



# Guías LivingNER: anotación, normalización y clasificación de especies, seres vivos y patógenos/enfermedades infecciosas

(LivingNER Guidelines: annotation, mapping and classification of pathogens, living organisms and infectious diseases)

**V3 [Abril 2022]**

## **AUTORES**

Eulàlia Farré-Maduell (BSC)  
Salvador Lima López (BSC)  
Gloria González (BITAC)  
Antonio Miranda-Escalada (BSC)  
Martin Krallinger (BSC)



**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>3</b>
<b>2. DEFINICIÓN DE LA TAREA</b>	<b>8</b>
<b>3. REGLAS DE ANOTACIÓN</b>	<b>11</b>
3.1 Reglas generales	11
3.2 Reglas positivas	17
TABLA 3. REGLAS POSITIVAS PARA SPECIES	17
TABLA 4. REGLAS POSITIVAS PARA HUMAN	20
TABLA 5. REGLAS POSITIVAS PARA ENFERMEDAD	22
3.3 Reglas negativas	26
TABLA 6. REGLAS NEGATIVAS PARA SPECIES	26
TABLA 7. REGLAS NEGATIVAS PARA HUMAN	27
TABLA 8. REGLAS NEGATIVAS PARA ENFERMEDAD	28
3.4 Reglas especiales y convenciones	30
TABLA 9. REGLAS ESPECIALES PARA SPECIES	30
TABLA 10. REGLAS ESPECIALES PARA HUMAN	35
TABLA 11. REGLAS ESPECIALES PARA ENFERMEDAD	36
TABLA 12. REGLAS MULTICLASE	37
<b>4. REGLAS DE NORMALIZACIÓN</b>	<b>38</b>
4.1 Reglas básicas	39
TABLA 13. REGLAS BÁSICAS DE NORMALIZACIÓN	39
TABLA 14. MODIFICADORES DE CÓDIGOS	40
4.2 Reglas específicas y convenciones	42
TABLA 15. CONVENCIONES DE NORMALIZACIÓN	42
TABLA 16. NORMALIZACIÓN DE CONCEPTOS INESPECÍFICOS	44
<b>5. REGLAS DE CLASIFICACIÓN</b>	<b>46</b>
TABLA 17. REGLAS GENERALES DE CLASIFICACIÓN DE ALIMENTACIÓN	46
TABLA 18. REGLAS GENERALES DE CLASIFICACIÓN DE ANIMALES DE COMPAÑÍA	47
TABLA 19. REGLAS GENERALES DE CLASIFICACIÓN DE ANIMALES CAUSANTES DE LESIÓN	48
<b>Bibliografía</b>	<b>51</b>

# 1. INTRODUCCIÓN

Esta guía describe el proceso de anotación y normalización del corpus **LivingNER**, una colección de 2.000 casos clínicos en castellano de 20 especialidades médicas (enfermedades infecciosas (incluidos casos de Covid-19), cardiología, neurología, oncología, otorrinolaringología, odontología, pediatría, endocrinología, atención primaria, alergología, radiología, psiquiatría, oftalmología, psiquiatría, urología, medicina interna, emergencias y medicina de cuidados intensivos, radiología, medicina tropical y dermatología) anotados con menciones de **especies** (incluyendo seres vivos y microorganismos) y **enfermedades infecciosas**. Cada una de las menciones del corpus ha sido normalizada usando la terminología NCBI Taxonomy<sup>1</sup>. Por último, las menciones de especies se han clasificado en unas clases de interés clínico para mejorar su aprovechamiento (animales de compañía, ataques de animales, alimentación, entidades nosocomiales y bacterias resistentes a antibióticos).

## La relevancia de las especies y enfermedades infecciosas

La detección de especies en la historia clínica es imprescindible por la carga de enfermedad que representan los patógenos en el caso de enfermedades infecciosas; seres vivos y alimentos en el caso de alergología. El corpus distingue también entre patógenos resistentes a los antibióticos e infecciones adquiridas en el hospital (nosocomiales), que representan una alta causa de morbilidad y mortalidad por la dificultad de encontrar tratamiento adecuado. También referencia y normaliza menciones a las distintas floras del organismo humano, en preparación para la literatura y casos clínicos relacionados con el microbioma. Se anotan también todas las formas de los ciclos parasitarios, para su uso en clínica y microscopía.

La detección y clasificación de especies y enfermedades infecciosas en literatura e informes médicos es crítica para múltiples aplicaciones que tienen se relacionan con disciplinas científicas como la medicina, biología, ecología, nutrición o agricultura. Algunas de las aplicaciones más relevantes están relacionadas con la extracción de información sobre

---

<sup>1</sup> <https://www.ncbi.nlm.nih.gov/taxonomy>

enfermedades sexuales y su modo de transmisión, sobre ataques de animales a humanos o sobre los animales con los que convive un humano con una patología. La correcta extracción de especies y enfermedades infecciosas permitiría, por ejemplo, la clasificación de las menciones de bacterias según si aparecen en un contexto de resistencia a antibióticos o no, o de patógenos y enfermedades según si son nosocomiales. Otra aplicación que depende de esta tarea puede ser la estructuración de la información relacionada con la comida (dietas, intoxicaciones, hábitos tóxicos, etc), con los desencadenantes de alergias, o con menciones de relevancia epidemiológica como pueden ser los contactos estrechos, convivientes y parientes.

## Trabajo previo

Debido a la relevancia de esta tarea, se han realizado esfuerzos de larga duración (más de 250 años) para definir las reglas de nomenclatura científica y las convenciones de denominación de las especies, seguidos de esfuerzos para recoger sistemáticamente las especies en catálogos, bases de datos y recursos terminológicos con relaciones taxonómicas jerárquicas. En este contexto cabe destacar recursos como la taxonomía del NCBI, la lista de nombres científicos de Thompson, el Catálogo de la Vida o la base de datos del Índice de Nombres Globales o el Catálogo del ITIS. Sin embargo, estos esfuerzos no se han alineado adecuadamente con el desarrollo de sistemas automáticos para el análisis semántico de las menciones de especies en el texto, especialmente cuando se consideran documentos más allá del inglés. Estas anotaciones semánticas tienen que hacer frente a los constantes cambios de denominación a lo largo del tiempo (nombres de especies obsoletos), a la homonimia con palabras de uso común (p. ej: "Spot" se refiere a la especie *Leiostomus xanthurus* o "Permit" a *Trachinotus falcatus*), abreviaturas (a veces muy ambiguas como EC, que puede utilizarse tanto para la bacteria "*Escherichia coli*" como para "células endoteliales" o "cáncer de endometrio"), casos incorrectos o nombres mal escritos (*Bacterium coli*, *Bacillus coli* y *Escheria coli* para *Escherichia coli*), coordinaciones y expresiones anidadas ("virus de la inmunodeficiencia humana tipos 1 y 2"), formas vernáculas (nombres comunes) o nombres de rol (por ejemplo por ejemplo, deportistas, respondedores).

Para solventar estas limitaciones, ya existen corpus y herramientas disponibles para la identificación de especies en la literatura biomédica en lengua inglesa y su normalización a

vocabularios controlados. Por ejemplo, LINNAEUS (Gerner et al., 2010) y la herramienta SPECIES (Pafilis et al., 2013) son capaces de detectar menciones de especies. Estos recursos fueron consultados para centrar aspectos tan relevantes como el objetivo de la anotación, la calidad mínima requerida y diferentes reglas de anotación iniciales. Sin embargo, la adaptación del corpus a un idioma y tipo de datos distintos ha requerido ejecutar modificaciones. Además, la adaptación de los sistemas automáticos publicados por estos autores es compleja, debido a su uso de sistemas basados en reglas y patrones. Por ello, las guías y corpus LivingNER son un recurso indispensable para detección y clasificación de especies y enfermedades infecciosas en literatura e informes médicos en castellano.

## **NCBI Taxonomy**

Las entidades manualmente anotadas en los textos son normalizadas a identificadores de NCBI Taxonomy. The National Center for Biotechnology Information (NCBI) Taxonomy incluye nombres de organismos clasificados principalmente en base a una jerarquía filogenética. NCBI Taxonomy es una base de datos universal, utilizada por la International Nucleotide Sequence Database Collaboration (INSDC), que incluye GenBank, el European Molecular Biology Laboratory (EMBL) y DNA Data Bank of Japan (DDBJ) como única fuente de clasificación taxonómica para mantener la consistencia entre bases de datos. En NCBI cada código único identifica un tipo específico de organismo (por ejemplo, Taxonomy ID: 5476 para *Candida Albicans*), o grupos de organismos (Taxonomy ID: 40674 para mamíferos). Por último, el corpus de LINNAEUS también emplea NCBI Taxonomy como vocabulario controlado para normalizar las menciones.

