



# Guías DISTEMIST: Anotación y normalización de enfermedades en textos clínicos

V1 [Abril 2022]

## AUTORES

Eulàlia Farré-Maduell (Barcelona Supercomputing Center)  
Luis Gascó Sánchez (Barcelona Supercomputing Center)  
Salvador Lima López (Barcelona Supercomputing Center)  
Antonio Miranda-Escalada (Barcelona Supercomputing Center)  
Martin Krallinger (Barcelona Supercomputing Center)



**Plan TL**  
Plan de Impulso de las  
Tecnologías del Lenguaje



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

<b>1. INTRODUCCIÓN</b>	<b>2</b>
<b>2. DEFINICIÓN DE LA TAREA</b>	<b>6</b>
<b>3. REGLAS DE ANOTACIÓN</b>	<b>9</b>
3.1 Reglas generales	9
3.2 Reglas positivas	16
TABLA 3. REGLAS POSITIVAS PARA ENFERMEDAD	16
3.3 Reglas negativas	20
TABLA 4. REGLAS NEGATIVAS PARA ENFERMEDAD	20
3.4 Reglas especiales y convenciones	24
TABLA 5. REGLAS ESPECIALES PARA ENFERMEDAD	24
TABLA 6. REGLAS ESPECÍFICAS PARA ONCOLOGÍA	26
<b>4. REGLAS DE NORMALIZACIÓN</b>	<b>29</b>

# 1. INTRODUCCIÓN

Esta guía describe el proceso de anotación y normalización del corpus **DISTEMIST**, 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 **enfermedades**. Cada una de las menciones del corpus ha sido normalizada usando la terminología SNOMED-CT.

## La relevancia de las enfermedades

La categorización y tratamientos de los pacientes se basan en los diagnósticos, y por tanto, en las enfermedades o “etiquetas” asignadas por los clínicos. Sin embargo, las enfermedades pueden aparecer en cualquier parte de los documentos clínicos, expresados de formas muy distintas (ej, ataque de corazón=infarto agudo de miocardio), y con distinto nivel de alcance (por ejemplo, el concepto podrá incluir si la enfermedad se presenta de forma súbita, si es una secuela, o es un trastorno congénito). A su vez, las terminologías a las que normalizaremos serán más o menos exhaustivas y granulares a la hora de definir las enfermedades.

La detección y clasificación de enfermedades en literatura e informes médicos es crítica para determinar:

1. La frecuencia de enfermedades, para planificación sanitaria y evaluación de los sistemas de salud
2. Distintas en prevalencias de algunas enfermedades entre zonas y países para poder investigar sus causas y analizar desigualdades
3. La emergencia de enfermedades desconocidas y re-emergencia de enfermedades poco frecuentes
4. Evaluación de campañas de vacunación, de salud, de hábitos saludables.

Además, la detección de enfermedades es clave para el indexado de información clínica y biomédica. Más del 20% de las consultas de búsqueda en PubMed están relacionadas con

enfermedades, trastornos y anomalías<sup>1</sup>. Esta categoría tiene una importante presencia tanto en los artículos científicos como en las narraciones clínicas<sup>2</sup>, por lo que el desarrollo de sistemas altamente eficientes capaces de hacer accesibles este tipo de entidades a los sistemas de búsqueda tiene un gran interés en el ámbito biomédico. Una anotación textual más granular puede mejorar los sistemas de indexación, dado que a la hora de indexar un documento textual, no todo el documento es relevante para este fin.

## SNOMED-CT

Las entidades manualmente anotadas en los textos son normalizadas a identificadores de SNOMED-CT<sup>3</sup>.

SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) es la terminología clínica integral, multilingüe y codificada de mayor amplitud, precisión e importancia desarrollada en el mundo. SNOMED CT es, también, un producto terminológico que puede usarse para codificar, recuperar, comunicar y analizar datos clínicos permitiendo a los profesionales de la salud representar la información de forma adecuada, precisa e inequívoca. La terminología se constituye, de forma básica, por conceptos, descripciones y relaciones. Estos elementos tienen como fin representar con precisión información y conocimiento clínico en el ámbito de la asistencia sanitaria.

Para la normalización de los datos se ha utilizado la versión de SNOMED-CT del 21/10/2021<sup>4</sup>.

