

language /'læŋɡwɪdʒ/
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: Basque Center for Language Technology

Annual Report
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 /'sentrə/
noun

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

INTRODUCTION



Language Technology, the next frontier of Artificial Intelligence

language /'læŋɡwɪdʒ/
noun

Language understanding is one of the main goals of artificial intelligence, and although much progress has been made in deep learning, it remains one of the main limitations of artificial intelligence, as computers will not be intelligent until they understand language.

In order to face the new challenges and opportunities of the digital transformation and artificial intelligence, it is essential to foster research in language technologies.

To this end, we created HiTZ, with a dual purpose: to be an international reference in language technology research, and the reference in the computational treatment of Basque. The activity of the center comprises five lines of action: training, research, innovation and transfer, fostering our culture and societal impact. This wide scope requires a broad interdisciplinary team, which brings together computer scientists, linguists, engineers, translators and sociologists, among others.

The work of the researchers from Aholab and Ixa, the two research groups making the HiTZ center, already bore fruit in its first year. This solid foundation allows us to continue with renewed energies, and, in cooperation with technology centers, companies and institutions, turn the Basque Country into an international hub for language technologies and artificial intelligence.

Eneko Agirre (Director of HiTZ) and German Rigau (Deputy Director of HiTZ)

technology /tɛk'nɒlədʒi/
noun

HiTZ IN NUMBERS

37

Projects

25

Transfer Contracts

78

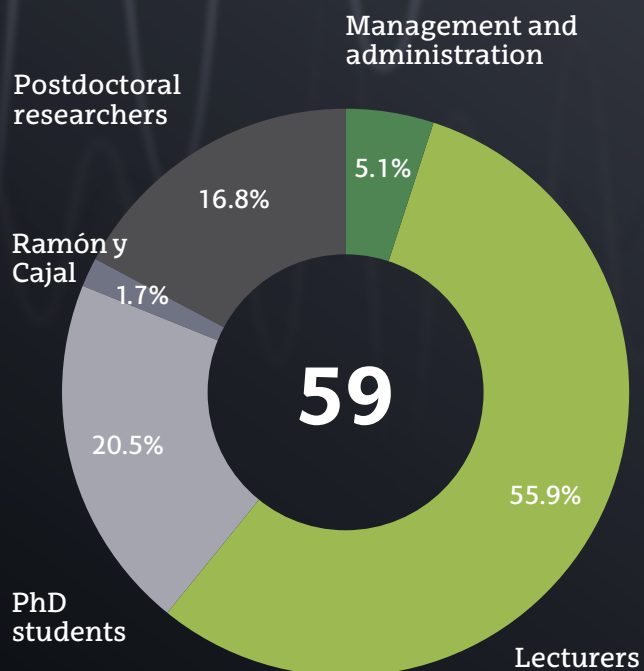
Publications

7

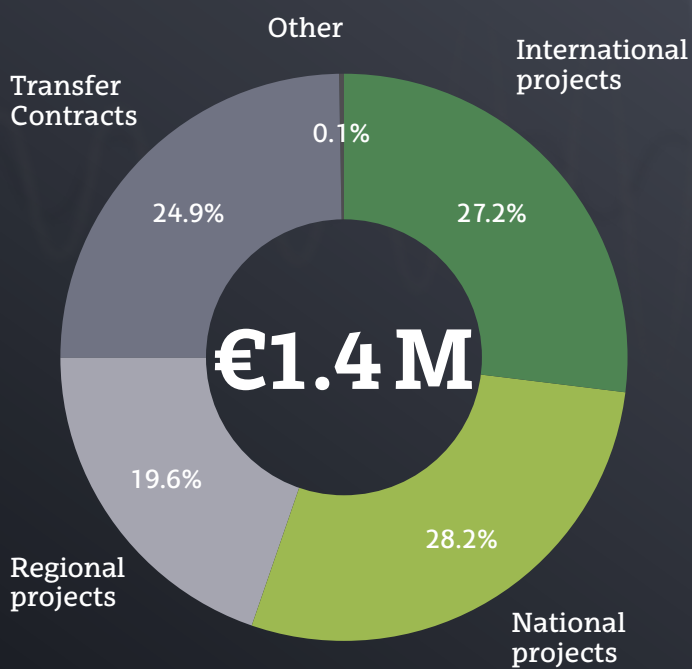
Thesis



People



Budget



ORGANIZATION

The research groups **Ixa** and **Aholab**, both from the University of the Basque Country (UPV/EHU), have been — since their creation in 1988 and 1998 respectively — the main tractors in the area of Language Technologies of the Basque Country.



ahōLAB



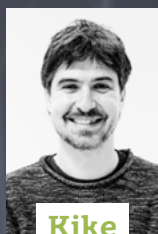
Eneko Agirre

Director



German Rigau

Subdirector



Kike Fernandez

Board member



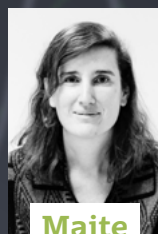
Inma Hernaez

Board member



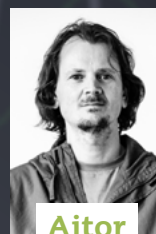
Oier Lopez de Lacalle

Board member



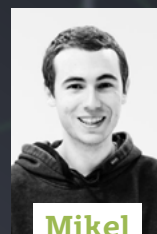
Maite Oronoz

Board member



Aitor Soroa

Board member



Mikel Artetxe

Board member

HiTZ is also a founding member of CLARIN-K Center (European Research Infrastructure for Language Resources and Technology), member of CLAIRE (key player on AI strategy in Europe) and full member of BDVA (Europe as the world leader in the creation of Big Data Value).



