

Herramientas digitales para el procesamiento y análisis de datos lingüísticos

Cristian R. Juárez

Instituto Federal de Tecnología de Zürich

juarezcristianr@gmail.com

1. Descripción

El objetivo de este taller es presentar y utilizar una serie de herramientas digitales, *Lameta*, *ELAN* y *FLEx*, diseñadas para la creación y el mantenimiento una base de datos lingüísticos. Vamos a trabajar sobre principios básicos para la sistematización de datos y su anotación mediante estas herramientas. Esto incluye el manejo de datos audiovisuales crudos hasta la creación de narraciones segmentadas y glosadas. Mediante la integración de estas herramientas digitales, vamos a lograr una base de datos robusta para tareas de largo plazo que son esenciales en prácticas de documentación y descripción lingüística. El corpus producido mediante estas herramientas será el insumo principal para el análisis lingüístico cualitativo y cuantitativo.

1.1. Repositorio en GitHub

Todo el material de este curso está disponible en el siguiente repositorio de GitHub: <https://github.com/JuarezRC/Herramientas-digitales-ling.git>

2. Cronograma de sesiones y actividades

Semana 1

- Lameta:
 - Creación de la colección principal.
 - Creación de sesiones y participantes.
 - Incluir contenido a las sesiones: título, palabras claves, descripción de sesión, etc.
 - Exportar base de datos.

Semana 2

- ELAN y Keyman:
 - Instalación de Keyman y el teclado AFI.

- Creación del archivo .eaf.
- Cómo crear segmentaciones y anotaciones.
- Cómo realizar búsquedas.
- Extraer audio y subtítulos.
- Integración con ELAN y FLEx.

Semanas 3

- ELAN - FLEx
 - Instalación de FLEx.
 - Funciones básicas de FLEX.
 - Exportar archivo ELAN a FLEx.
 - Segmentación y glosado en FLEx.
 - Creación de entradas léxicas y glosas.
 - Cómo exportar documento de FLEx.

Semanas 4

- FLEx - ELAN
 - Revisión de texto glosado y otras funciones de FLEX.
 - Cómo exportar documento de FLEx a ELAN.
 - Otros programas: IPAHelp, Phonology Assistant.

3. Datos y programas

Las sesiones de trabajo incluyen una breve descripción de cada herramienta digital seguida de práctica aplicada a datos lenguas Guaycurúes, mocoví/moqoit y toba. Si algunos participantes desean trabajar con sus propios datos, podríamos agregarlos al repositorio también. Cualquier sugerencia es bienvenida.

3.1. Lameta

Lameta (Hatton et al., 2021) nos ayuda a generar metadatos para mantener un registro exhaustivo de nuestra colección. Permite anotar detalles sobre el contenido de cada sesión de trabajo e incluir información relevante sobre participantes, además de muchos otros detalles que exploraremos en clase. Esta colección de sesiones forma parte de una base de datos fácil de exportar y manipular para la búsqueda de contenido específico.

- sitio web: <https://sites.google.com/site/metadatatooldiscussion/home/about-lameta>
- instalación: <https://github.com/onset/lameta/releases/tag/v2.3.16-beta>

3.2. Eudico Language Anotator (ELAN)

ELAN (Version 6.9) (2024) permite realizar anotaciones alineadas a audio y video. Las funcionalidades que ELAN brinda son múltiples, pero aquí solamente exploraremos las opciones de transcripción y traducción de los datos. Veremos también otras posibles integraciones con programas como Praat y su posible uso con R.

- sitio web: <https://archive.mpi.nl/tla/elan>
- instalación: <https://archive.mpi.nl/tla/elan/download>

3.2.1. Keyman

Es un teclado de acceso libre que permite utilizar caracteres del Alfabeto Fonético Internacional en cualquier otro programa o plataforma. Para entender cómo funciona este teclado, pueden leer más acá: https://help.keyman.com/keyboard/sil_ipa/2.0.1/sil_ipa#ipa109

- sitio web: <https://keyman.com/>
- descarga: <https://keyman.com/windows/>
- descarga del teclado AFI: <https://keyman.com/keyboards/h/ipa/>

3.3. Fieldworks Language Explorer (FLEX)

FLEX es un programa diseñado para la creación progresiva de diccionarios y análisis morfológico de textos. Nosotros vamos a utilizarlo principalmente para el análisis de textos y veremos de qué manera podemos integrar el resultado final a ELAN.

- sitio web: <https://software.sil.org/fieldworks/>
- instalación: <https://software.sil.org/fieldworks/download/>
- tutoriales: <https://software.sil.org/fieldworks/download/training-videos/>

3.4. Otros programas

- IPA Help: <https://software.sil.org/ipahelp2-1/>
- Phonology Assistant: <https://software.sil.org/phonologyassistant/>

Referencias

ELAN (Version 6.9) (2024). Max Planck Institute for Psycholinguistics, The language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>.

Hatton, J., G. Holton, M. Seyfennedipur, and N. Thieberger (2021). Lameta.