

# HiTZ - Basque Center for Language Technology

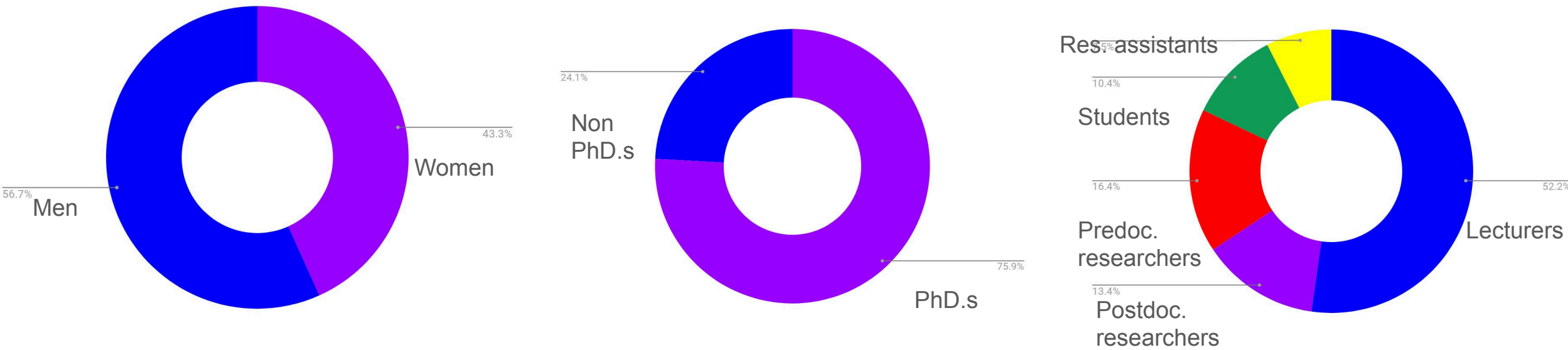
## Natural Language Processing Made in Europe

hitz.eus  
hit@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

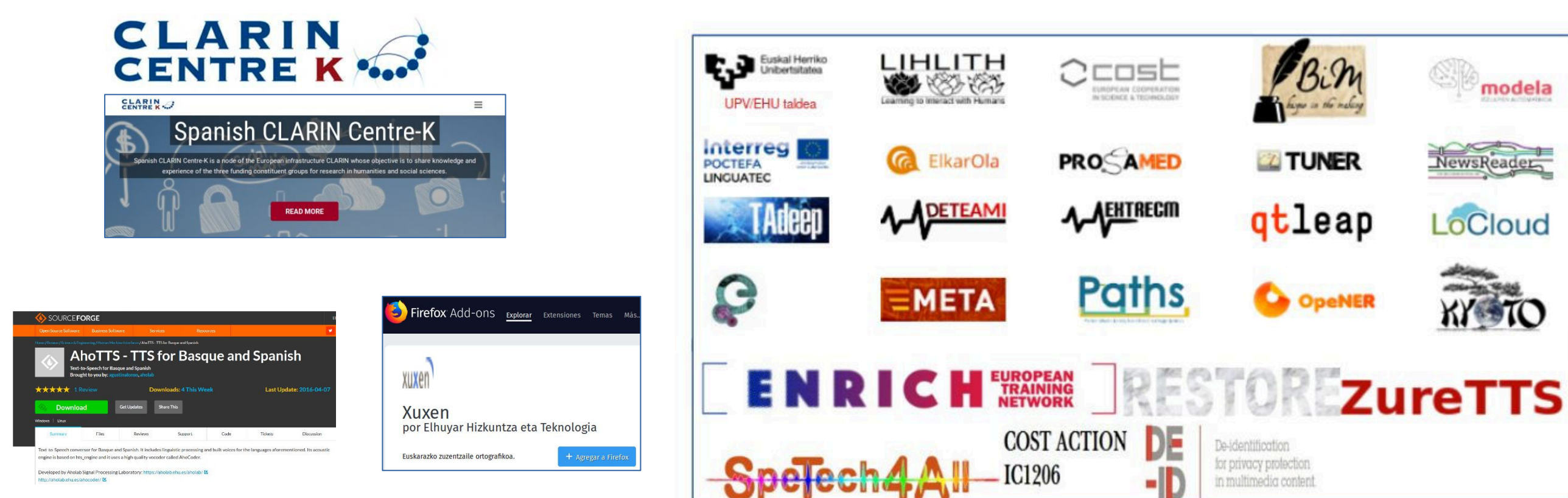
Human-centric AI: **Natural Language & Speech Processing**

