



# MEDDOPLACE Guidelines: Annotation, Normalization and Classification of Locations and Place-Related Information in Clinical Texts

V1 [March 2023]

## AUTHORS

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



This study has been carried out within the scope of the *Plan de Impulso de las Tecnologías del Lenguaje* with funding from the Secretaría de Estado para el Avance Digital, which does not necessarily share the contents expressed therein. These contents are the sole responsibility of their authors. All rights reserved. Copying and distribution by any means is permitted provided that the authors are acknowledged, that no commercial use is made of the works and that no modification is made to them.

## INDEX

---

### 1. INTRODUCTION

Motivation of the corpus and use cases  
Previous work in annotation and location detection  
The MEDDOPLACE corpus and guidelines

### 2. ANNOTATION OF ENTITIES

#### 2.1 DEFINITION

#### 2.2 ANNOTATION RULES AND EXTENSION

B1. General positive annotation rules	12
B2. General negative annotation rules	14
C1. Rules for "GPE"	16
C2. Rules for "GEO"	19
C3. Rules for "FAC"	22
C4. Rules for "COMMUNITY"	25
C5. Rules for "LANGUAGE"	26
C6. Rules for "DEPARTMENT"	27
C7. Rules for "TRANSPORT"	31

### 3. CLASSIFICATION OF ENTITIES

3.1 Clinical classification	35
3.2 Classification by referents	38
3.3 Polarity and temporality	40

### 4. NORMALIZATION

4.1 Normalization with GeoNames	43
4.2 Normalization with PlusCodes	46
4.3 Normalization with SNOMED-CT	48

### ANNEX 1. Summary Table of Annotation Rules

### BIBLIOGRAPHY

**57**  
**66**

## 1. INTRODUCTION

---

These guidelines describe the annotation of the MEDDOPLACE corpus, a set of 1,000 clinical cases in Spanish, intended for the training and evaluation of automatic systems for the detection, classification and normalization of locations and place-related information (communities, languages, clinical departments and mentions of movements from patients) in medical texts.

### Motivation of the corpus and use cases

In clinical practice, knowing where the patient has been is extremely relevant. For example, an episode of fever is investigated differently if during the previous month we have been in a malaria endemic area, while the probability of tuberculosis in a case of chronic cough is much higher in areas with a high prevalence of *Mycobacterium tuberculosis* infections. In the context of the COVID-19 pandemic, patient movements and locations visited have become even more relevant. Despite this, there are hardly any resources to categorize the locations extracted from the texts according to their clinical relevance.

With this in mind, the creation of MEDDOPLACE has been motivated by the potential use cases of geographic locations in cases and medical records.

One of the main use cases is the **association between a geographical location and health risk factors**. These risk factors can be: communicable diseases (malaria, sleeping sickness, Chagas disease, Dengue fever, tuberculosis, etc.), non-communicable diseases (genetic diseases, obesity, cardiovascular disease) and mental health diseases (violence and war trauma, traumatic migration, human trafficking, etc.). At the individual level, it will be essential to diagnose the patient with malaria in order to administer artesunate, for example. At the public health level, it is essential to detect **areas of emerging diseases** and **epidemic outbreaks**, e.g. *Salmonella enteritidis* infection in people who have eaten in a restaurant, or an outbreak of Marburg virus in Equatorial Guinea.

In the case of detection of departments, services and medical teams, another main use case is **patient routing** through unstructured text. For example, a patient is admitted through the *Emergency Department* referred by *Primary Care*, transferred to the *Cardiology Department*, develops a nosocomial infection,

requires admission to the *ICU*, then goes to the *Internal Medicine Department*, to be finally discharged. The analysis of these routes is useful for quality and iatrogenic control of medical care. Locations also warn us about the severity of the patient.

## Preliminary work on annotation and location detection

Location detection could be considered one of the classic tasks of natural language processing. In the Message Understanding Conferences (MUC) (Hirschman, 1998) of the 1990s, one of the pioneers in entity recognition, it was already defined as one of the three key types of entities to be recognized. Three categories were defined in these conferences: PERSON, ORGANIZATION and LOCATION. The LOCATION category has a very broad scope, including within it proper and common names of all kinds of places: cities, countries, rivers, mountains, airports, monuments, etc. Two extra categories are also considered: dates and numbers.

This annotation scheme was maintained in the datasets of the CoNLL 2002 (Sang and Tjong, 2002) and 2003 (Sang and De Meulder, 2003) conferences, where an extra category called MISC was added, which includes information such as demonyms or certain types of events. In 2002, a dataset was presented in Spanish and another in Dutch, and in 2003 one in English and another in German. All these datasets were composed of news texts. Due to the impact of the conference and the annotated data, these four categories became almost a standard in language processing. Many subsequent corpora, such as the Spanish and Catalan news corpora Ancora (Taulé, Martí and Recasens, 2008), have used the same scheme for general entity annotation. Currently, some popular NLP tools have available pre-trained models generated using the CoNLL corpora. For example, the StanfordNER tool<sup>1</sup> uses the 2003 edition data for its English version and spaCy<sup>2</sup> uses the 2002 edition data for its Spanish NER models.

The ACE (Automatic Concept Extraction; Doddington et al., 2004) project at the University of Pennsylvania introduced a more granular classification for entity recognition in a news corpus that included new classes and subclasses of locations. The project began in 2002 and grew progressively until 2008, creating data and annotation guidelines for English, Spanish, Arabic and Chinese. In ACE, locations were divided in a more granular way into three categories: geo-political entities ("geo-political entities" or GPE; countries, cities, etc.), facilities ("facilities" or FAC; buildings, airports, roads, etc.) and geographic entities in general ("locations"

---

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

<sup>2</sup> <https://spacy.io/>

or LOC; geographic features, celestial bodies, some locative phrases, etc.). It is important to note that within the GPEs there are different subclasses such as GPE-PER for demonyms or GPE-ORG for mentions referring to governments.

Although the MUC, CoNLL and ACE proposals have probably been the most successful and influential, there are more proposals worth commenting on. For example, the BBN Pronoun Coreference and Entity Type Corpus (Weischedel and Brunstein, 2005), which uses texts from the Wall Street Journal, includes 12 named entity types (Person, Facility, Organization, GPE, Location, Nationality, Product, Event, Work of Art, Law, Language and Contact-Info). At the location detection level, the BBN maintains the main geographic entity types and separates nationalities or demonyms into separate labels. It also includes annotations for "named entities" (separate from those mentioned using proper names) and numerical entities. All these entities in turn have subtypes, making a total of 64 possible labels. This corpus is the one used by the spaCy and Stanza tools for some of their English NER models and the basis for larger resources such as OntoNotes (Weischedel et al., 2017).

There are many more corpora in multiple languages that can be used to train location detection systems, although, according to our literature review, none of them have a clinical focus. Some of these corpora are the English collaborative corpus GUM (George University Multilayer; Zeldes, 2017), the German corpus Nosta-D (Dipper, Lüdeling, & Reznicek, 2013), the Norwegian corpus NorNE (Jørgensen et al., 2020), the Czech-language corpus CNEC (Ševčíková, Žabokrtský and Krůza, 2007) or the Finnish-language corpus FiNER (Ruokolainen et al., 2020).

On the other hand, for the use of the locations found in the text by other systems, detection is not enough. A subsequent step of normalization of the mentions found is necessary, either to a set of coordinates or to an identifier of some ontology, in order to be able to place them on a map. This is the objective of the toponym resolution task (also called *geoparsing* or *geocoding*). This task is usually part of *Geographic Information Retrieval* (GIR) systems or methods.

Some resources that stand out in this area are the multiple GeoCLEF tasks between 2005 (Gey et al., 2006) and 2008 (Mandl et al., 2009), aiming to evaluate GIR systems in news texts in different languages, or SemEval 2019 task 12 (Weissenbacher et al., 2019), in which participants were asked to recognize toponyms in English-language PubMed scientific articles and normalize them to GeoNames.

Beyond evaluation tasks, there are open published corpora such as the collection of news items normalized with GeoNames Local Global Lexicon (Lieberman et al., 2010), the WikiNews GeoVirus annotated disease news corpus (Gritta et al., 2018) or the GeoWebNews corpus (Gritta et al., 2020), which includes granular annotation for different types of toponyms and normalizes not only geopolitical entities but also some buildings and street names. We recommend reading Gritta et al. (2020) for an overview of general place name resolution and existing resources.

## The MEDDOPLACE corpus and guidelines

The aim of the corpus is the detection, normalization and classification of all the locations mentioned in the text, as well as four types of place-related information: clinical departments, mentions of patient movements, communities and languages. As a summary, Figure 1 shows the different possible options in brat (Stenetorp et al., 2012), the tool used to do the corpus annotation.

For textual annotation, we have 10 tags. These textual annotation categories are briefly explained below, with more details, examples and rules for each in section 2.

On the one hand, for places and locations we define six granular labels: GPE\_NOM, GPE\_GEN, GEO\_NOM, GEO\_GEN, FAC\_NOM and FAC\_GEN. Two resources explained in the previous section, ACE (Doddington et al., 2004) and the BBN corpus (Weischedel et al., 2017) inspire this splitting. From the former we take the granular division of places into three categories: geopolitical entities (GPE), facilities, constructions and buildings (FAC), and natural geographical features and other types of places (LOC, which we rename to GEO here). From the second we take the division between common names (GEN) and proper names (NOM), which allows us to be exhaustive in the annotation and to separate the annotation into two axes: the type of location and the type of name.

On the other hand, due to the clinical focus of the corpus, after inspecting some documents we decided to add four complementary place-related labels to the annotation:

- DEPARTMENT collects all mentions of specific sites within hospitals (*operating room*), as well as services (*emergency service, cardiology*), units (*ICU*), equipment and other clinical elements that represent the patient's journey through the healthcare system. It is a very frequent and varied label, as well as very specific and relevant to the clinical domain.

- TRANSPORT includes mentions related to movements of the patient (and other persons appearing in the text), both outside and inside the health care setting. For example, travel, ambulance transport, flying in an airplane, moving by car, etc.
- COMMUNITY collects sociodemographic information that is often (but not always) related to a person's place of origin or residence, including ethnicities, religions and ethnicities.
- LANGUAGE includes mentions of the languages mentioned in the text, as well as linguistic problems (language barriers).

The entities annotated in the corpus are characterized in a classification step to maximize their exploitability from a clinical point of view. All entities are classified according to their polarity and temporality, while locations are additionally labelled to reflect their possible clinical relevance, such as whether they refer to the patient's residence or to a site where they have received medical care. The classification also has its own rules, which are described in section 3.

