# NLPMEDTERM CORPUS

# ANNOTATION GUIDELINE

English version

Leonardo Campillos Llanos
Adrián Capllonch Carrión
Ana Valverde Mateos

Latest update: June 2020

# 1. ANNOTATION CRITERIA

The task consists in annotating 4 types of entities on Spanish texts from journal abstracts about clinical trial studies, published in the SciELO repository,[1] and clinical trial annoucements from EudraCT.[2] The entities belong to semantic groups from the Unified Medical Language System® (UMLS) (Bodenreider 2004) referring to pathologies (DISO), anatomic entities (ANAT), biochemical or pharmacological substances (CHEM) and lab tests, diagnostic or therapeutic procedures (PROC).

To check the type of entity or to understand each concept mention, the following sources are 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



Figura 1. Sample of the annotation

The goals of the annotation are **detecting** entities, as mencioned in the scientific literature or clinical texts, and also **normalize** the entities to the corresponding medical concepts. Consequently, some 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).

---

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

**SCOPE OF THE ANNOTATION**

In general, annotate full phrases if possible:

(1.) dosis convencionales en 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. E.g. 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% (S1729-519X2016000600017)
Annotate only *eventos adversos* (DISO)

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: administración de **vacunas vivas**   (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. *distinto, futuro, 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 quirúrgica 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 anestésico fue más corto después de aplicar la **técnica alveolar media**                    (S0213-12852017000100005)
Do not annotate *anestésico* because 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**
                                    (S0009-741115002339)
Annotate *procedimiento quirúrgico de citorreducción* (one entity) as PROC (C3850079 Cytoreductive Surgery)

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

(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. <u>anestesia con</u>, <u>analgesia con</u>, <u>cirugía en</u>, <u>tratamiento de</u>, <u>prevención de</u>, <u>control de,</u> 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 2 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)
(S0212-16112010000600008)

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):

(40.) pre y post-**implante del marcapasos**          (S1135-57272015000200004)
Annotate *implante del marcapasos* (PROC)

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

(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 entity *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
(S0716-10182019000100032)
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)

Inner entities incorrectly pre-annotated are to be deleted: e.g. *índice* (ANAT) in *Índice de masa corporal*.

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 (e.g. 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). It is not necessary to 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/](http://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 2 mentions: *umbral de dolor a la presión* and *UDP*

## NEGATED ENTITIES

Entities occurring in negated contexts are also annotated, but the negation cue is left for a future annotation task:

(67.) 301 no **vacunados**     (S1131-57682003000400010)
Annotate only *vacunados* (PROC)

(68.) sin **complicaciones clínicas**        DISO   (S1729-519X2016000500004)
Annotate only *complicaciones clínicas* (DISO)

(69.) destacando la ausencia de **edemas**        (0211-699504017174)
Annotate only *edemas* (DISO)

## ENTITIES EXPRESSING POSSIBILITY, PROBABILITY OR ASSERTION

Even though these types of entities are very relevant to interpretate clinical results, we do not annotate them in this project stage. Do not mark entities expressing presence of a pathology, possibility, probability, condition with regard to an outward circumstance, risk or hypothetical speculation:

(70.) puede mejorar los resultados y minimizar los posibles **efectos indeseables**
Annotate only *efectos indeseables* as DISO.       (S0212-16112010000600008)

(71.) **sepsis nosocomial** confirmada
Annotate only *sepsis nosocomial* as DISO        (S0325-00752014000400004)

(72.) recibían **fármacos** susceptibles de ser **prescritos** como **genéricos**
Annotate *fármacos* and *genéricos* as CHEM, and *prescritos* as PROC
(S0213-91112002000600008)

(73.) la duración prolongada del QRS está asociada a un aumento en el riesgo
de mortalidad (S0066-782X2010001600011)
Do not annotate anything.

