



DrugTEMIST

Guías DrugTEMIST: Anotación de fármacos en textos clínicos

V1 [Abril 2024]

AUTORES

Eulàlia Farré-Maduell (Barcelona Supercomputing Center)
Salvador Lima López (Barcelona Supercomputing Center)
Martin Krallinger (Barcelona Supercomputing Center)



Este estudio ha sido realizado dentro del ámbito del Plan de Impulso de las Tecnologías del Lenguaje con financiación de la Secretaría de Estado para el Avance Digital, que no comparte necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores. Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de las mismas.

ÍNDICE

1. INTRODUCCIÓN	3
2. DEFINICIÓN DE LA TAREA	5
3. REGLAS DE ANOTACIÓN	8
3.1 Reglas generales	8
3.2 Reglas positivas	13
TABLA 3. REGLAS POSITIVAS PARA FÁRMACO	13
3.3 Reglas negativas	15
TABLA 4. REGLAS NEGATIVAS PARA FÁRMACO	15
3.4 Reglas especiales y convenciones	17
TABLA 5. REGLAS ESPECIALES PARA FÁRMACO	17

1. INTRODUCCIÓN

Esta guía describe el proceso de anotación y normalización del corpus **DrugTEMIST**, una colección de 1.000 casos clínicos en castellano de distintas especialidades médicas como cardiología, oncología, otorrinolaringología, odontología, pediatría, atención primaria, alergología, radiología, psiquiatría, oftalmología y urología anotados con menciones de **fármacos**.

Este corpus es ‘hermano’ de los corpus DisTEMIST, MedProcNER y SympTEMIST, en el que están anotadas y normalizadas las enfermedades, procedimientos y hallazgos clínicos que aparecen en la misma colección de documentos.

La relevancia de los fármacos

Los fármacos son sustancias elaboradas en laboratorios farmacéuticos para el tratamiento y prevención de síntomas y enfermedades. La mayoría de condiciones médicas se tratan con fármacos. Por ejemplo, se utilizan antibióticos para tratar infecciones bacterianas, y fármacos hipolipemiantes para la prevención de arteriosclerosis coronaria. Hay muy pocas historias clínicas que no contengan menciones de fármacos. Incluso en casos de cirugía o el tratamiento post-quirúrgico encontramos múltiples menciones de fármacos, por ejemplo, para el control de la inflamación y el dolor.

Muchos servicios de salud tienen los fármacos en forma estructurada y codificada junto con las farmacias que dispensan estos fármacos. Sin embargo, la captura de los fármacos dentro de la historia clínica sigue siendo imprescindible. En muchos casos, se mencionan fármacos administrados en otro servicio de salud, en distintos centros de medicina privada, e incluso hay aún muchos fármacos que se obtienen sin receta médica, como analgésicos leves (aspirina, ibuprofeno, paracetamol), antihistamínicos e incluso corticoides no sistémicos (budesonida nasal).

Aunque la mayoría de los fármacos son de administración oral, esta guía incluye también fármacos administrados por otras vías: intravenosa, intramuscular, intranasal, rectal, etc.

Alcance de la guía

Esta guía ha sido formulada para la anotación y normalización exhaustiva de fármacos en la literatura biomédica, casos clínicos publicados e informes médicos reales. Para ello, se han utilizado documentos de diferentes disciplinas médicas clasificados manualmente por un oncólogo en ejercicio y revisados por un documentalista clínico con el fin de asegurar que estuvieran relacionados con el ámbito médico y similares al tipo de estructura y contenido que es relevante para procesar el contenido clínico.

Puesto que se desarrollaron de manera paralela, estas guías son muy similares a las del corpus DisTEMIST pero su contenido está adaptado al campo de los fármacos.

