



REVISTA COLOMBIANA DE PSQUIATRIA

www.elsevier.es/rcp



Metodología de investigación y lectura crítica de estudios

El procesamiento de lenguaje natural y su relación con la investigación en salud mental

The Natural Language Process and Its Relationship With Mental Health Research

Jhon Camacho^{a,b,*}, Socorro Moreno^a, Fernando Suarez-Obando^c,
Juan Carlos Puyana^d y Carlos Gómez-Restrepo^a

^aDepartamento de Epidemiología y Bioestadística, Pontificia Universidad Javeriana, Bogotá, Colombia

^bDepartment of Biomedical Informatics, University of Pittsburgh, Pennsylvania, Estados Unidos

^cInstituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Colombia

^dDepartment of Surgery, University of Pittsburgh, Pennsylvania, Estados Unidos

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 21 de febrero de 2013

Aceptado el 24 de abril de 2013

Introducción

Uno de los principales motivos para la introducción de computadores en ambientes clínicos fue la posibilidad de compilar y analizar los grandes volúmenes de información consignados en las historias clínicas. A través de la introducción de esta tecnología, se buscó responder a preguntas aparentemente sencillas como: ¿qué pacientes presentaron determinada condición?, ¿qué clase de ayudas diagnósticas se utilizaron para detectar la condición?, ¿cuáles fueron los resultados de las pruebas realizadas? o ¿cuál fue el tratamiento instaurado? Han pasado varias décadas desde que los computadores entraron a los hospitales, y son innegables los avances en los sistemas de información clínica. Sin embargo, en muchos casos aún es difícil responder a estas preguntas confiablemente.

Una de las razones detrás de esta dificultad es que **gran cantidad de la información clínica se registra en texto libre** (principalmente en notas de evolución) y la mayoría de los sistemas de información clínica actuales no son óptimos para trabajar con texto libre. Por ejemplo, luego de examinar a un paciente con fiebre, un médico podría anotar en la historia clínica: «Paciente en estado febril» o «El paciente presenta fiebre» o «Paciente con temperatura de 39,5». Para un computador estas tres opciones son completamente diferentes, mientras que para el médico tratante las tres notas son prácticamente equivalentes. Para un programa es mucho más fácil entender expresiones del tipo «Fiebre = verdadero». La situación se vuelve más compleja cuando se consideran aspectos como la temporalidad; por ejemplo, ante estas dos notas: «El paciente presenta fiebre» y «El paciente reporta haber tenido fiebre hace 2 meses», la mayoría de los sistemas de información clí-

*Autor para correspondencia.

Correo electrónico: jjcamachosanchez@gmail.com (J. Camacho).

nica solamente identifican que en ambas notas se menciona la palabra fiebre, sin identificar si la condición está en curso o sucedió con anterioridad.

Un campo de la computación llamado procesamiento de lenguaje natural (PLN) se está aplicando a documentos médicos redactados en texto libre para construir bases de datos que los programas puedan entender y analizar¹. Utilizando las técnicas de PLN, un programa puede leer un fragmento de una nota de evolución como el siguiente: «Paciente en estado febril, se rechaza neumonía. Historia familiar de cáncer de mama» y transformarlo en una serie de atributos como los que aparecen en la tabla 1.

Una vez obtenidos estos atributos a partir del texto libre, otros programas pueden utilizar esta información para realizar análisis estadísticos o guiar las decisiones clínicas².

Además de su utilidad para el apoyo de la asistencia clínica, el PLN permite realizar gran variedad de estudios clínicos basados en información textual²⁻¹⁵ que de otra forma requerirían que médicos expertos invirtieran gran cantidad de tiempo en leer y codificar los textos. Aun cuando el desarrollo de este tipo de sistemas usualmente requiere el concurso de expertos para anotar y clasificar documentos, su poder radica en que, una vez desarrollados, se pueden utilizar para analizar automáticamente nuevos documentos del mismo dominio. El PLN también se ha empleado para detectar y eliminar automáticamente información de identidad de historias clínicas, lo que permite generar conjuntos de datos anónimos¹¹ y la recuperación automática de información basada en literatura clínica^{4,6} y registros médicos¹⁵.

Este artículo presenta los conceptos básicos del PLN y plantea algunas de sus aplicaciones en investigación clínica, específicamente en el área de la salud mental, con lo que se busca estimular la aplicación de tecnologías informáticas que permitan aprovechar la información derivada de la compleja atención psiquiátrica.