## ENTITIES EXPRESSING ASPECT

In the current version, we do not annotate entities expressing recurrence or relapse;
these terms tend to be classified as Findings in the UMLS Metathesaurus:

(74.) Hubo 2 casos (3,8%) de recurrencia de **coagulopatía**
    Annotate only *coagulopatía* as DISO     (S0121-07932007000300002)

(75.) La duración del **tratamiento** disminuyó el número de recaídas
                                        (S1665-11462016000500309)
    Do not annotate *recaídas,* only *tratamiento* (PROC)

(76.) Analizamos las recidivas de la **pancreatitis** (S1137-66272003000300004)
    Annotate only *pancreatitis* as DISO.

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

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

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

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

## FOREIGN WORDS

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

(80.) 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:

(81.) 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:

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

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

## 2. ANNOTATED 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)
- DISO (pathologies)
- PROC (diagnostic and therapeutic )

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). In a future version, these data can be annotated with Evidence-Based Medicine (EBM) PICO elements: *Patients, Interventions, Comparators* and *Outcomes.*

**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:

> (84.) La permanencia fue mayor en la **pierna**     (S0325-00752018000100041)

> (85.) en la insuficiencia **renal** crónica          (0211-699504017174)

> (86.) 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:

> (87.) puntuación para cada **región del colon**     (S1022-51292016000400002)
> Annotate *región del colon* (C0447522 Region of colon)

> (88.) anestesia troncular de **1o dedo**     (S0210-1238201324)
> Annotate *1o dedo*

However, cardinal numbers are not annotated: e.g. annotate *ojos* instead of *dos ojos.*


**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:

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

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

   (91.) 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:

   (92.) **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:

   (93.) **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)


**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:

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

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

(96.) 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 if an ICD-10 category exists:

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

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

As an exception, in order to get more generalizable annotations, we do not annotate intermediate stages or degrees of the chronic kidney disease, which are often expressed in ranges or inexact values:

(99.) **ERC** estadio 3A-4                       (S1134-009615002090)
Annotate only *ERC* (C1561643 Chronic Kidney Disease, ICD-10. N18), even though more specific ICD-10 categories exist (N18.3 and N18.4)

(100.) **Insuficiencia renal** >3B
Annotate only *Insuficiencia renal*

Regarding entities with a severity modifier without an ICD-10 category, only the pathologic entity is annotated (not the modifier):

(101.) 72,3% fueron de **laceración** de primer grado (S0104-11692008000300007)
Annotate only *laceración* (C0043246 Laceration).

(102.) 61% de los pacientes refirieron **cefalea** leve (S1130-01082004000400004)
Annotate only *cefalea* (C0018681 Headache, ICD-10: R51)

(103.) cifras consideradas **dolor** moderado        (S0325-00752008000500004)
Annotate only *dolor* (C0030193 Pain, ICD-10: R52)

(104.) presentaban **obstrucción bronquial** leve o moderada
Annotate only *obstrucción bronquial* (C0221725 Bronchial Obstruction)

Likewise, we do not annotate in the current version classifications of physical or functional state of pathologies, neoplasia stages, or values of the assessment of the oncologic quality of life; the UMLS tends to classify them as Findings:

(105.) pacientes con **falla cardiaca** estado funcional II-IV por más de 6 meses
(S0120-56332018000200106)
Annotate only *falla cardiaca* as DISO*.*

(106.) clase funcional ii a iv de la New York Heart Association (NYHA)
(S0211-699519300529)
Do not annotate *clase funcional ii a iv de la New York Heart Association (NYHA)*

(107.) **toxicidad gastrointestinal** aguda <u>Grado 1-2</u> (S0185-106317300264)
Annotate only *toxicidad gastrointestinal* (C1142499 Gastrointestinal toxicity)

(108.) con edades entre 20-65 años, 50-100 kg, <u>estado físico ASA I-II-III</u>
Do not annotate *estado físico ASA I-II-III*          (S0034-70942014000300177)

(109.) Pacientes con **CVMI**, en estadio clínico <u>T2-T4a, N0, M0</u> (2017-004692-31)
Annotate only CVMI ('Carcinoma de vejiga músculo invasivo')

(110.) ECOG (Eastern Cooperative Oncology Group) 0-1  (2015-004816-39)
Do not annotate anything.

