Universidad del País Vasco
Euskal Herriko Unibertsitatea
eman ta zabal zazu

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

www.hitz.eus
@hitz_zentroa

Language technology, the next frontier of Artificial Intelligence

Annual Report
2020

INTRODUCTION

Language understanding is a main goal of artificial intelligence and, although much progress has been made in deep learning, it remains one of the main limitations of artificial intelligence because computers will not attain intelligence until they understand language. In order to face the new challenges and opportunities of artificial intelligence and the digital transformation, it is essential to foster research in language technologies. To this end, we created HiTZ with a dual purpose: to be an international referent in language technology research and to be the referent in the computational treatment of Basque. The center engages

in 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 that make up the HiTZ center, bore fruit this year. Of note were the successful creation of the first neural language models for Basque, leadership in the European Language Equality project that seeks to define a strategy for comprehensive access to digital language technology across Europe, advances in the European infrastructure CLARIN, and seven awards received at the most prominent international congresses and in various scientific competitions, including one organized by the United States government in response to the Covid-19 pandemic. This solid foundation allows us to continue with renewed energy 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)

HiTZ
IN NUMBERS

40
Projects

22
Transfer Contracts

71
Publications

3
Thesis

People ——— 106

Management and administration	3
Lecturers	39
PhD Grants	13
Postdoctoral researchers	4
Ramón y Cajal	1
HAP/LAP students	10
EMLTC students	2
PhD students	31
Doctoral theses defended	3

Budget ———

€1.25 M

Other — 2.6%

International projects — 32.8%

Transfer Contracts — 18.5%

Regional projects — 36.2%

National projects — 9.9%
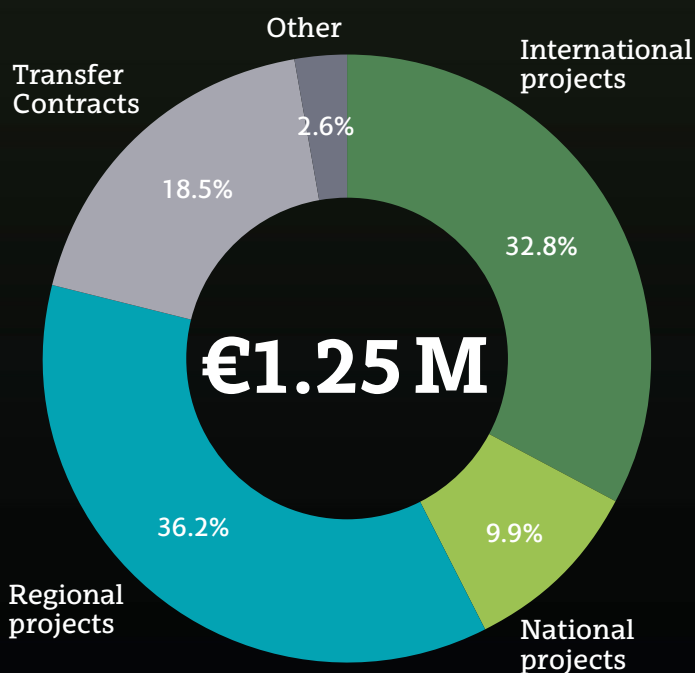
ORGANIZATION

HiTZ is a multidisciplinary research center on **Language-centric Artificial Intelligence** with members from seven departments of the University of the Basque Country. The objective of the center is to **investigate** language and speech technologies, with a significant effort towards the **transfer** of knowledge and technology to companies. It comprises two research groups Aholab and Ixa, both with extensive experience since 1993, performing basic research, creating linguistic resources and tools and launching different commercial products on the market. HiTZ is also a **founding member** of **CLARIN-K Center**, member of **CLAIRE** and full member of **BDVA** and **DAIRO**. Through CLAIRE and BDVA, we also participate in the European Partnership on Artificial Intelligence, Data and **Robotics**.