El corpus DrugTEMIST

El corpus **DrugTEMIST** ha sido anotado por dos especialistas clínicos de un hospital de referencia, junto con el apoyo de una médico, quien se encargó también de revisar las menciones para llegar a una versión final. El proceso de anotación del corpus tuvo lugar entre los años 2018 y 2020, con una duración aproximada de 14 meses usando la herramienta *brat*¹ y el proceso de normalización ocurrió en 2020 con una duración aproximada de 6 meses. El proceso de revisión de la anotación y de la normalización tuvo una duración aproximada de 2 meses cada uno. La anotación se hizo en paralelo a la del corpus DisTEMIST.

Antes de comenzar la anotación, se creó un primer borrador de estas guías basándonos en trabajos previos en el dominio. Las guías se refinaron mediante varias rondas de *inter-annotator agreement* (IAA) consistentes en la anotación paralela de un **10%** del corpus. Tras varias rondas, se llegó a una puntuación de IAA total de **95%** (calculado como el pairwise agreement entre dos anotadores independientes) para los fármacos.

Además, durante el proceso de anotación de DrugTEMIST tuvo lugar una discusión continua sobre el contenido del corpus, especialmente sobre casos difíciles y ambiguos, con el objetivo de conseguir la mayor calidad posible y refinar lo máximo posible estas guías.

¹ <https://brat.nlplab.org/>

El resto de esta guía está estructurado como sigue: el apartado 2 describe en mayor detalle el proceso de anotación (contenido a anotar, herramienta y etiquetas a usar); el apartado 3 describe las reglas a seguir para el marcado de entidades; por último, el apartado 4 explica cómo debe hacerse la normalización de las entidades.

2. DEFINICIÓN DE LA TAREA

El objeto de esta tarea es el marcado y normalización de menciones textuales (entidades) relacionadas con **fármacos** en casos clínicos. En términos generales, consideraremos que un fármaco es una sustancia elaborada en un laboratorio farmacéutico, que se administra a un paciente para el tratamiento o la prevención de enfermedades y síntomas.

Debido a la naturaleza especializada del corpus, la persona a cargo de la anotación de este tipo de recurso debe disponer de las siguientes capacidades:

- Conocimiento de biomedicina.
- Habilidades en la búsqueda de información científica y recursos terminológicos.

La tarea tiene dos fases: anotación y normalización. La fase de anotación consiste en el marcado directamente en el texto de entidades siguiendo las reglas de la sección 3 (“Reglas de Anotación”). Para ello, utilizaremos la herramienta de anotación online *brat*. En esencia, la mecánica de anotación consiste en:

1. Marcar con el cursor un determinado texto en el caso clínico (evidencia textual).
2. Clicar en el botón de edición que se visualiza de forma automática después de marcar un texto
3. Seleccionar la entidad FARMACO que aparece como una de las opciones en formato *radio buttons*.

La anotación no se hace sobre un texto vacío, sino sobre un texto con sugerencias de anotación para FARMACO que el anotador podrá aceptar, modificar o eliminar. Estas sugerencias provendrán de nuestros NER entrenados en estas categorías, además de las anotaciones ya realizadas en el corpus hasta el momento.

Las sugerencias no son equivalentes a las guías de anotación, que siempre tomarán precedencia. Hay que tener en cuenta que hay muchas entidades que se tendrán que anotar que no estarán sugeridas, y que muchas sugerencias se tendrán que modificar o eliminar.

La figura 1 muestra un ejemplo de un documento anotado utilizando la interfaz de *brat*.

derecha. A su llegada a urgencias presenta dos episodios convulsivos de crisis tónico-clónicas que ceden con **FARM** **diacepam** iv. Pocos minutos después sufre una tercera crisis, motivo por el cual se inicia tratamiento con **FARM** **Fenitoína** i.v. e ingresa en la Unidad de Cuidados Intensivos (UCI). Se realiza TC craneal objetivando hemorragia subaracnoidea mínima en región suprasilviana izquierda secundaria al TAC. ECG y Rx tórax sin alteraciones significativas.