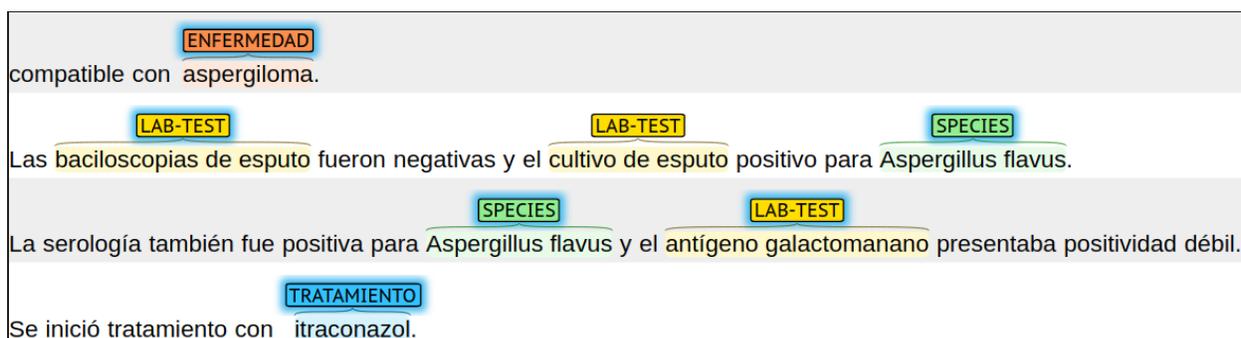
## **Alcance de la guía**

Para cubrir un rango de casos de uso tan ambiciosos como los planteados, esta guía ha sido usada para la anotación y normalización de especies en literatura, casos clínicos e informes médicos. Además, ha sido empleada en documentos de más de 10 disciplinas médicas. Por último, aunque la guía está diseñada para su uso en documentos en español, se han realizado pilotos en inglés, catalán y portugués.

## Limitaciones

A pesar de que el reconocimiento y clasificación de las menciones de especies y enfermedades infecciosas es muy relevante, para construir algunas herramientas de uso clínico o científico es necesario completarlo con otras entidades.

Dentro de estas, las que se han considerado más relevantes son los tests de laboratorio y los tratamientos, porque aportan información fundamental sobre la detección de los patógenos en la clínica y sobre el tratamiento que sigue el paciente. Estas entidades no se han incluido ni en el corpus, ni en la guía, debido a la complejidad que hubieran añadido al proyecto. Se plantea su inclusión en un proyecto posterior.



**Figura 1.** Ejemplo de la fase inicial de anotación del corpus LivingNER..

## El corpus LivingNER

El corpus LivingNER ha sido anotado y normalizado por una especialista en el dominio con el apoyo de una especialista clínica, quien se encargó también de revisar las menciones y sus códigos asociados para llegar a una versión final. El proceso de anotación y normalización del

corpus tuvo lugar entre los años 2020 y 2021, con una duración aproximada de 5 meses usando la herramienta *brat*<sup>2</sup>.

Antes de comenzar la anotación, se creó un primer borrador de estas guías basándonos en nuestras experiencias previas de anotación (MEDDOCAN (Marimon et al., 2020), CANTEMIST (Farré et al., 2020) o MEDDOPROF (Farré et al. 2020) entre otras) y trabajos previos en el dominio (Gerner et al., 2010, Pafilis et al., 2013). Las guías se refinaron mediante varias rondas de *inter-annotator agreement* (IAA) consistentes en la anotación paralela de un 5% del corpus. Tras varias rondas, se llegó a una puntuación de IAA total de 0.942 para las especies y 0.885 para las enfermedades infecciosas.

Además, durante el resto de la anotación de LivingNER se revisaron exhaustivamente un 10% aleatorio de los documentos para asegurar que la calidad se mantenía. También 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 NCBI Taxonomy para asignar un identificador a cada anotación manual.

La versión final del corpus LivingNER incluye 30886 menciones de especies, de las cuales el 43.9% corresponde a humanos, 4580 son únicas y 29411 están normalizadas a NCBI Taxonomy. Además, contiene 11841 menciones de enfermedades infecciosas, siendo 4093 de ellas únicas y estando 2283 normalizadas a NCBI Taxonomy. El total son 42727 menciones.

Por último, todas menciones de especies se han clasificado en unas clases de interés clínico para mejorar su aprovechamiento (animales de compañía, ataques de animales, alimentación, entidades nosocomiales y bacterias resistentes a antibióticos).

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>2</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 **especies** y **enfermedades infecciosas** en casos clínicos. Consideraremos especies a todos los seres vivos del reino animal y vegetal, con consideraciones especiales para seres humanos y patógenos.

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 la clasificación NCBI Taxonomy, así como de otras terminologías relacionadas de interés como SNOMED CT o CIE 10 MC.
- Conocimiento de la nomenclatura específica en microbiología.
- Conocimiento de los términos, abreviaturas y siglas utilizadas habitualmente en microbiología.
- Habilidades en la búsqueda de información científica y recursos terminológicos.

La tarea tiene tres fases: anotación, normalización y clasificación. 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 SPECIES/HUMAN/ENFERMEDAD que aparece como una de las opciones en formato *radio buttons* (sólo permite seleccionar 1)

La figura 2 muestra un ejemplo de un documento anotado utilizando la interfaz de *brat*.

12	La probabilidad de una prueba preliminar para <b>ENFERMEDAD</b> COVID-19 era entre moderada y alta a causa de la fiebre, disnea e hipoxia, así como las radiografías previas que mostraban opacidades de vidrio esmerilado periféricas.
13	Una prueba de RT-PCR para el <b>SPECIES</b> coronavirus del síndrome respiratorio agudo grave 2 ( <b>SPECIES</b> SARS-CoV-2) fue positiva en la visita inicial a urgencias y en nuestra institución.
14	Varios marcadores de enfermedad presentaban valores anómalos, como la cifra de leucocitos, la proteína C-reactiva y las enzimas cardíacas.
15	Una radiografía torácica mostró infiltrados bilaterales asociados a <b>ENFERMEDAD</b> neumonía atípica.

**Figura 2.** 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. Es relevante aclarar que durante el proceso de anotación original, especies y humanos se anotaron como una sola etiqueta. Gracias a la normalización, las menciones de humanos se separaron en un paso posterior extrayendo las menciones normalizadas como *Homo sapiens*.

TABLA 1. ETIQUETAS DISPONIBLES EN BRAT	
<b>[SPECIES]</b>	Comprende cualquier forma de organismo vivo a excepción de humanos (bacterias, virus, parásitos, hongos, animales, plantas, frutas, verduras...).
<b>[HUMAN]</b>	Menciones que hagan referencia a humanos, incluyendo entre otros: familiares, etapas de la vida, ocupaciones, ...
<b>[ENFERMEDAD]</b>	Enfermedades de etiología infecciosa, como absceso parotídeo, rubéola o sífilis.

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 NCBI

Taxonomy cuando la correspondencia entre entidad y taxonomía sea unívoca. Para este proceso deben seguirse las reglas y convenciones especificadas en la sección 4 (“Reglas de normalización”).

La normalización se realizó en dos pasos. En primer lugar, durante la anotación, se normalizaron directamente en la herramienta brat aquellas menciones que tuviera cualquier sospecha de ambigüedad (por ejemplo, las siglas y acrónimos) usando el apartado de comentarios. La segunda fase se llevó a cabo en una hoja de cálculo.

En esta segunda fase, el codificador dispone de la siguiente información: etiqueta, entidad, código anotado en la primera fase, modificadores de código anotados en la primera fase y sugerencias de normalización. Con ella, debe añadir en la hoja de cálculo el código final como se muestra en la Figura 3.

label	span_norm	Code	isN	isF	isO	isR	isI	isH	code_	code_sug	code_sug_link	code_sug_scientifi	code_sug_cc	FINAL_CODE
SPECIES	(htlv) I									1	https://www.ncbi root			11908
SPECIES	(htlv) I y II									1	https://www.ncbi root			11908 11909
SPECIES	(VHS) 1									1	https://www.ncbi root			10298
SPECIES	(VHS) 1 y 2									1	https://www.ncbi root			10298 10310
SPECIES	(vih)-1													11676
SPECIES	(vih)-1/2													11676 11709
SPECIES	2019-ncov								2697049		https://www.ncbi	Severe acute respiratory syndrome coro		2697049
SPECIES	A. actinomycetemcomitans								714		https://www.ncbi	Aggregatibacter actinomycetemcomitai		714
SPECIES	A. aphrophilus								732		https://www.ncbi	Aggregatibacter aphrophilus		732
SPECIES	A. baumannii-complex	N	T											909768 N
SPECIES	A. fumigatus								746128		https://www.ncbi	Aspergillus fumigatus		746128
SPECIES	A. lumbricoides								6252		https://www.ncbi	Ascaris lumbricoides	common roundv	6252
SPECIES	A. nomius								41061		https://www.ncbi	Aspergillus nomiae		41061
SPECIES	A. variegatum								112594 3461		https://www.ncbi	Aconitum variegatum	tropical bont ti	34610
SPECIES	a.fumigatus								746128		https://www.ncbi	Aspergillus fumigatus		746128
SPECIES	AAR			2 H						1	https://www.ncbi root			2 H
SPECIES	AAR									1	https://www.ncbi root			2 H
ENFERMEDAD	abcedificación													NOCODE

**Figura 3.** Ejemplo del proceso de normalización.

La tercera parte de la tarea corresponde a la clasificación de las entidades en clases de interés clínico. Todas menciones de especies se han clasificado en unas clases de interés clínico para mejorar su aprovechamiento (animales de compañía, ataques de animales, alimentación, entidades nosocomiales y bacterias resistentes a antibióticos).