Componentes del procesamiento de lenguaje natural

Un sistema de PLN se basa en el reconocimiento de conceptos en el texto y la comprensión de las relaciones entre esos conceptos. Este proceso consiste en una serie de tareas: segmentación del texto en palabras o frases, reconocimiento de conceptos médicos y extracción del contexto². En conjunto, estas tareas permiten que el programa detecte los conceptos clínicos a que se refiere el texto, los sujetos relacionados con dichos conceptos y aspectos del contexto, como la temporalidad de los eventos.

Segmentación

Para analizarlo, el texto debe ser segmentado en palabras (p. ej., depresión) o frases con sentido propio (p. ej., trastorno depresivo). Frecuentemente, esta tarea requiere preprocesar el texto para corregir la ortografía, expandir siglas y eliminar información no relevante, como tablas, gráficos o caracteres indeseados^{1,16}. La segmentación implica varios retos, dado que una palabra puede estar delimitada de muchas formas: espacios en blanco, tabulaciones, guiones y signos de puntuación, entre otros. Al mismo tiempo, algunos signos de puntuación pueden formar parte de palabras: por ejemplo, los acrónimos de algunos genes contienen puntos (p. ej., M03F4.2)¹. En cuanto a los límites de las frases, son difíciles de hallar; por ejemplo, la palabra «depresión» tiene sentido propio, pero también puede ser parte de la frase «depresión reactiva», en la que la palabra tiene un sentido relacionado pero particular a un contexto. La delimitación de las frases requiere procesos adicionales después de la segmentación de palabras, a menudo basados en el conocimiento del dominio al que pertenece el texto¹⁶.

Etiquetado de categorías gramaticales

El proceso de etiquetado asigna una categoría gramatical (artículo, sustantivo, pronombre, verbo, adjetivo, adverbio, preposición, conjunción o interjección) a cada palabra o frase¹⁶. Esta tarea depende del contexto y es importante, dado que la interpretación de la oración depende en gran medida de la asignación de esas categorías^{1,2,16}. Si se considera la palabra «consulta» en la siguiente frase: «se recomienda consulta por psiquiatría», es fácil para una persona utilizar el contexto de la oración para clasificarla como un sustantivo; sin embargo, para un computador puede tratarse tanto de un sustantivo como de un verbo. Un programa podría superar este dilema basándose en las categorías gramaticales de las palabras anteriores o posteriores, así que se puede crear reglas que asignen o modifiquen la categoría de acuerdo con el contexto. Una regla de este tipo puede estipular que: «se debe reemplazar verbo por sustantivo si la etiqueta anterior es un verbo»¹, de tal modo que «consulta» sólo se puede etiquetar como sustantivo, dado que la antecede una palabra etiquetada como verbo: «recomienda». Otra aproximación a este problema es el uso de modelos estadísticos que consideran las relaciones entre etiquetas, asignando probabilidades de pertenencia de una palabra a una categoría gramatical específica acorde con el contexto¹⁷.

Reconocimiento de entidades

Mediante reconocimiento de entidades, se detectan los conceptos clínicos en el texto y se cartografían a un concepto determinado en una ontología o un vocabulario controlado^{2,18}. Una forma para lograr este propósito es usar un diccionario que provea una lista de nombres para las entidades del dominio. Un diccionario de este tipo es el Metatesauro del Sistema Unificado de Lenguaje Médico (UMLS por sus siglas

Tabla 1 – Atributos obtenidos a partir del texto

Atributo	Valor
Fiebre	Verdadero
Cáncer de mama	Indeterminado
Neumonía	Falso
Historia familiar de cáncer de mama	Verdadero

en inglés), que agrupa gran cantidad de términos de conceptos médicos en varios idiomas, incluido el español^{19,20}. Sin embargo, el uso de esta estrategia se dificulta al analizar textos que no usan términos estándar para referirse a conceptos médicos o en los que hay gran variabilidad en su uso. Por ejemplo, en el apartado de «motivo de consulta» o en el cuerpo de una narración de experiencias de vida en las que se registran, con lenguaje coloquial, las razones o necesidades de asistencia clínica, es muy probable que no se encuentren equivalentes perfectos con los términos de un diccionario médico. Una alternativa para resolver este problema es **usar modelos estadísticos que relacionen los términos utilizados en el texto libre con los conceptos clínicos estandarizados**, por **ejemplo, estimando la probabilidad de que la palabra «tristeza» sea equivalente a «depresión» en ese contexto**⁶.

