

# Lexicografía de idiomas poco documentados

Lev Michael y Christine Beier  
University of California, Berkeley

Pontificia Universidad Católica del Perú  
11 y 13 de abril, 2023

# Bienvenidos

- Día 2:
  - El papel central del análisis gramatical en la lexicografía de idiomas poco documentados
  - Software de base de datos para lexicografía: *Fieldworks Language Explorer* (FLEX)
  - Recomendaciones de flujo de trabajo para el trabajo léxico en idiomas poco documentados
  - Observaciones y recomendaciones sobre metas realistas en la lexicografía de idiomas poco documentados

# El papel del análisis gramatical

- La documentación léxica de alta calidad de un idioma es inseparable del análisis gramatical detallado del idioma.
- El análisis gramatical es necesario para:
  - determinar la forma fonológica correcta de la entrada (encabezado), la raíz, etc.
  - identificar correctamente la categoría gramatical de la entrada.
  - determinar la transitividad y estructura argumental de los verbos.
  - identificar las clases flexivas relevantes y determinar la pertinencia de los encabezados a estas clases.
  - identificar las alternancias y alomorfos relevantes.

# Información gramatical relevante a las entradas

- Las categorías y propiedades que son específicas al idioma deben incluirse, por ejemplo:
  - Clases flexivas de los verbos
    - El matsigenka cuenta con dos clases de verbo, “clase a” y “clase i”, con distintos patrones de flexión.
  - Clases nominales
    - El májìkì cuenta con tres clases de nombres: masculinos, femeninos y neutros.
  - Alternancias verbales como pares activo/medio
    - El iquito cuenta con más de 60 pares con este tipo de alternancia.
  - Alternancias sistemáticas en las formas de las raíces verbales
    - El iquito cuenta con más de 130 raíces que muestran una alternancia entre la forma flexiva y la forma derivativa.
  - Formas plurales irregulares
    - El iquito cuenta con 12 sufijos plurales y más de 140 sustantivos con formas plurales no predecibles.

## Las categorías gramaticales

- La categoría gramatical de cada entrada solo puede ser identificada de manera confiable sobre la base del análisis gramatical.
- Categorías mayores:
  - Ciertas categorías gramaticales, como las de verbos y sustantivos, se encuentran en todos los idiomas; otras categorías, como las de adjetivos y adposiciones, deben ser identificadas a base del análisis gramatical.
  - Sobre todo, incluso para categorías como sustantivos y verbos, se necesitan pruebas gramaticales explícitas para adjudicar la categoría gramatical de la palabra en los casos difíciles.

## Las categorías menores

- Todos idiomas exhiben subclases de las principales clases de palabras y es de crucial importancia distinguir entre ellas en un diccionario.
- Ejemplos:
  - sustantivos alienables y inalienables (matsigenka)
  - sustantivos locativos (iquito)
  - verbos primarios y secundarios (májjiki)

# Sustantivos alienables y inalienables en el matsigenka

- Los sustantivos inalienables pueden ser marcados como poseídos simplemente por agregar un prefijo posesivo:
  - no-gito  
1SG-cabeza  
'mi cabeza'
- Los sustantivos alienables requieren además un sufijo de posesión alienable:
  - no-tfago-re  
1SG-flecha-POS.ALIEN  
'mi flecha'

- Los sustantivos inalienables tienden a referirse a entidades como términos de partes del cuerpo y términos de parentesco, y los alienables a entidades que pueden separarse más fácilmente de sus poseedores.
- Sin embargo, no se puede confiar completamente en criterios semánticos; por ejemplo, 'flecha' es alienable, pero 'canoa' es inalienable:
  - no-bitó  
1SG-canoa  
'mi canoa'
- Por lo tanto, la distinción entre los sustantivos alienables y los inalienables debe identificarse mediante criterios gramaticales y esta característica debe indicarse en las entradas relevantes.



