



HAL
open science

LINGUATEC: Development of cross-border cooperation and LINGUATEC: Development of cross-border cooperation and knowledge transfer in language technologie

Kepa Sarasola, Jean-Baptiste Coyos

► **To cite this version:**

Kepa Sarasola, Jean-Baptiste Coyos. LINGUATEC: Development of cross-border cooperation and LINGUATEC: Development of cross-border cooperation and knowledge transfer in language technologie. Nexus Linguarum Cost Action 1rst planary meeting, 2020, Praga, Czech Republic. artxibo-02494780

HAL Id: artxibo-02494780

<https://artxiker.ccsd.cnrs.fr/artxibo-02494780v1>

Submitted on 29 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LINGUATEC (2018-2020)

LINGUATEC: Development of cross-border cooperation and knowledge transfer in language technologies

Itziar Aldabe, Josu Aztiria, Francho Beltrán, Myriam Bras, Klara Ceberio, Itziar Cortes, Benaset Dazeas, Erramun Osa, Kepa Sarasola, Aure Séguier, Jean Sibille, Joseba Zabaleta



IDIOMAS Y PARTICIPANTES

Elhuyar Foundation (coordinator)

Non-profit foundation.

Basque in specialized fields (journal, research books, dictionaries...).

Research and NLP applications

LO CONGRÈS (Pau)

Regulating agency for Occitan.

Computer tools (lexicography, lexicology, terminology, neology, phonology, orthography, grammar and toponymy).

CNRS (Toulouse)

Midi-Pyrenees Regional Delegation.

Linguistics and NLP for Occitan and French.
Building resources for Occitan language.

Aragón Turismo (Zaragoza)

Tourism Promotion and Management Society

Entity in charge of the Aragonese language.

Collaboration with the Aragonese Technological Institute (ITA)

Euskaltzaindia – Acad. of Basque Language

Language resources and consultation service on Euskara in the north of the Basque Country

IXA taldea – Univ. of the Basque Country

Advice/experience of 30 years in NLP
Improved Spanish-Basque translation
Translation system for Basque-French.



RESOURCES TO BE CREATED

BASQUE

- Speech recognition: word recognition and classification
- Speech synthesis in Northern Basque .
- Improved translation of the Spanish-Basque pair
- Interactive digital tool on the rules of Basque Language Academy (Northern dialects).

ARAGONESE

- Speech synthesis in Aragonese.
- Improved machine translation in Spanish-Aragonese
- Online dictionary of Aragonese

OCCITAN

- Monolingual Occitan lexicon with flexed forms
- Bilingual lexicon with common words in other languages
- Morphosyntactic analysis and corpus construction
- Parsing and dependency corpus
- Speech synthesis for mobile devices
- Occitan Text Detector
- Text detector of text in Occitan variants
- Improved French-Occitan machine translation

APPLICATIONS TO BE CREATED

Applications to facilitate interoperability between the languages of the Pyrenees.

Planned applications and responsible partner:

- LO CONGRES

Downloadable translation bar for Websites

- ELHUYAR

MT application for CMS

- ELHUYAR

Translation app between the languages of the Pyrenees:

Basque-French, Basque Spanish, French-Occitan and Spanish-Aragonese

- EUSKALTZAINDIA

Handbook of the Unified Basque. "Euskara Eskuz Esku Digitala".

- TURISMO DE ARAGÓN

On-line dictionary of Aragonese.

- LO CONGRÈS

Multilingual Semantic Search Engine.

Measuring the vitality of Occitan, Basque and Aragonese

Information and contact

> <https://linguatec-poctefa.eu/proyecto/>

> j.aztiria@elhuyar.eus