

Multilingual Texts in the Digital Age

Hannah Alpert-Abrams
@hralperta

Slides available at:
halperta.com/conferences.html

Aquí comienza vn vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignas† domine seruum

¶ dorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

* nis redemptionis nostre:;

Aquí comienza vn vocabula-
rio en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignas† domine seruum

* nis redemptionis nostre:;

¶ Indorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

Tassado por el Audiencia Real en dos pesos y vn nombrón

Benson Latin American Collection

John Carter Brown Library (copy 1)

Gutenberg
Bible

Bay Psalm
Bible

Molina's
Vocabulario

Conquest of
Tenochtitlan

1450s

1521

1555

1640

Aquí comienza vn vocabula-
rio en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignatii domine seruum

Idorū nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

* nis redemptionis nostre:.

Aquí comienza vn vocabula-
rio en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignatii domine seruum

Indorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

Tassado por el Audiencia Real en dos pesos y vn nombrón

John Carter Brown Library (copy 1)

Benson Latin American Collection

Aquí comienza vn vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardiā del couéto del sant Antonio d'
Tetzcuco dla ordē delos frayles Menores.

* tuū Franciscum sig *



Signasti domine seruum

* nis redemptionis nostre::

dorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum specias, stigmata carne geris.

Aquí comienza vn vocabula-
rio en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardiā del couéto del sant Antonio d'
Tetzcuco dla ordē delos frayles Menores.

* tuū Franciscum sig *



Signasti domine seruum

* nis redemptionis nostre::

Indorum nimia te fecit prole parentem,
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum specias, stigmata carne geris.

Tassado por el Audiēcia Real en dos pesos y vn nombr

Benson Latin American Collection

John Carter Brown Library (copy 1)

06585

Aquí comienza un vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.



¶ dorum nimia te fecit prole parentem.
qui genuit moriens, quos pater almifoues,
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

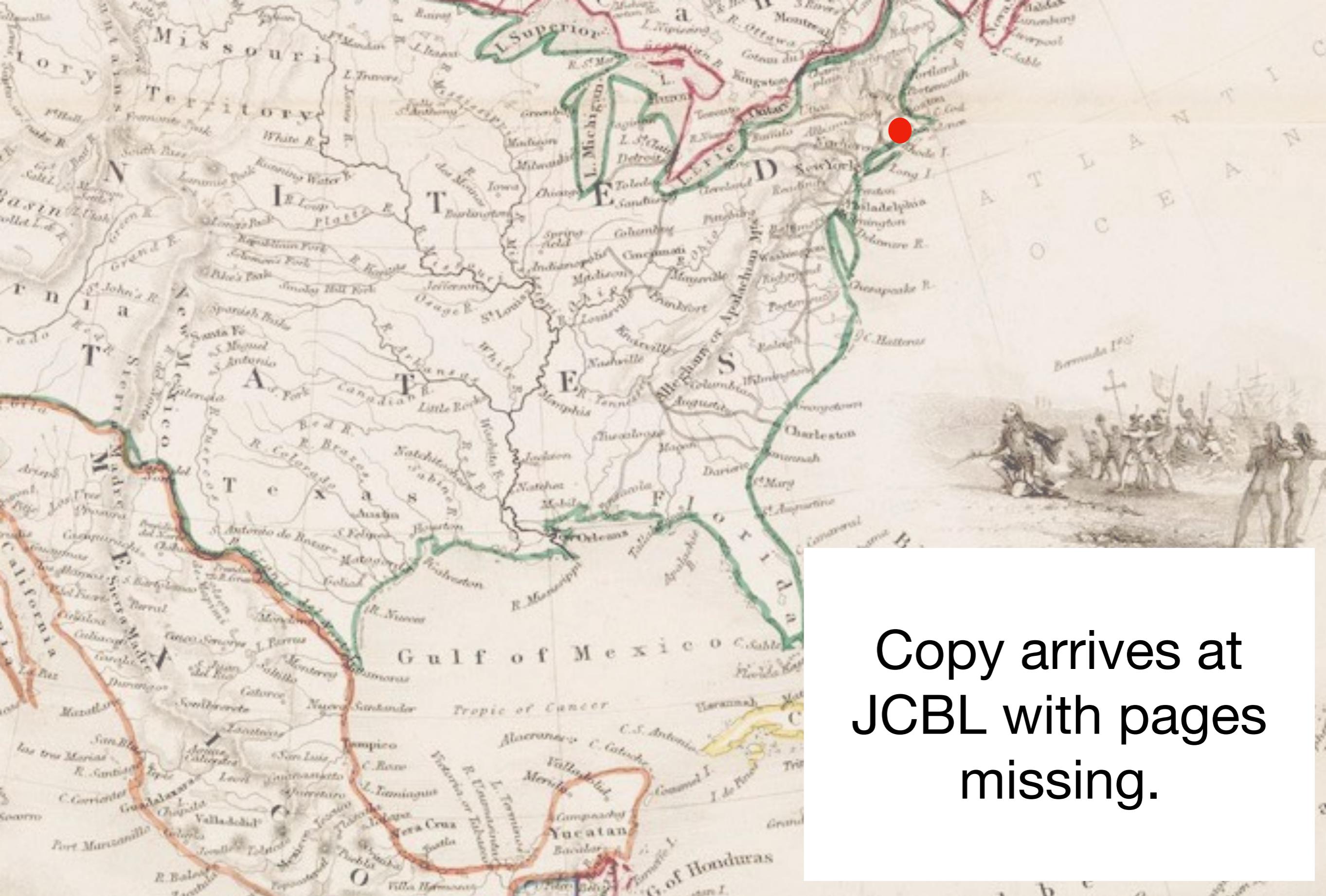
Benson Latin American Collection

This copy of Molina's *Vocabulario* 1535 is exceedingly rare, so much so that Mr. Ludwig in his work on American aboriginal languages doubts its existence.

The present copy was purchased of Astor No. Berlin, upon whose catalogue it appeared as perfect. Upon its receipt, however, I discovered that a considerable portion at the end was wanting. The title page and some preliminary matter was also deficient. Upon enquiring of Mr. Lenox, whom he possessed a copy, which he kindly loaned me that we might have a copy made of the portions which were deficient. Photographic copies were therefore made accordingly.

