

NEH Application Cover Sheet

Digital Humanities Start-up Grants

PROJECT DIRECTOR

Dr. Brook Danielle Lillehaugen
Assistant Professor of Linguistics
370 Lancaster Avenue
Haverford, PA 19041-1336
UNITED STATES

E-mail: blilleha@haverford.edu
Phone(W): 610-795-6138
Phone(H):
Fax:

Field of Expertise: Languages: Linguistics

INSTITUTION

Corporation of Haverford College
Haverford, PA UNITED STATES

APPLICATION INFORMATION

Title: *Ticha: a digital text explorer for Colonial Zapotec*

Grant Period: From 6/2014 to 11/2015

Field of Project: Languages: Linguistics; Interdisciplinary: Latin American Studies

Description of Project: Zapotec, an indigenous language family of Mexico, has a long record of alphabetic texts, the earliest dated 1565. Reading and interpreting these colonial documents can be extremely difficult because of the challenges of early Zapotec orthography, vocabulary, grammar, and printing conventions. Our project, *Ticha: a digital text explorer for Colonial Zapotec*, will provide a reader with an online tool to access and explore many interlinked layers of these texts, including images of the original documents, transcriptions, translations into English and modern Spanish, linguistic analysis, and commentary. This tool will make Colonial Zapotec texts accessible to both members of the modern Zapotec community and to scholars in diverse fields, including linguistics, anthropology, and history. We are seeking funding for the planning, design, and alpha-level implementation of *Ticha*.

BUDGET

Outright Request	\$29,995.00	Cost Sharing	\$0.00
Matching Request	\$0.00	Total Budget	\$29,995.00
Total NEH	\$29,995.00		

GRANT ADMINISTRATOR

Mr. Brian Jackson
Grant and Financial Accountant
370 Lancaster Avenue
Haverford, PA 19041-1336
UNITED STATES

E-mail: bdjackso@haverford.edu
Phone(W): 610-896-2920
Fax:

1. Table of contents

<i>Attachment</i>	<i>Description</i>	<i>page</i>
#		<i>number(s)</i>
1	Table of contents	1
2	List of participants	2
3	Abstract and statements of innovation and humanities significance	3
4	Narrative	4-6
5	Project budget and budget narrative	7-10
6	Biographies	11-12
7	Data management plan	13
8	Letters of commitment and support	
	-Broadwell letter of commitment	14
	-Oudijk letter of commitment	15
	-Allen letter of commitment	16
	-Valdivia letter of commitment	17
	-Harrison letter of support	18
	-Bas van Doesburg letter of support	19
9	Appendices	
	-Appendix A: Sample page, <i>Arte</i> 102r	20-21
	-Appendix B: Sample page, <i>Doctrina</i> 65r	22-23
	-Appendix C: TEI sample page	24
	-Appendix D: <i>Ticha</i> sample page, <i>Arte</i> 102r	25
	-Appendix E: Archive.org screenshot showing usage rights	26
	-Appendix F: Agenda for meeting in Oaxaca	27
	-Appendix G: Works cited	28

2. List of participants

Brook Danielle Lillehaugen
Haverford College

George Aaron Broadwell
University at Albany, SUNY

Michel R. Oudijk
Universidad Nacional Autónoma de México

Laurie Allen
Haverford College

Enrique Valdivia
University of Michigan

3. Abstract and statement of innovation and humanities significance

Abstract

Zapotec, an indigenous language family of Mexico, has a long record of alphabetic texts, the earliest dated 1565. Reading and interpreting these colonial documents can be extremely difficult because of the challenges of early Zapotec orthography, vocabulary, grammar, and printing conventions. Our project, *Ticha: a digital text explorer for Colonial Zapotec*, will provide a reader with an online tool to access and explore many interlinked layers of these texts, including images of the original documents, transcriptions, translations into English and modern Spanish, linguistic analysis, and commentary. This tool will make Colonial Zapotec texts accessible to both members of the modern Zapotec community and to scholars in diverse fields, including linguistics, anthropology, and history. We are seeking funding for the planning, design, and alpha-level implementation of *Ticha*.

Statement of innovation

Linguists, ethnohistorians, and other scholars may use the same archival texts, but methods of analysis differ. *Ticha* combines two encoding standards across multiple translations and analyses of texts by incorporating XML output from FLE_x (a common linguistic tool) with TEI (a standard for encoding historic texts), serving to connect compatible (but currently non-connected) encoding practices of distinct academic fields. A web-based interface makes this rich information available to a broad scholarly and general audience.

Statement of humanities significance

Zapotec is a major culture of Mexico, with a large, mostly inaccessible, corpus of archival manuscripts. The exposition of these Colonial Zapotec texts will allow scholars in e.g. anthropology, history, musicology, religion, and the interested public access to materials currently unusable to all but a few specialists. *Ticha* will unlock these texts in a multi-dimensional fashion allowing users to explore their language and content, depending on their needs and interests.

4. Narrative

Background. Zapotec is a language family indigenous to southern Mexico. Today, there are over 50 different Zapotec languages, most endangered, spoken primarily in the state of Oaxaca, Mexico, by a total of approximately 250,000 people within a much larger Zapotec ethnic community. The Zapotec language family is on par with the Romance language family in terms of time depth and diversity of member languages. The Zapotecs are one of the major civilizations of Mesoamerica with cultural traditions going back to 500 B.C. and distinct from the better known Nahuatl (Aztec) and Maya.

With the arrival of the Spanish in 1519, the Colonial period began. Alphabetic writing was introduced and quickly adopted by indigenous peoples. Zapotec has one of the longest records of alphabetic written documents for any indigenous language of the Americas. Over 900 documents written in Zapotec by native scribes have been identified, the earliest from 1565 (Oudijk 2008:230). The richest variety of colonial Zapotec documents are those written in the kind of Zapotec spoken in and around Oaxaca City, known as Valley Zapotec. There is an extensive dictionary (Córdova 1578b), grammar (Córdova 1578a), doctrine (Feria 1567), and over 200 native-speaker documents (mostly wills) written in Colonial Valley Zapotec. For this reason, we will focus our initial investigation on this variety of Zapotec.

Reading and translating these Colonial Valley Zapotec documents can be extremely difficult. The language in these documents is significantly different from modern Valley Zapotec. The orthography of such texts is very inconsistent, and there is no fully adequate Zapotec-Spanish or Zapotec-English dictionary for the Colonial varieties of the language (Broadwell and Lillehaugen 2013). The grammar of these documents is also different from that of modern Zapotec languages. Thus potential users of such documents cannot read them without extensive training.

A second set of problems comes from the fact that the colonial documents are either handwritten or printed according to conventions that are often opaque to contemporary users. Texts often use extensive abbreviations and may also contain printing errors (such as reversed letters and broken type) or contain hand-written corrections. Because of the difficulty in using the original manuscripts and in understanding the language, this corpus of documents written in Colonial Valley Zapotec has not been accessible outside of a small circle of specialists.

Enhancing the humanities through innovation. These documents hold untapped treasures of information for a wide range of academics, including linguists, anthropologists, historians, as well as the general public. They provide insight into the ethnic diversity, religious history, and familial, social and economic structures of Mexico for a 500 year period. Thus they create a bridge not only through time, but also across multiple cultural borders: a link between modern scholars, colonial priests, and Zapotec people. The large corpus of Colonial Nahuatl language material has proven useful to scholars across many disciplines. Colonial Zapotec is less-studied and is understood by vastly fewer scholars, thus linguistic analysis is particularly needed for these documents to help users understand the texts and to allow potential users to critically evaluate any translations of the original text.

