



Guías MEDDOPLACE: Anotación, normalización y clasificación de lugares e información relacionada en textos clínicos

V1 [Marzo 2023]

AUTORES

Salvador Lima López (Barcelona Supercomputing Center)
Eulàlia Farré-Maduell (Barcelona Supercomputing Center)
Vicent Briva-Iglesias (Dublin City University)
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

1. INTRODUCCIÓN	3
Motivación del corpus y casos de uso	3
Trabajo previo en anotación y detección de lugares	4
El corpus y las guías MEDDOPLACE	6
2. MARCADO DE ENTIDADES	9
2.1 DEFINICIÓN	9
2.2 REGLAS DE MARCADO Y EXTENSIÓN	12
B1. Reglas de marcado generales positivas	12
B2. Reglas de marcado generales negativas	14
C1. Reglas para “GPE”	16
C2. Reglas para “GEO”	19
C3. Reglas para “FAC”	22
C4. Reglas para “COMUNIDAD”	25
C5. Reglas para “IDIOMA”	26
C6. Reglas para “DEPARTAMENTO”	27
C7. Reglas para “TRANSPORTE”	31
3. CARACTERIZACIÓN DE ENTIDADES	35
3.1 Caracterización clínica	35
3.2 Clasificación por referentes	38
3.3 Polaridad y temporalidad	40
4. NORMALIZACIÓN	42
4.1 Normalización con GeoNames	43
4.2 Normalización con PlusCodes	46
4.3 Normalización con SNOMED-CT	48
ANEXO 1. Tabla Resumen de Reglas de Anotación	57
BIBLIOGRAFÍA	66

1. INTRODUCCIÓN

Estas guías describen la anotación del corpus MEDDOPLACE, un conjunto de 1000 casos clínicos en castellano pensado para el entrenamiento y evaluación de sistemas automáticos para la detección, caracterización y normalización de localizaciones e información relacionada (comunidades, idiomas, departamentos clínicos y menciones a movimientos del paciente) en textos médicos.

Motivación del corpus y casos de uso

En la práctica clínica, conocer los lugares dónde el paciente ha estado es extremadamente relevante. Por ejemplo, un episodio de fiebre se investiga de forma distinta si durante el mes anterior hemos estado en una zona endémica de malaria, mientras que la probabilidad de tuberculosis en un cuadro de tos crónica es mucho más alta en zonas con gran prevalencia de infecciones por *Mycobacterium tuberculosis*. En el contexto de la pandemia de la COVID-19, los movimientos y estancias de los pacientes han tomado aún más relevancia. Pese a esto, no existen apenas recursos que permitan categorizar los lugares extraídos de los textos según su relevancia clínica.

Con esto en mente, la creación de MEDDOPLACE ha estado motivada por los potenciales casos de uso de las localizaciones geográficas en los casos e historias clínicas.

Uno de los principales casos de uso es la **asociación entre un lugar geográfico y factores de riesgo para la salud**. Estos factores de riesgo pueden ser: enfermedades transmisibles (malaria, enfermedad del sueño, enfermedad de Chagas, Dengue, tuberculosis, etc.), enfermedades no transmisibles (enfermedades genéticas, obesidad, enfermedad cardiovascular) y enfermedades de salud mental (trauma de violencia y de guerra, emigración traumática, tráfico humano, etc.). A nivel individual, será esencial diagnosticar al paciente de malaria para administrarle artesunato, por ejemplo. A nivel de salud pública, es esencial para detectar **zonas de enfermedades emergentes** y de **brotes epidémicos**, por ejemplo, la infección por *Salmonella enteritidis* en personas que han comido en un restaurante, o un brote de virus Marburg en Guinea Ecuatorial.

En el caso de la detección de departamentos, servicios y equipos médicos, otro de los casos de uso principales es el **trazado de la ruta del paciente** a través del texto no estructurado. Por ejemplo, un paciente ingresa a través del *Servicio de Urgencias* remitido por *Atención Primaria*, es trasladado al *Servicio de Cardiología*, desarrolla una infección nosocomial, requiere ingreso en la *UCI*, pasa después a *Planta de Medicina Interna*, y se va de alta. El análisis de estas rutas son útiles para el control de calidad y de iatrogenia de la asistencia médica. Las localizaciones nos advierten también sobre la gravedad del paciente.

Trabajo previo en anotación y detección de lugares

La detección de lugares se podría considerar una de las tareas clásicas del procesamiento de lenguaje natural. En las Message Understanding Conferences (MUC) (Hirschman, 1998) de los años 90, una de las pioneras en reconocimiento de entidades, ya se definía como uno de los tres tipos de entidades clave a reconocer. En estas conferencias se definieron tres categorías: PERSON, ORGANIZATION y LOCATION. La categoría LOCATION tiene un alcance muy amplio, incluyéndose dentro de ella nombres propios y comunes de todo tipo de lugares: ciudades, países, ríos, montañas, aeropuertos, monumentos, ... También se suelen considerar dos categorías extra: fechas y números.

Este esquema de anotación se mantuvo en los datasets de las conferencias CoNLL 2002 (Sang y Tjong, 2002) y 2003 (Sang y De Meulder, 2003), en las que se añadió una categoría extra llamada MISC, que incluye información como gentilicios o ciertos tipos de eventos. En el año 2002 se presentó un dataset en castellano y otro en neerlandés, y en el año 2003 uno en inglés y otro en alemán. Todos estos datasets están compuestos por textos de noticias. Debido al impacto de la conferencia y de los datos anotados, estas cuatro categorías acabaron convirtiéndose prácticamente en un estándar en el procesamiento del lenguaje. Muchos corpus posteriores, como los corpus de noticias en castellano y catalán Ancora (Taulé, Martí y Recasens, 2008), han usado el mismo esquema para la anotación general de entidades. En la actualidad, algunas herramientas populares de NLP tienen modelos pre-entrenados disponibles generados usando los corpus de CoNLL. Por ejemplo, la herramienta StanfordNER¹ usa los datos de la edición del 2003 para su versión en inglés y spaCy² usa los del 2002 para sus modelos NER en castellano.

El proyecto ACE (Automatic Concept Extraction; Doddington et al., 2004), de la Universidad de Pensilvania, introdujo una clasificación más granular para el

¹ <https://nlp.stanford.edu/software/CRF-NER.shtml>

² <https://spacy.io/>

reconocimiento de entidades en un corpus de noticias en el que se incluían nuevas clases y subclases de lugares. El proyecto comenzó en el año 2002 y creció progresivamente hasta el año 2008, creando datos y guías de anotación para el inglés, castellano, árabe y chino. En ACE, los lugares se dividen de manera más granular en tres categorías: entidades geopolíticas (“geo-political entities” o GPE; países, ciudades, ...), instalaciones (“facilities” o FAC; edificios, aeropuertos, caminos, ...) y entidades geográficas en general (“locations” o LOC; accidentes geográficos, cuerpos celestiales, algunas frases locativas, ...). Es importante señalar que dentro de las GPE hay distintas subclases como GPE-PER para los gentilicios o GPE-ORG para los usos referidos a gobiernos.

Aunque probablemente las propuestas de MUC, CoNLL y ACE han sido las más exitosas e influyentes, existen más que merece la pena comentar. Por ejemplo, el BBN Pronoun Coreference and Entity Type Corpus (Weischedel y Brunstein, 2005), que usa textos del Wall Street Journal, incluye 12 tipos de entidades nombradas (Person, Facility, Organization, GPE, Location, Nationality, Product, Event, Work of Art, Law, Language y Contact-Info). A nivel de detección de lugares, el BBN mantiene los principales tipos de entidades geográficas y separa las nacionalidades o gentilicios en etiquetas independientes. También incluye anotaciones para “entidades nominales” (separadas de las que se mencionan usando nombres propios) y numéricas. Todas estas entidades tienen a su vez subtipos, llegando a un total de 64 etiquetas posibles. Este corpus es el que usan las herramientas spaCy y Stanza para algunos de sus modelos NER en inglés y la base de recursos más grandes como OntoNotes (Weischedel et al., 2017).

Existen muchos más corpus en múltiples idiomas que pueden usar para entrenar sistemas de detección de lugares, aunque, de acuerdo con nuestra revisión bibliográfica, ninguno de ellos tiene un enfoque clínico. Algunos de estos corpus son el corpus colaborativo en inglés GUM (George University Multilayer; Zeldes, 2017), el corpus en alemán Nosta-D (Dipper, Lüdeling y Reznicek, 2013), el corpus en noruego NorNE (Jørgensen et al., 2020), el corpus en checo CNEC (Ševčíková, Žabokrtský y Krůza, 2007) o el corpus en finlandés FiNER (Ruokolainen et al., 2020).

Por otro lado, para el aprovechamiento de los lugares encontrados en el texto por otros sistemas la detección no es suficiente. Es necesario un paso posterior de normalización de las menciones encontradas, bien a una serie de coordenadas o a un identificador de alguna ontología, para poder situarlos en un mapa. Ese es el objetivo de la tarea de resolución de topónimos (también llamado *geoparsing* o *geocoding*). Esta tarea suele formar parte de sistemas o métodos de extracción de información geográfica (en inglés *Geographic Information Retrieval* o GIR).

Algunos recursos que destacan en este ámbito son las múltiples tareas de GeoCLEF entre el año 2005 (Gey et al., 2006) y 2008 (Mandl et al., 2009), con el objetivo de evaluar sistemas GIR en textos de noticias en distintos idiomas, o la tarea 12 de SemEval 2019 (Weissenbacher et al., 2019), en la que se se pedía a los participantes reconocer topónimos en artículos científicos de PubMed en inglés y normalizarlos a GeoNames.

Más allá de tareas de evaluación, existen corpus publicados en abierto como la colección de noticias normalizadas con GeoNames Local Global Lexicon (Lieberman et al., 2010), el corpus de noticias sobre enfermedades anotadas con WikiNews GeoVirus (Gritta et al., 2018) o el corpus GeoWebNews (Gritta et al., 2020) que incluye una anotación granular para distintos tipos de topónimos y que normaliza no solo entidades geopolíticas sino también algunas construcciones y nombres de calles. Recomendamos la lectura de Gritta et al. (2020) para una visión general de la resolución general de topónimos y los recursos existentes.

El corpus y las guías MEDDOPLACE

El objetivo del corpus es la detección, normalización y caracterización de todos los lugares que se mencionan en el texto, así como de cuatro tipos de información relacionados: departamentos clínicos, menciones a movimientos del paciente, comunidades e idiomas. A modo de resumen, la Figura 1 muestra las diferentes opciones posibles en brat (Stenetorp et al., 2012), la herramienta usada para hacer la anotación del corpus.

Para la anotación textual contamos con un total de 10 etiquetas. Estas categorías de anotación textual se explican brevemente a continuación, con más detalles, ejemplos y reglas para cada una en el apartado 2.

Por un lado, para lugares y ubicaciones definimos un total de seis etiquetas granulares: GPE_NOM, GPE_GEN, GEO_NOM, GEO_GEN, FAC_NOM y FAC_GEN. Esta separación está inspirada en dos recursos explicados en la sección anterior, ACE (Doddington et al., 2004) y el BBN corpus (Weischedel et al., 2017). Del primero tomamos la división granular de lugares en tres categorías: entidades geopolíticas (GPE), instalaciones, construcciones y edificios (FAC), y elementos geográficos naturales y otros tipos de lugares (LOC, que aquí renombramos a GEO). Del segundo tomamos la división entre nombres comunes (GEN) y propios (NOM), que nos permite ser exhaustivos en la anotación y separar en dos ejes la anotación: el tipo de sitio y el tipo de nombre.

Por otro lado, debido al enfoque clínico del corpus, tras inspeccionar algunos documentos decidimos añadir a la anotación cuatro etiquetas complementarias a los lugares:

- DEPARTAMENTO recoge todas las menciones a sitios específicos dentro de hospitales (*quirófano*), así como de servicios (*urgencias*, *cardiología*), unidades (*UCI*), equipos y otros elementos clínicos que representan el viaje que hace el paciente por el sistema sanitario. Es una etiqueta muy frecuente y variada, además de muy específica y relevante para el dominio clínico.
- TRANSPORTE recoge menciones relacionadas con movimientos del paciente (y otras personas que aparezcan en el texto), tanto fuera como dentro del ámbito sanitario. Por ejemplo, viaje, transporte en ambulancia, volar en avión, desplazarse en coche, etc.
- COMUNIDAD recoge información sociodemográfica que muchas veces (pero no siempre) está relacionada con el lugar de origen o residencia de la persona, incluyendo gentilicios, religiones y etnias.
- IDIOMA recoge menciones de las lenguas que se mencionan en el texto, así como a problemas lingüísticos (barreras idiomáticas).

Las entidades anotadas en el corpus se caracterizan en un paso de clasificación para maximizar su aprovechamiento desde el punto de vista clínico. Todas las entidades se clasifican según su polaridad y temporalidad, mientras que a las ubicaciones, además, se les añaden etiquetas que reflejan su posible relevancia clínica como si se refieren a la residencia del paciente o a un sitio donde han recibido atención médica. La clasificación también tiene unas reglas propias, que se describen en el apartado 3.

