

HiTZ - Basque Center for Language Technology

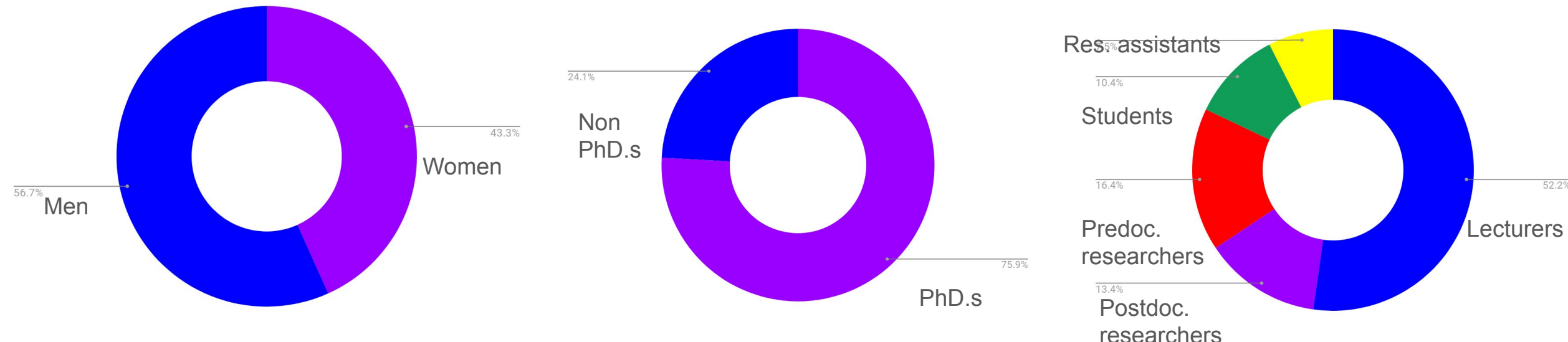
Natural Language Processing Made in Europe

hitz.eus
hitz@ehu.eus

- **industry**, including start-ups and big corporations
- **Multidisciplinary** staff (**67 researchers**): computer scientists, linguists, engineers, etc.

HiTZ Center

- Reference center on **Language Technologies** (LT)
- AIM: promote **research**, **training**, technological **transfer** and **innovation** in LTs
 - **Research breakthroughs** like unsup. Machine Translation, Semantic Textual Similarity
 - **Multilingual** resources and tools, including **low resource languages** like Basque
 - Broad portfolio of **transfer to**



Highlights

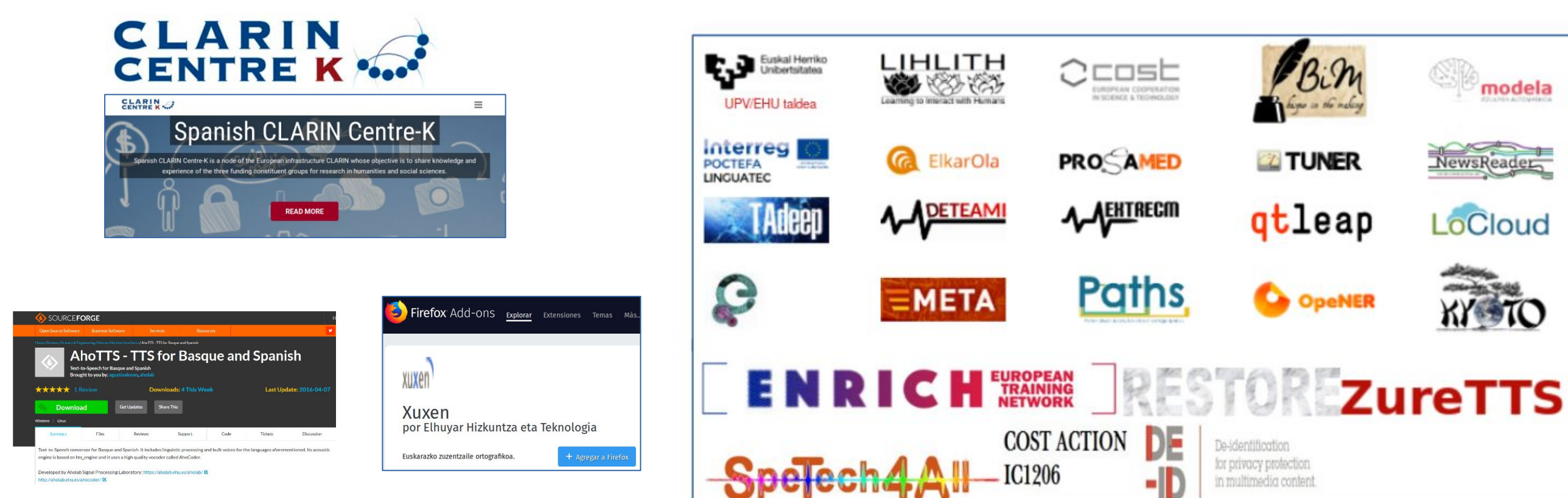
Human-centric AI: **Natural Language & Speech Processing**

