White Paper: Jonathan D. Amith
Mesolex: Lexicosemantic Resources for Mesoamerican Languages
NEH ODH award HAA-266482-19 (Level 1)

(provisionally at https://mesolex-drako-iano9.ondigitalocean.app/en)

Section 1

Overview and introduction
The present project has grown from a multifaceted portal for lexicosemantic resources into an integrated content management system (CMS) for these same varied resources. The portal is designed to handle the following materials pertinent to the study of Mesoamerican languages and cultures:
- Online dictionaries (§2)
- Continuous line-by-line playback of audio files with associated transcriptions and translations (annotated in ELAN and converted to HTML with Mesolex software) (§3)
- Textual resources (a library of published and unpublished material) pertinent to each of the languages represented in Mesolex (§4)
- Video links (§4)
- Electronic audio books (in this example, an e-book on native plant knowledge; §5)

A manual for the use of each resource category is included in this white paper, each in its own section.

Philosophy of Mesolex
Mesolex is built on the realization that accessibility of Mesoamerican audiovisual and textual resources to communities of interest—scholars and students, native speakers and communities, the general public—is dispersed and deficient. Thus Mesolex, a multifaceted content management system, was created:
- to gather multiple resources relevant to any particular community in a single integrated location;
- to develop new software to facilitate discovery, display, and utilization of lexicosemantic resources, particularly Indigenous language dictionaries and annotated audio recordings.

All Mesolex resources are freely available to users although for access to audio annotated with time-coded transcriptions and translations, users must agree to certain basic Terms and Conditions under a Creative Commons license.

Mesolex resources fall into the following categories:
- Published and unpublished textual material (§4)
- Videos relevant to native community language and culture (§4)
- An electronic book on comparative Indigenous knowledge of flora and fauna (§5)

- Dictionaries (§2) of Indigenous languages of Mesoamerica
- Annotated (transcribed and translated) audio recordings covering a wide range of discourse genres (§3)

The most innovative aspects of Mesolex are how it has developed new processes and software, specifically developed for this project, to enhance user experience and success in accessing a variety of resources.

Dictionaries: Dictionary look-up in any Mesoamerican language is challenging given both the absence of standardized orthographies and the pervasiveness of linguistic variation across closely related languages (such as the 31 Nahuatl languages represented in Glottolog). In regard to dictionaries and lexicons, Mesolex's goal is to enable end users to easily look-up entries:

1. In dictionaries that might use an orthography distinct from the one with which they are familiar (e.g., a dictionary that uses <s> instead of <z> or <w> instead of <u> or <hu>); or
2. In dictionaries of languages slightly distinct from the one that an end user speaks (e.g., a speaker of central Guerrero Nahuatl trying to find an entry in a dictionary of Northern Puebla Nahuatl).

Figure 1: Mesolex home page
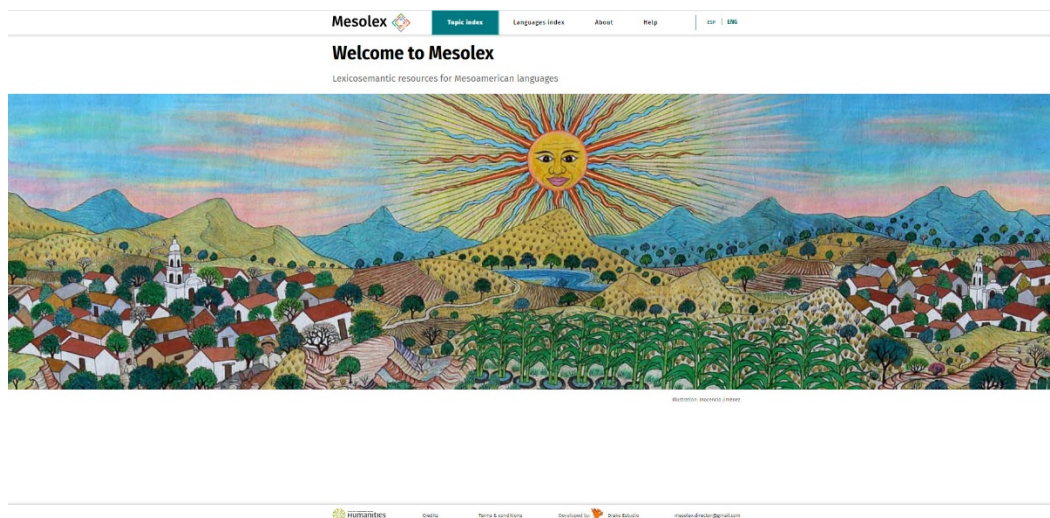
Mesolex accomplishes these goals by building language specific finite state transducers (FSTs) that are hosted on the Mesolex server and automatically generate equivalences across orthographic and linguistic variation. This FST utility is extensively discussed in §2 of this white paper.

Annotated (transcribed and translated) audio recordings: Scholars, students, and native speakers who work to document an Indigenous language often record an extensive amount of audio which, in the best case scenario, is transcribed and translated, often in state-of-the-art software such as ELAN (https://archive.mpi.nl/tla/elan) or, less commonly, in older software such as Transcriber. Making this audio and textual material available to a range of potentially interested end users, however, is challenging. At present delivery is extremely cumbersome: Users would need to download both the audio and text (transcription and translation in ELAN format), install ELAN on their local computer, link the audio and text files through ELAN, and learn to use ELAN for playback. This is not a trivial task for most potential users. Moreover, it is extremely inefficient as each user would have to repeat the same process on their own computer, file by file. Mesolex has developed software called SYLARD

(Synchronized Language Annotation Result Display) that converts ELAN files to HTML with synchronized line-by-line playback of audio and annotations (transcription/translation) that can be viewed through any browser. With the SYLARD software, ELAN (XML) to HTML conversion need take place only once and be uploaded to the Mesolex portal via the interface created for this purpose. Any user will be able to hear the audio (and download it) and follow the transcription and translation line by line. The SYLARD program is described in detail in a manual included in this White Paper (§3)

Paths to discovering and accessing Mesolex resources

Another particularly challenging aspect of language documentation is discovery: How end users find materials of potential interest. Mesolex presents two paths to its resources. The first is through a Topic index (fig. 2), which divides resources into five groups:

- Lexicons
- Library (published and unpublished textual material)
- Flora and fauna (a prototype eBook of traditional Indigenous knowledge, in this case about plants)
- Audio (accessed through SYLARD)
- Video (a catalogue of links to videos about Indigenous culture)

Figure 2: Topic index	Figure 3: Language index
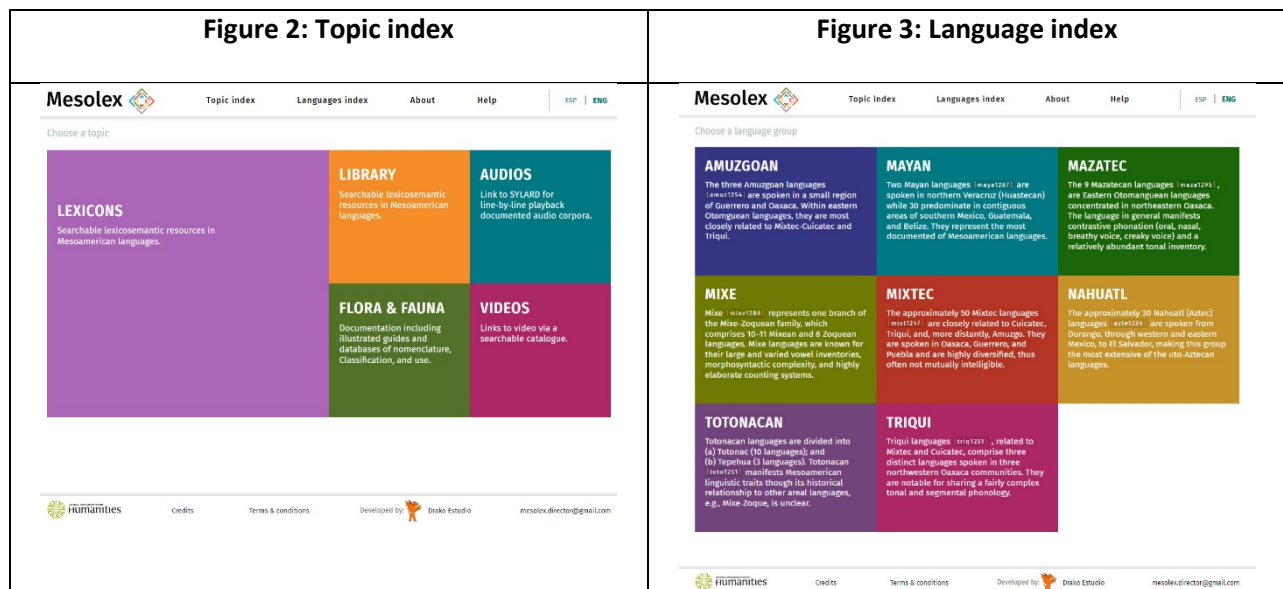
The second path is through a Language index (fig. 3) in which a user first selects a "language group" (e.g., Mixtec) and then what is called a "terminal language," usually a language spoken in one or several closely related communities. Thus a given user might first select Nahuatl. After clicking on the appropriate colored rectangle, a list of terminal Nahuatl languages appears, each with its own set of relevant resources.

The way in which one may use either of these two paths to discover Mesolex resources is described in the sections that follow.

Mesolex support functions

Mesolex is bilingual, English, and Spanish, so that users who speak either of these two languages enjoy easy and full access to the lexicosemantic resources accessible through this portal in their native language. The home page provides two support documents. The first is simply entitled About Mesolex (fig. 4). It presents much of the information that is included in this White Paper, a full discussion of the different resources, the paths to these resources, and the specific Mesolex software that supports online presentation of dictionaries and audio with line-by-line playback.

Figure 4: About Mesolex

The second support document is a help file. A user, by toggling between the English and Spanish options at the top right corner of the Mesolex page, can view either the English (fig. 5) or the Spanish (fig. 6) help manual. The sections of this present White Paper, Manuals covering all Mesolex resources, will be linked to this help document as well as made available through each particular subportal be it based on topic or language community.

Figure 5: Mesolex Help

Contact

Users may at any time write to the Mesolex administration for help with the use of any module or with any Mesolex other questions that might come up. Please, however, first read all pertinent help items in Mesolex. Allow at least 2 weeks for a response. The email address is mesolex.director@gmail.com

General help

The module that requires the most specific knowledge from end users is the Audio module managed through **SYLARD**. In this instance a separate User Manual is being developed and this should be consulted for help with SYLARD. The discussion that follows relates to help for the Library, Video, Flora and Fauna, and Lexicon modules.

Library

The Library module provides access to a wide range of textual material relevant to the terminal languages included in Mesolex resources. Five principal fields are searchable: Title, Author, Language, Community, and Keywords. Users may also choose to search across all fields. The results of all searches will be listed onscreen; the user can choose to display the results in ascending or descending order for any of the four principal fields: Title, Author, Language, or Community.

Videos

Videos are accessible through links to outside URLs, particularly on YouTube. The search interface and results display are the same as those used in the Library module.

Flora and Fauna

This is the prototype for an electronic book about Flora and Fauna in Mesoamerican Indigenous communities. As in print books, the species entries can be found via an index page (table of contents) that presents the family and species of all biotaxa included in the ebook. For each species, various native language texts (Nahuatl, Totonac, Mixtec among others) are offered in audio, transcription, and translation to English and Spanish. At present it is the table of contents that users will use to access the different documented species.

Lexicons

Lexicons are associated with terminal languages and each has its own data

Figure 6: Mesolex Ayuda

Mesolex ◈   Índice temático   Índice de lenguas   Acerca de   Ayuda   ESP | ENG

Ayuda | Contacto

Los usuarios pueden escribir en cualquier momento a los encargados de Mesolex para solicitar ayuda acerca de como usar cualquier módulo o para cualquier pregunta que pueda surgir. Sin embargo, antes de contactar la administración, favor de leer todos los documentos de apoyo pertinentes a los usuarios de Mesolex. Permitir por lo menos 2 semanas para recibir una respuesta. El correo para ayuda es mesolex.director@gmail.com

Ayuda en general

El módulo que requiere el nivel de conocimiento más alto de los usuarios es el de Audio, manejado por medio de SYLARD. En este caso un Manual de Usuario especial para accesar a los audios está siendo desarrollado y debe consultarse directamente via SYLARD. La presentación que sigue se enfoca en los módulos de Biblioteca, Video, Flora y Fauna, y Léxicos.

Biblioteca

El módulo de Biblioteca facilita el acceso a un gran número de textos relevantes a las lenguas terminales incluidas en *Mesolex*. Se pueden hacer búsquedas en cinco campos: Título, Autor, Lengua, Comunidad y Palabras claves. Los usuarios también pueden buscar en Título, Autor, Lengua, Comunidad y Palabras claves. O bien, pueden buscar en todos los campos juntos. El usuario puede elegir el orden en que se presentan los resultados de una búsqueda, escogiendo un orden ascendente o descendente de cualquier de los cuatro campos principales: Título, Autor, Lengua o Comunidad.

Videos

Los videos son accesibles mediante vínculos a URLs externos, particularmente en YouTube. La interfaz de búsqueda y la muestra de resultados son los mismos que en el módulo de Biblioteca.

Flora y Fauna

Éste es el prototipo de un libro electrónico acerca de la Flora y Fauna de pueblos indígenas mesoamericanos. Como ocurre en los libros impresos, las entradas de cada especie se pueden encontrar en un índice donde se presenta la familia y la especie de todos los biotaxa incluidos en este libro electrónico. Para cada especie hay uno o más audios en lenguas indígenas (náhuatl, totonaco, mixteco, entre otros) con transcripción y traducción al inglés y español. Por el momento la búsqueda y consulta por parte del usuario de las distintas especies presentadas en él libro es a través de un índice.

Section 2
Manual for the dictionary of Sierra Nororiental de Puebla Nahuat

The following figures and texts are pertinent to the Sierra Nororiental de Puebla Nahuat dictionary look-up and presentation module that has been incorporated into the Mesolex lexicon database. Problems of dictionary look-up affect all users, no matter the target language of a bilingual Indigenous language dictionary. The following challenges are just some of those that occur in regard to look-up in any Indigenous language dictionary:

- Orthographic variation: Often such languages have no standardized orthographies or completing standardized orthographies. For example, in Mexico some Indigenous language orthographies might use <ca> and others <ka> to represent the [ka] sound;
- Linguistic variation: In general an Indigenous language will be closely related to different variants or languages. The most use dictionary would be one that is accessible across these linguistic variants. Thus a word in one tonal language may have a certain pattern whereas a cognate in a related language may have another. Searches and discovery procedures would be most useful if the tone of a search term would discover cognates with different tones. Nahuatl languages are commonly divided into tl- and t- languages (e.g., *tlantli* in the first, *tanti* in the second). A flexible search interface would allow an input of *tlantli* to discover *tanti*.
- Translation variation: One lexicographer may translate an Indigenous word as 'opinion' and another might use 'outlook' or 'belief'. Look-up that focuses on specific words can hide relevant synonyms. Mesolex dictionary look-up includes the possibility of activating an online thesaurus creating what is known as an onomasiological search tool.