Extracción de contexto

Como se evidencia en las secciones anteriores, **el contexto es muy relevante para comprender el texto**. Por ejemplo, **es importante entender si el texto se refiere a que la entidad está presente o ausente**: «el paciente presenta depresión» en contraste con «se descarta depresión». **Lo mismo respecto a la temporalidad de la entidad**: «el paciente presentó depresión» en contraste con «el paciente ha presentado depresión en las últimas 3 semanas». Por otra parte, **es necesario entender quién es el sujeto o el objeto de la entidad**: «paciente con depresión» o «paciente con historia familiar de depresión». Finalmente, **el contexto también puede contener diversos grados de certidumbre sobre la presencia de la enfermedad**: «paciente con depresión mayor» frente a «se sospecha depresión mayor»².

Medidas de evaluación

Con frecuencia, **el desempeño de este tipo de sistema se evalúa comparando sus resultados con un patrón estándar predefinido**. Por ejemplo, **para evaluar el desempeño de un sistema diseñado para detectar pacientes con depresión a partir de los textos en sus historias clínicas, el primer paso es obtener una muestra de historias clínicas de pacientes con sospecha de diagnóstico de depresión. Un grupo de expertos (al menos dos expertos independientes) determinan en cada caso la presencia o ausencia de depresión evaluando las historias clínicas. Este conjunto de datos definido por los expertos constituye el estándar respecto al cual se llevan a cabo las mediciones. En un segundo paso, el sistema analiza las mismas historias clínicas; los resultados de ambos análisis se consignan en una tabla de contingencia similar a la que se emplea en la evaluación de pruebas diagnósticas (fig. 1). A partir de esa tabla se puede calcular la proporción de registros clasificados como «depresión», que pertenecen a esta categoría según el patrón de referencia**; esta proporción se conoce como exhaustividad^{21,22} y se puede interpretar de la misma manera que la sensibilidad de las pruebas diagnósticas. Un valor alto de exhaustividad indica que el sistema

		Patrón de referencia	
		Depresión	Sin depresión
Sistema PLN	Depresión	A	B
	Sin depresión	C	D

$$\text{Exhaustividad} = \frac{A}{A + C}$$

$$\text{Tasa de falsas alarmas} = \frac{B}{B + D}$$

Figura 1 – Tabla de contingencia empleada en la evaluación del desempeño de un sistema de procesamiento de lenguaje natural (PLN).

tiene buena capacidad para detectar los casos positivos (es decir, es altamente sensible). Asimismo se puede calcular la tasa de falsas alarmas¹², que corresponde a la proporción de registros clasificados como «depresión» que no pertenecen a esta categoría según el patrón de referencia; esta medida podría interpretarse como el complemento de la especificidad de pruebas diagnósticas. Un valor bajo en la tasa de falsas alarmas indica que el sistema tiene una probabilidad baja de clasificar como positivo un caso negativo. Por último, se puede estimar la precisión^{21,22}, que corresponde a la proporción de registros que pertenecen a la categoría «depresión» de acuerdo con el patrón de referencia de entre todos los registros que se clasificó como «depresión»; esta medida se interpreta igual que el valor predictivo positivo de las pruebas diagnósticas. Un valor alto en esta medida indica que un caso clasificado como positivo tiene una probabilidad alta de ser efectivamente el diagnóstico definido por el patrón de referencia.

Ejemplos de uso de PLN en investigación en salud mental

Recuperación de registros de asesoría psiquiátrica

PsychPark (<http://www.psychpark.org>) es un servicio de asesoría psiquiátrica en línea, proporcionado por la Asociación Taiwanesa de Informática en Salud Mental (Taiwan Association of Mental Health Informatics)²³. A través de este sitio web, las personas que sufren síntomas depresivos pueden obtener asesoría profesional y consultar las recomendaciones recibidas por pacientes anónimos con síntomas similares. A través de la consulta de los registros de asesoría psiquiátrica, las personas pueden darse cuenta de que otros también padecen los mismos síntomas y obtener información sobre cómo aliviar sus síntomas. Sin embargo, debido a que los pacientes no están familiarizados con la terminología médica, buscar entre los registros de otros pacientes es difícil y toma mucho tiempo. Además, una búsqueda basada exclusivamente en palabras clave excluiría aspectos importantes de los síntomas, como la temporalidad¹⁵.