- También resulta que los sufijos de posesión alienable exhiben alomorfía condicionada por factores prosódicos y léxicos, lo cual implica que esta información también tiene que agregarse a las entradas relevantes.
- Los sufijos de posesión alienable:
  - -ne
  - -te
  - -re
  - -e

## Sustantivos locativos

- El iquito exhibe un conjunto de sufijos nominales que brindan información espacial de marcos de referencia absoluta relativa al centro deíctico (CD).
  - Son distintos de las posposiciones del idioma, y incluyen los siguientes:
    1. *-ku* 'arriba/rio arriba del CD'
    2. *-ma* 'debajo/rio abajo del CD'
    3. *-kúùrà* 'horizontal/perpendicular al río con respecto al CD'
- Por ejemplo, a base de la palabra *nasi* 'chacra' se puede formar:
  1. *nasiku* 'chacra río arriba del CD'
  2. *nasima* 'chacra río abajo del CD'
  3. *nasikúùrà* 'chacra perpendicular al río con respecto al CD'

- Significativamente, hay al menos 66 raíces para las cuales el uso de sufijos locativos es obligatorio.
  - Por ejemplo: *aji* 'boca de río', *araa* 'tronera', *iyi* 'puesto', *kuri* 'puerto', *niiku* 'camino', *papa* 'hueco de palo', etc.
- Como tal, este conjunto de raíces forma una subcategoría especial de sustantivos que debe indicarse en el diccionario.
- Para esta clase de palabra en el iquito, utilizamos el nombre "sustantivo locativo".

## La transitividad

- La información sobre la transitividad de los verbos es fundamental en la mayoría de los diccionarios.
- En algunos idiomas, sin embargo, la transitividad de los verbos no es sencilla.
- En májìkì, por ejemplo, hemos determinado que si bien hay muchos verbos que son estrictamente transitivos y otros que son estrictamente intransitivos, también hay muchos verbos ambitransitivos: es decir, verbos que pueden aparecer opcionalmente con un objeto — como ‘comer’ en castellano y ‘eat’ en inglés.
- Debido a la elipsis y la expresión cero de los argumentos tópicos en tercera persona en májìkì, determinar rigurosamente si un verbo es intransitivo, transitivo o ambitransitivo no siempre es simple.

- Afortunadamente, resulta que las construcciones de verbos seriales (CVS) en májìkì exhiben restricciones estrictas sobre la transitividad de los verbos que pueden combinarse en un CVS.
- Al usar un verbo de transitividad conocida, uno puede probar la transitividad de otros verbos formando CVS con ellos.
- Usando este método, hemos identificado muchos verbos estrictamente transitivos o intransitivos, pero también descubrimos que muchos verbos en májìkì son de hecho ambitransitivos.

## La estructura argumental verbal

- La manera en que los roles temáticos pueden ligarse a las relaciones gramaticales no es 100% predecible a base de la semántica del verbo.
- Un ejemplo clásico es la diferencia en la relación entre los roles temáticos y las relaciones gramaticales entre *gustar* del castellano y *like* del inglés.



## Software para bases de datos léxicos

- La documentación léxica y el desarrollo de un diccionario son tareas lo suficientemente complejas como para recomendar el uso de un software de base de datos.
- Si bien, en principio, uno puede usar un programa como Word para la lexicografía — en verdad, diccionarios han sido creados con eso, hasta con herramientas aún más primitivas — aumenta mucho la dificultad de la tarea y limita mucho lo que uno puede lograr.



## Software para bases de datos léxicos

- Como la mayoría de los lingüistas documentales, utilizamos Fieldworks Language Explorer (FLEX), un programa (aplicación) de base de datos diseñado para desarrollar corpus lingüísticos analizados y bases de datos léxicas.
- De manera similar, aunque es posible utilizar un programa de base de datos más general, como, por ejemplo, Filemaker Pro, el uso de un programa diseñado específicamente para la investigación lingüística tiene numerosos beneficios.
- Tanto FLEX como el resto del software y los programas (aplicaciones) que recomendamos son de código abierto y/o son gratuitos para obtener y usar, y han tenido un soporte estable durante décadas.

