UniDive

2nd General Meeting

University of Naples "L'Orientale

Naples, Italy, 8-9 February 2024

https://unidive.lisn.upsaclay.fr/

ATHENA Research & Innovation Information Technologies

IEA LSP

The Philotis platform: Empowering low-resource languages processing

Vivian Stamou Vasileios Arampatzakis Dimitrios Karamatskos
Vasileios Sevetlidis Nicolaos Valeontis Stella Markantonatou George Pavlidis

Institute for Language and Speech Processing, Athena R.C.

{vistamou, vasilis.arampatzakis, dkaramatskos, vasiseve,
marks, gpavlid}@athenarc.gr, {nickvaleontis}@ssl-mail.com

Introduction

Philotis web-based platform:

- **Complete pipeline for recording and documenting living languages.**
- **Cutting-edge multi-modal technologies, integrating text, image, and audio data.**
- **Designed for linguists with varying technical expertise.**
- **Facilitates the development of both plain and annotated corpora from diverse multimodal sources.**

Key features

- **Language Agnostic:** Efficient functionality across linguistic variations.
- **Data Flexibility:** Selection of resources according to users' needs: enables multilingual model construction.
- **Writing System Support:** Supports the development of keyboards for different writing systems (via the Keyman service).
- **Active Annotation:** Supports incremental text fragment use for cycles of training and evaluation.

Resource Development

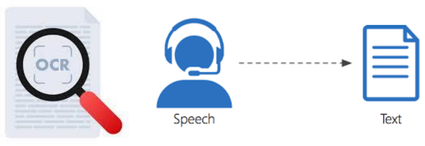
Metadata assignment

- Corpus name, language, dialect, description, license, administrator, and contact person.

Source Description

- Supports different modalities.

Optional processes:



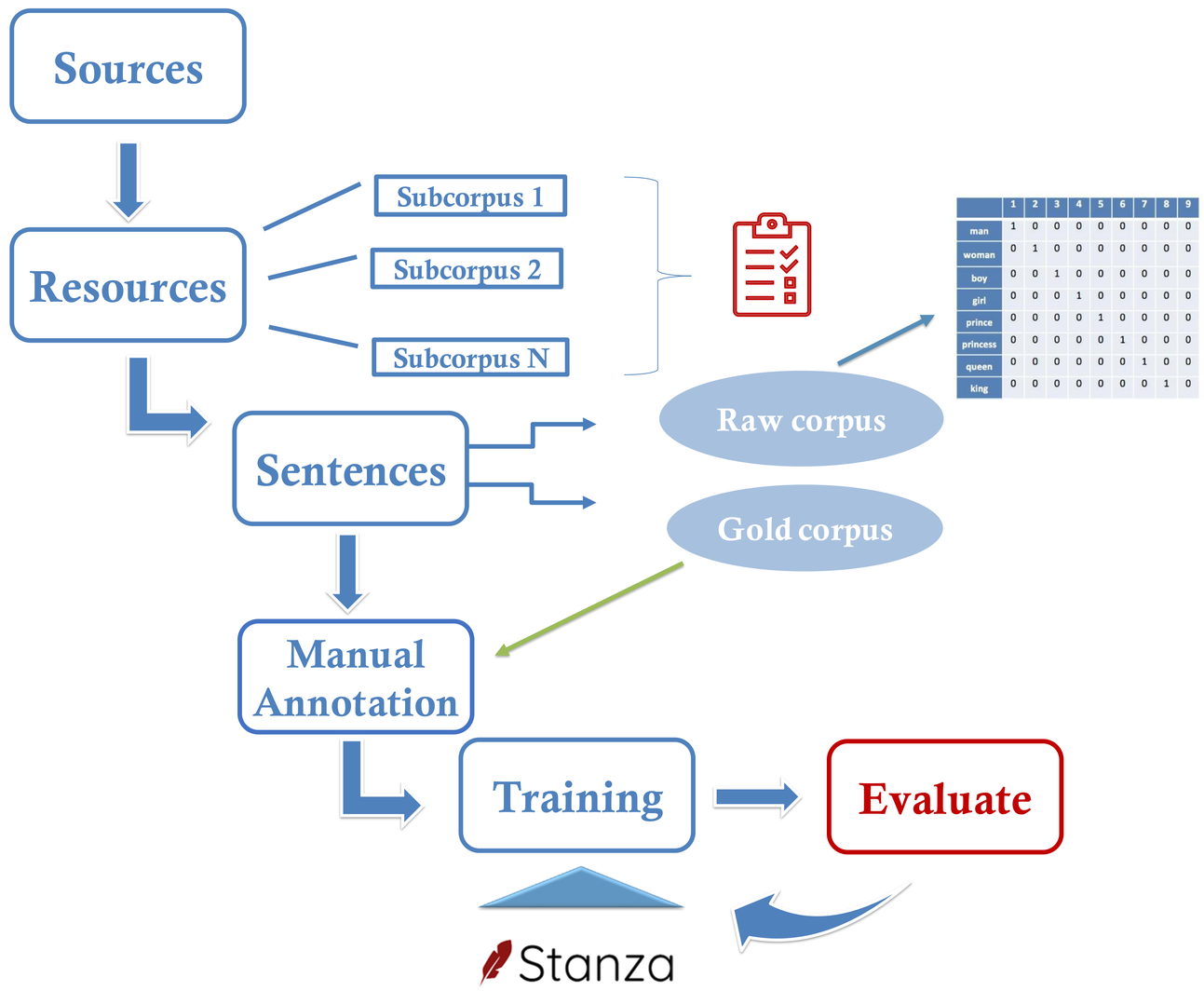
- **OCR: 'Tesseract-ocr'**
- **SST: 'Wav2vec XLS-R'**

- File Review Tab: Edit OCR and STT results manually.

Manual annotation

- According to the Universal Dependencies schema.
- Incorporates functionalities of the Arborator tool (CoNLL-U editing & visualization).

Pipeline



Backend

- Flask web framework for RESTful API endpoints.
- Docker platform for consistent behavior.

Frontend

- Developed using JavaScript/PHP webpages with Axios for handling asynchronous HTTP requests.
- Access to the database is handled by MySQL Workbench.

Funded by the European Union

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

WG3