La última capa de anotación es la normalización de las entidades anotadas. Debido a la diversidad de contenido, utilizamos tres fuentes distintas para la normalización: GeoNames, PlusCodes y SNOMED. Cada una de estas fuentes tiene una sección propia en el apartado 4 de las guías, con explicaciones sobre su funcionamiento, reglas y ejemplos.

En definitiva, esta guía se estructura como sigue: el apartado 2 explica las etiquetas y reglas para hacer el etiquetado textual del corpus; el apartado 3 recoge cómo hacer la clasificación de las entidades anotadas; a continuación, el apartado 4 describe el proceso, ontologías y reglas para la normalización de las entidades. En el Anexo I se incluye una tabla resumen de las reglas de anotación para su revisión rápida durante el proceso de anotación.

Con respecto al código de colores de esta guía, en los ejemplos que ilustran las distintas reglas se marca en **azul** lo que es correcto a la hora de anotar y en **rojo** lo que no. Las menciones en **negro** son ejemplos que deberían anotarse, pero que no tienen relación con la categoría o regla que se está describiendo.

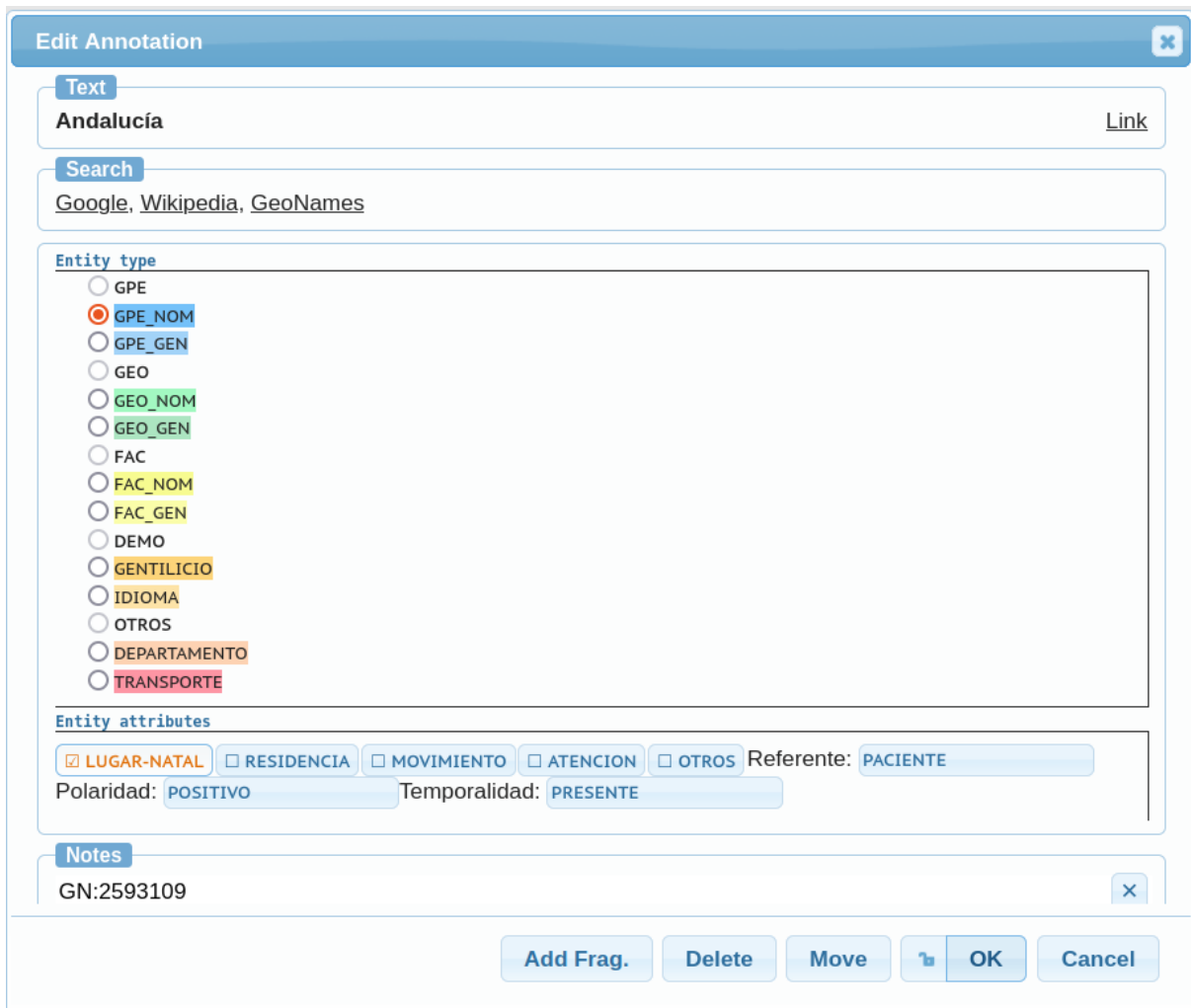


Figura 1. Ventana de anotación de brat con todas las categorías de anotación y clasificación. La normalización se presenta como comentario de cada anotación. Los archivos de configuración de brat se distribuyen junto al corpus.

2. MARCADO DE ENTIDADES

Esta sección describe las entidades que se deben marcar en el texto, así como las reglas que se deben seguir para su etiquetado. En el Anexo I, al final de estas guías, se incluye un resumen general de esta sección.

2.1 DEFINICIÓN

La siguiente tabla describe las distintas etiquetas junto a ejemplos para cada una:

A. DEFINICIÓN DE ETIQUETAS POSIBLES	
ETIQUETAS PARA LUGARES	
[GPE]	<p>Se usa para las llamadas “entidades geopolíticas” (<i>GeoPolitical Entities</i>), es decir, regiones geográficas definidas de manera política o social. Esto incluye continentes, países, ciudades, pueblos, islas y sitios poblados por humanos en general..</p> <p>Distinguimos dos subtipos:</p> <ol style="list-style-type: none"> 1. GPE-NOM³: menciones que hacen referencia a un sitio concreto con nombre propio (<i>Europa</i>). 2. GPE-GEN: menciones que hacen referencia a un sitio genérico usando un sustantivo común (<i>continente</i>).
<p>[ejemplos]</p> <ol style="list-style-type: none"> 1. (GPE-NOM) Procedente de la región nororiental del Perú. 2. (GPE-GEN) El omeprazol es uno de los fármacos más vendidos en los países europeos. 	

³ Al principio, en la guía distinguíamos entre GPE-NOM con sentido locativo y no locativo (es decir, Barcelona como ciudad frente a Barcelona como equipo de fútbol). Aunque definitivamente es relevante, decidimos eliminar esta distinción porque apenas encontramos menciones no locativas en el corpus.

[GEO]

Se usa para **entidades geográficas naturales**, en contraste con las entidades geopolíticas y administrativas (GPE). Se incluyen: accidentes geográficos en general (ríos, mares, montañas, selvas, ...), hábitats (zonas cálidas, frías, rurales, zonas de enfermedades endémicas...), ...

Distinguimos dos subtipos:

1. **GEO-NOM**: menciones que hacen referencia a un sitio concreto con nombre propio (*Río Guadalquivir*).
2. **GEO-GEN**: menciones que hacen referencia a un sitio genérico usando un sustantivo común (*río*).

[ejemplos]

3. (GEO-NOM) Visitó las **Cataratas de Iguazú**.
4. (GEO-GEN) Se cayó mientras daba un paseo por el **monte**.

[FAC]

Se usa para las **estructuras y espacios construidos por humanos (facilities)**: viviendas, comercios, sitios de ocio, lugares turísticos, estructuras de transporte (estaciones de autobuses y trenes), oficinas, granjas, escuelas, ...

Distinguimos dos subtipos:

3. **FAC-NOM**: menciones que hacen referencia a un sitio concreto con nombre propio (*Sagrada Familia*).
4. **FAC-GEN**: menciones que hacen referencia a un sitio genérico usando un sustantivo común (*monumento*).

[ejemplos]

5. (FAC-NOM) Se remitió el caso al **Hospital La Paz** en **Madrid**.
6. (FAC-GEN) Después de salir de **prisión** trabajó un año en el **aeropuerto**.

ETIQUETAS DEMOGRÁFICAS	
[COMUNIDAD]	Se anotan los gentilicios, es decir, adjetivos o sustantivos que hacen referencia a un lugar, así como etnias, religiones y elementos sociodemográficos del estilo.
[ejemplos]	
7. Mujer pakistaní de 16 años. 8. Varón de 36 años de nacionalidad ucraniana . 9. Paciente de origen español .	
[IDIOMA]	Esta categoría se usa para los idiomas que se mencionan en el texto.
[ejemplos]	
10. La paciente presenta problemas comunicativos, ya que no habla español 11. El hijo interpreta la conversación, ya que el paciente solo habla chino	
OTRAS ETIQUETAS	
[DEPARTAMENTO]	Entran dentro de esta categoría los diferentes departamentos y servicios de instituciones de salud (por ejemplo: Dermatología, Cardiología), así como las distintas partes dentro de hospitales y centros de salud.
[ejemplos]	
12. En planta , previo a quirófano , el paciente no presenta antecedentes personales de interés. 13. Seguimiento por cardiología pediátrica .	
[TRANSPORTE]	Entran dentro de esta categoría las menciones a transportes y movimientos en general, incluyendo la

palabra “viaje”, menciones a movimientos de pacientes y medios de transporte, entre otros.

[ejemplos]

14. El paciente acaba de volver de un **crucero**.

15. No refiere **viajes** recientes.

2.2 REGLAS DE MARCADO Y EXTENSIÓN

A continuación, se describen las reglas generales para el marcado de entidades. Estas reglas se aplican a todas las categorías, independientemente de su etiqueta:

B1. REGLAS DE MARCADO GENERALES POSITIVAS	
[GEN-P1] [ortografía]	Se deben anotar todas las menciones a lugares aunque incluyan errores ortográficos.
[ejemplos]	
1. Paciente natural de [sic] [nicaragua] _{GPE} .	
[GEN-P2] [separación]	Si se enumeran varias entidades geográficas de distinto o del mismo tipo, debe anotarse cada una de ellas por separado.
[ejemplos]	
2. Sus abuelos viven en Puente Genil (Córdoba) .	
3. Ha estado viajando por Europa (Noruega, Dinamarca y Alemania) .	
4. Estuvo durante 5 días en una [zona rural] _{GEO} de [Santa Cruz] _{GPE} .	
[excepciones]	
<ul style="list-style-type: none"> Las divisiones administrativas (<i>ciudad, provincia, ...</i>) que vayan junto a un 	

topónimo se anotan como una sola entidad (regla GPE-P2). Esto facilitará después la normalización ante casos ambiguos como *ciudad de Barcelona* o *provincia de Barcelona*.

5. Originaria del **estado de San Luis Potosí**, con residencia en el **barrio Acantzen** de la **localidad Tanchahuil** en el **municipio de San Antonio**, en el mismo **estado**.

[GEN-P3]
[anidación]

En el caso de que dos menciones se solapen, las anotaremos de manera anidada. Es decir, se anotarán ambas menciones de manera independiente desde principio a fin.

En el ejemplo de abajo, se usan corchetes para marcar hasta dónde llega cada anotación independiente.

[ejemplos]

6. Había estado en **[[América Central]_{GPE} y del Sur]_{GPE}**.

[GEN-P4]
[siglas]

Las siglas que aparezcan junto a su forma larga se anotan juntas.

[ejemplos]

7. Es llevada por la policía al **servicio de urgencias (SU)** después de que, al parecer, tratara de secuestrar un **autobús**.

[GEN-P5]
[adjetivos-prep]

Si se especifican características de una entidad _GEN mediante adjetivos o sintagmas preposicionales, se incluyen dentro de la mención.

[ejemplos]

8. Es dueña de una **casa** en el **piedemonte** de la **cordillera** de la **costa litoral central venezolana** (*GEO_GEN*)

[GEN-P6]

[compl-nombre]

Anotamos también los lugares que aparezcan dentro de un complemento del nombre.

[ejemplos]

9. Las pruebas de **laboratorio** apoyan la hipótesis inicial.
10. Recoge de la calle o hurta en el trabajo herramientas, publicidad, material de **papelería** y utensilios de limpieza.
11. Médico de **urgencias**

B2. REGLAS DE MARCADO GENERALES NEGATIVAS

[GEN-N1]

[det-prep]

No se incluyen dentro de las menciones determinantes ni preposiciones anteriores a la entidad.

[ejemplos]

12. Recién llegada de **un viaje a Tanzania**.
13. Trabajador afincado en **los Países Bajos**.

[excepciones]

- Si el determinante forma parte de un topónimo (suele ir en mayúsculas), sí que debe incluirse dentro de la mención.

14. Hizo un **viaje** desde **El Salvador** hasta **La Coruña**.

Esta excepción no afecta, sin embargo, a las contracciones “del” y “al” (véase ejemplo 10).

15. Procedente **del Cairo**.
Procedente del **Cairo**.

- Algunos GPE genéricos deben anotarse junto al determinante si van acompañados de un determinante posesivo (ver GPE-E2).