(111.) 24 pacientes con **cirrosis alcohólica** (8 en <u>estadio B de Child-Pugh</u> y 16 en <u>estadio C Child-Pugh</u>)          (S1130-01082008000800007)
Annotate only *cirrosis alcohólica* (C0023891 Liver Cirrhosis, Alcoholic, ICD-10: K70.30)

Besides, the UMLS includes the ***Findings*** category in the DISO semantic group. Despite their clinical relevance, most of these entity types **were excluded in the current version of the corpus**; e.g.:

- 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* (do not annotate *normoalbuminúrico*) (S0211-69952012000200010)
- Too general or vague terms, when they occur isolated: *anomalía, hallazgo, problema...*
- Terms such as *historia de, antecedentes de...*: e.g. *antecedentes de drogadicción* (annotate only *drogadicción*; S1134-80462013000600002)
- 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, esperanza de vida inferior a un año...*
- Findings related to medical devices: e.g. *portador de marcapasos, portadores de prótesis valvulares...*
- 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*

**We only annotate entities of type *Finding* if there is a category in the International Classification of Diseases vs 10 (ICD-10),** but 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**.

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

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

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

(115.) **embarazos múltiples**      (2013-003778-29)
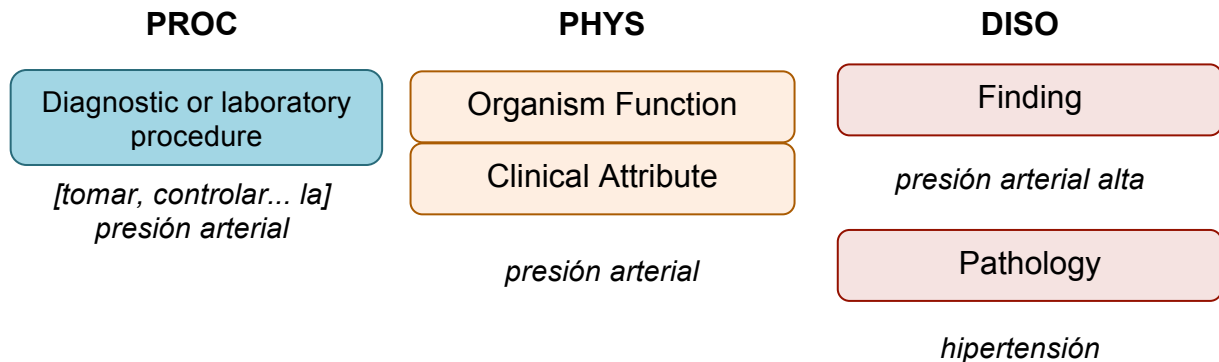(C0032989 Multiple Pregnancy; ICD-10: O30.90)

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)

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

(117.) 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:

(118.) 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*):

(119.) Sífilis **IgG+**               (2020-001565-37)
Annotate DISO en *Sífilis* y PROC en IgG (sin carácter +)

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:

(120.) 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:

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

(122.) 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.

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):

(123.) 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).* We do not annotate the gen or exon involved:

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

(125.) **liposarcoma mixoide** con **mutación** PIK3CA     (2013-005155-32)
Annotate DISO in *liposarcoma mixoide* and in *mutación*

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


**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:

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

(128.) 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:

(129.) 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:

(130.) 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):

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

(132.) 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:

(133.) **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*).

(134.) **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):

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

(136.) 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':

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

(138.) 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:

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

(140.) **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:

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

(142.) **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:

(143.) **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:

(144.) [**Estudio multicéntrico, aleatorizado, doble ciego, con control activo**] para [**evaluar la eficacia**] en la semana 52 de **secukinumab** (2015-004477-32)
Annotate 2 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...)