En caso de duda del anotador, se aconseja consultar recursos externos. Recomendamos usar sobre todo el navegador de la NCBI Taxonomy<sup>3</sup>, así como Wikipedia u otras páginas web especializadas como MedlinePlus.

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 patógenos, seres vivos y enfermedades infecciosas.

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	
<b>[G1]</b> [exhaustividad]	Deben marcarse todas las apariciones de una entidad, incluso si ya se ha marcado anteriormente en el mismo texto o frase.

<sup>3</sup> <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>

[ejemplos]	
1. Tuvieron que extraerle una <b>garrapata</b> . Se mandó analizar la <b>garrapata</b> al laboratorio de referencia	
<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>
[ejemplos]	
2. Lesión periorbitaria de <b>tinea facies</b>	
[En esta mención, además de enfermedad, <i>tinea facies</i> se podría considerar especie. Si no encontramos evidencia de que se refiera a una especie, la anotaremos solo como enfermedad y la apuntaremos para discutirla].	
<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>
[ejemplos]	
3. Frasco <b>anaerobios</b> -> <b>Enterococcus faecalis</b>	
<ul style="list-style-type: none"> <li>● Sensible a penicilina</li> </ul>	

[En este ejemplo nos interesaría anotar que el *E. faecalis* es sensible a penicilina. Sin embargo, como la información está en dos oraciones separadas, anotaremos solo “*Enterococcus faecalis*”.]

**[G5]**  
[errores-  
ortográficos]

Las menciones que incluyan algún tipo de error ortográfico (por ejemplo: letras de más o de menos, espacios incorrectos) también deben anotarse.

[ejemplos]

4. **Heptitis B** (*Hepatitis B*)
5. **Poliomelitis vacunal** (*Poliomielitis vacunal*)
6. **Osteo mielitis** (*Osteomielitis*)

**[G6]**  
[palabras-  
completas]

Todas las menciones deben estar compuestas por palabras enteras. No se pueden anotar palabras a medias.

[ejemplos]

**7. Absceso laterocervical**

**Absceso latero**cervical

**[G7]**  
[determinantes]

No se deben incluir dentro de una mención los determinantes que aparezcan al principio (por ejemplo: *el, la, este, ese*).

[ejemplos]

8. La **fiebre de las Rocky Mountains**

<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 style="text-align: center;">[ejemplos]</p> <p>9. <b>neumonía difusa (presuntivamente por <i>P. carinii</i>)</b></p> <p style="text-align: center;"><b>neumonía difusa (presuntivamente por <i>P. carinii</i>)</b></p> <p style="text-align: center;">[En este ejemplo hay dos anotaciones: una especie ("<i>P. carinii</i>") y una enfermedad ("neumonía difusa (presuntivamente por <i>P. carinii</i>)". Si no incluimos el paréntesis de cierre en la enfermedad, la anotación es incorrecta]</p> <p>10. (<b><i>P. jiroveci</i></b>)</p> <p>11. <b>flora bacteriana habitual</b>, cultivo de <b>Nocardia</b>: negativo, Ziehl-auramina: negativo, <b><i>Aspergillus</i></b>: negativo, <b><i>C. pneumoniae</i></b>: negativo, PCR <b>CMV</b> y <b>virus respiratorios</b>: negativo, antigenuria <b>neumococo</b>.</p> <p>12. La evaluación <b>médico-enfermera</b> constató...</p>	
<p><b>[G9]</b> [otros-idiomas]</p>	<p>Se anotan las menciones en todos los idiomas.</p>
<p style="text-align: center;">[ejemplos]</p> <p>13. <b>Lyme disease, AIDS</b></p>	

<p><b>[G10]</b> [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>Esta regla funciona conjuntamente a menudo con las reglas EN-P5 a EN-P8.</p>
<p>[ejemplos]</p> <p>14. <b>absceso drenado en territorio de fibrosa mitroaórtica</b> [palabras relevantes: “absceso” y “mitroaórtica”.] [Distancia: 5 tokens]</p> <p>15. <b>ENDOCARDITIS MITRAL sobre válvula con degeneración mixomatosa por S. AUREUS</b> [palabras relevantes: “endocarditis mitral” y “S. aureus”.] [Distancia: 6 tokens]</p> <p>16. <b>Miocarditis aguda de probable etiología vírica</b> [palabras relevantes: “Miocarditis aguda” y “vírica”.] [Distancia: 3 tokens]</p>	
<p><b>[G11]</b> [negaciones]</p>	<p>Las menciones que estén negadas también se anotan.</p>
<p>[ejemplos]</p> <p>17. Sin antecedentes de <b>enfermedades de transmisión sexual</b></p>	
<p><b>[G12]</b> [no-paciente]</p>	<p>Los menciones que no hagan referencia al paciente (por ejemplo: antecedentes familiares) también deben anotarse.</p>

[ejemplos]

18. **Paciente** del sexo **femenino**, de 34 años de edad, **pareja** de adicto, con dos **hijos** negativos para el **VIH**

**[G13]**  
[abreviaturas-  
siglas]

Una mención puede ser una abreviatura o una sigla de la entidad taxonómica.

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.

En el caso de abreviaturas que se pueden corresponder con más de un patógeno se adopta la más representada teniendo en cuenta el contexto clínico: E.coli (*Escherichia/Entamoeba coli*) será por defecto *Escherichia coli* si el texto no nos indica lo contrario.

Sin embargo, no anotaremos entidades en siglas que nieguen su carácter infeccioso (ver segundo ejemplo).

[ejemplos]

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

20. **ETNB** (*endocarditis trombótica no bacteriana*)

**[G14]**  
[menciones-  
anidadas]

Están permitidas las anotaciones anidadas (menciones que se solapan unas con otras) en los siguientes casos:

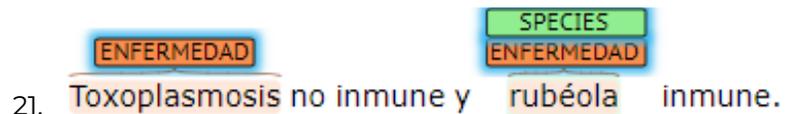
- a) Palabras o expresiones que hacen referencia tanto a una especie taxonómica como a una enfermedad (ver primer ejemplo).

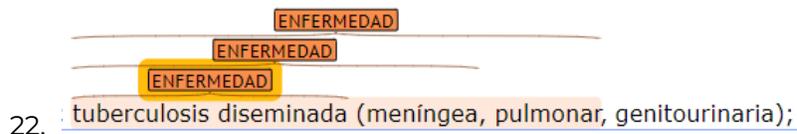
En estos casos, anotaremos cada entidad de manera independiente.

- b) Expresiones que contengan varios conceptos juntos expresados de manera coordinada (ver segundo ejemplo).

En estos casos, se deben anotar varias entidades solapadas hasta un máximo de tres. Cada entidad debe cubrir desde el principio de la expresión hasta el final del concepto, de manera que la primera anotación incluya el primer concepto, la segunda el primero más el segundo, ...

[ejemplos]

21. 

22. 

23. **Parainfluenza 1-4**

[Dos menciones: "Parainfluenza 1" y "Parainfluenza 1-4"]

## 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 SPECIES

<p><b>[SP-P1]</b> [patógenos- seres-vivos]</p>	<p>Se anotan con esta etiqueta tanto microorganismos (bacterias, virus, microbios) como animales y plantas.</p> <p>Las reglas SP-E8 a SP-E12 proporcionan más detalles sobre la anotación de microorganismos.</p>
<p>[ejemplos]</p> <p>24. Refería exposición a ganado <b>ovino</b> en la juventud</p>	
<p><b>[SP-P2]</b> [concretos- generales]</p>	<p>Se deben anotar tanto nombres de especies concretas (por ejemplo: “Aspergillus fumigatus”, “Klebsiella”, “Pseudomonas”, “Staphylococcus aureus”) como nombres generales de tipos de especies (por ejemplo: “Micobacterias”, “hongos”, “anaerobios”, “helmintos”).</p>
<p>[ejemplos]</p> <p>25. El cultivo para <b>bacterias</b> habituales, incluidas <b>Neiserias</b> y <b>Haemophilus</b>, ...</p> <p>26. Crecen dos cepas <b>Staphilococo coagulasa negativo</b> (identificado como <b>Staphylococcus epidermidis</b> y <b>Streptococcus alfa hemolítico</b>)</p> <p>27. Determinación sanguínea de anticuerpos (ELISA) y de existencia del <b>virus (HTLV-I)</b></p>	
<p><b>[SP-P3]</b> [raíz- taxonómica]</p>	<p>Se anotan las palabras que en la raíz contengan una palabra taxonómica de una especie independientemente del contexto, excepto cuando se refieran a fármacos específicos (penicilina, estreptomicina).</p>
<p>[ejemplos]</p>	

28. **Antiviral, antifúngico, antibrucella, antituberculoso, antirretroviral, antitetánica, TARV, tuberculostático**

**[SP-P4]**  
[derivadas]

Se anotan las palabras derivadas de entidades taxonómicas independientemente del contexto, excepto cuando se refieran a fármacos específicos (penicilina, estreptomicina).

