

hizkuntza

Language

Centre Basque des Technologies du Language

Rapport annuel
2021

INTRODUCTION



words

language

2

hizkuntza

La compréhension de la langue est l'un des principaux objectifs de l'intelligence artificielle et, bien que de nombreux progrès aient été réalisés dans l'apprentissage profond, elle reste l'une des principales limites de l'intelligence artificielle, car les ordinateurs ne seront pas intelligents tant qu'ils ne comprendront pas la langue. Afin de faire face aux nouveaux défis et opportunités de la transformation numérique et de l'intelligence artificielle, il est essentiel d'encourager la recherche dans les technologies du langage. À cette fin, nous avons créé HiTZ, avec un double objectif: être une référence internationale dans la recherche sur les technologies linguistiques et la référence dans le traitement automatique du basque. L'activité du centre comprend cinq lignes d'action: formation, recherche, innovation et transfert, favorisant notre

center

```

def assign_position
  # ...
end

def assign_position
  # ...
end

def assign_position
  # ...
end

```

```

has_attached_file :photo
  belongs_to :team1, class:
  belongs_to :team2, class:
  belongs_to :division
  validates :team1_id, :te
  has_many :match_events

RESULT_TYPES = %w[usual

def members
  team1.members ++ team2
end

```

culture et notre impact social. Ce vaste champ d'action nécessite une vaste équipe interdisciplinaire, qui réunit des informaticiens, des linguistes, des ingénieurs, des traducteurs et des sociologues, entre autres. Le travail des chercheurs d'Aholab et d'Ixa, les deux groupes de recherche qui constituent le centre HiTZ, a déjà porté ses fruits cette année. Parmi les faits marquants, citons la création réussie des premiers modèles de langue neuronale pour la langue basque, la direction du projet ELE qui définira la stratégie en matière de technologies linguistiques en Europe, les progrès réalisés dans l'infrastructure de recherche européenne CLARIN, et enfin, les 5 prix reçus lors des meilleurs congrès internationaux et aussi dans divers concours scientifiques, dont celui promu par le gouvernement américain sur le COVID. Les

deux Prix Nationaux de Recherche en Informatique que nous avons reçus sont particulièrement remarquables, de même que la reconnaissance de notre directeur comme l'un des 15 membres que compte la principale association de recherche dans notre domaine en Europe. Enfin, il convient également de mentionner que l'un de nos docteurs a reçu le prix de la meilleure thèse en Intelligence Artificielle en Europe (EurAI). Cette base solide nous permet de poursuivre avec des énergies renouvelées et, en coopération avec des centres technologiques, des entreprises et des institutions, de faire du Pays Basque un pôle international des technologies linguistiques et de l'intelligence artificielle.

