



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

language /'læŋgwɪdʒ/
noun

1. the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way.
- a non-verbal method of expression or communication.
2. a system of communication used by a particular country or community.

COMPUTING

- a system of symbols and rules for

speech.

HiTZ: Centro Vasco de Tecnología del Lenguaje

Informe anual
2019

technology /tɛk'nɒlədʒi/
noun

the application of scientific knowledge for practical purposes, especially in industry.

- machinery and equipment developed from the application of scientific knowledge.
- the branch of knowledge dealing with engineering or applied sciences

www.hitz.eus

 @Hitz_zentroa

center /'sɛntə/
noun

a place or group of buildings where a specified activity is concentrated.

INTRODUCCIÓN



La tecnología del lenguaje, la próxima frontera de la Inteligencia Artificial

language /'læŋgwɪdʒ/
noun

La comprensión del lenguaje es uno de los principales objetivos de la inteligencia artificial y, aunque se ha avanzado mucho en el aprendizaje profundo, sigue siendo una de las principales limitaciones de la inteligencia artificial, ya que los ordenadores no serán inteligentes hasta que no entiendan el lenguaje.

Para hacer frente a los nuevos desafíos y oportunidades de la transformación digital y la inteligencia artificial, es primordial fomentar la investigación en las tecnologías del lenguaje.

Para ello, hemos creado HiTZ, con un doble propósito: ser un referente internacional en la investigación de las tecnologías del lenguaje, y el referente en el tratamiento computacional del euskera. La actividad del centro comprende cinco líneas de actuación: formación, investigación, innovación y transferencia, fomentando nuestra cultura y el impacto en la sociedad. Este extenso ámbito requiere un amplio equipo interdisciplinar, que reúne a informáticos, lingüistas, ingenieros, traductores y sociólogos, entre otros.

El trabajo de los investigadores de Aholab e Ixa, los dos grupos de investigación que conforman el centro HiTZ, ya dio sus frutos en su primer año. Esta sólida base nos permite continuar con energías renovadas y, en colaboración con centros tecnológicos, empresas e instituciones, convertir al País Vasco en un centro internacional de tecnologías del lenguaje e inteligencia artificial.

Eneko Agirre (Director del HiTZ) y German Rigau (Subdirector del HiTZ)

technology /tɛk'nɒlədʒi/
noun

HITZ EN CIFRAS

37

proyectos

25

contratos de transferencia

78

publicaciones

7

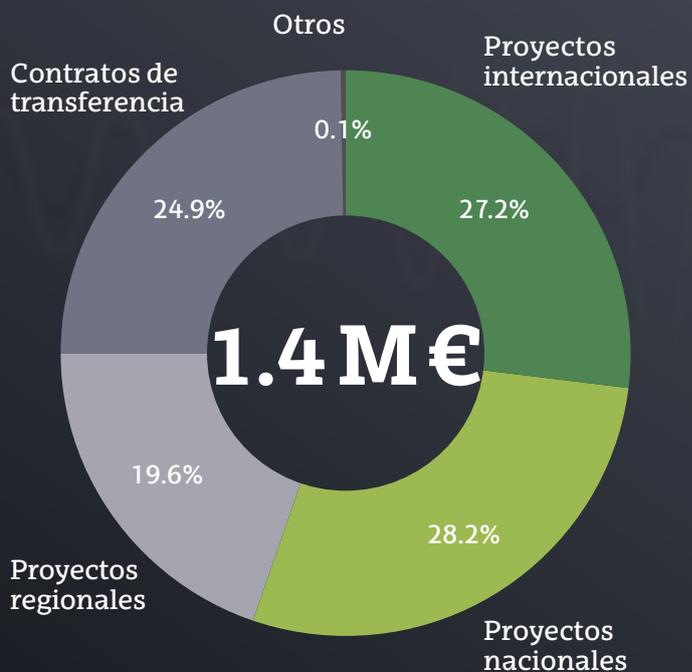
tesis



Miembros



Presupuesto



ORGANIZACIÓN

Los grupos de investigación **Ixa** y **Aholab**, ambos de la Universidad del País Vasco (UPV/EHU), han sido — desde su creación en 1988 y 1998 respectivamente — los principales tractores en el área de las Tecnologías del Lenguaje del País Vasco.



ahō LAB



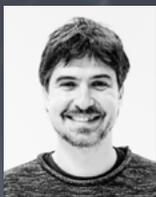
Eneko Agirre

Director



German Rigau

Subdirector



Kike Fernandez

Miembro de la Junta



Inma Hernaez

Miembro de la Junta



Oier Lopez de Lacalle

Miembro de la Junta



Maite Oronoz

Miembro de la Junta



Aitor Soroa

Miembro de la Junta



Mikel Artetxe

Miembro de la Junta

HiTZ es también miembro fundador del Centro CLARIN-K (Infraestructura Europea de Investigación para Recursos y Tecnología del Lenguaje), miembro de CLAIRE (actor clave en la estrategia de la IA en Europa) y miembro de pleno derecho de BDVA (Europa como líder mundial en la creación de Big Data Value).



ÁREAS DE INVESTIGACIÓN

Investigador/a principal:



German Rigau



Extracción y recuperación de información



Investigador/a principal:



Aitor Soroa



Análisis de texto



Investigador/a principal:



Eneko Agirre



Interacción persona-computadora



Investigador/a principal:



Arantza Casillas



Dominios médico y legal



Investigador/a principal:



Gorka Labaka



Traducción automática



Investigador/a principal:



Inma Hernaez



Síntesis y reconocimiento del habla



Investigador/a principal:



Ainara Estarrona



Recursos del habla y el lenguaje



Investigador/a principal:



Mikel Iruskieta



Humanidades digitales y educación



INFRAESTRUCTURA

14

servidores multiprocesadores GNU/Linux

4

servidores de SPARC Solaris

1

HPC Cluster con 128 núcleos

6

servidores con 27 GPU

Más de

120_{TB}

de capacidad de almacenamiento bruto en red

1

sala aislada acústicamente con equipo de audio para grabaciones profesionales

INVESTIGACIÓN Y TRANSFERENCIA



7

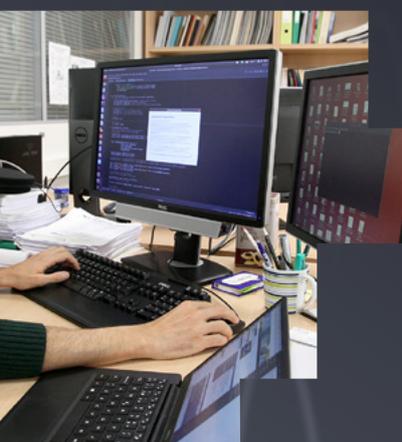
PROYECTOS INTERNACIONALES

12

PROYECTOS NACIONALES

18

PROYECTOS REGIONALES



25

CONTRATOS DE TRANSFERENCIA DE CONOCIMIENTO

3

OTROS PROYECTOS

7

TESIS DOCTORALES DEFENDIDAS (3 internacionales)

36

ARTÍCULOS DE REVISTAS (11 Q1)

34

ARTÍCULOS EN CONGRESOS (7 A, A+)

7

CAPÍTULOS DE LIBROS

1

LIBRO

language /'læŋgwɪdʒ/
noun

- 1. the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way.
- a non-verbal method of expression or communication.

technology
/tek'nɒlədʒi/

noun

- 2. a system of communication used by a particular country/region/culture.
- COMPUTING a system of symbols and rules for writing programs or algorithms.
- 3. the title of a piece of writing or speech.

