CT-EBM-SP CORPUS

Version 3

ANNOTATION GUIDELINE

English version

Leonardo Campillos Llanos
Adrián Capllonch Carrión
Ana Valverde Mateos

2021-24

Latest update: December 2024

1. ANNOTATION CRITERIA

The task consists in annotating medical entity types on Spanish texts from journal abstracts about clinical trial studies, published in the SciELO repository,[1] and clinical trial annoucements from EudraCT.[2] Eight entities belong to semantic groups from the Unified Medical Language System® (UMLS) (Bodenreider 2004) referring to:

- Anatomic entities (ANAT)
- Biochemical or pharmacological substances (CHEM)
- Medical devices (DEVI)
- Pathologic conditions (DISO)
- Genes and genetic material (GENE)
- Living beings and virus (LIVB)
- Physiological processes (PHYS)
- Lab tests, diagnostic or therapeutic procedures (PROC).

I addition, five types of temporal entities were annotated based on the TimeML scheme adapted to the clinical domain (Styler et al. 2014):

- Age entities: this class was not used in TimeML, but was needed for the task
- Date: e.g., *2022*
- Duration (also called *set* in the TimeML scheme): e.g., *2 horas* ('2 hours')
- Frequency: e.g., *mensualmente* ('monthly')
- Time or part of the day: e.g., *por la noche* ('at night')

The following types of medication information entities were also marked:

- Contraindicated: e.g., *contraindicación a aspirina* ('contraindication to aspirin')
- Dose or strength: e.g., *150 mg*
- Form: e.g., *gel, pastillas* ('pills')
- Route or administration mode: e.g., *oral*

Negation and speculated information were also annotated:

- Negation cue (Neg_cue) and Negated events: e.g., *sin fiebre* ('no fever')
- Speculation cue (Spec_cue) and Speculated events: e.g., *posible carcinoma* ('possible carcinoma')

The following miscellaneous medical entities were marked:

- Concept: e.g., *inicio del ensayo* ('at study start')
- Food or Dring: e.g., *alcohol*
- Observation or Finding: e.g., *normotenso* ('normal blood pressure')
- Quantifier or Qualifier: e.g., *grave* ('severe')
- Result or Value: e.g., *> 90 mmHg*

---

[1] www.scielo.org
[2] www.clinicaltrialsregister.eu

Lastly, event temporality and experiencer attributes were labeled with these values:

- Past event (History_of or Family_history_of), future event (Future) or Hypothetical.
- Patient (e.g., 'patient'), Family_member (e.g., 'mother') and Other ('surgeon')

To check the type of entity or to understand each concept mention, the following sources were recommended:

- CIMA (Centro de información online de medicamentos): https://cima.aemps.es
- International Clasification of Diseases vs. 10 (ICD-10), Spanish version, electronic edition (3ª ed, 2010), Ministerio de Sanidad, Consumo y Bienestar Social: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_10_mc.html
- Dictionary of Medical Abbreviations SEDOM: http://www.sedom.es/diccionario/
- DrugBank: https://www.drugbank.ca
- MedlinePlus: https://medlineplus.gov/spanish/
- UMLS Metathesaurus Browser (requires a free register as a user): https://uts.nlm.nih.gov//metathesaurus.html
- Vademecum: https://www.vademecum.es/
- Wikipedia
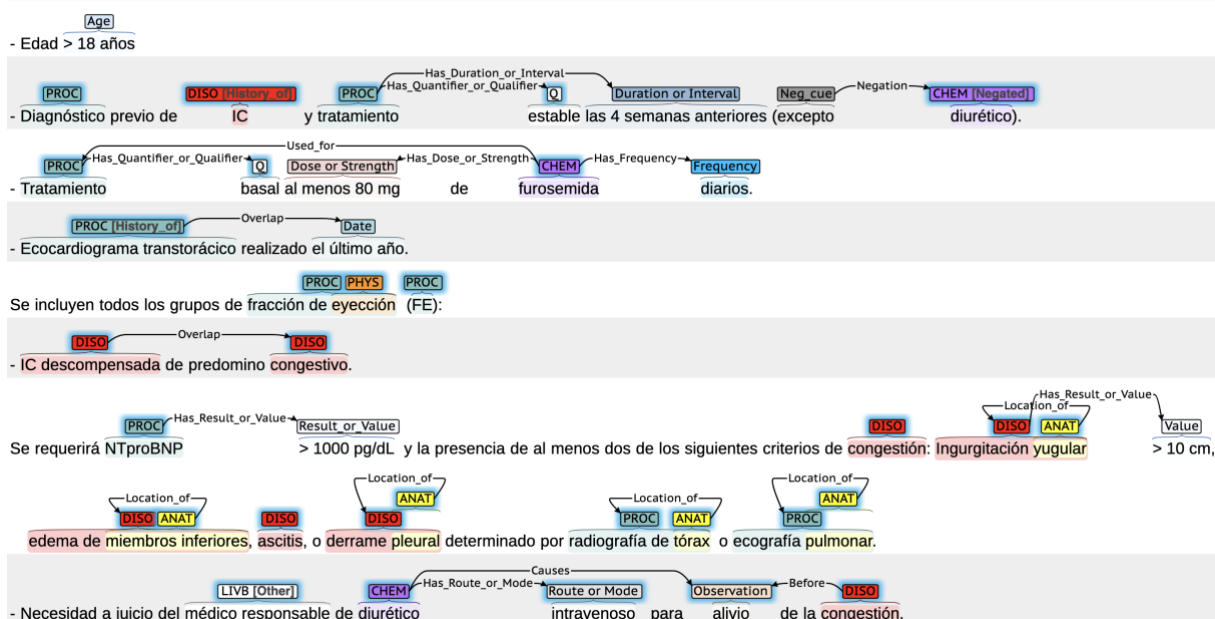
Figure 1. Sample of the annotation

The goals of the annotation are **detecting** entities and relations between them, as menctioned in the scientific literature or clinical texts; and also **normalize** the entities to the corresponding medical concepts. Consequently, annotation criteria were defined considering whether the potential annotation could be mapped to concepts or codes in standard terminologies in the UMLS Metathesaurus. We have especially followed the International Classification of Diseases vs 10 (ICD-10).

SCOPE OF THE ANNOTATION

In general, annotate full phrases if possible:

(1.) el **episodio hemorrágico** (S1130-01082008000800004)
Annotate *episodio hemorrágico* (DISO) and not only *hemorrágico.*

We do not annotate articles and cardinal numbers (*dos, tres...*), nor punctuation marks (comma, dot...) at the begnining and the end of the entity. For example, annotate *episodio hemorrágico* and not *el episodio* hemorrágico; or *eventos adversos* (not *147 eventos adversos*; the semicolon is not annotated either):

(2.) reportaron 147 **eventos adversos**; tuvieron alguna relación causal 13,6%
Annotate only *eventos adversos* (DISO)   (S1729-519X2016000600017)

The start or end of the annotacion cannot overlap with the beginning or the end of other annotation; only some entities are completely allowed inside other entity (see Section Annotation of nested entities):

(3.) DO NOT ANNOTATE: <u>administración de **vacunas vivas**</u> (2018-002521-30)
    ANNOTATE: [**administración**] de [**vacunas vivas**]
        administración (PROC) and *vacunas vivas* (CHEM)

Only the more specific entitites will be annotated, including modifiers expressing medical concepts, adjacent to the medical term:

(4.) **carcinoma renal metastásico**   (S0004-06142006000100006)
Annotate *carcinoma renal metastásico* as DISO*,* and not *carcinoma renal*.

Try to annotate the more specific entity (the wider span):

(5.) el **déficit de calcitriol**   DISO   (0211-699500012698)
Annotate *déficit de calcitriol,* not only *déficit*.

If annotating an adjective or modifier is unclear, check the scope of the term in a terminological resource (e.g., ICD-10 or UMLS Metathesaurus). A list of common modifiers to be annotated (or not) are included in Appendix 1. **Modifiers expressing general concepts are not annotated**: e.g., *próximo, siguiente* (excepting full entities registered in the UMLS: see list in Appendix):

(6.) **hipertensión** futura   DISO   (S0025-76802004000500002)
Annotate only *hipertensión*.

(7.) diferentes **fármacos**   CHEM   (S0034-70942013000300003)
Annotate only *fármacos*.

(8.) **miembros superiores**   ANAT   (S0034-70942013000300004)
Annotate *miembros superiores* (UMLS CUI: C1140618)

In some contexts, derivational adjectives expressing a medical concept are annotated, isolated from the head term (although this might not be annotated). These cases are annotated only **to avoid missing unannotated information**:

(9.) acompañada de mejoría **endoscópica**   (S0025-76802004000500002)
  Annotate *endoscópica* (PROC), but not *mejoría endoscópica*.

(10.) fue llevado a cabo en una sala <u>quirúrgica</u> ortopédica. En total, 60 pacientes ASA I-III seleccionados para **artoplastia electiva total de rodilla** fueron incluidos
                                            (S0034-70942012000300003)
  Do not annotate *quirúrgica* because we already annotate *artroplastia electiva total de rodilla* (PROC)

(11.) El inicio del efecto <u>anestésico</u> fue más corto después de aplicar la **técnica alveolar media**   (S0213-12852017000100005)
  Do not annotate *anestésico* since we already annotate *técnica alveolar media* (PROC)

Very long noun phrases are not annotated, but only the **head of the phrase with its modifiers (not several prepositional groups)**:

(12.) **Estudio comparativo** del uso de mitomicina C en el pterigium recidivante
                                  (S0864-21762008000200004)
  Annotate *estudio comparativo* as PROC, *mitomicina C* as CHEM, and *pterigium recidivante* as DISO

In entities of type procedure (PROC), the prepositional phrase is annotated or not according to the need for understanding the context, or if the full term is found in a reference terminology. **Annotating very long terms is discouraged**. In general:

- Terms with the prepositional phrase are annotated if an UMLS CUI exists, or if they can be mapped to a category from the International Classification of Diseases v.10:

  (13.) dos días antes de la **cirugía de terceros molares**
  (S0034-70942012000400003)
  Annotate as PROC *cirugía de terceros molares* (C0577760 Surgical removal of third molar tooth)

  (14.) **examen en lámpara de hendidura**   (S0365-66912010000400002)
  Annotate the full entity as PROC (C0419360 Slit lamp biomicroscopy)

  (15.) mediante **procedimiento quirúrgico de citorreducción**
  Annotate *procedimiento quirúrgico de citorreducción* (one entity) as PROC (C3850079 Cytoreductive Surgery)   (S0009-741115002339)

  (16.) La **prevención de <u>diabetes</u>** y sus complicaciones es un desafío
  Annotate *prevención de diabetes* (C1659987 Diabetes prevention), *diabetes* (DISO) and *complicaciones* (DISO)   (S0104-11692012000300008)

  (17.) el **control de la hipertensión intracraneal**   (S1130-14732005000100001)
  Annotate *control de la hipertensión intracraneal* as PROC (C0150260 Intracranial Pressure Monitoring)

- Annotate the full term with the prepositional phrase if it helps to understand the meaning of the text; e.g., in entities of type procedure (PROC), each term might distinguish the *Intervention* from the *Comparator*:

(18.) Grupo I (n = 75): **funduplicatura de 360°, corta y holgada por laparoscopia** en el Hospital Universitario Dr. Peset de Valencia. -Grupo II (n = 28): **funduplicatura 360°, corta y holgada por vía abierta** (S1130-01082005000500005)
Annotate PROC in *funduplicatura de 360°, corta y holgada por laparoscopia* (Intervention), and in *funduplicatura de 360°, corta y holgada por vía abierta* (Comparator).

In cases different from the above, annotate each term of type PROC and annotate another entity in the term introduced by preposition (e.g., *anestesia con*, *analgesia con*, *cirugía en*, *tratamiento de*, *prevención de*, *control de*, etc.). Terms introduced by prepositions tend to refer to a different concept. If in doubt, each term is annotated separately if the full term lacks a category or code in the ICD-10 or UMLS MetaThesaurus:

(19.) **Tratamiento** de la **anorexia** en **hemodiálisis**
(S2254-28842013000500055)
Annotate separately *Tratamiento* (PROC), *anorexia* (DISO), *hemodiálisis* (PROC)

(20.) **analgesia epidural** posquirúrgica con **bupivacaína**
(S0121-07932008000600004)
Annotate separately *analgesia epidural* (PROC) and *bupivacaína* (CHEM)

(21.) **Cirugía** de la **enfermedad por reflujo gastroesofágico**
Annotate *cirugía* (PROC) and *enfermedad por reflujo gastroesofágico* (DISO).
(S1130-01082005000500005)

(22.) **tratamiento** del **pterigium recidivante** (S0864-21762008000200004)
Annotate *tratamiento* (PROC) and *pterigium recidivante* (DISO).

(23.) **aplicación de microcorriente** como **tratamiento** en las **úlceras venosas**
(S0104-11692012000400016)
Annotate *aplicación de microcorriente* (PROC, C2147192, *microcurrent electrical stimulation*), *tratamiento* (PROC), *úlceras venosas* (DISO) and *venosas* (ANAT)

(24.) **prevención** de **fenómenos embólicos** (S0066-782X2010001600013)
Annotate *prevención* (PROC) and *fenómenos embólicos* (DISO)

We tend to annotate in the same entity a general or underspecified term adjacent to a specific term from the same semantic type (introduced by preposition):

(25.) la cicatrización mediante **terapia con VAC** (S0009-741115002911)
Annotate as PROC *terapia con VAC* ('cierre asistido por vacío', C1956080 Vacuum-Assisted Closure); do not annotate *terapia* and *VAC* separately.

(26.) **cirugía de revascularización coronaria** (S0025-76802017000100001)
Annotate one entity, and not *cirugía* and *revascularización coronaria* separately.

When 2 head terms can be annotated with the same label, only annotate one (that with the shorter span):

(27.) aplicación de una **técnica de energía muscular**
Annotate *técnica de energía muscular* (PROC) and not *aplicación de una técnica de energía muscular*
(S1134-80462010000700004)

Likewise, regarding mentions of pathologies (DISO), each entity is annotated separately, provided that each one can be mapped to a different UMLS concept and the full combination of mentions lacks a CUI or they are different ICD-10 categories:

(28.) pacientes **obesos** con **nefropatías proteinúricas**

(S0211-69952009000500007)

Annotate two DISO entities: *obesos* (ICD-10: E66.9) and *nefropatías proteinúricas*

On the contrary, the full mention is annotated in the same entity if there is an ICD-10 category or a concept in the UMLS:

(29.) pacientes con **carcinoma diferenciado de tiroides con metástasis ganglionares cervicales** (S0009-741115002261)
Annotate from *carcinoma* to *cervicales*, because a specific concept exists in UMLS (C2981270).

(30.) **hiperparatiroidismo secundario de la insuficiencia renal**
Annotate one DISO entity (C0271847; Hyperparathyroidism due to renal insufficiency; ICD-10: N25.81)

Similarly, we tend to annotate in the same entity those mentions with terms *síntomas/signos/manifestaciones de* + pathology (DISO):

(31.) **síntomas de depresión** (S0104-11692016000100416)
Annotate one DISO entity (C0086132 Depressive Symptoms).

Long terms with coordinated or juxtaposed noun phrases may be ambiguous to parse; non-head phrases might be interpreted as modified or not by the head term. In this case, we annotate only the shortest entity (independent from the head term):

(32.) Se evaluaron diariamente los **síntomas de infección respiratoria**, **diarrea aguda** y **efectos secundarios** (S0120-41572014000100011)
Annotate 3 DISO entities: *síntomas de infección respiratoria, diarrea, efectos secundarios* (do not annotate one full, long DISO entity).

We annotate in the same way entities referring to an anatomical concept. We annotate the full mention if there is a CUI in the UMLS, or if annotating the full entity is needed to understand the text content and distinguish the most specific entity. Otherwise, we annotate each term separately:

(33.) área de corte de la **capa de fibras nerviosas de la retina**

(S0365-66912007000700004)

Annotate one ANAT entity (C0229219 Structure of nerve fiber layer of retina) rather than two independent entities (*capa de fibras nerviosas, retina*)

(34.) **fibras del músculo** del **ventrículo izquierdo** (S0066782X2010001600011)
Annotate 2 ANAT entities: *fibras del músculo* (C0242697 Muscle Fibers) and *ventrículo izquierdo* (C0225897 Left ventricular structure)

Regarding entities expressing the type of clinical trial, we annotate the full mention (criteria are detailed in the Section explaining the annotation of PROC entities):

(35.) **Estudio clínico prospectivo fase IV abierto y sin grupo control**
(S0716-10182019000100032)

(36.) **estudio fase IV no aleatorizado**   (S0716-10182019000100032)

Brand names of drugs are annotated only if they are mapped to an UMLS CUI or if they appear in CIMA, UMLS or DrugBank. Do not annotate the ® or TM characters:

(37.) los nuevos sellantes adhesivos, como **Tissucol**®
Do not annotate ®, but only *Tissucol,* as CHEM   (S0009-741115002261)

(38.) dieta-formula normocalórica e hiperproteica (Vegestart complet®)
Do not annotate *Vegestart complet* (not in UMLS or CIMA)
(S0212-1611201000600008)

As an exception, the ® character is annotated if it occurs in the middle of a multiword term (to avoid discontinuous annotations):

(39.) evaluar si el **Quantiferon® TB-gold** In Tube (QFT-GIT) puede contribuir
                                                    (S0211-69952011000200009)
Annotate *Quantiferon® TB-gold* (CHEM, C1875713 QUANTIFERON-TB GOLD) inside *Quantiferon® TB-gold In Tube* (PROC); also, annotate PROC in the abbreviation (*QFT-GIT*)

We annotated the terms after the hyphen in mentions with *pre-* or *post-* prefixes (or similar ones). Only the most frequent occurrences reported in the medical literature (*pre-anaesthesia, post-surgical*...) are annotated as full temporal entities (with the *pre/post-* prefix):

(40.) post-**CPRE**   (S1130-01082017000300002)
Annotate PROC in *CPRE* ('colangiopancreatografía retrógrada endoscópica')

(41.) **PMB** (**presión máxima basal**) pre-**tratamiento**
(S1130-01082004001200006)
Annotate PROC in *PMB, presión máxima basal* y *tratamiento*; and also annotate Time in *pre-tratamiento*.

(42.) sin-**CEC** vs. con-**CEC**   (S0025-76802017000100001)
Annotate *CEC* ('circulación extracorpórea') as PROC

ANNOTATION OF NESTED ENTITIES

Entities included in other entities with a wider scope will be anotated, provided that both do not have the same semantic group. For example, in the following context, the inner entities *nervio óptico* or *cadera* (ANAT) **will not be annotated** inside the wider entitity *cabeza del nervio óptico* or *músculos abductores de la cadera* (also ANAT):

(43.) cabeza del **nervio óptico**   ANAT   (S0365-66912007000700004)

(44.) músculos abductores de la **cadera**   ANAT
(S0325-00752016000600009)

Examples of nested entities to be annotated:

(45.) transaminasas hepáticas se conservaron en rangos normales
(S0716-1018201900010032)
Annotate *transaminasas hepáticas* (PROC), *transaminasas* (CHEM), *hepáticas* (ANAT)

(46.) medición de **glucosa** (S0212-16112011000300017)
Annotate *glucosa* (CHEM) inside *medición de glucosa* (PROC)

(47.) **hallux** valgus (S2306-41022015000200007)
Annotate *hallux* (ANAT) inside *hallux valgus* (DISO).

(48.) toxinas **urémicas** (S0211-69952013000700005)
Annotate *urémicas* (DISO) inside *toxinas urémicas* (CHEM)

When overlap occurs, nested entities cannot belong to two different entities with a wider span. Inner entities can only belong to one or another entity with a wider span, but not to two different ones. For example, in this context:

(49.) DO NOT ANNOTATE: cierre con sutura de la herida (S0009-741115002911)
ANNOTATE: [**cierre con sutura**] de la [**herida**]

Two annotations are possible: *cierre con sutura* (PROC) or *sutura de la herida* (PROC). However, annotations overlapping across different entities are discouraged, thus we annotate *cierre con sutura* (PROC) and *herida (DISO)*.