The last layer of annotation is the normalization of the annotated entities. Due to the diversity of content, we use three different sources for normalization: GeoNames, PlusCodes and SNOMED. Each of these sources has its own section in section 4 of the guidelines, with explanations on how they work, rules and examples.

In short, this guide is structured as follows: section 2 explains the tags and rules for textual annotation of the corpus; section 3 explains how to classify the annotated entities; section 4 describes the process, ontologies and rules for entity normalization. A summary table of the annotation rules is included in Annex I for quick review during the annotation process.

With respect to the color coding in this guide, the examples illustrating the various rules are marked in **blue** for what is correct to annotate and in **red** for what is not. Mentions in **black** are examples that should be annotated, but are not related to the category or rule being described.

Edit Annotation
✕

Text

**Andalucía** Link

Search

Google, Wikipedia, GeoNames

Entity type

- GPE
- GPE\_NOM**
- GPE\_GEN
- GEO
- GEO\_NOM
- GEO\_GEN
- FAC
- FAC\_NOM
- FAC\_GEN
- DEMO
- GENTILICIO
- IDIOMA
- OTROS
- DEPARTAMENTO
- TRANSPORTE

Entity attributes

LUGAR-NATAL
  RESIDENCIA
  MOVIMIENTO
  ATENCION
  OTROS
 Referente: PACIENTE

Polaridad: POSITIVO
 Temporalidad: PRESENTE

Notes

GN:2593109 ✕

Add Frag.
Delete
Move
↶
OK
Cancel

Figure 1. Brat annotation window with all annotation categories and classification. Normalization is presented as a comment for each annotation. The brat configuration files are distributed together with the corpus.



## 2. ANNOTATION OF ENTITIES

This section describes the entities to be marked in the text, as well as the rules to be followed for their labeling. A general summary of this section is included in Annex I at the end of these guidelines.

### 2.1 DEFINITION

The following table describes the various labels along with examples for each:

A. DEFINITION OF POSSIBLE LABELS	
LOCATION LABELS	
<b>[GPE]</b>	<p>It is used for so-called "<i>GeoPolitical Entities</i>", i.e., geographic regions defined politically or socially. This includes continents, countries, cities, towns, villages, islands and places populated by humans in general.</p> <p>We distinguish two subtypes:</p> <ol style="list-style-type: none"> <li>1. <b>GPE-NOM</b><sup>3</sup> : mentions referring to a specific location with its own name (<i>Europe</i>).</li> <li>2. <b>GPE-GEN</b>: mentions referring to a generic location using a common noun (<i>continent</i>).</li> </ol>
[examples].	
<ol style="list-style-type: none"> <li>1. (GPE-NOM) She comes from the <b>northeastern region of Peru</b>.</li> <li>2. (GPE-GEN) Omeprazole is one of the most widely sold drugs in <b>European countries</b>.</li> </ol>	
<b>[GEO]</b>	<p>It is used for <b>natural geographic entities</b>, in contrast to geopolitical and administrative entities (GPE). Included</p>

<sup>3</sup> Initially, in the guide we distinguished between GPE-NOM with locative and non-locative sense (i.e. Barcelona as a city vs. Barcelona as a soccer team). Although definitely relevant, we decided to eliminate this distinction because we hardly found any non-locative mentions in the corpus.



are: geographic features in general (rivers, seas, mountains, forests, etc.), habitats (hot, cold, rural areas, endemic disease areas...).

We distinguish two subtypes:

1. **GEO-NOM**: mentions that refer to a specific location with its own name (*Guadalquivir River*).
2. **GEO-GEN**: mentions that refer to a generic location using a common noun (*river*).

[examples]

3. (GEO-NOM) She visited the **Iguazú Falls**.
4. (GEO-GEN) He fell while taking a walk in the **mountain**.

**[FAC]**

It is used for **structures and spaces built by humans (facilities)**: dwellings, stores, leisure sites, tourist sites, transportation structures (bus and train stations), offices, farms, schools, etc.

We distinguish two subtypes:

3. **FAC-NOM**: mentions that refer to a specific location with its own name (*Sagrada Familia*).
4. **FAC-GEN**: mentions referring to a generic location using a common noun (*monument*).

[examples]

5. (FAC-NOM) The case was referred to **Hospital La Paz** in **Madrid**.
6. (FAC-GEN) After his release from **prison** he worked for a year at the **airport**.

### DEMOGRAPHIC LABELS

**[COMMUNITY]**

Demonyms, i.e., adjectives or nouns referring to a place, as well as ethnicities, religions and sociodemographic elements of the style are annotated with this label.

[examples].	
7. 16-year-old <b>Pakistani</b> woman. 8. Male, 36 years old, <b>Ukrainian</b> nationality. 9. Patient of <b>Spanish</b> origin.	
<b>[LANGUAGE]</b>	This category is used for the languages mentioned in the text.
[examples].	
10. The patient has communication problems, as she does not speak <b>Spanish</b> . 11. The son interprets the conversation, as the patient only speaks <b>Chinese</b> .	
<b>OTHER LABELS</b>	
<b>[DEPARTMENT]</b>	This category includes the different departments and services of health institutions (e.g. dermatology, cardiology), as well as the different parts of hospitals and health centers.
[examples]	
12. On the <b>ward</b> , prior to the <b>operating room</b> , the patient had no personal history of interest. 13. Follow-up by <b>pediatric cardiology</b> .	
<b>[TRANSPORT]</b>	This category includes mentions of transport and movements in general, including the word "travel", mentions of patient movements and means of transport, among others.
[examples]	
14. The patient has just returned from a <b>cruise</b> . 15. He does not refer to recent <b>trips</b> .	

## 2.2 ANNOTATION AND EXTENSION RULES

The general rules for annotating entities are described below. These rules apply to all categories, regardless of their label:

B1. GENERAL POSITIVE ANNOTATION RULES	
<p><b>[GEN-P1]</b> [spelling]</p>	<p>All mentions of places should be annotated even if they include misspellings.</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[examples]</p> <p>1. Patient native of [sic] [<b>nicaruaga</b>]<sub>GPE</sub> .</p>
<p><b>[GEN-P2]</b> [separation]</p>	<p>If several geographical entities of different or the same type are listed, each of them should be annotated separately.</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[examples]</p> <p>2. His grandparents live in <b>Puente Genil (Córdoba)</b>.</p> <p>3. He has been traveling around <b>Europe (Norway, Denmark and Germany)</b>.</p> <p>4. He stayed for 5 days in a [<b>rural area</b>]<sub>GEO</sub> of [<b>Santa Cruz</b>]<sub>GPE</sub> .</p> <hr style="border-top: 1px dashed black;"/> <p style="text-align: center;">[exceptions]</p> <ul style="list-style-type: none"> <li>The administrative divisions (<i>city, province, ...</i>) that go together with a toponym are annotated as a single entity (rule GPE-P2). This will facilitate later standardization in ambiguous cases such as <i>city of Barcelona</i> or <i>province of Barcelona</i>.</li> </ul> <p>5. Originally from the <b>state of San Luis Potosi</b>, with residence in the <b>Acantzen neighborhood</b> of the <b>town of Tanchahuil</b> in the <b>municipality of San Antonio</b>, in the same <b>state</b>.</p>

**[GEN-P3]**  
[nesting]

In the event that two mentions overlap, they will be annotated in a nested manner. That is, both mentions will be annotated independently from beginning to end.

In the example below, square brackets are used to mark how far each independent annotation goes.

[examples]

6. He had been to **[[Central]<sub>GPE</sub> and South America]<sub>GPE</sub>**.

**[GEN-P4]**  
[acronyms]

Acronyms that appear next to their long form are annotated together.

[examples]

7. She is taken by police to the **emergency department (ED)** after allegedly trying to hijack a **bus**.

**[GEN-P5]**  
[adjectives-prep].

If characteristics of a GEN entity are specified by means of adjectives or prepositional syntagms, they are included within the mention.

[examples]

8. She owns a **house** in the **foothills** of **Venezuela's central coast** in the **mountain range** (*GEO\_GEN*).

**[GEN-P6]**  
[compl-name]

We also annotate the places that appear within a name complement.

[examples]

9. **Laboratory** tests support the initial hypothesis.

10. Picks up tools, leaflets, and **stationery** supplies from the street or steals them at work.

11. **Emergency department** physician

## B2. GENERAL NEGATIVE ANNOTATION RULES

**[GEN-N1]**  
[det-prep]

The determiners and prepositions preceding the entity are not included in the mentions.

[examples]

12. Just back from **a trip to Tanzania**.

13. Worker based in **the Netherlands**.

[exceptions]

- If the determiner is part of a toponym (usually capitalized), it must be included in the mention.

14. He made a **trip** from **El Salvador** to **La Coruña**.

- Some generic GPEs must be annotated next to the determiner if a possessive determiner (see GPE-E2) accompanies them.

15. He resides in **our country**.

**[GEN-N2]**  
[generic-sites]

Nouns that refer to a place in a very, very generic way are not annotated if they are used alone: *site, place, zone, area, ...*

[Exception: DEPA-E2].

[examples]

16. He lives in the "local way", with the same hygienic-dietary measures as the locals of this **area**.

17. His **workplace** is located near a **river**, a **place** where he washes his tools.

**[GEN-N3]**  
[adv-dem]

Neither demonstratives (*here, there, there, there, ...*) nor adverbs (*in front, beside, behind, far, near, ...*) that are related with a location are annotated.

[examples]

18. Once **there**, he came into contact with several **farm** animals.

19. He sprained his ankle **in front of** the **ambulatory centre**.

**[GEN-N4]**  
[events]

No general mentions of locations where an event has occurred are annotated.

[Related rules: GEO-N2, GEO-E1].

[examples]

20. No evidence was found at the **accident location**.

21. The thief always returns to the **scene of the crime**.

**[GEN-N5]**  
[web-sites]

Websites in general and application names are not considered locations and are not annotated.

[examples].

22. He met his current partner on **Twitter**.

**[GEN-N6]**  
[species-procedure]

Names of locations that are part of or homonyms of the name of pathogens, clinical procedures or diseases are not annotated.

[examples].

23. The patient had a **Glasgow** score of 15.

24. Application of the test revealed a specific band for L. **braziliensis**, which is not detected in sera from individuals infected with L. **mexicana** or L. **amazonensis**.
25. In the course of the evolution, the two remaining **Amsterdam** criteria will be confirmed.