4

TECNOLOGÍAS TRANSFERIDAS

8

BECAS DE DOCTORADO

PRÁCTICAS

training
/'treɪnɪŋ/
noun

15 

ESTUDIANTES
EN EL MÁSTER

17 

ESTUDIANTES EN
EL PROGRAMA DE
DOCTORADO

5 

TFMs de EMLCT
finalizadas

10 

TFMs de HAP/LAP
finalizadas

57 

ESTUDIANTES EN 2 CURSOS
COMPLEMENTARIOS DE
APRENDIZAJE PROFUNDO

3

IKASIKER

8

PRÁCTICAS INTERNAS
Y EXTERNAS



ACTIVIDADES



language
/ˈlæŋɡwɪdʒ/
noun



25

SEMINARIOS

6

VISITANTES
INTERNACIONALES

technology

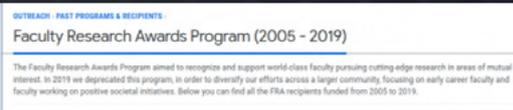
/tekˈnɒlədʒi/
noun

3

TALLERES
ORGANIZADOS

1

PREMIO
INVESTIGACIÓN
UNIVERSITARIA
DE GOOGLE



2019 award recipients	
Algorithms and Optimization	Augmented and Virtual Reality
Computational Neuroscience	Geo
Health	Human-computer Interaction
Information retrieval and real time content	Machine learning and data mining
Machine Perception	Machine Translation
Mobile	Natural Language Processing



1

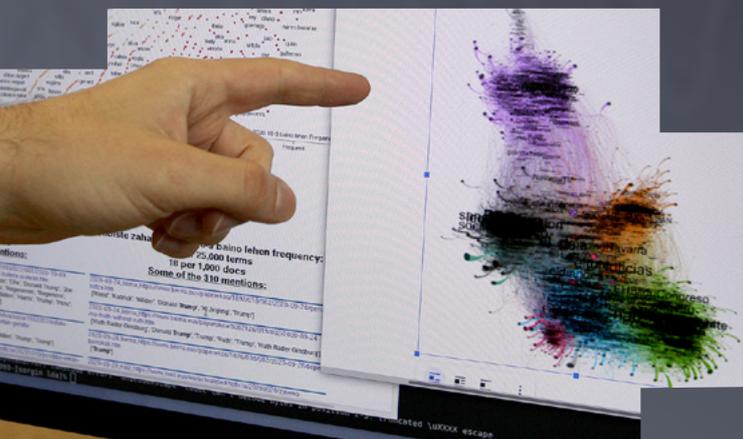
BECA DEL BBVA PARA
LA INVESTIGACIÓN
CIENTÍFICA EN BIG DATA

1

BECA DE
FACEBOOK

1

PREMIO MITXELENA
DE TESIS
DOCTORALES



2

PREMIOS
TFG/TFM

1

HOMENAJE DE
LA UPV/EHU

center

/ˈsɛntə/
noun

Investigación

Tesis

Publicaciones

Patentes

Proyectos y contratos

Tesis

1. *Speech recognition based strategies for on-line Computer Assisted Language Learning (CALL) systems in Basque/Hizketa-ezagutzan oinarritutako estrategiak, euskarazko online OBHI (Ordenagailu Bidezko Hizkuntza Ikaskuntza) sistemetarako.*
Odriozola Susaeta, Igor
Dirección: Hernáez, Inmaculada - Navas, Eva
2. *Técnicas para la mejora de la inteligibilidad en voces patológicas*
Serrano Garcia, Luis
Dirección: Hernáez, Inmaculada
3. *Euskal telebistaren sorrera, garapena eta funtzioa euskararen normalizazioaren testuinguruan*
Larrinaga Larrazabal, Asier
Dirección: Zabala, Miren Igone - Elorduy, Miren Agurtzane
4. *Aditza+izena unitate fraseologikoak gaztelaniatik euskarara: azterketa eta tratamendu konputazionala.*
Iñurrieta Urmeneta, Usoa
Dirección: Labaka, Gorka
5. *Multilingual sentiment analysis in social media*
San Vicente Roncal, Iñaki
Dirección: Rigau, German - Agerri, Rodrigo

6. *Sentimenduen analisi automatikorantz: oinarrizko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia-aldatzaileen identifikazioa/Towards the automatic analysis of sentiments in Basque: the creation of basic resources and the identification of valence shifters in different language levels.*

Alkorta Agirrezabala, Jon

Dirección: Gojenola, Koldo - Iruskieta Mikel

7. *Adverse drug reaction extraction on electronic health records written in Spanish*

Santiso Gonzalez, Sara

Dirección: Casillas, Arantza - Pérez, Alicia

Publicaciones

REVISTAS

1. Martin Cooke, Simon King, Valerie Hazan, Yannis Stylianou, Esther Janse, Deniz Baskent, Volker Hohmann, Axel Winneke, Inma Hernaez
Enriched communication across the lifespan
Procesamiento del Lenguaje Natural 63, 2019.
2. Raman, Sneha; Serrano, Luis; Winneke, Axel; Navas, Eva; Hernaez, Inma
Intelligibility and Listening Effort of Spanish Oesophageal Speech
APPLIED SCIENCES-BASEL Volumen: 9 Número: 16 Número de artículo: 3233
3. Sarasola, Xabier; Navas, Eva; Tavarez, David; Serrano, Luis; Saratzaga, Ibon; Hernaez, Inma
Application of Pitch Derived Parameters to Speech and Monophonic Singing Classification
APPLIED SCIENCES-BASEL Volumen: 9 Número: 15 Número de artículo: 3140
4. Tejedor,J; Toledano,DT; Lopez-Otero,P; Docio-Fernandez,L; Montalvo,AR; Ramirez,JM; Penagarikano,M; Rodriguez-Fuentes,LJ
ALBAYZIN 2018 spoken term detection evaluation: a multi-domain international evaluation in Spanish
Eurasip Journal on Audio Speech and Music Processing, 2019, 1, 16
5. Tejedor,J; Toledano,DT; Lopez-Otero,P; Docio-Fernandez,L; Penagarikano,M; Rodriguez-Fuentes,LJ; Moreno-Sandoval,A
Search on speech from spoken queries: the Multi-domain International ALBAYZIN 2018 Query-by-Example Spoken Term Detection Evaluation
Eurasip Journal on Audio Speech and Music Processing, 0, 13

6. Agata Savary, Silvio Ricardo Cordeiro, Timm Lichte, Carlos Ramisch, Uxoia Iñurrieta, Voula Giouli
Literal occurrences of Multiword Expressions: rare birds that cause a stir.
The Prague Bulletin of Mathematical Linguistics. NUMBER 112 APRIL 2019 5–54. ISSN: 1804-0462
7. Agerri,R; Rigau,G
Language independent sequence labelling for Opinion Target Extraction
Artificial Intelligence, 268, 85-95
8. Alberto Poncelas, Kepa Sarasola, Meghan Dowling, Andy Way, Gorka Labaka, Iñaki Alegria
Adapting NMT to caption translation in Wikimedia Commons for low-resource languages
Procesamiento del Lenguaje Natural 63, 2019.
9. Alvez,J; Hermo,M; Lucio,P; Rigau,G
Automatic white-box testing of first-order logic ontologies
Journal of Logic and Computation, 29, 5, 723-751
10. Alvez,J; Lucio,P; Rigau,G
A Framework for the Evaluation of SUMO-Based Ontologies Using WordNet
Ieee Access, 7, 36075-36093
11. Ander Sorraluze, Olatz Arregi, Xabier Arregi, Arantza Diaz de Ilarraza
EUSKOR: End-to-end coreference resolution system for Basque
PLoS ONE 14(9): e0221801.
12. Atutxa,A; Bengoetxea,K; de Ilarraza,AD; Iruskieta,M
Towards a top-down approach for an automatic discourse analysis for Basque: Segmentation and Central Unit detection tool
Plos One, 14, 9, e0221639
13. Atutxa,A; de Ilarraza,AD; Gojenola,K; Oronoz,M; Perez-de-Vinaspre,O
Interpretable deep learning to map diagnostic texts to ICD-10 codes
International Journal of Medical Informatics, 129, 49-59
14. Begoña Altuna
EUSKARAZKO DENBORA-EGITUREN AZTERKETA ETA CORPUSAREN SORRERA /ANALYSIS OF BASQUE TEMPORAL CONSTRUCTIONS AND THE CREATION OF A CORPUS.
UZTARO 110, 105-110 . ISSN: 1130-5738.
15. Blanco,A; Casillas,A; Perez,A; de Ilarraza,AD
Multi-label clinical document classification: Impact of label-density
Expert Systems with Applications, 138
16. Casillas,A; Ezeiza,N; Goenaga,T; Perez,A; Soto,X
Measuring the effect of different types of unsupervised word representations on