ANOTATION OF COORDINATED ENTITIES

In some contexts, coordinated or juxtaposed terms, or enumerations, need to be included in the same annotated entity, **provided that they occur in the same sentence**. **Several entities are not annotated if occurring across sentences**, or if **they include too many words** (e. g., more than 10).

(50.) **artroplastia de cadera o rodilla** (2018-002465-18)
Annotate one PROC entity (C0086511 Knee Replacement Arthroplasty; C0186193 Arthroplasty of hip)

Annotate in one entity coordinated mentions with *y/o*:

(51.) la incidencia de **náuseas y/o vómitos** (S1134-80462016000200002)
Annotate one DISO entity (C0027498 Nausea and vomiting)

However, in some contexts it is not possible to annotate the full mention (for example, because several words or noisy items appear). A solution can be annotating only the common head word, or the most general term:

(52.) **AIJ**: 5 tenían la forma pauciarticular; 9, la poliarticular y 6, la sistémica.
(S0034-75312010000100001)
Annotate only as DISO *AIJ* ('artritis idiopática juvenil', C1444845, Juvenile idiopathic arthritis)

In some contexts, part of the entity is elliptic (very often, with words such as *versus* or *vs*). To avoid annotating mentions that could not be generalized, only the full entity is annotated (the modifier of the elliptic word is not marked). We will consider the fact that the full, non-elliptic entity can be annotated in other context of the same text:

(53.) Calidad de vida de pacientes con cáncer de próstata en tratamiento con **bloqueo androgénico continuo** vs intermitente: estudio prospectivo (...) Fundamento: El tratamiento con **bloqueo androgénico intermitente** (BAI) pretende mejorar la calidad de vida. (S1137-66272015000200003)
Annotate *cáncer de próstata* (DISO), *bloqueo androgénico continuo* (PROC) y *bloqueo androgénico intermitente* (PROC).
Do not annotate *vs intermitente,* because the entity referring to that concept is annotated elsewhere in the text.

In the case of PROC entities that express a measurement or analitic procedure, two (or more) different entities are annotated when occurring in coordinated or juxtaposed structures:

(54.) Tomamos datos de **niveles de hemoglobina** y **albúmina**
(S0212-16112016000200036)
Annotate two PROC entities (*niveles de hemoglobina*; *albúmina*), but not only one entity

SPELLING AND TYPOGRAPHICAL ERRORS

Entities with spelling, typographical and tokenization errors are **also annotated,** although they do not follow the academic or normative use:

(55.) náusea y vomito **post operatorio** (S0066-782X2011000300008)
Annotate DISO, although the normative forms are *postoperatorio* o *posoperatorio*

We did not annotate abreviations with spelling errors, which are not common in the research community and cannot be generalized to other contexts (e.g., *EICA* instead of *IECA*).

ABBREVIATIONS AND ACRONYMS

Abbreviations and acronyms of medical entities are also annotated:

(56.) diabéticos tipo 2 (**DM2**) (S0212-16112017000300532)
Annotate *diabéticos tipo 2* and acronym *DM2* as DISO

(57.) con índice de masa corporal (**IMC**) de 30,7 (S0211-69952012000200010)
Annotate as PROC *índice de masa corporal* and *IMC*

We also annotate English abbreviations and acronyms, even though they are not translated or adapted to Spanish:

(58.) marcadores de inflamación (**PCR, PAI-1, TNF-alpha** ...
Annotate *PCR*, *PAI-1* and *TNF-alpha* (S0212-16112014001200013)

Do not annotate one-letter acronyms or abbreviations: e.g., *P* (standing for 'phosphorus') is not annotated.

Abbreviations specifying biological entities are annotated: e.g., entities referring to lymphocytes (ANAT), biochemical receptors (CHEM), etc.

    (59.) **linfocitos TCD4**   ANAT   (S0716-10182019000100032)

Abbreviations are annotated as full words. Parts of a given word will not be annotated, although they are abbreviations:

    (60.) saturación periférica de oxígeno (SpO2) (S0716-10182019000100032)
    Annotate *SpO2* (PROC), not *O2*.

Abbreviations and acronyms of proper names are not annotated:

    (61.) consumieron dos fórmulas en cuatro ocasiones (Glucerna SR(r) Laboratorios Abbott C.A [FG] y Enterex Diabetic(r), Victus, C.A [FE]
    Do **not** annotate *SR, C.A, FG* or *FE*.   (S0212-16112017000300532)

Spelling errors in abbreviations are not annotated:

    (62.) el grupo tratado con EICA (...) pacientes inicialmente tratados con **IECA**
    Annotate only *IECA* (CHEM), but not *EICA*   (0211-699505018925)

Abbreviations/acronyms are not annotated if they are not commonly used in the scientific litterature. We do not annotate them when they are used as *ad-hoc* anaphoras to shorten a text. We recommend looking up abbreviations/acronyms in resources such as SEDOM (www.sedom.es/diccionario/), or checking their frequency in publications available at Google Books or Google Scholar (see list in Appendix).

    (63.) solución glucosada (**SG**)
    In this context, *SG* is annotated because it is registered in SEDOM.

    (64.) Se incluyeron a 30 pacientes, 14 en el grupo arterial (GA), 10 en el grupo venoso (GV)   (S0066-782X2009001000010)
    In this context, we do not annotate abbreviations because they are just used in this text, they are not widespread in the scientific literature.

When an abbreviation or acronym to be annotated occurs in the middle of a multiword term, the full entity is marked:

    (65.) la **enfermedad de Crohn (EC) fistulosa**   (S1130-01082004000800004)
    Annotate as DISO *enfermedad de Crohn (EC) fistulosa* (C3888864 Fistulising Crohn's disease)

If the abbreviation occurs at the end of the full term, it is annotated separately:

    (66.) el **umbral de dolor a la presión** (**UDP**) (S1134-80462010000700004)
    Annotate as PROC two mentions: *umbral de dolor a la presión* and *UDP*

ENTITIES EXPRESSING ASPECT

We annotate entities expressing recurrence or relapse as Observation (see Sect. Miscellaneous medical entities), since the UMLS classifies them as Findings:

(67.) Hubo 2 casos (3,8%) de recurrencia de **coagulopatía**
Annotate *coagulopatía* (DISO) and *recurrencia*   (Observation)
(S0121-07932007000300002)

(68.) La duración del **tratamiento** disminuyó el número de recaídas
(S1665-11462016000500309)
Annotate *tratamiento* (PROC) and *recaídas* (Observation)

(69.) Analizamos las recidivas de la **pancreatitis** (S1137-66272003000300004)
Annotate *pancreatitis* as DISO and *recidivas* as Observation.

We only annotate DISO in multiword entities that cannot be split up (e.g., *recurrencias isquémicas, recidivas tumorales*) or if there is an ICD-10 class:

(70.) pacientes adultos con **mieloma múltiple recidivante**   (2017-001465-24)
Annotate DISO (C2349261 Relapse multiple myeloma, ICD-10: C90.02)

(71.) portadores de **pterigium recidivante**   (S0864-21762008000200004)
Annotate DISO (C0155158 Recurrent pterygium, ICD-10: H11.069)

(72.) **Aborto recurrente**   (2017-001878-42)
Annotate DISO (C0000809 Recurrent abortion; ICD-10: N96)

FOREIGN WORDS

Entities adapted straigthaway from English, without being translated to Spanish, are also annotated:

(73.) Se realizó **bypass gástrico** en 54 pacientes...
Annotate *bypass gástrico* as PROC   (S0212-16112006000800002)

In contexts where both the Spanish and the foreign word occur, we only annotate the Spanish term and the abbreviation:

(74.) realizar una **queratoplastia endotelial de membrana de Descemet** (**DMEK**:
Descemet membrane endothelial keratoplasty)
(S0365-66912009000500004)
Annotate *queratoplastia endotelial de membrana de Descemet* and *DMEK* both as PROC; do not annotate *Descemet membrane endothelial keratoplasty.*

ANAPHORIC ENTITIES

Do not annotate pronouns, determiners or adjectives expressing a referential meaning with regard to another entity:

(75.) uso adecuado de **betabloqueantes** o **intolerantes** a ellos
(S0066-782X2010001600003)
Annotate only *betabloqueantes* (CHEM) and *intolerantes* (DISO)

(76.) una técnica útil en el **tratamiento** de dichas **fracturas**
(S1130-05582017000300143)
Annotate *tratamiento* (PROC) and *fracturas* (DISO)

2. ANNOTATED ENTITIES

UMLS Entities

We annotate entities corresponding to a subset of semantic groups from the Unified Medical Language System (UMLS) (Bodenreider 2004); namely the following ones:

- ANAT (anatomic entities)
- CHEM (pharmaco-chemical entities)
- DEVI (medical devices)
- DISO (pathological conditions)
- GENE (genes and genetic material)
- LIVB (living beings and virus)
- PHYS (physiological processes)
- PROC (procedures)

We chose this subset of semantic groups since they fit more adequately the type of data annotated (abstracts of clinical trials studies and trial protocols announcements).

ANAT

Entities expressing parts of the body. We include here entities from the following UMLS semantic types:
- Anatomical Structure: e.g., *cisura*
- Body Location or Region: e.g., *epigastrio*
- Body Part Organ or Organ Component: e.g., *músculo*
- Body Space or Junction: e.g., *muñeca*
- Body Substance: e.g., *sangre*
- Body System: e.g., *sistema urinario*
- Cell Component: e.g., *neurotúbulo*
- Cell: e.g., *basófilo*
- Embryonic Structure: e.g., *neuroectodermo*
- Fully Formed Anatomical Structure: e.g., *cuerpo*
- Tissue: e.g., *tejido nervioso*

Examples:

(77.) La permanencia fue mayor en la **pierna** (S0325-00752018000100041)

(78.) en la insuficiencia **renal** crónica   (0211-699504017174)

(79.) contaminación con **heces**   (S0210-48062010000300010)

Some anatomic entities follow words such as *región, porción, cabeza...,* or *primer, segundo,* etc. We include all terms in the annotetion whenever an UMLS concept exists:

(80.) puntuación para cada **región del colon** (S1022-51292016000400002)
Annotate *región del colon* (C0447522 Region of colon)

(81.) anestesia troncular de **1o dedo**   (S0210-1238201324)
Annotate *1o dedo*

CHEM

We include here entities from the following UMLS semantic types:

- Amino Acid, Peptide, or Protein: e.g., *serina*
- Antibiotic: e.g., *penicilina*
- Biologically Active Substance: e.g., *factor I, neurotransmisor*
- Carbohydrates
- Chemical: e.g., *ozono*
- Chemical Viewed Functionally: e.g., *odorante*
- Chemical Viewed Structurally: e.g., *isopropilo*
- Clinical Drug: e.g., *tacrólimus*
- Element, Ion, or Isotope: e.g., *carbono*
- Enzyme: e.g., *lipasa*
- Hazardous or Poisonous Substance: e.g., *cianuro*
- Hormone: e.g., *progesterona*
- Immunologic Factor: e.g., *vacuna nonavalente*
- Indicator, Reagent, or Diagnostic Aid: e.g., *eluyente*
- Inorganic Chemical: e.g., *fosfato*
- Nucleic Acid, Nucleoside, or Nucleotide: e.g., *ADN*
- Organic Chemical: e.g., *carbonato, ácido graso*
- Pharmacological Substance: e.g., *opioide*
- Receptor: e.g., *receptor de la progesterona*
- Vitamin: e.g., *retinol*

Examples:

(82.) mantenimiento con **adalimumab**   (S1130-01082008001100002)

(83.) pacientes que reciben **acetato cálcico**   (0211-699500012698)

(84.) no recibían **ácido fólico** ni **vitamina B6**   (0211-699501013259)

We annotate together compound chemical and pharmacological entities expressing a standard pharmacological preparation, or referring to compounds that do not exist separately:

(85.) **amoxicilina-clavulánico**   (S0210-4806200600090000)
Annotate in one entity (UMLS C0054066, *Amoxicillin / Clavulanate*)

On the contrary, we annotate separately 2 CHEM entities in contexts where the entities do not refer to a mixture or substance preparation:

(86.) **Varfarina** o **aspirina** en la prevención de fenómenos embólicos
                                        (S0066-782X2010001600013)

Biochemical substances or elements (e.g., *calcio, fósforo, bicarbonato*) might be ambiguous in contexts where the chemical substance can also be interpreted as

laboratory analysis (e.g., when lab results are reported). E.g., *calcio: 8,8 ± 0,5 mg/dl*. Annotation criteria are explained in Section 2 (Criteria regarding ambiguities)

DEVI

Entities referring to medical instruments or devices:

> (87.) El UDP se midió, por medio de un **algómetro analógico**
>   Annotate *algómetro analógico*   (S1134-80462010000700004)

These types of entities can be part of a larger PROC entity: *colocar una sonda* ('probe placement')*, alimentación por sonda* ('probe feeding')*...*

> (88.) garantiza la óptima **colocación del catéter** (S0211-69952010000300014)
> Annotate *colocación del catéter* as PROC; and *catéter* as DEVI.

> (89.) post **extracción de sonda**   (S0004-06142008000700005)
> Annotate *extracción de sonda* as PROC; and *sonda* as DEVI.

⚠️ Note the difference in the annotation: verbs such as *sondar, sondaje, o cateterizar* are anotated PROC (procedures using a medical device such as a probe or catheter); however, *sonda* ('probe') is annotated as DEVI.

The scope of the annotation of DEVI entities might cause doubts; terminological resources (e.g., UMLS) will be used to annotate the wider scope, as registered in a thesaurus. Models, brand or registered names will not be annotated if no CUI exists:

> (90.) Una **sonda nasogástrica** 20 F fue insertada
> Annotate *sonda nasogástrica* (C0085678) but not *sonda nasogástrica 20 F* (the full entity is not registered in the UMLS).   (S0212-16112011000100009)

When an entity can be interpreted as a procedure (PROC) or a medical device (DEVI), we prioritized annotating PROC:

> (91.) evaluar el uso sistemático de **enemas**   (S1022-51292011000400003)
> Annotate *enemas* as PROC and not DEVI

DISO

We include here entities from the following UMLS semantic types:

- Acquired Abnormality: e.g., *callo*
- Anatomical Abnormality: e.g., *arruga*
- Cell or Molecular Dysfunction: e.g., *gen mutado*
- Congenital Abnormality: e.g., *dismorfia*
- Disease or Syndrome: e.g., *diabetes*
- Experimental Model of Disease: e.g., *sarcoma de rous*
- Injury or Poisoning: e.g., *traumatismo*
- Mental or Behavioural Dysfunction: e.g., *bulimia nerviosa*
- Pathologic Function: e.g., *shock alérgico*

- Neoplastic Process: e.g., *tumor*
- Sign or Symptom: e.g., *fiebre*

Examples:

(92.) adalimumab en la **enfermedad de Crohn**   (S1130-01082008001100002)

(93.) pacientes **hipertensos** con **diabetes tipo 2** y **albuminuria**
(0211-699501013956)

(94.) los **signos y síntomas de depresión** se evaluaron
(S0104-11692016000100416)
Annotate in the same span *signos y síntomas de depresión* (C4716638 Signs and symptoms of depression)

In the current version, severity degrees of pathologies are only annotated within the DISO entity if an ICD-10 category exists:

(95.) Indicación pública: **miastenia grave**   (2013-003589-15)
Annotate *miastenia grave* (C0026896 Myasthenia Gravis; ICD-10: G70.0)

(96.) **deterioro cognitivo leve**   (2015-004238-85)
(C1270972 Mild cognitive impairment; ICD-10: G31.84)

Stages, severity modifiers or degrees of disorders are marked with the Quantifier_or_Qualifier label (see section Miscellaneous medical entities):

(97.) **ERC** estadio 3A-4   (S1134-009615002090)
Annotate *ERC* as DISO and *estadio 3A-4* as Quantifier_or_Qualifier

(98.) **Insuficiencia renal** >3B
Annotate *Insuficiencia renal* as DISO

(99.) 72,3% fueron de **laceración** de primer grado (S0104-11692008000300007)
Annotate *laceración* (DISO; C0043246 Laceration) and *de primer grado*
(Quantifier_or_Qualifier)

(100.) 61% de los pacientes refirieron **cefalea** leve (S1130-01082004000400004)
Annotate *cefalea* (DISO; C0018681 Headache, ICD-10: R51) and *leve*
(Quantifier_or_Qualifier)

(101.) cifras consideradas **dolor** moderado   (S0325-00752008000500004)
Annotate *dolor* (DISO; C0030193 Pain, ICD-10: R52) and *moderado*
(Quantifier_or_Qualifier)

(102.) presentaban **obstrucción bronquial** leve o moderada
Annotate *obstrucción bronquial* (DISO; C0221725 Bronchial Obstruction) and *leve o moderada* (Quantifier_or_Qualifier)

Likewise, we do not annotate in the DISO entity the physical or functional state of disorders, neoplasia stages, or values of the assessment of the oncologic quality of life. These are marked as Quantifier_or_Qualifier or Result_or_Value:

(103.) pacientes con **falla cardiaca** estado funcional II-IV por más de 6 meses
(S0120-56332018000200106)
Annotate *falla cardiaca* (DISO), *estado funcional* (PROC) and *II-IV* (Result_or_Value).

(104.) clase funcional ii a iv de la New York Heart Association
(S0211-699519300529)
Annotate *clase funcional ii a iv* (Result_or_Value) and *New York Heart Association* (PROC*)*

(105.) **toxicidad gastrointestinal** Grado 1-2 (S0185-106317300264)
Annotate *toxicidad gastrointestinal* (DISO; C1142499 Gastrointestinal toxicity) and *Grado 1-2* (Quantifier_or_Qualifier)

(106.) estado físico ASA I-II-III   (S0034-70942014000300177)
Annotate *estado físico ASA* (PROC) and *I-II-III* (Result_or_Value)

(107.) ECOG superior a 2   (2017-002238-21)
Annotate *ECOG* (CONC) and *superior a 2* (Result_or_Value).

(108.) 24 pacientes con **cirrosis alcohólica** (8 en estadio B de Child-Pugh)
                        (S1130-01082008000800007)
Annotate *cirrosis alcohólica* (DISO; C0023891 Liver Cirrhosis, Alcoholic, ICD-10: K70.30), *pacientes* (LIVB), *Child-Pugh* (PROC) and *estadio B* (Result_or_Value)

Besides, the UMLS includes the ***Findings*** category in the DISO semantic group. Because of their clinical relevance, these entity types **were included in the Observation category** (Sect. Miscellaneous medical entities).

We only annotate as DISO the entities of semantic type *Finding* if there is a category in the International Classification of Diseases vs 10 (ICD-10). We generally **exclude** categories from the **Z** class referring to Histories, Factors or Social situations that affect the health state. We also annotate as DISO those entities that, in annotators' opinion, **can be interpreted as signs or symptoms, pathologic functions, and/or are key to understand the text content**.

