

Panorama de las Tecnologías del Lenguaje en México

Luis Villaseñor Pineda
Editor



ACADEMIA MEXICANA DE COMPUTACIÓN, A, C.

Panorama de las Tecnologías del Lenguaje en México

Editor: Luis Villaseñor Pineda.

En colaboración con la Academia Mexicana de Computación:

Coordinador: Luis Enrique Sucar Succar.

Colaboradores capítulo 1: Héctor Jiménez, Luis A. Pineda Cortés, Luis Villaseñor Pineda.

Colaboradores capítulo 2: Hiram Calvo, Luis A. Pineda Cortés, Luis Villaseñor Pineda.

Colaboradores capítulo 3: Hiram Calvo, Noé A. Castro Sánchez, Alexander Gelbukh, Yasmín Hernández Pérez, Héctor Jiménez, Manuel Montes y Gómez, Luis A. Pineda Cortés, David Pinto Avendaño, Fernando Sánchez Vega, Grigori Sidorov, Luis Villaseñor Pineda.

Colaboradores capítulo 4: Gabriela Ramírez de la Rosa, Esaú Villatoro Tello.

Colaboradores capítulo 5: Hiram Calvo, Ángel Hernández Castañeda, Javier Sánchez Junquera.

Colaboradores capítulo 6: Juan Martínez Miranda, Humberto Pérez Espinosa.

Colaboradores capítulo 7: Héctor Jiménez, Luis Villaseñor Pineda.

Segunda edición: 2019

Academia Mexicana de Computación, A. C.

Todos los derechos reservados conforme a la ley.

ISBN:978-607-97357-4-6

Corrección de estilo: Luis Enrique Sucar Succar.

Diseño de portada: Mario Alberto Vélez Sánchez.

Cuidado de la edición: Luis Enrique Sucar Succar.

Este libro se realizó con el apoyo del CONACyT, Proyecto I1200/28/2019. Queda prohibida la reproducción parcial o total, directa o indirecta, del contenido de esta obra, sin contar con autorización escrita de los autores, en términos de la Ley Federal del Derecho de Autor y, en su caso, de los tratados internacionales aplicables.

Impreso en México.

Panorama de las Tecnologías del Lenguaje en México

Colaboradores

(en orden alfabético)

Hiram Calvo

Noé A. Castro Sánchez

Alexander Gelbukh

Ángel Hernández Castañeda

Yasmín Hernández Pérez

Héctor Jiménez Salazar

Juan Martínez Miranda

Manuel Montes y Gómez

Humberto Pérez Espinosa

Luis Alberto Pineda Cortés

David Pinto Avendaño

Gabriela Ramírez de la Rosa

Juan J. Sánchez Junquera

Fernando Sánchez Vega

Grigori Sidorov

Luis Villaseñor Pineda

Esau Villatoro Tello

Agradecimientos

Agradecemos a la Academia Mexicana de Computación y a la Asociación Mexicana de Procesamiento de Lenguaje Natural, en especial a todos aquellos miembros de la comunidad que contribuyeron con el desarrollo de este libro. En especial, agradecemos a Luis Enrique Sucar Succar por coordinar este valioso esfuerzo.

Agradecemos a la Coordinación de Ciencias de la Computación del INAOE por las facilidades otorgadas para la elaboración de este libro.

Finalmente, también agradecemos al CONACyT por apoyar la creación de esta obra.

Prólogo

El objetivo principal de las Tecnologías del Lenguaje (TL) es permitir que una computadora comprenda y produzca información en una o varias lenguas humanas. Si recordamos que casi toda actividad humana implica la transmisión de información, ya sea a través de un medio escrito u oral, el potencial de las TL es muy amplio. De hecho, ellas están presentes desde hace años en infinidad de actividades que cotidianamente realizamos con una computadora. Ellas están en el verificador de ortografía o corrector gramatical dentro de nuestro procesador de textos, en el sistema de dictado de nuestro teléfono celular, así como en el imprescindible motor de búsqueda en Internet. Ahora bien, gracias al gran avance tecnológico, el abanico de oportunidades de las TL ha crecido considerablemente. Las TL permiten a las personas colaborar más fácilmente entre ellas, hacer negocios, compartir conocimientos, participar en la vida política, utilizar dispositivos usando la voz, etc., sin importar sus habilidades para manejar una computadora.

El desarrollar las TL abrirá posibilidades para la creación de innovadores productos y servicios en diferentes sectores como comercio, administración, educación, salud, entretenimiento o turismo, entre otros. Aunado a esto, el tamaño del mercado es enorme; y no sólo porque México es el país hispanohablante más poblado, sino que el español es una de las lenguas más habladas en el mundo.

El tratamiento del idioma español por medios automáticos es claramente una necesidad para nuestra población, y trabajar en ello, no sólo se justifica por el enorme mercado sino porque una lengua es un reflejo social y cultural. Bajo este escenario, cada vez cobra más importancia proveer los elementos necesarios para impulsar las Tecnologías del Lenguaje dentro del ámbito nacional.

El presente libro se orienta a estudiantes o investigadores afines a las ciencias computacionales que desean descubrir las oportunidades y retos de las TL. Para ello, este libro presenta un panorama general de lo que son las TL, sus conceptos básicos, y discute la importancia de su desarrollo en el contexto de nuestro país. Además, encamina al lector en el área al explicar varias de las investigaciones realizadas por grupos de investigación en México.

Luis Villaseñor Pineda
Tonatzintla,
México, 2019

Abreviaturas

ACL	<i>Association for Computational Linguistics</i>
AMEXCOMP	Academia Mexicana de Computación
AMPLN	Asociación Mexicana de Procesamiento de Lenguaje Natural
BoW	<i>Bag of Words</i>
BUAP	Benemérita Universidad Autónoma de Puebla
CENIDET	Centro Nacional de Investigación y Desarrollo Tecnológico
CIC	Centro de Investigación en Computación
CICESE	Centro de Investigación Científica y de Educación Superior de Ensenada
COLING	<i>International Conference on Computational Linguistics</i>
CONACYT	Consejo Nacional de Ciencia y Tecnología
EMPLN	<i>Empirical Methods in Natural Language Processing</i>
IA	Inteligencia Artificial
IIMAS	Instituto de Investigaciones en Matemáticas y Sistemas
INAOE	Instituto Nacional de Astrofísica, Óptica y Electrónica
INEEL	Instituto Nacional de Electricidad y Energías Limpias
IPN	Instituto Politécnico Nacional
LIWC	<i>Linguistic Inquiry and Word Count</i>
LDA	<i>Latent Dirichlet Allocation</i>
MTD	Matriz término-documento
NB	<i>Naïve Bayes</i>
PLN	Procesamiento de Lenguaje Natural
RedTTL	Red Temática en Tecnologías del Lenguaje
SEPLN	Sociedad Española de Procesamiento de Lenguaje Natural
SMV	<i>Support Vector Machine</i>
TL	Tecnologías del Lenguaje
UAEMex	Universidad Autónoma del Estado de México
UAM	Universidad Autónoma Metropolitana
UNAM	Universidad Nacional Autónoma de México

Índice general

Índice general	XI
1. Introducción	I
2. Conceptos básicos	7
2.1. Lingüística Computacional	7
2.2. Modelos computacionales	9
2.2.1. Nivel fonético y fonológico	9
2.2.2. Nivel de prosodia y entonación	11
2.2.3. Nivel léxico y morfológico	11
2.2.4. Nivel sintáctico	12
2.2.4.1. Enfoque de constituyentes	14
2.2.4.2. Enfoque de dependencias	15
2.2.5. Nivel semántico	16
2.2.6. Nivel pragmático	16
2.3. La ambigüedad	20
3. La Investigación en México en TL	23
3.1. Lingüística de corpus	24

3.1.1.	Modelos del lenguaje	25
3.1.2.	Corrección ortográfica	27
3.2.	Reconocimiento de voz	27
3.3.	Sistemas conversacionales en lenguaje natural	29
3.4.	Procesamiento de textos	31
3.4.1.	Análisis de opinión	31
3.4.2.	Detección de ironía	32
3.4.3.	Semejanza entre palabras y diccionarios de ideas afines	33
3.4.4.	Desambiguación del sentido de las palabras	34
3.4.5.	Detección de engaño	35
3.4.6.	Atribución de autoría	35
3.4.7.	Reconocimiento de paráfrasis e implicación textual	36
Referencias	38
4.	Identificación de personalidad	45
4.1.	Introducción	45
4.1.1.	Perfilado de autores	47
4.2.	Definición del problema	48
4.2.1.	El modelo de personalidad <i>Big Five</i>	50
4.3.	La personalidad de estudiantes mexicanos	52
4.3.1.	Metodología para la construcción del corpus TxPI-u	53
4.3.2.	Descripción de los datos	55
4.3.3.	Partición estratificada	57
4.4.	Desempeño de sistemas automáticos	61
4.4.1.	Métricas de evaluación	62
4.4.2.	Configuración experimental	62
4.4.3.	Resultados	63

4.5.	Conclusiones y panorama futuro	65
	Referencias	69
5.	Detección de engaño en texto	75
5.1.	Introducción	75
5.2.	Reconocimiento del engaño	78
5.3.	Enfoques de identificación	83
5.4.	Conjuntos homogéneos	86
5.5.	Detección intercultural del engaño	89
5.6.	Aprendizaje de multitudes	92
5.7.	Modelos probabilísticos	96
5.7.1.	Clasificación de dominios mezclados	96
5.7.2.	Clasificación de dominio cruzado	98
5.7.3.	Palabras y tópicos predominantes	100
	Referencias	101
6.	Sistemas Interactivos Emocionales	107
6.1.	Introducción	108
6.2.	Sistemas de interacción afectivos	111
6.2.1.	Detección de emociones en voz	114
6.2.2.	Modelado de emociones en sistemas artificiales	117
6.3.	Aplicaciones de sistemas afectivos	123
6.3.1.	Creación de bases de datos de voz emocional	123
6.3.2.	Análisis paralingüístico en niños	125
6.3.3.	Estimación de la calidad de interacción de adultos ma- yores	127
6.3.4.	Agentes virtuales conversacionales en salud mental	130

6.4. Conclusiones	135
Referencias	137
7. Retos y Perspectivas	147

Capítulo I

Introducción

Héctor Jiménez Salazar, UAM-Cuajimalpa.
Luis Alberto Pineda Cortés, IIMAS/UNAM.
Luis Villaseñor Pineda, INAOE.

Las Tecnologías del Lenguaje Humano (TLH ó TL) nacen de la necesidad de lograr que la gente pueda comunicarse con las máquinas utilizando una comunicación natural, basada en el lenguaje humano. Crear un sistema automático capaz de interactuar con una persona, de tal forma que parezca natural, es una idea perseguida desde los inicios de la Computación. La primera propuesta en el entorno científico y académico para construir y programar una máquina capaz de conversar en lenguaje natural con seres humanos aparece en el artículo *Computing Machinery and Intelligence* (Turing 1950), en el que Alan Turing propuso el programa de investigación para la Inteligencia Artificial (IA). Es también en este artículo donde se presenta el “Juego de imitación”, mejor conocido como la “Prueba de Turing”, donde se propone que, si una máquina es capaz de conversar de manera natural con un ser humano en inglés o español, por ejemplo, se tendrá que considerar “inteligente”.

La búsqueda de soluciones al problema del entendimiento automático del lenguaje humano ha sufrido un accidentado camino. Este problema, mucho más complejo de lo que originalmente se pensó, enfrenta un amplio conjunto de retos, el cual sólo puede desafiarse desde una perspectiva multidisciplinaria. Crear un sistema que interactúe debidamente con una persona requiere del entendimiento de la estructura simbólica del lenguaje, así como de los fenómenos acústicos que lo acompañan; además, de dominar las estrategias usadas por las personas para alcanzar una comunicación efectiva. Todo esto bajo una arquitectura computacional lo suficientemente robusta para combinar los muy variados estímulos propios de la comunicación humana y proveer una respuesta natural, en tiempo real, basada en la comprensión del mensaje recibido. De ahí, que esta área aglutine los esfuerzos de expertos en lingüística, psicología, estadística, ingeniería y ciencias computacionales.

Lo que es más, dada la complejidad de la empresa, se recurrió a proponer soluciones a subáreas específicas. Esta fragmentación, en ocasiones, obedeciendo más a razones pragmáticas que teóricas, ha permitido importantes logros en el área. En particular, dentro de las ciencias computacionales, los esfuerzos se han dividido para incidir en tres grandes campos de aplicación: i) la gestión de información, ii) las interfaces hombre-máquina; y iii) la traducción automática. Es así como se han creado y desarrollado diferentes teorías y marcos metodológicos, herramientas y recursos lingüísticos, todos ellos orientados al tratamiento automático del lenguaje. Las Tecnologías del Lenguaje (TL) reúnen todos estos elementos propuestos desde los diferentes frentes; sin olvidar que la motivación final es lograr que una máquina comprenda o manipule el lenguaje natural para alcanzar una fácil comunicación con un ser humano.

De manera general, a continuación se describen los tres grandes campos de aplicación donde las TL inciden: *la gestión de información, la traducción automática y las interfaces humano-computadora.*

La *gestión de información* cubre todo aquello que implica almacenar, recuperar, organizar y extraer información de grandes colecciones de documentos. Muchas técnicas se han propuesto para ello, y a pesar de ser parciales, han demostrado su utilidad incidiendo en nuestras actividades cotidianas. El ejemplo obvio es el buscador de Google. A través de él, se puede conseguir información casi de cualquier cosa en un par de segundos. Ahora bien, las TL no se limitan al almacenamiento y recuperación de grandes colecciones de documentos, intentan ir más allá al organizar y extraer información contenida en dichas colecciones. Una colección puede organizarse o categorizarse desde diferentes puntos de vista con distintas utilidades. Por ejemplo, podríamos desear recuperar notas periodísticas cuyo tema principal es la política durante el sexenio pasado, discriminándolas del resto; o podríamos, desear clasificar las notas en función de si ésta menciona cierta empresa o renombrado personaje. Lo que es más, las TL pueden ayudarnos a observar aspectos subjetivos de la información, por ejemplo, al determinar la aceptación o rechazo de un cierto producto, evento o personaje, al distinguir las opiniones positivas o negativas de miles de mensajes circulando en las redes sociales.

La segunda gran área de aplicación es la *traducción automática*. No es sorprendente que los últimos estudios comprueben que los usuarios de Internet prefieren acceder a ella usando su lengua materna, y lo que es más, están de acuerdo en que los sitios realizados en su región deban presentarse en las len-

guas oficiales de dicha región¹. Este mismo estudio, realizado en la comunidad europea, muestra que el 90 % de los usuarios de Internet prefieren realizar compras electrónicas usando su lengua. Cabe resaltar que esto ocurre en Europa, una región de gran interacción lingüística entre sus ciudadanos.

Ahora bien, el escenario lingüístico en nuestro país difiere de la realidad europea. Por un lado, el español mexicano es la lengua más importante a pesar de su variedad de acentos y dialectos entre regiones y estados. El dominio de otras lenguas en la población mexicana es limitado. Según la consulta ² en 2013 el 13 % de la población mexicana dicen hablar otro idioma además del español, y más de 9 de cada 10 adultos que dicen hablar otro idioma hablan inglés (aproximadamente 9.5 millones de mexicanos mayores de 18 años). Si consideramos que el inglés sigue siendo el idioma dominante en la Internet, la traducción automática es una necesidad real para una gran parte de nuestra población. El proveer herramientas automáticas de traducción ampliará las posibilidades de cualquier usuario de la red. El fin último es que cualquier comunicación escrita –u oral– no esté condicionada ni limitada por la barrera del idioma.

Por otro lado, en México existe una gran diversidad de lenguas indígenas la cual fue reconocida apenas en el 2001, a través de la Ley de los Derechos Lingüísticos. Este tardío reconocimiento aunado a la marginación social en que viven la mayoría de los cerca de 6 millones y medio de hablantes de alguna lengua indíge-

¹Flash Eurobarometer 313. (2011). User language preference online. European Commission. The Gallup Organization. http://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_313_en.pdf. Consultado 20/04/2018.

²Campos, R., & Hernández, A. M. (enero, 2013). Mexicanos y los idiomas extranjeros. México: Consulta Mitofsky. <http://consulta.mx/index.php/estudios-e-investigaciones/mexico-opina/item/823-mexicanos-y-los-idomas-extranjeros>. Consultado 04/05/2018.

na, los ha alejado de las ventajas de esta sociedad de la información. Paradójicamente, las Tecnologías del Lenguaje pueden brindar mecanismos innovadores para favorecer a estos grupos relegados. Por ejemplo, reconocer automáticamente la lengua de un hablante monolingüe sería de gran utilidad en emergencias médicas o jurídicas.

La tercera gran área de aplicación son las *interfaces humano-computadora*. Bajo este rubro caen las aplicaciones que permiten interactuar a las personas con su ambiente utilizando el lenguaje, incluyendo la interacción humano-humano mediada por computadora. La interacción actual con una computadora es efectiva, pero la tuvimos que aprender y adaptarnos a ella. Lo que realmente se desea es una comunicación natural, de manera que cualquier persona pueda comunicarse con una computadora. Esto implica que un sistema automático no sólo debe descifrar el contenido del mensaje e interpretarlo adecuadamente para realizar la tarea solicitada, sino integrar otros elementos propios de la comunicación humana.

En estas situaciones la exigencia de que la computadora domine el lenguaje humano es ineludible. La idea de convertir una máquina en un asistente *inteligente*, orientado a resolver o asistir en la realización de una tarea, ha sido un motor importante para el crecimiento de las Tecnologías del Lenguaje. Los dispositivos actuales, como los teléfonos inteligentes y las tabletas digitales, empiezan a tener funcionalidades en esta dirección. El sistema Siri de Apple es un buen ejemplo de este tipo de aplicaciones. Dotar de capacidades a una máquina para establecer un diálogo natural tiene un sin fin de aplicaciones, pero en casos como el de un robot de servicio, éste quedaría *incompleto* si no es capaz de responder de esta manera. Bajo este escenario, se desea llegar lo más cercano posible a la comunicación humana, cubriendo aspectos contextuales y no reser-

vándose únicamente al contenido del mensaje mismo. La forma y el momento en que se emite el mensaje nos da elementos no sólo para la correcta interpretación del mensaje, sino información sobre nuestro interlocutor, cómo su estado de ánimo o la urgencia de la petición. Un *sistema conversacional* completo deberá aprehender el significado del mensaje, el propósito y las emociones de su interlocutor para responder en consecuencia.

Cabe remarcar que las Tecnologías del Lenguaje buscan soluciones, en estas tres grandes áreas, al analizar el lenguaje usado en escenarios específicos. Es decir, se aplican enfoques que capturan las regularidades del lenguaje al observar grandes colecciones de datos, ya sean escritos u orales. Por ejemplo, los sistemas de traducción más populares realizan su tarea al comparar la frase a traducir con miles de frases previamente traducidas.³ La calidad de la producción depende en gran medida de la cantidad del corpus disponible. De ahí la importancia de contar con grandes colecciones para tener éxito usando estas técnicas. No obstante, contar con datos suficientes no lo es todo, los métodos estadísticos ignoran las propiedades intrínsecas de una lengua. De ahí, que dependiendo del problema es posible proveer soluciones híbridas basadas en la mezcla de métodos estadísticos –calculados en base a corpus suficientemente grandes– y métodos sustentados en procesamientos lingüísticos de la lengua.

³Es importante mencionar que en los últimos años, los enfoques tradicionales han sido superados por aquellos basados en redes neuronales profundas (*deep learning*). El trabajo de la comunidad internacional se ha volcado sobre este tipo de algoritmos logrando éxitos importantes en problemas como la traducción automática o el reconocimiento de voz.

Capítulo 2

Conceptos básicos

Hiram Calvo, CIC/IPN.

Luis Alberto Pineda Cortés, IIMAS/UNAM.

Luis Villaseñor Pineda, INAOE.

El presente capítulo no pretende ser exhaustivo y presentar todo el ámbito de las Tecnologías del Lenguaje, éste únicamente pretende presentar algunos de los conceptos más relevantes del área, en específico, los formulados bajo la Lingüística Computacional.

2.1. Lingüística Computacional

Al igual que la lingüística descriptiva tradicional, la Lingüística Computacional aborda una gama muy amplia de fenómenos del lenguaje, tanto hablado como escrito, pero con la restricción de que los modelos se deben caracterizar mediante sistemas de reglas formales y/o procedimientos computacionales, lo que impone una restricción muy severa a los tipos de modelos admisibles. En este enfoque, el objeto de investigación es la representación computacional del conocimiento lingüístico, así como los procesos que lo utilizan. Por lo mismo, esta

disciplina caracteriza los diferentes niveles de representación lingüística, así como sus relaciones. En general se reconocen seis niveles, como sigue: i) el fonético y fonológico, ii) prosódico y entonativo, iii) léxico y morfológico, iv) sintáctico, v) semántico y vi) pragmático.

Por la magnitud de la empresa y con el fin de abordar sus objetivos específicos, la Lingüística Computacional se ha desarrollado históricamente en varias especialidades, con metáforas, teorías y metodologías diferentes, y cada una de éstas ha representado un esfuerzo de investigación y desarrollo tecnológico de dimensiones colosales. Entre las más prominentes podemos mencionar: i) Reconocimiento del habla y síntesis de voz (1952), ii) Traducción automática entre lenguajes naturales (1954), iii) Procesamiento del lenguaje natural (1964) y iv) Lingüística de corpus (1967).

Los programas y dispositivos que se presentaron inicialmente –como prueba de concepto– generaron grandes expectativas y se pensó que la construcción de la máquina del lenguaje iba a ser sólo cuestión de tiempo. Sin embargo, las propuestas no se materializaron y estas disciplinas pasaron por varios ciclos con un fuerte impulso inicial hasta su agotamiento a lo largo de la segunda mitad del siglo XX. A pesar de ello, durante este período se creó un acervo de conocimiento que se reflejó en la aparición de varias revistas emblemáticas donde se detalla la historia de la disciplina, principalmente *Computational Linguistics*, *Artificial Intelligence* y otras más, en libros de texto de procesamiento de lenguaje natural y reconocimiento del habla y en varias conferencias internacionales de gran prestigio. Asimismo, a finales de la década de los noventa y principios de este siglo aparecieron dos textos que son ahora referencia central para los estudios de posgrado en la especialidad.

Con el fin de contextualizar las contribuciones de la comunidad mexicana a esta disciplina, en la Sección 2.2 se describen los seis niveles de representación lingüística principales, aunque sin comprometerse, en la medida de lo posible, con ninguna corriente de pensamiento o teoría del lenguaje en particular. Con este antecedente en ellos siguientes capítulos se describen los principales estudios teóricos y aplicaciones realizados por la comunidad mexicana.

2.2. Modelos computacionales de la estructura del lenguaje

Los estudios computacionales de la estructura del lenguaje se iniciaron con la publicación de *Syntactic Structures*, por Noam Chomsky, quien propuso por primera vez que la estructura sintáctica del lenguaje se puede modelar a través de un sistema de reglas completamente mecanizadas o formales. Esta teoría abrió la puerta para modelar dentro del paradigma computacional no sólo la sintaxis sino los diversos niveles de representación lingüística y tuvo repercusiones en una gama muy amplia de disciplinas científicas, tecnológicas y en las humanidades. Los niveles principales de estructura lingüística son como sigue:

2.2.1. Nivel fonético y fonológico

Las unidades básicas de la estructura del lenguaje hablado se conocen como fonemas. Los fonemas se reconocen por sus características combinatorias o sintagmáticas, y de contraste o paradigmáticas, y corresponden, hasta cierto punto, con los símbolos del alfabeto. Por ejemplo, la palabra *casa* está constituida por cuatro fonemas en secuencia o en relación sintagmática, es decir *c*, *a*, *s* y *a*, y contrasta paradigmáticamente con las palabras *tasa* y *cara* en la primera y terce-

ra posición respectivamente, por lo que dichas palabras se distinguen entre sí y tienen significados diferentes. El sistema fonológico se caracteriza por las reglas que permiten o restringen las secuencias y contrastes entre estas unidades, y que generan el conjunto de palabras actuales y potenciales de una lengua, como el español o el inglés.

Por su parte, cada fonema da lugar a una realización acústica específica, aunque hay fonemas que tienen diferentes realizaciones. Por ejemplo, en el español de México la *ch* en la palabra *Chihuahua* se pronuncia de manera diferente por hablantes del centro y del norte del país (africada sorda y fricativa respectivamente) pero dicha alteración no cambia su significado, por lo que el fonema es el mismo pero tiene dos realizaciones acústicas o alófonos diferentes. Mientras la fonología estudia la estructura de estas unidades abstractas, la fonética estudia las propiedades físicas de sus realizaciones acústicas. Las unidades acústicas de los lenguajes humanos se codifican en el alfabeto fonético internacional (AFI), el cual se concibió originalmente para representar la pronunciación de las palabras en los diccionarios bilingües y permitir la comunicación entre hablantes de lenguas maternas diferentes. Por otra parte, cada región lingüística utiliza un conjunto de unidades acústicas específicas y el estudio de la fonética de su dialecto se centra en la definición de su alfabeto fonético. Los alófonos se pueden representar por un símbolo en un alfabeto fonético computacional, el cual se puede asociar a la representación de las características físicas de la señal de audio. Por razones prácticas, se han desarrollado alfabetos fonéticos para cada dialecto para habilitar el reconocimiento de voz y la traducción computacional entre lenguajes hablados.

2.2.2. Nivel de prosodia y entonación

En este nivel se representa la acentuación de las palabras y la estructura tonal de los enunciados. Las unidades sobre las que recaen los tonos son las sílabas y la entonación se caracteriza por la variación de la frecuencia fundamental o F_0 en la elocución, la cual permite distinguir las oraciones declarativas, interrogativas y admirativas. Al igual que en el nivel fonético-fonológico, la entonación se puede estudiar desde la perspectiva de la estructura tonal abstracta o en términos de su realización física. Este nivel de representación es necesario para reconocer los tipos de intenciones expresadas en el habla y es fundamental para la creación de voces sintéticas de calidad, que pongan el acento de las palabras en la sílaba correcta y que trasmitan la intención de la elocución de manera clara.

2.2.3. Nivel léxico y morfológico

El siguiente nivel de representación se enfoca al análisis de la estructura abstracta de las palabras. Esta estructura se puede pensar en términos de una lista de palabras o “lexicón” con una entrada por cada palabra; cada entrada a su vez contiene una secuencia de fonemas, la cual se asocia a uno o varios alófonos así como a la representación de su significado convencional. Intuitivamente, el primer paso para comprender el lenguaje hablado es el reconocimiento de la voz, que consiste en “alinear” a las secuencias de unidades acústicas en la elocución con secuencias de realizaciones de palabras en el lexicón y, a través del mismo, con sus significados convencionales.

Es necesario considerar que la gran mayoría de las palabras tienen variaciones debido a partículas como prefijos, infijos y sufijos que las transforman para especificarlas en algún sentido predecible y reconocible por los hablantes de la lengua; por ejemplo las palabras *inmaterial*, *corredor* y *materialista* se forman

respectivamente a partir del prefijo de negación *in*, el infijo *do* y el sufijo *ista*. El primero niega la propiedad adscrita por el adjetivo que se modifica (*material*), la segunda crea el nombre de quien realiza una actividad a partir del verbo que la nombra, y la tercera crea el nombre de quien cree o hace algo a partir de la propiedad que adscribe el adjetivo modificado (*Juan es un materialista*). Si pensamos que el lexicon contiene una forma nuclear de cada palabra, asociada a un significado básico, la morfología computacional estudia las reglas formales que producen las transformaciones posibles del núcleo, que a su vez producen una alteración correspondiente en la representación de su significado. Esta ampliación dinámica del lexicon permite que se puedan reconocer no sólo las palabras actuales sino también las potenciales en conjunto con sus significados convencionales.

2.2.4. Nivel sintáctico

La sintaxis caracteriza la estructura de las frases y las oraciones. En este nivel se considera que dentro de la oración hay palabras cuya función es central en la oración y que las otras tienen un papel subordinado. Estas relaciones se representan como una jerarquía cuyo nodo superior o cabeza es el núcleo gramatical y los nodos inferiores son los constituyentes. Dicha estructura puede ser simple o muy compleja dependiendo de los fenómenos involucrados y del formato de representación y se conoce como “estructura sintáctica”. Su relevancia se debe a que es “la portadora” del significado convencional de la oración. Mientras que el significado de las palabras se codifica directamente en el lexicon, el significado convencional de la oración depende de los significados de las palabras y de cómo se combinan en la estructura sintáctica. Por esta razón, desde la propuesta original de Chomsky, inducir la estructura sintáctica a partir de las reglas de la

gramática y el estímulo lingüístico se considera como una de las tareas centrales del procesamiento del lenguaje. Este proceso se conoce como “parseo” y su estudio ha sido también sujeto de una investigación de grandes dimensiones en la Lingüística Computacional.

La estructura sintáctica debe tomar en cuenta que tanto las palabras básicas como las derivadas morfológicamente pueden sufrir transformaciones adicionales cuando se ponen en el contexto de una frase o una oración, como las inflexiones debidas al género y al número de los sustantivos, que en el español deben concordar con las inflexiones de los verbos. Por lo mismo, aunque las inflexiones son también partículas que modifican a las palabras su carácter es sintáctico y no morfológico. Un fenómeno de características morfo-sintáctico muy singular y frecuente en nuestra lengua es el de los llamados pronombres clíticos, como *se* y *lo* en *se lo comió* o *dáselo*. Lo interesante de estas partículas, independientemente de la determinación de sus referencias, es que pueden aparecer como morfemas o como palabras independientes, tanto adelante como atrás del verbo, de manera muy flexible, además de que llevan implícito su caso (acusativo o dativo) que indica si su referencia recibe la acción verbal directamente o es beneficiario de la misma: en *se lo comió*, *lo* es un pronombre acusativo que denota el objeto de comer, *lo comido*, y *se* es un pronombre dativo, cuya referencia es el beneficiario de dicha acción, es decir, quien realizó la acción de comer.

