

# Research

Thesis

Publications

Projects and Knowledge Transference Contracts

## Thesis

1. *Application of singing synthesis techniques to bertsoaritza*  
**Xabier Sarasola Aramendia**  
Supervisor: Eva Navas Cordon
2. *Medidas de distância entre línguas baseadas em corpus*  
**José Ramom Pichel Campos**  
Supervisors: Iñaki Alegria Loinaz and Pablo Gamallo Otero
3. *Itzulpen automatiko gainbegiratu gabea*  
**Mikel Artetxe Zurutuza**  
Supervisors: Eneko Agirre Bengoa and Gorka Labaka Intxauspe

## Publications

### JOURNALS

1. Abel Camacho, Mikel Iruskietta, A. Latatu, P. Lonbide  
*UEUren Online ikaskuntzarako eredu pedagogikoaren sorrera eta garapena*  
UZTARO 118, 5-38

2. Itziar Aduriz, Jose Mari Arriola  
*Testu-corpusen informazio morfosintaktikoaren etiketatze automatikoa hizkuntz ezagutzan oinarrituz: zenbait arazo, hainbat erronka*  
Fontes Linguae Vasconum 50 urte. Ekarpen berriak euskararen ikerketari / Nuevas aportaciones al estudio de la lengua vasca
  
3. Mikel Iruskietza, Amaia Arroyo-Sagasta, Abel Camacho, Montse Maritxalar  
*Teknologia, testuinguru digitala eta konpetentzia digitalak hezkuntzan*  
Euskonews 748. ISSN: 1139-3629
  
4. Ainara Estarrona, Izaskun Etxeberria, Ricardo Etxepare, Manuel Padilla-Moyano, Ander Soraluze  
*Sintaktikoki etiketatutako euskarazko corpus historikoa eraikitzen*  
Fontes Linguae Vasconum 50 urte. Ekarpen berriak euskararen ikerketari. Nuevas aportaciones al estudio de la lengua vasca
  
5. Rebecka Weegar, Alicia Pérez, Arantza Casillas, Maite Oronoz  
*Recent advances in Swedish and Spanish medical entity recognition in clinical texts using deep neural approaches*  
BMC Medical Informatics and Decision Making
  
6. Garazi Hormaetxe, Mikel Iruskietza  
*Parekoen behaketarekin komunikazio-gaitasuna ebaluatzen: zer dute nahiago ikasleek, errubrika tradizionala ala bideo-behaketa*  
e-Hizpide 96
  
7. Sara Santiso, Alicia Pérez, Arantza Casillas, Maite Oronoz  
*Neural negated entity recognition in Spanish electronic health records*  
Journal of Biomedical Informatics
  
8. Itziar Aduriz, Jose Mari Arriola, Xabier Artola, Zuhaitz Beloki, Nerea Ezeiza, Koldo Gojenola  
*Morfeus+: Word Parsing in Basque beyond Morphological Segmentation*  
WORD STRUCTURE 13.3, 283-315

9. Pablo Gamallo José Ramom Pichel, Iñaki Alegria  
*Measuring Language Distance of Isolated European Languages*  
MDPI Information 2020, 11(4), 181
  
10. Uxoa Inurrieta, Itziar Aduriz, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola  
*Learning about phraseology from corpora: A linguistically motivated approach for Multiword Expression identification*  
PLoS ONE 15(8)
  
11. Uxoa Iñurrieta  
*Identification and translation of verb+noun multiword expressions: a Spanish-Basque study*  
Procesamiento del Lenguaje Natural, 64, pp. 123-126.
  