(109.) Treinta y cinco sujetos **asintomáticos** (S1134-80462010000700004)
   Annotate *asintomáticos* as DISO

(110.) **vómitos en "posos de café"**   (2016-003339-39)
   (C0278002 Vomit contains coffee grounds finding; ICD-10: K92.0)

(111.) **infiltrados radiológicos**   (2020-001445-39)
   (C0235896 Pulmonary Infiltrate, ICD-10: R91.8)

(112.) **embarazos múltiples**   (2013-003778-29)
   (C0032989 Multiple Pregnancy; ICD-10: O30.90)
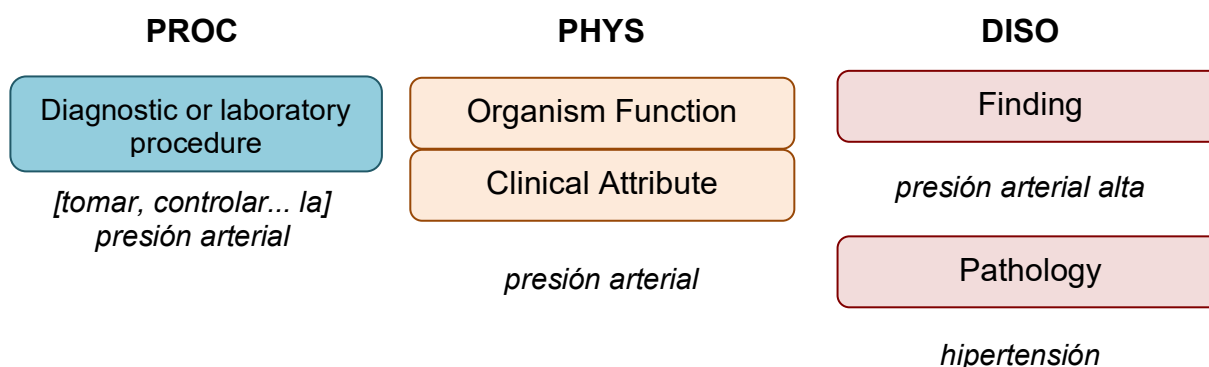
Likewise, we also annotate as Findings the clinical signs revealing a pathologic state and observed in a diagnostic procedure or laboratory analysis; e.g.:

- Neurologic focality (abnormalities in a neurological exploration): e.g., *signo de babinsky, signos meníngeos...*
- Abdominal guarding: e.g., *peritonismo*
- Heart murmurs (abnormalities in a cardiopulmonary auscultation)

(113.) La **kaliemia** no se modificó**.**   (S0211-69952012000200010)
We interpretate that the measurement result (*potassium level*) did not change (*Observation/Finding*), rather than the change in the analysis method (*potassium measurement*)

(114.) Valorar los cambios en el umbral del **dolor** a la presión
                                        (S1134-80462010000700004)
Annotate *umbral del dolor a la presión* (PROC) and *dolor* (DISO)

The following scheme can help to understand the UMLS classification:

PROC	**PHYS**	**DISO**
Diagnostic or laboratory procedure	Organism Function	Finding
[tomar, controlar... la] presión arterial	Clinical Attribute	*presión arterial alta*
	presión arterial	Pathology
		hipertensión

Entities expressing a clinical attribute or organic function are ambiguous (*tensión arterial, frecuencia cardiaca, frecuencia respiratoria, saturación de oxígeno,* etc.). In contexts where these entities co-occur with a number or figure, or with terms such as *medir, valorar,* etc.*,* we interpretate them as 'mesurement' of that attribute and annotate them as PROC:

(115.) midieron la **presión arterial promedio** (**PAP**) y la **frecuencia cardíaca** (**FC**)
Annotate as PROC: *presión arterial promedio, PAP, frecuencia cardíaca, FC*

Contexts such as *serología positiva, prueba de IgG+,* etc. are classed as Laboratory or Test Result in the UMLS Metathesaurus (PHYS semantic group). Some cases are also classed as Findings. Owing to this ambiguity, we only annotate PROC in the term referring to the analytic test (we do not annotate +, *positivo* or *negativo*):

(116.) Sífilis **IgG+**   (2020-001565-37)
Annotate DISO in *Sífilis* and PROC in IgG (without the + character)

However, exceptions are terms such as *VIH+* or *VIH positivo,* which have an ICD-10 category (C4759637 HIV positive status) and are anotated as DISO:

(117.) Tratamiento antiviral o **VIH+**   (2020-001565-37)
Annotate DISO in *VIH+* (also, *tratamiento antiviral* is annotated as PROC)

In this stage of the project, we do not annotate text fragments requiring medical knowledge to infer a pathological state:

(118.) peso al nacer menor o mayor de 800 g (…) **prematuros** con peso al nacimiento entre 500 y 1.250 g   (S0370-41062014000300006)
Only annotate DISO in *prematuros.*

(119.) Quince pacientes con (...) **frecuencia cardíaca** > 65 lpm
Annotate *frecuencia cardiaca* as PROC. The text span expressing the measurement value (*frecuencia cardíaca > 65 lpm*) is **not** annotated as DISO because it requires inferring medical knowledge. Instead, it is annotated as Result_or_Value.

We recommend checking the UMLS Metathesaurus when hesitating about whether to annotate a Finding or not. **Do not annotate anything if the context is unclear**. Do not annotate either coordinate or juxtaposed structures where Findings are separated by several words expressing measurements or laboratory values (increased or reduced):

(120.) la **fosforemia** (5,9 ± 0,6 a 5,0 ± 1,4 mg/dL, p = 0,049), el **producto calciofósforo** (59,8 ± 5,8 a 48,6 ± 12,5 mg2/dL2, p = 0,01) y el **colesterol** (191 ± 29 a 167 ± 33 mg/dL, p = 0,02) <u>disminuyeron</u>.   (0211-699503016284)
Annotate *fosforemia*, *producto calciofósforo* and *colesterol* as PROC.

With regard to mutations, in the current version we only annotate the term *mutación* (C1705285 Mutation Abnormality; Cell or Molecular Dysfunction) or *deleción* (C1511760 Deletion Mutation; Cell or Molecular Dysfunction), or the mutation subtype (e.g., *mutación molecular).* The gen or exon involved is annotated GENE:

(121.) **mutaciones de resistencia** mayores a cualquiera de los fármacos
Annotate *mutaciones de resistencia* (DISO) and *fármacos* (CHEM)   (2016-005226-11)

(122.) **liposarcoma mixoide** con **mutación** PIK3CA (2013-005155-32)
Annotate DISO in *liposarcoma mixoide* and in *mutación*; and annotate *PIK3CA* as GENE

(123.) **deleción** del exón 19 o **mutación** L858R en el exón 21 (2014-002565-30)
Annotate DISO in *deleción* and in *mutación.*

GENE

Entities referring to genes and genetic material.

(124.) **mutación** en uno de los cinco genes HMR (ASXL1, EZH2, SRSF2 e IDH1/2) (2014-004928-21)
Annotate *mutación* (DISO) and *genes, ASXL1, EZH2, SRSF2, IDH1/2* (each with a different GENE entity)

LIVB

We have annotated the most adequate entities for the task, namely following:

- **Microorganisms or pathogens**: viruses, fungi or bacteria (including probiotics). **In contexts of metonymy** (where the mention of a virus refers to the disease: e.g., *virus de la hepatitis B, virus de inmunodeficiencia humana*) we annotate **DISO**, if it can be **interpretated as patological state** (e.g., *VIH+*):

(125.) **Enfermedad hepática** de etiología diferente al **VHC**
Annotate DISO in *enfermedad hepática* and *VHC* ('virus de la hepatis C') (S1025-028X2011000300001)

(126.) pacientes con **VIH/SIDA** naïve
(S0716-10182019000100032)

In contexts such as *infección por VIH* or *infección VHC,* annotate DISO in the full entity and LIVB (as a nested entity) for the name of the virus or bacteria:

(127.) **pacientes** con **infección por VIH**/**SIDA**
Annotate DISO in *infección por VIH* y SIDA; and LIVB in *pacientes* y *VIH*

- **Population groups:** we annotate entities referring to the participants in the clinical trial: *paciente, hombres, mujeres...*

(128.) **Varones** y **mujeres** con **18 años** de edad   (2013-003645-42)
Annotate LIVB in *varones, mujeres,* and Age in *18 años*

Entities referring to **patient groups** (i.e., suffering from a **pathology**: e.g., *diabético, infértil*) are annotated as **DISO**; likewise, patient groups experiencing a **physiological process** (e.g., *embarazadas*) are marked **PHYS**:

(129.) seis **diabéticos** consumieron dos fórmulas (S0212-16112017000300532)
Annotate DISO In *diabéticos,* and CHEM in *fórmulas*

(130.) **anemia ferropénica** en la **embarazada**   (S0212-16112017000300532)
Annotate DISO in *anemia ferropénica,* and PHYS in *embarazadas*

If possible, annotate PHYS entities inside LIVB entities:

(131.) **embolia pulmonar** en las **mujeres** embarazadas   (2012-001505-24)
Annotate DISO in *embolia pulmonar;* LIVB in *mujeres embarazadas* (Pregnant Women, C0033011) and PHYS in *embarazadas*

- **Age group:** *adolescente, sietemesino, adulto, anciano...* These entities also express age and are also marked with Age attribute and a Experiencer value.

(132.)
CHEM   LIVB [Age][Patient]
fibra en   adultos

(133.) **niños** y **adolescentes** de **7 a 18 años** de edad   (2008-005356-25)
Annotate LIVB with Age attribute in *niños, adolescentes*;
and annotate Age in *7 a 18 años*

- **Entities expressing participants' demographic data:** *género femenino, género masculino, caucásico...* The entity with the wider scope will be marked:

(192.) **participantes masculinos y femeninos**   (2018-003711-21)
Annotate LIVB in the full entity (do not annotate only *participantes*)

(193.) **Género femenino**   (2013-000577-77)
Annotate LIVB (C0043210 Woman).

(193.) de **ambos sexos**, edad **entre 18 y 65 años**, de origen **caucásico**
                                                    (2013-001315-71)
Annotate LIVB in *ambos sexos* and *caucásico*, and Age in *entre 18 y 65 años*

Do not annotate LIVB in very specific intervention groups (containing the name of the drug or procedure: e.g., *grupo lidocaína*), except in generalizable expression with a CUI (e.g., *grupo intervención*: C2986530 Intervention Group; *grupo control,* C0009932 Control Groups, *grupo experimental,* C5205743 Experimental Group):

(193.) 53 **pacientes** en el grupo de **Cefazolina**   (S1022-51292011000400003)
Annotate *pacientes* (LIVB) and *Cefazolina* (CHEM) (not *grupo de Cefazolina*)

(192.) el **grupo de tratamiento** recibió **40 mg** de **telmisartán**
Annotate *grupo de tratamiento* (LIVB), *tratamiento* (PROC), *40 mg* (Dose) and *telmisartán* (CHEM)

PHYS

We annotate entities of the following UMLS semantic types:

- Biologic Function: e.g., *ingestión*
- Cell Function: e.g., *meiosis*
- Genetic Function: e.g., *herencia autosómica*
- Molecular Function: e.g., *glucólisis*
- Organ or Tissue Function: e.g., *lagrimeo, tacto*
- Organism Function: e.g., *recuperación, sueño*
- Physiologic Function: e.g., *menopausia, puerperio*

(134.) no se habían realizado **pruebas** en este **embarazo**
(S1020-49892016001200455)
Annotate *pruebas* (PROC) and *embarazo* (PHYS)

(135.) Pacientes en periodo de **menstruación** (2016-000371-24)
Annotate *menstruación* as PHYS

We also annotate PHYS in expressions such as *efecto antiinflamatorio* (C1516019, Physiologic Function) or *efecto adrenérgico* (C3179174):

(136.) La ingesta de omega-3 se ha asociado con **efectos antinflamatorios**
Annotate *efectos antinflamatorios* as PHYS   (S0212-16112014001200013)

We also annotate PHYS in modifiers, adjetives or participle forms expressing pacient groups: e.g., *embarazadas*: están *embarazadas, quedarse embarazadas*...:

(137.) **embolia pulmonar** en las **mujeres embarazadas**   (2012-001505-24)
Annotate DISO in *embolia pulmonar;* LIVB in *mujeres embarazadas* (Pregnant Women, C0033011) and PHYS in *embarazadas*

We did not mark the following semantic types: *Organism Attribute* and *Clinical Attribute*: e.g., *postura* ('Body position', C1262869), *sexo* ('Sex', C1522384)...

We do not mark PHYS in names of institutions, medical departments or organisms:

(138.) Instituto del Sueño   (S0066-782X2011000400006)

PROC

We include here entities from the following UMLS semantic types:

- Diagnostic Procedure: e.g., *colonoscopia*
- Health Care Activity: e.g., *examen clínico*
- Laboratory Procedure: e.g., *hemograma*
- Molecular Biology Research Technique: e.g., *secuenciación del ADN*
- Research Activity: e.g., *farmacovigilancia.* This category mainly includes entities expressing medical research tasks. Statistical analysis (e.g., *t de student, ANOVA, regresión múltiple*, *coeficiente de correlación,* etc.) are classified as concept entities (CONC) in the UMLS; therefore, we do not annotate them as PROC. Consequenty, we do not annotate as PROC other semantically-similar terms, even though they can be classed as PROC in UMLS: e.g., *análisis estadístico* (C0871424).
- Therapeutic or Preventive Procedure: e.g., *diálisis de riñón*

Examples:

(139.) mujeres **hospitalizadas** (S0104-11692016000100416)

(140.) complicaciones en el **cateterismo venoso** (S0104-11692016000100435)

Entities expressing measurements of chemical substantes or laboratory analyses (PROC) may include a CHEM entity. We annotate both the outer and inner entities:

(141.) Los **niveles de PTH** < 120 pg/ml en pacientes en diálisis
Annotate *niveles de PTH* (PROC) and *PTH* (CHEM) (0211-699500012506)

In general, the term *placebo* is annotated as PROC:

(142.) al grupo B se le indicó un **placebo** (S0104-11692016000100434)

General words such as *técnica, aplicación, uso...* are only annotated when they are modified with an adjective or noun phrase specifying or describing the type of procedure; otherwise, do not annotate them (annotate only the more specific terms):

(143.) utilizando la misma **técnica quirúrgica** (S0009-741115001620)
Annotate *técnica quirúrgica* as PROC.

(144.) Su uso en pacientes con **sedación profunda** es seguro y la técnica no se ve modificada (S1729-519X2016000500004)
Annotate *sedación profunda* as PROC, but not *técnica*.

With regard to chemotherapy regimens combining several drugs, each pharmacological substance is annotated separately (CHEM), even though the UMLS Metathesaurus often includes a CUI for the compound term:

(145.) **Nivolumab** más **Ipilimumab** en sujetos con **melanoma** (2016-001941-26)
Annotate as CHEM *nivolumab* and *ipilimumab,* and *melanoma* as DISO.

General verbs expressing analysis, study or data evaluation are not annotated as PROC when they occur without object or complement: e.g., *analizar, comparar, estudia, evaluar, valorar*... We only annotate these verbs with the object specifying the type of procedure or analysis (e.g., *valorar la eficacia y seguridad*).

(146.) **Analizar la eficacia y la seguridad** de la **trombólisis intraarterial**
(S0210-56912010000600003)
Annotate *Analizar la eficacia y la seguridad* (PROC) and *trombólisis intraarterial* (DISO)

Entities such as *ensayo, estudio, análisis, investigación,* etc. are not annotated in isolation; we annotate them in compound terms detailing the type of trial, analysis or study (see more details below):

(147.) **Análisis sanguíneos** que indiquen función hepática anormal
(2019-002626-67)
Annotate *análisis sanguíneos* (PROC) and *función hepática anormal* (DISO)*.*

(148.) Sin contraindicación mayor a ninguno de los **fármacos** en estudio
Do not annotate *estudio.*   (2019-002626-67)

Only annotate *análisis* or *ensayo* without any complement in contexts where they refer to 'laboratory analysis' or 'immunoassay':

(149.) El **ensayo** anti-CCP2 mostró un mejor balance sensibilidad (48,5%) y especificidad (98,0%)   (S1729-519X2018000400540)

(150.) valor de paciente normal para el laboratorio que realiza el **análisis**
(2016-002554-21)

Entities describing or specifying the type of clinical trial are fully annotated, including the phase of the trial, the use of control intervention or placebo, whether it is randomized or not, simple or double blind, unicentric or multicentric:

(151.) **ensayo clínico aleatorizado y controlado (Fase-IIb) a doble ciego**
(S1729-519X2016000500004)

(152.) **ensayo clínico controlado con placebo, aleatorizado y a doble ciego**
(S1025-028X2011000300001)

However, we tend to avoid annotating temporal expressions or the number of participants; that is, we do not annotate specific data that may not generalize enough to other contexts:

(153.) **ensayo clínico prospectivo no aleatorio** de dos grupos de pacientes
Do not annotate *de dos grupos de pacientes.* (S1130-01082005000500005)

(154.) **Estudio prospectivo de intervención no controlado ni aleatorizado** de tres años de seguimiento   (S2011-69952012000200010)
Do not annotate *de tres años de seguimiento.*

Likewise, we do not annotate in the entity expressing the type of trial other entities referring to the intervention therapy or the pathology considered:

(155.) **Estudio de evaluación de la seguridad y la eficacia** en pacientes de **pegunigalsidasa alfa** (**PRX-102**) con la **enfermedad de Fabry** (2018-001148-67)
Annotate as PROC: *Estudio de evaluación de la seguridad y la eficacia*
Also, annotate (out of the PROC entity): *pegunigalsidasa alfa* (CHEM), PRX-102 (CHEM) and *enfermedad de Fabry* (DISO).

To avoid long annotations, we tend to annotate in one PROC entity the type of trial or study, and we annotate the goal or objective in a different PROC entity:

(156.) [**Estudio multicéntrico, aleatorizado, doble ciego, con control activo**] para [**evaluar la eficacia**] en la semana 52 de **secukinumab** (2015-004477-32)
Annotate two PROC entities: *Estudio multicéntrico, aleatorizado, doble ciego, con control activo*, y *evaluar la eficacia*; also, annotate *secukinumab* (CHEM)

The following is a non-exhaustive list of the types of analyses or studies we annotate:

- Efficacy analysis, safety and tolerability studies
- Pharmacokinetic and/or pharmacodynamic studies
- Survival analysis and/or follow-up studies
- Cost-effectiveness analyses
- Intention-to-treat analyses (C2718028)
- Descriptive studies, comparative, prospective or retrospective analyses
- Observational and/or exploratory studies, intervention studies, and pilot studies
- Extension trials and open-label studies (C1709323)
- Comparative and non-comparative studies (two or more cohort groups, nested clinical trials...)

MEDICATION INFORMATION

Mode or administration route (Route)

We mark as Route those entities expressing modes o administration routes of drugs, substances or foof (note that the UMLS classifies these entities as CONC):

(157.) recibieron 5 mg orales de ácido fólico (0211-699501013259)

(158.) infusión de dexmedetomidina intravenosa (S1134-80462015000100002)

(159.) nutrición parenteral (S0370-41062014000300006)
Annotate PROC in *nutrición parenteral,* and Route in *parenteral*

Note that *oral* is annotated as ANAT in the following context:

(160.) acumulación de bacterias en la cavidad **oral**
(S0716-10182013000400003)

Inside PROC entities, Route is marked if it expresses the drug administration route:

(161.) **infusiones intraperitoneales** de **anestésicos locales**
(S0034-70942007000400002)
Annotate *infusiones intraperitoneales* as PROC (C0031146 intraperitoneal infusion), *intraperitoneales* (Route) and *anestésicos locales* as CHEM (C0002934; Local Anesthetics; Organic Chemical; Pharmacologic Substance).

If an entity can be labeled as a procedure (PROC) or the administration route (Route), we annotate PROC:

(162.) recibieron una **infusión** de **ocitocina**   (S0034-70942012000500003)
Annotate *infusión* as PROC and *ocitocina* as CHEM

Note that we do not annotate ANAT nor Route in words expressing a surgical approach, which is marked with the CONC label (see sect. Miscellaneous entities):

(163.) **Cirugía** <u>**transnasal**</u> <u>**transesfenoidal**</u> **endoscópica**
(S1130-14732005000100004)
Annotate *cirugía transnasal transesfenoidal endoscópica* as PROC; *transnasal* and *transesfenoidal* are annotated as Concept (CONC, surgical approach)

(164.) la realización de **biopsia** <u>**prostática**</u> **transrectal ecodirigida**
(S0004-06142009000500003)
Annotate *biopsia prostática transrectal ecodirigida* (PROC), *prostática* (ANAT) and *transrectal* (Concept; surgical approach)

Annotate CHEM in the drug or pharmacological substance, and Route in the entity expressing administration mode or route:

(165.) **pantoprazol** por <u>**vía intravenosa**</u>   (S1130-01082008000800004)
Annotate *pantoprazol* as CHEM and *vía intravenosa* as Route

(166.) el consumo de **anestésicos** <u>**venosos**</u>   (S0034-70942004000100013)
Annotate *anestésicos* as CHEM y *venosos* as Route

(167.) **citrato de fentanilo** <u>**oral transmucoso**</u>   (S1134-80462010000300002)
Annotate CHEM in *citrato de fentanilo oral transmucoso* (C3853047; Transmu-cosal Oral Fentanyl Citrate, Organic Chemical) and *oral transmucoso* as Route

Dose, concentration or strength

Entities indicating the quantity of the drug supplied (e.g., in milligrams, percentage...):

(168.) **tintura de propóleos** <u>**al 10%**</u>   (S0034-75072007000300006)
Annotate *tintura de propóleos* as CHEM and *al 10%* como Dose.