But unfortunately the title of Mr. Lenox's copy was mutilated, the central

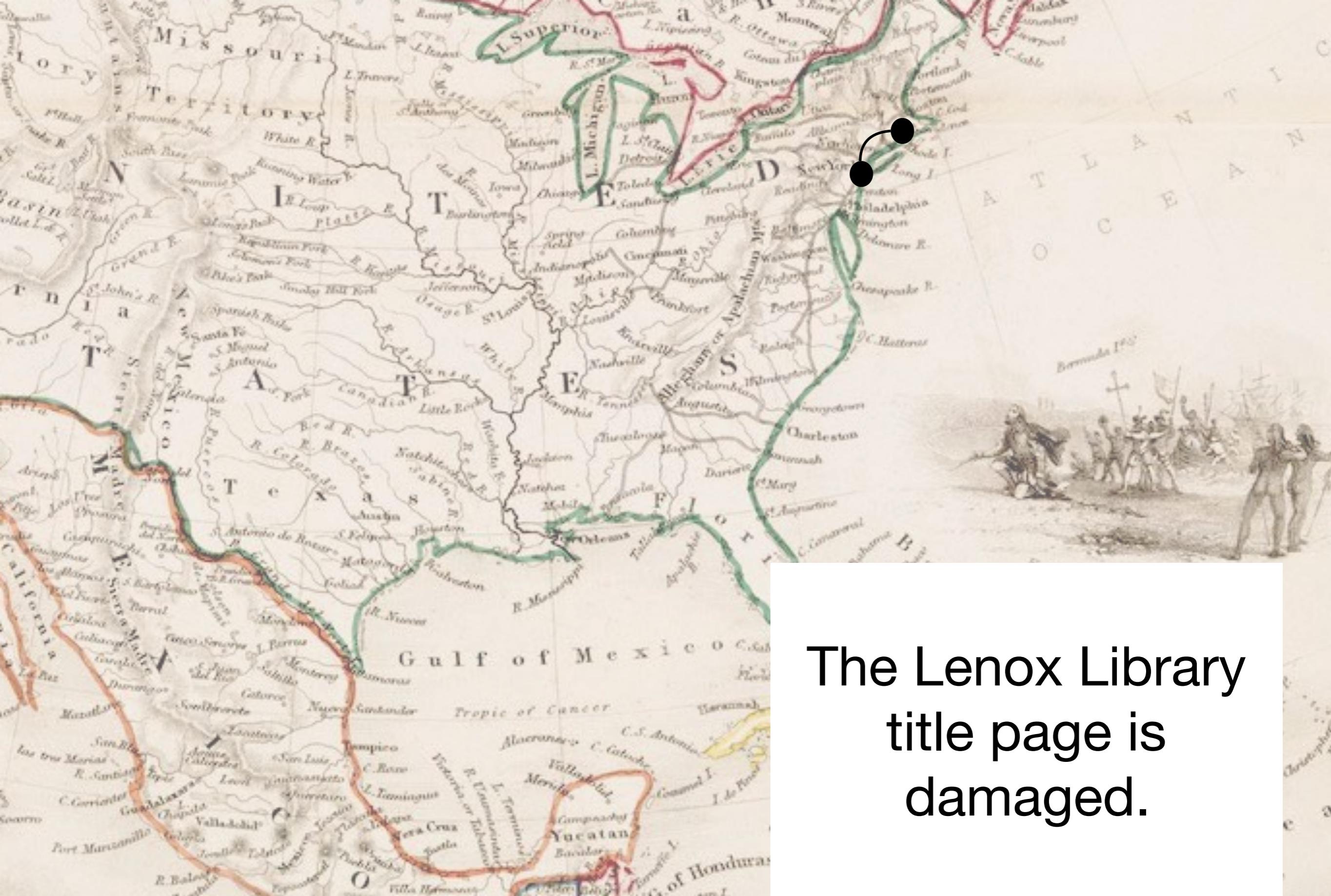
John Carter Brown Library



Copy arrives at
JCBL with pages
missing.



John Russell
Bartlett writes to the
Lenox Library.



The Lenox Library
title page is
damaged.



The British Museum
& Icazbalceta title
pages are damaged.

Aquí comienza vn vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Mol na: Guardiā del couéto del sant Antonio d'
Tetzcuco dla ordē delos frayles Menores.

* tuū Franciscum sig

* nis redemptionis

A substitute
woodcut is located

Benson Latin American Collection

Arte de la lengua Nōxri
canay Castellana , compuesta por el muy Re
uerendo padre fray Alonso de Mo
lina dela orden de Señor
sant Francisco.

JOHN CARTER BROWN



En Mexico en casa de Pedro Ocharte. 1571

John Carter Brown Library

Aquí comienza un vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Mol na: Guardiā del couéto del sant Antonio d'
Tetzcuco dla ordē delos frayles Menores.

* tuū Franciscum sig

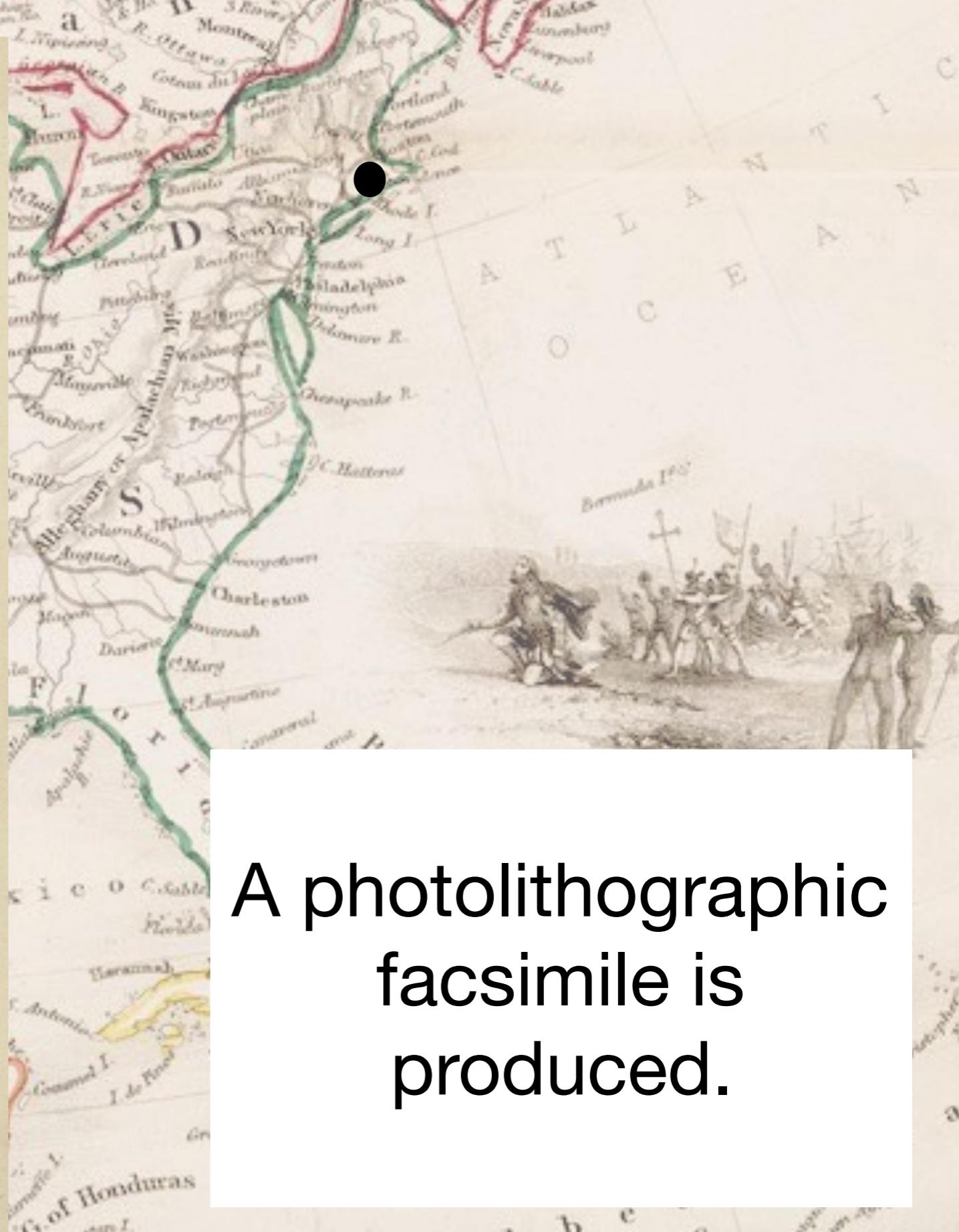


* misericordia domine seruum

* filis redēptionis nostre::

¶ dorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues,
Confixus viuis, langues: cum mente revoluta,
vulnera, cum speceras, stigmata carne geris.

Benson Latin American Collection





The facsimile is sent
to Icazbalceta.



Icazbalceta's
collection is
acquired by Texas.

Aquí comienza vn vocabula-
no en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignas† domine seruum

¶ dorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

* nis redemptionis nostre::

Aquí comienza vn vocabula-
rio en la lengua Castellana y Mexicana, Compuesto
por el muy reuerendo padre fray Alonso de
Molina: Guardián del couéto del sant Antonio d'
Tetzcuco dla ordé de los frayles Menores.