Medical Named Entity Recognition

International Journal of Medical Informatics, 129, 100-106

17. Agirre E, Otegi A, Pradel C, Rosset S, Peñas A, Cieliebak M
LIHLITH: Learning to Interact with Humans by Lifelong Interaction with Humans
Procesamiento del Lenguaje Natural 63, 2019.
18. Etxeberria,I; Alegria,I; Uria,L
Weighted finite-state transducers for normalization of historical texts
Natural Language Engineering, 25, 2, 307-321
19. Gamallo, Pablo, Susana Sotelo, José Ramom Pichel, Mikel Artetxe
Contextualized Translations of Phrasal Verbs with Distributional Compositional Semantics and Monolingual Corpora
Computational linguistics, first online
20. Jon Alkorta, Koldo Gojenola, Mikel Iruskietea
SentiTegi: Semi-manually Created Semantic Oriented Basque Lexicon for
Computación y Sistemas, 22 (4)
21. Joseba Fernandez de Landa, Rodrigo Agerri, Iñaki Alegria
Large Scale Linguistic Processing of Tweets to Understand Social Interactions among Speakers of Less Resourced Languages: The Basque Case
Information, vol. 10, 6.
22. José Ramom Pichel, Pablo Gamallo, Iñaki Alegria
Measuring diachronic language distance using perplexity. Application to English, Portuguese and Spanish.
Natural Language Engineering, 26(4), 433-454.
23. José Ramom Pichel, Pablo Gamallo, Iñaki Alegria
Cross-lingual Diachronic Distance: Application to Portuguese and Spanish.
Procesamiento del Lenguaje Natural 63, 2019.
24. Juan J. Lastra-Díaz, Josu Goikoetxea, Mohamed Ali Hadj Taieb, Ana García- Serrano, Mohamed Ben Aouicha, Eneko Agirre
Reproducibility dataset for a large experimental survey on word embeddings and ontology-based methods for word similarity
DATA IN BRIEF, VOL 26
25. Lastra-Diaz,JJ; Goikoetxea,J; Taieb,MAH; Garcia-Serrano,A; Ben Aouicha,M; Agirre,E
A reproducible survey on word embeddings and ontology-based methods for word similarity: Linear combinations outperform the state of the art
Engineering Applications of Artificial Intelligence, 85, 645-665

26. Lopez-Gazpio,I; Maritxalar,M; Lapata,M; Agirre,E
Word n-gram attention models for sentence similarity and inference
Expert Systems with Applications, 132, 1-11
27. Maria Jesus Aranzabe, Aitziber Atutxa, Kepa Bengoetxea, Arantza Diaz de Ilarraza, Iakes Goenaga, Koldo Gojenola, Larraitx Urria.
Dependentzia Unibertsalen ereduraren egokitutako euskarazko zuhaitz-bankua
Ekaia, 35, 2019, 291-307. ISSN 0214-9001 e-ISSN:2444-3255
28. Mikel Artetxe, Gorra Labaka, Eneko Agirre
Unsupervised Neural Machine Translation, a new paradigm solely based on monolingual text
Procesamiento del Lenguaje Natural 63, 2019.
29. Mikel Artetxe, Holger Schwenk
Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond
Transactions of the Association for Computational Linguistics 7 (2019): 597-610
30. Ona de Gibert, Nora Aranberri.
Estrategia multidimensional para la selección de candidatos de traducción automática para posesión
Linguamática, 11(2), 3-16.
31. Santiso,S; Casillas,A; Perez,A
The class imbalance problem detecting adverse drug reactions in electronic health records
Health Informatics Journal, 25, 4, 1768-1778
32. Santiso,S; Casillas,A; Perez,A; Oronoz,M
Word embeddings for negation detection in health records written in Spanish
Soft Computing, 23, 21, 10969-10975
33. Santiso,S; Perez,A; Casillas,A
Exploring Joint AB-LSTM With Embedded Lemmas for Adverse Drug Reaction Discovery
Ieee Journal of Biomedical and Health Informatics, 23, 5, 2148-2155
34. Santiso,S; Perez,A; Casillas,A
Smoothing dense spaces for improved relation extraction between drugs and adverse reactions
International Journal of Medical Informatics, 128, 39-45

35. Soto,X; Perez-de-Vinaspre,O; Labaka,G; Oronoz,M
Neural machine translation of clinical texts between long distance languages
Journal of the American Medical Informatics Association, 26, 12, 1478-1487
36. Itziar Aldabe, Josu Aztiria, Franchó 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: Desarrollo de recursos lingüísticos para avanzar en la digitalización de las lenguas de los Pirineos
Procesamiento del Lenguaje Natural 63, 2019.

PROCEEDINGS

1. Diez Gaspon, I., Saratxaga, I., Lopez de Ipiña, K.
Deep Learning For Natural Sound Classification
INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 2019; Pages 5683 - 5692
2. Serrano, L., Raman, S., Tavárez, D., Navas, E., Hernaez, I.
Parallel vs. Non-Parallel Voice Conversion for Esophageal Speech.
Proc. Interspeech 2019, pages 4549--4553.
3. Sneha Raman, Inma Hernández, Eva Navas, Luis Serrano
A multifaceted enrichment of oesophageal speech
Proc. of the 23rd International Congress on Acoustics, pages 5739-5741
4. Aitor Ormazabal, Mikel Artetxe, Gorka Labaka, Aitor Soroa and Eneko Agirre
Analyzing the Limitations of Cross-lingual Word Embedding Mappings
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4990-4995
5. Amir Zeldes, Debopam Das, Erick G. Maziero, Juliano D. Antonio, Mikel Iruskieta.
The DISRPT 2019 Shared Task on Elementary Discourse Unit Segmentation and Connective Detection
Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 144–152.
Minneapolis, MN, June 6, 2019. ACL
6. Andrea Amelio Ravelli, Oier Lopez de Lacalle, Eneko Agirre.
A comparison of representation models in a non-conventional semantic
Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy.
7. Arantza Diaz de Ilarraza, Mikel Iruskieta.
Ayuda de las tecnologías lingüísticas en la investigación en Humanidades Digitales.
XVI Simposio Internacional de Comunicación Social (XVI-SICS)

8. Begoña Altuna, Maria Jesus Aranzabe, Arantza Diaz de Ilarraza
EusTimeBank-TL corpora: denbora-informaziodun testuetatik denbora lerroetara.
Olatz Arbelaitz, Urtzi Etxeberria, Ainhoa Latatu, Miren Josu Ormaetxebarria (arg.), III. Ikergazte.
Nazioarteko Ikerketa Euskaraz, Giza Zientziak eta Artea (1. liburukia), 83-90. Udako Euskal
Unibertsitatea (UEU). Bilbo. ISBN: 978-84-8438-682-7
9. Damien Sileo, Camille Pradel, Guillermo Echegoyen, Anselmo Peñas, Arantxa Otegi, Jan Deriu,
Mark Cieliebak, Ander Barrena, Eneko Agirre.
Matching Words and Knowledge Graph Entities with Meta-Embeddings.
Proceedings of CAP2019, Toulouse (France) pages 34-39.
10. Eneko Agirre, Anders Jonsson, Anthony Larcher
Framing Lifelong Learning as Autonomous Deployment: Tune Once Live Forever.
Dialogue Systems and Lifelong Learning special session at Tenth International Workshop on
Spoken Dialogue Systems Technology (IWSDS).
11. Gorka Urbizu, Ander Soraluze, Olatz Arregi
*Deep Cross-Lingual Coreference Resolution for Less-Resourced Languages: The Case
of Basque.*
Proceedings of the 2nd Workshop on Computational Models of Reference, Anaphora and
Coreference (CRAC 2019), co-located with NAACL 2019. ISBN: 978-1-948087-97-1
12. Gorka Urbizu, Ander Soraluze, Olatz Arregi
Neurona-sareetan oinarritutako euskararako korreferentzia-ebazpena.
III. Ikergazte: Nazioarteko ikerketa euskaraz. pp. 141-147, Baiona. ISBN 978-84-8438-686-5
13. Iakes Goenaga, Sergio Santana, Sara Santiso, Koldo Gojenola, Alicia Pérez, Arantza Casillas.
*IxaMed at eHealth-KD Challenge 2019 Using Different Paradigms to Solve Clinical
Relation Extraction*
"Proceedings of the Iberian Languages Evaluation Forum co-located with 35th Conference of the
Spanish Society for Natural Language Processing, IberLEF@SEPLN 43-50. ISSN 1613-0073 "
14. Iratxe Esparza, Mikel Iruskietia.
*Haur Hezkuntzako ipuin-bilduman emakumearen irudia eta funtzioa: hurbilpen
narratologiko eta eLiterarioa.*
XX Congreso Internacional de Didáctica de la Lengua y la Literatura: Hizkuntzaren eta
Literaturaren didaktika testuinguru eleaniztunetan. SEDLL, 2019.
15. Itziar Gonzalez-Dios, German Rigau
*Textual genre based approach to use wordnets in language-for-specific-purpose
classroom as dictionary*
Proceedings of the Tenth Global Wordnet Conference, pp 222--227. ISBN 978-83-7493-108-3
16. Izaskun Aldezabal, Jose Mari Arriola, Ainara Estarrona
A modular grammar-helping tool for Basque: work in progress