(169.) **calcio** en dosis de <u>**1.000 mg**</u> **cada 24 horas**
(S0212-16112015000500025)
Annotate *calcio* as CHEM; *1000 mg* as Dose; and *cada 24 horas* as Frequency.

Annotate also the number of medication intakes:

(170.) dosis recomendadas de <u>**5 frascos**</u>   (S0121-07932007000300002)

(171.) <u>**múltiples dosis**</u> de **insulina**   (2015-002963-40)
Annotate as Dose: *múltiples dosis*; and *insulina* (CHEM)

(172.) <u>**3 inyecciones**</u> preprandiales de **análogo de insulina rápida**
(2015-002963-40)
Annotate as Dose: *3 inyecciones*; annotate also *preprandiales* (Time), *inyecciones* (PROC) and *análogo de insulina rápida* (CHEM)

Dose expressions in ranges, enumerations or coordinated phrases are marked in one entity:

(173.) **1, 2-4 o más de 4 <u>comprimidos</u> al día** (S1130-63432014000400005)
Annotate as Dose the full entity (*1, 2-4 o más de 4 comprimidos*); annotate also *comprimidos* (Form) and *al día* (Frequency)

(174.) **<u>1-2 inyecciones</u>** diarias de **insulina Detemir** (2015-002963-40)
Annotate as Dose the full entity (*1-2 inyecciones*); annotate also *inyecciones* (PROC), *diarias* (Frequency) and *insulina Detemir* (CHEM)

(175.) **AT** por **vía oral** a razón de **50-100 mg** (S0025-76802004000100001)
Annotate one Dose entity (*50-100 mg*); annotate also *vía oral* (Route) and *AT* (CHEM)

Some unspecified dose entities are also annotated:

(176.) quimioterapia en **dosis altas** (2017-002238-21)
Annotate *dosis altas* as Dose

(177.) **anticoagulantes** a **dosis terapéuticas** (2013-000619-26)
Annotate *dosis terapéuticas* as Dose and *anticoagulantes* as CHEM

However, expressions of dose variation (increase or decrease) are annotated PROC (they are interpreted as a therapeutic intervention):

(178.) Dexametasona (**4mg** cada 8h, re-evaluación a las 72h y **disminución progresiva de dosis**) (S1130-14732009000400003)
Annotate *4mg* as Dose and *disminución progresiva de dosis* as PROC (C1827449 Drug Tapering; Therapeutic or Preventive Procedure)

(179.) permite **disminución la dosis** de morfina
Annotate *disminución de dosis* as PROC (C2348004 Agent Dose Decreased)

(180.) **aumentos de dosis** de **remifentanilo** (S0034-70942014000200073)
Annotate *aumentos de dosis* (C1707810 Dose Increased, PROC).

Drug dosage form

Annotate entities such as: *aerosol, cápsulas, comprimidos, crema, formulación, gel, gotas, tabletas...*

(181.) comparar la efectividad de **800 UI/día** de **D2** (**<u>gotas</u>**) y **D3** (**<u>comprimidos</u>**)
(S0025-76802012000300002)
Annotate *D2* (C0014695 Ergocalciferol) and D3 (C0008318 Cholecalciferol) as CHEM; *gotas, comprimidos* as Form; *800 UI/día* (Dose), *800 UI* (Dose); and */día* (Frequency).

(182.) **12,5 gr** de <u>gel</u> con **250 mgr** de **hidrocloruro de lidocaína intra-rectal**
(S0004-06142009000500003)
Annotate *hidrocloruro de lidocaína* as CHEM; *gel* as Form; *12,5 gr* and *250 mgr* as Dose; and *intra-rectal* as Route.

(183.) **montelukast** en <u>tabletas masticables</u> de **5 mg**
(S0120-41572012000300010)
Annotate *montelukast* as CHEM; *tabletas masticables* as Form; and *5 mg* as Dose.

The word *solución* ('solution') is annotated CHEM in compound entities indicating the drug substance (e.g., *solución glucosada*), or if we can interpret that *solución* implies a solution containing a diluted drug:

(184.) uso de la **solución glucosada** por **vía oral** (S0325-00752008000500004)
Annotate *solución glucosada* as CHEM, and *vía oral* as Route.

(185.) Las **soluciones hipotónicas** se han vinculado a la producción de **hiponatremia iatrogénica** (S0325-00752013000400005)
Annotate CHEM in *soluciones hipotónicas* (C0020677 Hypotonic Solutions) and DISO in *hiponatremia iatrogénica*

Contraindicated

Pharmacological entities (CHEM) or procedures (PROC) are marked with the attribute Contraindicated, generally when introduced with words such as *contraindicado, contraindicación,* etc.:

(186.) contraindicación conocida a **corticoesteroides** (2020-001278-31)
Annotate *corticoesteroides* (CHEM, attribute Contraindicated)

(187.) Pacientes con contraindicación para **anestesia espinal** (2017-001430-25)
Annotate *pacientes* (LIVB) and *anestesia espinal* (PROC, Contraindicated)

Contraindicated is also marked in other PROC or CHEM entities introduced by words or expressions implying that the use of the medication is not recommended (e.g., because it would be harmful to the patient):

(188.) Enfermedades que limiten el uso de **ibuprofeno** y/o **paracetamol** (2019-000902-31)
Annotate CHEM in *ibuprofeno* and *paracetamol* with attribute Contraindicated

(189.) Uso de **medicación** prohibida en el estudio (2014-005419-18)
Annotate *medicación* (PROC) with attribute Contraindicated

TEMPORAL ENTITIES

We followed the TimeML[3] standard (TIMEX3 version) to annotate time expressions (*TIMEX*). Nevertheless, there are some differences with regard to the scheme of the THYME corpus (Styler et al. 2014):

1) Entity subtypes are annotated as classes and not as attributes of a general temporal expression class.
2) Entities of type *Quantifier* are not annotated.
3) Entities of type *Prepostexp* (with prefix *pre/post*-: e.g., *preoperative*, *postprandial*) are annotated as Date or part of the day (Time).
4) Entities expressing age are annotated (Age).
5) Except for dates, coordinated expressions are annotated in the same entity: e.g., *3, 6 and 12 months* (annotate one Duration entity).

---

[3] https://www.cs.brandeis.edu/~cs112/cs112-2004/annPS/TimeML12wp.htm [Accessed: 26/12/2024]

In summary, we have adapted TimeML standard to our task. For all classes:

1) Only annotate entities with words expressing temporal information (*día, año, mes, minuto,* or numbers of dates: e.g., 2005). Do not annotate patologies or procedures (DISO or PROC) inside a temporal entity. In the TimeML schema, those are considered EVENTS and are linked using temporal relations (TLINK):

(190.) se registró el total de movimientos **al primer minuto** después de la **intubación**
(S0034-70942014000600425)
Annotate *al primer minuto* but not *después de la intubación* (only *intubación* is annotated as PROC)

(191.) **5min** después de la **intervención** y **24horas** después de la **intervención**
(S1134-80462010000700004)
Annotate two Duration entities: *5min* and *24 horas* (*intervención* is PROC)

2) Do not annotate prepositions or adverbs introducing the temporal expression, except in case of ambiguitity (e.g., *mañana,* 'tomorrow' vs. *por la mañana*, 'in the morning'):

(192.) En **1994** se publicaron los resultados   (S0370-41062001000100013)
Annotate *1994* (no *en*)

(193.) comunicación **diaria** por teléfono hasta **el 3º día**
Annotate *el 3º día,* without *hasta* (Date) and *diaria* (Frequency)
(S0034-70942010000200002)

3) Segmentation errors are also annotated:

(194.) **5min** después de la intervención
Annotate *5min*

4) Expressions with ellipsis of temporal words are only annotated if some information to understand the text would be lost if not annotated:

(195.) la frecuencia de sesiones con las que el paciente inicia HD (**una o 2 sesiones por semana**) es menor que en la HD convencional (**3 por semana**). Dicha frecuencia aumenta (de una a 2, y de 2 a 3)   (S0211-699518301425)
Annotate Frequency only in *una o 2 sesiones por semana* and in *3 por semana*

(196.) dexametasona **la primera semana** y fluorometalona **las tres siguientes**
(S0365-66912005000100004)
Annotate *la primera semana* y *las tres siguientes* (otherwise, information is lost)

Age

Age is annotated both in years and months (for child population). Note that the word *edad* is not annotated:

(197.) Diez varones, con una edad media de **59 años**
(S0210-56912010000600003)

(198.) pacientes pediátricos de **12 meses** de edad
(S0120-41572012000300010)

When age is expressed in ranges, only one entity is marked:

(199.) Hombres y mujeres de **18-75 años** de edad   (2013-003588-73)
Annotate one Age entity

Modifiers (*mayor/menor de*) or characters such ≥ as are also annotated:

(200.) mujeres **mayores de 25 años**   (S0185-106317300069)

(201.) Edad : **≥18 años y < 90 años**   (2020-001541-39)
Annotate one Age entity

Translation errors are not marked; e.g., do not annotate ≥ *18 years.*

Entities expressing age groups (e.g., *adolescente, joven, sietemesino, mayor de edad...*) are annotated LIVB with the Age attribute:

(202.) **adolescentes** (edades de **12 a 18 años**)   (2015-003314-24)
One Age entity (*12 a 18 años*) and *adolescentes* as LIVB (with *Age* attribute).

Date

This class gathers entities expressing dates or indefinite temporal location longer than one day. It is not necessary to include the introductory adverb or preposition in these expressions:

(203.) En **1994** se publicaron los resultados   (S0370-41062001000100013)
Annotate *1994* (without *en*)

(204.) Analizados desde el **1/1/1995**   (S0370-41062001000100013)
Annotate *1/1/1995* (without *desde*)

(205.) desde **el tercer mes**   (0211-699505018925)
Annotate *el tercer mes* (without *desde*)

In entities with *hace* + temporal expression ('ago'), *hace* is included in the full expression (so that it is not interpreted as duration):

(206.) **FA hace menos de un año**   (S0066-782X2010001600013)
Annotate *hace menos de un año* (Date) and *FA* (DISO)

Do not annotate as Date some vague expressions such as *al final, al principio...*

(207.) <u>Al final</u> de cada intervención no se observaron cambios
Do not annotate any temporal entity.   (S0212-16112014001200013)

Only annotate vague expressions such as the following ones:

(208.) fecha límite aproximada, **finales de marzo de 2013**   (2013-003866-15)

(209.) un 50% presenta **actualmente** valores negativos
(S0212-71992005001000008)

In the following cases, according to the TimeML criteria, two entities are marked, rather than one:

(210.) entre **Octubre 2001** y **Junio 2013**   (S2306-41022014000400005)
Annotate two Date entities: *Octubre 2001* and *Junio 2003*.

(211.) durante el periodo **2001-2003**   (S0036-36342004000400008)
Annotate two Date entities Date: *2001* and *2003*.

(212.) entre **enero** y **diciembre de 2012**   (S1561-31942015000400005)
Annotate two Date entities.

(213.) estableciendo los controles **al día siguiente, a la semana, al mes y a los 3 meses** de la cirugía   (S0365-66912010000800003)
Annotate 4 Date entities: *al día siguiente, a la semana, al mes, a los 3 meses*

Pre-/post- expressions are generally underspecified and it is not always possible to link them to a date or time frame. These cases will be interpreted according to the context:

(214.) **náuseas y vómito** en el **postoperatorio**   (S0034-70942015000300191)
Annotate *postoperatorio* as Time; and *náuseas y vómito* as DISO

(215.) **pravastatina preoperatoria**   (S0185-106317300069)
Annotate *pravastatina* (CHEM) and *preoperatoria* as Time

Duration

This class is used for entities expressing how long the event lasts.

(216.) tras un seguimiento de **dos años**   (S1698-69462006000200015)
Annotate *2 años*

The preposition or adverb introducing the expression is not marked: e.g., annotate *2 horas* (not *durante dos horas*).

(217.) se dio seguimiento a 88 pacientes durante **4 años**
Annotate *4 años* as Duration   (S0212-16112016000200036)

(218.) Cefazolina 1 g tid EV. x **7 dias**   (S1022-51292011000400003)
Annotate *7 dias* as Duration

Duration is also marked in entities expressing a time interval:

(219.) Se analizaron los datos a los **6 meses** (S0211-699519300529)
Annotate *6 meses* as Duration

(220.) estado vital a **40 semanas**   (S0124-0064_PMID23250319)
Annotate *40 semanas* as Duration

Note that we annotate together in one entity coordinated duration expressions (given that we do not normalize time expressions to time points in our task):

(221.) el seguimiento se realizó a los **12 y 24 meses**
Annotate one Duration entity   (S1137-66272013000300007)

We do not annotate adjetives such as *antiguo, moderno...* Only adjetives such as *anterior, posterior, siguiente,* etc. are marked in relative time expressions:

(222.) en **los siguientes doce meses**, el **clopidogrel** no se administraba
(S0210-56912007000400001)
Annotate *los siguientes doce meses* (Duration) and *clopidogrel* (CHEM)

We also annotate indefinite modifiers or quantifiers such as *y más*, *al menos*...:

(223.) AZT vía oral durante **un mes o más**   (S0370-41062001000100013)
Annotate *un mes o más* as Duration

(224.) **al menos dos días** con pirosis   (S1130-01082006000300003)
Annotate *al menos dos días* as Duration

In contexts of ambiguity, Duration shall be noted, especially when the period or time interval cannot be associated with a chronological date, or is associated to a drug or treatment.

(225.) Después de **dos meses** se midieron los niveles de hemoglobina
Annotate *dos meses* as Duration   (S0120-41572013000300003)

(226.) han iniciado recientemente la anticoagulación oral (**menos de 1 mes** antes de la selección)   (2018-000753-45)
Annotate *menos de 1 mes* as Duration.

(227.) Uso de un fármaco en investigación en los **30 días** previos a la visita (set67/2017-000250-19)
Annotate *30 días* as Duration and *fármaco* (CHEM)

(228.) después de **6 horas** de **bloqueo**   (S0034-70942010000200004)
Annotate *6 horas* (Duration) and *bloqueo* (PROC).

Frequency

Entities that express how often the event occurs, including the frequency of taking or administering a drug. In some annotation standards, this is called Set.

(229.) Las medidas se realizaron **cada dos minutos**
(S0365-66912009000100004)

(230.) Los efectos adversos se dieron **a menudo**   (S0034-70942010000200003)

(231.) placebo (n = 43) **1 vez/día**   (S0212-16112016000400031)

(232.) **800 UI/día**   (S0025-76802012000300002)
Annotate *800 UI* (Dose); and */día* (Frequency).

(233.) Raltegravir (**1200 mg QD**)   (2017-000986-60)
Annotate Frequency in *QD* ('quaque die', 'once a day') and *1200 mg* (Dose)

Abbreviations of frequency expressions not related to the intake of a pharmacological substance or the administration of a therapeutic procedure are not annotated (we do not consider them necessary for the task).

(234.) el filtrado glomerular estimado se redujo (69,8 ± 29,7 vs. 60,25 ± 23,0 ml/min/m2) (S0211-69952012000200010)
Do not annotate */min*

(235.) flujo arterial medio 354,16 ml/min   (S0211-69952010000300018)
Do not annotate */min*

(236.) El grupo I (**infusión**) recibió **0,5 mg/kg/min** de **esmolol**
(S0034-70942014000400247)
Note that we do annotate /min (Frequency) in this context because it refers to the infusion frecuency (PROC) of *esmolol* (CHEM), with a dose of *0,5 mg/kg* (Dose)

(237.) recibió **lidocaína 1.5 mg/kg/h**   (S2448-87712017000200009)
Note that we do annotate */h* (Frequency) because it refers to the frecuency of lidocaine administration (*lidocaína*, CHEM) with a dose of *1.5 mg/kg* (Dose)

Time

Entities expressing hours or parts of the day.

(238.) un ayuno **nocturno** de **12 horas**   (S0034-75072007000400001)
Annotate *nocturno* (Time) and *12 horas* (Duration)

(239.) **10 g por la mañana**   (S0212-16112011000100009)
Annotate *10 g* (Dose) and *por la mañana* (Time). Do not annotate *mañana* as an isolate expression because it is ambiguous with other temporal expressions.

Temporal entities with prefix *intra-,* and expressions such as *preprandial* or *postprandial,* are interpreted as Time:

(240.) El uso de **opiáceos** en el **período intraoperatorio**
(S0034-70942015000300191)
Annotate *período intraoperatorio* as Time, and *opiáceos* as CHEM

(241.) un pico de **glicemia postprandial**   (S0212-16112014000300013)
Annotate *glicemia postprandial* (PROC) and *postprandial* (Time)

However, according to TimeML, Duration is marked in cases where time is measured in hours (e.g., *2 horas, 24 horas*) and if expressions cannot be associated with a part of the day or a chronological time:

(242.) evaluación a **las 4 horas** del **postoperatorio** (S0120-33472007000400003)
Annotate *las 4 horas* (Duration) and *postoperatorio* (Time)

(243.) El **torniquete** fue retirado después de **1 min** (S0034-70942014000300169)
Annotate *1 min* (Duration) and *torniquete* (DEVI)

(244.) a **las 2, 6 y 24 horas**   (S1134-80462015000500003)
Annotate one Duration entity

(245.) El hematocrito **posoperatorio** a las **24, 48 y 72 h**
Annotate one Duration entity; annotate also *posoperatorio* (TIME)
(S1134-80462015000500003)

MISCELLANEOUS MEDICAL ENTITIES

Concept

We annotate concepts or ideas (CONC) that are essential to understand the text and not to lose information. We mark several subtypes:

- Temporal concepts not considered in TimeML:

  fin del estudio (C2983670; Study End Date; Temporal Concept)
  periodo/ fase de lavado (C1710661; Washout Period; Temporal Concept)

- Evaluation or assesment scales not considered PROC in the UMLS:

  Clasificación TNM (C1515169; TNM Staging System; Classification)
  DAS28 (C2732493; Disease activity score 28 joint in rheumatoid arthritis; Intellectual Product)

- Surgical approaches:

  abordaje abierto (C0348025; Open approach; Spatial Concept)
  abordaje laparoscópico (C0393360; Laparoscopic approach; Spatial Concept)

- Statistical or research methods conducted on the clinical trial:

  ANOVA (C0002780; Analysis of Variance; Quantitative Concept)
  coeficiente de correlación de Spearman (C0242929; Spearman Rank Correlation Coefficient; Quantitative Concept)

  ⚠️ Note that quantitative concepts such as *average, mean, median,* etc. are not considered necessary for the task.

- Other concepts to be marked to understand the text and not lose information:

  esperanza de vida (C0023671; Life Expectancy; Group Attribute)
  variables demográficas (C0011292; Demographic Factors; Idea or Concept)

Food or Drink

We mark entities referring to food or drink, which cannot be annotated as a chemical substance, but are essential for interpreting the text: *plátano* ('banana')*, huevo* ('egg')*, clara de huevo* ('egg-white')*, leche* (C2825079; Milk - dietary; Food), etc.

(246.) Consumo de **alcohol** o analgésicos
Annotate *alcohol* as Food_or_Drink, and *analgésicos* as CHEM.