16. Residente en **nuestro país**.

<p>[GEN-N2] [sitios-genéricos]</p>	<p>No se anotan sustantivos que hacen referencia a un lugar de manera muy muy genérica si van solos: <i>sitio, lugar, zona, área, ...</i></p> <p>[Excepción: DEPA-E2]</p>
<p>[ejemplos]</p> <p>17. Vive a la "manera local", con las mismas medidas higiénico-dietéticas de los naturales de la zona.</p> <p>18. Su centro laboral se encuentra próximo a un río, lugar donde lava sus herramientas.</p>	
<p>[GEN-N3] [adv-dem]</p>	<p>No se anotan los demostrativos (<i>aquí, ahí, allí, ...</i>) ni adverbios (<i>delante, al lado, detrás, lejos, cerca, ...</i>) de lugar.</p>
<p>[ejemplos]</p> <p>19. Una vez llegó allí estuvo en contacto con varios animales de granja.</p> <p>20. Se torció el tobillo delante del ambulatorio.</p>	
<p>[GEN-N4] [eventos]</p>	<p>No se anotan menciones generales a sitios donde ha ocurrido algún evento.</p> <p>[Reglas relacionadas: GEO-N2, GEO-E1]</p>
<p>[ejemplos]</p> <p>21. No encontraron pruebas en el lugar del accidente.</p> <p>22. El ladrón siempre vuelve a la escena del crimen.</p>	
<p>[GEN-N5] [sitios-web]</p>	<p>Los sitios web en general, nombres de aplicaciones, ... no se consideran lugares y no se anotan.</p>
<p>[ejemplos]</p>	

23. Conoció a su actual pareja en Twitter .	
[GEN-N6] [especies-proced]	No se anotan los nombres de localizaciones que formen parte o sean homónimos del nombre de patógenos, procedimientos clínicos o enfermedades.
[ejemplos]	
24. La paciente presentaba un Glasgow de 15.	
25. La aplicación del test reveló una banda específica para L. braziliensis , la cual no es detectada en sueros de individuos infectados por L. mexicana o L. amazonensis .	
26. Según su árbol genealógico, muchos familiares tenían Corea de Huntington	
27. A lo largo de la evolución, se confirmarán los dos criterios de Ámsterdam restantes	

Las siguientes tablas explican las reglas específicas para cada una de las entidades geográficas:

C1. REGLAS PARA “GPE”	
REGLAS POSITIVAS	
[GPE-P1] [ejemplos]	Se anotan con esta etiqueta regiones geográficas habitadas por personas y que están influenciadas por la política, la economía, la cultura y la historia de su entorno. Esto incluye países, regiones, ciudades, ...
[ejemplos]	
28. El paciente reside en Madrid pero por motivos de trabajo se había traslado en los últimos días a Huelva . (NOM)	
29. Llegó de la India cuando tenía 11 años. (NOM)	
30. Niña de seis años, natural y procedente de Pozuzo, departamento de Pasco (selva central) , admitida en un hospital de la ciudad de Lima en agosto de 2006. (NOM)	

31. Paciente de 18 años procedente de **Guinea Conakry**, sin antecedentes de interés, recién llegado en **patera** a **España** hace un mes y medio. (NOM)
32. Casada y con un hijo sano, vive en un **área rural** cercana a la **capital de la isla de Gran Canaria**. (NOM)
33. Independiente para las ABVD que vive solo en un **barrio con nivel socioeconómico medio**. (GEN)
34. Parasitosis prevalente de **[[países tropicales] y subtropicales]**, causada por diversos helmintos. (GEN)
35. Recientemente había estado trabajando en el **extranjero**. (GEN)
36. Había recibido antibióticos por infección urinaria (pielonefritis), sin mejoría, en un **establecimiento** de **su tierra natal**. (GEN)

[GPE-P2]
[divisiones-1]

Las divisiones administrativas ("distrito", "provincia"), regiones y zonas específicas ("zona noroeste"), puntos cardinales, etc. se incluyen dentro de la mención si aparecen junto a un nombre propio, anotándose todo como GPE_NOM.

[ejemplos]

37. Natural de la **provincia de Punjab**.
38. El gazpacho es un plato típico del **sur de España**.

[GPE-P3]
[divisiones-2]

Las divisiones administrativas que aparezcan en solitario o con un sintagma adjetival o preposicional se anotan como GPE-GEN.

[ejemplos]

39. Este agosto planea hacer un recorrido por varias **[capitales asiáticas]** COMUNIDAD GPE-GEN.
40. Se han visto casos similares de picaduras de garrapata en la **provincia**.

[GPE-P4]
[barrios]

Anotaremos los barrios y vecindarios como GPE, considerándolos uno de los tipos de divisiones administrativas más pequeños.

[ejemplos]

41. A 5 personas del **vecindario** en que reside el paciente se les había diagnosticado COVID-19
42. Al llegar a **Sevilla** se alojó en un **piso compartido** de un **barrio marginal**.
43. ...mapeo de activos de salud comunitaria preexistente en el **barrio de La Pau**.
44. Antecedentes epidemiológicos: hombre heterosexual, residente en **barrio de Antímamo**, ubicado al **suroeste caraqueño**.

REGLAS ESPECIALES

[GPE-E1]
[extranjero]

La palabra “extranjero” se anota como GPE-GEN si se usa para referirse a un lugar.

(Ver también regla COMU-E1)

[ejemplos]

45. Hace quince años que viajó por última vez al **extranjero**.

[GPE-E2]
[pos-más-país]

A modo de excepción de la regla GEN-N1, si un determinante posesivo (“mi”, “su”, “nuestro”, ...) acompaña a una división administrativa (“país”, “pueblo”, “localidad”, “distrito”, ...) se anota todo como una sola mención con la etiqueta GPE_GEN.

[ejemplos]

46. Se presenta el primer caso en **nuestro país** de coronavirus.

<p>[GPE-E3] [sprep-demo]</p>	<p>Los complementos del nombre que hagan referencia al lugar natal o residencia de alguien deben incluirse en la mención.</p> <p>Esta regla suele combinarse con la GPE-E2.</p>
<p>[ejemplos]</p> <p>47. Calendario vacunal actualizado en su país natal.</p> <p>48. Fue operado en su país de residencia.</p>	
<p>[GPE-E4] [coordenadas]</p>	<p>Deben anotarse como GPE-NOM las coordenadas si aparecen en el texto (también cuando aparezcan junto a una entidad que sí anotaríamos).</p>
<p>[ejemplos]</p> <p>49. Procedente de la Sierra de Parima (2° 47' 21" N, 64° 13' 55" W).</p>	

C2. REGLAS PARA "GEO"	
REGLAS POSITIVAS	
<p>[GEO-P1] [ejemplos]</p>	<p>Se anotan con esta etiqueta los accidentes geográficos, sitios naturales en general, tipos de ecosistemas y ambientes, zonas del mundo, ...</p>
<p>[ejemplos]</p> <p>50. El grupo permaneció gran parte del tiempo en la playa a pies descalzos y se bañaban en la desembocadura de un río (GEN)</p> <p>51. En la última década existe el consenso mayoritario de un aumento de la prevalencia e incidencia de esta enfermedad en Europa, en la cuenca mediterránea. (NOM)</p> <p>52. Exposición a polvos de cenizas volcánicas por su cercanía al radio de emisión del volcán Tungurahua (NOM)</p> <p>53. Trabajador de la construcción, con exposición habitual al asbesto y durante 3 años en una mina de carbón (GEN)</p>	

54. La patología se incrementa cuando nos acercamos al **ecuador** (NOM)
 55. Realiza habitualmente excursiones por el **campo** en bicicleta (GEN)
 56. Ascendió hasta la cima de un **cerro de 1.828 metros**, bajo intenso calor ambiental (GEN)
 57. Había viajado a la **zona rural** de **Dagua (Valle del Cauca)**. (GEN)
 58. Relató haber acudido a una **zona de aguas termales** al aire libre en el área próxima a su **municipio granadino** (GEN)
 59. Antecedente de viajes a **zonas endémicas de tungiasis** (GEN)

[GEO-P2]
[acc-nom]

Si aparece el nombre del accidente geográfico junto a un topónimo se anota todo como una sola mención con la etiqueta GEO_NOM.

[ejemplos]

60. Este verano se bañó en el **mar Mediterráneo** y en el **océano Atlántico**.
 Este verano se bañó en el mar **Mediterráneo** y en el océano **Atlántico**.

[GEO-P3]
[cardinales]

Los puntos cardinales, si no especifican una localización dentro de una entidad tipo GPE (ej: "sur de España"), se anotan como GEO-GEN.

[ejemplos]

61. El gazpacho es un plato típico del **sur**.

[GEO-P4]
[adjetivos-prep]

Si se especifica el tipo de una entidad GEO-GEN mediante adjetivos o sintagmas preposicionales, se incluye dentro de la mención.

[ejemplos]

62. Es dueña de una **casa** en el **piedemonte** de la **cordillera** de la **costa litoral central venezolana**.

<p>[GEO-P5] [hábitats]</p>	<p>Los sustantivos “región”, “zona” y “área” seguidos de un adjetivo que exprese temperatura (frío, cálido, ...), tipo de ecosistema (selvático, rural, urbano, ...) o similar se anotan como GEO-GEN. Nos referiremos a este tipo de menciones como “hábitats” o “ecosistemas”.</p> <p>(Ver también: regla GEO-N3)</p>
<p>[ejemplos]</p> <p>63. Trabajaba en una región selvática del país.</p> <p>64. Una mujer joven proveniente de área endémica.</p>	
<p>[GEO-P6] [endémicas]</p>	<p>Cuando se mencionen en el texto zonas endémicas, epidémicas, de alto riesgo, ... de enfermedades o patógenos concretos, extenderemos la mención hasta el nombre del patógeno o enfermedad.</p>
<p>[ejemplos]</p> <p>65. Estuvo de visita en una región endémica de babesiosis.</p>	
<p>REGLAS NEGATIVAS</p>	
<p>[GEO-N1] [celestes]</p>	<p>No se anotan los objetos astronómicos: planetas, estrellas, constelaciones, ...</p>
<p>[ejemplos]</p> <p>66. El paciente presenta delirios en los que cree que trabaja como astronauta y vive con su familia en Neptuno.</p>	
<p>[GEO-N2] [aguas]</p>	<p>Aunque puedan parecer sitios, no anotaremos los tipos de agua donde alguien se puede bañar.</p> <p>[Ver regla relacionada: GEN-N4]</p>

[ejemplos]

67. No refería conductas de riesgo específicas durante el **viaje** salvo baños en **agua dulce** en **Bamako**.

REGLAS ESPECIALES

[GEO-E1]

[lugar-consumo]

Como excepción a la regla GEN-N4, decidimos anotar la mención “lugares de consumo”.

[Ver regla relacionada: GEN-N4]

[ejemplos]

68. Se prescribe control de los factores de riesgo para evitar recaídas (control del manejo de dinero, evitación de amistades consumidoras y **lugares de consumo**, evitar el estar y salir sólo).

C3. REGLAS PARA “FAC”

REGLAS POSITIVAS

[FAC-P1]

[ejemplos]

Se anotan con esta etiqueta espacios construidos por humanos, lo cual incluye muchos sitios comunes que encontraríamos en una ciudad: establecimientos, tiendas, estructuras, monumentos, ... así como distintos tipos de centros de salud.

[ejemplos]

69. Refiere haber trabajado en una **tienda de bolsos** cuando tenía 16 años durante un año, en una **pescadería** y en una **gasolinera** (GEN)

70. Comienza de manera regular a acudir al **gimnasio**. (GEN)

71. Se matriculó al finalizar bachillerato en una **academia de informática y mecanografía**, que no finalizó (GEN)

72. El paciente declara no ser fumador y trabajar en una **fábrica de baterías para coches**. (GEN)
73. El minucioso examen del **vehículo**, una vez depositado en el **desguace**, permitió la recogida de varios cartuchos (GEN)
74. Los padres fueron ingresados en una **Residencia Psicogeriátrica** por Orden Judicial (GEN)
75. Posteriormente, el paciente fue referido al **Instituto Nacional de Cancerología** para su valoración (NOM)
76. Trabajaba en la restauración de un **edificio escolar**, en la **parroquia de Santa Rosalía** de **Caracas, Venezuela** (GEN/NOM)
77. La cepa fue enviada al **Servicio de Micología** del **Centro Nacional de Microbiología** para identificación definitiva (NOM)
78. Ingresó en una **casa de acogida** cuando la paciente tenía 13 meses (GEN)
79. Fue ingresada en julio de 2008 (verano en el **Hemisferio Norte**) al **Hospital San Juan de Dios**, con el antecedente de haber regresado 8 días atrás de un **viaje** al **Norte del Estado de La Florida**, en los **Estados Unidos** -donde estuvo paseando en un **bosque** conocido por tener una alta población de garrapatas (NOM)

[FAC-P2]
[tipo-edificio]

Si se menciona el tipo de *facility* junto a su nombre propio se anota todo como una sola mención y se anota como FAC-NOM.