The paucity of accessible and analyzed Colonial Zapotec documents has left a hole in the scholarship of Mexico; addressing this absence allows us to take advantage of tremendous tools available within the digital humanities in ways that have not yet been combined. We are engaged in a long-term project of creating a digital tool, which we have titled *Ticha: a digital text explorer for Colonial Zapotec*. This tool will allow the reader to explore Colonial Valley Zapotec documents through an online viewer which actively connects images of original documents, annotated transcriptions of such documents, translations into English and modern Spanish, as well as linguistic analysis and commentary. (An early mock-up of what this might look like can be seen in Appendix D.) This tool will take advantage of modern, digital modes of publication to make Colonial Valley Zapotec texts accessible to members of the Zapotec community, the general public, and scholars. There will be multiple entry points and customizable viewing options depending on the user's need.

Ticha is innovative in bringing together data analyzed in FLEx (Fieldworks Language Explorer, fieldworks.sil.org) a system for lexical and grammatical analysis, with current TEI standards (Text Encoding Initiative, tei-c.org) for paleographic and translational representations of texts. By participating

in both the robust community of TEI encoding, and working within the FLEx community, the *Ticha* project has the opportunity to produce an interface that could not have existed even ten years ago. This will allow the robust community of historical, literary, and other humanistic scholars to access the documents at a deeper level than would be available without the encoding.

We seek funding from the NEH Digital Humanities start-up grant program (Level 1) for project planning and design, and the development of an alpha-level prototype of this online resource. There will be three main outcomes in this phase of the project. The first will be the linguistic analysis, translation, and TEI encoding of two Colonial Zapotec texts, parts of which we will make available through *Ticha* at the alpha-level stage. The texts are difficult enough to understand that even without full implementation of the text explorer, the component non-interactive resources will be useful to some researchers. Moreover, by working through the process of analyzing, translating, transcribing, and encoding two texts we will better understand the challenges and options available to us. Secondly, and mainly, we will develop a plan for subsequent steps and future development. Our seed funding (June-Sept 2013) allowed us to prove that our concept was viable. The NEH support would give us time and resources to work out the details of the design of the online tool. During this time we will consult with potential users, including U.S. and international scholars and Zapotec community members, to better understand how they may want to use this tool. We will also seek out long-term relationships with potential XSLT and TEI consultants for the future large scale implementation stage of our project. The third outcome will be a publically available version of the scripting, in order to transform the XML generated by FLEx into a baseline TEI. Connecting these two resources is necessary for our project, but our results can be adapted by others for their respective projects, potentially facilitating them to transport their linguistic corpuses from a FLEx database to an interactive website.

For the purposes of developing the optimal functionality for *Ticha*, we concentrate on the transcription and analysis of texts from different genres for the alpha-level prototype. The first is the *Arte en lengva zapoteca*, a grammar of Colonial Valley Zapotec written by Dominican priest Fray Juan de Córdova in 1578 (Córdova 1578a). The second is the *Doctrina christiana*, a bilingual Spanish-Zapotec explanation of Christian doctrine (Feria 1567). The *Doctrina* is the longest text in Colonial Valley Zapotec and in contrast to the *Arte*, provides many long connected passages of Zapotec. In subsequent phases of the project, we plan to expand to include other types of documents, particularly wills.

Both the *Arte* and *Doctrina* are of linguistic and cultural significance, but are very difficult to use. Color-scanned PDF versions are available online through the John Carter Brown library via Internet Archive (archive.org), but this alleviates only one of the challenges to using the manuscripts, i.e. physical access. (See Appendix E for general permissions to use these materials.) Several problems remain. For one, the texts are not fully searchable nor are they indexed. If one wants to find information on a particular topic (e.g. names, religious holidays or passive forms of verbs) one must look page by page.

The *Arte* presents a particular challenge. Córdova presented his understanding of the Zapotec language in the Latin grammatical framework he was familiar with. Thus the *Arte* includes discussions of grammatical categories, such as case declensions, that do not exist in Zapotec, and neglects other important features of Zapotec that are not present in Latin. The organization of the *Arte* makes it difficult to locate any particular piece of information. A searchable, marked up version of Córdova would greatly enhance the usability of this resource. (An portion of the *Arte* in TEI markup can be seen in Appendix C.) Finally, most researchers do not know the Zapotec language well enough to understand the Zapotec examples given, which may show linguistic complexities not discussed by Córdova.

Linguistic analysis of the Zapotec language portions sheds light on the depth of information locked in the text. For example, the Zapotec examples often show something other than (or in addition to) what Córdova is using them to illustrate. Consider the entry for the number 138 in Appendix A. Five possible expressions of this number are presented, and linguistic analysis – not simple translation – reveals that these forms all relate to a base twenty system of counting, providing important information about native mathematical concepts. (See Appendix A for details.) Another example of the value added by the linguistic analysis can be seen in a passage from the *Doctrina* where the meanings of the parallel Zapotec and Spanish sentences differ in some important ways from each other, see Appendix B. These

are but two examples of the type of information contained in the Zapotec language itself which can be revealed through morphological analysis and translation. A complete linguistic analysis of the Zapotec textual sources would be a substantial resource for researchers in many fields, including linguists, historians, and anthropologists, as well as for the Zapotec community.

Environmental scan. Among indigenous languages of the Americas, there are few comparable projects. Our project complements the available Nahuatl texts at the Early Nahuatl Library (<http://enl.uoregon.edu/>) and will add to Colonial Mexican materials available. Our project differs from the Nahuatl site in providing extensive linguistic analysis for the Zapotec texts included and are encoded in TEI for richer use by non-linguistics scholars. The text repositories from other parts of the world, such as the Indic corpus at <http://sarit.indology.info/top.shtml> and the Sumerian corpus at <http://etcsl.orinst.ox.ac.uk/>, include large searchable corpora with transcriptions and translations, but no linguistic analysis. There are also a number of examples of projects designed to take advantage of TEI as a tool for interaction with texts, such as <http://www.petrusplaoul.org/>. However, these projects do not address multiple translations of a text, nor do they include linguistic analysis.

History and duration of the project. After early (online) discussions and grant applications, we received an internal digital humanities pilot grant for \$7807.50 in **June 2013**, which funded an initial team meeting in Mexico City. This initiated our summer 2013 *Ticha* pilot on selected *Arte* pages, with development of transcription, analysis and encoding standards. Between **Sept. 2013 and May 2014** we will continue the lexical and grammatical development of the Colonial Valley Zapotec FLEx project, the transcription and translation of documents, and the TEI markup. The NEH funded period of the project will occur between **June 2014 and Nov. 2015**. (See "Staff and Work Plan" below.) Prior to **Nov. 2015** we hope to have secured additional funding to begin the larger implementation phase of the project, which would include a functional webpage with multiple entry points, customizable viewing options, a Spanish interface, and an expanded corpus of Colonial Valley Zapotec documents.

Staff and work plan. The interdisciplinary team consists of **Broadwell** and **Lillehaugen**, linguists; **Oudijk**, a philologist and ethnohistorian, and **Allen** and **Valdivia**, TEI and digital scholarship experts.