* tuū Franciscum sig



†ignas† domine seruum

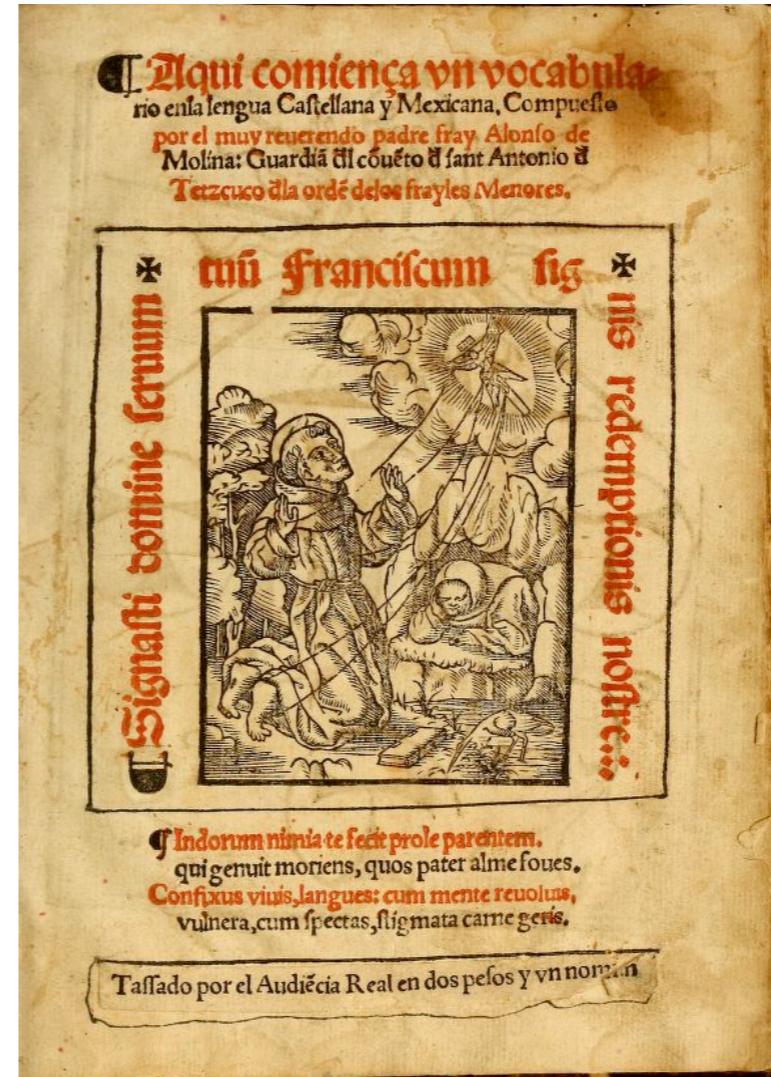
¶ Indorum nimia te fecit prole parentem.
qui genuit moriens, quos pater alme soues.
Confixus viuis, langues: cum mente revolus,
vulnera, cum speceras, stigmata carne geris.

Tassado por el Audiencia Real en dos pesos y vn nombrón

John Carter Brown Library (copy 1)

Benson Latin American Collection

Digital facsimile



Digital facsimile

Photolithographic facsimile



Digital facsimile

Photolithographic facsimile

“Photoshopped” image

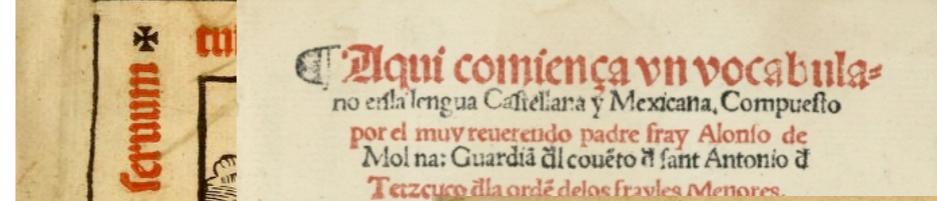


Digital facsimile

Photolithographic facsimile

“Photoshopped” image

Circulating woodcut



Books move
across borders.

Copies circulate
information.

Copying is
creative.



Multilingual Texts in the Digital Age

Introduction: Molina's *Vocabulario*

Case Study: Copying by Hand

Case Study: Copying by Machine

Conclusion: Copying for
the Collective Good

Transcription:

The sequential replication of texts across surfaces.

Transcription:

This strange and banal exercise.

- Arlette Farge

Transcription:

It is through this action that meaning is discovered.

- Arlette Farge

Yrarios que los tenjades guarda
dos grandes mercedes, y grandes
beneficios, aueys hecho a este pue-
blo, ya esta gente quelos aueys
hablado, como y padres asus
hijos, aueys hecho el deuer para
con vuestro pueblo, y los aueys
declarado, y manifestado los
secretos de otras coracones,
y ellos anoydo, gres cabido:
yo anuestro señor, que los sien-
tan; y entiendo dan ylopon
gan por obra, adonde quiera
que fueren, y estuieren. Plega
adios, que con lagrimas se acuer-
dan deste beneficio, y conselso
consuelen, quando hisieren al
guna cosa, que no conviene:
Señor nuestro, y reynuestro,
señores senadores, y suzes,
poruentura ya osdoy pena
contra proxidad de mis pala-
bras, seays muy bien aventu-
rados, deos nuestro señor dios
muchapaz, y asosiego, y viuays
por muchos años, regiendo, y
gouernando, y gaudando
auestro señor, con vuestros
oficios: dqual es invisable, y
ynpalpable.



Honon mattinemj, maco-
nijmijchuenj injnemja, inj-
gujecan, auhinjuesia: a
uhimac valchocas, maic-
oal mellaquaoas iniquac
itla ipan dioloto, in noma-
lauh, innomotecujn. A
motzon tecotzin amelohi
quijutzin njgueoa: tleá
quijomoma dutia, ma amedz
motlamatca Halili mb: auh
maximotlacotilican, maxi-
motequittilican, maxicmo-
nanamjgujlican inHoq in
naoq que, mioalli, inhecatl.

Capitulo desisiete,
del razonamiento, lleno
de muy buena doctrina
enlomoral, que el señor
hacia asus hijos, quando
ya auja llegado alosa
nos dedis crecion: exor-
tandolos ahuyr, losvi-
cios, ya que se diesen,
alos exercicios denoble-
za, y de virtud.

Tic cax tolli omome ca-
pitulo, vncan moteneoa,
centhamati cerca qualli, le
nonotzaliztlatolli, neremj
liztilony: injc qujn nonotba-
ja, i pilhoan tlatolanj: inj-
quacie ixtlamati, ietlaca
quj, qujn tlaqua uhma-
raia, injc quj tlalcahujs
que, injc qujch in aqualli,
mnaiectli. Auh in jecuhtla
quauhtzitzgujsque, in pilte-
gajt, in tlatocatequjtl: auh

Culiba aparecieron gente de mas cuenca y
sus principales. Los unos y los otros vini-
eron á la Laguna de Mexico. Los de ^{aut.}Cul-^{hua} Laguna de
Mexico
liba entraron por la parte de Oriente, y
edificaron un pueblo que se dice Tulancin. ^{Pueblo}
Tulaningo
co, diez y siete leguas de Mexico, y de alle
Tollan(hoy Tula)
fueron a Tula, doce leguas de Mexico, á la

Culhuas

tollantrino hoy
tulancingo

tollan (hoy Tula)

Texcoco

Texcoco

Texcoco

Texcoco
Festato, que es la orilla del agua de la La-
guna de Mexico, cinco leguas de ^{esta}
Texcoco
y ocho de Bojio. Tescoco está la parte de
Oriente, y Mexico al Occidente, la laguna
en medio. Algunos quieren decir que ^{de}
Tescoco se dice Culiba por respeto de estos

que allí poblaron. Despues el Señorío de
Texcoco fue tan grande como el de Mexico.

De allí de Tescoco vinieron á edificar ^{co-}
Co-

PIANOS THAT SERVE



Smithsonian Transcription Center:
National Baptist Metoka and Galeda Bible Class Magazine, September 1917

PIANOS THAT SERVE



[[boxed advertisement]]

PIANOS THAT SERVE

[[double line]]

[[image - small floral sketch]]

Smithsonian Transcription Center:
National Baptist Metoka and Galeda Bible Class Magazine, September 1917



Santalac.

M

Makino Herbarium
Tokyo Metropolitan University

12880

Buckleya lanceolata Miq.

Smithsonian Transcription Center:
Herbarium Project: Santalaceae

TRANSCRIPTION FORM

INSTRUCTIONS

1) General 2) Collector Details 3) Location

Collector #

Collection Date

single date

Collector

Notes on Transcribing this page (optional)

Makino Herbarium 12880
Tokyo Metropolitan University

Buckleya lanceolata Miq.

UNITED STATES
NATIONAL HERBARIUM
2413877

santalac.