Proceedings of the NoDaLiDa 2019 Workshop on Constraint Grammar - Methods, Tools and Applications, 30 September 2019, Turku, Finland. ISBN 978-91-7929-918-7

17. Javier Álvez, Itziar Gonzalez-Dios, German Rigau
Commonsense Reasoning Using WordNet and SUMO: a Detailed Analysis
Proceedings of the Tenth Global Wordnet Conference, pp 197--205. ISBN 978-83-7493-108-3
18. Jon Alkorta, Koldo Gojenola, Mikel Iruskietia
Towards discourse annotation and sentiment analysis of the Basque Opinion Corpus.
Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 144–152.
Minneapolis, MN, June 6, 2019. ACL
19. Jon Alkorta, Koldo Gojenola, Mikel Iruskietia
Sentimenduen tratamendu konputazionalerantz: gramatika maila ezberdinetako sentimendu balentzia aldatzaileen bila.
Olatz Arbelaitz, Urtzi Etxeberria, Ainhoa Latatu, Miren Josu Ormaetxebarria (arg.), III. Ikergazte. Nazioarteko Ikerketa Euskaraz, Giza Zientziak eta Artea (1. liburukia), 39-46. Udako Euskal Unibertsitatea (UEU). Bilbo. ISBN 978-84-8438-682-7
20. Jon Ander Campos, Arantxa Otegi, Aitor Soroa, Jan Deriu, Mark Cieliebak, Eneko Agirre.
Conversational QA for FAQs
NeurIPS 3rd Conversational AI Workshop: "Today's Practice and Tomorrow's Potential"
21. Joseba Fernandez de Landa, Rodrigo Agerri, Iñaki Alegria.
Euskaldun gazte eta helduen harremanak Twitterren
III. Ikergazte. Nazioarteko ikerketa euskaraz. Kongresuko artikulu bilduma. Gizarte Zientziak eta Zuzenbidea. 2, pp. 83 – 90. ISBN 978-84-8438-681-0
22. Manex Agirrezabal, Begoña Altuna, Lara Gil-Vallejo, Josu Goikoetxea, Itziar Gonzalez-Dios.
Creating vocabulary exercises through NLP
"Proceedings of the Digital Humanities in the Nordic Countries 4th Conference, CEUR-WS, vol. 2364, pp. 18-32. ISSN:1613-0073. <http://ceur-ws.org/Vol-2364/> "
23. Mikel Artetxe, Gorka Labaka, Eneko Agirre
Bilingual Lexicon Induction through Unsupervised Machine Translation
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5002-5007.
24. Mikel Artetxe, Gorka Labaka, Eneko Agirre
An Effective Approach to Unsupervised Machine Translation.
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 194-203
25. Mikel Artetxe, Holger Schwenk
Margin-based Parallel Corpus Mining with Multilingual Sentence Embeddings

- Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3197-3203.
26. Mikel Iruskieta.
CLARIN Europako sarea: eHumanitateak eta zientzia sozialak lankidetzarako behar duten hizkuntza-azpiegitura sortzen.
Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation.
27. Mikel Iruskieta, Chloé Braud
EusDisParser: improving an under-resourced discourse parser with cross-lingual data
Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 62–71. Minneapolis, MN, June 6, 2019. ACL
28. Mikel Iruskieta, Kepa Bengoetxea, Aitziber Atutxa, Arantza Diaz de Ilarraza.
Multilingual segmentation based on neural networks and pre-trained word embeddings.
Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 125-133. Minneapolis, MN, June 6, 2019. ACL.
29. Rodrigo Agerri.
Doris Martin at SemEval-2019 Task 4: Hyperpartisan News Detection with Generic Semi-supervised Features
Proceedings of SemEval@NAACL-HLT 2019: 944-948. ISBN: 978-1-950737-06-2
30. Udane Beaskoetxea, Mikel Iruskieta
Ipuin-moldaketa herri-hizkerara egokitzeko, aldatzeko eta modu esanguratsuan kontatzeko markaketa: Ahozko komunikazioa lantzen eta aztertzen Haur Hezkuntzako gelan
Tantak 31(1), 53--80
31. Atutxa U, Iruskieta M, Ansa O
Laburpenaeskolan: estrakzioaren eta abstrakzioaren arteko zubia
XX Congreso Internacional de Didáctica de la Lengua y la Literatura: Hizkuntzaren eta Literaturaren didaktika testuinguru eleaniztunetan. SEDLL, 2019.
32. Xabier Lahuerta, Iakes Goenaga, Koldo Gojenola, Aitziber Atutxa, Maite Oronoz.
IxaMed at PharmacoNER Challenge 2019
"Proceedings of the Iberian Languages Evaluation Forum co-located with 35th Conference of the Spanish Society for Natural Language Processing IberLEF@SEPLN. ISSN 1613-0073 "
33. Xabier Soto, Olatz Perez de Viñaspre, Maite Oronoz, Gorra Labaka.
Leveraging SNOMED CT terms and relations for machine translation of clinical texts from Basque to Spanish

Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation.

34. Y Yaghoobzadeh, K Kann, TJ Hazen, E Agirre, H Schütze
Probing for Semantic Classes: Diagnosing the Meaning Content of Word Embeddings.
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 5740–5753

LIBROS

Iruskieta M, Maritxalar M, Arroyo-Sagasta A, Camacho A
IKTak eta konpetentzia digitalak hezkuntzan
UPV/EHU – UEU, 2019

CAPITULO DE LIBRO

1. Begoña Altuna, María Jesús Aranzabe, Arantza Díaz de Ilarraza
Euskarazko denbora-informazioaren azterketa tratamendu automatikorako
In Itziar Aduriz eta Ruben Urizar (ed.), Hizkuntzalari euskaldunen III. topaketa. Zer berri?, 135-148. Bilbo: Udako Euskal Unibertsitatea (UEU).
2. Itziar Gonzalez-Dios
Nautikako terminologia biltzen testu-generoak abiapuntu: nabigazio egunerokoen ereduak
Hizkuntzalari euskaldunen III. topaketa. Zer berri?. ed. Itziar Aduriz, Ruben Urizar. 79-91. Udako Euskal Unibertsitatea
3. Jennifer Isasi, Mikel Iruskieta.
Humanitate Digitalak eta haien ikerketaren nondik norakoak.
In "IKTak eta konpetentzia digitalak hezkuntzan" Eds: Iruskieta, M. Maritxalar, M. Arroyo, A. Camacho, A. 307-332. UEU eta UPV/EHU. Bilbo.
4. Itziar Aldabe, Montse Maritxalar
Corpusak eta beren aplikazioak
In "IKTak eta konpetentzia digitalak hezkuntzan" Eds: Iruskieta, M. Maritxalar, M. Arroyo, A. Camacho, A. UEU eta UPV/EHU. Bilbo.
5. Oier Lopez de Lacalle
Bistaraketa teknikak konpetentzia digitaletan
In "IKTak eta konpetentzia digitalak hezkuntzan" Eds: Iruskieta, M. Maritxalar, M. Arroyo, A. Camacho, A. UEU eta UPV/EHU. Bilbo.