Con el objetivo de proveer un mecanismo de búsqueda efectivo para recuperar este tipo de registros de asesoría, Yu

et al¹⁵ implementaron un sistema de PLN por el que el usuario de PsychPark puede expresar sus síntomas en lenguaje natural y obtener un listado de registros de asesoría relevantes. A partir de la descripción hecha por el paciente, el sistema reconoce automáticamente síntomas correspondientes en la **escala de Hamilton para la evaluación de depresión**²⁴. Empleando los síntomas reconocidos, el sistema construye una serie de escenarios teniendo en cuenta las relaciones temporales, de causa y efecto y de concomitancia. El mismo proceso se aplica a los registros de asesoría. Finalmente, el sistema calcula la relevancia de un registro de asesoría con respecto a los síntomas del sujeto realizando una suma ponderada de la correspondencia entre las características del cuadro de consulta (síntomas y relaciones entre los síntomas) y el descrito en el registro.

Para extraer los síntomas, el sistema utiliza un proceso de etiquetado, segmentación y extracción de contexto, mediante el cual transforma la descripción en lenguaje natural en un grupo de parámetros. Tomando un ejemplo del estudio de Yu et al¹⁵, a partir de la descripción «Con frecuencia me preocupo por cosas pequeñas», el sistema extrae los siguientes parámetros: *experimentador* = «yo me preocupo», *tiempo* = «frecuentemente me preocupo», *foco* = «me preocupo por cosas», *propiedad* = «cosas pequeñas». Finalmente, estos parámetros se aplican a un modelo probabilístico que calcula la verosimilitud de cada síntoma según la escala de Hamilton y entrega el síntoma de mayor verosimilitud (para este caso, ansiedad)²⁵.

El sistema también extrae las relaciones de causalidad y temporalidad entre los síntomas. Usando otro ejemplo del mismo estudio¹⁵, a partir de la descripción: «Durante los últimos años, me he sentido siempre muy deprimido, debido a lo cual he tratado de suicidarme varias veces. Después de eso, he tenido problemas para quedarme dormido cada noche. Además, en los últimos meses a menudo estallo sin ninguna razón». El sistema extrae los síntomas: «depresión», «intento de suicidio», «insomnio y ansiedad» y las relaciones entre los términos «depresión causa intento de suicidio», «depresión luego insomnio luego ansiedad» e «intento de suicidio luego insomnio luego ansiedad». Con esta información el sistema construye tres escenarios: en el primer escenario, los cuatro síntomas son concomitantes y no se consideran las relaciones; en el segundo, se considera únicamente la relación de causalidad entre depresión e intento de suicidio; por último, el tercer escenario considera exclusivamente las relaciones temporales. Finalmente, se comparan los tres escenarios con los definidos en cada registro de asesoría para calcular la relevancia.

Los autores evaluaron el sistema comparándolo contra herramientas de búsqueda basadas exclusivamente en palabras clave. Durante la evaluación, se utilizaron 50 consultas de prueba y se seleccionaron los 100 registros de mayor relevancia entregados por cada sistema. Posteriormente, un grupo de tres psiquiatras expertos evaluó cada registro y lo clasificó como relevante o irrelevante para la consulta de prueba. El sistema propuesto por Yu et al alcanzó mayor precisión, lo que indica que **la incorporación de los síntomas y sus relaciones tiene un impacto significativo en la eficacia de este tipo de procesos de búsqueda.**

COGNO: el albor de un sistema de retroalimentación para pacientes

La identificación de pensamientos disfuncionales es uno de los fundamentos de la terapia cognoscitiva. La identificación automática de estos pensamientos podría ser el **paso inicial para el desarrollo de un sistema de apoyo** para esta aproximación terapéutica a partir de las descripciones que proveen los pacientes en lenguaje natural.

El **objetivo del proyecto COGNO**¹², desarrollo conjunto de la Northern Illinois University y la DePaul University en Chicago, es **implementar un sistema en el cual un computador sea capaz de tomar un pensamiento disfuncional, clasificarlo dentro de una de las categorías propuestas por Beck**²⁶ (una de las clasificaciones más utilizadas por los terapeutas cognoscitivos) **y proveer al paciente una retroalimentación adecuada que le permita reestructurar dicho pensamiento.**

El paso inicial de este ambicioso sistema fue el desarrollo de un módulo capaz de determinar qué tipo de pensamiento disfuncional se encuentra presente en los textos redactados por un paciente. El sistema emplea reglas derivadas de marcadores del lenguaje identificados en una muestra de 149 pensamientos disfuncionales reunidos a partir de una exhaustiva revisión de literatura.

