

## Medical Lexicon for Spanish (MedLexSp) - LICENSE AGREEMENT

Licensers: Consejo Superior de Investigaciones Científicas (CSIC), Laboratorio de Lingüística Informática (Universidad Autónoma de Madrid), and Unidad de Terminología Médica (Real Academia Nacional de Medicina de España, UTM-RANME

Licensee: \_\_\_\_\_ agree as follows:

### 1. Definition

The Medical Lexicon for Spanish (MedLexSp) is a unified lexicon for the Spanish language, featuring terms and inflected word forms mapped to Unified Medical Language System (UMLS®) Concept Unique Identifiers (CUIs), semantic types and groups. The lexicon contains linguistic information such as part-of-speech tag and morphological data (number, gender, verb tense and person, and information about affix/abbreviation data). MedLexSp is aimed at Natural Language Processing (NLP) tasks in the Spanish language and the health domain.

### 2. Status of Licensee

Licensee confirms that Licensee is an academic, educational or research institution. MedLexSp is distributed freely for research or educational purposes.

### 3. Product

The Medical Lexicon for Spanish gathers term entries with the lemma and inflected forms, along with its UMLS® CUIs, semantic types and groups. Terms come from several sources:

- Terminologies in the Spanish language in the UMLS® Metathesaurus: the Medical Subject Headings, the World Health Organization (WHO) Adverse Drug Reactions terminology, the International Classification of Primary Care (ICPC), the Medical Dictionary for Regulatory Activities Terminology (MedDRA) and the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT®). MedLexSp also includes terms mapped to Spanish through the UMLS from the English versions of the International Classification of Diseases vs. 10 (ICD-10), the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5®) and the Online Mendelian Inheritance in Man (OMIM®) catalog of genes and genetic disorders.
- A subset of terms from the *Diccionario panhispánico de términos médicos (DPTM)* developed by Unidad de Terminología Médica (Real Academia Nacional de Medicina de España), in collaboration with Asociación Latinoamericana de Academias Nacionales de Medicina, España y Portugal (ALANAM).
- Supplementary terminological sources such as the Anatomical Therapeutical Classification from the WHO, the National Cancer Institute (NCI) Dictionary of Cancer Terms, OrphaData from INSERM, or the Nomenclátor de Prescripción from AEMPS.
- Corpus-based terms derived from the Spanish version of MedlinePlus (supported by the National Library of Medicine) and Summaries of Product Characteristics in Spanish (used in the Spanish Drug Effect database, by Universidad Carlos III de Madrid).
- Terms derived from annotated corpora: the Chilean Waiting List Corpus (CWLC by Universidad de Chile) and the Clinical Trials for Evidence-Based Medicine in Spanish (CT-EBM-SP) corpus (by CSIC & Universidad Autónoma de Madrid).
- Terms derived from corpora used in shared tasks organized by the Biomedical Text Mining Unit: Cancer Text Mining Shared Task (CANTEMIST), Clinical Case Coding in Spanish Shared Task (CODIESP) and Pharmacological Substances, Compounds and proteins and Named Entity Recognition (PharmaCoNER) challenge.
- Linguistic information derived from the DELAS electronic dictionaries for Spanish and the SPACCC POS-TAGGER.

The version of MedLexSp freely available does not include terms or coding data from terminological sources with copyright rights; only the subset of data in MedLexSp without usage restrictions is accessible.

Licenser and Licensee recognize that the copyrights of original terms and source data belong to the Institutions who provided them (see **Appendix**). All rights in and to MedLexSp, including copyright, which are not granted to the Licensee in this Agreement, are retained by Consejo Superior de Investigaciones Científicas (CSIC), Laboratorio de Lingüística Informática (Universidad Autónoma de Madrid), Unidad de Terminología Médica (UTM-RANME) and the source providers.

MedLexSp has been developed under the NLPMedTerm project, which received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie COFUND grant agreement n°. 713366 (InterTalentum UAM); and under the CLARA-MeD project, (PID2020- 116001RA-C33) funded by MCIN/AEI/ 10.13039/501100011033/, in project call: "Proyectos I+D+i Retos Investigación". The LLI-UAM and UTM-RANME signed a collaboration agreement (2019-2020).

#### **4. License**

Licenser grants Licensee a non-exclusive license to use the MedLexSp lexicon. Licensee agrees:

- to use the lexicon only for non-commercial, non-profit research purposes;
- to make no changes to the lexicon;
- and to acknowledge the use of the lexicon in all publications reporting on results produced with the help of MedLexSp.

This license is granted by Licenser to Licensee free of charge. You agree that the owner has the right to publicly identify you as a User/Licensee of MedLexSp. This Agreement and the appendix hereto embody the entire understanding between the parties relating to the subject matter hereof, and there are no terms or conditions hereof express or implied written or oral. This Agreement supersedes all prior oral or written representations, agreements, promises or other communications, concerning or relating to the subject matter of this Agreement. Use of the lexicon or use of data derived from the lexicon for any commercial purposes requires explicit written agreement of Licenser.

#### **5. Non-Disclosure and Requirement to Exercise Professional Care**

The Medical Lexicon for Spanish will be held in confidence by Licensee. The Licensee shall not copy, publish, distribute or otherwise give access to any third party to the whole or any part of MedLexSp. The Licensee shall ensure that its employees understand and abide by this restriction and shall supervise their activities with respect to the lexicon. The Licensee agrees to MedLexSp in confidence and treats such Materials with at least the same degree of care that it treats its own most confidential information. The Licensee may not assign, transfer, lease, sell, rent, charge or otherwise encumber MedLexSp.