Of the preceding three challenges, the first two are language group specific (i.e., tools must be developed for each language group): both orthographic and linguistic variation will be different for each language. The third challenge is general (solving it for one language will provide a solution for many languages). Thus, a Spanish-language thesaurus could be used with any bilingual (native language <> Spanish) dictionary.

This manual details how to access and use the Sierra Nororiental de Puebla Nahuat dictionary.

Figure 1 is the Topic index home page and figure 2 is the list of dictionaries planned for Mesolex. They are presented to the user as shown in figure 2 once the user has clicked on the category Lexicons as presented in figure 1.

Figure 1: Topic index with "Lexicons" as one catagory

Figure 2: List of lexicons that are awaiting incorporation into Mesolex

Clicking on the Highland Puebla Nahuatl link (above in brown) takes the user to the Sierra Nororiental de Puebla Nahuat homepage (fig. 7).

Figure 3: Language index main page with language groups

A second way to access a dictionary is through the Language index, which begins (fig. 3) with a presentation of what are here called "language groups," higher level nodes in a hierarchy of historical linguistic relations. The level chosen as the node to be displayed in the Mesolexis Language index is in some senses arbitrary: instead of "Nahuatl" the higher level "Uto-Aztecan" could have been chosen; instead of Mixe the higher level Mixe-Zoquean could have been chosen. An example of Uto-Aztecan taxonomy can be found on Glottolog at the expandable tree found at https://glottolog.org/resource/languoid/id/utoa1244.

Clicking on any language group will take the users to a seconday page in which the terminal languages included in the activated Mesolex language group are represented. Figure 4 shows the terminal languages included in the Nahuatl language group. Here, as in the language group presentation (fig. 3) a short text describes the language group. The short text in figure 3 is continued and amplified in a slightly larger text on the page dedicated exclusively to Nahuatl (fig. 4).

Finally, the user can access terminal language resources by clicking on any one of the languages listed in the language group detail (fig. 4). In this case, three Nahuatl languages are represented in the Nahuatl language group page: (1) Highland Puebla Nahuatl; (2) Balsas tonogenetic Náhuatl; and (3) Nawat (Pipil) of El Salvador. Clicking on the first square—Highland Puebla Nahuatl (Sierra Nororiental de Puebla)— remits the user to the relevant terminal language page (fig. 5). Then, clicking on Lexicons on the terminal language page takes the user to the same Sierra Nororiental de Puebla Nahuat dictionary interface (fig. 7) that was shown to be accessible through the Topic Index > Lexicons > Highland Puebla Nahuat (fig. 2).

Figure 4: Language group main page with language group description and list of terminal languages

Figure 5: Terminal language page: Highland Puebla Nahuat

Figure 6: Terminal language homepage: San Pedro Amuzgos with final design for access to: (1) Lexicon, (2) Audios (via SYLARD), (3) Videos; (4) Biblioteca; and (5) Flora and fauna

The final design of the terminal language page will have a series of clickable resources: Lexicons, Audios (via SYLARD), Videos, Biblioteca (Library), and Flora and Fauna. This design has been implemented in the San Pedro Amuzgos terminal language page (fig. 6). Clicking on Lexicons/Lexicos remits the user to the lexicon or lexicons relevant to the active terminal language. At present, for Highland Puebla Nahuat this interface is pictured in figure 7 (a new design has been created by not yet fully implemented for Mesolex lexicons.

Figure 7: Search engine for Highland Puebla Nahuat (Náhuat de la Sierra Nororiental de Puebla)
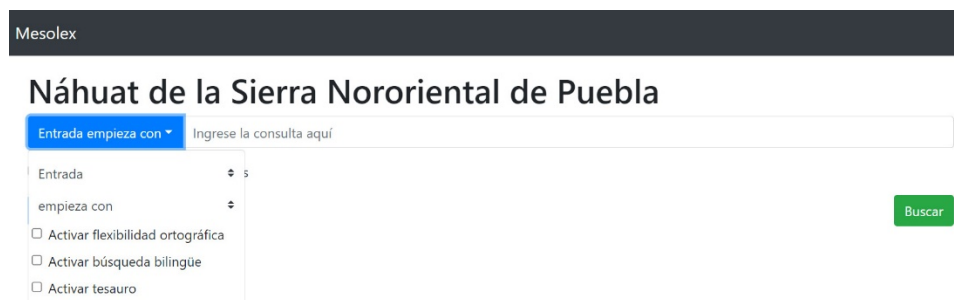
Figure 8: Search engine for Highland Puebla Nahuat: Fields available for searching

Figure 9: Search engine for Highland Puebla Nahuat: Operators for searching field contents

The present search engine presents three search catagories arranged vertically in one column; the search terms that users enter are typed into the thin rectangular space (second column). In the redesigned search interface (see appendix), the two search criteria are represented horizontally, each with a drop-down menu.

Present design

Fields to be searched (row 1 in the fig. 9 drop-down menu)	Headword (entrada in fig. 8) Gloss Root Semantic field …
Operators that control search (row 2 in the fig. 9 drop-down menu)	Begins with (empieza con in fig. 9) Ends with Contains sequence Contains word …
Character string to be searched (rectangular space to the right)	[here the user enters the characters to be searched]
Server-based functionalities (checkboxes rows 3, 4, and 5)	Activate flexible orthography Activate bilingual search Activate thesaurus

The user can alter the first two rows as he or she wishes in a mix-and-match style, e.g, selecting Extended meaning (significado extendido) and then Contains secuence [fill in characters]. After selecting these two fields, the user then fills in a sequence of characters (e.g., 'moon'): "Extended meaning contains secuence moon". The user-friendly redesign (see appendix to this present manual) is pending incorporation into Mesolex. It converts the rows into columns:

Field to be searched	Terms (operators) of search	Text to be searched
Headword	Begins with	[type in characters]
Gloss	Ends with	
Root	Contains sequence	
Semantic field	Contains word	
…	…	

Integrated into the bottom rows of the single drop-down menu of the present interface (and offered in separate items below the search field in the new design) are three server-side functionalities that will facilitate discovery in a single dictionary or, eventually, across multiple among Indigenous dictionaries. They are activated by clicking on the square to the left of each server-based functionality:

- ☐ Activate flexible orthography (see below)
- ☐ Activate bilingual searches (planned for the future)
- ☐ Activate thesaurus (activated as in fig. 10 and discussed below)

The following pages discuss and demonstrate each of the abovementioned functionalities along with, at the conclusion, the use of regular expression searches, an option that is presented for advanced users (see fig. 9 above, bottom of the pull-down menu).

Bilingual searches
At present, the ability to activate bilingual searches is only contemplated in the search interface but not integrated into the back-end search engine. It is planned for a future version of Mesolex when users might want to search across dictionaries with distinct L2 languages (e.g., Spanish and English). For example, if Mesolex has integrated Nahuatl-Spanish and Nahuatl-English dictionaries, activating bilingual searches will permit searches across multiple dictionaries with distinct L2 languages. Thus a user could discover the Nahuatl word for 'cat' in a Nahuatl-English dictionary even though the user had typed gato into the search condition Gloss contains word.

Thesaurus (onamasiological searches)
Figure 10 shows that the user has activated the thesaurus and typed in the word culebra ('snake') for the condition Gloss begins with (Glosa empieza con). The result includes the Nahuat word *kowa:t*, which was glossed as 'serpiente' (note, however, that although the first sense definition includes both 'serpiente' and 'culebra' the search was only in the gloss field so the appearance of *kowa:t* in the results is based on the synonymy established in the back-end thesaurus). Hidden to the user, the server-based thesaurus includes 'culebra', 'serpiente', and 'víbora' as synonyms. The synonymy works across all forms: the user could type in 'culebra' or 'serpiente' or 'víbora' and get the same results if the server-based thesaurus has been activated.

Figure 10: Mesolex search engine with Thesaurus activated

The innovative use of a server-based thesaurus responds to the challenges created by three situations. First, users who are looking for an Indigenous word for 'serpiente' would certainly want to include words glossed as 'culebra' or 'víbora'. Those looking for 'ponder' would undoubtedly be happy to get words returned from such a search glossed as 'think about', 'contemplate', and 'consider' as well as, probably, 'reflect upon' and 'envision' among many others.

Second, the glossing or sense definition of any Indigenous word is somewhat arbitrary. Different lexicographers might easily create different glosses such as 'snake' or 'serpent' for the same Indigenous language word such as Nahuatl *kowātl*, Mixtec *ko¹o⁴*, or Totonac *lūwaʼ*. Lexicographers from different regions will have their own areal "bias" (think of 'soda', 'pop' and 'coke', words that dominate in different sections of the United States), a "bias" that will inhibit consistency across glosses and definitions. A back-end thesaurus will mostly eliminate this problem of synonymous glosses.

Third, the user will undoubtedly utilize collocational or phrasal expressions that call for specific words in his or her language whereas the Indigenous language is more flexible: think of *envision a problem*, and *cut a deal*. A user who searches for 'envision' and 'cut' without an active back-end thesaurus will probably not find a useful result. 'Envision' might well not be in the dictionary and a search for 'cut' would probably find the Indigenous word used for objects such as wood, fruit, and paper and not figurative uses such as 'cut a deal'. Of course the user could run through a series of potential substitutes—'foresee' and 'make', respectively—an enter each one into the search field. But an automated back-end thesaurus for onomasiological look-up will avoid the cumbersome practice of users having to enter multiple synonyms. Mesolex provides such a solution.

Figure 11 illustrates the results that are delivered from a search for Gloss contains the sequence considerar. Only a couple of results are visible in this truncated screenshot but note how one gloss includes 'mirar' (lit., 'look at') while the second includes 'encontrar' (lit., 'find' or 'encounter'). Again the back-end thesaurus finds related words that might best fit the word that the user seeks. However, in the

case of a onomasiological search for 'considerar' a total of 173 results were returned, many useful but others some distance from the needed translation.

Caveat: Unwanted search results. At present, therefore, one problem of the back-end thesaurus is that it produces results that are often overly expansive. With dictionaries of 2,000–8,000 entries, this is often not a significant problem as users themselves can evaluate the results. But in future developments of Mesolex lexicon resources, attempts will be made to limit the results.

<div align="center">

Figure 11: Mesolex search engine with Thesaurus activated: Search term <u>considerar</u>

</div>

Flexible orthography searches

The final back-end functionality that is available for Mesolex lexicons is what is called "flexible orthography", an orthography that is generated by a back-end or server-side finite state transducer (FST). An FST create builds rule-based associations between two distinct forms, in linguistics often an underlying and a surface form. In Balsas Nahuatl, for example, an underlying affricate (e.g., /ch/ and /ts/) is realized as a fricative (/x/, as in English /sh/, and /s/) before occlusives such as /t/ and /k/. This change

does not occur in all Nahuatl languages. It is only one of many such variations that affect distinct Nahuatl languages. Note the following examples from Oapan:

Meaning	Unpossessed form	First person possessed ('my')
ground maize (nixtamal)	tixtli	notix
young man	tēlpōxtli	notēlpōch
metal	tepostli	notepos
thorn	wistli	nowits

Both *tixtli* and *tēpōxtli* have a surface /x/ before the absolute ending *-tli.* But underlyingly each surface /x/ is different, a difference that appears in the unpossessed forms *notix* and *notēlpōch*. The surface /x/ in *tixtli* reflects an underlying /x/, so it appears in the possessed form: *notix*. But the surface /x/ in *tēlpōxtli* represents an underlying /ch/, so it is a /ch/ that appears in the possessed form: *notēlpōch*. An equivalent variation occurs with /ts/ and /s/, as evidenced in the final two rows.

This changes from underlying /ch/ to surface /x/ or from underlying /ts/ to surface /s/ do not occur in all Nahuatl languages. If a speaker who speaks a Nahuatl language that loses /ch/ and /ts/ in surface forms (to /x/ and /s/) types in his or her pronunciation, e.g., *tēlpōxtli*, without the flexible orthography function activated, the search will not find *tēlpōchtli*. Activating the FST-based flexible orthography will automatically find linguistic equivalents among most Nahuatl languages.

The preceding are *linguistic* variations, some of many regional differences that affects Nahuatl. Orthographic variations represent another challenge: Even equivalent surface forms across various Nahuatl languages may be spelled differently. Thus the word for 'fat' is often [tomāwak] but distinct orthographic conventions may spell this as /tomāwak/, /tomahuac/, /tomauak/, /tomauac/, etc. By designing a finite state transducer that links all these spelling variations, a user can type in any one of the many potential spelling and find the entry /toma:wak/ in the Highland Puebla Nahuat dictionary.

Linguistic (e.g., absolute suffixes *-tli* vs. *-ti*) and orthographic (e.g, /h/ vs. /j/) variations may create dozens of possibilities with even the simplest words:

ohtli   ohtle   otli   ótlí   ohti   ojtli   ojtle   ojti

The advantages of back-end, language specific finite state transducers to address issues of linguistic and orthographic variation are immense. For Mixtec the representation of tone, which varies greatly among languages both phonologically and orthographically, can be neutralized. For Totonac the same can be done for laryngealization and for Triqui a back-end FST can eliminate the need for users to distinguish fortis from lenis consonants. Figures 12 and 13 illustrate the utility of flexible orthography for searches in the Highland Puebla Nahuat dictionary.

Figure 12 shows a search for /tlacatl/, the spelling that most Nahuatl speakers would use, at least as the most common default. The FST generates a series of alternative forms and submits these forms to the lexical database, finding the entry for *ta:kat*. The finite state transducer generates the linguistic alternation /tl/ <> /t/ and the orthographic alternations /c/ <> /k/ and /a/ <> /a:/.

Figure 13 offers another example of Mesolex's flexible orthography. The user entered a search for /otli/, which is the spelling found in Molina's classical Nahuatl dictionary (Molina did not represent the glottal stop after the initial /o/) and the pronunciation that occurs in Ameyaltepec, Guerrero, where the underlying /h/ disappears in the surface form. Thus a student of classical Nahuatl and a speaker from

Ameyaltepec would both probably search for /otli/. The back-end FST would, however, make the necessary linguistic and orthographic conversions and find Highland Puebla /ohti/.

Figure 12: Mesolex search engine with Flexible orthography activated: Search for <u>tlacatl</u>

Figure 13: Mesolex search engine with Flexible orthography activated: Search for <u>otli</u>

Regular expression (regex) searches

Regular expressions are powerful tools for creating strings of characters that match patterns in text. This Mesolex functionality allows advanced users, including linguists, to develop highly specific searches for patterns in the lexicon. For example, a regex search for headword = ^tl?[ie] will find any headword that begins with /t/, followed by an optional /l/ and then a front vowel, /i/ or /e/.