## Acciones básicas en el programa FLEx

- Abrir un proyecto:  
Archivo → Abrir → Elija un proyecto
- Respalidar un proyecto:  
Archivo → Administración de proyectos → Respalidar este proyecto → Guardar en: [ ]
- Seleccionar el idioma de la interfaz:  
Herramientas → Opciones → General → La interfaz de usuario (seleccionar)
- Cambiar fuente y tamaño:  
Formato → Estilos → Fuente (seleccionar)

# Acciones básicas en el programa FLEx

- Buscar una Entrada:  
Editar → Buscar entrada léxica → Writing System  
(seleccionar), → [ ]
- Crear nueva Entrada:  
Insertar → Entrada → Categoría (seleccionar), Forma Lexema  
[ ], Glosa [ ],
- Crear campo personalizado:  
Herramientas → Configurar → Campos personalizados →  
Añadir → Ubicación (seleccionar), Nombre de campo [ ]

## Partes clave del programa FLEx

→ Cada parte de FLEx ofrece múltiples funciones ubicadas al lado izquierdo; mencionamos las más imprescindibles:

- Léxico
  - Editar léxico
- Textos & Palabras
  - Textos interlineales
  - Concordancia
- Listas
  - Tipos de variante
- Gramática
  - Editar categoría
- Cuaderno

# Partes clave en la ventana de «Editar léxico»

- Lista de Entradas
  - Columnas
  - Filtros
  - Ordenar la lista
- Entrada
  1. Nivel de Entrada
  2. Acepciones
  3. Variantes
  4. Alomorfos
  - Campos por defecto
  - Campos de notas
  - Mostrar campos escondidos
  - Campos personalizados

## Funciones básicas en «Editar léxico»

→ Ver la siguiente diapositiva

- Cómo seleccionar las columnas
- Cómo filtrar en las columnas
- Cómo mostrar la entrada en Concordancia
- Cómo mostrar una acepción en Concordancia
- Cómo crear campos personalizados
  - Nuestra reglita de oro: ¡solo **un** tipo de información por campo!

Haga clic aquí para seleccionar columnas

Haga clic aquí para mostrar la entrada en Concordancia

Haga clic en el encabezado de la columna para cambiar el orden de las palabras

Haga clic aquí para acceder a los filtros para cada columna

Editar léxico

Examinar

Diccionario

Recoger Palabras

Diccionario clasificado

Edición masiva de entradas

Léxicos inversos

Edición masiva de entradas inversas

Archivo Enviar/Recibir Editar Vista Datos Insertar Formato Herramientas Analizador Ventana Ayuda

Spanish

Mostrar campos escondidos

Palabra principal	Glosas (Spn)	Información gramatical (compl.)
anunamija = jata	Crear nueva entrada	Buscar
júttàà <sub>1</sub>	que	complementizer
= jina <sub>2</sub>	COMP.NF	complementizer
júttàà <sub>1</sub>	como.(similar.a)	conjunction
júttàà <sub>1</sub>	como.(dado.que)	conjunction
júttikàri <sub>5</sub>	cuando.conj	conjunction
yaamáákùjì	porque	conjunction
jáátààraata <sub>1</sub>	ya.que	conjunction
júttààraata <sub>1</sub>	como.(similar)	conjunction
júttààraata <sub>1</sub>	ya.que	conjunction
iyaamiáákùjì	porque	conjunction
júttikàrii	cuando.conj	conjunction
jáátààraataa <sub>2</sub>	como.ENCR	connective
júttàà <sub>2</sub>	cuando	connective
sààkaa	que.ENCR	connective
júttààraataa <sub>2</sub>	como.ENCR	connective
tii	COP	copular verb
jútti	COPX	copular verb
táánà	COP REP	copular verb

Forma del lexema *teu* **táàjà**

Tipo de morfema **stem**

Forma de cita *teu*

Tipo de variante **prepausal form**

Variante de **táà**

Componentes

Pronunciación *teu* **táà hà**

Patrón CV **evv.cv 2**

Tono **y**

Remisiones

JMS edits **JMS 01oct2020 T**

EKL edit **EVS 2019.08.02**

Ireg Poss

CMBcheck

Ireg Pl *teu*

Entry History

**CMB 12sep2015 CE, 2015CHCKJPI; 2015TD make prepausal var of tAa; LDM 14jul2016 WED; LDM 03aug2016 confirmed lx; LDM 26jan2020 táaja > táàjà**

Deriv Root *teu*

ActiveMiddle

Residuos de import

Fecha de creación **sábado, 12 de septiembre de 2015 20:03**

Ordenadas por Grammatical Info (Full)

342/55

## Funciones clave en «Editar léxicos»

# Funciones básicas de «Textos & Palabras»

- Textos interlineales
  - Info
  - Texto básico
  - Analizar
- Concordancia
  - Especificar los criterios
  - Resultados
  - Enlaces a los textos de fuente



## Ayuda y capacitación en FLEx

- El programa incluye una sección de «Ayuda», pero está escrito en inglés.
- Hay materiales y videos de capacitación disponibles en inglés en:  
<https://software.sil.org/fieldworks/getting-help-and-training/>
- Hay algunos materiales de capacitación disponibles en castellano en: <https://mexico.sil.org/es/capacitacion/>
- Siempre podemos hacer una búsqueda en Google para ver si hay informaciones disponibles en el internet.