[ejemplos]

29. Antiviral, antifúngico, antibrucella, antituberculoso, antirretroviral, antitetánica, TARV, tuberculostático

30. **bacteriuria** abundante

31. crecimiento **polimicrobiano**

32. Con la sospecha de **endocarditis bacteriana subaguda**

[En este caso, se anota “endocarditis bacteriana subaguda” como enfermedad y “bacteriana” como especie. Abajo igual.]

33. Sospechando una **glomerulonefritis (GN) aguda**, probablemente **postinfecciosa**

**[SP-P5]**  
[vulgares]

Se deben anotar tanto los nombres taxonómicos como los vulgares.

En el caso de que el nombre vulgar y el taxonómico en latín aparezcan juntos, se anotarán separadamente.

[ejemplos]

34. **Staphylococcus, Trypanosoma brucei** (nombres taxonómicos)

35. **Estafilococo, parásito de la enfermedad del sueño** (nombres vulgares)

36. ...veneno de **avispa (Vespula spp.)** < 0.1 kU/L

<p><b>[SP-P6]</b> [figurativas]</p>	<p>Las menciones que hagan referencia a una especie y que estén usadas de manera figurativa también deben anotarse.</p> <p>Las menciones de este tipo deben normalizarse añadiendo el sufijo “ F” (más detalles en la sección 4).</p>
<p>[ejemplos]</p> <p>37. A la exploración presenta múltiples lesiones faciales con distribución en alas de <b>mariposa</b>, de aspecto plano</p> <p>38. Es considerada como la "<b>oveja</b> negra" de la <b>familia</b>.</p> <p>39. ...en capas de <b>cebolla</b>...</p>	
<p><b>[SP-P7]</b> [comida]</p>	<p>Se anotarán los vegetales y animales que puedan ser alimentos.</p>
<p>[ejemplos]</p> <p>40. Recomendamos una dieta de <b>patata</b> hervida o en puré</p>	

<p><b>TABLA 4. REGLAS POSITIVAS PARA HUMAN</b></p>	
<p><b>[HM-P1]</b> [humanos]</p>	<p>Deben marcarse las palabras que hagan referencia a humanos de manera general: “humano”, “paciente”, “persona”, ... así como a grupos de personas: “familia”.</p> <p>Excepción: ver regla HM-N1</p>
<p>[ejemplos]</p> <p>41. <b>Paciente</b> de 36 años...</p>	

<p><b>[HM-P2]</b> [familiares]</p>	<p>Deben marcarse los sustantivos y adjetivos que hagan referencia a relaciones familiares (“madre”, “hermana”, “abuela”, “sobrina”, ...), así como posiciones dentro de la familia (por ejemplo: “el pequeño”).</p> <p>Si aparece la rama de la familia (materna/paterna), debe incluirse dentro de la mención.</p>
[ejemplos]	
<p>42. <b>Lactante</b>, de raza caucasiana, la primera <b>hija</b> de <b>padres</b> jóvenes (<b>madre</b> de 28 años y <b>padre</b> de 30 años).</p> <p>43. El <b>tío paterno</b> había sido diagnosticado de...</p> <p>44. El <b>mayor</b> de tres <b>hermanos</b> presentó convulsiones.</p>	
<p><b>[HM-P3]</b> [etapas-vida]</p>	<p>Deben marcarse las distintas etapas de la vida cuando aparezcan en forma sustantiva: “juventud”, “infancia”, “adolescencia”, “anciano”, ...</p>
[ejemplos]	
<p>45. Acude con <b>lactante</b> de tres meses por erupción vesiculiforme.</p>	
<p><b>[HM-P4]</b> [sexo-género]</p>	<p>Deben marcarse las palabras que hagan referencia al sexo o género de una persona.</p> <p>En el caso de que una de estas palabras aparezca junto a la palabra “paciente”, se anotarán ambas como una única entidad.</p>
[ejemplos]	
<p>46. <b>Varón</b> de 56 años que acude por...</p>	

47. <b>Paciente femenina</b> de 62 años de edad, obesa, diabetes e hipertensa de un año de evolución.	
<b>[HM-P5]</b> [estado-civil]	Deben marcarse los sustantivos que hagan referencia al estado civil de una persona (“soltero”, “viudo”, “divorciado”, ...).
[ejemplos]	
48. Acude a consulta una <b>viuda</b> de 81 años con...	
<b>[HM-P6]</b> [profesiones]	Deben marcarse las profesiones en forma sustantiva, así como equipos y grupos de profesionales.  Las menciones de este tipo deben normalizarse añadiendo el sufijo “ O” (más detalles en la sección 4).
[ejemplos]	
49. Trabaja como <b>cardióloga</b> en un hospital de la comarca.	
50. Empezó a trabajar como <b>barrendero</b> el pasado octubre.	
51. Está pendiente de unas pruebas <b>médicas</b> .	
52. El <b>equipo de oncólogos</b> tomó la decisión de...	

**TABLA 5. REGLAS POSITIVAS PARA ENFERMEDAD**

<b>[EN-P1]</b> [infecciosas]	Deben marcarse todas las enfermedades causadas por infecciones, incluso si el nombre no lo indica implícitamente..
---------------------------------	--

[ejemplos]	
53. En la <b>artritis reactiva</b> es habitual encontrar en su presentación inflamación articular junto a <b>conjuntivitis</b> o <b>uretritis</b> .	
<b>[EN-P2]</b> [generales]	Deben marcarse también las menciones muy generales (por ejemplo: "neumonía atípica", "absceso bacteriano", "diarrea del viajero").
[ejemplos]	
54. Se administra ceftriaxona para tratar el proceso <b>infeccioso</b>	
55. <b>faringoamigdalitis de probable etiología viral</b>	
56. <b>gastroenteritis aguda</b>	
<b>[EN-P3]</b> [concretas]	Deben marcarse todas las enfermedades causadas por gérmenes específicos (por ejemplo: "aspergilosis pulmonar", "filariasis", "malaria", "neumonía pneumocócica", "cirrosis por VHC").
[ejemplos]	
57. Entre los comensales aparecieron varios casos de <b>brucelosis</b>	
58. <b>Toxoplasmosis</b> no inmune y <b>rubéola</b> inmune	
59. Se diagnostica de <b>tuberculosis genitourinaria</b> con <b>tuberculoma vesical</b> y afectación renoureteral izquierda	
60. hepatomegalia a 2 cm del borde costal y <b>condilomas vulvares</b> .	
<b>[EN-P4]</b> [nosocomiales]	Se deben anotar todas las infecciones nosocomiales, es decir, las adquiridas durante un ingreso hospitalario o a través de un procedimiento médico.

	<p>Las menciones de este tipo deben normalizarse añadiendo el sufijo “ N” (más detalles en la sección 4).</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p>61. Con el diagnóstico de <b>endocarditis sobre catéter</b> se inició tratamiento</p> <p>62. Ingresa en urología con el diagnóstico de <b>infección de herida operatoria</b></p> <p>63. Se aísla <b>Staphylococcus aureus resistente a meticilina</b> en muestra hospitalaria</p> <p style="text-align: center;">64. <b>Neumonía intrahospitalaria por Aspergillus spp</b></p> <p>[Se anota “neumonía intrahospitalaria por Aspergillus spp” como enfermedad con el código 5065 N y también “Aspergillus spp.” como especie con el mismo código 5065 N]</p>
<p><b>[EN-P5]</b> [agente-causal]</p>	<p>Siempre que sea posible las menciones deben incluir el agente causal.</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p>65. Los cultivos y la patología confirman <b>infección por Mucor</b>.</p> <p>[en este caso infección por Mucor tendría doble anotación: enfermedad: infección por Mucor y especie: Mucor]</p>
<p><b>[EN-P6]</b> [localización]</p>	<p>Siempre que sea posible las menciones deben incluir la localización anatómica.</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[ejemplos]</p> <p style="text-align: center;">66. <b>Tuberculosis pulmonar y cutánea</b></p>

<p><b>[EN-P7]</b> [ordenamiento]</p>	<p>Siempre que sea posible las menciones deben incluir las fases de enfermedades expresadas mediante ordinales (“primaria”, “secundaria”, ...).</p>
<p>[ejemplos]</p> <p>67. <b>Tuberculosis primaria</b></p> <p>68. <b>Sífilis secundaria</b></p>	
<p><b>[EN-P8]</b> [localización]</p>	<p>Siempre que sea posible las menciones deben incluir el curso de la enfermedad (“aguda”, “recurrente”, “crónica”, ...).</p>
<p>[ejemplos]</p> <p>69. <b>Herpes zóster recurrente</b></p> <p>70. <b>Hepatitis B crónica</b></p> <p>71. <b>Neumonía aguda bacteriana</b></p>	
<p><b>[EN-P9]</b> [manifestaciones]</p>	<p>Se deben anotar también las manifestaciones clínicas claramente relacionados con procesos infecciosos.</p>
<p>[ejemplos]</p> <p>72. Se presenta en urgencias con <b>otalgia, exantema truncal y facial</b></p> <p>73. Acude por empeoramiento de <b>exantema vesiculoso sugestivo de varicela</b></p> <p>74. La madre refiere <b>tos pertusoide, febrícula y micropetequias faciales</b></p>	