**Figure 14: Regular expression search for words of the pattern CVCV
in which the consonants are occlusives**

Figure 14 shows the results (the first of 23 words) of a regular expression search for *^[ptk]w?[aeiou]:?[ptk]w?[aeiou]:?$*. This "translates" as 'field begins with (^) either p, t, or k, followed by an optional w, then any vowel optionally long (:?). This pattern repeats again until reaching the end of the field ($).

The preceding regular expression search demonstrates how such searches can find patticular patterns of interest to a user. For example, another search might be developed not for CVCV words but for CVCCVC, i.e., two closed syllables in a two-syllable word.

Summary

This manual demonstrates how the Mesolex dictionary search interface is particularly able to faciliate dictionary look-up by users from a range of backgrounds. Most importantly, this includes speakers and students of related languages with a phonology and morphology slightly distinct from the one that is used in the online Mesolex dictionary.

Section 3
Manual for SYLARD

Introduction: SYLARD (Synchronized Language Annotation Result Display) is a program designed to allow line-by-line playback of digital audio that has been annotated in the ELAN (EUDICO Linguistic Annotator; https://archive.mpi.nl/tla/elan) time-coded format. It does this by converting the ELAN XML code to online HTLM and linking this HTML file to an .mp3 version of the original uncompressed .wav audio file. The .wav file is the format used for permanent archiving whereas the .mp3 is used for online playback through a browser. Note that use of SYLARD requires registration, which is requested to assure that all those who access SYLARD have agreed to the terms and conditions of fair use established by a Creative Commons license (https://creativecommons.org/).

SYLARD allows for two types of registration:
   Visitors **:** Can freely access and play back all materials (.mp3 audio and annotations [including transcriptions and translations]). Registration as a Visitor is automatic and quick and is simply used to assure that Visitors have agreed to the Terms and Conditions.
   Contributors **:** Can create a collection that they control and administer. A Contributor must have independently uploaded an .mp3 version of the annotated audio to a secure (https) server. SYLARD allows the Contributor to create the collection metadata and to upload ELAN files for conversion to HTML while linking the annotations to the .mp3 file for line-by-line playback.

The following details are for Contributors. Visitors can register simply by sliding the selection in the top left corner to Visitor instead of Contributor (see fig. 4)

Stage 1: Beginning. Go to the Mesolex home page (fig. 1; presently at https://mesolex-drako-iano9.ondigitalocean.app/en) and then click on Topic Index to access five different paths to Mesolex data (fig. 2: Lexicons, Library, Flora and Fauna, Audios, Videos). SYLARD is accessed by clicking on the Audios square.

Figure 1: Mesolex home page: Path to SYLARD

Figure 2: Mesolex Topic Index home page (Path to SYLARD via Audios)
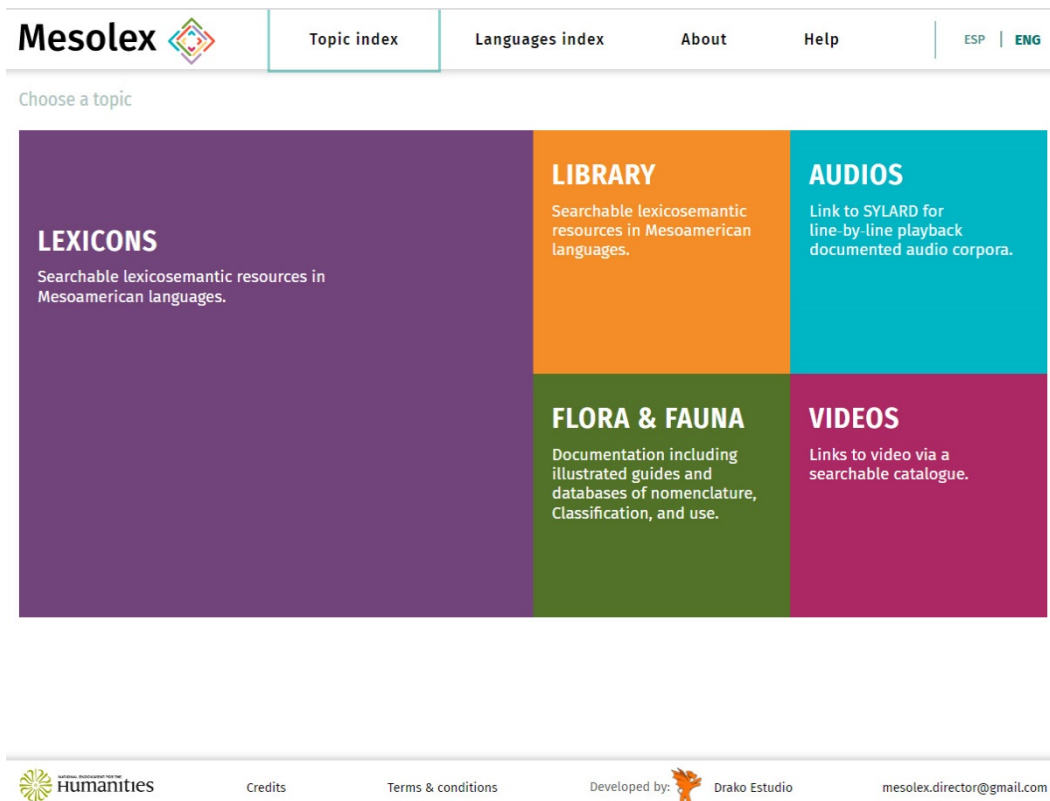
Stage 2: Accessing SYLARD. After clicking on Audios in the Mesolex Topic Index screen (fig. 2), the user is taken to the SYLARD home page (fig. 3). On this page there is a dark blue-green header band that brings users to any one of several functionalities and documents: Catalogue, About SYLARD, User manual (this document), Source code (on GitHub), Terms and Conditions, and Credits. For those who have already registered (as either Visitor or Contributor) there is a Login option (upper right). For those new to SYLARD there is a Sign-up (upper right) or Registration option (the latter is one of three clickable options in mid-page). Clicking either Sign-up or Register takes the new user to the Registration page (see figs. 4, 5, 6).

Figure 3: SYLARD home page
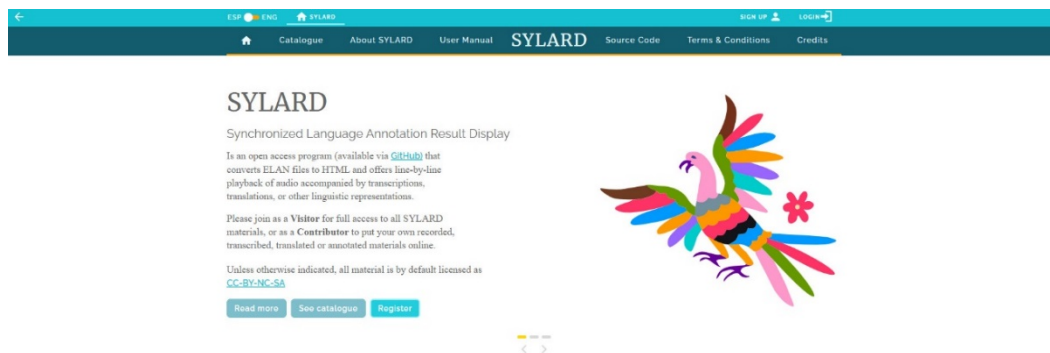
Stage 3: Registration. Registration is accomplished through the Registration panel. The lefthand rectangle with blue background contains a link to the Terms and Conditions that Visitors and Contributors agree to by completing their registration.

At the top left of the righthand section (white background) users are asked to choose the status for which they wish to register (Visitor or Contributor). The first steps are virtually the same for both but the Contributor status differs in that this status must (1) be authorized by the SYLARD administrator (figs. 10, 11, and 12) and (2) allows the user to create collections and populate them with annotated audio accessible through any browser. As the Visitor status is limited to viewing only, administrative authorization is not necessary.

Figure 4: Register as either Visitor or Contributor

The first panel for Contributor registration is a request to answer the question "Tell us why you want a Contributor account". Visitors are not asked this type of question. For Contributors the responses to this question allow the SYLARD team to better meet the needs of Contributors. The rest of the entry form is fairly standard:
- name
- email
- password and password confirmation
- country of origin
- languages spoken or studied (remember to click on the plus sign so that the language is added to the following box: Languages spoken or studied added)
- personal statement (optional; again this question is presented to allow the SYLARD team to better understand the interests of Contributors)

Figure 5 shows the completed registration form with three languages spoken or studied having been added by clicking the + sign next to the Languages spoken or studied rectangle. After the form is completed, the potential Contributor will need to click the Register button at the bottom right of the Registration screen. A pop-up acknowledgement appears (fig. 6) and the Contributor must click Accept. At this point, the Contributor is notified that an email has been sent to them (fig. 7). The Contributor

should then go to the email address associated with their SYLARD account and confirm the request and email by clicking on the link in the email (fig. 8). This triggers another notification (fig. 9) that advises the potential Contributor that their request has been forwarded to the SYLARD administrator (fig. 10), who must then authorize the requested Contributor status of the applicant. Once the administrator has authorized the Contributor status the administrator will see confirmation of this authorization (fig. 11) and the Contributor will be sent an email confirmation of the authorization (fig. 12).

Please continue to figure 13 and the immediately preceding text to continue reading this manual.

Figure 5: SYLARD registration screen completed by potential Contributor

Figure 6: After clicking Register a pop-up screen of Terms and Conditions appears and must be accepted