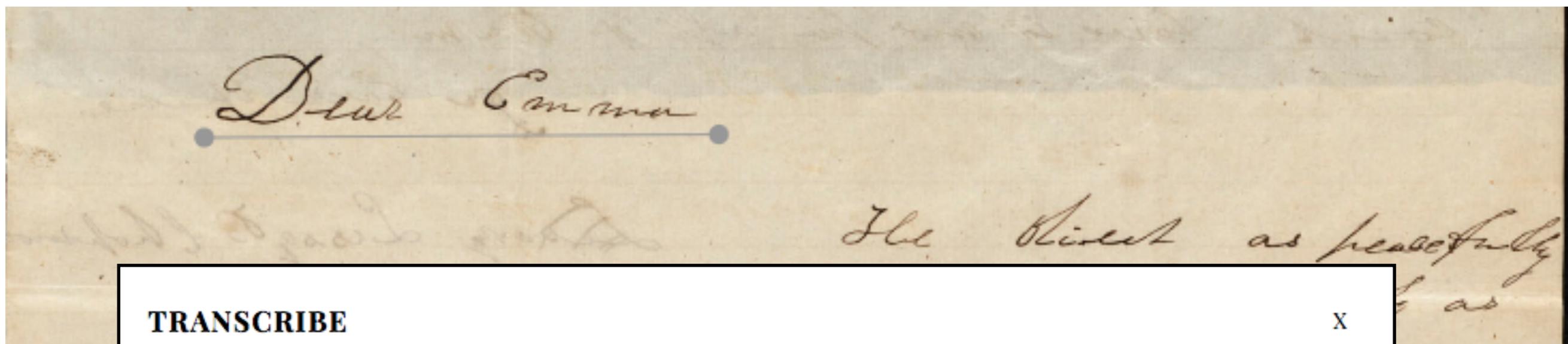
Smithsonian Transcription Center:
Herbarium Project: Santalaceae

Dear Emma

He slept as peacefully
as an infant He looks contentfully at

Anti-Slavery Manuscripts
antislaverymanuscripts.org

Zooniverse & Boston Public Library



TRANSCRIBE

X

Please transcribe all of the words in the line of text.

TEXT MODIFIERS [insertion] [deletion] [unclear] [underline]

Dear Emma

Delete

Cancel

Done

Anti-Slavery Manuscripts
antislaverymanuscripts.org

Zooniverse & Boston Public Library

هذا كتاب الدر النقي في فن الموسيقى
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

al-Durr al-naqī fī fann al-mūsīqī الدر النقي في الفن الموسيقي Mawsilī, Ahmad ibn ‘Abd al-Rahman
[2v] (27/70) موصلي، أحمد بن عبد الرحمن

هذا كتاب الدر النقي في فن الموسيقى
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

- [1] هذا كتاب الدر النقي في فن الموسيقى
- [2] بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

al-Durr al-naqī fī fann al-mūsīqī Mawsilī, Ahmad ibn ‘Abd al-Rahman
موصلی، أَحْمَدُ بْنُ عَبْدِ الرَّحْمَنِ [2v] (27/70)

What is language?

What is text?

What is information?



TICHA

A DIGITAL TEXT EXPLORER FOR COLONIAL ZAPOTEC

Ticha allows users to access and explore many interlinked layers of texts from a corpus of **Colonial Valley Zapotec** manuscripts and printed books, including images of the original documents, transcriptions, translations, and linguistic analysis, including morphological interlinearization. Ticha seeks to make this corpus of Colonial Zapotec texts accessible to scholars in diverse fields, Zapotec community members, and the general public.

[\[more\]](#)

SEARCH THE ZAPOTEC DICTIONARY

pichina

Enter a Zapotec word (try 'pichina') to find a translation or [click here](#) to view all entries.

TRANSCRIBE MANUSCRIPTS

91/204 (45%) of manuscripts have transcriptions. However, 113/204 (55%) of manuscripts are still in need of transcriptions. We would love your help!

Tweets by @TichaProject

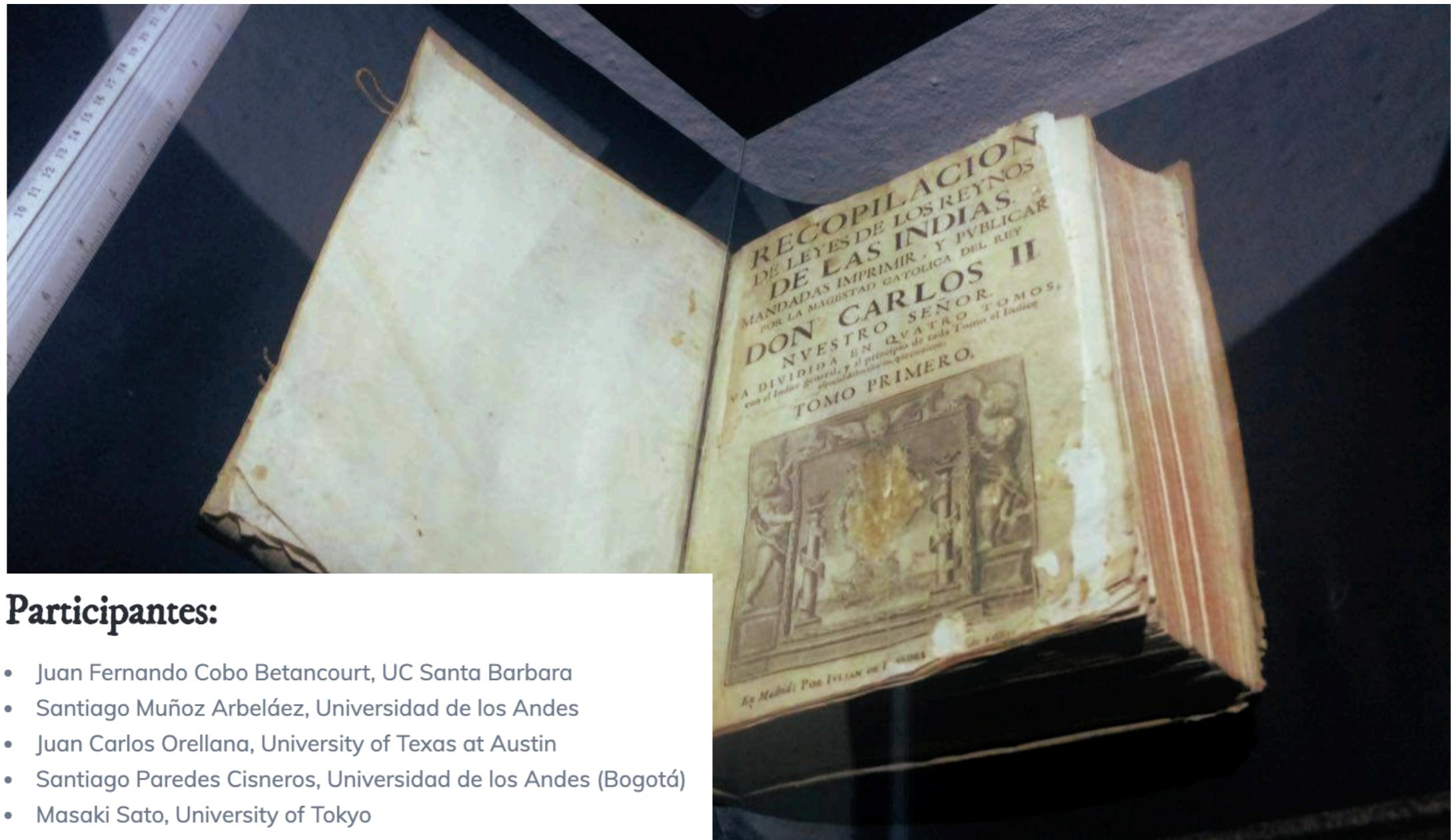
Ticha Project Retweeted

Brook Danielle Lillehaugen
@blillehaugen

Talking about the [@TichaProject](#) today at Penn [@PennPriceLab](#)! Thanks for great conversation!