De forma muy general el análisis sintáctico se ha abordado desde dos enfoques principalmente: constituyentes y dependencias. Ambos se han estudiado durante más de 40 años y representan alternativas teóricas y metodológicas para el estudio y las aplicaciones de la Lingüística Computacional, como se verá a continuación.

2.2.4.1. Enfoque de constituyentes

En este enfoque, presentado originalmente por Chomsky, un constituyente es una palabra o grupo de palabras que cumplen una función específica en la oración. Estos grupos se conocen como “categorías gramaticales”. El proceso sintáctico consiste en encontrar a las categorías de una oración segmentándola en sus partes de manera recurrente hasta que las partes sean las palabras básicas o derivadas en el lexicon. Aunque el número de palabras en el lexicon y el número de reglas sintácticas sea finito, la aplicación sistemática de las reglas sintácticas y morfológicas genera un número infinito de oraciones. En la Figura 2.1 se ilustra una estructura sintáctica por constituyentes, donde *O* representa a la oración, *GN* al grupo nominal y *GV* al grupo verbal. Las estructuras jerárquicas se pueden representar linealmente poniendo la expresión o subexpresión entre paréntesis seguida de un subíndice que indica la categoría del constituyente, como se indica en la parte superior de la Figura 1.1.

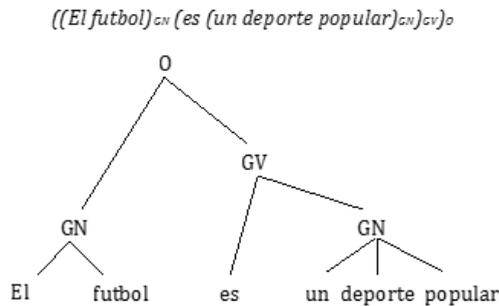


Figura 2.1: Análisis por constituyentes de *El futbol es un deporte popular*.

2.2.4.2. Enfoque de dependencias

Por su parte el enfoque de dependencias iniciado por Lucien Tesnière en 1959, establece que la estructura sintáctica consiste en relaciones de dependencia entre pares de palabras donde una es la principal, rectora o cabeza, y la otra está subordinada. Si cada palabra de la oración tiene una palabra propia rectora (cabeza), la oración entera se ve como una estructura jerárquica de diferentes niveles, o como un “árbol de dependencias”. La única palabra que no está subordinada es la raíz del árbol. En la Figura 2.2 se ilustra la estructura de la oración en la Figura 2.1 pero en términos de dependencias. A diferencia del enfoque de constituyentes, en el enfoque de dependencias no se postulan categorías sintácticas abstractas por lo que las estructuras sólo contienen unidades léxicas concretas. Por lo mismo, la oposición entre el enfoque por constituyentes y el enfoque por dependencias proviene a su vez de una oposición más profunda entre la hipótesis de que la estructura sintáctica es una representación abstracta *versus* la hipótesis de que ésta tiene un carácter concreto.

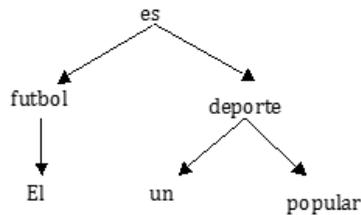


Figura 2.2: Representación en dependencias de *El fútbol es un deporte popular*.

2.2.5. Nivel semántico

El propósito de la semántica es caracterizar el significado convencional o independiente del contexto de emisión o interpretación de una unidad lingüística, ya sea básica, como las palabras, o compuesta, como las frases y las oraciones. La semántica computacional se basa en el llamado “Principio de Composicionalidad”, que establece que el significado de una estructura compuesta es función del significado de sus partes y de su modo de combinación gramatical. En su formulación más general la semántica de una oración se caracteriza como una función proposicional, donde la palabra principal en la estructura es la función y las otras sus argumentos. Por ejemplo, la representación semántica o estructura de argumentos de *Juan se comió el pan* es la predicación $\text{comer}(E, \text{juan}, \text{pan}) \& \text{pasado}(E) \& \text{beneficiario}(E) = \text{Juan}$, cuya interpretación es que E es un evento de comer que ocurrió en el pasado, que el agente de dicho evento fue Juan, quien comió, y el paciente (el ente que es pasivo en la acción) fue el pan, lo comido, mientras que se es un pronombre dativo que duplica al sujeto e indica que éste se benefició con la acción de comer. Estas representaciones pueden ser de carácter concreto, como en el presente ejemplo, pero también se pueden referir a abstracciones muy profundas. Posiblemente el formalismo más general para caracterizar la semántica de los lenguajes naturales de forma computacional es la Semántica de Montague aunque existe una gran variedad de formatos representacionales, como redes semánticas, primitivas conceptuales, etc., que son materia de la representación del conocimiento.

2.2.6. Nivel pragmático

Toda unidad lingüística, ya sea hablada o escrita, se realiza e interpreta en relación a un “contexto”. La noción de contexto se puede ilustrar con una ana-

logía pictórica: todo cuadro tiene una imagen saliente, “la figura”, que ocurre siempre en relación con una escena o “fondo”. La unidad lingüística es la figura y el contexto el fondo. En general se pueden distinguir dos tipos de contextos: i) el lingüístico, que consiste en información aportada por el lenguaje mismo a lo largo de la conversación o el discurso y ii) el situacional, que consiste en información extra-lingüística y que tiene como parámetros al hablante, al oyente y posiblemente a uno o más terceros, con sus expectativas e intenciones, así como la situación espacial y temporal en la que se realiza o interpreta la elocución.

El contexto se puede apreciar, como una primera aproximación, en relación a las palabras que funcionan como índices o variables cuya referencia o denotación cambia en cada situación de uso. Los índices más directos y evidentes son los pronombres. Si éstos se interpretan en relación al contexto lingüístico, es decir a lo que se ha dicho en la conversación o el discurso, se llaman anafóricos; por su parte, si se interpretan en relación al contexto no lingüístico, como la situación espacial o temporal, se llaman deícticos o indexicales. Asimismo, las inferencias para encontrar los referentes de los pronombres se conocen como anafóricas e indexicales respectivamente. Esta distinción es esencial no sólo para interpretar el lenguaje en uso, sino también para interpretar representaciones multimodales, como los mapas o diagramas con sus anotaciones textuales.

Para ilustrar la complejidad de la inferencia anafórica considere el discurso compuesto por las oraciones *Pepe vio a Juan comprar el pan, se lo comió*. El problema es determinar quiénes son los referentes de los pronombres *se* y *lo*, para lo cual es necesario identificarlos y correferenciarlos con sus antecedentes. Una posibilidad es que se tenga como antecedente a *Pepe*, quien comió, y *lo* al pan, lo comido. Sin embargo, también puede ser el caso que quien se comió el pan haya sido Juan e incluso que quien comió haya sido Pepe y lo comido Juan. La

inferencia anafórica es sumamente compleja y ha sido también objeto de un esfuerzo de investigación de grandes dimensiones en lingüística computacional. Por su parte, los pronombres indexicales tienen que tomar su referente directamente del mundo en relación al hablante y al oyente, como en la interpretación de *yo* en *yo vi a Juan comerse el pan*, cuyo referente será quien profiera la elocución en el contexto particular. Hay también otras palabras como los adverbios *aquí*, *allá*, *ahora* o *ahorita* que se interpretan en relación a una locación espacial o un momento o un intervalo en el tiempo y tienen una connotación indexical pura.

Sin embargo, los pronombres que se usan normalmente como anafóricos también se pueden utilizar como indexicales, en cuyo caso el proceso de interpretación se vuelve más complejo. Por ejemplo, en una situación en que el pan está sobre la mesa y Pepe le dice a Juan *¿me lo pasas?* al tiempo que señala al pan con su dedo índice, la interpretación de *me* es quien profirió la elocución y es beneficiario de la acción, es decir Pepe, y de *lo* es el pan, donde el contexto relevante es el mundo y no el discurso. Aunque en principio se puede distinguir al contexto lingüístico y al que se establece por la situación en el mundo, en la práctica los contextos interactúan y el problema de interpretación es en general sumamente complejo. Un caso particular de índices son los nombres propios como Pepe, Juan y Pedro, que se refieren en cada caso al individuo en el contexto, aunque haya miles de individuos que han tenido, tienen y tendrán estos nombres.

La interpretación del lenguaje en cada situación de uso, en oposición a la interpretación convencional estudiada por la semántica, se complica aún más cuando se toman en cuenta las intenciones y las creencias de los interlocutores, así como la entonación de las elocuciones. En español las oraciones declarativas

que expresan una proposición tienen una entonación relativamente plana, las interrogativas tienen una curva ascendente al final, y las imperativas como las órdenes enfatizan los tonos iniciales. Por lo mismo, es posible identificar si una elocución es un enunciado, una pregunta o una orden, sin siquiera interpretar el sentido convencional de la oración. Se dice que todo enunciado en cada situación de uso es un “acto del habla”, y que, si la entonación corresponde al tipo del acto del habla, éste es directo, pero si esta relación se cambia, el acto del habla es indirecto. Por ejemplo, enunciar *¿Me puedes pasar el pan?* en un contexto normal de interpretación no es una solicitud de información –si el interlocutor tiene la capacidad de llevar a cabo dicha acción– sino una directiva de acción, y en ciertos contextos puede ser incluso una orden. Los actos del habla indirectos ocurren muy frecuentemente en la conversación cotidiana y su estudio es también objeto de la pragmática.

Los actos del habla indirectos pueden también alterar tanto el material léxico como la forma sintáctica del enunciado; por ejemplo, si Pepe y Juan están en una habitación donde hace mucho frío y el primero le dice al segundo, quien está junto a la ventana, que está abierta, *¿Qué tal tu veranito?*, la intención es realmente la directiva de acción o incluso la orden de que Juan cierre la ventana. Este ejemplo ilustra cómo el contexto o situación de interpretación espacial y temporal, que involucra también a las expectativas e intenciones de los interlocutores, es indispensable para interpretar al lenguaje en uso.

En resumen, el propósito del análisis pragmático computacional es inferir la intención o significado de una elocución de manera automática a partir de su significado convencional y del contexto. Asimismo, la salida del módulo pragmático se puede conceptualizar como la especificación de las acciones que el intérprete tiene que realizar, entre un conjunto posiblemente muy amplio de tipos

de acciones, como respuesta a la intención expresada por su interlocutor. Estas acciones pueden ser mentales, motoras e incluso emotivas. Mental, como consultar y reportar información, o hacer una inferencia conceptual, como determinar si dos expresiones son sinónimas, o deliberativa, como hacer un diagnóstico, tomar una decisión o inducir un plan; motora, como tomar un objeto y entregarlo a un destinatario; y emotiva, como ponerse muy contento al recibir una buena noticia.

2.3. La ambigüedad

Un problema que aqueja a los modelos computacionales de la interpretación del lenguaje natural es la ambigüedad. Este fenómeno se puede apreciar directamente en el lexicon, ya que hay muchas palabras que tienen más de un significado convencional, como banco o gato, y su resolución consiste en inferir qué es lo que quiso decir quien las expresó en la situación de uso, tanto en el lenguaje hablado como en el escrito. La ambigüedad aparece también en el nivel sintáctico; por ejemplo, en *Pepe vio a Juan en el banco comiéndose el pan*, es ambiguo quién estaba en el banco, quién estaba comiéndose el pan, y qué quiere decir banco. Se invita al lector a visualizar las diversas escenas descritas por esta oración, incluyendo la escena en que los dos están comiendo pan sentados en un banco en el banco. El problema es más agudo si se toma en cuenta que la ambigüedad puede también surgir en el nivel fonético y la entonación, y frecuentemente se sostiene que puede ocurrir en todos los niveles de representación lingüística.

Los modelos generativos o basados en reglas seleccionan una entrada léxica para cada palabra y asignan una estructura sintáctica particular a cada interpreta-

ción posible, que se refleja como una predicación particular en la semántica, por lo que la aplicación sistemática de estos procesos produce un número significativo de estructuras sintácticas y, consecuentemente, de interpretaciones para cada elocución, aunque normalmente sólo una es la apropiada en el contexto. La resolución de la ambigüedad se ha abordado tradicionalmente tanto con la inclusión de restricciones que prevengan la generación de interpretaciones incorrectas como con la generación de todas las interpretaciones para después eliminar las incorrectas, o mediante combinaciones de estas dos estrategias. Sin embargo, si sólo se toma en cuenta el significado convencional de las palabras o las oraciones, el problema es realmente muy complejo, y frecuentemente se sostiene que éste es el mayor problema de la lingüística computacional y que todo lenguaje bien regimentado debería excluir completamente a la ambigüedad, como sucede en los lenguajes formales.

Sin embargo, la ambigüedad es también un recurso expresivo que se puede capitalizar para expresar varias figuras lingüísticas, como la generalización y la ironía, entre muchas otras, y normalmente los seres humanos no generamos interpretaciones incorrectas, cuando menos conscientemente; más aún, somos capaces de apreciar los diferentes sentidos posibles y entender el chiste. Esto se debe a que a diferencia de los modelos formales que utilizan tan sólo significados convencionales, los seres humanos interpretamos el lenguaje en relación al contexto. Considere que si las palabras u oraciones se expresan en un contexto espacial o temporal específico, digamos por Pedro, quien profiere *Pepe vio a Juan en el banco comiéndose el pan* al tiempo que ve la escena cuando ocurre, o viendo una fotografía del evento, la información extralingüística previene parcial o totalmente la generación de hipótesis incorrectas.

Capítulo 3

La Investigación en México sobre Tecnologías del Lenguaje

Hiram Calvo, CIC/IPN.

Noé A. Castro Sánchez, CENIDET.

Alexander Gelbukh, CIC/IPN.

Yasmín Hernández Pérez, INEEL.

Héctor Jiménez Salazar, UAM-Cuajimalpa.

Manuel Montes y Gómez, INAOE.

Luis Alberto Pineda Cortés, IIMAS/UNAM.

David Pinto Avendaño, BUAP.

Fernando Sánchez Vega, INAOE.

Grigori Sidorov, CIC/IPN.

Luis Villaseñor Pineda, INAOE.

La oportunidad que brinda la cantidad de información disponible en la Internet es enorme. Casi todo lo que vemos, leemos, escribimos u oímos es recolectado y puesto a nuestra disposición a través de ella. Por supuesto, recuperar un documento o grabación específica dentro de este mar de datos es una tarea compleja. Entre los varios desafíos a enfrentar, y posiblemente el más difícil, es lograr que una máquina comprenda el contenido de estos documentos –escri-

tos, orales, multilingües y/o multimodales— de tal forma que cualquier persona tenga las respuestas apropiadas. Es justo bajo esta problemática que las Tecnologías del Lenguaje (TL) trabajan.

La comunidad científica mexicana viene trabajando en Tecnologías del Lenguaje (TL) desde hace varios años. Los trabajos realizados son muy variados e impactan de diferente manera las tres grandes áreas de aplicación. Las contribuciones de la comunidad mexicana van desde la creación de recursos lingüísticos —transversales a las tres áreas— como contribuciones específicas a tareas particulares. La presente sección presenta de manera general una muestra de las áreas de investigación desarrolladas por la comunidad mexicana.

3.1. Lingüística de corpus

La lingüística de corpus se enfoca a la creación de recursos lingüísticos, tanto en la modalidad hablada como en la textual. Los corpus son recursos empíricos codificados en formatos computacionales para el estudio de todos los niveles de representación lingüística, así como para el desarrollo de aplicaciones diversas. El desarrollo de corpus exige procedimientos y criterios rigurosos por su magnitud y para su diseño, recolección y organización, de manera que sean confiables y apropiados para las tareas de interés.

La práctica de recabar corpus se inició en el entorno computacional de manera muy temprana con la recopilación del conjunto de textos de diferentes géneros del *Brown Corpus*¹ y más tarde con la construcción de textos anotados sintácticamente, llamados *Treebanks*², particularmente con el *Penn Treebank Project* (Taylor, Marcus y Santorini 2003). Un corpus muy conocido para el lenguaje

¹https://en.wikipedia.org/wiki/Brown_Corpus

²<https://en.wikipedia.org/wiki/Treebank>

hablado es el *Switchboard*.³ Para el español, un corpus textual anotado o transcrito en varios niveles es el *Cast3LB Corpus* (Navarro, Izquierdo y Saiz-Noeda 2004)

Los corpus textuales de primera generación contenían alrededor de un millón de palabras, mientras que los mega corpus actuales contienen más de cien millones de palabras; esto es en parte posible gracias a la Web. Los corpus deben prepararse para su tratamiento computacional, por ejemplo, para la búsqueda de ocurrencias de determinado tipo y número de palabras, la extensión oracional, etc. A continuación, se muestran diversas herramientas de utilidad para este tratamiento y en especial para corpora de gran magnitud.

3.1.1. Modelos del lenguaje

Una tarea básica en el análisis del lenguaje es la predicción de las unidades que se siguen en una elocución o un texto (por ejemplo, alófonos o palabras) dadas las unidades que se han observado. Esta predicción se puede hacer utilizando modelos probabilísticos, llamados modelos de n-gramas. Un n-grama es una secuencia de n tokens o instancias de palabras; un 2-grama (o bigrama) es una secuencia de dos palabras como favor de, de guardar o guardar silencio; un 3-grama (trigrama) es una secuencia de tres palabras como favor de guardar o de guardar silencio. Un modelo de lenguaje consiste en la asignación de una probabilidad para todos los n-gramas diferentes en relación al corpus de referencia, utilizando para este efecto un algoritmo, normalmente de carácter probabilístico. Por ejemplo, la probabilidad de un unigrama es simplemente la probabilidad de ocurrencia de cada palabra en relación al corpus, y la probabilidad de un bigrama es la probabilidad de que la palabra al final ocurra dada la probabilidad de

³<https://catalog.ldc.upenn.edu/ldc97s62>

que la palabra inicial ocurre. Este procedimiento se aplica de manera recurrente para obtener la probabilidad de n-gramas de orden mayor.

Es muy importante que el corpus sea representativo o informativo del dominio de interés ya que su calidad impacta directamente en la calidad de los modelos del lenguaje. Por lo mismo es común utilizar medidas entrópicas para evaluarlo. Una medida común es la perplejidad, que cuantifica el promedio de las secuencias del lenguaje que pueden seguir a una secuencia dada. Mientras menos sean éstas más predecible e informativo es el corpus, y su entropía tiene un valor más bajo, por lo que el corpus será de mayor utilidad.

Por su parte, en todos los corpus hay contextos que no aparecen de manera contingente aunque puedan ocurrir en el lenguaje, por lo que sus n-gramas tendrán un valor de cero. Para enfrentar este problema es común reasignar la masa de probabilidad de los n-gramas, quitándole a los que más tienen y distribuyéndola entre los que menos tienen, de forma que la masa total se conserve. Este proceso se conoce como “suavizado” y hay diversas estrategias y algoritmos para llevarlo a cabo. Por ejemplo, si un trigramo tiene probabilidad cero, se le puede asignar la misma probabilidad del trigramo que menos probabilidad tiene en el corpus, al tiempo que se les quita esta probabilidad de manera proporcional a los que más tienen. Este tipo de estrategias son muy útiles en la práctica ya que en un corpus suavizado todos los n-gramas tienen un valor, a pesar de que sus partes constitutivas no lo tengan.

Los modelos de lenguaje tienen una gran variedad de aplicaciones, como la corrección de errores gramaticales, reconocimiento de voz, análisis de opiniones, generación de lenguaje natural, medida de semejanza entre palabras, identificación de autoría, entre muchas otras, como se explica más adelante en este texto.

3.1.2. Corrección ortográfica

Supongamos por ejemplo que la frase *en quince monitos* ocurre en un texto. El error se puede detectar si la probabilidad del 3-grama *en quince monitos* es cero, pero, al mismo tiempo, la probabilidad del 2-grama *en quince* y del 1-grama *monitos* no lo son; por lo mismo, *monitos* en dicho contexto debe ser un error. Por otra parte *monitos* se parece a *minutos*, lo cual se puede evaluar por algún tipo de medida de similitud entre palabras; supongamos adicionalmente que el 3-grama *en quince minutos* tiene una probabilidad diferente de cero. Consecuentemente es posible corregir el error substituyendo *monitos* por *minutos*. Por supuesto, esta explicación es sumamente simplificada, pero se incluye para dar una idea de la intuición subyacente a este tipo de algoritmos. Por ejemplo en (Hernández y Calvo 2014) se propone la corrección de errores gramaticales mediante un modelo de lenguaje de trigramas sintácticos continuos (Sidorov y col. 2013) y no continuos (Sidorov 2013) obtenidos de un corpus de texto de dimensiones muy significativas.

3.2. Reconocimiento de voz

El proceso de traducir el habla a su representación textual por medio de un proceso computacional se conoce como “reconocimiento de voz”. Este proceso se conceptualiza en relación a los niveles fonético-fonológico, léxico y morfológico. El reconocimiento se inicia con el análisis de la señal del habla para identificar la secuencia de unidades acústicas de la elocución. El proceso propiamente consiste en alinear dicha secuencia con una secuencia de pronunciaciones de un conjunto de palabras en el lexicón. Como el lexicón contiene a las palabras representadas como secuencias de fonemas, posiblemente agrupados en morfemas,

se requiere adicionalmente traducir dicha forma abstracta a su realización ortográfica, lo cual se puede hacer conceptualmente mediante reglas que traduzcan fonemas y morfemas a grafemas (o símbolos ortográficos). Los algoritmos de reconocimiento de voz son normalmente probabilísticos, por lo que cada elocución puede dar lugar a un número muy significativo de hipótesis, cada una asociada a un valor de preferencia o *score*.

Por otra parte, los modelos de lenguaje asignan una probabilidad a la elocución, la cual se puede tomar como su probabilidad a priori. Mediante este recurso adicional la hipótesis preferida por el reconocedor de voz será aquella cuyo producto de su *score* acústico y la probabilidad que le asigne el modelo del lenguaje sea mayor.

La creación de sistemas de reconocimiento de voz requiere de dos grandes esfuerzos de investigación: i) desarrollar los algoritmos de reconocimiento propiamente, incluyendo los algoritmos para crear los modelos del lenguaje y ii) crear los lexicones o diccionarios fonéticos de la lengua en cuestión para lo cual se requiere crear los modelos acústicos de todas las realizaciones de los fonemas del dialecto para el que se construye el reconocedor. Para esto último es necesario contar con un alfabeto fonético específico para el dialecto en el que se incluya un símbolo para cada alófono de cada fonema. Asimismo, se requiere contar con una gran cantidad de instancias de la realización acústica de cada alófono, recolectadas en los contextos acústicos en los que pueda ocurrir, para crear su modelo a nivel de la señal. También es necesario crear los diccionarios fonéticos propiamente, los cuales deben incluir todas las palabras del dialecto, normalmente en las decenas de miles, así como las diferentes formas en que cada palabra se puede pronunciar. Además, es necesario contar con corpus de grandes dimensiones en los dominios lingüísticos en los que se utilizará el reconocedor para crear los mo-

3.3. SISTEMAS CONVERSACIONALES EN LENGUAJE NATURAL²⁹

delos del lenguaje correspondientes. La creación de estos recursos lingüísticos es también parte de la lingüística de corpus.

Con el fin de contar con un recurso lingüístico de esta naturaleza para el español de la Ciudad de México se desarrolló el alfabeto Mexbet (Cuétara 2004), el cual hizo posible diseñar, coleccionar y transcribir el corpus DIMEx100 (L. A. Pineda, Castellanos y col. 2010). Asimismo, este corpus se utilizó para construir un diccionario fonético que incluye las realizaciones de cada palabra, transcritas adicionalmente en tres niveles de granularidad, y se validó con la construcción de un número muy significativo de reconocedores de voz utilizando los algoritmos del sistema Sphinx producido por la Universidad de Carnegie Mellon.⁴

3.3. Sistemas conversacionales en lenguaje natural

Desde la propuesta inicial de Turing el gran reto de la lingüística computacional ha sido la construcción de sistemas que puedan conversar con los seres humanos a través del lenguaje, especialmente hablado. Los esfuerzos en esta línea de investigación se iniciaron a mediados de la década de los sesenta con el programa Eliza⁵ que simulaba ser un psicoterapeuta Rogeriano, y un poco más tarde por el programa SHRDLU,⁶ ambos desarrollados en el MIT. Esta tradición se continuó en la década de los noventa con sistemas interactivos capaces ya de interactuar en inglés hablado, apoyados por máquinas inferenciales, principalmente para hacer planes, como los sistemas de la serie TRIPS y TRAINS desarrollados en la Universidad de Rochester (Ferguson y Allen 1998).

⁴https://es.wikipedia.org/wiki/CMU_Sphinx

⁵<https://en.wikipedia.org/wiki/ELIZA>

⁶<https://en.wikipedia.org/wiki/SHRDLU>

La construcción de sistemas conversacionales o de diálogo en lengua je natural se ha abordado en México en el contexto del Proyecto Diálogos Inteligentes Multimodales en Español (DIME).⁷ Este proyecto se enfocó originalmente en la creación del Corpus DIME consistente en un conjunto de diálogos colaborativos para la solución de tareas, los cuales se etiquetaron ortográficamente, así como en los niveles fonético, prosódico y entonativo, léxico, sintáctico y pragmático. Para el nivel sintáctico se desarrolló una gramática del español de México con énfasis en el sistema de clíticos dentro de la perífrasis (L. Pineda y Meza 2005). Para el nivel pragmático se desarrolló el esquema de análisis DIME-DAMLS (L. Pineda, Castellanos y col. 2006) mediante el cual se obtuvo la estructura de los actos del habla, directos e indirectos, relativos a la tarea propiamente, a la administración de la tarea y a la administración de la comunicación.

Estas ideas se desarrollaron en paralelo con un modelo para la administración de diálogos computacionales enfocado a la interpretación de los actos del habla en relación a un contexto, el cual se caracteriza como una gráfica recursiva de situaciones llamada “modelo de diálogo”, donde cada situación se define en términos de las expectativas y acciones intencionales del agente computacional. En este modelo una situación puede también embeber a un modelo de diálogo subordinado, de tal forma que la estructura del grafo corresponde a la estructura de la conversación. Dichas ideas dieron lugar a la creación del lenguaje de programación SitLog (L. A. Pineda, Salinas y col. 2013), para la especificación e interpretación de modelos de diálogo en sistemas conversacionales y se incorporaron al Proyecto Golem,⁸ para el desarrollo de robots de servicio capaces de comunicarse con los seres humanos en lenguaje hablado, tanto en español como en inglés.

⁷<http://turing.iimas.unam.mx/luis/DIME/>

⁸<http://golem.iimas.unam.mx/>

3.4. Procesamiento de textos

El procesamiento de textos en México se ha trabajado desde muy diversas tareas y aplicaciones desde recuperación de información (Villatoro-Tello y col. 2010), búsqueda de respuestas (Montes, Villaseñor-Pineda y López-López 2008), (Téllez-Valero y col. 2011) y minería de texto (Ramírez-De-La-Rosa y col. 2013), (Ortega-Mendoza, Villaseñor-Pineda y Montes-y-Gómez 2007) hasta especificaciones de software (Díaz y col. 2005).

El enfoque más usado para procesar texto escrito es a través de la búsqueda de regularidades en los documentos de interés. Estas regularidades expresadas como patrones ayudarán a describir y/o discriminar documentos de diferentes tipos. Para identificar aquellas regularidades de interés, se han usado lexicones y/o tesoros, así como métodos estadísticos ó algoritmos de aprendizaje máquina. Por supuesto, la caracterización de los documentos cobra gran importancia, pues de ello dependerá la identificación de patrones útiles. Esta caracterización se puede realizar a diferentes niveles: morfológico, léxico, sintáctico, semántico, etc.

En las siguientes secciones se presentan varias de la tareas en procesamiento de textos realizadas por investigadores en México.

3.4.1. Análisis de opinión

El análisis de opinión, también conocido como extracción de opiniones, minería de opiniones, análisis subjetivo o análisis de sentimiento (*sentiment analysis*), ayuda a conocer la percepción de la comunidad acerca de productos o servicios. Este análisis se basa frecuentemente en la información textual disponible

en las redes sociales, como *Facebook* y *Twitter*, incluyendo los llamados “emoticones”.

Para este análisis se consideran la fuente o emisor, el objetivo o receptor y el tipo de actitud o polaridad, que puede ser positiva o negativa. El análisis de opinión puede ser simple, complejo o avanzado, dependiendo respectivamente de si sólo se reporta la polaridad, de si además se reporta el grado de la actitud, o si también se reporta la fuente y el objetivo, posiblemente con otras características, como teléfonos y fechas, así como los llamados *hashtags* en *Twitter*.

La metodología consiste en crear corpus anotados a diferentes niveles de granularidad con los parámetros de opinión y utilizar estos recursos para crear clasificadores mediante toda la gama de algoritmos de aprendizaje automático, para clasificar o anotar los textos analizados. Esta tarea se puede realizar también a través de lexicones de opinión⁹ o mediante la medición de distancias semánticas a diversos conceptos (Calvo 2015), incluso éstos se pueden crear de manera automática a partir de corpus anotados (Hu y Liu 2004).