The following tables explain the specific rules for each of the geographic entities:

C1. RULES FOR "GPE".	
POSITIVE RULES	
<p><b>[GPE-PI]</b> [examples]</p>	<p>Geographic regions inhabited by people and influenced by the politics, economy, culture and history of their environment are annotated with this label. This includes countries, regions, cities, etc.</p>
<p>[examples]</p>	
<p>26. The patient resides in <b>Madrid</b> but for work reasons had <b>moved</b> to <b>Huelva</b> in the last few days (<i>NOM</i>).</p> <p>27. He arrived from <b>India</b> when he was 11 years old (<i>NOM</i>).</p> <p>28. Six-year-old girl from <b>Pozuzo, department of Pasco (central jungle)</b>, admitted to a <b>hospital</b> in <b>Lima</b> in August 2006 (<i>NOM</i>).</p> <p>29. 18-year-old patient from <b>Guinea Conakry</b>, with no previous history of interest, recently arrived in <b>Spain</b> on a <b>small boat</b> a month and a half ago (<i>NOM</i>).</p> <p>30. Married with a healthy son, she lives in a <b>rural area</b> near the <b>capital of the island of Gran Canaria</b> (<i>NOM</i>).</p> <p>31. Independent for his daily life tasks, is living alone in a <b>neighborhood with medium socioeconomic level</b> (<i>GEN</i>).</p> <p>32. Parasitotic prevalent in <b>[tropical and [subtropical countries]]</b>, caused by various helminths. (<i>GEN</i>)</p> <p>33. He had recently been working <b>abroad</b> (<i>GEN</i>).</p>	



34. He had received antibiotics for urinary tract infection (pyelonephritis), without improvement, at a **medical facility** in **his homeland**. (GEN)

**[GPE-P2]**  
[divisions-1]

Administrative divisions ("district", "province"), specific regions and zones ("northwest zone"), cardinal points, etc. are included in the annotation if they appear next to a proper name, all annotated as GPE\_NOM.

[examples]

35. Native of the **province of Punjab**.  
36. Gazpacho is a typical dish of **southern Spain**.

**[GPE-P3]**  
[divisions-2]

Administrative divisions appearing alone or with an adjective or a preposition are annotated as GPE-GEN.

[examples]

37. This August, he plans to tour several **[Asian]<sub>COMMUNITY</sub> capitals** ]<sub>GPE-GEN</sub> .  
38. Similar cases of tick bites have been seen in the **province**.

**[GPE-P4]**  
[neighborhoods]

We will annotate neighborhoods as GPEs, considering them as one of the smallest types of administrative divisions.

[examples]

39. Five people in the patient's **neighborhood** have been diagnosed with COVID-19.

40. When he arrived in **Seville**, he stayed in a **shared apartment** in a **marginal neighborhood**.

41. ...mapping of pre-existing community health assets in the **neighborhood of La Pau**.

42. Epidemiological background: heterosexual man, resident in the **Antímamo neighborhood**, located in the **southwest of Caracas**.

<b>SPECIAL RULES</b>	
<p><b>[GPE-S1]</b> [abroad]</p>	<p>The word "abroad" is annotated as GPE-GEN if it is used to mention a location.</p> <p>(See also rule COM-S1)</p>
<p>[examples]</p> <p>43. It has been fifteen years since he last traveled <b>abroad</b>.</p>	
<p><b>[GPE-S2]</b> [post-more-count ry]</p>	<p>As an exception to the GEN-N1 rule, if a possessive determiner ("my", "their", "our", ...) accompanies an administrative division ("country", "village", "town", "locality", "district", etc.), we annotate everything as a single mention with the label GPE_GEN.</p>
<p>[examples]</p> <p>44. The first case of coronavirus in <b>our country</b> is presented.</p>	
<p><b>[GPE-S3]</b> [sprep-demo]</p>	<p>Genitive constructions that refer to someone's place of birth or residence should be included in the citation.</p> <p>This rule is usually combined with GPE-S2.</p>
<p>[examples].</p> <p>45. Up-to-date immunization schedule in <b>her home country</b>. 46. He underwent surgery in <b>his country of residence</b>.</p>	
<p><b>[GPE-S4]</b> [coordinates]</p>	<p>Coordinates should be annotated as GPE-NOM if they appear in the text (also when they appear next to an entity that we would annotate).</p>
<p>[examples]</p> <p>47. From the <b>Sierra de Parima (2° 47' 21" N, 64° 13' 55" W)</b>.</p>	

## C2. RULES FOR "GEO".

### POSITIVE RULES

#### [GEO-P1] [examples]

Geographical features, natural locations in general, types of ecosystems and environments, areas of the world, etc.

[examples]

48. The group spent most of the time on the **beach** in bare feet and bathed at the **mouth of a river** (*GEN*).
49. In the last decade, there is a majority consensus of an increase in the prevalence and incidence of this disease in **Europe**, in the **Mediterranean basin** (*NOM*).
50. Exposure to volcanic ash dust due to proximity to the **Tungurahua volcano** emission radius (*NOM*)
51. Construction worker, with regular exposure to asbestos and 3 years of work experience in a **coal mine** (*GEN*)
52. Pathology increases as we get closer to the **equator** (*NOM*)
53. Regularly goes on cycling excursions in the **countryside** (*GEN*)
54. He climbed to the top of a **1,828-meter hill**, under intense environmental heat (*GEN*).
55. He had traveled to the **rural area** of **Dagua (Valle del Cauca)** (*GEN*).
56. He reported having visited an outdoor **hot springs area** in the area near his **municipality of Granada** (*GEN*).
57. History of travel to **tungiasis endemic areas** (*GEN*)

#### [GEO-P2] [acc-nom]

If the name of the geographical feature appears together with a toponym, it is annotated as a single mention with the label GEO\_NOM.

[examples]

58. This summer he bathed in the **Mediterranean Sea** and the **Atlantic Ocean**.
- This summer he bathed in the **Mediterranean** Sea and the **Atlantic Ocean**.

<p><b>[GEO-P3]</b> [cardinal]</p>	<p>Cardinal points, if they do not specify a location within a GPE-type entity (e.g., "southern Spain"), are annotated as GEO-GEN.</p>
<p style="text-align: center;">[examples]</p> <p style="text-align: center;">59. Gazpacho is a typical <b>southern</b> dish.</p>	
<p><b>[GEO-P4]</b> [adjectives-prep]</p>	<p>If the type of a GEO-GEN entity is specified by means of adjectives or prepositional phrases, it is included within the mention.</p>
<p style="text-align: center;">[examples].</p> <p style="text-align: center;">60. She owns a <b>house</b> in the <b>foothills</b> of the <b>Venezuelan central coastal mountain range</b>.</p>	
<p><b>[GEO-P5]</b> [habitats]</p>	<p>The nouns "region", "zone" and "area" followed by an adjective expressing temperature (cold, warm, etc.), type of ecosystem (forest, rural, urban, etc.) or similar elements are annotated as GEO-GEN. We will refer to this type of mentions as "habitats" or "ecosystems".</p> <p>(See also: GEO-N3 rule)</p>
<p style="text-align: center;">[examples].</p> <p style="text-align: center;">61. He worked in a <b>jungle region</b> of the <b>country</b>. 62. A young woman from an <b>endemic area</b>.</p>	
<p><b>[GEO-P6]</b> [endemic]</p>	<p>When the text mentions endemic, epidemic, or high risk areas of specific diseases or pathogens, we will extend the mention to the name of the pathogen or disease.</p>
<p style="text-align: center;">[examples].</p>	

63. He was visiting a **babesiosis endemic region**.

### NEGATIVE RULES

**[GEO-N1]**  
[astronomical]

Astronomical objects: planets, stars, constellations, etc., are not annotated.

[examples].

64. The patient presents with delusions in which he believes he works as an astronaut and lives with his family on **Neptune**.

**[GEO-N2]**  
[waters]

Although they may look like locations, we will not note the types of water where someone can bathe.

[See related rule: GEN-N4]

[examples].

65. He did not report any specific risk behaviors during the **trip**, except for bathing in **fresh water** in **Bamako**.

### SPECIAL RULES

**[GEO-S1]**  
[place-consumption]

As an exception to the GEN-N4 rule, we decided to annotate the mention of "places of consumption".

[See related rule: GEN-N4]

[examples] [examples]

66. Control of risk factors is prescribed to avoid relapses (control of money management, avoidance of consumer friends and **places of consumption**, avoidance of being and going out alone).

## C3. RULES FOR "FAC".

## POSITIVE RULES

**[FAC-P1]**  
[examples]

Spaces built by humans are annotated with this label, which includes many common locations that we would find in a city: establishments, stores, structures, monuments, etc., as well as different types of health centers.

[examples].

67. He reports having worked in a **handbag store** when he was 16 years old for a year, in a **fish market** and at a **gas station** (*GEN*).
68. Started to regularly go to the **gym** (*GEN*).
69. Enrolled at the end of high school in a **computer and typing academy**, which he did not complete (*GEN*).
70. The patient declares that he is a non-smoker and works in a **car battery factory** (*GEN*).
71. A thorough examination of the **vehicle**, once deposited at the **scrapyard**, allowed the collection of several cartridges (*GEN*).
72. The parents were admitted to a **psychogeriatric home** by court order (*GEN*).
73. Subsequently, the patient was referred to the **National Cancer Institute** for evaluation (*NOM*).
74. He was working on the restoration of a **school building** in the **Santa Rosalía parish** of **Caracas, Venezuela** (*GEN/NOM*).
75. The strain was sent to the **Mycology Service** of the **National Microbiology Center** for definitive identification (*NOM*).
76. Admitted to **foster care** when the patient was 13 months old (*GEN*).
77. She was admitted in July 2008 (summer in the **Northern Hemisphere**) to **San Juan de Dios Hospital**, with the background of having returned 8 days earlier from a **trip to North Florida**, in the **United States**, where she was walking in a **forest** known to have a high tick population (*NOM*).



