

Hizkuntzaren teknologia, adimen artifizialaren hurrengo jomuga

Urteko Txostena
2020



Hizkuntza ulertzea da adimen artifizialaren helburu nagusietako bat, eta nahiz eta ikasketa sakonari esker aurrerabide handia egin den, adimen artifizialaren ahulgune nagusietakoa izaten jarraitzen du. Izan ere, ordenagailuak ez dira adimentsuak izango harik eta hizkuntza ulertzen duten arte. Transformazio digitalaren eta adimen artifizialaren erronka eta aukera berrien aurrean, ezinbesteko bihurtu da hizkuntza teknologien ikerketan sakontzea. Horretarako sortu genuen HiTZ, helburu bikoitz batekin: hizkuntza teknologien arloan nazioarteko erreferentzia izatea, eta Euskararen tratamendu konputazionalen erreferentziatzko

zentro bihurtzea. Horretarako, zentroak bost ildotan du jarduera: prestakuntza, ikerketa, berrikuntza eta transferentzia, gure kultura bultzatzea, eta gizartean eragitea. Ereku zabal horrek disziplina arteko talde aniztuna eskatzen du, besteak beste, informatikariak, hizkuntzalariak, ingeniariak, itzultzaileak eta soziologoak biltzen dituenak. Hitz zentroa osatzen duten Aholab eta Ixa taldeetako ikertzaileen lanak, fruituak eman ditu aurtengoan ere. Azpimarratzekoak dira euskararako lehen hizkuntza eredu neuronal arrakastatsuak sortu izana, 2025erako hizkuntza teknologiarako estrategia definituko duen ELE proiektu europarraren lidergoan egotea,

CLARIN ikerkuntza-azpiegitura europarrean emandako aurrerapausoak, eta nazioarteko kongresu hoberenetan eta hainbat lehiaketa zientifikotan (COVID-aren inguruan Estatu Batuetako gobernuak bultzatu zuena barne) jasotako 7 sariak. Oinarri sendo horrek indarberrituta jarraitzeko aukera ematen digu, eta era berean, bestelako zentro teknologiko, enpresa eta erakundeekin elkarlanean, Euskadi hizkuntza-teknologiaren eta adimen artifizialaren nazioarteko erdigune bihurtzeko abagunea eskaintzen digu.

Eneko Agirre (HiTZeko zuzendaria) eta German Rigau (HiTZeko zuzendariordea)

HiTZ ZENBAKITAN

40

proiektu

22

transferentzia-
kontratu

71

argitalpen

3

tesi

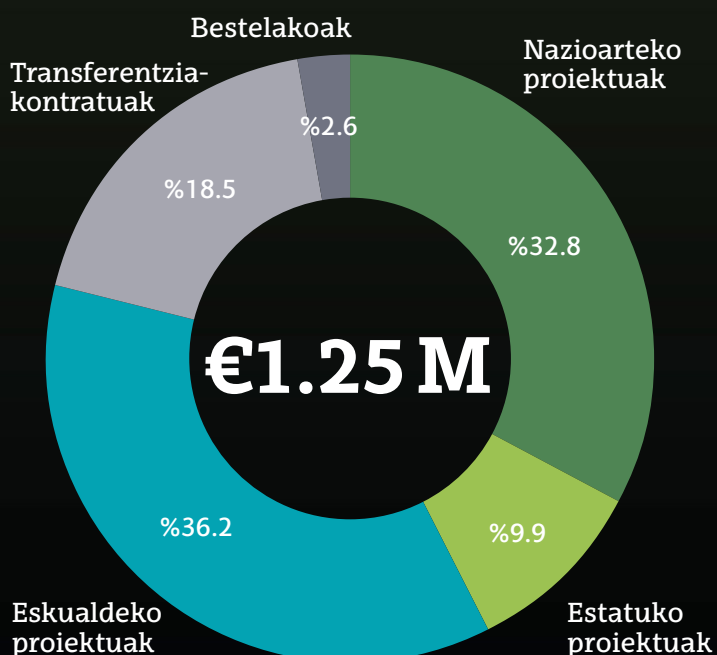
4



Kideak — 106



Aurrekontua —



ANTOLAKUNTZA

Hizkuntza ardatz duen adimen artifizialean diharduen ikerketa-zentroa da HiTZ. Euskal Herriko Unibertsitateko zazpi sailetakoko kideek osatzen dutelarik, diziplina anitzeko taldea da. Hizkuntzaren eta hizketaren teknologiak ikertzea da zentroaren helburua, enpresei ezagutza eta teknologia transferitzeko ahalegin nabarmena eginez. Bi ikerketa taldek osatzen dute: **Aholabek** eta **Ixak**. Bi taldeok esperientzia handia dute, 1993az geroztik oinarritzko ikerketa egin, hizkuntza-baliabideak eta -tresnak sortu eta hainbat produktu merkaturatu dituztela. **CLARIN-K Center**-eko kide sortzaileetako bat da HiTZ zentroa, CLAIRE-ko kidea da, baita BDVA eta DAIROko eskubide osoko kidea ere. Claire eta BDVAren bidez, European Partnership on Artificial Intelligence, Data and Robotics-en ere parte hartzen du.