3.4.2. Detección de ironía

Una tarea complementaria al análisis de opinión es el análisis de ironía, ya que esta figura retórica consiste precisamente en expresar una proposición mediante su negación, pero esperando que el interlocutor haga la interpretación correcta. El problema es que si una proposición se expresa irónicamente el análisis de opinión anotará las opiniones positivas como negativas y viceversa. La detección de ironía se ha abordado dentro del procesamiento de textos mediante modelos basado en ngramas simples de categorías gramaticales, de palabras frecuentemente utilizadas en textos con tono humorístico, tales como aquellas

⁹<https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html>

relacionadas con sexualidad, relaciones humanas, parentesco, así como de medidas de palabras usualmente clasificadas como positivas o negativas (Reyes, Rosso y Veale 2013). También se han considerado características de afectividad usando el recurso de WordNet-affect (Strapparava, Valitutti y col. 2004), y una medida de complacencia basada en diccionarios de términos afectivos (Reyes y Rosso 2012).

3.4.3. Semejanza entre palabras y diccionarios de ideas afines

Los diccionarios de ideas afines o tesauros (*thesaurus* en inglés) son recursos lingüísticos que organizan las palabras de acuerdo a la relación que guardan entre sí, tales como sinonimia, antonimia, hiperonimia, hiponimia, meronimia, holonimia, etc. Mediante estas relaciones es posible establecer “distancias” entre palabras (Ortega y col. 2011); por ejemplo, las palabras más cercanas a *cebra* son *jirafa*, *búfalo*, *hipopótamo*, *rinoceronte*, *gacela* y *antilope*; las próximas a *excepción* son *exención*, *limitación*, *exclusión*, *instancia*, *modificación*; y a *jarrón* son *tazón*, *sartén*, *jarra*, *contenedor*, *platillo* y *taza*.

Los tesauros tienen muchas aplicaciones dentro del procesamiento de lenguaje natural, tales como el análisis sintáctico (Calvo y Gelbukh 2004), la interpretación de conjunciones, la resolución de anáforas, la medición de cohesión en textos, la desambiguación de sentidos de palabras (ó *WSD* por sus siglas en inglés), la corrección ortográfica y el reconocimiento de voz, entre muchas otras. Algunos de los tesauros más utilizados son: WordNet,¹⁰ Roget (Roget y Dutch 1962), WASPS (Kilgarriff 2003), Word Sketches (Kilgarriff y col. 2004) y Medical Subject Headings (MeSH) (Rogers 1963).

¹⁰WordNet puede consultarse en línea en wordnet.princeton.edu

WordNet es una base de datos organizada de manera jerárquica que contiene un tesoro en línea junto con un diccionario para el idioma inglés. A través del proyecto EuroWordNet también está disponible para otros idiomas, como el español, italiano, alemán, francés, sueco, checo y estonio. WordNet define los sentidos utilizando un concepto llamado *synset* (conjunto de sinónimos). El *synset* contiene un conjunto de palabras que están relacionadas de manera cercana con la definición de la palabra, como se ilustra en los ejemplos mencionados.

3.4.4. Desambiguación del sentido de las palabras

La desambiguación del sentido de las palabras se aborda desde tres enfoques diferentes: i) la desambiguación basada en conocimiento que usa fuentes léxicas externas tales como diccionarios y tesauros (Calvo, Gelbukh y Kilgarriff 2005), (Tejada-Cárcamo, Calvo y Gelbukh 2008) aunado al uso de propiedades del discurso; ii) la desambiguación supervisada, la cual utiliza algoritmos de aprendizaje de máquina y se requiere contar con un corpus en el que se etiquete el sentido apropiado de las palabras (Rosso y col. 2005) y iii) la desambiguación no supervisada, que requiere un conjunto de entradas codificadas, pero no necesariamente el etiquetado del sentido apropiado (Calvo 2008). También hay enfoques mínimamente supervisados y el sentido más frecuentemente utilizado (Cárcamo, Gelbukh y Calvo 2008). En los enfoques basados en aprendizaje de máquina se construye un vector de características que representan al contexto (Tejada-Cárcamo, Calvo, Gelbukh y Hara 2010). En particular para el español, se evalúan los trabajos a través de concursos internacionales como SENSEVAL-2, SENSEVAL-3, que después dieron lugar a SEMEVAL.¹¹

¹¹<http://www.senseval.org>

3.4.5. Detección de engaño

El problema de detección de engaño se ha estudiado ampliamente, para determinar si una opinión es verdadera o falsa, es decir que tiene por fin engañar al lector. En esta tarea, el texto se representa también como vectores de características que se utilizan para entrenar clasificadores mediante técnicas de aprendizaje automático (Fusilier y col. 2015). En la mayoría de los trabajos se usan herramientas que requieren información específica del dominio, tales como n-gramas sintácticos y diccionarios léxicos de emociones. Sin embargo, otros trabajos consideran únicamente aspectos distribucionales del texto (Hernández-Castañeda y col. 2017), por ejemplo mediante el uso de algoritmos de modelado de tópicos dentro de un entorno probabilístico. Estos trabajos combinan sus características con otras obtenidas a partir de diversas fuentes, como representación de espacios de palabras, categorías gramaticales o diccionarios de aspectos psicológicos de las palabras (Sánchez-Junquera y col. 2018). Para efectos de evaluación existen diversos conjuntos de datos enfocados a dominios específicos. Aunque se ha tratado de encontrar un detector universal de engaño con alto desempeño, esto todavía no se ha logrado. Para más detalles, ver el capítulo 5 en este libro.

3.4.6. Atribución de autoría

La identificación de autor es otro problema que se aborda dentro del procesamiento de textos. Por ejemplo, es necesario saber qué autor escribió un libro anónimo, identificar la autoría de las notas de un posible criminal, etc. Esta tarea puede ser muy compleja cuando se realiza en un entorno abierto, es decir, cuando no se tiene un conjunto predefinido de posibles autores. Es por ello que se han creado tareas más específicas donde se incluyen varios documentos de un autor conocido (Coyotl-Morales y col. 2006), (Escalante, Solorio y Montes-y-

Gómez 2011) y un documento y un autor desconocido. Entre más cortos son los textos, la tarea se dificulta más, pues existen menos pistas que permitan distinguir el estilo de un autor. En casos reales como el campo forense, difícilmente se cuenta con textos largos. Otra tarea asociada con esta problemática es la identificación del perfil del autor, cuyo objetivo es inferir el género, lugar de origen, rango de edad e incluso rasgos de personalidad del autor a partir de los temas sobre los que versa el texto y su estilo (López-Monroy y col. 2015). Se han considerado diversos enfoques para obtener características más informativas basadas en el estilo; también es posible generar características al extraer información léxica, sintáctica o semántica. La información léxica usualmente se limita a conteos de palabras y ocurrencias de palabras comunes. Por otra parte, mediante la información sintáctica es posible obtener, hasta cierto punto, el contexto de las palabras. Algunos trabajos usan información semántica léxica para encontrar características que permitan discriminar los textos mediante modelos probabilísticos de distribuciones latentes.

3.4.7. Reconocimiento de paráfrasis e implicación textual

Otro reto del procesamiento de textos es el parafraseo, el cual consiste en reconocer si dos expresiones significan lo mismo, o reformular una expresión con otra que tenga el mismo significado. El reconocimiento de paráfrasis puede ser utilizado en aplicaciones como extracción de información, sistemas de pregunta y respuesta, generación de resúmenes de múltiples documentos, detección de plagio, etcétera.

Considere las oraciones E₁: Carlos aprecia la comida francesa; E₂: A Carlos le gusta la cocina francesa y E₃: Carlos aprecia la comida francesa picante. A pesar de que E₁ y E₂ son oraciones compuestas por distintas palabras la idea que

expresan es la misma y por lo tanto se deben considerar como paráfrasis. Por su parte, aunque E₂ y E₃ transmiten la idea de que a Carlos le gusta la comida francesa, no se pueden considerar como paráfrasis ya que E₃ es más específica que E₂.

De forma general el procesamiento de paráfrasis puede ser dividido en tres grandes tareas: i) extracción, ii) generación y iii) reconocimiento. La extracción tiene como objetivo obtener el conjunto más grande posible de pares de expresiones que conforman paráfrasis; esto se realiza a partir de un corpus de referencia. La generación, tiene como objetivo generar el mayor conjunto de expresiones en lenguaje natural que sean paráfrasis de la expresión objeto de análisis. Las expresiones generadas deben tener los menos errores posibles. Finalmente, el reconocimiento tiene por objetivo determinar si dos expresiones de entrada son o no paráfrasis (Calvo, Segura-Olivares y García 2014). Esta tarea se requiere a su vez para abordar otras, como la detección de plagio (Sánchez-Vega y col. 2013).

Una tarea relacionada es el reconocimiento de implicación textual. Ésta consiste en identificar si un determinado texto, denominado hipótesis, se implica o se puede inferir a partir de otro texto. Existen trabajos que se basan en enfoques de reconocimiento léxicos, sintácticos y semánticos, los cuales se complementan con diversas técnicas como lematización, eliminación de palabras sin contenido, manejo de negación, relaciones semánticas y semejanza entre palabras (Segura-Olivares, García y Calvo 2013). Estos métodos se basan en medir la razón o porcentaje de cobertura de la hipótesis con respecto al texto dado. Para efectos de evaluación se utiliza el marco de referencia PASCAL de reconocimiento de implicación textual (Giampiccolo y col. 2007).

Referencias

- Calvo, Hiram (2008). “Augmenting word space models for Word Sense Discrimination using an automatic thesaurus”. En: *International Conference on Natural Language Processing*. Springer, págs. 100-107.
- (2015). “Opinion analysis in social networks using antonym concepts on graphs”. En: *International Conference on Future Data and Security Engineering*. Springer, págs. 109-120.
- Calvo, Hiram y Alexander Gelbukh (2004). “Acquiring selectional preferences from untagged text for prepositional phrase attachment disambiguation”. En: *International Conference on Application of Natural Language to Information Systems*. Springer, págs. 207-216.
- Calvo, Hiram, Alexander Gelbukh y Adam Kilgarriff (2005). “Distributional thesaurus versus WordNet: A comparison of backoff techniques for unsupervised PP attachment”. En: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, págs. 177-188.
- Calvo, Hiram, Andrea Segura-Olivares y Alejandro García (2014). “Dependency vs. constituent based syntactic n-grams in text similarity measures for paraphrase recognition”. En: *Computación y Sistemas* 18.3, págs. 517-554.
- Cárcamo, Javier Tejada, Alexander Gelbukh e Hiram Calvo (2008). “An innovative two-stage WSD unsupervised method”. En: *Procesamiento del lenguaje natural* 40, págs. 99-105.
- Coyotl-Morales, Rosa María y col. (2006). “Authorship attribution using word sequences”. En: *Iberoamerican Congress on Pattern Recognition*. Springer, págs. 844-853.

- Cuétara, Javier (2004). “Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla”. Tesis doct. Tesis para obtener el título de Maestro en Lingüística Hispánica.
- Díaz, Isabel y col. (2005). “Integrating natural language techniques in OO-method”. En: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, págs. 560-571.
- Escalante, Hugo Jair, Thamar Solorio y Manuel Montes-y-Gómez (2011). “Local histograms of character n-grams for authorship attribution”. En: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, págs. 288-298.
- Ferguson, George y James F Allen (1998). “TRIPS: An integrated intelligent problem-solving assistant”. En: *AAAI/IAAI*, págs. 567-572.
- Fusilier, Donato Hernández y col. (2015). “Detecting positive and negative deceptive opinions using PU-learning”. En: *Information processing & management* 51.4, págs. 433-443.
- Giampiccolo, Danilo y col. (2007). “The third pascal recognizing textual entailment challenge”. En: *Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing*. Association for Computational Linguistics, págs. 1-9.
- Hernández, S David e Hiram Calvo (2014). “CoNLL 2014 shared task: Grammatical error correction with a syntactic n-gram language model from a big corpora”. En: *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, págs. 53-59.
- Hernández-Castañeda, Ángel y col. (2017). “Cross-domain deception detection using support vector networks”. En: *Soft Computing* 21.3, págs. 585-595.

- Hu, Minqing y Bing Liu (2004). "Mining and summarizing customer reviews". En: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, págs. 168-177.
- Kilgarriff, Adam (2003). "Thesauruses for natural language processing". En: *International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003*. IEEE, págs. 5-13.
- Kilgarriff, Adam y col. (2004). "Itri-04-08 the sketch engine". En: *Information Technology* 105, pág. 116.
- López-Monroy, A Pastor y col. (2015). "Discriminative subprofile-specific representations for author profiling in social media". En: *Knowledge-Based Systems* 89, págs. 134-147.
- Montes, Manuel, Luis Villaseñor-Pineda y Aurelio López-López (2008). "Mexican experience in spanish question answering". En: *Computación y Sistemas* 12.1, págs. 40-64.
- Navarro, Borja, Rubén Izquierdo y Maximiliano Saiz-Noeda (2004). "Exploiting semantic information for manual anaphoric annotation in Cast3LB corpus". En: *Proceedings of the 2004 ACL Workshop on Discourse Annotation*. Association for Computational Linguistics, págs. 65-71.
- Ortega, Rosa María y col. (2011). "Hacia la identificación de relaciones de hiponimia/hiperonimia en Internet". En: *Revista signos* 44.75, págs. 68-84.
- Ortega-Mendoza, Rosa M, Luis Villaseñor-Pineda y Manuel Montes-y-Gómez (2007). "Using lexical patterns for extracting hyponyms from the web". En: *Mexican International Conference on Artificial Intelligence*. Springer, págs. 904-911.
- Pineda, Luis A, Hayde Castellanos y col. (2010). "The Corpus DIMEx100: transcription and evaluation". En: *Language Resources and Evaluation* 44.4, págs. 347-370.

- Pineda, Luis A, Lisset Salinas y col. (2013). "Sitlog: a programming language for service robot tasks". En: *International Journal of Advanced Robotic Systems* 10.10, pág. 358.
- Pineda, Luis, Hayde Castellanos y col. (2006). "Balancing transactions in practical dialogues". En: *CICLing, LNCS. Springer Verlag*.
- Pineda, Luis e Ivan Meza (2005). "The Spanish pronominal clitic system." En: *Procesamiento del lenguaje natural* 34.
- Ramírez-De-La-Rosa, Gabriela y col. (2013). "A document is known by the company it keeps: neighborhood consensus for short text categorization". En: *Language resources and evaluation* 47.1, págs. 127-149.
- Reyes, Antonio y Paolo Rosso (2012). "Making objective decisions from subjective data: Detecting irony in customer reviews". En: *Decision support systems* 53.4, págs. 754-760.
- Reyes, Antonio, Paolo Rosso y Tony Veale (2013). "A multidimensional approach for detecting irony in twitter". En: *Language resources and evaluation* 47.1, págs. 239-268.
- Rogers, Frank B (1963). "Medical subject headings." En: *Bulletin of the Medical Library Association* 51.1, págs. 114-116.
- Roget, Peter Mark y Robert A Dutch (1962). *The Original Roget's Thesaurus of English Words and Phrase*. New York: Longmans, Green & Co./Dell Publishing Co., Inc.
- Rosso, Paolo y col. (2005). "Two web-based approaches for noun sense disambiguation". En: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, págs. 267-279.

- Sánchez-Junquera, Javier y col. (2018). "Character N-Grams for Detecting Deceptive Controversial Opinions". En: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, págs. 135-140.
- Sánchez-Vega, Fernando y col. (2013). "Determining and characterizing the reused text for plagiarism detection". En: *Expert Systems with Applications* 40.5, págs. 1804-1813.
- Segura-Olivares, Andrea, Alejandro García e Hiram Calvo (2013). "Feature Analysis for Paraphrase Recognition and Textual Entailment." En: *Research in Computing Science* 70, págs. 119-144.
- Sidorov, Grigori (2013). "N-gramas sintácticos no-continuos". En: *Polibits* 48, págs. 69-78.
- Sidorov, Grigori y col. (2013). "Rule-based system for automatic grammar correction using syntactic n-grams for english language learning (l2)". En: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, págs. 96-101.
- Strapparava, Carlo, Alessandro Valitutti y col. (2004). "Wordnet affect: an affective extension of wordnet." En: *LREC*. Vol. 4. 1083-1086, pág. 40.
- Taylor, Ann, Mitchell Marcus y Beatrice Santorini (2003). "The Penn treebank: an overview". En: *Treebanks*. Springer, págs. 5-22.
- Tejada-Cárcamo, Javier, Hiram Calvo y Alexander Gelbukh (2008). "Improving unsupervised WSD with a dynamic thesaurus". En: *International Conference on Text, Speech and Dialogue*. Springer, págs. 201-210.
- Tejada-Cárcamo, Javier, Hiram Calvo, Alexander Gelbukh y Kazuo Hara (2010). "Unsupervised WSD by finding the predominant sense using context as a dynamic thesaurus". En: *Journal of Computer Science and Technology* 25.5, págs. 1030-1039.

- Téllez-Valero, Alberto y col. (2011). "Learning to select the correct answer in multi-stream question answering". En: *Information Processing & Management* 47.6, págs. 856-869.
- Villatoro-Tello, Esaú y col. (2010). "A probabilistic method for ranking refinement in geographic information retrieval". En: *Procesamiento del lenguaje natural* 44, págs. 123-130.

Capítulo 4

Identificación de rasgos de personalidad a través del lenguaje escrito

Gabriela Ramírez de la Rosa, UAM-Cuajimalpa.

Esaú Villatoro Tello, UAM-Cuajimalpa.

4.1. Introducción

Actualmente, el uso cotidiano de servicios y aplicaciones de Internet genera una gran cantidad de información inexplorada sobre nuestro comportamiento. En su mayoría, dicha información está compuesta por documentos como son libros, artículos, correos electrónicos, revisiones de productos en línea, opiniones generadas en blogs y video-blogs (conocidos también como vlogs) e incluso información más detallada de los usuarios, como son el tipo y forma de interacciones en redes sociales, etc. Gracias a esta inmensa acumulación de información se ha vuelto posible, a través de técnicas de Procesamiento de Lenguaje Natural

(PLN), realizar de forma automática análisis e inferencias más detallados que involucran directamente características inherentes de los usuarios de internet.

Por ejemplo, mediante el texto que los usuarios comparten en redes sociales o revisiones de productos en sitios como Amazon es posible determinar la polaridad de sus opiniones (negativa/positiva/neutral) (Leon-Martagón y col. 2013), los sentimientos involucrados (Pang y Lee 2008), identificación el estado de ánimo de las personas (Golder y Macy 2011), predecir las fluctuaciones en la bolsa de valores (Bollen, Mao y Zeng 2011), identificar a pedófilos en chat-rooms (Escalante, Villatoro-Tello, Juárez y col. 2013), así como la obtención de información relacionada a la salud mental de los usuarios (Villatoro-Tello, Ramírez-de-la-Rosa y Jiménez-Salazar 2017), entre muchos otros.

Específicamente, el perfilado de autor, una sub-disciplina del PLN, se interesa en el problema de identificar a través de analizar el texto que escribe un usuario, características demográficas del autor de ese texto, por ejemplo: genero, edad, lenguaje nativo, preferencias políticas o religiosas, etc. Sin embargo, existen otros aspectos demográficos que son de interés no solo a la comunidad de computación, sino también a áreas de las Ciencias Sociales, particularmente a la Psicología; por ejemplo: la identificación de rasgos de personalidad, depresión, anorexia, etc., mismos que se consideran una dimensión más al problema de perfilado de autor (Losada, Crestani y Parapar 2017). En las siguientes secciones se describe con más detalle el problema del perfilado de autor, y específicamente describiremos la metodología adoptada para el diseño de un sistema computacional orientado a la identificación de rasgos de personalidad a partir de un texto producido por el usuario.

4.1.1. Perfilado de autores

El perfilado de autor (PA) es uno de los retos recientes que ha llamado la atención de la comunidad científica, en particular de áreas como el procesamiento de lenguaje natural, ciencias forenses, estrategias de marketing y seguridad en internet. El objetivo principal del perfilado de autor es distinguir una característica o conjunto de características demográficas del autor de un texto; no tiene que ver con identificar al autor específico de dicho texto, este último escenario corresponde al problema conocido como atribución de autoría (Stamatatos 2009). Así entonces, la tarea de perfilado de autor busca modelar a través de atributos sociolingüísticos más generales a grupos de autores, dichos atributos son además indicadores de cómo los distintos grupos de autores emplean el lenguaje dependiendo de su género, edad y/o lenguaje nativo (Argamon y col. 2003).

Uno de los primeros trabajos que enfrentaron el problema de PA fueron los propuestos en (Argamon y col. 2003; Koppel, Argamon y Shimoni 2002), donde se mostró, a través de técnicas estadísticas, que el análisis sobre el uso de las palabras en distintos documentos permite determinar el género, edad, idioma nativo e incluso la personalidad del autor. A partir de entonces, muchos trabajos se han propuesto resolver el problema de perfilado de autores, ejemplos de estas investigaciones son (Schler y col. 2006; Burger y col. 2011; Peersman, Daelemans y Van Vaerenbergh 2011; Nguyen y col. 2013; López-Monroy y col. 2015; Escalante, Villatoro-Tello, Garza y col. 2017). En varios de estos trabajos se ha enfatizado el uso y análisis de representaciones textuales, las cuales han mostrado ser bastante eficientes cuando los documentos que se quieren clasificar son escritos formales (por ejemplo: artículos de noticias, libros, etc.). Sin embargo, cuando se trata de textos informales (por ejemplo: blogs, chats, tuits), las representaciones tradicionales tienen problemas determinando el perfil de los autores. Esto

se debe en gran parte a la dificultad que representa analizar textos informales, los cuales contienen frecuentemente muchos errores ortográficos, variado uso de jerga específica de los medios sociales, así como el uso excesivo de emoticonos y palabras fuera de vocabulario (Sepúlveda-Barrera y col. 2016). A partir de esto, surge la necesidad de proponer formas de representación más robustas, las cuales tienen como principal objetivo capturar tanto elementos léxicos y estilísticos de los documentos, como información extraída de distintas modalidades (e.g., audio, vídeo, imágenes), de forma que ayuden a resolver el problema de perfilado de autores. A este tipo de representaciones, conocidas como multi-modales, en años recientes se les ha empezado a considerar en el PA (Alvarez-Carmona y col. 2018), especialmente en problemas que involucran aspectos subconscientes de los seres humanos como es la identificación automática de su personalidad.

4.2. Definición del problema

La personalidad se refiere a los patrones distintivos de comportamiento, incluyendo emociones y pensamientos. Estos patrones caracterizan la forma en que cada individuo se adapta a una situación particular (Mischel, Shoda y Ayduk 2007; Funder 2001). Usualmente, estos patrones de comportamiento se mantienen estables independientemente de la situación en la que un individuo se encuentre. Desde el área de la psicología, estudiar e identificar la personalidad de un individuo puede ayudar en la detección de aspectos como el nivel de bienestar de las personas analizadas. Por ejemplo, se ha encontrado que la personalidad tiene un impacto en el estudio de la salud física y mental de los sujetos; ayudando en la prevención de conductas o enfermedades mentales, o la identificación de trastornos de la personalidad detonadas por algún evento inesperado

(Ozer y Benet-Martinez 2006). Además de esto, existen otros trabajos que han mostrado que existe una correlación entre la personalidad de un individuo y la forma en que se adapta a la tecnología (Svendsen y col. 2013), diferentes estrategias de aprendizaje (Pavalache-Ilie y Cocorada 2014). Por lo tanto, conocer nuestra personalidad puede beneficiar directamente nuestro nivel de bienestar en diferentes aspectos de la vida cotidiana.

Sin embargo, identificar la personalidad de un individuo no es una tarea sencilla, barata ni rápida. Tradicionalmente esta tarea es realizada por psicólogos expertos que utilizan instrumentos como auto-reportes o métodos de observación para tal efecto. De esta forma, para contar con sistemas computacionales que aprovechen esta información, se vuelve necesario construir métodos automáticos para identificar la personalidad de un individuo, y en el caso ideal, apoyar de manera confiable a la decisión del experto al estudiar otros aspectos del comportamiento y salud mental de cada persona.

Algunos trabajos han mostrado que es posible, hasta cierto punto, identificar la personalidad de los sujetos por medio de conocer las actividades que éstos realizan (Matthews, Deary y Whiteman 2003). Por ejemplo: la forma de presentarse ante otros (Batinca y col. 2011), su comportamiento en reuniones informales (Kalimeri 2013), la forma de expresarse verbal y de forma escrita (Mairesse y col. 2007; Pennebaker, Francis y Booth 2001), el uso de sus redes sociales (Adali y Golbeck 2012; Ortigosa, Carro y Quiroga 2014), entre otras. Sin embargo, poder contar con sistemas confiables que requieran poca información en un ambiente poco o nada controlado es cada vez más deseable. En este sentido, la hipótesis general del trabajo que hemos desarrollado durante los últimos años establece que a través de un ejercicio de escritura a mano de un texto libre, es posible identificar la personalidad del autor de dicho texto. Esto a través de ana-

lizar dos de las actividades más comunes realizadas por los humanos como son: el uso del lenguaje y la forma de escritura a mano.

4.2.1. El modelo de personalidad *Big Five*

Los modelos psicológicos más eficientes para medir los aspectos en la vida de las personas están basados en rasgos (Goldberg 1993). Estos rasgos de personalidad son disposiciones internas que se manifiestan, de cierta forma, en los procesos de pensar, sentir, o actuar frente a situaciones específicas con resultados esperados. Uno de los modelos basados en rasgos es conocido como *Big Five* (BF) o FFM (Five-Factor Model). El Big Five es el paradigma dominante en la investigación de la personalidad, además es uno de los modelos que más ha influenciado la psicología actualmente (McCrae 2002).

El modelo Big Five consta de cinco rasgos o factores cada uno con dos polaridades (positiva y negativa):

- Extroversión (Extroversion) es asociado con la energía, emociones positivas, asertividad, sociabilidad y expresividad; su polo negativo es introversión.
- Neuroticismo (Neuroticism) es la tendencia a experimentar emociones no placenteras como enojo, ansiedad, depresión o vulnerabilidad. A veces, es referido como su otro: Estabilidad emocional que está asociado con controlar el impulso.
- Amabilidad (Agreeableness) se refiere a la tendencia de ser compasivo y cooperativo. En su polo negativo hace referencias a la desconfianza y apatía hacia los demás.

- Responsabilidad (Conscientiousness) es la tendencia a mostrar autodisciplina, actuar de forma leal, alcanzar objetivos y a ser planificador, organizado y confiable. Su polo negativo refiere a comportamientos espontáneos.
- Apertura (Openness to experience) es asociado con la apreciación a ideas inusuales, personas imaginativas y curiosas. El polo negativo de este rasgo está asociado a seres inflexibles al cambio y poco imaginativos.

Usualmente la personalidad de un sujeto, definida por estos cinco factores, se mide mediante la aplicación de baterías (cuestionarios) estandarizadas (Batería BFQ) (McCrae y Costa Jr 1997). Este tipo de instrumentos son denominados auto-reportes, los cuales siguieren al individuo indicar su grado de acuerdo o desacuerdo a ciertas afirmaciones. Por lo regular, estos cuestionarios son costosos de obtener y en general no son reutilizables. Algunos de los cuestionarios más populares para esta tarea son el NEO-Personality-Inventory Revised (NEO-PI-R con 240 preguntas), el NEO Five Factor Inventory (NEO-FFI con 60 preguntas) y el Big-Five Inventory (BFI con 44 preguntas).

Enfoques más recientes proponen la realización de ejercicios de escritura en los cuales el sujeto describe algún evento particular (i.e., un texto sobre un tema específico). A través de este escrito se intenta determinar algunos rasgos de personalidad (Mairesse y col. 2007). En ambos casos, es necesario que psicólogos expertos evalúen y analicen los instrumentos (i.e., baterías y ensayos) para determinar los factores de cada rasgos de cada sujeto. Aunque hasta el momento, este proceso es el más efectivo para identificar la personalidad es un proceso lento y costoso, tanto para los psicólogos que evalúan como para los sujetos que son evaluados.

4.3. La personalidad de estudiantes mexicanos

Con la finalidad de proponer métodos automáticos para la identificación de la personalidad, distintos grupos de investigación se han dado a la tarea de recolectar sus propios conjuntos de datos. Algunos de los corpus que han sido bien aceptados por la comunidad científica son: *i)* Essays corpus, el cual es un corpus de 2479 ensayos recolectado por Pennebaker (Pennebaker y King 1999) y Mairesse (Mairesse y col. 2007), *ii)* my Personality corpus el cual es una colección de posts de Facebook, y, *iii)* el corpus del PAN-AP-15, el cual es un corpus recolectado en el marco de la competencia de PA (Rangel Pardo y col. 2015) desarrollada en PAN-15¹, donde datos de varios usuarios de Twitter fue proporcionada en varios idiomas, incluyendo Español. Estos corpus tiene diferencias importantes en sus características, por un lado *Essays* es un corpus con textos extensos por cada sujeto, mientras que *my Personality* y PAN-AP-15 son ejemplos de textos cortos recolectados masivamente de redes sociales.

Sin embargo, a pesar de que el corpus PAN-AP-15 tiene una pequeña partición de datos es español, es parte de nuestros objetivos de investigación conocer la información relevante sobre la norma culta léxica de una muestra de la comunidad en México y observar las variaciones léxicas que podrían posibilitar la elección, así como la selección, de ciertos patrones léxicos, sintácticos, e incluso estilísticos para la identificación de la personalidad a través de métodos automáticos. A partir de esta motivación, nos dimos a la tarea de recolectar un corpus representativo del Español de México para la identificación automática de personalidad.