12. Jose Mari Arriola, Josu Goikoetxea, Mikel Iruskieta  
*Hizkuntza-teknologiak hizkuntzen ikas-irakaskuntzan: zenbat aukera, hainbat erronka*  
ehizpide 95: 1-21
  
13. Alberto Blanco, Alicia Pérez, Arantza Casillas  
*Extreme multi-label ICD classification: sensitivity to hospital service and time*  
IEEE Access, Volume 8, 183534-183545
  
14. Alberto Blanco, Alicia Pérez, Arantza Casillas, Daniel Cobos  
*Extracting Cause of Death from Verbal Autopsy with Deep Learning interpretable methods*  
IEEE Journal of Biomedical and Health Informatics
  
15. Begoña Altuna, María Jesús Aranzabe, Arantza Díaz de Ilarraza  
*EusTimeML: A mark-up language for temporal information in Basque*  
Research in Corpus Linguistics 8: 86-104. ISSN 2243-4712

16. Mikel Artetxe, Gorka Labaka, Noe Casas, Eneko Agirre  
*Do all roads lead to Rome? Understanding the role of initialization in iterative back-translation*  
Knowledge-Based Systems, Volume 206
17. José Ramom Pichel, Pablo Gamallo, Marco Neves, Iñaki Alegria  
*Distância diacrónica automática entre variantes diatópicas do português e do espanhol*  
Linguamática, Vol. 12 N. 1, 117–126 ISSN: 1647–0818
18. Eneko Agirre  
*Cross-Lingual Word Embeddings (Book Review)*  
Computational Linguistics 46 (1), 245-248
19. Naiara Pérez, Pablo Accuosto, Àlex Bravo, Montse Cuadros, Eva Martínez-Garcia, Horacio Saggion, German Rigau  
*Cross-lingual semantic annotation of Biomedical literature: experiments in Spanish and English*  
Bioinformatics, 36, 6, 1872-1880, ISSN 1367-1880
20. Gorka Azkune, Aitor Almeida, Eneko Agirre  
*Cross-environment activity recognition using word embeddings for sensor and activity representation*  
Neurocomputing
21. Arantxa Otegi, Aitor Soroa, Eneko Agirre, Jon Ander Campos  
*Cómo gestionar la sobrecarga de información científica sobre COVID-19*  
The Conversation. ISSN 2201-5639
22. Nora Aranberri  
*Can translationese features help users select an MT system for post-editing?*  
Revista Procesamiento del Lenguaje Natural, 64, 93-100

23. Alberto Blanco, Olatz Perez de Viñaspre, Alicia Pérez, Arantza Casillas  
*Boosting ICD multi-label classification of health records with contextual embeddings and label-granularity*  
Computer Methods and Programs in Biomedicine, Volume 188, 105264
24. Kepa Bengoetxea, Itziar Gonzalez-Dios, Amaia Aguirregoitia  
*AzterTest: Open source linguistic and stylistic analysis tool*  
Procesamiento del Lenguaje Natural, 64, 61-68
25. Begoña Altuna  
*Análisis de estructuras temporales en euskera y creación de un corpus*  
Procesamiento del Lenguaje Natural 64
26. Sara Santiso  
*Adverse Drug Reaction extraction on Electronic Health Records written in Spanish*  
Procesamiento del Lenguaje Natural
27. Jose R. Pichel, Pablo Gamallo, Iñaki Alegria, Marco Neves  
*A Methodology to Measure the Diachronic Language Distance between Three Languages Based on Perplexity*  
Journal of Quantitative Linguistics.
28. Juan J. Lastra-Díaz, Josu Goikoetxea, Mohamed Ali Hadj Taieb, Ana Garcia-Serrano, Mohamed Ben Aouicha, Eneko Agirre, David Sánchez  
*A large reproducible benchmark of ontology-based methods and word embeddings for word similarity*  
Information Systems
29. Amaia Aguirregoitia Martinez, Kepa Bengoetxea Kortazar, Itziar Gonzalez-Dios  
*Are CLIL texts too complicated? A computational analysis of their linguistic characteristics*  
Journal of Immersion and Content-Based Language Education (Available online)

