

# Recursos para procesamiento del lenguaje médico en español



CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Leonardo Campillos Llanos

ILLA Instituto de Lengua, Literatura y Antropología



Proyecto CLARA-MeD financiado por la AEI/MICIN/10.13039/501100011033 en la convocatoria RETOS 2020 (PID2020-116001RA-C33)





# Recursos léxicos

- **MedLexSp\*:**

+100000 entradas,  
+300000 formas  
variantes

\*Resultado de  
proyecto anterior

NLP MED TERM



- **SimpMedLexSp:**

equivalentes  
técnicos y  
simplificados.

(en construcción)

```
<LexicalEntry id="C0011849_diabético" partOfSpeech="adjective_or_noun">
  <Lemma writtenForm="diabético">
    <WordForm grammaticalGender="masculine" grammaticalNumber="singular">diabético</WordForm>
    <WordForm grammaticalGender="feminine" grammaticalNumber="singular">diabética</WordForm>
    <WordForm grammaticalGender="masculine" grammaticalNumber="plural">diabéticos</WordForm>
    <WordForm grammaticalGender="feminine" grammaticalNumber="plural">diabéticas</WordForm>
  </Lemma>
  <SourceDC sourceName="UMLS" sourceId="C0011849" />
  <SemanticType val="Disease_or_Syndrome" />
  <SemanticGroup val="DISO" />
</LexicalEntry>
<LexicalEntry id="C0011849_diabetes_mellitus" partOfSpeech="noun">
  <Lemma writtenForm="diabetes mellitus">
    <WordForm grammaticalGender="feminine" grammaticalNumber="singular_and_plural">diabetes mellitus</WordForm>
  </Lemma>
  <SourceDC sourceName="UMLS" sourceId="C0011849" />
  <SemanticType val="Disease_or_Syndrome" />
  <SemanticGroup val="DISO" />
</LexicalEntry>
<LexicalEntry id="C0011849_dm" partOfSpeech="noun">
  <Lemma writtenForm="dm" VariantType="acronym">
    <WordForm grammaticalGender="feminine" grammaticalNumber="singular">dm</WordForm>
  </Lemma>
  <SourceDC sourceName="UMLS" sourceId="C0011849" />
  <SemanticType val="Disease_or_Syndrome" />
  <SemanticGroup val="DISO" />
</LexicalEntry>
```

L. Campillos-Llanos (2023) MedLexSp – a medical lexicon for Spanish medical natural language processing. *Journal of Biomedical Semantics*

<https://doi.org/10.1186/s13326-022-00281-5>

<https://digital.csic.es/handle/10261/270429>

```
CUI|TECHNICAL_TERM|PATIENT_TERM
C0040423|amigdalectomía|operación de anginas; operación de las anginas
C0040652|transcutáneo|a través de la piel
C0175668|secundario a|a consecuencia de
C0231221|asintomático|sin síntomas
C0277797|apirético|sin fiebre
C0522523|percutáneo|a través de la piel
```

**+12000 formas variantes**

## 1. CT-EBM-SP: 1200 textos (abstracts + anuncios ensayos), entidades UMLS.

**Fase actual:** extensión a más tipos de entidades y anotación relaciones

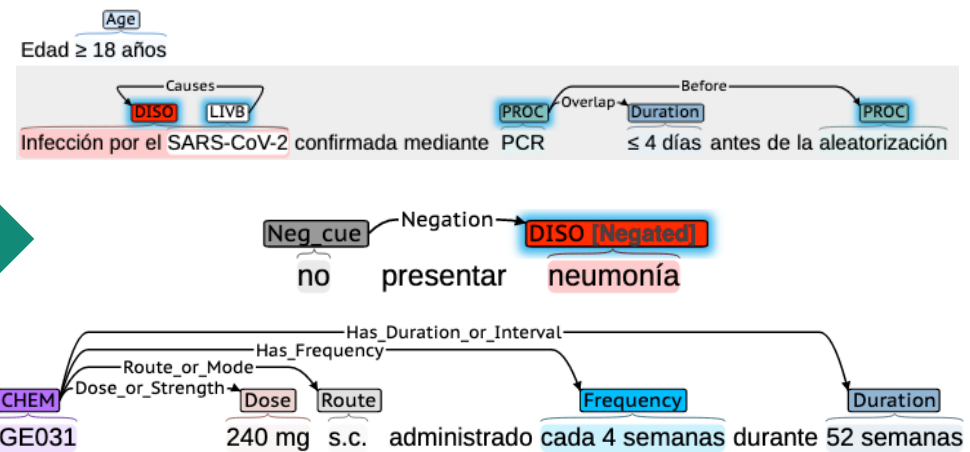
Título científico: Estudio de Fase 3 Aleatorizado para Evaluar la Seguridad y la Actividad Antiviral de Remdesivir (GS-5734™) en

