

### Parte A. DATOS PERSONALES

	Fecha del (	CVA		12/07/2024
e				
		Edad	//1	

Nombre y apellidos   Gorka Labaka Intxauspe					
DNI/NIE/pasaporte	34100230-Q		Edad	41	
Núm. identificación del	Researcher ID	G-8236-2011			
Num. identificación del	iiivesiigauoi	Código Orcid	0000-00	003-4611-2502	

A.1. Situación profesional actual

Organismo	Universidad del País Vasco (UPV/EHU)				
Dpto./Centro	Lenguajes y sistemas informáticos/Escuela de Ingeniería de Bilbao				
Dirección	Paseo Rafael Moreno "Pitxitxi", Bilbao				
Teléfono	946014486	36 correo electrónico gorka.labaka@ehu.eus			
Categoría profesional	Profesor Titula	Profesor Titular			01/07/2024
Espec. cód. UNESCO	570109; 120304				
Palabras clave Procesamiento del Lenguaje Natural, Traducción automática					

A.2. Formación académica (título, institución, fecha)

Licenciatura/Grado/Doctorado	Universidad	Año
Ingeniería Informática	Universidad del País Vasco (UPV/EHU)	2004
Doctorado en Informática	Universidad del País Vasco (UPV/EHU)	2010

## A.3. Indicadores generales de calidad de la producción científica

- Citas totales: Google Scholar: 4956 (3995 desde 2019)
- Promedio de citas/año últimos 5 años: Google scholar: 799
- Índice h: Google scholar: 24 (19 desde 2019)

#### Parte B. RESUMEN LIBRE DEL CURRÍCULUM

El Dr. Gorka Labaka es Profesor Titular del Departamento de Lenguajes y Sistemas Informáticos de la Escuela de Ingeniería de Bilbao de la UPV/EHU y miembro del grupo de investigación IXA. Se incorporó al grupo IXA en 2001, cuando aún estaba estudiando la carrera, y obtuvo su doctorado en Informática en la UPV/EHU en 2010. Sus principales intereses de investigación son el Procesamiento del Lenguaje Natural (PLN) y la Traducción Automática.

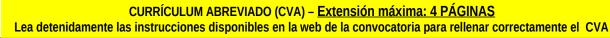
Actualmente participa en los proyectos ILENIA (2022/TL22/00215335), DeepR3 (TED2021-130295B-C31) y TRAIN (PID2021-123988OB-C31). Ha publicado más de 22 artículos en revistas y congresos, ha participado en 23 proyectos regionales, nacionales y europeos y ha registrado tres propiedades intelectuales. Tiene experiencia en docencia tanto a nivel de grado como de máster, dentro del máster LAP-EMLCT, y ha tutorizado varios proyectos fin de carrera y tesis de máster. Junto con miembros del grupo IXA, ha diseñado y desarrollado varias herramientas de PNL y traducción automática.

Recientemente, ha comenzado a trabajar en traducción de lengua de signos a través de su participación en el proyecto SignON de la Comisión Europea (2020-2023). Siguiendo esta línea de investigación, actualmente es co-IP del subproyecto coordinador de TRAIN. Este proyecto está dedicado a la traducción de lenguas con muy pocos recursos e incluye la traducción de la Lengua de Signos Española (LSE) al español.

# Parte C. MÉRITOS MÁS RELEVANTES (ordenados por tipología) C.1. Publicaciones

Adrian Núñez-Marcos, Olatz Perez-de-Viñaspre and Gorka Labaka. 2023. *A survey on Sign Language machine translation*. Expert Systems with Applications, Volume 213, Part B, 2023, 118993, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118993 (revista JCR Q1 - 5 citas en Google Scholar)

Aitor Ormazabal, Mikel Artetxe, Aitor Soroa, Gorka Labaka, and Eneko Agirre. 2022. Principled Paraphrase Generation with Parallel Corpora. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1621–1638, Dublin, Ireland. (conferencia con clasificación A\* en CORE; GGS Class 1)





Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label Verbalization and Entailment for Effective Zero and Few-Shot Relation Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 1199–1212, Punta Cana, Dominican Republic (conferencia con clasificación A en CORE; GGS Class 1; 20 citas en Google Scholar)