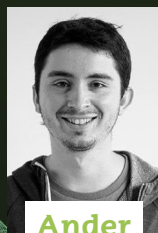
Eneko Agirre

Zuzendaria



German Rigau

Zuzendariordea



Ander Salaberria

Batzordeko kidea



Maite Oronoz

Batzordeko kidea

**CLARIN
K CENTRE**



CLAIRE



BDV

BIG DATA VALUE
ASSOCIATION

Zentroko kideak nazioarteko erreferenteak dira beren jakintza-arloetan. Gaur egun **60 kidek baino gehiagok** osatzen dute zentroa: informatikariek, hizkuntzalariek eta 3 ikerketa-teknikariek. Azken bost urteetan, 200 argitalpen zientifiko baino gehiago argitaratu dituzte zentroa osatzen duten ikertzaileek. Ikaskuntza sakoneko teknikak hizkuntzaren prozesamenduan aplikatzen liderra da taldea, eta azken bi urteetan, **4.000 aipu baino gehiago** izan ditu arlo horretan egin duen lanak. Zentroko kideek **aholkulari** gisa lagundu dute Espainiako Hizkuntzaren Teknologien Plan Nazionala sortzen, eta, gaur egun, Eusko Jaurlaritza aholkatzen ari dira antzeko lanetan.



ahōLAB



**Kike
Fernandez**

Batzordeko
kidea



**Itziar
Aldabe**

Batzordeko
kidea



**Inma
Hernaez**

Batzordeko
kidea



**Aitor
Soroa**

Batzordeko
kidea

Bai Ixa bai Aholab errendimendu handiko ikerketa-talde izendatu ditu Eusko Jaurlaritzaren Zientzia Kontseiluak, ikerketa ebaluatzeko azken ekitaldian. Historian zehar, taldeek 200 **ikerketa-proiektutan** baino gehiagotan parte hartu dute, eskualde-proiektuetatik hasi eta Europakoetaraino. Halaber, teknologia industriara transferitzeko helburuarekin, 100 **industria-kontratutan** baino gehiagotan parte hartu dute.

Erasmus Mundus+ European Masters Program Hizkuntzaren eta Komunikazioaren Teknologien Europako Masterreko (LCT) kide ere bada HiTZ. Industriak eta ikerketak hizkuntzaren teknologiaren alorrean dituzten

beharrak asetzeko diseinatuta dago Masterra, azkar hazten ari den esparrua baita. HiTZek, halaber, **Doktorego Programa** bat eskaintzen du Hizkuntzaren Analisi eta Prozesamenduz.

Euskal Herriko Unibertsitatea (UPV/EHU) Euskal Autonomia Erkidegoko irakaskuntza- eta ikerketa-erakunde nagusia da. UPV/EHU munduko 400 unibertsitate onenen artean dago, Shanghaiko rankingaren arabera, eta Espainiako Gobernuak Nazioarteko Bikaintasun Campus izendatu du. Euskal Herriko Unibertsitateak, 30 urteko erakunde biziak, 45.000 ikasle ditu, mundu mailako 5.000 langile akademiko eta abangoardiako instalazioak, bere hiru campusetako 20 zentrotan banaturik.



Informazioaren Erauzketa eta Berreskurapena



Ikertzaile nagusia:



Aitor Soroa



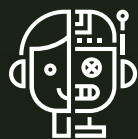
Itzulpen Automatikoa



Ikertzaile nagusia:



Gorka Labaka



Gizakiaren eta Ordenagailuaren arteko elkarreragina



Ikertzaile nagusia:



Eneko Agirre



Hizketa- eta Hizkuntza-Baliabideak



Ikertzaile nagusia:



Ainara Estarrona

AZPIEGITURA

14

GNU/Linux zerbitzari multiprozesadore

4

SPARC Solaris zerbitzariak

1

HPC Clusterra, 128 nukleoduna



Testuen Analisia



Ikertzaile nagusia:



**Rodrigo
Agerri**



Hizketa- teknologiak



Ikertzaile nagusia:



**Inma
Hernaez**



Hizkuntzaren Prozesamen- dua Medi- kuntzan eta Zuzenbidean



Ikertzaile nagusia:



**Arantza
Casillas**



Humanitate Digitalak eta Hezkuntza



Ikertzaile nagusia:



**Mikel
Iruskietia**

1

Akustikoki isolaturiko
gela, grabazio
profesionalerako audio-
ekipo batez hornitua

31

7 zerbitzari,
GPUdunak

260

TBtik
gora

sareko biltegiatze
ahalmen gordina

IKERKUNTZA ETA TRANSFERENTZIA



40

ikerkuntza
proiektuak

22

ezagutza
transferitzeko
kontratuak

7

bestelako
proiektuak

3

doktoretza-
tesiak
defendaturik

29

aldizkari-
artikuluak
(8 Q1)

38

artikuluak
biltzarretan
(8 A o A+)

1

liburu-
kapitulua

3

liburua

13

doktoretza-
bekak



12 ikasleak
masterrean31 ikasleak
doktorego
programanPRAKTIKAK 2 MALak amaituta
EMLCTn10 MALak amaiturik
HAP/LAPen91 ikasleak, sakoneko
ikaskuntza bi ikas-
taro osagarritan

3

Ikasikerrak

8

barne- eta
kanpo-praktikak

JARDUERAK



25

mintegiak



3

12
tailer
antolatuak

7

sariak

Google Faculty Research Awards

eHealth-2020

CodiEsp-D. Main evaluation metric:
Mean Average Precision

Capitel@IberLEF 2020 shared task on
Named Entity Recognition (NER) for
Spanish lehiaketaren irabazlea

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

Ohorezko aipamena EMNLP kongresuan

"Outstanding Paper" saria COLING-
en (International Conference on
Computational Linguistics)