## PROCEEDINGS

1. Nora Aranberri  
*With or without you? Effects of using machine translation to write flash fiction in the foreign language*  
Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, p. 165–174, Lisboa, Portugal, November 2020.
2. Adrián Nuñez-Marcos, Gorka Azkune, Eneko Agirre, Diego López-de-Ipiña, Ignacio Arganda-Carreras  
*Using External Knowledge to Improve Zero-shot Action Recognition in Egocentric Videos*  
International Conference on Image Analysis and Recognition (ICIAR)
3. Ivana Kvapilíková, Mikel Artetxe, Gorka Labaka, Eneko Agirre, Ondřej Bojar  
*Unsupervised Multilingual Sentence Embeddings for Parallel Corpus Mining*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Pages 255-262
4. Mikel Artetxe, Gorka Labaka, Eneko Agirre  
*Translation Artifacts in Cross-lingual Transfer Learning*  
Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). (Pages 7674–7684).
5. Javier Álvez, Itziar Gonzalez-Dios, German Rigau  
*Towards Word Sense Disambiguation by Reasoning*  
Vampire 2018 and Vampire 2019. The 5th and 6th Vampire Workshops. EPiC Series in Computing. Pages 19-29. ISSN: 2398-7340
6. Bernardo Magnini, Begoña Altuna, Alberto Lavelli, Manuela Speranza, Roberto Zanolli  
*The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases*  
In Johanna Monti, Felice Dell'Orletta and Fabio Tamburini (eds.), Proceedings of the Seventh Italian Conference on Computational Linguistics. Associazione Italiana di Linguistica Computazionale. Bologna, Italy, 2020.

7. Jan Deriu, Don Tuggener, Pius von Däniken, Jon Ander Campos, Alvaro Rodrigo, Thiziri Belkacem, Aitor Soroa, Eneko Agirre, Mark Cieliebak  
*Spot The Bot: A Robust and Efficient Framework for the Evaluation of Conversational Dialogue Systems*  
Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). (Pages 3971–3984). Honorable Mention Paper (Top 1%).
  
8. Gorka Urbizu, Ander Soraluze, Olatz Arregi  
*Sequence to Sequence Coreference Resolution*  
Proceedings of the 3rd Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2020), pages 39–46, Barcelona, Spain (online), December 12, 2020.
  
9. Xabier Soto, Dimitar Shterionov, Alberto Poncelas, Andy Way  
*Selecting Backtranslated Data from Multiple Sources for Improved Neural Machine Translation*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp: 3898–3908.
  
10. C. Pradel, D. Sileo, A. Rodrigo, A. Peñas, E. Agirre  
*Question Answering when Knowledge Bases are Incomplete?*  
Proceedings of Conference and Labs of the Evaluation Forum.
  
11. Rodrigo Agerri, German Rigau  
*Projecting Heterogeneous Annotations for Named Entity Recognition*  
In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020). Winner of the CAPITEL@IberLEF task on Spanish NER.
  
12. Thierry Declerck, Itziar Gonzalez-Dios, German Rigau (editors)  
*Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMWN-2020)*  
European Language Resources Association (ELRA), Paris.
  
13. Mikel Artetxe, Sebastian Ruder, Dani Yogatama  
*On the cross-lingual transferability of monolingual representations*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

14. Elena Zotova, Rodrigo Agerri, Manuel Nuñez and German Rigau  
*Multilingual Stance Detection in Tweets: The Catalonia Independence Corpus*  
Resources and Evaluation Conference (LREC 2020)
  
15. Iker de la Iglesia, Mikel Martinez-Puente, Alexander Platas, Iria San Miguel, Aitziber Atutxa, Koldo Gojenola  
*MEDIA team at the CLEF-2020 Multilingual Information Extraction Task*  
Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum Thessaloniki, Greece, September 22-25, 2020.
  
16. Andrea Horbach, Itziar Aldabe, Marie Bexte, Oier Lopez de Lacalle and Montse Maritxalar  
*Linguistic Appropriateness and Pedagogic Usefulness of Reading Comprehension Questions*  
Proceeding of 12th Edition of its Language Resources and Evaluation Conference (LREC2020).  
Marseille, France
  
17. Itziar Aldabe, Josu Aztiria, Francho Beltrán, Myriam Bras, Klara Ceberio, Itziar Cortes, Jean-Baptiste Coyos, Benaset Dazeas, Louise Esher, Gorka Labaka, Igor Leturia, Kepa Sarasola, Aure Séguier, Jean Sibille  
*LINGUATEC: Development of cross-border cooperation and knowledge transfer in language technologies*  
Workshop INTELE : INfraestructura de TECnologías del LEnguaje CLARIN DARIAH-EU.
  