## Alternativas a FLEx

- Una pregunta común es: ¿Existen buenas alternativas a FLEx?
- La respuesta corta: No.
- La respuesta larga: FLEx tiene sus problemas y hemos visto una serie de iniciativas a fin de desarrollar alternativas (por ejemplo, LingSync). Sin embargo, ninguna ha sido ampliamente adoptada y ninguna se ha mantenido durante un período de tiempo significativo.
- Hasta ahora, FLEx es lo mejor que hay.

# Flujo del trabajo

- En esta sección, brindamos algunos consejos concretos sobre cómo administrar el trabajo en un proyecto en equipo que se lleva a cabo durante varias temporadas de campo.
- Específicamente:
  - El uso de los historiales de entradas
  - La verificación de consistencia entre entradas

## Historial de entrada

- La mayoría de nuestros proyectos de documentación léxica comparten dos características (que probablemente comparten con muchos otros proyectos de documentación):
  1. Involucran colaboraciones entre múltiples investigadores.
  2. Se llevan a cabo durante muchos años, a menudo con largos intervalos entre períodos de trabajo.
- Bajo estas circunstancias, es muy importante mantener un registro de los cambios realizados en una entrada; a ese registro lo llamamos el **«Entry History»** o **«Historial de Entrada»**.

## Historial de entrada

- En este campo (personalizado), comunicamos tanto a los otros miembros del equipo como a nosotros mismos en el futuro.
- Detallamos:
  - las preguntas, las dudas, etc. que tenemos sobre la entrada,
  - cómo hemos cambiado la entrada,
  - por qué hemos hecho estos cambios,
  - y sobre qué base empírica.
  - También, es el lugar donde indicamos el trabajo que aún queda por hacer.

# Ejemplo 1 de un «Historial de entrada»

## *m̄isaji*

- Noun
  1. woman
  2. female

LDM 23.09.2006 added lx [new lexeme Entry];  
LDM 23.10.2006 JPI now says 'tiene pausa' so recheck hw [headword];  
LDM 24.10.2006 mod [modified] hw *m̄isáji* > *m̄isáji*, expanded def [definition] with JPI, now 2 senses;  
BGG Praat [segment lengths annotated];  
CMB 12sep2015 RNLT [removed non-lexical tone mark];  
2015CHCK confirm tone is non-lexical;  
2015TD WED [write English definitions];  
LDM 25sep2015 conf [confirmed] non-lex tone, WED [wrote English definitions];

## *itimira*

- Noun
  - irregular plural of: *m̄isaji*

LDM 23.09.2006 added lx;  
LDM 28.09.2006 mod hw *itimira* > *itimira*;  
LDM 23.10.2006 conf lx with JPI;  
2015CHCK all;  
LDM 13oct2015 hw *itimira* > *itimira*, RNLT, WED;  
2015TD conf for humans only? yes: *pisiki m̄isajika*;  
LDM 15jul2017 ELY has n/poss [possessed vs. non-possessed] alt, nu-iltim+ra;  
LDM 10jun2018 confirmed, added to IrregPoss [field];  
2019CHECKJPI GrammNote: confirm and add what to use in non-human cases