¹<https://pan.webis.de/clef15/pan15-web/author-profiling.html>

4.3.1. Metodología para la construcción del corpus TxPI-u

El corpus TxPI-u (Text for Personality Identification of Undergraduates, por sus siglas en Inglés) tiene como principal finalidad servir como un recurso lingüístico que permita a la comunidad científica de PLN, proponer y evaluar métodos automáticos de identificación de personalidad. El corpus está conformado de textos escritos en Español por alumnos de nivel universitario. En lo que resta de la sección se describirá brevemente la metodología empleada para su recolección.

Para la conformación de la muestra se seleccionó a la Universidad Autónoma Metropolitana Unidad Cuajimalpa (UAM-C). Los alumnos invitados a participar en el proyecto fueron los alumnos de nuevo ingreso. Específicamente, durante el 2016, la UAM-C recibió al rededor de 600 alumnos de 10 distintos programas académicos. Durante las primeras semanas de clases, se visitó a los distintos grupos, se les comentó el objetivo del proyecto, y se les invitó a participar. Un total de 417 alumnos decidieron participar en el proyecto, la tabla 4.1 muestra la distribución de los alumnos de acuerdo al programa académico en el que están inscritos.

Nótese que el conjunto está balanceado en términos de género, y hay participantes de todos los programas académicos ofrecidos por la UAM-C. Respecto a la edad de los participantes, debido a que todos son alumnos de nuevo ingreso, la edad de los participantes es de 19 a 21 años.

Para la construcción del corpus TxPI-u, cada alumno respondió un instrumento el cual tiene dos objetivos principales: *i)* determinar la personalidad del sujeto de acuerdo al modelo de personalidad Big Five, y *ii)* recolectar una muestra de texto escrito del sujeto en la cual expresara detalles de sus experiencias

Tabla 4.1: Participantes que ingresaron en 2016 a la UAM-C, divididos por género, y programa académico

Carrera	Masculino	Femenino	Total
Administración	6	17	23
Humanidades	19	25	44
Estudios Socio-territoriales	12	12	24
Ciencias de la Comunicación	30	30	60
Diseño	15	37	52
Tecnologías y Sistemas de Información	28	16	44
Ing. en Computación	45	15	60
Matemáticas aplicadas	14	5	19
Ing. Biológica	20	23	43
Biología Molecular	19	29	48
Total	208	209	417

personales más relevantes. Así entonces, el instrumento diseñado contiene tres secciones principales:

1. Información general. En esta sección se le pidió a cada sujeto información como nombre, edad, género, correo electrónico, carrera, nombres de usuarios de redes sociales como Facebook y Twitter.
2. Cuestionario de personalidad. En esta sección se encuentra el *test* de personalidad, en específico se utilizó el TIPI (Ten Item Personality Inventory) (Gosling, Rentfrow y Swann Jr 2003) en su versión ajustada al Español de Latinoamérica (Renau y col. 2013).

3. Ensayo. Esta sección solicitó a los participantes redactar en un texto libre una historia describiendo algún evento que ellos consideraran relevante en su vida.

4.3.2. Descripción de los datos

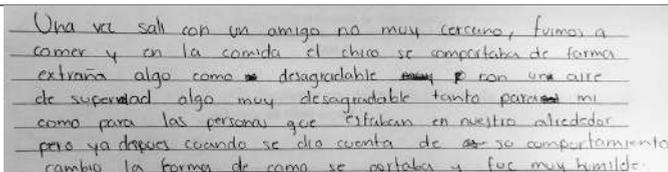
Debido a que el ensayo que los sujetos proporcionaron fue escrito a mano, hubo un proceso de transcripción de la información durante el cual se observaron distintos fenómenos de escritura, por ejemplo, modificaciones al texto, faltas de ortografía, inserción de palabras o letras, dibujos y/o silabificación. Como una hipótesis adicional, creemos que la recurrencia de estos fenómenos de escritura podrían estar relacionados con uno o varios rasgos de personalidad.

A partir de estas observaciones, se definieron un total de siete etiquetas para nombrar a los distintos fenómenos de escritura: <FO:palabra correcta>(error ortográfico), <D:descripción>(dibujo), <IN>(inserción de una letra o palabra en el flujo del texto), <MD>(modificación de una palabra, una corrección sobre la palabra misma), <DL>(eliminación de una palabra), <NS>(cuando dos palabras son escritas juntas; e.g. *Yosoy* en lugar de *Yo soy*) y, SB (silabificación). Un ejemplo de cómo se realizó el etiquetado se muestra en la tabla 4.2.

Se realizó un análisis de correlación de Pearson del porcentaje de aparición de las etiquetas en los ensayos, los cuáles pueden verse en la tabla 4.3. A pesar de los valores de correlación bajos, es importante resaltar que las etiquetas más correlacionadas son *NS* y *FO* con 0.08, y *NS* con *MD* con 0.07.

Para la segunda parte del instrumento, i.e., el test de personalidad, se obtuvo el valor numérico correspondiente a cada rasgo de acuerdo a las normas del TIPI, el cual determina un valor entre 1 y 7 para cada rasgo. La figura 4.1 muestra la distribución de los valores obtenidos para los 5 rasgos.

Tabla 4.2: Ejemplo de un ensayo escrito. Debajo de la figura se muestra la correspondiente transcripción con sus respectivas etiquetas de los distintos fenómenos de escritura.



Una vez salí con un amigo no muy cercano, fuimos a comer y en la comida el chico se comportaba de forma extraña algo como desagradable con un aire de superioridad algo muy desagradable tanto para mi como para las personas que estaban en nuestro alrededor pero ya después cuando se dio cuenta de su comportamiento cambio la forma de como se portaba y fue muy humilde.

Transcripción
manual

Una vez sali <FO:salí> con un amigo no muy cercano, fuimos a comer y en la comida el chico se comportaba de forma extraña algo como <DL> desagradable <DL> <DL> con un <MD> aire de superioridad <MD> algo muy desagradable tanto para <DL> mi <FO:mí> como para las personas que estaban en nuestro alrededor pero ya despues <FO:después> cuando se dio cuenta de <DL> su comportamiento cambio <FO:cambió> la forma de como <FO:cómo> se portaba y fue muy humilde.

Con la finalidad de determinar si existe alguna correlación entre los rasgos de personalidad reportados por los sujetos, en la tabla 4.4 se muestran los valores de la correlación de Pearson. De esta tabla podemos observar que los rasgos con una mayor correlación son: estabilidad emocional y amabilidad (con un valor de correlación de 0.34). De manera similar estabilidad emocional (Sta) aparece correlacionado positivamente con responsabilidad (Con). A pesar que los valores de correlación son relativamente bajos, podemos decir que los sujetos emocionalmente estables son, hasta cierto punto, personas amables y responsables. Además de las anteriores, existe una correlación entre apertura (Ope) y extro-

Tabla 4.3: Correlación entre etiquetas de fenómenos de escritura a mano; donde las etiquetas FO, D, IN, MD, DL and NS significan error ortográfico, dibujo, inserción de letra, modificación de alguna palabra, eliminación de palabra, no separa dos palabras, y silabificación, respectivamente

	FO	D	IN	MD	DL	NS
D	-0.04					
IN	-0.01	-0.01				
MD	0.00	0.01	0.00			
DL	0.02	-0.02	-0.03	0.05		
NS	0.08	-0.03	-0.03	0.07	0.06	
SB	-0.04	-0.02	-0.03	0.02	-0.06	0.00

versión (Ext), sugiriendo que los sujetos que son abiertos a nuevas experiencias son también personas extrovertidas.

Por otro lado, la normativa definida por el instrumento TIPI (Gosling, Rentfrow y Swann Jr 2003) utilizado clasifica a cada rasgo en 4 categorías: alto, medio alto, medio bajo y bajo. De acuerdo a esta normativa, los 417 sujetos participantes en el estudio fueron clasificados en una categoría por cada rasgo. La tabla 4.5 muestra el número de sujetos en cada categoría de cada rasgo de personalidad.

4.3.3. Partición estratificada

En su forma original, el corpus TxPI-u es útil para realizar el análisis de los rasgos de personalidad de forma independiente, sin embargo, dificulta el análisis combinado de más de un rasgo. Es por esta razón que se decidió hacer una

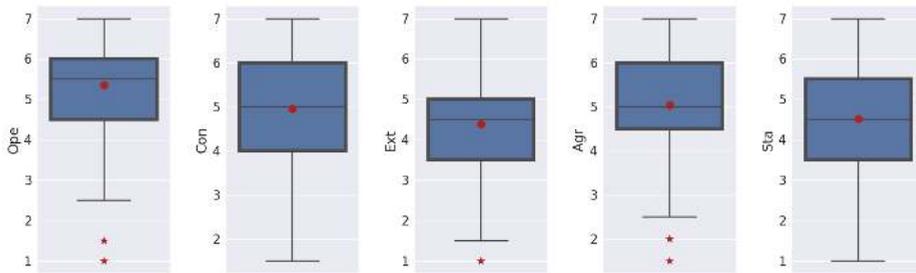


Figura 4.1: Distribución de los valores numéricos para cada rasgo de acuerdo a las respuestas proporcionadas por los sujetos en el test TIPI. El valor máximo es de 7, mientras que el mínimo es 1. El valor medio aparece dentro de la caja representado por un círculo rojo.

Tabla 4.4: Valores de correlación entre los diferentes rasgos de personalidad de acuerdo a las respuestas proporcionadas en el test TIPI. Ext, Agr, Con, Sta y Ope significan Estroversión, Amabilidad, Responsabilidad, Estabilidad Emocional (Neuroticismo) y Apertura respectivamente

	Ext	Agr	Con	Sta	Ope
Ext	1.00				
Agr	-0.06	1.00			
Con	0.08	0.23	1.00		
Sta	0.09	0.34	0.28	1.00	
Ope	0.27	0.15	0.16	0.07	1.00

partición estratificada del corpus de forma que cada rasgo tenga un conjunto de ejemplos significativos tanto del polo positivo (Alto) como del polo negativo (Bajo), así como un conjunto de ejemplos de control.

Tabla 4.5: Número de sujetos por clase de cada rasgo de acuerdo a la normativa del TIPI.

Rasgo	Alto	Medio	Medio	Bajo
		Alto	Bajo	
Apertura	91	145	116	64
Responsabilidad	19	150	138	109
Extraversión	72	137	169	38
Amabilidad	60	115	151	90
Estabilidad Emocional	34	151	151	80

En la partición estratificada del corpus TxPI-u, la muestra de control contiene a todos los sujetos cuyos valores para todos los rasgos fueron “medio-alto” o “medio-bajo”. En otras palabras, todos los sujetos en el grupo de control no tienen ningún rasgo predominante (“alto” o “bajo”).

Los sujetos con rasgos representativos, i.e., sujetos etiquetados con las clases “alto” o “bajo” en solo uno de los rasgos son seleccionados para la partición estratificada. La tabla 4.6 muestra el número de hombres y mujeres presentes en cada uno de los rasgos, así como el número de sujetos en el grupo de control. Como se puede observar, esta partición resulta en un corpus reducido (apenas la mitad del corpus original), sin embargo permite realizar un análisis más fino de las diferencias presentes en los cinco rasgos.

Finalmente, se hizo un análisis respecto a la presencia de los distintos fenómenos de escritura dentro de la partición estratificada. La tabla 4.7 muestra el

Tabla 4.6: Número de sujetos por género y número de sujetos por clase, por rasgo, en la partición estratificada.

Rasgo	Género		Clases		Total
	Hombre	Mujer	Alto	Bajo	
Apertura	18	14	16	16	32
Responsabilidad	11	17	3	25	28
Extraversión	15	10	17	8	25
Amabilidad	12	20	10	22	32
Estabilidad Emocional	7	7	6	8	14
Control	44	39	-	-	83
Total	107	107	52	79	214

porcentaje de aparición de estos fenómenos. Es importante resaltar que la etiqueta FO (error ortográfico) es la que predomina en los datos.

La figura 4.2 muestra un análisis más detallado sobre la distribución de la etiqueta FO en los datos. Note la diferencia del porcentaje de la etiqueta FO entre los polos (positivo y negativo) del mismo rasgo. Por ejemplo, para el rasgo de extraversión, existe casi el mismo porcentaje de errores ortográficos para los sujetos en el polo negativo (“low”), mientras que hay una mayor distribución para los que están en el polo positivo (“high”). Esto puede ser un indicador de una posible correlación entre la etiqueta FO con la personalidad predominante de un sujeto.

Tabla 4.7: Porcentaje de etiquetas presentes en los ensayos de la partición estratificada del corpus TxPI-u.

Etiqueta	media	dev-estd	min	max
FO	1.46	1.71	0	9.68
D	0.01	0.14	0	1.92
IN	0.01	0.09	0	1.22
MD	0.99	1.99	0	20.64
DL	0.22	0.72	0	6.33
NS	0.42	0.96	0	6.04
SB	0.14	0.58	0	5.56

4.4. Desempeño de sistemas automáticos

El objetivo principal de esta sección es la de proveer un panorama del desempeño de muchas de las técnicas tradicionales de clasificación de textos en el problema de identificación automática de la personalidad. Para los experimentos reportados en esta sección se utilizó la partición estratificada del corpus TxPI-u.

Para realizar los experimentos se adoptaron técnicas tradicionales de clasificación de textos. Los resultados obtenidos tienen la finalidad de servir como un punto de comparación para los métodos de representación más sofisticados que se puedan proponer en un futuro próximo. Como estrategias de representación se emplearon n-gramas de palabras, n-gramas de caracteres, n-gramas de etiquetas de parte de la oración (POS) en combinación con los algoritmos de aprendizaje más populares en el área de clasificación de textos, e.g., naive bayes, máquinas de vectores de soporte, y árboles de decisión.

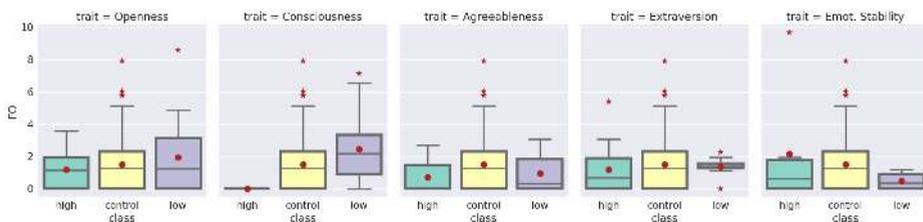


Figura 4.2: Distribución de la etiqueta de error ortográfico (FO) en los cinco rasgos dentro de la partición estratificada del corpus. El punto rojo dentro de las cajas muestra el valor medio de cada muestra y los puntos más allá de los límites de las cajas representan los valores atípicos (“outliers”).

4.4.1. Métricas de evaluación

La métrica de evaluación empleada para reportar los resultados es la medida F-macro F_1 , también conocida como medida F. La medida F permite obtener valores confiables respecto al desempeño del sistema, en particular para casos donde las clases son altamente desbalanceadas, que es el caso del corpus TxPI-u. Finalmente, es conveniente mencionar que para todos los experimentos reportados se utilizó como estrategia de validación una validación cruzada de diez pliegues.

4.4.2. Configuración experimental

Se definieron cinco problemas de clasificación, uno por cada rasgo. Cada problema tiene tres posibles clases: “alto”, “bajo” y “control”. Cada ensayo fue representado utilizando tres diferentes tipos de representación: n-gramas de palabras, n-gramas de caracteres, y n-gramas de etiquetas POS. Para cada tipo de n-grama se utilizaron diferentes tamaños de n : 1, 2 y 3 para palabras, y POS,

mientras que para los n-gramas de caracteres se emplearon tamaños de 3, 4, 5. Agregado a esto, para cada experimento se emplearon tres distintos algoritmos de clasificación: Naive Bayes, J48 y SMO₃.

Un modelo de espacio vectorial fue empleado para representar cada texto; así, para cada ensayo se tiene un vector multi-dimensioal. Para el vector construido se considerando tres esquemas diferentes de pesado: booleano (o de aparición), frecuencia de término tf (número de veces que aparecen los términos del vector), y frecuencia relativa $tf - idf$ (donde la importancia de un término es dada por su frecuencia y su frecuencia inversa).

Agregado a los anterior, se realizó un conjunto de experimentos adicionales empleando una estrategia de representación basada en un vocabulario cerrado. Para estos experimentos se utilizó una representación de tipo bolsa-de-palabras empleando las categorías psico-lingüísticas existentes en la versión en español del recurso LIWC (*Linguistic Inquiry and Word Count* (Pennebaker, Francis y Booth 2001)). La hipótesis de fondo en este conjunto de experimentos establece que por medio de emplear un vocabulario previamente identificado y asociado a distintos procesos cognitivos, el desempeño de un sistema de clasificación será más efectivo.

4.4.3. Resultados

La tabla 4.8 muestra los resultados obtenidos al emplear como forma de representación la bolsa de palabras construida con las categorías de LIWC. Observe que no se reportan los resultados bajo el esquema de pesado $tf - idf$, esto es debido a que el desempeño de éste es muy similar al esquema tf . En lo que respecta al resto de los experimentos, un total de 405 experimentos fueron realizados (5 rasgos, 9 representaciones, 3 esquemas de pesado, y 3 algoritmos de

aprendizaje). En la tabla 4.9 se muestran los resultados de estos experimentos, omitiendo nuevamente esquema de pesado $tf - idf$.

Tabla 4.8: Resultados de clasificación empleando una representación basada en LIWC.

	NB		SMO		J48	
	bool	tf	bool	tf	bool	tf
Ope	0.31	0.33	0.35	0.26	0.34	0.34
Con	0.30	0.36	0.29	0.33	0.26	0.36
Ext	0.31	0.34	0.35	0.34	0.34	0.33
Agr	0.34	0.40	0.31	0.27	0.43	0.36
Sta	0.35	0.40	0.27	0.30	0.31	0.37

En general, el desempeño de los distintos modelos evaluados es relativamente bajo, siendo $F = 0.49$ el valor máximo obtenido, mostrando que la tarea de identificación automática de personalidad es un problema difícil. Sin embargo, a pesar de los resultados, podemos obtener algunas conclusiones importantes. Por ejemplo, la representación basada en etiquetas POS resulta ser la mejor para los rasgos de apertura (Ope) con un $F = 0.49$ para $n = 1$, responsabilidad (Con) con un $F = 0.39$ con $n = 2$, y estabilidad emocional (Sta) con $F = 0.40$ con $n = 1$. Por otro lado, los n-gramas de caracteres son los que permiten identificar en mejor medida los rasgos de extraversión (Ext) y amabilidad (Agr) con un $F = 45$ con $n = 5$ para ambos casos.

Otra observación importante es que usar un esquema de representación basado en vocabulario cerrado no es la mejor estrategia para este problema. En

general, los resultados obtenidos con la representación de n-gramas de etiquetas POS y de caracteres, vocabulario abierto, son mucho mejores que los obtenidos bajo el esquema de vocabulario cerrado (tabla 4.8). Es evidente que aún queda mucho trabajo por hacer de forma que el desempeño de los sistemas automáticos de identificación de personalidad pueda mejorar significativamente. En los experimentos reportados en este trabajo no se utilizó la información de las etiquetas de los fenómenos de escritura. A partir del análisis inicial de correlación de éstas etiquetas con los distintos rasgos, creemos que emplear esta información como atributos adicionales permitirá mejorar el desempeño de los clasificadores.

4.5. Conclusiones y panorama futuro

Este trabajo describe brevemente la compilación de un corpus multi-modal en Español etiquetado con información de género, edad, programa académico, y rasgos de personalidad de una parte de la población estudiantil de la Universidad Autónoma Metropolitana Unidad Cuajimalpa. El corpus, nombrado TxPI-u (Text for Personality Identification of undergraduate) contiene hasta el momento información de 416 sujetos, los cuales además de la información anterior, han redactado un ensayo describiendo lo que consideran experiencias relevantes de su vida.

El proceso de transcripción manual de los ensayos producidos por los sujetos participantes permitió la identificación de distintos fenómenos de escritura, como son: errores ortográficos, inserciones, modificaciones, dibujos, etc. En el análisis realizado no se observó una correlación clara entre los distintos fenómenos de escritura, sin embargo, contar con esta información en el corpus es algo que no se ha visto en ningún otro corpus de personalidad disponible. Además

de esto, el análisis de correlación entre los fenómenos de escritura y los rasgos de personalidad mostró que existe una posible relación entre algunas etiquetas con rasgos particulares, por ejemplo, los sujetos en el polo alto de responsabilidad tienen cero faltas ortográficas. Creemos que incorporar la información de los fenómenos de escritura en la forma de representación permitirá mejorar el desempeño de los clasificadores.

Como una contribución importante debemos mencionar el proceso exhaustivo de evaluación experimental realizado. Para esto, se emplearon técnicas tradicionales de clasificación de textos y de perfilado de autores. Además, se comparó el desempeño de un esquema basado en vocabulario cerrado utilizando la información psico-lingüística del recurso LIWC. En general, los experimentos muestran que una representación basada en vocabulario abierto es mejor. Sin embargo, estos primeros experimentos representan un punto de referencia para trabajos futuros.

Finalmente, el corpus construido representa una contribución importante para la comunidad de Procesamiento de Lenguaje Natural en México realizando investigación en áreas de perfilado de autores, y en especial de identificación de personalidad. El corpus TxPI-u tiene información extra como género, edad, carrera, redes sociales, e incluso la imagen del texto escrito a mano; información que en su conjunto podría ser considerada a través de técnicas de representación multi-modal para incorporarla a los algoritmos de aprendizaje automático.

Como trabajo futuro nos interesa explorar la pertinencia de la información que se pueda extraer de un análisis de elementos que aparecen en la escritura manual, es decir, de los distintos atributos que se puedan extraer de la imagen del texto escrito por los sujetos. Hasta el momento, no se han encontrado trabajos que además del estudio independiente de la escritura a mano se combine con otro tipo de atributos. Es posible que si se combina el análisis de los trazos escritos a mano con el contenido de ese texto se logre obtener un mejor desempeño en la identificación automática de la personalidad.

Referencias

- Adali, Sibel y Jennifer Golbeck (2012). "Predicting personality with social behavior". En: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, págs. 302-309.
- Alvarez-Carmona, Miguel A y col. (2018). "A visual approach for age and gender identification on Twitter". En: *Journal of Intelligent & Fuzzy Systems* 34.5, págs. 3133-3145.
- Argamon, Shlomo y col. (2003). "Gender, genre, and writing style in formal written texts". En: *Text-The Hague Then Amsterdam Then Berlin-* 23.3, págs. 321-346.
- Batrinca, Ligia Maria y col. (2011). "Please, tell me about yourself: automatic personality assessment using short self-presentations". En: *Proceedings of the 13th international conference on multimodal interfaces*. ACM, págs. 255-262.
- Bollen, Johan, Huina Mao y Xiaojun Zeng (2011). "Twitter mood predicts the stock market". En: *Journal of computational science* 2.1, págs. 1-8.
- Burger, John D y col. (2011). "Discriminating gender on Twitter". En: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, págs. 1301-1309.
- Escalante, Hugo Jair, Esaú Villatoro-Tello, Sara E Garza y col. (2017). "Early detection of deception and aggressiveness using profile-based representations". En: *Expert Systems with Applications* 89, págs. 99-111.
- Escalante, Hugo Jair, Esaú Villatoro-Tello, Antonio Juárez y col. (2013). "Sexual predator detection in chats with chained classifiers". En: *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, págs. 46-54.
- Funder, DC (2001). "Personality". En: *Annual Review of Psychology* 52.1, págs. 197-221.

- Goldberg, Lewis R (1993). “The structure of phenotypic personality traits.” En: *American psychologist* 48.1, pág. 26.
- Golder, Scott A y Michael W Macy (2011). “Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures”. En: *Science* 333.6051, págs. 1878-1881.
- Gosling, Samuel D, Peter J Rentfrow y William B Swann Jr (2003). “A very brief measure of the Big-Five personality domains”. En: *Journal of Research in personality* 37.6, págs. 504-528.
- Kalimeri, Kyriaki (2013). “Towards a dynamic view of personality: multimodal classification of personality states in everyday situations”. En: *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, págs. 325-328.
- Koppel, Moshe, Shlomo Argamon y Anat Rachel Shimoni (2002). “Automatically categorizing written texts by author gender”. En: *Literary and linguistic computing* 17.4, págs. 401-412.
- Leon-Martagón, Gilberto y col. (2013). “Análisis de polaridad en twitter”. En: *Journal of Research in Computing Science* 62, págs. 69-78.
- López-Monroy, A Pastor y col. (2015). “Discriminative subprofile-specific representations for author profiling in social media”. En: *Knowledge-Based Systems* 89, págs. 134-147.
- Losada, David E, Fabio Crestani y Javier Parapar (2017). “eRISK 2017: CLEF lab on early risk prediction on the internet: experimental foundations”. En: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, págs. 346-360.

- Mairesse, François y col. (2007). "Using linguistic cues for the automatic recognition of personality in conversation and text". En: *Journal of artificial intelligence research* 30, págs. 457-500.
- Matthews, Gerald, Ian J Deary y Martha C Whiteman (2003). *Personality traits*. Cambridge University Press.
- McCrae, Robert R (2002). "Cross-cultural research on the five-factor model of personality". En: *Online readings in psychology and culture* 4.4.
- McCrae, Robert R y Paul T Costa Jr (1997). "Personality trait structure as a human universal." En: *American psychologist* 52.5, pág. 509.
- Mischel, Walter, Yuichi Shoda y Ozlem Ayduk (2007). *Introduction to personality: Toward an integrative science of the person*. John Wiley & Sons.
- Nguyen, Dong y col. (2013). "'How old do you think I am?': A study of language and age in Twitter". En: *Seventh International AAAI Conference on Weblogs and Social Media*.
- Ortigosa, Alvaro, Rosa M Carro y José Ignacio Quiroga (2014). "Predicting user personality by mining social interactions in Facebook". En: *Journal of computer and System Sciences* 80.1, págs. 57-71.
- Ozer, Daniel J y Veronica Benet-Martinez (2006). "Personality and the prediction of consequential outcomes". En: *Annual Review of Psychology* 57, págs. 401-421.
- Pang, Bo y Lillian Lee (2008). "Opinion mining and sentiment analysis". En: *Foundations and Trends® in Information Retrieval* 2.1-2, págs. 1-135.
- Pavalache-Ilie, Mariela y Sorin Cocorada (2014). "Interactions of students' personality in the online learning environment". En: *Procedia-Social and Behavioral Sciences* 128, págs. 117-122.
- Peersman, Claudia, Walter Daelemans y Leona Van Vaerenbergh (2011). "Predicting age and gender in online social networks". En: *Proceedings of the 3rd*

- international workshop on Search and mining user-generated contents*. ACM, págs. 37-44.
- Pennebaker, James W, Martha E Francis y Roger J Booth (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates.
- Pennebaker, James W y Laura A King (1999). "Linguistic styles: Language use as an individual difference." En: *Journal of personality and social psychology* 77.6, pág. 1296.
- Rangel Pardo, Francisco Manuel y col. (2015). "Overview of the 3rd Author Profiling Task at PAN 2015". En: *CLEF 2015 Evaluation Labs and Workshop Working Notes Papers*, págs. 1-8.
- Renau, Vanessa y col. (2013). "Translation and validation of the ten-item-personality inventory into Spanish and Catalan". En: *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport* 31.2.
- Schler, Jonathan y col. (2006). "Effects of age and gender on blogging." En: *AAAI spring symposium: Computational approaches to analyzing weblogs*. Vol. 6, págs. 199-205.
- Sepúlveda-Barrera, Diana M y col. (2016). "Importancia del lenguaje coloquial y de los símbolos de puntuación en el perfilado de autores". En: *Advances in Natural Language Processing and Computational Linguistics* 115, págs. 43-56.
- Stamatatos, Efstathios (2009). "A survey of modern authorship attribution methods". En: *Journal of the American Society for information Science and Technology* 60.3, págs. 538-556.
- Svendsen, Gunnvald B y col. (2013). "Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model". En: *Behaviour & Information Technology* 32.4, págs. 323-334.

- Villatoro-Tello, Esaú, Gabriela Ramírez-de-la-Rosa y Héctor Jiménez-Salazar (2017).
“UAM’s Participation at CLEF eRisk 2017 task: Towards Modelling Depressed Bloggers.” En: *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation forum, CEUR Workshop proceedings*. Vol. 1866.

Capítulo 5

Detección de engaño en texto

Hiram Calvo, CIC/IPN.

Ángel Hernández Castañeda, UAEMex.

Juan Javier Sánchez Junquera, INAOE.

5.1. Introducción

La detección de texto engañoso mediante algoritmos de aprendizaje automático consiste en identificar las características que definen a estos textos de acuerdo a un conjunto de datos previo llamado conjunto de entrenamiento. De esta manera, los algoritmos pueden analizar un nuevo texto (no contenido en el conjunto de entrenamiento) e identificar si pertenece a alguna clase específica.

Los investigadores han estudiado la detección de engaño mediante el uso de diferentes fuentes de características. Algunos de estos métodos están basados en herramientas de bolsa de palabras (BoW por sus siglas en inglés), en las cuales no se considera el orden de las palabras; otros añaden información sintáctica como características. En algunos casos, se buscan algunas señales generales de engaño

(DePaulo y col. 2003), tales como el uso de palabras únicas, autorreferencias o modificadores, entre otros.

En general, los enfoques de estilo lingüístico (como n-gramas sintácticos, también llamados *NS-GRAMAS*) analizan las relaciones entre las palabras; en cambio, los enfoques de bolsa de palabras (LDA, MTD, LIWC, n-gramas) ignoran la gramática e incluso el orden de las palabras, pero siguen contando el número de instancias de cada palabra.

Para detectar textos engañosos, una técnica comúnmente aplicada es usar n-gramas. Este método puede extraer características de un texto basado en diferentes elementos, por ejemplo palabras, sílabas, fonemas, letras, etc.