## 2. 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...):

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

  (146.) 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 interpreted as PROC: *medición de* + CHEM (e.g. *medición de calcio*) CHEM *total* (e.g. *calcio total*), *valores de* + CHEM, *cifras de* + CHEM...

  (147.) 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).

  (148.) 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 interpretated as the therapeutic procedure. The word *vacuna* is annotated as CHEM when it refers to the medical drug or the pharmacological substance:

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

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

(151.) 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:

(152.) 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.

(153.) 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':

(154.) 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)

(155.) 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:

(156.) 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:

(157.) **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)


**Entities referring to drug mode of administration**

These types of entities are very relevant clinically. The UMLS classifies them as CONC (concept entities expressing modes or routes of administration related to drugs or procedures). We do not annotate them in the current version, but it would be very positive to label them in a future version:

(158.) recibieron 5 mg <u>orales</u> de ácido fólico         (0211-699501013259)

(159.) la infusión de dexmedetomidina <u>intravenosa</u>
(S1134-80462015000100002)

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

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

We annotate them only adjacent to PROC entities in the current version:

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

(162.) tratamiento con **bolo**         (S1130-01082008000800004)
Annotate *tratamiento con bolo* as PROC.

We do not annotate ANAT in procedure entities expressing the administration route; the ANAT label is only marked in mentions referring to anatomic sites:

(163.) **Cirugía <u>transnasal</u> <u>transesfenoidal</u> endoscópica**
                                        (S1130-14732005000100004)
Annotate only *cirugía transnasal transesfenoidal endoscópica* as PROC

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

We do not annotate entities expressing administration route adjacent to pharmacological entities. Although the full term may exist in the UMLS, we only annotate CHEM in the substance or drug. **Exceptionally, we annotate it if the entity referring to the administration route occurs inside the drug entity** (when the full mention cannot be split):

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

(166.) el consumo de **anestésicos** venosos  (S0034-70942004000100013)
Annotate only *anestésicos* as CHEM

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

**Entities expressing drug concentration, strength and/or dosage**

Likewise, at this stage of the project, we do not annotate these types of entities, but only the pharmacological or chemical substance:

(168.) **tintura de propóleos** al 10%       (S0034-75072007000300006)
Annotate only *tintura de propóleos* as CHEM.

(169.) **calcio** en dosis de 1.000 mg cada 24 horas (S0212-16112015000500025)
Annotate only *calcio* as CHEM

**Entities expressing drug dosage forms**

We do not annotate at this stage the following types of terms: *aerosol, cápsulas, comprimidos, crema, formulación, gel, gotas, tabletas...*

(170.) comparar la efectividad de 800 UI/día de **D2** (gotas) y **D3** (comprimidos)
                                          (S0025-76802012000300002)
Annotate only *D2* (C0014695 Ergocalciferol) and D3 (C0008318 Cholecalciferol) as CHEM

(171.) 12,5 gr de gel con 250 mgr de **hidrocloruro de lidocaína** intra-rectal
                                          (S0004-06142009000500003)
Annotate only *hidrocloruro de lidocaína* as CHEM.

(172.) **montelukast** en tabletas masticables de 5 mg
                                          (S0120-41572012000300010)
Annotate only *montelukast* as CHEM.

We annotate the term *solución* as CHEM in compound terms referring to the type of pharmacological drug used in the preparation (e.g. *solución glucosada*), or in cases in which the term *solución* implies that a pharmacological or chemical substance is diluted in water or another drug diluent:

(173.) el uso de la **solución glucosada** por vía oral (S0325-00752008000500004)
Annotate *solución glucosada* as CHEM.