Para el desarrollo de este módulo, los investigadores derivaron una batería de marcadores lingüísticos que permiten clasificar los pensamientos disfuncionales. Entre estos marcadores, **se incluyeron variados aspectos del lenguaje** como **palabras clave de contenido** (p. ej., éxito, fracaso, logro, etc.), **palabras clave derivadas de categorías gramaticales** (p. ej., adverbios de excesiva frecuencia temporal, adverbios de exclusividad, verbos emocionales, verbos indicadores de evaluación interpersonal, etc.), **marcos sintácticos** (es decir, combinaciones de pronombres, verbos y adjetivos; p. ej., «yo soy un bruto») y el **uso del tiempo futuro.**

La revisión de literatura arrojó 188 pensamientos individuales, que se encontraban clasificados en las 10 categorías de Beck; se conservaron 149 debido a que existen algunas categorías (filtro mental y visión de túnel) en que la discriminación es prácticamente imposible debido a que las estructuras gramaticales, el fraseo y las expresiones conceptuales se superponen. Con los 149 ítems que se conservaron, se realizó un análisis textual para identificar marcadores lingüísticos asociados con cada categoría que, a su vez, fueran discriminativos entre ellas. Finalmente, se diseñaron reglas de clasificación que incorporaban combinaciones de marcadores lingüísticos.

Utilizando estas definiciones, el sistema COGNO clasificó correctamente el 77% de los pensamientos disfuncionales, lo que los autores interpretan como evidencia de que las reglas y los marcadores capturaron las regularidades del lenguaje presentes en los 149 ítems empleados para entrenar el sistema. **Posteriormente, un experto en terapia cognoscitiva extrajo 112 nuevos pensamientos disfuncionales de un libro de texto.** Al alimentar estos nuevos pensamientos al sistema COGNO, la categoría aplicada coincidió con la del autor del libro de texto en un 77%. Esta cifra resulta promisorio, pues indica que las reglas excedieron las particularidades lingüísticas del primer conjunto de pensamientos y se generalizaron a nuevos ejemplos.

El sistema mostró mejor desempeño (alta exhaustividad y baja tasa de falsas alarmas) al clasificar pensamientos de las categorías: «debería», «etiquetado», «razonamiento emocional» y «predicciones negativas», mientras que, en el caso de las categorías «magnificación» y «descarte de lo positivo», los autores refieren la necesidad de incluir marcadores adicionales para alcanzar mejor discriminación.

Este trabajo demuestra que **es factible construir un sistema que discrimine al menos algunos tipos de pensamientos disfuncionales**. Sin embargo, para alcanzar un desempeño mejor, la investigación en este campo debe avanzar en el descubrimiento de marcadores lingüísticos que permitan discriminar pensamientos de las categorías excluidas. Un sistema como COGNO será muy útil para instruir a los pacientes en el reconocimiento de sus pensamientos disfuncionales y dar apoyo a las actividades terapéuticas formuladas entre sesiones. Asimismo se podría emplear para el entrenamiento de estudiantes y residentes.

Human Brain Functional Mapping Knowledge Base

El objetivo del proyecto *Human Brain Functional Mapping Knowledge Base* (HBFMKB) es construir un modelo de asociación entre funciones cognoscitivas y conductuales humanas y las áreas del cerebro a partir de la literatura médica incluida en MEDLINE⁴. Dado que la nomenclatura de las funciones cognoscitivas no ha alcanzado consenso, los investigadores de la Universidad Nacional de Taiwán emplearon un método de procesamiento de lenguaje natural para extraer las tareas conductuales y las funciones cognoscitivas y realizar un mapa aproximado con los conceptos del UMLS.

La base de conocimiento que alimenta el sistema se creó a partir de estudios clínicos en los que se usaron imágenes de resonancia magnética funcional (fMRI, por sus siglas en inglés). El sistema emplea una sencilla estrategia de búsqueda en MEDLINE —(fMRI AND human) NOT animal— para obtener mensualmente los resúmenes de los artículos publicados a partir de 1985. Al momento de la publicación, la base comprendía 124.450 resúmenes de 3.156 revistas.