En el trabajo de Hernandez-Fusilier et al. (Fusilier y col. 2015) los n-gramas de palabras se compararon con n-gramas de letras. Estos últimos han demostrado tener un mejor rendimiento en el conjunto de datos OpSpam. Aunque los n-gramas logran resultados aceptables por sí mismos, generalmente se complementan con otras técnicas de procesamiento de lenguaje natural (PLN) debido al hecho de que la combinación de características ha demostrado mejorar los resultados.

Otro enfoque de BoW consiste en utilizar el diccionario linguistic inquiry and word count (LIWC), que incluye una clasificación de palabras y una herramienta de recuento. Al analizar las categorías de palabras de LIWC, el equipo de los creadores de LIWC (Newman y col. 2003), encontró que los mentirosos usan menos autoreferencias y usan palabras de emociones negativas. Este trabajo sentó las bases para que la herramienta LIWC fuera ampliamente utilizada por otros investigadores (Schelleman-Offermans y Merckelbach 2010; Toma y Hancock 2012).

Hauch et al. (Hauch y col. 2012) introdujo un meta-análisis de varios trabajos de investigación de identificación de textos engañosos. Este meta-análisis se centró en categorías lingüísticas específicas, por ejemplo, las contenidas en LIWC. Los hallazgos de la investigación sugieren que los mentirosos usan ciertas categorías lingüísticas en una tasa diferente que los que dicen la verdad.

La detección de engaño se ha aplicado en diversas situaciones particulares. En Williams et al. (Williams y col. 2014), se hizo una comparación entre las mentiras dichas por niños y las mentiras dichas por adultos. La investigación se llevó a cabo con el objetivo de detectar el engaño en los tribunales donde los niños testifican. Los autores eligieron a 48 niños y 28 adultos para generar un conjunto de datos; la mitad de los niños y adultos contaron mentiras y la mitad de ellos dijeron la verdad. De esta forma, se utilizó la herramienta LIWC para generar las muestras para la clasificación. Los resultados mostraron que existen diferencias significativas entre los textos verdaderos y los falsos, principalmente en variables lingüísticas tales como las autorreferencias singulares (por ejemplo, *yo, mi, a mí*), plurales (por ejemplo, *nosotros, nuestro, a nosotros*) y las emociones negativas. Además, los resultados de la investigación mostraron que las variables lingüísticas se encontraron en distintas proporciones dependiendo de si la mentira fue contada por un niño o por un adulto.

Los estudios desarrollados mediante el uso de herramientas de bolsa de palabras han tenido éxito; sin embargo, en un esfuerzo por mejorar los resultados, se ha tenido en cuenta el contexto de las oraciones, por ejemplo al analizar las relaciones sintácticas de las palabras con el uso de árboles de dependencias (Xu y Zhao 2012). En general, el uso de las relaciones sintácticas no ha mostrado un rendimiento sobresaliente en la tarea de clasificar el texto engañoso, aunque el

complementar este método con un enfoque de bolsa de palabras puede mejorar los resultados.

En el estudio de Mihalcea y Pérez-Rosas (Pérez-Rosas y Mihalcea 2014b), las características fueron generadas utilizando diferentes enfoques (por ejemplo, etiquetas gramaticales (PoS), gramáticas libres del contexto (CFG), unigramas, LIWC y combinaciones de éstos). Las autoras predijeron, con una exactitud de entre 60 % y 70 %, si una persona de sexo femenino o masculino había escrito un texto engañoso. En los resultados mostrados, el uso de categorías gramaticales y gramáticas libres de contexto no mostró una mejora significativa en la exactitud con respecto a los unigramas y LIWC. Esto sugiere que, al menos para esta tarea, los enfoques de bolsa de palabras pueden tener un desempeño similar a los enfoques de estilo lingüístico.

Para probar la eficiencia de los algoritmos de aprendizaje se recurre a conjuntos de datos llamados corpus (singular) o corpora (plural). Los conjuntos de datos se pueden dividir en dos subconjuntos: el ya mencionado conjunto de entrenamiento que le da el conocimiento previo al clasificador, y el conjunto de prueba, el cual consiste en documentos que se clasificarán con base en el conocimiento previo.

5.2. Exploración del reconocimiento automático del engaño

En el trabajo de investigación de Mihalcea y Strapparava (Mihalcea y Strapparava 2009) se abordó la detección de texto engañoso y texto veraz mediante el enfoque de aprendizaje automático. De esta manera, se pretendían capturar características subyacentes de forma automática en el texto, que permitieran iden-

tificar si un texto es engañoso o veraz. Los autores se basaron solamente en la obtención de características lingüísticas, aunque otros trabajos incluyen características no lingüísticas (por ejemplo, características acústicas), esto debido a que la mayor parte de información que se encuentra en la web son textos.

El objetivo de la investigación fue saber si los textos engañosos y veraces realmente eran separables, es decir, si existen características que solo se pueden encontrar o que se encuentran con mayor probabilidad en los textos engañosos. Para resolver lo anterior, los autores crearon un conjunto de datos de tópicos controvertidos; este conjunto contiene textos cortos engañosos y veraces.

De forma específica el conjunto de datos que generaron consta de tres tópicos controvertidos: el aborto, la pena de muerte y la opinión sobre los mejores amigos. Para poder evaluar el enfoque propuesto, fue necesario que el conjunto de datos fuera etiquetado, es decir, que cada texto tuviera su respectivo identificador para saber si era engañoso o veraz. Para realizar la labor de etiquetado, los autores optaron por usar el servicio del turco mecánico de Amazon, el cual es una herramienta web que permite a diferentes usuarios ayudar en una tarea a cambio de una aportación monetaria. El conjunto se formó mediante dos dinámicas. Primero, en el caso de los tópicos acerca de aborto y pena de muerte la tarea consistió en que los usuarios debieron escribir su opinión real acerca de cada uno de estos dos tópicos; en seguida se les pidió escribir la opinión contraria a su opinión real. De esta forma, el resultado es un texto con información veraz y un texto con información engañosa. En segundo lugar, para el caso del tópico referente a los mejores amigos, se les pidió a los usuarios que pensarán en su mejor amigo y describieran las razones de su amistad; después se les pidió que pensarán en una persona a la que no pudieran soportar y la describieran como si fuera su mejor amigo. Como resultado, en el primer caso (aborto y pena

de muerte) los participantes mienten acerca de sus creencias, mientras que en el segundo caso (mejor amigo) los participantes mienten acerca de sus sentimientos hacia otra persona. Finalmente, fueron recolectados 100 textos veraces y 100 engañosos por tópico, sumando un total de 600 textos.

Con el propósito de obtener las características más relevantes del texto engañoso y del texto veraz, los autores idearon un método para valorar en qué medida las palabras se inclinan hacia una u otra clase. Tal método se basa en las siguientes medidas: la cobertura del texto engañoso (Fórmula 5.1), la cual obtiene la fracción del número de ocurrencias (*tokens*) de la palabra w_i en el corpus de textos engañosos (D) entre el total de tokens del corpus D . De la misma forma, para la cobertura del texto veraz (Fórmula 5.2) se realiza el mismo cálculo, pero esta vez para (T).

$$Cobertura_D(C) = \frac{\sum_{w_i \in C} Frecuencia(w_i)}{Tamaño_D} \quad (5.1)$$

$$Cobertura_T(C) = \frac{\sum_{w_i \in C} Frecuencia(w_i)}{Tamaño_T} \quad (5.2)$$

$$Predominio_D(C) = \frac{Cobertura_D(C)}{Cobertura_T(C)} \quad (5.3)$$

Finalmente, se calcula el valor de predominio (Fórmula 5.3) de la clase C (clases incluidas en LIWC en su versión 2001), la cual es la proporción entre la cobertura del texto veraz y la cobertura del texto engañoso, calculado para la misma clase. Los tres tópicos fueron mezclados para formar dos conjuntos de datos generales: el conjunto T y el conjunto D , de texto veraz y engañoso, respectivamente. Sobre estos conjuntos generales fueron calculados los valores de

Tabla 5.1: Clases de LIWC dominantes del texto engañoso y veraz

Clase	Puntaje	Ejemplo de palabras
Texto engañoso		
metáfora	1.71	dios, muerte, sagrado, compasión, pecado, muerte, infierno, alma, señor.
formalidad	1.53	tú, usted
ref. a otros	1.47	ella, a ella, ellos, su, a ellos, a él, por sí mismo, por sí misma, por ellos mismos
personas	1.31	persona, niño/a, humano(s), bebé, hombre, mujer, individuo(s), persona, adulto
certeza	1.24	siempre, todo(s), muy, realmente, completamente, totalmente
Texto veraz		
optimalidad	0.57	mejor, listo, esperanza, acepta, aceptar, determinado, aceptado, ganado, súper
Autorreferencias	0.59	yo, yo mismo, mi(s)
amigos	0.63	amigo, compañía, grupo
sí mismo	0.64	nuestro(s), a mí, mi
visión	0.65	creo, pienso, sé, veo, entiendo, encuentro, me parece, siento, admito que

Tabla 5.2: Exactitud obtenida por tópico usando una validación cruzada de 10 pliegues

Tópico	Naïve Bayes	SVM
Aborto	70.00 %	67.50 %
Pena de muerte	67.40 %	65.90 %
Mejor amigo	75.00 %	77.00 %
Promedio	70.80 %	70.10 %

predominio de cada clase C . De esta manera, un resultado cercano a 1 indicará que existe una distribución similar de palabras en la clase C tanto de texto veraz como de texto engañoso. Por otra parte, si el resultado es menor a 1, significará que la clase es predominante en el texto veraz, en cambio, si el resultado es superior a 1, significará que la clase es predominante del texto engañoso. En la Tabla 5.1 se muestran algunos valores de clases dominantes obtenidos por los autores.

En los experimentos se clasificaron los textos mediante un clasificador bayesiano (Naïve Bayes) y una máquina de soporte vectorial (SVM, por sus siglas en inglés — *Support Vector Machine*). Primero se clasificaron los tres tópicos por separado; los resultados se muestran en la Tabla 5.2, en la cual se especifican los valores de exactitud obtenidos por cada clasificador; asimismo, se muestra la exactitud promedio de los tres tópicos. Se observa también que el mejor resultado en promedio fue obtenido por el clasificador bayesiano.

Siguiendo el trabajo de Mihalcea y Strapparava, también se realizó un estudio para detectar texto engañoso escrito en español: Almela et al. (Almela, Valencia-García y Cantos 2012) recopiló un nuevo conjunto de datos con temas que abarcan la adopción homosexual, opiniones sobre la tauromaquia y sentimientos acerca de un mejor amigo. Cien documentos falsos y cien verdaderos

Tabla 5.3: Exactitud, precisión (P), exhaustividad (R), y medida F (F) de la identificación de texto engañoso realizada por humanos

Participante	Exactitud	Veraz			Engañoso		
		P	R	F	P	R	F
Juez 1	61.90 %	57.9	87.5	69.7	74.4	36.3	48.7
Juez 2	56.90 %	53.9	95	68.8	78.9	18.8	30.3
Juez 3	53.10 %	52.3	70	59.9	54.7	36.3	43.6

fueron recogidos para cada tema con un promedio de 80 palabras por documento. Se utilizaron diferentes dimensiones de LIWC para lograr una clasificación más precisa por medio de una máquina de soporte vectorial (SVM).

5.3. Enfoques para la identificación del engaño

El objetivo de la investigación de Ott et al. (Ott y col. 2011) fue identificar las opiniones engañosas, que define como opiniones ficticias que han sido deliberadamente escritas para sonar auténticas.

Para lograr el objetivo los autores exploran tres enfoques, el primero es ver el problema como una tarea de categorización de texto mediante el uso de n-gramas y clasificadores, esto con el fin de etiquetar los documentos como veraces o engañosos. El segundo fue optar por un enfoque psicolingüístico, en el cual se pretende verificar cuáles son los estados psicológicos, tales como las emociones negativas, relacionados con el engaño. Por último, en el tercero se abordó la tarea como un problema de género literario, mediante el cual se percibe la escritura de textos engañosos como *imaginativa* y la escritura de textos veraces como *informativa*.

Mediante la clasificación de un conjunto de 400 textos engañosos y 400 textos veraces los autores identificaron que los clasificadores de aprendizaje automático, entrenados con características obtenidas mediante enfoques psicolingüísticos y de identificación de género, sobrepasan estadísticamente el nivel de los enfoques de categorización de texto, tal como los n-gramas. Además, remarcaron que las capacidades humanas para diferenciar entre un texto engañoso y uno veraz, como se muestra en la Tabla 5.3, son limitadas. Esto lo ratificaron mediante una clasificación, hecha por tres participantes, de un subconjunto de 160 documentos. En la tabla también se muestran otras medidas además de la exactitud: precisión (P), exhaustividad (R) y medida F (F).

Los autores también dan importancia a que la identificación de texto tiene que ser afrontada mediante el contexto y la motivación que condujo al engaño, y no solamente mediante la identificación de un conjunto de señales universales de engaño. Por esta razón, proponen generar características basadas en categorías gramaticales (PoS tags, por sus siglas en inglés part of speech tags); esto, con base en que algunas investigaciones (Biber y col. s.f.; Rayson, Wilson y Leech 2002) han probado que la frecuencia y distribución de las etiquetas gramaticales tienen cierta relación con la identificación de género. De esta manera, los autores construyen vectores de características basadas en la frecuencia de cada etiqueta gramatical.

Por otra parte, y para comparar los resultados del enfoque que propusieron, usaron LIWC en su versión 2007, generando una característica por cada una de las 80 dimensiones incluidas en este diccionario.

Finalmente, con referencia a las fuentes de características, usaron n-gramas debido a que mediante ellos se puede capturar el contenido y contexto de los do-

Tabla 5.4: Resultados de la aplicación de diferentes fuentes de generación de características y aplicación de dos clasificadores.

Características	Exactitud	Veraz			Engañoso		
		P	R	F	P	R	F
Identificación de género							
PoS/SVM	73.00 %	75.3	68.5	71.7	71.1	77.5	74.2
Enfoque psicolingüístico							
LIWC/SVM	75.80 %	77.2	76	76.6	76.4	77.5	76.9
Categorización de texto							
Unigramas/SVM	88.40 %	89.9	86.5	88.2	87	90.3	88.6
Bigramas/SVM	89.60 %	90.1	89	89.6	89.1	90.3	89.7
LIWC+bigramas/SVM	89.80 %	89.8	89.8	89.8	89.8	89.8	89.8
Trigramas/SVM	89.00 %	89	89	89	89	89	89
Unigramas/NB	88.40 %	92.5	83.5	87.8	85	93.3	88.9
Bigramas/NB	88.90 %	89.8	87.8	88.7	88	90	89
Trigramas/NB	87.60 %	87.7	87.5	87.6	87.5	87.8	87.6

cumentos. Asimismo, generaron características a partir de unigramas, bigramas y trigramas.

Para realizar la clasificación de documentos usaron un clasificador bayesiano (Naïve Bayes) y una máquina de soporte vectorial (SVM) con una validación cruzada de cinco pliegues. Los resultados se muestran en la Tabla 5.4, en la cual podemos observar que la combinación de características de LIWC y bigramas, clasificados con una SVM, obtuvieron el mejor resultado. Además, con respecto a exactitud y medida F, todas las características generadas y clasificadas con aprendizaje automático superan la clasificación humana (ver Tabla 5.3).

Aunque los autores proponen la identificación de texto mediante un enfoque de identificación de género, se puede observar en la Tabla 5.4 que la generación de características mediante la frecuencia de las etiquetas gramaticales no fue la mejor opción para producir vectores suficientemente informativos; y puesto que los autores argumentan que en los experimentos se probaron todas las combinaciones de fuentes de características, y dado que no se muestran más combinaciones del enfoque PoS, se puede inferir que combinar el enfoque de etiquetas gramaticales con otro enfoque resultó perjudicial para la clasificación.

5.4. Conjuntos de sujetos homogéneos en el análisis de texto engañoso

En la investigación de Fornaciari y Poesio (Fornaciari y Poesio 2012) se aborda la identificación de texto engañoso mediante la creación de subconjuntos a partir de un conjunto general o corpus. La agrupación de subconjuntos homogéneos de sujetos se da, por ejemplo, con la separación de género, es decir, si el sujeto que escribió el texto es hombre o mujer. La agrupación de conjuntos fue

realizada tanto de forma automática, mediante técnicas de agrupamiento automático, y también mediante el uso de los metadatos contenidos en los conjuntos de datos. De esta manera, las agrupaciones, como argumentan los autores, podrían generar la eficiencia en la detección de texto engañoso, basados en que el comportamiento similar en la escritura de los autores podría influir en la detección del engaño.

En este estudio se generó un conjunto de datos llamado *DeCour* (Deception in Court), el cual es un conjunto de datos de transcripciones de treinta y cinco audiencias llevadas a cabo en cuatro cortes de Italia. Los textos se componen de preguntas hechas en la corte a los testigos que defienden al acusado, como resultado el conjunto de datos contiene etiquetas que informan si la oración o respuesta transcrita es veraz o engañosa.

Los autores crearon vectores de características con tres tipos de características: la primera agrega información básica como el tamaño de las oraciones y el número de palabras que tienen más de seis caracteres; la segunda incluye información léxica mediante el uso de LIWC 2001, agregando al vector de características las 80 dimensiones con las que cuenta dicha versión; la última agrega frecuencias de n-gramas de lemas y de etiquetas gramaticales. Los autores tomaron en cuenta en los experimentos desde unigramas hasta pentagramas (grupos de cinco palabras).

Mediante la obtención de n-gramas, con valores de n desde 1 hasta 5, se generaron dos listas de los n-gramas sobresalientes, una correspondiente al texto engañoso y otra al texto veraz. Dichas listas se compararon para evitar que una característica igual se presentara en ambas, para así generar solamente información particular que permitiera una clasificación eficiente.

Se realizaron tres experimentos usando el conjunto de datos creado. El primer experimento consistió en usar todo el conjunto de datos, estableciendo un subconjunto de entrenamiento y uno de prueba.

En el segundo experimento se generaron subconjuntos del corpus completo, de forma que se eliminaron aquellos vectores que estuvieran demasiado alejados de los grupos (*outliers*). Estos grupos fueron establecidos mediante un método no supervisado de agrupamiento de patrones.

En el tercer experimento se hizo uso de los metadatos que se recabaron con el corpus. En principio el corpus cuenta con la siguiente información: el género (masculino o femenino), el lugar de nacimiento y la edad al momento de la declaración. Finalmente, los autores decidieron tomar en cuenta sólo el género de los sujetos, en específico, sólo aquellos participantes de género masculino. Esta restricción impactó fuertemente sobre el conjunto de entrenamiento y no a tal grado el conjunto de prueba.

El conjunto de prueba consta de 426 oraciones de las cuales 190 son engañosas y 236 veraces. Los resultados de la clasificación en los tres diferentes experimentos se muestra en la Tabla 5.5. Se puede observar que el peor resultado, respecto al promedio de medida F, se obtiene cuando sólo se contempla el género masculino; esto podría deberse a que el conjunto de entrenamiento se redujo drásticamente. Por otra parte, el mejor resultado se obtuvo en el experimento donde se eliminaron los outliers, lo que pudo permitir un grupo más homogéneo.

Tabla 5.5: Resultados de la clasificación del corpus DeCour. Se muestran: precisión (P), exhaustividad (R) y medida F (F), así como el promedio de medidas F.

Experimento	Promedio F	Veraz			Engañoso		
		P	R	F	P	R	F
Todo el corpus	60.1	62.9	94	75.3	80.8	31.0	44.9
Sin outliers	64.0	66.8	93.8	77.9	81.0	36.2	50.0
Género masculino	59.5	67.8	94.2	78.9	74.4	27.4	40.0

5.5. Detección intercultural del engaño

En la sección anterior se consideró que la creación de subconjuntos basados en el género y en los outliers tendría cierto impacto en la eficiencia de la clasificación. Por otra parte, en la investigación de Pérez-Rosas y Mihalcea (Pérez-Rosas y Mihalcea 2014a) se muestra el resultado de experimentos con textos escritos por personas de diferentes países para verificar hasta qué punto es posible identificar el engaño, por ejemplo, usando los textos de un país para identificar engaño en los textos de otro país diferente.

La mayoría de estudios de detección de engaño están enfocados a documentos de determinado país e idioma, principalmente al idioma inglés. En este sentido los autores argumentan que los textos, principalmente aquellos que se encuentran en los sitios web, tienden, en cierta proporción, a ser escritos por personas que tienen diferente cultura o idioma; esto implica que las personas pueden tener diferentes creencias y valores morales. En consecuencia, se pone en tela de juicio si los enfoques que se basan en un solo idioma y cultura podrán ser igual-

mente aplicados a una mezcla de textos interculturales e incluso de diferente idioma (este último, usando herramientas de traducción).

En este estudio se abordó la detección de engaño con textos escritos en tres diferentes culturas: Estados Unidos de América, México, y la India. Por lo cual, los autores generaron tres conjuntos de datos que abordan, para cada uno, tres tópicos controvertidos: aborto, pena de muerte y mejores amigos, tal y como se manejó en el estudio presentado en la sección 5.2, excepto para los textos recabados en español de México; estos últimos se obtuvieron mediante una página creada por los autores debido a la poca participación recibida mediante el turco mecánico de Amazon.

Las fuentes de generación de características para este estudio fueron dos. Como en la mayoría de las investigaciones descritas en este capítulo, los unigramas son requeridos para generar una parte del vector de características, por lo que se puede inferir que las palabras por sí solas son una importante fuente informativa para detectar el engaño. Para complementar el vector de características se utilizó LIWC, herramienta que hemos visto se ha vuelto muy popular en la detección de texto engañoso. Recordemos que esta herramienta está basada en un análisis psicolingüístico.

Dos tipos de experimentos fueron realizados: el primero es una clasificación sobre los conjuntos de datos con documentos del mismo país; el segundo se trata de un cruce de dominios, es decir, tomar dos conjuntos (de los tres disponibles) para formar el conjunto de entrenamiento y usar el conjunto restante para generar el conjunto de prueba. Este último experimento se hizo tanto de tópicos mezclado, donde se aplica el dominio cruzado sobre los tres tópicos sobre cada cultura por separado; como para cultura cruzada, donde se aplica dominio cruzado sobre las tres culturas. Este último experimento, con el objetivo de probar

Tabla 5.6: Experimentos realizados sobre los corpora de cada país, por separado.

Tópico	LIWC	Unigramas	
		En dominio	Dominio cruzado
Inglés: EUA			
Aborto	73	63.8	80.4
Mejor amigo	73	74.5	60.8
Pena de muerte	58.1	58.1	77.2
Inglés: India			
Aborto	56	46	50
Mejor amigo	71.4	60.5	57.2
Pena de muerte	63.5	57.5	54
Español: México			
Aborto	62.2	52.5	57.7
Mejor amigo	75.3	66.7	50.5
Pena de muerte	62.2	54.9	63.4

si las características que sirven para identificar texto engañoso en un determinado lenguaje de un país, son equivalentes o útiles al buscar engaño en el lenguaje manejado por otro país diferente (de diferente cultura).

En la Tabla 5.6 se muestran los resultados de la clasificación, de forma individual, de cada cultura mediante la generación de características con el enfoque psicolingüístico y mediante el uso de unigramas. Sobre este último se aplica la clasificación en dominio, en la cual se clasifica cada tópico por separado (aborto, mejor amigo, y pena de muerte) y la clasificación de dominio cruzado, en la cual se usan dos tópicos como entrenamiento y un tercer tópico como prueba.

En la Tabla 5.7 se muestran los resultados de la clasificación de dominio cruzado mediante el uso de dos lenguajes como conjunto de entrenamiento y el lenguaje restante como conjunto de prueba.

5.6. Identificación del engaño mediante el método de aprendizaje mediante multitudes

Las opiniones engañosas en la venta de libros, de acuerdo al estudio de Fornaciari y Poesio (Fornaciari y Poesio 2014), ocurren “cuando los autores escriben una opinión brillante sobre sus propios libros”. Con el propósito de analizar las opiniones engañosas presentes en la venta de libros a través de Amazon, los autores crearon un nuevo conjunto de datos que etiquetaron con base en algunas señales de engaño, la presencia de dichas señales determinó si la opinión fue considerada engañosa o veraz. El enfoque usado para el etiquetado fue el aprendizaje de multitudes propuesto por Raykar (Raykar y col. 2010); además se evaluó la efectividad de diferentes métodos de etiquetado de acuerdo con el rendimiento de los modelos generados para detectar opiniones engañosas.

Los autores argumentan que la gran desventaja de otros trabajos que tratan la detección de engaño es que carecen de un conjunto de datos con textos reales, en cambio son recreados artificialmente; por ejemplo, cuando se les pide a los participantes mentir acerca de una opinión real. El punto es que, al darle permiso a una persona de mentir, es posible que las características de engaño no se reflejen en el texto, o se reflejen de forma parcial; inclusive cabe la posibilidad de que lo que se esté capturando no sea el mismo fenómeno, por lo tanto, la identificación y uso del modelo no será congruente con los textos engañosos en fenómenos reales.

Tabla 5.7: Clasificación de cultura cruzada con LIWC y unigramas como fuente de características.

Tópico	LIWC	Unigramas
Entrenamiento: Inglés: EUA, prueba: Inglés: India		
Aborto	52.3	57.9
Mejor amigo	59.5	51
Pena de muerte	53.5	59
Entrenamiento: Inglés: India, prueba: Inglés: EUA		
Aborto	62.5	55.5
Mejor amigo	55.8	53.2
Pena de muerte	39.2	50.7
Entrenamiento: Inglés: EUA, prueba: español: México		
Aborto	53.9	61.5
Mejor amigo	67.7	65
Pena de muerte	62.2	59.8
Entrenamiento: Inglés: India, prueba: español: México		
Aborto	43.6	55.1
Mejor amigo	60.8	67.2
Pena de muerte	59.8	51.2

Por lo anterior, los autores propusieron un nuevo método para identificar opiniones engañosas en Amazon. Este sistema se basa en dos procesos que se definen a continuación.

En el primero se colaboró con un experto en detección de opiniones engañosas para identificar una serie de criterios para encontrar opiniones no genuinas. De algunas de estas opiniones ya se tenía, de antemano, la seguridad de que eran engañosas debido a que se obtuvieron confesiones posteriores de personas dedicadas a la escritura de las mismas.

En el segundo, los autores desarrollaron un enfoque que identificó la veracidad de las opiniones usando indicadores potenciales de veracidad. Finalmente, usaron el algoritmo propuesto por Raykar para asignar cada opinión del corpus a una clase.

Después de crear el corpus, los autores identificaron un conjunto de señales, cuya presencia sugería que los textos eran engañosos. Las señales identificadas fueron las siguientes: (1) libros sospechosos: de acuerdo a la investigación, son libros que seguramente o muy probablemente recibieron opiniones engañosas; (2) el tiempo: Sandra Parker, una escritora de opiniones engañosas relató que las agencias para las que trabajaba le daban máximo 48 horas para escribir las opiniones, por lo cual sugirió que se pusiera especial atención en libros que recibieran opiniones en cortos periodos de tiempo; (3) registro: Amazon permite crear cuentas de usuario usando los nombres reales de los usuarios, por tanto es menos probable que usaran su nombre real cuando escribieran opiniones engañosas; (4) libro comprado: otra información que se puede obtener de Amazon es si el libro fue comprado o no mediante dicho sitio, por lo tanto la ausencia de compra del libro fue considerada una señal de engaño.

Tabla 5.8: Tabla de clasificación del conjunto DeRev que muestra el rendimiento de los métodos de etiquetado

Método	Exactitud	Engañoso		
		P	R	F
Voto de clases	75.40 %	83.3	63.6	72.1
Algoritmo de Raykar et al.	76.30 %	78.7	72	75.2

Fueron llevados a cabo dos experimentos. En el primero, la asignación de clases se realizó con base en una votación de las señales de engaño. Aquellas opiniones con hasta 2 señales de engaño fueron consideradas veraces, y aquellas con más de dos señales fueron consideradas engañosas. En el segundo, el método de aprendizaje mediante multitudes (*crowdsourcing*) fue usado.

Para realizar la clasificación se generó un vector de características por cada opinión. Dichas características se basaron en unigramas, bigramas, y trigramas de lemas y de etiquetas gramaticales. Entonces, se generaron dos listas de las frecuencias de cada lista de las fuentes de generación de características, una lista para opiniones engañosas y otra para opiniones veraces. Con esto se buscaron las características más frecuentes de cada clase. Para esto, solamente los n-gramas que aparecieran más de 300 veces, en cada lista de frecuencias, fueron tomados en cuenta.

En la Tabla 5.8 se muestra la clasificación del conjunto de datos DeRev mediante una máquina de soporte vectorial. La tabla muestra también la comparación entre métodos de etiquetado, teniendo un mayor rendimiento, aunque no significativo, el uso del algoritmo de Raykar et al.

5.7. Detección de engaño basada en modelos probabilísticos

A pesar de que hemos visto un buen desempeño con las características anteriores, es posible utilizar modelos probabilísticos para clasificar el engaño. En esta sección presentamos resultados detallados sobre la identificación del engaño con base en distintos métodos de generación de características. Hernández-Castañeda, Calvo y col. 2017, han realizado experimentos sobre los corpora presentados en secciones anteriores, utilizando los siguientes métodos de extracción de características: el modelo probabilístico Latent Dirichlet allocation (LDA), el modelo de espacio de palabras Matriz Término Documento (MTD), n-gramas sintácticos (NS-GRAMAS) y linguistic inquiry and word count (LIWC); así como las combinaciones de éstos. Encontramos que la clasificación de textos engañosos podía realizarse mejor con características basadas en modelos probabilísticos (LDA) junto con la matriz término-documento (MTD), como se muestra en la Tabla 5.9.