En la analítica presenta Ca iónico: 2,3 mg/dl (LN 3,9-5,2); 25-hidroxivitamina D: 10 ng/ml (LN 11-40), Magnesio 0,5 mg/dl; Sodio 125 meq, Potasio 2,8 meq. Se inició tratamiento IV con sueroterapia para rehidratación, así como tratamiento iv con **FARM** **sulfato magnésico**, **FARM** **gluconato cálcico** y **FARM** **cloruro potásico** para la corrección de los déficit electrolíticos. Durante su estancia en UCI no presentó nuevos episodios de crisis ni focalidad neurológica, y al alta en este servicio los niveles de magnesio, calcio y potasio se habían normalizado. Se solicita entonces interconsulta a la Unidad de Nutrición objetivando en la exploración física un peso de 63 Kg (peso habitual un año antes: 74 Kg), Talla 1.68 m; IMC 22.3 Kg/m2. junto con signos de disminución de masa magra y masa grasa.

Durante las 3 semanas que el paciente estuvo hospitalizado se trató con medidas dietéticas tales como: dieta astringente, baja en grasas y azúcares simples, reposo de 30 minutos tras las comidas, ingesta de líquidos fuera de las comidas. Además, se instauró tratamiento con **FARM** **sales de rehidratación oral** (**FARM** **Sueroral®**), **FARM** **Loperamida** (14 mg/día), **FARM** **Codeína** (90 mg/día), suplementos de **FARM** **Magnesio** oral (hasta 60 meq/día, repartidos en 4 tomas), de **FARM** **Carbonato cálcico** (500 mg) + **FARM** **Colecalciferol** (400 UI) cada 8 horas.

Figura 1. Ejemplo de marcado de entidades en la herramienta *brat*.

La tabla 1 enumera las etiquetas de entidades que aparecen en el corpus junto a una pequeña descripción.

TABLA 1. ETIQUETAS DISPONIBLES EN BRAT	
[FARMACO]	Cualquier producto, compuesto o sustancia específica con una composición molecular definida y un fin terapéutico o preventivo. Se excluyen las familias y grupos generales de fármacos (p. ej. estatinas, antibióticos).

Por ejemplo: *ribavirina, beclometasona, aciclovir.*

En general, los fármacos específicos recibirán la etiqueta de Fármaco (por ejemplo, *ritonavir, calcio, oxígeno y amoxicilina-clavulánico*). Sin embargo, utilizaremos la etiqueta de Procedimientos para familias de fármacos, como por ejemplo *beta lactámicos, corticoides, antidepresivos*.

En caso de duda del anotador, se aconseja consultar recursos externos. Recomendamos usar sobre todo el navegador de SNOMED-CT (<https://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=MAIN/SNOMEDCT-ES/2024-03-31&release=&languages=es,en>), el navegador de la Agencia Española de Medicamentos y Productos Sanitarios (<https://www.aemps.gob.es/la-aemps/quienes-somos/#medicamento-de-uso-humano>), así como Wikipedia u otras páginas web especializadas. Para abreviaturas en castellano recomendamos utilizar el Diccionario de Siglas Médicas de la Sociedad Española de Documentación Médica (SEDOM, <http://diccionario.sedom.es/>)².

NOTA IMPORTANTE: los medicamentos se anotarán y normalizarán tanto si aparece su nombre farmacéutico como el nombre comercial. En caso que el texto presente un nombre comercial, éste se anotará, pero se normalizará el principio activo del fármaco. Por ejemplo, si en el texto encontramos Ventolín®, lo anotamos y normalizamos su principio activo, *salbutamol*.

Los ejemplos de las secciones que siguen usan el siguiente código de colores: las anotaciones correctas relacionadas con cada regla se marcan en **azul**; las anotaciones correctas que no tienen relación con la regla se marcan en **negrita**; en **rojo** se marcan las anotaciones incorrectas. Para facilitar la explicación de algunas reglas, también es posible que algunos ejemplos aparezcan en forma de capturas de pantalla de la interfaz de *brat*.

² <http://diccionario.sedom.es/>

3. REGLAS DE ANOTACIÓN

En esta sección se detallan las reglas y convenciones a seguir para hacer el marcado textual de signos, síntomas y hallazgos.