Observation or Finding

Entities referring to health states that are not pathological states nor symptoms. They cannot be classified in any other way, but we need to annotate them to avoid information loss. Measurement results (including verbs of change: e.g., *improved, decreased*) are generally not annotated. We especially annotate Observation or Finding on entities that have clinical relevance; or if they convey information about the criteria that patients must meet in order to participate in the trial.

(247.) **sujetos normotensos** (2018-001148-67)
Annotate *sujetos* (LIVB) and *normotensos* (Observation_or_Finding; C2712122; Normal blood pressure; Finding)

The UMLS clusters the Finding category with the DISO semantic group. We marked with the Observation label the following types of mentions (not exhaustive):

- Findings referring to aspects addressed in the anamnesis: e.g., *divorciado*, C0086170; *alto*, C0241240...
- Findings expressing neuter information (not necessarily a pathologic condition): e.g., *el 25% de los pacientes se hizo normoalbuminúrico* (annotate *normoalbuminúrico* as Observation_or_Finding) (S0211-69952012000200010)
- Too general or vague terms, when they occur isolated: *anomalía, hallazgo, problema...*
- Variations in values of a measurement or lab test that are beyond upper or lower limits of normal: e.g., *aumento de la fosfatasa alcalina* (0211-699503016284; do not annotate *aumento de la fosfatasa alcalina*), *descenso de colesterol LDL* (0211-699505017919X), *reducción de la presión arterial* (S0211-69952012000800013), *disminución de hemoglobina*, etc.
- Abnormal results in laboratory analysis or examinations: e.g., *bazo no palpable, examen neurológico anormal, anomalías de laboratorio...*
- Findings of improvement or worsening patterns: *progresión del dolor torácico* (S0375-090615000543; annotate only *dolor torácico* as DISO). Two exceptions may appear:
  1) *progresivo* (adjetive) is annotated together with a term with an ICD-10 category: e.g., *parálisis supranuclear progresiva* (C0038868, ICD-10: G23.1)
  2) *progresión* or *progresivo* is annotated with terms of oncologic processes, when it expresses the meaning of change in the disease or tumor stage: e.g., *progresión del tumor* (C0178874 Tumor Progression; Neoplastic Process)
- Findings related to life expectancy: e.g., *muerte inminente*
- Findings related with nutritional or functional state, or lifestyle (s*edentario, vegetariano...*). However, we annotate *tabaquismo, fumador* or *uso de drogas* because there are ICD-10 categories (F17.200 and F19.20)
- Findings related with good or bad response to treatment, therapeutic failure... E.g., *fracaso del tratamiento*

Quantifier or Qualifier

We annotate as qualifiers those modifiers or adjectives that are considered essential for understanding the text. In addition, we annotate quantifying expressions needed to specify the intervention or pathology, which are not noted according to TimeML (e.g., *cuatro veces,* 'four times'). Note that figures expressing the size of the study population or the experimental results of the trial do not need to be recorded.

The following is a non-exhaustive list of the quantifiers or qualifiers we annotate:

- Quantifying expressions, including those not considered in the TimeML standard:

    (248.) vitamina B12 mensual por **4 veces**
    (249.) **al menos dos episodios** neumónicos

- Qualifiers or adjetives expressing the status or severity degree of a pathology:

    (250.) efectos adversos **graves**
    (251.) fibrosis **avanzada** (C0205179; Advanced phase; Qualitative Concept)

- Adjectives or numeral entities expressing cancer staging or status:

    (252.) melanoma **irresecable** o metastásico, en **estadio III o IV**

- Classes or degrees of pathological states according to evaluation scales:

    (253.) **clase II a IV** de la New York Heart Association
    Annotate *New York Heart Association* as PROC and *clase II a IV* as Quantifier_or_Qualifier

    (254.) pacientes ASA **I-III**
    Annotate *ASA* as PROC and *I-III* as Quantifier_or_Qualifier

    (255.) Clase funcional **II**
    Annotate *Clase funcional* as PROC and *II* as Quantifier_or_Qualifier

- *Negative,* when it refers to a laboratory test (it is not a negation cue)

    (256.) cultivo posquirúrgico final **negativo**
    Annotate *cultivo posquirúrgico* as PROC and *negativo* as Quantifier_or_Qualifier

- Other modifiers that are essential to avoid information loss:

    (257.) diabetes **mal controlada**
    (C3853134; Poorly controlled; Functional Concept)

    (258.) Cáncer **activo**   (C0679217; Active State; Idea or Concept)

Result or Value

We use this class to mark all mentions to results of diagnostic procedures or laboratory values:

    (259.) *ECOG de 0 o 1*
    (260.) *PSA mayor de 2 ng/ml*

(261.) *los niveles de LDL están dentro del rango local recomendado*

We also mark *negativo* ('negative') or *positivo* ('positive') referred to a test value:

(262.) *positivo para el test de la tuberculina*

NEGATION AND SPECULATION

Negated entities

Entities occurring in negated contexts are also annotated, and the negation cue is marked as *Neg_cue.* Unlike other corpora (e.g., Vincze et al. 2008; Lima et al. 2020), the full scope is not marked. Only the event or negated concept is marked with the attribute Negated:

(263.) 301 **no vacunados** (S1131-57682003000400010)
Annotate *vacunados* (PROC) with attribute Negated; and *no* as Neg_cue

(264.) **sin complicaciones clínicas** (S1729-519X2016000500004)
Annotate *sin* (Neg_cue) and *complicaciones clínicas* (DISO, Negated)

(265.) **no** hubieran recibido **tratamiento antibiótico**
(S1022-51292011000400003)
Annotate *no* (Neg_cue) and *tratamiento antibiótico* (DISO, Negated)

In cases of lexical negation (e.g., *asymptomatic*) only the Negated attribute is marked:

(266.)

With *ninguno/a(s),* the negation is marked in the noun, the event or both:

(267.) **ninguno falleció** o comenzó **hemodiálisis** (S1729-519X2013000300020)
Annotate *ninguno* (Neg_cue), *falleció* (DISO, Negated) and *hemodiálisis* (PROC, Negated)

(268.) **ningún paciente** con **efectos secundarios** (S0120-41572006000500020)
Annotate *ningún* (Neg_cue), *paciente* (LIVB, Negated, Patient) and *efectos secundarios* (DISO, Negated)

In negated coordinated contexts (e.g., *no... ni...*), only annotate the negation cue that is syntactically closer to the negated entity:

(269.) no se observaron cambios en variables antropométricas **ni efectos adversos**
(S0212-16112014001200013)
Annotate *ni* (Neg_cue) and *efectos adversos* (DISO [Negated])

(270.) sin evidencia clínica **ni** de laboratorio de **infección**
(S1022-51292011000400003)
Annotate *ni* (Neg_cue) and *infección* (DISO [Negated])

Negation is not marked in elliptic entities:

(271.) Se formaron 2 grupos: en los que se utilizó **Tissucol®** **(**grupo caso) y en los que <u>no</u> (grupo control)   (S0009-741115002261)
Do not annotate negation, only *Tissucol* (CHEM, without atttribute Negated).

(272.) 94 **pacientes** fueron evaluados para su elegibilidad, de los cuales 54 fueron <u>excluidos</u>   (S0009-741115002261)
Do not annotate negation, only *pacientes* (LIVB, without atttribute Negated).

In laboratory tests results, ***negativo*** ('negative', modifier) might indicate a negative value with regard to a **pathology or medical condition**. In this case, the adjective should be marked both as Result_or_Value and Neg_cue, and the event as Negated. However, it is not annotated (nor the Negated attribute in the entity referring to the test or procedure) if the context could be interpreted as a procedure not conducted:

(273.) <u>test de</u> **embarazo** **negativo** en sangre u orina
                                        (S1022-51292011000400003)
Annotate *negativo* (Result_or_Value and Neg_cue) referred to *embarazo* (PHYS [Negated]); *test de embarazo* (PROC) is annotated, without attribute Negated.

(274.) **Cultivo de orina** <u>negativo</u> al inicio del estudio
                                        (S1022-51292011000400003)
Annotate *cultivo de orina* (PROC); *negativo* is not Neg_cue but Result_or_Value

Do not annotate negation in doubly negated contexts that can be interpreted as a positive statement:

(275.) <u>sin excluir</u> a <u>ningún</u> paciente   (S0210-48062010000300010)
Do not annotate negation (only *paciente,* LIVB); the sentence is interpreted as: 'all patients were included'.

(276.) **ninguna** necesidad de **vasoactivos** (<u>excepto</u> **dopamina**)
                                        (S0210-56912010000700003)
Only annotate *ninguna* (Neg_cue) and Negated in vasoactivos (CHEM). Do not annotate *excepto* as Neg_cue nor the attribute Negated in *dopamina* (CHEM)

Speculated entities

We annotate as Spec_cue ('Speculation cue') those words introducing uncertain or probable events, outward circumstance or risk (e.g., pathologies or medical conditions). The full scope is not marked; the attribute Speculated in only marked on the focus (event or concept: CHEM, DISO, LIVB, PROC...):

(277.) puede mejorar los resultados y minimizar los <u>posibles</u> **efectos indeseables**
Annotate *posible* (Spec_cue) and *efectos indeseables* (DISO, Speculated).
                                        (S0212-16112010000600008)

(278.) **podría** ser útil para la **prevención**   (S0025-76802004000500002)
Annotate *podría* (Spec_cue) and *prevención* (PROC [Speculated]).

Speculated or Spec_cue are not marked in other modality expressions not expressing uncertainty (e.g., *es necesario,* 'it is necessary').

(279.) <u>es necesaria</u> la supresión **ácida** máxima para disminuir el **resangrado**
(S1130-01082008000800004)
Do not annotate Spec_cue; only annotate *ácida* (CHEM) and *resangrado* (DISO)

(280.) El uso de **morfina intratecal** es una <u>opción terapéutica</u> segura
(S1130-01082008000800004)
Do not annotate Spec_cue in *opción terapéutica*; only annotate *morfina intratecal* (CHEM)

We mark Spec_cue and Speculated in contexts expressing both uncertainty and negation:

(281.) **pueden o no** acompañarse de **neutropenia** (S0370-41062002000100005)
Annotate *pueden o no* (Spec_cue) and *neutropenia* (DISO, [Speculated])

Conditional subordinate clauses (*si, en caso de que...* + verbo) express condition or hypothesis and are not marked as speculated. These contexts are marked with the Hypothetical attribute (see section Event temporality attributes):

(282.) la **cirugía** mejora la **AV**, siendo más efectiva <u>si</u> se asocia **ICG**
(S0365-66912006001000005)
Annotate *ICG* (CHEM with attribute Hypothetical), *cirugía* (PROC) and AV (C0200150, Visual acuity testing, PROC*)*

From a linguistic perspective, we annotate contexts corresponding to the epistemic modality, which expresses "the degree of probability of the state of affairs according to the knowledge of the speaker" (Herrero Zorita, 2017:41). We do not annotate as Speculated those contexts where *poder* ('can') expresses a capacity or a real possibility (not uncertain nor speculated):

(283.) La Ropivacaína se puede utilizar cómo anestésico local (S0210-1238201324)

(284.) Los pacientes pueden dar su consentimiento informado

Ambiguous contexts (i.e., can express deontic modality) will not be annotated.

EVENT TEMPORALITY ATTRIBUTES

History_of

The attribute History_of is annotated in entities expressing past medical conditions (DISO) or procedures (PROC); i.e., entities that are not necessarily present at the moment of the writing of the text or the participation in the clinical trial:

(285.) La **hiperglucemia** sin antecedentes previos de **diabetes mellitus** (**DM**)
(S0212-16112015000500025)
Annotate *hiperglucemia* (DISO), *diabetes mellitus* (DISO, with attributes Negated and History_of) and *DM* (DISO, [Negated][History_of])

(286.) Haber recibido cualquier tipo de **vacunación** (2018-000497-30)
Annotate *vacunación* (PROC) with attribute History_of

(287.) Pacientes que hayan recibido un **tratamiento** previo (2016-004441-82)
Annotate *tratamiento* (PROC) with attribute History_of

(288.) Antecedente de **cirugía ocular** (2017-000250-19)
Annotate *cirugía ocular* (PROC) with History_of; annotate also *ocular* (ANAT)

(289.) **Alergia** o **intolerancia** conocida al **producto de investigación**
(2020-001379-34)
Annotate *alergia* and *intolerancia* (both DISO) with attribute History_of; annotate also *producto de investigación* (CHEM).

(290.) **seguimiento** en un **estudio clínico** previo   (2016-004365-16)
Annotate *estudio clínico* (PROC) with attribute History_of; annotate also *seguimiento* (PROC).

In general, History_of is marked only in contexts introduced by words such as *antecedentes, historia de, previo, anterior...* In case of doubt, do not annotate anything (including when the procedure or the patology might happen in the future):

(291.) sufre una **reacción adversa** durante los estudios M16-067 o M16-065
(2016-004676-22)
Annotate *reacción adversa* as DISO, without History_of; it is unclear if the adverse event occurred in the past or might occur in a future study.

We both mark the attributes History_of and Negated in contexts indicating the absence of any medical history:

(292.) Pacientes sin **tratamiento** previo   (2018-001423-40)
Annotate *tratamiento* (PROC) with attributes History_of and Negated; *Pacientes* (LIVB) is also annotated

Likewise, History_of can be annotated with the Speculated attribute:

(293.) Historia <u>sospechada</u> o conocida de **tuberculosis**   (2020-001162-12)
Annotate *tuberculosis* (DISO) with attributes History_of and Speculated

Family_history_of

Entities expressing family history medical conditions are annotated with the attribute Family_history_of.

(294.) Antecedentes familiares de **enfermedad de Alzheimer** (2017-004454-42)
Annotate *enfermedad de Alzheimer* (DISO) with attribute Family_history_of

In disjunctive concepts, both Family_history_of and History_of can be marked:

(295.) Antecedentes personales o familiares de **carcinoma medular de tiroides**
(2012-004413-14)
Annotate DISO with attributes History_of and Family_history_of

Future

We annotate this attribute in entities referring to an event that will take place after the writing or reading of the text. We mark this when explicit context words appear: for example, *previsión de, programado, plan de, futuro…*

(296.) Previsión de **vacunación** con **agentes virales vivos** en los próximos 7 días
(2014-004471-23)
Annotate PROC (with attribute Future) in *vacunación* and CHEM (with attribute Future) in *agentes virales vivos*.

(297.) **Pacientes** programadas para **cesárea**   (2017-001430-25)
Annotate PROC in *cesárea* with attribute Future. *Pacientes* is annotated as LIVB (with attribute Patient).

(298.) **prevención** de una **hipertensión** futura   (S0025-76802004000500002)
Annotate DISO in *hipertensión* (with attribute Future) and PROC in *prevención*.

Hypothetical

We mark this attribute in entities referring to events expressed in contexts introduced with subordinate conjunctions such as *en caso de que, a menos que* o *si* (only if *si* introduces a conditional subordinate clause):

(299.) **Agentes citotóxicos**, a menos que hayan transcurrido **al menos 4 semanas**
                                        (2018-000340-26)
Annotate Duration with attribute Hypothetical in *al menos 4 semanas*; and annotate CHEM in *Agentes citotóxicos* (but without attribute Hypothetical)*.

(300.) Si es **mujer**, debe cumplir los requisitos de **métodos anticonceptivos**.
(2017-003238-96)
Annotate *mujer* as LIVB with attributes Patient and Hypothetical; and annotate PROC in *métodos anticonceptivos* (but without attribute Hypothetical)

EXPERIENCER ATTRIBUTES

We annote this attribute **only in LIVB entities referring to human beings,** and with the following values.

- PATIENT: when the LIVB entity refers to the patient or clinical trial participant, who generally experiments a medical condition, procedure or intervention:

  (301.) **niños** y **adolescentes** con **psoriasis en placas**   (2015-003314-24)
  Annotate LIVB in *niños* and *adolescentes,* both with attribute Patient; Annotate DISO en *psoriasis en placas*.

  We also annotate the value Patient in entities speficied with an adjective such as *sano* ('healthy') or *normal* (they can be a participant in a trial):

  (302.) estudio observacional transversal en 100 **personas** con malaria y 100 **controles** sanos   (S0716-10182014000500006)
  Annotate *personas* (LIVB [Patient]), *controles* (LIVB [Patient]), *malaria* (DISO), *sanos* (Quantifier_or_Qualifier) and *estudio observacional transversal* (PROC)

- FAMILY_MEMBER: we mark this value in LIVB entities referring to a patient's or trial participant's relative; this information is essential to extract information about family history of disorders:

  (303.) Los **sujetos** recibieron 0,2 ml de calostro de su **madre**
                                        (S0212-16112016000200007)
  Annotate LIVB in *sujetos* (with value Patient) and *madre* (with value Family_member); we also mark Dose in *0,2 ml* and ANAT in *calostro*.

- OTHER: we use this value in any other LIVB entity referring to human beings that do not allude to patients, trial participants nor relatives: e.g., healthcare professionals, legal agents, researchers, study sponsors…

(304.) El **sujeto** (y su **padre**/**madre**/**representante legal**, cuando proceda) deben entender y firmar voluntariamente un DCI
Annotate LIVB in *sujeto* (with attribute value Patient), in *padre* and *madre* (both with attribute value Family_member) and in *representante legal* (with attribute value Other); *DCI* is also annotated as CONC.

We do not mark any Experiencer attribute in LIVB entities referring to plants, virus, bacteria and other disease-causing microorganisms.

3. RELATIONS

⚠️ In the current version, relationships between entities in different sentences are not annotated, only within the same syntactic unit. The end of a sentence is marked with a full stop, an ellipsis or a line break.

(305.) se incluyeron 20 **pacientes** en el grupo 0 y 21 en el grupo 1. No se encontraron diferencias entre ambos grupos en cuanto al sexo, edad, **hábito tabáquico**
Do not annotate a relation between *pacientes* (LIVB) and *hábito tabáquico* (DISO)

(306.)
- Historial de **enfermedad pulmonar intersticial**
- **Sujetos** con **enfermedad autoinmune** activa, conocida o **sospechada**.
pulmonar (ANAT) → LOCATION_OF → *enfermedad pulmonar intersticial* (DISO)
Sujetos (LIVB) → EXPERIENCES → *enfermedad autoinmune* (DISO)
sospechada (Spec_cue) → SPECULATION → *enfermedad pulmonar intersticial* (DISO)
Do not annotate a relation between *Sujetos* and *enfermedad pulmonar intersticial*

⚠️ **Only one relation is annotated between two entities.** Where more than one relation can be annotated, the most appropriate one will be chosen according to the relevant criteria.
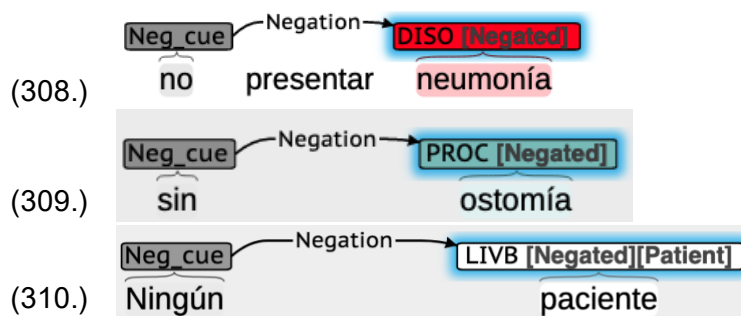
(307.)
complicaciones de la **raquianestesia**
raquianestesia (PROC) → CAUSES → *complicaciones* (DISO)
Do not annotate relation: *raquianestesia* → BEFORE → *complicaciones*

3.1. Relations of negation and speculation

Negation

A relation between a negation cue (Neg_cue, e.g., *no, sin, ningún*) and an annotated entity (DISO, PROC, LIVB...). The full scope of the negation (Scope) is not marked. The Negated attribute is marked on the negated entity that appears in a negated context.

(308.)

Neg_cue — Negation → DISO [Negated]
no presentar neumonía

(309.)

Neg_cue — Negation → PROC [Negated]
sin ostomía

(310.)