En cada uno de los resúmenes, el sistema detecta oraciones en las que concurren el nombre anatómico de un área cerebral y una tarea conductual o la función cognoscitiva. En un primer paso, el sistema utiliza las definiciones del Atlas Talairach²⁷ (un sistema de coordenadas del cerebro humano que describe la localización de las estructuras cerebrales independientemente del tamaño del cerebro) para detectar oraciones en las que se mencionan áreas del cerebro. Posteriormente, el sistema aplica a estas oraciones un proceso de reconocimiento de conceptos para detectar tareas conductuales y funciones cognoscitivas. Este paso es especialmente difícil debido a la falta de consenso sobre la nomenclatura de estos conceptos. Por último, el sistema cartografía las tareas comportamentales o funciones cognoscitivas detectadas hasta términos UMLS y completa el modelo asociándoles el área cerebral de acuerdo con el Atlas Talairach.

Para abordar el problema de la falta de consenso en la nomenclatura, los investigadores emplearon un sistema de reglas construidas manualmente que reconoce las entida-

des que extraer a partir de las categorías gramaticales a su alrededor. El siguiente ejemplo, tomado del artículo original⁴, muestra la aplicación de estas reglas:

- Oración: *A novel event-related potential (ERP) elicited by a visuospatial recognition memory task was recorded in 20 patients with temporal lobe epilepsy using depth electrodes sited in the temporal lobes.*
- Regla de extracción:
 - Precondición: [verbo, preposición].
 - Poscondición: *task*.
 - Extensión límite: 6.
- Resultado de la extracción: *Visuospatial recognition memory*.

El sistema emplea la regla para fijar un límite izquierdo (precondición), que en este caso corresponde a la aparición de un verbo seguido de una preposición: *elicited by*, y un límite derecho (poscondición), que corresponde a la palabra *task*, y por último una extensión límite de seis palabras. En el siguiente paso, el concepto extraído, *visuospatial recognition memory*, se cartografía respecto al concepto UMLS «*Memory, short-term*».

El resultado final es un sistema de presentación y visualización interactivo que localiza las regiones relevantes del cerebro cuando el usuario realiza una búsqueda por una tarea conductual o una función cognoscitiva específica. El sistema también presenta los coeficientes de ocurrencia que indican el número de oraciones en que se hallaron la tarea conductual o la función cognoscitiva y el área cerebral. Mediante este sistema, el usuario puede leer el texto de los resúmenes de donde se extrajeron las asociaciones e incluso acceder al texto completo de los artículos vía PubMed.

Este sistema permite a estudiantes, residentes, médicos y neurocientíficos explorar las relaciones entre las estructuras del cerebro y los resultados de la investigación en funciones cognoscitivas y tareas conductuales, así como reducir el tiempo de búsqueda y recuperación de literatura relevante tanto para aplicación clínica como para investigación.

Otros posibles usos del PLN para la investigación en salud mental

Los anteriores ejemplos demuestran la versatilidad del PLN para la identificación de categorías lingüísticas equivalentes a categorías diagnósticas, lo que permite su aplicación en diversas formas de investigación documental. Asimismo, debido a que sus características operativas son similares a las empleadas en el estudio de pruebas diagnósticas, el PLN abre la posibilidad de nuevos campos de investigación, al mismo tiempo que permite la aplicación a la investigación clínica de metodologías de evaluación familiares. Por ejemplo, **el PLN puede utilizarse para analizar las notas de las historias clínicas y así verificar criterios diagnósticos e identificar casos clínicos relevantes, signos de alerta que merezcan una intervención urgente e incluso errores en diagnósticos o tratamientos**²⁸.

En los estudios observacionales basados en el análisis de narraciones de la enfermedad o de etnografías clínicas, la generación de categorías de análisis es un trabajo dispendioso que, además, no cuenta en todos los casos con métodos de validación robustos. El PLN se puede emplear para generar resúmenes y categorías a partir de grandes volúmenes de texto; estas categorías podrían ser objeto de análisis comparativos contra patrones de referencia surgidos tanto del trabajo de expertos como de manuales diagnósticos como el DSM-IV.