Phase One: June 1, 2014 - January 12, 2015. During the beginning stages of the grant, the linguists will be completing the linguistic analysis of the *Arte* and advancing on the analysis of the *Doctrina*. These texts will be used evaluate a possible workflow for moving documents through analysis, transcription, translation, checking, and encoding to ensure that important research questions are highlighted. Allen and Valdivia will continue to check TEI, and ensure that the encoding matches the texts and includes pointers to all versions. Mockups describing interface features and highlighting different aspects of the projects will be created. The team will prepare questions to explore with scholars and community members in Oaxaca during January. Monthly online meetings will be held.

Phase Two: Meeting in Oaxaca, January 12-16, 2015. During a five day meeting in Oaxaca the team will meet in person to evaluate the direction of the project and test project assumptions with Mexican scholars and community members. The draft agenda for this meeting is Appendix F.

Phase Three: January 17 - November 30, 2015. We will use the information gathered from our discussions and meetings with potential users to revise the first phase models prepared in phase one. Broadwell, Lillehaugen, and Oudijk will revise necessary portions of the *Arte* and Lillehaugen will provide second-level review for *Doctrina* portions. The linguist team will double-check for accuracy of transcription and consistency in analysis. The team will evaluate the outcomes of the NEH funded phase of the project, specifically in regards to the points listed below.

Final product and dissemination. The final products will be 1) an alpha-level prototype of this online resource plus specifications for functionality, 2) all the component resources (transcriptions, analyses, markups, and translations), 3) the XSLT resources for transforming FLEx XML to TEI, 4) a development plan for future expansion and resources, and 5) a publically available "white paper" that evaluates how well these technologies enhance the accessibility of the texts, specifying areas of success and remaining challenges. We will make the prototype version available to other researchers for testing and feedback and put all the component and XSLT resources on publicly available archives.

5. Budget



Budget Form

OMB No 3136-0134

Expires 7/31/2015

Applicant Institution: *Haverford College*Project Director: *Brook Lillehaugen*Project Grant Period: *06/01/2014-11/30/2015*[click for Budget Instructions](#)

	Computational Details/Notes	(notes)	Year 1	(notes)	Year 2	Project Total
			06/01/2014 - 05/31/2015		06/01/2015 - 11/30/2015	
1. Salaries & Wages						
Student research assistants	Spanish and English language proofreading and editing. \$9/hr for 220 hrs in Year 1 and 110 hrs in Year 2.		\$1,980		\$990	\$2,970
2. Fringe Benefits						
3. Consultant Fees						
Linguistic consulting (Broadwell)	based on 1 month of work per year		\$3,000		\$3,000	\$6,000
Ethnohistorical consulting (Oudijk)	based on 1 month of work per year		\$3,000		\$3,000	\$6,000
TEI consulting (Valdivia)	based on 2 months of work per year		\$3,000		\$3,000	\$6,000
						\$0
4. Travel						
NEH Planning Meeting in DC			\$340			\$340
Meeting in Oaxaca, Mexico			\$8,685			\$8,685
5. Supplies & Materials						
						\$0

5. Budget

6. Services					\$0
7. Other Costs					\$0
8. Total Direct Costs	Per Year		\$20,005	\$9,990	\$29,995
9. Total Indirect Costs	Per Year		\$0	\$0	\$0
10. Total Project Costs	(Direct and Indirect costs for entire project)				\$29,995
11. Project Funding	a. Requested from NEH		Outright:		\$0
			Federal Matching Funds:		\$0
			TOTAL REQUESTED FROM NEH:		\$0
	b. Cost Sharing		Applicant's Contributions:		\$0
			Third-Party Contributions:		\$0
			Project Income:		\$0
			Other Federal Agencies:		\$0
			TOTAL COST SHARING:		\$0
12. Total Project Funding					\$0

Total Project Costs must be equal to Total Project Funding ----> (\$29,995 ?)
 Third-Party Contributions must be
 greater than or equal to Requested Federal Matching Funds ----> (#REF!:emptyRang ?)

5b. Budget narrative

1. Salaries and Wages.

Student research assistants. Student workers will be used to do first pass translations of the Spanish portions of texts into modern Spanish and English, general editing, and proofreading of both English and Spanish. Others will also be employed to do first pass encoding in TEI. Students will be undergraduates at Haverford College or its sister schools Bryn Mawr College or Swarthmore College; this pay rate is fixed by the Haverford College Business Office at \$9/hr during the academic year. 220 hrs in Year 1 and 110 hrs in Year 2. **\$2,970 total.**

3. Consultant Fees

Linguistic consulting (Broadwell). Checking 88 pages of first draft transcriptions of the *Arte* (done by student researchers); completing first pass linguistic analysis of the Zapotec portions of 132 pages of the *Arte*; checking 88 page of first pass linguistic analysis done by Lillehaugen. Providing transcriptions and first pass analyses of folia 10-64 of the *Doctrina*. Providing background content on the Colonial Valley Zapotec language for non-linguistic users of the website (2-5 pages). Participating in two in-person team meetings and additional monthly online meetings. Consulting fee based on approximately one month of work during two summers each, for a total of two months of work. **\$6,000 total.**

Ethnohistorical consulting (Oudijk). Checking 88 pages of first draft transcriptions of the *Arte* (done by student researchers); checking 88 page of first pass linguistic analysis done by Lillehaugen and Broadwell. Providing background content on the historical and philological context of Colonial Valley Zapotec language documents for non-expert users of the website (5-10 pages). Participating in two in-person team meetings and additional monthly online meetings. Based on one month of work during two summers each, for a total of two months of work. **\$6,000 total.**

TEI consulting (Valdivia). Creating TEI encoding for 40 pages of the original text. Checking TEI encoding on 240 pages of original text. Creating technical documentation on encoding practices and decisions. Providing background content on TEI and the encoding process for non-expert users of the website (5-10 pages). Based on two months of work during two summers each, for a total of four months of work. **\$6,000 total.**

cont. on next page

4. Travel

Washington D.C.: 1 person (PI Lillehaugen). **TOTAL \$340.**

Attend planning meeting at the NEH offices in Washington, D.C. (1 day). One additional day of travel budgeted for.

Transportation. \$102 roundtrip between PHL and Washington, D.C.

Lodging. 1 night, \$167 / night *

Meals and incidental expenses. 1 day, \$71/day*

*based on US Department of State General Service Administration per diem rates for fiscal year 2014

Oaxaca City: 5 people. **TOTAL \$8,685.**

Attend working/planning meeting and outreach trip in Oaxaca City, Mexico from Jan 12-16, 2015 (5 days). One additional travel day budgeted for.

Airfare (\$3500)

-airfare for 4 from US (various cities) @\$800 each

-airfare for 1 from Mexico, @300 each

Lodging (\$3060)

-5 people / 6 nights, \$102/nt*

Meals and incidental expenses (\$1950)

-5 people / 6 days / \$65 day*

Meeting Space (\$175)

-\$35 / day for 5 days

* based on US Department of State foreign per diem rates 9/1/2013

6. Biographies

Brook Danielle Lillehaugen is Assistant Professor of linguistics at Haverford College, with joint appointments at Bryn Mawr and Swarthmore Colleges. She received her Ph.D. in linguistics from the University of California, Los Angeles in 2006 and has been working on Colonial and Modern Valley Zapotec since 1999. She is co-PI on the largest extant Colonial Valley Zapotec database. She is co-author of *Cali Chiu? A Course on Tlacolula Valley Zapotec* and publishes on the grammar of Zapotec in both in its modern and colonial forms. In collaboration with other linguists, the Living Tongues Institute and native Zapotec speakers, she is developing an online "talking dictionary" on Valley Zapotec (talkingdictionary.swarthmore.edu/zapotec).