## Alcance de la guía

Esta guía ha sido formulada para la anotación y normalización exhaustiva de enfermedades en la literatura biomédica, casos clínicos publicados e informes médicos reales. Para ello, se

---

<sup>1</sup> Islamaj Dogan, R., Murray, G. C., Névéol, A., & Lu, Z. (2009). Understanding PubMed® user search behavior through log analysis. Database, 2009.

<sup>2</sup> Gasco, L., Nentidis, A., Krithara, A., Estrada-Zavala, D., Murasaki, R. T., Primo-Peña, E., ... & Krallinger, M. (2021). Overview of BioASQ 2021-MESINESP track. Evaluation of advance hierarchical classification techniques for scientific literature, patents and clinical trials. CEUR Workshop Proceedings.

<sup>3</sup> <https://www.snomed.org/>

<sup>4</sup>

<https://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=MAIN/SNOMEDCT-ES/2021-10-31&release=&languages=es,en>

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 se parecieran al tipo de estructura y contenido que es relevante para procesar el contenido clínico.

## El corpus DISTEMIST

El corpus DISTEMIST ha sido anotado y normalizado por dos especialistas clínicos de un hospital de referencia, junto con el apoyo de una médico, quien se encargó también de revisar las menciones y sus códigos asociados 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*<sup>5</sup> 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.

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 **82.3** (calculado como el pairwise agreement entre dos anotadores independientes) para las enfermedades.

Además, durante el proceso de anotación de DISTEMIST 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.

Para garantizar el aprovechamiento de las menciones del corpus, se ha utilizado la terminología SNOMED-CT para asignar un identificador a cada anotación manual.

La versión final del corpus DISTEMIST incluye 10666 menciones de enfermedades, de las cuales 10318 son únicas y están normalizadas a Snomed-CT.

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

---

<sup>5</sup> <https://brat.nlplab.org/>

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 **enfermedades** en casos clínicos. En términos generales, consideraremos que una ENFERMEDAD es un trastorno en la salud de las personas que tiene una persistencia temporal.

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 ENFERMEDAD 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 ENFERMEDADES 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*.

1	Paciente varón de 73 años de edad con antecedentes de <span style="border: 1px solid green; padding: 2px;">ENFERMEDAD</span> brucelosis, <span style="border: 1px solid green; padding: 2px;">ENFERMEDAD</span> hiperuricemia e <span style="border: 1px solid green; padding: 2px;">ENFERMEDAD</span> hipercolesterolemia que ingresa en nuestro Servicio para cirugía programada de <span style="border: 1px solid green; padding: 2px;">ENFERMEDAD</span> hiperplasia prostática benigna.
2	En el estudio diagnóstico presentó un tacto rectal que evidenció una próstata de características adenomatosas, muy aumentada de tamaño (G III/IV).
3	Se practica adenomectomía retropúbica según técnica de T.
4	Millin, enucleándose un gran <span style="border: 1px solid green; padding: 2px;">ENFERMEDAD</span> adenoma de 170 gramos de peso.

**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	
<b>[ENFERMEDAD]</b>	<p>Trastorno en la salud de las personas que tiene una persistencia temporal.</p> <p>Por ejemplo: neumonía, adenocarcinoma gástrico, cáncer de próstata, cistitis infecciosa, depresión, diabetes mellitus, alergia a la penicilina, ...</p>

La distinción entre los conceptos **enfermedad** y **síntoma** resulta frecuentemente arbitraria. Como regla general, entenderemos los síntomas como observaciones o hallazgos clínicos sobre el estado de salud de las personas que pueden ser tanto normales como anormales y que, al contrario que las enfermedades, son generalmente *transitorios*. Aún así, hay enfermedades transitorias [por ejemplo, una infección respiratoria de las vías altas (=enfermedad en SNOMED CT) puede durar menos de 48 horas] y síntomas crónicos [incontinencia de esfuerzo (=síntoma en SNOMED CT) puede durar meses]. Para evitar discrepancias entre anotadores, las guías acotan la categoría enfermedad y los atributos que podemos incluir en la mención.

La segunda fase de la tarea está centrada en la normalización de las entidades marcadas en la primera fase. Esto consiste en asignar a cada entidad un código de la terminología SNOMED-CT que será exacto o un término hiperónimo. Para este proceso deben seguirse las reglas y convenciones especificadas en la sección 4 (“Reglas de normalización”).

