composición, proceso de recolección, pre-procesamiento, limpieza, etiquetado, usos, distribución y mantenimiento del conjunto de datos RxPI. Estas secciones siguen la estructura de la ficha técnica para conjuntos de datos propuesta por Gebru et al. (2021) como una forma de estandarizar la información sobre los datos utilizados en la comunidad de aprendizaje automático. Esta ficha documenta la creación y uso del conjunto de datos con la finalidad de transparentar el proceso de recolección y así influir en la mitigación de los sesgos sociales no deseados, entre otros beneficios.

### 5.1.1. Motivación

#### **Propósito (*For what purpose was the dataset created?*)**

El conjunto de datos RxPI fue creado para la investigación sobre identificación automática de rasgos de personalidad en una población con español como lengua materna, particularmente para el español mexicano. El objetivo de su creación es usarlo en la generación de análisis y/o métodos automáticos para la identificación del polo —alto o bajo—, o el valor numérico de cada uno de los cinco rasgos de personalidad del modelo *Big-Five*.

#### **Creadores (*Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?*)**

El conjunto de datos RxPI fue creado por tres investigadores pertenecientes al grupo de investigación Lenguaje y Razonamiento de la Universidad Autónoma Metropolitana unidad Cuajimalpa (UAM–C) en la Ciudad de México. Los investigadores son Gabriela Ramírez de la Rosa, Esaú Villatoro Tello<sup>1</sup>, y Héctor Jiménez Salazar.

#### **Financiamiento (*Who funded the creation of the dataset?*)**

El financiamiento para la recolección, transcripción y etiquetación provino de 2 fuentes: i) Consejo Nacional de Ciencia y Tecnología (CONACyT): con el apoyo al proyecto número 258588 y a la Red Temática de Tecnologías del Lenguaje con número de proyecto 281795; y ii) Universidad Autónoma Metropolitana Unidad Cuajimalpa (UAM–C): a través del presupuesto del Departamento de Tecnologías de la Información.

---

<sup>1</sup>Durante la creación del conjunto de datos Villatoro Tello estaba afiliado a la UAM, actualmente se encuentra afiliado al Idiap Research Institute.

### 5.1.2. Composición

#### **¿Qué representan las instancias? (*What do the instances that comprise the dataset represent?*)**

Las instancias en el conjunto de datos RxPI son documentos de texto simple escritos por estudiantes de la UAM-C. Estos documentos son transcripciones manuales de ensayos escritos a mano en una hoja de papel. Cada ensayo (instancia) tiene como instrucción detonante: “Cuéntanos sobre ti, por ejemplo algo sobre tu historia familiar o el(los) evento(s) que te venga a la mente que consideres relevante(s) en tu vida”. Y cada instancia de texto está asociado a la personalidad de cada autor de acuerdo al modelo *Big-Five*. Por cada uno de los cinco rasgos en este modelo, la instancia contiene información sobre el valor numérico y la clase nominal arrojados por el instrumento TIPI (*Ten Item Personality Measure*) (Gosling et al., 2003) en español (Renau et al., 2013) utilizado para obtener la personalidad autoreportada de cada participante. La Figura 5.1 muestra un ejemplo de una instancia etiquetada con extroversión baja —o introversión— y el valor numérico de 2.5 en el rasgo de *Extroversión*.

#### **Tamaño (*How many instances are there in total (of each type, if appropriate)?*)**

El corpus RxPI contiene 1484 instancias, cada instancia tiene asociados valores para cada uno de los rasgos de personalidad del modelo *Big-Five*. La Tabla 5.1 muestra el número de instancias por rasgo dependiendo de la clase nominal calculada de acuerdo a las normas del instrumento de personalidad TIPI. Y la Figura 5.2 muestra la distribución numérica de cada rasgo de personalidad obtenido directamente del instrumento TIPI; este número está dentro del rango [1, 7] (en la Sección 5.1.3, en el apartado *Información de cada instancia* se pueden ver más detalles sobre el instrumento TIPI).

#### **Población muestra (*Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?*)**

El conjunto de datos contiene instancias de texto producidas por un subconjunto de estudiantes de primer trimestre de la Universidad Autónoma Metropolitana Unidad Cuajimalpa que ingresaron en el semestre de otoño de los años 2016 al 2018. La muestra consistió de personas inscritas en algún programa de estudio que ofrece la UAM-C que decidieron participar en el estudio. Se trató de incluir todas las disciplinas que se imparten en la UAM-C y el mayor grupo de estudiantes durante las primeras semanas del primer trimestre. La Tabla 5.2 muestra información referente al número de participantes por cada uno de los 11 programas académicos que ofrecía la UAM-C durante los años de recolección. No se realizaron pruebas para determinar la representación de la población del alumnado de la UAM-C.

## Capítulo 5. Colecciones de datos

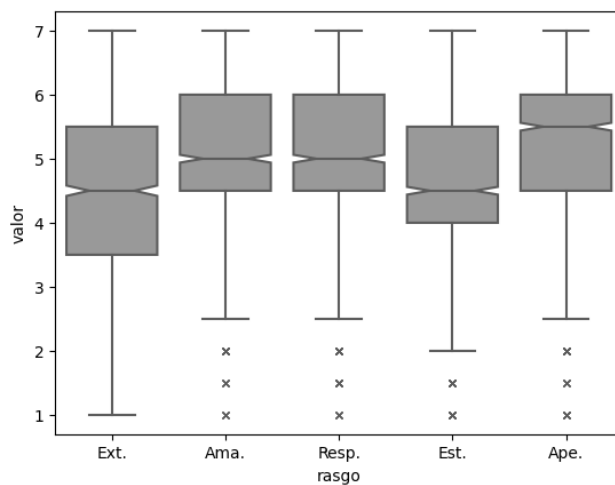
**Figura 5.1:** Un ejemplo de una instancia (con id: b18-0123) del corpus RxPI con extroversión baja —o introversión— y con un valor numérico 2.5 para el rasgo de Extroversión.

```

1 Mi familia consta de mi madre, hermana, abuela y prima y la considero asi <F0:asi> porque son las personas que
  convivi <F0:convivi> toda <MD> mi vida
2
3 Siempre me a gustado el futbol <F0:fútbol> y lo practico desde los 7 años, muchas veces en torneos escolares, y
  de otros lugares. Se liderar un equipo pero no es algo que prefiero
4
5 Soy callado cuando me encuentro con alguien que no conozco sin embargo una vez que siento confianza soy una
  personas divertida, amigable, confiable.
6
7 Me gusta viajar y mi lugar favorito es guanajuato <F0:Guanajuato> que fue uno de los lugares que he visitado
  con amigos y familia.
  
```

Rasgo	Bajo	Medio-Bajo	Medio-Alto	Alto	Total
Extroversión	156	532	557	239	1484
Amabilidad	241	532	485	226	1484
Responsabilidad	360	471	583	70	1484
Estabilidad Emocional	237	508	594	145	1484
Apertura a Nuevas Experiencias	207	418	500	359	1484
<b>Total</b>	<b>1201</b>	<b>2461</b>	<b>2719</b>	<b>1039</b>	

**Tabla 5.1:** Número de instancias de cada rasgo de personalidad para cada clase del corpus RxPI.



**Figura 5.2:** Distribución de las instancias en el corpus RxPI de acuerdo al valor numérico de cada uno de los cinco rasgos del *Big-Five*.

### Composición de una instancia (*What data does each instance consist of?*)

Cada instancia consiste de un documento de texto plano asociado con 10 valores: 5 valores reales con rango [1, 7], y 5 valores nominales:  $clases = \{bajo, medio\ bajo, medio\ alto, alto\}$ . Cada valor corresponde a cada uno de los 5 rasgos de personalidad del modelo psicológico *Big-Five*. Formalmente cada instancia está compuesta por una tripleta:  $I = \langle t, V, C \rangle$ , donde  $t$  es texto,  $V$  es un conjunto de 5 valores numéricos (uno por cada rasgo de personalidad) y  $C$  es un conjunto de 5 valores nominales. La Tabla 5.3 muestra los valores asociados para los vectores  $V$  y  $C$  de la tripleta  $I$  de nueve instancias en RxPI.

## 5.1 RxPI—un recurso para el análisis de la personalidad

Programa académico	Género			Total
	Hombre	Mujer	Sin info.	
Administración	46	55		101
Humanidades	44	78		122
Estudios Socio-Territoriales	35	28		63
Derecho	23	30		53
Ciencias de la Comunicación	112	104	1	217
Diseño	78	114		192
Tecnologías y Sistemas de Información	117	42	1	160
Ingeniería en Computación	141	34		175
Matemáticas Aplicadas	46	20	1	67
Ingeniería Biológica	67	86		153
Biología Molecular	81	100		181
<b>Total</b>	<b>790</b>	<b>691</b>	<b>3</b>	<b>1484</b>

**Tabla 5.2:** Número de participantes que aportaron ensayos para el corpus RxPI por programa académico de la UAM-C y género reportado.

Instancia ID	Valores numéricos (V)					Valores nominales (C)				
	Ext.	Ama.	Resp.	Est.	Ape.	Ext.	Ama.	Resp.	Est.	Ape.
a16-0020	6.0	4.5	4.5	1.0	7.0	a	mb	mb	b	a
a16-0021	6.0	5.5	5.0	4.5	5.5	a	ma	mb	mb	ma
a16-0022	5.0	5.5	4.5	5.5	6.0	ma	ma	mb	ma	ma
b18-0123	2.5	5.0	6.5	5.5	5.0	b	mb	ma	ma	mb
b18-0124	2.5	6.0	4.0	4.5	6.0	b	ma	b	mb	ma
b18-0125	4.0	6.5	7.0	6.5	6.5	mb	a	a	a	a
i17-0208	2.5	6.0	6.0	4.0	6.5	b	ma	ma	mb	a
i17-0209	5.0	5.5	4.5	5.0	5.0	ma	ma	mb	ma	mb
i17-0210	4.0	4.5	4.5	4.5	6.0	mb	mb	mb	mb	ma

**Tabla 5.3:** Valores asociados a nueve instancias del corpus RxPI. Cinco valores numéricos y cinco valores nominales, obtenidos a través del instrumento TIPI, correspondientes a cada uno de los 5 rasgos de personalidad del *Big-Five*. En los valores nominales, *a* indica alto, *b* indica bajo, *ma* y *mb* indican medio-alto y medio-bajo, respectivamente.

Los documentos de textos son transcripciones manuales de texto escrito a mano en una hoja de papel. La transcripción fue realizada por una persona y posteriormente revisada por otra distinta. Dentro de la transcripción propusimos representar la forma en que usualmente se escribe sobre una hoja de papel; es decir, respetando espacios grandes, dejando varios renglones entre párrafos, etc. Conocer este tipo de fenómenos puede usarse como elementos adicionales al texto para explorar la relación de su uso con los rasgos de personalidad. Hacer este análisis con RxPI es uno de los trabajos a explorar en el futuro. Así, se introdujeron manualmente 7 etiquetas especiales con la finalidad de indicar fenómenos de la escritura a mano (la Figura 5.1 muestra algunos ejemplos), tales como:

- <FO:palabra-correcta>: Errores ortográficos con la palabra corregida dentro de la etiqueta

## Capítulo 5. Colecciones de datos

---

y adelante de la palabra a corregir. Las palabras se corrigieron dependiendo del contexto y a juicio de quien transcribió. Por ejemplo, *cuervo* <FO:cuervo>.

- <D:descripción>: Para indicar la inserción de un *dibujo* o gráfico dentro del ensayo. Las descripciones pueden ser multipalabra.
- <IN>: Para indicar *inserciones* de una letra o palabra por arriba o por abajo de lo ya escrito. Estas inserciones usualmente se marcan en la escritura a mano por una flecha, o símbolos <sup>∨</sup> o <sup>^</sup>. Las inserciones se transcriben en el lugar donde se puso la marca en la hoja de papel. Por ejemplo, *Anoche vi la televisión un programa* <IN:que se llama el calabozo de ambrocio>y me gustó.
- <MD>: Para indicar que una palabra ha sido *modificada* en alguna letra o letras sin eliminar lo que se quiere modificar. Por ejemplo, se escribe “Univercidad” y sobre la “c” se escribe una “s”, entonces se mantiene la corrección en la transcripción, pero se inserta adelante la etiqueta; es decir, *Universidad* <MD>.
- <DL>: Para indicar palabras tachadas o rayadas que indican que quisieron *borrarse* del texto. Por ejemplo, *La* <DL>*cabaña*, indica que cuando se escribió la palabra *cabaña* se había escrito algo más antes de que el autor cambiara de parecer.
- <NS>: Para indicar que las dos palabras que aparecen antes y después de esta etiqueta se escribieron juntas en la hoja de papel. Por ejemplo, para transcribir los caracteres *yano* y quien transcribe considera que de acuerdo al contexto son dos palabras, entonces se transcribe como *ya* <NS>*no*.
- <SB>: Para indicar que la palabra ha sido dividida en dos partes por un guion —*silabación*— al final de un renglón. La etiqueta se escribe adelante de la palabra separada. Por ejemplo en el texto se encuentra *pala-* al final de un renglón y en el inicio de siguiente renglón aparece *bra*, entonces en la transcripción se debe escribir *palabra* <SB>.

### Clases/Objetivo (*Is there a label or target associated with each instance?*)

Las clases son derivadas de las normas del instrumento TIPI (Gosling et al., 2003) utilizado para obtener la personalidad de cada participante. Estas normas son independientes para cada rasgo de personalidad y establecen qué rango de valores numéricos corresponden a cada una de las *clases* = {*bajo*, *medio bajo*, *medio alto*, *alto*}.

El conjunto de datos también tiene asociado los valores numéricos u objetivos extraídos directamente del instrumento de personalidad TIPI en un rango de [1, 7]. Estos valores numéricos son calculados por cada rasgo de acuerdo a las respuestas a dos preguntas del instrumento TIPI. La

## 5.1 RxPI—un recurso para el análisis de la personalidad

---

Tabla 5.3 muestra ejemplos las clases y objetivos de 9 instancias del conjunto de datos RxPI. Para una mejor claridad de cómo utilizar el instrumento, calcular los valores numéricos y encontrar las normas asociadas a cada rasgo se puede consultar el Apéndice A.2.

### **Información faltante (*Is any information missing from individual instances? This does not include intentionally removed information, but might include, e.g., redacted text.*)**

Existe un porcentaje de ensayos que contienen información como nombre completo, correo electrónico, dirección, etc. Por lo tanto, en el proceso de transcripción, se sustituyeron las palabras que contenían información personal por etiquetas genéricas que indican el tipo de información suprimida (*redacted text*). Las etiquetas usadas en este proceso son: <NAME> para suprimir el nombre de la persona cuando este viene acompañado de sus apellidos, <EMAIL> para el correo electrónico, <PHONE> para el número de teléfono, y <ADDRESS> para su dirección cuando incluye detalles como la calle y número. Estas etiquetas se agregaron a las instancias en el proceso de transcripción manual, por lo que no es un proceso automático.

### **Relaciones entre instancias (*Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?*)**

No existen relaciones explícitas entre instancias individuales.

### **Particiones recomendadas (*Are there recommended data splits (e.g., training, development/validation, testing)?*)**

RxPI contiene instancias de 1484 participantes y cada participante presenta los cinco rasgos de personalidad de acuerdo al modelo *Big-Five*. Esta información puede ser útil para analizar rasgos de forma independiente. Sin embargo, para entender las diferencias entre la producción de lenguaje dentro de un mismo rasgo de personalidad propongo una versión estratificada de RxPI: RxPI-estratificado.

El conjunto de datos *RxPI-estratificado* contiene tres *clases* = {*bajo*, *control*, *alto*}. El objetivo de esta partición es considerar instancias provenientes de participantes con un rasgo dominante. En este sentido, una persona con rasgo dominante es aquella que se registra en los extremos de solo un rasgo. Por ejemplo, una persona con el rasgo de *Extroversión* dominante es aquella que tiene asignada o la clase *baja* o la clase *alta* en el rasgo de *Extroversión*, y es indispensable que tenga asignada la clase *medio-baja* o la clase *medio-alta* en los cuatro rasgos restantes. Todas las instancias con clases *medio-alto* y *medio-bajo* en los cinco rasgos se concentran en una clase llamada *control*. Es decir, los participantes considerados en el grupo de control no tienen algún

## Capítulo 5. Colecciones de datos

---

rasgo dominante y es el mismo conjunto para los cinco rasgos. Por ejemplo, de las instancias de la Tabla 5.3 las instancias a16-0021 y b18-0123 forman parte de RxPI-estratificado para el rasgo de *Extroversión*; mientras que las instancias a16-0022, i17-0209 y i17-0210, forman parte de la clase de *control*. Las cuatro instancias restantes en la tabla no forma parte de RxPI-estratificado pues no tienen un rasgo dominante, por ejemplo, a16-0020 registra tres rasgos en sus polos extremos: *Extroversión*, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*.

En RxPI-estratificado, la instancia de una persona puede usarse para estudiar solamente un rasgo ya sea que aparezca en el polo *alto* o el polo *bajo* (bajo estas condiciones, en RxPI-estratificado no existen participantes con dos rasgos dominantes). La idea de este tipo de partición ya se ha usado en conjuntos de datos para identificar personalidad; como en la investigación realizada por Oberlander and Gill (2006) se indica que un corpus estratificado en 3 clases permite analizar características entre dimensiones únicas. La Tabla 5.4 muestra el tamaño de RxPI-estratificado y su considerable reducción de instancias en comparación con RxPI completo.

### **Errores (*Are there any errors, sources of noise, or redundancies in the dataset?*)**

El conjunto de datos contiene elementos que pueden considerarse errores; estos son descritos en el apartado *Composición de una instancia* de la Sección 5.1.2.

### **Autocontenido (*Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?*)**

El conjunto de datos RxPI es autocontenido.

### **Información confidencial (*Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)?*)**

Las personas participantes fueron informadas sobre la actividad que estaban realizando y sobre el uso que se daría de la información recolectada, por lo que consideramos que el conjunto

**Tabla 5.4:** Número de participantes en RxPI-estratificado en dos clases correspondientes a los polos alto y bajo de cada rasgo de personalidad del modelo *Big-Five*, más el conjunto control.

	Alto	Bajo	Control	Total
Extroversión	41	30	335	406
Amabilidad	42	52	335	429
Responsabilidad	10	84	335	429
Estabilidad Emocional	23	30	335	388
Apertura a Nuevas Experiencias	82	52	335	469



## 5.1 RxPI—un recurso para el análisis de la personalidad

---

de datos no contiene información confidencial. Adicionalmente, suprimimos la información personal como se explica en la Sección 5.1.2 en el apartado *Información faltante*.

**Lenguaje sensible** (*Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?*)

Algunos ensayos pueden contener descripciones de incidentes que pueden causar ansiedad.

**Identificación de subpoblaciones** (*Does the dataset identify any subpopulations (e.g., by age, gender)?*)

La información solicitada a las personas participantes contiene información sobre su edad y género, como muestra la Tabla 5.2, pero esta información no pertenece al conjunto de datos.

**Identificación de individuos** (*Is it possible to identify individuals (i. e., one or more natural persons), either directly or indirectly (i. e., in combination with other data) from the dataset?*)

Como el apartado *Información faltante* (Sección 5.1.2) indica, información como el nombre con apellidos, correo electrónico, dirección y número de teléfono fueron sustituidos manualmente por etiquetas genéricas. Sin embargo, se podría identificar a una persona que proporciona su nombre de pila (que no se sustituye por la etiqueta <NAME>) y el lugar de donde es originario, particularmente cuando este lugar es una localidad pequeña o donde el participante haya escrito otros detalles de su familia como sus nombres (sin apellidos). Es posible que alguien que ya conozca al participante lo pueda identificar mediante estas narraciones. Sin embargo, este tipo de identificación es poco probable que pueda realizarse por medios automáticos.

**Datos sensibles** (*Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?*)

Dada la pregunta detonante usada para la escritura del ensayo, es posible que el ensayo resulte en una narración que contenga datos sensibles (el instrumento completo de recolección del conjunto de datos puede verse en el Apéndice A.1). Sin embargo, propusimos mecanismos para suprimir información que pudiera identificar a quienes proporcionaron sus ensayos para RxPI (ver el apartado *Identificación de individuos* de la Sección 5.1.2).

### 5.1.3. Proceso de recolección

#### **Información de cada instancia (*How was the data associated with each instance acquired?*)**

La información asociada a cada instancia se obtuvo mediante el instrumento TIPI (*Ten Item Personality Measure*) propuesto por Gosling et al. (2003). El instrumento es una encuesta con 10 preguntas, dos por cada rasgo de personalidad del modelo *Big-Five*. La encuesta se responde mediante una escala Likert de 7 valores de acuerdo. De estas respuestas se obtiene el valor numérico asociado a cada instancia; la clase se infiere del valor previamente calculado. Para determinar la clase, TIPI proporciona normas que corresponden a rangos independientes de cada rasgo (Gosling et al., 2003). El Apéndice A.2 muestra a detalle cómo se calcularon tanto los valores numéricos como las clases nominales.

#### **Mecanismo de recolección (*What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs?)*)**

El proceso de recolección de ensayos e información asociada a cada instancia fue obtenida de forma analógica, en papel. Las personas participantes se encontraban dentro de las instalaciones de la UAM-C, específicamente dentro de las aulas de clase. El Apéndice A.1 muestra el instrumento de recolección utilizado.

#### **Estrategia de muestreo (*If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities?)*)**

Para la obtención del corpus, consideramos a toda la población de primer trimestre de la UAM-C en sus 11 programas académicos a nivel licenciatura durante los años de ingreso 2016 a 2018. Sin embargo, el número de instancias colectadas es menor que el número de estudiantes admitidos en ese periodo; esto se puede deber a diferentes factores fortuitos (como estar presentes durante la aplicación del instrumento) o debido a que algunas personas declinaron la participación en el conjunto de datos.

#### **Personas involucradas (*Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid?)*)**

Las personas encargadas de la aplicación del instrumento de recolección del conjunto de datos fueron los trabajadores de la UAM-C del departamento de Servicios Escolares (y estudiantes haciendo su servicio social dentro de este departamento). La aplicación de este instrumento

## 5.1 RxPI—un recurso para el análisis de la personalidad

---

formó parte de una batería de diversos instrumentos para investigaciones desarrolladas dentro de la UAM–C.

Para las transcripciones manuales de todas las instancias se contó con la participación de cuatro personas becadas bajo financiamiento de la Red Temática de Tecnologías del Lenguaje dentro del programa “Creación de Recursos Lingüísticos”. El monto de la beca fue de \$10,000.00 pesos mexicanos (cantidad estándar en la UAM–C) por persona durante 4 meses.

Adicionalmente, dos personas becadas participaron en la revisión de todas las transcripciones. Todo el proceso de transcripción y revisión fue organizado y supervisado por M. C. Gabriela Ramírez de la Rosa. El Apéndice A.3 describe el material proporcionado a las personas que transcribieron y revisaron los ensayos.

### **Tiempo (*Over what timeframe was the data collected?*)**

El instrumento se aplicó en el trimestre de Otoño (de septiembre a diciembre) en cada año desde 2016 a 2018. La transcripción manual y las revisiones se realizaron durante el periodo de 2017 al 2019.

### **Proceso de revisión ética (*Were any ethical review processes conducted (e.g., by an institutional review board)?*)**

La UAM no contaba con un proceso de revisión ética antes ni durante el tiempo de recolección del conjunto de datos. Debido a que se obtuvo financiamiento público del entonces Consejo Nacional de Ciencia y Tecnología (CONACyT), se siguieron las disposiciones que este organismo indicó para el pago de las becas a las personas involucradas en la actividad de creación del conjunto de datos RxPI.

### **Recolección directa o indirecta (*Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?*)**

Los datos fueron obtenidos directamente de las personas participantes.

### **Notificación de participación (*Were the individuals in question notified about the data collection?*)**

Dado que los datos fueron obtenidos directamente de las personas participantes para un propósito específico, estas fueron notificadas de la creación del conjunto de datos antes de iniciar el proceso de recolección. El instrumento completo puede consultarse en el Apéndice A.1.

### **Consentimiento (*Did the individuals in question consent to the collection and use of their data?*)**

Las personas participantes convinieron en participar en el estudio de forma voluntaria y consintieron que sus datos podrían ser usados para fines académicos. El documento que los participantes firmaron se puede ver en el Apéndice A.1.

### **Exclusión voluntaria (*If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?*)**

No existe un mecanismo de exclusión voluntaria en forma explícita. Sin embargo, a todas las personas que aportaron su correo electrónico, al enviarles sus resultados de personalidad se les hizo hincapié en que la participación en el estudio es voluntaria. Se les comunicó que si no deseaban participar en el estudio podrían enviar un correo directamente a los responsables de la recolección para que sus datos fueran eliminados. De esta forma se eliminaron instancias digitales que no forman parte del conjunto de datos RxPI.

### **Análisis de protección de datos (*Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?*)**

No existe un análisis de protección de datos para RxPI.

#### **5.1.4. Preprocesamiento/limpieza/etiquetado**

##### **Tipos de procesos realizados (*Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)*)**

En la transcripción de las instancias se agregaron etiquetas para marcar fenómenos de la escritura manual. También, se usan etiquetas para remover información sensible que identifique al participante. El proceso se describe en detalle en la Sección 5.1.2.

##### **Datos originales (*Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?*)**

Existe una versión digitalizada (imagen) de los ensayos escritos a mano. Un subconjunto las imágenes con sus respectivas transcripciones manuales forman parte de una partición llamada HWxPI *Hand-Written teXts for Personality Identificación* y consiste de 711 instancias. Este con-

## 5.1 RxPI—un recurso para el análisis de la personalidad

---

junto de imágenes no tiene información ofuzcada como el nombre completo, dirección, correo electrónico o número de teléfono.

### **Software utilizado** (*Is the software that was used to preprocess/clean/label the data available?*)

No aplica. Los procesos de etiquetación y limpieza de las instancias fueron realizado de forma manual.

### **5.1.5. Usos**

#### **Tareas** (*Has the dataset been used for any tasks already?*)

Existen dos tareas en las cuales se ha utilizado parte del conjunto RxPI. El dataset *RxPI-estratificado* se ha utilizado en la tarea de clasificación binaria de personalidad (Ramírez-de-la-Rosa et al., 2023). Mientras que el dataset *HWxPI* se ha utilizado para la tarea de clasificación binaria durante el taller “2018 ICPR *Multimedia Information Processing for Personality & Social Networks Analysis Challenge*” llevado a cabo dentro de la conferencia ICPR 2018–*International Conference on Pattern Recognition* (Ramírez et al., 2019).

#### **Repositorios** (*Is there a repository that links to any or all papers or systems that use the dataset?*)

El corpus *HWxPI* puede ser accedido via CodaLab<sup>2</sup> o descargarse directamente en el repositorio del conjunto de datos en <https://github.com/lyr-uam/HWxPI-Track-ICPR2018>. Dado que este corpus fue usado en un taller científico, existe un repositorio de artículos que hacen uso del corpus y que está reportado por Ramírez et al. (2019).

#### **Otras tareas** (*What (other) tasks could the dataset be used for?*)

Además de las tareas de clasificación binaria en que se ha utilizado parte del conjunto de datos RxPI, el corpus RxPI y sus particiones también pueden ser utilizados para la tarea de regresión usando los valores numéricos disponibles para cada rasgo.

Adicionalmente, es posible utilizar el conjunto de datos para análisis lingüístico y la relación entre rasgos de personalidad y ensayos escritos. También, este corpus puede usarse en tareas de transferencia de aprendizaje —*transfer learning*— tomando en cuenta el contexto y población en la que se obtuvieron los datos.

---

<sup>2</sup>[https://competitions.codalab.org/competitions/18362#learn\\_the\\_details-overview](https://competitions.codalab.org/competitions/18362#learn_the_details-overview)

### **Uso futuro (*Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?*)**

El conjunto de datos RxPI contiene información de personas que son estudiantes de licenciatura con características demográficas similares: edad, distribución de género, que habita la Ciudad de México y con lenguaje nativo español mexicano; por lo tanto, los futuros usos de este corpus deberán considerar el impacto de la composición de la población participante (se pueden encontrar más detalles sobre la composición en la Sección 5.1.2).

### **Tareas a evitar (*Are there tasks for which the dataset should not be used?*)**

El conjunto de datos RxPI fue recolectado bajo condiciones particulares (como se describió en la Sección 5.1.2) que pueden influenciar la generalización. Los resultados de clasificación automática y regresión solo pueden aplicarse a una población similar a la que componen RxPI y sus subparticiones.

## 5.1.6. Distribución

### **¿A quiénes? (*Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?*)**

El dataset completo RxPI puede ser distribuido a investigadores o estudiantes que investigan la relación entre personalidad y texto escrito. RxPI puede ser distribuido fuera de la UAM-C únicamente para investigación y no se puede utilizar en aplicaciones comerciales.

### **¿Cómo? (*How will the dataset will be distributed (e.g., tarball on website, API, GitHub?)*)**

La descripción del conjunto de datos RxPI puede ser consultado en <https://github.com/lyr-uam/RxPI> y el conjunto de datos puede ser solicitado mediante el correo electrónico proporcionado.

### **¿Cuándo? (*When will the dataset be distributed?*)**

El conjunto de datos RxPI estará disponible a partir de febrero de 2024.

## 5.1 RxPI—un recurso para el análisis de la personalidad

---

### **Licencia (*Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?*)**

El corpus RxPI tiene una licencia Creative Commons CC BY-NC-SA 4 (*Attribution-NonCommercial-ShareAlike 4.0 International*).<sup>0</sup> La licencia puede consultarse en <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

### **Restricciones (*Have any third parties imposed IP-based or other restrictions on the data associated with the instances?*)**

No aplica.

### **Controles de exportación (*Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?*)**

No aplica.

#### **5.1.7. Mantenimiento**

### **Persona encargada del conjunto de datos (*Who will be supporting/hosting/maintaining the dataset?*)**

La persona encargada del conjunto de datos RxPI es Gabriela Ramírez de la Rosa.

### **Información de contacto (*How can the owner/curator/manager of the dataset be contacted (e.g., email address)?*)**

La descripción del conjunto de datos RxPI puede ser consultado en <https://github.com/lyr-uam/RxPI> y el conjunto de datos puede ser solicitado mediante el correo electrónico proporcionado en el archivo README del repositorio del corpus RxPI.

### **Fe de erratas (*Is there an erratum?*)**

Las fé de erratas se publicarán en el repositorio del conjunto de datos RxPI: <https://github.com/lyr-uam/RxPI>.

## Capítulo 5. Colecciones de datos

---

**Actualización** (*Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?*)

Es posible que existan nuevas versiones, si eso sucede, se publicará la nueva versión en el repositorio del corpus RxPI (<https://github.com/lyr-uam/RxPI>).

**Límites de retención** (*If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?*)

No existe un límite de tiempo para mantener en uso el conjunto de datos.

**Versiones antiguas** (*Will older versions of the dataset continue to be supported/hosted/maintained?*)

Todas las versiones, si existieran más de una, estarán disponibles.

**Contribuciones al conjunto de datos** (*If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?*)

No existe un mecanismo establecido para contribuciones al conjunto de datos RxPI.

## 5.2. Otras colecciones utilizadas

### 5.2.1. Conjunto de datos Auto–descripciones

*Auto–descripciones* es un conjunto de datos recolectado para el análisis de aspectos psicológicos en adolescentes con diferente orden de nacimiento. Este corpus pertenece a un estudio hecho por Cruz-Sánchez et al. (2017). El conjunto de datos cuenta con solo 121 participantes (63 mujeres y 58 hombres) que fueron etiquetados con sus rasgos de personalidad del modelo *Big-Five*.

Las instancias consisten de transcripciones manuales de una autopresentación grabada en video con una duración aproximada de 3 minutos hecha en español. Esta actividad se llevó a cabo dentro de una escuela secundaria en la zona rural del estado de Tlaxcala, México. En este estudio, los rasgos de personalidad de cada participante fueron obtenidos con el instrumento BFQ–25, el cual asigna diferentes categorías (*bajo*, *medio* y *alto*) a los valores numéricos obtenidos en la prueba de personalidad. La Tabla 5.5 muestra el número de instancias por clase en cada rasgo de personalidad del conjunto de datos Auto–descripciones.



## 5.2 Otras colecciones utilizadas

	Alto	Medio	Bajo	Total
Extroversión	49	56	16	121
Amabilidad	21	56	44	121
Responsabilidad	20	61	40	121
Estabilidad Emocional	36	43	42	121
Apertura a Nuevas Experiencias	45	60	15	121

**Tabla 5.5:** Número de instancias por clase en el conjunto de datos Auto-descripciones.

De forma similar que con RxPI, cada instancia está etiquetada con los 5 rasgos de personalidad de acuerdo a las normas establecidas en el instrumento BFQ-25. El acceso a este conjunto de datos fue obtenido únicamente para ser usados en esta investigación a través de la doctora responsable de la recolección, Dr. Verónica Reyes Meza.

### 5.2.2. Conjunto de datos Essays

*Essays* es un recurso para el análisis de la personalidad que consiste de ensayos escritos en inglés mayoritariamente por estudiantes de psicología a quienes se les dio la instrucción de escribir lo que les viniera a la mente por 20 minutos ininterrumpidos sin preocuparse por la ortografía o la gramática. El primer conjunto de ensayos fue recopilado por Pennebaker and King (1999) durante 1997 y 1999 y consistió de aproximadamente 1200 ensayos. Este primer conjunto fue incrementado por Mairesse et al. (2007), con ensayos escritos hasta el 2004, para formar el corpus que se conoce actualmente como Essays.

El corpus Essays contiene 2468 ensayos escritos en inglés pero no se cuenta con información completa sobre el género y la edad de los participantes. Sin embargo, Pennebaker and King (1999) presentan información parcial de 1200 participantes de los primeros ensayos compilados. El corpus Essays está etiquetado con los cinco rasgos del *Big-Five* y para cada rasgo contiene valores nominales ‘yes’ para indicar el polo alto del rasgo, y ‘no’ para indicar el polo bajo del rasgo. Es decir, no contiene información sobre los valores numéricos de los rasgos de personalidad de cada participante. La Tabla 5.6 muestra el número de participantes (o ensayos) en cada rasgo por clase en el conjunto de datos Essays. Cabe destacar que igual que con el corpus RxPI, cada instancia del corpus está etiquetada con los 5 rasgos de personalidad.

	Yes	No	Total
Extroversión	1277	1191	2468
Amabilidad	1310	1158	2468
Responsabilidad	1254	1214	2468
Estabilidad Emocional	1235	1233	2468
Apertura a Nuevas Experiencias	1271	1196	2467

**Tabla 5.6:** Número de instancias por clase en el conjunto de datos Essays.



# 6

## Método propuesto

Dada como entrada una **intención comunicativa** —un mensaje cuyo objetivo es transmitir información de una persona emisora a otras receptoras mediante el lenguaje escrito u oral— esta tesis propone una metodología para identificar automáticamente la personalidad de la persona emisora. Así, propongo incorporar la disponibilidad léxica en la representación de la intención comunicativa para generar un modelo de clasificación automática para la identificación de rasgos de personalidad. El método propuesto, también busca determinar de forma automática, categorías de términos disponibles adaptados a cada rasgo de personalidad. Esta incorporación de la disponibilidad léxica en la representación se basa en 2 aspectos de las palabras contenidas en la intención comunicativa: *pertenencia* y *orden*.

Las contribuciones mostradas en este capítulo corresponden a las contribuciones 2–4 listadas en el Capítulo 1.3:

2. Adaptación de la prueba de disponibilidad léxica que en lugar de elicitar listas de palabras, se realiza mediante intenciones comunicativas,
3. Representación de texto basada en la disponibilidad léxica y
4. Método automático para seleccionar categorías de términos que permitan mejorar la representación de textos para la identificación automática de la personalidad.

Al final del capítulo presento el esquema general en el que se utilizan las tres contribuciones descritas.

### 6.1. Adaptación de la prueba de disponibilidad léxica

Tradicionalmente, la prueba de disponibilidad léxica se lleva a cabo elicitando a un conjunto de participantes una lista de *palabras respuesta* que primero vengan a la mente dado un *centro de interés* o *tema detonante*. Este conjunto de participantes tiene características particulares dependiendo del estudio y la variable a estudiar (como se describió en trabajos del Capítulo 2).

Un ejemplo es el uso de la disponibilidad léxica para estudiar el uso de un lenguaje en una segunda lengua en comunidades provenientes de dos países distintos; donde el objetivo es comparar la disponibilidad de palabras o términos obtenidos entre las comunidades de cada país. El procedimiento para realizar esta actividad es que una vez que cada participante haya producido la lista de palabras del centro de interés dado, se revisan todas las listas generadas, y si existen palabras repetidas solo se mantiene en las listas la primera repetición de una palabra.

