

# Construcción de recursos de texto para la identificación automática de información clínica en narrativas no estructuradas

PABLO BÁEZ<sup>1,a</sup>, FABIÁN VILLENAL<sup>1,2,b</sup>, KAREN ZÚÑIGA<sup>3,c</sup>,  
NATALIA JONES<sup>3,c</sup>, GUSTAVO FERNÁNDEZ<sup>3,c</sup>,  
MANUEL DURÁN<sup>1,d</sup>, JOCELYN DUNSTAN<sup>1,2,e</sup>

## Construction of text resources for automatic identification of clinical information in unstructured narratives

**Background:** A significant proportion of the clinical record is in free text format, making it difficult to extract key information and make secondary use of patient data. Automatic detection of information within narratives initially requires humans, following specific protocols and rules, to identify medical entities of interest. **Aim:** To build a linguistic resource of annotated medical entities on texts produced in Chilean hospitals. **Material and Methods:** A clinical corpus was constructed using 150 referrals in public hospitals. Three annotators identified six medical entities: clinical findings, diagnoses, body parts, medications, abbreviations, and family members. An annotation scheme was designed, and an iterative approach to train the annotators was applied. The F1-Score metric was used to assess the progress of the annotator's agreement during their training. **Results:** An average F1-Score of 0.73 was observed at the beginning of the project. After the training period, it increased to 0.87. Annotation of clinical findings and body parts showed significant discrepancy, while abbreviations, medications, and family members showed high agreement. **Conclusions:** A linguistic resource with annotated medical entities on texts produced in Chilean hospitals was built and made available, working with annotators related to medicine. The iterative annotation approach allowed us to improve performance metrics. The corpus and annotation protocols will be released to the research community.

(Rev Med Chile 2021; 149: 1014-1022)

**Key words:** Data Curation; Data Mining; Medical Informatics; Natural Language Processing; Supervised Machine Learning.

<sup>1</sup>Centro de Informática Médica y Telemedicina, Facultad de Medicina, Universidad de Chile. Santiago, Chile.

<sup>2</sup>Centro de Modelamiento Matemático, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile. Santiago, Chile.

<sup>3</sup>Escuela de Medicina, Universidad de Chile. Santiago, Chile.

<sup>a</sup>Microbiólogo, PhD en Ciencias Biomédicas.

<sup>b</sup>Cirujano Dentista, Magíster en Informática Médica.

<sup>c</sup>Estudiante de Medicina, Universidad de Chile.

<sup>d</sup>Médico Cirujano, Estudiante de Magíster en Informática Médica.

<sup>e</sup>Física, PhD en Matemática Aplicada y Física Teórica.

PBB es financiado por el proyecto U-INICIA VID 2019 UI-004/19 centro de costo 7487 y del Proyecto ICM P09-015F. JD, FV y MD reciben financiamiento del CMM-ANID AFB 170001. FV y JD reciben apoyo del centro de costos 570111 - CIMT-CORFO. JD agradece a su Fondecyt 11201250.

Las organizaciones que financiaron este trabajo no tuvieron influencia en el diseño del estudio; en la recolección, análisis o interpretación de los datos ni en la preparación, revisión o aprobación del manuscrito.

Recibido el 7 abril de 2020, aceptado el 28 de abril de 2021.

Correspondencia a:  
Jocelyn Dunstan

Avenida Independencia  
1027, Independencia, Región  
Metropolitana, Chile.  
jdunstan@uchile.cl

El texto libre, es decir, sin estructura predefinida, representa una proporción importante de la información en la ficha clínica de los pacientes<sup>1</sup>. Debido a la complejidad de estos datos no estructurados, su información es a menudo descartada en proyectos que intentan automatizar la gestión médica o el uso secundario de los datos<sup>2</sup>.

El texto clínico tiene desafíos particulares debido al uso extensivo de abreviaciones, la presencia de negación, la variabilidad del lenguaje clínico entre especialidades médicas y por diferencias geográficas, además de su disponibilidad restringida por razones de privacidad<sup>1</sup>. Dado que la mayoría de los recursos para el análisis de texto están disponi-

bles para el idioma inglés, centrarse en el análisis automatizado de texto clínico producido en Chile representa una enorme oportunidad<sup>3</sup>.

El análisis automatizado de texto se hace actualmente por medio de herramientas de Procesamiento del Lenguaje Natural (PLN), una rama de la inteligencia artificial que permite la interacción humano-máquina a través del lenguaje, ya sea por medio de la generación o reconocimiento de voz, o el entendimiento y producción de texto<sup>4</sup>.