# Ejemplo 2 de un historial de entrada

## **awasi**

- Noun  
digit, finger, or toe

LDM 05.09.2006 added lx;  
LDM 08.11.2006 lx def;  
LDM 06jun2016 JPI conf lx, irreg.pl.;  
LDM 18jun2016 JPI conf def, WED;  
LDM 28jul2016 ELY has toneless awAsi for nposs [non-possessed] form;  
2016TD add this variant;  
2016CHCKJPI to see if he has same? he does not;  
2017CHCKELY again on her variants;  
23jun2017 ELY confirms nposs form is toneless, poss form is Awasi;  
2017TD discuss w/CMB what to do in such cases where JPI does not show a variation;  
LDM 24jun2017 hw áwasi > awasi, created hw for poss form áwasi;  
2017CHCKELY plurals for poss, nposs forms;  
LDM 06jul2017 ELY has awasi no lextone for nposs, but nAwasi, nawAsika for poss;  
2017TD deal with how to annotate JPI variants? added to GramNote [field];  
LDM 15jul2017 ELY confirmed n/poss alt, added to IrregPoss [field];  
CMB 13nov2019 added 3.poss.Var [populated Variant field to generate a Minor Entry]

## Cómo mantener el «Historial de entrada»

- Cada vez que se hace un cambio significativo, se agrega un comentario al «historial».
- Usamos el siguiente formato:
- Primero van las iniciales de la persona que es autor del comentario, seguido por la fecha, por ejemplo:  
*LDM 13apr2023.*
- Después va un comentario que describe el cambio, la pregunta, la duda, etc. — con los iniciales del hablante con quien has trabajado, si es relevante.



## Trabajo para el futuro

- También, utilizamos el «historial de la entrada» para registrar o describir trabajo que queda para hacer.
- Para organizar tal trabajo, utilizamos dos códigos clave, los cuales ponemos antes de la descripción del trabajo pendiente:
  - CHECK: señala una tarea que requiere trabajo con un hablante.
  - TODO: señala una tarea que no requiere trabajo con un hablante.

## Trabajo para el futuro

- Utilizando estos códigos junto con la función de filtro de FLEx, podemos identificar todas las entradas que llevan CHECK o TODO.
- Después de que se haya resuelto la tarea del CHECK o TODO, borramos las vocales del código (CHECK > CHCK, TODO > TD).
  - Este cambio sirve para excluirlos de filtro en el futuro a la vez que mantiene un registro de todas las preguntas, cambios, tareas, etc. que ha sufrido la entrada.

# Ejemplo de los códigos en uso

## *m̄isaji*

- Noun
  1. woman
  2. female

LDM 23.09.2006 added lx [new lexeme Entry];  
LDM 23.10.2006 JPI now says ‘tiene pausa’ so recheck hw [headword];  
LDM 24.10.2006 mod [modified] hw *m̄isáji* > *m̄isáji*, expanded def [definition] with JPI, now 2 senses;  
BGG Praat [segment lengths annotated];  
CMB 12sep2015 RNL [removed non-lexical tone mark];  
2015CHCK confirm tone is non-lexical;  
2015TD WED [write English definitions];  
LDM 25sep2015 conf [confirmed] non-lex tone, WED [wrote English definitions];

## *itimira*

- Noun
  - irregular plural of: *m̄isaji*

LDM 23.09.2006 added lx;  
LDM 28.09.2006 mod hw *itímira* > *itimira*;  
LDM 23.10.2006 conf lx with JPI;  
2015CHCK all;  
LDM 13oct2015 hw *itimira* > *itimira*, RNL, WED;  
2015TD conf for humans only? yes: *pisiki m̄isajika*;  
LDM 15jul2017 ELY has n/poss [possessed vs. non-possessed] alt, nu-itim+ra;  
LDM 10jun2018 confirmed, added to IrregPoss [field];  
2019CHECKJPI GrammNote: confirm and add what to use in non-human cases

## Ejemplo del uso de CHECK en el «historial» de una entrada

## Verificaciones de consistencia entre entradas

- Por lo general, la manera de trabajar en un diccionario es entrada por entrada, donde completamos todos los distintos campos de una entrada, volviendo a ella de vez en cuando para hacer correcciones.
- En esta modalidad, es fácil que surjan diversos tipos de inconsistencias.
- A pesar de los esfuerzos por trabajar de manera consistente:
  - diferentes personas pueden trabajar de maneras ligeramente diferentes, tomando diferentes tipos de decisiones.
  - con el tiempo, una persona puede cambiar la manera en que trabaja.