George Aaron Broadwell is Professor in the Department of Anthropology and the Program in Linguistics and Cognitive Science at the University at Albany, State University of New York. His research focuses on the documentation of Native American languages, particularly in the southeastern United States and Oaxaca, Mexico. He has worked with Zapotec languages since 1989, with research on three modern varieties (Santa Ana del Valle, San Dionisio Ocotepéc, and Macuiltianguis) as well as Colonial Valley Zapotec materials. He is the author of numerous publications on Zapotec and also author/editor of *A Choctaw Reference Grammar*; *The origin of the sun and moon: A Copala Triqui legend*; and *Nana nagan' rihaan nij sii chihaan': Words of counsel for the Triqui people*.

Michel R. Oudijk is an historian/philologist whose main interest is Zapotec historiography and its relationship to present day communities. Since 1992 he has worked in three of the four main Zapotec regions and is renowned for his analysis of pictographic and alphabetic documents alike, relating historical information to present day landscapes and oral tradition. His work on the indigenous participation in the "Spanish" conquest of Mexico has opened up a whole new field of study called the *New Conquest History*. In 2000 he received his Ph.D. at Leiden University, after which he worked for a year in the National Archives of the Indies in Seville, Spain. From there he became full professor at the University of Copenhagen, Denmark, and since 2004 he is a researcher at the Institute of Philological Studies at the Universidad Nacional Autónoma de México (the National Autonomous University of Mexico). He has published seven books and over 40 articles and chapters in journals and volumes all over the world, and is a well-known speaker at international academic meetings and in local indigenous communities events.

Laurie Allen is Coordinator for Digital Scholarship & Services at Haverford College. She earned a Bachelor's Degree from Bard College and an MS in Library and Information Science from Simmons College. She worked for 6 years at the Social Science Data and Research & Instructional Services Department at the University of Pennsylvania before joining the Haverford College Libraries in 2008. Since 2011, Allen has led a team of students and staff within the Haverford Library as Coordinator for Digital Scholarship & Research Services. The Digital Scholarship Group supports faculty and student research projects with technical and infrastructural support. Allen and her team support several Drupal sites, databases, and mapping projects including The Global Terrorism Research Project and others, visible

at <http://library.haverford.edu/services/digital-scholarship/projects/>. Allen has experience developing and training in Drupal, XML and TEI, and XSLT.

Enrique Valdivia is a graduate student in the Masters of Information Science program at the University of Michigan School of Information. He also reviews and edits TEI for the Early English Books Online Text Creation Partnership (EEBO-TCP). In 2012, he earned his Bachelor's degree in anthropology from the University of Nevada, with honors research focusing on language and religious contact between colonial powers and indigenous cultures. His research interests include issues of digital dissemination of archival manuscripts for both preservation and access, as well as ethnohistorical approaches to these documents. His training beyond traditional coursework includes a 20 hour course at the Mexican National Museum of Anthropology, *Introduccion a los Codices Mesoamericanos* (Introduction to the Mesoamerican Codices). He has done extensive work in Mexico, including internships at the Archivo Nacional de la Nacion (AGN, the General Archive of the Nation) and the archival branch of the Fundacion Ingenieros Civiles Asociados (ICA), Mexico's largest civil engineering firm, both in Mexico City. He developed digital preservation policies, including drafting digital preservation plans and guidelines, for both the AGN and ICA, which will be the foundation for these institutions' digital preservation programs. He plans to pursue a Ph.D. in anthropology, where his research will focus on the analysis of Mesoamerican, colonial, and pre-colonial era manuscripts.

7. Data management plan

During this planning phase, two distinct but overlapping kinds of data will be produced. The first kind of data is planning data for the larger *Ticha* project, and will include process documentation, wireframes, user stories, research questions, and technical specifications. The second set includes data specifically tied to the texts themselves and will be of four types: digital images of the primary texts; transcriptions and translations of the texts into Modern Spanish and English; linguistic analysis of the Zapotec words within the texts; and TEI markup of the transcribed, translated, and analyzed texts. As part of the production of those data, we will also produce associated data, including workflow documents, XML schema and XSLT transformation documents, and Zapotec lexicon.

Because the *Ticha* site will be designed to make in depth analysis of these texts available to a variety of disciplinary audiences and members of the general public, we will be attentive to common data standards throughout. We will rely on the open standard of XML for encoding, sharing, and storing our data.

The translations will be created in Microsoft Word; the linguistic analysis will be created in SIL Fieldworks Language Explorer Program (FLEX); and the TEI encoding and associated schema and xslt files will be created in Oxygen XML Editor.

The primary working repository for versions of all *Ticha* data will be a shared Dropbox folder and monthly archiving of all Dropbox files to the Haverford Storage Network. Data on the Haverford storage server is redundantly spread across multiple san servers for easy backup. Tape archives of the data are also archived, and the college maintains both local and off-site copies of these tapes as a normal function of disaster preparedness.

As we anticipate completing the translation of Cordova's *Arte en Lengua Zapoteca* during this phase, we will also produce the finished versions of this text as encoded and analyzed. While we do not anticipate having the interface ready by the end of the project, we will ensure that the translations, encoded texts, and analyses are available in multiple formats on the *Ticha* website (ticha.haverford.edu). They will be stored for long term preservations as PDF/A documents (for unmarked-up transcription, and translation) and XML files (for TEI encoded versions and linguistic analysis) in Triceratops (<http://triceratops.brynmawr.edu/dspace/>), the Haverford College instance of Dspace, an open source repository. We will also store, and make available copies of the TEI schema used in the project, and the XSLT files used to transform the linguistic analysis from FLEX XML to TEI. The header information for each of the XML files will include rights information describing the Creative Commons license under which the data is being distributed, and requiring attribution and share-alike.

The linguistic analysis and TEI are data outputs of the project, but we expect that the analyses and encoding will also yield scholarly publications beyond the data themselves. These publications will be made available in open access formats to the extent possible under publishing and licensing agreements.



September 9, 2013

To whom it may concern,

I am committed to work on NEH-DH project "*Ticha*: a digital text explorer for Colonial Zapotec." I will check 88 pages of first draft transcriptions of the *Arte* (done by student researchers); complete first pass linguistic analysis of the Zapotec portions of 132 pages of the *Arte*; and check 88 page of first pass linguistic analysis done by Lillehaugen. I will provide transcriptions and first pass analyses of folia 10-64 of the *Doctrina*. I will provide background content on the Colonial Valley Zapotec language for non-linguistic users of the website (2-5 pages) and participate in two in-person team meetings and additional monthly online meetings. My consulting fee based on approximately one month of work during two summer each, for a total of two months of work is \$3000 per year, for a total of \$6,000.

Sincerely,

A handwritten signature in blue ink that reads "George Aaron Broadwell".