Aitor Ormazabal, Mikel Artetxe, Aitor Soroa, Gorka Labaka, and Eneko Agirre. 2021. *Beyond Offline Mapping: Learning Cross-lingual Word Embeddings through Context Anchoring.* In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 6479–6489 (conferencia con clasificación A\* en CORE; GGS Class 1; 6 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, Noe Casas, Eneko Agirre (2020) *Do all roads lead to Rome? Understanding the role of initialization in iterative back-translation.* Knowledge-Based Systems, Volume 206, 2020, 106401, ISSN 0950-7051,

https://doi.org/10.1016/j.knosys.2020.106401. (revista JCR Q1 - 5 citas en Google Scholar)

Uxoa Inurrieta, Itziar Aduriz, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola (2020) Learning about phraseology from corpora: A linguistically motivated approach for Multiword Expression identification. PLoS ONE 15(8): e0237767.

https://doi.org/10.1371/journal.pone.0237767 (revista JCR Q1 - 5 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2019. *An Effective Approach to Unsupervised Machine Translation*. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 194–203, Florence, Italy. (conferencia con clasificación A\* en CORE; GGS Class 1; 131 citas en Google Scholar)

Xabier Soto, Olatz Perez de Viñaspre, Gorka Labaka, Maite Oronoz. 2019. *Neural Machine Translation of clinical texts between long distance languages*. Journal of the American Medical Informatics Association, Volume 26, Issue 12, December 2019, Pages 1478–1487, https://doi.org/10.1093/jamia/ocz110 (11 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. *Unsupervised Statistical Machine Translation*. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3632–3642, Brussels, Belgium. (conferencia con clasificación A en CORE; GGS Class 1; 204 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, Iñigo Lopez-Gazpio, and Eneko Agirre. 2018. *Uncovering Divergent Linguistic Information in Word Embeddings with Lessons for Intrinsic and Extrinsic Evaluation.* In Proceedings of the 22nd Conference on Computational Natural Language Learning, pages 282–291, Brussels, Belgium. (conferencia con clasificación A en CORE; GGS Class 2; 41 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. *A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings.* In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 789–798, Melbourne, Australia. (conferencia con clasificación A\* en CORE; GGS Class 1; 504 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, Eneko Agirre. 2018. *Generalizing and Improving Bilingual Word Embedding Mappings with a Multi-Step Framework of Linear Transformations*. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18) pages 5012-5019. (conferencia con clasificación A\* en CORE; GGS Class 1; 202 citas en Google Scholar)

Mikel Artetxe, Gorka Labaka, Eneko Agirre, Kyunghyun Cho. 2018. *Unsupervised Neural Machine Translation*. 6th International Conference on Learning Representations, ICLR 2018. Vancouver, Canada. (conferencia con clasificación A\* en CORE; GGS Class 1; 705 citas en Google Scholar)



Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. *Learning bilingual word embeddings with (almost) no bilingual data*. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 451–462, Vancouver, Canada. (conferencia con clasificación A\* en CORE; GGS Class 1; 496 citas en Google Scholar)

C.2. Proyectos

Título: LENIA: Impulso de las lenguas en Inteligencia Artificial

Referencia: 2022/TL22/00215335

Entidad financiadora y convocatoria: Ministerio de Transformación Digital

Investigador Principal: Marta Villegas

**Duración, desde**: 01/01/2023 hasta: 31/12/2025

Tipo de participación: Investigador

**Título**: TRAIN: Extremely low-resourced machine translation

**Referencia**: PID2021-123988OB-C31

Entidad financiadora y convocatoria: Ministerio de Ciencia e Innovación (Generación De

Conocimiento)

**Investigador Principal**: Gorka Labaka y Eneko Agirre **Duración, desde**: 01/09/2022 **hasta**: 31/08/2025

Tipo de participación: Investigador Principal

Título: Ixa Taldea: tecnología del lenguaje al servicio de la sociedad

Referencia: IT1570-22

Entidad financiadora: Gobierno Vasco

Investigador Principal: Eneko Agirre Bengoa

**Duración, desde**: 01/01/2022 hasta: 31/12/2026

Tipo de participación: Investigador

Título: SignOn: Sign Language Translation Mobile Application and Open Communications