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

words

words

hizkuntza

HITZ EN CHIFFRES

Recherche et transfert

34

Projets

22

Contrats de
transfert

4

65

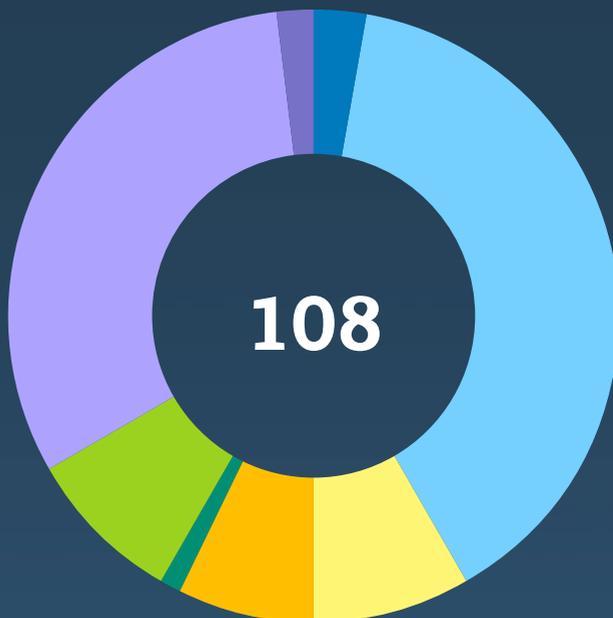
Publications

techn

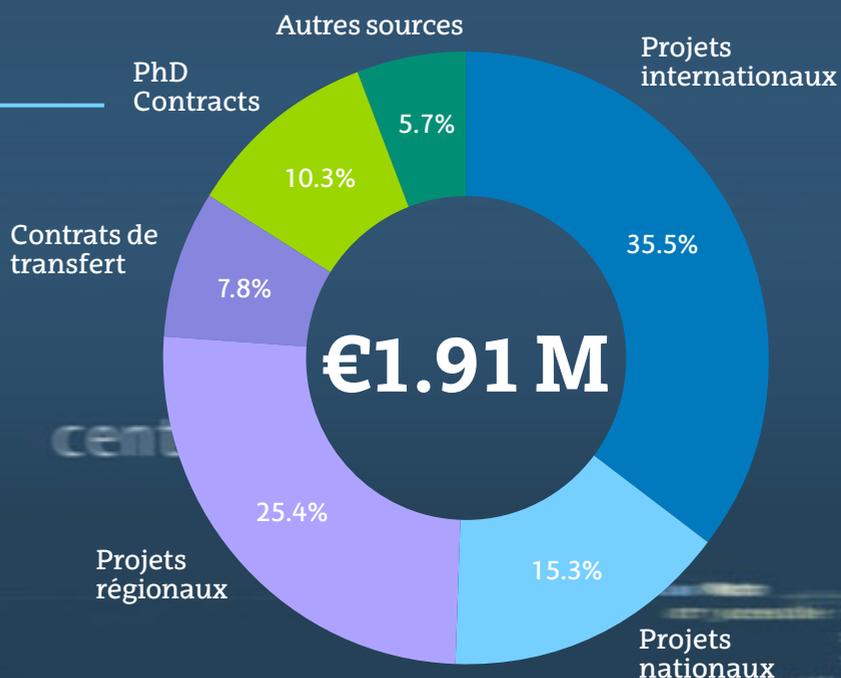


Personnes — 108

	Gestion et administration	3
	Enseignants	42
	Bourses de doctorat	9
	Chercheurs postdoctoraux	8
	Ramón y Cajal	1
	Étudiants HAP/LAP	9
	Étudiants en doctorat	34
	Thèses de doctorat soutenues	2



Budget



ORGANISATION

HiTZ est un centre de recherche multidisciplinaire sur **l'intelligence artificielle centrée sur la langue**, dont les membres proviennent de sept départements de l'Université du Pays Basque. L'objectif du centre est **d'étudier** les technologies du langage et de la parole, avec un effort important pour le **transfert** de connaissances et de technologies aux entreprises. Il comprend deux groupes de recherche, Aholab et Ixa, tous les deux dotés d'une grande expérience depuis 1993, qui effectuent de la recherche fondamentale, créent des ressources et des outils langagiers et lancent différents produits commerciaux sur le marché. HiTZ est également un **membre fondateur** du **Centre CLARIN-K**, membre de **CLAIRE** et membre à part entière de **BDVA** et **DAIRO**. Par l'intermédiaire de CLAIRE et de BDVA, nous participons également au European Partnership on Artificial Intelligence, Data and Robotics (partenariat européen sur l'intelligence artificielle, les données et la robotique).



Eneko Agirre

Directeur



German Rigau

Sous-directeur



Ander Salaberria

Membre du conseil



Maite Oronoz

Membre du conseil

Les membres du centre sont des référents internationaux dans leurs domaines scientifiques. Il est actuellement composé de **plus de 60 membres**, dont des informaticiens, des linguistes et 3 techniciens de recherche. Au cours des cinq dernières années, les chercheurs du centre ont publié plus de 200 publications scientifiques. Le groupe est un leader dans l'application des techniques d'apprentissage approfondi au traitement des langues, et ses travaux récents dans ce domaine ont été **cités plus de 4 000 fois** au cours des deux dernières années. Les membres du centre ont été **conseillers** dans la création du plan national pour les technologies de la langue espagnole et conseillent actuellement l'homologue du gouvernement basque.

**CLARIN
K CENTRE**

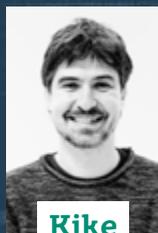


CLAIRE





ahō LAB



**Kike
Fernandez**

Membre du
conseil



**Itziar
Aldabe**

Membre du
conseil



**Inma
Hernaez**

Membre du
conseil



**Aitor
Soroa**

Membre du
conseil

Tant Ixa comme Aholab ont été évalués comme des groupes de recherche performants lors du dernier exercice d'évaluation de la recherche par l'agence scientifique du gouvernement basque. Au cours de leur histoire, les groupes ont participé à plus de 200 **projets de recherche** allant de projets régionaux à des projets européens. Ils ont également participé à plus de 100 **contrats industriels** dans le but de transférer des technologies dans l'industrie.

HiTZ est également membre du programme **Erasmus Mundus+ European Masters Program** in Language and Communication Technologies (LCT). Il est conçu pour répondre aux demandes de l'industrie et de la recherche dans le domaine en pleine expansion des technologies linguistiques.

HiTZ propose également un **programme doctoral** en analyse et traitement des langues.