#### **6. Termination**

This Agreement will take effect from the date of signature and will continue until terminated either by Licensee or Licenser. Licenser may terminate the agreement forthwith if the Licensee commits any material breach of this agreement or if the copyright holder raises an objection against the terms of the license agreement. Licensee may terminate this agreement within 30 days of written notice. On termination howsoever caused, the Licensee shall erase, or otherwise

destroy, all full and partial copies of MedLexSp. A duly authorized officer of the Licensee shall promptly verify in writing to the contact person of Consejo Superior de Investigaciones Científicas (CSIC) that the Licensee has complied with this obligation.

## **7. Disclaimer**

This license is granted under the premise that the copyright holder of the raw data accepts the terms of the agreement. If the copyright holder raises an objection against the terms of the agreement, Licenser may terminate the agreement and Licensee shall then erase all copies of MedLexSp. No liability is accepted for any damage caused by the use of this lexicon. Licenser hereby informs Licensee that the distributed version of MedLexSp is work in progress, and despite that large efforts are spent on removing errors, some of them still remain in the lexicon.

---

## **LICENSEE**

Institution: \_\_\_\_\_

Name: \_\_\_\_\_

Title: \_\_\_\_\_

Signature:

Date:

## Appendix to the MedLexSp License agreement

Some thesauri included in MedLexSp were obtained through a distribution and usage agreement from the corresponding institutions that develop and update them.

Some material in the UMLS® Metathesaurus is from copyrighted sources of the copyright holders. Users of the UMLS® Metathesaurus are solely responsible for compliance with any copyright, patent or trademark restrictions and are referred to the copyright, patent or trademark notices appearing in the original sources, all of which are hereby incorporated by reference.

The Licenser acknowledges the intellectual property rights of the institutions who develop the sources from which subsets of terms were extracted to compile the lexicon, and who gave permission (or provide a license to reuse their data) to distribute these subsets of terms:

- The Medical Terminology Unit (Unidad de Terminología Médica) from the Real Academia Nacional de Medicina de España develops the *Diccionario panhispánico de términos médicos (DPTM)* in collaboration with the Asociación Latinoamericana de Academias Nacionales de Medicina, España y Portugal (ALANAM).
- The National Library of Medicine maintains the MedlinePlus resource and the Medical Subject Headings, and BIREME/OPS (Latin-American and Caribbean Center on Health Sciences Information) is in charge of the Spanish translation (Descriptores en Ciencias de la Salud, DeCS). The Spanish translation of the Medical Subject Headings is available thanks to a 3-year license.
- The National Cancer Institute publishes the Dictionary of Cancer Terms and maintains its Spanish translations.
- The American Psychiatric Association publishes the *Diagnostic and statistical manual of mental disorders* (5th ed.) and the related English psychiatric terms.
- The McKusick-Nathans Institute of Genetic Medicine (John Hopkins University, Baltimore, MD) supports the Online Mendelian Inheritance in Man (OMIM®) catalog of genes and genetic disorders.
- The French National Institute of Health and Medical Research (INSERM) supports OrphaNet and gathers the information provided in OrphaData.
- The World Health Organization (WHO) produces the Adverse Drug Reactions terminology, the International Classification of Diseases vs. 10 (ICD-10), and the Anatomical Therapeutical Classification.
- The World Organization of Family Doctors supports the Spanish translation of the International Classification of Primary Care (ICPC).
- The Spanish Agency of Drugs and Food Products (AEMPS) publishes the Nomenclátor de prescripción.
- Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), under Plan de Tecnologías del Lenguaje, funded the development of the SPACCC POS Tagger by Felipe Soares; and funded the evaluation campaigns CANTEMIST, CODIESP and PharmaCoNER, led by the team of Dr. Martin Krallinger.
- Dr. Xavier Blanco collected the DELAS electronic dictionaries for Spanish, as explained in the following article: X. Blanco (2000) Les dictionnaires électroniques de l'espagnol (DELASs et DELACs). *Linguisticae Investigationes*, 23(2):201–18.
- Dr. Isabel Segura Bedmar and Dr. Paloma Martínez Fernández collected the Spanish Drug Effect database (Universidad Carlos III de Madrid).
- Dr. Jocelyn Dunstan and her team (Universidad de Chile) collected the Chilean Waiting List Corpus (CWLC)
- Dr. Leonardo Campillos-Llanos and his collaborators collected the Clinical Trials for Evidence-Based Medicine in Spanish (CT-EBM-SP) corpus.

MedLexSp includes a subset of the Spanish version of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT®), used by permission of the International Health Terminology Standards Development Organization (IHTSDO; all rights reserved). SNOMED CT® was created by The College of American Pathologists. The distributed version of MedLexSp does not include terms coming solely from SNOMED-CT®, because of copyright restrictions.

MedLexSp also gathers some terms from the Spanish version of the Medical Dictionary for Regulatory Activities (MedDRA), which is maintained by the Maintenance and Support Services Organization (MSSO). However, the distributed version of MedLexSp does not include terms coming solely from the MedDRA sources, because of copyright restrictions.