6. Mikel Iruskieta, Aitziber Iturricastillo
Idatzizko eta ahozko hizkuntza-irakaskuntza online
In "IKTak eta kompetentzia digitalak hezkuntzan" Eds: Iruskieta, M. Maritxalar, M. Arroyo, A. Camacho, A. 279-294 UEU eta UPV/EHU. Bilbo.
7. Mikel Iruskieta, Arantxa Otegi, Larraitx Uria, Arantza Diaz de Ilarraza, Amaia Artolazabal.
Zer i(ra)kas dezakegu geure corpusekin "jolastuz"?
Traineru bete lagun: Iñaki Gaminde omenduz. UPV/EHU. 35-66 or.

Patentes

1. *TADEEP: Sistema traducción automática neuronal para español-inglés y español-euskera*
Iñaki Alegria, Nora Aranberri, Mikel Artetxe, Gorra Labaka, Kepa Sarasola
2019
2. *Sistema automático para la detección de entidades de dominio médico en español*
Aitziber Atutxa, Arantza Casillas, Koldo Gojenola, Maite Oronoz, Alicia Pérez, Olatz Perez de Viñaspre, Sara Santiso
2019
3. *IXA pipes*
Rodrigo Agerri, Itziar Aldabe, Begoña Altuna, Nora Aranberri, Zuhaitz Beloki, Ainara Estarrona, German Rigau , Aitor Soroa, Larraitx Uria, Ruben Urizar
2019
4. *EUSRST_TB*
Jon Alkorta, Maxux Aranzabe, Arantza Díaz de Ilarraza, Kike Fernandez, Itziar Gonzalez-Dios, Mikel Iruskieta, Mikel Lersundi, Oier Lopez de Lacalle , Esther Miranda
2019

Proyectos y contratos

Tipo	Convocatoria	Proyecto	IP	Inicio	Fin	Monto	2019
EU	H2020-MSCA-ITN 2014 (2014)	(H2020-MSCA-ITN14/04) Enriched communication across the lifespan	COOKE , MARTIN	2016	2020	684.934	170.965
EU	Interreg V17/06	(EFA_227/16) Linguatex: Desarrollo de la cooperación transfronteriza y la transferencia de conocimiento en tecnologías de la lengua	Sarasola Gabiola, Kepa	01/01/2018	01/12/2020	159.750	53.250
EU	INT-NOCORE 2019 (2019)	(INT NOCORE19/08) Better Extraction from Text Towards Enhanced Retrieval	Eneko Agirre Bengoa	01/01/2018	01/12/2020	363.590	33.822
EU	INT-NOCORE 2019 (2019)	(INT NOCORE19/03) Accessing FAQ and CQA sites via dialogue	Eneko Agirre Bengoa	01/05/2019	30/04/2020	79.654	53.103
EU	Appel à projets générique 2017, PRC. Défi 8"	BIM: Corpus histórico del euskera anotado sintácticamente	Ricardo Etxepare	04/12/2017	03/12/2021	101.529	67.686
EU	INT-EURORÉGION AQUITAINE-EUSKADI-NAVARRA 2018 (2018)	Multilingüismo e identidades glocales en las redes sociales entre los jóvenes vascos	ELORDUY URQUIZA, MIREN AGURTZANE	20/07/2018	30/10/2019	9.000	5.850
EU	European Commission - Cost Action	(COST Action CA16105) enetCollect: Una nueva red europea para el aprendizaje de idiomas y el crowdsourcing	Verena Lyding	07/03/2017	06/03/2021	-	-
ES	MINECOR 2015 RETOS PROYECTOS (2015)	(MINECOR15/P51) Restauración, Almacenamiento y Rehabilitación de la voz	Inmaculada Hernáez Rioja	01/01/2016	31/12/2018	145.442	0
ES	Proyectos de Investigación I+D+i RETOS INVESTIGACIÓN Jóvenes Investigadores. (2019)	(RTI2018-098082-J-I00) Creación de un Corpus Histórico Anotado Sintácticamente para el Euskera (SAHCOBA)	Ainara Estarrona Ibarloza	02/09/2019	02/09/2022	101.160	71.000
ES	MINECOG 2018 GENERACION CONOCIMIENTO ACCIONES (2018)	(MINECOG18/A07) Red estratégica para la promoción de las infraestructuras de tecnologías del lenguaje en eHumanidades y ciencias sociales	German Rigau Claramunt	2019	2021	-	-
ES	Fundaciones Privadas 2018 (2018)	(FP18/77) BigKnowledge for Text Mining	Eneko Agirre Bengoa	30/04/2019	30/04/2021	71.500	23.833
ES	MINECOR 2018 RETOS PROYECTOS (2018)	(MINECOR18/P79) Reconocimiento e interacción con Personas basado en biometría blanda para la mejora de los Servicios de asistencia multilingüe robótica basados en percepción multisensorial	SIERRA ARAUJO, BASILIO	01/01/2019	31/12/2021	101.640	48.787
ES	MINECOR 2018 RETOS PROYECTOS (2018)	(MINECOR18/P75) DeepReading: Mining, Understanding, and Reasoning with Multilingual Content	German Rigau Claramunt	01/01/2019	31/12/2021	146.894	108.702

Tipo	Convocatoria	Proyecto	IP	Inicio	Fin	Monto	2019
ES	MINECOG 2018 GENERACION CONOCIMIENTO PROYECTOS (2018)	(MINECOG18/P36) Traducción Automática Neuronal, en DOMinio, NO supervisada	Kepa Sarasola Gabiola	01/01/2019	31/12/2021	81.070	38.914
ES	MINECOG 2017 GENERACION CONOCIMIENTO EXPLORA (2017)	(MINECOG17/01) Traducción Automática Neuronal no Supervisada: un nuevo paradigma basado solo en textos monolingües	Eneko Agirre Bengoa	01/11/2018	31/10/2020	48.400	27.588
ES	MINECOR 2016 RETOS PROYECTOS (2016)	(MINECOR16/P58) PROCesamiento Semántico textual Avanzado para la detección de diagnósticos, procedimientos, otros conceptos y sus relaciones en informes MEDicos	Koldo Gojenola Gallettebeitia	30/12/2016	29/12/2019	154.638	42.525
ES	Acciones de Programación Conjunta Internacional (MINECO) 2017	(APCI-MINECO17/01) Aprendiendo a Interactuar con personas mediante interaccioncontinuada con personas	AGIRRE BENGOA, Eneko	01/12/2017	30/11/2020	105.000	35.000
ES	MINECOR 2015 RETOS PROYECTOS (2015)	(MINECOR15/P85) Análisis de Personas con biometria blanda paRA Servicios Inteligentes Multilingües de asistencia sobre Percepción multisensorial y robótica mOvi	SIERRA ARAUJO, BASILIO	01/01/2016	30/09/2019	89.903	2.341
ES	MINECOG 2015 GENERACION CONOCIMIENTO EXPLORA (2015)	(MINECOG15/E12) Generación Automática de Procesadores Semánticos	German Rigau Claramunt	01/05/2017	30/04/2019	44.770	0
Eus	PROYECTOS ESTRATÉGICOS 2018 (2018)	(PES18/87) DESARROLLO DE SISTEMAS DE VERIFICACIÓN DE LA LENGUA Y DEL LOCUTOR	Amparo Varona Fernández	18/12/2018	17/12/2022	2.185	547
Eus	PROYECTOS DE INVESTIGACION BASICA V/O APLICADA 2018 (2018)	(PIBA18/25) Aprendizaje profundo aplicado a voces patológicas	Inmaculada Hernáez Rioja	01/09/2018	01/06/2021	50.000	26.143
Eus	PROYECTOS ESTRATÉGICOS 2017 (2017)	(PES17/26) Tecnologías de voz	Eva Navas Cordón	2017	2021	106.653	31.549
Eus	PROYECTOS ESTRATÉGICOS 2017 (2017)	(PES17/15) Consolidación de Recursos, Aplicaciones y Técnicas de Alineamiento de Voz y Texto	German Bordel García	2017	2021	52.248	13.071
Eus	ELKARTEK 2019 (2019)	(ELKARTEK19/76) Deep Learning aplicado al Procesamiento del Lenguaje Natural como apoyo a los ámbitos del RIS3	Inmaculada Hernáez Rioja	2019	2020	9.002	4.501
Eus	PROYECTOS UNIVERSIDAD-SOCIEDAD 2018 (2018)	(US18/22) Voz infantil sintética para comunicación aumentativa y alternativa	Eva Navas Cordón	2018	2020	7.200	5.092
Eus	ELKARTEK 2019 (2019)	(ELKARTEK19/56) Traducción Automática Neuronal del Habla	Eva Navas Cordón	01/04/2019	31/12/2020	58.296	23.979
Eus	GRUPOS DE INVESTIGACION UPV/EHU 2016 (2016)	(GIU16/32) Aholab Signal Processing Laboratory	Inmaculada Hernáez Rioja	2016	2019	33.750	10.948

