CURRICULUM VITAE (CVA)

IMPORTANT – The Curriculum Vitae cannot exceed 4 pages. Instructions to fill this document are available in the website.

CV date	09-03-2022

Part A. PERSONAL INFORMATION

First name	Maite	
Family name	Oronoz	
e-mail	maite.oronoz@ehu.eus	URL Web
Open Research and Contributor ID (ORCID)(*)	0000-0001-9097-6047	

(*) Mandatory

A.1. Current position

Position	Personal Docente Investigador (Category: Profesor Agregado)		
Initial date	23/10/2001		
Institution	University of the Basque Country (UPV/EHU)		
Departament/Center	Lenguajes y Sistemas Informáticos		
Country	Spain	Teleph. number	943015298
Key words	NLP, medical domain, SNOMED CT translation, Electronic Health Record translation, Medical Entity Recognition		

Part B. CV SUMMARY *(max. 5000 characters, including spaces)*

Bachelor Degree in Computer Science and Doctor in Computer Science from the University of the Basque Country and member of the Department of Computer Languages and Systems. Since its creation in February 2019, I am a committee member of HiTZ (Basque Center for Language Technology) and coordinator of the master Language Analysis and Processing from 2018 (also in the Academic committee of the Erasmus Mundus master LCT). As coordinator of the master in NLP I have guided students (around 60 national and international students in the last 3 years) in terms of research topics, proposing ideas for their research work in the master's thesis and coordinating with possible thesis supervisors.

My research focuses on Artificial Intelligence within the field of Natural Language Processing (NLP) and Understanding (NLU) applied to Clinical Text Mining (CTM) and machine translation. CNAI recognised my first six-year research period in 2015 and I am currently applying for the second one.

I have published more than 90 publications in national and international conferences and journals (http://ixa.si.ehu.es/node/1394/104) in two areas: automatic grammatical error detection and correction and applications of NLP to the medical domain. From these papers 10 are indexed in Journal Citation Reports with 5 publications in the first quartile (Q1), 4 in the second one (Q2) and 1 in Q3. I have an h-index in Scopus of 10 and of 15 in Google Scholar. In our group we take active part in international shared tasks and challenges connected with CTM and we got awarded in CLEF-2018. I am actively involved in the review of indexed journal articles within my field (e.g. Expert Systems and applications, CMPB, BMC).

I have been advisor and co-advisor of 2 international PhD theses in the last ten years. One of these students is currently a lecturer at the Computer Science School at the UPV/EHU. I

was also co-advisor of the master's degree project of both researchers. Both doctoral thesis were related to the medical domain, i) one about the automatic translation of the medical lexicon SNOMED CT into Basque (2017) and, ii) the other one about neural machine translation of electronic health records between Spanish and Basque (2021). Currently I am co-director of a PhD thesis in medical named recognition for ICD coding and co-advisor of 5 master's thesis, 4 of then in the application of NLP in the medical domain (e.g. federated machine learning for medical named entity detection, argumentation in the medical domain). The scientific contribution has been carried out within the framework of the Participation in 45 Research projects, funded by: 3 European Commission (Antidote, LoCloud, Meaning); 7 Spanish Ministry; 1 BBVA; 20 Basque Government; and the rest University and local council. The ANTIDOTE (ArgumeNtaTIon-Driven explainable artificial intelligence fOr digiTal mEdicine) CHIST-ERA project is aligned with the OBSER-MEN project regarding explainability.

With regard to the transference of Knowledge to the society: I contributed with 5 intellectual property registrations, one of them related to CTM. I have been the main researcher in two main contracts with Osakidetza. One of them for the automatic translation into Basque of the International Classification of disorders (14.534,04 €) and in the other one we won a public bid to develop and neural machine translation system to translate electronic health records between Spanish and Basque (191.469,58 €). In this period my relationship with workers and some directors of the Basque Health Service has been continuous. Previously I was PI in contracts with INDRA S.A. (90.604,19 €) and Elhuyar (8.748,48 €). As the coordinator of the master's degree, I have an ongoing relationship with companies in the area of AI and NLP to find and offer internships to students of both master's degrees.