[ejemplos]

80. Trabaja como camarero en el **restaurante La Barca**.

[FAC-P3]
[pisos]

Anotamos como FAC-GEN las partes dentro de un edificio ("primer piso", "planta baja", "sótano", "azotea", ...), así como los distintos tipos de habitaciones ("cocina", "dormitorio", "baño", ...).

[ejemplos]

81. **Vivienda:** Vive en un **primer piso** de un **edificio** de **Tarragona**, ante el **Parque Central**.
82. Sufre un episodio de pérdida de conocimiento estando en el **lavabo** preparándose para ir a trabajar
83. Disminuye la tendencia a aislarse en su **dormitorio**, pasando la mayor parte del día en el **salón común**.

[FAC-P4]
[nombres-hospitales]

Algunos hospitales concretos incluyen dentro de su nombre el sitio al que pertenecen. Es el caso de lugares con nombres muy genéricos como "Hospital Universitario", "Hospital General", En estos casos, anotaremos todo junto como entidad FAC_NOM (incluyendo el nombre del sitio) y también anotaremos una entidad GPE-NOM de manera anidada.

[ejemplos]

84. Fue ingresado en una **unidad de cuidados intensivos pediátricos** en el **[Hospital Maternoinfantil de la [provincia de Hubei]] (Wuhan, China)**.
85. Los **[Centros de Control y Prevención de Enfermedades de [Pekín] (CDC)]** confirmaron, mediante RT-PCR en tiempo real, que el paciente tenía COVID-19
86. Realiza seguimiento en el **[Hospital Clínic de [Barcelona]]**

[excepciones]

- No lo anotaremos así si se especifica el sitio donde está el hospital, pero no forma parte de su nombre (ver último ejemplo).
87. Traslado al **Hospital Niño Jesús en Madrid**.
88. Se remitió para seguimiento a las **consultas externas** de su **hospital de origen (Alto Guadalquivir de Andújar)**.
- Tampoco se aplica esta regla cuando se hable de algún sitio de manera genérica o poco específica. Estos casos se anotan por separado (FAC_GEN + GPE).
89. En 2015 ingresó en un **hospital** de **Málaga** de donde se escapa a los pocos días de su ingreso.

<p>[FAC-P5] [residencia]</p>	<p>También se anotan como FAC las menciones al lugar de residencia del paciente.</p>
<p>[ejemplos]</p> <p>90. Ya se había planteado la necesidad de incluir al paciente en recurso residencial</p> <p>91. Al alta, volvió a su domicilio</p>	
<p>REGLAS NEGATIVAS</p>	
<p>[FAC-N1] [organizaciones]</p>	<p>En algunos contextos, hay nombres que parecen de Facilities pero que en realidad se refieren a organizaciones. De manera general, estos casos no deben anotarse.</p>
<p>[ejemplos]</p> <p>92. Se traslada al medio residencial de FAISEM (Fundación Pública Andaluza para la Integración Social de Personas con Enfermedad Mental), continuando seguimiento en esta Comunidad Terapéutica.</p>	
<p>REGLAS ESPECIALES</p>	
<p>[FAC-E1] [ambulatorio]</p>	<p>Anotamos la palabra <i>ambulatorio</i> también en expresiones como “seguimiento ambulatorio” o “control ambulatorio”..</p>
<p>[ejemplos]</p> <p>93. Es dada de alta para seguimiento ambulatorio.</p>	

<p>C4. REGLAS PARA “COMUNIDAD”</p>	
<p>REGLAS POSITIVAS</p>	
<p>[COMU-P1] [nombre-adj]</p>	<p>Se deben anotar todos los gentilicios, tanto si aparecen como adjetivos como si aparecen como sustantivos.</p>

[ejemplos]

94. Hombre **español** de 17 años de edad que viene a **consulta** por...

95. Fue atendido en un **[centro hospitalario [español]COMU]FAC**.

[COMU-P2]
[anidados]

Los gentilicios que aparezcan dentro de una mención de tipo geográfico también se anotan..

[ejemplos]

96. Pasó su infancia en una **[ciudad [africana]COMU]GPE**.

REGLAS ESPECIALES

[COMU-E1]
[extranjero]

Por convención, anotaremos la palabra “extranjero” usando esta etiqueta cuando se refiere a una persona.

(Ver también regla GPE-E1)

[ejemplos]

97. La paciente es de nacionalidad **extranjera** aunque desconocemos exactamente de dónde

[COMU-E2]
[religiones]

Debido a su posible relevancia clínica, aprovechamos esta etiqueta para anotar también religiones y etnias en general..

[ejemplos]

98. Paciente **hindú** de 21 años...

99. De origen **judío**

C5. REGLAS PARA “IDIOMA”

REGLAS POSITIVAS

<p>[IDI-P1] [idioma]</p>	<p>Se anotan todas las menciones a idiomas.</p>
<p>[ejemplos]</p> <p>100. El paciente no habla castellano y no conseguimos localizar al traductor.</p>	
<p>[IDI-P2] [de-habla]</p>	<p>La expresión “de habla...” más un adjetivo que haga referencia a un idioma se anota con esta etiqueta, sin incluir la preposición.</p>
<p>[ejemplos]</p> <p>101. Se trata de algo muy común en los países de habla inglesa.</p>	
<p>REGLAS ESPECIALES</p>	
<p>[IDI-E1] [anglosajón]</p>	<p>Por convención, anotaremos la palabra “anglosajón” usando la etiqueta de idioma.</p>
<p>[ejemplos]</p> <p>102. Paciente anglosajón de 76 años residente en España desde hace 22.</p>	
<p>[IDI-E2] [barrera idiomática]</p>	<p>Por convención, anotaremos la expresión “barrera idiomática” y similares con esta etiqueta..</p>
<p>[ejemplos]</p> <p>103. No entra en el estudio por motivos de barrera idiomática.</p>	

C6. REGLAS PARA “DEPARTAMENTO”

REGLAS POSITIVAS

<p>[DEPA-P1] [departamentos]</p>	<p>Anotamos los departamentos, servicios, secciones y unidades dentro de los hospitales y otros centros de salud.</p> <p>En este tipo de anotaciones normalmente incluimos la palabra “pediátrica” dentro de la misma mención, ya que se refiere a un tipo de servicio concreto.</p>
<p>[ejemplos]</p> <p>104. La paciente fue tratada por el departamento de enfermedades infecciosas del Hospital La Paz.</p> <p>105. Paciente vista por primera vez en el servicio de Endocrinología Pediátrica del Hospital Materno-Infantil de Málaga, a los 8 años y 6 meses de edad, enviada con diagnóstico de ADS.</p> <p>106. Se realiza ingreso en la UVI y posteriormente se decide ingreso en planta de hospitalización convencional.</p>	
<p>[DEPA-P2] [especialidades]</p>	<p>Anotamos los nombres de especialidades médicas cuando se refieran a secciones o equipos dentro de un centro médico.</p>
<p>[ejemplos]</p> <p>107. Estudian el caso Medicina Interna y Neurología y no encuentran datos patológicos justificativos en las pruebas complementarias</p> <p>108. A pesar de su baja FE, cardiología no consideró que se hallara en insuficiencia cardíaca aguda</p> <p>109. Pedimos consulta a neurología tras episodio convulsivo.</p>	
<p>[DEPA-P3] [salas]</p>	<p>Anotamos las salas (p. ej. de espera o rehabilitación) y plantas concretas (p. ej. de especialidades o de hospitalización) dentro de un centro médico, incluyendo habitaciones y camas.</p>

[ejemplos]

110. Dos semanas después de su ingreso en la **planta de Aparato Digestivo**, el cuadro se complicó con una hemorragia digestiva grave

111. El paciente se mantuvo con soporte ventilatorio en el **área de observación de urgencias** a la espera de un posible fallecimiento o donación de órganos

112. Se llevó al paciente de la **ambulancia** a una **sala de flujo laminar** (con antecámara) en la **primera planta** de la **unidad**

[DEPA-P4]
[consulta]

Anotamos la palabra “consulta” por sí sola cuando se refiera al sitio físico, así como tipos específicos de consultas (externas, especialidades).

Como la palabra “consulta” por sí sola puede ser muy ambigua (entre otros, puede significar “visita” o “pregunta”), un buen truco para saber si debe anotarse es reemplazarla por el sinónimo “consultorio”.

[ejemplos]

113. La paciente fue enviada desde la **Consulta de Otorrinolaringología (ORL)** a la **Consulta Psiquiátrica** en abril de 2019

114. Paciente de 30 años de edad que **consulta** por dolor abdominal en el **Servicio de Urgencia** del **Hospital San José, Santiago de Chile**.

115. Acude a la **consulta de un psiquiatra** llevado por su esposa.

116. Motivo de la **consulta**: accidente laboral.

117. Un año después, el paciente retorna a **consulta** por presentar cólicos abdominales de fuerte intensidad.

[DEPA-P5]
[equipos]

Anotamos también los equipos de especialidades del hospital, ya que implican que el paciente ha pasado por dicha especialidad.

[ejemplos]

118. Es seguida desde febrero de 2012 por el **Equipo de Atención Primaria (EAP)** para control de síntomas y observación de su enfermedad neoplásica.

119. Finalmente, el **Equipo de Valoración de Incapacidades** del **Instituto Nacional de la Seguridad Social**, reunido para valorar su caso resuelve Incapacidad Permanente Total derivada de enfermedad profesional

120. Comentado el caso en **Comité de Tumores de Otorrinolaringología**, se decide la realización de una tomografía por emisión de positrones para valorar la naturaleza de las adenopatías

REGLAS NEGATIVAS

[DEPA-N1]
[centros]

No se anotan con esta categoría los centros médicos de especialidades, policlínicas, dispositivos de rehabilitación, ... Este tipo de menciones se anotan con la clase FAC, puesto que suelen ser edificios enteros.

[ejemplos]

121. No ha sido tratado en ningún otro **centro médico**

122. Fue tratada en una **clínica de salud mental y adicciones**

[DEPA-N2]
[pruebas]

No se anotan con esta categoría nombres de pruebas o aspectos biológicos, aunque coincidan con el nombre de alguna especialidad médica.

[ejemplos]

123. Con respecto a la **microbiología**, se han descrito una alta variabilidad de patógenos como responsables de esta patología.

124. A las 24 horas de la **cirugía**, el paciente estaba afebril por primera vez desde su llegada al **hospital**

125. La **anatomía patológica** informa de inflamación granulomatosa necrotizante

<p>126. Pruebas realizadas de urgencia a su llegada</p> <p>127. Digestivo: Ausencia de signos de peritonismo, puño percusión renal derecha positiva</p> <p>[Aquí, “Digestivo” hace referencia a un tipo de exploración, no al departamento clínico]</p>	
REGLAS ESPECIALES	
<p>[DEPA-E1] [específicos]</p>	<p>Anotamos con esta etiqueta los siguientes sitios dentro de un centro médico: quirófano, incubadora, pabellón,</p>
<p>[ejemplos]</p> <p>128. Ingresa directamente en quirófano para extracción de la bala.</p> <p>129. Permanece en la incubadora hasta el día 17 después del parto.</p> <p>130. Cuarenta y cinco minutos después de su llegada, ingresó a pabellón, con el diagnóstico de herida por arma de fuego abdominal complicada</p>	
<p>[DEPA-E2] [servicio]</p>	<p>Anotamos las palabras “servicio” y “unidad” también cuando van solas, siempre que se refieran a una localización.</p>
<p>[ejemplos]</p> <p>131. Ante la reaparición de sintomatología consultó en nuestro Servicio.</p>	

C7. REGLAS PARA “TRANSPORTE”	
REGLAS POSITIVAS	
<p>[TRA-P1] [viajes]</p>	<p>La palabra viaje y sus posibles sinónimos se anotan siempre que aparezcan (excepción: regla TRA-N2).</p> <p>Algunos posibles sinónimos son: <i>visita</i> (siempre que haga referencia a algún desplazamiento y no a una consulta</p>

médica), *recorrido*, *travesía* o *desplazamiento*, entre otros elementos.

[ejemplos]

- 132. Tenía planeado un **viaje** por **Europa del Este** con su familia.
- 133. Se estaba gestionando la **aeroevacuación** a **Italia**.
- 134. Comentó que sus problemas habían empezado 3 meses antes, al regresar de una **visita** a **Japón**.
- 135. Negaba picaduras de algún tipo, aunque algunos de sus compañeros de **excursión** en **Nueva York** sí habían sufrido picaduras de garrapata

[TRA-P2]
[tipos-viajes]

Se incluyen en la mención los complementos del nombre que describan el tipo o motivo del viaje: *intercontinental*, *de turismo*, *laboral*, ... No se deben incluir, sin embargo, adjetivos de tiempo o frecuencia.

En el caso de que se describa más a fondo alguna característica del viaje ("viaje turístico tipo 'mochilero'") debemos anotar solo el primer calificador.

[ejemplos]

- 136. Refería haber realizado múltiples **viajes intercontinentales**, sin problemas de salud aparentes.
- 137. Historia de **viajes anteriores** a **Tailandia**, **Marruecos** y **México**..
- 138. Realizó un **viaje turístico tipo "mochilero"** con su pareja

[TRA-P3]
[medios]

Los medios de transporte también se anotan con esta etiqueta.