El proceso de normalización se ha diseñado en base a un modelo secuencial. En primer lugar, separamos las menciones únicas, es decir, las que requieren un único código (por ejemplo: cáncer de próstata) de las menciones compuestas, que se pueden descomponer y van a requerir dos o más códigos (por ejemplo: ganglios axilares y de mediastino=1. Ganglios axilares+2. Ganglios de mediastino). Después verificamos los términos que encontramos exactos en la terminología (las palabras exactas, o sinónimos exactos (por ejemplo: linfadenopatía maligna axilar = ganglio canceroso axilar). En un tercer paso, determinamos si los términos que no son exactos son hipónimos o hiperónimos. Estas clasificaciones agilizan y disminuyen el tiempo de normalización manual.

En caso de duda del anotador, se aconseja consultar recursos externos. Recomendamos usar sobre todo el navegador de SNOMED-CT de la versión del 21/10/2021, así como Wikipedia u otras páginas web especializadas como MedlinePlus. Para abreviaturas en castellano recomendamos utilizar el Diccionario de Siglas Médicas de la Sociedad Española de Documentación Médica (SEDOM)<sup>6</sup>.

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*.

### 3. REGLAS DE ANOTACIÓN

En esta sección se detallan las reglas y convenciones a seguir para hacer el marcado textual de enfermedades.

---

<sup>6</sup> <http://diccionario.sedom.es/>



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	
<p><b>[G1]</b> [exhaustividad]</p>	<p>Deben marcarse todas las apariciones de una entidad, incluso si ya se ha marcado anteriormente en el mismo texto o frase.</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p>1. El paciente presentaba numerosos <b>quistes</b> de contenido seroso. Dichos <b>quistes</b> mostraban...</p>
<p><b>[G2]</b> [menciones- dudosas]</p>	<p>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>
<p><b>[G3]</b> [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><b>[G4]</b> [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>[ejemplos]</p> <p>2. <b>Sepsis abdominal</b>. Causa: Enterococcus faecalis</p> <p>[En este ejemplo nos interesaría anotar que las sepsis abdominal está causada por E. faecalis. Sin embargo, como la información está en dos oraciones separadas, anotaremos solo “Sepsis abdominal”.]</p>	
<p><b>[G5]</b> [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>3. <b>Heptitis B</b> (<i>Hepatitis B</i>)</p> <p>4. <b>Poliomelitis vacunal</b> (<i>Poliomiелitis vacunal</i>)</p> <p>5. <b>Osteo mielitis</b> (<i>Osteomiелitis</i>)</p> <p>6. La radiografía revela <b>numonía del lóbulo superior</b> (<i>Neumonía</i>)</p>	
<p><b>[G6]</b> [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>7. <b>Absceso laterocervical</b></p>	

<p><b>Absceso latero</b>cervical</p> <p>8. <b>Ca. pulmonar</b></p> <p><b>Ca. pulmonar</b></p> <p>9. <b>TEP</b> (<i>TromboEmbolia Pulmonar</i>)</p> <p><b>TEP</b></p>	
<p><b>[G7]</b> [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>).</p>
<p>[ejemplos]</p> <p>10. La <b>fiebre de las Rocky Mountains</b></p>	
<p><b>[G8]</b> [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>11. Se objetivó un extenso <b>hematoma, de 12 cm de diámetro máximo, en la celda renal izquierda</b>, sin <b>líquido libre intraperitoneal</b>; la suprarrenal izquierda...</p>	
<p><b>[G9]</b> [otros-idiomas]</p>	<p>Se anotan las menciones en todos los idiomas.</p>



[ejemplos]

12. **Lyme disease, AIDS**

**[G10]**  
[límite-  
palabras]

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.

[ejemplos]

13. **absceso drenado en territorio de fibrosa mitroaórtica**

[palabras relevantes: “absceso” y “mitroaórtica”.]  
[Distancia: 5 tokens]

14. **ENDOCARDITIS MITRAL sobre válvula con degeneración mixomatosa por S.  
AUREUS**

[palabras relevantes: “endocarditis mitral” y “S. aureus”.]  
[Distancia: 6 tokens]

**[G11]**  
[negaciones-1]

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”).

Excepción: regla EN-E7.