## Comparaciones sistemáticas

- Uno de los mejores métodos de detectar y corregir inconsistencias de varios tipos es comparar sistemáticamente un solo campo en todas las entradas de la base de datos.
- FLEx facilita tales comparaciones al hacer posible seleccionar y ordenar un campo en particular en todas las entradas en la base de datos.
- Por ejemplo, en la siguiente imagen, tenemos columnas para los campos «Palabra principal», «Glosa (Spn)» y «Información gramatical (completa)» y luego hemos ordenado las Entradas por la columna de «Información gramatical»; esto nos permite comparar las formas categorizadas como “conjunction” y “connective”.

Archivo Enviar/Recibir Editar Vista Datos Insertar Formato Herramientas Analizador Ventana Ayuda

Spanish

**éxico** Mostrar campos escondidos

Palabra principal	Glosas (Spr)	Información gramatical (compi.)
nunamija = jata		
jáítàà <sub>1</sub>	que	complementizer
= jina <sub>2</sub>	COMP.NF	complementizer
jáítàà <sub>1</sub>	como.(similar.a)	conjunction
jáítàà <sub>1</sub>	como.(dado.que)	conjunction
jáítikári <sub>3</sub>	cuando.conj	conjunction
yaamááküji	porque	conjunction
jáátàaraata <sub>1</sub>	ya.que	conjunction
jáítàaraata <sub>1</sub>	como.(similar)	conjunction
jáítàaraata <sub>1</sub>	ya.que	conjunction
iyaamiááküji	porque	conjunction
jáítikárii	cuando.conj	conjunction
jáátàaraataa <sub>2</sub>	como.ENCR	connective
jáítàà <sub>2</sub>	cuando	connective
sáàkaa	que.ENCR	connective
jáítàaraataa <sub>2</sub>	como.ENCR	connective
tüi	COP	copular verb
jáíti	COPX	copular verb
ráámà	COP REP	copular verb

Forma del lexema **teu táàjà**

Tipo de morfema **stem**

Forma de cita **teu**

Tipo de variante **prepausal form**

Variante de **táà**

Componentes

Pronunciación **teu táà.hà**

Patrón CV **cvv.cv 2**

Tono **y**

Remisiones  
JMS edits **JMS 01oct2020 T**  
EKL edits **EVS 2019.08.02**

Ireg Poss

CMBcheck

Ireg Pl **teu**

Entry History  
**CMB 12sep2015 CE, 2015CHCKJPI; 2015TD make prepausal var of táà; LDM 14jul2016 WED; LDM 03aug2016 confirmed lx; LDM 26jan2020 táaja > táàjà**

Deriv Root **teu**

ActiveMiddle

Residuos de import

Fecha de creación **sábado, 12 de septiembre de 2015 20:03**

Ordenadas por Grammatical Info: (Full)

2/Sep/2015 02/May/2021 Cola: (-/-) No está cargado el analizador 342/513

## Filtro en «Información gramatical»

## Espacios en blanco

- FLEx también facilita la búsqueda de campos vacíos, lo que puede revelar información faltante.
- Buscamos los «espacios en blanco» en el menú desplegable del filtro en en campo relevante.
- En la siguiente imagen, hemos buscado las entradas que carecen de información en el campo «Glosa (spn)».
- Esto revela 36 Entradas que carecen de glosas en el castellano y nos permite averiguar si debemos agregar esta información.

Archivo Enviar/Recibir Editar Vista Datos Insertar Formato Herramientas Analizado Ventana Ayuda

Mostrar campos escondidos

Palabra principal	Glosas (Spn)	Información gramatical (abr.)
Mostrar todos	Espacios en blanco	Mostrar todos
= jáájà		
= jítíí		
isayíikíitííni		
árákà		
jiimàná		adj
suwaata		adv
suwa		adv
suwa		adv
jítíri		det
jimatííni		i.v
iwííni		i.v
kutitííni		i.v
tíitíírá		indef.pron
tíitíí		indef.pron
saakaaya <sub>2</sub>		interrog
kijawaja		intrj
kamjííita <sub>2</sub>		loc.dem.pron
namjííita		loc.dem.pron

Forma del lexema Igu **jáájà**

Tipo de morfema **enclitic**

Forma de cita Igu

Variante de Pronunciación

Remisiones JMS edits

EKL edits

Irreg Poss

CMBcheck

Irreg Pl Igu

Entry History **LDM 11mar2023 CE to account for one of the many jaas; 2023EXCLUDE**