Apr 16, 2019



Participantes:

- Juan Fernando Cobo Betancourt, UC Santa Barbara
- Santiago Muñoz Arbeláez, Universidad de los Andes
- Juan Carlos Orellana, University of Texas at Austin
- Santiago Paredes Cisneros, Universidad de los Andes (Bogotá)
- Masaki Sato, University of Tokyo
- Natalie Cobo, University of Oxford
- María José Afanador Llach, Universidad de los Andes

Recopilación

Edición digital de la Recopilación de leyes de los reinos de las Indias

neogranadina.org



TICHA

A DIGITAL TEXT EXPLORER FOR COLONIAL ZAPOTEC

Ticha allows users to access and explore many interlinked layers of texts from a corpus of Colonial Valley Zapotec manuscripts and printed books, including images of the original documents, transcriptions, translations, and linguistic analysis, including morphological interlinearization. Ticha seeks to make this corpus of Colonial Zapotec texts accessible to scholars in diverse fields, Zapotec community members, and the general public.

[more]

SEARCH THE ZAPOTEC
DICTIONARY

Q pichina

Enter a Zapotec word (try [pichina](#)) to
find a translation or [click here](#) to view
all entries.

Intellectual

TRANSCRIBE
MANUSCRIPTS

122/449 (27%) of manuscripts have
transcriptions. However, 113/204 (55%)
of manuscripts are still in need of
transcriptions. We would love your
help!

Community-Centered

122/449 (27%) of manuscripts have
transcriptions. However, 113/204 (55%)
of manuscripts are still in need of
transcriptions. We would love your
help!

Retweets by @TichaProject

Ticha Project Retweeted



Brook Danielle Lillehaugen
@blillehaugen

Talking about the @TichaProject today at Penn
@PennPriceLab! Thanks for great
conversation!



Apr 16, 2019

Multilingual Texts in the Digital Age

Introduction: Molina's *Vocabulario*

Case Study: Copying by Hand

Case Study: Copying by Machine

Conclusion: Copying for
the Common Good



¶ Nic. ca ypampa inic yehuantin quicui lozque ininemilitzin ynictlaltecpacmo nemitico in totecuyo Jesu xpo: intleyn quimochihuilico in quimotemachtilico: inictehuantin tictotepotztoquilizque y nitemachtiltzin. ¶ No. Catlehualt ino quicuiloque in nahuintin in Euangeli-tame.

¶ Nic. La yehuatl ynauhtlamantli yn Euangilio yn ipan mopia yn ixquich totech monequi in ticneltocazque yn tic chihuazq. Yhuá intleyn tictlalcahuizq intic telchihuazque. Yhuan intley tichi huazq yye yxqch totech monequi. Tluh yntlacamo ticneltocazque in tlein qui-cuilotiaque çan nimá ahueltitomaquiax-tizqui.

¶ No. nicmatiznequi yn ic tiazque yn il-huicac cuix ça yeyyo ticneltocazque yn oquicuilotiaque.

¶ Nic. ca itlacamo tictopielicā tictone milizticā initeotenahuatiltzin, yyehuá tin inquicauhquiaque caniman ahuel tiaz que yn ilhuicac.

b ij

¶ Nic. ca ypampa inic yehuantin quicui lozque ininemilitzin yn ictlaltecpacmo nemitico in totecuyo Jeju x~po: intleyn quimochihuilico in quimotemachtilico: inictehuantin tictotepotztoquilizque y nitemachtiltzin. ¶ No. Catlehualt ino quicuiloque in nahuintin in Euangeli-tame.

¶ Nic. Ca yehuatl yn auhtlamantli yn Euangilio yn ipan mopia yn ixquich totech monequi in ticneltocazque yn tic chihuaz~q. Yhu~a intleyn tictlalcahuiz~q intic telchihuazque. Yhuan intley tichi huaz~q yye yx'qch totech monequi. Tluh yntlacamo ticneltocazque in tlein qui-cuilotiaque çan nim~a ahueltitomaquiax-tizqui.

¶ No. nicmatiznequi yn ic tiazque yn il-huicac cuixça yeyyo ticneltocazque yn oquicuilotiaque.

¶ Nic. ca ~itlacamo tictopielic~a tictone milizticā initeotenahuatiltzin, yyehu~a tin inquicauhquiaque caniman ahuel tiaz que yn ilhuicac.

b ij

Chic. ca yampampa inic yehuantin quicui lozque ininemtiltzin ynictlalticpacmo nemitico in totecuygo Jesu xpo:intleyn quimochihuilico in quimotemachtilico: inictehuantin tictotepotztoquilizque y nitemachtiltzin. **C**Mo. Catlehuatl ino quicuiloque in naquintin in Euangeliſtame.

Chic. La yehuatl ynauhlamantli yn Euangilio ynipan mopia ynixquich totech monequi in ticneltocazque yn tic chihuazq. Yhuā intleyn tictlalcahuizq intic telchihuazque. Yhuan intleytichi huazq yxe yxqch totech monequi. Ah yntlacamo ticneltocazque in tlein qui cuiotiaque can nimā ahueltitomaquiax tizqui.

CMo. nicmatiznequi ynictiazque yn il huicac cuiç ga yeyyo ticneltocazque yn oquicuilotiaque.

Chic. ca itlacamo tictopielicā tictone militzican initeotenahuatiltzin, y yehua tin inquicauhtiaque caniman ahueltiaz que yn il huicac.

b ij

o — M r^ -^ v^SO |>»00 Cv O — «
r^ -«t- «^lo r^-oo Oso — r^
vo SO vO \0 vO *0 vO vO nO so \0
sOvO vososOnOvOvOvosOvOvO

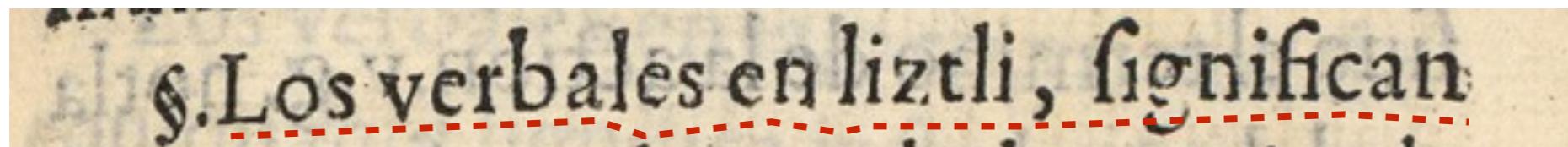
00 (7s O — «* < '♦N ^ w^sO r->00
(7s o «- n r^ 't- «^VO f>.00 Cv o
w^ w^vO sovovosovovovo>ovo r**
r^r^i^i^c^r>.c^r>. r**00

VO r**00 Cv o — *^ «^ -^ «^SO
t>i30 J\ o — r«

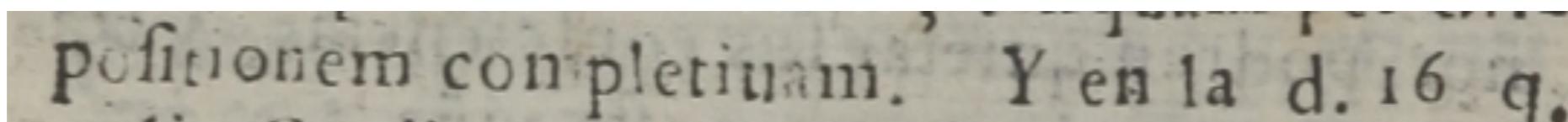
o — M r^ -^ v^SO |>»00 Cv O — «
r^ -«t- «^lo r^-oo Oso — r^
vo SO vO \0 vO *0 vO vO nO so \0
sOvO vososOnOvOvOvosOvOvO

00 (7s O — «* < '♦N ^ w^sO r->00
(7s o «- n r^ 't-

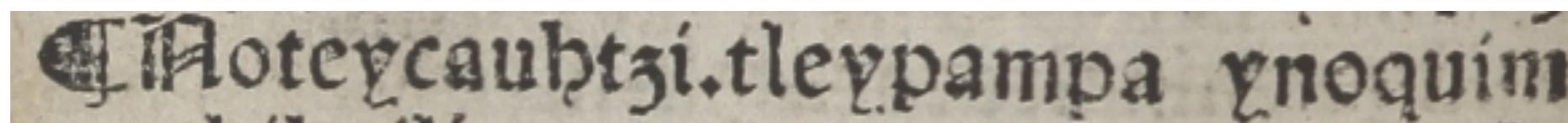
Wandering baseline



Uneven inking



Unfamiliar Typefaces



præfertim **virgoete** **causa**

ligature

non-standard
spelling

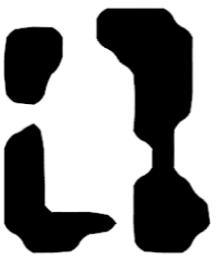
character
ellipsis

obsolete
character

praesertim **urgente** **causa**

¶ Quid dicendum de Ocnamacaque. s. de
las pulqueras, o taberneros, qui vendunt vi-²⁷
num Indorum, Indis, quod dicitur Octli :&
etiam de ijs qui eis vendunt viñu verum His.