George Aaron Broadwell

Professor

Dept. of Anthropology



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

INSTITUTO DE INVESTIGACIONES FILOLÓGICAS

SEMINARIO DE LENGUA INDÍGENA

September 6, 2013

To whom it may concern,

I am committed to work on the NEH-DH project *Ticha: a digital text explorer for Colonial Zapotec*. Within this project I will 1) check the transcriptions; 2) transform the sixteenth century Spanish texts into readable modern Spanish texts; 3) based on my knowledge of colonial administrative and religious Zapotec texts, verify the linguistic analysis previously done by Lillehaugen and Broadwell; 4) provide information on the historical context in which the *Arte* was formed; and 5) write a brief description of the historical and philological contexts of Zapotec alphabetical documentation. Furthermore will I participate in regular meetings, both on-line and in person, with Brook Lillehaugen, Aaron Broadwell, Laurie Allen, and Enrique Valdivia. As such, I will be committing 10% of my time over the 18 month period (approximately 320 hours).

Sincerely,

Dr. Michel R. Oudijk
Researcher
Seminar of Indigenous Languages
Institute of Philological Investigations
National Autonomous University of Mexico
mroudyk@hotmail.com

Haverford

C O L L E G E

The Library

September 9, 2013

To whom it may concern,

I look forward to working on NEH-DH project "Ticha: a digital text explorer for Colonial Zapotec." My work as Coordinator for Digital Scholarship in the Library is to support projects like this one, and I am committed to spending significant time developing this project through to its next phase. I will take the lead on the development and production of TEI associated with the two sample texts, in collaboration with the rest of the team, and will develop the XSLT to transform FLEXML to TEI. I will also participate in regular meetings with Lillehaugen, Broadwell, Oudijk, and Valdivia both on-line and in person.

Sincerely,



Laurie Allen
Coordinator for Digital Scholarship & Research Services
Haverford College Libraries
610-642-2817

lallen@haverford.edu

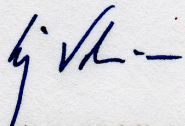


September 7, 2013

To whom it may concern,

I am committed to work on the NEH-DH project "Ticha: a digital text explorer for Colonial Valley Zapotec." I will contribute TEI encoding for forty pages of original text, review and edit TEI encoding for 240 pages of original text, create technical documentation on encoding practices and decisions, and provide five to ten pages of background content on TEI and the encoding processes for non-expert users of the website. I will participate in regular meetings with Lillehaugen, Broadwell, Oudijk, and Allen both on-line and in person. I am committing 22% of my time over the 18 month period (approximately 320 hours).

Sincerely,



Enrique Valdivia

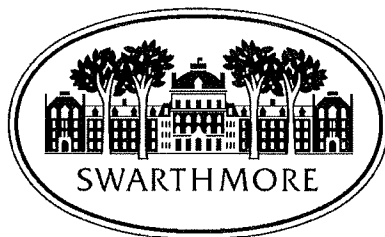
University of Michigan, School of Information Masters of Information Science Candidate, 2014

valdiven@umich.edu

348 Larkspur St.

Ann Arbor, MI

48105



Linguistics Department

Telephone: 610-328-8421/8422

Fax: 610-690-6846

09 September 2013

Email: linguistics@swarthmore.edu

Dear Colleagues,

I am writing to enthusiastically support "***Ticha: a digital text explorer for Colonial Zapotec***" for NEH funding.

This is an exemplary project in several respects.

First, it directly serves the efforts by an indigenous Mexican community to reclaim and recover aspects of their history that have been long obscured and inaccessible.

Second, it supports efforts by the Zapotecs to revitalize their endangered language. A key ingredient in language revitalization is linguistic pride, and the existence of a centuries long written record in the language should be a matter of great pride to the community. It serves to help reestablish the rightful place of Zapotec in Mexican culture, as being prior to Spanish and independently attested.

Thirdly, this project will provide new insights for many disciplines, including Linguistics, Anthropology, History, Literary Studies, Latino Studies, Religious Studies, and Translation Theory.

Finally, this project will provide a valuable training environment for undergraduate students, who will encounter disciplines such as Paleography and Zapotec Linguistics for the first time in the context of this work. Prof. Lillehaugen has had great success mentoring undergraduates and also bringing them to Mexico to do Linguistic fieldwork with the Zapotec community (four students from Haverford and Swarthmore went there with her this past summer to do fieldwork).

This project is thus a logical extension of the work that is already successfully being carried out by a collaborative effort of Zapotec language activists and US-based scholars and students.

It will have multiple positive effects, in terms of scholarship, scientific discovery, and social impact. It will help an ancient language community in crossing the digital divide.

Sincerely,

K. David Harrison, PhD
Associate Professor and Chair



September 9, 2013

To whom it may concern:

I am pleased to write in support of the NEH Digital Humanities Start Up Grant: *Ticha: a digital text explorer for Colonial Zapotec* from Lillehaugen and her collaborators George Aaron Broadwell, Michel Oudijk, Laurie Allen, and Enrique Valdivia. Their proposal is extremely worthwhile and of great use to scholars in distinct but related fields, and to contemporary speakers of the Zapotecan languages. The known trajectory of the participants in the field of Zapotec studies is a guarantee that this project will be carried out in a skillful and intelligent way.

Traditionally, the post-contact history of Spanish America has been studied and explored on the basis of the copious and abundant Spanish language documents. These Spanish colonial documents are generally accessible, but were invariably prepared by or under the supervision of Spanish administrators and clerics. This inevitably biased the view we have developed of the post-contact history of indigenous America. The decades since 1980 have seen an increasing interest in the literary production of the indigenous peoples of Mesoamerica themselves. Oaxaca holds a special place in this regard because of the enormous production of texts in several indigenous languages between the mid-sixteenth century and early-nineteenth century. Archival research has unearthed hundreds of documents written in Mixtec, Zapotec, Chochón, and Nahuatl. These texts do not only have value as sources for linguistic studies; together with the better known printed sources (of which many are nonetheless very rare), these texts give a unique insight into the lives of the ancestors of the current speakers of these languages. They reveal the often complex ways in which indigenous communities coped with the colonial reality and how they actively strove to adapt to that setting without losing group-identities and the control over cultural reproduction. These texts often provide insights into the indigenous point of view not available from the Spanish documents alone.

However, the inaccessibility of these texts is currently one of the major obstacles for their use by both scholars and speakers. Printed editions are often very rare and most of the archival documents are scattered over remote archives and deposits. Also, the changes experienced by the indigenous languages over the last few hundred years and the distinct orthographic systems used over time make the reading and understanding of these texts notably difficult. This is especially the case for scholars that do not have a linguistic formation or for untrained current speakers. Reading these texts has become the domain of a small group of specialized scholars.

Centro Académico y Cultural San Pablo - Callejón Antiguo de San Pablo - Independencia 904 – 68000,
Oaxaca, México. Tel. +52-951-5018803.



The proposed project will do a great service to the scholars in the fields of history, anthropology, and linguistics and to current speakers by making available several texts in an accessible way. Unlocking the primary sources is of course at the basis of any advancement in the study of the languages and cultures of the indigenous groups of the Oaxaca area. The team that has come together to propose this project is amply qualified for this task. All members have notable expertise in their field. I am grateful that this team wishes to undertake the analysis of the Zapotec materials, which I am sure will be done with a serious knowledge of the subject. For these reasons, I highly recommend the team and their project.