Otras técnicas de investigación cualitativa utilizadas en psiquiatría podrían verse beneficiadas con la aplicación del PLN. Por ejemplo, el PLN podría emplearse para la identificación de patrones de respuesta en el análisis de grupos focales, la descripción de tendencias en el análisis de preguntas abiertas en encuestas o la detección de categorías conceptuales para la construcción de teorías fundamentadas. **El uso de esta herramienta sería de gran utilidad en una investigación que pretendiera identificar, por ejemplo, signos tempranos de enfermedad** o las relaciones entre funciones cognoscitivas y variables clínicas de psicosis no afectivas que los familiares o los cuidadores de los pacientes percibieran.

La caracterización de la enfermedad psiquiátrica según diferentes grupos etarios es una tarea especialmente compleja en la población pediátrica. Asimismo, la percepción y la carga de la enfermedad mental en niños y adolescentes están compuestas de conceptos complejos difíciles de dilucidar y categorizar. Estos conceptos surgen tanto de la combinación de las categorías médicas asignadas a los pacientes como de su interpretación por los afectados y sus familiares²⁹⁻³¹, y constituyen casi una taxonomía personalizada de la enfermedad surgida del contexto social en que se desarrolla. El empleo del PLN podría contribuir al análisis de la percepción de la enfermedad, por ejemplo, identificando categorías diagnósticas estandarizadas (como DSM-IV), pero expresadas en el lenguaje coloquial con que se denominan las enfermedades, y posteriormente identificando atributos y cualidades que pacientes y familiares asignan a los padecimientos. De esta forma se podría construir un puente entre la clasificación «oficial» y las taxonomías surgidas de la indexación social de la enfermedad^{32,33}.

Conclusiones

En gran medida, la comprensión, el diagnóstico y el manejo de las enfermedades psiquiátricas se basa en el análisis de narraciones y descripciones que se categorizan o clasifican en marcos conceptuales estandarizados, tales como los manuales diagnósticos. Sin embargo, el registro de las narraciones en forma de lenguaje natural hace difícil su análisis, tanto por la extensión de los textos como por la complejidad que supone la propia narración. La aplicación de las técnicas del PLN hace posible el desarrollo de investigaciones que analicen estas narraciones en profundidad al automatizar los procesos de extracción de información. El PLN también se puede aplicar a la literatura médica y permite extraer información que de otra forma requeriría inversiones en tiempo y recursos prohibitivas.

BIBLIOGRAFÍA

1. Friedman C, Johnson SB. Natural language and text processing in biomedicine. En: Shortliffe EH, Cimino JJ, editores. Biomedical informatics: computer applications in health care and biomedicine [Internet]. 3.^a ed. New York: Springer; 2006. p. 312-43.
2. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Informatics. 2009;42:760-72.
3. Brown SH, Speroff T, Finkelstein EM, Bauer BA, Wahner-Roedler DL, Greevy R, et al. eQuality: electronic quality assessment from narrative clinical reports. Mayo Clin Proc. 2006;81:1472-81.
4. Hsiao M-YY, Chen D-YY, Chen J-HH. Constructing human brain-function association models from fMRI literature. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:1188-91.
5. Irvine A, Haas S. Natural language processing and temporal information extraction in emergency department triage notes [Internet]. 2008. Disponible en: <http://etd.ils.unc.edu/dspace/bitstream/1901/511/1/IrvineAnnMP.pdf>
6. Jin Y, McDonald RT, Lerman K, Mandel MA, Carroll S, Liberman MY, et al. Automated recognition of malignancy mentions in biomedical literature. BMC bioinformatics. 2006;7:492.
7. Jorge-Botana G, Olmos R, León JAA. Using latent semantic analysis and the predication algorithm to improve extraction of meanings from a diagnostic corpus. Span J Psychol. 2009;12: 424-40.
8. Pamer C, Serpi T, Finkelstein J. Analysis of Maryland poisoning deaths using classification and regression tree (CART) analysis. AMIA Annu Symp Proc. 2008:550-4.
9. Pestian J, Nasrallah H, Matykievicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: a content analysis. Biomed Inform Insights. 2010;2010:19-28.
10. Tsai RT-HT, Chou W-CC, Su Y-SS, Lin Y-CC, Sung C-LL, Dai H-JJ, et al. BIOSMILE: a semantic role labeling system for biomedical verbs using a maximum-entropy model with automatically generated template features. BMC Bioinform. 2007;8:325.
11. Velupillai S, Dalianis H, Hassel M, Nilsson GH. Developing a standard for de-identifying electronic patient records written in Swedish: precision, recall and F-measure in a manual and computerized annotation trial. Int J Medic Inform. 2009;78.
12. Wiemer-Hastings K, Janit AS, Wiemer-Hastings PM, Cromer S, Kinser J. Automatic classification of dysfunctional thoughts: a feasibility test. Behav Res Methods Instrum Comput. 2004;36:203-12.
13. Wolf M, Sedway J, Bulik CM, Kordy H. Linguistic analyses of natural written language: unobtrusive assessment of cognitive style in eating disorders. Int J Eat Disord. 2007;40:711-7.
14. Yang S, Kadouri A, Révah-Lévy A, Mulvey EP, Falissard B. Doing time: a qualitative study of long-term incarceration and the impact of mental illness. Int J Law Psychiatry. 2009;32:294-303.
15. Yu L-CC, Wu C-HH, Jang F-LL. Psychiatric consultation record retrieval using scenario-based representation and multilevel mixture model. IEEE transactions on information technology in biomedicine? IEEE Eng Med Biol Soc. 2007;11:415-27.
16. Ferreira A, Kotz G. ELE-Tutor Inteligente: Un analizador computacional para el tratamiento de errores gramaticales en español como lengua extranjera. Signos. 2010;43:211-36.
17. Smith L, Rindflesch T, Wilbur WJ. MedPost: a part-of-speech tagger for biomedical text. Bioinformatics. 2004;20:2320-1.