Una tarea común en PLN clínico es el reconocimiento de entidades médicas, es decir, que el computador identifique en una narrativa clínica cuando se están nombrando piezas específicas de información, por ejemplo, enfermedades o medicamentos. La extracción automática de esta información permite detectar frecuencias y dosis de medicamentos<sup>5</sup>, levantar información epidemiológica de existencia de enfermedades<sup>6</sup> o factores de riesgo<sup>7</sup>, entre otros.

El reconocimiento de entidades médicas se hizo tradicionalmente usando reglas y diccionarios. En la actualidad, se obtienen mejores resultados con algoritmos de aprendizaje de máquinas, en particular, aprendizaje profundo, en donde los computadores aprenden a extraer y modelar automáticamente un entendimiento del lenguaje médico a partir de muchos ejemplos previamente anotados por humanos<sup>8</sup>.

Los textos clínicos anotados por humanos se utilizan como insumos tanto para entrenar como para verificar la automatización. La escasa disponibilidad pública de estos textos, sobretodo para idiomas distintos del inglés, ha enlentecido el avance del área<sup>9</sup>. Un texto o *corpus* anotado corresponde al resultado de haber analizado cada palabra para identificar si ésta pertenece a un grupo de entidades médicas. Para el lenguaje clínico producido en Chile no existen recursos de texto abiertos a la comunidad científica.

El primer paso en la creación de un *corpus* es determinar el contexto de su creación, las necesidades lingüísticas que se intenta resolver y finalmente cuál será su utilidad y aplicación. Con esto definido, se puede establecer cuáles serán las fuentes del texto que conformarán el *corpus*, las cuales pueden ser tan variadas como libros, transcripciones de conversaciones o texto de redes sociales<sup>10</sup>.

En la práctica, los algoritmos de reconocimiento automático necesitan un *corpus* anotado

manualmente con conceptos, entidades, eventos, relaciones y/o atributos<sup>11</sup>. El proceso de anotación manual implica que una persona (el anotador) con un nivel de experiencia suficiente en el tema, revise los textos que componen el *corpus*. Éste sigue unas directrices, identifica un segmento de texto y lo asigna a un tipo de entidad. Para aumentar la precisión, estas tareas de anotación requieren idealmente que más de un anotador trabaje independientemente el mismo *corpus* y se logre un nivel adecuado de acuerdo entre ellos<sup>12</sup>.

Aunque la anotación manual por expertos es demandante, el proceso es necesario por al menos las siguientes razones: 1) el análisis de datos anotados por humanos enfoca y aclara los requerimientos de un algoritmo computacional, 2) sirve de *gold standard* para evaluar los resultados obtenidos empleando sistemas automáticos, y 3) proporciona datos para el desarrollo de sistemas automáticos<sup>13</sup>.

Este artículo presenta la metodología empleada para anotar un *corpus* clínico de interconsultas de la lista de espera no cubiertas por las Garantías Explícitas de Salud (GES). La construcción y uso de este *corpus* resulta estratégico, ya que esta lista de espera es un foco de alta atención en salud pública y la automatización del análisis de su información es relevante para apoyar a tomadores de decisión. Se explicará qué entidades médicas fueron seleccionadas, las distintas etapas del entrenamiento de anotadores y las métricas para calcular el acuerdo entre ellos, así como mostrar ejemplos de texto anotado y discutir sus posibles usos.

## Datos y Método

### *Corpus de texto clínico construido*

El *corpus* clínico se construyó a partir de datos de la lista de espera no cubierta por el Plan GES para nueva consulta de especialidad, los cuales fueron solicitados a través de Ley de Transparencia ([www.portaltransparencia.cl](http://www.portaltransparencia.cl)). Estos datos son almacenados centralmente en el sistema integrado de gestión de tiempos de espera (SIGTE)<sup>14</sup> y cuya sospecha diagnóstica se registra como texto no estructurado.

De las 29 solicitudes de Ley de Transparencia realizadas, correspondientes a la totalidad de Servicios de Salud del país, 17 fueron respondidas,

enviando datos de lista de espera entre los años 2008 y 2018, con las cifras por región mostradas en la Figura 1. Se emplearon 2.592.925 interconsultas de 50 especialidades en la construcción del *corpus*. Los detalles del preprocesamiento del texto de la sospecha diagnóstica fueron descritos en una publicación previa<sup>2</sup>.