[ejemplos]

- 139. Le arrolló un **autobús** al cruzar la **carretera**.
- 140. Llegada a **España** en **patera** a **Canarias**.
- 141. Al paso de su **vehículo** estalla un artefacto explosivo de gran potencia.

REGLAS NEGATIVAS

<p>[TRA-N1] [verbos]</p>	<p>Los verbos de movimiento no se anotan. (Excepción: regla TRA-E2)</p>
<p>[ejemplos]</p> <p>142. Viajó recientemente a casa de sus padres en Tailandia.</p> <p>143. Hace tres años voló de Marruecos a España</p> <p>144. Había viajado a la India hace unos meses</p>	
<p>[TRA-N2] [viaje-figurado]</p>	<p>Las menciones a viajes no se anotan si no hacen referencia real a uno. Es decir, no incluimos viajes de drogas, objetos específicos para usarlos durante viajes, menciones dentro de ocupaciones, ...</p>
<p>[ejemplos]</p> <p>145. Trabaja como agente de viajes.</p> <p>146. Usa medias de viaje para evitar la trombosis.</p> <p>147. Acabó en Urgencias tras un mal viaje de ácido.</p> <p>148. Una mujer de 47 años de viaje turístico por la Argentina, habitando con su compañero de viaje una casa rodante</p>	
<p>REGLAS ESPECIALES</p>	
<p>[TRA-E1] [silla-ruedas]</p>	<p>Anotamos con esta etiqueta también las menciones a sillas de ruedas y andadores.</p>
<p>[ejemplos]</p> <p>149. Precisa silla de ruedas para acudir a la Unidad.</p> <p>150. Neuropatía periférica que con el paso de los años se tornó invalidante (año 2008) y la obliga a desplazarse en silla de ruedas.</p> <p>151. Marcha con andador.</p>	

[TRA-E2]

[traslado]

Anotaremos la palabra “traslado”, tanto en forma de sustantivo como verbal, cuando se refiera al movimiento de pacientes entre entidades y establecimientos de salud.

[ejemplos]

152. Se va a proceder a su **traslado** de forma urgente al **Hospital Nakasero** en **Kampala**.
153. En la última consulta, al palpar el abdomen su médico de **atención primaria** valoró el **traslado** a **servicio de Urgencias** del **hospital**.

[TRA-E3]

[visita]

Anotaremos la palabra “visita” cuando vaya acompañada de adjetivos u otros complementos del nombre.

[ejemplos]

154. [...] facilitada por el impulso de la **visita telemática** instaurada durante la pandemia de COVID-19.
155. Realizamos **visita a domicilio**.

3. CARACTERIZACIÓN DE ENTIDADES

En esta sección se describen los atributos con los que caracterizaremos las entidades marcadas durante la primera fase de la anotación. Estos atributos expresan distintos tipos de información: información clínica, referente o persona asociada, polaridad y temporalidad.

3.1 CARACTERIZACIÓN CLÍNICA

Hay ciertos aspectos relacionados con lugares que interesan especialmente para el ámbito biomédico. Con el fin de rescatar cierto tipo de información y pensando en los distintos casos de uso, hemos definido cinco clases que nos permitirán caracterizar las entidades que marcamos. Estas categorías se usan solo para las entidades de lugares (GPE, GEO, FAC).

La siguiente tabla expone las categorías para la caracterización clínica de entidades:

D1. ETIQUETAS DE CARACTERIZACIÓN CLÍNICA PARA ENTIDADES	
[LUGAR NATAL]	Se usará para las entidades que hagan referencia al lugar de procedencia de una persona: dónde nació, cuál es su país de origen, nacionalidad, etc.
[ejemplos]	
1. Mujer de 52 años, natural de Estados Unidos .	
[RESIDENCIA]	Se usará para las entidades que hagan referencia al lugar de residencia actual o pasado de una persona.
[ejemplos]	
2. Residencia actual en Galicia desde hace 7 años.	

[MOVIMIENTO]	Se usará para las entidades que hagan referencia a lugares que son el origen o el destino de un desplazamiento. Las estancias también entran dentro de esta categoría.
<p style="text-align: center;">[ejemplos]</p> <p>3. Dos meses antes, había viajado a Mauritania.</p> <p>4. Tras regresar del viaje a la India</p> <p>5. El viaje tuvo motivos religiosos, con estancia principalmente en el estado de Rajasthan (noroeste del subcontinente indio)</p>	
[ATENCIÓN]	Se usará para las entidades que hagan referencia al lugar en el que el paciente ha tenido algún tipo de atención médica.
<p style="text-align: center;">[ejemplos]</p> <p>6. Requirió atención médica en Manaus</p> <p>7. En 1988, intervenida de bocio de Argelia.</p> <p>8. Acudió a un centro médico en Turquía.</p>	
[OTROS]	Se usará para entidades que no encajen dentro de ninguna de las categorías anteriores.
<p style="text-align: center;">[ejemplos]</p> <p>9. El cuadro comenzó en Senegal.</p> <p>10. Escolarizado sin problemas en Ecuador.</p>	

La tabla siguiente describe las reglas de clasificación semántica clínica de entidades geográficas:

D2. REGLAS DE CARACTERIZACIÓN CLÍNICA DE ENTIDADES	
REGLAS GENERALES	
<p>[CLAS-G1] [multietiqueta]</p>	<p>Las entidades geográficas pueden tener más de una etiqueta semántica.</p> <hr/> <p>[ejemplos]</p> <p>11. Estuvo de vacaciones en Ibiza, de donde era natural. [ETIQUETAS: lugar natal, movimiento]</p>
<p>[CLAS-G2] [temporalidad]</p>	<p>Las etiquetas semánticas se atribuyen sin tener en cuenta la temporalidad de la frase.</p> <hr/> <p>[ejemplos]</p> <p>12. Estuvo de vacaciones en Kenia en el año 83. [ETIQUETAS: movimiento]</p> <p>13. Planea hacer un recorrido el año que viene por Sudamérica: Chile, Argentina, Uruguay. [ETIQUETAS: movimiento]</p>
<p>[CLAS-G3] [polaridad]</p>	<p>Las etiquetas semánticas se atribuyen independientemente de si la entidad está afirmada, negada o especulada.</p> <hr/> <p>[ejemplos]</p> <p>14. Los posibles destinos de su viaje incluyen Noruega, Estonia y Turquía. [ETIQUETAS: movimiento]</p>
REGLAS MULTICLASE	
<p>[CLAS-M1] [estancias]</p>	<p>Las menciones a estancias se etiquetan normalmente como MOVIMIENTO.</p>

[ejemplos]

15. Realizará un recorrido tanto por **[[zonas urbanas]_{GEO} como rurales]_{GEO}**, incluyendo una estancia en la **selva** de 2 a 4 semanas.

3.2 CLASIFICACIÓN POR REFERENTES

En la clasificación por referentes intentaremos relacionar cada entidad anotada a la persona con la que está asociada. Para asignar esta etiqueta, debemos leer la frase donde aparece la entidad e inferir del contexto a quién está asociada. Por el tipo de dominio con el que trabajamos, en la mayoría de frases la etiqueta por defecto será la de PACIENTE.

La tabla a continuación explica las categorías posibles para la anotación de referentes:

E1. ATRIBUTOS PARA REFERENTES	
[PACIENTE]	Se usará para las entidades que hablen sobre el paciente. Es la etiqueta por defecto para la mayoría de entidades.
[ejemplos]	
16. Mujer de 36 años, de nacionalidad ucraniana , residente en España .	
[FAMILIAR]	Se usará para las entidades que hablen sobre personas que son familiares del paciente.
[ejemplos]	
17. Casado en el 2009 con un varón natural de Málaga .	
[OTROS]	Se usará para las entidades que no hagan referencia a personas (otros lugares, eventos, ...).

[ejemplos]

18. Se describe el primer caso de anisakis en **Colombia**.

A continuación, se describen algunas reglas específicas para esta clasificación:

E2. REGLAS DE CLASIFICACIÓN POR REFERENTES	
REGLAS GENERALES	
[REF-G1] [etiqueta-única]	Cada entidad puede tener solo una etiqueta de tipo referente.
REGLAS MULTICLASE	
[REF-M1] [pct-y-alguien]	Las entidades que puedan anotarse como PACIENTE más otra clase se deben anotar como PACIENTE.
[ejemplos]	
19. Se bañó en las [cataratas]_{PAC} del [País Dogón]_{PAC} junto a su pareja de viaje.	
[REF-M2] [bebé-y-madre]	En algunos casos de pediatría tanto el bebé como la madre son tratados como pacientes. En estos casos, anotaremos ambos como PACIENTE.
[ejemplos]	
20. Un día tras el parto, madre y bebé reciben el alta y vuelven a [su hogar]_{PAC}	

3.3 POLARIDAD Y TEMPORALIDAD

Esta capa de anotación nos proporciona información extra-proposicional sobre las entidades que anotamos, permitiéndonos conocer si han ocurrido o no y cuándo. Marcamos dos aspectos separados: polaridad (tabla F1) y temporalidad (tabla F2).

F1. ATRIBUTOS DE POLARIDAD	
[POSITIVO]	Se usará para las entidades que estén afirmadas. Es la etiqueta por defecto para la mayoría de entidades.
[ejemplos]	
21. Está de residente en anatomía patológica .	
[NEGATIVO]	Se usará para las entidades que estén negadas.
[ejemplos]	
22. No refiere viajes recientes.	
[INCIERTO]	Se usará para las entidades sobre las que se exprese algún tipo de duda sobre si han ocurrido o no.
[ejemplos]	
23. La abuela del paciente no está segura de si ha estado en alguna zona pantanosa .	

F2. ATRIBUTOS DE TEMPORALIDAD

[PRESENTE]

Se usará para las entidades que estén en presente. Es la etiqueta por defecto para la mayoría de entidades.

[ejemplos]

24. Está viviendo en una **residencia de ancianos**.

[PASADO]

Se usará para las entidades que estén en pasado.

[ejemplos]

25. Trabajó durante 7 años en las **minas** de **Sudáfrica**.

[FUTURO]

Se usará para las entidades que estén en futuro, incluyendo planes.

[ejemplos]

26. Necesita profilaxis porque viajará a **Papúa**.

[HIPOTÉTICO]

Se usará para las entidades que no ocurran en ningún punto temporal determinado, si no que se mencionan de manera hipotética o figurativa.

[ejemplos]

27. Parece que la crioterapia se utiliza más en **Rusia**.

4. NORMALIZACIÓN

La normalización de las entidades es un paso crucial para el aprovechamiento óptimo de la información extraída de los textos. La división de las entidades geográficas está hecha teniendo en mente este paso, ya que cada tipo debe normalizarse a un recurso diferente. La normalización se hará a la vez que el marcado y clasificación, introduciendo el código correspondiente en la sección de comentarios que proporciona brat para cada entidad.

Para la normalización usaremos tres fuentes distintas. Estas fuentes son: GeoNames, PlusCodes y SNOMED-CT. Cada una de ellas tiene sus características propias que la hacen adecuada para una situación u otra. En las diferentes subsecciones de este apartado se explican las peculiaridades de cada una.

De este modo, cada tipo de entidad debe normalizarse de la siguiente manera:

GI. FUENTES DE NORMALIZACIÓN PARA CADA TIPO DE ENTIDAD	
Entidad	Ontología
GPE-NOM	GeoNames
GPE-GEN	SNOMED-CT
GEO-NOM	GeoNames
GPE-GEN	SNOMED-CT
FAC-NOM	PlusCodes
FAC-GEN	SNOMED-CT
COMUNIDAD	SNOMED-CT
IDIOMA	SNOMED-CT
TRANSPORTE	SNOMED-CT

Al hacer la anotación en brat escribiremos cada código precedido de un prefijo para distinguir de dónde viene el código. Los prefijos posibles son estos:

G2. PREFIJO PARA CADA FUENTE DE NORMALIZACIÓN	
Fuente	Prefijo
GeoNames	GN:
PlusCodes	PC:
SNOMED-CT	SCTID:
ISO	ISO:

Las menciones que no se puedan normalizar de ningún modo se marcarán con el término “**NOCODE**”.

En las secciones siguientes, se especifica el protocolo a seguir para normalizar con cada una de las fuentes.

4.1 NORMALIZACIÓN CON GEONAMES

GeoNames es una base de datos geográfica gratuita con más de 21 millones de entradas. Cada una de las entradas lleva unos metadatos asociados (como la latitud, la longitud, la altitud, la población, ...), además de un código identificador único. Ese código identificador es el que usaremos para normalizar las menciones.

Dentro de GeoNames podremos encontrar desde países y continentes a pueblos y algunos barrios, pasando por accidentes geográficos nombrados. Sin embargo, no encontraremos muchos establecimientos concretos del tipo FACILITY, por lo que intercalamos esta base de datos con PlusCodes (4.2).

GeoNames puede accederse directamente desde brat al seleccionar un trozo de texto⁴ (Figura 2) o entrando en <http://www.geonames.org/>.

