

Part A. PERSONAL INFORMATION		CV date		03/02/2019	
First and Family name	Gorka Labaka Intxauspe				
Social Security, Passport, ID number	34100230-Q	Age	38		
Researcher numbers	Researcher ID	G-8236-2011			
	Orcid code	0000-0003-4611-2502			

A.1. Current position

Name of University/Institution	University of the Basque Country				
Department	Computer Languages and Systems / Engineering School				
Address and Country	Paseo Rafael Moreno "Pitxitxi", Bilbao, Spain				
Phone number	946014486	E-mail	Gorka.labaka@ehu.eus		
Current position	Profesor adjunto (Assistant Professor)	From	04/05/2018		
Espec. cód. UNESCO	570109; 120304				
Palabras clave	Natural Language Processing, Machine Translation				

A.2. Education

PhD	University	Year
B.Sc Computer Science	University of the Basque Country	2004
PhD in Computer Science	University of the Basque Country	2010

A.3. JCR articles, h Index, thesis supervised...

- Citations: **Google Scholar**: 695 (488 since 2014)
- Average number of citations/year last 5 years: **Google scholar**: 97
- h-index: **Google scholar**: 19 (11 since 2014)

Part B. CV SUMMARY (max. 3500 characters, including spaces)

Dr. Gorka Labaka is an assistant lecturer at the Engineering School of the UPV/EHU since 2018. He received his PhD in Computer Science from the University of the Basque Country in 2010 and has been Post-doctoral researcher at the Faculty of Computer Science since then. He has published over 25 international refereed articles and conference papers in Machine Translation and cross-lingual word embeddings, focusing on highly inflected languages and hybridization techniques. He has participated in two European projects (NewsReader and QTLeap) and several national (OpenMT, OpenMT-2, TACARDI, Tadeep, etc.) and regional (Quales, Modela, Ber2Tek, Berbatek, etc.) research projects. Since 2012, he has been teaching Machine Translation courses at the HAP/LAP and EMLCT master degrees.

Recently, Dr Labaka has worked on crosslingual mapping of word embeddings. Working both to improve the quality of embeddings (Artetxe, 2016; Artetxe, 2018a) and to reduce the amount of bilingual information needed for mapping (Artetxe, 2017), until the need for a bilingual dictionary is completely eliminated (Artetxe, 2018c). Using these embeddings to generate for the first time an unsupervised machine translation system without a parallel training corpus (Artetxe, 2018b).

Part C. RELEVANT MERITS

C.1. Publications (including books)

Mikel Artetxe, Gorka Labaka, Eneko Agirre. 2018. *Unsupervised Statistical Machine Translation*. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018) (CORE A conference)

Mikel Artetxe, Gorka Labaka, Iñigo Lopez-Gazpio, Eneko Agirre. 2018. *Uncovering Divergent Linguistic Information in Word Embeddings with Lessons for Intrinsic and Extrinsic Evaluation*. Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL 2018) (CORE A conference)

Mikel Artetxe, Gorka Labaka, Eneko Agirre. 2018. *A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings*. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 789–798. Melbourne, Australia (CORE A* conference)

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. (CORE A* conference)

Mikel Artetxe, Gorka Labaka, Eneko Agirre. 2017. *Learning bilingual word embeddings with (almost) no bilingual data*. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), pp. 451-462, Vancouver, Canada. | ISBN 978-1-945626-75-3 (CORE A* conference)

Uxo Iñurrieta, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola, Itziar Aduriz, John Carroll. 2016. *Using Linguistic Data for Verb-Noun Combination Identification*. Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016), pp. 857-867, Osaka, Japan. | ISBN: 978-4-87974-702-0 (CORE A conference)

Mikel Artetxe, Gorka Labaka, Eneko Agirre. 2016. *Learning principled bilingual mappings of word embeddings while preserving monolingual invariance*. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), pp. 2289-2294, Austin, Texas. | ISBN: 978-1-945626-25-8 (CORE A conference)

Eneko Agirre, Iñaki Alegria, Nora Aranberri, Mikel Artetxe, Ander Barrena, António Branco, Arantza Diaz de Ilarraza, Koldo Gojenola, Gorka Labaka, Arantxa Otegi and Kepa Sarasola. 2015 *Lexical semantics, Basque and Spanish in QLeap: Quality Translation by Deep Language Engineering Approaches* Procesamiento del Lenguaje natural, Volumen 55, pp. 169-172. ISSN: 1135-5948

Mikel Artetxe, Gorka Labaka, Kepa Sarasola. 2015 *Building hybrid machine translation systems by using an EBMT preprocessor to create partial translations* Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT2015), pp. 11-18, Antalya, Turkey (CORE B conference)

Gorka Labaka, Cristina España-Bonet, Lluís Màrquez, Kepa Sarasola 2014 *A hybrid machine translation architecture guided by syntax* Machine Translation Journal. Volume 28 Issue 2 (2014) pp. 91-125. DOI: 10.1007/s10590-014-9153-0

Iñaki Alegria, Unai Cabezón, Unai Fernández de Betoño, Gorka Labaka, Aingeru Mayor, Kepa Sarasola and Arkaitz Zubiaga 2013 *Reciprocal Enrichment between Basque Wikipedia and Machine Translators*. 'The People's Web Meets NLP: Collaboratively Constructed Language Resources', book edited by Iryna Gurevych and Jungi Kim, Springer, Book series 'Theory and Applications of Natural Language Processing'. April 30, 2013 | ISBN-10: 3642350844 | ISBN-13: 978-3642350849.