Sincerely,

Bas van Doesburg, PhD
Coordinación de Humanidades,
Universidad Nacional Autónoma de México /
Centro Académico y Cultural San Pablo
Fundación Alfredo Harp Helú Oaxaca
www.basvandoesburg.wordpress.com

Appendix A: Sample Page, *Arte* 102r

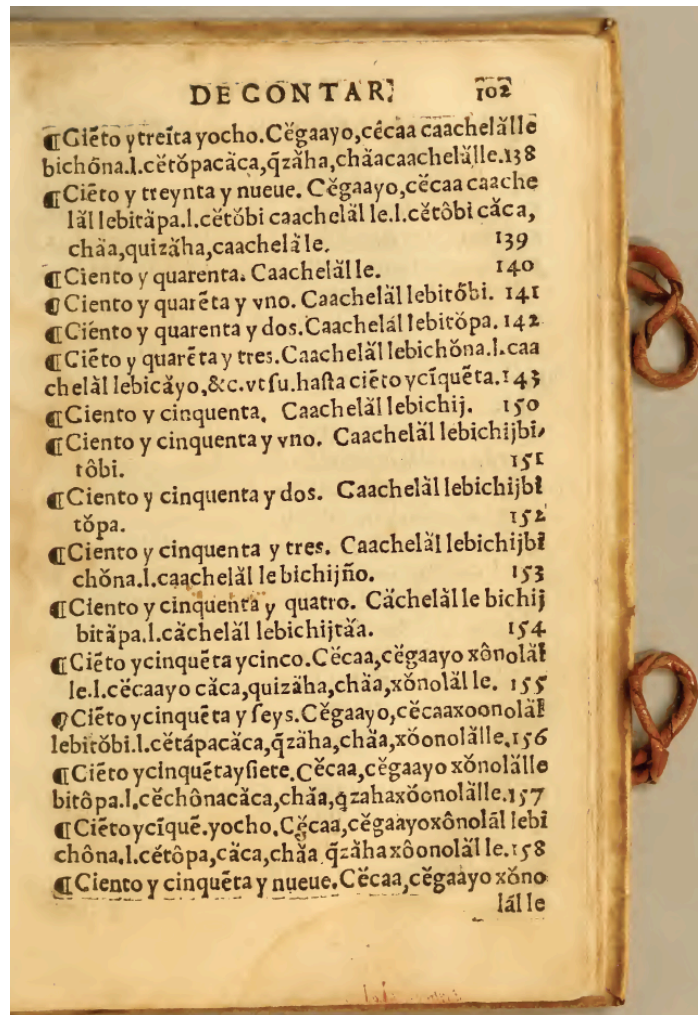


Figure 1. Original (Córdoba 1578a: 102r)

Transcription of lines 1-2

¶ Ciēto y treita yocho. Cēgaayo, cécaa caachelälle
bichõna.l.cētõpacäca, qzãha, chãacaachelälle. 138

Translation and linguistic analysis of lines 1-2

One hundred and thirty-eight. [1] *Cēgaayo, cécaa caachelälle bichõna* or [2] *cētõpa cäca, qzãha, chãa caachelälle*. 138

General notes

Brackets. Square brackets are used to indicate omitted letters, e.g. those omitted through abbreviation.

Comma. The comma usually represents a choice between the words on either side of the comma. So X,Y A B stands for X A B or Y A B.

Spacing. The spacing in the original document is based on many factors, linguistic and otherwise. In my analysis, I use spacing to indicate (linguistic) word boundaries.

[1] *Cēgaayo, cécaa caachelälle bichõna*

- [1a] *Cě-gaayo* *caache-lälle* *bi-chōna*
 DEF¹-five seven-twenty and-three
 'another five (until) 7x20 and 3' = 138

Grammatical note: The following schema represents the expression of this value, beginning in the center with *caache-lälle* 'seven twenties', which is 140. The next calculation incorporates the first word in the phrase, *Cě-gaayo* 'another five'. When applied to 140, this means 'another five will be 140' or 'five shy of 140', thus resulting in 135. The final calculation is complete with the last word, *bi-chōna* 'and three'; adding three to 135 results in 138.

In sum: [7x20 is 140; 5 shy of 140 is 135; plus three more is 138; =**138**]

<i>Cě-gaayo</i> DEF-five	<i>caache-lälle</i> seven-twenty 140	<i>bi-chōna</i> and-three
	135	
138		

- [1b] *cé-caa* *caache-lälle* *bi-chōna*
 DEF-five(S)² seven-twenty and-three
 'another five (until) 7x20 and 3 [=138]'

Grammatical note: [7x20 is 140; 5 shy of 140 is 135; plus three more is 138; =**138**]

[2] *cětōpa cāca, qzāha, chāa caachelälle*

- [2a] *cě-tōpa* *c-āca* *caache-lälle*
 DEF-two IRR-be seven-twenty
 'two more will be 7x20 [=138]'

Grammatical note: [7x20 is 140; 2 shy of 138; =**138**]

- [2b] *cě-tōpa* *q[ui]-zāha* *caache-lälle*
 DEF-two IRR-walk seven-twenty
 'two more will walk to (arrive at) 7x20 [=138]'

Grammatical note: [7x20 is 140; 2 shy of 138; =**138**]

- [2c] *cě-tōpa* *chāa* *caache-lälle*
 DEF-two IRR.fill_up seven-twenty
 'two more will complete 7x20 [=138]'

Grammatical note: [7x20 is 140; 2 shy of 138; =**138**]

¹ The following abbreviations are used in the glossing of the Zapotec:

DEF: definite aspect. (When attached to numbers has the meaning 'another #' or '# more'; Munro and Sonnenschein 2007.)

IRR: irrealis aspect.

S: short form.

² *Caa* 'five' seems to be a shortened form of *gaayo* 'five'.