(174.)  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*

## Entities referring to diagnostic measurement scales

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

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

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

(177.) **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*

We do not annotate terms classified as scale or classification in the UMLS, but not
classified as diagnostic procedure (e.g. *escala* or *puntaje*):

(178.) presentaron menor escore de sedación según la escala de Ramsay
(S0066-782X2011000300008)
Do not annotate *escala de Ramsay* (concept entity, CONC: C0559295 Ramsay
sedation scale; Intellectual Product)

(179.) Impresión Clínica Global                    (S0211-57352011000100002)
Do not annotate (C3639708 Clinical global impression scale, CONC)

(180.) Inventario de Discapacidad de Sheehan    (S0211-57352011000100002)
Do not annotate (C4720840 Sheehan disability scale, CONC)

## Entities referring to Associations, Medical specialties or Hospital units

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

(181.) en la Unidad de Digestivo del Complejo Hospitalario de Ciudad Real
Do not annotate *Unidad de Digestivo* nor *Digestivo*
(S0212-71992005001000008)

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


**Entities referring to medical devices**

In the current version, entities from the DEVI semantic group are not annotated (they could be annotated in future versions of the corpus):

(183.) Una <u>sonda nasogástrica 20 F</u> fue insertada  (S0212-16112011000100009)
Do not annotate *sonda nasogástrica 20 F*.

(184.) El UDP se midió, por medio de un <u>algómetro analógico</u>
      Do not annotate *algómetro analógico*        (S1134-80462010000700004)

These entities would only be annotated if they are included in other PROC entities: e.g. *colocar una sonda, alimentación por sonda...*

(185.) garantiza la óptima **colocación del catéter** (S0211-69952010000300014)
Annotate *colocación del catéter* as PROC.

(186.) evaluó la continencia de orina a los 14, 30 y 60 días post **extracción de sonda**
Annotate *extracción de sonda* as PROC.        (S0004-06142008000700005)

⚠️ Note the difference: *sondar, sondaje, o cateterizar* are annotated as PROC (they express a procedure); however, *sonda* is not annotated if the term appears alone (it is a considered medical device).


**Entities expressing microorganisms or pathogenic agents**

In general, we do not annotate virus, hongos or bacteria. **Excepcionally**, we annotate mentions of **virus** (e.g. *virus de la hepatitis B, virus de inmunodeficiencia humana*) **in contexts where the entity refers to the disease**, or if the entity can be **interpretated as pathologic state** (e.g. *VIH+* implying *enfermo de SIDA*):

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

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

We also annotate DISO in contexts such as *infección por VIH, infección VHC…* In these cases, we only annotate the full entity if it is not very long. Past participles (e.g. *causado por, producido por…*) are not annotated if the entity is very long and cannot be normalized to a CUI:

(189.) **leishmaniasis cutánea** producida por Leishmania (V.) panamensis
                    (S0212-16112018000100162)
      Annotate only *leishmaniasis cutánea* (C0023283 Leishmaniasis, Cutaneous)

On the contrary, if the entity with a past participle (*causado, producido...*) is not very long, and can be normalized to a CUI, we annotate the full entity (to avoid discontinuous annotations):

(190.) manejo de **infecciones producidas por Bacteroides fragilis**
(S0716-10182015000300001)
Annotate DISO (C2205936 Infection by Bacteroides fragilis)

## Entities expressing genes

In the current stage, we do not annotate these types of entities (the UMLS classifies them in the GENE semantic group).

(191.) **mutación** en uno de los cinco genes <u>HMR</u> (<u>ASXL1, EZH2, SRSF2</u> e <u>IDH1/2</u>) (2014-004928-21)
Annotate only *mutación* (DISO)

# 1. APPENDIX

**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*):

> (192.) El grado final de **envenenamiento** fue leve en 13 (24,5%), moderado en 30 (56,6%) y grave (S0121-07932007000300002)
> Annotate only *envenenamiento* (DISO); do not label *grave*, *leve* or *moderado*.