Figure 7: On the SYLARD page: Notification to a potential Contributor that an email confirming the application (fig. 8) has been sent to their account
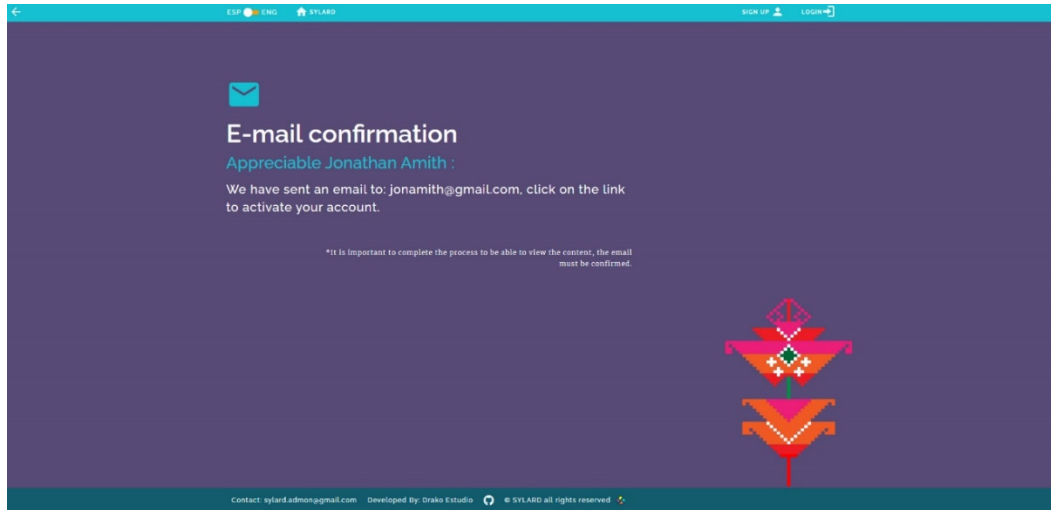
Figure 8: SYLARD generated email sent to the user requesting activation of a Contributor account
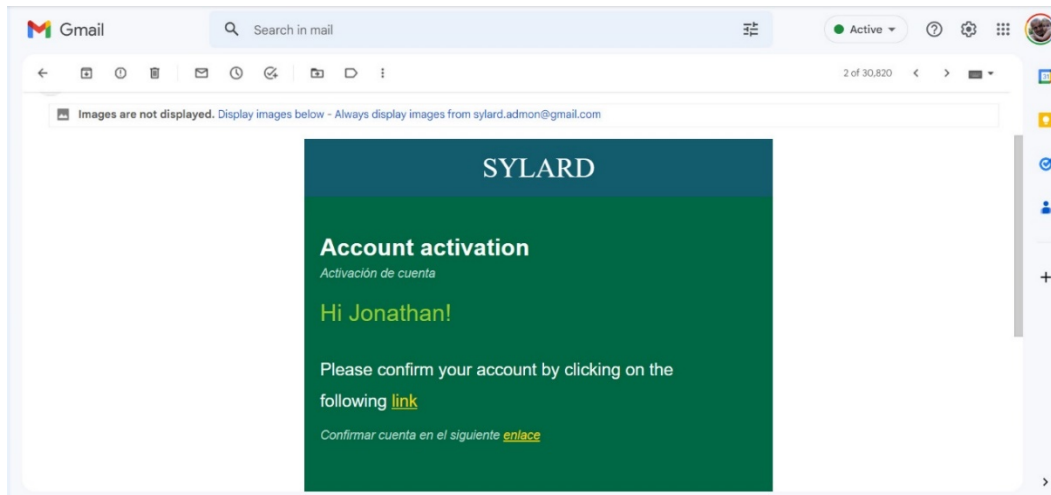
Figure 9: Confirmation on the SYLARD page that an application for Contributor status is being processed
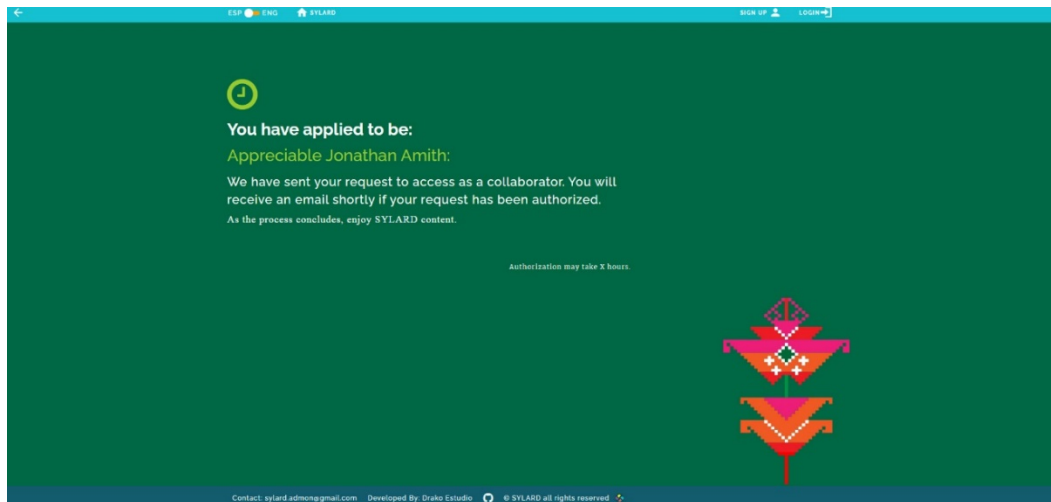
Figure 10: Administrator's panel (seen only by Administrator) that a user has requested Contributor status
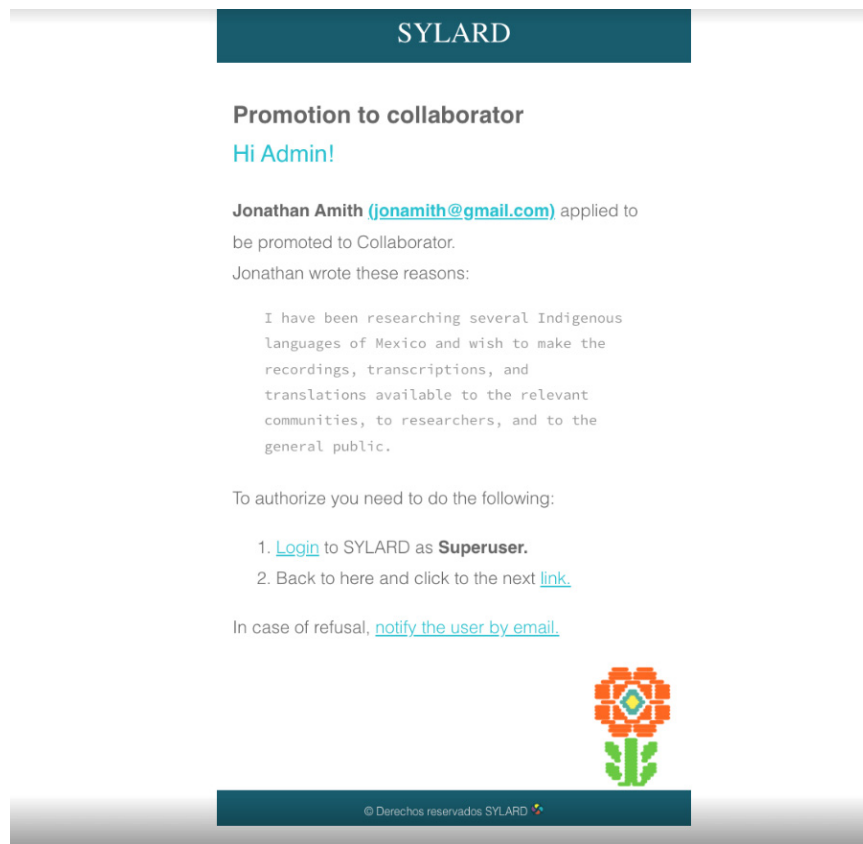
Figure 11: Administrator's panel (seen only by Administrator) once the Administrator has approved a request for Contributor status
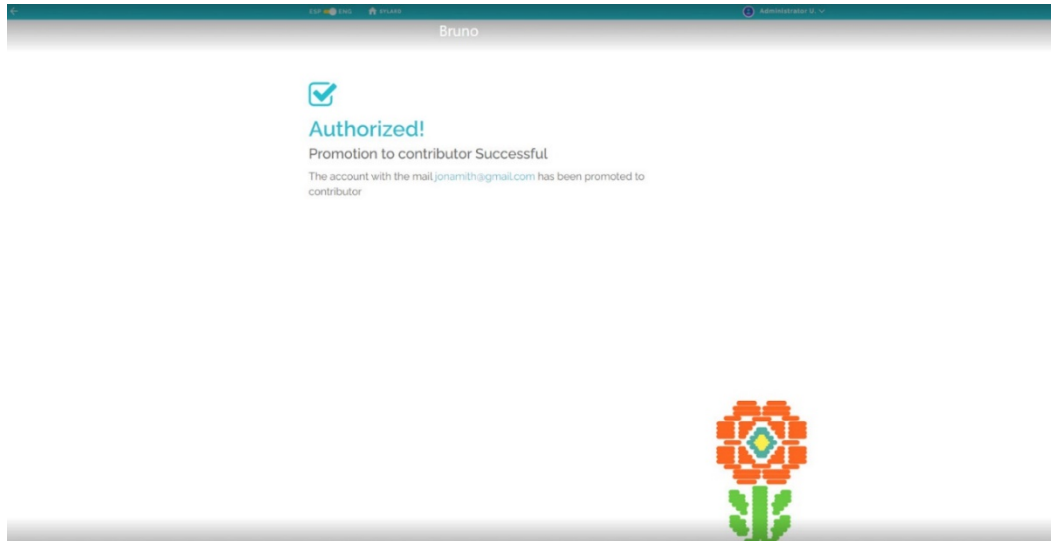
Figure 12: Notification via email to an applicant for Contributor status that their application has been approved
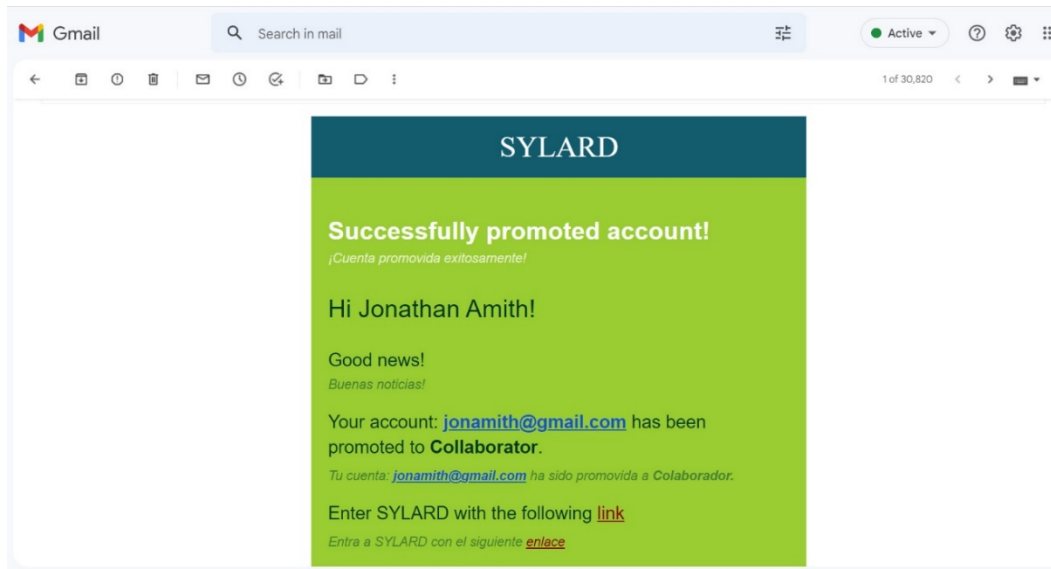
Stage 4: Log-in and reset password options. Once a User has been authorized as a Contributor they can now log in with this new status. To do so, they must go to the SYLARD home screen (fig. 13) and log in through the link in the upper right corner of the topmost light blue header band. By clicking on the Login link the Contributor is taken to the login page (fig. 14) where they are asked for their registered email and password. However, it often occurs that a User (either Visitor or Contributor) forgets their password. In such cases SYLARD has an integrated system to obtain a temporary password and then reset a permanent password. To do this they must first click on Forgot password as in figure 14, lower left corner of the white portion of the screen.

Figure 13: SYLARD home page for login
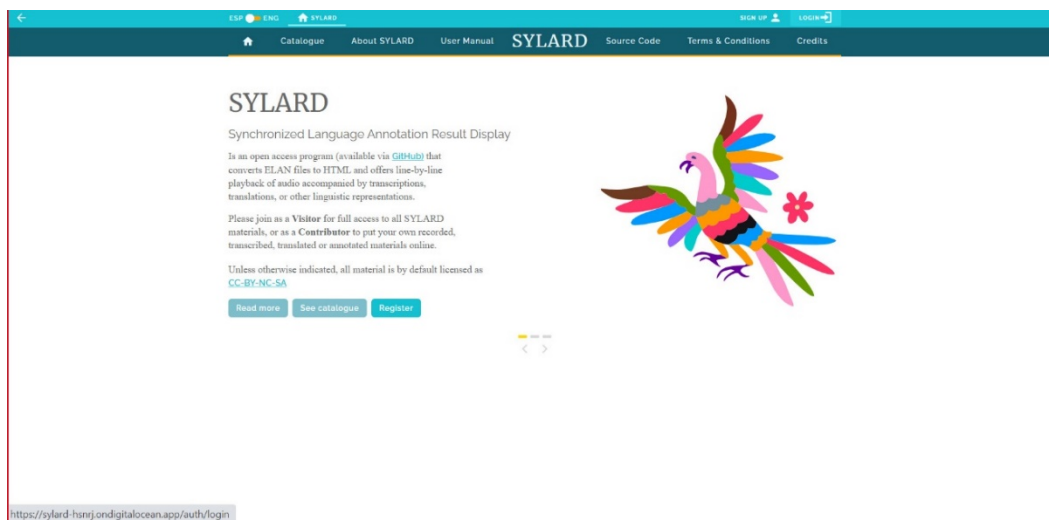
Figure 14: Login panel in SYLARD with "Forgot password" option
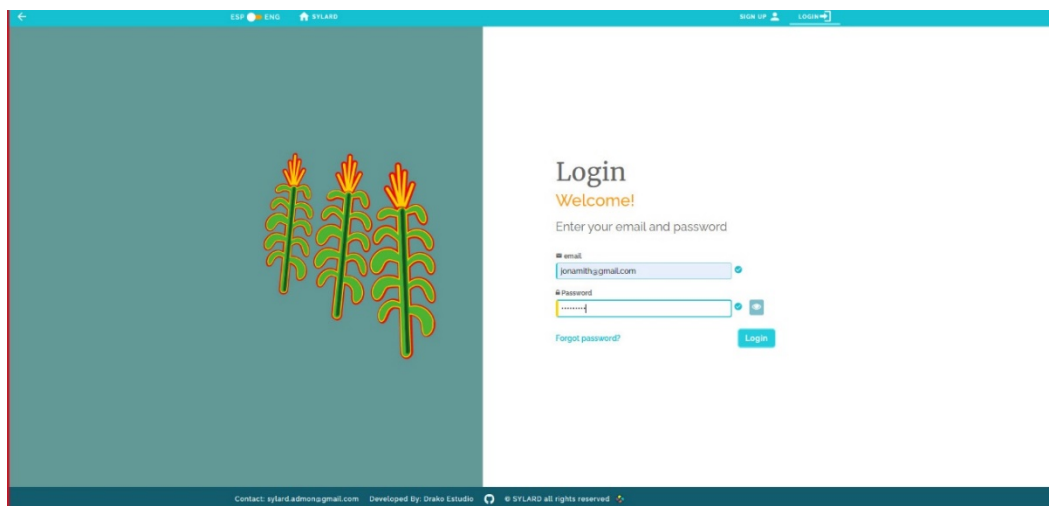
After clicking on the Forgot password? text in blue at the bottom of figure 14, the User is taken to the online password reset module seen in figure 15. The User must first enter their registered email and then click Submit. A temporary password is then sent to the registered email (fig. 16). The Contributor

must use this temporary password to log in (fig. 17). Once logged in, the User can go to the drop-down menu in the upper right corner of figure 18 and select the Change password option. Selecting this option takes the User to the change password screen (fig. 19) where the new password must be entered and reentered to confirm.

The text of this manual continues immediately above figure 20.