Aingeru Mayor, Iñaki Alegria, Arantza Díaz de Ilarraza, Gorka Labaka, Mikel Lersundi, Kepa Sarasola 2011 *Matxin, an open-source rule-based machine translation system for Basque*. Machine Translation Journal: Volume 25, Issue 1 (2011), pp. 53-82. ISSN: 0922-6567. DOI: 10.1007/s10590-011-9092-y

Alberto Barrón-Cedeño, Paolo Rosso, Eneko Agirre, Gorka Labaka 2010 *Plagiarism Detection across Distant Language Pairs* Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pp. 37-45, Beijing, China. (CORE A conference)

C.2. Research projects and grants

Title: LINGUATEC: Development of cross-border cooperation and knowledge transfer in language technologies.

Reference: EFA227/16

Funding entity and call: European Commission, POCTEFA

Principal Investigator: Kepa Sarasola

Duration, from: 01/01/2018

To: 31/12/2020

Amount of the grant: 797875 euro

Type of participation: Resercher

Title: UnsupNMT: Traducción Automática Neuronal no Supervisada: un nuevo paradigma basado solo en textos monolingües.

Reference: TIN2017-91692-EXP

Funding entity and call: MINECO (Conv. Explora)

Principal Investigator: Eneko Agirre

Duration, from: 01/07/2018

To: 30/06/2020

Amount of the grant: 48.400 euro

Type of participation: Resercher

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

Reference: KK-2018/00087

Funding entity and call: Dep. Industria del Gob. Vasco (Conv. ELKARTEK)

Principal Investigator: Kepa Sarasola

Duration, from: 01/01/2018

To: 31/12/2019

Amount of the grant: 148.640 euro

Type of participation: Resercher

Title: TADEEP: Traducción automática en profundidad

Reference: TIN2015-70214-P

Funding entity and call: MINECO-FEDER

Principal Investigator: Kepa Sarasola

Duration, from: 01/01/2016

To: 31/12/2018

Amount of the grant: 96.800 euro

Type of participation: Resercher

Title: QUALES: Aprendizaje Automático mediante Supervisión Modulable para Estimación Automática de Calidad de Traducción

Reference: KK-2017/00094

Funding entity and call: Dep. Industria del Gob. Vasco (Conv. ELKARTEK)

Principal Investigator: Kepa Sarasola

Duration, from: 01/07/2017

To: 31/12/2018

Amount of the grant: 90.000 euro

Type of participation: Resercher

Title: MODELA: Modelado Estadístico y Deep Learning para Traducción Automática de Alta Calidad

Reference: KK-2016/00082

Funding entity and call: Dep. Industria del Gob. Vasco (Conv. ELKARTEK)

Principal Investigator: Kepa Sarasola

Duration, from: 01/01/2016

To: 31/12/2017

Amount of the grant: 99.838,02 euro

Type of participation: Resercher

Title: QTLeap: Quality Translation by Deep Language Engineering Approaches

Reference: FP7-ICT-2013.4.1-610516

Funding entity and call: European Commission, FP7

Principal Investigator: Antonio Branco (FCUL)

Duration, from: 1/11/2013

To: 31/10/2016

Amount of the grant: 3.003.540 euro

Type of participation: Resercher

Title: TACARDI: Traducción automática en contexto y aumentada con recursos dinámicos de internet.

Reference: TIN2012-38523-C02-01

Funding entity and call: Ministerio de Educación y Ciencia

Principal Investigator: Kepa Sarasola (UPV/EHU)

Duration, from: 01/01/2013

To: 31/12/2015

Amount of the grant: 193.260 euro

Type of participation: Investigador

Title: NewsReader: Building structured event indexes of large volumes of financial and economic data for decision making

Reference: FP7-ICT-2011-8-316404

Funding entity and call: European Commission, FP7

Principal Investigator: Piek Vossen (VUA)

Duration, from: 01/01/2013

To: 31/12/2015

Amount of the grant: 2.800.000 euro

Type of participation: Investigador

Title: Ber2Tek - Tecnologías de la Lengua, de Voz y Multimedia al Servicio de la Industria de las Lenguas

Reference: IE12-333

Funding entity and call: Gobierno Vasco, ETORTEK

Principal Investigator: Igor Leturia (Elhuyar Fundazioa)

Duration, from: 01/01/2012

To: 31/12/2014

Amount of the grant: 1.028.479 euro

Type of participation: Investigador

Title: IXA: Grupo Consolidado de Investigación (Categoría A)

Reference: IT344-10

Funding entity and call: Gobierno Vasco.

Principal Investigator: Arantza Diaz de Ilarraza (UPV/EHU)

Duration, from: 01/01/2010

To: 31/12/2015

Amount of the grant: 766.000 euro

Type of participation: Investigador

C.3. Contracts

Title: Matxin, traducción automática en código abierto para euskera-castellano: Consultoría desarrollo motor estadístico.

Companies: ELEKA INGENIARITZA LINGUISTIKOA

Principal Investigator: Iñaki Alegria (UPV/EHU)

Contract Type: Research Project **Begin:** 01/07/2009 **End:** 31/12/2009

Amount: 14.974,12

C.4. Patents

Inventors: Iñaki Alegria, Aitziber Atutxa, Arantza Casillas, Gorka Labaka, Mikel Lersundi, Aingeru Mayor, Kepa Sarasola

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

Application number: SS-409-08

Priority country: Spain

Priority date: 22/12/2008

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

Title: EUSMT: Traducción automática estadística para el euskara

Application number: SS-19-12

Priority country: Spain

Priority date: 18/01/2012