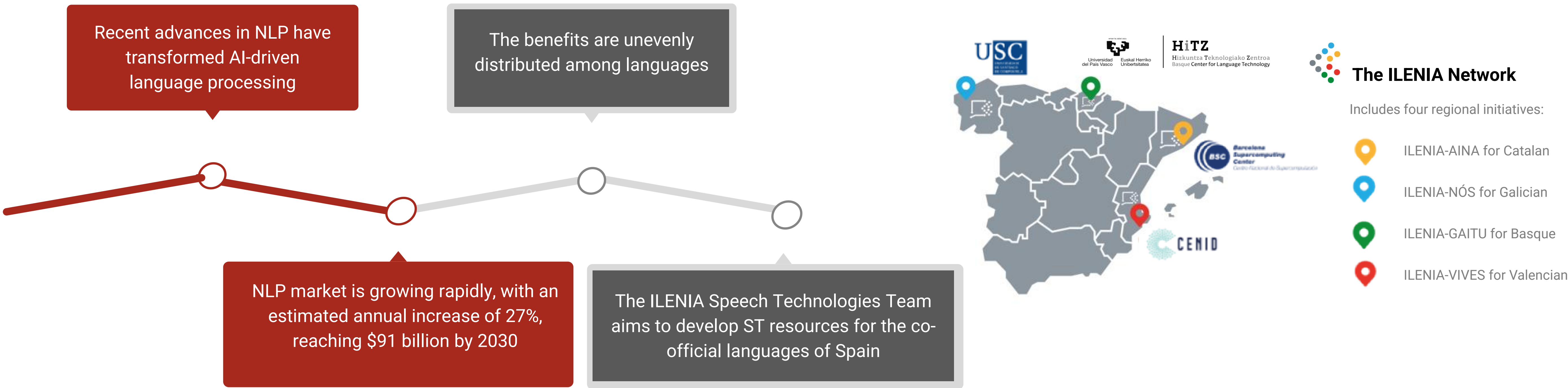




Speech Technologies in the ILENIA Project: Generating Resources to Develop Voice Applications in the Official Languages of Spain

B. Külebi¹, I. Hernández², E. Fernández Rei³, A. Montoyo⁴, S. Solito¹, C. Armentano-Oller¹, J. Hernando¹, E. Navas², C. Magariños³, A. Vladu³, I. Saratzaga², J. Sánchez², V. García Romillo², A. Herranz², C. Souganidis², N. García³, A. Moscoso Sánchez³, X.L. Regueira³, F. Dubert³, Y. Gutiérrez⁴

¹Barcelona Supercomputing Center, ²HITZ Center-Aholab, UPV/EHU, ³Instituto da Lingua Galega, USC, ⁴Dept. of Software and Computing Systems, UA