- Language resources and **multilinguality**
- **Deep learning**
- **Big data** and **Social media** monitoring
- **IE & IR**
- (unsupervised) **Neural Machine Translation**
- **BioNLP** and LegalTech
- Dialogue systems, **chatbots**, QA
- **eLearning** and assistive technologies
- eXplainable NLP
- **Commonsense Reasoning**



Outcomes

- Published more than 500 scientific papers, 60 per year
- **CLARIN-K** centre
- **CLAIRE**, EUROPEANA, Flarenet and national research networks
- More than 15 national and regional awards, and 2 **Google Research Awards** (2016 & 2019)
- More than 30 public and private entities using our **products**
- More than 100 **projects** in 30 years: Paths, OpeNER, NewsReader, QTLeap, Lihlith, ...



Training

Master's degrees

- Language Analysis and Processing
- **Erasmus Mundus** Language and Communication Technologies in EU consortium

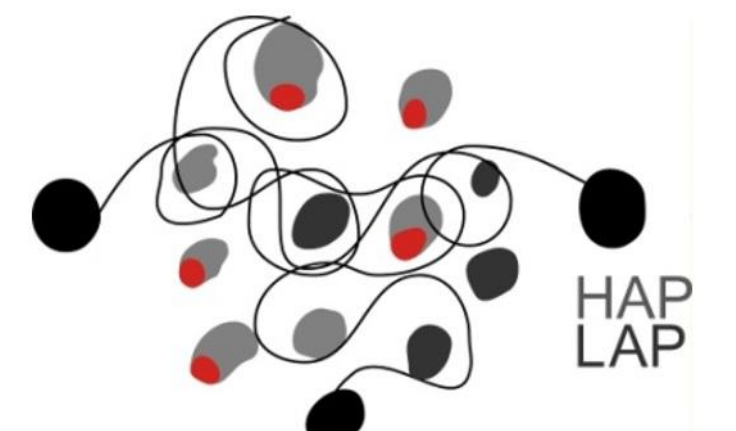
Doctorate

Language Analysis and Processing programme

- More than 60 **PhD.s**
- 25 ongoing

Life-long learning courses

Deep Learning for NLP & MT

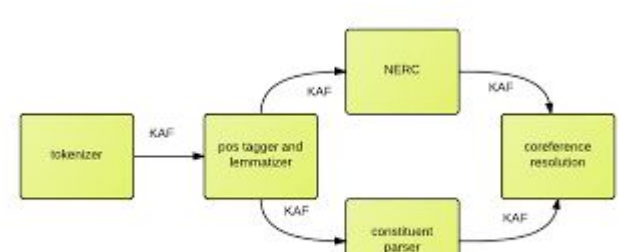


Ethics

- Effort on developing **open source software** and **resources**
- Turning **FAIR** (Findable, Accessible, Interoperable and Reusable) into reality

Our multilingual tools

Ixa pipes: text processing tools



MCR: multilingual knowledge bases

Multilingual Central Repository (MCR) - **WALSLEY**
It is a central repository for multilingual knowledge bases, providing a common platform for researchers to access and share data across different languages and domains.

TTS: Text to speech synthesiser



UKB: Graph Based Word Sense Disambiguation and Similarity

VecMap: cross-lingual word embedding mappings

FreelingMed: analysis + medical entity (body parts, disorders, drugs...) recogniser

ZureTTS Voice Banking: web interface to allow users to get their own synthetic voice

AHOcoder: a speech coder for statistical speech processing tasks

Industrial and public partners



AI Strategies

European AI plan

- Focusing on **Human-centric AI**, Cognition, NLP, Vision...
- Advanced Computing and **Big Data**

Spanish Plan for the Advancement of Language Technologies

- HiTZ members as advisors
- **Health** and Justice as main domains
- Focusing on **Iberian & Latin American languages**
- Promoting NLP, Machine Translation, Speech and Dialogue systems...
- Aims: technological transfer, improve linguistic infrastructures
- Public administration as main promoter
- 90M€ - 5 years



Basque Plan

- Being designed by HiTZ members
- Coordinated with the Spanish Plan



eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