| Annotate | Do not annotate |
|---|---|
| Specific temporal modifiers or adjectives in multi-word terms, when they can be mapped to a CUI or ICD-10 category:<br><br>*crónico:* e.g. *fisura anal crónica* (ICD-10: K60.1, C0349071 Chronic anal fissure)<br><br>*postoperatorio*: e.g. *dolor postoperatorio* (C0030201 postoperative pain; ICD-10: G89.18)<br><br>*postpunción*: e.g. *cefalea post-punción lumbar* (C2875393 Headache due to lumbar puncture; ICD-10: G97.1)<br><br>*postprandial*: e.g. *glicemia postprandial* (C4700958 postprandial glucose in serum or plasma) | Vague temporal qualifiers:<br>*actual, anterior, futuro, inicial, inmediato, precoz, previo, siguiente, posterior, reciente...*<br><br>Exceptionally, we annotate these adjectives if they are necessary to understand the context, or if the adjective distinguishes different concepts (CUI exist):<br><br>E.g. *inflamación persistente* (C1265824 Persistent inflammation)<br><br>E.g. *eyaculación precoz* (C0033038, ICD-10: F52.4)<br><br>E.g. *seguimiento a largo plazo* (C1517942 Long-term Follow-up)<br><br>We do not annotate adjetives expressing 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:<br><br>*alta*: e.g. en *endoscopia alta* (C0017195 upper endoscopy)<br><br>*central*: e.g. *sistema nervioso central* | Past participles without specific details or descriptive information about the term:<br><br>*administrado*<br>*asociado*<br>*conocido*<br>*empleado*<br>*especializado*<br>*establecido* |

| | |
|---|---|
| (C1269563 central nervous system)<br><br>*local*: e.g. *anestésico local* (C0002934 Local Anesthetics)<br><br>*superior* (e.g. *miembro superior*, C1140618 Upper Extremity)<br><br>*unilateral*: e.g. *dolor unilateral* (C0747150 pain unilateral; Sign or Symptom) | *generado*<br>*realizado*<br>*reconocido*<br>*relacionado*<br>*sometido*<br>*tratado* |
| 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:<br><br>*clínico*: e.g. *ensayo clínico* (C0008976 Clinical Trial)<br><br>*medible*: ej, *enfermedad medible* (C1513041 Measurable Disease), *lesión no medible* (C1334988 Non-Measurable Lesion) | Vague qualifiers and adjetives not describing specific details:<br><br>*diferente*<br>*distinto*<br>*diverso*<br>*mismo*<br>*nuevo*<br>*principal*<br>*solo*<br>*único* |
| Adjetives describing the type of procedure (the *Comparator*), or when they are needed to understand the context:<br><br>*convencional:* e.g. *cirugía convencional* (C1521745 Conventional Surgery)<br><br>*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:<br><br>E.g. *dolor refractario* (C0030200 Refractory Pain) | *refractario a* + phrase: we do not annotate it when other phrases modify it (to avoid long annotations):<br><br>E.g. *hipertensión intracraneal (HIC) <u>refractaria a</u> las medidas de primer nivel*<br><br>Annotate only *hipertensión intracraneal* |
| *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: | *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. **hiperparatiroidismo secundario** *en diálisis*<br>Annotate *hiperatiroidismo secundario* (C0494305 Secondary hyperparathyroidism, ICD-10: E21.1) | E.g. **rectorragia** *crónica* <u>*secundaria a*</u> *la* **proctopatía por radiación**<br>Annotate two different terms: *rectorragia* and *proctopatía por radiación* |
| 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:<br><br>*recidivante*: e.g. *pterigium recidivante* (C0155158 Recurrent pterygium, ICD-10: 11.06)<br><br>*recurrente*: e.g. *estomatitis aftosa recurrente* (C2937365 Recurrent aphthous stomatitis; ICD-10: K12.0)<br><br>*remitente*: e.g. *esclerosis múltiple remitente-recurrente* (C0751967 Multiple Sclerosis, Relapsing-Remitting) | Evaluative adjetives:<br><br>*beneficioso*<br>*bueno*<br>*efectivo*<br>*fundamental*<br>*ideal*<br>*incómodo*<br>*malo*<br>*útil*<br>*viable* |
| 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:<br><br>*activo*: annotate only in terms with an UMLS CUI:<br>E.g. *Inmunización activa* (C0042196 Active immunization)<br><br>E.g. *EC fistulosa* <u>*activa*</u><br>Annotate only *EC fistulosa* (C3509259 crohn's disease with fistula)<br><br>*abierto*: e.g. *cirugía abierta* (C4283938 Open Surgical Procedure)<br><br>*basal*: e.g. *tasa metabólica basal* (C0542499 Measurement of basal metabolic rate)<br><br>*continuo*: e.g. *perfusión continua* (C0444889 Continuous infusion)<br><br>*descompensado*: e.g. *EPOC descompensado* (C2887446 Decompensated COPD; ICD-10: J44.1) | Comparative adjetives:<br><br>*inferior* (*que*)<br>*mayor* (*que*)<br>*mejor* (*que*)<br>*peor* (*que*)<br>*superior* (*que*)<br><br>Except in contexts where the adjective needs to be annotated with the main term, because there is an ICD-10 category or UMLS CUI:<br><br>*Cirugía mayor*<br>(C0679637 major surgery)<br><br>*Trastorno depresivo mayor*<br>(C1269683 Major Depressive Disorder, ICD-10: F32.9) |