⁴ Se debe añadir la siguiente línea al archivo *tools.conf* de brat en el apartado *[search]*:
GeoNames <URL>:http://www.geonames.org/search.html?q=%s&country=

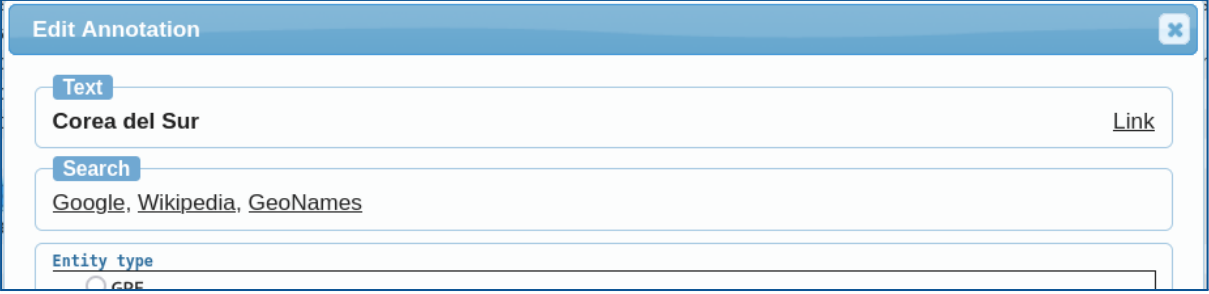


Figura 2. Botón de búsqueda directa en GeoNames en brat

Al hacer una búsqueda en GeoNames, debemos ver las opciones que nos proporciona y seleccionar la ubicación que sea la adecuada. Para ello, haremos clic en el nombre y entraremos en la página de la ubicación. El código es la parte que sale a la derecha con un fondo azul oscuro (Figura 3).




Figura 3. Entrada de una ubicación en GeoNames

La manera más fácil de seleccionarla para copiar y pegar el código es haciendo clic encima de la pestaña Feature, justo encima del nombre de la ubicación. En la tabla que aparece podremos seleccionar el código en el campo *id* con un doble clic para copiarlo y pegarlo en brat.

En la codificación geográfica hay una alta ambigüedad y es común que un mismo nombre se use para referirse a distintos sitios (por ejemplo, Barcelona puede ser el nombre de una provincia y de una ciudad). Por suerte, GeoNames tiene una estructura interna de identificadores que indican la granularidad de cada elemento. Estos identificadores consideran desde un continente, hasta un pequeño pueblo de pocos habitantes, pasando por regiones, comunidades autónomas o provincias. En la Figura 3, el identificador es el código que aparece en negrita bajo el nombre del sitio.

A la hora de elegir códigos, es importante tener en cuenta este identificador para elegir siempre la granularidad adecuada. En la tabla de reglas de codificación (más abajo) hay reglas específicas sobre qué tipo de identificador se debe elegir en caso de ambigüedad.

A modo de referencia, algunos de los identificadores de la jerarquía de GEONAMES más frecuentes son los siguientes:

H1. EJEMPLOS JERARQUÍA GEONAMES		
ID	Nombre	Ejemplo
CONT	continent	Europa (GN:6255148)
RGN	region	Europa del Sur (GN:9408658)
PCLI	independent political entity	España (GN:2510769)
ADM1	first-order administrative division	Comunidad Valenciana (GN:2593113)
PPLA	seat of a first-order administrative division	-
ADM2	second-order administrative division	Valencia [provincia] (GN:2593113)
PPLA2	seat of a second-order administrative division	Valencia [ciudad] (GN:2509954)
ADM3	third-order administrative division	-
PPLA3	seat of a third-order administrative division	-
ADM4	fourth-order administrative division	-
PPLA4	seat of a fourth-order administrative division	-
PPL	Population	Meliana (GN:2513949)

A continuación se presentan las reglas para la codificación con GeoNames:

H2. REGLAS DE CODIFICACIÓN CON GEONAMES

REGLAS GENERALES	
[GN-G1] [granularidad]	<p>A veces puede haber ambigüedad entre varias localizaciones que tienen el mismo nombre y pueden referirse a distintas divisiones administrativas (por ejemplo: Málaga ciudad y Málaga provincia).</p> <p>Si el texto no especifica de cuál se habla, se debe escoger la más general (es decir, la de la región más grande). En la tabla anterior, esto significa que para normalizar “Valencia” deberemos coger el código con el ID ADM2 (provincia).</p>
[GN-G2] [cardinales-1]	<p>Los puntos cardinales dentro de continentes se normalizarán usando el código correspondiente siempre que exista.</p> <p>Sin embargo, no existen códigos en GeoNames para los puntos cardinales dentro de países o ciudades, por lo que usaremos solamente el código de dicho país o ciudad.</p>
[GN-G3] [cardinales-2]	<p>Muchas veces no encontraremos en GeoNames códigos para dos puntos cardinales a la vez (<i>noroeste, sudeste</i>). Estos casos deben normalizarse usando o bien el código para la región norte, o bien el del sur.</p>
[GN-G4] [españa-defecto]	<p>En el caso de que aparezcan topónimos de sitios que existen tanto en España como en algún otro lugar del mundo (por ejemplo: existe una ciudad llamada Córdoba tanto en España como en Argentina y en Colombia), en este corpus escogeremos por defecto la ubicación en España si no se especifica en el texto cuál de todas es.</p>

4.2 NORMALIZACIÓN CON PLUSCODES

PlusCodes, también conocido como OpenLocationCodes, es un sistema de códigos para identificar ubicaciones concretas en la Tierra con bastante precisión. Para crear estos códigos, se divide el mundo en pequeñas cuadrículas, cada una

con un código. Cada código se compone de una serie de letras y números, calculados a partir de la latitud y longitud, que representan una ubicación específica en la superficie de la Tierra.

El sistema de PlusCodes fue creado por investigadores de Google y los códigos pueden ser utilizados en diferentes servicios y aplicaciones de mapas y GPS. En nuestro corpus usaremos este sistema para normalizar las menciones de edificios y *facilities* concretas referenciadas en el texto (es decir, las entidades de tipo FAC_NOM) que no aparecen en otras ontologías.

Para la normalización necesitaremos la versión larga de los PlusCodes. Este código está disponible al realizar búsquedas en el mapa que ofrece la web de PlusCodes (plus.codes/map). Esta web a veces nos ofrecerá la versión corta del código en la parte de abajo, debemos desplegar el menú de la izquierda para obtener la versión larga (ver figura de abajo).

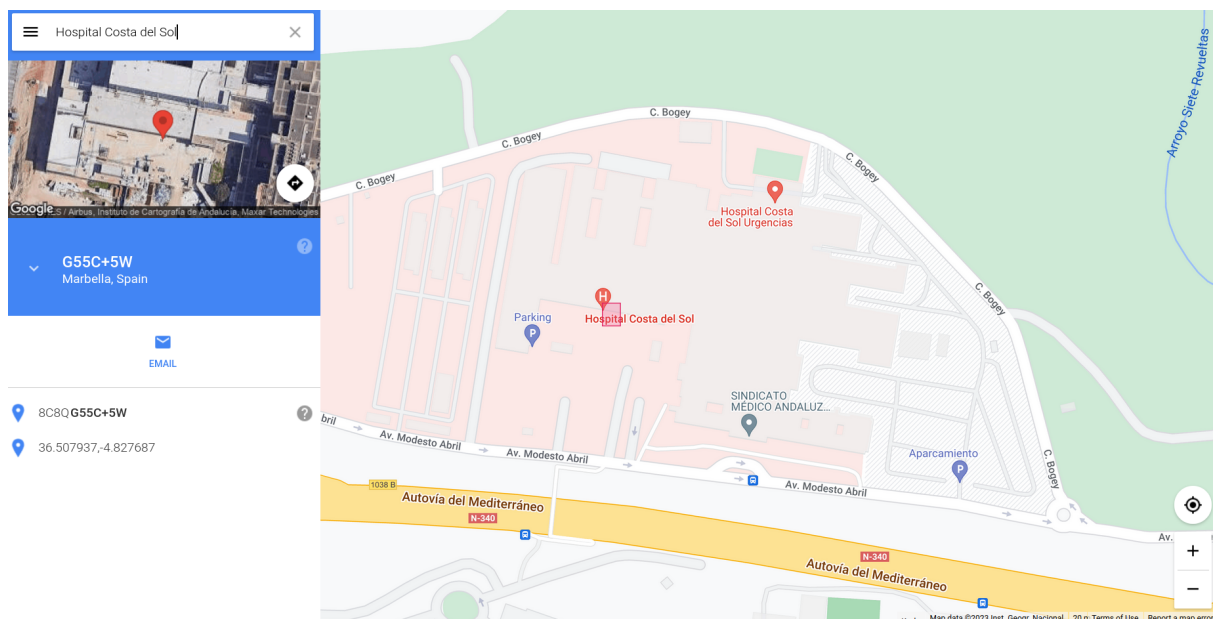


Figura 4. Ejemplo de búsqueda en el mapa de PlusCodes (plus.codes/maps) para el Hospital Costa del Sol en Marbella, Málaga. La versión corta del código es el identificador de cuatro caracteres + dos caracteres en la parte azul. El código largo, el cual debemos usar, es el que se muestra en la parte blanca encima de las coordenadas. Este código es igual que el corto pero con cuatro caracteres extra al principio que ayudan a situar el lugar en la cuadrícula.

Otra alternativa para encontrar el PlusCode de un sitio es hacer una búsqueda en Google Maps y seleccionar la ubicación que corresponda. Dentro de la ventana de

información que se abrirá a la izquierda podremos ver varios datos sobre dicha ubicación, entre ellos el PlusCode (ver Figura 4).

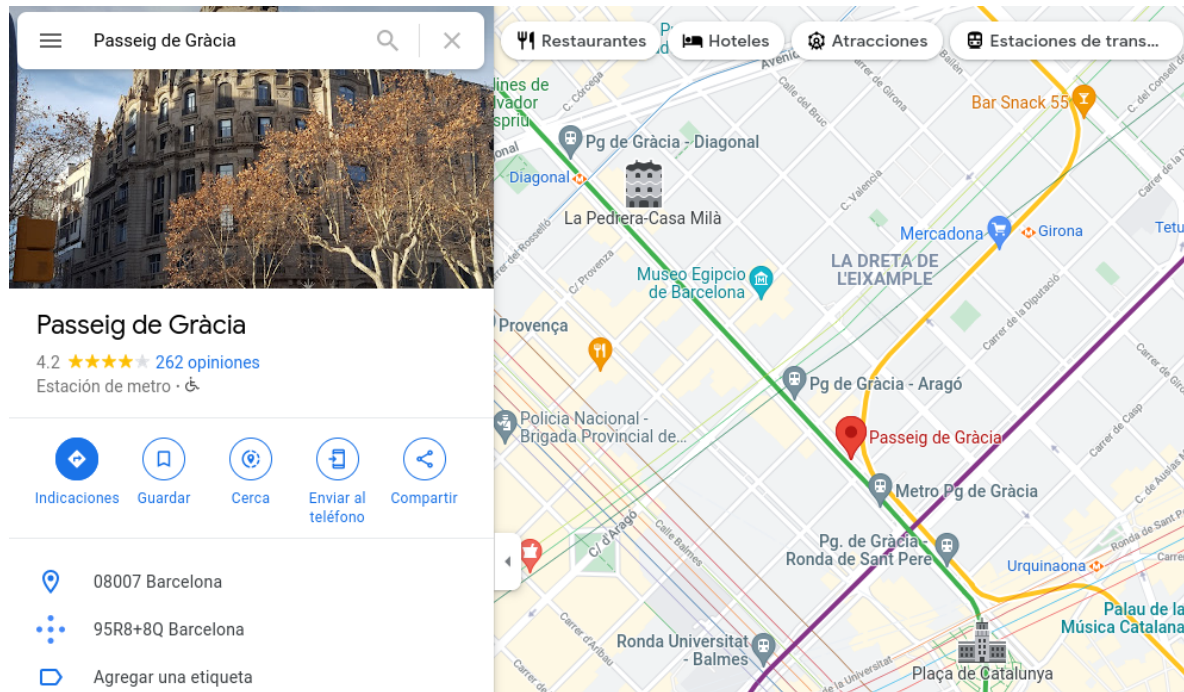


Figura 5. Ejemplo de búsqueda en Google Maps para la estación de metro Passeig de Gràcia en Barcelona. El PlusCode aparece en la línea precedida por un símbolo hecho por cinco puntos. En este caso sería: 95R8+8Q Barcelona.

El código proporcionado por Google Maps se trata de una versión corta acompañada del nombre de una localidad o ciudad. Este código no nos sirve para la normalización (ya que para convertirlo a la versión larga de manera automática hay que usar la API de Geocoding de Google que es de pago), pero podemos usar plus.codes/map para extraer el código largo introduciendo el PlusCode corto con ubicación en el buscador. Debido a este paso extra, recomendamos usar Google Maps solo cuando tengamos problemas para encontrar algo en el mapa de PlusCodes.

El procesamiento posterior de los PlusCodes se puede hacer con la librería oficial de OpenLocationCodes: <https://github.com/google/open-location-code>.