<p><b>[FAC-P2]</b> [building-type]</p>	<p>If the type of <i>facility</i> is mentioned together with its proper name, it is annotated as a single mention and annotated as FAC-NOM.</p>
<p style="text-align: center;">[examples].</p> <p style="text-align: center;">78. He works as a waiter at <b>La Barca restaurant</b>.</p>	
<p><b>[FAC-P3]</b> [floors]</p>	<p>We annotate as FAC-GEN the parts within a building ("second floor", "first floor", "basement", "rooftop", etc.), as well as the different types of rooms ("kitchen", "bedroom", "bathroom", etc.).</p>
<p style="text-align: center;">[examples]</p> <p>79. <b>House</b>: He lives on the <b>second floor</b> of a <b>building</b> in <b>Tarragona</b>, in front of the <b>Central Park</b>.</p> <p>80. Suffers a blackout episode while at the <b>toilet</b> getting ready for work</p> <p>81. Decreases the tendency to isolate themselves in their <b>bedroom</b>, spending most of the day in the <b>common room</b>.</p>	
<p><b>[FAC-P4]</b> [hospital-names]</p>	<p>Some specific hospitals include within their name the site to which they belong. This is the case of sites with very generic names such as "University Hospital", "General Hospital", etc. In these cases, we will annotate all together as FAC_NOM entity (including the site name) and we will also annotate a nested GPE-NOM entity.</p>
<p style="text-align: center;">[examples]</p> <p>82. He was admitted to a <b>pediatric intensive care unit</b> at the <b>[Maternal and Children's Hospital of [Hubei Province]] (Wuhan, China)</b>.</p> <p>83. <b>[[Beijing] Centers for Disease Control and Prevention (CDC)]</b> confirmed by real-time RT-PCR that the patient had COVID-19.</p> <p>84. Follow-up at <b>[Hospital Clínic de [Barcelona]]</b>.</p>	
<p style="text-align: center;">[exceptions]</p>	

- We will not annotate if the location of the hospital is specified, but it is not part of its name (see last example).

85. Transfer to **Hospital Niño Jesús in Madrid**.

86. He was referred for follow-up to the **outpatient clinic** of his **hospital of origin (Alto Guadalquivir in Andújar)**.

- This rule also does not apply when a location is mentioned in a generic or non-specific manner. These cases are annotated separately (FAC\_GEN + GPE).

87. In 2015 he was admitted to a **hospital** in **Malaga** from where he escaped a few days after his admission.

**[FAC-P5]**  
[residence]

Mentions of the patient's place of residence are also annotated as FAC.

[examples]

88. The need to include the patient in a **residential resource** had already been raised.

89. Upon discharge, he returned **home**

### NEGATIVE RULES

**[FAC-N1]**  
[organizations]

In some contexts, there are names that look like Facilities but actually refer to organizations. In general, these cases should not be annotated.

[examples].

90. She is transferred to the **residential environment** of **FAISEM (Andalusian Public Foundation for the Social Integration of Persons with Mental Illness)**, continuing to be monitored in this **Therapeutic Community**.

### SPECIAL RULES



<b>[FAC-S1]</b> [ambulatory]	We also note the word <i>ambulatory</i> in expressions such as "ambulatory follow-up" or "ambulatory control".
[examples]	
91. She is discharged for <b>ambulatory</b> follow-up.	

C4. RULES FOR "COMMUNITY"	
POSITIVE RULES	
<b>[COM-P1]</b> [name-adj]	All demonyms should be annotated, whether they appear as adjectives or nouns.
[examples]	
92. 17 year old <b>Spanish</b> male who comes to the <b>hospital</b>	
93. He was treated in a <b>[[Spanish]<sub>COMU</sub> hospital<sub>FAC</sub>]</b>	
<b>[COM-P2]</b> [nested]	Demonyms that appear within a geographic mention are also annotated.
[examples]	
94. He spent his childhood in an <b>[[African]<sub>COM</sub> city]<sub>GPE</sub></b>	
SPECIAL RULES	
<b>[COM-S1]</b> [foreigner]	By convention, we will annotate the word "foreigner" using this label when it refers to a person.  (See also rule GPE-S1)
[examples]	

95. The patient is a **foreigner**, although we do not know exactly where she is from.

**[COM-S2]**  
[religions]

Due to its possible clinical relevance, we also use this label to note religions and ethnicities in general...

[examples]

96. A 21 year old **Indian** patient...

97. Of **Jewish** origin

## C5. RULES FOR "LANGUAGE"

### POSITIVE RULES

**[LAN-P1]**  
[language]

All language mentions are annotated.

[examples]

98. The patient does not speak **Spanish** and we were unable to find a translator.

**[LAN-P2]**  
[de-talk]

The expression "-speaking" plus an adjective referring to a language is annotated with this tag, without including the preposition.

[examples]

99. This is very common in **English-speaking countries**.

### SPECIAL RULES

**[LAN-S1]**  
[Anglo-Saxon]

By convention, we will annotate the word "Anglo-Saxon" using the language tag.

[examples]

100. **Anglo-Saxon** patient, 76 years old, resident in Spain for 22 years.

**[LAN-S2]**  
[language  
barrier]

By convention, we will annotate the expression "language barrier" and the like with this label....

[examples]

101. Not included in the study due to **language barrier**.

## C6. RULES FOR "DEPARTMENT".

### POSITIVE RULES

**[DEP-P1]**  
[departments]

We annotate the departments, services, sections and units within hospitals and other health centers.

In this type of annotations, we normally include the word "pediatric" within the same mention, since it refers to a specific type of service.

[examples]

102. The patient was treated by the **infectious diseases department** of **Hospital La Paz**.
103. Patient seen for the first time in the **Pediatric Endocrinology Service** of the **Maternal and Child Hospital of Malaga**, at 8 years and 6 months of age, referred with a diagnosis of ADS.
104. The patient was admitted to the **ICU** and later it was decided to admit him to a **conventional hospital ward**.



**[DEP-P2]**  
[specialties]

We note the names of medical specialties when they refer to sections or teams within a medical center.

[examples]

- 105. The case was studied by **Internal Medicine** and **Neurology** and no pathological data were found in the complementary tests.
- 106. Despite his low EF, **cardiology** did not consider him to be in acute heart failure.
- 107. We requested **neurology** consultation after a seizure episode.

**[DEP-P3]**  
[rooms]

We annotate the specific rooms (e.g. waiting or rehabilitation) and floors (e.g. specialty or hospitalisation) within a medical facility, including rooms and beds.

[examples]

- 108. Two weeks after his admission to the **gastrointestinal ward**, his condition was complicated by a severe gastrointestinal hemorrhage.
- 109. The patient was maintained on ventilatory support in the **emergency observation area** awaiting possible death or organ donation.
- 110. The patient was taken from the **ambulance** to a **laminar flow room** (with antechamber) on the **second floor** of the **unit**.

**[DEP-P4]**  
[consultation]

We annotate the word "consultation" when referring to the physical location (i.e. the consultation room), as well as specific types of consultations (outpatient, specialty).

Since the word "consultation" by itself can be very ambiguous (among others, it can mean "visit" or "question"), a good trick to know if it should be annotated is to replace it with the synonym "ward".

[examples]

111. He goes to a **psychiatrist's consultation** with his wife.

112. Reason for **consultation**: occupational accident.

113. One year later, the patient returned to the **consultation room** with severe abdominal cramps.

**[DEP-P5]**  
[equipment]

We also annotate the hospital's specialty teams, since they imply that the patient has gone through that specialty.

[examples]

114. She has been followed since February 2012 by the **Primary Care Team (EAP)** for symptom control and observation of her neoplastic disease.

115. Finally, the **Disability Assessment Team** of the **National Institute of Social Security**, which met to assess your case, resolved Total Permanent Disability due to occupational disease.

116. After discussing the case in the **Otorhinolaryngology Tumor Committee**, it was decided to perform a positron emission tomography to assess the nature of the adenopathies.

### NEGATIVE RULES

**[DEP-N1]**  
[centers]

Specialty medical centers, polyclinics, rehabilitation facilities, etc., are not annotated in this category, since they are usually entire buildings.

[examples]

117. Has not been treated at any other **medical center**

118. She was treated in a **mental health and addiction clinic**.

**[DEP-N2]**  
[tests]

The names of tests or biological aspects are not annotated under this category, even if they coincide with the name of a medical specialty.

[examples]

119. With respect to **microbiology**, a high variability of pathogens have been described as responsible for this pathology.
120. Within 24 hours of **surgery**, the patient was afebrile for the first time since his arrival at the **hospital**.
121. **Pathological anatomy** reports necrotizing granulomatous inflammation.
122. **Emergency** tests performed upon arrival
123. **Digestive**: Absence of signs of peritonism, positive right renal fist percussion.

[Here, "Digestive" refers to a type of examination, not to the clinical department.]

### SPECIAL RULES

**[DEP-S1]**  
[specific]

We annotate with this label the following sites within a medical center: operating room, incubator, ward,

[examples]

124. Admitted directly to the **operating room** for removal of the bullet.
125. It remains in the **incubator** until day 17 after birth.
126. Forty-five minutes after his arrival, he was admitted to the **ward** with the diagnosis of a complicated abdominal gunshot wound.

**[DEP-S2]**  
[service]

We annotate the words "service" and "unit" also when they stand alone, whenever they refer to a location.

[examples]

127. After the reappearance of symptoms, she consulted our **department**.

## C7. RULES FOR "TRANSPORT".



## POSITIVE RULES

### [TRA-P1] [trip]

The word **trip** and its possible synonyms are annotated whenever they appear (exception: rule TRA-N2).

Some possible synonyms are: *visit* (as long as it refers to a trip and not to a medical consultation), *tour*, *journey* or *trip*, among others.

#### [examples]

128. He had planned a **trip** to **Eastern Europe** with his family.  
129. **Air evacuation** to **Italy** was being arranged.  
130. He said that his problems had started 3 months earlier, when he returned from a **visit** to **Japan**.  
131. He denied bites of any kind, although some of his fellow **hike** companions in **New York** had been bitten by ticks.

### [TRA-P2] [travel-types]

The noun affixes that describe the type or reason for the trip are included in the mention: *intercontinental*, *tourism*, *work*, etc. However, adjectives indicating time or frequency should not be included.

In the case of a more in-depth description of a travel feature ("backpacking trip"), only the main qualifier should be annotated.

#### [examples]

132. He reported having made multiple **intercontinental trips**, with no apparent health problems.  
133. History of **previous trips** to **Thailand**, **Morocco** and **Mexico**...  
134. He went on a **"backpacking" trip** with his partner

### [TRA-P3] [means-transport ]

Means of transportation are also annotated with this label.



[examples]

135. He was hit by a **bus** while crossing the **road**.  
 136. Arrival in **Spain** by **boat** in the **Canary Islands**.  
 137. A high explosive device detonates in the path of his **vehicle**.

### NEGATIVE RULES

#### [TRA-N1]

[verbs]

Verbs of motion are not annotated.

(Exception: TRA-S2 rule)

[examples]

138. He recently **travelled** to his parents' home in **Thailand**.  
 139. Three years ago he **flew** from **Morocco** to **Spain**  
 140. He had **travelled** to **India** a few months ago

#### [TRA-N2]

[trip-figured]

Mentions of trips are not annotated if they do not actually refer to one. That is to say, we do not include drug trips, specific objects to use during trips, mentions within occupations, etc.

[examples]

141. He works as a **travel agent**.  
 142. She wears **travel** stockings to avoid thrombosis.  
 143. He ended up in the ER after a bad acid **trip**.  
 144. A 47-year-old woman on a **tourist trip** in **Argentina**, living with her **travel** companion in a **mobile home**.