Part C. RELEVANT MERITS *(sorted by typology)*

C.1. Publications *(see instructions)*

[1] (4/4): *X. Soto, O. Perez-de-Viñaspre, G. Labaka, **M. Oronoz** (2019). Neural Machine Translation of clinical texts between long distance languages. JAMIA. IF: 4.112, **JCR-Q1***

[2] *(4/4) S.Santiso, A. Pérez, A. Casillas, **M. Oronoz** (2020). Neural negated entity recognition in Spanish electronic health records. Journal of Biomedical Informatics (JBI). 2019 IF 3.526, **JCR-Q2***

[3] *(4/5) A. Atutxa, A.Diaz de Ilarraza, K.Gojenola, **M. Oronoz**, O. Perez de Viñaspre (2019). Interpretable Deep Learning to Map Diagnostic Texts to ICD10 Codes. International Journal of Medical Informatics. 2019 IF :3.025, **JCR-Q2***

[4] *(4/4) R. Weegar, A. Pérez, A. Casillas, **M. Oronoz** (2020).* Recent advances in Swedish and Spanish medical entity recognition in clinical texts using deep neural approaches. *BMC Medical Informatics and Decision Making. 2018 IF: 2.067, **JCR-Q3***

[5] *(3/5)A. Casillas, A. Pérez., **M. Oronoz**, K.Gojenola, S. Santiso (2016) Learning to extract adverse drug reaction events from electronic health records in Spanish. Expert Syst. Appl., 61, 2016. JCR 2016: 3.928, 5-Year IF: 3.526, **JCR-Q1**.*

[6] *(5/6) A. Pérez, R. Weegar, A. Casillas, K. Gojenola, **M. Oronoz**, H. Dalianis.* Semi-supervised medical entity recognition: A study on Spanish and Swedish clinical corpora**,** *Journal of Biomedical Informatics, Volume 71, 16-30, 2017. IF: 2.753, **JCR-Q1**.*

[7] (1/5**)** ***M. Oronoz**, K. Gojenola, A. Pérez, A. Díaz de Ilarraza, A. Casillas 2015. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. Journal of Biomedical Informatics, Volume 56, 318–332, IF: 2.194, **JCR-Q1**.*
[8] (3/4): *A. Pérez, K. Gojenola, A. Casillas, **M. Oronoz,** A. Díaz de Ilarraza:Computer*

aided classification of diagnostic terms in Spanish. Expert Syst. Appl. 42(6): 2949-2958 (2015) **JCR-Q1**

[9] *(2/2) O. Perez-de-Viñaspre, M. Oronoz 2015.* SNOMED CT in a language isolate: an algorithm for a semiautomatic translation. *BMC Medical Informatics and Decision Making 2015, 15(2):S5. IF: 1.830,***JCR-Q2**

[10] (4/4) *S.Santiso, A. Casillas, A. Pérez, M. Oronoz (2018) Word embeddings for negation detection in health records written in Spanish. Soft Computing (2018). pp-1.7, 2018. IF: 2.367* **JCR-Q2**

C.2. Congress

[1] Rachel Bawden, Giorgio Maria Di Nunzio, Cristian Grozea, Inigo Jauregi Unanue, Antonio Jimeno Yepes, Nancy Mah, David Martinez, Aurélie Névéol, Mariana Neves, Maite Oronoz, Olatz Perez-de-Viñaspre, Massimo Piccardi, Roland Roller, Amy Siu, Philippe Thomas, Federica Vezzani, Maika Vicente Navarro, Dina Wiemann and Lana Yeganova (2020). *Findings of the WMT 2020 Biomedical Translation Shared Task: Basque, Italian and Russian as New Additional Languages*. Fifth Conference on Machine Translation (WMT20). Shared Task: Biomedical Translation Task

[2] Xabier Soto, Olatz Perez-de-Viñaspre, Gorka Labaka, Maite Oronoz (2020). Ixamed's submission description for WMT20 Biomedical shared task: benefits and limitations of using terminologies for domain adaptation. Proceedings of the Fifth Conference on Machine Translation, pp: 873--878.