Figure 15: Enter email and submit to obtain a temporary password (fig. 16)
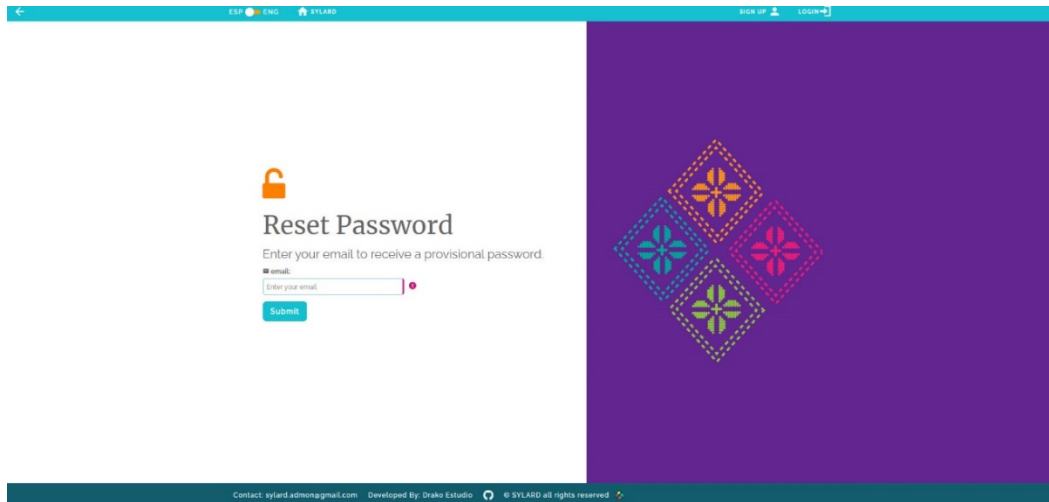
Figure 16: Email receipt of temporary password: Copy and paste this (fig. 17) to log in to SYLARD
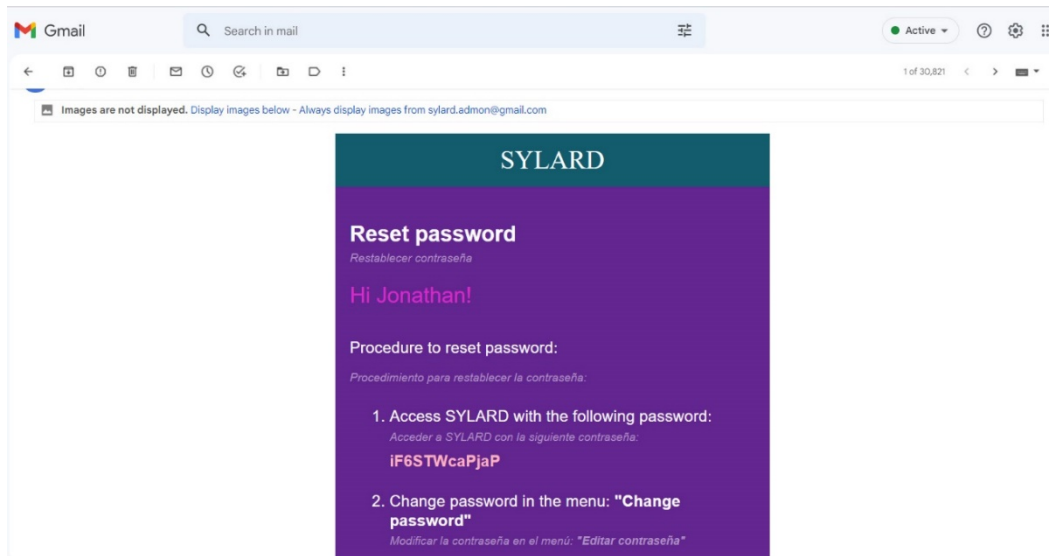
Figure 17: Copy and paste temporary password to log in to SYLARD
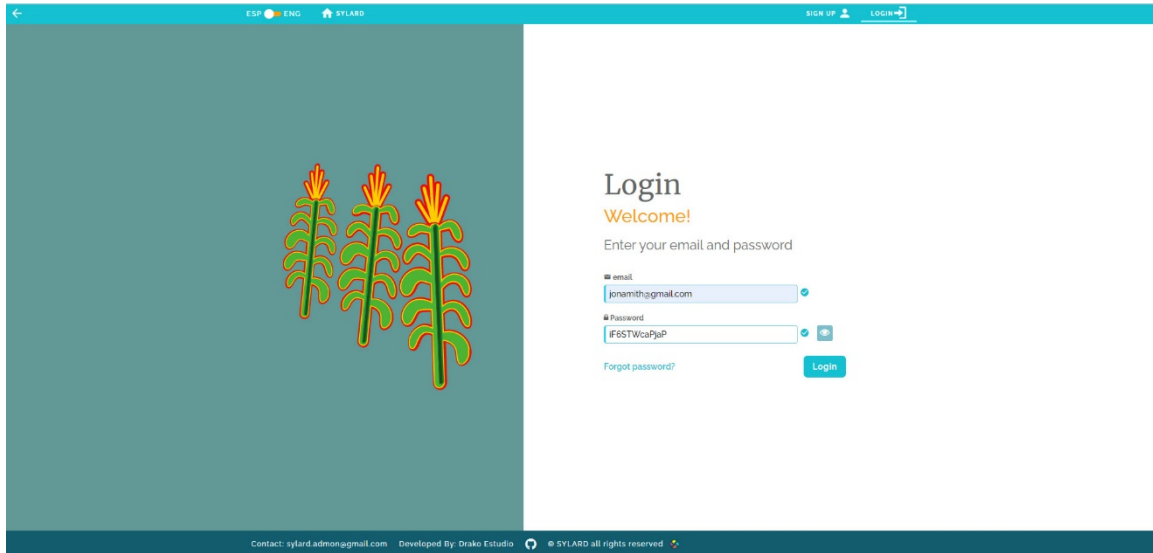
Figure 18: Once logged in with temporary password: Select Change Password through drop-down menu
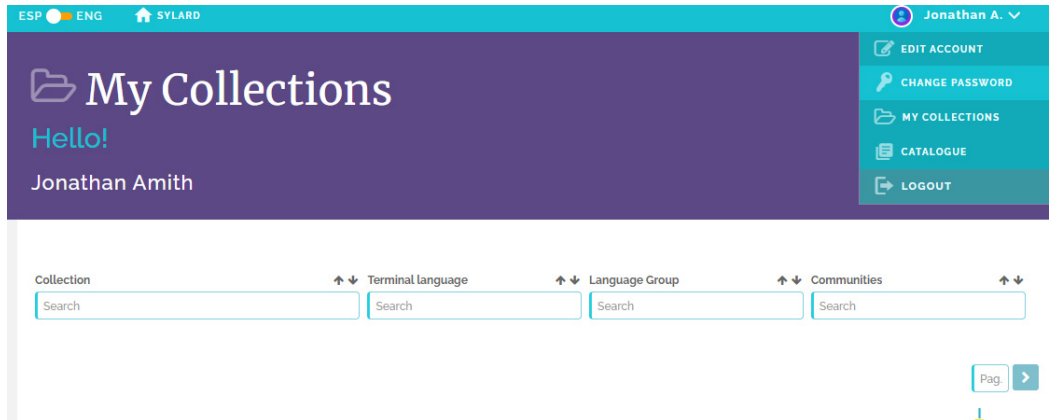
Figure 19: Enter and confirm new password

Stage 5: Creating a collection: Once logged in (either with the original or changed password) a Contributor will be taken to the Catalogue page (fig. 20) if they have no collections already in SYLARD. If they have already uploaded annotated audio in one or more collections they will instead see their My Collections page with the collections listed (fig. 21). At this point, to add a new collection a new Contributor can click on Add Collection from the drop-down menu (fig. 20) whereas Contributors with a collection or collections already in SYLARD can click the Add Collection tab located in the purple header on their My Collections page (fig. 21).

The text of this manual continues above figure 23.

Figure 20: Catalogue page through which Contributors access their Collections

Figure 21: My Collections page presented when a Contributor with collections in SYLARD logs in
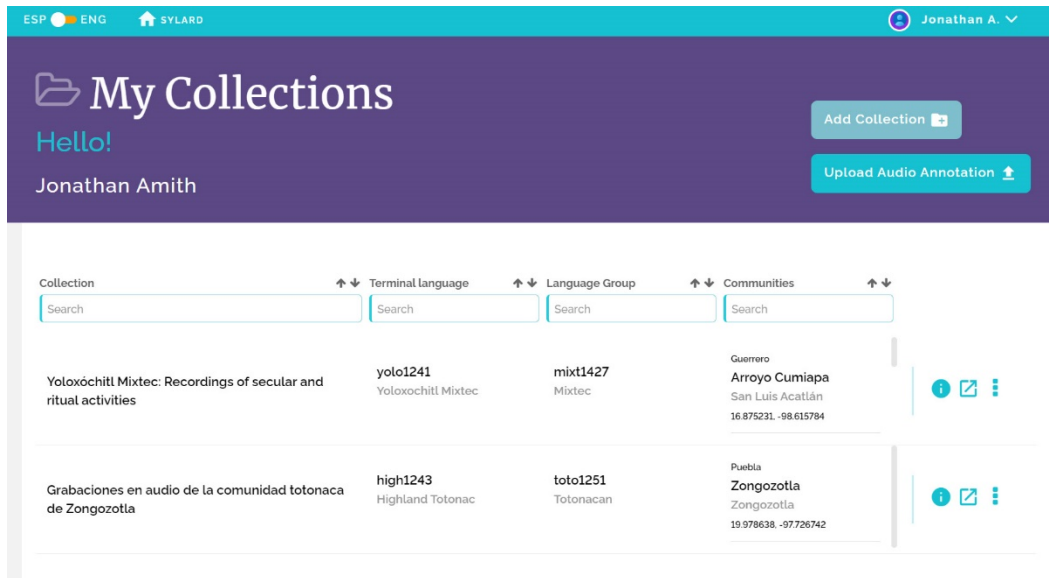
Figure 22: My Collections page accessed by new Contributors from the drop-down menu on figure 20

After clicking on the Add collection option in either figure 21 (for totally new Contributors) or the Add collection option in figure 22 (for Contributors with collections already integrated into SYLARD), the Contributor will be taken to the online Add collection template, figure 23. The following pages of this manual explain how this template should be filled in.

Figure 23: Add collection template as it first appears

Title and Description: The first step is to give the collection a short name and description. For reasons of economy of space both number of characters in the name (500 characters) and description (1000 characters) are limited. The name and description, along with other information in a My Collection metadatabase can be edited at any time (see figure 37: Edit metadata).

Note that the collection description is available to any user by clicking on the circled capital I (for Information) to the right of the three vertical dots next to the project name in the catalogue (see fig. 33).

Once the name and description are entered, the Contributor then proceeds to enter the language data (language group, terminal language, communities included). This process is illustrated in figures 25 to 31. The final step is attaching a license to the new SYLARD collection (fig. 32) At present only two Creative Commons licenses are available. In the future other licenses will be added, including TK (Traditional Knowledge) licenses (see https://localcontexts.org/licenses/traditional-knowledge-licenses/).

The language and community metadata that are added in screenshot figures 25 to 31 are selected from controlled vocabularies that are derived from Glottolog (https://glottolog.org/) and Mexico's INEGI (Instituto Nacional de Estadística y Geografía) database. This allows auto-fill functionality from a drop-down menu for all languages and admininistrative units relevant to Mexico.

The Contributor begins by adding what is called the Terminal language, i.e., the lowest level classification of the language that is documented in the collection being added to SYLARD.

Terminal language: As the contributor starts to type in the terminal language, in this present case a Mixtec language, after four characters are entered SYLARD provides a list of Mixtec languages taken from Glottolog (fig. 25). The Contributor can then simply click on any of the terminal language names in the drop-down menu.

The text of this manual continues immediately above figure 28.

Figure 24: Add collection template with Collection name and Collection description filled in

Figure 25: Auto-fill of terminal languages based on the Glottolog catalogue of world languages

Figure 26: Auto-fill of language group based on the Glottolog catalogue of world languages

Figure 27: Automatic selection of country—Mexico—after Terminal language and Language group are selected

Language group: Once a terminal language is entered, the language group list for the pull-down menu is automatically filtered to only show Glottolog language nodes above the selected terminal language. Figure 26 shows the higher level nodes for Yoloxóchitl Mixtec: Guerrero Mixtec, Mixtec, Mixtec-Cuicatec, Mixtecan, Amuzgo-Mixtecan, Eastern Otomanguean, Otomanguean. The Contributor is free to select the node they wish, though usually a lower-level node is preferable as it groups fairly similar terminal languages together. **NOTE:** It is necessary to click the plus sign next to Language group to write Terminal language and Language group to the database (see fig. 28: Languages added to the collection).

Geography: This set of metadata items includes Country, Entity (in this case, state), Municipality, Locality, and Latitude and Longitud. For Mesolex the country is automatically filled in as "Mexico". Eventually and if needed Guatemala will be added. As the country is preselected, the Entity drop-down menu (fig. 28) provides a list of states that can be selected by clicking on the appropriate entity. In turn, once the state is selected the Municipality drop-down menu (fig. 29) is automatically populated by the municipalities in the state. Finally, after the Municipality is selected the pull-down menu on locality is filtered for those localities (villages, agencias, etc.) that INEGI has established for the selected municipality. Once the locality is selected from the drop-down list, the latitude and longitude are automatically inserted into the metadata template for this collection. **NOTE:** It is necessary to click the plus sign next to Longitude to add the geographical information to the database (see fig. 32: Communities added to the collection).
    The text of this manual continues immediately below figure 31.

Figure 28: Drop-down presentation of states in Mexico

Figure 29: Municipality drop-down menu that appears after the State is selected and entered

Figure 30: Locality options in the drop-down menu after municpality has been selected

Figure 31: Locality selected prompting insertion of geographical references (lattitude and longitude)

Licensing: The final data point to be entered are the terms for the Content license (see fig. 32). At present there are only two options, both are Creative Commons licenses. As development continues, Traditional knowledge licenses will be added.

Adding multiple communities to the collection database: The steps illustrated in screens 25 to 32 can be repeated within the same collection to add multiple communities. It is, indeed, often the case that even collections from individual researchers will include material from multiple nearby communities. SYLARD can handle this situation. Figure 34 illustates how this will appear once a total of four communities have been added to the Yoloxóchitl Mixtec collection. Figures 35 and 36 illustrate how the multiple communities will appear in the catalogue, in the Communities column to the right. The sliding grey bar to the right of this fourth column can be moved to show the four communities, listed in alphabetical order: Arroyo Cumiapa, Buena Vista, Cuanacaxtitlan, and Yoloxóchitl.

Collection metadata editing and adding new audio: After the collection has been populated with annotated audio, it is still possible to edit the collection metadata (for editing annotated audio metadata, see fig. 37). Collection metadata can be edited by a Contributor at any time simply by logging in to SYLARD and going to the relevant collection (see, e.g., fig. 37). By clicking on the Edit metadata option

(the option is in a pop-up window that appears when the 3 vertical dots are clicked) the Contributor is taken to the Collection metadata template (fig., 32) from which point the metadata can be edited and saved. New annotated audio can also be added by clicking on the blue rectangle that reads Upload Audio Annotation. This process is discussed below.

The text of this manual about uploading annotations continues below figure 37.

Figure 32: Collection description online template: Choosing the collection license

Figure 33: Online display of collection description
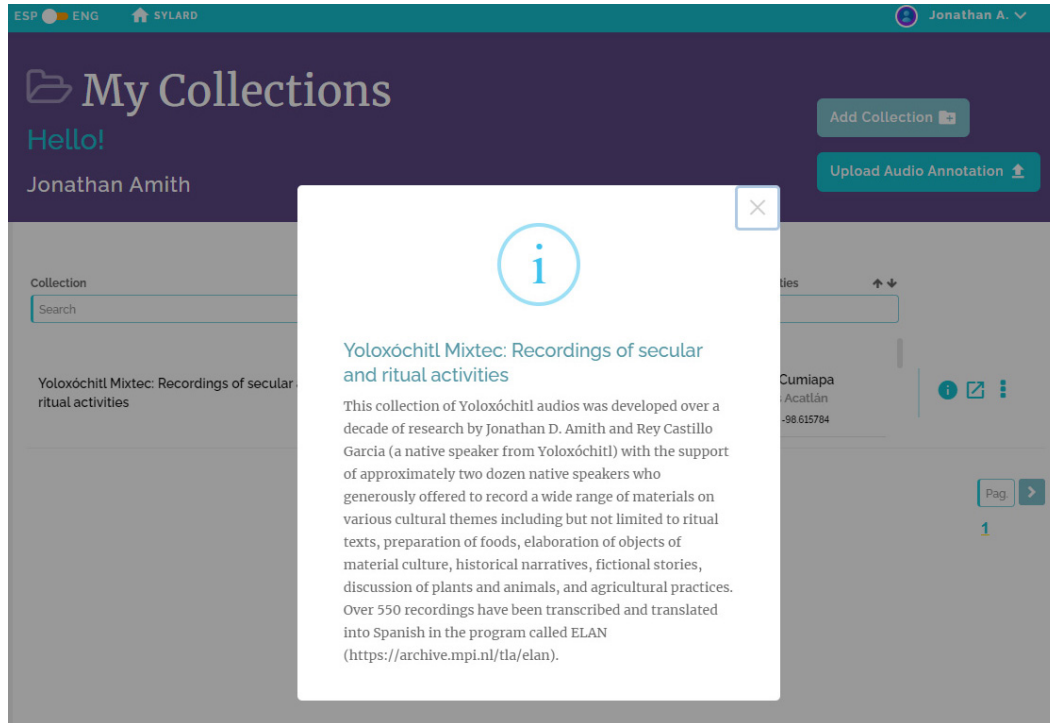
Figure 34: Collection with multiple communities within one terminal language