Appendix B: Sample Page, *Doctrina 65r*

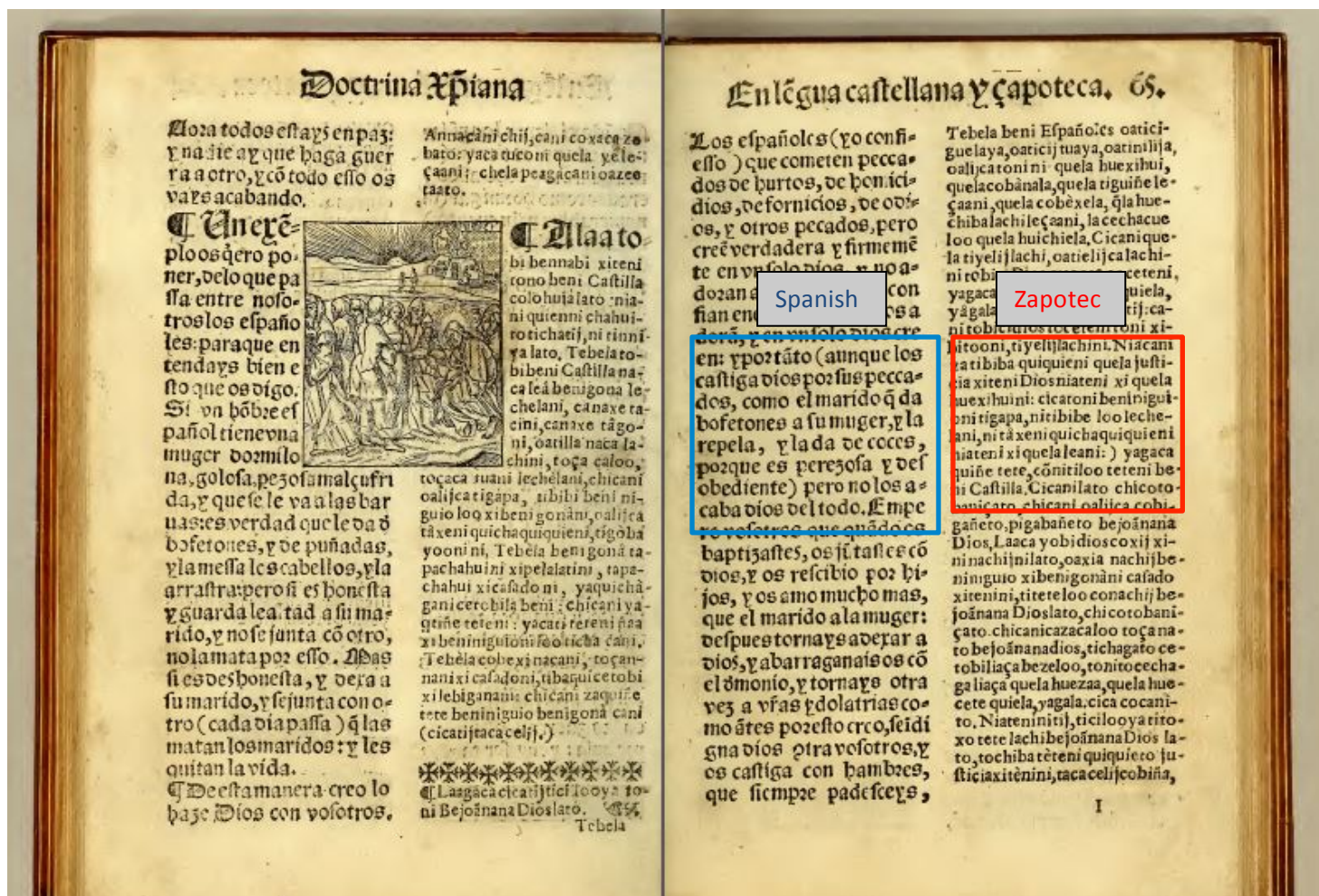


Figure 2. Original (Feria 1567: 65r)

Transcription of f65, third sentence of Zapotec

Niacani zatitiba quiquieni quela justicia xiteni Dios niateni xi quela huexihuini: cicatoni beniniguioni tigapa nitibibe loo lechelani, ni tãxeniquichaquiquieni niateni xiquela leani:) yagaca quife tete, cõnitiloo teteni beni Castilla.

Analysis and translation in FLEx

The screenshot shows the FieldWorks Language Explorer (FLE) interface. The title bar reads "cvz23 - FieldWorks Language Explorer". The menu bar includes "File", "Send/Receive", "Edit", "View", "Data", "Insert", "Format", "Tools", "Parser", "Window", and "Help". The toolbar contains various icons for file operations and editing. The main window is titled "Text" and shows a document titled "Doctrina 064v-65r (The abusive husband)". The text is displayed in a grid format, with morphemes and their glosses. A pop-up window is visible over the text, showing a list of morphemes: "cicatoni", "cica", "t-", "oni", "like", "HAB", and "do". The status bar at the bottom indicates "08/Sep/2013", "Queue: (-/-) No Parser Loaded", and "Sorted by Title | Filtered 206/340".

Text commentary. The Spanish and Zapotec texts are discrepant with each other in the following ways:

- 1) The Spanish sentence ends with the verb *abandonar* 'abandon' with 'them' as its object, while the corresponding Zapotec sentence ends with a different verb 'forgive' with 'the Spanish person' as object.
- 2) The comparison in Spanish is between God and the abusive husband, while the Zapotec comparison seems to be between those who come under God's judgment and the abusive husband.

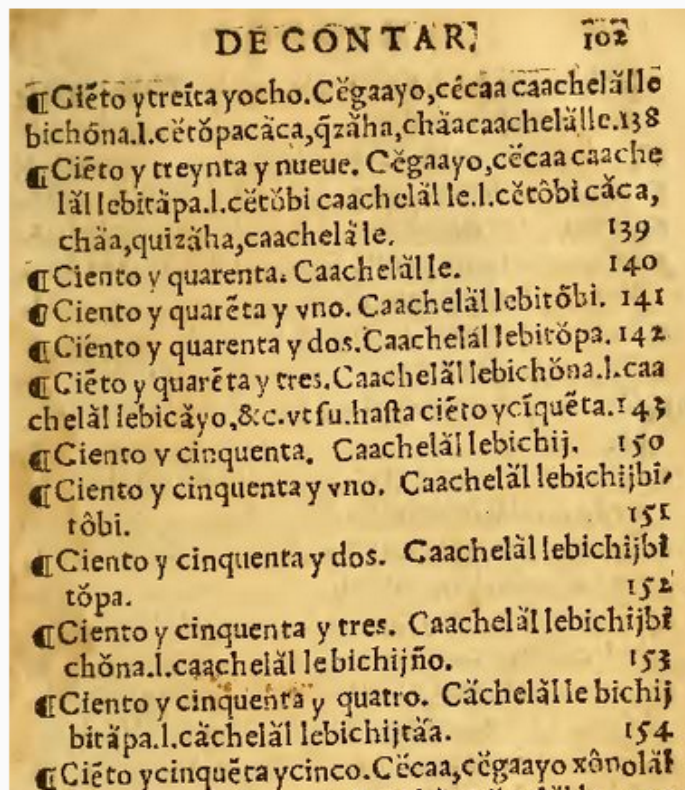
Appendix C: Example TEI in process, from the *Arte*

```
<head>
  <hi rend="case(allcaps)">Indicativo modo</hi>
</head>

<div5 n="Tempore Prefenti" xml:id="Arte_1.14.1.1.1">
  <head>
    Tempore Prefenti.
  </head><lb/>
  <p>
    <foreign xml:lang="cvz">Svm</foreign>, yo soy. <choice>
      <abbr>f.</abbr>
      <ex xml:lang="la">scillicet</ex>
    </choice>
    <foreign xml:lang="cvz">Tăcaya</foreign>
    <choice>
      <abbr>.l.</abbr>
      <ex xml:lang="la">vel</ex>
    </choice>
    <foreign xml:lang="cvz">năcaya</foreign> , efte,
    <foreign xml:lang="cvz">tăca</foreign>, las<lb/>
    menos vezes firue de <foreign xml:lang="cvz">fum</foreign> eft fuy, porque finoes<lb/>
    alguna vez en el prefente, todo lo demas tocan<lb break="no"/>
    te a efte fignificado fe lo dexa a, <foreign xml:lang="cvz">năca</foreign>
    <pc>,</pc> el qual tomă<lb break="no"/>do en fu ayuda
    un pronõbre paffa fu carrera por to-<lb break="no"/>dos los
    tiempos de fum eft <damage><choice>
      <orig>fu</orig>
      <corr>fuy</corr>
    </choice></damage>. Como parecera en<lb/> ehta conjugaciõ, pues
    dize afsi. <foreign xml:lang="cvz">Tăca</foreign><choice>
      <abbr>.l.</abbr>
      <ex xml:lang="la">vel</ex>
    </choice>
    <foreign xml:lang="cvz">năca</foreign>, yo foy.<lb/>
```

Appendix D: Example mock up of page from *Ticha* for navigating *Arte 102r*

This mock-up page demonstrates how a site like the one proposed could allow users to access and navigate through multiple layers of information and multiple versions of a text.