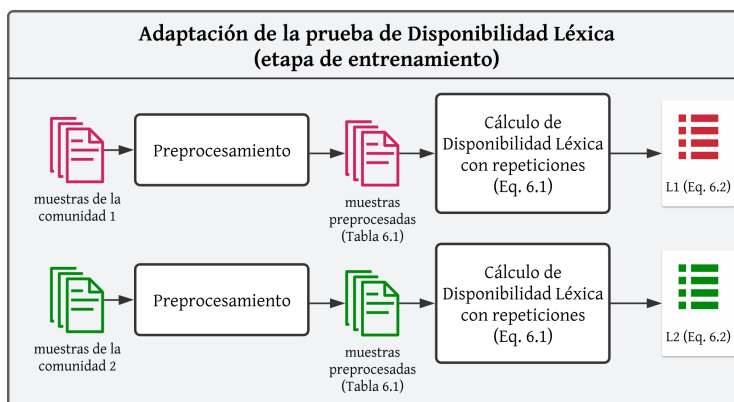
En esta investigación propongo usar *intenciones comunicativas* de los participantes en las comunidades a estudiar en lugar de elicitar listas de palabras. Donde las dos comunidades a estudiar a lo largo de la investigación están compuestas de personas que tienen la personalidad en polos opuestos de un mismo rasgo —por ejemplo: comunidad de extrovertidos y comunidad de introvertidos, para los polos opuestos del rasgo de *Extroversión*—. En estas comunicaciones orales o escritas es muy probable que existan repeticiones de palabras. Como se mencionó en el Capítulo 5, las intenciones comunicativas que se usan en esta investigación tienen como propósito comunicar información personal sobre los participantes. Así, una forma diferente de ver la hipótesis de esta tesis es que **la primera información que venga a la mente en estos ejercicios revelará aspectos importantes relacionados con el rasgo de personalidad de un participante.**

Considerando estas diferencias entre la elicitación tradicional y comunicaciones orales o escritas, las siguientes secciones describen los procesos involucrados en obtener la disponibilidad léxica de los términos usados en intenciones comunicativas: i) preprocesamiento del texto, y ii) cálculo de la disponibilidad léxica que considera repeticiones de palabras. La Figura 6.1 muestra un diagrama de esta contribución. La última parte de esta sección muestra ejemplos de listas de disponibilidad léxica en las colecciones de datos usadas, i. e. RxPI, Auto-descripciones, y Essays, para el rasgo de *Extroversión*. Estas listas son ejemplos que corresponden a L1 y L2 en el diagrama de la Figura 6.1.

#### 6.1.1. Preprocesamiento de textos

El preprocesamiento de textos realizado tiene como objetivo mantener los términos que tienen más probabilidades de pertenecer a las palabras descriptivas de la personalidad. Dado un documento de texto que corresponda a una comunicación escrita o hablada (transcrita) se

## 6.1 Adaptación de la prueba de disponibilidad léxica



**Figura 6.1:** Diagrama del módulo de Adaptación de la prueba de disponibilidad léxica.

requieren pasos de normalización para eliminar información que no es relevante para identificar la personalidad. Mi adaptación propone como pasos de normalización<sup>1</sup>: eliminación de palabras vacías o palabras de función (por ejemplo, artículos y preposiciones), eliminación de números y símbolos, y transformación de todas las palabras a minúsculas. Además, a través del software FreeLing (Carreras et al., 2004), se detectan multi-palabras (como colocaciones o nombres propios como nombres de ciudades) y se reconocen entidades nombradas (o NER).

La Tabla 6.1 muestra un ejemplo del preprocesamiento realizado con FreeLing. Primero aparece un extracto de la transcripción de una persona con extroversión alta, luego puede verse la transcripción después del preprocesamiento realizado, y finalmente se muestra la misma transcripción en forma de lista que correspondería a las listas obtenidas en la elicitación tradicional de la prueba de disponibilidad léxica.

### 6.1.2. Cálculo de la disponibilidad léxica con repeticiones

De forma tradicional, para obtener la información de disponibilidad léxica de una comunidad se analizan las palabras producidas por todas las personas de dicha comunidad. Para esto, primero se obtiene un vocabulario único de la comunidad, este vocabulario contendrá el conjunto unión de las palabras (términos) elicidadas por todas las personas participantes. Luego, para obtener el valor de disponibilidad léxica por cada término  $t_j$  en el vocabulario de la comunidad se utiliza la Ecuación 6.1, propuesta por Callealta Barroso and Gallego Gallego (2016), que considera todas las posiciones donde el término  $t_j$  es mencionado dentro de los ensayos proporcionados por esa comunidad, así como la frecuencia de su uso en cada una de esas posiciones.

$$DL_{w,k,m}(t_j) = \sum_{i=1}^n w^{\left(\frac{i-1}{k-1}\right)^m} \times \frac{f_{ji}}{I} \quad (6.1)$$

<sup>1</sup>La normalización propuesta puede cambiar de acuerdo al problema específico a resolver.

## Capítulo 6. Método propuesto

---

donde

- $t_j$  representa el término léxico (palabra) del cual se desea encontrar el valor de disponibilidad léxica;
- $i$  es una posición en el ensayo, este valor itera desde 1 hasta  $n$ ;
- $n$  es la última posición donde se encuentra el término  $t_j$  al considerar todos los ensayos de la comunidad<sup>2</sup>;
- $f_{ji}$  es el número de veces en que el término  $t_j$  se encuentra en la posición  $i$ ;
- $k$  indica la posición en la que el valor de disponibilidad léxica es igual a  $w$ ;
- $w$  es el valor deseado para el  $k$ -ésimo término en la lista  $L$  ( $w$  puede tomar valores entre 0 y 1, donde el valor deseado para el  $k$ -ésimo término es cercano a cero);
- $m$  modula el decaimiento del valor de disponibilidad léxica entre los términos de  $L$  de la Eq 6.2 (entre mayor es el valor de  $m$ , menos pronunciado es el decaimiento);
- $I$  es un valor de normalización (en la propuesta original, este valor es el número total de participantes en el estudio); y
- el rango de  $DL_{w,k,m}(t_j)$  es  $(0, 1]$ .

En el cálculo de disponibilidad léxica tradicional solo se considera la primera aparición de palabras repetidas, cuando las hay, pues el objetivo es relacionar el tema detonante (centro de interés) con sus respuestas tan pronto como estas emerjan dentro de la mente de cada participante. Sin embargo, como se ha discutido, en una intención comunicativa las repeticiones de palabras son inevitables, por lo que para considerar qué tanto estos términos aparecen en la mente de las personas al hablar, propongo adaptar el cálculo de disponibilidad léxica para considerar repeticiones. Como resultado, la propuesta cambia el valor de normalización  $I$  en Eq. 6.1 a la frecuencia máxima alcanzada por cualquier término en los ensayos dentro de la comunidad de estudio. En contraste, en Callealta Barroso and Gallego Gallego (2016) el valor de normalización  $I$  se establece como el número máximo de participantes en el estudio.

La salida de este proceso es una lista de pares de términos que pertenecen al vocabulario de la comunidad con su respectivo valor de disponibilidad léxica. Esta lista se puede ordenar de mayor a menor valor de disponibilidad léxica para facilitar su interpretación. La lista final queda definida como lo indica la Ecuación 6.2.

---

<sup>2</sup>El valor de  $n$  varía para cada término  $t_j$ . Para posiciones mayores a  $n$  la frecuencia  $f_{ji}$  es cero.

## 6.1 Adaptación de la prueba de disponibilidad léxica

**Tabla 6.1:** Un extracto de una transcripción del corpus RxPI que corresponde a una persona con extroversión alta. La transcripción original responde a la instrucción “Cuéntanos sobre ti, por ejemplo algo sobre tu historia familiar o el(los) evento(s) que te venga a la mente que consideres relevante(s) en tu vida”. La transcripción después del preprocesamiento se muestra en forma de lista que correspondería a la lista elicitada de las pruebas de disponibilidad léxica tradicional.

Transcripción original	<i>Estoy bien emocionalmente con mi familia, mis padres apoyan mis preferencias y decisiones siempre y cuando no me afecte en un futuro, tengo un hermano con el que me llevo bien, tengo dos abuelas a las cuales respeto y quiero y admiro. Mis tios son muy amables conmigo, no he tenido algún problema familiar en el cual me concierne directamente a mi.</i>
Transcripción después del preprocesamiento con FreeLing	<i>bien emocionalmente familia padres apoyan preferencias decisiones siemprecuando afecte futuro hermano llevo bien dos abue- las cuales respeto quiero admiro tíos amables conmigo algún problema familiar concierne directa- mente</i>
Transcripción después del preprocesamiento (en forma de lista)	<i>bien emocionalmente familia padres apoyan preferencias decisiones siemprecuando afecte ...</i>

$$L = \{\langle t_j, DL_\theta(t_j) \rangle \mid t_j \in V\} \quad (6.2)$$

donde  $\theta = \{w, k, m\}$  contiene los parámetros para el cálculo del valor de disponibilidad léxica. Siguiendo la sugerencia de los autores de la Eq. 6.1, para realizar una comparación significativa entre la disponibilidad léxica de dos comunidades, se recomienda que se utilicen los mismos valores para cada componente de  $\theta$  en la generación de las listas en ambas comunidades. En mi adaptación propuesta, el vocabulario de la comunidad,  $V$ , se forma por la unión de los conjuntos de términos (o palabras) en las intenciones comunicativas (ensayos) después del preprocesamiento.

### 6.1.3. Disponibilidad léxica en diferentes colecciones

Las listas obtenidas de la prueba de disponibilidad léxica permiten revelar información adicional sobre el conjunto de términos disponibles que tiene cada comunidad en los diferentes conjuntos de datos. En esta sección se presentan ejemplos de términos con mayor disponibilidad léxica para el rasgo de *Extroversión* en los tres conjuntos de datos utilizados en esta investigación: RxPI, Auto-descripciones y Essays. Las listas  $L$  en esta sección se calcularon con un subconjunto aleatorio de ensayos de cada conjunto de datos.

## Capítulo 6. Método propuesto

---

La Tabla 6.2 presenta un ejemplo de los primeros 20 términos en las listas de disponibilidad léxica  $L$  de la Eq. 6.2 obtenidas con el conjunto de datos RxPI para el rasgo de *Extroversión* (primeras columnas). Como se puede ver en la tabla, existen valores de  $DL$  muy altos: 0.90 y 0.85 para el término más accesible en el polo alto y bajo, respectivamente. Sin embargo, no es el mismo término disponible en ambas comunidades. Para el polo alto, el término más disponible es ‘familia’ y para el polo bajo es el término ‘gusta’; siendo ‘familia’ un término que es más disponible (con un valor de disponibilidad léxica de 0.90) que ‘gusta’ (con un valor de 0.85). Además, se puede apreciar que 9 de los 20 términos son términos comunes en ambas comunidades, esto habla de lo parecido de los temas que se tocan en este conjunto.

La sección media de la Tabla 6.2 muestra un ejemplo de los primeros 20 términos en las listas de disponibilidad léxica  $L$  calculadas a partir del conjunto de datos Auto–descripciones para el rasgo de *Extroversión*. Aquí, el término ‘gusta’ es el término más disponible en ambas comunidades —polo alto y polo bajo—; pero tienen una ligera diferencia en el valor de disponibilidad léxica, 0.62 y 0.66 para el polo alto y el polo bajo, respectivamente. También se puede notar un descenso del valor de  $DL$  a partir de la tercera posición en ambas listas, lo cual podría indicar que los dos primeros términos son usados por la mayoría de los participantes en las primeras posiciones —lo primero que les viene a la mente— una posible causa es que las comunicaciones en este corpus fueron obtenidas de forma oral, donde el término ‘pues’ es usado como muletilla en muchos de los casos. En cuanto a términos comunes, en este ejemplo podemos observar un mayor número de términos iguales en ambas listas, 13 de 20. Esta intersección entre términos disponibles en ambas listas reitera la intuición que la mayoría de los participantes hablan de forma similar y de temas similares.

Las últimas columnas de la Tabla 6.2 muestran un ejemplo de los primeros 20 términos en las listas de disponibilidad léxica  $L$  calculadas a partir del conjunto de datos Essays para el rasgo de *Extroversión*. Para el caso de Essays, se observa que los tres primeros términos disponibles en ambas listas son iguales —aunque en diferente orden—. Con valores de disponibilidad léxica mayores para el polo alto que para el polo bajo. También es notoria la intersección de términos entre ambas listas, donde 18 de 20 son compartidos en ambas comunidades. Aunque las palabras con mayor disponibilidad léxica en ambas comunidades pueden considerarse como palabras de uso frecuente, por ejemplo: ‘like’, ‘think’, ‘going’, ‘want’, etc; tienen sentido en el contexto de la pregunta detonadora del corpus, ‘lo primero que se les venga a la mente en 20 minutos’.



## 6.1 Adaptación de la prueba de disponibilidad léxica

**Tabla 6.2:** Ejemplos de listas de disponibilidad léxica para los dos polos del rasgo de *Extroversión* en las colecciones de datos RxPI, Auto-descripciones y Essays.

RxPI corpus			Auto-descripciones corpus			Essays corpus		
polo-alto	polo-bajo		polo-alto	polo-bajo		polo-alto	polo-bajo	
$t_j$	$t_j$	$DL(t_j)$	$t_j$	$t_j$	$DL(t_j)$	$t_j$	$t_j$	$DL(t_j)$
familia	gusta	0.856	gusta	gusta	0.626	gusta	like	0.927
años	familia	0.756	gusta	gusta	0.663	gusta	really	0.919
vida	años	0.597	gusta	gusta	0.185	gusta	's	0.894
gusta	persona	0.562	gusta	gusta	0.178	gusta	think	0.697
carrera	vida	0.541	gusta	gusta	0.164	gusta	know	0.607
siempre	personas	0.517	gusta	gusta	0.157	gusta	time	0.538
persona	siempre	0.51	gusta	gusta	0.134	gusta	going	0.519
estudiar	madre	0.496	gusta	gusta	0.123	gusta	want	0.469
mas	cosas	0.485	gusta	gusta	0.122	gusta	people	0.452
hermanos	padres	0.427	gusta	gusta	0.109	gusta	feel	0.383
tiempo	mamá	0.418	gusta	gusta	0.092	gusta	good	0.351
yaque	ser	0.414	gusta	gusta	0.092	gusta	right	0.333
amigos	considero	0.398	gusta	gusta	0.091	gusta	need	0.315
cosas	hacer	0.376	gusta	gusta	0.083	gusta	things	0.286
bien	yaque	0.348	gusta	gusta	0.081	gusta	school	0.279
hermano	diferentes	0.33	gusta	gusta	0.078	gusta	wonder	0.272
considero	sido	0.306	gusta	gusta	0.073	gusta	friends	0.265
quero	hace	0.293	gusta	gusta	0.073	gusta	class	0.259
así	año	0.291	gusta	gusta	0.071	gusta	love	0.251
uam	hermana	0.282	gusta	gusta	0.07	gusta	guess	0.25
			gusta	gusta	0.087	gusta	home	0.218

Comparar las listas de disponibilidad léxica de las comunidades de esta forma ha servido a los investigadores de diversos estudios a analizar e interpretar diferencias del uso de términos entre dos comunidades. La información que contienen las listas va más allá del orden de disponibilidad que tengan el mismo término en ambas, también permite saber qué valor de disponibilidad tienen tales términos. Por ejemplo, el término ‘familia’ del corpus Auto-descripciones en la Tabla 6.2 es el sexto término más disponible en la comunidad del polo–alto y el octavo más disponible en la comunidad del polo–bajo; pero el valor de disponibilidad léxica es más alto para el octavo término en el polo–bajo que para el sexto término en el polo–alto: 0.162 y 0.157, respectivamente. Estas sutiles diferencias pueden ayudar a diferenciar entre el uso del lenguaje en ambas comunidades para ayudar a distinguir e identificar el polo de los rasgos de personalidad.

## 6.2. Representación de texto basada en disponibilidad léxica

La segunda contribución mostrada en este capítulo consiste en generar una representación vectorial basada en la información de disponibilidad léxica obtenida de dos comunidades a analizar. En términos de aprendizaje supervisado, cada comunidad corresponde a una clase. Se proponen dos formas de utilizar la disponibilidad léxica calculada de cada clase: en una representación básica que solo incluye la *pertenencia* de los términos dentro de la lista  $L$ : DL–base; y en una representación que incluye además el *valor de accesibilidad* (es decir, el valor de disponibilidad léxica) de cada término en el vector generado: DL–va. La Figura 6.2 muestra un diagrama de cómo se obtiene el vocabulario (o la lista de términos con disponibilidad léxica combinada) para las dos variantes propuestas. Este vocabulario (o lista) se usa para generar los vectores que representarán a cada instancia tanto en el entrenamiento como en la etapa de evaluación.

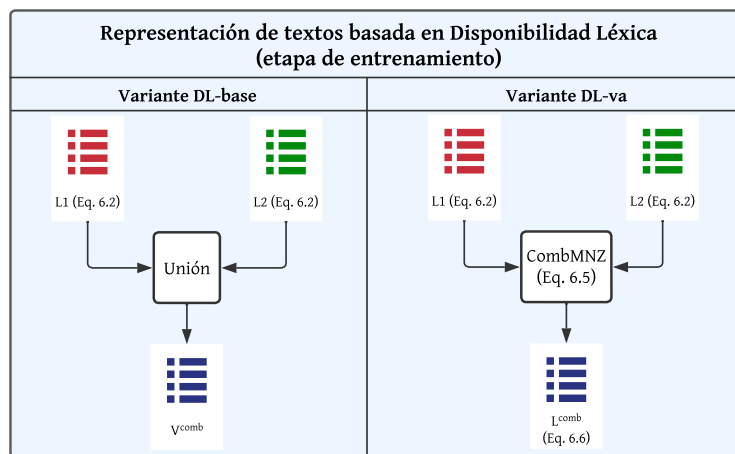


Figura 6.2: Diagrama de las dos variantes propuestas para la representación de textos basada en disponibilidad léxica.

### 6.2.1. Representación básica: DL–base

El objetivo de la representación básica DL–base es considerar únicamente aquellos términos con alta disponibilidad léxica en cada una de las comunidades de estudio; en contraste con utilizar términos frecuentes como lo hace un enfoque basado en representaciones tradicionales de bolsa de palabras. Esto implica considerar términos que *pertenecen* al lexicón mental obtenido por medio de la disponibilidad léxica.

Así, el vocabulario resultante consiste de términos en el conjunto de entrenamiento a los que se calcularon el valor de disponibilidad léxica. Dependiendo del parámetro  $k$  en  $\theta$  de la Eq. 6.1 es posible generar vocabularios compactos en comparación con una representación típica de bolsa de palabras.

Dadas dos listas de disponibilidad léxica  $L^1$  y  $L^2$ , correspondientes a la comunidad 1 y 2, se genera una representación vectorial donde el vocabulario de términos en la representación se define como la unión de los términos en  $L^1$  y en  $L^2$ . Entonces, el vector de una instancia tiene la forma  $\mathbf{d} = (w_1, w_2, \dots, w_{|L^1 \cup L^2|})$ ; donde el valor de  $w$  se determina con alguno de los esquemas de pasado tradicionales usados en tareas de clasificación automática de textos, a saber:

- Booleano: asigna 1 si el término  $t_j$  está presente en la instancia  $d$ ; 0 en caso contrario.
- Frecuencia del Término (TF): asigna la frecuencia del término  $t_j$  en la instancia  $d$ ; 0 si el término no está presente.
- Frecuencia del Término Modificada (MTF) (Chen et al., 2021): es una variante de TF que considera la longitud promedio de las instancias en la colección como un factor de normalización; se define como:

$$MTF(t_j, d) = \sqrt{f_j \cdot \log_2\left(1 + \frac{avg\_l}{l(d)}\right)} \quad (6.3)$$

donde  $avg\_l$  es la longitud promedio de las instancias en el conjunto de entrenamiento,  $l(d)$  es la longitud de la instancia a representar  $d$  y  $f_j$  corresponde a la frecuencia del término  $t_j$  en  $d$ .

- Frecuencia del Término – Frecuencia Inversa del documento (TF–IDF) (Salton and Buckley, 1988): es una variación de TF que considera un factor de normalización llamado IDF o frecuencia inversa de documento; donde entre menos aparezca el término en diferentes instancias de la colección, más importante será en la instancia representada; se define como:

$$TF - IDF(t_j, d) = f_j \times \log_2\left(\frac{n}{d(t_j)}\right) \quad (6.4)$$

donde  $n$  es el número de instancias en el conjunto de entrenamiento,  $d(t_j)$  es el número de instancias donde aparece  $t_j$  y  $f_j$  corresponde a la frecuencia del término  $t_j$  en  $d$ .

### 6.2.2. Representación basada en el valor de disponibilidad léxica: DL-va

La propuesta de representación DL-va, además del vocabulario disponible, considera el valor de disponibilidad léxica de ambas comunidades. Así, el *valor de accesibilidad* para un término dado considera los valores calculados de ese término en cada comunidad. Esto implica considerar tanto los términos que *pertenecen* al lexicón mental como considerar el *orden* de estas palabras dentro de las listas producidas.

Por lo tanto, la representación DL-va consiste de dos procesos:

1. El cálculo de un vocabulario compartido entre las dos comunidades. Esto es relevante porque al usar el mismo tema detonante en ambas comunidades, se tiende a usar el mismo vocabulario. En consecuencia, es altamente probable que las listas generadas en las dos clases a clasificar tengan una alta intersección de términos.
2. El pesado de términos en el vector de representación que incorpore este valor de disponibilidad combinado de las dos clases.

Dadas dos listas de disponibilidad léxica  $L^1$  y  $L^2$ , correspondientes a la comunidad 1 y 2, se considera una representación vectorial donde el conjunto vocabulario a utilizar en la representación se define como la unión de los términos en  $L^1$  y en  $L^2$ , es decir  $V^{comb}$ . Y donde el valor de disponibilidad léxica combinado ( $DL^{comb}$ ) se calcula con una estrategia de fusión de listas ordenadas, específicamente con el método de CombMNZ que define la Eq. 6.5 (Fox and Shaw, 1994). El objetivo de utilizar un método de fusión es calcular un sólo valor de disponibilidad léxica para aquellos términos que aparecen en ambas listas. La ecuación general de CombMNZ está definida para múltiples listas, pero esta investigación la utiliza para combinar pares de ellas.

$$DL^{comb}(t_j) = CombMNZ(t_j) = \sum_{i=1}^2 DL_{\theta}(t_j)^{L^i} \times |DL_{\theta}(t_j)^{L^i} > 0| \quad (6.5)$$

donde  $DL_{\theta}(t_j)^{L^i}$  es el valor de disponibilidad léxica del término  $t_j$  en la lista  $L^i$ , y  $|DL_{\theta}(t_j)^{L^i} > 0|$  es el número de valores diferentes de cero dado el término  $t_j$  en cualquier lista, es decir en

cuántas listas aparece  $t_j$ . Como resultado se obtiene una lista combinada, definida en la Ecuación 6.6 que se utilizará en el problema de clasificación.

$$L^{comb} = \{ \langle t_j, DL^{comb}(t_j) \rangle \mid t_j \in V^{comb} \} \quad (6.6)$$

En cuanto al esquema de pesado que utiliza información del valor de disponibilidad léxica se propone una combinación de los esquemas tradicionales de pesados con el valor de accesibilidad de cada término, como sigue:

- Booleano con disponibilidad (va–bool): asigna  $1 \times DL^{comb}(t_j)$  si el término  $t_j$  está presente en la instancia a representar  $d$ ; 0 en caso contrario.
- Frecuencia del Término con disponibilidad (va–TF): asigna la frecuencia del término  $TF(t_j) \times DL^{comb}(t_j)$  en la instancia  $d$ ; 0 si el término no está presente.
- Frecuencia del Término Modificada con disponibilidad (va–MTF): asigna  $MTF(t_j) \times DL^{comb}(t_j)$ .

### 6.3. Autoselección de categorías

Una forma de incluir el lexicón mental en una representación textual para la identificación de rasgos de personalidad es analizando las categorías semánticas que conforman los textos analizados. Como se ha establecido en el Capítulo 2, la información semántica que se puede encontrar mediante el lexicón mental da pautas para entender cómo accedemos a aspectos de nuestra mente y pensamiento. Por lo que analizar las categorías semánticas que se forman de los términos en las pruebas de disponibilidad léxica será relevante para generar representaciones más finas y con el propósito específico para la identificación automática de rasgos de personalidad. Así, dichas categorías permiten analizar características de los rasgos de personalidad desde un ángulo psicolingüístico.

Dado que el tema detonante en una intención comunicativa es el mismo para ambas comunidades, los términos en ambas listas de disponibilidad léxica son virtualmente los mismos. Sin embargo, como se puede ver en la Sección 6.1.3, la disponibilidad léxica de los términos en común tienden a diferenciarse. Con el objetivo de entender mejor las diferencias entre las comunidades estudiadas, se toma como referencia la metodología utilizada en el área de la Disponibilidad Léxica. En esta subdisciplina de la Lingüística se busca clasificar las palabras generadas por las comunidades; y así describir y analizar las categorías semánticas derivadas del tema detonante en la prueba de disponibilidad léxica.

## Capítulo 6. Método propuesto

---

Para automatizar este proceso de análisis se propone utilizar un conjunto de términos organizados en categorías semánticas que sean relevantes desde la perspectiva de la identificación de rasgos de personalidad. Una vez se cuente con este recurso al que llamaré *Diccionario de Categorías*, todos los términos de la lista de disponibilidad léxica quedarán categorizados para su posterior análisis.

El **Diccionario de Categorías (DC)** deberá tener como mínimo las siguientes características:

1. Un conjunto de  $M$  categorías,  $C = \{c_1, c_2, \dots, c_M\}$ , en las cuales ubicar uno o más términos, y
2. Un conjunto de términos,  $T = \{t \mid t \in c_i, \text{ para alguna } c_i\}$ , que pertenezcan a una o más categorías en  $C$ .

Luego, dada una lista de disponibilidad léxica  $L$  y el DC se produce una salida que consiste en sublistas de  $L$ , donde una sublista  $L_c$  está definida de manera general por la Ecuación 6.7. De modo que cada sublista  $L_c$  contendrá todos los términos de  $L$  que estén presentes en el DC en la categoría  $c$ . Estos términos en  $L_c$  están ordenados de mayor a menor valor en disponibilidad léxica y este orden está representado por el valor del elemento  $i$  en la terna, así los valores de  $i$  serán:  $i = 1$  para el  $t_j$  con el valor más grande de disponibilidad léxica,  $i = 2$  para el elemento con el segundo valor más grande de disponibilidad léxica, y así sucesivamente.

$$L_c = \{\langle i, t_j, DL(t_j) \rangle \mid (t_j \in L \wedge t_j \in c), i \in \mathbb{N}\} (\forall c \in C) \quad (6.7)$$

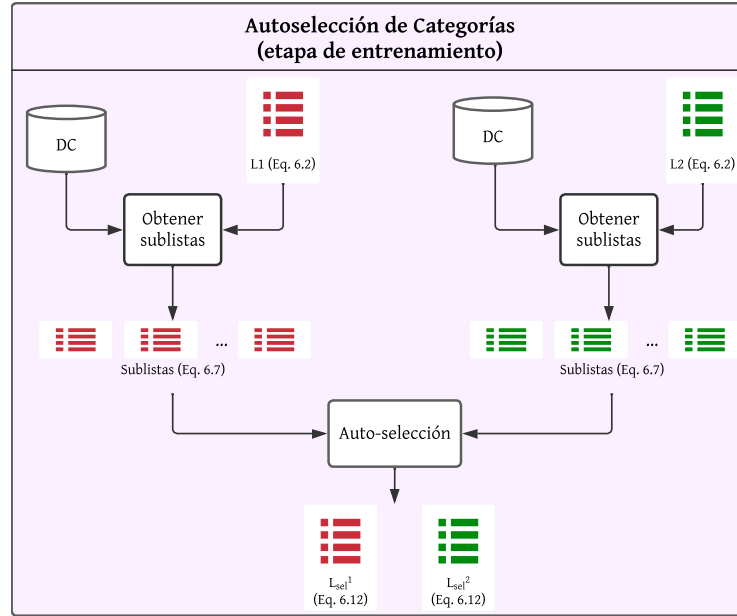
Una vez organizadas las dos listas de disponibilidad léxica en sublistas de categorías de términos podemos analizar cuáles de estas categorías son más relevantes para cada problema de clasificación. Para ello, en la siguiente sección presento dos formas para autoseleccionar categorías relevantes derivadas de estas las sublistas de disponibilidad léxica (esto se ilustra en la Figura 6.3).

### 6.3.1. Método de autoselección de categorías relevantes

Existe evidencia que al construir métodos automáticos de identificación de rasgos de personalidad no todas las categorías de términos usados muestran una alta correlación con cada rasgo de personalidad. Por ejemplo, Mairesse et al. (2007) expresan que “es probable que las características varíen en función del lenguaje origen y del método de evaluación de la personalidad”<sup>3</sup>.

---

<sup>3</sup>La cita en inglés es: “features are likely to vary depending on the source of language and the method of assessment of personality”



**Figura 6.3:** Diagrama del método de Autoselección de categorías basado en el valor de disponibilidad léxica.

Es por esto que propongo un método de autoselección de las categorías más relevantes para cada rasgo considerando la disponibilidad léxica de los términos en las categorías del DC. El objetivo de este proceso es seleccionar las sublistas de términos que pertenezcan a las categorías con mayores diferencias en el valor de disponibilidad léxica de los primeros  $n$  términos entre las dos comunidades.

Particularmente, dado el conjunto de comunidades  $\Omega = \{polo-alto, polo-bajo\}$  y un conjunto de categorías  $C = \{c_1, c_2, \dots, c_M\}$ , las sublistas por comunidad,  $L_c^\omega$ , quedan definidas como se indica en la Ecuación 6.8, donde  $\omega \in \Omega$ .

$$L_c^\omega = \{\langle i, t_j, DL(t_j) \rangle \mid (t_j \in L^\omega \wedge t_j \in c), i \in \mathbb{N}\} (\forall c \in C) \quad (6.8)$$

Ahora, para cada categoría  $c \in C$  se construye un vector ordenado de acuerdo a los valores de disponibilidad léxica  $DL(t_j)$  con el objetivo de seleccionar aquellas categorías que tengan valores de disponibilidad léxica más distantes en las mismas posiciones en las sublistas  $L_c^\omega$ .

Para cada comunidad  $\omega \in \Omega$ , se construyen  $M = |C|$  vectores de la forma:

$$\mathbf{v}_c^\omega = (DL(t_j)_1, \dots, DL(t_j)_q) \quad (6.9)$$

donde  $q = \min(|L_c^{polo-alto}|, |L_c^{polo-bajo}|)$ , y cada elemento en estos vectores corresponde al

## Capítulo 6. Método propuesto

---

orden en que se listan los términos de mayor a menor disponibilidad léxica, el índice  $i$  de la Ecuación 6.8.

Dado que son dos comunidades  $\omega \in \Omega$ , se calcula el error absoluto medio (o *mean absolute error* MAE) para cada par de vectores de la misma categoría  $c$ ,  $\mathbf{v}_c^\omega$ . Finalmente, se seleccionan las categorías,  $C_{sel}$ , con un mayor valor en MAE. Lo que implica que seleccionamos las categorías en los que sus términos tienen valores de disponibilidad léxica más distantes entre las dos comunidades como se describe en la Eq. 6.10.

$$C_{sel} = \arg \max_{c \in C} MAE(\mathbf{v}_c^{polo-alto}, \mathbf{v}_c^{polo-bajo}) \quad (6.10)$$

donde el valor de MAE para cualesquiera dos vectores  $\mathbf{x}$  y  $\mathbf{y}$  se define por la Ecuación 6.11.

$$MAE(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^q |y_i - x_i|}{q} \quad (6.11)$$

Los términos que pertenezcan a este subconjunto de categorías seleccionadas (como se describe en la Eq. 6.12) son utilizados en la representación de instancias ya sea utilizando la variante DL-base o DL-va.

$$L_{sel}^\omega = \{\langle i, t_j, DL(t_j) \rangle \mid (t_j \in L^\omega \wedge t_j \in c), i \in \mathbb{N}\} (\forall c \in C_{sel}) \quad (6.12)$$

### 6.3.2. Diccionario de Categorías

Para seleccionar categorías de forma automática asumo que existe un Diccionario de Categorías que tiene como mínimo un conjunto de categorías y una lista de términos que pertenecen a una o más categorías. En esta subsección, hablaré de dos tipos de Diccionarios de Categorías: un diccionario predefinido y un diccionario que se construye de forma automática basado en la generación automática de tópicos.

#### Uso de recursos léxicos predefinidos

Existen muchos recursos léxicos predefinidos, también llamados Diccionarios Definidos por el Usuario o DDU. Dependiendo del problema y el lenguaje estudiado, los DDU pueden ser muy específicos para un problema, digamos, depresión; o generales. Para el caso de la identificación automática de la personalidad el recurso léxico más utilizado en la literatura es LIWC (*Linguistic Inquiry and Word Count*) (Pennebaker et al., 2001) que es considerado un DDU general.



### 6.3 Autoselección de categorías

El diccionario LIWC está organizado en cuatro dimensiones principales: I) Dimensiones Lingüísticas Estándar, II) Procesos Psicológicos, III) Inquietudes Personales, y IV) Categorías del habla. La Tabla 6.3 muestra la organización del diccionario LIWC como se utiliza en esta investigación. Dado que el objetivo en esta tesis es distinguir personalidad, propongo el uso de las 6 categorías de palabras en la dimensión II (Procesos psicológicos) al mismo nivel que el resto de las dimensiones de LIWC (en la tabla las categorías están numeradas en la primera columna bajo #).

Cada categoría en LIWC tiene un conjunto de subcategorías como una organización más fina. De esta forma, cada término/palabra en LIWC puede asignarse a una o más subcategorías, y por consiguiente, a una o más categorías. Cabe hacer notar que existe una décima categoría donde se encontrarán todos los términos/palabras que no están en LIWC pero sí en los vocabularios de las comunidades a contrastar. Esta última categoría tiene como objetivo incluir en cierta medida vocabulario abierto —vocabulario que no pertenece a un recurso léxico cerrado como LIWC—. En total se agrupan términos dentro de 10 categorías.

#	Dimensión (Categoría)	Subcategorías
1	<b>I. Dimensiones lingüísticas estándar</b>	Palabras de función, Pronombres personales, 1ra. singular, 1ra. plural, 3ra. singular, 3ra. plural, Pronombres impersonales, Artículos, Verbos, Verbos auxiliares, Tiempo pasado, Tiempo presente, Tiempo futuro, Adverbios, Preposiciones, Conjunciones, Negaciones, Cuantificadores, Números, Maldiciones, ...
<b>II. Procesos psicológicos</b>		
2	Procesos sociales	Familia, Amigos, Humanos
3	Procesos afectivos	Emociones positivas, Emociones negativas, Ansiedad, Enfado, Tristeza
4	Procesos cognitivos	<i>Insight</i> , Causa, Discrepancia, Tentativos, Certeza, Inhibición, Inclusión, Exclusión
5	Procesos perceptuales	Ver, Oír, Sentir
6	Procesos biológicos	Cuerpo, Salud, Sexual, Ingerir
7	Relatividad	Movimiento, Espacio, Tiempo
8	<b>III. Inquietudes personales</b>	Trabajo, Logro, Placer, Hogar, Dinero, Religión, Muerte
9	<b>IV. Categorías del habla</b>	Asentir, No-fluido, Rellenos
10		OOV (términos fuera del vocabulario)

**Tabla 6.3:** Organización del Diccionario de Categorías LIWC para el idioma español (versión 2007). El conjunto de términos/palabras se pueden encontrar en Pennebaker et al. (2001). En el último renglón se muestra una categoría adicional OOV (*out of vocabulary*) que se utiliza en esta investigación pero que no forma parte de LIWC.

### Uso de tópicos generados automáticamente

Dado que no siempre se cuenta con recursos léxicos predefinidos que puedan ser útiles para los problemas de identificación de procesos mentales, presento una opción automática de construir un diccionario de categorías. Este proceso automático consiste en identificar tópicos en un conjunto de documentos de la tarea a resolver. Posteriormente, el conjunto de los tópicos generados con sus respectivos términos/palabras asociados se pueden usar como el recurso léxico. A este recurso generado automáticamente lo llamaré **Diccionario de Tópicos** (DT) para diferenciarlo de los DC predefinidos; pues los últimos asignan nombres a las categorías con un significado semántico preestablecido. Para generar los tópicos de forma automática se propone el uso de *Latent Dirichlet Allocation*.

*Latent Dirichlet Allocation* o LDA es un modelo generativo probabilístico propuesto por Blei et al. (2003) que utiliza un conjunto de documentos y asume que estos están representados por una distribución aleatoria de tópicos latentes, donde cada tópico se representa por una distribución de términos. Para generar el modelo, hay que definir varios parámetros, entre el más relevante está el valor de tópicos a generar, este parámetro se puede establecer de forma empírica.

Igual que con LIWC, cada tópico generado contiene un conjunto de términos y cada término puede pertenecer a uno o más tópicos. A diferencia de LIWC, los términos pertenecen a un tópico con cierta probabilidad. Una vez que se generan los tópicos, el Diccionario de Tópicos resultante se utiliza de la misma forma que se utilizaría el Diccionario de Categorías por el método propuesto en esta investigación.

Aunque los tópicos generados no tienen un nombre específico como es el caso de las categorías de LIWC, un humano podría generarles nombres partiendo de los términos que cada tópico contiene. La Tabla 6.4 muestra un ejemplo de 4 tópicos del modelo generado para el conjunto de datos RxPI en el rasgo de *Extroversión* y 10 tópicos en total.

Como puede verse al comparar la Tabla 6.3 y la Tabla 6.4 el tipo de información que se obtiene en el Diccionario de Tópicos es muy diferente a la información de un Diccionario de Categorías como LIWC. Una de las principales diferencias entre estos dos tipos de Diccionarios es que con LIWC, los términos dentro de cada categoría tienen un significado semántico. Es decir, todos los términos dentro de la categoría Procesos Afectivos, tendrán relación con este tipo de proceso psicológico de acuerdo con las personas que organizaron y generaron el recurso.