Deriv Root Igu

ActiveMiddle

Residuos de import

Fecha de creación sábado, 11 de marzo de 2023 15:40

Fecha de modificac sábado, 11 de marzo de 2023 15:41

Acepción 1

Glosa Eng **EXH.PRED**

Spn

Qse

Biz

11/Mar/2023 11/Mar/2023 Cola: (-/-) No está cargado el analizador Ordenadas por Grammatical Info: (Abbr.) Filtrado 1/36

Espacios en blanco para las «Glosas (Spn)»



## Colaborando por FLEx

- En muchos proyectos de documentación léxica, la colaboración es imprescindible: simplemente hay demasiado trabajo para una sola persona.
- El deseo de colaborar enfrenta ciertos retos para el flujo de trabajo — sobre todo, cómo mantener el control de las versiones de la base de datos si hay más de una persona trabajando en ella.
- Si no es factible limitar el trabajo a una versión de la base de datos, ubicada en una sola computadora, entonces hay que utilizar algún método de control de las versiones.
- Felizmente, FLEx tiene una buena funcionalidad para eso: la función de «Send/Receive Project (Mandar/Recibir Proyecto)».



## Mandar/Recibir Proyecto por el internet

- Para fusionar versiones de FLEx por el internet, hay que registrarse y crear un proyecto en Language Depot, que es el sitio donde se sube las versiones distintas de FLEx para fusionarlas y sincronizarlas.
  - Language Depot funciona un poco como GitHub, pero es específicamente para proyectos de FLEx.
- Una vez que el proyecto y los usuarios son registrados en Language Depot, todos pueden utilizar la función de Mandar/Recibir para sincronizar todas las versiones individuales del proyecto.
- Siempre que usted y otros usuarios no hayan editado exactamente el mismo campo de la misma entrada, las dos versiones del proyecto FLEx se fusionarán sin problemas.

## Conflictos entre las versiones

- Generalmente, podemos evitar conflictos entre las versiones de FLEx por comunicar explícitamente entre todos los colegas sobre quién está haciendo cuáles cambios en cuáles campos de la base de datos.
- Sin embargo, con tiempo suficiente pueden surgir conflictos entre las versiones; en en tales casos:
  - FLEx conserva una de las versiones.
  - FLEx emite un mensaje de error que detalle cada conflicto, que permite al usuario escoger cual versión desea conservar de los datos en conflicto.

## Mandar/Recibir por USB

- En vez de utilizar internet, se introduce un USB en la compu y se hace clic en la opción de Mandar/Recibir, lo cual resulta que FLEx pone una versión de la base de datos en el USB.
- Sacando el USB, luego se introduce este USB en la otra compu con FLEx, se hace clic en la función de Mandar/Recibir y esto resulta en la sincronización del proyecto en la segunda compu.
- Finalmente, uno cierra el ciclo, introduciendo el USB en la primera compu, que resulta en la sincronización del proyecto en la primera compu.

## Para publicar su diccionario

- En diferentes tiempos, FLEx ha ofrecido diferentes opciones de exportación como bases para la publicación.
- Cuando empezamos a pensar en publicar el diccionario iquito-inglés (en 2017), las opciones de exportación de FLEx solo brindaron un control relativamente limitado sobre el orden y el formato de los campos, por lo que elegimos una opción ‘hecha en casa’:
  - Usamos un código de Python que ingiere el contenido de una exportación XML de FLEx y lo formatea como un archivo LaTeX, al que pudimos aplicar más operaciones de formato.
    - LaTeX es un lenguaje de marcado (*mark-up language*) para la preparación de documentos y es utilizado ampliamente en las ciencias y en ciertos campos de la lingüística.
  - Tiene la ventaja de permitir un control bastante fino sobre el formato, pero requiere que alguien sea capaz de escribir y modificar el script de Python y, por supuesto, que sea familiarizado con LaTeX.



# Publicar en línea: Dictionaria y Webonary

- También hay opciones para publicar diccionarios en línea:
  - *Dictionaria*: Una 'revista' dirigida por Martin Haspelmath y colegas que presenta diccionarios en forma CLDF para un público de lingüistas.
    - *Dictionaria* proporciona revisión por pares y transforma la base de datos de FLEx al formato CLDF requerido.
  - *Webonary*: Una iniciativa del ILV que publica diccionarios en formato HTML.
    - Hay opciones de exportación de Webonary en FLEx.