Las reglas de anotación están estructuradas de la siguiente forma:

- 3.1 Reglas generales
- 3.2 Reglas positivas
- 3.3 Reglas negativas
- 3.4 Reglas especiales y convenciones

3.1 Reglas generales

En este apartado se explican algunos aspectos básicos relacionados con la anotación en forma de reglas. Estas reglas afectan a todos los tipos de etiquetas.

TABLA 2. REGLAS GENERALES	
[G1] [exhaustividad]	Deben marcarse todas las apariciones de una entidad, incluso si ya se ha marcado anteriormente en el mismo texto o frase.
[ejemplos]	
1. El nuevo régimen de antibióticos consistió en vancomicina + gentamicina . La gentamicina se sustituyó posteriormente por ceftriaxona .	
[G2] [menciones-dudosas]	No se deben anotar las menciones sobre las que tengamos dudas sobre su pertenencia al corpus, elección de etiqueta, ... Estas menciones deben recogerse (nombre del documento, número de línea y texto) para comentarse y discutir las entre el anotador y los organizadores del corpus.

<p>[G3] [revisión- guías]</p>	<p>Si se detectan tipos de menciones o casos especiales que podrían ser de interés y las guías no especifican su anotación, se deben reportar estos casos junto con ejemplos para refinar las reglas de anotación.</p>
<p>[G4] [anotaciones- entre-frases]</p>	<p>Las anotaciones no pueden atravesar más de una oración.</p> <p>Esta regla puede implicar que no se anote información relevante. No obstante, su aplicación simplifica y acelera la anotación y facilita el tratamiento computacional posterior.</p>
<p>[G5] [errores- ortográficos]</p>	<p>Las menciones que incluyan algún tipo de error ortográfico (por ejemplo: letras de más o de menos, espacios incorrectos) también deben anotarse.</p>
<p>[ejemplos]</p> <p>2. Bomuro de ipratropio inhalado (<i>Bromuro de ipratropio</i>)</p> <p>3. Infiximab y corticoides (<i>Infliximab</i>)</p>	
<p>[G6] [palabras- completas]</p>	<p>Las menciones deben estar compuestas por todas las palabras tal y como están incluidas en el texto (palabras completas, abreviaturas, siglas). No se pueden anotar palabras a medias.</p>
<p>[ejemplos]</p> <p>4. Insulina Humalog</p> <p>Insulina Humalog</p> <p>5. NTG</p> <p>NTG</p>	

<p>[G7] [determinantes]</p>	<p>No se deben incluir dentro de una mención los determinantes que aparezcan al principio (por ejemplo: <i>el, la, este, ese</i>). Se incluyen dentro de esta regla los determinantes numerales, por lo que no se deben anotar los números antes de las entidades.</p>
<p>[ejemplos]</p> <p>6. Se retira el suero glucosado al 5%</p>	
<p>[G8] [puntuación]</p>	<p>No se deben incluir dentro de una mención los signos de puntuación que aparezcan al final (comas, puntos, etc.).</p> <p>En las menciones que aparezcan enteras entre paréntesis o comillas, dejaremos fuera estos signos. Como excepción, si dentro de la mención aparece un paréntesis o unas comillas de apertura intentaremos incluir también el signo de cierre.</p> <p>En el caso de las abreviaturas sí que se debe incluir el punto que acompaña a la mención.</p>
<p>[ejemplos]</p> <p>7. revirtiendo a RS tras PC de amiodarona.</p>	
<p>[G9] [otros-idiomias]</p>	<p>Se anotan las menciones en todos los idiomas.</p>
<p>[ejemplos]</p> <p>8. MRI (RMN)</p> <p>9. ASA (AAS)</p>	