La calidad del DT dependerá de los conjuntos de datos que se utilicen para generar el diccionario. En esta propuesta se usan ensayos muy parecidos (documentos homogéneos) y por lo tanto, los tópicos generados pueden contener muchos términos en común. Sin embargo, una ventaja de los DT es que se puede hacer uso de la mayoría de los términos que realmente aparecen en el problema que se intenta resolver; es decir, en un DT se combina el uso del vocabulario abierto

## 6.4 Esquema general de las contribuciones

Tópico (posible nombre)	Términos con mayor probabilidad
Tópico 1 (empezando vida universitaria)	familia, ser, vida, hermanos, gusto, mude, aproximadamente, universidad, trabajo, papa, mejores, siempre, mayor, día, comencé, forma, madre, si, aunque, tampoco, ...
Tópico 2 (estudios)	siempre, cosas, familia, enseñado, ayudar, gusta, carrera, hacer, preparatoria, estudiando, diferentes, estudios, puedo, nuevo, cambio, padres, considero, ser, relevante, ello, ...
Tópico 3 (mención del núcleo familiar)	bien, persona, hermana, familia, gusta, sido, personas, madre, cosas, actualmente, mamá, considero, padres, mayor, ser, abuelitos, demasiado, padre, unida, dos, ...
Tópico 4 (mención de gustos)	momentos, siempre, ver, pasado, mismo, año, así, mayor, diferentes, vida, día, voces, gusta, chica, algunas-veces, nombre, seguir, varios, padres, cosas, ...

**Tabla 6.4:** Primeros 4 tópicos del diccionario de tópicos construido con documentos del conjunto de datos RxPI-estratificado para el rasgo de *Extraversión*. Para la generación de este modelo se utilizó la biblioteca de python *gensim*

con un enfoque de vocabulario cerrado.

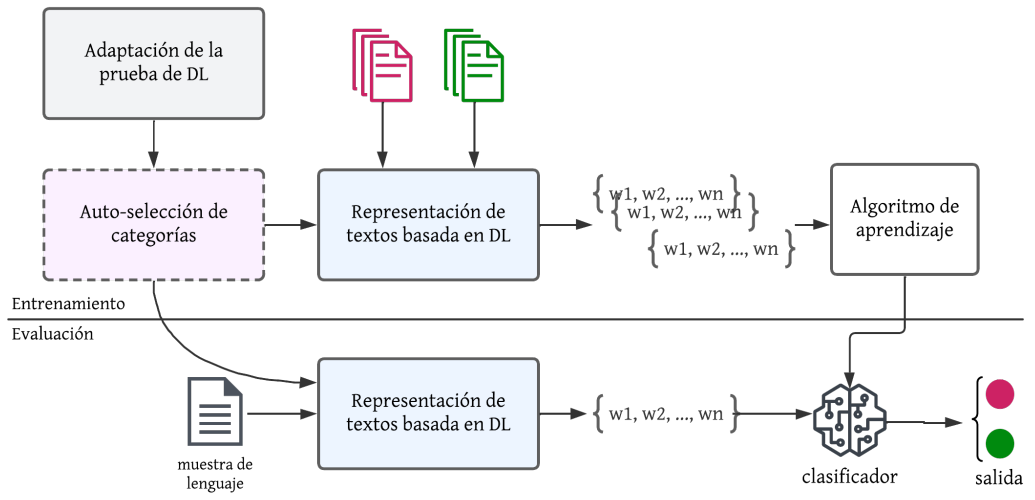
### 6.4. Esquema general de las contribuciones

La Figura 6.4 muestra cómo las contribuciones listadas en este capítulo interactúan para producir un modelo automático que dada una muestra de su lenguaje escrito permite identificar rasgos de la personalidad de una persona.

De forma general, el flujo de desarrollo de un modelo automático consiste en dos etapas: entrenamiento y evaluación. Durante el entrenamiento, después del módulo de *adaptación de la prueba de disponibilidad léxica* (descrito en la Sección 6.1) se puede optar por utilizar el módulo de *autoselección de categorías* (Sección 6.3) o se puede pasar directamente al módulo de *representación de textos basada en disponibilidad léxica* (Sección 6.2).

Durante la etapa de entrenamiento, el módulo de *representación de textos basada en disponibilidad léxica* encuentra el vocabulario que formará parte de las dimensiones de los vectores de cada instancia; en la etapa de evaluación, se representan las muestras de lenguaje en el vocabulario establecido y el pesado o valores del vector corresponden a lo descrito en la Sección 6.2.1 y la Sección 6.2.2, respectivamente.

## Capítulo 6. Método propuesto



**Figura 6.4:** Esquema general de un método automático para la identificación de rasgos de personalidad

# 7



## Identificación automática de la personalidad: experimentación y análisis

La hipótesis principal en esta tesis establece que a través del uso de una representación de texto basada en el análisis del lexicón disponible extraído de muestras de lenguaje, y bajo un esquema de aprendizaje supervisado, es posible identificar rasgos de la personalidad de una persona. Para comprobar esta hipótesis, realicé un conjunto de experimentos de clasificación supervisada con el método propuesto descrito en el Capítulo 6.

En estos experimentos una comunidad corresponde a una clase. Así, construí un clasificador automático para distinguir entre dos clases de un mismo rasgo de personalidad —polos opuestos del mismo rasgo: alto y bajo—. Por lo tanto, los experimentos realizados corresponden a un problema de clasificación binaria por cada rasgo de personalidad en el modelo *Big-Five*, i. e. se evalúa el método propuesto en 5 problemas binarios.

La primera sección este capítulo muestra una descripción de la configuración experimental, como el preprocesamiento realizado a las muestras, los parámetros particulares que se utilizaron para el cálculo de la disponibilidad léxica, las estrategias de evaluación utilizadas y se lista los algoritmos de aprendizaje con los que se construyeron los clasificadores binarios.

La segunda sección del capítulo presenta el primer experimento realizado. Este experimento tiene como objetivo determinar la pertinencia del uso del léxico disponible para representar un documento de texto en la tarea de la identificación automática de la personalidad, así como determinar el impacto del uso del valor de disponibilidad de cada término dentro de la representación vectorial.

Y en la tercera y última sección se presenta el segundo experimento que consiste en determinar si al seleccionar de forma automática términos en categorías de un Diccionario de Categorías es posible mejorar el desempeño de un clasificador automático para identificar personalidad. La última parte de esta sección permite comparar el desempeño de clasificación al utilizar dos tipos

de diccionario de categorías: predefinido y autogenerado.

### 7.1. Configuración experimental

#### Preprocesamiento

Durante el preprocesamiento de los tres conjuntos de datos utilizados en la evaluación, la normalización de los textos se realizó mediante el *software* FreeLing (Carreras et al., 2004) tanto para español (en RxPI–estratificado y Auto–descripciones) como para inglés (en Essays).

Dado que el objetivo es distinguir entre el polo de un rasgo de personalidad, se eliminó información que no es relevante para el problema. Específicamente, cuando se utiliza el modelo de rasgos que emplea la hipótesis léxica (Goldberg, 1993) las palabras de función no aportan información relevante. Por ello, durante el preprocesamiento se eliminan números, símbolos y palabras vacías (o de función).

Además, mediante Freeling se identifican multipalabras. Por ejemplo, si se identifica la cadena *a pesar de* se convierte en *a-pesar-de*, nombres propios como *Los Angeles*, se convierten en *los-angeles*. Esto tiene como objetivo preservar cadenas de más de una palabra que tienen un significado más allá de las palabras vacías que usualmente forman parte de ellas. Si el problema a resolver es distinto a la identificación de personalidad este paso de normalización puede adaptarse con conocimiento *a priori* al problema a tratar.

#### Cálculo de la disponibilidad léxica

Para el cálculo de la disponibilidad léxica los parámetros  $w$  y  $m$  de  $\theta$  en la Fórmula 6.1 se fijan empíricamente en  $w = 0.01$  and  $m = 1$ . El parámetro  $k$  de la misma fórmula se determina de acuerdo con el conjunto de entrenamiento y se establece como un valor constante para ambas clases, los valores considerados son  $k = 100$ ,  $k = 500$ , y  $k = 1000$ ; dependiendo del tamaño de los conjuntos de datos se eligió el valor correspondiente. Esto permitirá hacer comparaciones menos sesgadas entre las listas de disponibilidad léxica de las comunidades/clases a contrastar.

#### Estrategias de evaluación

La estrategia de evaluación depende del tamaño del corpus utilizado. Por un lado, para los conjuntos de datos en español (con aproximadamente 100 instancias por rasgo) se utiliza *leave-one-out*. En esta estrategia, si  $n$  es el tamaño del corpus, entonces se usan  $n - 1$  instancias para entrenar y se evalúa en la restante; esto se repite hasta evaluar en las  $n$  instancias que conforman el corpus. Por otro lado, para el conjunto de datos en inglés que tiene más de 2000 instancias por

---

## 7.2 Experimento 1: Disponibilidad léxica para una representación textual

---

rasgo, se utiliza una estrategia de evaluación cruzada de diez pliegues (o *10 fold cross validation*). Las métricas de evaluación consideradas son exactitud (Ex.), macro precisión (P), macro recuerdo (R) y macro f-score (F). Sin embargo, para comparar entre métodos se utiliza macro f-score, principalmente por el desbalance que existe en los conjuntos de datos —especialmente en los conjuntos de datos pequeños—.

### Algoritmos de aprendizaje

Durante el entrenamiento de los clasificadores binarios utilicé diferentes tipos de algoritmos provistos por la biblioteca `scikit-learn`<sup>1</sup> de Python, a continuación se listan los 7 algoritmos usados con sus parámetros respectivos:

- Complement Naive Bayes (CNB)
- Multinomial Naive Bayes (MNB)
- Perceptron classifier (Perc): with `random_state = 42`
- Multi-layer Perceptron classifier (MLP): with `solver = 'lbfgs'`, `alpha = 1e-5`, `hidden_layer_sizes = (5, 2)`, `random_state = 42`,
- Linear Support Vector Classifier (LSVC): with `random_state=42`,
- Decision Tree classifier (DT): with `criterion='entropy'`, `random_state=42` y
- Logistic Regression classifier (LR): with `random_state=42`.

## 7.2. Experimento 1: Disponibilidad léxica para una representación textual

Este primer conjunto de experimentos tiene dos objetivos:

- (a) Determinar la pertinencia del uso del léxico disponible —extraído a través de la prueba de disponibilidad léxica— para representar documentos de texto en la tarea de identificación del polo de un rasgo de personalidad. Esto es, la variante de representación básica o DL-base, descrita en la Sección 6.2.1. DL-base considera como vocabulario aquellas palabras que aparecen en la unión de las listas de disponibilidad léxica de las dos clases; es decir, aquellos términos disponibles calculados durante la etapa de entrenamiento del modelo.

---

<sup>1</sup><https://scikit-learn.org>

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

- (b) Determinar el impacto del uso del valor de disponibilidad de cada término como parte de la representación vectorial. Es decir, la variante de representación DL-va, descrita en la Sección 6.2.2, que no solo considera los términos disponibles, sino también considera el valor de accesibilidad/disponibilidad.

Cada uno de estos objetivos corresponde a una variante del método de representación basada en disponibilidad léxica como se puede ver en la Figura 6.2. Como referencias (*baselines*) para comparar el impacto del uso de la disponibilidad léxica en la representación de los textos utilicé un método de asignación de clases y dos métodos de representación en el esquema general presentado en el Capítulo 6. Del más simple al más complejo, son:

- Clase mayoritaria. La clase asignada a cada instancia de evaluación es la clase que aparece en la mayoría de las instancias de entrenamiento en el conjunto de datos.
- BoW (bolsa de palabras). Esta representación consiste en la utilización de todo el vocabulario disponible en las instancias de entrenamiento, después del preprocesamiento, como dimensiones del vector de cada instancia.
- Representación basada en arquitecturas *transformers*. Para esta representación se omitió el paso de preprocesamiento; para el idioma inglés utilicé el modelo pre-entrenado BERT (Devlin et al., 2018) mientras que para el idioma español usé BETO (Cañete et al., 2020).

Los dos primeros métodos son comúnmente usados en tareas de clasificación automática de textos como *baselines*; mientras que el último utiliza una de las nuevas representaciones usadas en tareas recientes de clasificación automática de textos.

La Tabla 7.1 presenta los resultados de este primer conjunto de experimentos. La columna *inc.* de la tabla muestra el incremento absoluto (positivo o negativo) que cada método tiene en comparación con el *baseline* BoW. En las siguientes secciones se describen los resultados obtenidos en los tres conjuntos de datos.

### RxPI-estratificado

De los resultados de la Tabla 7.1, vemos que utilizar una representación basada en el léxico disponible (DL-base y DL-va), en el corpus RxPI-estratificado, obtiene mejores resultados de clasificación que utilizar cualesquiera de los métodos de referencia —incluyendo la representación basada en *Transformers* (del enfoque de aprendizaje de características)—.

Por un lado, la variante DL-base obtiene una mejora en el desempeño de clasificación en el macro *f*-score desde 0.03 (para el rasgo de *Extroversión*) y hasta arriba del 0.10 (para los rasgos



## 7.2 Experimento 1: Disponibilidad léxica para una representación textual

---

de *Amabilidad* y *Responsabilidad*). Por otro lado, al utilizar el valor de la disponibilidad, i. e. DL-va, hay un incremento de todos los rasgos en relación con BoW. Estos incrementos van desde el 0.03 y 0.04 en el valor de f-score para los rasgos de *Extroversión*, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*; hasta un incremento del 0.11 para el rasgo de *Amabilidad*.

Para el caso de rasgos con clases altamente desbalanceadas como es el rasgo de *Responsabilidad*, que cuenta con 10 y 84 instancias en cada clase (como se indica en la Tabla 5.4), utilizar DL-base y DL-va provee un incremento en el macro f-score del 0.13 y 0.07, respectivamente. Mientras que el método basado en *Transformers* obtiene un desempeño similar al método de clasificación de la clase Mayoritaria, esto indica que ninguno de estos dos métodos de referencia pueden generalizar con tan limitada información.

### Auto-descripciones

Para el corpus Auto-descripciones, en la Tabla 7.1 se observa que la variante DL-base mejora el desempeño en 3 de los cinco rasgos de personalidad con un incremento en f-score de entre el 0.02 (de 0.52 a 0.54) para *Amabilidad* y 0.15 para el rasgo de *Estabilidad Emocional* (de 0.52 a 0.67) y para *Apertura a Nuevas Experiencias* (de 0.59 a 0.74) en comparación con BoW.

Para dos rasgos —*Extroversión* y *Responsabilidad*— no existe evidencia de que un único método de referencia es mejor. Por ejemplo, para el rasgo de *Extroversión*, el mejor resultado se obtiene con el uso de BETO (modelo *transformer* para español) —el incremento absoluto en relación con BoW es de 0.05 en f-score—. Mientras tanto, DL-base tiene un decremento de 0.02 en relación con BoW y de 0.07 en relación con BETO. Para el caso del rasgo de *Responsabilidad*, ningún método es mejor que el basado en BoW; pero solo la representación basada en léxico disponible DL-va tiene un desempeño equivalente a BoW (ambos con un valor de f-score de 0.61).

### Essays

Los resultados en la Tabla 7.1 con el conjunto de datos en inglés Essays muestran que no existe una clara evidencia que algún método es mejor que otro en los cinco rasgos, a diferencia de lo que indican los resultados para los conjuntos de datos anteriores.

En tres de los 5 rasgos, el mejor desempeño se observa al utilizar la representación basada en BERT (*Amabilidad*, *Responsabilidad*, y *Estabilidad Emocional*); sin embargo la mejora al utilizar BERT contra solamente usar una representación más simple basada en BoW es mínima, el incremento está entre el 0.01 y el 0.03. Esto propone cuestionarse si esta pequeña mejora es suficiente para utilizar un método complejo y costoso.

**Tabla 7.1:** Resultados de clasificación de personalidad obtenidos en los conjuntos de datos RxPI-estratificado, Auto-descripciones y Essays. Se compara la representación basada en el léxico disponible (DL-base y DL-va) contra una representación basada en bolsa de palabras (BoW) y modelos basados en *transformers* (BERT y BETO). Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonalidades de verde indican un incremento de F en comparación con el mejor resultado de referencia.

Método	RxPI-estratificado					Auto-descripciones					Essays								
	Clasif.	Ex.	P	R	F	inc.	Clasif.	Ex.	P	R	F	inc.	Clasif.	Ex.	P	R	F	inc.	
Ext.	Mayoritaria	-	0.57	0.28	0.50	0.36	-0.28	-	0.75	0.37	0.50	0.42	-0.22	-	0.51	0.25	0.50	0.34	-0.23
	BoW	DT-bool	0.67	0.66	0.65	0.65		DT-bool	0.70	0.63	0.65	0.64		MNB-ff	0.57	0.57	0.57	0.57	
	Transformer	MLP-beto	0.64	0.64	0.64	0.64	-0.01	DT-beto	0.75	0.68	0.71	0.69	+0.05	LR-bert	0.56	0.56	0.56	0.56	-0.01
	DL-base	DT-ff	0.69	0.68	0.68	0.68	+0.03	DT-bool	0.69	0.61	0.63	0.62	-0.02	CNB-mtf	0.58	0.58	0.58	0.58	+0.01
	DL-va	DT-bool	0.69	0.68	0.68	0.68	+0.03	LR-ff	0.72	0.59	0.56	0.57	-0.07	RF-mtf	0.56	0.56	0.55	0.55	-0.02
Ana.	Mayoritaria	-	0.55	0.27	0.50	0.35	-0.17	-	0.67	0.33	0.50	0.40	-0.12	-	0.53	0.26	0.50	0.34	-0.19
	BoW	DT-ff	0.53	0.52	0.52	0.52		MLP-ff	0.61	0.53	0.52	0.52		LR-mtf	0.53	0.53	0.53	0.53	
	Transformer	MLP-beto	0.59	0.59	0.59	0.59	+0.07	MLP-beto	0.58	0.51	0.51	0.51	-0.01	LR-bert	0.56	0.56	0.56	0.56	+0.03
	DL-base	DT-ff	0.64	0.64	0.64	0.64	+0.12	DT-ffidf	0.60	0.54	0.54	0.54	+0.02	CNB-mtf	0.56	0.56	0.56	0.56	+0.03
	DL-va	DT-ff	0.63	0.63	0.63	0.63	+0.11	MLP-ff	0.74	0.70	0.66	0.67	+0.15	CNB-bool	0.55	0.54	0.54	0.54	+0.01
Resp.	Mayoritaria	-	0.89	0.44	0.50	0.47	-0.08	-	0.66	0.33	0.50	0.40	-0.21	-	0.50	0.25	0.50	0.33	-0.23
	BoW	DT-ffidf	0.75	0.55	0.59	0.55		DT-ff	0.63	0.61	0.62	0.61		LR-ffidf	0.56	0.56	0.56	0.56	
	Transformer	LR-beto	0.89	0.44	0.50	0.47	-0.08	DT-beto	0.63	0.57	0.56	0.56	-0.05	RF-bert	0.57	0.57	0.57	0.57	+0.01
	DL-base	DT-ffidf	0.88	0.69	0.67	0.68	+0.13	LR-ff	0.63	0.57	0.56	0.56	-0.05	CNB-ff	0.56	0.56	0.56	0.56	=
	DL-va	DT-mtf	0.86	0.63	0.61	0.62	+0.07	Perc-ff	0.65	0.61	0.61	0.61	=	CNB-mtf	0.56	0.56	0.56	0.56	=
Est.	Mayoritaria	-	0.56	0.28	0.50	0.36	-0.22	-	0.53	0.26	0.50	0.35	-0.33	-	0.49	0.49	0.49	0.49	-0.07
	BoW	DT-ff	0.62	0.61	0.59	0.58		DT-mtf	0.69	0.69	0.68	0.68		LR-ffidf	0.56	0.56	0.56	0.56	
	Transformer	MLP-beto	0.56	0.55	0.55	0.55	-0.03	LR-beto	0.51	0.50	0.50	0.50	-0.18	LR-bert	0.59	0.59	0.59	0.59	+0.03
	DL-base	MLP-ff	0.58	0.59	0.59	0.58		DT-mtf	0.73	0.73	0.73	0.73	+0.05	CNB-bool	0.58	0.58	0.58	0.58	
	DL-va	CNB-bool	0.62	0.62	0.62	0.62	+0.04	DT-mtf	0.65	0.65	0.65	0.65	-0.03	LR-bool	0.56	0.56	0.56	0.56	=
Ape.	Mayoritaria	-	0.61	0.30	0.50	0.37	-0.20	-	0.75	0.37	0.50	0.42	-0.17	-	0.51	0.25	0.50	0.34	-0.29
	BoW	LSVC-bool	0.60	0.58	0.57	0.57		DT-bool	0.72	0.60	0.59	0.59		MNB-ffidf	0.63	0.63	0.63	0.63	
	Transformer	DT-beto	0.62	0.60	0.59	0.59	+0.02	RF-beto	0.75	0.37	0.50	0.42	-0.17	LR-bert	0.61	0.61	0.61	0.61	-0.02
	DL-base	RF-ff	0.67	0.66	0.62	0.62	+0.05	DT-mtf	0.82	0.76	0.72	0.74	+0.15	CNB-mtf	0.62	0.62	0.62	0.62	-0.01
	DL-va	RF-ff	0.64	0.61	0.60	0.60	+0.03	DT-mtf	0.77	0.69	0.69	0.69	+0.10	CNB-bool	0.61	0.61	0.61	0.61	-0.02

## 7.2 Experimento 1: Disponibilidad léxica para una representación textual

---

Cabe mencionar que las diferencias entre los métodos contrastados, BoW, basados en *Transformers* y las dos variantes del método propuesto, varían en incrementos o decrementos de 0.01 en el valor de f-score en tres rasgos: *Extroversión*, *Responsabilidad* y *Apertura a Nuevas Experiencias*. Dado estas mínimas diferencias, se intuye que, el corpus Essays en particular tiene características que hacen aún más compleja la identificación de la personalidad. Estos resultados también permiten concluir que al incluir información del valor de disponibilidad léxica en la representación, mediante la variante DL-va, no refleja una mejora en el desempeño en ninguno de los cinco rasgos con relación a la variante DL-base para este corpus.

De los resultados obtenidos con el conjunto de datos Essays y de los resultados de investigaciones previas en este mismo conjunto de datos (ver Tabla 4.1 de la Sección 4.2.2), es posible realizar algunas comparaciones en el desempeño de métodos para identificar rasgos de personalidad en Essays. Al utilizar métodos complejos basados en redes neuronales (últimos renglones en la Tabla 4.1) es posible obtener valores ligeramente mejores en la clasificación binaria a los que se obtienen con el método propuesto basado en disponibilidad léxica (ver Tabla 7.1). Estas mejoras en la métrica de exactitud —columna Ex.— varían entre apenas un 0.02 (para *Estabilidad Emocional*) y hasta un 0.09 (para *Apertura a Nuevas Experiencias*).

Estas mismas diferencias son las que se observan al comparar esos métodos complejos basados en redes neuronales contra el trabajo descrito en Mairesse et al. (2007); excepto por el rasgo de *Extroversión*, donde el método basado en disponibilidad léxica, particularmente la variante DL-base, consigue un valor más grande de exactitud al obtenido por Mairesse et al. (2007) (0.58 vs 0.52). De aquí se puede observar que la identificación de personalidad, específicamente en el corpus de Essays, no es una tarea sencilla. En la siguiente sección se analizarán a más detalles los resultados mostrados hasta ahora.

### 7.2.1. Análisis sobre el uso de la disponibilidad léxica

Los resultados del primer experimento permiten derivar conclusiones parciales. Primero, los resultados muestran que utilizar la disponibilidad léxica como método de representación apoya la clasificación automática de textos para identificar rasgos de personalidad. Además, una de las ventajas de utilizar una representación basada en disponibilidad léxica sobre utilizar una representación basada en bolsa de palabras es que el vocabulario utilizado está basado en instrumentos psicolingüísticos que tiene un significado dentro de la psicología y puede ayudar a los especialistas a entender mejor el lenguaje producido; pues la lista de disponibilidad léxica es interpretable por expertos.

Segundo, cuando existe poca información para identificar la personalidad basada en textos, realizar una representación basada en un modelo cognitivo como es la disponibilidad léxica,

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

es más importante. Puede notarse que con los corpora pequeños RxPI–estratificado y Auto–descripciones, la mejora de utilizar DL–base o DL–va es más grande que en el corpus de ensayos (Essays). Hay dos posibilidades para que la disponibilidad léxica no tenga un efecto más grande en este último corpus, por un lado la cantidad de información disponible (2000% más instancias que en los corpora en español) ayuda a métodos de representación más tradicionales como BoW o más actuales como BERT que requieren una mayor cantidad de información para generar un clasificador confiable; por otro lado la diversidad de temas abordados por los textos en Essays es mayor.

Con respecto a la diversidad temática es importante recordar cómo están construidos los conjuntos de datos. Mientras que RxPI y Auto–descripciones tienen instrucciones acotadas, i. e. contar *un* evento familiar relevante, o una autopresentación en *3 minutos*; Essays fue construido con una instrucción más general, i. e. escribir lo primero que se te venga a la mente durante 20 minutos. RxPI y Auto–descripciones son obtenidos dentro de una institución educativa, lo que provoca pensamientos relacionados con el entorno en el que se encuentran. Para Essays, la actividad puede realizarse en cualquier lugar y a cualquier hora, por lo que el tren de pensamiento de más de 2000 participantes pueden verse afectados por estas variables. Por lo tanto, es probable que los corpora con instrucciones más específicas contengan poca variación de temas abordados sin importar el rasgo de personalidad que cada persona exhiba.

Históricamente, métodos de clasificación textual han tenido dificultades en tareas donde no existe mucha variación temática entre las instancias de las diferentes clases; especialmente cuando dentro de la representación, solo se considera el contenido de los textos. Para medir cómo se ve afectado el contenido temático de cada conjunto de datos propongo una medida de dispersión temática. Esta medida de dispersión temática podría ayudar a determinar cuándo utilizar un enfoque basado en la disponibilidad léxica.

### 7.2.2. Dispersión temática

La dispersión temática de un conjunto de datos la defino en esta tesis como el promedio del número de temas —o tópicos— que se abordan en cada comunicación escrita de dicho conjunto. Un conjunto de datos con una dispersión temática alta incluye instancias que abordan un mayor número de temas. Mientras que un conjunto con una dispersión temática baja, indica que las instancias en ese corpus hablan de pocos temas. El valor de la dispersión temática calculada será el promedio de temas que se abordan por documento en el conjunto de datos.

Para medir la dispersión temática propongo utilizar un modelo automático de generación de tópicos, i. e. *Latent Dirichlet Allocation* (LDA) (Blei et al., 2003). Así, construyo un modelo de tópicos con un conjunto de datos disponible  $\mathcal{D}$ . Dado el modelo de tópicos construido con

## 7.2 Experimento 1: Disponibilidad léxica para una representación textual

$i$  tópicos,  $\mathfrak{M} = \{t_1, t_2, \dots, t_i\}$ , la cantidad de tópicos  $nt_d$  por cada documento  $d \in \mathcal{D}$  será igual al número de tópicos en  $d$  con una probabilidad mayor o igual a  $p$ ; donde  $p = 1/i$ . Luego la dispersión temática,  $DispTem$ , del conjunto de documento,  $\mathcal{D}$ , queda definida como el promedio del número de tópicos en los documentos del conjunto, como lo expresa la Ecuación 7.1.

$$DispTem_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} nt_d \quad (7.1)$$

Para validar el modelo de tópicos generado realicé una búsqueda de los mejores valores para los 3 parámetros del algoritmo LDA — $i$ ,  $\alpha$ , y  $\beta$ — (Blei et al., 2003):  $i \in \{2, \dots, 10\}$ ,  $\alpha \in \{0.01, 0.31, 0.61, 0.91\}$  y  $\beta \in \{0.01, 0.31, 0.61, 0.91\}$  considerando la perplejidad del modelo como la métrica de desempeño para seleccionar el mejor (un menor valor de perplejidad indica un desempeño de generalización mejor). Este proceso se repitió de forma independiente para cada conjunto de instancias por rasgo por conjunto de datos. Así, la Tabla 7.2 presenta información generada con el modelo con mejor perplejidad ( $i = 10$ ,  $\alpha = 0.91$  y  $\beta = 0.01$ ) para cada rasgo. También se muestra nuevamente el número de instancias por cada rasgo en cada conjunto de datos y la longitud promedio de ellas; en la última columna se muestra el valor de dispersión temática calculado. En general, dado que el número de tópicos usado es  $i = 10$  para todos los casos, es posible hacer comparaciones directas entre los tres corpora.

		modelo LDA			Tamaño del corpus		$DispTem$
		$i$	Perplex.	$c_v$	# Instancias	Long. prom.	
RxPI-estr.	Ext.	10	-8.79	0.36	71	39.59 ±24.14	1.55 ±0.85
	Ama.	10	-8.97	0.36	94	42.56 ±22.15	1.66 ±0.81
	Res.	10	-8.78	0.35	94	37.47 ±23.08	1.84 ±1.05
	Est.	10	-8.66	0.38	53	34.16 ±18.47	1.64 ±0.09
	Ape.	10	-8.90	0.32	134	40.29 ±21.11	1.83 ±1.02
Auto-desc.	Ext.	10	-7.45	0.33	65	99.50 ±49.43	1.47 ±0.79
	Ama.	10	-7.52	0.35	65	102.72 ±50.64	1.61 ±0.96
	Res.	10	-7.51	0.27	60	100.21 ±47.94	1.48 ±0.72
	Est.	10	-7.48	0.39	78	101.52 ±48.74	1.52 ±0.83
	Ape.	10	-8.91	0.32	61	97.57 ±49.58	1.63 ±0.94
Essays	Ext.	10	-8.89	0.28	2467	243.19 ±98.11	2.35 ±1.04
	Ama.	10	-8.94	0.28	2467	243.19 ±98.11	2.33 ±1.04
	Res.	10	-8.88	0.27	2467	243.19 ±98.11	2.34 ±1.06
	Est.	10	-8.94	0.28	2467	243.19 ±98.11	2.36 ±1.00
	Ape.	10	-8.97	0.28	2467	243.19 ±98.11	2.40 ±1.05

**Tabla 7.2:** Dispersión temática en tres conjuntos de datos y 5 rasgos de personalidad. En la columna *modelo LDA*,  $i$  es el número de tópicos generados, *Perplex.* el valor de perplejidad obtenido por el modelo y  $c_v$  es el valor de cohesión del modelo.

La Tabla 7.2 muestra que el tamaño de las instancias (longitud promedio de términos en cada instancia) es diferente para los 3 conjuntos de datos. Por un lado, en el corpus RxPI–estratificado, la longitud promedio de las comunicaciones obtenidas es de menos de 50 términos y una  $DispTem$  entre 1.55 y 1.84; es decir, cada instancia de RxPI–estratificado aborda en promedio menos de 2 temas. Por otro lado, las instancias en Auto–descripciones tienen en promedio más del doble de términos que RxPI–estratificado pero un valor de  $DispTem$  ligeramente menor, entre 1.47 y 1.63.

Los corpora RxPI–estratificado y Auto–descripciones tiene en común el reducido número de instancias por rasgos, menos de 100. En contraste, el conjunto Essays tiene 2467 instancias para cada rasgo con una longitud promedio de 243 términos —casi 6 veces más que RxPI–estratificado y 1.5 veces más que Auto–descripciones— y valores de  $DispTem$  entre 2.33 y 2.40; esto es, en promedio en cada instancia de Essays están presentes más de dos tópicos.

### 7.3. Experimento 2: Autoselección de categorías para mejorar la representación textual

El segundo conjunto de experimentos tiene dos objetivos:

- (a) Determinar si al seleccionar de forma automática términos, en categorías específicas dentro de un diccionario de categorías, para la representación de muestras del lenguaje es posible mejorar el desempeño de métodos automáticos para identificar el polo de cada rasgo de personalidad.
- (b) Comparar el desempeño de la autoselección con dos tipos de diccionarios de categorías: predefinido y autogenerado.

Estos objetivos corresponden a validar la utilidad del módulo de autoselección de categorías, basada en disponibilidad léxica, que de acuerdo al esquema general del método propuesto, puede o no utilizarse (como lo indica la línea punteada en la Figura 6.4).

Para este conjunto de experimentos se representan los textos utilizando DL–base y DL–va (como en los experimentos anteriores) pero enfatizando la selección de las categorías semánticas relevantes dependiendo del problema, como se describe en la Sección 6.3.1, con dos tipos de Diccionarios de Categorías (DC).

Así, la Sección 7.3.1 presenta experimentos utilizando como DC el recurso LIWC —un diccionario predefinido—; y luego la Sección 7.3.2 presenta experimentos utilizando un DT —recurso auto-generado por el algoritmo LDA—. Los experimentos que se realizaron muestran que, de forma

## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

general, el desempeño de clasificación obtenido con el diccionario autogenerado es menor en comparación con los resultados obtenidos al utilizar el diccionario predefinido LIWC. Finalmente, la Sección 7.3.3 presenta una discusión sobre las diferencias entre los dos tipos de Dicionarios de Categorías.

### 7.3.1. Autoselección de categorías predefinidas en LIWC

Las categorías semánticas que se utilizan en esta sección pertenecen al recurso psicolingüístico LIWC, descrito en el Capítulo 6, que son categorías predefinidas con términos fijos dentro de cada categoría. Las Tablas 7.3, 7.4, 7.5, 7.6 y 7.7 muestran los resultados de clasificación de seleccionar categorías relevantes mediante el orden de los términos en las listas de disponibilidad léxica de las dos clases a identificar en cada rasgo de personalidad. La columna  $n$  lista el valor del  $n$ -avo percentil utilizado para seleccionar las categorías. Entre mayor es el valor de  $n$  se filtran más categorías (se eligen menos). Cuando el valor de esa columna es 0 indica que todas las categorías presentes en el conjunto de entrenamiento son utilizadas, i. e. corresponde al resultado de DL-base y DL-va de la Tabla 7.1, por lo tanto sirven de referencia para comparar si la Autoselección de categorías mejora el desempeño de clasificación. A continuación se describen los resultados por cada conjunto de datos.

#### RxPI-estratificado

Para el conjunto de datos RxPI-estratificado es posible obtener mejoras en el desempeño de clasificación en 4 de los 5 rasgos de personalidad —*Extroversión*, *Amabilidad*, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*— con aumentos máximos por rasgo en el valor de  $f$ -score desde 0.05 (para *Apertura a Nuevas Experiencias*) a 0.17 (para *Estabilidad Emocional*).

Además, para 3 de los 5 rasgos —*Extroversión*, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*— la mejora sobre no realizar una selección de categorías se puede alcanzar en diversos valores del porcentaje de filtrado ya sea DL-base o DL-va. Por ejemplo, para *Extroversión* se obtienen mejores resultados en 5 de 9 opciones para DL-base y 4 de 9 tasas de filtrado para DL-va (resaltados en diferentes tonos de verde en la Tabla 7.3). Esto implica que la Autoselección de categorías es importante para este conjunto de datos.

En contraste, para el rasgo de *Responsabilidad* (ver Tabla 7.5) no se logra mejorar el desempeño de clasificación del método sin el filtrado de categorías —resaltado en gris en la tabla—; es decir, no hay mejoras en el desempeño cuando se utiliza Autoselección de categorías. Esto puede ser debido al alto desbalance de las instancias en cada clase y al reducido número de instancias de la clase *alta* (10 instancias en total para la clase *alta* vs. 84 instancias para la clase *baja*). Con la limitada cantidad de información en este rasgo en particular, es difícil generalizar con un

vocabulario limitado.

### Auto-descripciones

Con respecto al conjunto de datos Auto-descripciones, la Autoselección de categorías ayuda a mejorar el desempeño de clasificación de 3 de los 5 rasgos: en *Extroversión* con un aumento del 0.04 en f-score en comparación con no usar la Autoselección de categorías llegando a un valor de f-score de 0.73, en *Amabilidad* con un aumento de 0.03 (para llegar a un valor de f-score de 0.70) y en *Responsabilidad* con un incremento del 0.08 (llegando a 0.69 de f-score).

En los rasgos *Estabilidad Emocional* y *Apertura a Nuevas Experiencias* no se logra superar el desempeño de cuando se emplea la información de disponibilidad léxica sin el método de Autoselección de categorías. Sin embargo, el desempeño del método que solo utiliza información de la disponibilidad léxica —sin la Autoselección de categorías— es de 0.73 y 0.74 de f-score, respectivamente; que en perspectiva, es mejor que los valores de desempeño obtenidos con los métodos de referencia.

### Essays

En el caso del conjunto de datos Essays, como en el primer experimento, el método propuesto no mejora el desempeño de los métodos de referencia como en las representaciones basadas en BoW y BERT. Por lo que es de esperarse, la Autoselección de categorías tampoco permite mejorar los resultados registrados en la Tabla 7.1. En el mejor de los casos se logra obtener un desempeño igual al mejor resultado del primer experimento, esto para los rasgos de *Extroversión*, *Amabilidad* y *Responsabilidad* y únicamente cuando se utiliza la opción DL-base en la representación (estos valores están resaltados en azul en las tablas correspondientes de resultados de cada rasgo).

La cantidad de términos usados para la representación es una ventaja importante del método de Autoselección de categorías en este conjunto de datos, además de la identificación clara de los términos en sus respectivas categorías semánticas. Por ejemplo, para *Responsabilidad* usar la Autoselección de categorías, con un número reducido de términos ( $n = 60$ ), logra un desempeño igual que al utilizar una representación basada en *transformers* (con un valor de f-score de 0.57).



## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

**Tabla 7.3:** Resultados de clasificación del rasgo de *Extraversión* obtenidos en los conjuntos de datos RxPI-estratificado, Auto-descripciones y Essays, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonálidades de verde indican un incremento de F en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 7.1.

n	RxPI					Auto-descripciones					Essays							
	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
0	DT-f	0.69	0.68	0.68	0.68	=	DT-bool	0.69	0.61	0.63	0.62	-0.07	CNB-mtf	0.58	0.58	0.58	0.58	*
10	DT-f	0.66	0.65	0.65	0.65	-0.03	LR-f	0.72	0.59	0.56	0.57	-0.12	CNB-mtf	0.58	0.58	0.58	0.58	=
20	DT-bool	0.63	0.63	0.63	0.63	-0.05	MLP-f	0.74	0.62	0.57	0.58	-0.11	CNB-mtf	0.58	0.58	0.58	0.58	=
30	DT-bool	0.70	0.70	0.69	0.70	+0.02	DT-f	0.68	0.57	0.58	0.57	-0.12	MNB-mtf	0.58	0.58	0.58	0.58	=
40	DT-f	0.77	0.77	0.77	0.77	+0.09	LR-f	0.74	0.62	0.57	0.58	-0.11	CNB-mtf	0.57	0.57	0.57	0.57	-0.01
50	DT-f	0.77	0.77	0.76	0.77	+0.09	DT-ffidf	0.74	0.65	0.66	0.65	-0.04	MNB-bool	0.56	0.56	0.56	0.56	-0.02
60	DT-f	0.77	0.77	0.77	0.77	+0.09	DT-f	0.66	0.56	0.57	0.56	-0.13	MNB-bool	0.56	0.56	0.56	0.56	-0.02
70	Perc-bool	0.68	0.68	0.68	0.68	=	DT-ffidf	0.72	0.63	0.63	0.63	-0.06	CNB-iffidf	0.55	0.55	0.55	0.55	-0.03
80	RF-bool	0.70	0.70	0.69	0.69	+0.01	Perc-ffidf	0.62	0.49	0.49	0.49	-0.20	RF-mtf	0.55	0.55	0.55	0.55	-0.03
90	Perc-f	0.49	0.61	0.55	0.44	-0.24	Perc-f	0.57	0.49	0.48	0.48	-0.21	MNB-f	0.54	0.54	0.54	0.54	-0.04
0	DT-bool	0.69	0.68	0.68	0.68	*	DT-bool	0.68	0.60	0.62	0.60	-0.09	RF-mtf	0.56	0.56	0.56	0.55	-0.03
10	DT-bool	0.69	0.68	0.68	0.68	=	LR-f	0.74	0.63	0.60	0.60	-0.09	CNB-bool	0.56	0.56	0.56	0.56	-0.02
20	DT-f	0.68	0.67	0.67	0.67	-0.02	DT-mtf	0.68	0.54	0.53	0.53	-0.16	CNB-bool	0.56	0.56	0.56	0.56	-0.02
30	DT-f	0.76	0.76	0.74	0.75	+0.07	DT-mtf	0.72	0.60	0.58	0.59	-0.10	CNB-bool	0.55	0.55	0.55	0.55	-0.03
40	DT-f	0.79	0.78	0.78	0.78	+0.10	DT-mtf	0.71	0.55	0.53	0.53	-0.16	CNB-bool	0.55	0.55	0.55	0.55	-0.03
50	DT-f	0.76	0.76	0.74	0.75	+0.07	LSVC-f	0.72	0.59	0.56	0.57	-0.12	CNB-bool	0.56	0.56	0.56	0.56	-0.02
60	DT-f	0.73	0.73	0.73	0.73	+0.05	DT-f	0.66	0.58	0.59	0.58	-0.11	CNB-bool	0.55	0.55	0.55	0.55	-0.03
70	DT-bool	0.69	0.68	0.68	0.68	=	LSVC-f	0.71	0.59	0.57	0.58	-0.11	CNB-bool	0.54	0.54	0.54	0.54	-0.04
80	DT-bool	0.66	0.66	0.66	0.66	-0.02	MLP-bool	0.74	0.62	0.57	0.58	-0.11	CNB-bool	0.55	0.55	0.55	0.55	-0.03
90	CNB-bool	0.58	0.29	0.50	0.37	-0.31	Perc-f	0.83	0.81	0.70	0.73	+0.04	CNB-bool	0.54	0.54	0.54	0.54	-0.04

**Tabla 7.4:** Resultados de clasificación del rasgo de *Amabilidad* obtenidos en los conjuntos de datos RxPI-estratificado, Auto-descripciones y Essays, comparando el método de Autoselección de categorías con utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonalidades de verde indican un incremento de F en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 7.1.

n	RxPI				Auto-descripciones				Essays									
	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
0	DT- <i>tf</i>	0.64	0.64	0.64	0.64	*	DT- <i>tfidf</i>	0.60	0.54	0.54	0.54	-0.13	CNB- <i>mtf</i>	0.56	0.56	0.56	0.56	=
10	DT- <i>tf</i>	0.64	0.64	0.64	0.64	=	LR- <i>tf</i>	0.63	0.55	0.54	0.54	-0.13	CNB- <i>mtf</i>	0.55	0.55	0.55	0.55	-0.01
20	DT- <i>tf</i>	0.65	0.65	0.65	0.65	+0.01	MLP- <i>tfidf</i>	0.65	0.56	0.54	0.53	-0.14	CNB- <i>mtf</i>	0.56	0.55	0.55	0.55	-0.01
30	DT- <i>mtf</i>	0.55	0.54	0.54	0.54	-0.10	MLP- <i>mtf</i>	0.62	0.55	0.54	0.54	-0.13	CNB- <i>mtf</i>	0.56	0.56	0.56	0.56	=
40	LR- <i>bool</i>	0.55	0.55	0.55	0.55	-0.09	LR- <i>tf</i>	0.65	0.58	0.56	0.57	-0.10	CNB- <i>tf</i>	0.56	0.56	0.56	0.56	=
50	DT- <i>tf</i>	0.59	0.58	0.58	0.58	-0.06	MLP- <i>bool</i>	0.62	0.55	0.54	0.54	-0.13	CNB- <i>bool</i>	0.55	0.55	0.55	0.55	-0.01
60	Perc- <i>tf</i>	0.59	0.58	0.57	0.56	-0.08	MNB- <i>tf</i>	0.71	0.66	0.61	0.61	-0.06	CNB- <i>idf</i>	0.55	0.55	0.55	0.55	-0.01
70	CNB- <i>bool</i>	0.54	0.55	0.55	0.54	-0.10	DT- <i>mtf</i>	0.66	0.61	0.61	0.61	-0.06	CNB- <i>idf</i>	0.54	0.54	0.54	0.54	-0.02
80	Perc- <i>bool</i>	<b>0.80</b>	<b>0.80</b>	<b>0.79</b>	<b>0.79</b>	+0.15	LR- <i>tf</i>	0.72	0.68	0.66	0.67	=	MNB- <i>mtf</i>	0.54	0.53	0.53	0.53	-0.03
90	LSVC- <i>tf</i>	0.57	0.60	0.59	0.57	-0.07	Perc- <i>bool</i>	0.69	0.63	0.59	0.59	-0.08	CNB- <i>bool</i>	0.53	0.52	0.52	0.52	-0.04
0	DT- <i>tf</i>	0.63	0.63	0.63	0.63	-0.01	MLP- <i>tf</i>	0.74	0.70	0.66	0.67	*	CNB- <i>bool</i>	0.55	0.54	0.54	0.54	-0.02
10	DT- <i>tf</i>	0.65	0.65	0.65	0.65	+0.01	MLP- <i>bool</i>	0.71	0.66	0.63	0.64	-0.03	CNB- <i>bool</i>	0.55	0.55	0.55	0.55	-0.01
20	DT- <i>tf</i>	0.65	0.66	0.66	0.65	+0.01	CNB- <i>tf</i>	0.66	0.59	0.56	0.56	-0.11	CNB- <i>bool</i>	0.54	0.54	0.54	0.54	-0.02
30	Perc- <i>bool</i>	0.62	0.62	0.59	0.57	-0.07	Perc- <i>tf</i>	0.62	0.61	0.63	0.60	-0.07	CNB- <i>bool</i>	0.54	0.54	0.54	0.54	-0.02
40	CNB- <i>bool</i>	0.61	0.60	0.60	0.60	-0.04	Perc- <i>tf</i>	0.65	0.63	0.65	0.63	-0.04	CNB- <i>bool</i>	0.55	0.54	0.54	0.54	-0.02
50	CNB- <i>bool</i>	0.61	0.61	0.61	0.61	-0.03	LR- <i>tf</i>	0.69	0.64	0.62	0.63	-0.04	CNB- <i>mtf</i>	0.54	0.53	0.53	0.53	-0.03
60	DT- <i>tf</i>	0.55	0.55	0.55	0.55	-0.09	LSVC- <i>bool</i>	0.71	0.66	0.62	0.63	-0.04	LR- <i>tf</i>	0.54	0.54	0.53	0.53	-0.03
70	Perc- <i>mtf</i>	0.62	0.65	0.58	0.55	-0.09	DT- <i>mtf</i>	0.65	0.59	0.59	0.59	-0.08	CNB- <i>mtf</i>	0.53	0.53	0.53	0.53	-0.03
80	Perc- <i>bool</i>	0.70	<b>0.83</b>	0.67	0.64	=	LSVC- <i>tf</i>	<b>0.77</b>	<b>0.76</b>	<b>0.68</b>	<b>0.70</b>	+0.03	CNB- <i>bool</i>	0.54	0.54	0.54	0.54	-0.02
90	Perc- <i>bool</i>	0.57	0.56	0.56	0.55	-0.08	CNB- <i>bool</i>	0.62	0.52	0.52	0.51	-0.16	Perc- <i>tf</i>	0.51	0.51	0.51	0.51	-0.05

## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

**Tabla 7.5:** Resultados de clasificación del rasgo de *Responsabilidad* obtenidos en los conjuntos de datos RxPI-estratificado, Auto-descripciones y Essays, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonálidades de verde indican un incremento de F en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 7.1.

n	RxPI					Auto-descripciones					Essays							
	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
0	DT-fidf	0.88	0.69	<b>0.67</b>	<b>0.68</b>	*	LR-ff	0.63	0.57	0.56	0.56	-0.05	CNB-ff	0.56	0.56	0.56	0.56	-0.01
10	DT-mtf	0.87	0.65	0.62	0.63	-0.05	DT-ff	0.67	0.64	0.65	0.64	+0.03	CNB-bool	0.57	0.57	0.57	0.57	=
20	DT-fidf	0.84	0.62	0.65	0.63	-0.05	Perc-fidf	0.70	0.66	0.63	0.63	+0.03	MNB-ff	0.57	0.57	0.57	0.57	=
30	MLP-mtf	0.89	0.70	0.54	0.55	-0.13	Perc-ff	0.58	0.55	0.55	0.55	-0.06	MNB-bool	0.57	0.57	0.57	0.57	=
40	DT-fidf	0.85	0.61	0.61	0.61	-0.07	DT-ff	0.60	0.56	0.56	0.56	-0.05	MNB-bool	0.57	0.57	0.57	0.57	=
50	DT-mtf	0.83	0.58	0.60	0.59	-0.09	CNB-ff	0.65	0.57	0.54	0.52	-0.09	MNB-mtf	0.57	0.57	0.57	0.57	=
60	DT-fidf	0.85	0.61	0.61	0.61	-0.07	DT-fidf	0.63	0.60	0.60	0.60	-0.01	MNB-bool	0.57	0.57	0.57	0.57	=
70	LSVC-ff	0.86	0.60	0.57	0.58	-0.10	DT-mtf	0.62	0.55	0.55	0.55	-0.06	CNB-idtf	0.56	0.56	0.56	0.56	-0.01
80	Perc-fidf	0.85	0.61	0.61	0.61	-0.07	DT-mtf	0.58	0.56	0.56	0.56	-0.05	CNB-mtf	0.55	0.55	0.55	0.55	-0.02
90	Perc-mtf	0.86	0.63	0.61	0.62	-0.06	DT-ff	<b>0.75</b>	<b>0.73</b>	<b>0.68</b>	<b>0.69</b>	+0.08	LR-mtf	0.55	0.55	0.55	0.55	-0.02
0	DT-mtf	0.86	0.63	0.61	0.62	-0.06	Perc-ff	0.65	0.61	0.61	0.61	*	CNB-mtf	0.56	0.56	0.56	0.56	-0.01
10	Perc-mtf	0.85	0.58	0.56	0.57	-0.11	MLP-ff	0.67	0.61	0.59	0.59	-0.02	MNB-mtf	0.56	0.56	0.56	0.56	-0.01
20	DT-mtf	0.86	0.63	0.61	0.62	-0.06	MLP-mtf	0.63	0.58	0.58	0.58	-0.03	CNB-mtf	0.56	0.56	0.56	0.56	-0.01
30	DT-mtf	0.86	0.63	0.61	0.62	-0.06	LSVC-ff	0.58	0.54	0.54	0.54	-0.07	CNB-mtf	0.56	0.56	0.56	0.56	-0.01
40	Perc-mtf	0.86	0.60	0.57	0.58	-0.10	LSVC-ff	0.67	0.61	0.59	0.59	-0.02	CNB-mtf	0.56	0.56	0.56	0.56	-0.01
50	MLP-ff	<b>0.90</b>	<b>0.79</b>	0.59	0.63	-0.05	DT-mtf	0.58	0.54	0.54	0.54	-0.07	LSVC-ff	0.55	0.56	0.55	0.55	-0.02
60	MLP-bool	0.89	0.70	0.54	0.55	-0.13	DT-mtf	0.52	0.49	0.49	0.48	-0.13	RF-mtf	0.55	0.55	0.55	0.55	-0.02
70	LSVC-ff	0.87	0.58	0.53	0.54	-0.14	DT-bool	0.65	0.62	0.63	0.62	+0.01	LSVC-ff	0.55	0.55	0.55	0.55	-0.02
80	Perc-mtf	0.83	0.61	0.64	0.62	-0.06	MLP-mtf	0.72	0.72	0.60	0.59	-0.02	LSVC-mtf	0.54	0.54	0.54	0.54	-0.03
90	Perc-bool	0.77	0.56	0.60	0.56	-0.12	DT-ff	<b>0.75</b>	<b>0.73</b>	<b>0.68</b>	<b>0.69</b>	+0.08	LR-ff	0.55	0.54	0.54	0.54	-0.03

**Tabla 7.6:** Resultados de clasificación del rasgo de *Estabilidad Emocional* obtenidos en los conjuntos de datos R<sub>XPI</sub>-estratificado, Auto-descripciones y Essays, comparando el método de Auto-selección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonalidades de verde indican un incremento de F en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 7.1.

n	R <sub>XPI</sub>					Auto-descripciones					Essays							
	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
0	MLP-ff	0.58	0.59	0.59	0.58	-0.04	DT-mtf	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>	*	CNB-bool	0.58	0.58	0.58	0.58	-0.01
10	MLP-ff	0.55	0.55	0.55	0.55	-0.07	DT-mtf	0.62	0.61	0.61	0.61	-0.12	MNB-mtf	0.58	0.58	0.58	0.58	-0.01
20	MLP-ff	0.64	0.64	0.64	0.64	<b>+0.02</b>	DT-mtf	0.67	0.66	0.66	0.66	-0.07	CNB-ff	0.56	0.56	0.56	0.56	-0.03
30	DT-bool	0.62	0.62	0.63	0.62	=	DT-ffidf	0.63	0.64	0.63	0.63	-0.10	MNB-ffidf	0.56	0.56	0.56	0.56	-0.03
40	DT-bool	0.58	0.59	0.59	0.58	-0.04	DT-ffidf	0.67	0.67	0.67	0.67	-0.06	CNB-mtf	0.55	0.55	0.55	0.55	-0.04
50	DT-ff	0.62	0.63	0.63	0.62	=	DT-ffidf	0.59	0.59	0.59	0.59	-0.14	MNB-ff	0.56	0.56	0.56	0.56	-0.03
60	RF-mtf	0.66	0.65	0.65	0.65	<b>+0.03</b>	DT-mtf	0.60	0.60	0.60	0.60	-0.13	MNB-ff	0.56	0.56	0.56	0.56	-0.03
70	CNB-ff	0.66	0.69	0.68	0.66	<b>+0.04</b>	DT-mtf	0.56	0.56	0.56	0.56	-0.17	MNB-ff	0.55	0.55	0.55	0.55	-0.04
80	Perc-bool	0.68	0.69	0.69	0.68	<b>+0.06</b>	DT-mtf	0.55	0.55	0.54	0.54	-0.19	DT-mtf	0.53	0.53	0.53	0.53	-0.06
90	CNB-bool	0.49	0.73	0.55	0.41	-0.21	MNB-bool	0.51	0.51	0.51	0.50	-0.23	LR-ffidf	0.54	0.54	0.54	0.54	-0.05
0	CNB-bool	0.62	0.62	0.62	0.62	*	DT-mtf	0.65	0.65	0.65	0.65	-0.08	LR-bool	0.56	0.56	0.56	0.56	-0.03
10	Perc-bool	0.64	0.65	0.65	0.64	<b>+0.02</b>	DT-mtf	0.60	0.60	0.60	0.59	-0.14	MNB-bool	0.56	0.56	0.56	0.56	-0.03
20	Perc-bool	0.62	0.65	0.64	0.62	=	DT-mtf	0.59	0.59	0.59	0.59	-0.14	CNB-ff	0.54	0.55	0.54	0.54	-0.05
30	Perc-mtf	0.68	0.67	0.68	0.68	<b>+0.06</b>	DT-mtf	0.56	0.57	0.57	0.56	-0.17	CNB-bool	0.54	0.54	0.54	0.54	-0.05
40	Perc-ff	0.62	0.63	0.63	0.62	=	RF-ff	0.59	0.59	0.58	0.57	-0.16	CNB-mtf	0.55	0.55	0.55	0.55	-0.04
50	DT-ff	0.70	0.70	0.70	0.70	<b>+0.08</b>	RF-ff	0.56	0.56	0.55	0.53	-0.20	MNB-mtf	0.56	0.56	0.56	0.56	-0.03
60	MLP-ff	0.66	0.66	0.66	0.66	<b>+0.04</b>	DT-mtf	0.67	0.66	0.66	0.66	-0.07	RF-mtf	0.55	0.55	0.55	0.55	-0.04
70	CNB-ff	0.66	0.69	0.68	0.66	<b>+0.04</b>	DT-mtf	0.58	0.57	0.57	0.57	-0.16	LR-ff	0.55	0.55	0.55	0.55	-0.04
80	Perc-mtf	<b>0.79</b>	<b>0.81</b>	<b>0.81</b>	<b>0.79</b>	<b>+0.17</b>	DT-ff	0.58	0.58	0.58	0.58	-0.15	LR-mtf	0.54	0.54	0.54	0.54	-0.05
90	Perc-mtf	0.47	0.73	0.53	0.37	-0.25	MLP-bool	0.51	0.51	0.51	0.51	-0.22	CNB-mtf	0.55	0.55	0.55	0.55	-0.04

## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

**Tabla 7.7:** Resultados de clasificación del rasgo de *Apertura a Nuevas Experiencias* obtenidos en los conjuntos de datos RxPI-estratificado, Auto-descripciones y Essays, comparando el método de Autoselección de categorías con utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). Tonalidades de verde indican un incremento de F en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 7.1.

n	RxPI					Auto-descripciones					Essays						
	Classifier	Ex.	P	R	inc.	Classifier	Ex.	P	R	inc.	Classifier	Ex.	P	R	inc.		
0	RF-ff	0.67	0.66	0.62	0.62	DT-mtf	<b>0.82</b>	<b>0.76</b>	<b>0.72</b>	<b>0.74</b>	*	CNB-mtf	0.62	0.62	0.62	-0.01	
10	LSVC-bool	0.63	0.61	0.61	0.61	DT-mtf	0.79	0.71	0.70	0.71	-0.03	MNB-mtf	0.62	0.63	0.63	0.62	-0.01
20	LSVC-bool	<b>0.69</b>	<b>0.68</b>	0.67	<b>0.67</b>	DT-mtf	0.77	0.69	0.67	0.68	-0.06	CNB-mtf	0.62	0.62	0.62	0.62	-0.01
30	MLP-bool	<b>0.69</b>	0.67	0.67	<b>+0.05</b>	DT-ff	0.72	0.62	0.61	0.62	-0.12	CNB-mtf	0.62	0.62	0.62	0.62	-0.01
40	DT-mtf	0.66	0.65	0.66	+0.03	DT-mtf	0.77	0.69	0.67	0.68	-0.06	CNB-bool	0.62	0.62	0.62	0.62	-0.01
50	DT-mtf	0.66	0.65	0.66	+0.03	DT-ff	0.75	0.66	0.66	0.66	-0.08	MNB-bool	0.62	0.62	0.62	0.62	-0.01
60	DT-mtf	0.64	0.63	0.63	+0.01	DT-ff	0.79	0.71	<b>0.72</b>	0.72	-0.02	MNB-bool	0.61	0.61	0.61	0.61	-0.02
70	DT-mtf	0.66	0.65	0.64	+0.03	Perc-mtf	0.75	0.66	0.63	0.64	-0.10	CNB-bool	0.62	0.62	0.62	0.62	-0.01
80	Perc-mtf	0.55	0.54	0.55	-0.08	Perc-mtf	0.75	0.65	0.61	0.62	-0.12	MNB-bool	0.61	0.61	0.61	0.61	-0.02
90	Perc-bool	0.59	0.62	0.62	-0.03	Perc-bool	0.72	0.57	0.55	0.54	-0.20	MNB-mtf	0.60	0.60	0.60	0.60	-0.03
0	RF-ff	0.64	0.61	0.60	-0.02	DT-mtf	0.77	0.69	0.69	0.69	-0.05	CNB-bool	0.61	0.61	0.61	0.61	-0.02
10	RF-bool	0.63	0.61	0.59	-0.03	DT-mtf	0.77	0.69	0.67	0.68	-0.06	CNB-bool	0.61	0.61	0.61	0.61	-0.02
20	RF-mtf	0.66	0.64	0.62	0.62	DT-ff	0.79	0.71	0.70	0.71	-0.03	CNB-bool	0.61	0.61	0.61	0.61	-0.02
30	RF-bool	0.65	0.62	0.61	0.61	DT-ff	0.72	0.62	0.61	0.62	-0.12	RF-ff	0.61	0.61	0.61	0.61	-0.02
40	DT-mtf	0.65	0.65	0.65	0.64	DT-ff	0.72	0.62	0.61	0.62	-0.12	RF-mtf	0.61	0.60	0.60	0.60	-0.03
50	DT-mtf	0.65	0.65	0.65	0.64	DT-ff	0.72	0.63	0.64	0.63	-0.11	CNB-bool	0.60	0.60	0.60	0.60	-0.03
60	DT-mtf	0.66	0.65	0.66	+0.03	DT-ff	0.75	0.67	0.68	0.68	-0.06	CNB-bool	0.60	0.60	0.60	0.60	-0.03
70	DT-mtf	0.64	0.62	0.62	0.62	DT-ff	0.77	0.69	0.67	0.68	-0.06	LSVC-bool	0.60	0.60	0.60	0.60	-0.03
80	Perc-ff	<b>0.69</b>	0.67	<b>0.68</b>	<b>+0.05</b>	LSVC-bool	0.75	0.65	0.59	0.60	-0.14	RF-mtf	0.60	0.60	0.60	0.60	-0.03
90	Perc-ff	0.46	0.43	0.43	-0.19	Perc-bool	0.80	0.80	0.80	0.80	-0.10	LR-mtf	0.60	0.59	0.59	0.59	-0.04

### ¿Cuántas categorías deben seleccionarse?

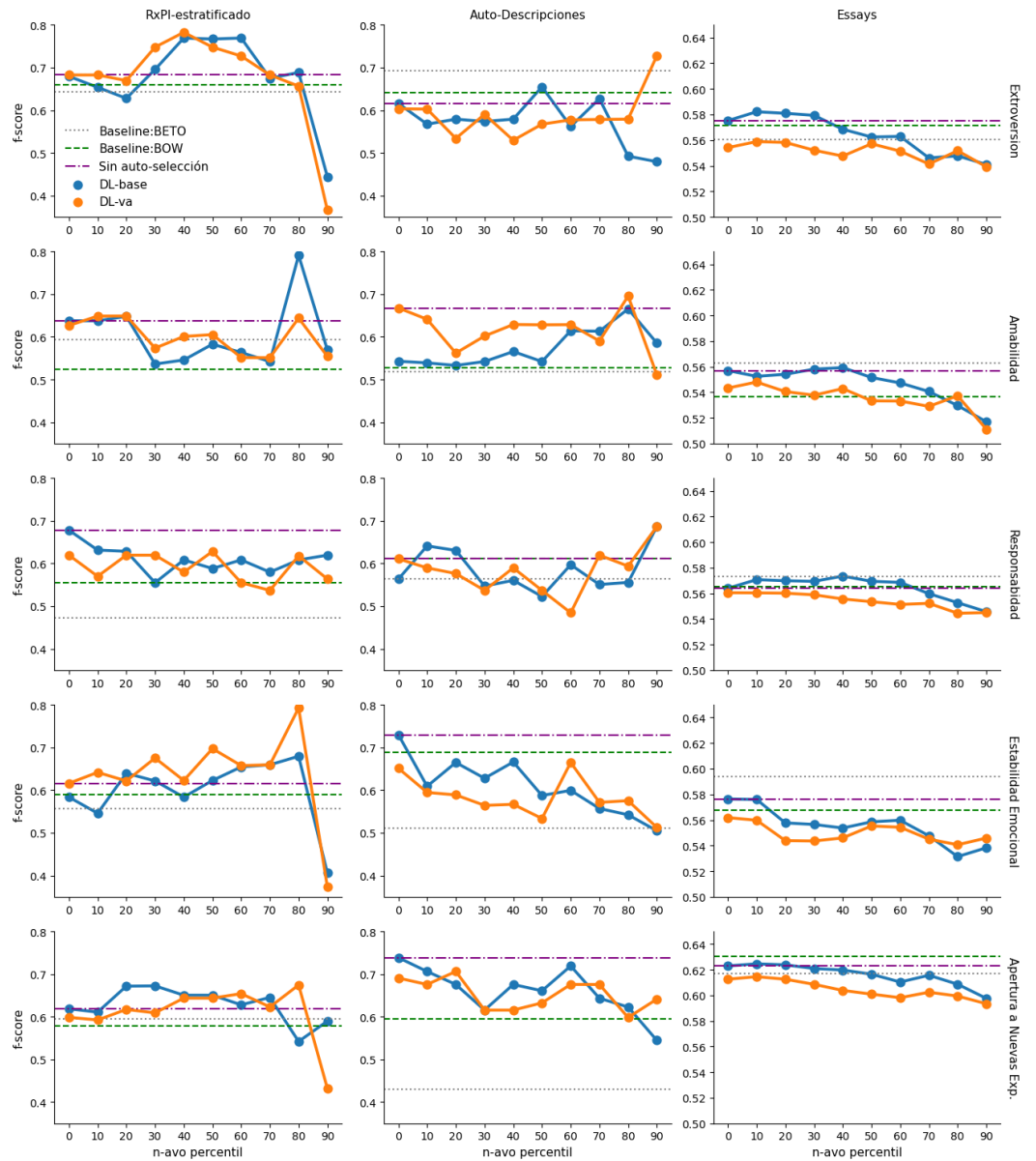
Un parámetro del método de Autoselección de categorías es la cantidad de categorías elegidas para la representación. Estas categorías se eligen en función del error absoluto medio (MAE), como se describe en el apartado 6.3.1, donde se calcula un valor de MAE para cada categoría en el recurso léxico. Este valor de MAE se ordena de menor a mayor, y luego se seleccionan las categorías con un valor de MAE arriba del  $n$ -avo percentil. Así, el percentil 0 en las Tablas de 7.3 a la 7.7 indica que no se hace ningún filtrado y por tanto se eligen *todas* las categorías existentes en el conjunto de entrenamiento. Conforme el valor de  $n$  se hace más grande, menos categorías del total de categorías presentes en el conjunto de entrenamiento son elegidas como importantes.

La Figura 7.1 muestra gráficas del desempeño de clasificación que permiten visualizar la mejora (o no) del uso de la disponibilidad léxica en la representación textual de las instancias, así como el impacto del conjunto de categorías semánticas seleccionadas de forma automática con el método propuesto. Cada gráfica condensa la información mostrada en las tablas de resultados anteriores. Las líneas horizontales corresponden a tres diferentes métodos de referencia, a decir: BoW, una representación basada en BETO y la mejor configuración de utilizar la disponibilidad léxica en la representación. Esta configuración depende del rasgo de personalidad y son los mismos valores marcados en la Tabla 7.1 que corresponde a usar todas las categorías semánticas presentes en el conjunto de entrenamiento. Finalmente, las líneas azul (DL-base) y naranja (DL-va) muestran el  $f$ -score para diferentes valores del  $n$ -avo percentil. Entre más grande es  $n$ , el método de Autoselección de categorías selecciona menos categorías.

Para diferentes conjuntos de datos los clasificadores de cada rasgo de personalidad tienen comportamientos distintos; por lo tanto, cada problema de clasificación debería tratarse de forma independiente y no generalizar los resultados encontrados para un conjunto de datos específico. Esta falta de generalización puede verse como una limitante del método propuesto; sin embargo, también habla sobre la complejidad del fenómeno que se trata de identificar y de los límites de los conjuntos de datos utilizados. Por lo que la parametrización que mejor funciona para un rasgo en un conjunto de datos puede ser distinta para ese rasgo en otro corpus.

De forma general, estas gráficas de desempeño ilustran que el uso de la disponibilidad léxica (en al menos alguna de sus configuraciones) es mejor al de los métodos de referencia. Por ejemplo, es posible notar que el método representado con la línea morada (Sin auto-selección) es el método de referencia más fuerte en los 5 rasgos de RxPI-estratificado y en 4 de 5 rasgos de Auto-descripciones. Aunque el *baseline* que corresponde a no realizar Autoselección de categorías en Essays solo es mejor en el rasgo de *Extroversión*, es el segundo mejor *baseline* en los otros 4 rasgos de este corpus. Cabe destacar que las diferencias en el valor de  $f$ -score para Essays es más pequeño que de los conjuntos de datos anteriores (note la escala del eje  $y$  en la última columna de la Figura 7.1 contra el mismo eje en las dos primeras columnas).

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual



**Figura 7.1:** Desempeño de modelos de clasificación generados con diferentes subconjuntos de términos organizados en categorías semánticas del recurso LIWC. Se muestran tres baselines: la representación basada en BoW (línea verde con guiones), la representación basada en *transformers* (línea gris punteada), y el mejor desempeño del uso de la disponibilidad léxica sin el filtrado de categorías semánticas (línea morada con líneas y puntos). Las gráficas de líneas continuas comparan DL-base y DL-va, para 10 conjuntos de categorías semánticas a usar en la representación.

En el conjunto de datos RxPI-estratificado, la mejora de la Autoselección de categorías es notoria para los rasgos de *Extroversión*, *Estabilidad Emocional* y en cierta medida para *Apertura a Nuevas Experiencias* (Figura 7.1, primera columna, renglones 1, 4 y 5, respectivamente). Principalmente en *Extroversión* es posible observar la curva de crecimiento y luego decrecimiento del valor de *f-score* para valores de *n* entre 20 y 70. Es decir, existe un límite en el filtrado de categorías que generan el mejor desempeño.

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

Para el caso de *Estabilidad Emocional*, el comportamiento del desempeño incrementa hasta  $n = 80$ . Este incremento se observa con una pendiente menor para *Apertura a Nuevas Experiencias*. En todos los casos del conjunto de datos RxPI–estratificado, hacer un filtrado extremo de categorías no es indicado; ya que cuando se usa un valor de  $n = 90$ , por lo general, se conserva solo una categoría. Para los rasgos *Amabilidad* y *Responsabilidad*, es posible notar que la mayoría de valores de  $f$ -score para diferentes valores de  $n$  caen por debajo del modelo que usa todas las categorías durante la representación —cuando no se realiza el filtrado de categorías— y por arriba de otros baselines utilizados: BoW y BETO.

En el conjunto de datos Auto–descripciones, si se utiliza la información del lexicón mental sin filtrar categorías semánticas se obtienen valores de  $f$ -score relativamente altos —en comparación con los métodos de referencia y con el comportamiento general en el corpus de RxPI–estratificado—; particularmente, para los rasgos de *Estabilidad Emocional* y *Apertura a Nuevas Experiencias* (Figura 7.1, segunda columna). Por lo tanto eliminar categorías en estas condiciones tiende a decrementar el desempeño de un clasificador binario (como puede verse en las gráficas).

Para el conjunto de datos Essays, no existe una indicación clara de la mejora en el desempeño del método al utilizar la Autoselección de categorías. Las gráficas de la Figura 7.1 (tercera columna) muestran una mejora mínima en el rasgo de *Extroversión* para valores de  $n = 10$ ,  $n = 20$  y  $n = 30$ ; pero que en términos prácticos este desempeño es igual al mejor método de referencia —DL–base sin Autoselección de categorías—, como se puede ver en la Tabla 7.3.

Para los rasgos *Extroversión*, *Amabilidad* y *Responsabilidad* del corpus Essays, utilizar el filtrado de categorías semánticas genera un desempeño de clasificación similar al mejor método de referencia (que puede ser sin la Autoselección de categorías o la representación basada en *transformers*). Para los otros dos rasgos, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*, la Autoselección de categorías disminuye el desempeño del *baseline*, i. e. representaciones basadas en *Transformers* y BoW, respectivamente, en 0.01 de  $f$ -score.

De los resultados en el corpus Essays, se puede observar que a pesar de que no existe una mejora cuantitativa clara, el utilizar un número pequeño de términos en la representación así como la agrupación de estos en categorías semánticas tienen el potencial de proporcionar información relevante a las personas especialistas, del área de psicología, para la toma de decisiones. De modo que los productos generados por el método propuesto —términos y grupos semánticos— son más simples de interpretar que miles de términos descontextualizados o vectores numéricos proporcionados por BoW y BERT, respectivamente.



## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

### ¿Qué categorías semánticas de LIWC son importantes?

La Figura 7.2 muestra las categorías usadas en cada configuración experimental. Cada gráfica es un mapa de calor del porcentaje de uso de cada categoría en el total de experimentos realizados, dado los esquemas de evaluación: *leave one out* para RxPI y Auto-descripciones, y el esquema de evaluación *10 cross-fold-validation* para Essays. Cuando en el eje  $x$  el valor de  $n = 0$  se utilizan todas las categorías disponibles en el diccionario de categorías: LIWC. Conforme el valor del  $n$  aumenta, menos categorías son elegidas por el método de Autoselección de categorías. Al lado derecho de cada mapa de calor se muestra el número absoluto de términos utilizado en los experimentos por cada categoría de LIWC. Cada conjunto de datos tiene longitudes de listas de disponibilidad léxica diferentes, por lo que el rango del número de término por categoría también es diferente para cada corpus.

De las gráficas en la Figura 7.2, la categoría nombrada OOV —que agrupa los términos que no están en el recurso léxico pero que son utilizados en el conjunto de datos— se usa solo en los primeros valores de  $n$  (excepto para *Apertura a Nuevas Experiencias* en RxPI-estratificado y para *Extroversión* en Auto-descripciones), indicando que la diferencia entre el valor de disponibilidad léxica de términos en las clases *alta* y *baja* de esos rasgos es muy pequeña y por lo tanto el valor de MAE no es el mayor de entre todas las categorías. Esto puede deberse a que son los mismos términos fuera de vocabulario en ambas comunidades contrastadas, lo cual también habla de la cohesión de temas que se abordan en estos conjuntos de datos.

### RxPI-estratificado

Del conjunto de Experimentos 1, descritos en la Sección 7.2, el mejor desempeño en el rasgo de *Extroversión* para RxPI-estratificado ocurre para  $n = 40$  (como puede verse tanto en la Figura 7.1 como en la Tabla 7.3). Como se observa en el mapa de calor (de la Figura 7.2 en la gráfica del primer renglón de la primera columna) las categorías seleccionadas por el método de Autoselección de categorías para ese valor de  $n = 40$  son: Dimensiones lingüísticas, Categorías del habla y 4 de los 6 procesos dentro de la categorías de Procesos psicológicos, estas son: Procesos sociales, afectivos, cognitivos y perceptuales.

Para los rasgos de *Amabilidad* y *Estabilidad Emocional* el mejor desempeño se obtiene para  $n = 80$ , donde se eligen dos categorías: Procesos biológicos y Categorías del habla. Sin embargo, el número de términos seleccionados, en promedio, considerando estas dos categorías son solo 2 y 3 términos, respectivamente para cada rasgo de personalidad. Por lo que es adecuado considerar que la representación de textos utilizando únicamente 2 o 3 términos no es recomendable. En el contexto de la composición del conjunto de datos RxPI-estratificado, es importante reflexionar si considerar o no la Categoría del habla. Especialmente cuando en RxPI no existe una gran

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

variedad de términos que puedan clasificarse dentro de la Categoría del habla de LIWC, ya que el corpus se obtuvo de forma escrita y naturalmente existen pocas incidencias de términos de este tipo. Así, aunque el desempeño del modelo de clasificación entrenado es alto para  $n = 80$  en estos dos rasgos, se recomienda tener presente la naturaleza del conjunto de datos y el número de términos que se consideran en la representación.

Para *Apertura a Nuevas Experiencias* de RxPI–estratificado el mejor desempeño se obtiene para un valor de  $n = 20$  con el esquema DL–base y con un valor de  $n = 80$  con el esquema DL–va. La diferencia entre DL–base y DL–va es la incorporación del valor de disponibilidad léxica dentro del esquema de pesado utilizado. En este sentido, DL–base necesitó de más categorías de términos para diferenciar entre un polo y otro; mientras que la incorporación del valor de disponibilidad léxica permitió mejorar el desempeño de clasificación con una menor cantidad de categorías en la representación. Particularmente, para  $n = 20$  se utilizan las categorías: Dimensiones lingüísticas, 5 de 6 procesos psicológicos —Procesos afectivos, cognitivos, perceptuales, biológicos y relatividad— y términos fuera del vocabulario. Para  $n = 80$  se utilizan únicamente 2 categorías: Procesos biológicos, y afectivos; y en algunos experimentos también se autoselecciona la categoría Relatividad.