| | |
|---|---|
| *primario*: e.g. *hipertiroidismo primario,* C0221002 Hyperparathyroidism, Primary; ICD-10: E21.0) <br><br> *sistémico*: e.g. *esclerosis sistémica,* (C0477589 systemic sclerosis, ICD-10: M34.8) <br><br> *total*: e.g. *anestesia intravenosa total* (C0473965, Total intravenous anesthesia) | |
| Descriptors of severity or pathological stage: we only annotate them when a full term has an ICD-10 category: <br><br> *grave:* e.g. *miastenia grave* (C0026896 Myasthenia Gravis; ICD-10: G70.0) <br><br> In general, we do not annotate other vague or imprecise adjectives such as *serio, importante,* etc.: <br><br> E.g. *eventos adversos serios* Annotate only *eventos adversos* <br><br> E.g. **afectación neurológica** *importante* Annotate only *afectación neurológica* | Adjetives or past participles expressing possibility, probability, assertion, certainty or uncertainty, or disease activity: <br><br> *activo* <br> *confirmado* <br> *posible* <br> *probable* <br> *sospechoso* <br> *sugestiva* <br> *susceptible* <br><br> E.g. *sepsis nosocomial confirmada o altamente probable* Annotate only *sepsis nosocomial.* <br><br> E.g. *tuberculosis activa* Annotate only *tuberculosis.* |
| Adjetives specifying the type of measurement or average values measured: we annotate them if a concept (CUI) exists in the UMLS: <br><br> *medio*: <br><br> E.g. *presión arterial media* (C0428886 Mean blood pressure) <br><br> E.g. *volumen corpuscular medio* (C0863148 Mean corpuscular volume analyses, PROC) <br><br> *promedio*: <br> 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: <br><br> *medio* <br> E.g. *pérdida de peso media* (annotate only *pérdida de peso*) <br><br> *promedio* <br> E.g. *IMC promedio* (annotate only *IMC*) |

## 2. REFERENCES

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

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.

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.
Disponible en: 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 (Fecha de consulta: 26/02/2020)

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 (Fecha de consulta: 17/02/2020)

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/ (Fecha de consulta: 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 (Fecha de consulta: 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. of the 56th Annual Meeting of the Association for Computational Linguistics,* Melbourne, Australia, Julio 2018, pp. 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, 318-332.

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

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 (Fecha de consulta: 17/02/2020)

Temporal Histories of Your Medical Event (THYME) Project. https://clear.colorado.edu/TemporalWiki/index.php/Main_Page  (Fecha de consulta: 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. The *Concept Annotation Guidelines* are available at: https://www.i2b2.org/NLP/Relations/assets/Concept Annotation Guideline.pdf (Fecha de consulta: 17/02/2020)