4.3 NORMALIZACIÓN CON SNOMED-CT

SNOMED-CT es una terminología clínica multilingüe y exhaustiva para codificar múltiples elementos de la historia clínica. Entre ellos, encontramos menciones a servicios clínicos, grupos étnicos y medios de transporte.

Para explorar SNOMED usaremos el navegador online que pone a disposición la ontología⁵. En MEDDOPLACE, se ha revisado que todos los códigos usados pertenezcan a la versión en castellano (Spanish Edition) del 31-10-2022.

Para normalizar con SNOMED usaremos el código único de cada entrada. Para encontrarlo, debemos buscar el concepto en el navegador de SNOMED, hacer clic en la entrada adecuada y poniendo el cursor sobre la flecha en el recuadro azul, clicar en “copy concept id” (en amarillo en el ejemplo, ver Figura 5).

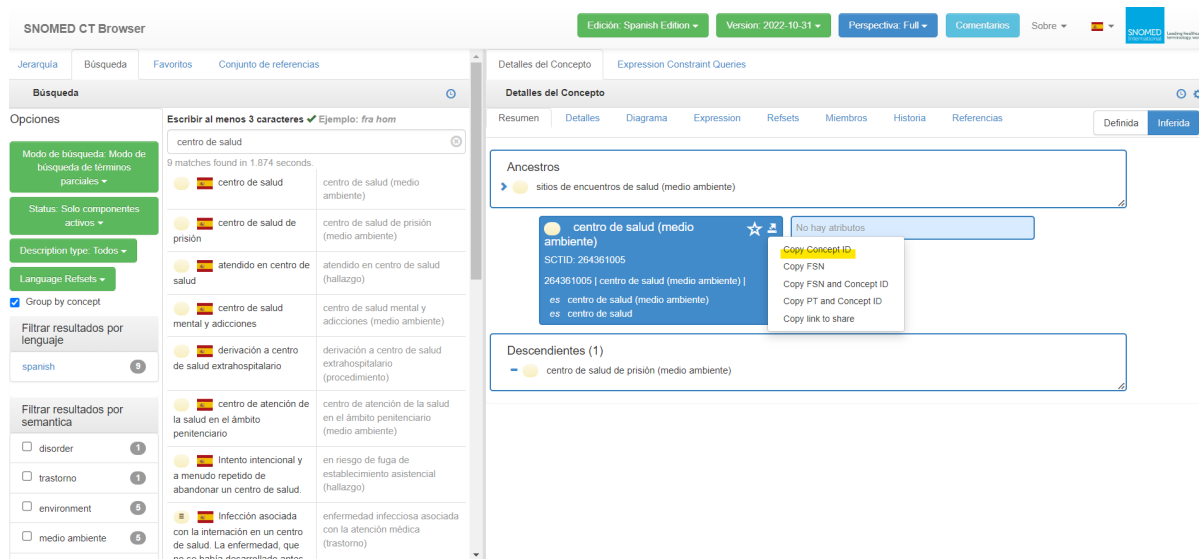


Figura 6. Ejemplo de búsqueda en SNOMED-CT. A la izquierda, los resultados de búsqueda. A la derecha, los detalles de cada entrada.

A continuación se listan las reglas y convenciones definidas para la normalización con SNOMED:

II. CÓDIGOS PARA GPE-GEN	
<p>[1] [país de origen, nacimiento, residencia]</p>	<p>Para menciones como “países de <u>origen</u>” usaremos el código para “país de origen (entidad observable)” (SCTID: 276205001).</p>

⁵ Disponible en:

<https://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=MAIN/SNOMEDCT-ES/2022-10-31&release=&languages=es,en>

II. CÓDIGOS PARA GPE-GEN

	<p>Por otro lado, para “países de <u>residencia</u>” usaremos el código para “país de residencia (entidad observable)” (SCTID: 416647007).</p> <p>Para “país de <u>nacimiento</u>” o “país natal” se usará el código para “país de nacimiento (entidad observable)” (SCTID: 315354004)</p>
<p>[2] [divisiones administrativas]</p>	<p>Debido a la falta de códigos, para menciones de divisiones administrativas como “distrito” o “provincia” usaremos siempre el código para “región (atributo)” (SCTID: 410673009).</p>

II. CÓDIGOS PARA GEO-GEN

<p>[1] [hábitats]</p>	<p>Definimos hábitats a las menciones que hablan de zonas/áreas/regiones más un adjetivo descriptivo. Son las menciones que describe la regla GEO-E1 de la tabla C2.</p> <p>Para estas menciones usaremos los códigos de la rama <i>medio ambiente</i>.</p>
<p>[2] [zonas endémicas]</p>	<p>El código asignado a menciones a regiones endémicas para enfermedades específicas (paludismo, Chagas, tripanosomiasis, etc.) será</p> <p>224779005 medio ambiente de alto riesgo (medio ambiente) </p>

III. CÓDIGOS PARA FAC-GEN

<p>[1] [atención a la]</p>	<p>Cuando no se encuentre un código idéntico o la mención se encuentre aislada utilizaremos los siguientes códigos:</p>
--	---

<p>salud]</p>	<ol style="list-style-type: none"> 1. para los centros de tipo hospitalario y cuando la palabra clínica aparezca sin más especificaciones, usaremos el código "hospital (medio ambiente)" (SCTID: 22232009). 2. "hospital (medio ambiente)" (SCTID: 22232009) también se utilizará para las menciones aisladas "sala" y "planta" cuando no se acompañen de adjetivos más específicos, y siempre que se refieran a una sala o planta situada en un establecimiento hospitalario. 3. para otros tipos de centros médicos usaremos el código "centro de salud (medio ambiente)" (SCTID: 264361005). Atención, la mención <i>centro</i> con frecuencia se refiere a centros educativos, que se normalizan con sus códigos correspondientes. 4. ambulatorio se normalizará como 35971002 centro médico ambulatorio (medio ambiente) 5. unidad: 43741000 unidad de atención médica (medio ambiente) 6. servicio: 224891009 servicios de atención de la salud (calificador)
<p>[2] [pensiones y albergues]</p>	<p>Usaremos el mismo código para pensiones y albergues, ya que no existe código para el primero: "albergue (medio ambiente)" (SCTID: 257638006).</p>
<p>[3] [negocios]</p>	<p>Para tiendas y negocios, intentaremos usar los códigos hijos de "medio ambiente de comercio y servicios (medio ambiente)" (SCTID: 284438004).</p> <p>En el caso de que no haya código para un tipo de negocio específico, usaremos por defecto el código "negocio (medio ambiente)" (SCTID: 257701001).</p>
<p>[4] [cronicidad]</p>	<p>Las menciones a sanatorios, ingresos en hospitales para enfermos crónicos e ingresos a largo plazo se normalizan con 32074000 hospital de atención prolongada (medio ambiente) </p>

<p>[5] [hospital materno infantil]</p>	<p>Los hospitales materno infantiles se codifican con 82242000 hospital de niños (medio ambiente) </p> <p>Hospitales y servicios maternos/obstétricos se normalizan con 52668009 maternidad hospitalaria (medio ambiente) </p>
<p>[6] [institución académica]</p>	<p>Las menciones a universidades e instituciones de ámbito académico biomédico las normalizamos como</p> <p>405608006 centro médico académico (medio ambiente) </p>
<p>[7] [servicios sociales]</p>	<p>Normalizamos con el código 310134006 servicios sociales (calificador) a los servicios sociales gubernamentales y también a las organizaciones cívicas para la mejora de la salud y la condición social de pacientes y personas de toda clase. Por ejemplo: <i>Cáritas, Instituto de Mayores y Servicios Sociales, Asociación de Ayuda a Ludópatas.</i></p>

14. CÓDIGOS PARA DEPARTAMENTO	
<p>[1] [consultas]</p>	<p>La mención “consulta” se normalizará con el código 83891005 consultorio particular (medio ambiente) , que incluye entre los sinónimos <i>consultorio médico</i>.</p>
<p>[2] [especialidades]</p>	<p>Los consultorios de especialidades se normalizan con la rama que parte de <i>consultorio especializado en...</i>Ejemplo:</p> <p>Ancestros</p> <p>> ● consultorio especializado (medio ambiente)</p> <div data-bbox="587 1659 1043 1921"> <p>● consultorio especializado en asma (medio ambiente) ☆</p> <p>SCTID: 702823004</p> <p>702823004 consultorio especializado en asma (medio ambiente) </p> <p>es consultorio especializado en asma</p> <p>es consultorio especializado en asma (medio ambiente)</p> </div>

<p>[3] [servicio salud]</p>	<p>Conceptos generales tipo “servicio de salud”, “servicios clínicos”, “servicios sanitarios” se normalizan a:</p> <p>43741000 unidad de atención médica (medio ambiente) </p>
<p>[4] [servicio vs departamento]</p>	<p>Para los departamentos y servicios hospitalarios, así como para los hospitales de especialidades, intentaremos usar siempre que sea posible el código más específico de la categoría <i>medio ambiente</i>.</p> <p>Por ejemplo, en la mención servicio de pediatría encontramos (1) 310066004 servicio de pediatría (calificador) y (2) 309945009 departamento de pediatría (medio ambiente) .</p> <p>En la mención Hospital Cardiológico, utilizamos 309915006 servicio de cardiología (medio ambiente) </p> <p>Cuando no se encuentre equivalente en la categoría medio ambiente se escogerá la única que nos ofrezca SNOMED, por ejemplo, en la mención psiquiatría infanto-juvenil sólo existe 310117003 servicio de psiquiatría de niños y adolescentes (calificador) .</p> <p>Para especialidades en las que no se especifica si es un departamento, servicio, equipo, ... (ej: “Cardiología”, “Digestivo”) usaremos siempre el código para departamento.</p>
<p>[5] [atención primaria]</p>	<p>Las menciones a servicios de atención primaria de salud (médico de cabecera, medicina comunitaria, médico de familia) se normalizarán como:</p> <p>702855004 consultorio especializado en medicina familiar (medio ambiente) </p> <p>Cuando se mencione <i>ambulatorio</i>, se normalizará como</p> <p>35971002 centro médico ambulatorio (medio ambiente) </p>

<p>[6] [estancia intermedia]</p>	<p>Las menciones a ingresos en <i>media estancia</i> o <i>estancia intermedia</i> se normalizan con el código genérico de hospital</p> <p>22232009 hospital (medio ambiente) </p>
<p>[7] [asociación]</p>	<p>Las menciones a asociaciones, fundaciones etc. NO se normalizan como departamentos clínicos si no como</p> <p>310134006 servicios sociales (calificador) </p>
<p>[8] [desintoxicación drogas y OH]</p>	<p>Las menciones a centros de deshabituación de drogas, como las Comunidades Terapéuticas, se normalizan con el código</p> <p>20078004 centro de tratamiento contra el abuso de sustancias (medio ambiente) </p> <p>Si es un centro exclusivo de tratamiento de alcoholismo, utilizaremos 274517002 centro de desintoxicación de alcohol (medio ambiente) </p>
<p>[9] [psiquiatría]</p>	<p>En menciones de consultas de psiquiatría utilizamos:</p> <p>702833007 consultorio especializado en psiquiatría infantojuvenil (medio ambiente) </p> <p>702914003 consultorio especializado en psiquiatría (medio ambiente) </p> <p>En casos de urgencias e ingresos en servicios de psiquiatría utilizamos:</p> <p>309959002 servicio de psiquiatría del niño y adolescente (medio ambiente) </p> <p>309958005 departamento de psiquiatría (medio ambiente) </p>

15. CÓDIGOS PARA TRANSPORTE

[1] [viajes]	Todos los tipos de viajes se normalizan siempre usando el código para “viaje (evento)” (SCTID: 420008001).
[2] [vacaciones]	Cuando aparezca el término vacaciones, utilizaremos también el código “viaje (evento)” (SCTID: 420008001).
[3] [cruceros]	Debido a que no existe un código específico, para los cruceros usaremos el código 262044003 barco de pasajeros (objeto físico) .
[4] [pateras]	Debido a que no existe ningún código para este concepto, usaremos el código para “balsa (objeto físico)” (SCTID: 86118007).
[5] [ambulancia medicalizada]	Las referencias a servicios de ambulancia medicalizadas (con soporte vital avanzado también) se codifican con: 11424001 atención médica en ambulancia (medio ambiente)

16. CÓDIGOS PARA COMUNIDAD

[1] [gentilicio, origen o lugar de nacimiento]	<p>Las anotaciones con la etiqueta COMUNIDAD pueden hacer referencia a gentilicios, países de origen o lugar de nacimiento. Seguiremos la siguiente jerarquía:</p> <ol style="list-style-type: none"> 1. “europeo” y cualquier otro gentilicio se anotará de manera preferente como “europeo (grupo étnico)” (SCTID: 414152003) 2. En el caso de no encontrar el código dependiente de la rama “grupo étnico”, como en el caso de “estadounidense”, utilizaremos 160516002 origen norteamericano (hallazgo) 3. Si no se encuentra ninguno de los anteriores, podemos también utilizar códigos de localización geográfica, por ejemplo, si la mención es <i>magrebí</i>,
---	--



16. CÓDIGOS PARA COMUNIDAD

	<p>223509000 país norafricano (localización geográfica) ; si es <i>camerunés</i>, 223521001 Camerún (localización geográfica) .</p> <p>4. Utilizaremos “nacido en” cuando sea una certeza: 315425009 nacido en Fidji (hallazgo) </p>
<p>[2] [grupo religioso]</p>	<p>Los grupos religiosos anotados con la etiqueta comunidad recibirán el código de la rama “persona”:</p> <p>1. “judío” se normalizará como “judío (persona)” (SCTID: 160543007)</p>

ANEXO 1. Tabla Resumen de Reglas de Anotación

En este anexo se incluye un resumen de las reglas de anotación de entidades expuestas en el apartado 2 de las guías. Esta sección pretende ser una referencia para las personas anotadoras y no un reemplazo de las reglas como tal. Recomendamos leer en primer lugar el apartado 2, ya que se incluyen muchos ejemplos, así como explicaciones detalladas de casos especiales y excepciones.