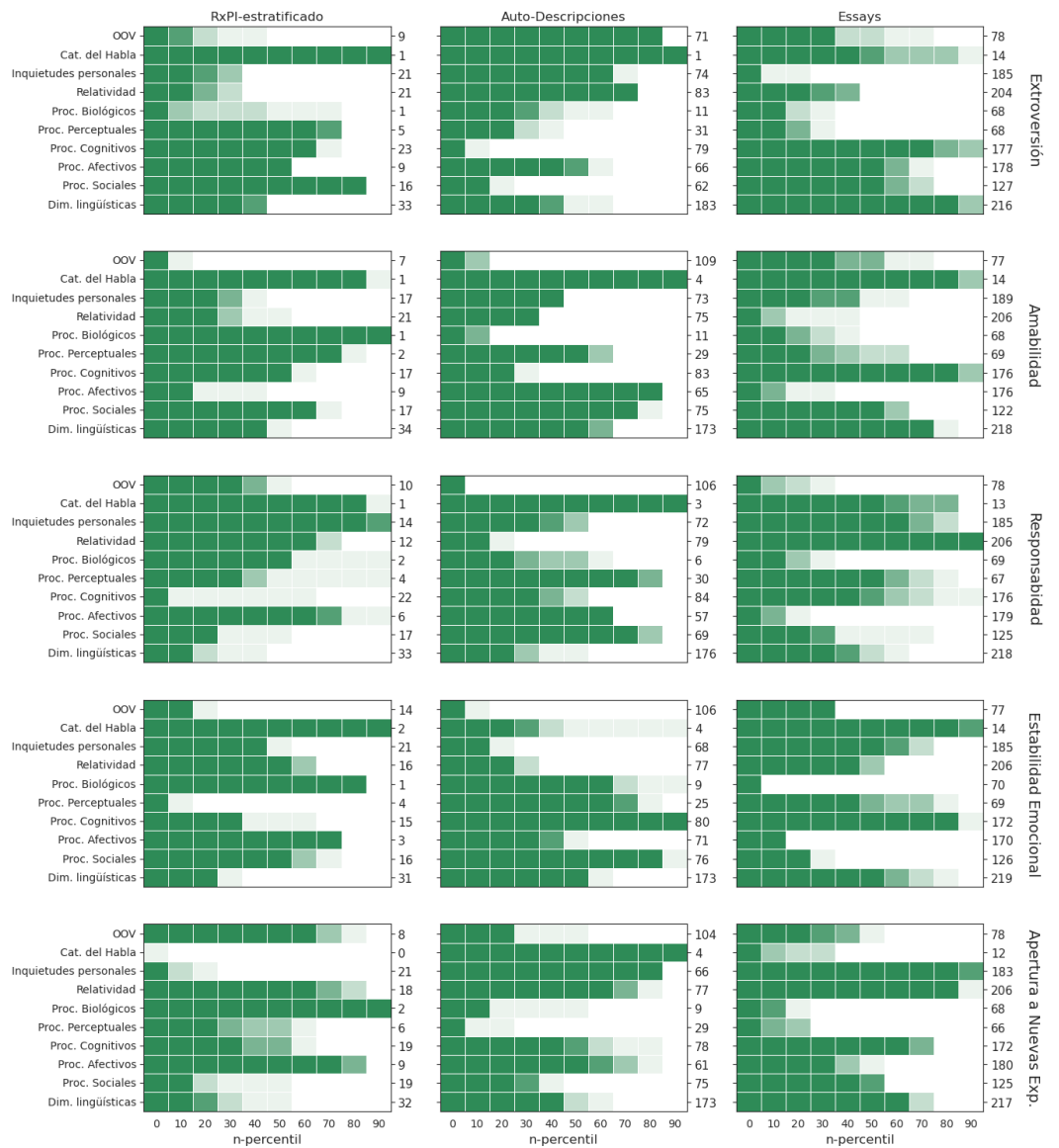
En contraste con los rasgos de *Amabilidad* y *Estabilidad Emocional*, el número de términos seleccionados por el método de Autoselección de categorías para *Apertura a Nuevas Experiencias* con el mismo valor de  $n = 80$  es mayor a 2 o 3 términos, lo que supone una representación más diversa en función del número de dimensiones en el vector de características, pero más importante aún —desde el punto de vista psicológico— que las categorías seleccionadas son relevantes para el problema de identificación de personalidad dentro del corpus.

De las gráficas mostradas en la Figura 7.1, la primera columna muestra una caída en el desempeño del clasificador cuando se utiliza el percentil  $n = 90$ ; es decir, cuando se selecciona (por lo general) solo una categoría en estos experimentos. Esta caída puede relacionarse con la categoría seleccionada pero también con el número de términos usados. Las gráficas muestran las caídas más notables para  $n = 90$  en los rasgos de *Extroversión*, *Estabilidad Emocional* y *Apertura a Nuevas Experiencias*, donde se selecciona la Categoría del habla y/o únicamente 1 o 2 términos son usados por la categoría seleccionada, como se muestra en los mapas de calor de la Figura 7.2.

### Auto–descripciones

Para el conjunto de datos Auto–descripciones, dado que no hay muchos ejemplos donde se pueda notar una mejora en el desempeño de los métodos de clasificación en relación a utilizar todas las categorías ( $n = 0$ ), se analizarán los mejores valores de  $n$  independientemente si mejoran o no el método de referencia. Así, dos casos que son relevantes para este corpus son

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual



**Figura 7.2:** Mapas de calor de las categorías de LIWC seleccionadas por el método de Auto-selección de categorías por cada conjunto de datos en los 5 rasgos de personalidad. La categoría OOV (*out of vocabulary*) agrupa todos los términos que no se encuentran en el recurso LIWC.

los resultados de *Extroversión* y *Responsabilidad* con el mejor resultado obtenido para  $n = 90$  (primer y tercer renglón en la segunda columna de la Figura 7.1) y de acuerdo a la información de los mapas de calor, la categoría elegida en ambos casos es solo 1: Categorías del habla. Similarmente que con los resultados en el corpus RxPI-estratificado, se recomienda tomar con precaución resultados que consideran 1 categoría con 1 término para usar en la representación. Entonces, para Auto-descripciones no existe una mejora en el desempeño de clasificación al filtrar categorías; sin embargo, las categorías seleccionadas, aun sin mejorar el desempeño de

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

los métodos de referencia, pueden ser relevante para un especialista que es capaz de dar una interpretación al conjunto de categorías elegidas.

Por ejemplo, para el rasgo de *Responsabilidad*, un valor de  $n = 10$  o incluso  $n = 20$  tiene una ligera mejora en el valor de *f-score* en los 3 métodos bases considerados (como se muestra en la gráfica del tercer renglón de la columna central en la Figura 7.1). Al elegir las categorías con un valor de percentil de  $n = 20$ , el método de Autoselección de categorías selecciona 8 categorías, las 5 que corresponden a Procesos psicológicos, Dimensiones lingüísticas, Inquietudes personales, y Categorías del habla; mientras que las categorías que no se utilizan son Relatividad y palabras fuera del vocabulario (OOV).

Con respecto al rasgo de *Amabilidad* del corpus de Auto-descripciones, en las gráficas de desempeño (Figura 7.1) se observa una ligera mejora en la medida *f-score* al utilizar un valor de  $n = 80$  que incluye 2 categorías: Procesos afectivos (aportando 65 términos) y Categorías del habla (que aporta solo 4 términos al vocabulario de la representación). Esto implica que para *Amabilidad*, la información de vocabulario que se clasifican en Emociones negativas y positivas, ansiedad, enojo y tristeza son importantes.

### Essays

Para el corpus Essays, es interesante analizar casos donde el desempeño es igual al mejor método de referencia, en este caso para los rasgos de *Extroversión*, *Amabilidad* y *Responsabilidad* (estos casos están marcados de color azul claro en las Tablas 7.3, 7.4 y 7.5, respectivamente).

De acuerdo con los resultados de clasificación, para el rasgo de *Extroversión* utilizar todas las categorías en LIWC obtiene desempeños similares al del mejor método de referencia; pero también es posible obtener resultados equivalentes eliminando la categoría Inquietudes personales (cuando  $n = 10$ ), la categoría de Procesos biológicos (para  $n = 20$ ), y la categoría de Procesos perceptuales (para  $n = 30$ ), como muestran los mapas de calor en la Figura 7.2. Es decir, mantener o no estas categorías en la representación no altera, en gran medida, el resultado de clasificación. Sin embargo, eliminar categorías diferentes a estas 3 sí afecta de forma negativa el desempeño.

Para *Amabilidad*, las categorías interesantes son Procesos biológicos y Procesos perceptuales. Cuando estas dos categorías no se seleccionan (para  $n = 30$  y  $n = 40$ ) el desempeño del método es similar al mejor método de referencia. Además de estas dos categorías, en  $n = 10$  y  $n = 20$ , donde el desempeño es ligeramente menor que el mejor *baseline*, son también relevantes las categorías de Dimensiones lingüísticas, Procesos sociales, Procesos cognitivos, Asuntos personales, Categorías del habla y palabras fuera del vocabulario (OOV). Por otro lado, al eliminar Asuntos personales (en  $n = 50$ ) y OOV ( $n = 60$ ) se aprecia un decremento del desempeño en la clasificación. Así, la curva de desempeño en *Amabilidad* para Essays va de  $n = 10$  a  $n = 50$

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

---

alcanzando su punto máximo en  $n = 40$ .

Finalmente, para el rasgo de *Responsabilidad*, se obtienen resultados similares al método de referencia con valores de  $n = 10$  a  $n = 60$ . Aquí el conjunto de términos seleccionados comprende un rango grande de categorías; sin embargo, al observar los mapas de calor se concluye que eliminar información de las categorías Procesos cognitivos y Procesos perceptuales deterioraron el desempeño del clasificador.

#### ¿Cuándo es relevante la Autoselección de categorías?

Como reflexión final, es esencial hacer una distinción acerca de cuándo considerar relevante la Autoselección de categorías. No es recomendable hacer una selección adicional de las mejores categorías cuando: el mejor resultado es obtenido al seleccionar categorías que se sabe *a priori* no tienen información en el conjunto de datos particular; por ejemplo, para RxPI las categorías del habla son irrelevantes; o cuando el conjunto de términos utilizados es tan pequeño que muy probablemente la representación esté sesgada por el conjunto de entrenamiento; por ejemplo, en Auto-descripciones las categorías del habla son relevantes pues la colección del corpus se hizo de forma hablada, pero la transcripción manual no representó fenómenos del habla de forma consistente.

#### 7.3.2. Autoselección de categorías autogeneradas con LDA

En la sección anterior utilizo el recurso léxico LIWC como Diccionario de Categorías para evaluar el desempeño del método de Autoselección de categorías propuesto en esta tesis. En esta sección se explora la utilización de un Diccionario de Tópicos construido de forma automática mediante el algoritmo LDA. A diferencia de LIWC que consiste de un recurso léxico predefinido (como se describe en la Sección 6.3.2), un recurso con categorías autogeneradas, al que llamo Diccionario de Tópicos (DT), se crea con el conjunto de instancias en el conjunto de entrenamiento. En esta sección se muestran los resultados de clasificación al utilizar el método de Autoselección de categorías con el Diccionario de Tópicos.

Como se explica en la Sección 6.3.2, mediante el algoritmo LDA genero un conjunto de tópicos utilizando el conjunto de entrenamiento disponible para cada problema de clasificación. Posteriormente, cada término se categoriza dentro de uno o más de los tópicos generados. En la construcción final del DT se consideran términos en un tópico si la probabilidad de pertenencia del término  $j$  en el tópico  $i$  es mayor o igual a la media de las probabilidades de todos los términos  $j$  en  $i$ . El objetivo es generar el Diccionario de Tópicos con los términos más probables de pertenecer a cada tópico. Es decir, se busca generar el recurso léxico con los términos más representativos de cada tópico.

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

**Tabla 7.8:** Resultados del rasgo de *Extroversión* obtenidos en los conjuntos de datos RxPI-estratificado y Auto-descripciones, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). El mejor resultado de la Tabla 7.1 está sombreado en gris.

	<i>n</i>	RxPI					Auto-descripciones						
		Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
DL-base	0	DT-tf	0.69	0.68	0.68	0.68	-0.01	DT-bool	0.69	0.61	0.63	0.62	-0.07
	10	DT-tf	0.70	0.70	0.69	0.70	+0.01	DT-bool	0.72	0.63	0.63	0.63	-0.06
	20	DT-tf	0.70	0.70	0.69	0.69	=	DT-bool	0.68	0.60	0.62	0.60	-0.09
	30	DT-tf	0.65	0.64	0.63	0.63	-0.06	DT-bool	0.72	0.65	0.67	0.65	-0.04
	40	DT-bool	0.65	0.64	0.65	0.64	-0.05	DT-bool	0.72	0.67	0.71	0.68	-0.01
	50	DT-bool	0.65	0.64	0.65	0.64	-0.05	DT-bool	0.69	0.61	0.63	0.62	-0.07
	60	DT-tf	0.66	0.65	0.65	0.65	-0.04	DT-bool	0.71	0.61	0.62	0.61	-0.08
	70	DT-bool	0.69	0.68	0.67	0.68	-0.01	DT-mtf	0.72	0.60	0.58	0.59	-0.10
	80	DT-bool	0.58	0.57	0.57	0.57	-0.12	Perc-bool	0.74	0.60	0.55	0.55	-0.14
	90	LR-tf	0.61	0.59	0.59	0.59	-0.10	DT-tf	0.69	0.61	0.63	0.62	-0.07
DL-va	0	DT-tf	0.70	0.70	0.69	0.69	*	DT-bool	0.68	0.60	0.62	0.60	-0.09
	10	DT-tf	0.66	0.65	0.64	0.65	-0.04	DT-bool	0.74	0.68	0.72	0.69	=
	20	DT-bool	0.66	0.65	0.65	0.65	-0.04	DT-bool	0.72	0.65	0.67	0.65	-0.04
	30	MLP-tf	0.65	0.64	0.63	0.63	-0.06	DT-tf	0.74	0.65	0.66	0.65	-0.04
	40	DT-tf	0.69	0.69	0.66	0.67	-0.02	DT-bool	0.68	0.60	0.62	0.60	-0.09
	50	DT-tf	0.69	0.68	0.67	0.67	-0.02	DT-bool	0.66	0.59	0.61	0.59	-0.10
	60	DT-bool	0.61	0.60	0.60	0.60	-0.09	DT-bool	0.69	0.60	0.61	0.60	-0.09
	70	DT-bool	0.62	0.61	0.60	0.60	-0.09	DT-bool	0.72	0.64	0.65	0.64	-0.05
	80	Perc-mtf	0.63	0.62	0.62	0.62	-0.07	DT-bool	0.68	0.54	0.53	0.53	-0.16
	90	LSVC-tf	0.63	0.62	0.62	0.62	-0.07	DT-tf	0.68	0.61	0.64	0.62	-0.07

Para la construcción del modelo de generación automática de tópicos sigo el mismo procedimiento descrito en la Sección 7.2.2, el algoritmo LDA con parámetros:  $i = 10$ ,  $\alpha = 0.91$  y  $\beta = 0.01$ . En este conjunto de experimentos la estrategia de evaluación también es *leave-one-out*, por lo que se generan tantos modelos de tópicos como instancias en cada conjunto de datos, y estos modelos son independientes entre un experimento y otro.

Las Tablas 7.8 a 7.12 muestran los resultados de la clasificación automática de los conjuntos de datos RxPI-estratificado y Auto-descripciones al utilizar el método de Autoselección de categorías que emplea el DT autogenerado. Los resultados de los métodos de referencia son los mismos que se muestran en la Tabla 7.1.

### RxPI-estratificado

Para el conjuntos de datos RxPI-estratificado en el rasgo de *Amabilidad* (Tabla 7.9), utilizar categorías autogeneradas, dentro del método propuesto en esta investigación, obtiene resultados inferiores que utilizar el recurso LIWC con categorías predefinidas. Con la utilización de LIWC se obtiene un valor de f-score igual a 0.79 (Tabla 7.4) mientras que con el Diccionario de Tópicos se alcanza un valor máximo de f-score de 0.68 (con la variante DL-base). Sin embargo, como se discutió antes, el mejor resultado obtenido utilizando LIWC se debe tratar con cautela ya que ese

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

**Tabla 7.9:** Resultados del rasgo de *Amabilidad* obtenidos en los conjuntos de datos RxPI-estratificado y Auto-descripciones, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). El mejor resultado de la Tabla 7.1 está sombreado en gris.

	<i>n</i>	RxPI					Auto-descripciones						
		Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
DL-base	0	DT-tf	0.64	0.64	0.64	0.64	*	DT-tfidf	0.60	0.54	0.54	0.54	-0.13
	10	DT-tf	0.67	0.67	0.67	0.67	+0.03	MLP-bool	0.62	0.52	0.52	0.51	-0.16
	20	DT-tf	0.66	0.66	0.66	0.66	+0.02	MLP-tf	0.68	0.61	0.59	0.59	-0.08
	30	DT-tf	0.68	0.68	0.68	0.68	+0.04	MLP-bool	0.68	0.60	0.56	0.56	-0.11
	40	DT-tf	0.66	0.66	0.66	0.66	+0.02	MLP-tf	0.55	0.48	0.48	0.48	-0.19
	50	DT-tf	0.67	0.67	0.67	0.67	+0.03	MLP-bool	0.63	0.54	0.53	0.52	-0.15
	60	DT-tf	0.63	0.63	0.63	0.63	-0.01	LSVC-bool	0.60	0.52	0.52	0.52	-0.15
	70	DT-tf	0.57	0.57	0.57	0.57	-0.07	MLP-mtf	0.63	0.55	0.54	0.54	-0.13
	80	DT-tf	0.59	0.58	0.58	0.58	-0.06	LR-mtf	0.62	0.55	0.54	0.54	-0.13
	90	DT-tfidf	0.52	0.52	0.52	0.52	-0.12	LR-tf	0.62	0.56	0.55	0.55	-0.12
DL-va	0	DT-tf	0.63	0.63	0.63	0.63	-0.01	MLP-tf	0.74	0.70	0.66	0.67	*
	10	DT-tf	0.65	0.65	0.65	0.65	+0.01	MLP-bool	0.71	0.66	0.61	0.61	-0.06
	20	DT-tf	0.66	0.66	0.66	0.66	+0.02	MLP-tf	0.75	0.76	0.64	0.65	-0.02
	30	DT-tf	0.66	0.66	0.66	0.66	+0.02	CNB-tf	0.69	0.63	0.60	0.60	-0.07
	40	DT-tf	0.60	0.60	0.60	0.60	-0.04	MNB-tf	0.69	0.63	0.60	0.60	-0.07
	50	DT-tf	0.67	0.67	0.67	0.67	0.03	MNB-tf	0.69	0.63	0.60	0.60	-0.07
	60	DT-tf	0.61	0.61	0.61	0.61	-0.03	CNB-tf	0.68	0.62	0.61	0.62	-0.05
	70	CNB-bool	0.57	0.57	0.57	0.56	-0.08	MNB-tf	0.71	0.66	0.63	0.64	-0.03
	80	DT-tf	0.55	0.55	0.55	0.55	-0.09	MNB-tf	0.71	0.66	0.63	0.64	-0.03
	90	Perc-bool	0.64	0.65	0.61	0.60	-0.04	MNB-tf	0.68	0.62	0.61	0.62	-0.05

resultado se obtuvo cuando la mayoría de las categorías semánticas son filtradas por el método de Autoselección de categorías (es decir cuando  $n = 80$  en la Tabla 7.4). En contraste, al utilizar el DT se alcanzan mejores resultados con un valor de  $n = 30$ .

Para *Estabilidad Emocional* (Tabla 7.11), también es posible mejorar el desempeño de la clasificación si se usa el método de Autoselección de categorías independientemente del diccionario utilizado. Al utilizar LIWC se alcanzan mejoras al método de referencia en 10 de 20 diferentes configuraciones, como se puede ver en la Tabla 7.6, de forma similar, al usar el DT se obtienen mejoras al método de referencia en 5 de 20 configuraciones. Sin embargo, los resultados obtenidos con LIWC son más altos que los obtenidos con el DT, de 0.02 a 0.17 incrementos absolutos vs. de 0.02 a 0.04 en la medida f-score, respectivamente.

Para el rasgo de *Responsabilidad* (Tabla 7.10) se observa un desempeño similar al utilizar ambos tipos de diccionarios; esto es, en ninguna configuración se logra superar al método base que consiste de no usar el método de Autoselección de categorías, donde el mejor desempeño obtiene un valor de f-score de 0.68.

En *Extroversión*, utilizar el DT (Tabla 7.8) no logra mejorar el desempeño del método de referencias excepto cuando  $n = 10$  con la ganancia mínima de 0.01 de f-score sobre el método base. En contraste con las mejoras en el desempeño que se obtienen al usar LIWC como Diccionario de

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

<i>n</i>	Classifier	RxPI					Auto-descripciones						
		Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	
DL-base	0	DT-tfidf	.88	0.69	0.67	0.68	*	LR-tf	0.63	0.57	0.56	0.56	-0.06
	10	DT-tfidf	0.85	0.63	0.65	0.64	-0.04	Perc-tfidf	0.72	0.68	0.64	0.64	+0.03
	20	DT-tfidf	0.87	0.66	0.66	0.66	-0.02	MLP-tfidf	0.72	0.72	0.60	0.59	-0.02
	30	DT-mtf	0.86	0.63	0.61	0.62	-0.06	Perc-tfidf	0.75	0.73	0.68	0.69	+0.08
	40	DT-mtf	0.87	0.65	0.62	0.63	-0.05	Perc-tf	0.63	0.60	0.60	0.60	-0.01
	50	DT-tfidf	0.85	0.58	0.56	0.57	-0.11	Perc-tf	0.67	0.64	0.65	0.64	+0.03
	60	DT-mtf	0.87	0.65	0.62	0.63	-0.05	MLP-tf	0.75	0.74	0.66	0.67	+0.06
	70	DT-tfidf	0.81	0.53	0.54	0.54	-0.14	Perc-tf	0.63	0.60	0.60	0.60	-0.01
	80	LSVC-bool	0.88	0.62	0.54	0.55	-0.13	MNB-tf	0.65	0.60	0.59	0.59	-0.02
	90	DT-mtf	0.84	0.52	0.51	0.51	-0.17	MLP-tf	0.67	0.63	0.64	0.63	+0.02
DL-va	0	DT-mtf	0.86	0.63	0.61	0.62	-0.06	MLP-mtf	0.67	0.62	0.61	0.61	*
	10	DT-mtf	0.87	0.65	0.62	0.63	-0.05	DT-tf	0.67	0.65	0.66	0.65	+0.04
	20	DT-mtf	0.86	0.63	0.61	0.62	-0.06	Perc-tf	0.65	0.61	0.61	0.61	=
	30	DT-mtf	0.85	0.61	0.61	0.61	-0.07	Perc-tf	0.65	0.61	0.61	0.61	=
	40	DT-mtf	0.90	0.76	0.64	0.67	-0.01	MLP-tf	0.67	0.61	0.59	0.59	-0.02
	50	DT-mtf	0.87	0.65	0.62	0.63	-0.05	MLP-mtf	0.67	0.61	0.60	0.60	-0.01
	60	DT-mtf	0.87	0.65	0.62	0.63	-0.05	DT-tf	0.67	0.63	0.64	0.63	+0.02
	70	MLP-bool	0.87	0.62	0.58	0.59	-0.09	Perc-tf	0.65	0.61	0.61	0.61	=
	80	DT-bool	0.83	0.55	0.55	0.55	-0.13	CNB-tf	0.63	0.59	0.59	0.59	-0.02
	90	Perc-mtf	0.81	0.56	0.58	0.57	-0.11	CNB-tf	0.63	0.59	0.59	0.59	-0.02

**Tabla 7.10:** Resultados del rasgo de *Responsabilidad* obtenidos en los conjuntos de datos RxPI-estratificado y Auto-descripciones, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro *f-score* (F). El mejor resultado de la Tabla 7.1 está sombreado en gris.

Categorías en el método de Autoselección de categorías, donde se obtienen mejoras en diferentes valores de *n* (como lo muestra la Tabla 7.3).

Finalmente, para el rasgo de *Apertura a Nuevas Experiencias* la diferencia entre el uso de los dos Diccionarios de Categorías es que con LIWC se obtienen mejoras en el desempeño del clasificador binario en diversas configuraciones del método propuesto, llegando a tener un incremento absoluto —en comparación con el método de referencia— de 0.05 de valor en la medida *f-score* para valores de *n* = 20 y *n* = 30 con DL-base (Tabla 7.7). Esto es, no se tienen que filtrar demasiadas categorías para alcanzar este pico. Por el contrario, utilizar el DT obtiene la mejor ganancia cuando *n* = 80 (Tabla 7.12), esto indica que se utilizan menos tópicos y por consiguiente menos términos en la representación.

### Auto-descripciones

El conjunto de datos Auto-descripciones, en 2 de los 5 rasgos: *Extroversión* y *Amabilidad*, el método que utiliza el diccionario predefinido LIWC logra un ligero aumento en el desempeño del método de clasificación; esto es, incrementos absolutos sobre el método base de 0.04 para *Extroversión* (Tabla 7.3) y 0.03 para *Amabilidad* (Tabla 7.4) en *f-score*. En contraste, al utilizar el Diccionario de Tópicos no se logra mejorar el método de referencia, como puede verse en las



## 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual

**Tabla 7.11:** Resultados del rasgo de *Estabilidad Emocional* obtenidos en los conjuntos de datos RxPI-estratificado y Auto-descripciones, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro f-score (F). El mejor resultado de la Tabla 7.1 está sombreado en gris.

	<i>n</i>	Classifier	RxPI					Auto-descripciones					
			Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.
DL-base	0	DT-tf	0.62	0.61	0.61	0.61	-0.01	DT-mtf	0.73	0.73	0.73	0.73	*
	10	DT-tf	0.62	0.62	0.60	0.59	-0.03	DT-mtf	0.67	0.67	0.66	0.66	-0.07
	20	DT-tf	0.66	0.66	0.64	0.64	+0.02	DT-mtf	0.64	0.64	0.64	0.64	-0.09
	30	DT-tf	0.62	0.61	0.61	0.61	-0.01	DT-tfidf	0.67	0.67	0.67	0.67	-0.06
	40	MLP-tfidf	0.64	0.65	0.61	0.59	-0.03	DT-mtf	0.69	0.69	0.69	0.69	-0.04
	50	DT-tf	0.60	0.59	0.59	0.59	-0.03	DT-mtf	0.67	0.66	0.66	0.66	-0.07
	60	MLP-tfidf	0.58	0.57	0.57	0.57	-0.05	DT-mtf	0.62	0.61	0.61	0.61	-0.12
	70	MLP-tf	0.64	0.64	0.64	0.64	+0.02	DT-mtf	0.68	0.68	0.68	0.68	-0.05
	80	DT-mtf	0.58	0.59	0.59	0.58	-0.04	DT-mtf	0.68	0.68	0.67	0.68	-0.05
	90	Perc-tfidf	0.55	0.54	0.54	0.54	-0.08	DT-mtf	0.58	0.58	0.58	0.58	-0.15
DL-va	0	DT-tf	0.64	0.63	0.62	0.62	*	DT-mtf	0.65	0.65	0.65	0.65	-0.08
	10	DT-tf	0.68	0.69	0.65	0.65	+0.03	DT-mtf	0.67	0.66	0.66	0.66	-0.07
	20	DT-tf	0.66	0.65	0.64	0.65	+0.03	DT-mtf	0.65	0.65	0.65	0.65	-0.08
	30	CNB-bool	0.62	0.62	0.62	0.62	=	DT-mtf	0.67	0.66	0.66	0.66	-0.07
	40	CNB-bool	0.60	0.60	0.60	0.60	-0.02	DT-mtf	0.71	0.70	0.70	0.70	-0.03
	50	MLP-tf	0.62	0.61	0.61	0.61	-0.01	DT-mtf	0.64	0.64	0.63	0.63	-0.10
	60	CNB-bool	0.60	0.60	0.60	0.60	-0.02	DT-mtf	0.67	0.66	0.66	0.66	-0.07
	70	Perc-mtf	0.62	0.65	0.64	0.62	=	DT-mtf	0.67	0.67	0.67	0.67	-0.06
	80	Perc-tf	0.66	0.66	0.66	0.66	+0.04	DT-mtf	0.67	0.66	0.66	0.66	-0.07
	90	MLP-bool	0.62	0.62	0.63	0.62	=	DT-mtf	0.58	0.58	0.58	0.58	-0.15

Tablas 7.8 y 7.9, respectivamente.

Para el rasgo de *Responsabilidad* (Tabla 7.10) se observa un mejor rendimiento para diferentes valores de *n* al usar el DT que al usar LIWC; aunque para ambos casos el mejor desempeño se obtiene con 0.69 de f-score. Cabe destacar que a pesar de lograr el mismo resultado, cuando se usa LIWC este mejor desempeño se alcanza con un valor de *n* = 90, lo que indica un alto filtrado de categorías; contrastado con el valor *n* = 30 cuando se usa el DT.

Finalmente, para los rasgos de *Estabilidad Emocional* (Tabla 7.11) y *Apertura a Nuevas Experiencias* (Tabla 7.12) ambas opciones de recursos léxicos generan valores de clasificación similares, es decir, el método de Autoselección de categorías no funciona para estos rasgos, sin importar el Diccionario de Categorías que se use.

### 7.3.3. Comparación del uso de los diccionarios generados: manualmente y automáticamente

La Figura 7.3 muestra gráficas donde se visualiza el incremento (o decremento) del desempeño de clasificación cuando se utiliza la Autoselección de categorías con un Diccionario de Tópicos en comparación con tres métodos de referencia: representaciones basadas en BoW, BETO y

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

<i>n</i>	RxPI						Auto-descripciones						
	Classifier	Ex.	P	R	F	inc.	Classifier	Ex.	P	R	F	inc.	
DL-base	0	RF-tf	0.67	0.66	0.62	0.62	*	DT-mtf	0.82	0.76	0.72	0.74	*
	10	Perc-mtf	0.63	0.61	0.60	0.60	-0.02	DT-mtf	0.70	0.60	0.60	0.60	-0.14
	20	RF-tf	0.66	0.64	0.60	0.60	-0.02	DT-mtf	0.77	0.69	0.69	0.69	-0.05
	30	MLP-bool	0.63	0.61	0.61	0.61	-0.01	DT-mtf	0.77	0.69	0.69	0.69	-0.05
	40	LSVC-bool	0.63	0.61	0.61	0.61	-0.01	DT-mtf	0.69	0.61	0.64	0.62	-0.12
	50	LSVC-bool	0.63	0.61	0.60	0.60	-0.02	DT-tf	0.79	0.71	0.68	0.69	-0.05
	60	RF-tfidf	0.65	0.62	0.60	0.60	-0.02	DT-tfidf	0.69	0.63	0.66	0.63	-0.11
	70	RF-bool	0.66	0.64	0.62	0.63	+0.01	DT-tfidf	0.70	0.59	0.58	0.58	-0.16
	80	RF-bool	0.69	0.68	0.65	0.66	+0.04	DT-tfidf	0.66	0.59	0.61	0.59	-0.15
	90	RF-mtf	0.66	0.63	0.61	0.61	-0.01	DT-tf	0.67	0.57	0.58	0.58	-0.16
DL-va	0	RF-tf	0.64	0.61	0.60	0.60	-0.02	DT-mtf	0.77	0.69	0.69	0.69	-0.05
	10	DT-bool	0.63	0.62	0.62	0.62	=	DT-mtf	0.72	0.63	0.64	0.63	-0.11
	20	RF-mtf	0.66	0.64	0.60	0.60	-0.02	DT-mtf	0.74	0.66	0.67	0.66	-0.08
	30	CNB-mtf	0.61	0.58	0.58	0.58	-0.04	DT-mtf	0.75	0.68	0.70	0.69	-0.05
	40	DT-bool	0.61	0.59	0.59	0.59	-0.03	DT-mtf	0.72	0.63	0.64	0.63	-0.11
	50	CNB-mtf	0.61	0.59	0.58	0.58	-0.04	DT-tf	0.77	0.68	0.65	0.66	-0.08
	60	CNB-mtf	0.60	0.58	0.57	0.58	-0.04	DT-tf	0.72	0.61	0.59	0.60	-0.14
	70	RF-bool	0.67	0.65	0.63	0.63	+0.01	DT-mtf	0.69	0.57	0.57	0.57	-0.17
	80	RF-bool	0.66	0.63	0.62	0.62	=	DT-bool	0.67	0.57	0.58	0.58	-0.16
	90	RF-mtf	0.64	0.61	0.59	0.59	-0.03	DT-tf	0.66	0.56	0.57	0.56	-0.18

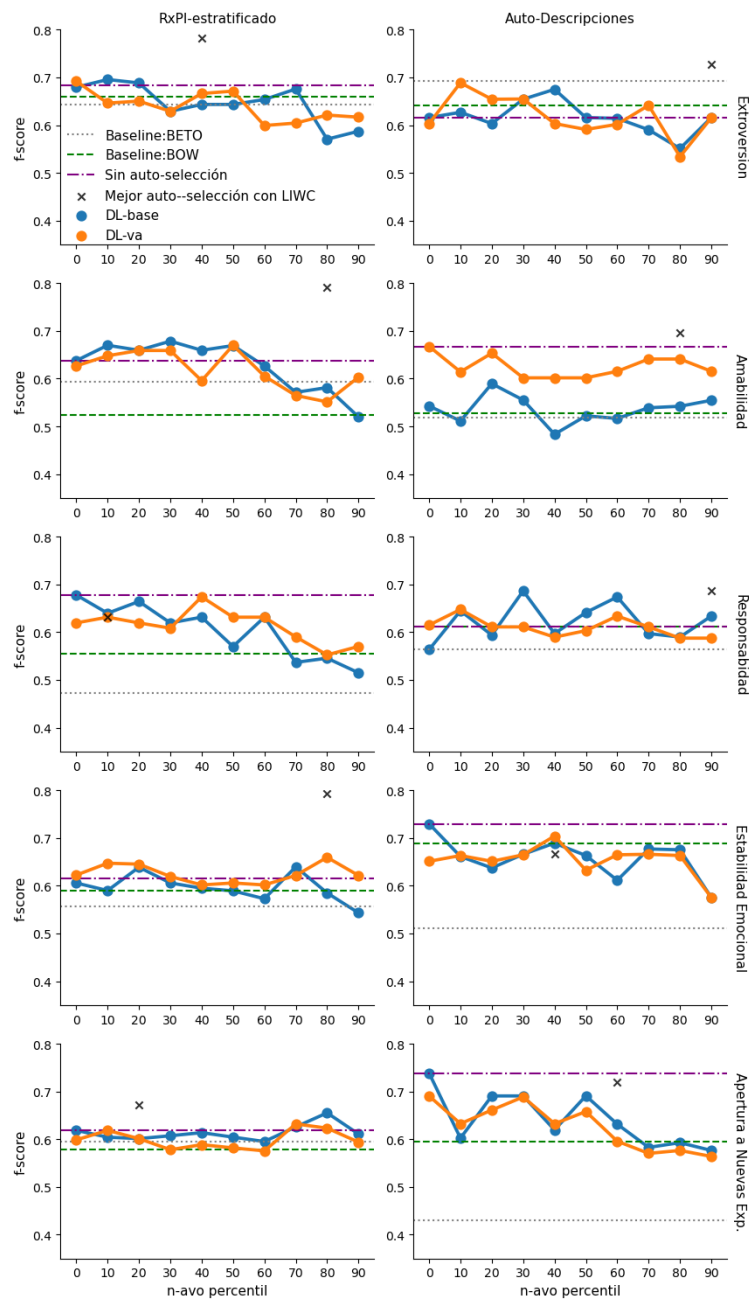
**Tabla 7.12:** Resultados del rasgo de *Apertura a Nuevas Experiencias* obtenidos en los conjuntos de datos RxPI-estratificado y Auto-descripciones, comparando el método de Autoselección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), macro precisión (P), macro recuerdo (R), y macro *f-score* (F). El mejor resultado de la Tabla 7.1 está sombreado en gris.

disponibilidad léxica sin Autoselección de categorías. Además, las gráficas incluyen la configuración del mejor desempeño obtenido con Autoselección de categorías con el Diccionario de Categorías LIWC (marcada con un cruz).

Dado que los métodos de referencias son los mismos que los utilizados en la Figura 7.1, es posible ver que, de forma general, el desempeño de clasificación obtenido con el uso del Diccionario de Tópicos es menor en comparación con los resultados obtenidos con el uso del diccionario predefinido LIWC. Esto puede indicar que las categorías de LIWC —construidas con conocimiento de expertos y por tanto predefinidas— son más útiles para la identificación de personalidad; señalando que la organización de términos dentro de categorías psicolingüísticamente relevantes es más importante que la organización de términos generados con el limitado vocabulario de las pocas instancias dentro de los corpora utilizados, de tal modo que reflejen aspectos de la personalidad de acuerdo al *Big-Five*.

Esta diferencia en la organización de los dos tipos de Diccionarios de Categorías se puede ver con más detalle en las Tablas 6.3 y 6.4, que muestran ejemplos de los diferentes tipos de organización bajo LIWC y bajo un Diccionario de Tópicos. La organización que se obtiene de un DT es generalmente orientado a temas, de ahí el nombre de tópicos, en lugar de grupos de términos que indican alguna característica sobre procesos psicológicos como emociones, formas de percepción, procesos cognitivos, por mencionar algunas, como se hace en LIWC.

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual



**Figura 7.3:** Desempeño de modelos de clasificación generados con diferentes subconjuntos de términos organizados en tópicos generados con LDA. Se muestran tres baselines: la representación basada en BoW (línea verde con guiones), la representación basada en la arquitectura BETO (línea gris punteada), y el mejor desempeño del uso de la disponibilidad léxica sin el filtrado de categorías semánticas (línea morada con líneas y puntos). Las gráficas de líneas continuas comparan DL-base y DL-va, para 10 conjuntos de categorías semánticas a usar en la representación.