Eneko Agirre
Director

German Rigau
Subdirector

Ander Salaberria
Board member

Maite Oronoz
Board member

The members of the center are international referents in their scientific areas. At the moment, it is formed by **more than 60 members**, including computer scientists, linguists and 3 research technicians. In the last five years, the researchers now in the center have published more than 200 scientific publications. The group is a leader in applying deep learning techniques to language processing and its recent work in the area has been **cited more than 4,000 times** in the last two years.

The members of the center have been **advisors** in the creation of the National Plan for Spanish Language Technologies and are currently advising the Basque Government's equivalent counterpart.

CLARIN K CENTRE

CLAIRE

BDV BIG DATA VALUE ASSOCIATION

ixa

ahôLAB

Kike Fernandez

Board member

Itziar Aldabe

Board member

Inma Hernaez

Board member

Aitor Soroa

Board member

Both IXA and Aholab have been evaluated as high-performance research groups in the last research evaluation exercise by the science agency of the Basque Government. During their history, the groups have participated in more than 200 **research projects** ranging from regional to European projects. It has also participated in more than 100 **industrial contracts** with the aim of transferring technology into the industry.

HiTZ is also a member of **Erasmus Mundus+ European Masters Program** in Language and Communication Technologies (LCT) **program**. It is designed to meet the demands of industry and research in the rapidly growing field of language technology. HiTZ also offers a **Doctoral Programme** in Language Analysis and Processing.

The University of the Basque Country (UPV/EHU) is the leading teaching and research institution in the Basque Country, a prosperous region stretching along the Atlantic coast of northern Spain. The UPV/EHU is among the best 400 universities in the world according to the Shanghai ranking, and has been recognized as an International Excellence Campus by the Spanish Government. The University of the Basque Country, a vibrant 30-year-old institution with 45,000 students, 5,000 world-class academic staff and state-of-the-art facilities distributed throughout 20 centers in its three campuses.

RESEARCH AREAS

Click on this symbol
for more information

Information Extraction and Information Retrieval
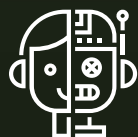
Main Researcher:

Aitor Soroa

Machine Translation

Main Researcher:

Gorka Labaka

Human-Computer Interaction

Main Researcher:

Eneko Agirre

Speech and Language Resources

Main Researcher:

Ainara Estarrona

INFRASTRUCTURE

14
multiprocessor
GNU/Linux servers

4
SPARC
Solaris servers

1
HPC Cluster
with 128 cores

Text Analysis
ⓘ

Main Researcher:

Rodrigo Agerri

Speech Technologies
ⓘ

Main Researcher:

Inma Hernaez

Medical and Legal domains
ⓘ

Main Researcher:

Arantza Casillas

Digital humanities and education
ⓘ

Main Researcher:

Mikel Iruskieta

1

acoustically isolated room with audio equipment for professional recordings

7 servers with

31

GPUs

Over

260 TB

of raw Network storage capacity

RESEARCH & TRANSFER ⓘ

40
Research projects

22
Knowledge transference projects

7
Other projects

3
Doctoral Thesis Defended

29
Journal papers
(8 Q1)

38
Conference papers
(8 A or A+)

1
Book chapter

3
Books

13
PhD grants

TRAINING

12 ⓘ
Students in masters

31 ⓘ
Students in doctoral program

2 ⓘ
EMLTC Master Thesis finalized

10 ⓘ
HAP/LAP Master Thesis Finalized

91 ⓘ
Students in 2 Deep Learning complementary courses

3
Ikasiker

8
Internal and external internships

ACTIVITIES

25
Seminars

3
Workshops

7
Awards

Google Faculty Research Awards

eHealth-2020 Tournament

CodiEsp-D. Main evaluation metric: Mean Average Precision

Winner of the Capitel@IberLEF 2020 shared task on Named Entity Recognition (NER) for Spanish

COVID-19 Open Research Dataset Challenge (CORD-19)

Honorable mention paper award (top 1%) at EMNLP (Conference on Empirical Methods in Natural Language Processing).

Outstanding Paper award (top 2%) at COLING (International Conference on Computational Linguistics).

BASQUE CENTER FOR LANGUAGE TECHNOLOGY