[ejemplos]

15. Sin antecedentes de **enfermedades de transmisión sexual**

**[G12]**  
[negaciones-2]

Si aparecen dos enfermedades seguidas con algún tipo de negación entre medias, se deben anotar como entidades separadas (excepto que la negación forme parte del nombre de la enfermedad).

[ejemplos]	
<p>16. <b>Neumonía sin afectación pleural.</b></p> <p>17. <b>Apendicitis no complicada</b></p> <p>18. <b>Esteatosis hepática no alcohólica</b></p>	
<p><b>[G13]</b> [no-paciente]</p>	<p>Los menciones que no hagan referencia al paciente (por ejemplo: antecedentes familiares) también deben anotarse.</p>
[ejemplos]	
<p>19. Refiere <b>enfermedad de Crohn</b> en un hermano gemelo.</p>	
<p><b>[G14]</b> [abreviaturas- siglas]</p>	<p>Una mención puede ser una abreviatura o una sigla de una enfermedad.</p> <p>Únicamente se debe anotar aquellas abreviaturas y siglas que pueden ser unidas a entidades clínicas según el conocimiento del anotador y descripciones directamente incluidas en el texto (abreviación ad-hoc) sin ningún tipo de duda.</p> <p>En el caso de abreviaturas que se pueden corresponder con más de una enfermedad se adopta la más representada teniendo en cuenta el contexto clínico: OI será “ojo izquierdo” en oftalmología, y “oído izquierdo” en otorrinolaringología..</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: “&lt;abreviatura&gt;: &lt;forma larga&gt;”.</p> <p>Ver también: regla G15.</p>
[ejemplos]	



20. Debido a este hallazgo, a pesar de que la paciente no tenía factores de riesgo para **TBC**, ...

21. **EPOC** (*enfermedad pulmonar obstructiva crónica*)

**[G15]**  
[menciones-  
anidadas]

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 el corpus DISTEMIST, 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).

Esto no significa, sin embargo, que haya que anotar conceptos claramente separados como una única entidad (ver segundo ejemplo).

Excepción: regla G16.

[ejemplos]

22. **hematoma perirrenal y pararrenal anterior y posterior, así como en gotiera parietocólica**

23. Imagen de **tumor esofágico y gástrico**

**[G16]**  
[abreviaturas-  
anidadas]

Las abreviaturas que aparezcan entre paréntesis junto a su forma larga se deben anotar también separadamente (sin incluir los paréntesis).

[ejemplos]

24. **Neoplasia (CaCu) maligna de cérvix**

[En este ejemplo hay dos anotaciones:  
“neoplasia (CaCu) maligna de cérvix” y “CaCu”]

## 3.2 Reglas positivas

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

TABLA 3. REGLAS POSITIVAS PARA ENFERMEDAD	
<p><b>[EN-P1]</b> [generales]</p>	<p>Términos muy generales como “enfermedad”, “problema”, “síndrome”, ... deben anotarse si van acompañados de calificadores como localizaciones anatómicas, nombres de personas, organismos/sustancias/productos farmacéuticos causantes, ...</p> <p>Excepción: ver regla...</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p style="text-align: center;">25. <b>Síndrome de Budd-Chiari</b></p> <p style="text-align: center;">26. <b>Enfermedad renal poliquística</b></p> <p style="text-align: center;">27. Remisión completa de la <b>enfermedad</b></p>
<p><b>[EN-P2]</b> [localización]</p>	<p>Se incluye en las menciones la zona anatómica donde ocurre la enfermedad, evitando puntos de referencia del tipo “proximal”, “distal”, “anteroposterior”, ...</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p style="text-align: center;">28. <b>L3-L4, donde impresiona de tumor de partes blandas indurado</b></p> <p style="text-align: center;">29. ...de una <b>polineuropatía-axonal sensitivo-motora de tipo mixto con afectación de las cuatro extremidades...</b></p>



30. **Rotura de una cuerda tendinosa a nivel de la cabeza de un vientre del músculo papilar posteromedial**

31. **Absceso cerebeloso posterolateral.**

**[EN-P3]**  
[lateralidad]

La lateralidad de una enfermedad se incluye dentro de la mención cuando se hable del corazón, colon, hipocondrio, fosa ilíaca, flanco y toda patología intracraneal (vasos craneales a partir de la arteria carótida interna, así como lo que ocurre dentro de la calota craneal, sea afectación global o localizada).