Analizando de forma independiente los rasgos de personalidad, se observa que existen rasgos donde la utilización del DT tiene un comportamiento que merece discutirse. Por ejemplo, para el caso de *Amabilidad* en RxPI-estratificado, utilizando el DT con un valor de  $n = 30$  se logra un desempeño de clasificación mayor que el *segundo* mejor resultado obtenido con LIWC. Aunque

## Capítulo 7. Identificación automática de la personalidad: experimentación y análisis

---

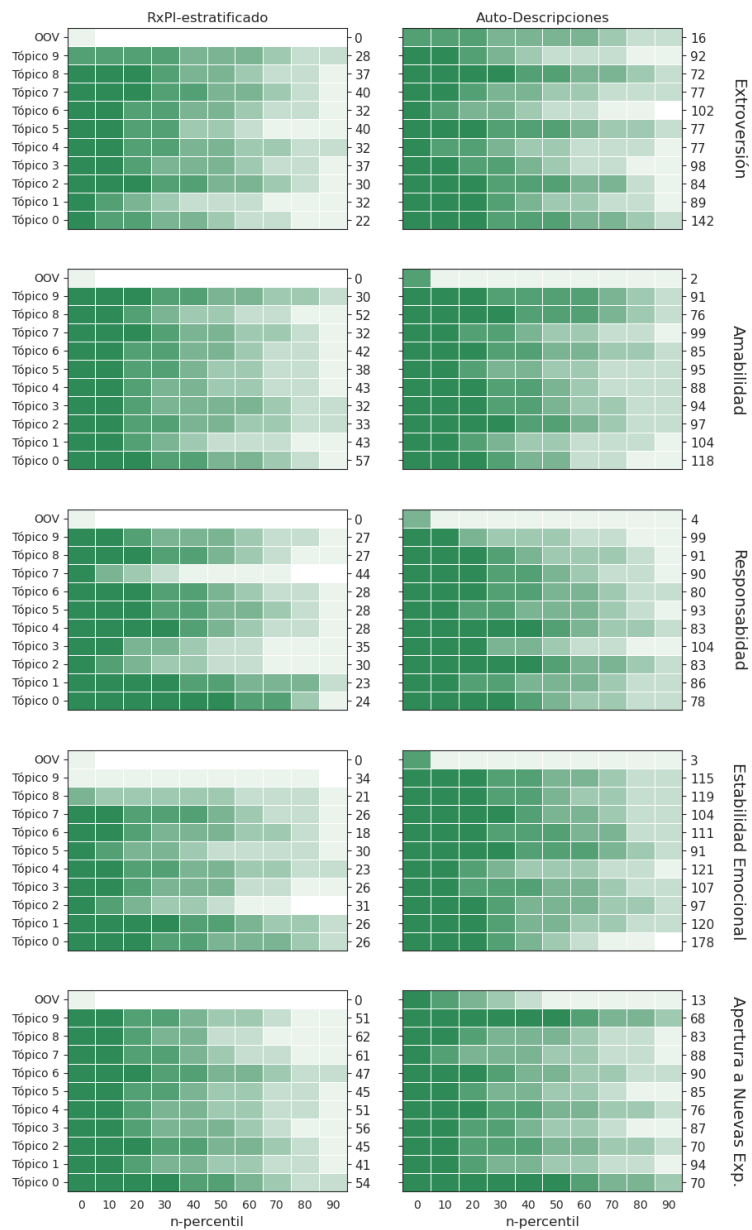
el mejor resultado obtenido con el Diccionario LIWC llega a casi 0.8 de *f*-score (indicado con la cruz en la gráfica de la Figura 7.3), como se discutió antes, este resultado conserva una cantidad muy limitada de términos en la representación lo que puede llevar a un sobreajuste. Por lo tanto, la comparación con el segundo mejor resultado obtenido con LIWC es más interesante. Con un DT, para *Amabilidad* en RxPI-estratificado se obtienen 5 configuraciones que sobrepasan el mejor baseline (para  $n = 10$  a  $n = 50$  con la variante DL-base y valores de *f*-score entre 0.66 a 0.68, en comparación con el valor del mejor método de referencia de 0.64); mientras que con el uso de LIWC, además del resultado ya discutido, únicamente 2 configuraciones tienen el mismo desempeño que el mejor baseline (con valores de *f*-score de 0.65 en ambos casos).

Un caso similar sucede con *Extroversión* en el conjunto de datos Auto-descripciones, mientras que el mejor resultado se obtiene con LIWC para  $n = 90$ , si se compara con el resto de los resultados obtenidos con LIWC, ninguna opción se acerca al mejor método de referencia —la representación basada en BETO—, como se puede ver en la segunda columna de la Figura 7.1. Sin embargo, al utilizar el DT hay dos puntos donde los resultados son comparables con el obtenido con la representación basada en *transformers* (primera gráfica la segunda columna en la Figura 7.3): para DL-base con  $n = 40$  (con una diferencia de 0.01 en el valor de *f*-score: 0.68 vs. 0.69); y con DL-va con  $n = 10$  (con un valor de *f*-score similar: 0.69).

Para el caso del rasgo de *Responsabilidad* en Auto-descripciones, utilizando el DT se logra superar el mejor método de referencia con 5 valores de  $n \in \{10, 30, 50, 60, 90\}$  con la variante DL-base (segundo renglón de la segunda columna en la Figura 7.3). Mientras que con el uso del recurso LIWC se supera en solo 3 valores de  $n \in \{10, 20, 90\}$  (como se puede ver en la gráfica correspondiente de la Figura 7.1).

De los resultados del uso del método de Autoselección de categorías con dos diccionarios distintos se puede ver un cambio más sutil en el desempeño cuando se usa el DT, no hay picos pronunciados a lo largo de las líneas mostradas en las gráficas de la Figura 7.3 y existen secciones planas en las líneas de desempeño; por ejemplo, la sección de  $n = 0$  a  $n = 60$  en los rasgos de *Estabilidad Emocional* y de *Apertura a Nuevas Experiencias* en RxPI-estratificado. Lo cual sugiere que los tópicos generados con LDA para estos casos no son lo suficientemente variados como para que la selección de diferentes subconjuntos de tópicos obtengan desempeños diferentes. Este fenómeno también puede verse en los mapas de calor de la Figura 7.4. A pesar de haber construido el Diccionario de Tópicos con términos con las mayores probabilidades de pertenecer a cada tópico, es posible que exista un conjunto pequeño de términos que están distribuidos en la mayoría de los tópicos y que por lo tanto, seleccionar uno u otro, no afectaría la representación final de las instancias de forma significativa.

### 7.3 Experimento 2: Autoselección de categorías para mejorar la representación textual



**Figura 7.4:** Mapas de calor de los tópicos seleccionados por el método de Autoselección de categorías por cada conjunto de datos en los 5 rasgos de personalidad. La categoría OOV (*out of vocabulary*) agrupa todos los términos que no se encuentran en el Diccionario de Tópicos.



# 8



## Identificación automática de Demencia y Depresión: experimentación y análisis

He presentado que existe evidencia de la utilidad del uso de un léxico seleccionado cuidadosamente con base en teorías lingüístico–cognitivas, como lo es el lexicón mental, para identificar los polos en los rasgos de personalidad de una persona basado en una muestra del lenguaje. El objetivo de este capítulo es determinar si el método propuesto para la identificación de rasgos de personalidad es útil para identificar otros aspectos psicológicos; específicamente la depresión y la demencia. Tanto la depresión como la demencia han sido estudiadas a través del lenguaje producido (Gremillion and Martel, 2014; Dick et al., 2018; Cardona et al., 2021).

Los siguientes experimentos pretenden establecer la utilidad del léxico disponible para la identificación de *Demencia* y *Depresión*. Los tres corpora utilizados han sido empleados en diversas investigaciones sobre la identificación automática de *Demencia* y *Depresión*; los corpora se describen en la Sección 8.1. Los experimentos de este capítulo son análogos a los realizados para personalidad, se trata de una tarea de clasificación binaria donde las clases corresponden a dos comunidades: la comunidad que presenta la característica de interés ya sea *Depresión* o *Demencia*, y la comunidad de control (aquella que no presenta esta característica).

### 8.1. Corpora para la identificación de *Demencia* y *Depresión*

Un corpus reúne textos de participantes con demencia, mientras que dos corpora son textos de participantes con depresión. Los tres conjuntos de datos consisten de transcripciones de muestras de habla espontánea en inglés. Cada corpus tiene particiones fijas de instancias para utilizarse durante el entrenamiento y durante la evaluación. Otra característica compartida de los corpora es el número limitado de instancias (entre 150 y 200 instancias) pues como es común en el estudio de fenómenos que tienen que ver con la salud mental, la compilación de datos es uno de los desafíos más difíciles que existen. La Tabla 8.1 muestra estadísticas de los tres corpora.

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

Corpus	Partición	Clases		Total
		Positiva	Control	
ADReSS	Entrenamiento	54	54	108
	Evaluación	24	24	48
DAIC-WOZ	Entrenamiento	30	77	107
	Evaluación	14	33	47
E-DAIC	Entrenamiento	37	126	163
	Evaluación	17	39	56

**Tabla 8.1:** Número de instancias para las clases positiva y control en tres corpora: ADReSS, DAIC-WOZ y E-DAIC

### 8.1.1. ADReSS: *Alzheimer's Dementia Recognition through Spontaneous Speech*

ADReSS es un conjunto de datos recolectado por Luz et al. (2020). Fue usado durante la conferencia INTERSPEECH en un taller para identificar *Demencia* de Alzheimer mediante el uso de habla espontánea. ADReSS está completamente balanceado en el número de instancias por clase en las dos particiones: entrenamiento y evaluación y en total consiste de 156 instancias. La Tabla 8.2 muestra 4 instancias de este corpus. En la revisión de investigaciones que han usando este corpus, Vrindha et al. (2023) encontraron que de los 5 trabajos que realizan análisis sobre el texto, la exactitud reportada va de 0.77 a 0.88 con representaciones basadas en LIWC hasta representaciones basadas en BERT.

### 8.1.2. *Distress Analysis Interview Corpus*

*Distress Analysis Interview Corpus* (DAIC) es un corpus diseñado para ayudar en el diagnóstico de depresión, ansiedad y trastorno de estrés postraumático. Este corpus fue recolectado por Gratch et al. (2014) y consiste de entrevistas semiestructuradas; actualmente tiene dos versiones. La primera versión, llamada DAIC-WOZ, se compone de entrevistas llevadas a cabo mediante un entrevistador virtual animado pero controlado por un humano en otra habitación (de ahí las siglas WOZ o *Wizard of Oz*) (DeVault et al., 2014). La segunda versión es E-DAIC (*Extended DAIC dataset*), una versión extendida de DAIC-WOZ, que fue utilizada en la competencia AVEC 2019 (Ringeval et al., 2019). Ambas versiones están desbalanceadas en el número de instancias por clase y en total cuentan con 154 y 219 instancias, respectivamente.

Mientras que DAIC-WOZ y E-DAIC se componen de una variedad de recursos como los videos de las entrevistas y los audios de las personas participantes, en los experimentos presentados en este capítulo utilizo las transcripciones automáticas y revisadas por un humano de las entrevistas semi-estructuradas. La Tabla 8.3 muestra ejemplos de las instancias en estos corpora.



## 8.2 Configuración experimental

Clase	ID	Muestra
Positiva	S083	<i>okay . we see a a little boy climbed up on a stool reaching for the cookie jar . and the stool is falling . he is trying to get a cookie for himself and also one for his sister . his sister is telling him to be very quiet . she's going &lt;UNK&gt; . can't write that down . telling him to be quiet . and let's see . in the meantime the mother is washing dishes and the sink has overflowed and is pouring . I don't think I see anything else . okay .</i>
	S101	<i>taking some cookies . and falling over . and mother's drying dishes . and things falling sink's running over . and two cups and saucer on the sink . and and the girl's windows on the curtains .and the spigot's on the sink . and well that's about it .</i>
Control	S036	<i>well for one thing this boy's on the stool getting cookies .and his stool's about to fall . and and the little girl's reaching for some cookies . she got her finger in her mouth . the mother's drying dishes and water's running over . yeah and she drying dishes water'srunning over . she's standing in water . there's a sink there . plate and a cup . outside you can see outside there the the walk+way and hedges ,a window, tree . and the cookie jar would hafta be inthe cupboard . and there's a the counter on in the kitchen. that's and the curtains . and the window .</i>
	S007	<i>boy taking cookies outof a cookie jar . the stool is falling . the little girl is reaching . water is running out of the faucet . the water is overflowing the sink . woman is drying dishes . there's nothing to indicate and I don't see any more action .</i>

**Tabla 8.2:** Muestras de lenguaje del corpus ADReSS, con dos instancias positivas y dos instancias de la clase de control.

## 8.2. Configuración experimental

### Preprocesamiento

Durante el preprocesamiento de los tres conjuntos de datos utilizados en los experimentos, la normalización de los textos se realizó mediante el software FreeLing (Carreras et al., 2004). Opté por solo considerar información que pueda ser encontrada dentro de alguna categoría semántica, por lo tanto los números, símbolos y palabras vacías (o de función) fueron eliminados. Además, mediante FreeLing se identifican multipalabras.

El cálculo de la disponibilidad léxica mantiene los mismos parámetros de  $w = 0.01$  y  $m = 1$  usados previamente. El valor de  $k = 100$  para ADReSS y de  $k = 500$  para las dos versiones de DAIC se estableció de forma experimental; los valores de  $k$  evaluados fueron 100, 500 y 1000.

### Estrategias de evaluación

Los tres conjuntos de datos utilizados están organizados en particiones de Entrenamiento y Evaluación; por lo tanto, utilicé una estrategia de evaluación simple: utilizar el conjunto de Entrenamiento de cada corpus para calcular la disponibilidad léxica y entrenar el modelo de clasificación, y luego evaluar el desempeño de la identificación automática utilizando la partición de Evaluación de cada corpus.

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

	DAIC-WoZ	E-DAIC
<b>Positiva</b> id: 319	<ul style="list-style-type: none"> <li>— sure</li> <li>— mm okay</li> <li>— los angeles</li> <li>— um</li> <li>— well &lt;laughter&gt;that's a good question</li> <li>— um i like the familiarity with everything i know where everything is in the city</li> <li>— mm</li> <li>— that's all i can say right now</li> <li>— so spread out so hard to get places</li> <li>— is there anything else</li> <li>— my dream job</li> <li>— i don't think there is a dream job</li> <li>— um</li> <li>— for me no there isn't one</li> <li>— mm</li> </ul>	<ul style="list-style-type: none"> <li>— and please</li> <li>— are you okay</li> <li>— yes I'm alright with this</li> <li>— doing pretty good little bit tired</li> <li>— I'm from Los Angeles</li> <li>— the weather and sports</li> <li>— you can't find jobs right now</li> <li>— I know</li> <li>— not really</li> <li>— I studied architecture and Chicano studies in a little bit of History</li> <li>— I would like to draw and then I'd like to learn about history</li> <li>— I'm trying to I'm going I'm working and going to school I'm trying to get my finish getting my degree in</li> <li>— playing in the NFL</li> <li>— what was that</li> <li>— I don't know what that is</li> </ul>
<b>Control</b> id: 318	<ul style="list-style-type: none"> <li>— yes</li> <li>— i'm alright</li> <li>— los angeles california</li> <li>— yep</li> <li>— um the lights big city it's always something going on</li> <li>— the traffic and that's it</li> <li>— uh business administration and business management</li> <li>— uh not that i'm doing something else i'm doing networking at the moment</li> <li>— &lt;sigh&gt;</li> <li>— hmm</li> <li>— uh to open up a big clothing line and just supply the whole world with clothing &lt;laughter&gt;</li> <li>— yeah</li> <li>— definitely</li> <li>— uh</li> <li>— 'cause i'm all about myself it's all about me &lt;laughter&gt;</li> </ul>	<ul style="list-style-type: none"> <li>— okay</li> <li>— okay alright</li> <li>— okay you</li> <li>— Inglewood California</li> <li>— yeah</li> <li>— no</li> <li>— that's okay</li> <li>— give me my girlfriend and my daughter</li> <li>— yes</li> <li>— are OKC okay cool</li> <li>— yes I do</li> <li>— well I'm football recruiter so I travel all around the world to to sign different football players for sports management company</li> <li>— I'm just the Independence about it being alone being able to be free and not tied down to one to one thing you know it's great</li> <li>— I'm I remember traveling to</li> <li>— and take them to clubs and this and that and we're not that type of organization</li> </ul>

**Tabla 8.3:** Muestras de lenguaje del corpus DAIC-WoZ y E-DAIC, con una instancia positiva y una instancia de la clase de control.

### 8.3 Experimento 1: Disponibilidad léxica para una representación textual

---

Para determinar el mejor desempeño de los métodos evaluados en este capítulo comparo el valor de  $f$ -score de la clase positiva. Pues, identificar correctamente la clase positiva en contextos de salud es más importante que identificar correctamente la clase de control; particularmente para casos donde es común un desbalance en el número de instancias por clase. Sin embargo, las tablas de resultados también muestran el valor de  $\text{macro } f$ -score y medidas como exactitud, precisión y recuerdo. Finalmente, los métodos de referencia utilizados son los mismos que los descritos en la Sección 7.2.

### 8.3. Experimento 1: Disponibilidad léxica para una representación textual

Igual que en el problema de identificación de personalidad, este primer experimento tiene dos objetivos:

- (a) Determinar la pertinencia del uso del léxico disponible —extraído a través de la prueba de disponibilidad léxica— para representar documentos de texto en la tarea de identificación de demencia y depresión. Esto es, la variante de representación básica o DL-base, descrita en la Sección 6.2.1. DL-base considera como vocabulario aquellas palabras que aparecen en la unión de las listas de disponibilidad léxica de las dos clases; es decir, aquellos términos disponibles calculados durante la etapa de entrenamiento del modelo.
- (b) Determinar el impacto del uso del valor de la disponibilidad de cada término como parte de la representación vectorial. Es decir, la variante de representación DL-va, descrita en la Sección 6.2.2, que no solo considera los términos disponibles, sino también considera el valor de accesibilidad/disponibilidad.

La Tabla 8.4 presenta los resultados de clasificación para los tres conjuntos de datos utilizados sobre la partición de Evaluación respectiva.

De los tres métodos de referencia utilizados, la representación basada en bolsa de palabras (BoW) es la que consistentemente tiene el mejor rendimiento en todos los experimentos; por lo tanto, considero BoW como referencia para indicar incremento o decremento absoluto de los desempeños de los otros métodos evaluados.

#### ADReSS

De la Tabla 8.4 para el conjunto de datos ADReSS el método basado en BoW obtiene 0.87 de  $f$ -score para la clase positiva, mientras que el método basado en la representación BERT

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

**Tabla 8.4:** Resultados de la clasificación en los conjuntos de datos ADReSS, DAIC-WOZ, and E-DAIC. Se compara la representación basada en el léxico disponible (DL-base y DL-va) contra una representación basada en bolsa de palabras (BoW) y modelos basados en *transformers* (BERT). Se reporta la exactitud (Ex.), precisión de la clase positiva (P+), recuerdo de la clase positiva (R+), *f-score* de la clase positiva (F+), y el macro *f-score* (F). Tonalidades de verde indican un incremento de F+ en comparación con el resultado de BoW.

	Método	Clasificador	Ex.	P+	R+	F+	(F)	inc.
ADReSS	Mayoritaria	–	0.50	0.0	0.0	0.0	(0.33)	–0.87
	BoW	LR-bool	0.85	0.79	<b>0.96</b>	0.87	(0.85)	
	Transformer	LSVC-bert	0.81	0.86	0.75	0.80	(0.81)	–0.07
	DL-base	RF-bool	<b>0.89</b>	<b>0.88</b>	0.92	<b>0.90</b>	<b>(0.90)</b>	+0.03
	DL-va	RF-tf	0.88	0.85	0.92	0.88	(0.87)	+0.01
DAIC-WOZ	Mayoritaria	–	<b>0.70</b>	0.00	0.00	0.00	(0.41)	–0.48
	BoW	DT-mtf	0.68	0.47	0.50	0.48	(0.63)	
	Transformer	DT-bert	0.30	0.30	<b>1.00</b>	0.46	(0.23)	–0.02
	DL-base	DT-mtf	<b>0.70</b>	<b>0.50</b>	0.57	<b>0.53</b>	<b>(0.66)</b>	+0.05
	DL-va	DT-mtf	<b>0.70</b>	<b>0.50</b>	0.50	0.50	(0.64)	+0.02
E-DAIC	Mayoritaria	–	0.70	0.00	0.00	0.00	(0.41)	–0.55
	BoW	LSVC-tf	0.77	0.67	0.47	0.55	(0.70)	
	Transformer	DT-bert	0.30	0.30	<b>1.00</b>	0.46	(0.23)	–0.09
	DL-base	LSVC-tf	<b>0.80</b>	<b>0.75</b>	0.53	<b>0.62</b>	<b>(0.74)</b>	+0.07
	DL-va	LSVC-tf	0.75	0.67	0.35	0.46	(0.65)	–0.09

solamente alcanza un valor de 0.80 de *f-score* para la misma clase. Ambas opciones del método propuesto DL-base y DL-va obtienen una mejora del valor de *f-score* sobre BoW de 0.03 (obteniendo 0.90 de *f-score* para la clase positiva) y 0.01 (llegando a un valor de 0.88 de *f-score* para la clase de interés), respectivamente. El método DL-base es el que obtiene la mejor precisión para la clase positiva que el resto de los métodos comparados. Como puede verse en la primera sección de la Tabla 8.4 todos los métodos alcanzan valores de desempeño por arriba del 0.8 en exactitud, precisión, recuerdo y *f-score*, salvo para el método de clase mayoritaria.

### DAIC-WOZ y E-DAIC

Para los conjuntos de datos de *Depresión*, DAIC-WOZ y E-DAIC, se obtiene una mejora en el desempeño de clasificación si se utiliza una representación basada en la disponibilidad léxica mediante el método propuesto, ya sea con DL-base o con DL-va (ver la Tabla 8.4). Particularmente, para el corpus DAIC-WOZ la mejora con relación al método basado en BoW es de 0.05 (de 0.48 a 0.53 en el valor de *f-score* para la clase positiva) y 0.02 para DL-base y DL-va, respectivamente. En el corpus E-DAIC, el mejor desempeño lo logra el método DL-base con un incremento de 0.07 del valor de *f-score* con relación al método basado en BoW (pasando de 0.55 a 0.62 en la clase positiva).

Ambos corpora tienen un desbalance en el número de instancias por clase. La clase positiva tiene una menor cantidad de instancias que la clase de control (14 vs. 33 para DAIC-WOZ y 17 vs. 39 para E-DAIC, como se muestra en la Tabla 8.1). Cuando existe este desbalance, usualmente se obtienen buenos resultados con la clase con más instancias —la clase de control, en este caso—;

## 8.4 Experimento 2: Autoselección de categorías para mejorar la representación

---

esto se refleja en el valor del macro *f-score* (columna *F* en la Tabla 8.4). Sin embargo, el método propuesto tiene el mejor desempeño considerando el macro *f-score* en ambos conjuntos de datos, lo que indica que no solamente clasifica mejor la clase *positiva* sino también la clase de *control* en comparación con los métodos de referencia.

Otro aspecto de interés es que a pesar de que el método basado en *transformers* recupera todas las instancias positivas (como lo indica el valor de recuerdo (*R+*) de 1.00), el macro *f-score* es de apenas 0.23; esto usualmente sucede cuando el clasificador tiene una *precisión* muy baja. De manera general, el método basado en BERT es mejor clasificando la clase positiva que la clase de control, pero no lo suficiente para alcanzar los resultados obtenidos con el método propuesto basado en la disponibilidad léxica, específicamente con la variante DL-base.

### 8.4. Experimento 2: Autoselección de categorías para mejorar la representación

El segundo conjunto de experimentos tiene dos objetivos:

- (a) Determinar si al seleccionar de forma automática términos, en categorías específicas dentro de un diccionario de categorías, para la representación de muestras del lenguaje es posible mejorar el desempeño de métodos automáticos para identificar Demencia y Depresión.
- (b) Comparar el desempeño de la autoselección con dos tipos de diccionarios de categorías: predefinido y autogenerado.

A continuación se describen y discuten los resultados obtenidos por cada conjunto de datos.

#### ADReSS

Como muestra la Tabla 8.5, para *Demencia*—del conjunto de datos ADReSS— es posible mejorar el desempeño de la clasificación automática con el método de Autoselección de categorías, obteniendo un valor máximo de *f-score* para la clase positiva de 0.92 cuando se utiliza la variante DL-va con un valor de  $n = 30$ . La variante DL-va incorpora el valor de disponibilidad léxica de cada término en el esquema de pesado. Dado que el desempeño general de clasificación en este corpus es relativamente bueno, las ganancias entre utilizar Autoselección o no son pequeñas y solo se obtienen para 2 de 18 variantes (renglones con diferentes valores de  $n$ ). Sin embargo, al utilizar el método de Autoselección de categorías se logra mejorar la *precisión* de la clase positiva de 0.88 a 0.92 (comparando contra el resultado DL-base de la Tabla 8.4),

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

**Tabla 8.5:** Resultados de la clasificación para el problema de identificación de *Demencia* del conjunto ADReSS. Se compara el método de Auto-selección de categorías contra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Acc.), precisión de la clase positiva (P+), recuerdo de la clase positiva (R+), f-score de la clase positiva (F+), y macro f-score (F). Tonalidades de verde indican un incremento de F+ en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 8.4.

	<i>n</i>	Clasificador	Acc	ADReSS				inc.
				P+	R+	F+	(F)	
DL-base	0	RF-bool	0.90	0.88	0.92	0.90	(0.90)	*
	10	LR-bool	0.88	0.82	<b>0.96</b>	0.88	(0.87)	-0.02
	20	RF-tfidf	0.85	0.87	0.83	0.85	(0.85)	-0.05
	30	RF-tf	0.85	0.84	0.88	0.86	(0.85)	-0.04
	40	RF-tf	0.90	0.91	0.88	0.89	(0.90)	-0.01
	50	DT-tf	0.81	0.83	0.79	0.81	(0.81)	-0.09
	60	DT-bool	0.85	0.87	0.83	0.85	(0.85)	-0.05
	70	DT-mtf	0.90	0.91	0.88	0.89	(0.90)	-0.01
	80	DT-mtf	0.88	0.88	0.88	0.88	(0.88)	-0.02
	90	LSVC-tf	0.81	0.83	0.79	0.81	(0.81)	-0.09
DL-va	0	RF-tf	0.88	0.85	0.92	0.88	(0.87)	-0.02
	10	LR-mtf	0.85	0.90	0.79	0.84	(0.85)	-0.60
	20	RF-mtf	0.85	0.87	0.83	0.85	(0.85)	-0.05
	30	RF-tf	<b>0.92</b>	0.92	0.92	<b>0.92</b>	<b>(0.92)</b>	+0.02
	40	RF-tf	0.90	0.88	0.92	0.90	(0.90)	=
	50	DT-tf	0.81	0.83	0.79	0.81	(0.81)	-0.09
	60	DT-bool	0.85	0.87	0.83	0.85	(0.85)	-0.05
	70	RF-bool	0.83	0.86	0.79	0.83	(0.83)	-0.07
	80	DT-bool	0.85	0.90	0.79	0.84	(0.85)	-0.06
	90	RF-tf	0.81	<b>0.94</b>	0.67	0.78	(0.81)	-0.12

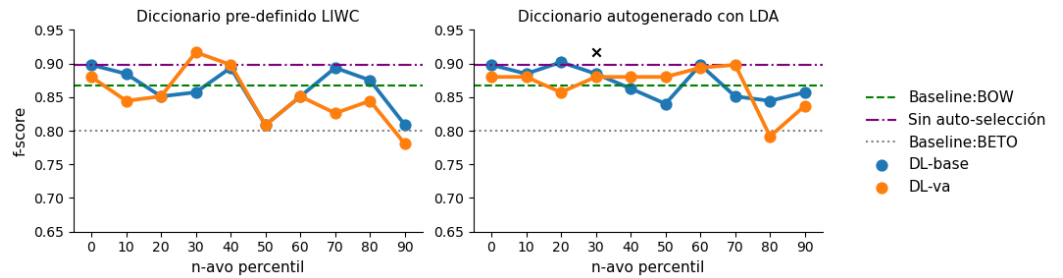
esto se logra sin decrementar los valores de recuerdo y f-score obtenidos al no utilizar la Autoselección de categorías.

La Figura 8.1 muestra gráficas del desempeño de clasificación con la comparación de todos los resultados obtenidos en el corpus ADReSS. Estas gráficas permiten visualizar qué tanto se mejora o no la clasificación al utilizar la disponibilidad léxica; también permite observar el impacto del conjunto de categorías seleccionadas por el método de Autoselección de categorías. La gráfica de la izquierda muestra resultados al utilizar LIWC; mientras que la gráfica de la derecha muestra los resultados obtenidos cuando se utiliza el Diccionario de Tópicos (DT). Este diccionario basado en LDA (el DT) se construye igual que como se describió en la Sección 7.3.2. Los valores indicados en el eje *y* de ambas gráficas corresponde al valor de f-score de la clase positiva.

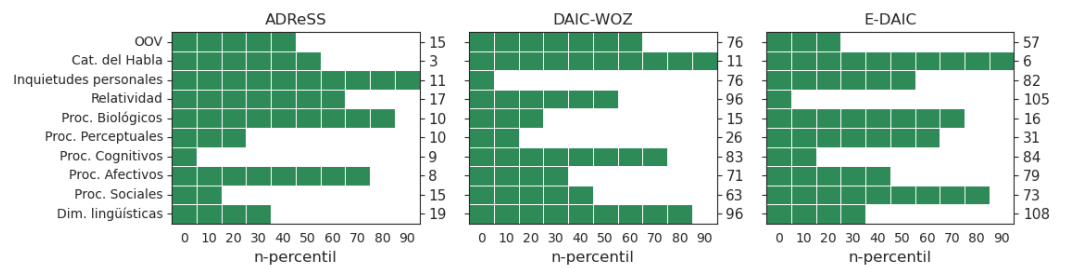
En este experimento se observan comportamientos similares que los obtenidos para los problemas de identificación de personalidad. Ya que las líneas en la gráfica de la derecha tienen cambios menos pronunciados que los observados al utilizar el Diccionario de Categorías LIWC, se puede inferir que los términos contenidos en el Diccionario de Tópicos no varían de forma importante para afectar la representación de las instancias y por consiguiente el desempeño

## 8.4 Experimento 2: Autoselección de categorías para mejorar la representación

**Figura 8.1:** Desempeño de modelos de clasificación generados con diferentes subconjuntos de términos organizados en categorías semánticas del recurso LIWC y DT para *Demencia* del corpus ADReSS. Las gráficas de líneas continuas comparan DL-base y DL-va, para 10 conjuntos de categorías semánticas a usar en la representación.



**Figura 8.2:** Mapas de calor de las tópicos seleccionados por el método de Autoselección de categorías para ADReSS, DAIC-WOZ y E-DAIC. La categoría OOV (*out of vocabulary*) agrupa todos los términos que no se encuentran en el Diccionario de Categorías.



del clasificador. Este comportamiento es más notorio al usar la variante DL-va, que incluye el valor de disponibilidad léxica en el esquema de pesado. Si bien puede que no haya mucha variedad entre los términos entre diferentes tópicos, también es posible inferir que los valores de disponibilidad léxica no son tan diversos que puedan apoyar en la clasificación.

Finalmente, el mejor resultado se logra con la utilización de LIWC como Diccionario de Categorías con  $n = 30$ . La Figura 8.2 muestra las categorías de LIWC seleccionadas para cada valor de  $n$ . De ahí podemos notar que las categorías que aportaron mayor información para obtener el mejor resultado son: Dimensiones lingüísticas (términos como pronombres, artículos, verbos, preposiciones, etc.), Procesos afectivos (que incluyen términos sobre emociones positivas o negativas, ansiedad, enojo y tristeza), Procesos biológicos (donde se incluyen términos sobre el cuerpo y la salud), Relatividad (que incluye términos relacionados con el movimiento, el espacio y el tiempo), Inquietudes personales (términos que indican placer, satisfacción, hogar, dinero, etc.), Categorías del habla y OOV. Interesantemente, los términos organizados dentro de la categoría Procesos cognitivos no son lo suficientemente distintos entre las dos clases para incluirlos dentro de la representación que obtuvo el mejor desempeño.

### DAIC-WOZ y E-DAIC

La Tabla 8.6 muestra los resultados de clasificación para el problema de la identificación de *Depresión*, tanto para el corpus DAIC-WOZ como para E-DAIC. El método propuesto basado en

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

disponibilidad léxica ya alcanza el desempeño más alto que los tres métodos de referencia utilizados (esto se ve gráficamente en la Figura 8.3), como se encontró en el Experimento 1 (descrito en la Sección 8.3). Este incremento se mejora aun más al utilizar el método de Autoselección de categorías basado en LIWC.

Particularmente, en DAIC–WOZ se logra una mejora absoluta de 0.07 con un  $n = 20$ , esto es un valor de *f*-score de 0.60 para la clase positiva vs. el valor obtenido de 0.53 sin utilizar la Autoselección de categorías. Si comparamos estas dos configuraciones ( $n = 0$  vs.  $n = 20$ ), con la Autoselección de categorías se recuperan menos instancias positivas (de 0.57 a 0.43 en valor de recuerdo), pero todas las recuperadas son identificadas correctamente, aumentando la precisión de la clase positiva de 0.5 a 1.0. En cuanto a las categorías de LIWC seleccionadas con esta mejor configuración (Figura 8.2) están Dimensiones lingüísticas, 4 de 5 categorías que corresponden a Procesos psicológicos: Procesos sociales, Procesos afectivos, Procesos cognitivos y Procesos biológicos; además de la categorías Relatividad, Categorías del habla y OOV. Con esta selección, no se consideran términos que pertenecen a las categorías de Procesos perceptuales e Inquietudes personales. Esto no significa que los participantes en estos corpora no hayan mencionado estos términos o que no tengan alta disponibilidad léxica, sino que la diferencia entre la disponibilidad léxica de estos y el resto de los términos ordenados de mayor a menor no son tan distintos en las clases.

En el corpus E–DAIC se logra mejorar el desempeño de clasificación en 0.03 para  $n = 10$  con la variante DL–base. Con esta configuración se logra un valor de *f*-score de 0.65 en contraste con el valor de 0.62 alcanzado sin el uso de la Autoselección de categorías. En este corpus, el incremento del *f*-score en la clase positiva resulta de que se mejora el recuerdo a pesar de que disminuye la precisión —ambos para la clase positiva— esto contrasta con el comportamiento que se observa para DAIC–WOZ. Una posible explicación es que el conjunto de entrenamiento para E–DAIC es aún más desbalanceado que para DAIC–WOZ y por lo tanto hay muchos más ejemplos de instancias en la clase de control que la clase positiva (como se puede ver en la Tabla 8.1). Que se *recuerden* más instancias positivas (0.53 vs. 0.59) indica que a pesar de haber más instancias de control, el método de Autoselección de categorías es capaz de identificar un mayor número de instancias positivas; a pesar de tener menor precisión para la clase positiva, en cuyo caso es relativamente mejor identificar incorrectamente a una persona sin depresión como persona con depresión que identificar a una persona con depresión como de control.

Casi todas las categorías de LIWC fueron útiles para obtener el mejor resultado en E–DAIC. Para  $n = 10$ , el método de Autoselección de categorías eligió 9 de las 10 categorías totales, incluyendo OOV (Figura 8.2, tercera gráfica). Otro aspecto importante de estos resultados es que la variante DL–va no proporciona información adicional para el clasificador. Es decir, en ningún caso se mejora el desempeño obtenido en el Experimento 1, cuando no se utiliza el método de Autoselección de categorías.