Figure 35: Catalogue display (four columns of metadata) with multiple communities displayed in the fourth column with sliding bar
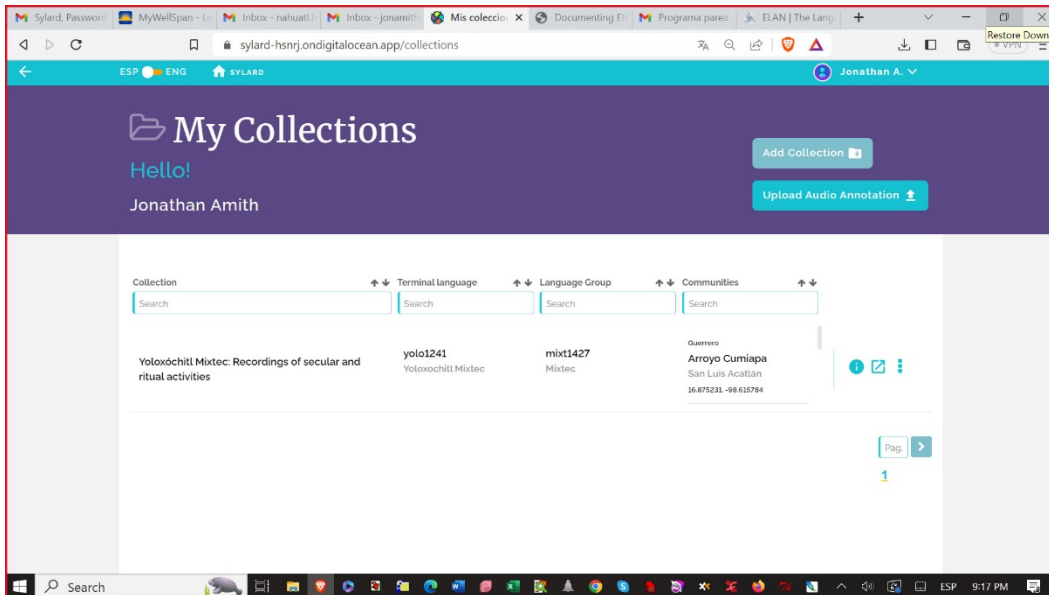
Figure 36: Catalogue display with multiple communities displayed in fourth column with sliding bar
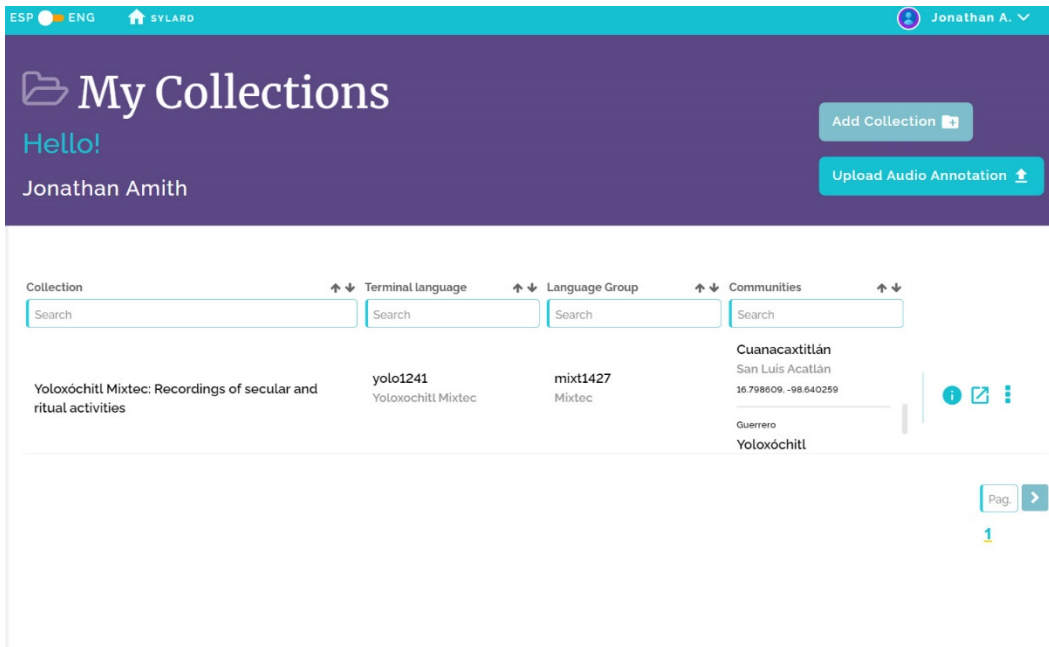
Figure 37: Collection metadata editing option: Click on 3 vertical dots for pop-up window
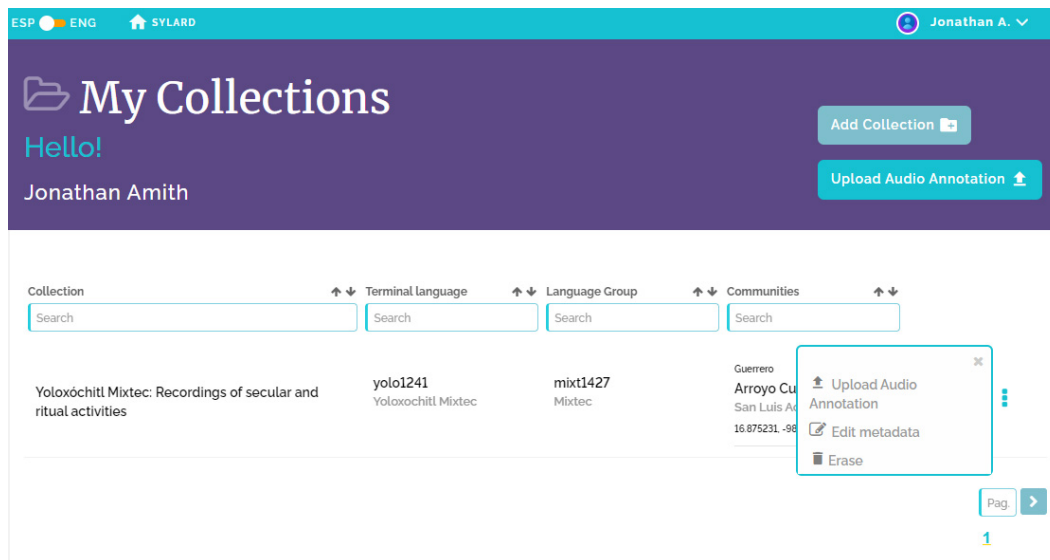
Stage 6: Uploading annotated audio to a collection: The final step in creating a collection is uploading time-coded ELAN files to the SYLARD server (which converts the XML ELAN format to HTLM for online display and line-by-line playback) and at the same time linking the annotation to the corresponding audio in mp3 format on a secure (https) server. The home screen for uploading and converting ELAN files (XML to HTML) and linking to .mp3 audios is accessed by clicking the Upload Audio Annotation button on the My Collections page (fig. 37, lower right corner in the purple banner). If there are multiple collections the upload is begun by clicking on the three vertical dots to the right of the collection to which audio will be added (fig. 37, pop-up window that offers the Contributor three options: Upload Audio Annotation, Edit metadata, Erase).

Step 1: Selecting an ELAN file for uploading: The first step is to choose an ELAN file on a local computer: Click Choose File (fig. 38) and navegate to the desired file and click Open (fig. 39). This will produce and lead the Contributor to the screen displayed in figure 40 into which the Contributor will enter the title, description, metadata and default display of the line-by-line playback of the selected audio annotation.

Figure 38: Home screen for creating line-by-line playback of annotated (transcribed, translated) audio

Figure 39: Selection ELAN annotation for uploading to SYLARD
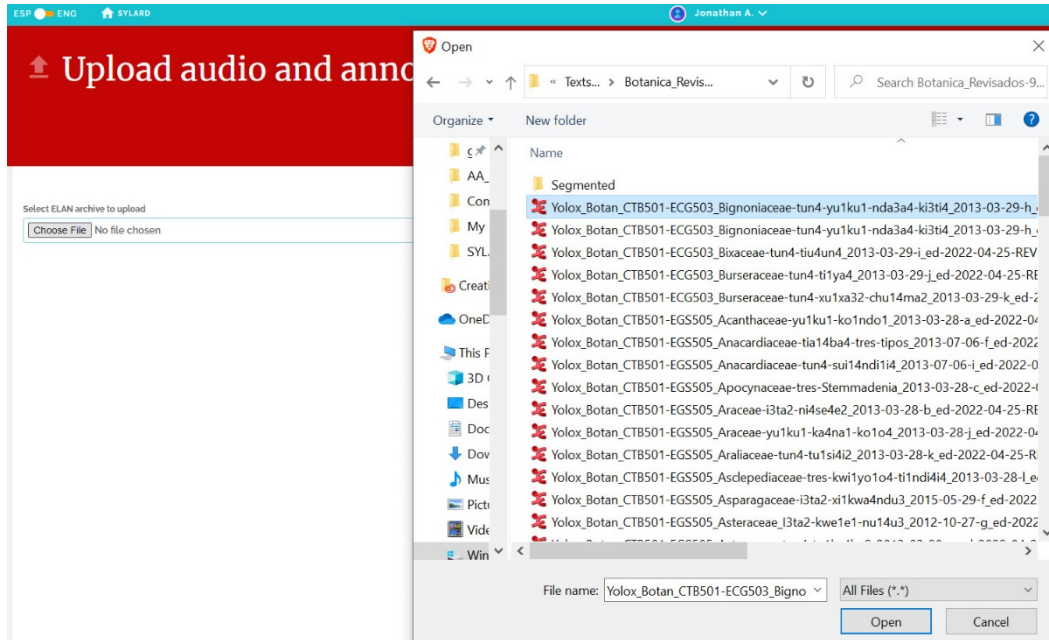
Step 2: Linking the ELAN annotation to an mp3 file, adding a title and description for the playback screen:
Once the ELAN file is uploaded to SYLARD and the corresponding Upload audio and annotation screen is
generated (fig. 40), the second step is for the Contributor to enter the complete URL path to the mp3 file
annotated in the uploaded ELAN file. The path must be complete and the .mp3 sound file must be on a
secure (https) server. Note that even though the ELAN annotation was from a .wav file, for Internet
playback a derived .mp3 should be used. The Contributor must then type in a Title of the audio (limited
to 120 characters) and a Description of the audio (limited to 200 characters).

Note that the Contributor can include HTML codes (e.g., <i/> and <b/>) in both the Title and Description
of the audio. These will be represented when the audio and annotation are viewed online (see fig. 45,
for example). Note that the Mixtec name, <i>itun³ yu¹ku¹ nda'³a⁴ ki³ti⁴</i> was entered in roman text
with the HTML tag <i/> surrounding the Yoloxóchitl Mixtec name. It is also important to note that all
special characters such as superscript numbers or letters must be entered as Unicode characters. For
example, superscripts [1], [2], [3], and [4] have the following hex codes in Unicode: 00b9, 00b2, 00b3, 2074.

Figure 40: Screenshot of upload template after having selected and uploaded an ELAN file

Step 3: A second set of metadata includes the following (fig. 40)*:*
- Collection in which the audio is included;
- Community of the speakers;
- Terminal language of the speakers;
- Genre of the recording.

The Contributor first selects the Collection in which the audio is included. Many if not most of the Contributors will have uploaded only one collection. For cases in which a Contributor has multiple collections, the drop-down menu for this metadata item will display all the collections that the Contributor has created. Contributors may add audio to a collection at any time, a particularly relelvant functionality given that Contributors may continually generate new audio for any given collection.

Once the "Collection in which the audio is included" is selected, the following two metadata fields will be provided with a drop-down menu that contains the communities and languages that were included in the Collection metadata when the collection was originally created. As evident in figure 42, the drop-down menu displays the same 4 communities that were originally included (fig. 34). The Contributor can click on the community relevent to the uploaded audio. Likewise, the Terminal language will have a drop-down list when applicable. However, most collections will have only one terminal language.

The final bit of metadata is taken from a controlled vocabulary of the genres of the recordings. The options, which are still only a preliminary list, are visible in the drop-down menu at the mid-right section

of figure 41. For the file about *Godmania aesculifolia*, the genre Ethnobotany/Etnobotánica was selected. In the next version of SYLARD this genre metadata will allow users to suggest new classifications. SYLARD will also allow users to filter content by particular topics across multiple language collections in SYLARD, creating a theme-specific multilanguague corpus.

Text of this manual continues below figure 42.

Figure 41: Screenshot of top half of the upload template (with metadata)

Figure 42: Metadata for audio annotations: Collection, Community, Terminal language, Genre

Step 4: Creation of default display of annotations (e.g., transcriptions) for line-by-line playback: The final step in uploading audio annotations is the selection of a default format for (1) where the annotations are displayed (one-line display or scrolling box) and (2) for the font colors. The template for this selection process is created during the upload of the original ELAN file: server-based SYLARD software reads the file and extracts the speakers and the tiers of annotation (fig. 41: Screenshot of upload template (with metadata). Note the following:

- Speakers: The top line of each group of tiers is the name of the speaker for the group of tiers listed immediately below. In the present case the two speakers are Esteban Castillo García and Constantino Teodoro Bautista. This control allows the Contributor to withhold from the default display the complete annotations for any speaker. It will also allow users to hide any speaker by using the Visual options (see below).
- Tiers: In the present case E. Castillo García has five tiers of annotation (visible in fig. 41) while C. Teodoro Bautista has four (not visible in the fig. 41 screenshot). The Contributor can individually select whether to show or hide each tier in the default view, the view that users see when they access line-by-line playback. Note that these same users (be they Visitors or Contributors) can activate the Visual options (fig. 46) and change, for their viewing, which tiers are shown or hidden, regardless of whether the tier was hidden or displayed in the default presentation. The user-activated Visual options control is discussed below in reference to figure 46.
- View in: The SYLARD display or play-back template comprises two sections:
  - One-line display
  - Scrolling
  Figure 45 illustrates these two display sections: One-line display and Scrolling. The one-line display is located below the sound bar and between two thin horizontal orange lines. In this particular case there are two Mixtec orthographies, the top is a practical orthography and the bottom is a surface linguistic orthography. The scrolling section contains the Spanish translation of the Mixtec text. As the recording progresses the one-line display changes as the recording reaches each new line of annotation. In the scrolling section, as the audio progresses the grey shaded highlighting moves down the translation, also line-by-line.
- Font color: Finally, the Contributor can select the color of the text annotation of each tier. There are two ways to select font color. The first is to enter the hex value of the color in the rectangle next to the color square, which changes color as the hex values are changed. The second way to select font color is through a color picker (see fig. 44). The possibilities with the picker are more

limited. But the advantage of the color picker is the ease with which it allows consistency in font color across many different audio annotations uploaded by a Contributor at different times.

Step 5: Once all the metadata and default display selections are entered, the Contributor needs to click on Save (bottom right of figs. 43 or 44). This preserves the metadata used in the default playback view (fig. 45) and available to users who wish to take advantage of the Visible options functionality (fig. 46).

Figure 43: Screenshot of upload template, bottom half, for default display of the annotations

Stage 7: This final stage presents the user experience in viewing and playing back annotated (transcribed, translated, and annotated) audio through SYLARD. Any registered user (Visitor or Contributor) can access the SYLARD catalogue, select a collection and audio to hear, and go to the playback page (fig. 45). The initial appearance of the line-by-line playback module is determined by how the Contributor created the metadata and display format discussed in Stage 6: Uploading annotated audio to a collection.