Transcription
Modern Spanish
English

¶Ciēto y treita yocho. Cēgaayo, cécaa caachelälle bichõna. l. cētõpacāca, qzāha, chāacaachelälle. 138

¶Ciēto y treynta y nueue. Cēgaayo, cécaa caache läl lebitāpa. l. cētõbi caacheläl le. l. cētõbi cāca, chāa, quizāha, caacheläl le. 139

¶Ciento y quarenta. Caacheläl le. 140

¶Ciento y quarēta y vno. Caacheläl lebitõbi. 141

¶Ciento y quarenta y dos. Caacheläl lebitõpa. 142

¶Ciēto y quarēta y tres. Caacheläl le bichõna. l. caa cheläl le bicāyo, &c. vt fu. hasta ciēto y ciquēta. 143

¶Ciento y cinquenta. Caacheläl le bichij. 150

¶Ciento y cinquenta y vno. Caacheläl le bichijbi tõbi. 151

¶Ciento y cinquenta y dos. Caacheläl le bichijbi tõpa. 152

¶Ciento y cinquenta y tres. Caacheläl le bichijbi chõna. l. caacheläl le bichijñõ. 153

¶Ciento y cinquenta y quatro. Cācheläl le bichij bitāpa. l. cācheläl le bichijtāa. 154

¶Ciēto ycinquēta y cinco. Cēcaa, cēgaayo xānoläl

Transcription
Modern Spanish
English

¶Ciēto y treita yocho. Cēgaayo, cécaa caachelälle bichõna.

l. cētõpacāca, qzāha, chāaca

¶Ciēto y treynta y nueue. [1a]Cē-gaayo caache-lälle bi-chõna x

Cē-gaayo	caache-lälle	bi-chõna
DEF-five	seven-twenty	and-three

'another five (until) 7×20 and 3' = 138

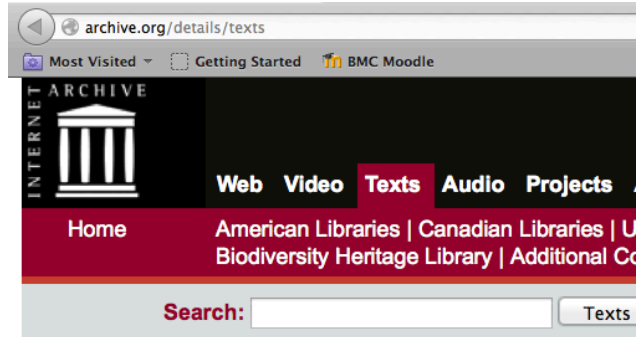
¶Ciento y cinquenta. [1b]cé-caa caache-lälle bi-chõna

cé-caa	caache-lälle	bi-chõna
DEF-five(S)	seven-twenty	and-three

'another five (until) 7×20 and 3 [=138]'

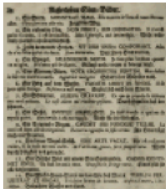
Grammatical note: The following schema represents the expression of this value, beginning in the center with *caache-lälle* 'seven twenties', which is 140. The next calculation incorporates the first word in the phrase, *Cē-gaayo* 'another five'. When applied to 140,

Appendix E: Archive.org screenshot showing usage rights



Ebook and Texts Archive

Spotlight Item



[DEVISES ET EMBLEMES
ANCIENNES & MODERNES,
TIREES DE PLUS CELEBRES
AUTEURS](#)

McGeary & Nash. Emblem
books at the University of
Illinois

About the Archive

[Background](#)

[Frequently Asked Questions](#)

Rights

Free to read, download, print,
and enjoy.

Some have restrictions on bulk
re-use and commercial use,
please see the collection or the
sponsor of a book.

By providing near-unrestricted
access to these texts, we hope
to encourage widespread use
of texts in new contexts by
people who might not have
used them before.

Contributors

Welcome to Ebook an

Download free books and te
fiction, popular books, child

Try the archive's site [Open](#)
links to other online book c

This collection is open to th
using [Creative Commons li](#)

Browse texts by keyword

[Open Source Books has I](#)
though.)

Recent [Internet Archive blo](#)

- [How Montana State
Archive](#)
- [1790-1930 U.S. Cer](#)
- [Go Books in Brows](#)
- [New BookReader!](#)
- [Books in Browsers I](#)
- [Ted Nelson Launch](#)
- [Banned Books Wee](#)
- [The Awesomeness](#)
- [MegaReader iPhon](#)
[a Personalized Rea](#)

Please feel free to [contribu](#)
license as part of the uploa
thanks!)

Sub-Collections

Figure 3. Usage rights from archive.org (bottom left)

Appendix F: Agenda for *Ticha* team meeting in Oaxaca

Draft agenda
Ticha planning, design, and outreach meeting
January 12-16, 2015
Oaxaca City, Oax; Mexico

Attendees: Brook Lillehaugen, George Aaron Broadwell, Michel Oudijk, Laurie Allen, and Enrique Valdivia.

Potential invitees for workshop/discussion sections: Bas van Doesburg (UNAM), Michael Swanton (UNAM), Francisco Arellanes (UNAM), Rosa Maria Rojas (INALI), Aurea Lopez Cruz, Rosemary Beam de Azcona (ENAH), Elizabeth Merrill (SIL), Cheryl and Andrew Black (SIL), and others to be determined.

Day one

Overview of current issues, logistics, and planning for public discussions and demos

Day two

Workshop/discussion session with potential users—Mexican academics

Day three

Workshop/discussion session with potential users – Zapotec community members

Day four

Evaluation of workshop feedback, identification of system errors or shortcomings, additional development goals

Day five

Phase 3 planning (task identification, prioritization, and assignment; strategies for team communication; future meetings and events; and publication and presentation opportunities)

Appendix G: Works Cited

- Broadwell, George Aaron and Brook Danielle Lillehaugen. 2013. Building an electronic database for Colonial Valley Zapotec. Presented at *International Conference on Mesoamerican Linguistics*, Fullerton, CA.
- Córdova, Fr. Juan de. 1987 [1578b]. Vocabulario en lengua çapoteca. México: Ediciones Toledo (INAH).
- Córdova, Fr. Juan de. 1886 [1578a]. Arte del idioma zapoteco. Morelia: Imprenta del Gobierno.
- Feria, Pedro de. 1567. Doctrina christiana en lengua castellana y çapoteca. Mexico City: En casa de Pedro Ocharte. <http://www.archive.org/details/doctrinachristia00feri>. (5/6/2013.)
- Munro, Pamela, and Aaron Huey Sonnenschein. 2007. "Four Zapotec Number Systems." In *Bernard Comrie Festschrift*: ms.
- Oudijk, Michel R. 2008. "El texto más antiguo en zapoteca." *Tlalocan*. 15:227-240. México, D.F.: UNAM.