En el resto de zonas anatómicas no debe incluirse excepto que la mención pierda el sentido si no se incluye.

La bilateralidad sí que debe anotarse siempre.

[ejemplos]

32. Presenta **focos isquémicos en hemisferio cerebral derecho.**

33. **Dextrocardia, hígado en hemiabdomen izquierdo y burbuja gástrica en hemiabdomen derecho.**

34. Se diagnostica de **neumonía pneumocócica bilateral** y se inicia tratamiento intravenoso con...

**[EN-P4]**  
[causa]

Cuando se exprese la causa u origen junto a una mención de enfermedad, se debe incluir dentro de la mención.

Cuando se anote patógenos como causa nos limitaremos a la forma más breve posible (género+especie) y excluirémos sensibilidades, resistencias, serotipos, etc.

Regla relacionada: EN-N8.





[ejemplos]

35. Paciente de 31 años con **lesión medular traumática**.
36. Los cultivos y la patología confirman **sepsis secundaria a Staphylococcus hominis**.
37. **Enfermedad renal crónica por enfermedad poliquistica** en tratamiento con hemodiálisis.
38. Presentó **obstrucción intestinal secundaria a enteritis rádica**.

**[EN-P5]**  
[patogénesis]

Cuando se exprese la patogénesis de una enfermedad (por ejemplo, "inmunitario", "hormonal", "hereditario", ...) debe incluirse en la mención..

[ejemplos]

39. Se diagnostica de **trastorno bipolar hereditario** y se inicia tratamiento con sales de litio

**[EN-P6]**  
[ordenamiento]

Siempre que sea posible las menciones deben incluir el ordenamiento de las enfermedades ("primaria", "secundaria", ...).

[ejemplos]

40. TAC Craneal: **tumor secundario** que desplaza línea media

**[EN-P7]**  
[curso]

Siempre que sea posible las menciones deben incluir el curso de la enfermedad ("aguda", "recurrente", "crónica", "terminal", ...).

[ejemplos]

41. Profilaxis con penicilina benzatina por **faringoamigdalitis estreptocócica recurrente**



<p><b>[EN-P8]</b> [inicio]</p>	<p>Siempre que sea posible las menciones deben incluir el <i>onset</i> o inicio de la enfermedad (“inicio adulto”, “inicio pediátrico”, “infantil”, “precoz”, “embrionario”, “puerperal”, ...)</p> <p>Ver también: regla EN-N2 [temporalidad].</p>
<p>[ejemplos]</p> <p>42. <b>Parkinson de comienzo temprano</b></p> <p>43. <b>Deficiencia de alfa-1,4-glucosidasa de comienzo tardío</b></p> <p>44. <b>Acné juvenil</b></p> <p>45. <b>Atrofia uretral postmenopáusica</b></p>	
<p><b>[EN-P9]</b> [severidad]</p>	<p>Las menciones pueden incluir SOLO los siguientes adjetivos de severidad: “leve”, “moderado”, “severo” y “grave”.</p>
<p>[ejemplos]</p> <p>46. Antecedentes de <b>pneumonía severa</b> que requirió ingreso y <b>fibrobioma masivo intervenido a los 46 años. de la conducta moderada</b></p>	
<p><b>[EN-P10]</b> [respuestas]</p>	<p>Se deben incluir adjetivos que hagan referencia a la respuesta a un tratamiento cuando especifiquen un subtipo de una enfermedad (por ejemplo: “refractario”).</p> <p>No se deben anotar sin embargo las respuestas a tratamientos por si solas.</p>
<p>[ejemplos]</p> <p>47. Antecedentes de <b>quistes renales</b>, colecistectomía e <b>hipertensión refractaria</b></p> <p>48. Se reinicia precozmente el segundo ciclo por su <b>buena tolerancia al tratamiento</b></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 ENFERMEDAD	
<p><b>[EN-N1]</b> [estadios]</p>	<p>No se deben incluir dentro de la mención estadios, escalas o clasificaciones clínicas.</p> <p>Se excluyen las numeraciones que formen parte del nombre de la enfermedad (como en los tipos de diabetes)..</p> <p>Excepción: regla EN-04.</p>
<p>[ejemplos]</p> <p>49. <b>Infección por VIH estadio C3</b></p> <p>50. <b>Carcinoma renal de células claras con infiltración capsular grado 2 de Fuhrman</b></p> <p>51. <b>Diabetes mellitus II</b></p> <p>52. <b>Glucogenosis tipo IV</b></p>	
<p><b>[EN-N2]</b> [temporalidad]</p>	<p>No se deben incluir dentro de la mención el contexto temporal (edades, “antigua”, “reciente”, “intermitente”, ...), así como tampoco cuantificadores de frecuencia ni ratios.</p> <p>Excepción: ver regla EN-P8 [inicio]</p>
<p>[ejemplos]</p>	