### Proceso de anotación

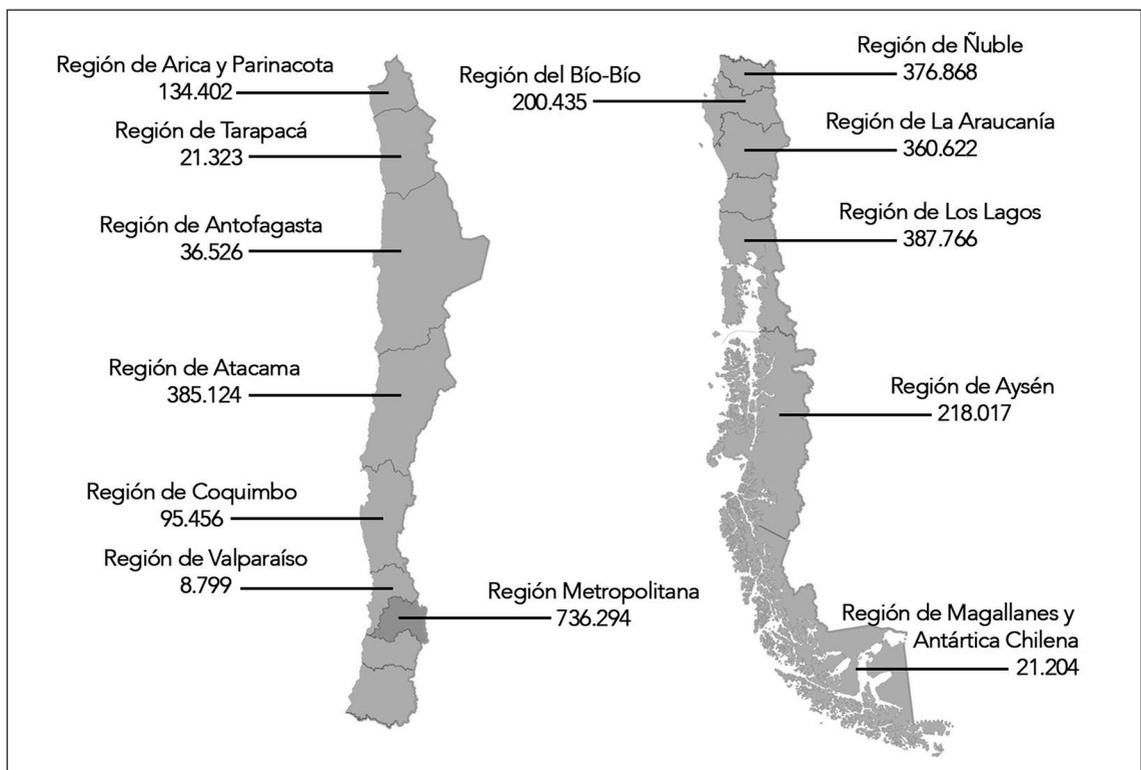
En este trabajo se utilizó BRAT<sup>15</sup> que es una plataforma abierta ampliamente utilizada por la comunidad de la lingüística computacional. Una vez que se decide qué entidades serán anotadas, estas opciones se configuran en la plataforma de BRAT tal como se muestra en la Figura 2. El anotador accede a la plataforma en donde selecciona segmentos de los documentos para ser anotados con una o más de las entidades configuradas. Como se muestra en esta figura, al seleccionar una o más palabras dentro de una narrativa, el *software* automáticamente ofrece un menú con entidades médicas y atributos.

### Esquema y guía de anotación

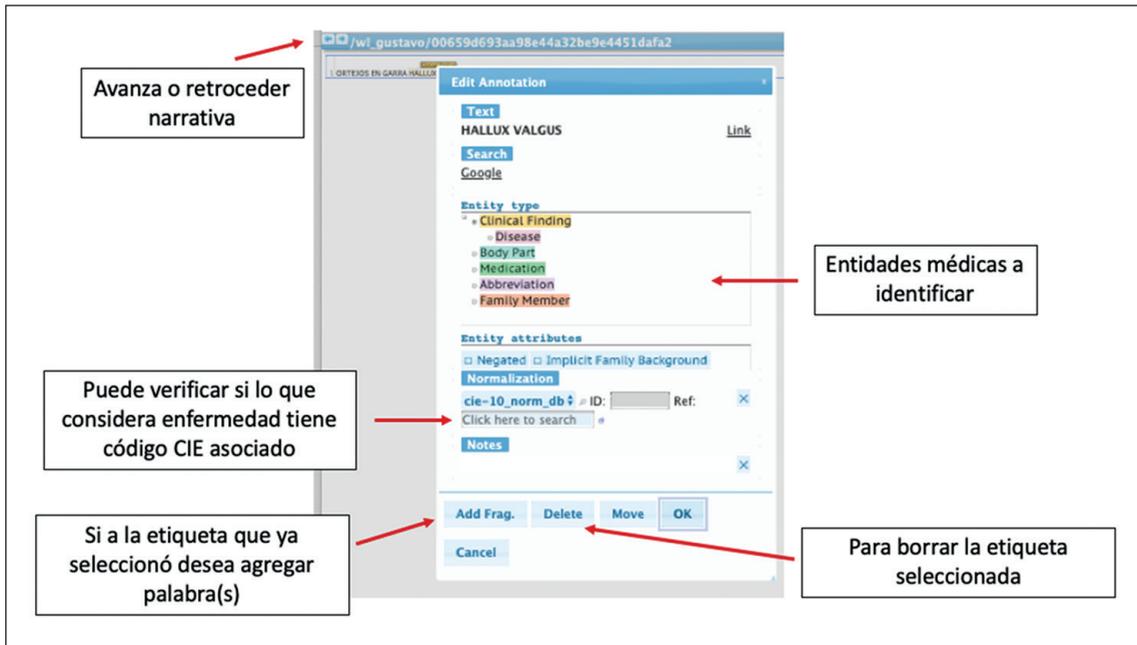
El proceso de anotación de un *corpus* puede dividirse en tres etapas: I) Precampaña: que corresponde a la etapa de construcción y prueba de las guías de anotación mediante la anotación de una muestra pequeña del *corpus*; II) Anotación: en la que se concentran la mayor parte de los recursos humanos y se realizan actualizaciones a las guías de anotación; III) Finalización: en la que se cierra la campaña de anotación y se evalúa la calidad de las anotaciones. La Figura 3 ilustra el proceso simplificado.