RESEARCH AREAS

Main Researcher:



**German
Rigau**



**Information
Extraction and
Information
Retrieval**



**Machine
Translation**

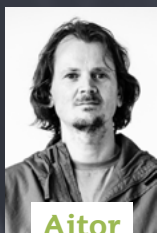


Main Researcher:



**Gorka
Labaka**

Main Researcher:



**Aitor
Soroa**



**Text
Analysis**



**Speech Synthesis
and Speech
Recognition**

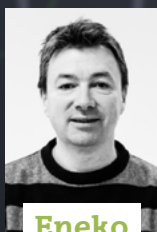


Main Researcher:



**Inma
Hernaez**

Main Researcher:



**Eneko
Agirre**



**Human-
Computer
Interaction**



**Speech and
Language
Resources**



Main Researcher:



**Ainara
Estarrona**

Main Researcher:



**Arantza
Casillas**



**Medical and
Legal domains**



**Digital
humanities
and education**



Main Researcher:



**Mikel
Iruskieta**

INFRASTRUCTURE

14

multiprocessor
GNU/Linux
servers

4

SPARC
Solaris
servers

1

HPC Cluster
with 128
cores

6

servers
with 27
GPUs

Over

120_{TB}

of raw Network
storage
capacity

1

acoustically isolated room
with audio equipment for
professional recordings

RESEARCH & TRANSFER



7

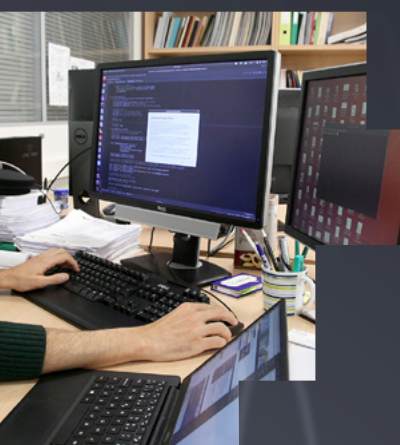
INTERNATIONAL
PROJECTS

12

NATIONAL
PROJECTS

18

REGIONAL
PROJECTS



25

KNOWLEDGE
TRANSFERENCE
CONTRACTS

3

OTHER
PROJECTS

7

DOCTORAL THESIS
DEFENDED
(3 International)

36

JOURNAL PAPERS
(11 Q1)

34

CONFERENCE PAPERS
(7 A, A+)

7

BOOK
CHAPTERS

1

BOOK

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

1. a system of communication used by a particular country or community.
- COMPUTING
a system of symbols and rules for writing programs or algorithms.
3. the style of a piece of writing or speech.

4

TRANSFERRED
TECHNOLOGIES

8

PHD GRANTS

TRAINING

training
/'treɪnɪŋ/
noun

15 ⓘ

STUDENTS IN
MASTERS

17 ⓘ

STUDENTS IN
DOCTORAL PROGRAM

5 ⓘ

EMLTC MASTER
THESIS
FINALIZED

10 ⓘ

HAP/LAP
MASTER THESIS
FINALIZED

57 ⓘ

STUDENTS IN 2 DEEP
LEARNING COMPLEMENTARY
COURSES

3

IKASIKER

8

INTERNAL AND
EXTERNAL INTERNSHIPS



ACTIVITIES



language
/ˈlæŋɡwɪdʒ/
noun



25 6

SEMINARS

INTERNATIONAL
VISITORS

technology

/tekˈnɒlədʒi/
noun

3

WORKSHOP
ORGANIZATION

1

GOOGLE FACULTY
RESEARCH
AWARD

OUTREACH: PAST PROGRAMS & RECIPIENTS
Faculty Research Awards Program (2005 - 2019)

The Faculty Research Awards Program aimed to recognize and support world-class faculty pursuing cutting-edge research in areas of mutual interest. In 2019 we deprecated this program, in order to diversify our efforts across a larger community, focusing on early career faculty and faculty working on positive societal initiatives. Below you can find all the FRA recipients funded from 2005 to 2019.

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 |

Christopher Potts, Stanford University
Dan Jurafsky, Stanford University
Eniko Agocs, University of the Basque Country
Karl Stratos, Rutgers University

1

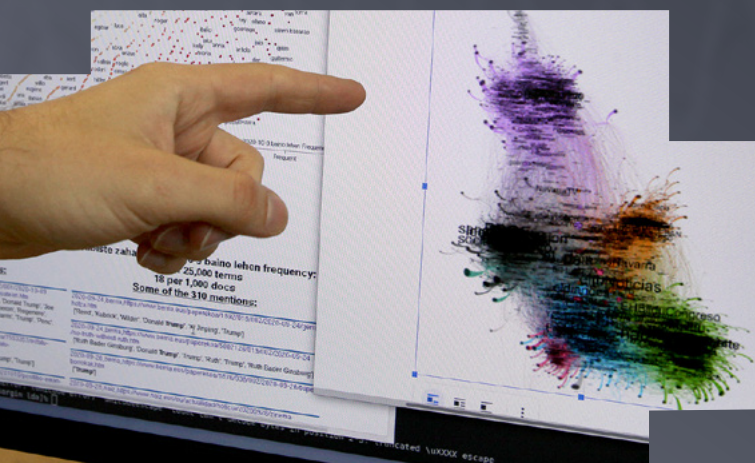
BBVA GRANT FOR
SCIENTIFIC RESEARCH
IN BIG DATA

1

FACEBOOK
GRANT

1

MITXELENA AWARD
FOR PHD THESIS



2

DEGREE/MASTER
THESIS PRIZES

1

UPV/EHU
TRIBUTE

center

/ˈsɛntə/
noun