- 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

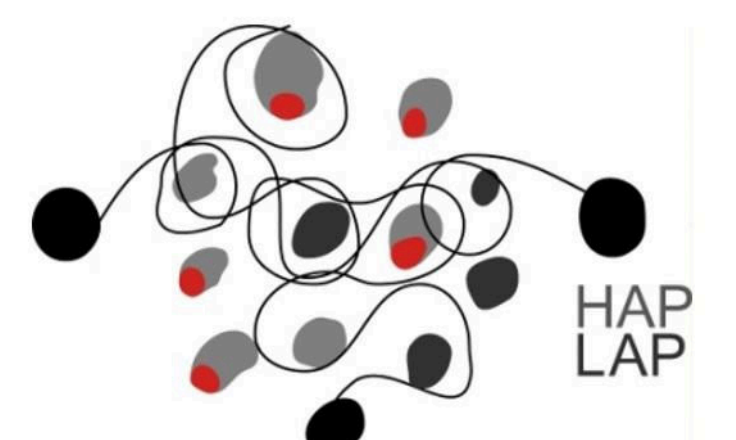
#### Doctorate

Language Analysis and Processing programme

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

#### Life-long learning courses

**Deep Learning** for NLP & MT

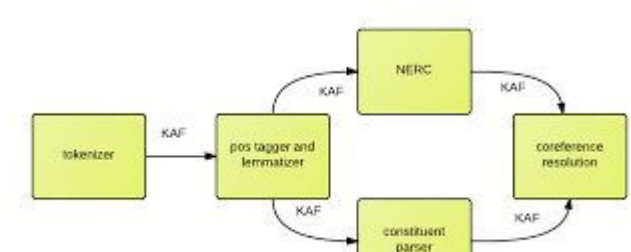


### Ethics

- 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

IB 30-04468005-n  
12 train\_1 railroad\_train\_1  
12 train\_2  
12 train\_3  
12 train\_4  
12 train\_5  
12 train\_6  
12 train\_7  
12 train\_8  
12 train\_9  
12 train\_10  
12 train\_11  
12 train\_12  
12 train\_13  
12 train\_14  
12 train\_15  
12 train\_16  
12 train\_17  
12 train\_18  
12 train\_19  
12 train\_20  
12 train\_21  
12 train\_22  
12 train\_23  
12 train\_24  
12 train\_25  
12 train\_26  
12 train\_27  
12 train\_28  
12 train\_29  
12 train\_30  
12 train\_31  
12 train\_32  
12 train\_33  
12 train\_34  
12 train\_35  
12 train\_36  
12 train\_37  
12 train\_38  
12 train\_39  
12 train\_40  
12 train\_41  
12 train\_42  
12 train\_43  
12 train\_44  
12 train\_45  
12 train\_46  
12 train\_47  
12 train\_48  
12 train\_49  
12 train\_50  
12 train\_51  
12 train\_52  
12 train\_53  
12 train\_54  
12 train\_55  
12 train\_56  
12 train\_57  
12 train\_58  
12 train\_59  
12 train\_60  
12 train\_61  
12 train\_62  
12 train\_63  
12 train\_64  
12 train\_65  
12 train\_66  
12 train\_67  
12 train\_68  
12 train\_69  
12 train\_70  
12 train\_71  
12 train\_72  
12 train\_73  
12 train\_74  
12 train\_75  
12 train\_76  
12 train\_77  
12 train\_78  
12 train\_79  
12 train\_80  
12 train\_81  
12 train\_82  
12 train\_83  
12 train\_84  
12 train\_85  
12 train\_86  
12 train\_87  
12 train\_88  
12 train\_89  
12 train\_90  
12 train\_91  
12 train\_92  
12 train\_93  
12 train\_94  
12 train\_95  
12 train\_96  
12 train\_97  
12 train\_98  
12 train\_99  
12 train\_100

**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