<p>[G10] [límite- palabras]</p>	<p>A veces encontraremos en el texto menciones largas que nos interesará anotar como una única entidad incluso si aparecen palabras “irrelevantes” dentro. En estos casos, para poder anotar la entidad completa, la distancia entre las dos partes relevantes que se quieran incluir debe ser igual o inferior a 6 tokens.</p>
<p>[ejemplos]</p> <p>10. Amoxicilina en este caso reforzada con ácido clavulánico [palabras relevantes: “amoxicilina” y “ácido clavulánico”.]</p>	
<p>[G11] [negaciones-1]</p>	<p>Las menciones que estén negadas también se anotan. No deben incluirse en la mención, sin embargo, marcadores de negación (tales como el adverbio “no” o la preposición “sin”).</p>
<p>[ejemplos]</p> <p>11. Se descarta sintrom por antecedente hemorrágico</p>	
<p>[G12] [negaciones-2]</p>	<p>Si aparecen dos fármacos seguidos con algún tipo de negación entre medias, se deben anotar como entidades separadas.</p>
<p>[ejemplos]</p> <p>12. Piperacilina sin tazobactam</p>	
<p>[G13] [no-paciente]</p>	<p>Las menciones que no hagan referencia al paciente también deben anotarse.</p>
<p>[ejemplos]</p> <p>13. Padre recibe digoxina por insuf.cardíaca</p>	

<p>[G14] [abreviaturas- siglas]</p>	<p>Una mención puede ser una abreviatura o una sigla.</p> <p>Únicamente se debe anotar aquellas abreviaturas y siglas que se reconozcan, y de descripciones directamente incluidas en el texto (abreviación ad-hoc) que no presenten ningún tipo de duda.</p> <p>A ser posible, se debe incluir en la sección de Notas de brat la forma larga de la abreviatura en el siguiente formato: "<abreviatura>: <forma larga>".</p> <p>Ver también: regla G15.</p>
<p>[ejemplos]</p> <p>14. Se administra NTG sublingual</p>	
<p>[G15] [menciones- anidadas]</p>	<p>Algunas menciones textuales pueden incluir más de un concepto usando un mismo sintagma nominal, por ejemplo para especificar más de una zona anatómica o múltiples etiologías. En este corpus, este tipo de mención debe anotarse como una única entidad. Es decir, no están permitidas las anotaciones anidadas (menciones que se solapan unas con otras).</p> <p>Esto no significa, sin embargo, que haya que anotar conceptos claramente separados como una única entidad (ver segundo ejemplo).</p> <p>Excepción: regla G16.</p>
<p>[ejemplos]</p> <p>15. Examen físico general y abdominal</p> <p>16. Se programa TAC abdominal y ECG</p>	

<p>[G16] [abreviaturas- anidades]</p>	<p>Las abreviaturas que aparezcan entre paréntesis junto a su forma larga se deben anotar también separadamente (sin incluir los paréntesis).</p>
<p>[ejemplos]</p> <p>17. Nitroglicerina (NTG) antes de angiografía</p>	

3.2 Reglas positivas

En este apartado se explican las reglas positivas, es decir, reglas generales sobre qué se debe incluir en la anotación. Estas reglas están divididas según el tipo de entidad a la que afectan.

TABLA 3. REGLAS POSITIVAS PARA FÁRMACO	
<p>[FARM-P1] [nombres- genéricos]</p>	<p>Se anotan todos los productos, compuestos o sustancias específicas con una composición molecular definida y un fin terapéutico o preventivo.</p>
<p>[ejemplos]</p> <p>18. diclofenaco</p> <p>19. insulina</p> <p>20. oxígeno</p> <p>21. calcio</p>	

<p>[FARM-P2] [nombres- comerciales]</p>	<p>Se anotan también los nombres comerciales de fármacos. Si aparece un nombre comercial junto a uno genérico, se anotan por separado.</p>
<p>[ejemplos]</p> <p>22. baclofeno (Lioresal ®)</p>	
<p>[FARM-P3] [combinaciones]</p>	<p>Las combinaciones comunes de fármacos se anotan como una única entidad cuando aparecen como una única palabra. Para saber si una combinación es común o no, podemos consultar SNOMED CT o la AEMPS (https://cima.aemps.es/cima/publico/lista.html)</p>
<p>[ejemplos]</p> <p>23. amoxicilina-clavulánico</p> <p>24. lopinavir/ritonavir</p> <p>25. Piperacilina-tazobactam</p>	
<p>[FARM-P4] [abreviaturas]</p>	<p>A veces encontraremos nombres de fármacos abreviados, en especial cuando se dan varios combinados. Estas abreviaturas también deben anotarse siempre que sepamos inequívocamente que se refieren a un fármaco.</p>
<p>[ejemplos]</p> <p>26. lopirito (lopinavir + ritonavir)</p> <p>27. L/r (lopinavir + ritonavir)</p> <p>28. Pipertazo (piperacilina + tazobactam)</p>	