C.3. Research projects

Participation within the research team in the following projects:

[1] **Title**: ANTIDOTE. https://univ-cotedazur.eu/antidote. **Funding Entity**: CHIST-ERA - INT-Acciones de Programación Conjunta Internacional (MINECO) 2020 (PCI2020-120717-2). **Length**: 01/04/2021 - 31/03/2024. **Amount**: 957.478 €. **PI**: Rodrigo Agerri

[2] **Title**: DOTT-HEALTH: Development Of Text-based Technology to support diagnosis, prevention and HEALTH institutions management (PID2019-106942RB-C31). **Funding Entity:** Ministerio de Ciencia, Innovación y Universidades. **Length**: 01/06/2020 - 31/05/2023. **Amount**: 160.000 €. **PIs**: Koldo Gojenola / Arantza Casillas

[3] **Title:** BigKnowledge for Text Mining. **Funding Entity**: Fundación BBVA. **Length:** 30/04/2019- 30/04/2021. **Amount:** 71500 €. **PI:** Eneko Agirre

[5] **Title**: DL4NLP: Deep Learning aplicado al Procesamiento del Lenguaje Natural como apoyo a los ámbitos del RIS3 (KK-2019/00045). **Funding Entity:** Gobierno Vasco/Eusko Jaurlaritza. **Length**: 01/01/2019- 31/12/2019. **Amount:** 43460 €. **PI:** Aitor Soroa

[6] **Title**: TADEEP: Traducción automática en profundidad (TIN2015-70214-P). **Funding Entity:** Ministerio de Economía y Competitividad. **Length:** 01/01/2016- 31/12/2018   . **Amount:** 117128 €. **PI:** Kepa Sarasola

[7] **Title:** OSAKU: OSAsun txostenen KUdeaketarako laguntza sistema/Sistema de ayuda para la gestión de informes médicos. **Funding Entity:** Basque Government. SAIOTEK. **Length**: 01/01/2011 - 31/12/2012. **Amount:** 29559 €. **PI: Maite Oronoz**

C.4. Contracts, technological or transfer merits

[1] **Title:** Itzulbide: Development and implementation of a neural machine translation system to translate electronic health records between Spanish and Basque http://info.osidonostialdea.eus/es/itzulbide-el-proyecto-de-traductor-automatico-neuronal-de-osakidetza-en-la-osi/. **Company/Health organization**: G/211/20/1/0199/o001/0000/012019 (public bid: Osakidetza) Osakidetza (Basque Public Health System). **Length:** 05/06/2019-05/06/2021. **Amount:** 191.469,58 €. **PI: Maite Oronoz**

[2] **Title**: Protágoras: Desarrollo de algoritmos de procesamiento de lenguaje natural para el desarrollo de un motor cognitivo. **Company:** INDRA S.A. **Length:** 15/12/2017- 15/12/2018. **Amount:** 90.604,19 €. **PI: Maite Oronoz**

[3] **Title:** Erreferentziazko terminologia klinikoa euskarara itzultzea/Translation of Clinical reference terminology into Basque. Company/Health organization: Osakidetza. **Length:** 03/11/2016- 02/05/2017. **Amount:** 14.534,04 €. **PI: Maite Oronoz**

[4] **Title**: Servicio para la desambiguación de agentes del sector I.D.I.. **Company:** Red.es/Elhuyar. **Length:** 26/04/2018- 25/01/2019. **Amount:** 8.748,48 €. **PI: Maite Oronoz**

[5] **Intellectual property registration:** *Sistema automático para la detección automática de entidades del dominio médico en español*; **Authors (by signature order) :** A. Atucha, A. Casillas, K. Gojenola, **M. Oronoz**, A. Pérez, O. Pérez de Viñaspre, S. Santiso. **Nº:** SS-23-19