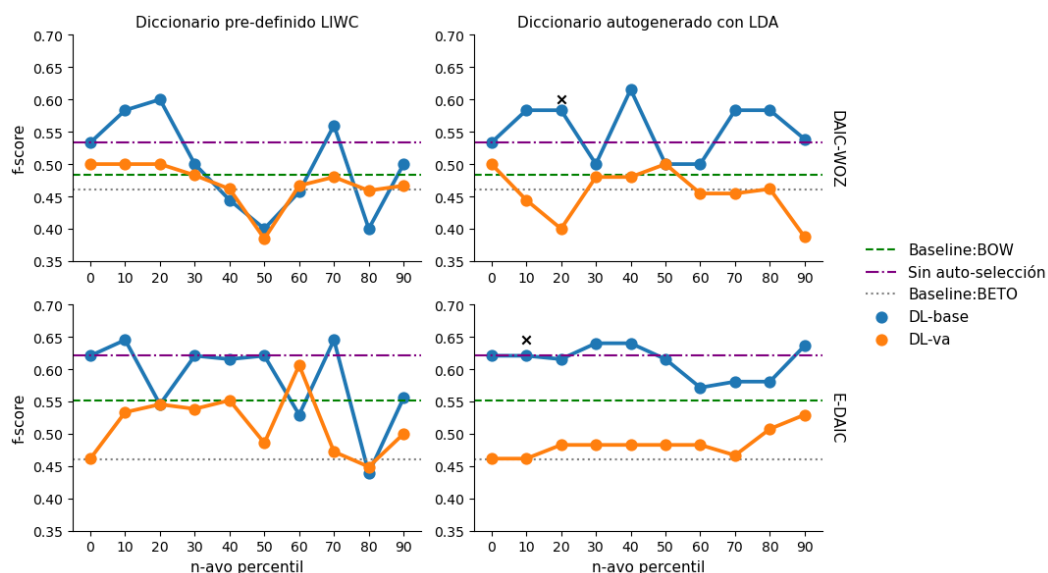
## 8.4 Experimento 2: Autoselección de categorías para mejorar la representación

**Tabla 8.6:** Resultados de la clasificación para el problema de identificación de *Depresión* de los conjuntos de datos DAIC-WOZ y E-DAIC. Se compara el método de Autoselección de categorías con tra utilizar todas las categorías presentes durante el entrenamiento del modelo. Se reporta la exactitud (Ex.), precisión de la clase positiva (P+), recuerdo de la clase positiva (R+),  $f$ -score de la clase positiva (F+), y macro  $f$ -score (F). Tonalidades de verde indican un incremento de F+ en comparación con el mejor resultado de referencia. En gris se presenta el mejor resultado de la Tabla 8.4

$n$	DAIC-WOZ						E-DAIC							
	Clasif.	Ex.	P+	R+	F+	inc.	Clasif.	Ex.	P+	R+	F+	inc.		
0	DT-mtf	0.70	0.50	0.57	0.53	(0.66)	*	LSVC-tf	0.80	0.75	0.53	0.62	(0.74)	*
10	MNB-tf	0.79	0.70	0.50	0.58	(0.72)	+0.05	LSVC-tf	0.80	0.71	0.59	<b>0.65</b>	( <b>0.75</b> )	+0.03
20	CNB-mtf	<b>0.83</b>	<b>1.00</b>	0.43	<b>0.60</b>	( <b>0.75</b> )	+0.07	LR-tf	0.73	0.56	0.53	0.55	(0.68)	-0.07
30	MLP-bool	0.79	0.83	0.36	0.50	(0.68)	-0.03	LSVC-bool	0.80	0.75	0.53	0.62	(0.74)	=
40	MLP-bool	0.79	<b>1.00</b>	0.29	0.44	(0.66)	-0.09	LR-bool	<b>0.82</b>	<b>0.89</b>	0.47	0.62	( <b>0.75</b> )	=
50	CNB-bool	0.74	0.67	0.29	0.40	(0.62)	-0.13	MLP-bool	0.80	0.75	0.53	0.62	(0.74)	=
60	Perc-tf	0.60	0.38	0.57	0.46	(0.57)	-0.07	DT-bool	0.71	0.53	0.53	0.53	(0.66)	-0.09
70	CNB-bool	0.77	0.64	0.50	0.56	(0.70)	+0.03	DT-bool	0.80	0.71	0.59	<b>0.65</b>	( <b>0.75</b> )	+0.03
80	CNB-tf	0.55	0.33	0.50	0.40	(0.52)	-0.13	Perc-tfidf	0.59	0.38	0.53	0.44	(0.56)	-0.18
90	DT-mtf	0.66	0.44	0.57	0.50	(0.62)	-0.03	CNB-tf	0.57	0.41	0.88	0.56	(0.57)	-0.06
0	DT-mtf	0.70	0.50	0.50	0.50	(0.64)	-0.03	LSVC-tf	0.75	0.67	0.35	0.46	(0.65)	-0.16
10	DT-mtf	0.70	0.50	0.50	0.50	(0.64)	-0.03	LSVC-tf	0.75	0.62	0.47	0.53	(0.68)	-0.09
20	DT-mtf	0.70	0.50	0.50	0.50	(0.64)	-0.03	DT-mtf	0.73	0.56	0.53	0.55	(0.68)	-0.07
30	DT-mtf	0.68	0.47	0.50	0.48	(0.63)	-0.05	LSVC-bool	0.79	0.78	0.41	0.54	(0.70)	-0.08
40	MLP-mtf	0.55	0.36	0.64	0.46	(0.54)	-0.07	DT-tf	0.77	0.67	0.47	0.55	(0.70)	-0.07
50	LR-tf	0.66	0.42	0.36	0.38	(0.57)	-0.15	Perc-mtf	0.36	0.32	<b>1.00</b>	0.49	(0.31)	-0.13
60	Perc-mtf	0.32	0.30	<b>1.00</b>	0.47	(0.26)	-0.06	DT-bool	0.77	0.63	0.59	0.61	(0.72)	-0.01
70	MLP-mtf	0.72	0.55	0.43	0.48	(0.65)	-0.05	Perc-mtf	0.32	0.31	<b>1.00</b>	0.47	(0.26)	-0.15
80	Perc-mtf	0.30	0.30	<b>1.00</b>	0.46	(0.23)	-0.07	CNB-tf	0.52	0.34	0.65	0.45	(0.51)	-0.17
90	DT-mtf	0.66	0.44	0.50	0.47	(0.61)	-0.06	CNB-mtf	0.54	0.37	0.76	0.50	(0.53)	-0.12

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis

**Figura 8.3:** Desempeño de modelos de clasificación generados con diferentes subconjuntos de términos organizados en categorías semánticas del recurso LIWC y DT para *Depresión* de los corpora DAIC-WoZ y E-DAIC. Se muestran tres *base-lines*: la representación basada en BoW, la representación basada en BERT y el mejor desempeño del uso de la disponibilidad léxica sin el filtrado de categorías. Las gráficas de líneas continuas comparan DL-base y DL-va, para 10 conjuntos de categorías semánticas a usar en la representación.



Además de los experimentos utilizando el Diccionario de Categorías predefinido LIWC, realicé experimentos con un Diccionario de Categorías autogenerado basado en LDA ó Diccionario de Tópicos (DT). Las gráficas de la derecha de la Figura 8.3 muestran los resultados obtenidos con el DT para los dos corpora de *Depresión*: DAIC-WOZ y E-DAIC. Estas gráficas también muestran, con una cruz, el mejor resultado obtenido con el Diccionario de Categorías LIWC (que corresponde al valor más alto alcanzado en las gráficas de la izquierda).

Para DAIC-WOZ, se observa que existe una ligera mejora al utilizar el Diccionario de Tópicos que con el uso de LIWC con valores de *f-score* (0.60 vs. 0.61) con la variante DL-base. Por otro lado, para E-DAIC, el comportamiento es consistente a los obtenidos previamente, donde la variación entre diferentes valores de *n* no es tan evidente y/o no existe un incremento en el desempeño del clasificador al utilizar el Diccionario basado en LDA.

### 8.4.1. Comparación entre DL-base y DL-va en tareas de identificación de trastornos psicológicos

El método propuesto para identificación de personalidad que se basa en la idea de utilizar información del léxico mental en la representación de las muestras de lenguaje es útil también para la identificación de demencia y depresión. Sin bien, la experimentación no es exhaustiva, sí se han utilizado 3 conjuntos de datos que han sido empleados para validar otros métodos de clasificación automática en estos dos problemas.

## 8.4 Experimento 2: Autoselección de categorías para mejorar la representación

---

De las variantes del método propuesto con Autoselección de categorías, la variante DL–base es mejor que DL–va para *Depresión* en los 2 corpora. La principal diferencia entre estas dos variantes es que la primera solo considera el conjunto de términos en las listas de disponibilidad léxica; mientras que la segunda variante considera además el valor de disponibilidad léxica de tales términos. Que funcione mejor la variante DL–base, en DAIC–WOZ y E–DAIC, implica que los valores de disponibilidad léxica en las dos clases no varían de forma significativa. Lo que a su vez indica que los participantes en las dos clases además de usar un vocabulario muy parecido, acceden a los mismos términos con valores de disponibilidad similares en ambas clases.

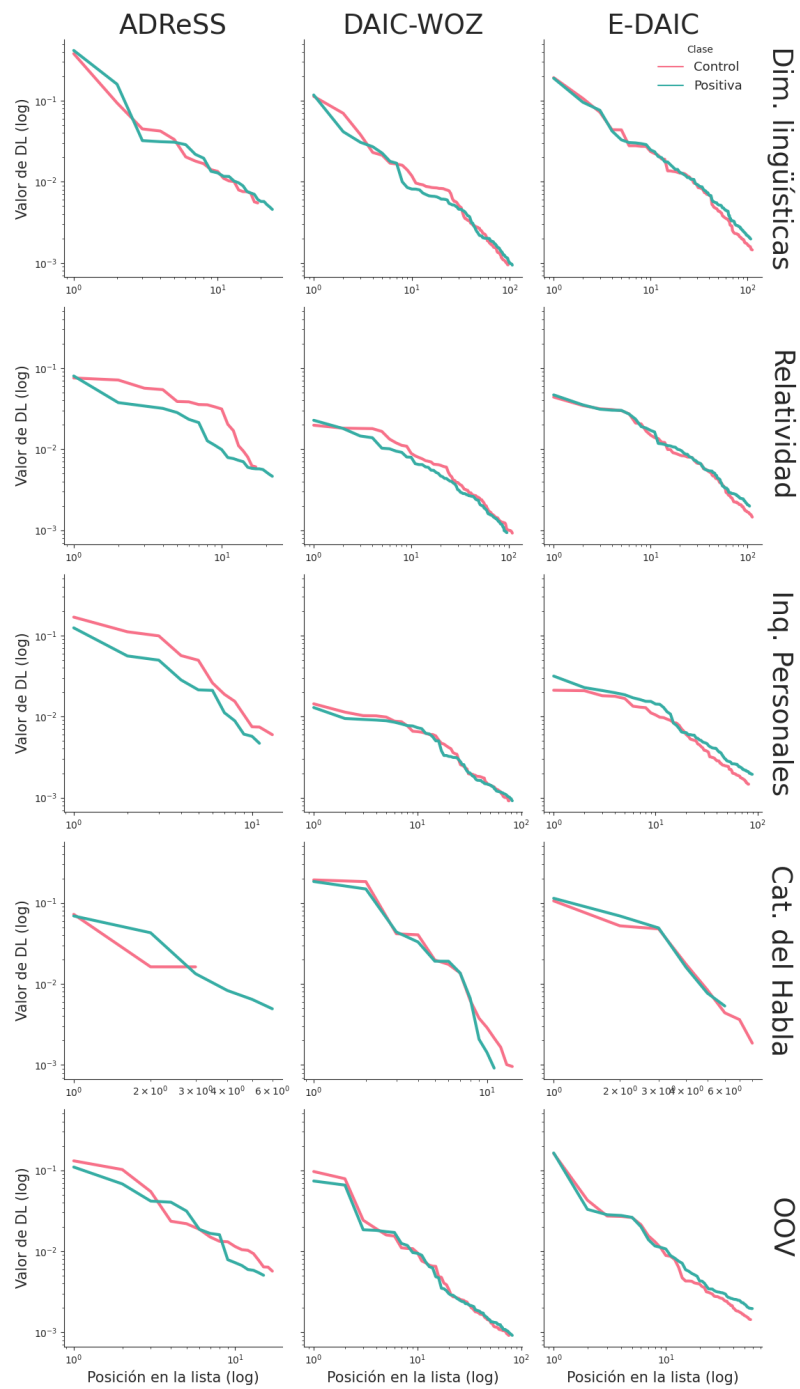
Para ilustrar la diferencia entre DL–base y DL–va para estos problemas de clasificación, la Figura 8.4 muestra cómo se ven los valores de disponibilidad léxica en los 3 conjuntos de datos para las 10 categorías semánticas de LIWC. Estas gráficas muestran el valor de disponibilidad léxica del *i*-vo término en cada categoría, estas líneas son las que se utilizó para calcular el MAE que sirve de métrica para seleccionar una categoría por el método de Autoselección de categorías como lo describe la Sección 6.3.1.

En los experimentos en el corpus ADReSS la variante DL–va es más útil que la variante DL–base al utilizar con el método de Autoselección de categorías y como puede verse en las gráficas de la Figura 8.4 los valores de disponibilidad léxica de los términos en la mayoría de las categorías son más distantes —las líneas están más separadas entre sí— para ADReSS que para los corpora de *Depresión* DAIC–WOZ y E–DAIC.

### 8.4.2. Resumen

Así, los experimentos de identificación de *Depresión* y *Demencia* muestran evidencia de que el método propuesto ayuda a identificar vocabulario relevante para ambas tareas. Más aún, la autoselección de categorías semánticas cuando se utiliza un diccionario manualmente generado como LIWC en lugar de utilizar un diccionario automáticamente generado como el Diccionario de Tópicos generado con LDA, puede ayudar a seleccionar un subconjunto de términos que apoyen la clasificación automática de *Depresión* y *Demencia* en los corpora utilizados. Esto apoya la intuición de que información del dominio es importante. Similarmente que con el caso de personalidad, el método propuesto proporciona elementos que pueden ayudar al experto a interpretar los resultados obtenidos de forma automática.

## Capítulo 8. Identificación automática de Demencia y Depresión: experimentación y análisis



**Figura 8.4:** Valores de disponibilidad léxica de los términos por clase en las 10 categorías de LIWC consideradas por el método de Autoselección de categorías, en los conjuntos de datos ADReSS, DAIC-WOZ y E-DAIC.

A horizontal decorative bar with a repeating purple pattern of overlapping, rounded, shell-like shapes. The text "Parte IV" is written in white, bold, serif font on the left side of the bar.

**Parte IV**

## **Conclusiones generales**



# 9

## Conclusiones

La investigación realizada en esta tesis está en la intersección de Psicología, Sociolingüística y Computación. De Psicología es el dominio del tema que se quiere entender, identificar rasgos de personalidad (y trastornos mentales como demencia y depresión); de Sociolingüística se obtiene el marco teórico que permitió aproximar la solución de representación de muestras de lenguaje; y de Computación el marco para la generación de modelos automáticos para identificar rasgos de personalidad. En este contexto la hipótesis planteada propone que la representación de textos basada en el análisis del léxico disponible entre dos comunidades con características en común, es posible identificar rasgos de personalidad en un enfoque de clasificación supervisada. Y mediante la experimentación se demostró que la hipótesis es correcta, pudiendo obtener desempeños comparables con métodos del estado del arte (para uno de los corpora); además de obtener mejores resultados, para ciertos rasgos, en comparación con los métodos de referencia utilizados. Se mostró evidencia de que la información extraída del lexicón mental es importante para la identificación automática de rasgos de personalidad (Capítulo 7), y demencia y depresión (Capítulo 8).

Para realizar estos experimentos, creamos un corpus (Capítulo 5) que permitiera estudiar el fenómeno, el corpus RxPI es ahora un recurso que otros investigadores pueden utilizar para entender más sobre la relación entre textos cortos derivados de narraciones personales y la personalidad del modelo *Big-Five*. A pesar de ser un corpus relativamente pequeño en comparación con el corpus Essays, RxPI fue construido con el propósito particular de estudiar la relación entre el lenguaje y la personalidad, por lo que la instrucción utilizada para capturar las narraciones cortas está respaldada por instrumentos psicológicos como la entrevista de historia de vida. La calidad de los instrumentos usados para identificar aspectos de la psicología humana es importante en el área de Psicología; por lo que el corpus RxPI puede resultar especialmente valioso dentro del área de la salud mental; y sirva para impulsar el uso de herramientas tecnológicas derivadas de

estos instrumentos y corpora en esta área.

El método propuesto de representación transforma muestras cortas de lenguaje en vectores numéricos (realiza una representación de textos) que serán empleados para aprender un modelo automático para identificar personalidad (Capítulo 6). El método propone obtener una lista del vocabulario o léxico disponible de las dos comunidades a contrastar. Este léxico disponible cuenta con información como la pertenencia y orden de términos dentro de este vocabulario. Luego, el método propuesto considera tanto la pertenencia (con la variante DL-base) como el orden (con la variante DL-va) para generar la representación de las muestras de lenguaje. La adaptación de la prueba de disponibilidad léxica para utilizarse en intenciones comunicativas —o narraciones— propone una forma de obtener el léxico disponible relevante a un evento personal. Si una palabra se usa constantemente en la narración puede indicar una mayor disponibilidad que además es relevante al evento narrado. En cuanto a la representación propuesta, es posible generar vectores de dimensiones moderadas y explicables; que puede ayudar a psicólogos a entender mejor la clasificación automática propuesta o a guiar la exploración subsecuente de la persona en cuestión.

Adicionalmente, el método propuesto utiliza un Diccionario de Categorías con la idea de mejorar la representación al seleccionar categorías semánticas relevantes con los rasgos de personalidad. Mientras que este Diccionario de Categorías puede ser arbitrario, los experimentos (Capítulo 7) emplean dos tipos: el recurso LIWC y también uno creado de forma automática a través de la generación de tópicos basada en LDA (Sección 6.3.2). Se mostró que los Diccionarios predefinidos por expertos son mejores que Diccionarios generados con el conjunto de datos disponible de cada problema a resolver. El método propuesto también propone un módulo de Autoselección de categorías (Sección 6.3) que determina automáticamente las categorías más relevantes que aparecen en el Diccionario de Categorías utilizado. Este método de Autoselección de categorías aporta información adicional a los expertos que decidan utilizarlo, pues genera una lista de categorías, que son útiles al clasificador, para identificar automáticamente rasgos de personalidad. Esta división del léxico disponible dentro de categorías predefinidas o autogeneradas puede ayudar a organizar la narración en análisis subsecuentes que pueden aportar mayor información al especialista.

### 9.1. Perspectivas y trabajo futuro

Los resultados experimentales muestran desempeños prometedores en la clasificación automática al utilizar el método propuesto para identificar personalidad y también para contextos de salud mental como demencia y depresión. Como una línea de investigación en el futuro considero analizar la relación de otros problemas de la salud mental con el léxico disponible y en general con el lexicón mental, por ejemplo, estrés y ansiedad. Es importante recalcar que dicho



análisis requiere de datos relacionados con estos problemas. Este requisito representará un reto en términos de recolección, pero permitirá estudiar la naturaleza de dichos datos en una manera similar al análisis de la Sección 7.2.1.

En ese sentido, un segundo trabajo a futuro será generar un léxico disponible *prototípico* de cada rasgo de personalidad que contenga información de las definiciones de cada rasgo, instrumentos especializados —como cuestionarios con un mayor número de reactivos— y creado en colaboración con especialistas. Con este *léxico disponible prototípico* será posible generar textos sintéticos pero ajustados a los léxicos disponibles de los conjuntos reales contenidos en RxPI. Si este método de aumento de datos resulta ser eficiente y confiable, como tercer trabajo a futuro, buscaré utilizarlo en otros problemas que cuenten con instrumentos especializados como depresión y demencia.

Finalmente, el corpus RxPI contiene etiquetas de fenómenos de la escritura a mano como modificaciones de palabras, borrado de palabras, uso de silabación, entre otros. Así, una representación que considere dichos fenómenos espontáneos podría contribuir al análisis e identificación automática de la personalidad.

## 9.2. Consideraciones éticas

El acelerado desarrollo de la tecnología en Inteligencia Artificial para la generación de modelos automáticos que simulan cada vez mejor el comportamiento humano, propuso un reto primero a la comunidad científica, luego a las entidades gubernamentales para prevenir usos que produzcan daño a la humanidad o al medio ambiente. En este contexto, en el 2021 La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) aprobó un documento que contiene una Recomendación sobre la Ética de la Inteligencia Artificial (UNESCO, 2021).

Así, los Estados Miembros deben aplicar esta Recomendación de manera voluntaria y también deben “señalar la Recomendación a la atención de autoridades, organismos, organizaciones universitarias y de investigación, instituciones y organización del sector público, privado y de la sociedad civil que participan en las tecnologías de la IA, para que el desarrollo y la utilización de estas tecnologías se guíen tanto por una investigación científica sólida como por un análisis y una evaluación éticos.” (UNESCO, 2021). Este conjunto de principios y guías deben considerarse en todo el ciclo de vida de los sistemas automáticos, desde la concepción hasta la terminación o desmantelamiento del sistema. En esta investigación, se han tenido presente los principios sugeridos, durante la construcción del corpus RxPI y hasta la evaluación del modelo construido. Puede verse que el *derecho a la privacidad, la protección de datos* es uno de los pilares importantes en la construcción de RxPI. Que la *autonomía y supervisión humana* es uno de los principales

## Capítulo 9. Conclusiones

---

puntos del modelo automático construido; así como la *transparencia y explicabilidad*.

El documento de Recomendación sobre la Ética de la Inteligencia Artificial está organizado en una serie de puntos divididos en 11 ámbitos de actuación, entre ellos el ámbito de actuación 8: Educación e Investigación y el ámbito de actuación 11: Salud y Bienestar social.

Dentro del Ámbito de actuación 8: Educación e Investigación, el punto 108 establece (UNESCO, 2021, p. 35):

Los Estados Miembros deberían velar por que los investigadores en IA reciban formación en ética de la investigación y exigirles que tengan en cuenta consideraciones éticas en sus concepciones, productos y publicaciones, especialmente en los análisis de los conjuntos de datos que utilizan, la forma en que estos se anotan y la calidad y el alcance de los resultados, así como las posibles aplicaciones.

A este respecto, la investigación en esta tesis, que contiene un conjunto de datos, RxPI, ha tenido presente desde su concepción la importancia del análisis de los datos que se utilizan. Por ejemplo, la inclusión en el Capítulo 5 de una descripción detallada de este corpus; ya que una de las prioridades del proyecto fue considerar conjuntos de datos de alta calidad obtenidos de forma ética.

Dentro del Ámbito de actuación 11: Salud y Bienestar social, el punto 123 establece (UNESCO, 2021, p. 38):

Los Estados Miembros deberían prestar particular atención a la regulación de las soluciones de predicción, detección y tratamiento médicos en las aplicaciones de la IA, mediante:

- a) la supervisión para minimizar y atenuar sesgos;
- b) la inclusión profesional, el paciente, el cuidador o el usuario del servicio [...] en todas las etapas pertinentes al elaborar los algoritmos;
- c) una debida atención a la privacidad [...] en materia de protección de datos;
- d) mecanismos eficaces para que las personas cuyos datos personales se están analizando sepan de la utilización el análisis de sus datos y den consentimiento informado al respecto, sin impedir el acceso a la atención de la salud;
- e) la garantía de que el cuidado humano y la decisión final sobre el diagnóstico y el tratamiento correspondan siempre a seres humanos, reconociendo al mismo tiempo que los sistemas de IA también pueden ayudarlos en su trabajo;

### 9.3 Colecciones de datos para investigación en el ámbito de la salud

---

- f) el examen, cuando sea necesario, de los sistemas de IA por un comité de investigación ética antes de su uso clínico.

En este contexto, al recolectar el conjunto de datos RxPI, fue requisito que quienes decidieron participar lo hicieran con el conocimiento del objetivo del estudio, el tipo de análisis a realizar y la importancia de su participación. Así, todas las instancias dentro del conjunto de datos son de personas que dieron su consentimiento explícito del uso de la información que proporcionaron; en el Apéndice A.1 puede leerse el instrumento de recolección de datos para la generación de RxPI.

Sobre la protección de datos, se ha descrito en el Capítulo 5 que los datos privados que puedan identificar a las personas participantes fueron ofuscados; además de que ningún otro dato sobre edad, nombres o otras redes sociales disponibles es parte del conjunto de datos RxPI. Particularmente, en la Sección 5.1.2, los apartados *información confidencial*, *identificación de las subpoblaciones*, *identificación de individuos* y *datos sensibles* detallan cómo se implementa la protección de datos; aunado con los apartados: *proceso de revisión ética*, *notificación de participación*, *consentimiento* y *exclusión voluntaria*, en la Sección 5.1.3, hablan sobre los mecanismos usados para obtener consentimiento informado de los participantes dentro de la recolección de RxPI.

En la elaboración del algoritmo propuesto en esta tesis, participaron dos psicólogos que dieron sus opiniones, ideas y análisis durante el proceso de creación. El trabajo conjunto con los dos psicólogos expertos aparece en (Ramírez-de-la-Rosa et al., 2023).

Finalmente, el método propuesto para identificar rasgos de personalidad —y que también puede ser usado para identificar demencia o depresión— tiene el objetivo de apoyar al especialista, no a dar un diagnóstico final.

### 9.3. Colecciones de datos para investigación en el ámbito de la salud

La construcción o recolección de conjuntos de datos para análisis y desarrollo de sistemas automáticos para salud y en particular para la salud mental no es simple, sencilla o barata. Debido a la limitante de obtener datos clínicos de alta calidad, los investigadores de ciencias computacionales y otras áreas han recurrido a la obtención de muestras de plataformas de redes sociales como Twitter (o X), Reddit, Facebook y otros foros. Mientras que este tipo de recursos permite obtener datos de forma rápida, los criterios de construcción no son del todo confiables, y los métodos para establecer casos positivos y negativos no son estándares (Chancellor and De Choudhury, 2020; Harrigian et al., 2021).

Desde el 2017, la comunidad científica en el área de Computación ha estado tratando de mitigar

la baja calidad de los datos obtenidos de redes sociales y que al mismo tiempo sean extraídos considerando los principios éticos básicos como el consentimiento informado. Uno de estos trabajos que trata de generar protocolos para la extracción de datos de redes sociales para investigación en la salud es el realizado por Benton et al. (2017) y más recientemente la ficha técnica para conjuntos de datos (*datasheets for datasets*) creada por Gebru et al. (2021). Sin embargo, dentro de la comunidad científica poner en práctica estos protocolos y recomendaciones no siempre se alinea con la tendencia de la comunidad de generar nuevos modelos de forma acelerada para competir con el nuevo resultado del estado de arte.

Afortunadamente, existen esfuerzos conjuntos de la comunidad científica para hacer frente a este contexto. Por ejemplo, en el 2020 Association for Computational Linguistics (2020) adoptó el código de ética de la Association for Computing Machinery (2018) y por consiguiente, en los envíos a revisión de artículos científicos bajo el esquema ACL *Rolling Review* se solicita un formulario con una lista de verificación que entre otras áreas se enfoca en cuestiones éticas como las limitaciones, riesgos potenciales de la investigación, protección de datos en el caso de uso de corpora, el uso de consentimiento informado, entre otros <sup>1</sup>.

### 9.4. Publicaciones científicas

A continuación, listo de los artículos científicos derivados de esta tesis de doctorado:

1. Ramírez-de-la-Rosa, G., Jiménez-Salazar, H., Villatoro-Tello, E., Reyes-Meza, V., and Rojas-Avila, J. (2023). A lexical–availability–based framework from short communications for automatic personality identification. *Cognitive Systems Research*
2. Villatoro-Tello, E., Ramírez-de-la-Rosa, G., Gática-Pérez, D., Magimai.-Doss, M., and Jiménez-Salazar, H. (2021b). *Approximating the Mental Lexicon from Clinical Interviews as a Support Tool for Depression Detection*, pages 557–566. Association for Computing Machinery, New York, NY, USA
3. Villatoro-Tello, E., Dubagunta, S. P., Fritsch, J., Ramírez-de-la-Rosa, G., Motlicek, P., and Magimai-Doss, M. (2021a). Late Fusion of the Available Lexicon and Raw Waveform-Based Acoustic Modeling for Depression and Dementia Recognition. In *Proc. Interspeech 2021*, pages 1927–1931
4. Ramírez, G., Villatoro, E., Ionescu, B., Escalante, H. J., Escalera, S., Larson, M., Müller, H., and Guyon, I. (2019). Overview of the multimedia information processing for personality & social networks analysis contest. In Zhang, Z., Suter, D., Tian, Y., Branzan Albu, A., Sidère,

---

<sup>1</sup>Puede obtenerse más información en <https://aclrollingreview.org/responsibleNLPresearch/>

- N., and Jair Escalante, H., editors, *Pattern Recognition and Information Forensics*, pages 127–139, Cham. Springer International Publishing
5. Ramírez-de-la Rosa, G., Villatoro-Tello, E., Jiménez-Salazar, H., Pinto, D., Singh, V. K., Villavicencio, A., Mayr-Schlegel, P., and Stamatatos, E. (2018). TxPI-u: A resource for personality identification of undergraduates. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 34(5):2991–3001



# A



## Creación del conjunto de datos RxPI

### A.1. El instrumento de recolección

Las Figuras A.1<sup>1</sup> y A.2 muestran las tres páginas del instrumento creado para recolectar el conjunto de datos RxPI. La primera página contiene información sobre la investigación para conocimiento de los participantes, y solicita información sobre datos generales, para uso interno en el proceso de recolección, que posteriormente se usa para generar estadísticas. Ninguna de la información solicitada bajo *Datos Generales* forma parte del conjunto de datos RxPI. La segunda página contiene el cuestionario TIPI usado para obtener la personalidad de cada participante. Y finalmente, la página 3 contiene la instrucción detonante del ensayo.

---

<sup>1</sup>Originalmente, el proyecto de recolección de datos contó con dos etapas; la primera etapa tenía como objetivo recolectar información textual; y la segunda parte tenía como objetivo recolectar información de audio y video. Esta es la razón por la que el título del instrumento hace referencia a un *corpus multimodal*. Desafortunadamente, la segunda etapa no se concluyó durante la realización de esta investigación.

## Apéndice A. Creación del conjunto de datos RxPI

Figura A.1: Instrumento usado para la recolección del corpus RxPI.

### Instrumento para la etapa 1 de la construcción de un corpus multimodal en español mexicano

Gabriela Ramirez de la Rosa, Esau Vilaloro Tello, Héctor Jiménez Salazar  
Profesores investigadores del departamento de Tecnologías de la Información  
División de Ciencias de la Comunicación y el Diseño  
UAM-C

El instrumento tiene como objetivo recolectar la información de referenda para la generación del corpus multimodal de personalidad.

Los objetivos particulares son:

1. Obtener la *personalidad de referenda* de los individuos mediante un instrumento ampliamente usado por Psicólogos (TIPI).
2. Obtener texto informal de la escritura de un ensayo sobre la vida propia de los participantes. La finalidad de este texto será buscar correlaciones entre el estilo de escritura y la personalidad de referenda obtenida en el objetivo anterior.

#### Información importante

El resultado de este estudio será muy importante y relevante para la comunidad científica en México, en particular para la comunidad UAM-Cuajimalpa. Con tus respuestas ayudarás a construir una pieza fundamental en esta investigación.

Si estas de acuerdo en participar en el estudio, por favor, contesta lo más honestamente posible a cada sección. Tu decisión de participar en el estudio es completamente voluntaria, quedando en entendido que: la información obtenida en el estudio se tratará de forma confidencial por parte de los investigadores, donde los datos personales serán reemplazados por identificadores alfa-numéricos. *Si tienes dudas o comentarios sobre esta actividad, ponte en contacto con la Mtra. Gabriela Ramirez de la Rosa, profesora del Departamento de Tecnologías de la Información, DCCD, UAM-C, Cubículo 602, Piso 6to, Torre III, Teléfono 55146500 Ext. 2948, correo electrónico: [gramirez@ccfco.uamh.unl.mx](mailto:gramirez@ccfco.uamh.unl.mx).*

Yo, \_\_\_\_\_, convengo en participar en este estudio voluntariamente. Y entiendo que los datos obtenidos en el estudio pueden ser usados en publicaciones con fines científicos.

Firma: \_\_\_\_\_ Fecha: \_\_\_\_\_

De antemano muchas gracias por tus respuestas. Por favor indicanos tu dirección de correo electrónico para enviarle los resultados de este estudio \_\_\_\_\_

#### Datos generales

Nombre completo: \_\_\_\_\_  
 Fecha de nacimiento: \_\_\_\_/\_\_\_\_/\_\_\_\_ (día/mes/año)  
 Coloca una X en tu género:    Mujer \_\_\_\_\_    Hombre \_\_\_\_\_    Prefiero no decir \_\_\_\_\_  
 Si tienes cuenta en Twitter, por favor colócala aquí: \_\_\_\_\_  
 Si tienes cuenta en Facebook, por favor colócala aquí: \_\_\_\_\_  
 Por favor, indica la carrera que estudias: \_\_\_\_\_

#### Cuestionario

Aquí encontrarás una serie de características de personalidad que pueden o no referirse a ti. Por favor, cruza una casilla de cada expresión con el fin de indicar en qué medida estás de acuerdo o en desacuerdo con ella. Deberás valorar el grado en que cada par de características (por ejemplo: extravertida y entusiasta) te describen mejor, aunque una pueda corresponder en mayor grado que la otra.

#### Escala:

- 1: Totalmente en desacuerdo
- 2: Bastante en desacuerdo
- 3: Un poco en desacuerdo
- 4: Ni de acuerdo ni en desacuerdo
- 5: Un poco de acuerdo
- 6: Bastante de acuerdo
- 7: Totalmente de acuerdo

Me veo a mí mismo/a como a una persona:	1	2	3	4	5	6	7	Totalmente de acuerdo
Extravertida, entusiasta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Colérica, disculidora	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiable, auto-disciplinada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ansiosa, fácilmente alterable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Abierta a nuevas experiencias, polifacética	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reservada, callada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comprensiva, amable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Desorganizadada, descuidada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Serena, emocionalmente estable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tradicional, poco imaginativa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>





## **A.2. Asignación de valores numéricos y clases nominales**

La Figura A.3 muestra el algoritmo que se utilizó para extraer los valores numéricos y clases asociadas a cada rasgo de personalidad de acuerdo a las normas establecidas por Gosling et al. (2003).

**Figura A.3:** Algoritmo para extraer los valores numéricos y nominales asociadas a cada rasgo de personalidad del instrumento TIPI.

### Cálculo de la clase de cada rasgo de acuerdo al instrumento TIPI.

Gabriela Ramírez de la Rosa - Nov 2018.

La descripción del instrumento TIPI puede encontrarse en:

<https://psjoling.ipsu.tecnas.edu/wp-content/uploads/2014/09/IRP-03-tipi.pdf>

Otros fuentes consultadas fueron tomadas de

<https://psjoling.ipsu.tecnas.edu/scales/webe-developed/ten-item-personality-measure.html>

Procedimiento para el cálculo de la Clase por rasgo  
Entrada: arreglo entero de respuestas a las preguntas: items[10] (índices de 1 a 10).  
El arreglo items contiene valores enteros entre 1 y 7.  
Salida: La clase nominal de cada uno de los rasgos.

1. Calcular el score para cada rasgo como sigue:
  - a.  $sExtraversion = (items[1] + (8-items[6])) / 2$
  - b.  $sAgradabilidad = ((8-items[2]) + items[7]) / 2$
  - c.  $sResponsabilidad = (items[3] + (8-items[8])) / 2$
  - d.  $sEstabilidad = ((8-items[4]) + items[9]) / 2$
  - e.  $sAbierto = (items[5] + (8-items[10])) / 2$
2. Asignar la clase de acuerdo con los siguientes intervalos de cada rasgo:
  - a. Para extraversión
    - i.  $if(sExtraversion >= 0 \text{ and } sExtraversion < 2.99)$
    - ii.  $Else \text{ if } (sExtraversion >= 2.99 \text{ and } sExtraversion < 4.44)$
    - iii.  $Else \text{ if } (sExtraversion >= 4.44 \text{ and } sExtraversion < 5.89)$
    - iv.  $Else$ 
      1.  $ClassExt = \text{media alta}$
  - b. Para Agradabilidad
    - i.  $if(sAgradabilidad >= 0 \text{ and } sAgradabilidad < 4.12)$
    - ii.  $Else \text{ if } (sAgradabilidad >= 4.12 \text{ and } sAgradabilidad < 5.23)$
    - iii.  $Else \text{ if } (sAgradabilidad >= 5.23 \text{ and } sAgradabilidad < 6.34)$
    - iv.  $Else$ 
      1.  $ClassAgr = \text{media alta}$
  - c. Para Responsabilidad
    - i.  $if(sResponsabilidad >= 0 \text{ and } sResponsabilidad < 4.08)$
    - ii.  $Else \text{ if } (sResponsabilidad >= 4.08 \text{ and } sResponsabilidad < 5.4)$
    - iii.  $Else \text{ if } (sResponsabilidad >= 5.4 \text{ and } sResponsabilidad < 6.72)$

- iv.  $Else$ 
  1.  $claseses = \text{media alta}$
- d. Para Estabilidad Emocional
  - i.  $if(sEstabilidad >= 0 \text{ and } sEstabilidad < 3.41)$
  - ii.  $Else \text{ if } (sEstabilidad >= 3.41 \text{ and } sEstabilidad < 4.83)$
  - iii.  $Else \text{ if } (sEstabilidad >= 4.83 \text{ and } sEstabilidad < 6.25)$
  - iv.  $Else$ 
    1.  $claseEst = \text{media alta}$
- e. Para Abierto a nuevas experiencias
  - i.  $if(sAbierto >= 0 \text{ and } sAbierto < 4.31)$
  - ii.  $Else \text{ if } (sAbierto >= 4.31 \text{ and } sAbierto < 5.38)$
  - iii.  $Else \text{ if } (sAbierto >= 5.38 \text{ and } sAbierto < 6.45)$
  - iv.  $Else$ 
    1.  $claseabi = \text{media alta}$

Resumen:

TIPI scale scoring (R' denotes reverse-scored items): Extraversion: 1, 6R; Agreeableness: 2R, 7; Conscientiousness: 3, 8R; Emotional Stability: 4R, 8; Openness to Experiences: 5, 10R.

Extraversión: LOOKUP(B11,(0.2,99,4.44,5.89))("Low","Medium Low","Medium High","High")  
Agradabilidad: LOOKUP(C1,(0.4,12,5.23,6.34))("Low","Medium Low","Medium High","High")  
Responsabilidad: LOOKUP(D1,(0.4,08,5.4,6.72))("Low","Medium Low","Medium High","High")  
Estabilidad Emocional: LOOKUP(E1,(0.3,41,4.83,6.25))("Low","Medium Low","Medium High","High")  
Abierto a nuevas experiencias: =LOOKUP(F1,(0.4,31,5.38,6.45))("Low","Medium Low","Medium High","High")

### A.3. Transcripción manual

Las Figuras A.4 y A.5 muestra el documento de instrucciones que se proporcionó a las personas encargadas de la transcripción manual de los ensayos obtenidos para el conjunto de datos RxPI.

El documento de Instrucciones hace referencia a tres recursos que se describen en las subsecciones posteriores: *Etiquetas.txt*, *FAQs* y *Formulario digital*.