Upon accessing an annotated audio, the user is taken to a page such as that in figure 45. The sound is played by clicking on the right-pointing triangle at the left of the playback bar. Clicking on the three vertical dots to the right of the playback bar allows users to control the playback speed and to download the mp3. A future enhancement will allow downloading of ELAN and HTML annotation files.

Visual options: SYLARD also allows users to personally configure the annotation display during playback. This is done by expanding the Visual options function at the top left of the playbar. In uploading an ELAN annotation file, a Contributor sets the default playback view. In the present case (fig. 45) two orthographic representations (practical and surface, in the reddish font transcriptions below the

playbar) appear in the one-line section of the SYLARD playback page. The Spanish translation, on the other hand, appears in the scroll box just under the one-line display of the Mixtec texts. The highlighted line of the translation changes as the audio progresses, and the Mixtec transcriptions change as well, line by line as the recording progresses.

Text of this manual continues below figure 45.

Figure 44: Default display: Choosing the font color for each tier

Figure 45: Initial view of the interface users see when accessing a audio annotation through the catalogue

Visual options can be personalized by each user. The changes in the visual presentation of annotated files are temporary and specific to the user who makes them. A future enhancement to SYLARD will be to create accounts in which users can store the display preferences that they have created, file by file.

The default values for playback are automatically displayed in the expanded Visual options screen (fig. 46). Note that three tiers were uploaded but hidden from display: Esteban García Castillo GLOSS, Esteban García Castillo PARSE, and Constantino Teodoro Bautista PARSE. Indeed, users can easily see which tiers have been uploaded but are not visible in the default view by opening the Visual options screen (fig. 46) and noting the tiers with Off in the Show column. Three tiers were hidden (uploaded but not displayed) while six lines of annotation appear in the default view (fig. 45, in which four of these tiers appear; not shown are the practical and surface orthographies of Contantino Teodoro Bautista, which will appear in the One-line view once the recording reaches his text). With the Visual options function, a user who wishes to see any of these lines of text, however, can slide the switch to On and then select where the newly revealed tier will be displayed (One-line display or Scrolling) and the font color in which it will appear.

Figure 46: User activated visual options for changing SYLARD display for a particular audio annotation

Figure 47 illustrates the immediately visible impact of user-defined custom views. In this image the user has opened the Visual options menu and has changed the display.

- Surface orthography is now hidden (both for Esteban Castillo García and Constantino Teodoro Bautista;
- The practical orthography transcription for both speakers is now represented in the Scrolling box;
- The translation for both speakers is now represented in the Scrolling box, immediately below the Mixtec transcription;
- The user has changed the font colors (cf. fig. 45 and fig. 47)

The changes are immediately reflected in the annotation display: Note that the Mixtec annotation and Spanish translation are intercalated one before the other in the scroll box, with new font colors.

As an example, many users might prefer the practical orthographic and translation to be viewed one immediately above the other, in order to easily associate one with the other. This would mean that both are displayed in the scrolling box and that the colors should be selected to maximize contrast. The user who reconfigured the display in the Visual options panel for figure 47 desired this effect; the change has been immediately implemented but only for this particular user. For other users, the default remains as the initial view upon accessing this audio annotation. As already noted, a planned future development will be user accounts in which personalized Visual options can be stored as desired so that in reopening a file at a later date, the reconfigured display appears as the "default".

Figure 47: Line-by-line playback display after Visual options have been implemented by a user

SUMMARY

SYLARD has been created to faciliate public dissemination (to native speakers and communities, to the general public, to researchers) of time-coded annotations of audio recordings in Indigenous languages. Often these audio recordings and annotations (e.g., transcriptions and translations, along with other linguistic analyses such as parsing and glossing) are not disseminated to those who might be most interested. Though in the present case SYLARD is part of the Mesolex portal, it is open source software and can be implemented anywhere.

Section 4
Manual for Library and Video Resources

Mesolex includes textual and video resources for each of the terminal language communities. Textual resources include both published and unpublished material and video resources online videos pertinent to Indigenous language and culture. There are two ways to access these recources:

- Via the Topic index by clicking on Library or Videos (see fig. 1)
- Via the Terminal language home page by clicking on (Biblioteca/Library or Videos) along the bottom of this page for any language (see fig. 3)

Figure 1: Topic index home page
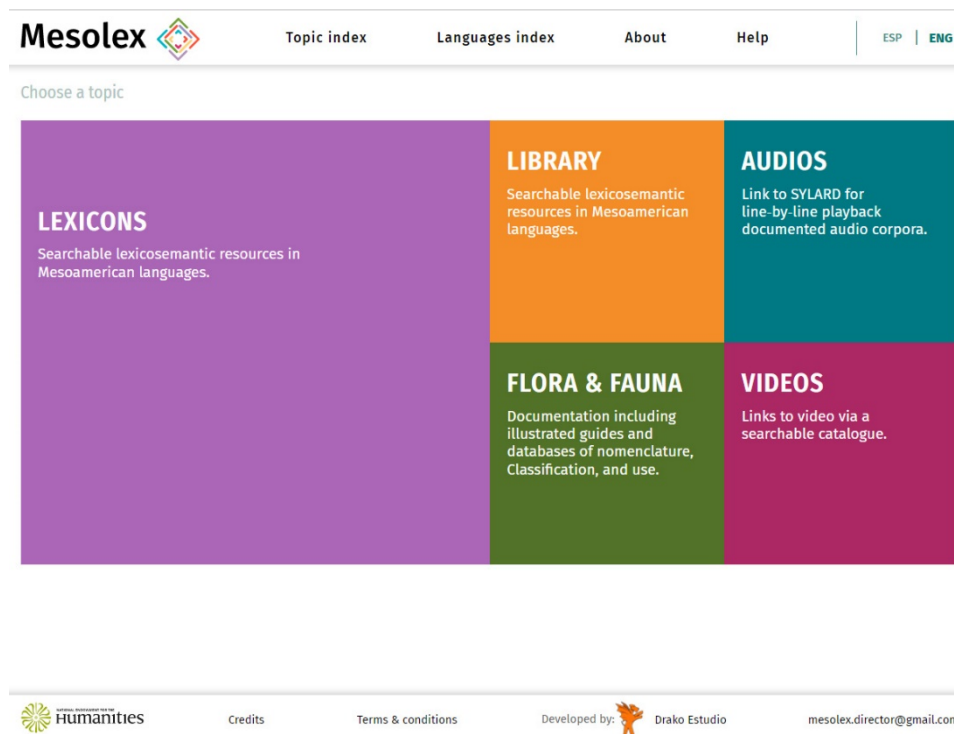
Access through the Topic index > Library path takes the user to the Library search interface (fig. 4), which is a list of all library resources (at present only 24 simply to demonstrate proof of concept). The user can then open up the drop-down menu and search in particular fields (title, author, community, language, keywords, or all fields) and then type in the search term.

Figure 2: Language index home page

Figure 3: Terminal language home page with filtered access to lexicosemantic resources (see fig. 6)

Figure 4: Library home page

Figure 5 demonstrates a search, generated by the progressive filtering of the results as the user types in a search term. In the example case, the user has selected the Keywords field and started to type in eth[nobotany]. After typing in the first three letters, the progressive filter of the search engine finds the only three resources that have this sequence in their keywords. The title of all three resources is in its original Spanish, which includes the word "etnobotánica," but the keywords are bilingual, in both English and Spanish depending on whether the user is working in the English or Spanish interface.

Sorting of library resources can be done on the content of any of the five columns. These same fields are those that can be searched for string matches:
- Title
- Author
- Language
- Community
- Keywords

The triangular "arrows" at the top of each column indicate whether the sorting is in ascending or descending order. The default sorting order is in ascending alphabetical order of the title (see fig. 4).

Mesolex will also have the capability, not yet integrated, to generate searches that are inclusive at the language group level. Language groups will often include multiple terminal languages, such as evidenced in the Nahuatl language group which lists 3 terminal languages (see §1, fig. 4). A search on any one of these languages will yield results such as those in figure 6 for Amuzgo. Next to the terminal language in the third column is a stacked book icon, which will be activated in the next generation of Mesolex

portals. By clicking on this stacked book icon the results of the library search will be expanded to all items related to the terminal language at the node established in Mesolex.

Figure 5: Library resources: Keyword search for <u>Ethnobotany</u>

Lbrary resources can also be accessed through terminal language pages (see fig. 3) by clicking on the Biblioteca/Library text in bold at the bottom of the terminal language page. Figure 6 illustrates the Mesolex library resources that have been generated by a search for Community = San Pedro Amuzgos through the regular interface Topic index > Biblioteca/Library > Community = San Pedro Amuzgos. Note again the autofull function integrated into the filtering system. The user has typed only <u>San P…</u> and the list has been created comprising the four items relevant to San Pedro Amuzgos.

As the first stage of Mesolex development is being completed, this same filter will be activated if a user goes to a terminal language homepage (see fig. 3 for a San Pedro Amuzgo example) and clicks on the Biblioteca/Library text along the bottom. Such a path automatically applies the relevant library search criteria for the terminal language community. Thus the screenshot in figure 6, of San Pedro Amuzgo resources, can be generated by accessing the library resource via the Topic index and searching/filtering at the Library catalogue page by typing in Community = San Pedro Amuzgos. It can also be accessed by going through the Language index to the terminal language page (Language Index > Amuzgoan > San Pedro Amuzgos) and clicking on the Biblioteca/Library text at the bottom of this page.

Figure 6: Library resources filtered for Community = San Pedro Amuzgos

Download resource and generate bibliographic references for library material in Mesolex
All library resources on Mesolex can be easily downloaded. The bibliographic reference can also be generated and copied to the clipboard from where it can be incorporated into a bibliography as desired.

Figure 7: Bibliographic reference for library resources on Mesolex

Figure 7 demonstrates how a bibliographic reference is generated, in this case for the first item on the San Pedro Amuzgos list. The user must only click on the information (circled i) icon at the right of the desired resource and the reference is generated as shown in figure 7. The download icon to the left of the information icon can also be clicked to immediately download the cited item, generally a .pdf file.

Videos
The video resources in Mesolex are at present all hosted elsewhere given the limited server resources available to Mesolex at this time. The search process for video is the same as that discussed above for library resources. Figure 8 illustrates the result of a search for videos from Yoloxóchitl. Again the autofill functionality generates the final list after typing only the first four letters of the language: Yolo[xóchitl Mixtec].

Figure 8: Video catalogue and search functionality

Section 5

Bilingual electronic book on the ethnobotany of Indigenous communities in Mexico

Among the resources offered in Mesolex is a bilingual (English-Spanish) book on the comparative ethnobotany of several Indigenous communities in Mexico. The prototype offers ethnobotanical information from: (1) several Nahuatl-speaking villages in the municipality of Cuetzalan del Progreso, Puebla; (2) Zongozotla, a Totonac-speaking municipal capital in the northern highlands of Puebla; and (3) Yoloxóchitl, a Mixtec community in the municipality of San Luis Acatlán along the Pacific Coast of Mexico.

An operational prototype of this book has been developed that can easily be expanded as additional content is developed, should funding become available. What follows is a description of the e-book and how it can be used. The prototype proof-of-concept model is accessible directly from the Mesolex home page (now at https://mesolex-drako-iano9.ondigitalocean.app/en) and then following the path Topic Index > Flora & Fauna.

Figure 1: Home page of ethnobotanical e-book (top)

Catálogo de etnobotánica

Contenido

The default e-book home page is a photographic index of the plants that will be included should further funding become available (fig. 1). An alternative textual index (fig. 2) is available by clicking on the square located at the top left corner of the photographic index home page, under Content/Contenido. This alternative view provides a textual table of contents organized by family (in bold, starting with Adoxaceae) and, within each family, by genus and species.

At present and for proof-of-concept, the content of only one species has been developed: *Pseudobombax ellipticum* (Kunth) Dugand, a well-known tree in the Malvaceae family. A user can click on the photo of the *P. ellipticum* flower (fig. 3) or on the species name (fig. 2) and be taken to the initial page of this species (fig. 4). Eventually there will be similar links for all the species (figs. 1 and 2). All the plants listed are common species amply distributed throughout Mexico. They are named in a wide range of Indigenous languages and often used locally for a variety of purposes.

The text of this manual continues between figures 3 and 4.

Figure 2: Textual representation of table of contents with botanical family and scientific species name
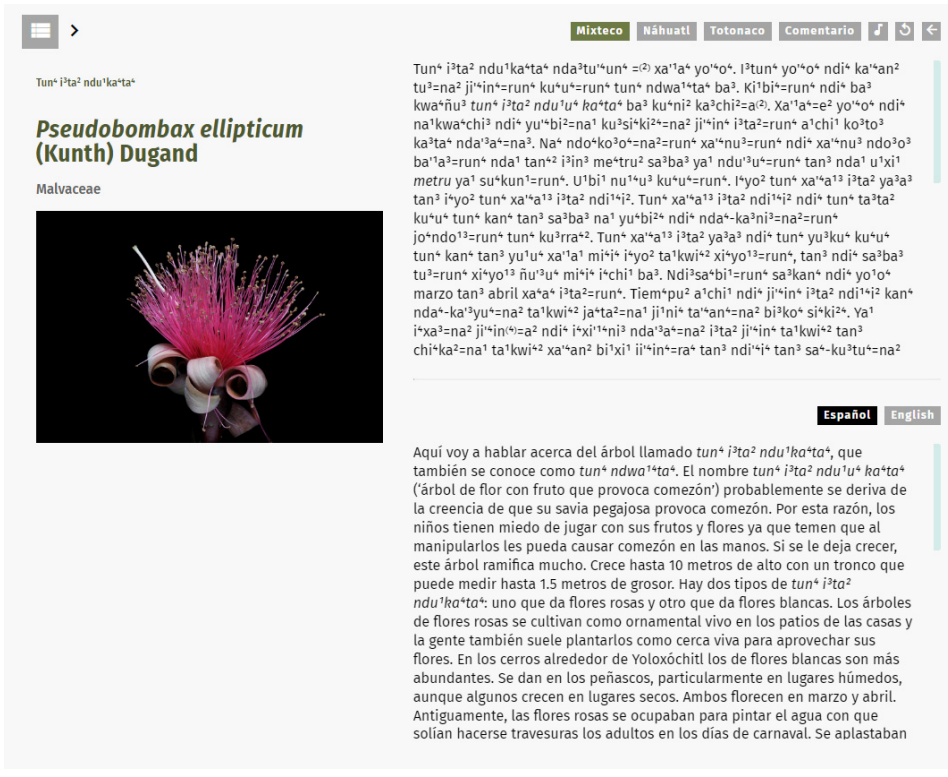

