



Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ

Hizkuntza Teknologiko Zentroa
Basque Center for Language Technology

www.hitz.eus

@hitz_zentroa



Centre Basque des Technologies du Language

Rapport annuel
2020

INTRODUCTION



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 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 neuronaux pour le 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 7 prix reçus lors des meilleurs congrès internationaux et aussi dans divers concours scientifiques, dont celui promu par le gouvernement américain sur le COVID. 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 (Directeur de HiTZ) et German Rigau (Directeur adjoint de HiTZ)

HiTZ EN CHIFFRES

40

Projets

22

Contrats de transfert

71

Publications

3

Thèses

4

CENTRE BASQUE DES TECHNOLOGIES DU LANGUAGE

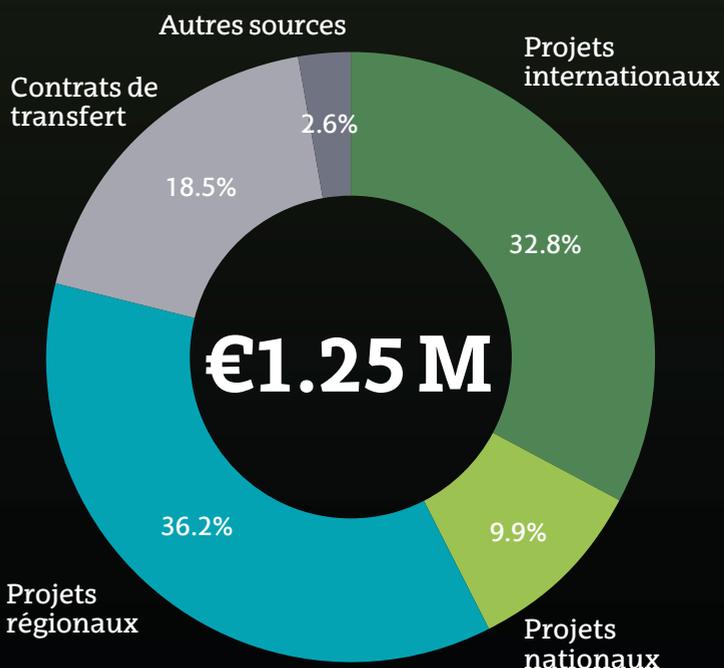


Personnes — 106

	Gestion et administration	3
	Enseignants	39
	Bourses de doctorat	13
	Chercheurs postdoctoraux	4
	Ramón y Cajal	1
	Étudiants HAP/LAP	10
	Étudiants EMLTC	2
	Étudiants en doctorat	31
	Thèses de doctorat soutenues	3



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



BDVA BIG DATA VALUE ASSOCIATION



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

14

serveurs multiprocesseurs GNU/Linux

4

serveurs SPARC Solaris

1

cluster HPC avec 128 noyaux

 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



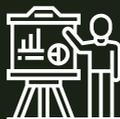
Domaines médical et juridique



Chercheuse principale :



Arantza Casillas



Sciences humaines numériques et éducation



Chercheur principal :



Mikel Iruskietia

1

salle acoustiquement isolée avec équipement audio pour des enregistrements professionnels

7 serveurs avec

31

GPUs

Plus de

260 TB

de capacité de stockage brute en réseau

RECHERCHE ET TRANSFERT



40

Projets de
recherche

22

Contrats de
transfert de
connaissances

7

Autres
projets

3

Thèses de
doctorat
soutenues

29

Articles de
revue
(8 Q1)

38

Communications
de congrès
(8 A or A+)

1

Chapitre
de Livre

3

Livre

13

Bourses de
doctorat





12

Étudiants en master



31

Étudiants en doctorat



FORMATION



2

Thèses du master EMLCT soutenues



10

Thèses du master HAP/LAP soutenues



91

Étudiants dans 2 cours complémentaires d'apprentissage profond



3

Ikasiker

8

Stages internes et externes

ACTIVITÉS



25

Séminaires



3

Workshops

7

Prix

Google Faculty Research Awards

Tournoi eHealth-2020

CodiEsp-D. Main evaluation metric:
Mean Average Precision

Gagnant de Capitel@IberLEF 2020 shared
task on Named Entity Recognition (NER)
for Spanish

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

Mention honorable pour le meilleur article
dans EMNLP (Conference on Empirical
Methods in Natural Language Processing)

Prix "Outstanding Paper" (top 2%) dans
COLING (International Conference on
Computational Linguistics)