3.3 Reglas negativas

En este apartado se explican las reglas negativas, es decir, reglas generales sobre qué no se debe incluir en la anotación. Estas reglas están divididas según el tipo de entidad a la que afectan.

TABLA 4. REGLAS NEGATIVAS PARA FÁRMACO	
[FARM-N1] [fines-no-terapéuticos]	<p>No se anotan las sustancias usadas con fines no preventivos o terapéuticos. Algunos ejemplos: compuestos y elementos utilizados con fines diagnósticos y de preparación de sustancias, desinfección de quirófanos, etc., medios de contraste, antisépticos no aplicados directamente al paciente, o sustancias que forman parte del cuerpo.</p>
	<p>[ejemplos]</p> <p>29. El día 5 se realizó una RMN cardíaca que confirmó la miopericarditis con realce tardío de gadolinio subepicárdico</p> <p>30. Los niveles de hierro del paciente son normales</p>
[FARM-N2] [tipos-generales]	<p>No se anotan como fármacos los tipos generales y familias de sustancias. Este tipo de menciones se deben anotar como procedimientos.</p>
	<p>[ejemplos]</p> <p>31. Actualmente en tto corticoides</p> <p>32. Tratamiento antiviral</p> <p>33. Profilaxis antibiótica adecuada</p>
[FARM-N3]	<p>No se incluye la potencia de los fármacos (g, mg, ...)</p>



[potencia]	
[ejemplos]	
34. Amlodipino 10 mg	
35. Bisoprolol normon 2,5 mg comprimidos recubiertos con película	
[FARM-N4] [dosificación]	No se incluye la dosificación de los fármacos.
[ejemplos]	
36. Azitromicina 250 mg, Una vez al día, en el desayuno	
37. Paracetamol kern pharma 1000 mg comprimidos efg, 40 comprimidos.1000 M G. Vía oral. Cada 8 horas. SI PRECISA	
[FARM-N5] [forma]	No se incluye la formulación del fármaco.
[ejemplos]	
38. pastilla, gragea, sobre, pomada	
[FARM-N6] [vía]	No se incluye la forma en la que se administra el fármaco.
[ejemplos]	
39. Oral, cutánea	
[FARM-N7] [elementos]	Atención a los elementos que forman parte de la fisiología humana, que tienen su equivalente en fármacos.

40. **Calcio** y **magnesio** por debajo los límites de la normalidad.

3.4 Reglas especiales y convenciones

En este apartado se explican las reglas especiales y convenciones, es decir, decisiones tomadas durante la anotación sobre casos muy específicos difíciles de generalizar.

TABLA 5. REGLAS ESPECIALES PARA FÁRMACO	
[FARM-E1] [porcentajes]	Como excepción a las reglas de potencias y dosificación, en los colirios y sueros sí que anotaremos el porcentaje de sustancias
[ejemplos]	
41. Tratamiento con Mitomicina-C 0,02% preoperatorio	
42. Administración de suero glucosado al 10%	
[FARM-E2] [inmuglobulinas- plasma]	La administración de inmunoglobulinas y plasma convaleciente se anota como fármaco
[ejemplos]	
43. Se administra inmunoglobulina intravenosa durante 5 días y se implanta el tratamiento indicado según el protocolo para pacientes COVID-19 con sulfato de hidroxiclороquina , antirretrovirales (lopinavir y ritonavir), antibiótico (amoxicilina), corticoides y oxigenoterapia de bajo flujo.	