Tipo	Convocatoria	Proyecto	IP	Inicio	Fin	Monto	2019
Eus	PROYECTOS ESTRATÉGICOS 2015 (2015)	(PES15/18) Procesamiento del habla	Eva Navas Cordón	2015	2019	132.594	0
Eus	GRUPOS DE INVESTIGACION UPV/EHU 2016 (2016)	(GIU16/68) Grupo Trabajo en Tecnologías Software	Amparo Varona Fernández	01/03/2017	01/03/2020	12.000	4.000
Eus	PROYECTOS ESTRATÉGICOS 2018 (2018)	(PES18/88) LEXIKOAREN BEHATOKIA IIII	Xabier Artola Zubillaga	2018	2022	423	106
Eus	PROYECTOS ESTRATÉGICOS 2018 (2018)	(PES18/28) Procesamiento automático de textos basado en arquitecturas avanzadas	Gorka Labaka Intxauspe	2018	2022	187.872	24.026
Eus	ELKARTEK 2019 (2019)	(ELKARTEK19/77) Deep Learning aplicado al Procesamiento del Lenguaje Natural como apoyo a los ámbitos del RIS3	Aitor Soroa Etxabe	28/02/2019	31/03/2020	42.547	21.274
Eus	ELKARTEK 2018 (2018)	(ELKARTEK18/114) Modelado neuronal avanzado para traducción de alta calidad	Kepa Sarasola Gabiola	22/02/2018	31/12/2019	74.320	45.212
Eus	GRUPOS DE INVESTIGACION UPV/EHU 2016 (2016)	(GIU16/16) IXA taldea	Arantza Diaz de Ilarraza Sanchez	22/12/2016	21/12/2019	135.000	43.791
Eus	PROYECTOS DIPUTACIÓN GIPUZKOA 2018 (2018)	(DG18/09) IXA azpiegitura: Ikasketa sakoneko teknikan ikertzeko GPU zerbitzaria	Arantza Diaz de Ilarraza Sanchez	25/07/2018	30/09/2019	12.000	4.975
Eus	PROYECTOS DIPUTACIÓN GIPUZKOA - ETORKIZUNA ERAIKIZ 2019 (2019)	(DGE19/05) JolasMATIKA	LARRAZA MENDILUZE, EDURNE	01/01/2019	31/12/2019	14.500	14.500
Cont	ELHUYAR FUNDAZIOA (Privada)	(2018.0788) LICENCIA AHOLAB ELHUYAR AotSS	Inmaculada Hernáez Rioja	2018	2038	0	0
Cont	DIDAKTIKER, S.A. (Privada)	(2010.0666) Licencia. Proyecto de cooperación c-Extractor, Plataforma de extracción automática de conocimiento a partir de fuentes de información estructuradas	Inmaculada Hernáez Rioja	2010	2020	0	0
Cont	PARLAMENTO VASCO / EUSKO LEGEBILTZARRA (Pública)	(2019.0127) Alineamiento automático de voz y texto para subtitulado de vídeos	German Bordel García	2019	2020	5.729	3.618
Cont	Mercedes-Benz España S.A. (Privada)	(2017.0056) ANÁLISIS DE LA CALIDAD ACÚSTICA DE LOS VEHÍCULOS PARA SU EVALUACIÓN AUTOMÁTICA	Ibon Saratxaga Couceiro	2017	2019	146.991	10.390
Cont	ELHUYAR FUNDAZIOA (Privada)	(2019.0419) Asesoría tecnológica AHOLAB	Inmaculada Hernáez Rioja	2018	2019	2.000	923
Cont	SHERPA EUROPE, S.L. (Privada)	(2019.0124) Asesoría científica en la construcción de un asistente personal	Eneko Agirre Bengoa	01/02/2019	31/01/2020	51.765	47.451

Tipo	Convocatoria	Proyecto	IP	Inicio	Fin	Monto	2019
Cont	HISPAVISTA LABS AIE (Privada)	(2019.0584) SERVICIO EN LA NUBE BASADO EN INTELIGENCIA ARTIFICIAL PARA LA RECOMENDACIÓN DE FORMACIÓN EN BASE A PERFIL Y NAVEGACIÓN DEL USUARIO	Olatz Arregi Uriarte	08/04/2019	30/09/2020	55.239	27.308
Cont	OSAKIDETZA - SERVICIO VASCO DE SALUD (Pública)	(2019.0426) Itzulbide: Testu klinikoak euskaratik eta euskarara egokitzeko itzultzaile automatiko baten garapen eta ezartzea	Maite Oronoz Anchordoqui	05/06/2019	05/06/2021	79.120	45.274
Cont	REPLAY LABS SPAIN,S.L. (Privada)	(2019.0134) Sistema de búsqueda de respuestas aplicado a las listas de FAQ	Aitor Soroa Etxabe	01/04/2019	01/06/2019	7.523	7.523
Cont	SHERPA EUROPE, S.L. (Privada)	(2019.0165) Fake News	Rodrigo Agerrri Gascón	01/04/2019	31/07/2019	26.400	26.400
Cont	MINISTERIO DE ECONOMÍA Y EMPRESA (Pública)	(2019.0166) Plan de formación de las tecnologías lingüísticas	Arantza Diaz de Ilarraza Sanchez	03/06/2019	30/09/2019	14.800	14.800
Cont	LEGÍTIMO SOLUTIONS, S.L. (Privada)	(2019.0296) Extracción de información documental	AZCUNE GALPARSORO, GORKA	01/06/2019	01/07/2019	4.518	4.518
Cont	EUSKALTZAINDIA-REAL ACADEMIA DE LA LENGUA VASCA (Privada)	(2019.0343) Observatorio lexical del euskera XI	EZEIZA RAMOS, NEREA	18/06/2019	18/08/2019	33.851	33.851
Cont	EUSKO JAURLARITZA- Hizkuntza Ikerketa eta Koordinaziorako Zuzendaritza. Hizkuntza Politikarako Sailburuordetza (Pública)	(2019.0488) Hizkuntza Teknologiak: EGOERAREN DIAGNOSTIKOA ETA AMIA egitea	Olatz Arregi Uriarte	02/09/2019	30/12/2019	12.390	12.390
Cont	EUSKO JAURLARITZA- Euskara Sustatzeko Zuzendaritza. Hizkuntza Politikarako Sailburuordetza (Pública)	(2019.0544) EUSKARA HTen ARLOAN SUSTATZEKO PROPOSAMENAK	Olatz Arregi Uriarte	01/10/2019	15/11/2019	8.300	8.300
Cont	DOMINION SMART SOLUTIONS, SAU (Privada)	(2019.0628) Revisión lingüística de un sistema de diálogo en Euskera	Aitor Soroa Etxabe	03/06/2019	30/09/2019	3.999	3.999
Cont	IKERLAN S.COOP. (Privada)	(2019.0720) Informe del estado del arte de sistemas de diálogo	Aitor Soroa Etxabe	03/06/2019	30/09/2019	15.000	15.000
Cont	EJIE Sociedad Informática del Gobierno Vasco (Privada)	(2019.0754) Tecnologías del lenguaje en proyectos transversales	Olatz Arregi Uriarte	03/06/2019	30/09/2019	12.390	12.390