### SPECIAL RULES

#### [TRA-S1]

[chair-wheels]

We also annotate with this label mentions of wheelchairs and walkers.





[examples]

145. He needs a **wheelchair** to go to the **unit**.
146. Peripheral neuropathy that over the years became disabling (year 2008) and forces her to move around in a **wheelchair**.
147. Walking with a **walker**.

**[TRA-S2]**

[transfer]

We will annotate the word "transfer", both as a noun and as a verb, when referring to the movement of patients between health care entities and facilities.

[examples]

148. He will be **transferred** urgently to **Nakasero Hospital** in **Kampala**.
149. At the last consultation, on palpation of the abdomen, his **primary care physician** assessed the **transfer** to the **hospital emergency department**.

**[TRA-S3]**

[visit]

We will annotate the word "visit" when it is accompanied by adjectives or other noun complements.

[examples]

150. [...] facilitated by the boost provided by the **telematic visit** introduced during the COVID-19 pandemic.
151. We make **home visits**.

### 3. CLASSIFICATION OF ENTITIES

---

This section describes the attributes with which we will characterize the entities annotated during the first phase annotation round. These attributes express different types of information: clinical information, referent or associated person, polarity and temporality.

#### 3.1 CLINICAL CLASSIFICATION

There are certain aspects related to places that are of special interest for the biomedical field. In order to rescue certain types of information and thinking about the different use cases, we have defined five tags that will allow us to classify the entities we annotate. These categories are used only for place entities (GPE, GEO, FAC).

The following table sets out the categories for the clinical classification of entities:

D1. CLINICAL CLASSIFICATION LABELS FOR ENTITIES	
<b>[BIRTHPLACE]</b>	To be used for entities that refer to a person's place of origin: where he/she was born, country of origin, nationality, etc.
[examples]	
1. A 52-year-old woman from the <b>United States</b> .	
<b>[RESIDENCE]</b>	To be used for entities that refer to a person's current or past place of residence.
[examples]	
2. She has been living in <b>Galicia</b> for 7 years.	

<b>[MOVEMENT]</b>	To be used for entities that refer to places that are the origin or destination of a movement. Stays also fall into this category.
<p style="text-align: center;">[examples]</p> <p>3. Two months earlier, he had traveled to <b>Mauritania</b>.</p> <p>4. After returning from the trip to <b>India</b></p> <p>5. The <b>trip</b> had religious motives, with a stay mainly in the <b>state of Rajasthan (northwest of the Indian subcontinent)</b>.</p>	
<b>[HEALTHCARE]</b>	To be used for entities that refer to the place where the patient received some type of healthcare.
<p style="text-align: center;">[examples]</p> <p>6. Required medical attention in <b>Manaus</b>.</p> <p>7. In 1988, she underwent surgery for goiter in <b>Algeria</b>.</p> <p>8. He went to a <b>medical center</b> in <b>Turkey</b>.</p>	
<b>[OTHERS]</b>	It will be used for entities that do not fit into any of the above categories.
<p style="text-align: center;">[examples]</p> <p>9. The picture began in <b>Senegal</b>.</p> <p>10. Schooled without problems in <b>Ecuador</b>.</p>	

The following table describes the clinical semantic classification rules for geographic entities:

## D2. RULES FOR CLINICAL CLASSIFICATION OF ENTITIES

### GENERAL RULES

**[CLAS-G1]**  
[multi-label]

Geographic entities can have more than one semantic label.

[examples]

11. He was on vacation in **Ibiza**, where he was born.  
[TAGS: birthplace, movement].

**[CLAS-G2]**  
[temporality]

Semantic labels are attributed without regard to the temporality of the sentence.

[examples]

12. He was on vacation in **Kenya** in '83.  
[TAGS: movement]
13. He plans to tour **South America** next year: **Chile, Argentina, Uruguay**.  
[TAGS: movement]

**[CLAS-G3]**  
[polarity]

Semantic labels are attributed regardless of whether the entity is affirmed, negated or speculated.

[examples]

14. Possible destinations include **Norway, Estonia** and **Turkey**.  
[TAGS: movement]

### MULTICLASS RULES

**[CLAS-M1]**  
[stays]

Mentions of stays are normally labeled MOVEMENT.

[examples]

15. You will tour both **[[urban areas]<sub>GEO</sub> and rural areas]<sub>GEO</sub>**, including a stay in the **jungle** for 2 to 4 weeks.

### 3.2 CLASSIFICATION BY REFERENCE

In the classification by referent we will try to relate each annotated entity to the person with whom it is associated. To assign this tag, we must read the sentence where the entity appears and infer from the context to whom it is associated. Because of the type of domain we are working with, in most sentences the default label will be PATIENT.

The table below explains the possible categories for referent annotation:

E1. ATTRIBUTES FOR REFERENCES	
<b>[PATIENT]</b>	It will be used for entities that talk about the patient. It is the default label for most entities.
[examples]	
16. 36-year-old woman, <b>Ukrainian</b> nationality, resident in <b>Spain</b> .	
<b>[FAMILY]</b>	It will be used for entities that talk about people who are relatives of the patient.
[examples]	
17. Married in 2009 to a man from <b>Malaga</b> .	
<b>[OTHERS]</b>	To be used for entities that do not refer to persons (other places, events, ...).
[examples]	

18. The first case of anisakis in **Colombia** is described.

Some specific rules for this classification are described below:

E2. CLASSIFICATION RULES BY REFERENTS	
<b>GENERAL RULES</b>	
<b>[REF-G1]</b> [single-tag]	Each entity can have only one referent type label.
<b>MULTICLASS RULES</b>	
<b>[REF-M1]</b> [pct-someone]	Entities that can be listed as PATIENT but also as another referent should be tagged as PATIENT.
[examples]	
19. He bathed in the <b>[waterfalls]<sub>PAC</sub></b> of the <b>[Dogon Country]<sub>PAC</sub></b> with his travel partner.	
<b>[REF-M2]</b> [baby-and-mother]	In some pediatric cases, both the infant and the mother are treated as patients. In these cases, we will annotate both as PATIENT.
[examples]	
20. One day after delivery, mother and baby are discharged and return to <b>[their home]<sub>PAC</sub></b>	

### 3.3 POLARITY AND TEMPORALITY

This annotation layer provides us with extra-propositional information about the entities we annotate, allowing us to know if and when they have occurred. We established two separate aspects: polarity (table F1) and temporality (table F2).

F1. POLARITY ATTRIBUTES	
<b>[POSITIVE]</b>	It will be used for entities that are affirmed. It is the default label for most entities.
[examples]	
21. He is a resident in <b>anatomic pathology</b> .	
<b>[NEGATIVE]</b>	It will be used for entities that are denied.
[examples]	
22. He does not refer to recent <b>trips</b> .	
<b>[UNKNOWN]</b>	It will be used for entities about which some kind of doubt is expressed as to whether or not they have occurred.
[examples]	
23. The patient's grandmother is not sure if he has been in any <b>swampy areas</b> .	

F2. TEMPORALITY ATTRIBUTES	
<b>[PRESENT]</b>	It will be used for entities that are in the present tense. It is the default tag for most entities.

[examples]	
24. He is living in a <b>nursing home</b> .	
<b>[PAST]</b>	It will be used for entities that are in the past tense.
[examples]	
25. He worked for 7 years in the <b>mines</b> of <b>South Africa</b> .	
<b>[FUTURE]</b>	It will be used for entities that are in the future, including plans.
[examples]	
26. He needs prophylaxis because he will travel to <b>Papua</b> .	
<b>[HYPOTHETICAL ]</b>	It will be used for entities that do not occur at any given point in time, but are mentioned in a hypothetical or figurative manner.
[examples]	
27. It seems that cryotherapy is used more in <b>Russia</b> .	

## 4 . NORMALIZATION

---

The normalization of the entities is a crucial step for the optimal use of the information extracted from the texts. The division of the geographical entities is made with this step in mind, since each type must be normalized to a different



resource. The normalization will be done at the same time as the annotation and classification by entering the corresponding code in the comments section provided by brat for each entity.

For normalization, we will use three different sources. These sources are: GeoNames, PlusCodes and SNOMED-CT. Each of them has its own characteristics that make it suitable for one situation or another. In the different subsections of this section the peculiarities of each one are explained.

Thus, each type of entity should be normalized as follows:

<b>G1. SOURCES OF NORMALIZATION FOR EACH TYPE OF ENTITY</b>	
Entity	Ontology
<b>GPE-NOM</b>	GeoNames
<b>GPE-GEN</b>	SNOMED-CT
<b>GEO-NOM</b>	GeoNames
<b>GPE-GEN</b>	SNOMED-CT
<b>FAC-NOM</b>	PlusCodes
<b>FAC-GEN</b>	SNOMED-CT
<b>COMMUNITY</b>	SNOMED-CT
<b>LANGUAGE</b>	SNOMED-CT
<b>TRANSPORT</b>	SNOMED-CT

When annotating in brat, we will write each code preceded by a prefix to distinguish where the code comes from. The possible prefixes are as follows:

<b>G2. PREFIX FOR EACH NORMALIZATION SOURCE</b>	
Source	Prefix
<b>GeoNames</b>	GN:
<b>PlusCodes</b>	PC:
<b>SNOMED-CT</b>	SCTID:

ISO	ISO:
-----	------

Mentions that cannot be normalized in any way will be marked with the term "**NOCODE**".

In the following sections, the protocol to be followed for normalizing with each of the sources is specified.

## 4.1 NORMALIZATION WITH GEONAMES

GeoNames is a free geographic database with more than 21 million entries. Each entry has associated metadata (such as latitude, longitude, altitude, population, etc.), as well as a unique identifier code. This identifier code is the one we will use to normalize the mentions.

Within GeoNames we will be able to find from countries and continents to towns and some neighborhoods, passing through named geographical features. However, we will not find many specific FACILITY type establishments, so we interleave this database with PlusCodes (4.2).

GeoNames can be accessed directly from brat by selecting a piece of text<sup>4</sup> (Figure 2) or by going to <http://www.geonames.org/>.