<p><b>[EN-P10]</b> [modificadores]</p>	<p>Las menciones que incluyan algún tipo de modificador sobre aspecto o certeza (“predominantemente”, “fundamentalmente”, “con características de”, “con rasgos de”, “de probable”, “con sospecha de”, ...) también deben anotarse..</p>
<p>[ejemplos]</p>	
<p>75. <b>Neumonía difusa presuntamente por P. carinii</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 6. REGLAS NEGATIVAS PARA SPECIES	
<p><b>[SP-N1]</b> [criaturas-fantásticas]</p>	<p>No se deben marcar referencias a criaturas fantásticas, entes sobrenaturales y religiosos, ....</p>
<p>[ejemplos]</p>	
<p>76. Su animal favorito es el <b>unicornio</b>.</p> <p>77. Tiene visiones con <b>ángeles</b> y <b>demonios</b>.</p>	
<p><b>[SP-N2]</b> [homónimos]</p>	<p>No se anotarán palabras homónimas (que se escriben y pronuncian igual) a especies cuando no quede claro que se refieran a un organismo.</p>

[ejemplos]	
78. Mancha color <b>naranja</b>	
[La palabra “naranja” puede ser un color o una fruta. Como en este caso se refiere al color, no se anota]	
79. <b>Tornillo</b>	
[Un “tornillo” puede referirse, entre otras cosas, a un implante médico o al nombre común de <i>Cedrelinga catenaeformis</i> . Si no está claro que se refiera al organismo, no se anota]	
<b>[SP-N3]</b> [toxinas]	<p>Las palabra “toxina” por sí sola puede referirse a varias bacterias y requiere excesiva inferencia por parte del anotador, por lo que no se anotarán. Sí se anota cuando va acompañada (por ejemplo: “toxina botulínica”)</p>
[ejemplos]	
80. No se investigaron <b>toxinas</b> ni <b>parásitos</b> .	

TABLA 7. REGLAS NEGATIVAS PARA HUMAN	
<b>[HM-N1]</b> [humanos]	<p>La palabra humano no se debe marcar cuando haga referencia a receptores, hormonas, ....</p>
[ejemplos]	
81. Factor de crecimiento epidérmico <b>humano</b>	
<b>[HM-N2]</b> [pronombres-adj tivos]	<p>No se deben anotar pronombres y adjetivos indefinidos que hagan referencia a personas (“algún”, “alguien”, “nadie”).</p>

[ejemplos]	
82. Aún no conoce a <b>nadie</b> en la ciudad.	
<b>[HM-N3]</b> [gentilicios]	No se deben anotar gentilicios ni referencias geográficas en general.
[ejemplos]	
83. <b>Paciente escocés</b> de 42 años...	
<b>[HM-N4]</b> [adj-estado-enfermedad]	No se deben anotar adjetivos que hagan referencia al estado de salud de una persona (incluyendo “embarazada”), a una enfermedad (“hipertenso”) o a un hábito (“fumador”).
[ejemplos]	
84. La <b>paciente</b> estaba <b>embarazada</b> pero era <b>fumadora</b> e <b>hipertensa</b>	
<b>[HM-N5]</b> [profesionales]	No se debe anotar la palabra “superior” en un contexto profesional.
[ejemplos]	
85. Sus <b>superiores</b> consideraron que el <b>paciente</b> no debía acudir a Urgencias.	

**TABLA 8. REGLAS NEGATIVAS PARA ENFERMEDAD**

<b>[EN-N1]</b> [estadios]	No se deben incluir dentro de la mención estadios o clasificaciones clínicas.
------------------------------	---

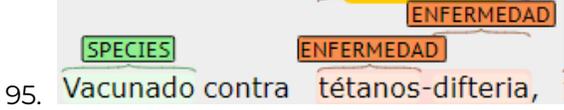
[ejemplos]	
86. <b>Infección por VIH</b> estadio C3	
<b>Infección por VIH estadio C3.</b>	
<b>[EN-N2]</b> [lateralidad]	No se deben incluir dentro de la mención modificadores relacionados con lateralidad: “izquierda”, “derecha”, “bilateral”, ...
[ejemplos]	
87. <b>Infección en seno esfenoidal</b> izquierdo	
<b>Infección en seno esfenoidal izquierdo</b>	
<b>[EN-N3]</b> [severidad]	No se deben incluir dentro de la mención modificadores relacionados con severidad: “fatal”, “severo”, “leve”, ...
[ejemplos]	
88. Lo trasladaron con el diagnóstico de <b>neumonía tuberculosa</b> severa	
Lo trasladaron con el diagnóstico de <b>neumonía tuberculosa severa</b>	
<b>[EN-N4]</b> [patología-onco]	No anotaremos patologías oncológicas de causa infecciosa.
[ejemplos]	
89. Por positividad a <b>HHV8</b> se estableció el diagnóstico de <b>Sarcoma de Kaposi</b> clásico.	
90. ... <b>linfomas</b> con la estimulación antigénica crónica por otras <b>infecciones</b> como <b>H. pylori, VEB o VHH8.</b>	

### 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. Estas reglas están divididas según el tipo de entidad a la que afectan. Se incluye una tabla extra “multiclase” para casos especiales que pueden etiquetarse con más de una etiqueta.

TABLA 9. REGLAS ESPECIALES PARA SPECIES	
<p><b>[SP-E1]</b> [menciones- sin-código]</p>	<p>Se anotan las siguientes menciones sin código específico en la NCBI Taxonomy: “cepas”, “colonias”, “virus hepatotropos”, “virus de la hepatitis”, “parásitos”, “bacilos”, “bacilos gram positivos”.</p> <p>La normalización de “virus hepatotropos” y “virus de la hepatitis” que no especifiquen el tipo de virus se hace de manera especial (más detalles en la sección 4).</p>
<p><b>[SP-E2]</b> [vectores]</p>	<p>Cuando la palabra “vector” se asocie a una especie, se anotarán ambas entidades separadamente. También anotaremos otras menciones de interés epidemiológico como: “huésped”, “huésped intermediario”, “vector”, “reservorio de patógenos”, “brotes” y sinónimos exactos.</p> <p>Este tipo de menciones debe normalizarse añadiendo el sufijo “ E” (más detalles en la sección 4).</p>
<p>[ejemplos]</p>	
<p>91. El <b><i>Plasmodium</i></b> se encontró en el mosquito <b>vector</b> (<b><i>Anopheles</i></b>).</p>	

<p><b>[SP-E3]</b> [drogas]</p>	<p>Se anotan las menciones de drogas de abuso que también sean plantas (marihuana, cannabis, tabaco), pero no las que precisen un proceso farmacológico para su producción (cocaína, heroína).</p>
<p>[ejemplos]</p> <p>92. Consumidor habitual de <b>cannabis</b>.</p>	
<p><b>[SP-E4]</b> [convenciones]</p>	<p>A modo de convención, se anotan las siguientes palabras: “patógenos”, “agentes patógenos”, “organismos”, “microorganismos”, “microbiota”, “microbioma”, “virología”, “bacteriología”.</p> <p>Del mismo modo, no se deben anotar: “microbiología”, “antibiótico”, “asepsia” y “antisepsia”, así como las palabras derivadas de estos términos.</p>
<p>[ejemplos]</p> <p>93. otros <u>agentes patógenos</u> tanto <u>bacterianos</u> como <u>fúngicos</u> y <u>parasitarios</u></p>	
<p><b>[SP-E5]</b> [vacunas]</p>	<p>Anotaremos la palabra “vacuna” y las formas del verbo “vacunar”, junto con el tipo de vacuna cuando conste.</p>
<p>[ejemplos]</p> <p>94. <u>Vacunas</u> sistemáticas al día, más <u>vacuna antineumocócica</u> (conjugada 13-valente).</p>	

	
<p><b>[SP-E6]</b> [pro/prebióticos]</p>	<p>Anotaremos la palabra probiótico (microorganismos vivos, predominantemente bacterias y hongos) y sus derivados., pero no la palabra prebiótico (componentes de la comida) y derivados.</p> <p>Este tipo de menciones se normaliza de manera especial (más detalles en la sección 4).</p>
<p>[ejemplos]</p> <p>96. <b>Diarrea</b> secundaria a antibióticos, por lo que se añadió un <b>probiótico</b>.</p>	
<p><b>[SP-E7]</b> [multi-resis-sensi]</p>	<p>Incluiremos en las menciones la palabra “multirresistente”, pero no la palabra “multisensible”, que consideramos como condición básica de un patógeno.</p> <p>También anotaremos: “estafilococo coagulasa negativo”, “estafilococo MARSÁ”.</p>
<p>[ejemplos]</p> <p>97. <b>Flora multirresistente</b></p>	
<p><b>[SP-E8]</b> [bacterias]</p>	<p>Con respecto a las <b>bacterias</b>, se debe anotar:</p> <p>a) Resistencias y sensibilidades antimicrobianas.</p> <p>(“Enterobacterias BLEE”, “SARM”, “morganella morganii fenotipo AmpC”, ...)</p>