Etiqueta	Descripción	Reglas	Ejemplo
Reglas Generales			
Todas	Reglas comunes para todas las etiquetas.	[GEN-P1] Ortografía	Paciente natural de [nicaragua]
		[GEN-P2] Separación	Ha estado viajando por Europa (Noruega, Dinamarca y Alemania)
		[GEN-P3] Anidación	Había estado en [[América Central] y del Sur]
		[GEN-P4] Siglas	Es llevada por la policía al servicio de urgencias (SU)
		[GEN-P5] Adjetivos-prep	[...] de la costa litoral central venezolana
		[GEN-P6] Compl-nombre	Médico de urgencias

		[GEN-N1] Det-prep	Recién llegada de un viaje a Tanzania
		[GEN-N2] Sitios-genéricos	Su centro laboral se encuentra próximo a un río, lugar donde lava sus herramientas
		[GEN-N3] Adv-dem	Una vez llegó allí estuvo en contacto con varios animales de granja
		[GEN-N4] Eventos	No encontraron pruebas en el lugar del accidente
		[GEN-N5] Sitios-web	Conoció a su actual pareja en Twitter
		[GEN-N6] Especies-proced	La paciente presentaba un Glasgow de 15
Reglas para ubicaciones			
GPE	Se anotan con esta etiqueta regiones geográficas habitadas por personas y que están influenciadas por la política, la economía, la cultura y la historia de su entorno. Esto incluye países, regiones, ciudades, ...	[GPE-P1] Ejemplos	Niña de seis años, natural y procedente de Pozuzo, departamento de Pasco (selva central) , admitida en un hospital de la ciudad de Lima en agosto de 2006
		[GPE-P2] Divisiones-1	Natural de la provincia de Punjab

		[GPE-P3] Divisiones-2	Este agosto planea hacer un recorrido por varias [capitales asiáticas]
		[GPE-P4] Barrios	Al llegar a Sevilla se alojó en un piso compartido de un barrio marginal .
		[GPE-E1] Extranjero	Hace quince años que viajó por última vez al extranjero
		[GPE-E2] Pos-más-país	Se presenta el primer caso en nuestro país de coronavirus
		[GPE-E3] Sprep-demo	Fue operado en su país de residencia
		[GPE-E4] Coordenadas	Procedente de la Sierra de Parima (2° 47' 21" N, 64° 13' 55" W).
GEO	Se anotan con esta etiqueta los accidentes geográficos, sitios naturales en general, tipos de ecosistemas y ambientes, zonas del mundo, ...	[GEO-P1] Ejemplos	Trabajador de la construcción, con exposición habitual al asbesto y durante 3 años en una mina de carbón
		[GEO-P2] Acc-nom	Este verano se bañó en el mar Mediterráneo
		[GEO-P3] Cardinales	El gazpacho es un plato típico del sur

		[GEO-P4] Adjetivos-prep	Es dueña de una casa en el pedemonte de la cordillera de la costa litoral central venezolana
		[GEO-P5] Hábitats	Trabajaba en una región selvática del país
		[GEO-P6] Endémicas	Estuvo de visita en una región endémica de babesiosis
		[GEO-N1] Celestes	El paciente presenta delirios en los que cree que trabaja como astronauta y vive con su familia en Neptuno
		[GEO-N2] Aguas	No refería conductas de riesgo específicas durante el viaje salvo baños en agua dulce en Bamako
		[GEO-E1] Lugar-consumo	Se prescribe control de los factores de riesgo para evitar recaídas (control del manejo de dinero, evitación de amistades consumidoras y lugares de consumo , evitar el estar y salir sólo).
FAC	Se anotan con esta etiqueta espacios contruidos por humanos, lo cual incluye muchos sitios comunes que encontraríamos	[FAC-P1] Ejemplos	Refiere haber trabajado en una tienda de bolsos cuando tenía 16

	en una ciudad: establecimientos, tiendas, estructuras, monumentos, ... así como distintos tipos de centros de salud.		años durante un año, en una pescadería y en una gasolinera
		[FAC-P2] Tipo-edificio	Trabaja como camarero en el restaurante La Barca
		[FAC-P3] Pisos	Disminuye la tendencia a aislarse en su dormitorio , pasando la mayor parte del día en el salón común
		[FAC-P4] Nombres-hospitales	Realiza seguimiento en el [Hospital Clínic de [Barcelona]] Se remitió para seguimiento a las consultas externas de su hospital de origen (Alto Guadalquivir de Andújar) .
		[FAC-P5] Residencia	Al alta, volvió a su domicilio
		[FAC-N1] Organizaciones	Se traslada al medio residencial de FAISEM (Fundación Pública Andaluza para la Integración Social de Personas con Enfermedad Mental) ,
		[FAC-E1] Ambulatorio	Es dada de alta para seguimiento ambulatorio

Reglas para etiquetas relacionadas con lugares			
Comunidad	Se anotan los gentilicios, es decir, adjetivos o sustantivos que hacen referencia a un lugar, así como etnias, religiones y elementos sociodemográficos del estilo	[COMU-P1] Nombre-adj	Hombre español de 17 años de edad que viene a consulta por...
		[COMU-P2] Anidados	Pasó su infancia en una [ciudad [africana]]
		[COMU-E1] Extranjero	La paciente es de nacionalidad extranjera aunque desconocemos exactamente de dónde
		[COMU-E2] Religiones	Paciente hindú de 21 años...
Idioma	Esta categoría se usa para los idiomas que se mencionan en el texto.	[IDI-P1] Idioma	El paciente no habla castellano y no conseguimos localizar al traductor
		[IDI-P2] De-habla	Se trata de algo muy común en los países de habla inglesa
		[IDI-E1] Anglosajón	Paciente anglosajón de 76 años residente en España desde hace 22.
		[IDI-E2] Barrera-idiomática	No entra en el estudio por motivos de barrera idiomática
Departamento	Entran dentro de esta categoría los diferentes departamentos y servicios de instituciones de salud (por ejemplo: Dermatología,	[DEPA-P1] Departamentos	Paciente vista por primera vez en el servicio de Endocrinología Pediátrica del Hospital

	Cardiología), así como las distintas partes dentro de hospitales y centros de salud.		Materno-Infantil de Málaga , a los 8 años y 6 meses de edad.
		[DEPA-P2] Especialidades	Estudian el caso Medicina Interna y Neurología y no encuentran datos patológicos justificativos en las pruebas complementarias
		[DEPA-P3] Salas	Dos semanas después de su ingreso en la planta de Aparato Digestivo , el cuadro se complicó con una hemorragia digestiva grave El paciente se mantuvo con soporte ventilatorio en el área de observación de urgencias a la espera de un posible fallecimiento o donación de órganos
		[DEPA-P4] Consulta	Un año después, el paciente retorna a consulta por presentar cólicos abdominales de fuerte intensidad
		[DEPA-P5] Equipos	Es seguida desde febrero de 2012 por el Equipo de Atención Primaria (EAP) para control de síntomas
		[DEPA-N1] Centros	Fue tratada en una clínica de salud

			mental y adicciones
		[DEPA-N2] Pruebas	Pruebas realizadas de urgencia a su llegada
		[DEPA-E1] Específicos	Ingresa directamente en quirófano para extracción de la bala
		[DEPA-E2] Servicios	Ante la reaparición de sintomatología consultó en nuestro Servicio
Transporte	Entran dentro de esta categoría las menciones a transportes y movimientos en general, incluyendo la palabra “viaje”, menciones a movimientos de pacientes y medios de transporte, entre otros.	[TRA-P1] Viajes	Tenía planeado un viaje por Europa del Este con su familia
		[TRA-P2] Tipos-viajes	Refería haber realizado múltiples viajes intercontinentales , sin problemas de salud aparentes
		[TRA-P3] Medios	Llegada a España en patera a Canarias
		[TRA-N1] Verbos	Hace tres años voló de Marruecos a España
		[TRA-N2] Viaje-figurado	Acabó en Urgencias tras un mal viaje de ácido
		[TRA-E1] Silla-ruedas	Precisa silla de ruedas para acudir a

			la Unidad
		[TRA-E2] Traslado	En la última consulta, al palpar el abdomen su médico de atención primaria valoró el traslado a servicio de Urgencias del hospital
		[TRA-E3] Visita	[...] facilitada por el impulso de la visita telemática instaurada durante la pandemia de COVID-19

BIBLIOGRAFÍA

- Dipper, S., Lüdeling, A., & Reznicek, M. (2013): "NoSta-D: A Corpus of German Non-Standard Varieties". In: Zampieri, Marcos & Diwersy, Sascha (Hrsg.): Non-standard Data Sources in Corpus-based Research. Aachen: Shaker Verlag. S. 69 – 76
- Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., & Weischedel, R. M. (2004). The automatic content extraction (ACE) program-tasks, data, and evaluation. In International conference on language resources and evaluation (vol. 2, p. 1).
- Gey, F., Larson, R., Sanderson, M., Joho, H., Clough, P., Petras, V. (2006). GeoCLEF: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview. In: , et al. Accessing Multilingual Information Repositories. CLEF 2005. Lecture Notes in Computer Science, vol 4022. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11878773_101
- Gritta, M., Pilehvar, M. T., Limsopatham, N., & Collier, N. (2017). What's missing in geographical parsing? *Language Resource Evaluation*, 52, 603–623.
- Gritta, M., Pilehvar, M. T., & Collier, N. (2020). A pragmatic guide to geoparsing evaluation: Toponyms, Named Entity Recognition and pragmatics. *Language resources and evaluation*, 54, 683-712.
- Hirschman, L. (1998). The evolution of evaluation: Lessons from the message understanding conferences. *Computer Speech & Language*, 12(4), 281–305.
- Jørgensen, F., Aasmoe, T., Ruud, A.S, Øvrelid, I., & Velldal, E.. (2020). NorNE: Annotating Named Entities for Norwegian. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4547–4556, Marseille, France. European Language Resources Association
- Lieberman, M. D., Samet, H., & Sankaranarayanan, J. (2010). Geotagging with local lexicons to build indexes for textually-specified spatial data. In 2010 IEEE 26th international conference on data engineering (ICDE 2010) (pp. 201–212). New York: IEEE.
- Mandl, T. et al. (2009). GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In: , et al. Evaluating Systems for Multilingual and Multimodal Information Access. CLEF 2008. Lecture Notes in Computer Science, vol 5706. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04447-2_106
- Sang, K., & Tjong, E. (2002). Introduction to the CONLL-2002 shared task: Language-independent Named Entity Recognition. Technical report. cs/0209010

- Sang, K., & De Meulder, F. (2003). "Introduction to the CONLL-2003 shared task: Language-independent Named Entity Recognition". In *Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003-Volume 4* (pp. 142–147). Stroudsburg: Association for Computational Linguistics
- Ševčíková, M., Žabokrtský, Z., Krůza, O. (2007). Named Entities in Czech: Annotating Data and Developing NE Tagger. In: Matoušek, V., Mautner, P. (eds.) *TSD 2007. LNCS (LNAI)*, vol. 4629, pp. 188–195. Springer, Heidelberg.
- Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S. and Tsujii, J. (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, Avignon, France. Association for Computational Linguistics.
- Ruokolainen, T., Kauppinen, P., Silfverberg, M., & Lindén, K. (2020). A Finnish news corpus for named entity recognition. *Language Resources and Evaluation*, 54(1), 247-272
- Taulé, M., Martí, A., & Recasens, M. (2008). "AnCora: Multilevel Annotated Corpora for Catalan and Spanish". In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Weischedel, R.M., & Brunstein, A. (2005). "BBN Pronoun Coreference and Entity Type Corpus LDC2005T33". Web Download. Philadelphia: Linguistic Data Consortium.
- Weischedel, R.M., Hovy, E.H., Marcus, M.P., & Palmer, M. (2017). "OntoNotes : A Large Training Corpus for Enhanced Processing."
- Weissenbacher, D., Magge, A., O'Connor, K., Scotch, M., & Gonzalez-Hernandez, G.. (2019). SemEval-2019 Task 12: Toponym Resolution in Scientific Papers. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 907–916, Minneapolis, Minnesota, USA. Association for Computational Linguistics
- Zeldes, A. (2017) "The GUM Corpus: Creating Multilayer Resources in the Classroom". *Language Resources and Evaluation* 51(3), 581–612.