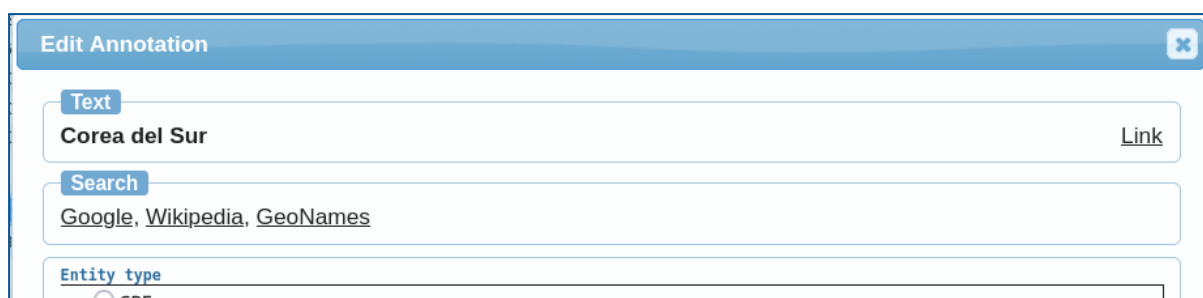


Figure 2. Direct search button in GeoNames in brat

When doing a search in GeoNames, we must see the options provided and select the location that is appropriate. To do this, we click on the name and enter the location page. The code is the part on the right with a dark blue background (Figure 3).

<sup>4</sup> The following line must be added to the brat *tools.conf* file in the *[search]* section:  
 GeoNames <URL>:http://www.geonames.org/search.html?q=%s&country=



Figure 3. Entering a location in GeoNames

The easiest way to select it to copy and paste the code is to click on the Feature tab, just above the location name. In the table that appears, we can select the code in the *id* field with a double click to copy and paste it into brat.

In geocoding there is a high ambiguity and it is common for the same name to be used to refer to different places (for example, Barcelona can be the name of a province and a city). Fortunately, GeoNames has an internal structure of identifiers that indicate the granularity of each element. These identifiers consider from a continent, to a small town of few inhabitants, to regions, autonomous communities or provinces. In Figure 3, the identifier is the code that appears in bold under the site name.

When choosing codes, it is important to take this identifier into account in order to always choose the right granularity. In the coding rules table (below) there are specific rules on which type of identifier to choose in case of ambiguity.

For reference, some of the most frequent GEONAMES hierarchy identifiers are as follows:

H1. EXAMPLES GEONAMES HIERARCHY		
ID	Name	Example
CONT	continent	Europe (GN:6255148)
RGN	region	Southern Europe (GN:9408658)
PCLI	independent political entity	Spain (GN:2510769)
ADM1	first-order administrative division	Valencian Community (GN:2593113)

PPLA	seat of a first-order administrative division	-
ADM2	second-order administrative division	Valencia [province] (GN:2593113)
PPLA2	seat of a second-order administrative division	Valencia [city] (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)

The following are the rules for coding with GeoNames:

<b>H2. CODING RULES WITH GEONAMES</b>	
<b>GENERAL RULES</b>	
<b>[GN-G1]</b> [granularity]	<p>Sometimes there may be ambiguity between several locations that have the same name and may refer to different administrative divisions (e.g. Malaga city and Malaga province).</p> <p>If the text does not specify which one is meant, the most general one (i.e. the largest region) should be chosen. In the table above, this means that to normalize "Valencia" we should take the code with the ID ADM2 (province).</p>

<p><b>[GN-G2]</b> [cardinal-1]</p>	<p>Cardinal points within continents will be normalized using the corresponding code whenever it exists.</p> <p>However, there are no codes in GeoNames for cardinal points within countries or cities, so we will use only the code for that country or city.</p>
<p><b>[GN-G3]</b> [cardinal-2]</p>	<p>Often we will not find in GeoNames codes for two cardinal points at the same time (<i>northwest, southeast</i>). These cases should be normalized by using either the code for the north or south region.</p>
<p><b>[GN-G4]</b> [spain-defect]</p>	<p>In the case of toponyms of places that exist both in Spain and somewhere else in the world (for example: there is a city called Córdoba both in Spain and in Argentina and Colombia), in this corpus we will choose by default the location in Spain if it is not specified in the text which of all of them it is.</p>

## 4.2 NORMALIZATION WITH PLUSCODES

PlusCodes, also known as OpenLocationCodes, is a system of codes to identify specific locations on Earth quite accurately. To create these codes, the world is divided into small grids, each with a code. Each code consists of a series of letters and numbers, calculated from latitude and longitude, that represent a specific location on the surface of the Earth.

The PlusCodes system was created by Google researchers and the codes can be used in different mapping and GPS services and applications. In our corpus, we will use this system to normalize mentions of specific buildings and *facilities* referenced in the text (i.e., entities of type FAC\_NOM) that do not appear in other ontologies.

For normalization, we will need the long version of the PlusCodes. This code is available when searching the map offered by the PlusCodes website ([plus.codes/map](https://plus.codes/map)). This site will sometimes offer us the short version of the code at

the bottom, we must drop down the menu on the left to get the long version (see Figure 4 below).

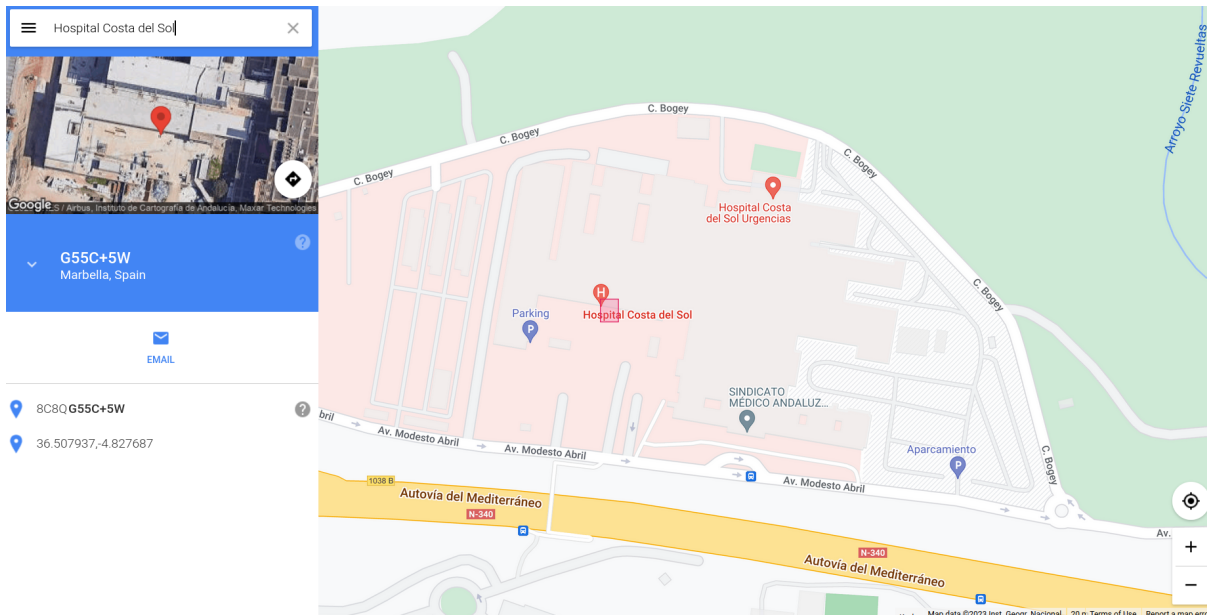


Figure 4. Example of a PlusCodes map search ([plus.codes/maps](https://plus.codes/maps)) for the Costa del Sol Hospital in Marbella, Malaga. The short version of the code is the four-character identifier + two characters in the blue part. The long code, which we must use, is the one shown in the white part above the coordinates. This code is the same as the short one but with four extra characters at the beginning to help locate the place in the grid.

Another alternative to find the PlusCode of a site is to search in Google Maps and select the corresponding location. In the information window that will open on the left, we will be able to see several data about that location, including the PlusCode (see Figure 5).

The code provided by Google Maps is a short version accompanied by the name of a town or city. This code is not useful for normalization (since to convert it to the long version automatically you have to use the Google Geocoding API which is paid), but we can use [plus.codes/map](https://plus.codes/map) to extract the long code by entering the short PlusCode with location in the search engine. Because of this extra step, we recommend using Google Maps only when we have trouble finding something on the PlusCodes map.

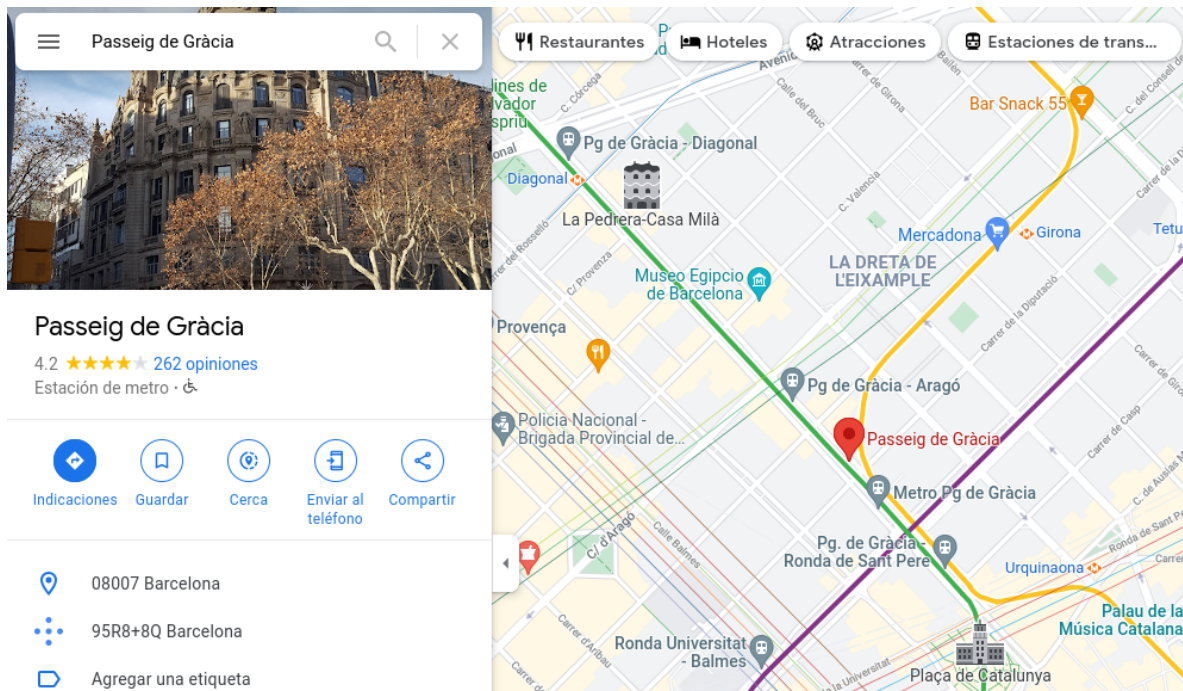


Figure 5. Example of a Google Maps search for the Passeig de Gràcia metro station in Barcelona. The PlusCode appears on the line preceded by a symbol made by five dots. In this case it would be: 95R8+8Q Barcelona.