Multilingual Transcription





a á b c d e é f g h i í j k l m n ñ o ó p q r s t u ú v w x y z 0 1 2 3 4 5 6 7 8 9 . , : ;

a o it

a o u

a c u

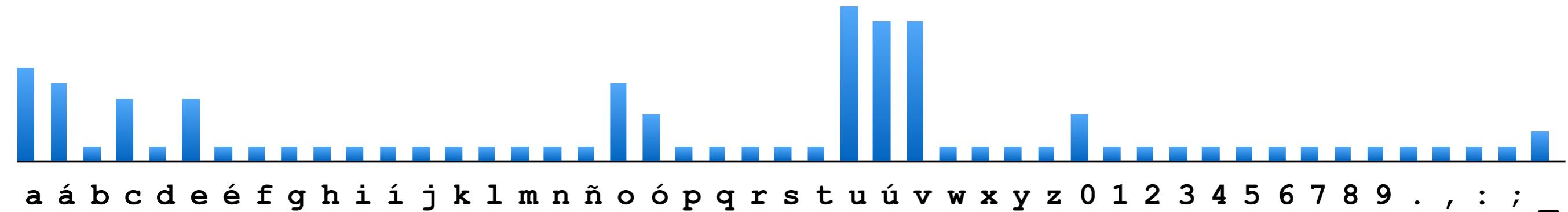
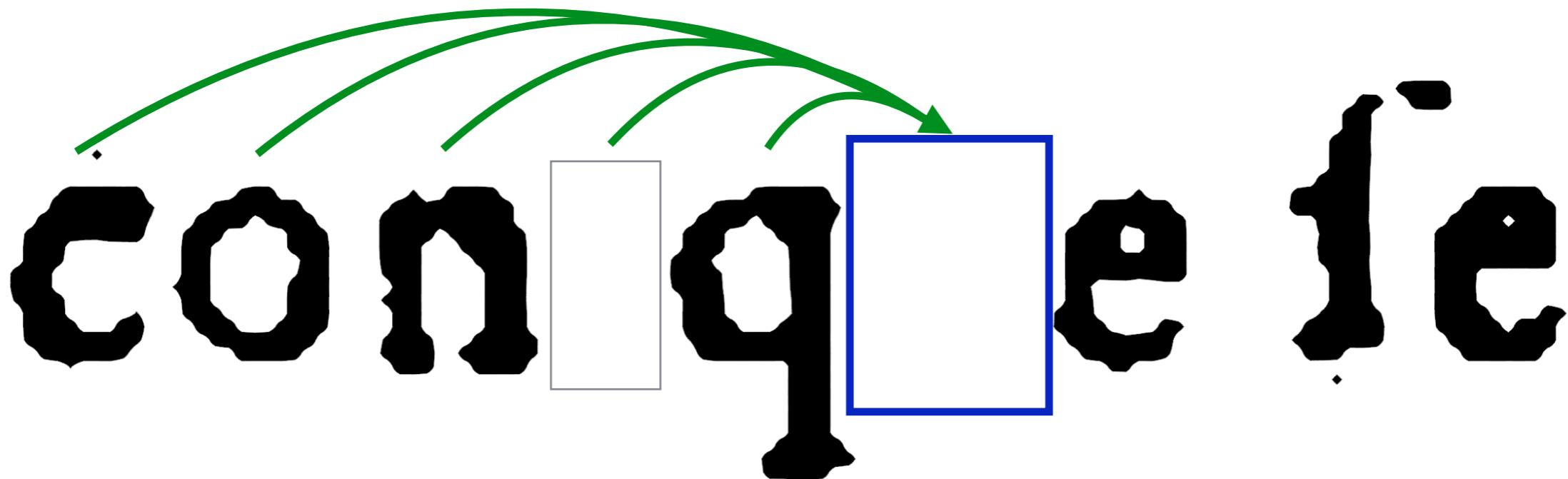
a o ll

a o ll
í

a c u

a á b c d e é f g h í j k l m n ñ ó p q r s t u ú v w x y z 0 1 2 3 4 5 6 7 8 9 . , : ;

Language Model



resp. por la mayor parte. Ca yehuatzin in

1. refp pool a may or parte. Ca yeftnat win fit
2. refponfortia mayor Paris Chairman analyfts
3. refpi por la mayor parte. Ca yehuatzin in

Reading the First Books

Project Team: Sergio Romero, Albert Palacios, Hannah Alpert-Abrams, Maria Victoria Fernández
(UT Austin)

Computer Science: Dan Garrette (U. of Washington), Taylor Berk-Kirkpatrick (UC Berkeley)

Mesoamerican Languages: Stephanie Wood (U. of Oregon), Kelly McDonough, Adam Coon (UT Austin)

eMOP and Primeros Libros: Laura Mandel, Anton Duplessis, Elizabeth Grumbach (Texas A&M University), Trey Dockendorf, Bryan Tarpley, Matt Christy

University of Texas Libraries: Aaron Choate

Aquí comienza un vocabula-
no es la lengua Castellana y Mexicana, Compuesto
por el muy reverendo padre fray Alonso de
Molna: Guardián del couéto del sant Antonio de
Tetzcuco dla ordé delos frayles Menores.

Sigilum Franciscum



¶ *dorum nimia te fecit prole parentem.*
qui genuit moriens, quos pater alme fous.
Confixus viuis, langues: cum mente reuoluta.
vulnera, cum speceras, stigmata carne geris.

niū Franciscum sig

mis redempcionis nostre:

A ante t, & ti. 29

Atronar se la muger, nino, cuitapan mauhtia.
Atronada muger, mocuitapan mauhtiqui.
Atronar a otro con ruido, nite, nacaztitza.
Atronado assi, tlancatzitztli.

C Audiencia delos juezes. tlacacoya, tlatzonte-
coya, tecutlatolya, teccalli.

Audiencia hazer, nitecutlatoa, nitla, caqui.
Auelo o aguelo, colli tecol.

Aullar, nite, coyota.

Aullador, tecoyouani.

Aullido, tecoyoualiztli.

Aun, noma, oc noma,

A vna hazerse, ticcemitoa, tictocentequilia.

A vna parte y a otra, necoc, nenecoc, necoccampa occa-
pa ixti, yyuccampaixti, yyuntlapalixti.

Aun aun, cuixçaoc, cuixcac.

Aun no, ayamo.

Aun no has buelto, ayate, cenza ayate, ayaiztiutz, tica-
ti.

Aunque, yntlanel, ymmanel, ymmaçonel, immaçonel-
iuh, maciui, maçoiui, maçoneliui, aço, yuh.

Ausentarse, canapa niyah, nino, yeltia, ni, cholca, nino,
tlatia, anixpa.

Ausencia, canapa yaliztli, netlatiliztli, ateixpa.

Ausente, ayac, canapayaqui, amo ixpa.

Autor hzedor dios, techuani, tepiquini, teyocoyani.

Autoridad de persona, teixmauhtiliztli, mauizticayotl.

Autoridad tener de persona, nite, ixmauhti, ni, mauiztic.

Autorizada persona, teixmauhti, teixmamaauhti.

Autoridad de scripture, tlaneltiloni, tlatolneltililoni.

Autorizada escriptura, tlaneltililli, tlatollaneltillili.

Autorizar escriptura, ni, tlatolneltilia, na, moxtlatolneltilia.

Autoridad tener para hazer algo, ni, nauatile.