5.7.1. Clasificación de dominios mezclados

En la búsqueda de la identificación de posibles características universales que pudieran identificar el engaño en diversos dominios, integramos los cinco conjuntos de datos en diferentes dominios. Con esto, el conjunto de entrenamiento contendría temas que también incluiría el conjunto de prueba (pero no de textos de un solo dominio). En este caso, nuevamente una combinación de características LDA y MTD produjo el mejor resultado; ver Tabla 5.10. La SVM genera una exactitud mayor que el clasificador NB.

Tabla 5.9: Comparación de nuestros resultados con otros estudios de los mismos corpora

Corpus	Estudios	Exactitud
OpSpam		
	Este estudio (LDA+MTD)	90.90 %
	Mile Ott et al. (LIWC + bigramas)	89.80 %
	Song Feng et al. (syntactic rel.+ unigramas)	91.20 %
	Donato et al.	90.20 %
DeRev		
	Este estudio (LDA+MTD)	94.90 %
	Fornaciari y Poesio	76.27 %
Aborto		
	Este estudio (LDA+MTD)	87.50 %
	Perez-Rosas y Mihalcea	80.30 %
	Song Feng et al. (syntactic rel. + unigramas)	77.00 %
Mejor amigo		
	Este estudio (LDA+MTD)	87.00 %
	Perez-Rosas y Mihalcea	75.90 %
	Song Feng et al. (syntactic rel. + unigramas)	85.00 %
Pena de muerte		
	Este estudio (LDA+MTD)	80.00 %
	Perez-Rosas y Mihalcea	77.20 %
	Song Feng et al. (syntactic rel. + unigramas)	71.50 %

Tabla 5.10: Porcentajes de exactitud, precisión (P), exhaustividad (R) y medida F (F) obtenidas con los corpora mezclados mediante SVM

Método	Exactitud	Veraz			Engañoso		
		P	R	F	P	R	F
LDA	65.8	67.6	60.9	64.1	64.4	70.8	67.4
MTD	73.5	73.6	73.5	73.5	73.5	73.6	73.5
LDA + MTD	76.3	76.8	75.6	76.2	75.9	77.1	76.5
LIWC	59.8	60.1	58.4	59.2	59.5	61.1	60.3
LIWC + MTD	73.7	71.9	77.8	74.7	75.8	69.7	72.6
LDA + LIWC	69.7	69.3	70.8	70.1	70.2	68.7	69.4
LIWC + WS + LDA	72.6	72.3	73.5	72.9	73	71.9	72.5

5.7.2. Clasificación de dominio cruzado

Finalmente, para comprobar qué tanto es posible detectar el engaño en un dominio completamente diferente del que se aprendieron las características, realizamos pruebas con clasificación de dominio cruzado. Para este experimento seleccionamos cada conjunto de datos una vez como conjunto de pruebas y usamos los otros conjuntos de datos restantes como un conjunto combinado de entrenamiento. De esta manera, el dominio del conjunto de prueba no se incluyó en el conjunto de entrenamiento.

La Tabla 5.11 muestra los resultados de la clasificación inter-dominio. En estos experimentos, a diferencia de los experimentos presentados en las secciones anteriores, la combinación de características de LDA y MTD no dio consistentemente la mejor exactitud, aunque en los casos en los que no fue la mejor, MTD sin LDA sí lo fue. Mostramos en cursiva la mejor precisión obtenida para cada

Tabla 5.11: Medida F obtenida en la clasificación de dominio cruzado. Se muestran resultados para los clasificadores NB y SVM.

		DeRev	OpSpam	Aborto	Mejor amigo	Pena de muerte	Promedio
LDA	SVM	52.1	48.8	57.5	50	53.5	50.43
	NB	43.2	49.8	56	51.5	56	46.5
MTD	SVM	53.3	52.8	55.5	55.5	58.5	53.05
	NB	50.8	53.8	58.5	56	48.5	52.3
LIWC	SVM	51.3	49.2	53.1	54.3	52.8	50.25
	NB	50.9	51.1	52.6	56.8	51.4	51
LIWC+MTD	SVM	54.6	53.8	54.5	55	56	54.2
	NB	47.8	52.5	55.5	59	51	50.15
LDA+MTD	SVM	59.3	50.6	57.5	64	55	54.95
	NB	58.8	52.3	55	59.5	52.5	55.55
LDA+LIWC	SVM	56.3	46.3	54.5	55.5	52	51.3
	NB	45.7	48.1	46.5	51.5	49.5	46.9
LDA+LIWC +MTD	SVN	52.1	52.6	57	57	54	52.35
	NB	52.9	53	58	62.5	53	52.95
	Mejor:	SVM	NB/SVM	NB	SVM	SVM	NB

conjunto de datos. Con respecto a los clasificadores, en la mayoría de los casos (3 de 5), el SVN superó a NB; sin embargo, con una configuración relativamente simple de una matriz término-documento, NB es capaz de mejorar la detección de engaño con las características aprendidas de otros conjuntos de datos.

5.7.3. Palabras y tópicos predominantes

Detectar el engaño en un texto es una tarea que se ha tratado de mejorar mediante distintas técnicas. Dicha tarea no es sencilla debido a que las características que distinguen a un texto engañoso pueden variar entre distintos conjuntos de datos. Algunas palabras que representan el texto falso en un conjunto de datos pueden representar al texto veraz en otro; esto indica que una palabra no es estrictamente representante de una clase u otra en todos los casos.

El hecho de que ciertas palabras sean contradictorias en conjuntos de datos distintos no indica que los métodos automáticos de reconocimiento de texto engañoso sean disfuncionales, sino que estas palabras pueden representar a un conjunto específico de datos, pero no figurarán al tener un conjunto más general, por ejemplo, al unir el conjunto OpSpam con el conjunto “Aborto”.

Se probó que la combinación de características LDA y MTD generó buena precisión en los tres tipos de conjuntos de datos estudiados. Esto implica que los modelos generados mediante esta combinación de características son más fiables para la detección de texto engañoso puesto que han sido probados en distintos casos de engaño y se han mantenido estables en cuanto a la exactitud obtenida. Por otra parte, LDA+MTD puede ser aplicada a otros idiomas debido a que no requiere de algún corpus para genera las características como LIWC o n-gramas sintácticos (NS-GRAMAS).

El uso de LDA resultó ser en gran medida un método alternativo más eficiente que el uso de LIWC. Esto se debe a que LDA genera tópicos y toma en cuenta todas las palabras contenidas en los documentos, de tal forma que no habría palabras que escapen del proceso estadístico de generación de grupos de palabras. Por otra parte, LIWC contiene un grupo de palabras preestablecidas de las cuales probablemente algunas no estén incluidas en los documentos a procesar. Por lo tanto, la generación de un modelo que represente al texto engañoso puede ser más exitoso si se usa LDA.

Para mostrar como la combinación LDA y MTD se complementan favorablemente, en la Tabla 5.12 se muestra un ejemplo de las palabras y tópicos más relevantes del texto engañoso y del texto veraz en el conjunto de datos “De-Rev”. Otros ejemplos pueden ser consultados en (Hernández-Castañeda y Calvo 2017). Un punto a destacar de este ejemplo es que los tópicos son características dominantes en la detección de los textos engañosos, es decir, está compuesto de más tópicos que palabras únicas. Mientras que, por otra parte, las palabras son características dominantes en el texto veraz. Se observó, en los conjuntos de datos analizados, que si una clase carece de características relevantes que acentúen una mayor diferencia entre clases, habrá tópicos que aumenten dichas características y mejoren la exactitud en la clasificación.

Referencias

Almela, Ángela, Rafael Valencia-García y Pascual Cantos (2012). “Seeing through deception: A computational approach to deceit detection in written communication”. En: *Proceedings of the Workshop on Computational Approaches to Deception Detection*. Association for Computational Linguistics, págs. 15-22.

Tabla 5.12: Los diez tópicos más relevantes y palabras representantes del texto engañoso y veraz en el conjunto de datos *DeRev*

Texto engañoso	Texto veraz
author	the was i they were some which had not during these even when if because character been most several
your	i the it book was t read but reading first all my had this that one so characters love
uno/un	history american native our americans country heart indians west day newspaper
thriller	the tone brown better known fully resist stolen classic
the and of a to is in that this for s book as it on with an are but and	edition I the movie book than seen made
intrigue	much movies version like cover film better more do it least half then
guide	version
drug	history
issue	again

- Biber, Douglas y col. (s.f.). *Longman grammar of spoken and written English*.
DePaulo, Bella M y col. (2003). "Cues to deception." En: *Psychological bulletin* 129.1, pág. 74.
- Fornaciari, Tommaso y Massimo Poesio (2012). "On the use of homogenous sets of subjects in deceptive language analysis". En: *Proceedings of the Workshop on Computational Approaches to Deception Detection*. Association for Computational Linguistics, págs. 39-47.
- (2014). "Identifying fake Amazon reviews as learning from crowds". En: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, págs. 279-287.
- Fusilier, Donato Hernández y col. (2015). "Detection of opinion spam with character n-grams". En: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, págs. 285-294.
- Hauch, Valerie y col. (2012). "Linguistic cues to deception assessed by computer programs: A meta-analysis". En: *Proceedings of the workshop on computational approaches to deception detection*. Association for Computational Linguistics, págs. 1-4.
- Hernández-Castañeda, Ángel e Hiram Calvo (2017). "Deceptive text detection using continuous semantic space models". En: *Intelligent Data Analysis* 21.3, págs. 679-695.
- Hernández-Castañeda, Ángel, Hiram Calvo y col. (2017). "Cross-domain deception detection using support vector networks". En: *Soft Computing* 21.3, págs. 585-595.
- Mihalcea, Rada y Carlo Strapparava (2009). "The lie detector: Explorations in the automatic recognition of deceptive language". En: *Proceedings of the*

- ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, págs. 309-312.
- Newman, Matthew L y col. (2003). "Lying words: Predicting deception from linguistic styles". En: *Personality and social psychology bulletin* 29.5, págs. 665-675.
- Ott, Myle y col. (2011). "Finding deceptive opinion spam by any stretch of the imagination". En: *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*. Association for Computational Linguistics, págs. 309-319.
- Pérez-Rosas, Verónica y Rada Mihalcea (2014a). "Cross-cultural deception detection". En: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, págs. 440-445.
- (2014b). "Gender differences in deceivers writing style". En: *Mexican International Conference on Artificial Intelligence*. Springer, págs. 163-174.
- Raykar, Vikas C y col. (2010). "Learning from crowds". En: *Journal of Machine Learning Research* 11.Apr, págs. 1297-1322.
- Rayson, Paul, Andrew Wilson y Geoffrey Leech (2002). "Grammatical word class variation within the British National Corpus sampler". En: *New Frontiers of Corpus Research*. Brill Rodopi, págs. 295-306.
- Schelleman-Offermans, Karen y Harald Merckelbach (2010). "Fantasy proneness as a confounder of verbal lie detection tools". En: *Journal of Investigative Psychology and Offender Profiling* 7.3, págs. 247-260.
- Toma, Catalina L y Jeffrey T Hancock (2012). "What lies beneath: The linguistic traces of deception in online dating profiles". En: *Journal of Communication* 62.1, págs. 78-97.

- Williams, Shanna Mary y col. (2014). "Is the truth in your words? Distinguishing children's deceptive and truthful statements". En: *Journal of Criminology* 2014.
- Xu, Qiongkai y Hai Zhao (2012). "Using deep linguistic features for finding deceptive opinion spam". En: *Proceedings of COLING 2012: Posters*, págs. 1341-1350.

Capítulo 6

Sistemas Interactivos Emocionales

Juan Martínez Miranda, CICESE-UT₃.
Humberto Pérez Espinosa, CICESE-UT₃.

El área de investigación conocida como interacción humano-computadora se ha enfocado tradicionalmente al estudio, diseño y desarrollo de metodologías y técnicas que faciliten una comunicación efectiva entre un sistema computacional y el usuario al que va dirigido. En años recientes, los objetivos de esta área de investigación han evolucionado pasando de concentrarse en el desarrollo de interfaces de usuario amigables, intuitivas y útiles, hacia la construcción de sistemas interactivos capaces de adaptar y personalizar su funcionalidad acorde a las características particulares de cada usuario con el fin último de motivarlo e involucrarlo plenamente en una tarea determinada. Un tipo particular de estos sistemas interactivos son aquellos que utilizan el lenguaje natural para comunicarse con el usuario. Esto facilita la ejecución de las acciones provistas por el sistema, pues emula una comunicación como la que se da de persona a persona. El uso del lenguaje natural para interactuar con un sistema computacional puede ser aprovechado no solo para dar y recibir instrucciones, también es un mecanismo de comunicación el cual puede ser utilizado para inferir ciertas características del

usuario tales como su estado emocional o anímico, y para transmitir la información de una manera empática que ayude a generar confianza en el usuario. En este capítulo se presenta una breve reseña sobre el desarrollo actual de los sistemas interactivos basados en lenguaje natural y algunas de las técnicas utilizadas para el reconocimiento de emociones a partir de la voz, así como el modelado de emociones en sistemas artificiales. Se presentan también algunos de los proyectos desarrollados en nuestro centro de investigación que utilizan estas técnicas para diferentes objetivos, desde la mejora de la interacción de niños y adultos mayores con sistemas basados en voz, hasta el desarrollo de sistemas interactivos para la ayuda en problemáticas del sector salud.

6.1. Introducción

El uso del lenguaje natural como un medio de interacción entre personas y los sistemas de cómputo ha tenido un enorme auge en los últimos años. Esto se debe en gran parte gracias al desarrollo de dispositivos personales y móviles cada vez más potentes los cuales se utilizan hoy en día para una infinidad de aplicaciones. Ya desde el desarrollo de las primeras computadoras personales a principios de los años 80's, dirigidas a usuarios sin conocimientos especializados sobre sistemas operativos o desarrollo de software, se hizo evidente la necesidad de diseñar interfaces de usuario *amigables* y *fáciles* de utilizar (Cockton 2001). Los primeros usuarios de estas computadoras eran expertos en otras áreas del conocimiento los cuales comenzaron a adoptar las computadoras como una herramienta indispensable para llevar a cabo sus tareas (Day y Boyce 1993). Desde entonces, el área de procesamiento de lenguaje natural se consideró como una manera de mejorar, o hacer más natural, la comunicación entre los usuarios y los sistemas de

cómputo. Incorporar capacidades de lenguaje natural a las interfaces de usuario, facilitaría aún más el uso de este tipo de sistemas no sólo a personas sin conocimientos especializados, además tendrían una enorme utilidad en situaciones en las cuales no es posible el uso de las herramientas tradicionales como el ratón y teclado o para personas con alguna discapacidad que les impida el uso de las mismas (Manaris 1998).

Aunque aún hay enormes retos por superar para tener una comunicación hombre-máquina similar a la que se da de persona a persona, en los últimos años ha habido un gran avance en el desarrollo de interfaces basadas en el lenguaje natural. Ejemplos de este tipo de interfaces son los asistentes basados en voz desarrollados por la grandes empresas de tecnología entre los que se incluye Siri de Apple, Cortana de Microsoft, Alexa de Amazon o el Assistant de Google. El uso del lenguaje natural en sistemas interactivos no se limita únicamente a mejorar la comunicación entre el usuario y el sistema, además tiene la capacidad de ser un mecanismo para inferir ciertas características del usuario tales como su estado emocional o anímico, el tipo de personalidad y alguna otra información particular que el sistema puede utilizar para adaptar y personalizar el estilo y contenidos de interacción. El ofrecer respuestas personalizadas acorde a las características propias de cada individuo puede mejorar de manera relevante la experiencia del usuario y maximizar la utilidad de estos sistemas.

Una línea de investigación relativamente nueva dedicada al estudio y diseño de sistemas de cómputo (hardware y software) capaces de reconocer, responder e influenciar los estados emocionales en los usuarios es la computación afectiva (Picard 1997). El diseño y desarrollo de sistemas de cómputo afectivos incluyen dos etapas fundamentales: la *identificación* de estados emocionales en el usuario, y la *generación* de respuestas afectivas por parte del sistema. Para el reconocimien-

to de emociones, actualmente se utilizan técnicas de inteligencia artificial para el análisis de señales y reconocimiento de patrones que son asociados a estados afectivos utilizando datos sobre las expresiones faciales, el habla, y señales fisiológicas como la respuesta galvánica de la piel entre algunos otros. Por otro lado, la generación y representación de reacciones afectivas en sistemas interactivos, tales como robots de servicio o agentes virtuales, son generadas por arquitecturas computacionales de emociones que emulan los procesos afectivos humanos. Estas arquitecturas tienen su fundamento teórico en modelos psicológicos, cognitivos, o neurológicos y son comúnmente implementadas como el mecanismo interno para la generación del comportamiento tanto en robots sociales o agentes virtuales conversacionales.

El uso del lenguaje natural es uno de los medios de comunicación más utilizados en este tipo de sistemas interactivos tanto para el reconocimiento de emociones en el usuario como para transmitir las respuestas emocionales de parte del sistema. El área de procesamiento de lenguaje natural se encarga precisamente de aplicar diferentes metodologías y técnicas para que los sistemas interactivos sean capaces de entender y manipular el lenguaje humano (ya sea mediante texto o voz) y llevar a cabo una comunicación más natural con el usuario. Para el reconocimiento de emociones, la comunicación basada en el habla tiene la capacidad de analizarse en términos lingüísticos y paralingüísticos con el fin de desarrollar algoritmos con una buena precisión en el reconocimiento de diferentes estados afectivos. De manera complementaria, el habla también es de vital importancia para comunicar diferentes reacciones emocionales del sistema hacia el usuario. Por ejemplo, en los agentes virtuales conversacionales, la comunicación verbal (las frases específicas) que el agente utiliza para responder al usuario son construidas a partir de la emoción generada en el agente. Uno de los retos actuales es

el desarrollo de sintetizadores de voz emocionales capaces de generar la entonación adecuada a estas respuestas verbales afectivas.

Este capítulo presenta una breve reseña sobre la incorporación del procesamiento de lenguaje natural en sistemas interactivos emocionales. En la primera parte (Sección 6.2) se presentan algunos de los trabajos más relevantes que se están llevando a cabo para el reconocimiento de emociones a través de la voz, en particular haciendo uso de información paralingüística (Sección 6.2.1); y una descripción de diferentes arquitecturas computacionales de emociones para la generación de reacciones afectivas en sistemas interactivos (Sección 6.2.2). La segunda parte del capítulo (Sección 6.3) presenta algunos ejemplos de trabajos de investigación realizados en CICESE-UT₃ los cuales utilizan estas técnicas de reconocimiento y generación de emociones para abordar problemas de diferentes grupos poblacionales.

6.2. Sistemas de interacción afectivos basados en lenguaje natural

La incorporación del lenguaje natural en sistemas interactivos emocionales guarda una estrecha relación con otra de las sub-disciplinas del área de interacción hombre-máquina: el diseño y desarrollo de sistemas basados en voz. La comunicación basada en voz es una de las más utilizadas entre las personas y es un medio fácil y rápido para interactuar con cualquier aplicación computacional (J. Pittermann, A. Pittermann y Minker 2010). Las funcionalidades de los sistemas interactivos basados en voz han evolucionado desde su uso para recuperación de información (por ejemplo para saber los horarios de sistemas de transporte como trenes o autobuses), hasta aplicaciones de soporte para resolver algún

problema específico (por ejemplo aplicaciones para resolver algún problema técnico a través del teléfono). Un sistema interactivo basado en voz usualmente implementa un flujo de comunicación entre diferentes componentes de software responsables de llevar a cabo una tarea determinada. Estos componentes son: i) un módulo de reconocimiento automático del habla; ii) un módulo de comprensión del lenguaje natural; iii) un sistema administrador de diálogo; iv) un modulo generador de lenguaje natural; y v) un módulo sintetizador de voz a partir del texto (Schmitt y Minker 2012).

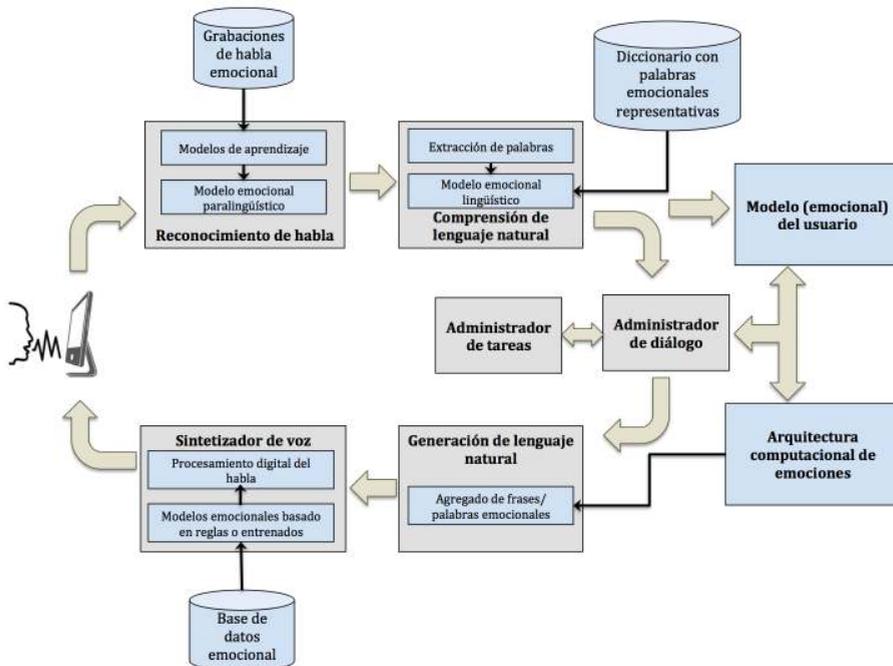


Figura 6.1: Arquitectura de un sistema interactivo basado en voz con componentes para el reconocimiento y generación de información afectiva.

El agregado de modelos y funcionalidades relacionadas con procesos afectivos a cada uno de estos componentes, permite el desarrollo de sistemas interac-

tivos basados en voz capaces de reconocer y comunicar diferentes estados emocionales. Esto puede verse en la Figura 6.1, la cual integra los diferentes componentes de un sistema basado en voz con el agregado de diferentes modelos emocionales tanto para el procesado y reconocimiento automático de emociones a partir de datos obtenidos de la voz del usuario; así como la generación de emociones y síntesis de voz afectiva utilizadas durante la interacción con el usuario. En particular, el reconocimiento del estado emocional del usuario se realiza a través del módulo de reconocimiento del habla agregando modelos de clasificación de *patrones acústicos paralingüísticos* y en el componente de comprensión del lenguaje natural a través de modelos de análisis de *información lingüística*. El estado emocional del usuario identificado a través de estos componentes, es utilizado para actualizar durante cada ciclo de interacción el *modelo –emocional– del usuario*. La información obtenida en el modelo del usuario, puede ser a su vez utilizada por la *arquitectura computacional de emociones* para generar una emoción concreta a utilizar como respuesta hacia el usuario (por ejemplo una emoción que represente una respuesta empática hacia el usuario). La manera de comunicar la emoción generada en el sistema se hace a través del módulo de generación de lenguaje natural (mediante frases y palabras concretas asociadas a la emoción) y a través del sintetizador de voz incorporando la entonación emocional adecuada. La siguiente subsección presenta algunos de los trabajos más representativos y retos abiertos en el reconocimiento de emociones a partir de información paralingüística. Para saber más sobre el reconocimiento automático de emociones utilizando información lingüística, se pueden consultar las siguientes referencias: (Binali, Wu y Potdar 2010; Shelke 2014; Strapparava y Mihalcea 2015).

6.2.1. Detección de emociones en voz

El desarrollo de un sistema capaz de reconocer estados emocionales de manera automática a partir del habla generalmente combina técnicas computacionales conocidas como procesamiento digital de señales (Vetterli, Kovačević y Goyal 2014) y aprendizaje automático (Mohri, Rostamizadeh y Talwalkar 2012). En esta área de investigación se abordan dos aspectos principales: el primero, es el desarrollo de técnicas para el procesamiento, análisis y caracterización de la señal de voz. En segundo lugar, se aborda la aplicación de diferentes técnicas de reconocimiento de patrones que modelan las propiedades de la información acústica extraída de la señal de la voz. Aunque se ha logrado un avance significativo en esta área de investigación, todavía hay un largo camino para obtener reconocedores de emociones en el habla robustos, con un buen desempeño en condiciones naturalistas.

Un factor que limita el óptimo rendimiento de estos sistemas en contextos reales es la escasez de bases de datos que contengan audios de emociones genuinas y espontáneas. De forma recurrente, los investigadores han trabajado con corpus de emociones actuadas que representan, en la mayoría de los casos, expresiones prototípicas e intensas que facilitan la búsqueda de correlación acústica y la clasificación automática posterior. Este tipo de base de datos generalmente se graba en un entorno controlado para evitar problemas en el procesamiento de la señal, como la variación en la distancia al micrófono, el ruido ambiental o la reverberación. Como consecuencia, la precisión obtenida por los modelos desarrollados no es lo suficientemente buena cuando estos modelos se transfieren a contextos reales o naturalistas (Steidl 2009). Esta falta de precisión en contextos poco controlados o no controlados es la razón principal detrás de la construcción reciente de recopilaciones de grabaciones en entornos naturalistas tales

como podcasts (Lotfian y Busso 2017), niños jugando (Lyakso y col. 2015) o programas de TV en vivo (Lubis y col. 2014).

Otro desafío en el reconocimiento de emociones a partir del habla es la identificación correcta de un conjunto de características acústicas para el reconocimiento de emociones en el habla espontánea. Varios trabajos se han centrado en características relacionadas con aspectos prosódicos, como el acento, la entonación y el ritmo. Sin embargo, se ha descubierto que cuanto más naturalista es el escenario, menos confiable es la prosodia como indicador del estado emocional del hablante (Batliner y col. 2003). Por lo tanto, es necesario identificar aquellas características que complementan la información proporcionada por el aspecto prosódico del habla, como la calidad de voz, información espectral y otros.

Un tercer desafío abierto en la etapa de caracterización es cómo encontrar el tamaño ideal de los segmentos del habla que se utilizarán para el análisis. El procedimiento habitual (análisis de tiempo corto) es dividir la señal de audio en ventanas pequeñas (alrededor de 25 milisegundos) donde se espera que la señal sea estacionaria y extraer las características acústicas de cada ventana (características locales). Sin embargo, para la captura de información paralingüística relevante (como emociones), es necesario analizar segmentos más largos, como palabras o frases, calculando estadísticas de todas las características del habla extraídas de una expresión (características globales). También se ha investigado la definición del tamaño de los segmentos y la mejor manera de capturar la información. En (Steidl 2009), se probaron tres niveles de segmentación: palabra (word), turno (turn) y segmento (chunk), pero la identificación de la mejor opción sigue siendo un tema de investigación abierto.

La selección y aplicación de técnicas de aprendizaje automático para entrenar a los modelos de clasificación o regresión para el reconocimiento de emoción

nes en el habla también es un paso fundamental. Hay dos enfoques principales en esta etapa. El procesamiento estático aplica funciones estadísticas en descriptores de bajo nivel en períodos específicos de tiempo, evitando el sobreajuste en el modelado fonético. Por otro lado, el procesamiento dinámico se refiere a cómo evolucionan las características acústicas a lo largo del tiempo con respecto a la evolución del estado emocional de los individuos. En el procesamiento estático, el reconocimiento de emociones se realiza a través de métodos de aprendizaje de máquina estática, como las máquinas de vectores de soporte, los árboles de decisión y las redes neuronales feed-forward. La clasificación se realiza a nivel de elocución o turno, y los segmentos analizados tienen diferentes tamaños. Las características se obtienen de la extracción de descriptores de bajo nivel, por ejemplo, la entonación, la energía, o los coeficientes espectrales. Posteriormente, se aplican funciones estadísticas, como la media, la desviación estándar, en las características, que dan como resultado vectores del mismo tamaño para todas las expresiones (Vogt y André 2009; Lee y col. 2011). En el procesamiento dinámico, características tales como el tono, la energía, los coeficientes cepstrales de frecuencia de mel y sus derivados se utilizan en combinación con métodos de aprendizaje automático dinámicos como las redes neuronales de memoria a corto y largo plazo (LSTM), los modelos ocultos de Markov (A. Pittermann y J. Pittermann 2006) o modelos Gaussianos de mezcla.

Una de las técnicas de aprendizaje automático más utilizada en años recientes son las conocidas como de aprendizaje profundo (Skansi 2018), mediante las cuales es posible obtener una mayor precisión en el reconocimiento de la emoción del habla (Fayek, Lech y Cavedon 2017). Los algoritmos basados en el aprendizaje profundo se aplican tanto para la clasificación como para la generación de características. En este enfoque, se modelan abstracciones de alto nivel median-

te el uso de arquitecturas de características jerárquicas compuestas por múltiples transformaciones de su representación. Una característica relevante del aprendizaje profundo es el paso de entrenamiento previo no supervisado del algoritmo de aprendizaje (generalmente una red neuronal). En esta etapa, se pueden usar grandes cantidades de datos sin etiquetar para construir características de alto nivel. El entrenamiento previo hace que la misma red neuronal obtenga mejores resultados que cuando se inicializa con parámetros aleatorios, como se hace tradicionalmente. Los algoritmos de aprendizaje profundo buscan explotar la estructura desconocida en la distribución de datos de entrada para descubrir buenas representaciones con características abstractas en niveles más altos de representación aprendidos de funciones de bajo nivel. Este hecho hace que sea más fácil separar los aspectos más descriptivos de los datos entre sí (Schuller 2015).

