

LANGUAGE TECHNOLOGY CHALLENGES OF A 'SMALL' LANGUAGE (CATALAN)

Maite Melero⁽²⁾, Gemma Boleda⁽¹⁾, Montse Cuadros⁽¹⁾, Cristina España-Bonet⁽¹⁾,
Lluís Padró⁽¹⁾, Martí Quixal⁽²⁾, Carlos Rodríguez⁽²⁾, Roser Saurí⁽¹⁾

⁽¹⁾ Centre de Recerca TALP

⁽²⁾ Barcelona Media

Universitat Politècnica de Catalunya

Centre d'Innovació

Speakers of 'small' languages may find themselves at a **comparative disadvantage** to speakers of bigger languages, given that companies often offer their LT products only in languages with large markets.

Official bilingualism in the Catalan territories entails that most residents are actually **bilingual**. There is a lack of interest from the industry to invest in Catalan, Spanish being a much more profitable market.

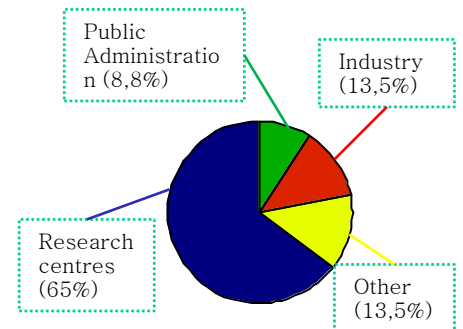


Good news! Most of the **basic LT tools and resources exist** but **research and development is often confined** to the **academic** world or to **altruistic** initiatives, with little connection with the real world. Oftentimes, the technology already exists, has been developed in some laboratory but needs the **final push** that would turn it into a real product.

Goals

1. Enhance communication and **collaboration** among developers from research groups, industry, and institutions.
2. Make a better and more **efficient use** of the already existing resources and tools
3. Give **visibility** to the research in the area.

Participation JPCC



Presented at JPCC:
14 research groups
+ 30 tools and resources

The 1st Workshop on the Computational Processing of Catalan (JPCC) (Barcelona, March 2009) succeeded in creating **community awareness**, provided us with a **snapshot of the state of affairs** in language technologies in Catalan and triggered a series of actions aimed at further **enhancing cohesion and communication** among the different actors in the Computational Processing of Catalan.



Visibility and **availability** of resources is the key issue for success. The "Harvesting Day" initiative wants to start a **decentralized effort** of resource description and launch an automatic, periodical information gathering routine. The collected data will build the seed of a **portal-catalogue-observatory of language resources in Catalan**.