18. Kepa Sarasola, Iñaki Alegria, Olatz Perez de Viñaspre  
*Language Technology for Language Communities: An Overview based on Basque Experience 2020*  
Symposium Academiaidd Technolegau Iaith Cymru 2020 -11-04 // Wales Academic Symposium on Language Technologies 2020-11-04
  
19. Rodrigo Agerri, German Rigau  
*Language independent sequence labelling for Opinion Target Extraction*  
International Joint Conference on Artificial Intelligence (IJCAI 2020)



20. Itziar Gonzalez-Dios, Kepa Bengoetxea, Amaia Aguirregoitia  
*LagunTest: A NLP Based Application to Enhance Reading Comprehension*  
1st Workshop on Tools and Resources to Empower People with REAding Difficulties (READI2020),  
pages 63–69. ISBN: 979-10-95546-44-3
21. Unai Atutxa, Mikel Iruskieta, Olatz Ansa  
*Laburpena eskolan: estrakzioaren eta abstrakzioaren arteko zubia eskolan*  
Hizkuntzaren eta Literaturaren Didaktika testuinguru eleaniztunetan: Hizkuntzaren eta  
Literaturaren Didaktikaren Nazioarteko XX. Kongresuko aktak. 57-66.
22. Xabier Soto, Olatz Perez-de-Viñaspre, Gorka Labaka, Maite Oronoz  
*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.
23. Jon Ander Campos, Kyunghyun Cho, Arantxa Otegi, Aitor Soroa, Eneko Agirre, Gorka Azkune  
*Improving Conversational Question Answering Systems after Deployment using  
Feedback-Weighted Learning*  
Proceedings of the 28th International Conference on Computational Linguistics (COLING), pages  
2561–2571. Outstanding Paper (Top 1%).
24. Sergio Santana, Alicia Pérez, Arantza Casillas  
*HapLap at eHealth-KD Challenge 2020*  
Proceedings of the Iberian Languages Evaluation Forum co-located with 36th Conference of the  
Spanish Society for Natural Language Processing, IberLEF@ SEPLN
25. Rodrigo Agerri, Iñaki San Vicente, Jon Ander Campos, Ander Barrena, Xabier Saralegi, Aitor Soroa,  
Eneko Agirre  
*Give your Text Representation Models some Love: the Case for Basque*  
Proceedings of LREC.

26. 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, Lana Yeganova  
*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
27. Jon Alkorta, Itziar Gonzalez-Dios  
*Exploring the Enrichment of Basque WordNet with a Sentiment Lexicon*  
Proceedings of the Workshop on Multimodal Wordnets (MMWN-2020), pages 20–24. ISBN: 79-10-95546-41-2
28. Oier Lopez de Lacalle, Ander Salaberria, Aitor Soroa, Gorka Azkune, Eneko Agirre  
*Evaluating Multimodal Representations on Visual Semantic Textual Similarity*  
Proceedings of the Twenty-third European Conference on Artificial Intelligence, ECAI 2020, June 8-12, 2020, Santiago Compostela, Spain
29. Jon Ander Campos, Arantxa Otegi, Aitor Soroa, Jan Deriu, Mark Cieliebak, Eneko Agirre.  
*DoQA - Accessing Domain-Specific FAQs via Conversational QA*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7302–7314
30. Oscar Sainz, Oier Lopez de Lacalle, Itziar Aldabe, Montse Maritxalar  
*Domain Adapted Distant Supervision for Pedagogically Motivated Relation Extraction*  
Proceeding of 12th Edition of its Language Resources and Evaluation Conference (LREC2020).  
Marseille, France
31. Piroska Lendvai, Sándor Darányi, Christian Geng, Moniek Kuijpers, Oier Lopez de Lacalle, Jean-Christophe Mensonides, Simone Reborá, Uwe Reichel  
*Detection of Reading Absorption in User-Generated Book Reviews: Resources Creation and Evaluation*  
Proceeding of 12th Edition of its Language Resources and Evaluation Conference (LREC2020).  
Marseille, France