Actualmente, los mejores resultados en el reconocimiento de emociones a partir del habla se logran principalmente mediante el uso de técnicas de aprendizaje automático semi-supervisadas y no supervisadas. Estos enfoques incluyen aprendizaje profundo y aprendizaje de transferencia de representaciones de datos compactas y tareas de reconocimiento automático. Una revisión de trabajos recientes también muestra una tendencia hacia un aprendizaje multitarea masivo y autónomo que requiere métodos de estimación de confianza, como por ejemplo el aprendizaje activo.

6.2.2. Modelado de emociones en sistemas artificiales

Una vez que el estado emocional del usuario ha sido automáticamente detectado, el sistema interactivo debe generar una respuesta emocional adecuada (acorde a los objetivos de la aplicación) y expresarla con claridad hacia el usuario. Actualmente existen diferentes modelos y arquitecturas computacionales

desarrolladas con el objetivo de emular los procesos de generación de emociones humanas e incorporarlos en sistemas artificiales. Algunas de estas arquitecturas son implementadas como el mecanismo interno para la generación del comportamiento cognitivo y emocional en robots sociales y agentes virtuales. En la siguiente sección se presenta una breve revisión de las arquitecturas computacionales de emociones existentes más representativas, poniendo énfasis en las teorías emocionales a partir de las cuales adaptan diferentes conceptos y procesos afectivos.

El desarrollo de diferentes modelos y arquitecturas computacionales de emociones tienen como objetivo representar o imitar en sistemas artificiales el funcionamiento de los procesos emocionales que se dan en sistemas biológicos (Sander y K. Scherer 2009). Las diferencias y particularidades entre cada una de las arquitecturas computacionales de emociones existentes son debidas principalmente a las características concretas de las teorías emocionales en las cuales se basan. Actualmente existen diversas teorías psicológicas de emociones las cuales difieren en los postulados que se asumen sobre los componentes intrínsecos de las emociones. Estos componentes incluyen entre otros, los procesos cognitivos, las respuestas fisiológicas, los cambios motivacionales y comportamientos conductuales asociados a las emociones, así como la relación e integración entre cada uno de ellos (Stacy Marsella, Jonathan Gratch y Petta 2010; K. R. Scherer 2010). Los diferentes postulados propuestos en cada teoría dificultan incluso tener una definición unificada sobre el concepto *emoción*. Algunas de las teorías más representativas en las cuales se basan las arquitecturas computacionales de emociones son las siguientes:

- *Teorías de valoración cognitiva*. Postulan que la activación de una emoción es producida por la valoración o evaluación (no necesariamente conscien-

te o controlada) de las situaciones y eventos que ocurren en el entorno del individuo. También postula que las respuestas para hacer frente a la causa que generó la emoción es guiada por los resultados del proceso de valoración realizado (Lazarus y Lazarus 1991; K. R. Scherer, Schorr y Johnstone 2001).

- *Teorías dimensionales.* Se caracterizan por definir las emociones no a través de valores discretos (por ejemplo felicidad, miedo, enojo, etc.) sino como puntos posicionados en un espacio dimensional (típicamente con dos o tres dimensiones). Las dimensiones más utilizadas en este tipo de teorías se refieren a la *valencia* (qué tan placentera o no es una emoción); el nivel de *activación*; y la *dominancia* –qué tan dominante es la naturaleza– de la emoción (Mehrabian y J. A. Russell 1974; J. A. Russell 2003; Barrett 2006).
- *Teorías de emociones básicas.* Postulan la existencia de un conjunto de emociones básicas (típicamente entre 7 y 14) y cada una tiene sus propias condiciones de activación, y sus particulares patrones de respuesta conductual, fisiológica y expresiva (Ekman 1992; Johnson-Laird y Oatley 1992).
- *Teorías anatómicas.* Se basan en reconstruir los lazos neuronales y los procesos que subyacen en las reacciones emocionales de los organismos (LeDoux 1996; Panksepp 2004).
- *Teorías basadas en una perspectiva racional.* Esta perspectiva se abstrae de los “detalles” sobre el proceso subyacente a la generación de emociones y se concentra en entender la funcionalidad adaptativa de las emociones

y como incorporar esta funcionalidad a un modelo *inteligente* (Sloman y Croucher 1981; Anderson y Lebiere 2003; Doyle 2006).

La selección de una teoría específica como base para el desarrollo de una arquitectura computacional de emociones depende en gran medida de los componentes específicos del proceso o procesos emocionales que el modelo computacional desea representar. Por ejemplo, las teorías dimensionales se utilizan frecuentemente en sistemas cuyo objetivo es reconocer comportamientos emocionales humanos. Algunos otros modelos computacionales combinan diferentes teorías para representar diferentes componentes y fases de procesos emocionales. Un ejemplo es el descrito en (K. R. Scherer 2010), en el cual las dimensiones *valencia* y *activación* de la teoría dimensional se ven como una representación del espacio en el cual las emociones básicas son posicionadas. También en el desarrollo de un sistema de interacción afectivo pueden combinarse diferentes teorías. Por ejemplo utilizando un modelo dimensional de emociones para el reconocimiento automático del estado emocional en el usuario, y la teoría de valoración cognitiva para la implementación del proceso de generación de emociones el cual generará la respuesta afectiva del sistema hacia el usuario. Una revisión más extensa sobre algunas de las arquitecturas computacionales de emociones existentes, y de las teorías en las cuales se basan, se puede consultar en (Stacy Marsella, Jonathan Gratch y Petta 2010). De ese trabajo se ha tomado la Figura 6.2, la cual presenta de manera visual estas arquitecturas computacionales.

Como se puede observar en la Figura 6.2, la teoría emocional sobre la que más se han desarrollado diferentes arquitecturas computacionales es la teoría de valoración cognitiva (Lazarus y Lazarus 1991; K. R. Scherer, Schorr y Johnstone 2001). El concepto principal de esta teoría es precisamente el proceso de identificación y valoración sobre los eventos que ocurren en el entorno. Esta va-

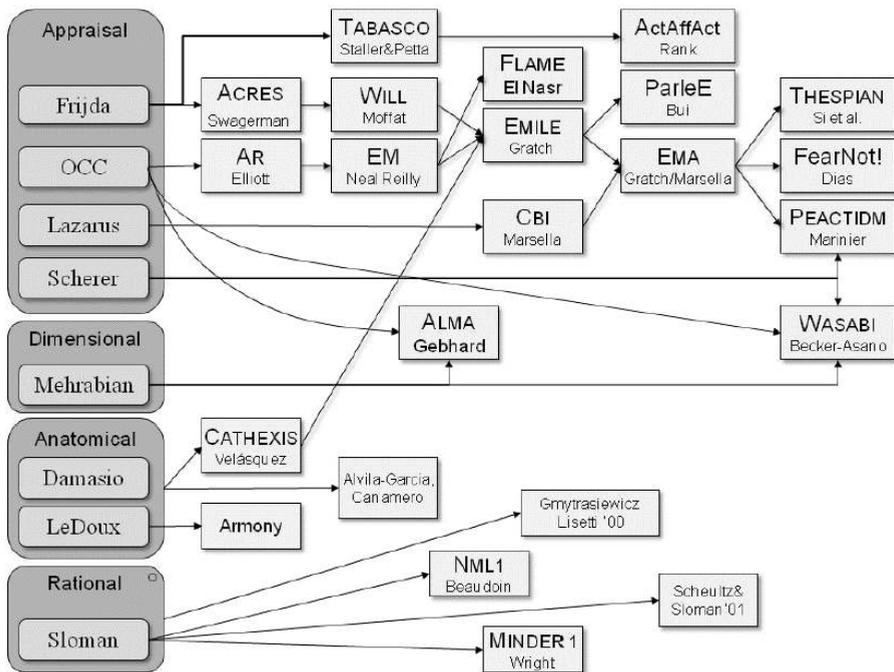


Figura 6.2: Algunas de las arquitecturas computacionales de emociones más representativas. Imagen tomada de (Marsella y Gratch 2009).

loración *cognitiva* produce una *respuesta emocional* (basada en la relevancia del evento evaluado sobre las metas, creencias y deseos del individuo). A su vez, esta respuesta emocional genera un *comportamiento específico* para hacer frente a las consecuencias del evento evaluado. Una de las razones por las cuales esta teoría sirve de base para la mayoría de las arquitecturas computacionales de emociones existentes es debido al énfasis y explicación que ofrece para la conexión entre la parte cognitiva y la parte emocional, lo cual ayuda a emular comportamientos humanos complejos en sistemas artificiales. Adicionalmente, esta teoría explica de una forma completa la complejidad de la generación de emociones y las acciones de respuesta, abarcando desde el proceso llevado a cabo para la valoración de

los eventos y situaciones, hasta la influencia que tienen los estados emocionales sobre el comportamiento generado como respuesta (K. R. Scherer 2010).

Diferentes arquitecturas computacionales de emociones se han venido desarrollando desde los años 90's, la mayoría de ellas implementando conceptos y componentes definidos en teorías emocionales de valoración cognitiva (Elliott 1992; Dias y Paiva 2005; Gebhard 2005; Marsella y Gratch 2009). Estas arquitecturas han sido principalmente desarrolladas para su incorporación en agentes virtuales conversacionales para hacer más realista su comportamiento y facilitar una mejor interacción con el usuario. La incorporación de un mecanismo que genere un comportamiento emocional coherente facilita el desarrollo de agentes virtuales más convincentes (Reilly 1996), creíbles (Mao y Jonathan Gratch 2006), o con mejores capacidades para inducir efectos sociales deseables en los usuarios (Paiva y col. 2005). Durante la interacción de un agente virtual con el usuario, una parte importante es la manera de comunicar el estado emocional generado por la arquitectura computacional, tanto de manera verbal (a través de diálogos concretos) como no verbal (mostrar la expresión facial y movimiento corporal correspondiente a dicha emoción). En términos de comunicación verbal, el componente de generación de lenguaje natural es el responsable de agregar las palabras o frases con la carga emocional que se va a comunicar. El otro componente importante para transmitir correctamente la emoción, es el sintetizador de la voz del agente, la cual debe tener la entonación adecuada. La generación de sintetizadores emocionales de voz es un tema abierto de investigación el cual presenta sus propios avances y retos, y la revisión de estos cae fuera del alcance de este capítulo, pero trabajos relevantes se pueden consultar en (Schröder 2001; Burkhardt y Campbell 2014; Crumpton y Bethel 2016).

6.3. Aplicaciones de sistemas afectivos basados en voz

6.3.1. Creación de bases de datos de voz emocional

Actualmente existen dos desafíos principales en el desarrollo de sistemas afectivos basados en voz. El primero es incrementar la precisión de los sistemas de reconocimiento de voz, y el segundo es el desarrollo e integración de métodos para el procesamiento de información paralingüística. Para abordar estos desafíos, los investigadores necesitan bases de datos con elocuciones genuinas con un gran número de sujetos y muestras. Una colección grande de datos ayuda a obtener una buena representación de los fenómenos bajo estudio y la independencia del hablante en los modelos acústicos entrenados. Al utilizar estos corpus, es más factible el desarrollo de sistemas con la capacidad de reconocer el estado emocional del usuario a partir de información paralingüística y diseñar aplicaciones que adapten automáticamente su comportamiento al estado detectado en los usuarios.

Para nuestro trabajo de investigación en el reconocimiento automático de emociones en el habla, hemos creado dos corpus de habla emocional. El primero se llama EmoWisconsin (Pérez-Espinosa, Carlos Aleberto Reyes-García y Villaseñor-Pineda 2011), que contiene voces de niños participando en un juego de cartas. El segundo se llama IESC-Child (Martínez-Miranda, Pérez-Espinosa y col. 2018) e incluye voces de niños que interactúan con un robot. En ambas bases de datos, inducimos emociones a través de la aplicación de estímulos positivos y negativos.

En la creación del corpus EmoWisconsin, se aplicó el Wisconsin Card Sorting Test (Grant y Berg 1948) con el objetivo de inducir diferentes estados emocionales en los participantes. Los psicólogos utilizan esta prueba para evaluar la

función cognitiva abstracta de los niños. Consiste en tarjetas impresas con figuras geométricas que varían según el color, la forma y la cantidad. El participante, a través de prueba y error, debe deducir el principio de clasificación que el facilitador está solicitando. La prueba fue aplicada por un grupo de psicólogos que modificaron el protocolo original para presentar la actividad como un juego, donde los niños se involucrarían y se expresarían verbalmente. Cada niño participó en dos sesiones: la primera incluye la participación de un facilitador afable, la segunda incluye la participación de un facilitador mal humorado. Durante la sesión con el facilitador afable, se esperaba que el niño experimentara estados emocionales positivos como serenidad, seguridad, motivación y alegría. Por otro lado, durante las sesiones con el facilitador malhumorado, se esperaba que el niño experimentara un estado emocional negativo con emociones como nerviosismo, inseguridad, agobio y frustración. El experimento involucró a 28 niños de entre 7 y 13 años y se registraron un total de 11:39 horas. El post-procesamiento se realizó mediante dos esquemas de anotación: el enfoque categórico y el enfoque continuo.

Para la creación del corpus IESC-Child, se diseñó un escenario de Mago de Oz en el que los niños interactuaron con un par de robots Lego al darles instrucciones para cumplir una misión. Se les explicó a los participantes que tenían que guiar al robot con comandos de voz a través de un laberinto, desde el inicio hasta el punto final. Además, el niño debía indicar al robot que ingresara a algunas estaciones intermedias para recolectar dulces. También, debían evitar algunos obstáculos para no perder los dulces ya obtenidos. Tanto las estaciones como los obstáculos fueron coloreados para representar cuántos dulces podían ganar o perder los niños. Se utilizaron dos robots con la misma apariencia pero personalidad diferente. En el primer robot, se simuló la personalidad agradable y su

comportamiento incluyó el seguimiento de todas las instrucciones proporcionadas por el niño además de felicitarlo cuando se recogían dulces. El otro robot fue modelado con la personalidad opuesta, desagradable, generando que el robot no siguiera algunas veces las órdenes de los niños al soltar algunos obstáculos o ignorar algunas estaciones intermedias para recoger los dulces. Además, el robot desagradable culpaba al niño por la pérdida de dulces.

En el experimento IESC-Child participaron un total de 174 niños entre 6 y 11 años de edad y se obtuvieron 87 horas de grabación de audio y video. Además, se recopilaron datos subjetivos mediante la aplicación de un cuestionario de opción múltiple a cada niño participante. El objetivo de este cuestionario fue recopilar información sobre las preferencias hacia los robots y las emociones que experimentaron los niños. Las elocuciones fueron etiquetadas con emociones discretas así como estados y actitudes mentales. Con la construcción de estas bases de datos del habla, nuestro objetivo es contribuir a la comunidad de investigadores con herramientas para llevar a cabo nuevos experimentos sobre el reconocimiento automático de emociones a partir del habla. Ambas bases de datos están disponibles, bajo pedido, para la comunidad de investigación.

6.3.2. Análisis paralingüístico en niños

Los sistemas interactivos basados en el habla tienen el potencial de identificar automáticamente información paralingüística relevante sobre el usuario, como la edad, el género, el estado emocional o de ánimo actual, la personalidad e incluso el estado de salud del usuario. La identificación automática de estas características puede facilitar la adaptación del contenido del sistema al estado particular del usuario. Por otra parte, los investigadores en tecnología interactiva han examinado cómo se desempeñan los niños como usuarios de las tecnologías

convencionales desarrolladas para adultos, y al hacerlo, han encontrado la necesidad de realizar diseños y adaptaciones particulares (Read y col. 2008). Estos sistemas interactivos adaptados a niños y adolescentes proporcionan un medio adecuado para los usuarios más jóvenes en aplicaciones específicas, como el acceso a diferentes contenidos en Internet y actividades de aprendizaje interactivo (Cucchiarini y Van Hamme 2013).

El éxito en la interacción entre los niños y los sistemas basados en el habla está estrechamente relacionado con la manera en que el sistema pueda identificar características relevantes en el niño y usar esta información para adaptar mejor los contenidos del sistema al perfil de cada niño. Una estrategia para inferir información relevante de los usuarios del sistema basado en el habla (incluidos los niños) es el uso de información paralingüística acústica obtenida del habla del usuario. El análisis de señales no verbales en una conversación, como risas, pausas llenas, silencios y habla traslapada, se utiliza actualmente para extraer información sobre el género, el rol, el tema de conversación, el modo de interacción y la personalidad.

Nuestro grupo de investigación ha trabajado en el desarrollo de algunos modelos de clasificación de información paralingüística. El trabajo presentado en (Pérez-Espinosa, Martínez-Miranda, Avila-George y col. 2018) usamos la base de datos de niños interactuando con un robot descrita en la Sección 6.3.1 para el reconocimiento de dos aspectos paralingüísticos: la clasificación de emociones y la clasificación de actitudes. Probamos tres conjuntos de características acústicas para caracterizar las grabaciones de habla y cuatro algoritmos de aprendizaje máquina para evaluar la dificultad de ambas tareas de clasificación. Para la clasificación de emociones el mejor resultado obtenido fue 55,5 de Medida-F y para la

clasificación de actitudes 72.9 de Medida-F, ambos resultados fueron obtenidos usando el clasificador *Random Forest*.

Por otra parte, en el trabajo reportado en (Pérez-Espinosa, Avila-George y col. 2018) se desarrollaron clasificadores para la identificación de la edad y sexo usando la base de datos descrita en la Sección 6.3.1. En este caso, como algoritmo de clasificación se usó una red neuronal convolutiva. Los parámetros de configuración de la red fueron afinados usando un objeto matemático llamado arreglos de cobertura. Para la clasificación de edad se obtuvo una medida F de 0.34 para las 6 edades (6-11). Al agrupar las edades en 3 y 2 clases, se obtuvo una medida F de 0.50 y 0.71, respectivamente. Para la clasificación del sexo se obtuvo una medida F de 0.71.

6.3.3. Estimación de la calidad de interacción de adultos mayores

Las aplicaciones de vida cotidiana asistida por el entorno (Ambient Assisted Living - AAL) pueden ayudar a personas mayores a realizar sus actividades diarias en su hogar y en su entorno habitual. Las aplicaciones AAL pueden incluir el monitoreo de patrones de comportamiento y el manejo de las actividades diarias en el hogar (Ullberg, Loutfi y Pecora 2014), promover la interacción social (Bisiani y col. 2013), apoyar la movilidad (Angeletou y col. 2013), prevenir y controlar las enfermedades crónicas relacionadas con la edad (Tabak y col. 2013), e identificar situaciones inseguras en el hogar (Moshtaghi, Zukerman y R. A. Russell 2015). La mayoría de estas aplicaciones requieren interacciones continuas y a largo plazo con los usuarios para recopilar datos relevantes y proporcionar recomendaciones útiles. Para aumentar la aceptación de este tipo de tecnología por parte de los usuarios y para aprovechar al máximo el potencial de las aplicaciones AAL, es necesario lograr interacciones más naturales. Por lo tanto, es crucial que

en el diseño de interfaces de usuario interactivas se consideren los cambios sensoriales, motores y cognitivos que ocurren durante el proceso de envejecimiento.

Las tecnologías del habla y el lenguaje pueden facilitar interacciones directas, ayudando así a abordar los problemas de comunicación que experimentan los usuarios con afecciones crónicas y problemas de motricidad fina. Existe evidencia de que entre los usuarios mayores, la interacción basada en el habla se considera menos complicada que las formas tradicionales de interacción, como usar un *mouse* o un teclado (Taveira y Choi 2009). Entre un grupo de adultos mayores se identificó el potencial de usar el habla para proporcionar una forma unificada de interactuar con diferentes servicios y dispositivos (Schlögl, Garschall y Tscheligi 2014). Sin embargo, los sistemas interactivos basados en voz también deben considerar la evolución de las interacciones y los problemas asociados con la comunicación basada en el habla, particularmente en adultos mayores.

En otro de los estudios llevados a cabo en nuestro laboratorio (Pérez-Espinosa, Martínez-Miranda, Espinosa-Curiel y col. 2017) se trabajó con un par de bases de datos de voces de adultos mayores interactuando con un sistema interactivo basado en voz. En este trabajo se plantearon dos objetivos: (i) analizar si existen diferencias entre adultos jóvenes y adultos mayores en términos de la información paralingüística acústica, y evaluar automáticamente la calidad de la interacción con el sistema mediante el uso de los fenómenos paralingüísticos extraídos de estos dos grupos de población; y (ii) explorar el uso de fenómenos paralingüísticos para evaluar automáticamente la calidad de la interacción entre adultos mayores y sistemas interactivos basados en voz. La adaptación personalizada y continua de los contenidos y el estilo de interacción es una característica clave para mejorar la aceptabilidad de este tipo de sistemas (Bresó y col. 2016). Además, el reconocimiento automático de fenómenos paralingüísticos durante las interacciones

puede ayudar a crear mejores modelos de comportamiento humano para aplicaciones basadas en HCI (Vinciarelli y col. 2015), por ejemplo, al mantener un modelo actualizado de los usuarios y su contexto, así como identificar características comunes en un grupo de usuarios (por ejemplo, la población de adultos mayores).

El método propuesto en este estudio para la estimación de la calidad de la interacción de los adultos mayores con el sistema basado en voz se divide en dos partes, en la primera se clasifican automáticamente fenómenos paralingüísticos encontrados en el habla y en el segundo, a partir de la presencia de fenómenos paralingüísticos, se estiman automáticamente algunos aspectos de la calidad de la interacción. El método reconoce automáticamente los fenómenos paralingüísticos (por ejemplo, gritos, hiper-articulación y vacilación) durante la interacción entre el usuario y el sistema. La interacción se caracteriza en función de la ocurrencia de estos fenómenos. Posteriormente, se entrena un modelo utilizando esta caracterización, con el objetivo de estimar la calidad de la interacción. En la etapa de clasificación de fenómenos paralingüísticos se obtuvieron resultados satisfactorios, alcanzando una medida-F de 75.9. En la etapa de estimación de la calidad de la interacción se estimaron tres métricas de calidad de interacción: calificación de la tarea, impresión general de la interacción e impresión general del sistema. El desempeño del sistema se midió calculando la correlación de Pearson entre los valores esperados y los valores estimados de las métricas. Para la calificación de la tarea se obtuvo un 0.78 de correlación, para la impresión general de la interacción se obtuvo 0.79 y para la impresión general del sistema se obtuvo 0.84.

6.3.4. Agentes virtuales conversacionales en salud mental

La principal característica de los agentes virtuales conversacionales es la capacidad de emular la conversación persona a persona a través de la combinación de diálogos generados en un personaje virtual con apariencia humana, y la reproducción de movimientos corporales y expresiones faciales que complementan la comunicación verbal durante la interacción con el usuario. El uso de estos agentes virtuales como asistentes para llevar a cabo actividades relacionadas al cuidado de la salud se ha incrementado en los últimos años. Actualmente existen agentes virtuales con capacidades de facilitar al usuario la búsqueda de información médica en línea (Bickmore y col. 2016); para la ayuda en el diagnóstico de trastornos del sueño (Philip y col. 2014); promover hábitos saludables (Sillice y col. 2018); o con funciones de cuidadores en casa de adultos mayores (Shaked 2017), por nombrar solo algunos. Una de las áreas en las que el desarrollo de agentes virtuales emocionales tienen un impacto relevante, es en el tratamiento terapéutico de trastornos afectivos (Hudlicka y col. 2008).

Agentes virtuales capaces de generar respuestas empáticas durante la interacción con usuarios diagnosticados con algún trastorno afectivo (por ejemplo depresión, ansiedad, síndrome de estrés post-traumático, etc.) pueden ser de gran ayuda como una herramienta tecnológica complementaria al tratamiento clínico. El modelado de reacciones empáticas en agentes con este objetivo presenta retos de investigación importantes. Durante las sesiones interactivas con usuarios que presentan algún trastorno afectivo, no es deseable que el agente muestre reacciones basadas en “*empatía natural*” a través de la adopción del mismo estado emocional –usualmente negativo– detectado en el usuario. La adopción de esta postura *puramente empática* puede ser contraproducente para el usuario, ya que puede ser interpretada como una expresión de condolencia con un sentido

implícito de acuerdo con la situación (negativa) del usuario (Clark 2014). Desde una perspectiva clínica, el tipo de reacción empática a mostrar hacia el usuario debería estar basada en una “*empatía terapéutica*”.

En este sentido, en nuestro laboratorio se están desarrollando un conjunto de agentes virtuales conversacionales con el objetivo de ser utilizados en la ayuda a la detección y tratamiento de trastornos afectivos. Concretamente, se está trabajando en modelos computacionales capaces de generar respuestas emocionales adecuadas respecto a la información que se recibe del usuario. Un ejemplo de estos modelos es el presentado en (Martínez-Miranda, Bresó y García-Gómez 2014b) (Martínez-Miranda, Bresó y García-Gómez 2014a) donde se describe un modelo de empatía terapéutica para generar respuestas empáticas adecuadas de un agente virtual durante la interacción con usuarios bajo tratamiento de depresión. El modelo desarrollado está basado en el concepto de empatía terapéutica definido en (Thwaites y Bennett-Levy 2007), el cual distingue este tipo de empatía respecto a la empatía natural generada en cualquier persona durante situaciones cotidianas. La principal diferencia está en el agregado de un componente cognitivo que ayuda a poner en perspectiva el componente emocional. Esto ayuda al terapeuta a conceptualizar de una manera más adecuada las preocupaciones o situaciones expresadas por el usuario en términos cognitivos, ofreciendo respuestas útiles de acuerdo a cada situación. En otras palabras, la empatía terapéutica ayuda al terapeuta a *involucrarse emocionalmente* para entender la situación del paciente, pero también a *desprenderse* de ese estado emocional para realizar una valoración *más objetiva* de dicha situación.

Basado en esta perspectiva, el modelo emocional de empatía terapéutica implementa el proceso de generación del estado emocional en el agente y el proceso de *desapego emocional* durante eventos concretos producidos en la sesión inter-

activa. La manera de llevar a cabo el *desapego emocional* fue a través de la implementación de un mecanismo para emular el proceso de *regulación* de emociones que se lleva a cabo en individuos, de manera consciente o inconscientemente, para incrementar, mantener o disminuir uno o más componentes de una respuesta emocional (Gross y Thompson 2007). Dos estrategias de regulación emocional, tomadas del modelo teórico propuesto en (Gross 2001) fueron implementadas: (i) la estrategia de *cambio cognitivo*, y (ii) la *modulación de la respuesta*. La estrategia de cambio cognitivo se activa cuando el usuario está reportando al agente virtual alguna situación o evento negativo (por ejemplo, un bajo estado de ánimo). Este evento es evaluado por el agente de acuerdo a sus metas y deseos pre-establecidos. Siguiendo con el ejemplo, si una de los deseos del agente virtual es que el usuario se encuentre con un estado de ánimo positivo, entonces la evaluación de este evento tendrá un resultado negativo y se activará de manera prospectiva una emoción negativa.

Sin embargo, antes de generar la respuesta emocional acorde a la emoción activada (la cual sería una emoción *empática* negativa), la estrategia de regulación emocional involucra la búsqueda de información adicional que pudiera cambiar el significado del evento detectado. Esta búsqueda se realiza en el historial de las sesiones anteriores, y si por ejemplo se encuentra que en las sesiones inmediatas anteriores ha habido una tendencia positiva en el estado de ánimo del usuario, la evaluación de este nuevo evento (o re-evaluación de la situación original) altera la intensidad de la emoción inicial o incluso puede cambiar el tipo de emoción con la que responderá el agente. Este proceso contribuye a realizar una evaluación más objetiva sobre los eventos generados por el usuario y en consecuencia generar una respuesta más útil al usuario. De manera complementaria, la estrategia de modulación de la respuesta se activa cuando la estrategia de

cambio cognitivo no consigue modular la intensidad o la valencia de la emoción original (debido a que no existe información que pueda servir para evaluar de manera más positiva el evento ocurrido). La modulación en la intensidad de la emoción a través de esta estrategia, genera en el agente virtual una reacción que evita el comunicar un sentido de condolencia hacia el usuario.

La emoción e intensidad generadas por el modelo de regulación emocional es utilizado para general el comportamiento verbal y no verbal en el agente que se mostrará durante su respuesta hacia el usuario. Siguiendo con el ejemplo, si al aplicar la estrategia de cambio cognitivo se encuentra una tendencia positiva en el estado de ánimo del usuario, la respuesta en lenguaje natural originada por la emoción e intensidad finalmente generadas se construye para comunicar algo como: “*Entiendo, parece ser que tu condición actual no es muy buena, sin embargo es mejor que en días pasados y esto es algo positivo*”. Este tipo de respuesta es más útil que aquella que pudo haber sido generada tras haber evaluado el evento original sin realizar una re-evaluación con información adicional y relacionada a lo que el usuario está reportando.

En otro trabajo se presenta la implementación de diferentes estrategias de regulación emocional llevadas a cabo en función del tipo de personalidad modelado en el agente virtual (Martínez-Miranda y Alvarado 2017). Este modelo está basado en la evidencia de que la selección y ejecución de las distintas estrategias para regular las emociones no son utilizadas de la misma manera por diferentes individuos y están basadas en los diferentes tipos de personalidad (John y Gross 2007). Las diferencias en el proceso de regulación de emociones generará también diferentes estilos de interacción de acuerdo al tipo de personalidad representado en el agente virtual. Este modelo actualmente en desarrollo se integrará en un conjunto de agentes virtuales que se están utilizando en una App



Figura 6.3: Aplicación móvil *HelPath* con agentes virtuales como principal medio de interacción con usuarios detectados con conducta suicida.

de prevención y detección de comportamientos suicidas (Figura 6.3), la cual es parte de un proyecto en ejecución financiado por el Fondo Sectorial en Salud y Seguridad Social de CONACyT (FOSISS-CONACyT). El uso de este tipo de aplicaciones facilitan tener un seguimiento remoto y continuo con este tipo de usuarios con el objetivo de minimizar la ocurrencia de episodios de suicidios. A través de los agentes virtuales se puede recabar información que ayude a valorar mejor la condición del usuario durante un periodo de tiempo; proveer información terapéutica como ayuda inmediata; y a detectar signos de alerta asociados a conductas suicidas que generen alertas a los familiares y/o especialistas para fomentar un contacto directo con el usuario (Martínez-Miranda 2017).