<p>53. <b>En pared abdominal se observó cicatriz antigua</b></p> <p>54. <b>hábito tabáquico importante (20-30 cigarrillos/día</b> desde hacía más de 10 años)</p>	
<p><b>[EN-N3]</b> [episodios]</p>	<p>Las referencias a la episodicidad de la entidad (por ejemplo: “episodio de”, “primer episodio”, “episodio antiguo”, etc.) no se deben incluir en las menciones excepto en el caso que vaya acompañado de un adjetivo calificador que describa suficientemente el tipo de episodio.</p>
<p>[ejemplos]</p> <p>55. <b>Segundo episodio de hematuria macroscópica</b></p> <p>56. Sin otros <b>episodios tromboembólicos</b> anteriores</p>	
<p><b>[EN-N4]</b> [inferencias]</p>	<p>No se deben hacer inferencias a la hora de anotar, por lo que no anotaremos hallazgos clínicos que indiquen la existencia de una enfermedad.</p>
<p>[ejemplos]</p> <p>57. <b>Fracción de eyección del 8-10%.</b></p> <p>[En este extracto de texto no existe una mención explícita de enfermedad, aunque una fracción de eyección menor del 15% es indicio de insuficiencia cardíaca congestiva.]</p>	
<p><b>[EN-N5]</b> [anáforas]</p>	<p>No anotaremos anáforas ni catáforas (referencias a una entidad usando pronombres).</p>
<p>[ejemplos]</p> <p>58. Antecedentes personales de resección de <b>fibroadenoma en mama derecha</b> 4 meses antes de acudir a nuestro Servicio. Con <b>éste</b> único antecedente...</p>	



<p><b>[EN-N6]</b> [incertidumbre]</p>	<p>No se anotarán juntas expresiones de (in)certidumbre como “sospecha de”, “sugestivo de”, “de probable origen”, “de aspecto”, “compatible con”, ...</p>
<p>[ejemplos]</p> <p>59. imagen sugestiva de <b>adenoma paratiroideo</b> 60. sospecha de <b>tumor endocrino pancreático</b> 61. <b>Colección epidural frontal de aspecto infeccioso</b></p>	
<p><b>[EN-N7]</b> [evolución]</p>	<p>No se incluyen en la mención cambios evolutivos del transcurso de las enfermedades (“mejoría”, “empeoramiento”, “aumento”, “persistencia”, ...).</p>
<p>[ejemplos]</p> <p>62. <b>Empeoramiento de la neumonía</b> 63. <b>Disminuye de tamaño la lesión pulmonar</b></p>	
<p><b>[EN-N8]</b> [patógenos]</p>	<p>No se deben anotar organismos patógenos (virus, hongos, bacterias, parásitos, priones, ...) que aparezcan por si solos.</p> <p>Regla relacionada: EN-P4.</p>
<p>[ejemplos]</p> <p>64. En el cultivo de secreción conjuntival crecen <b>gonococos</b> y se administra penicilina</p>	
<p><b>[EN-N9]</b> [determinantes]</p>	<p>No se deben anotar menciones de datos demográficos, etapas de la vida, factores de riesgo, estilos de vida, valoraciones geriátricas ni otros determinantes sociales de la salud.</p>
<p>[ejemplos]</p>	