32. Ainara Estarrona, Izaskun Etxeberria, Ricardo Etxepare, Manuel Padilla-Moyano, Ander Soraluze  
*Dealing with dialectal variation in the construction of the Basque historical corpus*  
Proceedings of the 7th Workshop on NLP for similar languages, varieties and dialects (VarDial2020 at COLING 2020).
33. Itziar Gonzalez-Dios.  
*Data statement of the Corpus of Basque Simplified Texts*  
Data Statements workshop
34. Arantxa Otegi, Aitor Agirre, Jon Ander Campos, Aitor Soroa, Eneko Agirre  
*Conversational Question Answering in Low Resource Scenarios: A Dataset and Case Study for Basque*  
Proceedings of The 12th Language Resources and Evaluation Conference, pp. 429–435. European Language Resources Association. ISBN: 979-10-95546-34-4
35. Arantxa Otegi, Jon Ander Campos, Gorka Azkune, Aitor Soroa, Eneko Agirre  
*Automatic Evaluation vs. User Preference in Neural Textual Question Answering over COVID-19 Scientific Literature*  
Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020
36. Alberto Blanco, Alicia Pérez, Arantza Casillas  
*Automatic Classification of Medical Records with Multi-label Classifiers and Similarity Match Coders*  
CEUR Workshop Proceedings, Vol 2696 - Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum
37. Jan Deriu, Katsiaryna Mlynchyk, Philippe Schläpfer, Alvaro Rodrigo, Dirk von Grünigen, Nicolas Kaiser, Kurt Stockinger, Eneko Agirre, Mark Cieliebak  
*A Methodology for Creating Question Answering Corpora Using Inverse Data Annotation*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 897-911.
38. Mikel Artetxe, Sebastian Ruder, Dani Yogatama, Gorka Labaka, Eneko Agirre  
*A Call for More Rigor in Unsupervised Cross-lingual Learning*  
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

39. Salvador Lima, Naiara Pérez, Montse Cuadros, German Rigau  
*NUBes: A Corpus of Negation and Uncertainty in Spanish Clinical Texts*  
Proceedings of the 12th Language Resources and Evaluation Conference (LREC'20). Marseille, France. 2020.
40. María Espinosa, Rodrigo Agerri, Roberto Centeno, Alvaro Rodrigo  
*DeepReading@SardiStance:Combining Textual, Social and Emotional Features.*  
Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020). Winners of the SardiStance@Evalita 2020 shared task
41. Eneko Agirre, Marianna Apidianaki, Ivan Vulić (Editors)  
*Proceedings of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*  
In conjunction with EMNLP. Association for Computational Linguistics
42. Itziar Gonzalez-Dios, Javier Álvez, German Rigau  
*Towards modeling SUMO attributes through WordNet adjectives: a Case Study on Qualities*  
Proceedings of the Workshop on Multimodal Wordnets (MMWN-2020), pages 1–6. ISBN: 979-10-95546-41-2

## BOOK CHAPTERS

1. Javier Álvez, Itziar Gonzalez-Dios, German Rigau  
*Applying the Closed World Assumption to SUMO-based FOL Ontologies for Effective Commonsense Reasoning*  
Frontiers in Artificial Intelligence and Applications. Giuseppe De Giacomo, Alejandro Catala, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, Jérôme Lang (eds.). Volume 325: ECAI 2020. Pages 585 - 592. IOS Press Ebooks