	<p>b) Morfologías, tinciones específicas, identificación bioquímica, requerimientos especiales de crecimiento.</p> <p>(“bacilos gram positivo/negativo”, “BAAR”, “estafilococo coagulasa negativo”, “S. epidermidis multirresistente”, “bacilo gramnegativo pleomorfo”, “bacilos gram+ y gram- anaerobios”, “bacilos gram-negativos aerobios facultativos”, “bacterias de crecimiento lento”, “bacterias piógenas”, “cocobacilo gramnegativo”, “catalasa positivo”, “oxidasa positivo”, “cocos gram positivos en racimos”, “gram positivo filamentoso ramificado”, ...)</p> <p>c) Clasificaciones no taxonómicas</p> <p>(“flora vaginal”, “flora bacteriana normal de la cavidad oral”, “flora saprófita mixta”, “gérmenes de la flora oral y vaginal”, ...)</p> <p>d) Genotipos, serotipos y cepas</p>
<p><b>[SP-E9]</b> [virus]</p>	<p>Con respecto a los <b>virus</b>, se debe anotar:</p> <p>a) Sistemas afectados.</p> <p>(“virus cardiotropos, neurotrópicos”, ...)</p> <p>b) Genotipos, serotipos y cepas</p>
<p><b>[SP-E10]</b> [hongos]</p>	<p>Con respecto a los <b>hongos</b>, se debe anotar:</p> <p>a) Resistencias y sensibilidades.</p> <p>(“candida albicans sensible a azoles”, ...)</p> <p>b) Morfología y estructuras micóticas</p>

	<p>(“hifas”, “hifas septadas”, “conidios”, “conidióforos”, “esporas”, “microconidio”, “micelio”, “levaduriforme”, “conidiación simpodial”, “colonia fúngica negra”, ...)</p> <p>c) Otras clasificaciones no taxonómicas</p> <p>(“oportunista”, “verdadero”, “dematiáceo”, “patógeno”, “filamentoso”, “hifomicetos”, ...)</p>
<p><b>[SP-E11]</b> [parásitos]</p>	<p>Con respecto a los <b>parásitos</b>, se debe anotar:</p> <p>a) Formas parasitarias de los ciclos biológicos.</p> <p>(“huevos”, “espécimen* adulto”, “quiste”, “trofozoito”, “miracidio”, “esquistocito”, “larva”, “amastigote”, “trypomastigote”, “cercaria”, “esporocisto”, “redia”, “ninfa”, ...)</p> <p>* Solo se anota para parásitos y acompañado de algún calificador</p> <p>b) Estructuras relacionadas con parásitos*</p> <p>(“Escólex”, “proglótide”, ...)</p> <p>* Se excluye: “muda”, “vestigio”, “apéndice”</p> <p>c) Tipo/forma de parasitación</p> <p>(“parásitos intraeritrocitarios, en anillo”, ...)</p>
<p><b>[SP-E12]</b> [priones]</p>	<p>Con respecto a los <b>priones</b>, se considerarán tanto en su forma de agente infeccioso y de enfermedad.</p>
<p>[ejemplos]</p>	

98. La demencia estaba causada por **priones**; los signos de demencia acelerada apuntaban a **enfermedad de Creutzfeldt-Jakob**

**TABLA 10. REGLAS ESPECIALES PARA HUMAN**

<p><b>[HM-E1]</b> [familiar]</p>	<p>Debe anotarse la palabra “familiar” en todos los contextos. Igual que en la regla HM-P2, si aparece la rama de la familia debe incluirse en la mención.</p>
<p>[ejemplos]</p> <p>99. Presentaba antecedentes <b>familiares</b> de diabetes.</p> <p>100. Rama <b>familiar materna</b></p>	
<p><b>[HM-E2]</b> [contacto]</p>	<p>La palabra “contacto”, de gran importancia en el contexto de enfermedades infecciosas, se anotará cuando aparezca se refiera a una persona.</p> <p>También anotaremos otros conceptos de interés epidemiológico como: “reservorios”, “portadores” (de patógenos únicamente), “contactos”, “poblaciones”, “viajeros”, ...</p>
<p>[ejemplos]</p> <p>101. Su <b>compañera</b> fue el <b>contacto</b> positivo que la obligó al aislamiento voluntario</p> <p>102. Su <b>familia</b> confirmó <b>contacto</b> previo dos días antes del ingreso</p>	
<p><b>[HM-E3]</b> [compuestas]</p>	<p>Anotaremos las palabras compuestas que estén formadas por dos o más palabras que hagan referencia a humanos..</p>

[ejemplos]

103. Hospital **maternoinfantil**.
104. Antecedentes **sociofamiliares**

**TABLA 11. REGLAS ESPECIALES PARA ENFERMEDAD**

<p><b>[EN-E1]</b> [relacionados- patógenos]</p>	<p>A modo de convención, deben anotarse conceptos que indican que existe una infección, tales como: “pus”, “purulento”, “piuria”, “séptico”, “infeccioso”, ....</p>
<p>[ejemplos]</p> <p>105. Tosía esputo <b>purulento</b> y hemoptoico.</p>	
<p><b>[EN-E2]</b> [epidemiológicas]</p>	<p>A modo de convención, deben anotarse menciones epidemiológicas como: “contagio”, “brote”, “epidemia” y “pandemia”, ....</p> <p>Este tipo de menciones debe normalizarse añadiendo el sufijo “ E” (más detalles en la sección 4).</p>
<p>[ejemplos]</p> <p>106. En situación de <b>pandemia</b> para el <b>virus influenza A</b>.</p>	
<p><b>[EN-E3]</b> [VDRL]</p>	<p>Anotaremos la prueba “VDRL”, ya que muchas veces se usa para referirse a la sífilis.</p>

**TABLA 12. REGLAS MULTICLASE**

<p><b>[MC-E1]</b> [cadáver]</p>	<p>Las palabras “cadáver”, “donante” y “convaleciente” se anotan usando la etiqueta de SPECIES si hacen referencia a un animal y con la etiqueta HUMAN si hacen referencia a un humano.</p> <p>La normalización de la palabra “cadáver” se hace de manera especial (más detalles en la sección 4).</p>
<p>[ejemplos]</p> <p>107. Trasplante de <b>donante</b> vivo</p>	
<p><b>[MC-E2]</b> [COVID]</p>	<p>La enfermedad COVID-19 y su agente etiológico el SARS-CoV-2 merecen una regla propia ya que muchas veces los dos conceptos se usan de manera intercambiable.</p> <ul style="list-style-type: none"> <li>a) Se anota siempre como SPECIES la forma abreviada “SARS-CoV-2” y la expandida “Síndrome de dificultad respiratoria aguda por coronavirus 2” cuando en el informe clínico se refiera al agente etiológico.</li> <li>b) Se anota siempre como ENFERMEDAD la mención COVID-19, aunque en el informe clínico se esté utilizando para referirse al agente etiológico</li> <li>c) Se anotan como entidad única de enfermedad las siguientes menciones: “neumonía por COVID19”, “neumonía a causa de COVID19”, ...</li> </ul>
<p>[ejemplos]</p> <p>108. Prueba PCR para el <b>Síndrome de dificultad respiratoria aguda por coronavirus 2 (SARS-CoV-2)</b> fue negativa [SPECIES]</p>	

109. Diagnosticado de <b>long COVID</b> tras persistir la fatiga después de 6 meses [ENFERMEDAD]	
<p><b>[MC-E3]</b> [servicios]</p>	<p>De manera general, no se anotan los servicios clínicos.</p> <p>Excepciones: “Unidad de <b>Neonatos</b>”, “Cuidados Intensivos <b>Neonatales</b>”, “Unidad de <b>Enfermedades Infecciosas</b>”, “Unidad de <b>Infectología</b>”, “Clínica de <b>Enfermedades Transmisibles, Enfermedades Importadas y Venéreas</b>”, “Unidad de <b>ETS (Enfermedades de Transmisión Sexual)</b>”.</p>
[ejemplos]	
110. <b>Unidad de Neonatología</b>	

## 4. REGLAS DE NORMALIZACIÓN

La segunda parte de la anotación del corpus LivingNER consiste en la normalización de las entidades asignándoles códigos de la NCBI Taxonomy. Además de un código, algunas menciones llevan asociadas un modificador que nos ayuda a separar distintos tipos de información o situaciones especiales de la normalización. Estos modificadores se incluyen siempre después del código y están compuestos por una barra vertical (“|”) más una letra en mayúsculas.

En los ejemplos de esta sección, los códigos usados y propuestos van precedidos del prefijo “NCBI:” y acompañados del descriptor del concepto entre paréntesis para facilitar su lectura.

Las reglas de normalización se distribuyen de la siguiente manera:

- 4.1 Reglas básicas
- 4.2 Reglas específicas y convenciones

## 4.1 Reglas básicas

En este apartado se enumeran las reglas básicas a seguir para hacer la normalización, además de los modificadores disponibles.