Participantes con infección grave por el COVID-19.

Indicación pública: Enfermedad por el Coronavirus 2019 (COVID-19)

Criterios de inclusión:

- 1) Disposición y capacidad para otorgar el consentimiento informado por escrito antes de realizar los procedimientos del estudio
- 2) Edad  $\geq$  18 años
- 3) Infección por el SARS-CoV-2 confirmada mediante PCR  $\leq$  4 días antes de la aleatorización
- 4) Hospitalización con fiebre, definida como una temperatura  $\geq$  36,6 °C en la axila,  $\geq$  37,2 °C en la boca y  $\geq$  37,8 °C en el recto



L. Campillos-Llanos et al. (2021) <https://doi.org/10.1186/s12911-021-01395-z>  
<https://zenodo.org/record/6059737>

## 2. Corpus CLARA-MeD: 24298 pares de textos técnicos/simplificados (Digital CSIC)

CIMA

*La administración concomitante de metazolol con metotrexato u otros antineoplásicos puede aumentar la toxicidad sanguínea de los antineoplásicos particularmente en pacientes de edad avanzada.*

*Si se administra conjuntamente con metotrexato u otros medicamentos para el tratamiento de los tumores (antineoplásicos), puede potenciar los efectos tóxicos en sangre de los antineoplásicos, sobre todo en pacientes de edad avanzada.*

L. Campillos-Llanos et al. (2022) <https://doi.org/10.26342/2022-69-16>

# Reconocedor de entidades médicas

## MEDSpaNER Herramienta de anotación semántica médica

Posología
  Expresiones temporales
  Negación
  Entidades anidadas
  Léxico UMLS
  Modelo neural UMLS
  Normalizar

La COVID-19 afecta de distintas maneras en función de cada persona . La mayoría de las personas que se contagian presentan síntomas de intensidad leve o moderada, y se recuperan sin necesidad de hospitalización .  
 Los síntomas más habituales son los siguientes: fiebre , tos seca y cansancio .  
 Otros síntomas menos comunes son los siguientes: Molestias y dolores , dolor de garganta , diarrea , conjuntivitis , dolor de cabeza , pérdida del sentido del olfato , erupciones cutáneas en los dedos de las manos o de los pies .  
 (www.who.int)

Patología - UMLS: C0015230; Exanthema; Sign or Symptom

- Evaluación por profesionales sanitarios.
- Pronto disponible código y modelos transformers.



Model card | Files and versions | Community 1 | Settings

**roberta-es-clinical-trials-ner**

This medical named entity recognition model detects 4 types of semantic groups from the Unified Medical Language System (UMLS) (Bodenreider 2004):

- ANAT: body parts and anatomy (e.g. *garganta*, 'throat')
- CHEM: chemical entities and pharmacological substances (e.g. *aspirina*, 'aspirin')
- DISO: pathologic conditions (e.g. *dolor*, 'pain')
- PROC: diagnostic and therapeutic procedures, laboratory analyses and medical research activities (e.g. *cirugía*, 'surgery')

The model achieves the following results on the evaluation set:

Downloads last month: 2,634

Hosted inference API

Token Classification

El ensayo clínico PROC con vacunas CHEM promete buenos resultados para la infección por SARS-CoV-2.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 7.342 s

El ensayo clínico PROC con vacunas CHEM promete buenos resultados para la infección por SARS-CoV-2 DISO.