Tipo	Convocatoria	Proyecto	IP	Inicio	Fin	Monto	2019
Cont	ELEKA INGENIARITZA LINGUISTIKOA (Privada)	(2019.0767) MultiNMT: Traducción automática neuronal multidireccional orientada al cliente	Gorka Labaka Intxauspe	01/07/2019	20/12/2019	9.000	9.000
Cont	TALIO TRAINING, S.L. (Privada)	(2019.0777) ZL-2019/00198 Plataforma para la generación, entrenamiento y gobierno de robots inteligentes	German Rigau Claramunt	01/09/2019	20/12/2019	30.150	30.150
Cont	GOBIERNO VASCO / EUSKO JAURLARITZA (Pública)	(FR60347) CURSO PREST GARA – LOTE: 12 TRATAMIENTO INTEGRAL DE LAS LENGUAS Y NORMALIZACIÓN 3 - CÓDIGO: 57 – “HIZKUNTZEN TRATAERA INTEGRATUA ESKOLAN” - (ONLINE 30 H.)	IBARLUZEA SANTISTEBAN, MIREN	2019	2019	2.685	2.685
Cont	GOBIERNO VASCO / EUSKO JAURLARITZA (Pública)	(FR60348) CURSO PREST GARA – LOTE: 12 TRATAMIENTO INTEGRAL DE LAS LENGUAS Y NORMALIZACIÓN 3 - CÓDIGO: 772 – “HIZKUNTZEN TRATAERA INTEGRATUA BIGARREN HEZKUNTZAKO IKASTETXEETAN” - (ONLINE 30 H.)	IBARLUZEA SANTISTEBAN, MIREN	2019	2019	2.685	2.685
Cont	HISPAVISTA LABS AIE (Privada)	(2018.0632) Nueva solución inteligente y universal para la búsqueda segura de información web orientada a niños	Olatz Arregi Uriarte	30/04/2018	30/09/2019	55.239	28.357
Cont	ELEKA INGENIARITZA LINGUISTIKOA (Privada)	(2018.0787) LICENCIA:CONTRATO PARA ESTABLECER LA PROPIEDAD, EL USO, Y LA EXPLOTACIÓN DE LA TECNOLOGÍA RESULTANTE DE MODELA	Iñaki Alegria Loinaz	2018	2019	0	0
Cont	ELHUYAR FUNDAZIOA (Privada)	(2018.0626) Servicio para la desambiguación de agentes del sector I D I.	Maite Oronoz Anchordoqui	26/04/2018	25/01/2019	7.230	482
Otro	EJ/GV (Pública)	(CGV19/27) 35th International Conference of the Spanish Society for Natural Language Processing SEPLN 2019	Inmaculada Hernández Rioja	2019	2020	6.350	1.676
Otro	UPV/EHU (Pública)	(OC19/17) SEDLen XX. Nazioarteko Kongresua: Hizkuntzaren eta Literaturaren Didaktika Testuinguru eleanizetan	ETXEARRIA LEJARRETA, AINTZANE	2019	2020	4.500	375

Formación

Tesis de máster

Tesis de
máster

EMLCT

1. *Speech Emotion Recognition using Machine Learning algorithms*
Natallia Chaiko
Dirección: Prof. Eva Navas - Prof. Roberto Zamparelli
2. *A multilingual approach towards improving the linguistic module of a TTS system: Case Navarro-Lapurdian dialect*
María Andrea Cruz Blandón
Dirección: Prof. Inma Hernaez Rioja - Prof. Denis Jouvét
3. *An investigation of accent conversion for non-native and native varieties of English*
Kenny Lino
Dirección: Dr. Claudia Borg - Eva Navas
4. *Data-Driven Pronunciation Generation for ASR*
Maria Obedkova
Dirección: Mgr. Ondřej Plátek - Eva Navas
5. *Cross-lingual Model Transfer for Neural Semantic Role Labeling*
Jovana Urošević
Dirección: Dr. Lonneke van der Plas - Dr. Eneko Agirre

HAP/LAP

1. *Robust Document Representations for Hyperpartisan and Fake News Detection*
Talita Anthonio, MA
Dirección: Rodrigo Agerri - Malvina Nissim
2. *Annotation Inconsistencies in Universal Dependencies: Identification, and Correction*
Akshay Aggarwal
Dirección: Daniel Zeman, Koldo Gojenola
3. *Universal Dependencies for Cape Verdean Creole*
Brandyn Emile Evora
Dirección: Aitziber Atutxa Salazar - Koldo Gojenola
4. *Automatic Stance Detection on Political Discourse in Twitter*
Elena Zotova
Dirección: Rodrigo Agerri - German Rigau
5. *A Common Semantic Space for Monolingual and Cross-Lingual Meta-Embeddings*
Iker García Ferrero
Dirección: German Rigau i Claramunt
6. *An implementation of PLTL-model-checking using context-based tableaux and SAT-solvers*
Luis Pérez Guerrero
Dirección: Montserrat Hermo - Paqui Lucio
7. *Exploring metrics for post-editing effort: and their ability to detect errors in machine translated output*
Cristina Cumbreño Díez
Dirección: Nora Aranberri
8. *Spanish-Swedish Neural Machine Translation for the Civil Engineering Domain*
Linda Emilia Hedberg
Dirección: Gorka Labaka - Koldo Gojenola

9. *Building a dialogue system for question-answer forum websites*

Jon Ander Campos

Dirección: Eneko Agirre - Arantxa Otegi

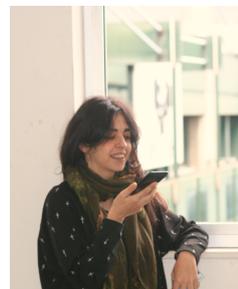
10. *Gazteak eta euskara sare sozialetan. Zer, nori, nork: euskarazko txio formal eta informalak sailkatuz eta konparatuz.*

Joseba Fernandez de Landa Agirre

Dirección: Iñaki Alegria - Rodrigo Agerri

Premios y menciones

- **ENEKO AGIRRE**, junto con el grupo de investigación IXA, recibirá uno de los premios de investigación de Google ([Google Faculty Research Awards](#)) de 80.000 dólares para desarrollar su investigación en inteligencia artificial, 'Accessing FAQ and CQA sites via dialogue' centrada en desarrollar sistemas de diálogo para poder conversar con las máquinas sobre la información existente en Internet. Más Información: [Campusa](#), [Diario Vasco](#)



- **EL GRUPO DE INVESTIGACIÓN IXA** obtiene una de las ayudas de investigación de La Fundación BBVA. De un total de 618 solicitudes, repartidas en cinco áreas, en concreto, en el área de la Biga Data, uno de los 5 proyectos seleccionados, de un total de 97 solicitudes, es el proyecto liderado por Eneko Agirre. El objetivo principal es utilizar Deep Learning y grandes colecciones de texto para transferir el conocimiento que ya existe en procesamiento de textos en inglés a otros idiomas, de forma que mejore la comprensión de textos en español, catalán, vasco y gallego. Esta área de investigación denominada Procesamiento del Lenguaje Natural forma parte de la Inteligencia Artificial, ya que pretende que los ordenadores lleguen a entender el significado de los textos. Esta área ha producido aplicaciones muy populares tales como la traducción automática o la búsqueda en Internet. Actualmente, el inglés es el idioma que más recursos tiene, es decir, existen programas que analizan esta lengua y extraen de grandes flujos de texto el conocimiento necesario para una aplicación en concreto. Este equipo de investigación quiere extrapolar esta comprensión al español, catalán, vasco y gallego. [Más información](#)

Los premios internacionales nos muestran que la investigación que se hace aquí es de alto nivel e interés

- **MIKEL ARTETXE** recibe una de las becas que concede Facebook para la realización de la tesis doctoral. Mikel está realizando el doctorado en la UPV/EHU, dentro del grupo de investigación IXA de la Facultad de Informática, tras cursar el máster en Análisis y Procesamiento del Lenguaje. Este año Facebook ha concedido 21 becas a doctorandos de entre más de 900 solicitantes de todo el mundo. La beca Facebook tiene una duración de dos años y está dotada con un aporte económico de 37.000 dólares anuales y una ayuda complementaria para la asistencia a congresos.



language /ˈlæŋɡwɪdʒ/
noun

1. the means of human communication, either spoken or written, consisting of the use of signs in a structured and conventional way.

technology /ˈteknɒlədʒi/ a non-verbal method of expression or communication

“La beca Facebook -explica Mikel Artetxe- supone un reconocimiento al trabajo realizado hasta ahora, así como un apoyo muy importante para continuar avanzando en mi carrera investigadora. A nivel personal estoy muy contento, pero no quiero perder la perspectiva de la precarización que sufre nuestro colectivo más en general. El mismo hecho de que sea una multinacional extranjera quien me va permitir continuar en unas condiciones dignas debería invitarnos a la reflexión.”