TABLA 13. REGLAS BÁSICAS DE NORMALIZACIÓN	
<b>[NORM-B1]</b> [modificadores]	Algunos códigos deben ir acompañados de un modificador. La lista completa de modificadores se incluye en la Tabla 14.
<b>[NORM-B2]</b> [multi-código]	Las menciones que contienen dentro múltiples conceptos se deben normalizar usando más de un código. Cada código debe separarse con una barra vertical (" ").
[ejemplos]	
<p>111. Flora de la cavidad oral del gato [NCBI:1227552 9685 (flora cavidad oral gato)]</p> <p>112. Parainfluenza 1-4 [NCBI:12730 1979160 11216 1979161]</p> <p>113. Echinococcus granulosus y multilocularis [NCBI:6210 6211]</p>	
<b>[NORM-B3]</b> [cod-específico]	Se debe utilizar el código más específico posible siempre que se pueda.
[ejemplos]	
<p>114. Fusobacterium nucleatum [NCBI:851]</p>	
<b>[NORM-B4]</b> [cod-padre]	En el caso de que no exista un código específico para la mención, debemos utilizar el código padre (codigo más general que está por encima en la jerarquía taxonómica) más cercano que podamos. Estos códigos están marcados por el modificador "IH".

[ejemplos]	
115. Mijo [NCBI:3193 ( <i>Embryophyta</i> , plantas de tierra)]	
116. Estreptococo no hemolítico [NCBI:1301 ( <i>Streptococcus</i> )]	
<b>[NORM-B5]</b> [sin-código]	Algunas menciones puede que no tengan código, bien por falta de cobertura de la taxonomía, bien porque la mención está fuera del scope de ella, o bien (en el caso de las enfermedades) porque no podemos asociar el código de ningún patógeno específico. Estos casos se marcan con el código NOCODE.
[ejemplos]	
117. Dermatitis abscesificada [NOCODE ya que no sabemos qué patógeno la causa]	
<b>[NORM-B6]</b> [convenciones]	Algunos conceptos deben normalizarse usando un código acordado durante el proceso de normalización (convenciones). Estos códigos se enumeran en la tabla 14.
<b>[NORM-B7]</b> [inespecíficos]	Algunos conceptos muy inespecíficos también deben normalizarse usando unos códigos concretos. Estos códigos se enumeran en la tabla 16.

**TABLA 14. MODIFICADORES DE CÓDIGOS**

<b> E</b> [epidemiológico]	Se debe utilizar para menciones de relevancia epidemiológica: reservorios, portadores, contactos, brotes, poblaciones, epidemias, pandemias, viajeros, vectores, ...
-------------------------------	--

[ejemplos]	
118. el <b>hermano</b> fue el <b>contacto</b> de Covid 119. ni tenido <b>contacto sexual</b> de riesgo en los últimos 12 120. el <b>portador nasal</b> de <b>MARSA</b>	
<b> F</b> [figurativo]	Se debe utilizar para menciones de especies usadas de forma figurativa (ver regla SP-P6).
[ejemplos]	
121. eritema en alas de <b>mariposa</b>	
<b> H</b> [hijo]	Se debe utilizar cuando se asigna un código más general que la mención y usamos un código padre.
[ejemplos]	
122. <b>Estafilococo meticilin resistente</b>	
<b>  </b> [inferido]	Hay determinadas pruebas de laboratorio que siempre están relacionadas con un patógeno concreto. En estos casos, podremos inferir el patógeno gracias a nuestro conocimiento del dominio.
[ejemplos]	
Inferimos <i>Treponema pallidum</i> de la mención <b>VDRL</b> [VDRL se anotará y etiquetará como <i>enfermedad</i> (VD significa Venereal Disease) y cómo <i>especie</i> (es una prueba muy común sinónimo de Sífilis)]	

<p><b>IN</b> [nosocomial]</p>	<p>Se debe utilizar para menciones de enfermedades y patógenos adquiridos en la estancia hospitalaria.</p>
<p>[ejemplos]</p> <p>123. persistencia de <b>infección</b> por <b><i>Clostridium difficile</i></b></p>	
<p><b>IO</b> [ocupación]</p>	<p>Se debe utilizar para menciones de humanos que sean ocupaciones o profesiones.</p>
<p>[ejemplos]</p> <p>124. <b>Jefe de una empresa de construcción</b></p>	

## 4.2 Reglas específicas y convenciones

En esta sección se explican algunas de las reglas y convenciones acordadas durante el proceso de anotación para determinados conceptos concretos y algunos muy inespecíficos.

TABLA 15. CONVENCIONES DE NORMALIZACIÓN	
<p><b>[ambiguos]</b></p>	<p>Para especies de microorganismos abreviadas que puedan ser ambiguas y para las que no dispongamos de suficiente contexto usaremos el código de la versión larga más frecuente.</p> <p>Por ejemplo, para el concepto “C. albicans” se asumirá que se refiere a la <i>Candida albicans</i> y se normalizará con el código NCBI: 5476.</p>

<b>[cadáver]</b>	Para el concepto “cadáver” deberemos usar el código del animal o especie correspondiente.
<b>[enfermedades]</b>	De manera general, las enfermedades causadas claramente por un patógeno específico se normalizarán usando el código de dicho patógeno. Si no podemos asociar la enfermedad a ningún patógeno, usaremos NOCODE.
<b>[esporas]</b>	Para las esporas usaremos el código NCBI: 4751 ( <i>fungi</i> ).
<b>[hepatotropos y hepatitis]</b>	Para los virus hepatotropos (primarios) y virus de la hepatitis sin especificar se usarán los códigos NCBI: 12092 10407 11103 12475 12461.
<b>[herpes]</b>	Para el herpes se usará el código NCBI:10294 (simplexvirus).
<b>[humanos]</b>	Para todos los humanos se usará el código NCBI:9606 ( <i>Homo Sapiens</i> ). Para las menciones que también sean ocupaciones o profesiones debemos recordar añadir el modificador  O.
<b>[microbiotas y microbiomas]</b>	Para las microbiotas y microbiomas en general se usará el código NCBI:646099 (human metagenome). Para las microbiotas y microbiomas intestinales, sin embargo, se usará el código NCBI:749906 (gut metagenome).
<b>[prion]</b>	Para los priones se usará el código NCBI: 36469
<b>[probióticos]</b>	Las menciones de probióticos y derivados se normalizan con dos código: NCBI: 2 4751 ( <i>bacteria / fungi</i> ).

<b>[tuberculosis]</b>	Para la tuberculosis se usará el código NCBI:77643 ( <i>Mycobacterium tuberculosis complex</i> ).
<b>[vacunas]</b>	<p>Las vacunas deben normalizarse solo cuando se indique qué tipo de vacuna es. Si no se indica, se marcará como NOCODE. En el caso de que se indique, usaremos el código de la especie con la que este relacionada.</p> <p>Por ejemplo, para “vacuna antineumocócica” usaremos el código NCBI: 1313 (<i>Streptococcus pneumoniae</i>).</p>

**TABLA 16. NORMALIZACIÓN DE CONCEPTOS INESPECÍFICOS**

<b>[aislamientos, colonias, cepas]</b>	Aislamientos microbiológicos, colonias y cepas sin especificar se normalizan con los códigos NCBI: 2 4751 ( <i>bacteria / fungi</i> ).
<b>[enfermedades causadas por bacterias y virus]</b>	Para las enfermedades que estén causadas principalmente por bacterias y virus usaremos la combinación de códigos NCBI: 2 10239 ( <i>bacteria, virus</i> ).
<b>[enfermedades causadas por bacterias, virus u hongos]</b>	Para las enfermedades que puedan estar causadas tanto por bacterias, virus y hongos usaremos la combinación de códigos NCBI: 2 4751 10239 ( <i>bacteria, fungi, virus</i> ).
<b>[enfermedades causadas por bacterias, virus, hongos o parásitos]</b>	Para agentes infecciosos, anti infecciosos, patógenos, microorganismos, oportunistas, especies, germen, enfermedades infecciosas, ETS, sepsis, shock séptico, enfermedades infecciosas (como prostatitis) causadas por bacterias, hongos, parásitos o virus

	usaremos la combinación de códigos NCBI: 2 2759 10239 H ( <i>bacteria, eukariota, virus</i> ).
<b>[gusanos]</b>	La palabra “gusano” (sin especificar) deben normalizarse usando tres códigos juntos: 6157 10232 6231 (platelmintos / acantocéfalos / nematodos).
<b>[huéspedes, vectores, reservorios]</b>	Conceptos como “huésped”, “huésped intermediario”, “vector”, “reservorio de patógenos” y sinónimos exactos se normalizan con el código NCBI: 33208 ( <i>animalia</i> ).
<b>[huevos]</b>	Los huevos de animales ( <u>no</u> de parásitos) se normalizan con el código NCBI: 33208 H ( <i>animalia</i> ).
<b>[microorganismo inespecífico]</b>	Cuando no se indique la especie de un microorganismo utilizaremos el código NCBI de su genus. Por ejemplo, para HPV usaremos el código NCBI: 10566 ( <i>Human papillomavirus</i> ).
<b>[parásitos y relacionados]</b>	Conceptos como “parásito”, “antiparasitario”, estadíos parasitarios inespecíficos (“larvas”, “huevos”), elementos flagelados y otros términos relacionados se normalizan con el código NCBI:2759 H ( <i>eukariota</i> ).
<b>[pez y pescado]</b>	Las menciones de peces y pescados inespecíficos se normalizan usando el código NCBI: 7898 ( <i>Actinopterygii</i> ).
<b>[reino vegetal]</b>	Las menciones de vegetales, hierbas, flores, plantas y árboles inespecíficas se normalizan con el código NCBI: 33090 ( <i>Viridiplantae</i> ).