Auaricia tener, ni, teoyeuacati, ni, tlatlameti, ni, tzotzoca
teuitzi, ni, tzotzocati, atle niccaualiztlamati, & per meta-
phoram, atle niquixcaya, aninocotonlani.

G

?? ??

Arronarfe la muger, ni no, cuitlapan mauhtia.
Atró nada muger, mocuitlapan mauhtiqui.
Atronar á otro con ruido, ni te nacaztitiza,
á tronado allí, tlanacaztititztli,
á[Audiencia de los Juezes. tlacaco ya
coya. tecutlatolya, teccalli,
Audiencia hazer, ni tecutlatoa. nitla-caqui.
Aue io o aguelo. colli tecol.
Aul lar, nite, coycua.
At illa dor. tecoyouatit-
Auilido. tecoyoualiztli,
Aun. Jsoma, oc noma.
A yn a hazerse, ticcemitoa, tictocentequilia,
Ayna parte y a otra, necoc, nenecoc, necocca
pa ixti yyuccampaixti yyuntlapalixti.
Aun aun̄cuix çä oc̄c97xçac.
Aun no. ayamo.
Aun no ha abuelto, ayateçcen casayateçayaíz
tli.
Aun que, yntla nel, ymmanel, ymmaçonel, ii
iuh maciui, maçoiui, maçoneltin, aço yuh.
Aufentat fe. canapa niyauh. nino-yeltia, ni-cl
tlatia. anñxpa,
Aufencia. canapa yaliztli, netlatiliztli, atéixpa,
Aufente. ayac. canapa yaqui amo ixpa.
Autor hzedor dios. techiuani. tepi quini, te yc
Autoridad de perfona, teixmauhiliztli. mau
Autoridad tener de perfona, ni tesix mauhti, ni
Autorizada perfona, teixmauhti, teixmamauj
Autoridad de fcriptura. tlaneltiloni, tlatolti
Autorizada ecriptura, tlaneltillili, tlatollanel
Autoriza referiptura, ni-tlatol neltilia. na, moxi
Autoridad tener para hazer algo, ni nauatile,
Auaticia tener, ni ste oyeuacatl, nist latjanteti,
teuitzti-ni, tzotzocati. atle niccaualiztlamati.
photam. atle niquixcatia. anñnocontlani.

Atronar se la muger. nino,cuitlapan mauhtia.
Atronada muger. mocuitlapan mauhtiqui.
Atronara otro con ruido, nite,nacaztitiza,
Atronado assi. tlanacaztititztli.
Audiencia delos juezes. tlacacoya.tlatzonte-
coya. tecutlatolya. teccalli.
Audiencia hazer, nitecutlatoa.nitla,caqui,
Auelo o aguelo. colli tecol.
Aullar. nite,coycua.
Aullador. tecoyouani.
Aullido. tecoyoualiztli.
Aun. noma. oc noma,
A vna hazerse, ticcemitoa, tictocentequilia.
A vna parte y a otra, necoc, nenecoc necoccampa occ-
pa ixti.yyuccampaixti.yyuntlapalixti.
Aun aun̄ cuixçaoç? cuixçac.
Aun no. ayamo.
Aun no has buelto. ayate? cenza ayate? ayaiztiitez? tica
tli.
Aunque. yntlanel, ymmanel, ymmaçonel, immaçonel-
iuh.maciui. maçoiui. maçoneliui. aço. yuh.
Ausentarse, canapa niyauh. nino,yeltia, ni,choloa. nino,
tlatia.anixpa.
Ausencia. canapa yaliztli.netlatiliztli. ateixpa.
Ausente. ayac.canapayaqui. amo ixpa.
Autor hzedor dios. techiuani.tepiquini. teyocoyani.
Autoridad de persona. teixmauhiliztli. mauizticayotl.
Autoridad tener de persona. nite,ixmauhti.ni,mauiztic.
Autorizada persona. teixmauhti.teixmamauhti.
Autoridad de scriptura. tlaneltiloni, tlatolneltiloni.
Autorizada ecriptura. tlaneltillili, tlatollaneltillili.
Autorizar ecriptura. ni,tlatol neltilia. na,moxtlatolneltilia
Autoridad tener para hazer algo. ni.nauatile.
Auaticia tener. ni,teoyeuacati. ni,tlatlameci. ni,tzotzoca
teuitzti.ni,tzotzocati.atle niccaualiztlamati. & permera
phoram. atle niquixcatia. anñnocontlani.

G

Atronada muger, **mocuitapan mauhtiqui.**

Atró nada muger,
mocuitapan mauhtiqui.

Atronará otro con ruido, **nite,nacaztitza.**

Atronar á otro con ruido,
ni te nacaztitza,

Atronado así. **tlanacaztitztli.**

á tronado allí,
tlanacaztitztli,

¶ Audiencia de los juezes. **tlacacoya**

ã[Audiencia de los Juezes.
tlacaco ya

Multilingual Transcription

| | | |
|---|----------------|---------------|
| | mentira | mérita |
| Without orthographic variation handling | mentira | merita |
| With orthographic variation handling | mentira | m̄etira |

Orthographic Variation

li yiollo, oquimmopanauili in cemana-

Norm: l̄ yollo, oquim~~mopan~~atili in cemana-
Hist: l̄ yollo, oquimmopanauili in cemana-

false
substitution

oac tlaca.

Mod: oac tlaca.
Hist: oac tlaca.

In iehoatzi sant Ioseph, imaceal omu

Norm: In iehoatzi sant joseph. ymaceal om~~o~~-
Hist: In iehoatzi fant fofeph. imaceal omu

correct
substitution!

Historical Orthography & Normalization



Initiative for Digital Humanities, Media, and Culture

Results Filter

| | | | | |
|---|-----|---------------------|-----|----------------|
| Ground Truth | All | Language | All | Add Conditions |
| Works | All | Print Font | All | |
| Collection | All | OCR completed date: | | |
| OCR Batch | All | From: | To: | |
| <input type="button" value="Reset Filter"/> <input type="button" value="Apply Filter"/> | | | | |

| Job Queue | |
|--------------------|-----|
| Scheduled: | 545 |
| In-Progress: | 0 |
| Await Postprocess: | 0 |
| Failed: | 9 |
| Ingested: | 33 |
| Ingest Failed: | 0 |
| TOTAL: | 587 |