### Precampaña

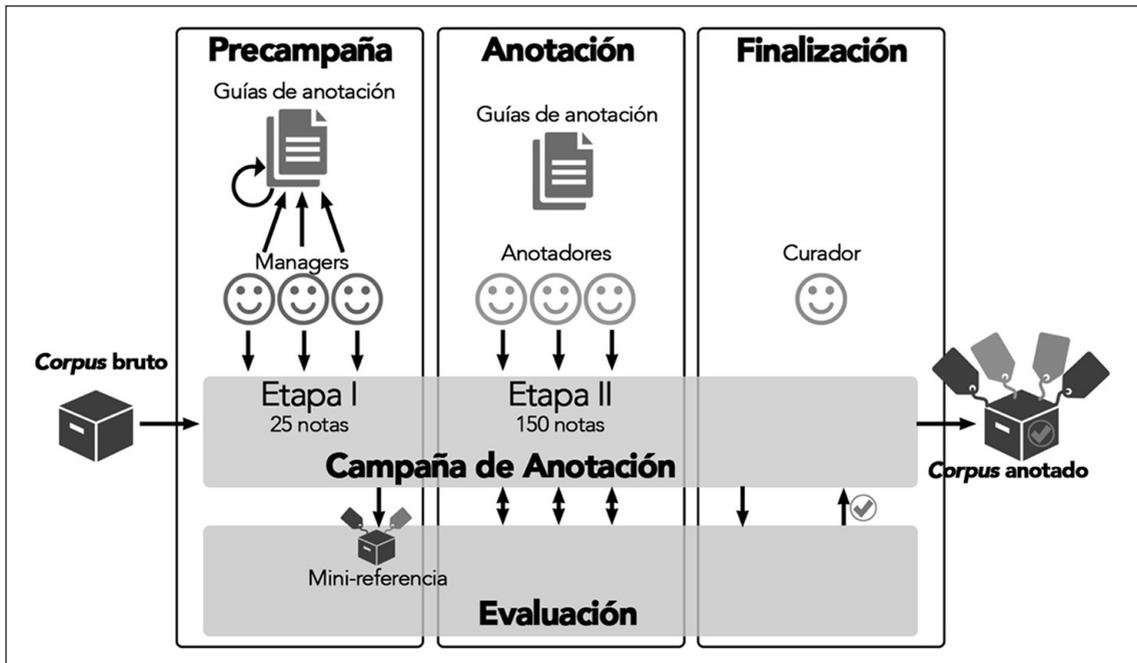
La anotación de un *corpus* requiere que los anotadores dispongan de directrices que clarifiquen el proceso y permitan obtener anotaciones consistentes. Estas directrices se registran en “guías de anotación” que pueden variar en contenido dependiendo del propósito de cada *corpus*. Por lo tanto, para cada proyecto se puede considerar escribir una guía nueva o adoptar una existente.



**Figura 1.** Número de interconsultas no-GES para nueva consulta de especialidad, totales por Región, enviadas desde 17 de los 29 Servicios de Salud de Chile.



**Figura 2.** Ejemplo del menú de anotaciones del software BRAT. Se observan las entidades médicas y atributos anotados durante esta investigación y se describen algunas instrucciones sobre su uso.



**Figura 3.** Representación del proceso de anotación. Los managers están encargados de la planificación y ejecución de la campaña. Los anotadores, quienes realizan la mayor parte de la anotación, pueden ser de tipo "experto" o "no experto", dependiendo de si son o no especialistas en un dominio. El experto curador, compara, mezcla y/o corrige las anotaciones antes de exportar y publicar la versión final del corpus anotado. Figura adaptada de [16].