 Data Management Speech Models Model Evaluation Prototypes

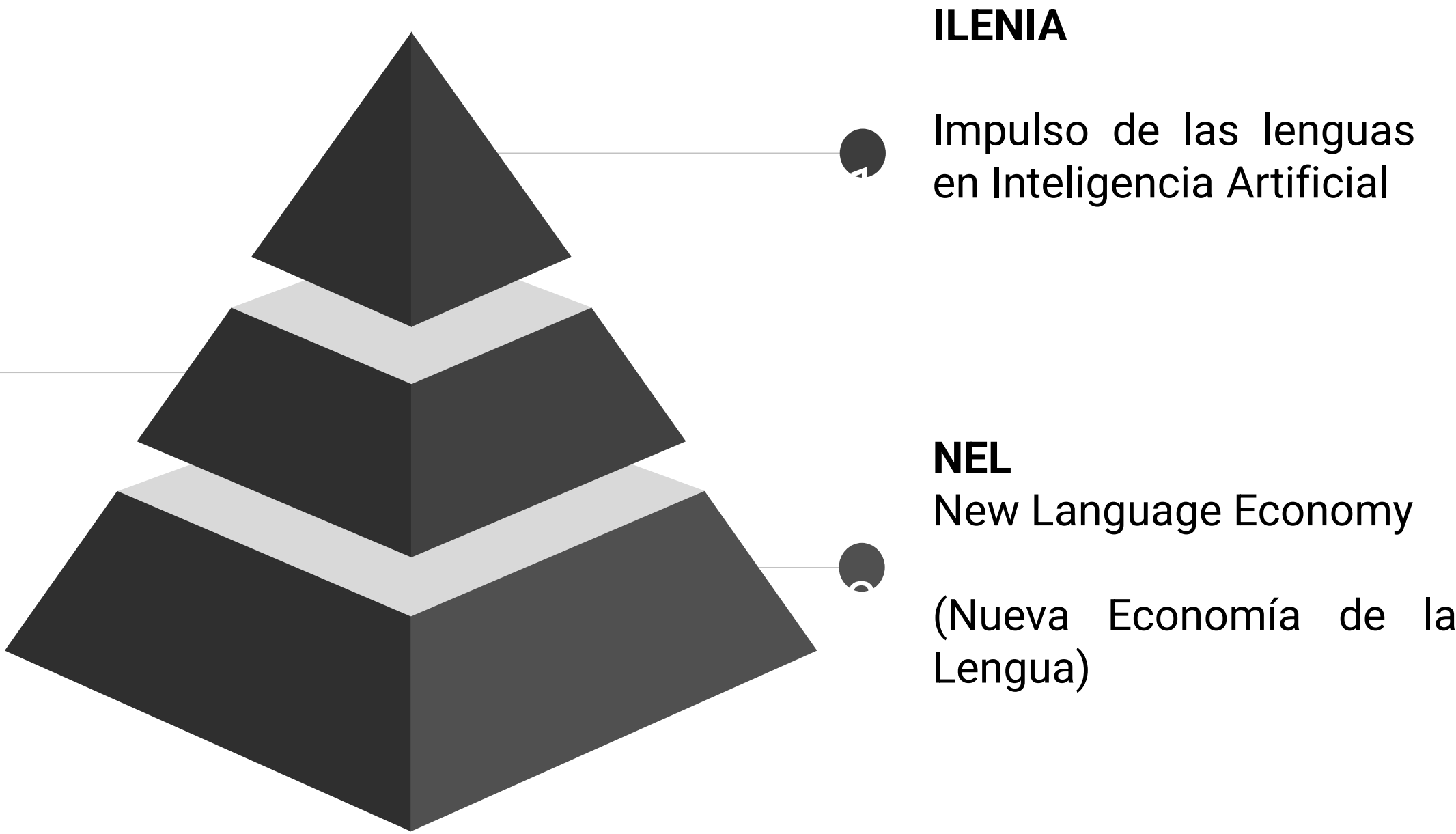
- **60,000+ Parallel Sentence Corpus**
Four Languages Aligned for Optimal Training
- **Catalan Multi-Accent Corpus**
Diverse accents captured from Parliamentary Sessions & YouTube Shows
- **Galician Media Corpus**
Sourced from TV, Radio & Podcasts for Authentic Language Use
- **Galician Text Corpus**
High-quality text for Read Speech Applications
- **Galician TTS Corpus**
Single-Speaker Database for TTS Modeling
- **700+ Hours of Basque Audio Corpus**

- **Multilingual TTS Model**
Supports TTS across Four Languages with Multiple Voices
- **Catalan Multi-Dialect ASR Model**
Tailored for Varied Catalan Accents
- **Basque ASR Model**
Optimized for Basque Language Recognition
- **Bilingual Basque-Spanish ASR Model**
Seamless Recognition Across Basque and Spanish
- **Single-Speaker Basque TTS Models**
High-Quality Speech Synthesis for Basque

- **Creation of Innovative Multilingual Benchmarks & Datasets**
Empowering Precise Evaluation and Advancement in Speech Model Performance Across Multiple Languages

- **Catalan and Galician Integration:** into the AhoMyTTS Speech Synthesizer and web portal, enhancing support for individuals with speech impairments (currently available in Basque and Spanish)
- **Open Voice OS Localization:** Adapting the Open Voice OS smart assistant to support all official languages of Spain, fostering inclusivity across digital interactions.

PERTE
Strategic Project for the Recovery and Economic Transformation
(Proyecto Estratégico para la Recuperación y Transformación Económica)



- **A common project, a shared Iberian vision**
- **Boost the Language-Centric AI Sector**
Elevating Spain's Competitiveness
Positioning Spain as a European and global leader in language-focused AI innovation
- **Facilitate AI Integration for Spanish Official Languages**
Accessible to All Businesses
Making AI profitable and appealing for companies of all sizes—from global multinationals to local SMEs
- **Empower Spanish Society in the Digital World**
Equal Language Access
Ensuring that every Spanish citizen can navigate and interact in any official language on par with leading world languages



Explore ILENIA Resources

Stay Connected with ILENIA