<p>65. Está en la semana 21 del <b>embarazo</b>.</p> <p>66. Se trata de un problema muy común en <b>lactantes</b>.</p> <p>67. Mujer de 48 años con <b>obesidad mórbida</b>, <b>hábito sedentario</b> y <b>poca adherencia a la dieta mediterránea</b>, consulta por <b>lesiones cutáneas tras la ingesta de moluscos</b></p> <p>68. Paciente de 84 años, <b>vive solo</b>, <b>IAVD</b>, acude por <b>fractura femoral</b> derecha.</p>	
<p><b>[EN-N10]</b> [edad]</p>	<p>Siempre que sea posible, se deben excluir referencias a la edad de las menciones.</p>
<p>[ejemplos]</p> <p>69. Persiste <b>enuresis nocturna hasta los 13 años</b></p>	
<p><b>[EN-N11]</b> [sujeto]</p>	<p>Siempre que sea posible, se debe excluir el sujeto de la frase o persona que sufre/reporta la enfermedad de las menciones..</p>
<p>[ejemplos]</p> <p>70. <b>Abuelo materno fallecido por glioblastoma</b></p> <p>71. <b>La madre consultó con ginecología debido a trastornos menstruales</b></p>	
<p><b>[EN-N12]</b> [códigos]</p>	<p>No se deben anotar códigos de ontologías si aparecen en el texto.</p>
<p>[ejemplos]</p> <p>72. <b>Síndrome de Kabuki (OMIM #147920)</b> o <b>síndrome de Niikawa-Kuroki</b></p>	

### 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. También se incluye una tabla con reglas específicas para los casos de oncología, que suelen ser más complejos y requieren reglas propias.

TABLA 5. REGLAS ESPECIALES PARA ENFERMEDAD	
<p><b>[EN-E1]</b> [convenciones- enf]</p>	<p>A modo de convención, deben anotarse como enfermedades los siguientes conceptos: “alergia”, “toxicidad”, “reacción adversa”, “adenopatía”, “infiltración tumoral”, “quiste”, “déficit de...”, “trombo”, “dilatación de...”.</p>
<p>[ejemplos]</p>	
<p>73. No <b>alergias medicamentosas</b> conocidas</p> <p>74. Ss palpan <b>adenopatías axilares y supraclaviculares</b></p> <p>75. <b>Masa quística abdominal</b></p> <p>76. <b>Déficit de hormona de crecimiento</b></p> <p>77. Imagen compatible con <b>trombo</b></p> <p>78. <b>Ventrículo derecho severamente dilatado</b></p> <p>79. <b>Dilatación de cavidades derechas</b></p>	
<p><b>[EN-E2]</b> [convenciones- sint]</p>	<p>A modo de convención, consideramos síntomas los siguientes conceptos y no deben anotarse en este corpus: “sano”, “toxicidad”, “parestesia”, “ensanchamiento mediastínico”, “síndrome miccional”, “síndrome constitucional”, “deterioro...”, “compromiso...”, así como conceptos relacionados con la muerte.</p>

[ejemplos]	
80. <b>Fallece a los 3 días de ingreso</b>	
81. <b>Muerte neonatal</b>	
82. <b>Deterioro de la función renal</b>	
83. <b>Deterioro cognitivo</b>	
84. <b>Compromiso medular</b>	
85. <b>Compromiso del retorno capilar en pie</b>	
<b>[EN-E3]</b> [hemograma]	En referencia los resultados de hemogramas, se considerarán siempre enfermedad los siguientes conceptos: “anemia”, “bicitopenia”, “tricitopenia”, “pancitopenia”.
<b>[EN-E4]</b> [mutaciones]	Las mutaciones genéticas no se anotarán excepto en aquellos casos que formen parte del nombre de la enfermedad.
[ejemplos]	
86. No presenta <b>mutación de EGFR</b> y en ese momento no se realizó la <b>translocación de ALK</b>	
87. <b>Síndrome de marfan con mutación en el gen FBNI</b>	
<b>[EN-E5]</b> [adicciones]	Las conductas adictivas se considerarán siempre enfermedad..
[ejemplos]	
88. <b>Ex-fumador</b>	
89. <b>Consume marihuana</b> todos los fines de semana	