Aquí optamos por construir una guía nueva tomando como referencia guías de anotación en el ámbito clínico publicadas anteriormente<sup>17-22</sup>. Se decidió anotar hallazgos clínicos (especificando si es un diagnóstico), partes del cuerpo, medicamentos, abreviaciones y miembros de la familia, entidades usuales en la literatura<sup>23</sup>.

Una vez obtenida la primera versión de la guía de anotación, se puso a prueba bajo un esquema de anotación piloto, en donde el equipo líder del proyecto anotó y evaluó 25 interconsultas para identificar problemas y realizar ajustes a la guía. Para descargar la guía de anotación refiérase a<sup>24</sup>.

### Anotación

La siguiente fase incluyó a tres estudiantes de medicina, los cuales anotaron 50 interconsultas idénticas por semana, en tres rondas de anotación. En un proceso de mejora iterativo, los anotadores fueron re-entrenados tras cada ronda de anotación, y las guías sufrieron modificaciones adicionales a fin de clarificar la tarea y mejorar la concordancia.

Las circunstancias de anotación se definieron como sigue: durante las dos primeras rondas, se optó por dar a los anotadores completa libertad respecto a los períodos y duración del proceso de anotación, así como también del espacio en el que decidieran trabajar. Para la segunda ronda se les sugirió trabajar en una sola sesión en cualquier entorno y finalmente en la tercera ronda se impuso una sesión única de anotación en un salón universitario con un investigador presente, disponible para aclarar dudas.

### Finalización

En este estudio la campaña de anotación se cerró evaluando el Acuerdo entre Anotadores (IAA, por *Inter-Annotator Agreement*). Existen diversas métricas del IAA que permiten medir la calidad del proceso de anotación. Aquí se optó por comparar las anotaciones por pares mediante el *F1-Score*<sup>25</sup>:

Se midieron dos *F1-Score*: “estricto” y “relajado”. En el caso estricto se requiere que las anotaciones coincidan exactamente en clase y extensión de palabras, mientras que en el relajado se requiere que la anotación tenga la misma clase, pero puede haber coincidencia parcial, con un solapamiento entre las palabras. Por ejemplo, para la expresión “cáncer de mamas”, si el anotador A1 marca solo “cáncer” como enfermedad, y el anotador A2

decide marcar la expresión completa “cáncer de mamas”, empleando la métrica estricta no existiría acuerdo entre A1 y A2, mientras que con la métrica relajada existe acuerdo ya que ambos anotadores incluyen la palabra “cáncer”.

### Resultados

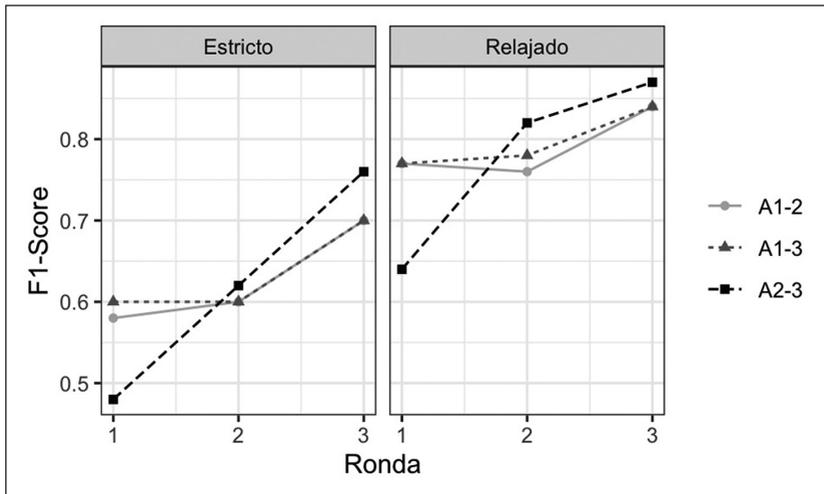
Las interconsultas analizadas presentaron un promedio de 38,8 *tokens*, donde *token* es una estructura lingüística con significado en sí misma, coincide con el concepto de palabra, pero también puede incluir elementos como números o signos de puntuación<sup>26</sup>. El *corpus* empleado está constituido por un total de 5.818 *tokens*, de los cuales 2.286 corresponden a *tokens* únicos (también conocido como vocabulario).

El 20% de las narrativas anotadas, que equivalen a 30 interconsultas, puede ser revisado en [https://pln.cmm.uchile.cl/brat-v1.3\\_Crunchy\\_Frog/index.xhtml/#/rev\\_med\\_chile/](https://pln.cmm.uchile.cl/brat-v1.3_Crunchy_Frog/index.xhtml/#/rev_med_chile/), donde cada entidad se muestra con distintos colores. Las anotaciones fueron consensuadas por los autores a partir del trabajo individual de los anotadores, por lo que este *corpus* corresponde a un patrón de oro (*gold standard*).