## Projects and Knowledge Transference Contracts

| Type | Call  | Project  | PI                                     | Start      | End        | Amount     | 2020    |
|------|---|--|--|------------|------------|------------|---------|
| EU   | Acciones de Programación Conjunta Internacional (MINECO) 2017 | (APCI-MINECO17/01) Aprendiendo a Interactuar con personas mediante interacción continuada con personas | Agirre Bengoa, Eneko                   | 01/12/2017 | 31/12/2020 | 105.000    | 34.054  |
| EU   | Interreg V17/06   | (EFA_227/16) LINGUATEC   | Sarasola Gabiola, Kepa                 | 01/01/2018 | 01/12/2020 | 159.750    | 48.813  |
| EU   | INT-NOCORE 2019   | GOOGLE-Better Extraction from Text Towards Enhanced Retrieval  | Agirre Bengoa, Eneko                   | 01/05/2019 | 30/04/2020 | 79.654,00  | 26.551  |
| EU   | INT-NOCORE 2019   | BETTER-Better Extraction from Text Towards Enhanced Retrieval  | Agirre Bengoa, Eneko                   | 01/09/2019 | 31/03/2023 | 363.590,00 | 101.467 |
| EU   | INT-NOCORE 2020   | GOOGLE 2020  | Agirre Bengoa, Eneko                   | 16/03/2020 | 15/03/2021 | 70.000     | 52.500  |
| EU   | INT-NOCORE 2020   | INT NOCORE20/02  | Agirre Bengoa, Eneko                   | 13/01/2020 | 31/07/2020 | 38.948     | 38.948  |
| EU   | INT-NOCORE 2020   | INT NOCORE20/01  | Labaka Intxauspe, Gorka                | 01/01/2020 | 31/12/2021 | 125.832    | 62.916  |
| EU   | European Commission - Cost Action                             | enetCollect: European Network for Combining Language Learning with Crowdsourcing Techniques            | Agerrri Rodrigo, Maritxalar Montserrat | 07/03/2017 | 06/03/2021 | 0          | 0       |
| EU   | Linköping University, Sweden                                  | TAILOR: Trustworthy AI - Integrating Learning, Optimisation and Reasoning                              | Rigau Claramunt, German                | 01/09/2020 | 30/08/2023 | 0          | 0       |
| EU   | H2020-MSCA-ITN 2014   | Enriched communication across the lifespan   | COOKE, MARTIN                          | 2016       | 2020       | 684.934    | 128.038 |
| ES   | MINECOG 2017 Generación Conocimiento Explora                  | (MINECOG17/01) Traducción Automática Neuronal no Supervisada: un nuevo                                 | Agirre Bengoa, Eneko                   | 01/01/2018 | 31/12/2020 | 48.400     | 0       |
| ES   | MINECOG 2018 Generación Conocimiento Proyectos                | (MINECOG18/P36)DOMINO  | Sarasola Gabiola, Kepa                 | 01/01/2019 | 31/12/2021 | 81.070     | 27.023  |
| ES   | MINECOR 2018 Retos Proyectos                                  | (MINECOR18/P75) DeepReading  | Rigau Claramunt, German                | 01/01/2019 | 31/12/2021 | 146.894    | 48.965  |