Neg_cue — Negation → LIVB [Negated][Patient]
Ningún paciente

When a negation cue refers to more or two annotated entities, one relation is annotated with each negated entity:

(311.)

Neg_cue — Negation → DISO [Negated]
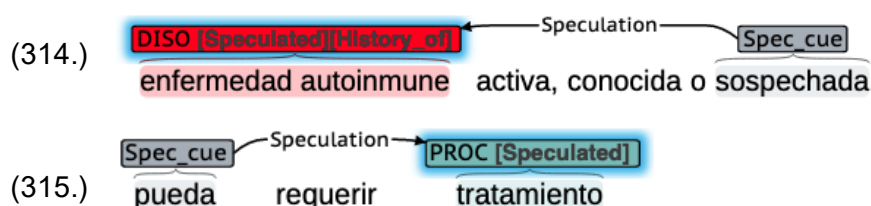sin — Negation → DISO [Negated]
síntomas/signos
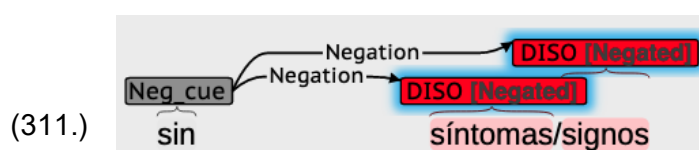
Negation relations are also annotated with concept entities (CONC) and Quantifier_or_Qualifier, in some cases to understand the context:

(312.) *no se encontraron diferencias significativas*
no (Neg_cue) → Negation → *diferencias* (CONC [Negated])

(313.) *casos en los cuales no sea viable la aplicación de injerto antólogo*
no (Neg_cue) → Negation → *viable* (Quantifier_or_Qualifier [Negated])

Speculation

The relationship between a speculation cue (Spec_cue, ej. *posible, sospechado*) and an entity within the uncertainty scope (DISO, PROC...), which is also annotated with the Speculated attribute.

(314.)

DISO [Speculated][History_of] ← Speculation — Spec_cue
enfermedad autoinmune activa, conocida o sospechada

(315.)

Spec_cue — Speculation → PROC [Speculated]
pueda requerir tratamiento

**3.2. Relations with entities expressing drug regimen**

Has_Dose_Strength

Annotated between a Dose entity (dose or concentration) and a CHEM entity.

⚠️ This relation is also annotated between Dose and a procedure entity (PROC)

Has_Route_or_Mode

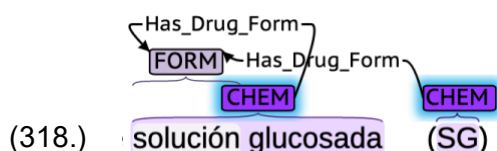Annotated between a Route entity (route or mode of administration) and CHEM.

(316.) QGE031 240 mg s.c. administrado cada 4 semanas durante 52 semanas
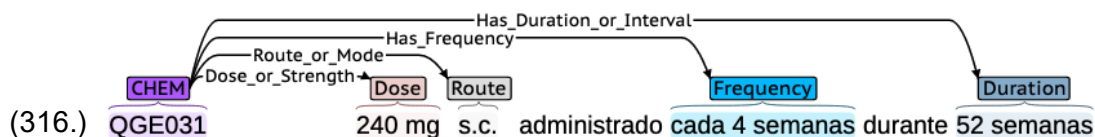
⚠️ This relation is also annotated between Route and a procedure entity (PROC):

(317.) ANALGESIA EPIDURAL

Has_Drug_Form

Annotated between a Form entity (dosage form) and a CHEM entity.

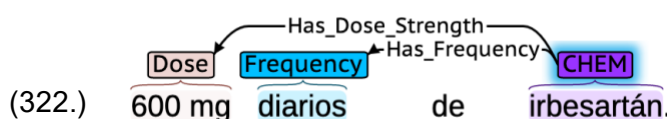(318.) solución glucosada (SG)

Has_Duration_or_Interval

Annotated between a Duration entity and a CHEM entity (example 316).

⚠️ This relation is also annotated between Duration and procedure (PROC), temporal concepts (CONC) or physiological processes (PHYS):

(319.) 3 años de seguimiento

(320.) menos de 12 semanas de gestacion

(321.) Esperanza de vida de al menos 3 meses
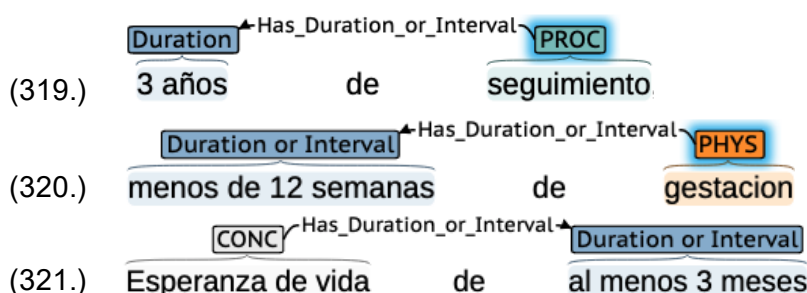
Has_Frequency

Annotated between a Frequency entity and a CHEM entity.

(322.) 600 mg diarios de irbesartán

⚠️ This relation is also annotated between a procedure (PROC) and Frequency:

(323.) podrían mejorar con diálisis diaria.

PROC ─Has_Frequency→ Frequency

⚠️ In some contexts, a drug regimen relation (e.g., Has_Frequency, Has_Dose, Has_Route) can be both annotated with regard to a PROC or a CHEM entity. We give priority to annotate the relation with the CHEM entity:

> (324.) *tratamiento multifactorial* basado en *600 mg* diarios de ***irbesartán***
> Annotate *irbesartán* → Has_Dose_or_Strength → *600 mg*
> Do not annotate Has_Dose_or_Strength with *tratamiento multifactorial* (PROC)

3.3. Temporal relations

The following temporal relations are annotated between these entity types:

1) **Relations between events** (PROC, DISO, CHEM, Observation, PHYS) and **temporal expressions** (TIMEX: Date, Duration, Frequency, Time; and Age).

2) **Relations between events** (PROC, CHEM, DISO, Observation...), **exclusively between those events that are needed to understand the trial methodology and avoid information loss**; especially, when temporal markers appear: *antes de, después de, previo a, anterior a*, etc.

> (325.) **inmunosupresores** tres meses antes de la **cirugía**
> *inmunosupresores* (CHEM) → Has_Duration_or_Interval → *tres meses* (Duration)
> *inmunosupresores* (CHEM) → Before → *cirugía* (PROC)

Temporal relationships are annotated between event type entities (DISO, PROC, CHEM, Observation...) and temporal expressions (TIMEX and Age). For example, duration or interval (Has_Duration_or_Interval) and frequency (Has_Frequency) relationships with respect to drugs (CHEM). A relation (Has_Age) between trial participants (LIVB) and their age (Age) is also marked. The sense of the temporal relationship goes from the event entity to the temporal entity (Age, Date, Duration, Frequency or Time). For example, it starts at the entity CHEM or PROC, and ends at the entity Frequency.

We mark **only the temporal relations between events that are necessary to avoid information loss**. The THYME (TimeML adapted to clinical texts) standard considers several types of temporal links (TLINKs): Before, Overlap, Contains, Starts_on and End_on. For the task of annotating clinical trial texts, we merged these relations in three classes, according to previous work (Sun et al., 2013):

1) **Before**: relations with previous events, and also Before_or_Overlap and Ends_on.
2) **After**: relations with subsequent events, and also After_or_Overlap and Starts_on.
3) **Overlap**: relations expressing the temporal coincidence between two events or between an event and a temporal expression, including temporal senses of simultaneity or full duration of the event along the TIMEX (Contains TLINK).

We do not annotate relations between temporal entities (TIMEX or Age):

> (326.) Entre **2012-2014**, se incluyeron y **siguieron** 40 pacientes
> *siguieron* (PROC) → after → *2012* (Date)
> *siguieron* (PROC) → before → *2014* (Date)
> Do not annotate a relation between *2012* and *2014.*

We consider the following temporal relations:

1) **Has_Age**, **Has_Duration_or_Interval**, and **Has_Frequency** (the two last relations are explained in the section about drug regimen relations).
2) **Before, After** and **Overlap**

⚠️ In case of doubt about how many relations to mark, redundant relations will not be annotated. For example, if we annotate event A → BEFORE → B, and event B → BEFORE → C, do not annotate A → BEFORE → C.

> (327.) **prevenir regresión** en **LASIK**
> Annotate: *prevenir* (PROC) → BEFORE → *regresión* (DISO)
> Annotate: *LASIK* (PROC) → BEFORE → *regresión* (DISO)
> Do not annotate: *prevenir* (PROC) → BEFORE → *LASIK* (PROC)
> Do not annotate: *LASIK* (PROC) → BEFORE → *prevenir* (PROC)

⚠️ If several entities may be annotated, the criterion is to annotate the relation with the syntactic head (e.g., a noun in a noun phrase, or a verb in a verb phrase). For example, in expressions such as <u>infusión de ocitocina,</u> <u>anestesia con isoflurano,</u> etc., the syntactic head has the PROC semantic group*.*

> (328.) **anestesia** con **isoflurano** para **cesárea**
> Annotate: *anestesia* (PROC) → BEFORE → *cesárea* (PROC)
> Annotate: *isoflurano* (CHEM) → USED_FOR → *anestesia* (PROC)
> Do not annotate: *isoflurano* (PROC) → BEFORE → *cesárea* (PROC)

Has_Age

Annotated between a LIVB entity (the participant in the trial) and an Age entity.

(329.) 249 pacientes con edades entre 15 y 44 años.

Before

A relation of precedence between an event (e.g., DISO or PROC) and a time expression (Age, Date, Duration or Time):

> (330.) aparición de las **migrañas** antes de los **50 años**
> *migrañas* (DISO) → BEFORE → *50 años* (Age)

(331.) Posibilidad de realizar **colecistectomía** antes de **dos meses**
colecistectomía (PROC) → BEFORE → *dos meses* (Duration)

(332.) uso de **métodos anticonceptivos** aceptables desde **al menos 2 semanas** antes del **día 1**
métodos anticonceptivos (PROC) → BEFORE → *día 1* (Date)
métodos anticonceptivos (PROC) → HAS_DURATION_OR_INTERVAL → *al menos 2 semanas* (Duration)

We annotate as Before those temporal relations of type Ends_on according to TimeML (THYME):

(333.) El **seguimiento** finalizó en **diciembre del 2016**
seguimiento (PROC) → BEFORE → *diciembre del 2016* (Date)

⚠️ If it is unclear whether to annotate a Before or After relationship:

1) Between an entity and a temporal expression: the arrow of the relation always ends on the temporal expression (TIMEX and Age).
2) Between two events: consider the linguistic marker (*antes de, después de*...) to choose the relation. When no temporal cue is available, we prioritize annotating BEFORE instead of AFTER.

(334.) El **UDP** se midió (..) antes de la **intervención**
Annotate: *UDP* (PROC) → BEFORE → *intervención* (PROC)
Do not annotate: *intervención* (PROC) → AFTER → *UDP* (PROC)

(335.) **operado** después de **quimiorradioterapia**
Annotate: *operado* (PROC) → AFTER → *quimiorradioterapia* (PROC)
Do not annotate: *quimiorradioterapia* (PROC) → BEFORE → *operado* (PROC)

(336.) **cicatrización de heridas**
Annotate: *heridas* (DISO) → BEFORE → *cicatrización de heridas* (PHYS)
Do not annotate: *cicatrización de heridas* (PHYS) → AFTER → *heridas* (DISO)

After

It is the inverse relation of Before:

(337.) **insulina** basal/**posprandial** de **dos y tres horas**
insulina (PROC) → OVERLAP → *posprandial* (Time)
insulina (PROC) → AFTER → *dos y tres horas* (Duration)
No relation is annotated between the temporal expressions (*posprandial* and *dos y tres horas*).

We also annotate as After the temporal relacions of type Starts_on according to TimeML (THYME):

(338.) **intervenidos** de **sustitución valvular aórtica** aislada por estenosis aórtica severa, desde **noviembre del 2013**
intervenidos → AFTER → *noviembre del 2013*
sustitución valvular aórtica → AFTER → *noviembre del 2013*

⚠️ With Duration_or_Interval entities, a relation of type Has_Duration_or_Interval is annotated when we interpretate a duration or continuous process; otherwise, Before or After are annotated.

Overlap

A temporal relation between an event and a temporal expression (TIMEX) or between two events (e.g., DISO, PHYS or PROC) with a temporal coincidence, being simultaneous or not:

(339.) **Posoperatorio** de **cirugía abdominal** (2014-000473-39)
cirugía abdominal (PROC) → OVERLAP → *posoperatorio* (Date)

(340.) **ecografía** del día de la **inducción de la ovulación**
ecografía (PROC) → OVERLAP → *inducción de la ovulación* (PROC)

We also annotate as Overlap those relations that TimeML classes as Contains or During. In these, an event (DISO or PROC) occurs completely during a temporal expression (ej. date or period of time, Date; or part of the day, Time).

(341.) **insulina posprandial** (S0212-16112017000300532)
Annotate *insulina* (PROC) → OVERLAP → *posprandial* (Time)

(342.) desarrollaron en **1963** un **radio-inmuno-ensayo**
(S1561-31942015000400005)
Annotate *radio-inmuno-ensayo* (PROC) → OVERLAP → *1963* (Date)

In summary, Overlap is used to annotate a relation between:

1) An event and a specific or imprecise date expression (Date):

(343.) trasplantado de hígado **hace más de tres años** (2017-004983-37)
trasplantado de hígado (PROC) → OVERLAP → *hace más de tres años* (Date)

(344.) uso **actual** de tratamientos (2015-004477-32)
tratamientos (PROC) → OVERLAP → *actual* (Date)

2) An event and an expression of time or part of the day (Time)

(345.) CPAP en el **periodo postoperatorio** (2017-004983-37)
CPAP (PROC) → OVERLAP → *periodo postoperatorio* (Time)

(346.) primera micción de **la mañana** (2016-002913-23)
micción (PHYS) → OVERLAP → *la mañana* (Time)

3) An event and an interval expression when no duration of the event is expressed, nor a relationship of anteriority or posteriority can be marked:

(347.) reducción septal durante **los 3 meses** anteriores a visita de selección
reducción septal (PROC) → OVERLAP → *los 3 meses*

4) Two events that coincide in time (completely or partially), only in cases where it is necessary to avoid information loss:

4a) With time markers such as *intra-, al momento de...*:

(348.) **urocultivo** intra y post **tratamiento**
urocultivo (PROC) → OVERLAP → *tratamiento* (PROC)
urocultivo (PROC) → AFTER → *tratamiento* (PROC)

(349.) **infección** al momento del **ingreso**
infección (DISO) → OVERLAP → *ingreso* (PROC)

4b) In contexts with markers expressing simultaneity of events: preposition *con,* or expressions such as *asociado a* (if it does not implies cause):

(350.) **fibrilación atrial** (**FA**) asociada a la **enfermedad valvar mitral**
fibrilación atrial (DISO) → OVERLAP → *enfermedad valvar mitral* (DISO)
FA (DISO) → OVERLAP → *enfermedad valvar mitral* (DISO)

(351.) **valvopatía mitral** con **fibrilación atrial** (**FA**)
valvopatía atrial (DISO) → OVERLAP → *fibrilación atrial* (DISO)

(352.) **pacientes** (..) con asma, que presentaban **obstrucción bronquial**
asma (DISO) → OVERLAP → *obstrucción bronquial* (DISO)

⚠️ Only relationships between events are annotated if the relations preserve the information regarding the trial methodology or protocol for patient inclusion, sample collection, or outcome assessment. If in doubt, or if advanced medical knowledge is required, no temporal relationship is marked.

⚠️ If in doubt regarding the type of temporal relation to mark, follow these criteria:

- **After / Before**: when the context includes linguistic markers of precedence (*antes de, previo, anterior...*) or posteriority (*después de, tras, posterior...*)
- **Has_Duration_or_Interval**: when we can interpretate a sustained or continuous event along the time expression: e.g., *seguidos durante 12 meses, tratamiento de 2 años*
- **Overlap**: in the rest of cases, e.g., if an event and a temporal expression occur with overlapping in the time line (not implying a continuous process): *insulina posprandial* (*insulina* [PROC] → Overlap → *posprandial* [Time])

⚠️ To unify the criteria, the sense of the Overlap relation is from left to right:

(353.) **enfermedad CoVid-19** durante el **embarazo**
Annotate: *enfermedad CoVid-19* (DISO) → OVERLAP → *embarazo* (PHYS)
Do not annotate: *embarazo* (PHYS) → OVERLAP → *enfermedad CoVid-19* (DISO)

3.4. Other relations

Combined_with

A relation between two entities referring to a therapeutic agent, procedure or device (CHEM, PROC or DEVI) when they are combined to treat a disorder. Typically, the context has expressions such as *combinado con, asociado a, más,* +, etc.

> (354.) Estudio que evalúa la combinación de **sirolimus** y **ciclofosfamida**
> (2013-005155-32)
> *sirolimus* (CHEM) → COMBINED_WITH → *ciclofosfamida* (CHEM)

Only annotate the relation between an entity and another which is syntactically loser; do not annotate all the posible relations between the combined entities:

> (355.) uso combinado de **hidroxicloroquina**, **azitromicina** y **tocilizumab**
> (set90/2020-001442-19)
> *hidroxicloroquina* (CHEM) → Combined_with → *azitromicina* (CHEM)
> *azitromicina* (CHEM) → Combined_with → *tocilizumab* (CHEM)
> **Do not annotate:** *hidroxicloroquina* (CHEM) → COMBINED_WITH → *tocilizumab* (CHEM).

> (356.) **dieta** y **ejercicio** en combinación con **metformina**   (2011-002605-29)
> *dieta* (PROC) → Combined_with → *ejercicio* (PROC)
> *ejercicio* (PROC) → Combined_with → *metformina* (CHEM)
> **Do not annotate:** *dieta* (PROC) → COMBINED_WITH → *metformina* (CHEM).

⚠️ To unify the criteria, the sense of the Combined_with relation is from left to rigth:

> (357.) esquema genérico **lamivudina**/**tenofovir** y **efavirenz**
> Annotate: *lamivudina* (CHEM) → COMBINED_WITH → *tenofovir* (CHEM)
> *tenofovir* (CHEM) → COMBINED_WITH → *tenofovir* (CHEM)
> Do not annotate: *tenofovir* (CHEM)→ COMBINED_WITH → *lamivudina* (CHEM)
> nor   *efavirenz* (CHEM)→ COMBINED_WITH → *tenofovir* (CHEM)

Experiences

Annotate this relation between entities expressed by trial participants (LIVB) and the pathology they suffer from (DISO), the physiological condition (PHYS: *embarazo, lactancia, menstruación...*) or the clinical findings experienced (Observation_or_Finding), expressed with both noun and adjective.

(358.) Mujeres infértiles que estén ovulando (2016-004114-99)

(359.) pacientes con procesos médicos (2014-000305-13)

Experiences is also annotated between LIVB (typically, the trial participants) and CHEM or PROC (which corresponds to the intervention):

(360.) Los **sujetos** deben haber recibido 4 ciclos de **quimioterapia**
(2015-002441-61)
Annotate: *sujetos* (LIVB) → EXPERIENCES → *quimioterapia* (PROC)

(361.) **pacientes** con indicación de **colonoscopía**
(S1729-519X2016000500004)
Annotate: *pacientes* (LIVB) → EXPERIENCES → colonoscopía (PROC)

(362.) Efecto de **abatacept** en **pacientes**   (2015-001275-50)
Annotate: *pacientes* (LIVB) → EXPERIENCES → *abatacept* (CHEM)

⚠️ With regard to the PROC entities, we only mark Experiences if the PROC entity expresses a therapeutic or surgical procedure, or a laboratory analysis. We do not annotate Experiences with health care activities such as *diagnóstico*, *seguimiento*, *ensayo clínico, aleatorización, consentimiento informado...*

This relationship is also noted in the case of negated entities:

(363.) Los **pacientes** con una causa **infertilidad excepción de** los **daños de trompas**
pacientes (LIVB) → EXPERIENCES → *infertilidad* (DISO)
excepción de (Neg_cue) → NEGATION → *daños de trompas* (DISO [Negated])
trompas (ANAT) → LOCATION_OF → *daños de trompas* (DISO [Negated])
pacientes (LIVB) → EXPERIENCES → *daños de trompas* (DISO [Negated]).