Further processing of PlusCodes can be done with the official OpenLocationCodes library: <<https://github.com/google/open-location-code>>.

### 4.3 NORMALIZATION WITH SNOMED-CT

SNOMED-CT is a multilingual and comprehensive clinical terminology for coding multiple elements of the medical record. Among them, we find mentions of clinical services, ethnic groups and means of transportation.

To explore SNOMED we will use the online browser provided by the ontology<sup>5</sup>. In MEDDOPLACE, we have checked that all the codes used belong to the Spanish version (Spanish Edition) of 31-10-2022.

To normalize with SNOMED we will use the unique code of each entry. To find it, we must search for the concept in the SNOMED browser, click on the appropriate

<sup>5</sup> Available at: <https://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=MAIN/SNOMEDCT-ES/2022-10-31&release=&languages=es,en>

entry and by placing the cursor over the arrow in the blue box, click on "copy concept id" (in yellow in the example, see Figure 6).

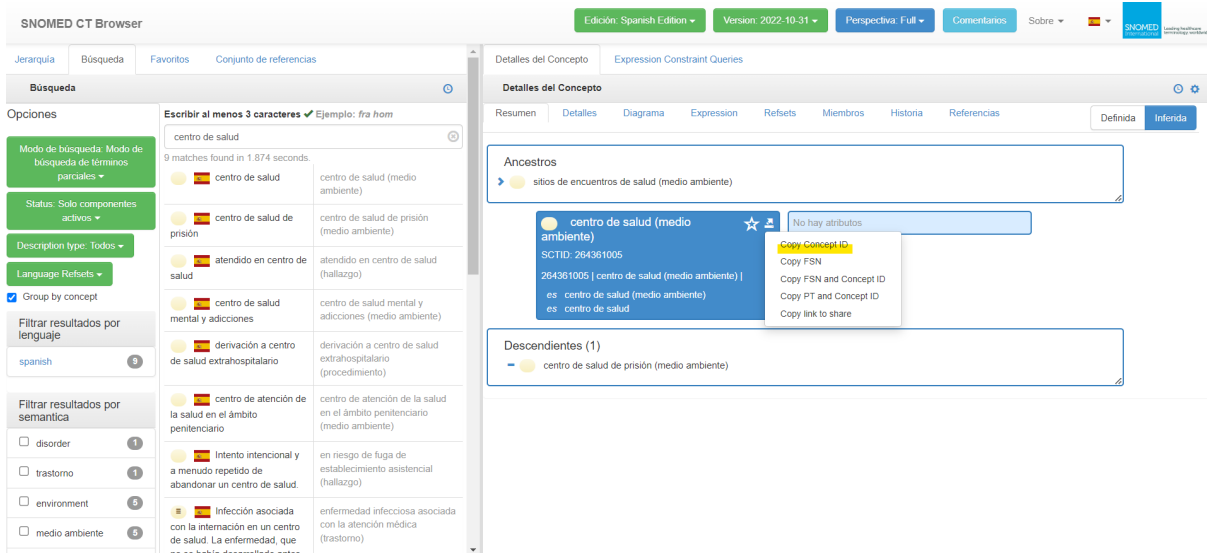


Figure 6. Example of a search in SNOMED-CT. On the left, the search results. On the right, the details of each entry.

Listed below are the rules and conventions defined for normalization with SNOMED:

II. CODES FOR GPE-GEN

<p><b>[1]</b> <b>[country of origin, birth, residence]</b></p>	<p>For mentions such as "country of <u>origin</u>" we will use the code for "country of origin (observable entity)" (SCTID: 276205001).</p> <p>On the other hand, for "countries of <u>residence</u>" we will use the code for "country of residence (observable entity)" (SCTID: 416647007).</p> <p>For "country of <u>birth</u>" or "country of birth" the code for "country of birth (observable entity)" (SCTID: 315354004) shall be used.</p>
--	--



## II. CODES FOR GPE-GEN

[2]  
[administrative  
divisions]

Due to the lack of codes, for mentions of administrative divisions such as "district" or "province" we will always use the code for "region (attribute)" (SCTID: 410673009).

## II.2. CODES FOR GEO-GEN

[1]  
[habitats]

Habitats are defined as mentions that refer to zones/areas/regions plus a descriptive adjective. These are the mentions described in the GEO-S1 rule of table C2.

For these mentions, we will use the codes of the *environment* branch.

[2]  
[endemic areas]

The code assigned to mentions of endemic regions for specific diseases (malaria, Chagas disease, trypanosomiasis, etc.) shall be as follows

224779005 : high risk environment (environment)

## III. CODES FOR FAC-GEN

[1]  
[health care]

When no identical code is found or the mention is isolated we will use the following codes:

1. for hospital type centers and when the word **clinic** appears without further specification, we will use the code "**hospital (environment)**" (SCTID: 22232009).
2. "hospital (environment)" (SCTID: 22232009) shall also be used for the isolated terms "**room**" and "**ward**" when not accompanied by more specific adjectives, and whenever they refer to a room or ward located in a hospital.
3. for other types of medical centers we will use the code "health center (environment)" (SCTID:



	<p>264361005). Note that the term <i>center</i> often refers to educational centers, which should be normalized with their corresponding codes.</p> <ol style="list-style-type: none"> <li>4. <b>outpatient clinic</b> will be standardized as 35971002  outpatient medical center (environment)</li> <li>5. <b>unit</b>: 43741000 : site of care (environment)</li> <li>6. <b>service</b>: 224891009 : healthcare services (qualifier)</li> </ol>
<b>[2] [guesthouses and hostels]</b>	<p>We will use the same code for guesthouses, hostels and lodges, since there is no code for the former: "lodging house (environment)" (SCTID: 257638006).</p>
<b>[3] [business]</b>	<p>For stores and businesses, we will try to use the child codes of "trade and service environment (environment)" (SCTID: 284438004).</p> <p>In case there is no code for a specific business type, we will use by default the code "shop (environment)" (SCTID: 257701001).</p>
<b>[4] [chronicity]</b>	<p>Mentions of sanatoriums, chronic hospital admissions and long-term admissions are normalized with 32074000  long-term care hospital (environment) .</p>
<b>[5] [mother and child hospital]</b>	<p>Mother and child hospitals are coded with 82242000  children's hospital (environment).</p> <p>Hospitals and maternal/obstetric services are standardized with 52668009  hospital-based birthing center (environment) </p>
<b>[6] [academic institution]</b>	<p>Mentions of universities and biomedical academic institutions are standardized as follows</p> <p>405608006  academic medical center (environment) </p>
<b>[7] [social services]</b>	<p>We normalize with code 310134006  social services (qualifier)  all governmental social services and also the civic organizations for the improvement of the health and social</p>



condition of patients and people of all kinds. For example: *Caritas, Instituto de Mayores y Servicios Sociales, Asociación de Ayuda a Ludópatas.*

**14. DEPARTMENT CODES**

<p><b>[1]</b> <b>[consultation]</b></p>	<p>The term "consultation" will be normalized with code 83891005 "solo practice private office (environment)", which includes among the synonyms "<i>medical consultation</i>".</p>
<p><b>[2]</b> <b>[specialties]</b></p>	<p>Specialty clinics are normalized with the branch starting from <i>clinic</i>.</p> <p>Example:</p> <div data-bbox="504 1034 1353 1368" style="border: 1px solid #ccc; padding: 10px; margin: 10px 0;"> <p>Parents</p> <ul style="list-style-type: none"> <li>&gt; ● Clinic (environment)</li> </ul> <div style="margin-top: 10px;"> <ul style="list-style-type: none"> <li>● Asthma clinic (environment) ☆ ⓘ No attributes</li> <li>SCTID: 702823004</li> <li>702823004   Asthma clinic (environment)</li> <li> </li> <li>en Asthma clinic (environment)</li> <li>en Asthma clinic</li> </ul> </div> </div>
<p><b>[3]</b> <b>[health service]</b></p>	<p>General concepts such as "health service", "clinical services", "sanitary services" are standardized to:</p> <p>43741000  site of care (environment)</p>
<p><b>[4]</b> <b>[service vs. department]</b></p>	<p>For hospital departments and services, as well as for specialty hospitals, we will try to use the most specific code of the <i>environment</i> category whenever possible.</p> <p>For example, under pediatrics service we find (1) 310066004  pediatrics service (qualifier)  and (2) 309945009  pediatrics department (environment) .</p>



	<p>In the mention Cardiology Hospital, we use 309915006 - Cardiology department (environment).</p> <p>When no equivalent is found in the environment category, the only one offered by SNOMED will be chosen, for example, in the mention of child and adolescent psychiatry (qualifier) there is only 310117003 - child and adolescent psychiatry service (qualifier).</p> <p>For specialties where it is not specified whether it is a department, service or team (e.g. "Cardiology", "Digestive"), we will always use the code for department.</p>
<p><b>[5] [primary care]</b></p>	<p>Mentions of primary health care services (general practitioner, community medicine, family doctor) will be standardized as:</p> <p>702855004  family medicine clinic (environment)</p> <p>When <i>ambulatory</i> is mentioned, it will be standardized as:</p> <p>35971002  ambulatory care site (environment) </p>
<p><b>[6] [intermediate stay]</b></p>	<p>Mentions of <i>mid-stay</i> or <i>intermediate stay</i> admissions are standardized with the generic hospital code.</p> <p>22232009  hospital (environment)</p>
<p><b>[7] [association]</b></p>	<p>Mentions of associations, foundations, etc. are NOT normalized as clinical departments but as 310134006  social services (qualifier) . are NOT normalized as clinical departments but as 310134006  social services (qualifier) .</p>
<p><b>[8] [drug rehabilitation]</b></p>	<p>Mentions of drug rehabilitation centers, such as Therapeutic Communities, are standardized with the code</p> <p>20078004  substance Abuse Treatment Center (Environment)</p> <p>If it is an exclusive alcohol treatment center, we will use 274517002  alcoholism detoxication center (environment) .</p>



<p><b>[9]</b> <b>[psychiatry]</b></p>	<p>In mentions of psychiatry consultations we use:</p> <p>702833007   Child and adolescent psychiatry clinic (environment) </p> <p>702914003  psychiatry clinic (environment)</p> <p>In cases of emergencies and admissions to psychiatric services we use:</p> <p>309959002  child and adolescent psychiatry department (environment)</p> <p>309958005  department of psychiatry (environment) </p>
---	--