18. Ananiadou S, Freidman C, Tsujii J. Introduction: named entity recognition in biomedicine. *J Biomed Inform.* 2004;37:393-5.
19. Metathesaurus [Internet] [citado 21 Feb 2013]. Bethesda: U.S. National Library of Medicine; 2009. Disponible en: http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html
20. Unified Medical Language System (UMLS)-Home [Internet] [citado 21 Feb 2013]. Bethesda: U.S. National Library of Medicine; 2009. Disponible en: <http://www.nlm.nih.gov/research/umls/>
21. Kowalski GJ. Information retrieval systems: theory and implementation (The Information Retrieval Series) [Internet]. 1.^a ed. New York: Springer; 1997. Disponible en: <http://www.springer.com/computer/database+management+%26+information+retrieval/book/978-0-7923-9926-1>
22. Gómez R. La evaluación en recuperación de la información. *Hipertext.net* [Internet]. 2003;1. Disponible en: http://www.upf.edu/hipertextnet/numero-1/evaluacion_ri.html
23. Bai Y, Lin C, Chen J, Liu W. Virtual psychiatric clinics. *Am J Psychiatry.* 2001;158:1160-1.
24. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56-62.
25. Wu CH, Yu LC, Jang FL. Using semantic dependencies to mine depressive symptoms from consultation records. *IEEE Intell Syst.* 2005;20:50-8.
26. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression (The Guilford Clinical Psychology and Psychopathology Series) [Internet]. 1.^a ed. The Guilford Press; 1987. Disponible en: <http://www.worldcat.org/isbn/0898629195>
27. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain: 3-D proportional system: an approach to cerebral imaging [Internet]. Thieme; 1988. Disponible en: <http://www.worldcat.org/isbn/0865772932>
28. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol Med.* 2012;42:41-50.
29. Esposito G, Nakazawa J, Venuti P, Bornstein MH. Perceptions of distress in young children with autism compared to typically developing children: a cultural comparison between Japan and Italy. *Res Development Disabil.* 2013;33:1059-67.
30. Wilkes TCR, Guyn L, Li B, Lu M, Cawthorpe D. Association of child and adolescent psychiatric disorders with somatic or biomedical diagnoses: do population-based utilization study results support the adverse childhood experiences study? *Perm J.* 2012;16:23-6.
31. Herbert LJ, Dahlquist LM. Perceived history of anaphylaxis and parental overprotection, autonomy, anxiety, and depression in food allergic young adults. *J Clin Psychol Med Set.* 2008;15: 261-9.
32. Smith CA, Wicks PJ. PatientsLikeMe: Consumer health vocabulary as a folksonomy. *AMIA Annu Symp Proc.* 2008;682-6.
33. Dixon BE, McGowan JJ. Enhancing a taxonomy for health information technology: an exploratory study of user input towards folksonomy. *Stud Health Technol Inform.* 2010;160 (Pt 2):1055-9.