## 5. REGLAS DE CLASIFICACIÓN

La última fase del corpus LivingNER consisten en la clasificación de las menciones de SPECIES en 4 categorías: alimentación, animales de compañía, animales causantes de lesión o nada.

Las reglas de clasificación se distribuyen de la siguiente manera:

- 5.1 REGLAS GENERALES DE CLASIFICACIÓN de alimentación
- 5.2 REGLAS GENERALES DE CLASIFICACIÓN de animales de compañía
- 5.3 REGLAS GENERALES DE CLASIFICACIÓN de animales causantes de lesión

TABLA 17. REGLAS GENERALES DE CLASIFICACIÓN DE ALIMENTACIÓN	
<p><b>[F1]</b> [alimentación- definición]</p>	<p>Se marcan con la etiqueta FOOD todos los alimentos etiquetados como SPECIES.</p>
<p>[ejemplos]</p>	
<p>125. Pocas semanas después estando manipulando un <b>melocotón</b> presentó un cuadro clínico similar.</p> <p>126. Durante la anamnesis refiere contacto con un <b>cordero</b> recién nacido dos semanas antes</p>	
<p><b>[F2]</b> [derivados- comida]</p>	<p>Se marcan con la etiqueta FOOD las SPECIES que produzcan derivados alimenticios mencionados en el texto.</p>
<p>[ejemplos]</p>	

<p>127. Su trabajo consiste no sólo en cuidar y ordeñar las cabras sino en la elaboración de queso de <b>cabra</b>.</p>	
<p><b>[F3]</b> [categorías- generales]</p>	<p>Se marcan como FOOD plantas que produzcan derivados comestibles</p>
<p>[ejemplos]</p>	
<p>128. Alergia a ácaros, polvo y <b>olivo</b></p>	

<b>TABLA 18. REGLAS GENERALES DE CLASIFICACIÓN DE ANIMALES DE COMPAÑÍA</b>	
<p><b>[A1]</b> [animales- compañía- definición]</p>	<p>Se marcan con la etiqueta PET las menciones de SPECIES con las que una persona (paciente o no) conviva o tenga contacto habitual (incluyendo mascotas, ganado, etc.)</p>
<p>[ejemplos]</p>	
<p>129. Convive con un <b>perro</b>.</p>	
<p>130. Varón de 53 años con antecedente de hipertensión arterial, es cabrero de profesión; Su trabajo consiste no sólo en cuidar y ordeñar las <b>cabras</b> sino en la elaboración de queso de <b>cabra</b>.</p>	
<p><b>[A2]</b> [animales- contacto- ocasional]</p>	<p>No se marcan con la etiqueta PET las SPECIES con las que el contacto sea puntual, inexistente o indeterminado.</p>
<p>[ejemplos]</p>	

131. Habiendo realizado de este modo el Camino de Santiago recientemente, teniendo contacto con algún animal, fundamentalmente **perros, cabras** y **vacas**, aunque sin contacto estrecho.

**[A3]**  
[compañía-vs-  
parásitos]

No se marcan con la etiqueta PET los parásitos que estén viviendo en el huésped.

[ejemplos]

132. Se encontró una **larva de Anisakis** y se la extrajo por endoscopia.

### TABLA 19. REGLAS GENERALES DE CLASIFICACIÓN DE ANIMALES CAUSANTES DE LESIÓN

**[AT1]**  
[animales-lesión-  
definición]

Se marcan con la etiqueta ATTACK las menciones de SPECIES en que se confirma que han causado algún daño al *paciente o persona mencionada en el caso clínico de forma directa* (picaduras, mordeduras, ...).

[ejemplos]

133. Previamente había recibido varias picaduras de **abejas** y de **avispas**, presentando sólo reacción local (...) Presentamos el caso de un niño de 6 años de edad, con antecedentes familiares de padre apicultor con hipersensibilidad a veneno de **Apis**

**Mellifera**

[en este segundo caso, no se marca Apis Mellifera como ATTACK, porque no se describe que haya atacado al niño, tan solo se da una definición]

134. Niño de 12 años de edad que recibe picadura de **himenóptero** (...) Se diagnostica de reacción sistémica grado III (grave) a veneno de **avispa común**

[se marca avispa común porque de la frase se sabe que el himenóptero atacante fue una avispa común]

135. Tungiasis causada por la **pulga hematófaga Tunga penetrans**
136. (...) *pequeña incisión en el orificio de entrada a la extracción natural de la larva, viva, siendo identificada como **Dermatoba hominis**.*
137. No recordaba picaduras de **insecto** en la zona.  
[No se marca insecto porque la frase es negativa]
138. La paciente negaba viajes al extranjero, picaduras de **insecto** ni haber bebido agua no potable.  
[No se marca insecto porque la frase es negativa]

**[AT2]**  
[sospecha]

Se marcan con la etiqueta ATTACK las menciones de SPECIES en que se sospecha que han causado algún daño al paciente o persona mencionada en el caso clínico *de forma directa*.

[ejemplos]

139. Picadura de **arácnido**, probablemente por **araña reclusa parda** con síntomas sistémicos asociados.
140. Juicio clínico de posible picadura de **artrópodo**

**[AT3]**  
[laboratorio-1]

Se incluyen pruebas de laboratorio que diagnostiquen al causante del ataque.

[ejemplos]

141. Identificándose como **Trips**, pertenecientes al **género Aelotrips Halliday (Aelothripidae, Thysanoptera)**

<p><b>[AT4]</b> [genéricos]</p>	<p>Se incluye la palabra genérica “animal” y “animales” cuando sea evidente que es la causa del ataque .</p>
<p>[ejemplos]</p> <p>142. <b>Larvas de L.obliqua</b> a los ejemplares del <b>animal</b> causante</p>	
<p><b>[AT5]</b> [diferencial]</p>	<p>Se excluyen alergias, parásitos y menciones parte de un diagnóstico diferencial o académico.</p>
<p>[ejemplos]</p> <p>143. Se realizaron prick tests para <b>almeja, mejillón, gamba, calamar, ostra, pulpo</b> y <b>caracol</b></p>	
<p><b>[AT6]</b> [contacto-consumo]</p>	<p>Se excluyen casos en que la persona tenga enfermedad por contacto con animales o por ingerir productos contaminados de animales (por ej, brucelosis por ingerir queso de cabra no pasteurizado, presencia de tinea corporis transmitida por gatos).</p>
<p>[ejemplos]</p> <p>144. Tras el diagnóstico de <b>tinea corporis</b> se recomendó también tratar a sus <b>gatos</b></p>	
<p><b>[AT7]</b> [laboratorio-2]</p>	<p>Se excluyen pruebas de laboratorio no destinadas a la identificación de especies causantes de ataque</p>
<p>[ejemplos]</p> <p>145. IgE específica: veneno de <b>aveja (Apis mellifera)</b>, veneno <b>avispa (Vespula spp.)</b>, veneno <b>abejorro (Vespa cabro)</b>.</p> <p>146. Determinación IgE específica, confirmándose de nuevo sensibilización frente</p>	

a diferentes **pescados blancos y azules**

## Bibliografía

- Martin Gerner, Goran Nenadic, and Casey M. Bergman. 2010. Linnaeus: A species name identification system for biomedical literature. BMC Bioinformatics.
- Evangelos Pafilis, Sune P. Frankild, Lucia Fanini, Sarah Faulwetter, Christina Pavloudi, Aikaterini Vasileiadou, Christos Arvanitidis, and Lars Juhl Jensen. 2013. The species and organisms resources for fast and accurate identification of taxonomic names in text. PLOS ONE, 8(6):1–6
- Farré, Eulàlia, González, Gloria, Mas, Toni, Miranda-Escalada, Antonio, & Krallinger, Martin. (2020). Cantemist guidelines: neoplasms morphology annotation and mapping to CIEO-3 (1.3). Zenodo. <https://doi.org/10.5281/zenodo.4121183>
- Eulàlia Farré-Maduell, Salvador Lima-López, Antonio Miranda-Escalada, Vicent Briva-Iglesias, & Martin Krallinger. (2021). MEDDOPROF guidelines. Zenodo. <https://doi.org/10.5281/zenodo.4720833>
- Marimon, Montserrat, Gonzalez-Agirre, Aitor, Intxaurrenondo, Ander, Rodríguez, Heidi, Lopez Martin, Jose Antonio, Villegas, Marta, & Krallinger, Martin. (2020). MEDDOCAN guidelines: Medical Document Anonymization (1.0). Zenodo. <https://doi.org/10.5281/zenodo.4279338>
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 102–107.
- Martin Krallinger, Miguel Vazquez, Florian Leitner, David Salgado, Andrew Chatr-Aryamontri, Andrew Winter, Livia Perfetto, Leonardo Briganti, Luana Licata, Marta Iannuccelli, et al. 2011. The protein-protein interaction tasks of biocreative iii: classification/ranking of articles and linking bio-ontology concepts to full text. BMC bioinformatics, 12(8):1–31.