Framework

Referencia: DLV-101017255

Entidad financiadora y convocatoria: Comisión Europea

**Investigador Principal**: Andrew Way

**Duración, desde**: 01/01/2021 **hasta**: 31/12/2023 **Tipo de participación**: Investigador Principal de la UPV/EHU

**Título**: MT4All: Unsupervised MT for Low-resourced language pairs

Referencia: INEA/CEF/ICT/A2019/1926910

Entidad financiadora y convocatoria: Comisión Europea (programa CEF)

Investigador Principal: Gorka Labaka

**Duración, desde**: 01/01/2020 hasta: 31/12/2021

Tipo de participación: Investigador Principal

Título MODENA: Modelado neuronal avanzado para traducción de alta calidad.

Referencia: KK-2018/00087

Entidad financiadora y convocatoria: Dep. Industria del Gob. Vasco (Conv. ELKARTEK)

Investigador Principal: Kepa Sarasola

**Duración, desde**: 01/01/2018 **hasta**: 31/12/2019

**Tipo de participación**: Investigador

Título: TADEEP: Traducción automática en profundidad

Referencia: TIN2015-70214-P

Entidad financiadora y convocatoria: MINECO-FEDER

Investigador Principal: Kepa Sarasola

**Duración, desde**: 01/01/2016 hasta: 31/12/2018

Tipo de participación: Investigador



## CURRÍCULUM ABREVIADO (CVA) - Extensión máxima: 4 PÁGINAS

Lea detenidamente las instrucciones disponibles en la web de la convocatoria para rellenar correctamente el CVA

Título: MODELA: Modelado Estadístico y Deep Learning para Traducción Automática de

Alta calidad

**Referencia**: KK-2016/00082

Entidad financiadora y convocatoria: Dep. Industria del Gob. Vasco (Conv. ELKARTEK)

Investigador Principal: Kepa Sarasola

**Duración, desde**: 01/01/2016 hasta: 31/12/2017

Tipo de participación: Investigador

Título: QTLeap: Quality Translation by Deep Language Engineering Approaches

Referencia: FP7-ICT-2013.4.1-610516

Entidad financiadora y convocatoria: European Commission, FP7

Investigador Principal: Antonio Branco (FCUL)

Duración, desde: 1/11/2013 hasta: 31/10/2016

Tipo de participación: Investigador

**Título**: TACARDI: Traduccion automatica en contexto y aumentada con recursos dinamicos

Referencia: TIN2012-38523-C02-01

Entidad financiadora y convocatoria: Ministerio de Educación y Ciencia

Investigador Principal: Kepa Sarasola (UPV/EHU)

Duración, desde: 01/01/2013 hasta: 31/12/2015

Tipo de participación: Investigador

# C.3. Contratos, méritos tecnológicos o de transferencia

Título: Matxin, traducción automática en código abierto para euskera-castellano: Consultoría

desarrollo motor estadístico.

Empresas: ELEKA INGENIARITZA LINGUISTIKOA Investigador/a principal: Iñaki Alegria (UPV/EHU)

Tipo contrato: Proyecto de investigación Inicio: 01/07/2009 Fin: 31/12/2009

Financiación total: 14.974,12

#### C.4. Patentes

Inventores: Iñaki Alegria, Aitziber Atutxa, Arantza Casillas, Gorka Labaka, Mikel Lersundi,

Aingeru Mayor, Kepa Sarasola

Título: Matxin: sistema de traducción automática de castellano a euskara basado en reglas

de transferencia

Número de solicitud: SS-409-08

País de prioridad: España Fecha de prioridad: 22/12/2008 Entidad titular: UPV/EHU

Inventores: Iñaki Alegria, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola

Título: EUSMT: Traducción automática estadística para el euskara

Número de solicitud: SS-19-12

País de prioridad: España Fecha de prioridad: 18/01/2012 Entidad titular: UPV/EHU

#### C.5. Organización de congresos

1. Organización del workshop MATMT-2008: "Mixing Approaches to Machine Translation"

2. Organización del workshop Workshop LIHMT-2011: "Using Linguistic Information for Hybrid Machine Translation"

3. Revisor en las principales conferencias de procesamiento del language natural (ACL, EMNLP, LREC) e inteligencia artificial (AAAI, ECAI).