| Type | Call  | Project   | PI                         | Start      | End        | Amount  | 2020   |
|------|---|---|----------------------------|------------|------------|---------|--------|
| ES   | MINECOG 2018 Generación Conocimiento Acciones | (MINECOG18/A07) Red estratégica para la promoción de las infraestructuras de tecnologías del lenguaje en eHumanidades y ciencias sociales | Rigau Claramunt, German    | 01/07/2019 | 31/12/2021 | 23.200  | 9.280  |
| ES   | MINECOR 2018 Retos Jóvenes Investigadores     | MINECOR18/J03   | Estarrona Ibarloza, Ainara | 01/01/2019 | 31/12/2021 | 140.360 | 46.787 |
| ES   | Red Española de Supercomputación (RES)        | Cross-language Natural Language Processing Models   | Agirre Bengoa, Eneko       | 20/02/2020 | 20/06/2020 | 0       | 0      |
| ES   | Fundaciones Privadas 2018                     | FP18/77 BBVA  | Agirre Bengoa, Eneko       | 30/04/2019 | 30/04/2021 | 71.500  | 34.320 |
| ES   | Fundaciones Privadas 2019                     | FP20/64 SUPERA COVID  | Agirre Bengoa, Eneko       | 01/09/2020 | 30/08/2021 | 46.500  | 15.500 |
| Eus  | Elkartek 2019                                 | (ELKARTEK19/77) Deep Learning aplicado al Procesamiento del Lenguaje Natural como apoyo a los ámbitos del RIS3                            | Soroa Echave, Aitor        | 01/01/2019 | 31/12/2020 | 42.547  | 21.274 |
| Eus  | Elkartek 2020                                 | ELKARTEK20/54 Deep Text   | Soroa Echave, Aitor        | 01/01/2020 | 31/12/2021 | 153.496 | 76.748 |
| Eus  | Elkartek 2020                                 | ELKARTEK20/49 TANDO   | Sarasola Gabiola, Kepa     | 01/01/2020 | 31/12/2021 | 79.208  | 39.604 |
| Eus  | Grupos Consolidados 2018 (Gv)                 | GIC18/37  | Agirre Bengoa, Eneko       | 01/01/2019 | 31/12/2021 | 282.900 | 94.300 |
| Eus  | Movilidad de Investigadores Gv 2020           | MOVGV20/01  | Agirre Bengoa, Eneko       | 20/08/2020 | 20/12/2020 | 12.189  | 12.189 |
| Eus  | Proy. Diputación Gipuzkoa 2020                | DG20/01   | Agirre Bengoa, Eneko       | 01/01/2020 | 30/09/2021 | 75.530  | 43.160 |
| Eus  | Proyectos Estratégicos 2018                   | (PES18/28) Arkitektura aurreratuetan oinarritutako testuetako prozesamendu-automatikoak   | Labaka Intxauspe, Gorka    | 01/09/2018 | 31/08/2022 | 51.423  | 12.856 |
| Eus  | Proyectos Estratégicos 2018                   | (PES18/88)  | Artola Zubillaga, Xabier   | 20/12/2018 | 19/12/2022 | 422     | 103    |
| Eus  | Proyectos Estratégicos 2020                   | (PES20/04)  | Perez De Viñaspre, Olatz   | 03/02/2020 | 02/02/2024 | 76.254  | 17.118 |
| Eus  | Ayuda Depart UPV/EHU 2020                     | DEP20/06  | Arregi Uriarte, Olatz      | 01/01/2020 | 31/12/2022 | 13.805  | 4.602  |
| Eus  | Ayuda Depart UPV/EHU 2020                     | DEP20/43  | Soroa Echave, Aitor        | 01/01/2020 | 31/12/2022 | 17.140  | 5.713  |
| Eus  | Infraestructura 2020                          | INF20/50  | Agirre Bengoa, Eneko       | 01/01/2020 | 31/12/2020 | 32.787  | 32.787 |

| Type | Call  | Project  | PI                        | Start      | End        | Amount  | 2020   |
|------|---|--|---------------------------|------------|------------|---------|--------|
| Eus  | UPV/EHU – COLAB19/19                                | Tools for the analysis of parliamentary discourses: polarization, subjectivity and affectivity in the post-truth era   | Agerri Gascón, Rodrigo    | 02/03/2020 | 01/03/2022 | 12.200  | 5.083  |
| Eus  | PROYECTOS ESTRATÉGICOS 2020                         | Aprendizaje profundo aplicado a las tecnologías de voz y audio   | Saratzaga Couceiro, Ibon  | 2020       | 2024       | 105.495 | 2.199  |
| Eus  | MINECOR 2019 Retos Proyectos (2019)                 | Restauración de voz con Interfaces EMG de Habla Silenciosa   | Hernáez Rioja, Inmaculada | 2020       | 2023       | 139.876 | 13.988 |
| Eus  | PROYECTOS ESTRATÉGICOS 2018                         | Desarrollo de Sistemas de Verificación de la Lengua y del Locutor  | Varona Fernández, Amparo  | 2018       | 2022       | 11.537  | 0      |
| Eus  | Proyectos Estratégicos 2017                         | Tecnologías de voz   | Navas Cordón, Eva         | 2017       | 2021       | 106.653 | 25.915 |
| Eus  | Proyectos Estratégicos 2017                         | Consolidación de Recursos, Aplicaciones y Técnicas de Alineamiento de Voz y Texto  | Bordel García, German     | 2017       | 2021       | 52.248  | 13.071 |
| Eus  | Proyectos de Investigación Básica y/o Aplicada 2018 | Aprendizaje profundo aplicado a voces patológicas  | Hernáez Rioja, Inmaculada | 2018       | 2021       | 50.000  | 23.357 |
| Eus  | PROYECTOS UNIVERSIDAD SOCIEDAD 2018                 | Voz infantil sintética para comunicación aumentativa y alternativa   | Navas Cordón, Eva         | 2018       | 2020       | 7.200   | 1.924  |
| Eus  | ELKARTEK 2019                                       | Traducción Automática Neuronal del Habla   | Navas Cordón, Eva         | 01/04/2019 | 31/12/2020 | 58.296  | 34.317 |
| Eus  | ELKARTEK 2019                                       | Deep Learning aplicado al Procesamiento del Lenguaje Natural como apoyo a los ámbitos del RIS3   | Hernáez Rioja, Inmaculada | 2019       | 2020       | 9.002   | 4.501  |
| Eus  | MINECOR 2019 Retos Proyectos                        | Metodologías no supervisadas para el aprovechamiento de datos de dominio público en reconocimiento automático del habla: de la abundancia a la escasez de recursos | Varona Fernández, Amparo  | 2020       | 2023       | 33.638  | 3.364  |
| Cont | Hazitek 2019  | Asesoría científica en la construcción de un asistente personal  | Agirre Bengoa, Eneko      | 01/02/2019 | 31/01/2020 | 55.882  | 51.225 |
| Cont | Ikerlan S.Coop.                                     | Asesoría sobre sistemas de diálogo voice2text  | Soroa Echave, Aitor       | 01/12/2020 | 31/12/2020 | 9.882   | 9.882  |