6.4. Conclusiones

En este capítulo se han presentado algunos de los recientes trabajos de investigación sobre el uso del lenguaje natural en el desarrollo de sistemas interactivos capaces de reconocer de manera automática el estado emocional en los usuarios, y de emular respuestas afectivas hacia ellos. A pesar de los avances en las dos áreas, el reconocimiento de emociones a partir de la voz y el modelado de respuestas emocionales, aún hay bastantes retos abiertos de investigación en ambas áreas. Por ejemplo, uno de los problemas en el reconocimiento de emociones a través de información paralingüística es el poder interpretar correctamente las emociones en escenarios naturalistas en los cuales las emociones no son generadas de manera prototípica, sino como una mezcla de emociones con mayor o menor intensidad. Algunos autores han demostrado que en ciertas situaciones, las personas pueden experimentar más de una emoción al mismo tiempo, lo cual se refleja en los distintos fenómenos paralingüísticos representando un reto importante para el reconocimiento automático de emociones.

Adicionalmente, las expresiones emocionales de las personas dependen en gran medida de su entorno, idiosincrasia y cultura (Picard 2003), sin embargo la mayoría de los trabajos de investigación se centran en un análisis monolingüe sin tomar en cuenta estas diferencias en los usuarios. Y aún en el mismo idioma, se presentan diferencias en el tipo de habla acorde a las diferentes regiones de origen, lo que significa que los modelos entrenados para hablantes de una región pueden no ser precisos para hablantes de una región diferentes. Algunos trabajos ya han explorado el reconocimiento de emociones en información paralingüística en más de un idioma (Pérez-Espinosa, Carlos A Reyes-García y Villaseñor-Pineda 2011; Brester, Semenkin y Sidorov 2016).

Por otro lado, la síntesis de comportamientos afectivos, por ejemplo en agentes virtuales o robots sociales, usualmente es llevada a cabo mediante el desarrollo de un modelo computacional basado en diversas teorías emocionales existentes muchas veces enfocadas a explicar diferentes perspectivas de los procesos afectivos. Por lo tanto, es esencial la elección del modelo o arquitectura computacional de emociones que mejor cumpla con los objetivos de la aplicación. Una desventaja de la proliferación de estos modelos basados en diferentes teorías es la dificultad de evaluar su funcionalidad independientemente del dominio de aplicación, así como el llevar a cabo un análisis comparativo entre los diferentes modelos existentes (Broekens, Bosse y S. C. Marsella 2013). Algunos autores proponen la *estandarización* del proceso de desarrollo dividiendo las teorías emocionales existentes en sus supuestos básicos y reformularlos en un marco conceptual común definido a través de un lenguaje formal (Reisenzein y col. 2013).

En la segunda parte del capítulo se presentan algunos de los trabajos que se están llevando actualmente en nuestro grupo de investigación en CICESE-UT₃. En particular, se describen los esfuerzos realizados en las dos vertientes principales para el desarrollo de sistemas interactivos emocionales, en los cuales el uso del lenguaje natural juega un rol fundamental. Los experimentos realizados para la recopilación de un conjunto de grabaciones de voz en niños, la identificación de emociones, actitudes, sexo y edad a partir de información paralingüística de estas grabaciones, así como la evaluación de la calidad de la interacción entre adultos mayores y sistemas interactivos basados en voz tienen como objetivo el desarrollo de mejores modelos de clasificación de estados afectivos que puedan ser integrados en sistemas de interacción para diferentes aplicaciones y usuarios. Por otro lado, el desarrollo de agentes virtuales emocionales involucra la creación y

adaptación de modelos para la generación y regulación de emociones capaces de generar diferentes tipos de reacciones y respuestas empáticas hacia los usuarios con el fin de promover la adherencia a este tipo de sistemas.

Una línea futura de investigación es incluir en los sistemas de detección de emociones una interfaz multi-modal que permita la recolección, fusión y análisis en tiempo real y en escenarios naturalistas de diferentes tipos de datos del usuario, no únicamente la voz, sino además las expresiones faciales y datos fisiológicos producidos durante la interacción. Trabajos recientes demuestran que el avance en técnicas de aprendizaje máquina mejoran la precisión en la detección de estados afectivos utilizando estos datos en conjunto (Mukeshimana y col. 2017). Por otro lado, el reto de desarrollar sistemas interactivos de ayuda al usuario en diversas tareas con los cuales mantenga una interacción a largo plazo y se maximicen los beneficios de estos sistemas involucra no únicamente la generación de respuestas empáticas y coherentes. También es necesario ir ajustando el comportamiento –incluido el emocional– de estos sistemas a partir de interacciones previas que les permitan adaptarse mejor a cambios relevantes detectados en el usuario, así como tener en cuenta características como edad, sexo, contexto cultural, educacional, etc. de los diferentes tipos de usuarios (Vinciarelli y col. 2015). El desarrollo de sistemas interactivos con este tipo de capacidades, facilitará la aceptabilidad por parte de los usuarios a los que van dirigidos no únicamente para ofrecer servicios asistenciales, sino para un largo conjunto de aplicaciones.

Referencias

Anderson, John R y Christian Lebiere (2003). “The Newell test for a theory of cognition”. En: *Behavioral and brain Sciences* 26.5, págs. 587-601.

- Angeletou, Angeliki y col. (2013). "I need to know, i cannot, i don't understand: older users' requirements for a navigation application". En: *Assistive Technology: From Research to Practice* 33, págs. 34-39.
- Barrett, Lisa Feldman (2006). "Are emotions natural kinds?" En: *Perspectives on psychological science* 1.1, págs. 28-58.
- Batliner, Anton y col. (2003). "How to find trouble in communication". En: *Speech communication* 40.1-2, págs. 117-143.
- Bickmore, Timothy W y col. (2016). "Improving access to online health information with conversational agents: a randomized controlled experiment". En: *Journal of medical Internet research* 18.1, e1.
- Binali, Haji, Chen Wu y Vidyasagar Potdar (2010). "Computational approaches for emotion detection in text". En: *4th IEEE International Conference on Digital Ecosystems and Technologies*. IEEE, págs. 172-177.
- Bisiani, Roberto y col. (2013). "Fostering social interaction of home-bound elderly people: the easyreach system". En: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, págs. 33-42.
- Bresó, Adrián y col. (2016). "A novel approach to improve the planning of adaptive and interactive sessions for the treatment of Major Depression". En: *International Journal of Human-Computer Studies* 87, págs. 80-91.
- Brester, Christina, Eugene Semenkin y Maxim Sidorov (2016). "Multi-objective heuristic feature selection for speech-based multilingual emotion recognition". En: *Journal of Artificial Intelligence and Soft Computing Research* 6.4, págs. 243-253.

- Broekens, Joost, Tibor Bosse y Stacy C Marsella (2013). "Challenges in computational modeling of affective processes". En: *IEEE Transactions on Affective Computing* 4.3, págs. 242-245.
- Burkhardt, Felix y Nick Campbell (2014). "Emotional speech synthesis". En: Oxford University Press New York, págs. 286-295.
- Clark, Arthur J (2014). *Empathy in counseling and psychotherapy: Perspectives and practices*. Routledge.
- Cockton, Gilbert (2001). "Usability Evaluation". En: *The encyclopedia of human-computer interaction*. Ed. por Mads Soegaard y Rikke Friis Dam. Interaction Design Foundation.
- Crumpton, Joe y Cindy L Bethel (2016). "A survey of using vocal prosody to convey emotion in robot speech". En: *International Journal of Social Robotics* 8.2, págs. 271-285.
- Cucchiari, C y H Van Hamme (2013). "The JASMIN speech corpus: recordings of children, non-natives and elderly people". En: *Essential Speech and Language Technology for Dutch*. Ed. por Peter Spyns y Jan Odijk. Springer, págs. 43-59.
- Day, Mary Carol y Susan J Boyce (1993). "Human Factors in Human-Computer System Design". En: *Advances in computers*. Vol. 36. Elsevier, págs. 333-430.
- Dias, João y Ana Paiva (2005). "Feeling and reasoning: A computational model for emotional characters". En: *Portuguese Conference on Artificial Intelligence*. Springer, págs. 127-140.
- Doyle, Jon (2006). *Extending Mechanics to Minds: The Mechanical Foundations of Psychology and Economics*. Cambridge University Press.
- Ekman, Paul (1992). "An argument for basic emotions". En: *Cognition & emotion* 6.3-4, págs. 169-200.

- Elliott, Clark (1992). "The Affective Reasoner: A Process Model of Emotions in a Multiagent System". Tesis doct. PhD thesis, The Institute for the Learning Sciences, Northwestern University.
- Fayek, Haytham M, Margaret Lech y Lawrence Cavedon (2017). "Evaluating deep learning architectures for Speech Emotion Recognition". En: *Neural Networks* 92, págs. 60-68.
- Gebhard, Patrick (2005). "ALMA: a layered model of affect". En: *Proceedings of the fourth international joint conference on Autonomous agents and multi-agent systems*. ACM, págs. 29-36.
- Grant, David A y Esta Berg (1948). "A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem." En: *Journal of experimental psychology* 38.4, págs. 404-411.
- Gross, James J (2001). "Emotion regulation in adulthood: Timing is everything". En: *Current directions in psychological science* 10.6, págs. 214-219.
- Gross, James J y Ross A Thompson (2007). "Emotion regulation: Conceptual foundations". En: *Handbook of Emotion Regulation*. Ed. por James J Gross. Guilford Publications, págs. 3-24.
- Hudlicka, Eva y col. (2008). "Panel on Artificial Agents for Psychotherapy." En: *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, págs. 60-64.
- John, Oliver P y James J Gross (2007). "Individual differences in emotion regulation". En: ed. por James J Gross, págs. 351-372.
- Johnson-Laird, Philip N y Keith Oatley (1992). "Basic emotions, rationality, and folk theory". En: *Cognition & Emotion* 6.3-4, págs. 201-223.
- Lazarus, Richard S y Richard S Lazarus (1991). *Emotion and adaptation*. Oxford University Press on Demand.

- LeDoux, Joseph (1996). *The emotional brain*. Simon & Schuster, New York.
- Lee, Chi-Chun y col. (2011). "Emotion recognition using a hierarchical binary decision tree approach". En: *Speech Communication* 53.9-10, págs. 1162-1171.
- Lotfian, Reza y Carlos Busso (2017). "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings". En: *IEEE Transactions on Affective Computing*.
- Lubis, Nurul y col. (2014). "Construction and analysis of Indonesian emotional speech corpus". En: *2014 17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA)*. IEEE, págs. 1-5.
- Lyakso, Elena y col. (2015). "EmoChildRu: emotional child Russian speech corpus". En: *International Conference on Speech and Computer*. Springer, págs. 144-152.
- Manaris, Bill (1998). "Natural language processing: A human-computer interaction perspective". En: *Advances in Computers*. Vol. 47. Elsevier, págs. 1-66.
- Mao, Wenji y Jonathan Gratch (2006). "Evaluating a computational model of social causality and responsibility". En: *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*. ACM, págs. 985-992.
- Marsella, S y J Gratch (2009). "Ema: A model of emotional dynamics". En: *Cognitive Systems Research* 10.1, págs. 70-90.
- Marsella, Stacy, Jonathan Gratch y Paolo Petta (2010). "Computational models of emotion". En: *A Blueprint for Affective Computing-A sourcebook and manual*. Ed. por K R Scherer, T. Bänziger y E B Roesch. Oxford University Press New York, NY, págs. 21-46.

- Martínez-Miranda, Juan (2017). “Embodied conversational agents for the detection and prevention of suicidal behaviour: current applications and open challenges”. En: *Journal of medical systems* 41.9, pág. 135.
- Martínez-Miranda, Juan y Matías Alvarado (2017). “Modelling Personality-based Individual Differences in the Use of Emotion Regulation Strategies”. En: *Canadian Conference on Artificial Intelligence*. Springer, págs. 361-372.
- Martínez-Miranda, Juan, Adrián Bresó y Juan Miguel García-Gómez (2014a). “Look on the bright side: a model of cognitive change in virtual agents”. En: *International Conference on Intelligent Virtual Agents*. Springer, págs. 285-294.
- (2014b). “Modelling two emotion regulation strategies as key features of therapeutic empathy”. En: *Emotion Modeling: Towards Pragmatic Computational Models of Affective Processes*. Springer, págs. 115-133.
- Martínez-Miranda, Juan, Humberto Pérez-Espinosa y col. (2018). “Age-based differences in preferences and affective reactions towards a robot’s personality during interaction”. En: *Computers in Human Behavior* 84, págs. 245-257.
- Mehrabian, Albert y James A Russell (1974). *An approach to environmental psychology*. the MIT Press.
- Mohri, Mehryar, Afshin Rostamizadeh y Ameet Talwalkar (2012). *The Foundations of Machine Learning*. the MIT Press.
- Moshtaghi, Masud, Ingrid Zukerman y R Andrew Russell (2015). “Statistical models for unobtrusively detecting abnormal periods of inactivity in older adults”. En: *User Modeling and User-Adapted Interaction* 25.3, págs. 231-265.
- Mukeshimana, Michele y col. (2017). “Multimodal Emotion Recognition for Human-Computer Interaction: A Survey”. En: *International Journal of Scientific & Engineering Research* 8, págs. 1289-1301.

- Paiva, Ana y col. (2005). "Learning by feeling: Evoking empathy with synthetic characters". En: *Applied Artificial Intelligence* 19.3-4, págs. 235-266.
- Panksepp, Jaak (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.
- Pérez-Espinosa, Humberto, Himer Avila-George y col. (2018). "Children Age and Gender Classification Based on Speech Using ConvNets". En: *Research in Computing Science* 147.4, págs. 23-35.
- Pérez-Espinosa, Humberto, Juan Martínez-Miranda, Himer Avila-George y col. (2018). "Analyzing children's affective reactions and preferences towards social robots using paralinguistic and self-reported information". En: *Journal of Intelligent & Fuzzy Systems* 34.5, págs. 3313-3324.
- Pérez-Espinosa, Humberto, Juan Martínez-Miranda, Ismael Espinosa-Curiel y col. (2017). "Using acoustic paralinguistic information to assess the interaction quality in speech-based systems for elderly users". En: *International Journal of Human-Computer Studies* 98, págs. 1-13.
- Pérez-Espinosa, Humberto, Carlos A Reyes-García y Luis Villaseñor-Pineda (2011). "Bilingual acoustic feature selection for emotion estimation using a 3d continuous model". En: *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, págs. 786-791.
- Pérez-Espinosa, Humberto, Carlos Aleberto Reyes-García y Luis Villaseñor-Pineda (2011). "EmoWisconsin: an emotional children speech database in Mexican Spanish". En: *International Conference on Affective Computing and Intelligent Interaction*. Springer, págs. 62-71.
- Philip, Pierre y col. (2014). "Could a virtual human be used to explore excessive daytime sleepiness in patients?" En: *Presence: teleoperators and virtual environments* 23.4, págs. 369-376.

- Picard, Rosalind W (1997). *Affective computing*. MIT Press.
- (2003). “Affective computing: challenges”. En: *International Journal of Human-Computer Studies* 59.1-2, págs. 55-64.
- Pittermann, Angela y Johannes Pittermann (2006). “Getting Bored with HTK? Using HMMs for Emotion Recognition from Speech Signals”. En: *8th international Conference on Signal Processing*. Vol. 1. IEEE.
- Pittermann, Johannes, Angela Pittermann y Wolfgang Minker (2010). *Handling emotions in human-computer dialogues*. Springer.
- Read, Janet C y col. (2008). “Child computer interaction”. En: *CHI’08 extended abstracts on Human factors in computing systems*. ACM, págs. 2419-2422.
- Reilly, W Scott (1996). “Believable Social and Emotional Agents”. Tesis doct. Carnegie-Mellon Univ Pittsburgh pa Dept of Computer Science.
- Reisenzein, Rainer y col. (2013). “Computational modeling of emotion: Toward improving the inter-and intradisciplinary exchange”. En: *IEEE Transactions on Affective Computing* 4.3, págs. 246-266.
- Russell, James A (2003). “Core affect and the psychological construction of emotion.” En: *Psychological review* 110.1, pág. 145.
- Sander, David y Klaus Scherer (2009). *Oxford companion to emotion and the affective sciences*. Oxford University Press.
- Scherer, Klaus R (2010). “Emotion and emotional competence: conceptual and theoretical issues for modelling agents”. En: *Blueprint for affective computing: A sourcebook*, págs. 3-20.
- Scherer, Klaus R, Angela Schorr y Tom Johnstone, eds. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

- Schlögl, Stephan, Markus Garschall y Manfred Tscheligi (2014). “Designing natural language user interfaces with elderly users”. En: *Workshop on Designing Speech and Language Interactions*.
- Schmitt, Alexander y Wolfgang Minker (2012). *Towards adaptive spoken dialog systems*. Springer Science & Business Media.
- Schröder, Marc (2001). “Emotional speech synthesis: A review”. En: *Seventh European Conference on Speech Communication and Technology*, págs. 561-564.
- Schuller, Björn (2015). “Deep learning our everyday emotions”. En: *Advances in neural networks: computational and theoretical issues*. Ed. por S Bassis, A Esposito y F Morabito. Springer, págs. 339-346.
- Shaked, Nava A (2017). “Avatars and virtual agents—relationship interfaces for the elderly”. En: *Healthcare technology letters* 4.3, págs. 83-87.
- Shelke, Nilesh M (2014). “Approaches of emotion detection from text”. En: *International Journal of Computer Science and Information Technology* 2.2, págs. 123-128.
- Sillice, Marie A y col. (2018). “Using relational agents to promote exercise and sun protection: Assessment of participants’ experiences with two interventions”. En: *Journal of medical Internet research* 20.2, e48.
- Skansi, Sandro (2018). *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer.
- Sloman, Aaron y Monica Croucher (1981). “Why robots will have emotions”. En: *Proc. of International Joint Conference on Artificial Intelligence, Vancouver, Canada*.
- Steidl, Stefan (2009). “Automatic classification of emotion related user states in spontaneous children’s speech”. Tesis doct. Universität Erlangen-Nürnberg.

- Strapparava, Carlo y Rada Mihalcea (2015). “Affect Detection in Texts”. En: *The Oxford handbook of affective computing*. Ed. por Rafael Calvo y col. Oxford University Press, USA, págs. 184-203.
- Tabak, Monique y col. (2013). “Acceptance and usability of an ambulant activity coach for patients with COPD”. En: *Proceedings of the IADIS International conference e-Health 2013*. IADIS press, págs. 61-68.
- Taveira, Alvaro D y Sang D Choi (2009). “Review study of computer input devices and older users”. En: *Intl. Journal of Human-Computer Interaction* 25.5, págs. 455-474.
- Thwaites, Richard y James Bennett-Levy (2007). “Conceptualizing empathy in cognitive behaviour therapy: Making the implicit explicit”. En: *Behavioural and Cognitive Psychotherapy* 35.5, págs. 591-612.
- Ullberg, Jonas, Amy Loutfi y Federico Pecora (2014). “A customizable approach for monitoring activities of elderly users in their homes”. En: *International Workshop on Activity Monitoring by Multiple Distributed Sensing*. Springer, págs. 13-25.
- Vetterli, Martin, Jelena Kovačević y Vivek K Goyal (2014). *Foundations of signal processing*. Cambridge University Press.
- Vinciarelli, Alessandro y col. (2015). “Open challenges in modelling, analysis and synthesis of human behaviour in human-human and human-machine interactions”. En: *Cognitive Computation* 7.4, págs. 397-413.
- Vogt, Thurid y Elisabeth André (2009). “Exploring the benefits of discretization of acoustic features for speech emotion recognition”. En: *Tenth Annual Conference of the International Speech Communication Association*.

Capítulo 7

Retos y Perspectivas

Héctor Jiménez Salazar, UAM-Cuajimalpa.

Luis Villaseñor Pineda, INAOE.

La utilidad de las Tecnologías del Lenguaje (TL) es clara, desde hace años las usamos cotidianamente integradas en infinidad de aplicaciones. Gracias a ellas podemos aprovechar la gran cantidad de información –expresada en lenguaje natural– disponible en Internet. Por un lado, estas tecnologías nos ayudan a gestionar esta gran cantidad de información permitiendo encontrar una aguja en un pajar; prácticamente todo el mundo con acceso a Internet conoce las posibilidades del buscador de Google. Del otro lado de la moneda, la gran cantidad de información acumulada es de enorme valor que al ser procesada y analizada permite descubrir patrones que permiten predecir y/o mejorar las prácticas de las organizaciones. Un ejemplo es el aprovechamiento de la gran cantidad de información circulando en las redes sociales como Twitter o Facebook que al aplicar las TL es posible observar tendencias de opinión sobre una marca, aceptación/rechazo de un nuevo producto o incluso de una nueva política gubernamental; o también, nos dan información sobre el autor de los mensajes, como

sus intereses o preferencias políticas, incluso se puede determinar su género o rasgos sobre su personalidad.

Por otro lado, gracias a las TL es posible enfrentar la barrera de los idiomas y nos permiten ir más allá del mundo hispanohablante. Es un hecho que en este mundo, cada vez más globalizado, la búsqueda de nuevos mercados es primordial. Nuevos mercados implican, por lo general, nuevos idiomas. La capacidad de acceder u ofrecer servicios o productos de/y a otros países, se ve obstaculizada por la barrera lingüística. Una simple búsqueda de un producto especializado por un empresario en México, se limita generalmente a proveedores que publicitan sus productos en español o inglés. Mecanismos que permitan una comunicación rápida y efectiva en otros idiomas son indispensables.

Paradójicamente, a pesar de que las TL están detrás de todas estas aplicaciones, existe un enorme desconocimiento de lo que son y de lo que estas tecnologías nos brindan. En buena medida, esto se debe a que, si estas tecnologías funcionan correctamente, ellas pasan inadvertidas. Desafortunadamente, esta situación ha jugado en su contra, evitando que tanto expertos de otras áreas como el empresariado mexicano, adviertan su importancia y se involucren en su desarrollo.

Impulsar las TL repercutirá en la creación de innovadores productos y servicios en infinidad de sectores, dado que la presencia del lenguaje es indispensable en toda actividad humana. Aunado a esto, el tamaño del mercado es enorme; y no sólo porque México es el país hispanohablante más poblado, sino que el español es una de las lenguas más habladas en el mundo, en 2019 se contaban poco mas de 460 millones de personas cuya lengua materna es el español –sólo superado por el chino mandarín– y de los cuales la enorme mayoría se concentra en Latinoamérica. El tratamiento del idioma español por medios automáticos

es claramente una necesidad para esta población (los usuarios de habla hispana ocupan el tercer puesto en el *ranking* de usuarios de internet por idioma¹), y trabajar en ello no sólo se justifica por el enorme mercado que representa sino también porque una lengua es un reflejo social y cultural propios de una comunidad.

Otro gran reto para el desarrollo de las TL en México, es la necesidad de trabajar en la constitución de recursos lingüísticos² representativos de nuestras lenguas. La carencia de recursos lingüísticos ha provocado un lento desarrollo, e incluso ha instado a que los investigadores mexicanos en el área, experimenten sus ideas sobre aquellas lenguas que si cuentan con dichos recursos lingüísticos. Claro está que los recursos constituidos para el español de España son una base importante, no obstante éstos deben adecuarse y enriquecerse con las muchas variantes existentes en nuestro país. Además, las iniciativas provenientes de otro país parten de contextos diferentes, de ahí que los problemas, y por ende, los recursos a constituir tengan prioridades distintas. Esta situación no es exclusiva de México, de hecho la agenda de investigación en TL ha estado subordinada a las necesidades y particularidades del idioma inglés. Esta situación ha provocado

¹Informe CIDIC. El español en la red. Centro de Investigación y Documentación del Instituto Cervantes. Informe en línea, fecha de acceso 20/agosto/2014: http://www.cervantes.es/imagenes/File/cidic/el_espanol_en_la_red.pps

²Por recursos lingüísticos nos referimos a toda aquella información que describe la estructura de una lengua. Habitualmente, los recursos lingüísticos se agrupan en tres grandes categorías: corpus, recursos léxicos y gramáticas. Tomando las definiciones de (Listerri 2003) un corpus puede definirse como un conjunto estructurado de textos que constituyen una muestra lo más realista posible del uso de la lengua. Los recursos léxicos son diccionarios electrónicos ya sea monolingües o multilingües. Por último, una gramática computacional puede entenderse como una descripción formalizada del conocimiento lingüístico que, en el marco del procesamiento del lenguaje, puede ser empleada tanto para realizar análisis como generación de textos.

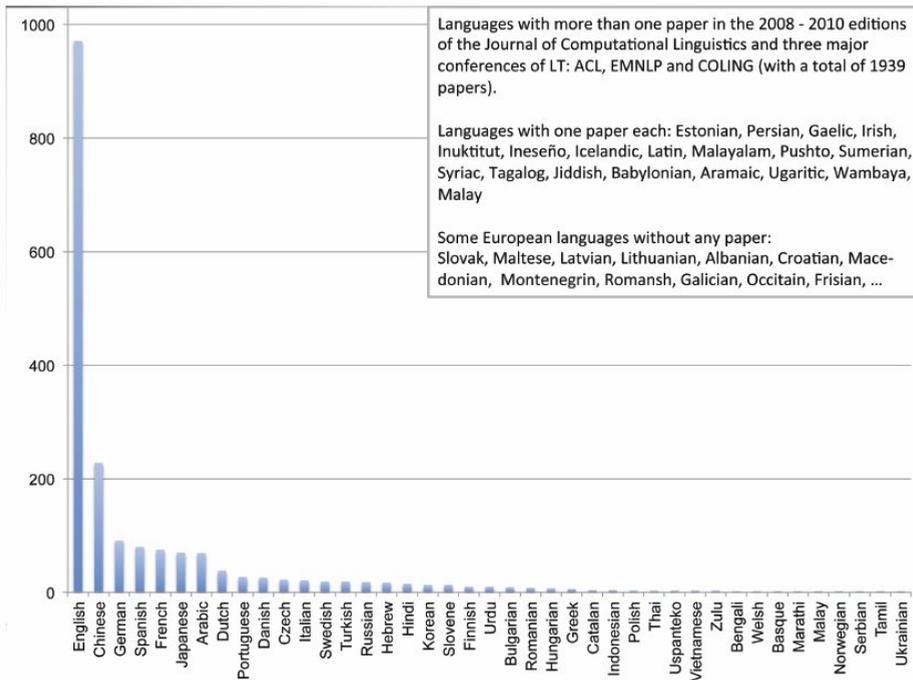


Figura 7.1: Número de publicaciones por idioma tratado. Trabajos publicados entre 2008-2010 en la revista Computational Linguistics y los congresos de la ACL, EMNLP y COLING. (Figura tomada de *Rehm, Georg y Hans Uszkoreit (2013). META-NET strategic research agenda for multilingual Europe 2020. Springer.*)

un rezago en el tratamiento automático del lenguaje en muchas otras lenguas, incluyendo el español (véase la Figura 7.1). Es así que recopilar y crear recursos lingüísticos para el español de México -y nuestras lenguas originarias- es una tarea primordial. El no contar con ellos, está dejando la puerta abierta para que la tarea de investigación -y consecuentemente el desarrollo de aplicaciones- sea retomada por equipos de investigación extranjeros.

Otro problema, no exclusivo de las TL, es la poca relación entre la academia y la industria. Es apenas en años recientes, con el impulso de programas gubernamentales,

mentales, que se han iniciado algunos acercamientos. Sin embargo, aún se está lejos de incidir verdaderamente en la industria nacional o de crear empresas de base tecnológica. No basta con que la industria subcontrate a un centro de investigación para resolver un problema concreto, sino que el problema a resolver implique una contribución científica. Para ello, es necesario adentrarse en las necesidades de la industria nacional y realizar investigación bajo esa problemática donde, por un lado exista posibilidad de avanzar la frontera de la ciencia, y por otro lado, los métodos propuestos estén orientados a solucionar necesidades concretas del contexto nacional.

Finalmente, cabe mencionar que la comunidad mexicana del área se ha organizado para enfrentar los retos anteriores. Para ello, se han utilizado varios mecanismos. Desde aquellos impulsados por el CONACYT, como el programa de Redes Temáticas, a través del cual la comunidad conformó la Red temática en Tecnologías del Lenguaje³; o a través de la participación y/o creación de asociaciones científicas como la misma Academia Mexicana de Computación⁴ y la Asociación Mexicana de Procesamiento de Lenguaje Natural⁵. Gracias a estas organizaciones se ha elaborado un diagnóstico del área, con la consecuente creación de un plan de trabajo incluyendo acciones concretas para desafiar dichos retos antes mencionados.

³redttl.mx

⁴amexcomp.org.mx

⁵ampln.mx

Panorama de las Tecnologías del Lenguaje en México,
se terminó el 30 de septiembre de 2019.

A partir de enero de Diciembre de 2019 está disponible en formato PDF en la
página de la Academia Mexicana de Computación:

<http://www.amexcomp.mx>