- **OLATZ PEREZ DE VIÑASPRE** logró el 8 de mayo, el VI Premio Koldo Mitxelena de Tesis en Euskera concedidos en la sede de Euskaltzaindia en Bilbao. Estos premios se dividen en cinco áreas de conocimiento: Ciencias; Ingeniería y Arquitectura; Ciencias de la Salud; Ciencias Sociales y Jurídicas y Arte y Humanidades. Olatz Pérez de Viñaspre, profesora e investigadora del grupo IXA, lo logró en el área de Ingeniería y Arquitectura.

En esta tesis se ha desarrollado y evaluado sistemas de traducción automática de términos en el dominio de ciencias de la salud. Para ello se ha partido de la SNOMED CT, ontología que engloba una terminología clínica amplia, y se ha desarrollado un sistema llamado EuSnomed para gestionar su traducción al euskera. EuSnomed implementa el algoritmo de cuatro pasos para obtener pagos en euskera de los términos: explotación de recursos lexicales, transliteración de términos neoclásicos (NeoTerm), traducción mediante términos anidados (KabiTerm) y adaptación a la medicina del traductor automático Matxin (MatxinMed).



- **AITOR ORMAZABAL** recibió el premio al mejor Trabajo de Fin de Grado en el congreso SEPLN2019 (Sociedad Española para el procesamiento del lenguaje natural) por su TFG "[Embedding eleartekoak sortzeko mapak eta metodoen mugen azterketa](#)" dirigido por los profesores Eneko Agirre y Aitor Soroa.



- JAVIER DE MARTIN GIL** ha recibido el segundo premio del "Premio Bizintek 2018 al Proyecto de Fin de Grado para Ingenieros" por su TFG "[Modelo basado en redes neuronales para la predicción de disponibilidad de un sistema urbano de préstamo de bicicletas](#)" dirigido por Ibon Saratxaga del grupo Aholab.



- La Universidad del País Vasco ha reconocido la trayectoria académica e investigadora de **INMA HERNAEZ** con motivo del Día Internacional de las Mujeres. Todas las premiadas son modelos de referencia en la universidad y en sociedad. El homenaje ha tenido lugar en el marco del Acto Institucional de la UPV/EHU con motivo del 8 de Marzo, Día Internacional de las Mujeres, organizado por la Dirección de Igualdad. [Más información](#)



Actividades

Seminarios

Estancias y visitas

Organización de workshops

Seminarios

1. *El Procesamiento del Lenguaje Natural (PLN) como recurso para el desarrollo de herramientas automáticas de ayuda a la redacción*

Iria Da Cunha

03/12/2019

2. *COST Action*

Itziar Aldabe

27/11/2019

3. *Métodos avanzados de Geoparsing para el reconocimiento de entidades nombradas y georreferenciación en texto libre*

Alejandro Molina

24/06/2019

4. *Automated and Assisted Content Scoring in Mono- and Cross-Lingual Educational Settings*

Andrea Horbach

17/09/2019

5. *Developing broad computational linguistic analyses of linguistic complexity - and using them to explore adaptivity in theory and practice*

Detmar Meurers

08/07/2019

6. *Deep Learning for Natural Language Processing*
Xavier Giro
05/07/2019
7. *EusDisParser*
Mikel Iruskiet
01/07/2019
8. *Deep Understanding for Automatic coding to the 2018 SEER/NAACCR Requirements and Comparison to Deep Learning Methods*
Jon Patrick
24/06/2019
9. *Social Psychology. Modern data analytics*
J.J. Pizarro, A. Mathias y L.N. Zumeta
18/06/2019
10. *DeepReading*
Rodrigo Agerri y German Rigau
17/06/2019
11. *Laburpen gaitasunaren garpena eta eskolako laburpen-testuen prozesamendua*
Unai Atutxa
03/06/2019
12. *Galderen sorkuntza eta testu ulermena*
Montse Maritxalar, Itziar Aldabe y Oier Lopez de Lacalle
27/05/2019
13. *Medikuntza bestelako lanak*
Aitziber Atutxa
20/05/2019
14. *Aholab signal Procesing Laboratory: Past, Present and Future activities*
Inma Hernaez
13/05/2019
15. *Euskarazko prentsa digitalerako garatutako hemeroteka-sistema*
Xabier Saralegi y Zuhaitz Beloki
06/05/2019

16. *Lingüistic Linked Open Data*

Arantxa Otegi

29/04/2019

17. *English-Irish Machine Translations*

Meghan Dowling

11/03/2019

18. *LIHLITH*

Arantxa Otegi

09/04/2019

19. *Lexicography and Digital Humanities*

Davide Lindemann

22/03/2019

20. *LINGUATEC*

Kepa Sarasola e Itziar Aldabe

04/03/2019

21. *Zertarako erabili dezakegu euskarako itzulpen automatikoa?*

Nora Aranberri

18/02/2019

22. *Itzulpen automatikoko proiektuak: Tadeep, Modena eta Domino*

Gorka Labaka

28/01/2019

23. *Lengoaia naturalaren prozesamendua eta medikuntza: gaztelaniazko txosten medikoen azterketa*

Koldo Gojenola

14/01/2019

24. *Osasun-txostenen itzulpen automatikoa euskararen eta gaztelararen artean*

Xabier Soto

07/01/2019

VISITAS

1. *Iria da Cunha*
Universidad Nacional de Educación a Distancia (UNED), Madrid, España
12/2019
2. *Alejandro Molina*
Consejo Nacional de Ciencia y Tecnología - CentroGeo, Yucatán, México
09//2019
3. *Andrea Orbach*
Language Technology Lab, Unversitat Duisburg Essen, Duisburg, Germany
09/2019
4. *Detmar Meurers*
Universität Tübingen, Tübingen, Germany
07/2019
5. *Jon Patrick*
Health Language Analytics, Sydney, Australia
06/2019
6. *Meghan Dowling*
Dublin City University, Dublin, Ireland
01/2019 - 04/2019

ESTANCIAS EN EL EXTRANJERO

1. *Olatz Arregi*
Queen Mary, University of London, UK
05/2019
2. *Xabier Soto*
Dublin City University, Irlanda
08/2019 - 10/2019
3. *Arantxa Otegi*
Université Paris-Sud, Francia
11/2019

4. *Jon Ander Campos*

Université Paris-Sud, Francia

11/2019 - 01/2020

ORGANIZACION DE WORKSHOPS

1. *Workshop on Discourse Relation Parsing and Treebanking*

Amir Zeldes, Debopam Das, Erick G. Maziero, Juliano A. Desiderato, Mikel Iruskieta

05/06/2019

2. *Dialogue Systems and lifelong learning satellite workshop Arantxa Otegi*

Université Paris-Sud, Francia

24/04/2019 - 26/04/2019

3. *First Crowdfest organized with the enetCollect COST action.*

Sophie Rosset, Eneko Agirre, Mark Cieliebak, Olivier Galibert, Sahar Ghannay, Arantxa Otegi, Anselmo Peñas, Camille Pradel, Anne Vilnat

23/01/2019 - 25/01/2019