| Type | Call  | Project  | PI                        | Start      | End        | Amount | 2020   |
|------|---|--|---------------------------|------------|------------|--------|--------|
| Cont | Eleka Ingeniaritza Linguistikoa                 | MultiNMT: Traducción automática neuronal Multidireccional orientada al cliente   | Labaka Intxauspe, Gorka   | 01/01/2020 | 31/12/2020 | 9.000  | 9.000  |
| Cont | Hispavista Labs Aie                             | PRIDAIAS: Hacia la PrIvacidad de los Datos en Aplicaciones de Inteligencia Artificial para la Salud  | Arregui Uriarte, Olatz    | 01/12/2020 | 01/04/2022 | 55.239 | 3.452  |
| Cont | Nymiz Software Company, S.L.                    | Asesoría científica en el área del procesamiento del lenguaje natural en el ámbito de la anonimización   | Rigau Claramunt, German   | 01/06/2020 | 30/11/2020 | 16.000 | 16.000 |
| Cont | Euskaltzaindia-Real Academia de la Lengua Vasca | Observatorio lexical del euskera XII   | Ezeiza Ramos, Nerea       | 26/11/2020 | 26/03/2021 | 48.156 | 4.682  |
| Cont | Elhuyar Fundazioa                               | LICENCIA AHOLAB ELHUYAR AoTSS  | Hernáez Rioja, Inmaculada | 2018       | 2038       | 0      | 0      |
| Cont | Parlamento Vasco / Eusko Legebiltzarra          | Alineamiento automático de voz y texto para subtítulo de vídeos  | Bordel García, German     | 2020       | 2021       | 2.282  | 176    |
| Cont | Parlamento Vasco / Eusko Legebiltzarra          | Alineamiento automático de voz y texto para subtítulo de vídeos  | Bordel García, German     | 2019       | 2020       | 6.345  | 2.338  |
| Cont | DIDARTIKER, S.A.                                | Licencia. Proyecto de cooperación cExtractor, Plataforma de extracción automática de conocimiento a partir de fuentes de información estructuradas | Hernáez Rioja, Inmaculada | 2010       | 2020       | 0      | 0      |
| Cont | Datik Información Inteligente, S.L.             | Síntesis de voz en sistemas embarcados en la flota de autobuses de Vitoria/Gasteiz   | Sanchez de la Fuente, Jon | 2020       | 2020       | 8.050  | 8.050  |
| Cont | Mercedes-Benz España, S.A.                      | Transferencia y soporte a la puesta en marcha de un sistema automático de detección y clasificación de eventos sonoros basado en redes neuronales  | Saratxaga Couceiro, Ibon  | 2020       | 2020       | 6.500  | 6.500  |
| Otro | UPV/EHU   | 35th International Conference of the Spanish Society for Natural Language Processing, SEPLN 2019   | Hernáez Rioja, Inmaculada | 2020       | 2020       | 3.750  | 3.750  |