15. CODES FOR TRANSPORT	
<p><b>[1]</b> <b>[trip]</b></p>	<p>All trip types are always normalized using the code for "trip (event)" (<i>SCTID: 420008001</i>).</p>
<p><b>[2]</b> <b>[vacation]</b></p>	<p>When the term vacation appears, we will also use the code "trip (event)" (<i>SCTID: 420008001</i>).</p>
<p><b>[3]</b> <b>[cruises]</b></p>	<p>Since there is no specific code, for cruise ships we will use the code 262044003  Passenger vessel (physical object) .</p>
<p><b>[4]</b> <b>[raft]</b></p>	<p>Since there is no code for this concept, we will use the code for "raft (physical object)" (<i>SCTID: 86118007</i>).</p>
<p><b>[5]</b> <b>[medicalized ambulance]</b></p>	<p>References to medicalized ambulance services (with advanced life support also) are coded with:</p> <p>11424001  ambulance-based care (environment)</p>

## 16. CODES FOR COMMUNITY

### [1] [ethnicity, origin or place of birth]

The annotations with the label COMMUNITY can refer to demonyms, countries of origin or place of birth. We will follow the following hierarchy:

1. "European" and any other ethnicity will be preferentially annotated as "European (ethnic group)" (SCTID: 414152003).
2. In the case of not finding the code dependent on the branch "ethnic group", as in the case of "American", we will use 160516002 |North American origin (finding).
3. If none of the above is found, we can also use geographic location codes, for example, if the mention is *Maghrebi*, 223509000 |North African country (geographic location); if *Cameroonian*, 223521001 |Cameroon (geographic location)|.
4. We will only use "born in" when it is a certainty: 315425009 |born in Fiji (finding)|

### [2] [religious group]

Religious groups annotated with the label community will receive the branch code "person":

1. "Jew" shall be standardized as "Jew (person)" (SCTID: 160543007).

## ANNEX 1. Summary Table of Annotation Rules

This annex includes a summary of the entity annotation rules set out in section 2 of the guidelines. This section is intended as a reference for annotators and not as a replacement for the rules as such. We recommend reading section 2 first, as it includes many examples, as well as detailed explanations of special cases and exceptions.

Label	Description	Rules	Example
<b>General Rules</b>			
<b>All</b>	Common rules for all labels.	[GEN-P1] Spelling	Patient from [ <b>nicaragua</b> ].
		[GEN-P2] Separation	He has been traveling in <b>Europe</b> ( <b>Norway</b> , <b>Denmark</b> and <b>Germany</b> ).
		[GEN-P3] Nesting	He had been in [ <b>Central and [South America]</b> ].
		[GEN-P4] Acronyms	She is taken by the police to the <b>emergency department (ED)</b> .
		[GEN-P5] Adjectives-prep	[...] of the <b>Venezuelan central coastal area</b>
		[GEN-P6] Compl-name	<b>Emergency department</b> physician
		[GEN-N1] Det-prep	Just back from <b>a trip to Tanzania</b>

		[GEN-N2] Generic-sites	His <b>workplace</b> is located near a <b>river</b> , where he washes his tools.
		[GEN-N3] Adv-dem	Once he arrived <b>there</b> , he came into contact with several <b>farm</b> animals.
		[GEN-N4] Events	No evidence was found at the <b>accident site</b>
		[GEN-N5] Websites	He met his current partner on <b>Twitter</b>
		[GEN-N6] Species-procedure	The patient had a <b>Glasgow</b> score of 15
<b>Rules for locations</b>			
<b>GPE</b>	Geographic regions inhabited by people and influenced by the politics, economy, culture and history of their environment are annotated with this label. This includes countries, regions, cities, etc.	[GPE-P1] Examples	Six-year-old girl from <b>Pozuzo</b> , <b>department of Pasco</b> ( <b>central jungle</b> ), admitted to a <b>hospital</b> in the <b>city of Lima</b> in August 2006.
		[GPE-P2] Divisions-1	Native of the <b>province of Punjab</b>
		[GPE-P3] Divisions-2	This August he plans to tour several <b>[[Asian] capitals]</b> .



		[GPE-P4] Neighborhoods	When he arrived in <b>Seville</b> , he stayed in a <b>shared apartment</b> in a <b>marginal neighborhood</b> .
		[GPE-S1] Abroad	It has been fifteen years since his last trip <b>abroad</b>
		[GPE-S2] Post-more-country	The first case of coronavirus in <b>our country</b> is presented
		[GPE-S3] Sprep-demo	He underwent surgery in <b>his country of residence</b>
		[GPE-S4] Coordinates	From the <b>Sierra de Parima (2° 47' 21" N, 64° 13' 55" W)</b> .
<b>GEO</b>	Geographical features, natural locations in general, types of ecosystems and environments, areas of the world, etc.	[GEO-P1] Examples	Construction worker with regular exposure to asbestos and 3 years in a <b>coal mine</b> .
		[GEO-P2] Acc-nom	This summer he bathed in the <b>Mediterranean Sea</b> .
		[GEO-P3] Cardinal	Gazpacho is a typical <b>southern</b> dish.
		[GEO-P4] Adjectives-prep	She owns a <b>house</b> in the <b>foothills</b> of the <b>Venezuelan central coastal mountain range</b> .

		[GEO-P5] Habitats	He was working in a <b>jungle region</b> of the <b>country</b> .
		[GEO-P6] Endemic	He was visiting a <b>babesiosis endemic region</b> .
		[GEO-N1] Astronomical	The patient presents with delusions in which he believes he works as an astronaut and lives with his family on <b>Neptune</b> .
		[GEO-N2] Waters	He did not report any specific risk behaviors during the <b>trip</b> except for bathing in <b>fresh water</b> in <b>Bamako</b> .
		[GEO-S1] Place-consumption	Control of risk factors is prescribed to avoid relapse (control of money management, avoidance of consumer friends and <b>places of consumption</b> , avoidance of being and going out alone).
<b>FAC</b>	Spaces built by humans are annotated with this label, which includes many common locations that we would find in a city: establishments, stores, structures, monuments, etc., as well as different types of health centers.	[FAC-P1] Examples	He reports having worked in a <b>handbag store</b> when he was 16 years old for a year, in a <b>fish market</b> and at a <b>gas station</b> .

	[FAC-P2] Building-type	He works as a waiter at <b>La Barca restaurant</b> .
	[FAC-P3] Floors	Decreases the tendency to isolate himself in his <b>bedroom</b> , spending most of the day in the <b>common room</b>
	[FAC-P4] Hospital-names	Follow-up at [ <b>Hospital Clínic de [Barcelona]</b> ].  He was referred for follow-up to the <b>outpatient clinic</b> of his <b>hospital of origin (Alto Guadalquivir de Andújar)</b> .
	[FAC-P5] Residence	Upon discharge, he returned <b>home</b>
	[FAC-N1] Organizations	She moves to the <b>residential environment</b> of <b>FAISEM (Andalusian Public Foundation for the Social Integration of People with Mental Illness)</b> ,
	[FAC-E1] Ambulatory	Discharged for <b>ambulatory</b> follow-up
<b>Rules for location-related labels</b>		

<b>Community</b>	The following are annotated: demonyms, i.e. adjectives or nouns that refer to a place, as well as ethnicities, religions and sociodemographic elements of the style.	[COM-P1] Name-adj	17 year old <b>Spanish</b> male who comes to <b>hospital</b>
		[COM-P2] Nested	Spent his childhood in an <b>[[African] city]</b> .
		[COM-S1] Foreigner	The patient is a <b>foreigner</b> , although we do not know exactly where she is from.
		[COM-S2] Religions	21 year old <b>Indian</b> patient...
<b>Language</b>	This category is used for the languages mentioned in the text.	[LAN-P1] Language	The patient does not speak <b>Spanish</b> and we are unable to locate a translator.
		[LAN-P2] De-talk	This is very common in <b>English-speaking countries</b> .
		[LAN-S1] Anglo-Saxon	<b>Anglo-Saxon</b> patient, 76 years old, resident in <b>Spain</b> for 22 years.
		[LAN-S2] Language barrier	Not included in the study due to a <b>language barrier</b>
<b>Department</b>	This category includes the different departments and services of health institutions (e.g. dermatology, cardiology), as well as the different parts of hospitals and	[DEP-P1] Departments	Patient seen for the first time at the <b>Pediatric Endocrinology Department</b> of the <b>Maternal and</b>

	health centers.		<b>Child Hospital of Malaga</b> , at 8 years and 6 months of age.
		[DEP-P2] Specialties	The case was studied by <b>Internal Medicine</b> and <b>Neurology</b> and no pathological data were found in the complementary tests.
		[DEP-P3] Rooms	Two weeks after his admission to the <b>gastrointestinal ward</b> , his condition was complicated by a severe gastrointestinal hemorrhage.  The patient was maintained on ventilatory support in the <b>emergency observation area</b> awaiting possible death or organ donation.
		[DEP-P4] Consultation	One year later, the patient returned to the <b>psychiatrist's consultation</b> for severe abdominal cramps.
		[DEP-P5] Equipment	She has been followed since February 2012 by the <b>Primary Care Team</b> for symptom control.
		[DEP-N1] Centers	She was treated in a <b>mental health</b>

			<b>and addiction clinic.</b>
		[DEP-N2] Tests	<b>Emergency</b> tests performed upon arrival
		[DEP-S1] Specific	Admitted directly to the <b>operating room</b> for bullet removal.
		[DEP-S2] Service	With the reappearance of symptoms, he consulted our <b>service.</b>
<b>Transportation</b>	This category includes mentions of transport and movements in general, including the word "trip", mentions of patient movements and means of transport, among others.	[TRA-P1] Trip	He was planning a <b>trip</b> to <b>Eastern Europe</b> with his family.
		[TRA-P2] Travel-types	He reported having made multiple <b>intercontinental trips</b> , with no apparent health problems.
		[TRA-P3] Means-transport	Arrival in <b>Spain</b> by <b>raft</b> in the <b>Canary Islands</b>
		[TRA-N1] Verbs	Three years ago he <b>flew</b> from <b>Morocco</b> to <b>Spain</b>
		[TRA-N2] Trip-figured	Ended up in the ER after a bad acid <b>trip</b>
		[TRA-S1] Wheel-chair	Requires a <b>wheelchair</b> to go to the <b>unit.</b>

	[TRA-S2] Transfer	At the last visit, after palpating the abdomen, his <b>primary care physician</b> assessed the <b>transfer</b> to the <b>hospital emergency department</b> .
	[TRA-S3] Visit	[...] facilitated by the boost provided by the <b>telematic visit</b> introduced during the COVID-19 pandemic.

## BIBLIOGRAPHY .

- 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](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](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.