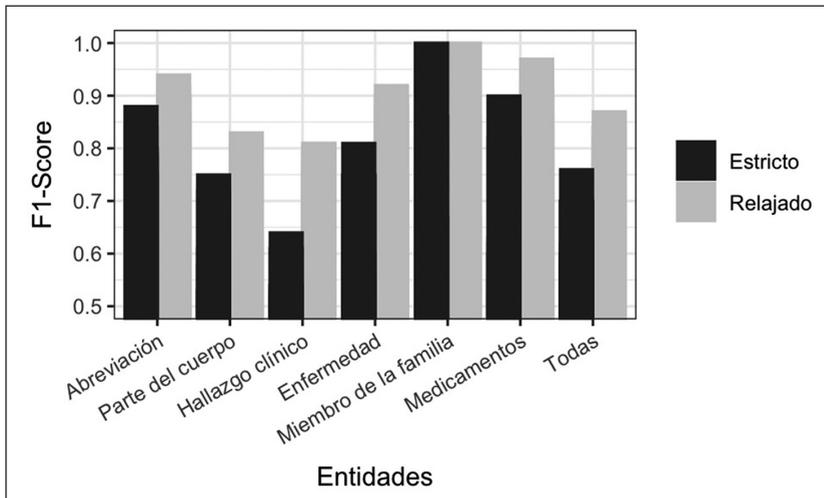
Durante la primera ronda de anotación se alcanzó en promedio un *F1-Score* estricto y relajado de 0,55 y 0,73, respectivamente. Se observó un incremento de 11% y 8% durante la segunda ronda y 23% y 14% respectivamente en la tercera (Figura 4). Durante la última ronda se alcanzó un máximo *F1-Score* estricto y relajado de 0,76 y 0,87, respectivamente.

En cuanto a la proporción de acuerdo por entidad médica, como se observa en la Figura 5, la anotación de hallazgos clínicos y partes del cuerpo presentó mayor discrepancia, mientras que en abreviaciones, medicamentos y miembros de la familia hay menos discrepancia.

Durante las dos primeras rondas, en donde los anotadores tuvieron completa libertad de tiempo y espacio para la anotación, se observó que efectivamente la dedicación fue variable con 4,3 sesiones en promedio y un número oscilante de anotaciones durante cada sesión. Durante la segunda ronda se observó una reducción en el número de sesiones empleadas para completar la tarea (una sesión larga y dos cortas para cada anotador) y finalmente la tercera ronda se implementó en una sesión única.



**Figura 4.** Cambios en el Acuerdo entre Anotadores (*F1-Score*) durante el período de entrenamiento. En esta figura cada color significa un par de anotadores siendo comparados, donde cada uno se designa simplemente con A1, A2 y A3.



**Figura 5.** Acuerdo entre Anotadores (*F1-Score*) para cada tipo de entidad anotada en la muestra del corpus.

## Discusión

El PLN es actualmente el área con más rápido progreso en la ciencia de los datos<sup>27</sup>, y en particular, el análisis de texto clínico ha sido identificado como un área prioritaria<sup>28</sup>. En nuestro país transitamos una etapa incipiente del PLN, y una pieza fundamental en la colaboración entre equipos clínicos e informáticos es la de producir insumos claves para el entrenamiento de algoritmos en los que las máquinas puedan apoyar la decisión clínica. La colaboración interdisciplinaria ofrece una sinergia cuyo impacto positivo en la optimización de procesos clínicos y administrativos en salud

pública, lo que se traduce en una mayor calidad y justicia en el sistema de salud<sup>29</sup>.

La implementación de soluciones de análisis automático basadas en PLN sobre bases de datos que gobiernan procesos sanitarios, tal como el Sistema de Gestión de Tiempos de Espera (SIGTE), puede ayudar con el monitoreo continuo del contenido no estructurado de estos sistemas por parte de los gestores de salud. En el caso del SIGTE en particular, podría mejorar la gestión de la lista de espera no-GES (detectando por ejemplo patologías que pueden resolverse por telemedicina) así como el uso de la información para fines epidemiológicos.

En el sistema de salud chileno, la correcta identificación de un paciente con un problema de salud candidato a ser resuelto mediante GES es vigilada por la Superintendencia de Salud. La utilización de recursos y herramientas basadas en PLN, como las que estamos desarrollando, podría disminuir la cantidad de errores. Más importante aún, al reducir el error en la detección de estos problemas, se puede impactar directa y positivamente la salud del paciente.

Este artículo da cuenta del proceso de construcción de un *corpus* clínico, las guías de anotación para un conjunto de entidades<sup>24</sup>, el proceso de entrenamiento de anotadores y la comparación entre éstos. Además, se entrega una parte del *corpus* anotado y consolidado por los autores.