<p><b>[EN-E6]</b> [ausencia-órganos]</p>	<p>La ausencia de órganos se anotará como enfermedad en las distintas formas que se exprese.</p>
<p>[ejemplos]</p> <p>90. En la ecografía <b>no se observaron trompas, ovarios ni útero</b></p> <p>91. El TAC cerebral mostró <b>ausencia de cuerpo calloso</b></p>	
<p><b>[EN-E7]</b> [lesiones]</p>	<p>La palabra “lesión” más un complemento del nombre (adjetivo, sintagma preposicional) se anotará como enfermedad, así como la palabra “lesión” por sí sola.</p>
<p>[ejemplos]</p> <p>92. <b>Lesión quística</b></p> <p>93. <b>LOE (lesión ocupante de espacio)</b></p>	
<p><b>[EN-E8]</b> [anomalía-morfo]</p>	<p>Se anotan como enfermedad las anomalías morfológicas que están clasificadas como trastorno en SNOMED-CT. Se debe incluir el color y la forma de las lesiones dermatológicas.</p>
<p>[ejemplos]</p> <p>94. Presenta <b>telangiectasias</b> de una semana de evolución</p> <p>95. Presentaba más de 8 <b>manchas café con leche en tronco y extremidades</b></p>	

#### TABLA 6. REGLAS ESPECÍFICAS PARA ONCOLOGÍA

<p><b>[EN-O1]</b> [tumoraciones]</p>	<p>Los tumores, quistes y neoplasias se anotarán como enfermedades.</p>
--	---

	<p>Las tumoraciones, neoformaciones, nódulos y masas por sí solos se considerarán síntomas. En este corpus, solo se anotan cuando van junto a un término que haga referencia a una enfermedad (por ejemplo: “masa quística”, “neoformación polipoidea”, ...).</p>
<p>[ejemplos]</p> <p>96. <b>tumor testicular tipo seminoma</b></p> <p>97. <b>masa quística retroperitoneal</b></p> <p>98. <b>Masa renal</b></p>	
<p><b>[EN-O2]</b> [metástasis-afectación]</p>	<p>Cuando se mencione la presencia de metástasis junto con el nombre de un tumor, se anotará todo junto.</p> <p>Además, cuando se hable de la zona a la que afecta la metástasis, se debe anotar la palabra “afectación” + la zona anatómica como enfermedad (“afectación” no se debe anotar si no se refiere a metástasis).</p>
<p>[ejemplos]</p> <p>99. <b>Tumores primarios tiroideos con metástasis a nivel de ganglios mediastínicos y parénquima pulmonar</b></p> <p>100. El estudio de extensión muestra <b>afectación laterocervical, cerebelosa y mastoidea.</b></p> <p>101. Tras el primer brote de <b>esclerodermia</b> presenta lesiones compatibles con <b>afectación articular y cutánea</b></p>	
<p><b>[EN-O3]</b> [forma-tamaño]</p>	<p>No se debe incluir en las menciones la descripción del tamaño y forma macroscópica de masas y tumoraciones.</p>
<p>[ejemplos]</p>	

102. <b>Quiste pulmonar masivo que desplaza el mediastino a la derecha</b>	
<b>[EN-O4]</b> [diferenciación]	Como excepción a la regla EN-N1, y a modo de convención, se anotan los siguientes términos usados para la diferenciación histológica (grados de diferenciación): “pobrementemente”, “moderado”, “bien”, “alto”, “bajo”.
[ejemplos]	
103. <b>Leiomioma fusocelular de alto grado</b>	
104. <b>Liposarcoma bien diferenciado</b>	
105. <b>Carcinoma ductal microinfiltrante pobrementemente diferenciado</b>	
106. <b>Carcinoma renal de células claras con infiltración capsular grado 2 de Fuhrman</b>	
107. <b>Adenocarcinoma Gleason 5 + 4</b>	
108. <b>Meduloblastoma clásico de localización vermiana grado IV de la OMS</b>	
<b>[EN-O5]</b> [recidiva]	Si un tumor es una recidiva, se debe incluir en la mención. La palabra “recidiva” por sí sola, sin embargo, no debe anotarse.
[ejemplos]	
109. <b>Prolactinoma gigante intervenido y recidivado</b> , pendiente de radioterapia	
<b>[EN-O6]</b> [proteínas]	No se anotan las proteínas expresadas por el tumor.
[ejemplos]	
110. El <b>tumor expresa queratina 7, TTF-1 y tiroglobulina</b>	

## 4. REGLAS DE NORMALIZACIÓN

La segunda parte de la anotación del corpus DISTEMIST consiste en asociar cada mención anotada a uno o más códigos de SNOMED-CT.

El proceso y reglas de normalización se añadirá en la siguiente versión de estas guías.