#### A.3.1. Etiquetas para anotadores

Un archivo de texto simple (*Etiquetas.txt*) estuvo disponible a todas las personas participantes en la recolección y anotación de RxPI, tanto personas anotadoras como responsables). Este archivo registraba las etiquetas utilizadas con la finalidad de unificar y reducir el número de descriptores dentro de las mismas. El archivo de texto se modificó constantemente al encontrarse con nuevas etiquetas o descriptores. La Tabla A.1 muestra el estado del archivo al final de la última ronda de transcripciones.

---

**Definición de etiquetas para fenómenos de escritura a mano.**

<D: descripción del gráfico>: Describe un gráfico o dibujo en el texto. Ver FAQ Q7.

<NS>: No Separó. Se escribe entre las palabras no separadas. Ver FAQ Q2.

<FO: palabra>: Falta Ortográfica. Se escribe adelante de la palabra mal escrita con la palabra corregida. Ver FAQ Q3.

<DL>: Palabra borrada. Ver FAQ Q9.

<IN: texto insertado>: Indica cuando un texto se inserta en el texto. Ver FAQ Q11.

<MD>: Adelante de la palabra cuando ha sufrido alguna modificación. Ver FAQ Q10.

<SB>: Adelante de la palabra cuando se ha separado por sílabas mediante un guión. Ver FAQ Q13.

---

**Precedencia de etiquetas.**

De mayor a menor precedencia: 1. FO; 2. SB; 3. NS; 4. MD

---

**Etiquetas usadas al momento:**

<D: emoji corazón>, <D: emoji cara triste>, <D: un rayón>,

<D: flecha a la derecha>, <D: emoji cara sonriente>, <D: firma>,

<D: viñeta>, <D: viñeta flecha>, <D: dos rayas>,

<D: caricatura hombre>, <D: caricatura mujer>, <D: línea asociativa>,

---

**Definición de etiquetas para ofuscar información personal.**

<NAME>: usar esta etiqueta cuando se escribió el nombre completo de una persona (uno o dos nombres mas uno o dos apellidos)

<EMAIL>: cuando se escribió una dirección de correo electrónico

<PHONE>: cuando se escribió un número de teléfono

<ADDRESS>: cuando se escribió una dirección existente, debe ser una dirección completa, si solo escribió el nombre de una ciudad, no necesita ofuscarse

---

Tabla A.1: Contenido del archivo *Etiquetas.txt*

**Figura A.4:** Documento que contiene instrucciones para las personas que realizaron la transcripción manual.

- Proyecto: "Transcripción de ensayos y compilación de pruebas de personalidad 2018"**
- El proyecto de "Transcripción de ensayos y compilación de pruebas de personalidad 2018" tiene como objetivo principal la digitalización de los test de personalidad de los sujetos que han decidido participar en el proyecto. Las actividades en las que estarás participando son tres:
1. Transcripción de ensayos
  2. Digitalización de ensayos (escaneo)
  3. Validación de transcripciones
- La primer actividad será la transcripción, misma para la cual se han formulado el siguiente conjunto de instrucciones y recomendaciones. En mensajes posteriores se te darán instrucciones precisas sobre las otras dos actividades restantes.
- Los documentos que se te han entregado son los test de personalidad de distintos sujetos. Cada documento está conformado de tres partes principales. Primero una sección de información personal del sujeto que está siendo evaluado. Segundo, una sección que corresponde al test de personalidad TIPI (Ten Item Personality Inventory). Y finalmente, una sección que corresponde a un ejercicio de escritura. El objetivo de esta primer actividad es contar con toda esta información en un formato electrónico, para lo cual se dan las siguientes recomendaciones.
- Recomendaciones para el proceso de transcripción del ejercicio de escritura:**
- Objetivo:** tener un archivo .txt con el texto que el sujeto escribió en su test de personalidad.
- La forma de **nombrar los archivos** es la siguiente: Para nombrar los documentos electrónicos deberás colocar tus iniciales (deben ser solo tres) seguidas de dos dígitos que refieren al año actual (18) un consecutivo de cuatro dígitos, por ejemplo, **GRR18\_0001.txt**. Si esta es la primera vez que capturas expedientes en este año, comienza con el consecutivo 0001. Este nombre deberás escribirlo en el expediente en papel (se sugiere hacerlo en la parte superior derecha de la primera página del expediente en papel) para que posteriormente sea fácil mantener la relación de expedientes electrónicos y expedientes físicos.
  - Una vez creado el archivo .txt deberás transcribir literalmente todo lo que el sujeto evaluado redactó.
  - Se recomienda usar un editor sencillo al momento de hacer la transcripción.
  - No debes de corregir errores ortográficos.
  - No debes agregar o quitar signos de puntuación.
  - Si el sujeto hizo una redacción ("más ordenada/estructurada") en más de un párrafo, esto debes de reflejarlo en tu transcripción, es decir poner un salto de línea (enter) donde termina un párrafo.
- Se deberán marcar los fenómenos de escritura con las 7 etiquetas descritas en el archivo *Ztriqueras.txt*.
  - Se deberán sustituir los datos confidenciales de cada sujeto de acuerdo a la descripción en el archivo *Ztriqueras.txt*.
  - **La transcripción debe ser una copia fiel del original, incluyendo errores ortográficos, símbolos de puntuación y organización del texto** (párrafos u oraciones aisladas)
  - Si utilizas un editor de textos tipo WORD, asegúrate de desactivar la "auto-corrección" **SE RECOMIENDA ALTAMENTE NO USAR WORD**, pues inserta símbolos "extraños" que hacen difícil el procesamiento automático posterior.
  - Una vez concluida tu transcripción, se recomienda hacer una revisión/comparación extra de lo escrito en papel con la transcripción generada en electrónico.
  - Para aclarar algunas dudas muy puntuales se compartirá un archivo en dropbox que explicará como atender diferentes fenómenos de escritura.
  - No olvides salvar el archivo!
- Recomendaciones para la construcción del expediente electrónico, información personal y TIPI:**
- Para vaciar la información personal de los sujetos evaluados, así como las respuestas que este proporcionó en el TIPI, deberás ingresar a la siguiente liga:  
<https://docs.google.com/forms/d/e/1FAIpQLSiKCPNCSL6GzxA8B0sU4HphKzXivImBNDRUSomusO/viewform>  
 Este link te llevara a un formulario de google donde deberás de colocar toda la información que el sujeto evaluado escribió en el papel.
- El primer filtro es verificar si el sujeto firmó el consentimiento de usar sus datos y si escribió algo en el ensayo. Si no firmó o no escribió algo, ese expediente físico no se digitalizará. Continúa con el siguiente.
- Después, la primera parte del formulario consiste en dar detalles de la información personal del sujeto, por ejemplo: nombre, año de nacimiento, género, carrera, etc. Es importante capturar toda esta información en el formulario y después dar click en "Next" (o Continuar).
- Una vez que avanzas a la siguiente hoja del formulario podrás ver la sección correspondiente al cuestionario del TIPI. En esta sección deberás, de igual manera, vaciar las respuestas que el sujeto evaluado indicó en el test en papel. Es conveniente mencionar que las preguntas puede que estén redactadas de manera distinta al instrumento en papel, sin embargo no debes preocuparte, pues el orden de las preguntas es el mismo, así como lo son también los valores en la escala de las posibles respuestas.
- Una vez concluido el vaciado de información en esta sección, deberás dar click en "Next". La siguiente sección del formulario consiste en dar detalles de la anotación que acaban de hacer. **Es muy importante que verifiquen que el identificador que está escrito en el**

**Figura A.5:** Documento que contiene instrucciones para las personas que realizaron la transcripción manual(Cont.)

expediente de papel corresponda al archivo de texto que guardaron con su transcripción y que sea el mismo que el que ingresarán en esta sección. Pues apartir de esta relación es que podremos saber quién escribió qué y dónde está su expediente.

Finalmente, hacer click en "Submit"

Ante cualquier duda usa el chat de whatsapp Proy: Corpus Personalidad.

### A.3.2. Preguntas frecuentes o FAQs

A continuación se listan las preguntas que surgieron durante la captura de expedientes electrónicos por las personas encargadas de la transcripción manual de los ensayos. Las respuestas fueron proporcionadas por las personas responsables de la recolección del corpus.

#### Sobre el ensayo escrito

**Q1.** *Si en el texto a transcribir hay mucho espacio entre palabras, ¿Cómo se captura esto en la transcripción electrónica?*

A1. Para mantener este fenómeno estilístico en la transcripción, se recomienda dejar los espacios de forma tácita en la transcripción, por ejemplo, dos espacios en lugar de solo uno.

**Q2.** *¿Qué hago si hay palabras juntas, sin espacio entre ellas?*

A2. Supongamos en el texto encontramos algo como esto: “nose”. La recomendación para transcribir esto en el formato electrónico será poner el marcador “<NS>” (NS significa “NO SEPARÓ”) entre las palabras que deberían ir separadas. Así entonces, en la transcripción “nose” se vería “no <NS>se”

**Q3.** *¿Qué hago si hay faltas de ortografía en las palabras?*

A3. Es necesario conservar las faltas ortográficas, indicando inmediatamente adelante de la palabra mal escrita la forma correcta de escribirla por medio del marcador “FO” (FO significa “FALTA ORTOGRÁFICA”). Así por ejemplo si se encuentra en el texto la palabra “cuerbo”, esta aparecerá en la transcripción como sigue: “cuerbo <FO:cuervo>”

**Q4.** *¿Qué hago si el signo de puntuación no esta tan marcado como todos los anteriores?*

A4. Debe conservarse el símbolo de puntuación en la transcripción, pues existe en efecto la intención de colocarlo.

**Q5.** *¿Qué hago con las i's que sí tienen el punto sobre la i, las que no lo tienen y las que van acentuadas? También hay i's donde el punto es claramente un punto, y otros que ponen una especie de “círculo” o “bolita” sobre la i ¿Se definirá un marcador para esto?*

A5. Hacer un etiquetado específico para cada forma distinta de puntuar las i's sería algo muy complejo. En su lugar se recomienda poner en las observaciones del formulario una nota haciendo referencia a este tipo de regularidades, por ejemplo: “El sujeto no utiliza puntos en las i's” o “El sujeto pone círculos sobre las i's”. En general si logras identificar alguna regularidad en la forma de escribir del usuario que creas vale la pena mencionar, puedes usar el campo de observaciones del formulario para expresarlo.

**Q6.** *¿Qué hago si en el texto aparecen dibujos (por ejemplo un corazón) o algunos emojis (por ejemplo caras sonriendo)?*

A6. Primero que nada hay que hacer una distinción entre emoji y emoticono. Un emoji corresponde a un dibujo hecho por el sujeto representando cualquier cosa, como los dibujos de *whatsapp*, mientras que el emoticono es cuando representas un dibujo con código ascii, por ejemplo cuando se quiere representar una cara sonriendo se utiliza un emoticono :) o :D. Entonces, cuando te encuentres en el texto un emoticono, este se debe conservar en la transcripción con su respectiva codificación ascii. Por otro lado, cuando te encuentras un emoji, este debe ser reemplazado por un código como se explica a continuación. Supongamos que en el texto encuentras el emoji de un corazón, o de un rostro sonriendo, si esto sucede, en la transcripción hay

## Apéndice A. Creación del conjunto de datos RxPI

---

que reemplazar el emoji por el siguiente marcador <D:'descripción del emoji'>, donde la 'descripción del emoji' puede ser algo como 'corazón' o 'cara sonriendo'. Asegúrate de usar siempre la misma descripción para los mismos emojis. Finalmente, cuando sea el caso de que hiciste alguna sustitución de un emoji por su respectiva descripción, asegúrate también de incluir en las observaciones del formulario en línea este hecho.

**Q7.** *¿Qué hago cuando un sujeto puso en su texto un “rayón” o una especie de firma al final de su texto o en cualquier otra parte del mismo?*

A7. Cuando esto suceda, debes colocar en la transcripción el marcador <D:'descripción del gráfico'>, donde en la 'descripción del gráfico' puedes colocar la palabra “rayas” o “firma” dependiendo del caso. Si llegarán existir dibujos u/o gráficos, puedes utilizar este mismo marcador (<D:desc>) para expresar el fenómeno en la transcripción. Igual que en el caso anterior, asegúrate de colocar este suceso en las observaciones del formulario en línea.

**Q8.** *¿Qué hago cuando el texto que se quiere transcribir tiene una letra poco legible?*

A8. Cuando esto suceda deberás intentar hacer la transcripción de lo que entiendas e indicar en las observaciones del formulario en línea el hecho de que el sujeto tiene “letra poco legible”.

**Q9.** *¿Qué hago cuando en el texto encuentro una palabra tachada (para indicar que se quiso borrar)?*

A9. Cuando esto suceda debes colocar en la transcripción el marcador <DL>por palabra tachada. Ejemplo: En el texto dice “La –casa– cabaña” donde la palabra casa es tachada en el texto, en la transcripción se deberá escribir “La <DL>cabaña”. DL significa *DELETED*.

**Q10.** *¿Qué hago cuando en el texto se encuentra una palabra que está modificada en alguna letra, por ejemplo el sujeto escribe “Univercidad” y sobre la “c” escribe una “s”?*

A10. Cuando esto suceda debes utilizar un marcador <MD>(MD significa Modificado) adelante de la palabra que sufrió una modificación. Ejemplo: “Univesidad <MD>”

**Q11.** *¿Qué hago cuando en el texto se encuentra algún tipo de inserción de otro texto? Este tipo de inserción se identifica cuando el que escribe quiere agregar algo aclaratorio, una nueva frase, o alguna palabra al texto que ya está escrito, pero no cabe, por lo que normalmente hace una marca y escribe el texto en una línea arriba o abajo de lo ya escrito, sin eliminar nada previo.*

A11. Cuando esto suceda, deberás usar una etiqueta con el texto insertado: <IN:texto insertado aquí>(IN significa Insertado). Ejemplo: “Anoche vi la televisión <IN:un programa que se llama el calabozo de ambrocio>y me gustó”

**Q13.** *¿Qué hago cuando una palabra se encuentra dividida en sílabas porque terminó el renglón (Syllabification)? Es decir, el sujeto usó un guion para separar sílabas en una palabra.*

A13. Cuando suceda esto se debe escribir la palabra completa y adelante de la palabra escribir la etiqueta <SB>. Ejemplo en el texto se cuenta “pala-bra” donde “pala-” aparece al final de un renglón y en el inicio de siguiente renglón aparece “bra” en la transcripción se debe escribir “palabra <SB>”.

**Q14.** *¿Qué hago si en un párrafo se dejan sangrías en las líneas que no son las primeras, o si no respetan el margen izquierdo?*



A14. Ignorar el espacio y escribir el fenómeno en las observaciones al final del formulario en línea. Por ejemplo, “No respeta el margen izquierdo”. Algo similar se deberá escribir en las observaciones si, por ejemplo escribe más allá de las líneas indicando los renglones, “No respeta el margen derecho”.

**Q15.** *¿Qué hago si a una palabra en el texto le aplican más de una etiqueta? Por ejemplo, una palabra sufrió una modificación, pero está mal escrita, entonces le corresponden las etiquetas <MD>y <FO:xxx>.*

A15. Cuando pase esto, se debe considerar la precedencia de las etiquetas. Esta precedencia se puede consultar en el archivo Etiquetas.txt. Se deberá conservar solo la etiqueta con mayor precedencia. En el caso del ejemplo, FO. Además se debe poner en las observaciones un texto como: “A la palabra XXXX en el párrafo 1 le aplicaban las etiquetas <FO:xxx>y <MD>”.

**Q16.** *¿Cómo trato al texto insertado, se pueden marcar los errores ortográficos, modificaciones, etc.?*

A16. Sí, el texto insertado se trata igual que el resto del texto, es decir, dentro de la inserción se pueden poner más etiquetas. Ejemplo el texto insertado es “ola quéhace?” se deberá escribir “<IN:ola <FO:hola>que <NS>hace?>”.

**Q17.** *¿Cómo ofusco la información personal del sujeto en la transcripción?*

A17. Usar las etiquetas definidas para cada tipo de información personal descrita en el archivo Etiquetas.txt. Por ejemplo, si el sujeto escribe “Mi abuelo Pedro Páramo que vive en Av. Vasco de Quiroga 4023 en la CDMX me dijo que marcara a su teléfono 5555555555”, la transcripción deberá quedar: “Mi abuelo <NAME>que vive en <ADDRESS>me dijo que marcara a su teléfono <PHONE>”. Luego indicar en las observaciones que se ofuscó información.

### **Sobre el cuestionario TIPI**

**Q12.** *¿Qué respuesta registro en el cuestionario si dos opciones están claramente marcadas?*

A12. Cuando esto suceda, debemos aprovechar las otras respuestas de las otras preguntas en el formulario para tratar de determinar cuál fue la respuesta correcta, por ejemplo: alguien muy “extrovertida” no puede ser al mismo tiempo muy “callada” . Si el texto escrito por el sujeto ayuda a definir la respuesta, se vale aprovecharlo. Es importante además, marcar en las observaciones del formulario esta situación, por ejemplo: “TIPI 1-marcó varias”, lo cual significa para nosotros que en la pregunta uno del cuestionario marco varias opciones el sujeto.

### **A.3.3. Formulario para transcriptores**

Para compilar toda la información recuperada con el instrumento de recolección, utilizamos un documento *Google Forms* donde se guía a los transcriptores a introducir toda la información correspondiente. Las Figuras A.6 a A.9 muestran la versión impresa del formulario; la versión en línea todavía puede accederse, pero ya no recibe respuestas y el hipervínculo está escrito en el documento de *Instrucciones*.

**Figura A.6:** Formulario usado por quienes realizaron la transcripción manual para digitalizar las respuestas del instrumento utilizado.

## Corpus de personalidad (2018) -transcription-

Antes de empezar, lee la guía/instrucciones para anotadores ([https://www.dropbox.com/2/revuew/Dudas\\_CorpusPersonalidad\\_2018/InstruccionesTranscripciones.docx](https://www.dropbox.com/2/revuew/Dudas_CorpusPersonalidad_2018/InstruccionesTranscripciones.docx))  
Luego, asigna un ID al documento (escríbelo en la parte superior derecha de la primera hoja). Finalmente, introduce la información correspondiente de los cuestionarios.

\* **Indicadores requeridos** *question*

1. ¿Firmó consentimiento? \*

Mark only one oval.

Sí Skip to question 3

No. ALTO. ¡No registres al sujeto!  
Skip to section 5 (Hemos terminado, cualquier duda, contáctame por el chat de WA.)

2. ¿Escribió el ensayo? \*

Mark only one oval.

Sí Skip to question 3

No. ALTO. ¡No registres al sujeto!  
Skip to section 5 (Hemos terminado, cualquier duda, contáctame por el chat de WA.)

Datos generales

3. Correo electrónico

4. Solo nombre \*

5. Solo apellidos \*

6. Fecha de nacimiento  
(si no introdujo fecha, dejar en blanco)

*Example: January 7, 2019*

7. Género \*

Mark only one oval.

Mujer

Hombre

Prefero no decir

8. Usuario en Twitter

Si tienes cuenta en Twitter, por favor colócala aquí; por ejemplo @usuario

9. Usuario de Facebook

Si tienes cuenta en Facebook, por favor colócala aquí

**Figura A.7:** Formulario usado por quienes realizaron la transcripción manual para digitalizar las respuestas del Instrumento utilizado (Cont.)

10. Licenciatura / Carrera \*  
 Por favor, indica aquí la carrera que te encuentras estudiando.  
 Mark only one oval.

Tecnologías y Sistemas de Información (LTS)  
 Diseño  
 Ciencias de la Comunicación  
 Administración  
 Derecho  
 Estudios Socioterritoriales  
 Humanidades  
 Biología Molecular  
 Ingeniería Biológica  
 Ingeniería en Computación  
 Matemáticas Aplicadas

11. Extraversión, entusiasta. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

12. Colérica, discutidora. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

13. Fiable, auto-disciplinada. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

14. Ansiosa, fácilmente alterable. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

15. Abierta a nuevas experiencias, polifacética. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

Me veo a mi mismo/a como a una persona:

Aquí encontrará una serie de rasgos de personalidad que pueden o no referirse a usted. Por favor, elija un número para cada expresión para indicar en qué medida está de acuerdo o en desacuerdo con ella. Debería valorar el grado en qué cada par de rasgos corresponde a su persona, aunque una pueda corresponder en más grado que otra.

11. Extraversión, entusiasta. \*  
 Mark only one oval.

1 2 3 4 5 6 7  
 Totalmente de acuerdo

Figura A.8: Formulario usado por quienes realizaron la transcripción manual para digitalizar las respuestas del Instrumento utilizado (Cont.)

16. Reservada, callada. \*  
*Mark only one oval.*

1 2 3 4 5 6 7

Toté        Totalmente de acuerdo

17. Comprensiva, amable. \*  
*Mark only one oval.*

1 2 3 4 5 6 7

Toté        Totalmente de acuerdo

18. Desorganizada, descuidada. \*  
*Mark only one oval.*

1 2 3 4 5 6 7

Toté        Totalmente de acuerdo

19. Serena, emocionalmente estable. \*  
*Mark only one oval.*

1 2 3 4 5 6 7

Toté        Totalmente de acuerdo

20. Tradicional, poco imaginativa. \*  
*Mark only one oval.*

1 2 3 4 5 6 7

Toté        Totalmente de acuerdo

Información sobre la anotación

21. Identificador del archivo \*  
 Iniciales del etiquetador más guión bajo más cuatro dígitos (GRR18\_0001)

---

22. Iniciales del anotador \*  
*Mark only one oval.*

BRR18  
 DSC18  
 IVT18  
 GRR18

**Figura A.9:** Formulario usado por quienes realizaron la transcripción manual para digitalizar las respuestas del Instrumento utilizado (Cont.)

23. Observaciones sobre este sujeto  
Detalles que se observaron y que llamaron la atención del etiquetador. Pueden ser situaciones que no se había atendido antes. Una observación importante es indicar si el texto contiene información privada del sujeto como: nombres completos de personas no públicas, direcciones exactas ("vivo en calle no etc."), correos electrónicos, número de teléfonos.

---

---

---

---

---

Hemos terminado, cualquier duda, contáctame por el chat de WA.

This content is neither created nor endorsed by Google.

Google Forms





- Adali, S. and Golbeck, J. (2012). Predicting personality with social behavior. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 302–309.
- Aitchison, J. (2003). *Words in the Mind*. Wiley.
- Allen, H. A., Liddle, P. F., and Frith, C. D. (1993). Negative features, retrieval processes and verbal fluency in schizophrenia. *The British Journal of Psychiatry*, 163(6):769–775.
- Allport, D. and Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 295(1077):397–410.
- Andreou, G. and Trott, K. (2013). Verbal fluency in adults diagnosed with attention-deficit hyperactivity disorder (adhd) in childhood. *ADHD Attention Deficit and Hyperactivity Disorders*, 5(4):343–351.
- Argamon, S., Dhawle, S., Koppel, M., and Pennebaker, J. W. (2005). Lexical predictors of personality type. In *In Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America*.
- Ashton, M. C. and Lee, K. (2005). A defence of the lexical approach to the study of personality structure. *European Journal of Personality*, 19(1):5–24.
- Association for Computational Linguistics (2020). ACL code of ethics.
- Association for Computing Machinery (2018). ACM code of ethics and professional conduct.
- Atkinson, R. (2007). The life story interview as a bridge in narrative inquiry. *Handbook of narrative inquiry: Mapping a methodology*, pages 224–245.
- Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., and Sebe, N. (2011). Please, tell me about yourself: Automatic personality assessment using short self-presentations. In *Proceedings of the 13th*

## Bibliografía

---

- International Conference on Multimodal Interfaces, ICMI '11*, pages 255–262, New York, NY, USA. ACM.
- Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.
- Benton, A., Coppersmith, G., and Dredze, M. (2017). Ethical research protocols for social media health research. In Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M., and Wallach, H., editors, *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 94–102, Valencia, Spain. Association for Computational Linguistics.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.
- Boyd, R. L. and Pennebaker, J. W. (2015). A way with words: Using language for psychological science in the modern era. In *Consumer psychology in a social media world*, pages 222–236. Routledge.
- Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. University of California Press.
- Callealta Barroso, F. J. and Gallego Gallego, D. J. (2016). Medidas de disponibilidad léxica: comparabilidad y normalización (measures of lexical availability: comparability and standardization). *Boletín de filología*, 51(1):39–92.
- Cardona, J. E., Grisales-Cardenas, J. S., Trujillo-Llano, C., Diazgranados, J. A., Urquina, H. E., Cardona, S., Torres, A., Torres, L. A., Gonzalez, L. M., Jaramillo, T., Cediell, J., Oñate-Cadena, N., Mateus-Ferro, G., and Marmolejo-Ramos, F. (2021). Semantic Memory and Lexical Availability in Parkinson’s Disease: A Statistical Learning Study. *Frontiers in Aging Neuroscience*, 13(July):1–11.
- Carreras, X., Chao, I., Padró, L., and Padró, M. (2004). Freeling: An open-source suite of language analyzers. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC’04)*.
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.
- Celli, F., Pianesi, E., Stillwell, D., and Kosinski, M. (2013). Workshop on computational personality recognition: Shared task. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, pages 2–5.



- Cerhan, J. H., Ivnik, R. J., Smith, G. E., Tangalos, E. C., Petersen, R. C., and Boeve, B. F. (2002). Diagnostic utility of letter fluency, category fluency, and fluency difference scores in alzheimer's disease. *The Clinical Neuropsychologist*, 16(1):35–42.
- Chancellor, S. and De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *npj Digital Medicine*, 3(1). Publisher Copyright: © 2020, The Author(s).
- Chen, L., Jiang, L., and Li, C. (2021). Using modified term frequency to improve term weighting for text classification. *Engineering Applications of Artificial Intelligence*, 101:104215.
- Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1):204–256.
- Costa Jr, P. T. and McCrae, R. R. (2008). *The Revised Neo Personality Inventory (neo-pi-r)*. Sage Publications, Inc.
- Cruz-Sánchez, K., Reyes-Meza, V., Hudson, R. E., Martínez-Gómez, M., and Amando, B.-O. (2017). Evaluación de la personalidad en adolescentes con distinto orden de nacimiento y número de hermanos. In Camacho, J., Cabrera, F., Zamora Arevalo, O., Martinez, H., and Irigoyen Morales, J., editors, *Aproximaciones al estudio del comportamiento y sus aplicaciones*, volume 1, pages 281–297. Universidad Autónoma de Tlaxaca.
- Darlansyah, A., Naeem, M. A., Mirza, F., and Pears, R. (2019). SENTIPEDE: A smart system for sentiment-based personality detection from short texts. *Journal of Universal Computer Science*, 25:1323–1352.
- De Deyne, S. and Storms, G. (2015). Word Associations. In Taylor, J. R., editor, *temp*. Oxford University Press.
- DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., and Morency, L.-P. (2014). Simsensei kiosk: a virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14*, page 1061–1068, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

## Bibliografía

---

- Deyne, S. D., Navarro, D. J., and Storms, G. (2003). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, 45:480–498.
- Dick, J., Fredrick, J., Man, G., Huber, J. E., and Lee, J. (2018). Sentence production in parkinson's disease. *Clinical Linguistics & Phonetics*, 32(9):804–822.
- Dimitrijević, N. R. (1981). A comparative study of the lexical availability of monolingual and bilingual schoolchildren. In *Studia Anglica Posnaniensia*, volume 13, pages 109–130. Adam Mickiewicz University.
- Fox, E. A. and Shaw, J. A. (1994). Combination of multiple searches. *NIST special publication SP*, 243.
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52(1):197–221.
- Galton F. (1884). Measurement of character. *Fortnightly Review*, 36(212):179–185.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., and Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12):86–92.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34.
- Gosling, S. D., Rentfrow, P. J., and Jr., W. B. S. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504 – 528.
- Gougenheim, G. (1954). Le français Élémentaire. *The French Review*, 27(3):217–220.
- Gougenheim, G., Michéa, R., Rivenc, P., and Sauvageot, A. (1964). *L'élaboration du français fondamental (1er degré): étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, volume 1. Chilton Books.
- Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D., Rizzo, S., and Morency, L.-P. (2014). The distress analysis interview corpus of human and computer interviews. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3123–3128, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Gremillion, M. L. and Martel, M. M. (2014). Merely misunderstood? receptive, expressive, and pragmatic language in young children with disruptive behavior disorders. *Journal of Clinical Child & Adolescent Psychology*, 43(5):765–776.
- Grupo de Investigación DispoLex (2003–2023). ¿qué es el proyecto panhispánico? Last accessed 2 December 2023.

- Harrigian, K., Aguirre, C., and Dredze, M. (2021). On the state of social media data for mental health research. In Goharian, N., Resnik, P., Yates, A., Ireland, M., Niederhoffer, K., and Resnik, R., editors, *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 15–24, Online. Association for Computational Linguistics.
- Hernández Muñoz, N. (2010). Social aspects of oral and lexical written production in spanish. *SKY Journal of Linguistics*, 23:101–123.
- Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., Barnes, P., and Mitchell, M. (2021). Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 560–575, New York, NY, USA. Association for Computing Machinery.
- Iacobelli, F. and Culotta, A. (2013). Too neurotic, not too friendly: structured personality classification on textual data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, pages 19–22.
- Iacobelli, F., Gill, A. J., Nowson, S., and Oberlander, J. (2011). Large scale personality classification of bloggers. In D’Mello, S., Graesser, A., Schuller, B., and Martin, J.-C., editors, *Affective Computing and Intelligent Interaction*, pages 568–577, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Jiang, H., Zhang, X., and Choi, J. D. (2020). Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(10):13821–13822.
- Jo, E. S. and Gebru, T. (2020). Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT\* '20, page 306–316, New York, NY, USA. Association for Computing Machinery.
- Jourdy, R. and Petot, J.-M. (2017). Relationships between personality traits and depression in the light of the “big five” and their different facets. *L’Évolution Psychiatrique*, 82(4):e27–e37.
- Kalimeri, K. (2013). Towards a dynamic view of personality: multimodal classification of personality states in everyday situations. In Epps, J., Chen, F., Oviatt, S., Mase, K., Sears, A., Jokinen, K., and Schuller, B., editors, *ICMI*, pages 325–328. ACM.
- Kampman, O., J. Barezi, E., Bertero, D., and Fung, P. (2018). Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In *Proceedings of the 56th*

## Bibliografía

---

- Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 606–611, Melbourne, Australia. Association for Computational Linguistics.
- Keirsey, D. (1998). *Please understand me II*. Del Mar, CA: Prometheus Nemesis Book Company.
- Kerz, E., Qiao, Y., Zanwar, S., and Wiechmann, D. (2022). Pushing on personality detection from verbal behavior: A transformer meets text contours of psycholinguistic features. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 182–194, Dublin, Ireland. Association for Computational Linguistics.
- Klein, D. N., Kotov, R., and Bufferd, S. J. (2011). Personality and depression: Explanatory models and review of the evidence. *Annual Review of Clinical Psychology*, 7(1):269–295. PMID: 21166535.
- Lázaro, J. and Rubino, M. (2020). Disponibilidad léxica en informantes culiacanenses: homosexualidad como centro de interés. *Lingüística Mexicana*, 2(2):29–48.
- Leonardi, S., Monti, D., Rizzo, G., and Morisio, M. (2020). Multilingual Transformer-Based Personality Traits Estimation. *Information 2020*, Vol. 11, Page 179, 11(4):179.
- López Morales, H. (2014). *Lexical Availability Studies*, pages 1–11. Springer Netherlands, Dordrecht.
- Luz, S., Haider, F., de la Fuente, S., Fromm, D., and MacWhinney, B. (2020). Alzheimer's dementia recognition through spontaneous speech: The ADReSS Challenge. In *Proceedings of INTERSPEECH 2020*, Shanghai, China.
- Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research (JAIR)*, pages 457–500.
- Majumder, N., Poria, S., Gelbukh, A., and Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79.
- Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.
- Matthews, G., Deary, I., and Whiteman, M. (2009). *Personality Traits*. Personality Traits. Cambridge University Press.
- McCrae, R. R. and Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American psychologist*, 52(5):509.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space.

- Mischel, W. (1993). *Introduction to Personality*. Harcourt Brace Jovanovich College Publishers.
- Myers, I. and Myers, P. (1995). *Gifts Differing: Understanding Personality Type*. Mobius.
- Oberlander, J. and Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes*, 42(3):239–270.
- Oberlander, J. and Nowson, S. (2006). Whose thumb is it anyway? classifying author personality from weblog text. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 627–634, Sydney, Australia. Association for Computational Linguistics.
- Ortigosa, A., Carro, R. M., and Quiroga, J. I. (2014). Predicting user personality by mining social interactions in facebook. *J. Comput. Syst. Sci.*, 80(1):57–71.
- Ozer, D. J. and Martínez, V. B. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, 57(1):401–421.
- Pennebaker, J. and King, L. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6):1296–1312.
- Pennebaker, J. W. (2011). *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury Press, New York, 1st edition.
- Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). *Linguistic Inquiry and Word Count*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Priestley, M., O'donnell, F., and Simperl, E. (2023). A survey of data quality requirements that matter in ml development pipelines. *J. Data and Information Quality*, 15(2).
- Pérez Durán, M. (2020). Análisis del léxico disponible del centro de interés del insulto en estudiantes de secundaria de San Luis Potosí, México. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 46(1):261–278.
- Ramírez, G., Villatoro, E., Ionescu, B., Escalante, H. J., Escalera, S., Larson, M., Müller, H., and Guyon, I. (2019). Overview of the multimedia information processing for personality & social networks analysis contest. In Zhang, Z., Suter, D., Tian, Y., Branzan Albu, A., Sidère, N., and Jair Escalante, H., editors, *Pattern Recognition and Information Forensics*, pages 127–139, Cham. Springer International Publishing.
- Ramírez-de-la Rosa, G., Villatoro-Tello, E., Jiménez-Salazar, H., Pinto, D., Singh, V. K., Villavicencio, A., Mayr-Schlegel, P., and Stamatatos, E. (2018). TxPI-u: A resource for personality identification of undergraduates. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 34(5):2991–3001.

## Bibliografía

---

- Ramírez-de-la-Rosa, G., Jiménez-Salazar, H., Villatoro-Tello, E., Reyes-Meza, V., and Rojas-Avila, J. (2023). A lexical–availability–based framework from short communications for automatic personality identification. *Cognitive Systems Research*.
- Renau, V., Oberst, U., Gosling, S., Rusiñol, J., and Chamarro, A. (2013). Translation and validation of the ten-item-personality inventory into spanish and catalan. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport*, 31(2).
- Rensbergen, B., Storms, G., and De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic bulletin & review*, 22(6):1717–1724.
- Ringeval, F., Schuller, B., Valstar, M., Cummins, N., Cowie, R., Tavabi, L., Schmitt, M., Alisamir, S., Amiriparian, S., Messner, E.-M., Song, S., Liu, S., Zhao, Z., Mallol-Ragolta, A., Ren, Z., Soleymani, M., and Pantic, M. (2019). Avec 2019 workshop and challenge: State-of-mind, detecting depression with ai, and cross-cultural affect recognition. In *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop, AVEC '19*, page 3–12, New York, NY, USA. Association for Computing Machinery.
- Rojas, D., Zambrano, C., and Salcedo, P. (2016). a Proposal for Predicting Academic Achievement Through a Vectorial Model of Lexical Availability. *ICERI2016 Proceedings*, 1:2799–2806.
- Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, USA, 3rd edition.
- Ríos González, G. (2008). Diferencias léxicas entre el hombre y la mujer en tres centros de interés: Saludos, temas de conversación y despedidas. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 33(1):151–166.
- Salcedo, P., del Valle, M., Quintanilla, A., and Zambrano, C. (2016). Lexical Availability of Students in Mathematics Education in Southern Chile. *ICERI2016 Proceedings*, 1(December):3676–3683.
- Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.*, 24(5):513–523.
- Stella, M. (2020). Multiplex networks quantify robustness of the mental lexicon to catastrophic concept failures, aphasic degradation and ageing. *Physica A: Statistical Mechanics and its Applications*, 554:124382.
- Sun, X., Liu, B., Cao, J., Luo, J., and Shen, X. (2018). Who am i? personality detection based on deep learning for texts. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6.

- Tackman, A., Sbarra, D., Carey, A., Donnellan, M., Horn, A., Holtzman, N., Edwards, T., Pennebaker, J., and Mehl, M. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology*, 116(5):817–834. Publisher Copyright: © 2018 American Psychological Association.
- Tamariz-Martel Mirêlis, M. (2005). *Exploring the Adaptive Structure of the Mental Lexicon*. PhD thesis, The University of Edinburgh.
- UNESCO (2021). Recomendación sobre la ética de la inteligencia artificial. Adopted on 23 November 2021.
- Villatoro-Tello, E., Dubagunta, S. P., Fritsch, J., Ramírez-de-la-Rosa, G., Motlicek, P., and Magimai-Doss, M. (2021a). Late Fusion of the Available Lexicon and Raw Waveform-Based Acoustic Modeling for Depression and Dementia Recognition. In *Proc. Interspeech 2021*, pages 1927–1931.
- Villatoro-Tello, E., Ramírez-de-la-Rosa, G., Gática-Pérez, D., Magimai.-Doss, M., and Jiménez-Salazar, H. (2021b). *Approximating the Mental Lexicon from Clinical Interviews as a Support Tool for Depression Detection*, pages 557–566. Association for Computing Machinery, New York, NY, USA.
- Vrindha, M. K., Geethu, V., Anurenjan, P. R., Deepak, S., and Sreeni, K. G. (2023). A review of alzheimer’s disease detection from spontaneous speech and text. In *2023 International Conference on Control, Communication and Computing (ICCC)*, pages 1–5.
- Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., Zhong, X., and Sun, J. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, 48(11):4232–4246.