⚠️ When several entities can be annotated, only the most specific one is marked:

(364.) **pacientes operados** de **tiroidectomía total con vaciamiento cervical**
Annotate: *pacientes* (LIVB) → Experiences → *tiroidectomía ... cervical* (PROC)
Do not annotate: *pacientes* (LIVB) → Experiences → *intervenidos* (PROC)

⚠️If several LIVB entities occur in a sentence, only annotate the relation with the LIVB entity that is more specific or more related with meaning, so that information is not lost. In the following example, only annotate Experiences with regard to *pacientes* (and not with regard to *varones,* which is a subgroup of patients):

(365.) **ingresaron** 17 **pacientes** (13 **varones**)
Annotate: *pacientes* (LIVB) → EXPERIENCES → *ingresaron* (PROC)
Do not annotate: *varones* (LIVB) → Experiences → *intervenidos* (PROC)

In contrast, in the following example, two relations of type Experiences can be annotated, both with *pacientes* and *grupo de intervención,* given that one entity class is not a subgroup of the other one:

(366.) **grupo de intervención** con 76 **pacientes** que iniciarán **HD**
Annotate: *grupo de intervención* (LIVB) → EXPERIENCES → *HD* (PROC)
Annotate: *pacientes* (LIVB) → EXPERIENCES → *HD* (PROC)

Likewise, in the following example, two relations of type Experiences can be annotated with each entity referring to *pacientes*:

(367.) sesenta **pacientes₁**, con edades entre 20-65 años, 50-100 kg, estado físico ASA I-II-III, programados para cirugía de mano, antebrazo y brazo con bloqueo supraclavicular, fueron aleatorizados en grupos de 30 **pacientes₂**
Annotate: *pacientes₁* (LIVB) → EXPERIENCES → *estado físico ASA*
Annotate: *pacientes₂* (LIVB) → EXPERIENCES → *estado físico ASA,* etc.

⚠️ In the case of CHEM and PROC entities in the same sentence, the criterion is to annotate the relation with the syntactic head of the noun phrase. This usually has the semantic group PROC in expressions such as *administración de misoprostol, infusión de ocitocina, enemas de solución fisiológica,* etc. In addition, the USED_FOR or COMBINED_WITH relationship will be annotated with respect to CHEM:

(368.) **tratamiento** con **benznidazol (BZ)** en **pacientes**
Annotate: *pacientes* (LIVB) → EXPERIENCES → *tratamiento* (PROC)
Annotate: *benznidazol* y *BZ* (CHEM) → USED_FOR → *tratamiento* (PROC)
Do not annotate: *pacientes* (LIVB) → Experiences → *benznidazol* y *BZ (CHEM)*

(369.) Efectos tisulares de la **glutamina** en **pacientes** con **cáncer de recto tratados con quimiorradioterapia**
Annotate: *pacientes* (LIVB) → EXPERIENCES → *tratados con quimiorradioterapia* (PROC)
glutamina (CHEM) → COMBINED_WITH → *tratados con quimiorradioterapia* (PROC)
Do not annotate: *pacientes* (LIVB) → Experiences → *glutamina (CHEM)*

(370.) **pacientes** que recibieron **medicación preanestésica** con **midazolam** por vía oral asociado con jarabe de **paracetamol**
Annotate: *pacientes* (LIVB) → EXPERIENCES → *medicación preanestésica* (PROC)
Annotate: *midazolam* (CHEM) → USED_FOR → *medicación preanestésica* (PROC)
Annotate: *paracetamol* (CHEM) → USED_FOR → *medicación preanestésica* (PROC)
Annotate: *midazolam* (CHEM) → COMBINED_WITH → *paracetamol* (PROC)
Do not annotate: *pacientes* (LIVB) → Experiences → *midazolam* ni *paracetamol (CHEM)*

In case that two procedure (PROC) entities are annotated with COMBINED_WITH, we prioritize annotating Experiences with the entity that is the head of the phrase:

(371.) **bloqueo simpático izquierdo** por **toracoscopia** en **pacientes** con **insuficiencia cardíaca**
Annotate: *pacientes* (LIVB) → EXPERIENCES → *bloqueo simpático izquierdo* (PROC)
Annotate: *bloqueo simpático izquierdo* (PROC) → COMBINED_WITH → *toracoscopia*
Annotate: *pacientes* (LIVB) → EXPERIENCES → *insuficiencia cardiaca* (DISO)
Do not annotate: *pacientes* (LIVB) → EXPERIENCES → *toracoscopia* (PROC)

Has_quantifier_or_qualifier

A relationship between a pathology (DISO), a procedure (PROC) or diagnostic scale (CONC) and an entity Quantifier_or_Qualifier (a quantitative value, expressed by a quantifier; or a qualitative value, expressed by adjectives or modifiers):

(372.) **Gonartrosis grado IV**   (2013-001303-36)
Gonartrosis (DISO) → HAS_QUANTIFIER_OR_QUALIFIER → *grado IV*

(373.) haber recibido **4 ciclos** de **quimioterapia**   (2015-002441-61)
quimioterapia (PROC) → Has_quantifier_or_qualifier → *4 ciclos* (Quantifier_or_Qualifier)

Has_result_or_value

This relationship is marked between an entity that refers to an analytical procedure or laboratory measurement (PROC) and an entity expressing the results (e.g., *positive, negative*) or the measured values (Result_or_Value):

(374.) **Prueba de embarazo positiva**   (2020-001606-33)
Prueba de embarazo → Has_result_or_value → *positiva* (Result_or_Value)

The relation can also be marked between a measurement scale (grouped in the CONC class) and the values obtained (Result_or_Value):

(375.) **ECOG 0 o 1**   (2020-001321-31)
ECOG (CONC) → Has_result_or_value → *0 o 1*

Location_of

A relation between an entity expressing anatomical localisation (ANAT) and a pathological condition (DISO) or a procedure (PROC):

(376.) **Cáncer** que afecta a los **ganglios linfáticos**   (2012-004053-88)
ganglios linfáticos (ANAT) → location_of → *cáncer* (DISO)

(377.) **Tratamiento quirúrgico** previo sobre la **cadera**   (2014-004120-21)
cadera (ANAT) → location_of → *tratamiento quirúrgico* (PROC)

Used_for

A relation between a pharmaceutical or chemical substance (CHEM) or a medical device (DEVI), used for a therapeutic or diagnostic procedure (PROC):

(378.) El uso de **CO2** en la **colonoscopía**   (S1729-519X2016000500004)
CO2 (CHEM) → Used_for → *colonoscopía* (PROC)

(379.) El **UDP** se midió, por medio de un **algómetro analógico**
                                        (S1134-80462010000700004)
algómetro analógico (DEVI) → Used_for → *UDP* (PROC)

The Used_for relation is also marked between a CONC entity expressing a diagnostic assessment scale and a diagnostic procedure (PROC):

(380.) **Estado funcional ECOG** 0 o 1   (2020-001321-31)
ECOG (CONC) → Used_for → *Estado funcional* (PROC)
ECOG (CONC) → Has_qualifier_or_quantifier → *0 o 1* (Quantifier_or_qualifier)

It is not necessary to annotate Used_for between the pharmacological entity (CHEM) and the PROC entity expressing the type of clinical trial (e.g., *Estudio observacional, Ensayo de fase II*, etc.):

(381.) Estudio comparativo del uso de mitomicina C en el pterigium recidivante
Do not annotate Used_for between *mitomicina C* and *Estudio comparativo*

In the case of several combinations of drugs used for a procedure, we only annotate a relation between the PROC entity and the CHEM entity that is syntactically closest:

(382.) esquema genérico **3TC/TDF/EFV** es efectivo y seguro en el **tratamiento**
EFV (CHEM) → Used_for → tratamiento (PROC)
3TC (CHEM) → Combined_with → TDF (CHEM)
TDF (CHEM) → Combined_with → EFV (CHEM)

⚠️ We do not annotate Used_for relations between PROC entities:

(383.) *el injerto conjuntival con limbo resultó más efectivo para el tratamiento*
Do not annotate Used_for between *tratamiento* and *injerto conjuntival con limbo*

Causes

A relation between LIVB (generally, microorganisms or virus), CHEM (drugs or chemical substances), PROC (procedures) or DISO (pathological states) that might cause a pathology (DISO) or physiological state (PHYS):
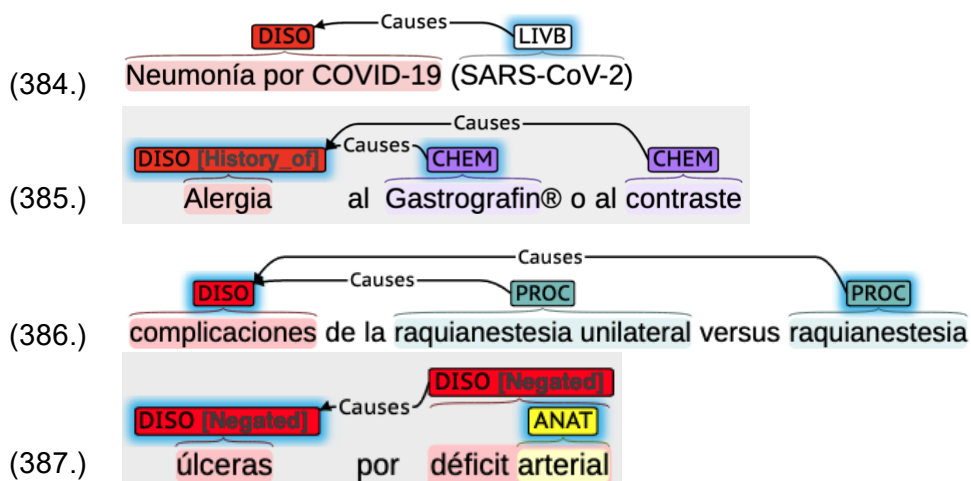
⚠️ In some contexts, both a temporal or a Causes relation can be annotated:

    a) **If the causal relation is unclear** (between a PROC, CHEM or LIVB entity and a pathologic entity or DISO), annotate a **temporal relation**.
    b) **If the causal relation is clear, do not annotate two relations** (Causes and Before), but rather annotate **Causes;** the temporal information can be inferred from the Causes relation (which implies precedence in time).

⚠️ In case of doubt about how many causal relations to annotate, avoid annotating redundant relations. For example, if we annotate event A → CAUSES → B, and event B → CAUSES → C, do not annotate A → CAUSES → C.

(388.) **Estrechamiento de las arterias coronarias** que puede causar **falta de aporte de oxígeno** al corazón produciendo **dolor torácico** e **infarto**
Annotate: *Estrechamiento de las arterias coronarias* (DISO) → CAUSES → *falta de aporte de oxígeno* (DISO)
Annotate: *falta de aporte de oxígeno* (DISO) → CAUSES → *dolor torácico* (DISO)
Annotate: *falta de aporte de oxígeno* (DISO) → CAUSES → *infarto* (DISO)
Do not annotate: *Estrechamiento de las arterias coronarias (DISO) → CAUSES → dolor torácico* (DISO)
Do not annotate: *Estrechamiento de las arterias coronarias (DISO) → CAUSES → infarto* (DISO)

4. CRITERIA FOR AMBIGUITIES

Chemical entities (CHEM) and Laboratory analysis (PROC)

Some entities might be annotated both as chemical substances (CHEM) and laboratory analysis or measurement of chemical substances (PROC). Although some contexts are often unclear, we mark only one semantic group to avoid noise when training the machine-learning algorithm. We tend to label CHEM in ambiguous contexts; the following are some criteria:

- CHEM: annotate only chemical entity when the context does not convey any meaning related to laboratory results, measurements of substances or procedures; in general, contexts referring to a pharmacological substance (the interventional drug, the administered medication...):

  (389.) reciben **carbonato cálcico** a dosis de 2.500 mg/día   (CHEM)
  (0211-699500012698)

  (390.) dosis de **EPO** (UI/semana) mensual   (0211-699500012698)
  Annotate *EPO* as CHEM.

- PROC: we annotate as PROC entities that can be interpreted as *analysis / measurement of* + chemical substance. Frequently, these contexts report results or values of lab procedures, and include terms such as *analítica* or verbs such as *se midió, se analizó, se cuantificó...* The following are other noun phrases often interpretated as PROC: *medición de* + CHEM (e.g., *medición de calcio*) CHEM *total* (e.g., *calcio total*), *valores de* + CHEM, *cifras de* + CHEM...

  (391.) En grupo I aumentaron **PTH** (basal: 38,75 ± 41, final: 99 ± 69, p < 0,01) y **fosfatasa alcalina**   (0211-699500012506)
  Annotate *PTH* and *fosfatasa alcalina* as PROC (Laboratory Procedure).

  (392.) analíticas: **hemograma**, **albúmina** y **creatinina séricas**
  (0211-699502014854)
  Annotate *hemograma*, *albúmina* and *creatinina séricas* as PROC.

Terms such as *inyección* or *vacunación* are annotated as PROC when they are interpreted as the therapeutic procedure. The word *vacuna* is annotated as CHEM when it refers to the medical drug or the pharmacological substance:

(393.) la **inyección** de propofol　　　　(S1025-028X2011000300001)
Annotate *inyección* as PROC and *propofol* as CHEM.

(394.) existía relación causal con la **vacunación**
Annotate *vacunación* as PROC　　　(S1025-028X2005000100002)

(395.) la reactogenicidad de la **vacuna** es escasa
Annotate *vacuna* as CHEM.　　　(S1025-028X2005000100002)

Ambiguity between chemical entities (CHEM) and body substances (ANAT)

We annotate chemical or pharmacological entities (CHEM) on mentions referring to substances that are administered as interventions or therapeutic procedures:

(396.) Aplicación de **sellante de fibrina**　　　(S0009-741115002261)
Annotate *aplicación* (PROC) and *sellante de fibrina* (CHEM) (*fibrina* is not ANAT)

On the contrary, we interpretate as anatomical entity (ANAT) from molecular level the el rest of macromolecules, cell components or body substances.

(397.) efecto genotóxico en **linfocitos humanos** (S0716-10182014000500006)
Annotate *linfocitos humanos* (ANAT) and *efecto genotóxico* (DISO)

We annotate the term *suero* as ANAT when we interpretate it as 'plasma', 'the liquid substance of the blood'; and we annotate it as CHEM when it has the meaning of 'therapeutic solution':

(398.) el nivel de betatrofina en **suero** fue más elevado (...) la betatrofina en **suero** desempeña una importante labor en síndrome metabólico
　　　　　　　　　　　　　　　　　　(S0212-16112016000200019)
Annotate *suero* (ANAT); *betatrofina* (CHEM) and *nivel de betatrofina en suero* (PROC)

(399.) pacientes deshidratados por diarrea aguda, menores de cinco años, recibieron **suero** oral　　　　　　　(S0036-36342002000100003)
Annotate *suero* as CHEM

Entities such as *plasma rico en plaquetas, plasma rico en factores de crecimiento* or *concentrados de hematíes* are interpretated as CHEM owing to the pharmacological processing they require to be produced:

(400.) tratamiento guiado con ecografía mediante **Plasma Rico en Plaquetas** (2018-003709-25) (C2362067 Platelet concentrate; Amino Acid, Peptide, or Protein; Pharmacologic Substance)

Regarding cell receptors, the UMLS ontological scheme classifies them as CHEM, even though these entities could be considered as ANAT:

(401.) **antagonistas del receptor AT1 de angiotensina II**, al bloquear más el **receptor AT1**, debería producir mayores beneficios (S0211-69952012000200010)
Annotate 2 CHEM entities: *antagonistas del receptor AT1 de angiotensina II* (C0521942, Angiotensin II receptor antagonist) and *receptor AT1* (C0529330, Receptor, Angiotensin, Type 1)

(402.) **pantoprazol** por vía intravenosa　　　(S1130-01082008000800004)
Annotate only *pantoprazol* as CHEM

(403.) el consumo de **anestésicos** <u>venosos</u>   (S0034-70942004000100013)
Annotate only *anestésicos* as CHEM

(404.) **citrato de fentanilo** <u>oral</u> **transmucoso** (S1134-80462010000300002)
Annotate as CHEM the full mention *citrato de fentanilo oral transmucoso* (C3853047;
Transmucosal Oral Fentanyl Citrate, Organic Chemical)

Entities referring to diagnostic measurement scales

Only entities found **in the UMLS and** with **PROC semantic type** will be annotated
(because they can be interpreted as diagnostic procedure; e.g., WAIS intelligence
test, CUI C2609456):

(405.) valoración subjetiva del dolor mediante la **escala visual analógica** (**EVA**)
(S2306-41022015000100005)
Annotate *escala visual analógica* and *EVA* (PROC, CUI C0042815)

(406.) evolución funcional por **escala de Rankin modificada**
(S0210-56912010000600003)
Annotate *escala de Rankin modificada* (PROC in the UMLS, CUI: C3888574)

(407.) **evaluación por Escala de Glasgow** para un **seguimiento**
Annotate PROC in *evaluación por Escala de Glasgow* (C0278349 Glasgow coma testing
and evaluation) and PROC in *seguimiento*

Terms classified as scale or classification in the UMLS (not classified as diagnostic
procedure, e.g., *escala* or *puntaje*) are annotated CONC rather than PROC:

(408.) presentaron menor escore de sedación según la <u>escala de Ramsay</u>
(S0066-782X2011000300008)
Do not annotate PROC in *escala de Ramsay* (concept entity, CONC: C0559295 Ramsay
sedation scale; Intellectual Product)

(409.) Impresión Clínica Global   (S0211-57352011000100002)
Do not annotate PROC (C3639708 Clinical global impression scale) but CONC

(410.) Inventario de Discapacidad de Sheehan   (S0211-57352011000100002)
Do not annotate PROC (C4720840 Sheehan disability scale) but CONC

Entities referring to Associations, Medical specialties or Hospital units

These types of entities (and the nested entities inside them) are not annotated:

(411.) en la <u>Unidad de Digestivo</u> del Complejo Hospitalario de Ciudad Real
Do not annotate *Unidad de Digestivo* nor *Digestivo*
(S0212-71992005001000008)

(412.) <u>American Diabetes Association</u> (S0211-69952012000200010)
Do not annotate *American Diabetes Association* nor *Diabetes.*

5. Appendix

5.1. List of modifiers

The following list might help to decide the types of modifiers to be annotated together with a core medical term; in general, we annotate an entity with modifiers:

- When the full term can be mapped to an ICD-10 category (in the case of DISO entities) or to an UMLS CUI (in the case of PROC entities).
- When it is necessary to understand the context or the content of the trial.

To avoid fragmented annotations, we do not annotate modifiers in attributive or copulative constructions (with the verb *ser*):

> (413.) El grado final de **envenenamiento** fue leve en 13 (24,5%), moderado en 30 (56,6%) y grave   (S0121-07932007000300002)
> Annotate only *envenenamiento* (DISO); annotate *grave*, *leve* or *moderado* as Quantifier_or_Qualifier.