L'Université du Pays basque (UPV/EHU) est le principal établissement d'enseignement et de recherche du Pays Basque, une région prospère qui s'étend le long de la côte atlantique du nord de l'Espagne. L'UPV/EHU fait partie des 400 meilleures universités du monde selon le classement de Shanghai, et a été reconnue comme un campus d'excellence international par le gouvernement espagnol. L'Université du Pays basque, une institution dynamique de 30 ans avec 45 000 étudiants, 5 000 membres du personnel académique de renommée mondiale et des installations de pointe réparties dans 20 centres sur ses trois campus.

DOMAINES DE RECHERCHE



**Extraction
et recherche
d'information**



Chercheur principal :



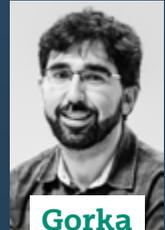
**Aitor
Soroa**



**Traduction
automatique**



Chercheur principal :



**Gorka
Labaka**



**Interaction
homme-
machine**



Chercheur principal :



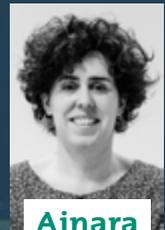
**Eneko
Agirre**



**Ressources
vocales et
langagières**



Chercheuse principale :



**Ainara
Estarrona**

INFRASTRUCTURE

15

serveurs
multiprocesseurs
GNU/Linux

4

serveurs
SPARC
Solaris

1

cluster HPC
avec 128
noyaux

1

salle acoustiquement isolée
avec équipement audio
pour des enregistrements
professionnels

 Si vous cliquez sur ce symbole vous pouvez obtenir plus d'informations



Analyse de textes



Chercheur principal :



Rodrigo Agerri



Technologies de la parole



Chercheuse principale :



Inma Hernaez



Domaines médical et juridique



Chercheuse principale :



Arantza Casillas



Sciences humaines numériques et éducation



Chercheur principal :



Mikel Iruskietia

hizkuntza

10 serveurs avec

45

GPUs

Plus de

260 TB

de capacité de stockage brute en réseau

1 1

Behringer 4x4 audio/MIDI interface

Quiet PC Sentinel Fanless i10

RECHERCHE ET TRANSFERT



34

Projets de
recherche

11

Contrats de
transfert de
connaissances

5

Autres
projects

2

Thèses de
doctorat
soutenues

25

Articles de
revue
(11 Q1)

36

Communications
de congrés
(8 A or A+)

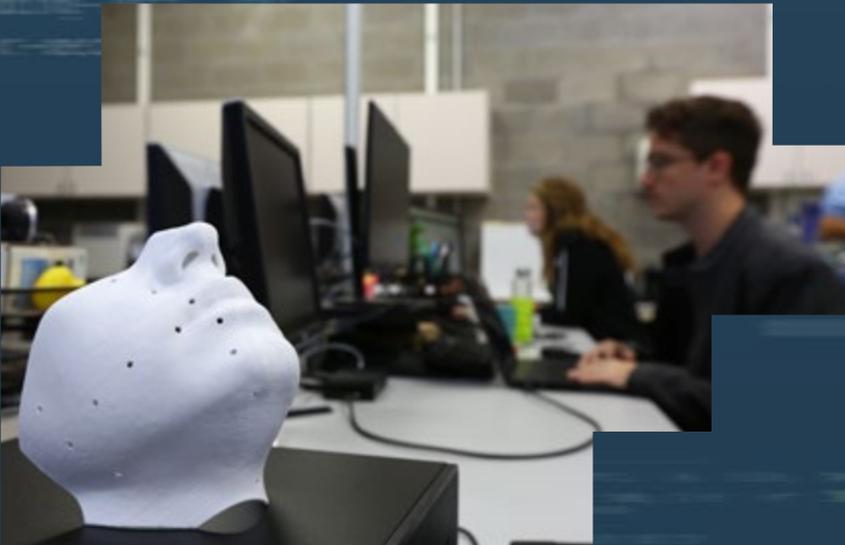
3

Chapitres
de livre

10

CENTRE BASQUE DES TECHNOLOGIES DU LANGUAGE

hitzak



Language

words



36 ⓘ

Étudiants en master

34 ⓘ

Étudiants en doctorat

FORMATION ⓘ

9 ⓘ

Thèses du master HAP/LAP soutenues

75 ⓘ

Étudiants dans 2 cours complémentaires d'apprentissage profond

6

Ikasiker

10

Stages internes et externes

ACTIVITÉS



40

Séminaires



12

Webinaires

7

Workshops

12

7

Prix

Fellow de l'Association of Computational Linguistics (ACL)

Prix de recherche national en informatique

Le prix des jeunes chercheurs en informatique

Prix SEPLN de la meilleure thèse de doctorat en traitement du langage naturel

2020 EurAI Doctoral Dissertation Award

Prix de recherche SCIE - Fondation BBVA: Jeunes Chercheurs en Informatique

Prix spécial Udalbiltza - Ikergazte 2021

CENTRE BASQUE DES TECHNOLOGIES DU LANGAGE