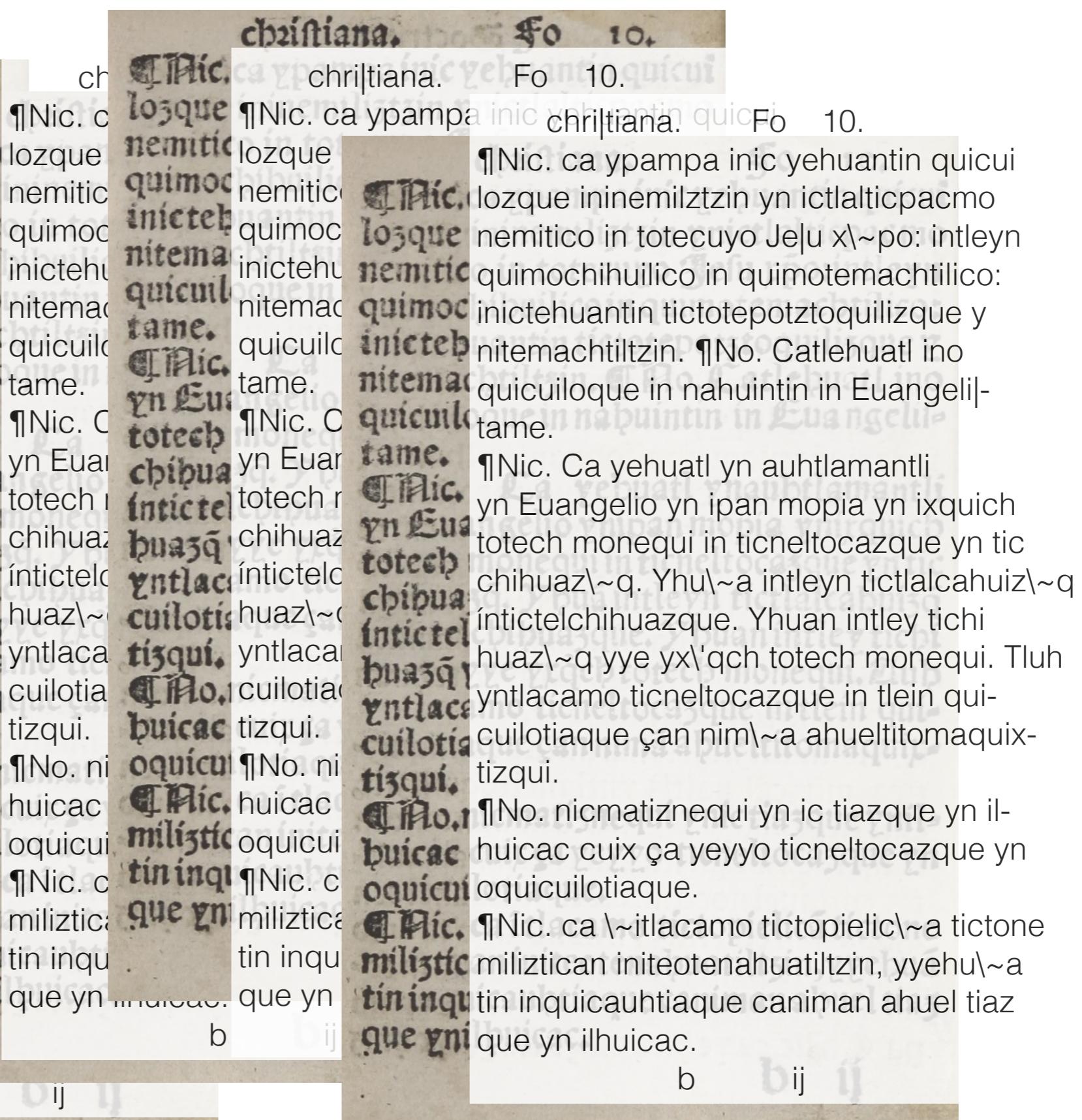
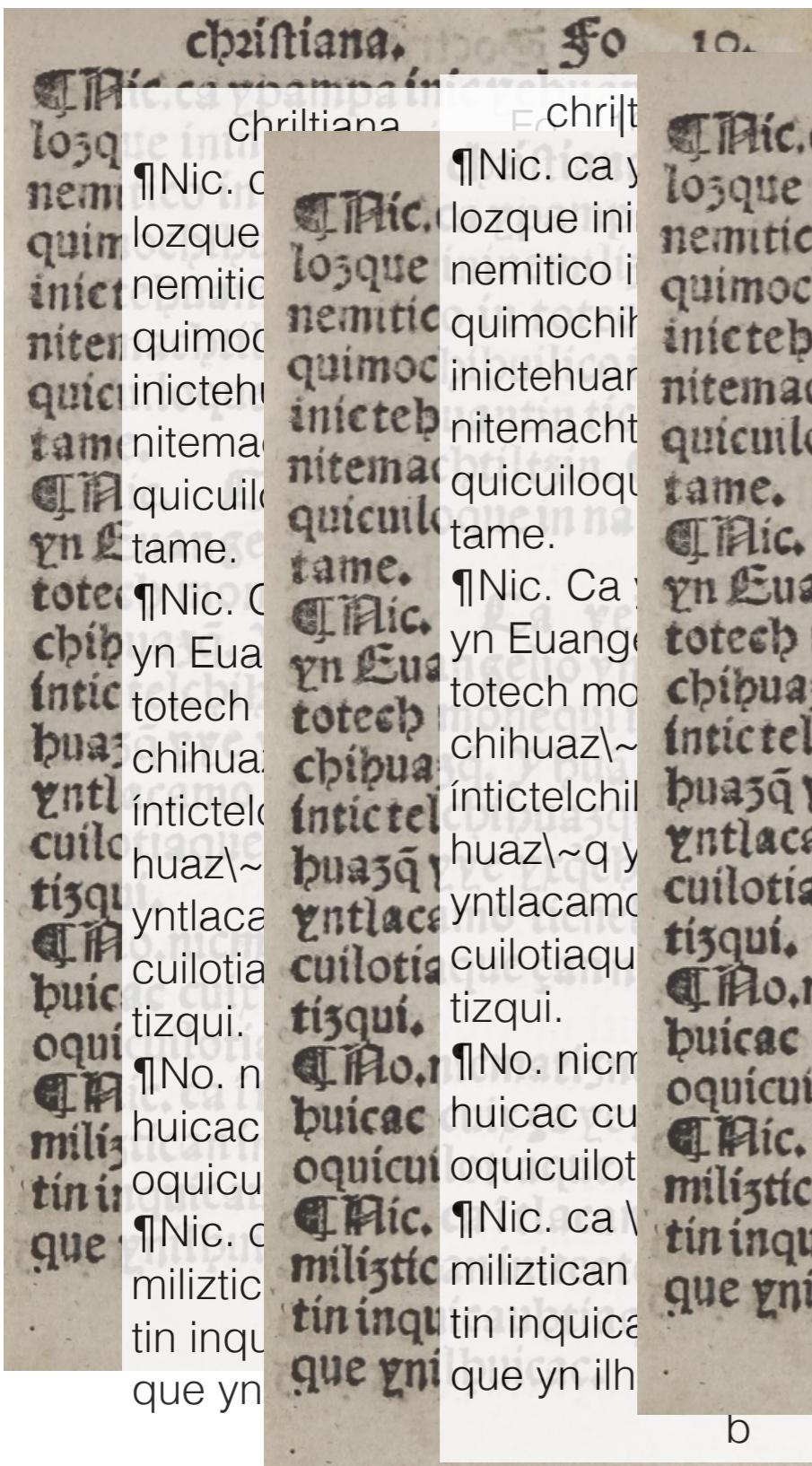
Select All | Schedule Selected | Schedule All | Set Print Font for Selected Works

Show 25 entries

| | Status | | Collection | ID | Book ID | GT Number | Language | Title | Author | Font | OCR Date | OCR Engine | OCR Batch | Juxta | RETAS |
|--------------------------|----------|--|------------|-----|-------------|-----------|----------|---|--------|-------------|----------|------------|---------------------|-------|-------|
| <input type="checkbox"/> | 10-0-0-0 | | FirstBooks | 101 | pl_plfx_008 | | mixtec | Vocabulario en lengua mixteca | | roman | | Ocular | 42: 01_testAllBooks | N/A | N/A |
| <input type="checkbox"/> | 10-0-0-0 | | FirstBooks | 102 | pl_plfx_009 | | mixtec | Vocabulario en lengua mixteca | | roman | | Ocular | 42: 01_testAllBooks | N/A | N/A |
| <input type="checkbox"/> | 10-0-0-0 | | FirstBooks | 103 | pl_tamu_006 | | mixtec | Doctrina cristiana en lengua mixteca | | blackletter | | Ocular | 42: 01_testAllBooks | N/A | N/A |
| <input type="checkbox"/> | 0-0-0-0 | | FirstBooks | 104 | pl_tecm_074 | | mixtec | Vocabulario en lengua mixteca | | roman | | | | N/A | N/A |
| <input type="checkbox"/> | 10-0-0-0 | | FirstBooks | 105 | pl_befk_001 | | nahuatl | Vocabulario en Lengua Castellana y Mexicana | | | | Ocular | 42: 01_testAllBooks | N/A | N/A |

Integrating with eMOP

Transcription



Many tools for digitization are
designed for modern, Anglophone
texts.

Making useable transcription is a
long-term, collaborative project.

Multilingual Texts in the Digital Age

Introduction: Molina's *Vocabulario*

Case Study: Copying by Hand

Case Study: Copying by Machine

Conclusion: Copying for
the Collective Good

Technology can facilitate the
movement of texts across borders.

Technology can be used to improve
access to cultural heritage

Technology can be used to build
community within and across
institutions.

Common Good:

The mandate to provide access to
information and facilitate the
creation of knowledge

Collective Good:

Prioritizing both the educational mandate of higher education and the desire to build new communities across sectors, borders, cultures, and institutions.

- Alpert-Abrams, Bliss, and Carbajal, 2019