Annotate in full entity	Do not annotate
Specific temporal modifiers or adjectives in multi-word terms, when they can be mapped to a CUI or ICD-10 category: temporal entities may also be annotated as nested inside the full entity: *crónico:* e.g., *fisura anal crónica*: Annotate *fisura anal crónica* as DISO (ICD-10: K60.1, C0349071 Chronic anal fissure) and also *crónica* (Duration) in it. *postoperatorio*: e.g., *dolor postoperatorio* Annotated DISO (C0030201 postoperative pain; ICD-10: G89.18) and also *postoperatorio* (Time) inside. *postpunción*: e.g., *cefalea post-punción lumbar* (C2875393 Headache due to lumbar puncture; ICD-10: G97.1). *postprandial*: e.g., *glicemia postprandial* Annotate DISO (C4700958 postprandial glucose in serum or plasma) and also *postprandial* (Time) inside.	Vague temporal qualifiers: *anterior, previo, siguiente, posterior...* Exceptionally, we annotate these adjectives in the full entity (e.g., DISO, PROC…) if they are necessary to understand the context, or if the adjective distinguishes different concepts (CUI exist): E.g., *inflamación persistente* (C1265824 Persistent inflammation) E.g., *eyaculación precoz* (C0033038, ICD-10: F52.4) E.g., *seguimiento a largo plazo* (C1517942 Long-term Follow-up) Adjetives expressing frequency are annotated with label Frequency: *diario, frecuente, periódico, semanal...*
Adjectives describing spatial position or localization, preferably only when an ICD-10 category or an UMLS CUI exists, or if they are needed to understand the text content: *alta*: e.g., en *endoscopia alta* (C0017195 upper endoscopy)	Past participles without specific details or descriptive information about the term are not annotated: *administrado* *asociado* *conocido* *empleado* *especializado*

central: e.g., *sistema nervioso central* (C1269563 central nervous system) *local*: e.g., *anestésico local* (C0002934 Local Anesthetics) *superior* (e.g., *miembro superior*, C1140618 Upper Extremity) *unilateral*: e.g., *dolor unilateral* Annotate DISO (C0747150 pain unilateral; Sign or Symptom) and also *unilateral* (Quantifier_or_Qualifier) inside.	*establecido* *generado* *realizado* *reconocido* *relacionado* *sometido*
Relational adjetives, descriptors of specific properties or features, or modifiers expressing a subtype or class of pathology, if an ICD-10 category or UMLS CUI exists: *clínico*: e.g., *ensayo clínico* (C0008976 Clinical Trial) *medible*: e.g., *enfermedad medible* (C1513041 Measurable Disease), *lesión no medible* (C1334988 Non-Measurable Lesion)	Vague qualifiers and adjetives not describing specific details: e.g., *diverso, variado...*
Adjetives describing the type of procedure (the *Comparator*), or when they are needed to understand the context: *convencional*: e.g., *cirugía convencional* (C1521745 Conventional Surgery) *estándar:* e.g., *tratamiento estándar* (C4684780 Standard Treatment)	
refractario: we only annotate it with the core term when this adjective is not modified by other phrases (to avoid long annotations) or when the full term has an UMLS CUI: E.g., *dolor refractario* (C0030200 Refractory Pain)	*refractario a* + phrase: we annotate only *refractario* when other phrases modify it (to avoid long annotations): E.g., *hipertensión intracraneal (HIC) refractaria a las medidas de primer nivel* Annotate *hipertensión intracraneal* (DISO) and *refractaria* (Quantifier_or_Qualifier)

secundario: we annotate it when it has the sense of 'caused by other factor' (e.g., a pathology) in constructions without phrases, and when the full term can be mapped to an ICD-10 category: E.g., ***hiperparatiroidismo secundario en diálisis*** Annotate *hiperatiroidismo secundario* (C0494305 Secondary hyperparathyroidism, ICD-10: E21.1)	*secundario a*: do not annotate it when other phrases or complements modify this adjective, except in contexts where the full term helps to understand the text content: E.g., ***rectorragia*** *crónica* <u>*secundaria a la*</u> ***proctopatía por radiación*** Annotate two different terms: *rectorragia* (DISO) and *proctopatía por radiación* (DISO)
Adjectives expressing recurrence or relapse: we only annotate them when the full term has an ICD-10 category or an UMLS CUI; or when the annotation is needed to understand the text contents: *recidivante*: e.g., *pterigium recidivante* (C0155158 Recurrent pterygium, ICD-10: 11.06) *recurrente*: e.g., *estomatitis aftosa recurrente* (C2937365 Recurrent aphthous stomatitis; ICD-10: K12.0) *remitente*: e.g., *esclerosis múltiple remitente-recurrente* (C0751967 Multiple Sclerosis, Relapsing-Remitting)	
Descriptors of the manner in which a procedure is done or applied, or adjectives expressing the extension of a pathological process: we only annotate terms having a CUI, or when the full term is needed to understand the context: *activo*: annotate only in terms with an UMLS CUI: E.g., *Inmunización activa* (C0042196 Active immunization) E.g., *EC fistulosa* <u>*activa*</u> Annotate *EC fistulosa* as DISO (C3509259 crohn's disease with fistula) and *active* (Quantifier_or_Qualifier) *abierto*: e.g., *cirugía abierta* (C4283938 Open Surgical Procedure) *basal*: e.g., *tasa metabólica basal* (C0542499 Measurement of basal metabolic rate)	Comparative adjetives: *inferior* (*que*) *mayor* (*que*) *mejor* (*que*) *peor* (*que*) *superior* (que) In contexts where the adjective needs to be annotated with the main term, we annotate the full entity, because there is an ICD-10 category or UMLS CUI: *Cirugía mayor* (C0679637 major surgery) Annotate PROC. *Trastorno depresivo mayor* (C1269683 Major Depressive Disorder, ICD-10: F32.9) Annotate DISO.

continuo: e.g., *perfusión continua* (C0444889 Continuous infusion) *descompensado*: e.g., *EPOC descompensado* (C2887446 Decompensated COPD; ICD-10: J44.1) *primario*: e.g., *hipertiroidismo primario,* C0221002 Hyperparathyroidism, Primary; ICD-10: E21.0) *sistémico*: e.g., *esclerosis sistémica,* (C0477589 systemic sclerosis, ICD-10: M34.8) *total*: e.g., *anestesia intravenosa total* (C0473965, Total intravenous anesthesia)	In other contexts, if needed, the adjective is annotated separately as Quantifier_or_Qualifier: *interacciones farmacológicas mayores* Annotate *interacciones farmacológicas* (CHEM) and *mayores* (Quantifier_or_Qualifier).
Descriptors of severity or pathological stage: we only annotate them when a full term has an ICD-10 category; otherwise, we annotate them as Quantifier_or_Qualifier: *grave:* e.g., *miastenia grave* (C0026896 Myasthenia Gravis; ICD-10: G70.0) In general, we annotate vague or imprecise adjectives such as *serio, importante,* etc., as Quantifier_or_Qualifier: E.g., *eventos adversos serios* Annotate *eventos adversos* (DISO) and *serios* (Quantifier_or_Qualifier) E.g., *afectación neurológica importante* Annotate *afectación neurológica* as DISO and *importante* as Quantifier_or_Qualifier	Adjetives or past participles expressing possibility, probability, assertion, certainty or uncertainty, or disease activity: annotate as Spec_cue. *posible* *probable* *sospechoso* *sugestiva* *susceptible* E.g., *sepsis nosocomial confirmada o altamente probable* Annotate *sepsis nosocomial* (DISO) and *probable* (Spec_cue).
Adjetives of type of measurement or average values measured: we annotate them if a concept (CUI) exists in the UMLS: *medio*: E.g., *presión arterial media* (C0428886 Mean blood pressure) E.g., *volumen corpuscular medio* (C0863148 Mean corpuscular volume analyses, PROC) *promedio*: E.g., *presión arterial promedio* (C0428886 Mean blood pressure, PROC)	Adjetives specifying the average values of measurements: we do not annotate them if an UMLS CUI does not exist: *medio* E.g., *pérdida de peso media* (annotate only *pérdida de peso*) *promedio* E.g., *IMC promedio* (annotate only *IMC*)

5.2. List of examples annotated as Observation_or_Finding

Example (sorted alphabetically)	Possible UMLS concept
actividad física	C4718384; Participates in physical activity; Finding
activos sexualmente, sexualmente activas	C0241028; Sexually active; Finding
adherencia inadecuada, adherencia apropiada, cumplidor, cumplimiento (con el tratamiento)	C0032646; Poor compliance; Finding C2364172; Adherence To Medication Regime; Finding
alteración del antígeno específico prostático	C0580555; Prostate specific antigen abnormal; Finding
anomalías de laboratorio, anomalía analítica, resultados anómalos de laboratorio	C3648370; laboratory tests nonspecific abnormal findings; Finding
anormal	C0205161; Abnormal; Finding
avidez por la fluorodesoxiglucosa (FDG)	C5239279; FDG-Avid Tumor; Finding
bajo nivel de conciencia	C0549249; Depressed Level of Consciousness; Finding
buen estado general, buena salud	C1277245; General health good; Finding
cambios en la microflora	C4047662; Altered microbiota; Finding
cambio metabólico	C4718639; Metabolic changes; Finding
capacitado, autorizado, habilitado	C1555472; Authorized; Finding
capaz de realizar las actividades básicas de la vida diaria	C0562507; Ability to perform activities of everyday life; Finding
cese del hábito de fumar	C0425310; Stopped smoking; Finding
Cognitivamente normal	C1846595; Normal cognition and intellectual function; Finding
condición (médica)	C4745084; Medical Condition; Finding
contaminación	C1879778; Bacterial Contamination of Device; Finding C2609326; Product contamination bacterial; Finding
cumplir con los requisitos de ser donante	C2585138; Eligible to be a donor; Finding
cumplir los requisitos del protocolo, tratamiento, etc.	C0582783; Ability to comply with treatment; Finding

desaturación de oxígeno	C4061338 Oxygen desaturation; Finding
descenso del segmento ST	C0520887; ST segment depression (finding); Finding
despierto	C0234422; Awake; Finding
deterioro de estado funcional	C4304688; Decline in functional status; Finding
diferenciación neuroendocrina	C1709218; Neuroendocrine Differentiation; Finding
elevación del segmento ST *elevación de ST*	C0520886; ST segment elevation (finding); Finding
en ayunas	C0015663; Fasting; Finding
esfuerzo respiratorio	C0425466 Respiratory effort; Finding
estabilidad hemodinámica	C4718759; Hemodynamic stability; Finding:
estables clínicamente *estabilidad clínica* *estabilizado*	C1282982; Patient's condition stable; Finding
estado de reposo	C0679218; Resting state; Finding
estado funcional no independiente	C4065055; functional limitations in capacity for independent living; Finding
estado nutricional	C0392209; Nutritional status; Finding
eutrófico	C0578019; Normal growth; Finding
factores de riesgo (médicos)	C0035648; risk factors; Finding C3468861; limitations and risks general; Finding
fallo virológico, *fracaso virológico*	C1699206; Virologic failure; Finding
focos	C4321394; Foci; Finding
frecuencia de heces	C0426642; Frequency of bowel action; Finding
hormona paratiroidea elevada	C0748089; Parathyroid hormone concentration above normal range; Finding
H. pylori+	C4763617; Helicobacter pylori Status; Finding
incumplimiento del tratamiento	C0184543; Noncompliance with therapeutic regimen; Finding
incremento del apetito	C0232461; Increased appetite; Finding
independiente (estado funcional de capacidad de autonomía)	C4273559; Functionally independent; Finding

institucionalizado	C0562359; Institutionalized; Finding
insuficiente control del idioma *barreras del idioma* *barrera idiomática* *no hablar el idioma* *desconocimiento del idioma*	C0237167; Language Barriers; Finding
en lista de espera	C0422768; Patient on waiting list; Finding
neutrofilia	C0151683; Neutrophilia (finding); Finding
normoalbuminuria *normoglicémicos* *normonutridos* *normotenso,* *normotermia* *peso normal*	C2712122; Normal blood pressure; Finding C0860865; Albumin normal; Finding C4717297; Provides appropriate nutrition; Finding C2712185; Normal weight; Finding C0231262; Temperature normal; Finding
preparación colónica deficiente *mala limpieza colónica* *mala preparación* *preparación deficiente* *colonoscopia mal preparada* *mala calidad de la limpieza* *colónica*	C1446338 Inadequate bowel preparation; Finding
presentación cefálica	C2979973; Cephalic Presentation; Finding
primípara	C0860096; Primigravida; Finding
prolongación del QTc *QTc prolongado* *QTc alargado*	C1969409; Prolonged QTc; Finding
pronóstico disminuido *pronóstico vital infausto* *situación terminal* *muerte inminente*	C0278252; Prognosis bad; Finding C0877369; Terminal state; Finding C2051439; patient condition is terminal; Finding
QRS prolongado	C0235475; Wide QRS complex; Finding
rasgos sarcomatoides	C1519183; Sarcomatoid Features; Finding
reclusión en centros *penitenciarios*	C0392751; In prison (finding); Finding
relaciones íntimas	C3244223; Intimate Interaction; Finding
remisión *remisión completa* *remisión clínica*	C0544452; Disease remission; Finding C0677874; In complete remission; Finding C4050094; Clinical Complete Response; Finding

remisión sintomática	C0871467; Symptom Remission; Finding
reordenación del gen de la cinasa del linfoma anaplásico	C4068350; ALK Gene Rearrangement; Finding
resistencia (al tratamiento)	C3542449; Drug Resistance Status; Finding
respuesta pobre *respuesta parcial* *respuesta buena* *pérdida de respuesta* *respuesta inadecuada* *respuesta al tratamiento* *respuesta clínica* *hayan respondido*	C1320680; Poor response to treatment; Finding C5400193; Loss of therapeutic response; Finding C1521726; partial response; Finding C4055223; Clinical Response; Finding
reversibilidad (de patología)	C0679253; Disease reversibility; Finding
respuesta virológica sostenida *respuesta virológica mantenida* *supresión viral,* *supresión virológica,* *virológicamente suprimido* *RVS*	C3888663; Sustained viral response; Finding
riesgo anestésico	C4544425; At risk of complication of anesthesia; Finding
riesgo obstétrico	C3533167; Obstetrical risk; Finding
satisfacción (con el tratamiento)	C3476649; Treatment satisfaction; Finding
sedentario	C1532253; Sedentary lifestyle; Finding
septo	C1708055; Fibrous Septa Present; Finding
seroconversión	C4042908; Seroconversion; Finding
refractario, refractariedad	C0438286; Absent response to treatment; Finding
tolerancia alimentaria	C1821263; Feeding tolerance; Finding
trabajan habitualmente en horario nocturno	C0555008; Night shift worker; Finding
uso indebido de medicamentos	C0581388 Prescription Drug Misuse; Finding
vía aérea difícil	C4273494; Unexpected difficult airway; Finding

6. REFERENCES

Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004 Jan 1; 32(Database issue): D267–D270. https://doi.org/10.1093/nar/gkh061

Campillos-Llanos, L, L. Deléger, C. Grouin, T. Hamon, A-L. Ligozat & A. Néveol (2018) A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMSI annOtated Text corpus (MERLOT). *Language Resources and Evaluation*, 52(2), pp 571–601. https://doi.org/10.1007/s10579-017-9382-y

Campillos-Llanos, L., Valverde-Mateos, A., Capllonch-Carrión, A., & Moreno-Sandoval, A. (2021). A clinical trials corpus annotated with UMLS entities to enhance the access to evidence-based medicine. *BMC medical informatics and decision making*, 21(1), 1-19. https://doi.org/10.1186/s12911-021-01395-z

Campillos-Llanos, L., Valverde-Mateos, A., & Capllonch-Carrión, A. (2025) Hybrid natural language processing tool for semantic annotation of medical texts in Spanish, *BMC Bioinformatics* (accepted). https://doi.org/10.1186/s12859-024-05949-6

Deléger, L., Campillos, L., Ligozat, A.-L., & Névéol, A. (2017) Design of an extensive information representation scheme for clinical narratives. *Journal of biomedical semantics,* 8 (1), 37. https://doi.org/10.1186/s13326-017-0135-z

Farré, E., González, G. Krallinger, M., Mas, T., & Miranda, A. (2019) *Guía de anotación de morfologías neoplásicas – Junio 2020, vs. 1.3 [CANTEMIST Challenge Annotation Guidelines].* Ministerio de Economía y Empresa, Plan de Impulso de Tecnologías del Lenguaje, área Sanidad. https://zenodo.org/record/3878179

Grouin, C., N. Grabar, R. Cardon & T. Hamon (2020) DEFT 2020: guide d'annotation. https://deft.limsi.fr/2020/guide-deft.html (Accessed: 26/02/2020)

Harkema, H., Dowling, J. N., Thornblade, T., & Chapman, W. W. (2009) ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics*, 42(5), 839-51. https://doi.org/10.1016/j.jbi.2009.05.002

Herrero Zazo, M., I. Segura-Bedmar, P. Martínez Fernández (2013) *Annotation Guidelines for the DDI Corpus*. Technical Report. Universidad Carlos III de Madrid. https://www.cs.york.ac.uk/semeval-2013/task9/data/uploads/annotation_guidelines_ddi_corpus.pdf (Accessed: 17/02/2020)

Herrero Zorita, C. (2017) *Modality in spoken Spanish and Japanese: a corpus-based study and automatic annotation*. PhD dissertation. Universidad Autónoma de Madrid.

Mas, T., A. Villarrubia, N. Vittini (2020) *Guía de anotación de Codificación e Indexado en CIE-10 [CODIESP Challenge Annotation Guidelines]*. https://temu.bsc.es/codiesp/index.php/2019/09/19/annotation-guidelines/ (Accessed: 31/05/2020)

Néveol, A., L. Déleger, A.-L. Ligozat, C. Grouin, & T. Hamon (2016) https://cabernet.limsi.fr/annotation_guide_for_the_merlot_french_clinical_corpus-Sept2016.pdf (Accessed: 5/12/2019)

Nye, B. J. J. Li, R. Patel, Y. Yang, I. J. Marshall, A. Nenkova, & B. C. Wallace (2018) A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature. *Proc. 56th Annual Meeting of the Association for Computational Linguistics,* Melbourne, Australia, July, 197–207

Ogren, P., G. Savova, & C. Chute (2008). Constructing evaluation corpora for automated clinical named entity recognition. *Proc. of the 6th Int. Conf. on Language Resources and Evaluation (LREC 2008)*. Marrakech, Mayo 2008, pp. 3143–50.

Oronoz, M., K. Gojenola, A. Pérez, A. D. de Ilarraza, & A. Casillas (2015). On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of biomedical informatics* 56, 318-32. https://doi.org/10.1016/j.jbi.2015.06.016

Rabal, O, A. Intxaurrondo & M. Krallinger (2018) *Guía de anotación y normalización de compuestos químicos. [PharmaCoNER Challenge Annotation Guidelines]*. Disponible en: https://bit.ly/395PLTE Ministerio de Economía y Empresa, Plan de Impulso de Tecnologías del Lenguaje, área Sanidad.

Real Academia Nacional de Medicina de España – RANME (2011) *Diccionario de términos médicos (DTM).* Madrid: Editorial Panamericana.

Roberts A, R. Gaizauskas, M. Hepple, G. Demetriou, Y. Guo, I. Roberts, A. Setzer (2009) Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42: 950–66. https://doi.org/10.1016/j.jbi.2008.12.013

Savova, G., W. Styler, D. Albright, M. Palmer, D. Harris, G. Zaramba, P. Haug, C. Clark, S. Wu, D. Ihrke (2012) *SHARP template annotations: Guidelines. Technical report*, Mayo Clinic. http://informatics.mayo.edu/sharp/images/0/08/SHARP_annotation_guidelines_March6_2012.pdf (Accessed: 17/02/2020)

Temporal Histories of Your Medical Event (THYME) Project. https://clear.colorado.edu/TemporalWiki/index.php/Main_Page (Accessed: 17/02/2020)

Uzuner, Ö., I. Solti, & E. Cadag (2010) Extracting medication information from clinical text. *Journal of the American Medical Informatics Association*,17(5):514–8.

Uzuner Ö., B. South, S. Shen, S. L. DuVall (2011) 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc.* 2011;18(5):552–6.   https://doi.org/10.1136/amiajnl-2011-000203   The *Concept Annotation Guidelines* are available at: https://www.i2b2.org/NLP/Relations/assets/Concept   Annotation   Guideline.pdf (Accessed: 17/02/2020)