Catálogo de etnobotánica

Contenido

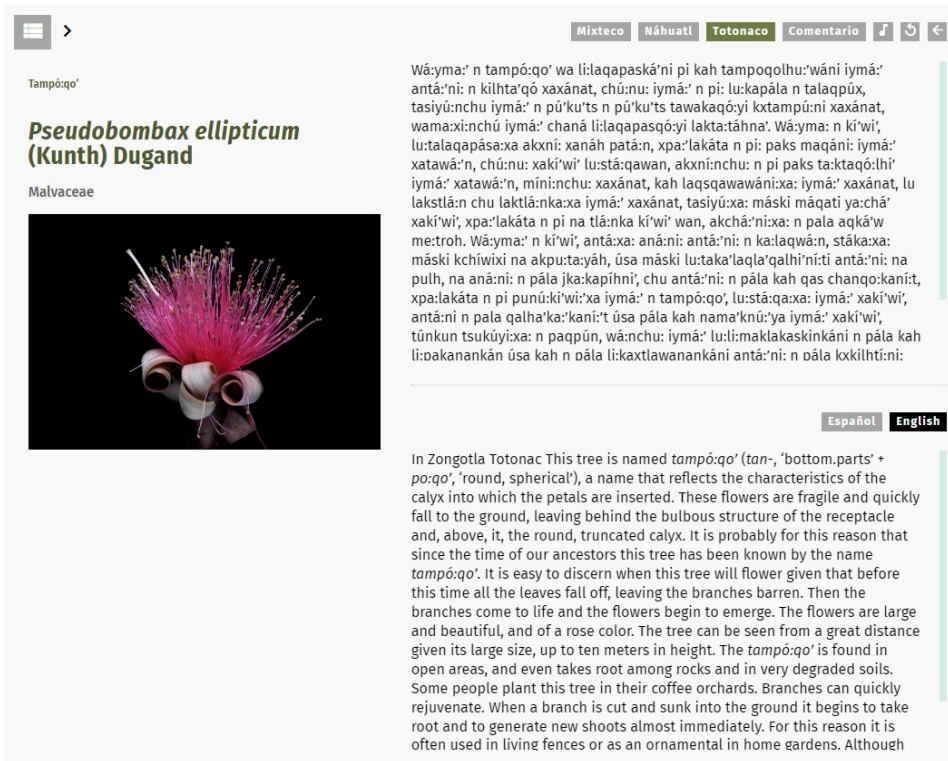|		
Sambucus nigra subsp. *canadensis* (L.) Bolli **Adoxaceae**	*Parmentiera aculeata* (Kunth) Seem. **Bignoniaceae**	*Licania platypus* (Hemsl.) Fritsch **Chrysobalanaceae**
Amaranthus hybridus L. **Amaranthaceae**	*Tabebuia rosea* (Bertol.) DC. **Bignoniaceae**	*Commelina erecta* L. **Commelinaceae**
Amaranthus spinosus L. **Amaranthaceae**	*Wigandia urens* (Ruiz & Pav.) Kunth **Boraginaceae**	*Tradescantia zebrina* hort. ex Bosse **Commelinaceae**
Spondias purpurea L. **Anacardiaceae**	*Bursera simaruba* (L.) Sarg. **Burseraceae**	*Cuscuta* spp. **Convolvulaceae**
Plumeria rubra L. **Apocynaceae**	*Rhipsalis baccifera* (J.S. Mueller) Stearn **Cactaceae**	*Ipomoea batatas* (L.) Lam. **Convolvulaceae**
Bidens odorata Cav. **Asteraceae**	*Celtis iguanaea* (Jacq.) Sarg. **Cannabaceae**	*Ipomoea indica* (Burm.f.) Merr. **Convolvulaceae**
Porophyllum ruderale (Jacq.) Cass. **Asteraceae**	*Trema micrantha* (L.) Blume **Cannabaceae**	*Merremia umbellata* (L.) Hallier f. **Convolvulaceae**
Tagetes erecta L. **Asteraceae**	*Canna indica* L. **Cannaceae**	*Costus* spp. **Costaceae**
Crescentia cujete L. **Bignoniaceae**	*Couepia polyandra* (Kunth) Rose **Chrysobalanaceae**	*Pseudobombax ellipticum* (Kunth) Dugand **Malvaceae**

Figure 3: Home page of ethnobotanical e-book (bottom)

After clicking on a particular species, in either the photographic or textual index, a user is taken to a page such as that represented in figure 4. The default representation seen in this screenshot is the leftmost Indigenous language (in this case Yoloxóchitl Mixtec) accompanied by a Spanish translation of the native language text. The interface offers the user a "mix-and-match" template in which the Indigenous languages for which textual and audio materials have been created are listed in the upper right corner in rectangular tabs. At present these tabs represent Yoloxóchitl Mixtec, Highland Puebla Nahuatl, and Highland Puebla Totonac. Clicking on any tab brings up the corresponding Indigenous text. Users can switch among Indigenous languages independently of the accompanying translation language. The language of the translation text is controlled by similar rectangular tabs for Español and English in the middle of the page. Figure 4 shows the Spanish translation of a Mixtec text; figure 5 shows the English translation of a Totonac text. All combinations of an Indigenous and a translation texts are possible.

For each Indigenous language text an audio rendition is publically available. The audio is activated (reproduced in the browser) by clicking on the musical note to the right of the Comentarios tab. Figure 6 shows the music note highlighted as the cursor was moved over it and just before it was clicked. The user can pause the audio at any time during playback. Figure 7 shows the pause button that appears as the audio is being rendered. Clicking it pauses the audio playback and reverts to the musical note icon. Clicking on this icon continues to play the audio where it had left off.

The text of this manual continues between figures 8 and 9.

Figure 4: Default home page with Mixtec text and Spanish translation

Mixteco | Náhuatl | Totonaco | Comentario

Tun⁴ i³ta² ndu¹ka⁴ta⁴

Pseudobombax ellipticum (Kunth) Dugand

Malvaceae

Tun⁴ i³ta² ndu¹ka⁴ta⁴ nda³tu'⁴un⁴ =⁽²⁾ xa'¹a⁴ yo'⁴o⁴. I³tun⁴ yo'⁴o⁴ ndi⁴ ka'⁴an² tu³=na² ji'⁴in⁴=run⁴ ku⁴u⁴=run⁴ tun⁴ ndwa¹⁴ta⁴ ba³. Ki¹bi⁴=run⁴ ndi⁴ ba³ kwa⁴ñu³ tun⁴ i³ta² ndu¹u⁴ ka⁴ta⁴ ba³ ku⁴ni² ka³chi²=a⁽²⁾. Xa'¹a⁴=e² yo'⁴o⁴ ndi⁴ na¹kwa⁴chi³ ndi⁴ yu'⁴bi²=na¹ ku³si⁴ki²⁴=na² ji'⁴in⁴ i³ta²=run⁴ a¹chi¹ ko³to³ ka³ta⁴ nda¹³a⁴=na³. Na⁴ ndo⁴ko³o⁴=na²=run⁴ xa'⁴nu³=run⁴ ndi⁴ xa'⁴nu³ ndo³o³ ba¹³a³=run⁴ nda¹ tan⁴² i³in³ me⁴tru² sa³ba³ ya¹ ndu¹³u⁴=run⁴ tan³ nda¹ u¹xi¹ *metru* ya¹ su⁴kun¹=run⁴. U¹bi¹ nu¹⁴u³ ku⁴u⁴=run⁴. I⁴yo² tun⁴ xa'⁴a¹³ i³ta² ya³a³ tan³ i⁴yo² tun⁴ xa'⁴a¹³ i³ta² ndi¹⁴i². Tun⁴ xa'⁴a¹³ i³ta² ndi¹⁴i² ndi⁴ tun⁴ ta³ta² ku⁴u⁴ tun⁴ kan⁴ tan³ sa³ba³ na¹ yu⁴bi²⁴ ndi⁴ nda⁴-ka³ni³=na²=run⁴ jo⁴ndo¹³=run⁴ tun⁴ ku³rra¹². Tun⁴ xa'⁴a¹³ i³ta² ya³a³ ndi⁴ tun⁴ yu³ku⁴ ku⁴u⁴ tun⁴ kan⁴ tan³ yu¹u⁴ xa'¹a¹ mi⁴i⁴ i⁴yo² ta¹kwi⁴² xi⁴yo¹³=run⁴, tan³ ndi⁴ sa³ba³ tu³=run⁴ xi⁴yo¹³ ñu'³u⁴ mi⁴i⁴ i⁴chi¹ ba³. Ndi³sa⁴bi¹=run⁴ sa³kan⁴ ndi⁴ yo'o⁴ marzo tan³ abril xa⁴a⁴ i³ta²=run⁴. Tiem⁴pu² a¹chi¹ ndi⁴ ji'⁴in⁴ i³ta² ndi¹⁴i² kan⁴ nda⁴-ka'³yu⁴=na² ta¹kwi⁴² ja⁴ta²=na¹ ji'¹ni⁴ ta'⁴an²=na² bi³ko⁴ si⁴ki²⁴. Ya¹ i⁴xa³=na² ji'⁴in⁽⁴⁾=a² ndi⁴ i⁴xi'¹⁴ni³ nda'³a⁴=na² i³ta² ji'⁴in⁴ ta¹kwi⁴² tan³ chi⁴ka²=na¹ ta¹kwi⁴² xa'⁴an² bi¹xi¹ ii'⁴in⁴=ra⁴ tan³ ndi'⁴i⁴ tan³ sa⁴-ku³tu⁴=na²

Español | English

Aquí voy a hablar acerca del árbol llamado *tun⁴ i³ta² ndu¹ka⁴ta⁴*, que también se conoce como *tun⁴ ndwa¹⁴ta⁴*. El nombre *tun⁴ i³ta² ndu¹u⁴ ka⁴ta⁴* ('árbol de flor con fruto que provoca comezón') probablemente se deriva de la creencia de que su savia pegajosa provoca comezón. Por esta razón, los niños tienen miedo de jugar con sus frutos y flores ya que temen que al manipularlos les pueda causar comezón en las manos. Si se le deja crecer, este árbol ramifica mucho. Crece hasta 10 metros de alto con un tronco que puede medir hasta 1.5 metros de grosor. Hay dos tipos de *tun⁴ i³ta² ndu¹ka⁴ta⁴*: uno que da flores rosas y otro que da flores blancas. Los árboles de flores rosas se cultivan como ornamental vivo en los patios de las casas y la gente también suele plantarlos como cerca viva para aprovechar sus flores. En los cerros alrededor de Yoloxóchitl los de flores blancas son más abundantes. Se dan en los peñascos, particularmente en lugares húmedos, aunque algunos crecen en lugares secos. Ambos florecen en marzo y abril. Antiguamente, las flores rosas se ocupaban para pintar el agua con que solían hacerse travesuras los adultos en los días de carnaval. Se aplastaban

Figure 5: Selection for Totonac text and English translation

Mixteco | Náhuatl | Totonaco | Comentario

Tampó:qo'

Pseudobombax ellipticum (Kunth) Dugand

Malvaceae

Wá:yma:' n tampó:qo' wa li:laqapaskáʼni pi kah tampoqolhu:'wáni iymá:' antá:'ni: n kilhtaʼqó xaxánat, chú:nu: iymá:' n pi: lu:kapála n talaqpúx, tasiyú:nchu iymá:' n púʼku'ts n púʼku'ts tawakaqó:yi kxtampú:ni xaxánat, wama:xi:nchú iymá:' chaná li:laqapasqó:yi lakta:táhna'. Wá:yma: n kíʼwi', lu:talaqapása:xa akxní: xanáh patá:n, xpa:'lakáta n pi: paks maqáni: iymá:' xatawá:'n, chú:nu: xakíʼwi' lu:stá:qawan, akxní:nchu: n pi paks ta:ktaqó:lhi' iymá:' xatawá:'n, míni:nchu: xaxánat, kah laqsqawawáni:xa: iymá:' xaxánat, lu lakstlá:n chu laktlá:nka:xa iymá:' xaxánat, tasiyú:xa: máski máqati ya:chá' xakíʼwi', xpa:'lakáta n pi na tlá:nka kíʼwi' wan, akchá:'ni:xa: n pala aqká'w me:troh. Wá:yma:' n kíʼwi', antá:xa: aná:ni: antá:'ni: n ka:laqwá:n, stáka:xa: máski kchíwixi na akpu:ta:yáh, úsa máski lu:takaʼlaqlaʼqalhiʼni:ti antá:'ni: na pulh, na aná:ni: n pála jka:kapíhni', chu antá:'ni: n pála kah qas chanqo:kaní:t, xpa:lakáta n pi punú:kiʼwi:'xa iymá:' n tampó:qo', lu:stá:qa:xa: iymá:' xakíʼwi', antá:ni n pala qalhaʼka:'kaní:'t úsa pála kah nama'knú:'ya iymá:' xakíʼwi', túnkun tsukúyi:xa: n paqpún, wá:nchu: iymá:' lu:li:maklakaskinkáni n pála kah li:pakanankán úsa kah n pála li:kaxtlawanankáni antá:'ni: n pála kxkilhtí:ni:

Español | English

In Zongotla Totonac This tree is named *tampó:qo'* (*tan-*, 'bottom.parts' + *po:qo'*, 'round, spherical'), a name that reflects the characteristics of the calyx into which the petals are inserted. These flowers are fragile and quickly fall to the ground, leaving behind the bulbous structure of the receptacle and, above, it, the round, truncated calyx. It is probably for this reason that since the time of our ancestors this tree has been known by the name *tampó:qo'*. It is easy to discern when this tree will flower given that before this time all the leaves fall off, leaving the branches barren. Then the branches come to life and the flowers begin to emerge. The flowers are large and beautiful, and of a rose color. The tree can be seen from a great distance given its large size, up to ten meters in height. The *tampó:qo'* is found in open areas, and even takes root among rocks and in very degraded soils. Some people plant this tree in their coffee orchards. Branches can quickly rejuvenate. When a branch is cut and sunk into the ground it begins to take root and to generate new shoots almost immediately. For this reason it is often used in living fences or as an ornamental in home gardens. Although

Figure 6: Sound file playing for Totonac text

Figure 7: Sound file paused for Totonac text

Figure 8: Comentario/Comentary page: English text

All plant species in the electronic book include a Commentary / Comentario tab that presents a descriptive text about the plant species starting from data on its geographic distribution and continuing on to a description of the plant, including its flowers and fruit, and ending with a summary of its use. The same text is offered in both English (fig. 8) and Spanish (fig. 9).

Finally, each species is iconically represented by a photo of its flower. Clicking on the flower takes the uder to an enlarged representation (fig. 10). In the future each species will include a gallery of photos, accessed by clicking on same the iconic image used in the photographic index and included at the left of each species page.

Figure 9: Comentario/Comentary page: Spanish text

Figure 10: Photo enlarged at any time by clicking on photo to the left of the screen