# Metas de los proyectos de documentación léxica

- En este cursillo, hemos intentado conseguir dos objetivos relacionados:
  - Compartir ideas y consejos sobre el desarrollo de recursos léxicos para idiomas poco documentados, basados en nuestra propia experiencia al realizar dicho trabajo.
  - Indicar formas en las que podemos, como una comunidad de investigadores, mejorar la calidad de la documentación léxica que producimos.
- A modo de conclusión del cursillo, queremos volver a la cuestión de los objetivos de los proyectos de documentación léxica.

- Idealmente, nos gustaría producir grandes diccionarios sobre la base de un excelente análisis gramatical, con muchos ejemplos y con extensas informaciones cultural y etnobiológica.
- Pero debemos recordar: **Lo perfecto es el enemigo de lo bueno.**

- Un diccionario “completo” es un trabajo de varias décadas.
- En muchos casos, tenemos que satisfacernos con “lo suficientemente bueno” — donde “suficiente” debe medirse con respecto a los objetivos del proyecto y del diccionario específico, por ejemplo:
  - lo que puede satisfacer a la comunidad para sus fines de revitalización o de educación.
  - lo que tiene cobertura satisfactoria para los estudios comparativos.
  - lo que llega a un punto de “saturación”: después de varios años de trabajo con los textos orales y la elicitación, dejarás (más o menos) de encontrar nuevas raíces o bases.

# Preguntas básicas al inicio de un proyecto lexicográfico

→ dan estructura al trabajo de investigación

→ dan estructura a la base de datos

- Marco de tiempo del proyecto
  - finito o delimitado
  - abierto
- Escala del proyecto
  - metas del investigador (o investigadores)
  - metas de los colaboradores de la comunidad
  - tipos de productos deseados
- Alcance del contenido del diccionario
  - tipos de información que cada entrada debe incluir
  - tipos de información adicional según el tipo de entrada
  - informaciones culturales, históricos, geográficos, sociolingüísticos, gramaticales, etc.

- En caso de que la meta de producir un diccionario comprensivo puede ser...
  - ...poco realista: no hay suficiente tiempo o interés...
  - ...innecesario: ya existe un diccionario adecuado...
- ...entonces, lo que puede constar de una contribución muy importante es un diccionario que se centre en un dominio semántico particular, por ejemplo:
  - La fauna y la flora (e.g., Zariquiey y Fleck (2014)).
  - El léxico de la percepción o de la emoción.
  - El léxico de las partes del cuerpo y la salud.

## Sugerencias concretas

- A base en nuestra experiencia con múltiples proyectos de documentación léxica, tenemos una serie de sugerencias para aquellos que están pensando en iniciar proyectos propios.

### 1. Trabajar en equipo:

- Más gente significa más horas de trabajo sobre el recurso léxico, lo que contribuye directamente a su alcance y calidad.
- Diferentes miembros del equipo pueden aportar diferentes habilidades e intereses al proyecto — por ejemplo, interés en morfosintaxis, en etnobiología, en la semántica léxica, etc.

2. Vincular el proyecto de documentación léxica a otros proyectos, porque aspectos de la lexicografía pueden surgir como productos de otros tipos de proyectos:

- Un estudio de la fonología de un idioma puede proporcionar información crucial para la representación de entradas.
- Buenas entradas sobre la flora y la fauna pueden resultar de un proyecto centrado en documentar el conocimiento etnobiológico.
- Buenas entradas sobre prácticas culturales tradicionales pueden resultar de un proyecto de revalorización cultural.
- Buenas descripciones de verbos que exhiben complementos clausales pueden resultar de una tesis sobre cláusulas complementarias.
- Un estudio sobre la posesión en un idioma puede arrojar luz sobre la (in)alienabilidad de una variedad de entradas en un diccionario.

3. Pensar en la documentación léxica como un proyecto a largo plazo, que corre paralelo a otras investigaciones sobre un idioma.
  
4. Pensar en la publicación de una variedad de productos intermedios, para que:
  - los miembros de la comunidad y otros investigadores no tengan que esperar una eternidad para beneficiarse de su trabajo.
  - puedes disfrutar de una sensación de progreso en medio de un largo esfuerzo.