Nuestros resultados son similares a los reportados anteriormente<sup>30,31</sup>, con valores de acuerdo bajos entre los anotadores durante las primeras rondas, que van incrementando con el paso del tiempo.

Evidencia reciente sugiere que además de establecer unas instrucciones de entrenamiento claras en las guías de anotación, debe prestarse especial cuidado a las circunstancias del proceso de anotación<sup>31</sup>. La duración de las sesiones de anotación, por ejemplo, deben ser de preferencia largas, para evitar que la comprensión de la anotación se deteriore con el tiempo entre múltiples sesiones de anotación. De igual forma se debe ofrecer un ambiente que propicie la concentración de los anotadores durante sus sesiones de anotación. Nuestros resultados apoyan la idea de que una única sesión de anotación en un ambiente de concentración conduce a mejores resultados.

En cuanto al acuerdo entre anotadores por entidad médica, nuestros resultados sugieren que el tipo de categoría a anotar presenta un efecto sobre el acuerdo entre anotadores tal como se ha reportado<sup>31,32</sup>. En este sentido, es fundamental que los anotadores reciban entrenamiento regularmente, manteniendo un balance entre el esfuerzo en la capacitación y la obtención de resultados consistentes. Para ello se debe mantener la anotación lo más simple y clara posible, minimizando la formación extra<sup>31</sup>.

Dentro de las fortalezas de nuestro trabajo se encuentra que cada texto fue revisado en forma independiente por tres anotadores y que la cantidad de entidades y relaciones es extensa comparada con otros artículos<sup>33</sup>. Una desventaja es que

nuestros anotadores tienen menos experiencia clínica que otros trabajos publicados, que usaron internos o becarios de especialidad<sup>33,34</sup>.

El texto clínico es una pieza clave en el ejercicio de la medicina ya que permite documentar los eventos de manera elocuente y natural, aclarando la línea de pensamiento y sirviendo como respaldo. Si bien la codificación de diagnósticos, medicamentos y procedimientos facilita el análisis, probablemente ésta nunca reemplace del todo al texto no estructurado, y es por ello que nos parece relevante impulsar esta área en Chile. En países como España, por ejemplo, la Real Academia Nacional de Medicina tiene personal técnico que colabora con grupos de lingüística computacional<sup>35</sup>.

Trabajo futuro incluye la anotación masiva de textos, el desarrollo de algoritmos de reconocimiento automático de entidades y la aplicación de éstos a datos masivos de lista de espera.

**Agradecimientos:** Los autores agradecen a Matías Rojas y Francisca Rojas por sus comentarios del manuscrito.

PBB es financiado por el proyecto U-INICIA VID 2019 UI-004/19, Proyecto ICM P09-015F y Fondecyt Postdoctorado 3210395. JD, FV y MD reciben financiamiento del CMM-ANID AFB 170001. FV y JD reciben apoyo del centro de costos 570111 - CIMT-CORFO. JD agradece a su Fondecyt de Iniciación 11201250. Esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02).

## Referencias

1. Dalianis H. Clinical text mining: Secondary use of electronic patient records. 2018: Springer.
2. Villena F, Dunstan J. Obtención automática de palabras clave en textos clínicos: una aplicación de procesamiento del lenguaje natural a datos masivos de sospecha diagnóstica en Chile. *Rev Med Chile* 2019; 147 (10): 1229-38.
3. Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than english: opportunities and challenges. *Journal of biomedical semantics* 2018; 9 (1): 12.
4. Jufarsky D, Martin JH. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Pearson education, Asia 2000.

5. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. *J Am Med Inform Assoc*. 2010; 17 (5): 514-8.
6. Lott JP, Boudreau DM, Barnhill RL, Weinstock MA, Knopp E, Piepkorn MW, et al. Population-Based Analysis of Histologically Confirmed Melanocytic Proliferations Using Natural Language Processing. *JAMA Dermatol*. 2018; 154 (1): 24-9.
7. Uzuner Ö, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc*. 2008; 15 (1): 14-24.
8. Gupta A, Banerjee I, Rubin D. Automatic information extraction from unstructured mammography reports using distributed semantics. *Journal of biomedical informatics* 2018; 78: 78-86.
9. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. *J Biomed Inform*. 2018; 88: 11-9.
10. Moreno A, Torre D, Valverde A, Campillos L. Estudio sobre documentos reutilizables como recursos lingüísticos en el marco del desarrollo del Plan de Impulso de las Tecnologías del Lenguaje, Procesamiento del Lenguaje Natural 2019; 63: 167-70.
11. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *Journal of Biomedical Informatics* 2018; 77: 34-49.
12. Hong N, Wen A, Mojarad MR, Sohn S, Liu H, Jiang G. Standardizing Heterogeneous Annotation Corpora Using HL7 FHIR for Facilitating their Reuse and Integration in Clinical NLP. *AMIA Annu Symp Proc*. 2018; 2018: 574-83.
13. Roberts A, Gaizauskas R, Hepple M, Davis N, Demetriou G, Guo Y, et al. The CLEF corpus: semantic annotation of clinical text. *AMIA Annu Symp Proc*. 2007: 625-9.
14. Alarcón G, Navarro C, Muñoz S, Cordero A, Tobar L, Pérez P, et al. Plan Nacional de Tiempos de Espera No GES en Chile en Modelo de Atención RISS. [Internet]. 2018. Disponible en: <https://www.minsal.cl/wp-content/uploads/2018/03/Plan-nacional-de-tiempos-de-espera-No-GES.pdf> [Consultado el 6 de abril de 2020].
15. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii Ji. BRAT: a web-based tool for NLP-assisted text annotation. En *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. 2012.
16. Fort K. Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects. 2016: John Wiley & Sons.
17. Mota E, Martín N, Moreno Á, Ferrete E, Santamaría J, Marimon M, et al. Guías de anotación de información de salud protegida [Internet] 2018. Disponible en: <https://temu.bsc.es/meddocan/wp-content/uploads/2019/02/guías-de-anotación-de-información-de-salud-protegida.pdf> [Consultado el 6 de abril de 2020].
18. Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *J Am Med Inform Assoc*. 2010; 17 (5): 519-23.
19. Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc*. 2011; 18 (5): 552-6.
20. Névéol A, Islamaj Doğan R, Lu Z. Semi-automatic semantic annotation of PubMed queries: A study on quality, efficiency, satisfaction. *Journal of Biomedical Informatics* 2011; 44 (2): 310-8.
21. Intxaurredo A, de la Torre JC, Rodríguez Betanco H, Marimon M, López-Martin JA, González-Agirre A, et al. Resources, guidelines and annotations for the recognition, definition resolution and concept normalization of Spanish clinical abbreviations: the BARR2 corpus. En *SEPLN*. [Internet]. 2018. Disponible en: <https://temu.bsc.es/BARR2/downloads/corpus-BARR2.pdf> [Consultado el 6 de abril de 2020].
22. Skeppstedt M, Kvist M, Nilsson GH, Dalianis H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study. *Journal of Biomedical Informatics* 2014; 49: 148-58.
23. Patel P, Davey D, Panchal V, Pathak P. Annotation of a large clinical entity corpus. En *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* 2018.
24. Báez P, Villena F, Durán M, Dunstan J. Protocolo de Anotación de Entidades en Textos Médicos [Internet] 2020. Disponible en: <https://plncmm.github.io/annodoc/>.
25. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc*. 2005; 12 (3): 296-8.
26. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc 2009.
27. Hirschberg J, Manning CD. Advances in natural language processing. *Science* 2015; 349 (6245): 261-6.
28. Marafino BJ, Park M, Davies JM, Thombley R, Luft HS, Sing DC, et al. Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing

- of Electronic Health Record Data. *JAMA Netw Open* 2018; 1 (8): e185097.
29. Villena F, Pérez J, Lagos R, Dunstan J. Supporting the Classification of Patients in Public Hospitals in Chile by Designing, Deploying and Validating a System Based on Natural Language Processing, 19 November 2020, PREPRINT (Version 1) disponible en *Research Square* [<https://doi.org/10.21203/rs.3.rs-108491/v1>].
  30. Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Setzer A, et al. Semantic annotation of clinical text: The CLEF corpus. En *Proceedings of the LREC 2008 workshop on building and evaluating resources for biomedical text mining*. 2008.
  31. Savkov A, Carroll J, Koeling R, Cassell J. Annotating patient clinical records with syntactic chunks and named entities: the Harvey Corpus. *Lang Resour Eval* 2016; 50: 523-48.
  32. Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I, et al. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics* 2009; 42 (5): 950-66.
  33. Steinkamp JM, Bala W, Sharma A, Kantrowitz JJ. Task definition, annotated dataset, and supervised natural language processing models for symptom extraction from unstructured clinical notes. *J Biomed Inform.* 2020; 102: 103354.
  34. He B, Dong B, Guan Y, Yang J, Jiang Z, Yu Q, et al. Building a comprehensive syntactic and semantic corpus of Chinese clinical texts. *J Biomed Inform.* 2017; 69: 203-17.
  35. Sandoval AM, Llanos LC, Herrero C, Zorita JM, Martínez AG, Samy D. An Online Tool for Enhancing NLP of a Biomedical Corpus, En Almeida FA, Ortega I, Quintana E, Sánchez M, Editores, *Input a Word, Analyze the World: Selected Approaches to Corpus Linguistics* 2016; 25.