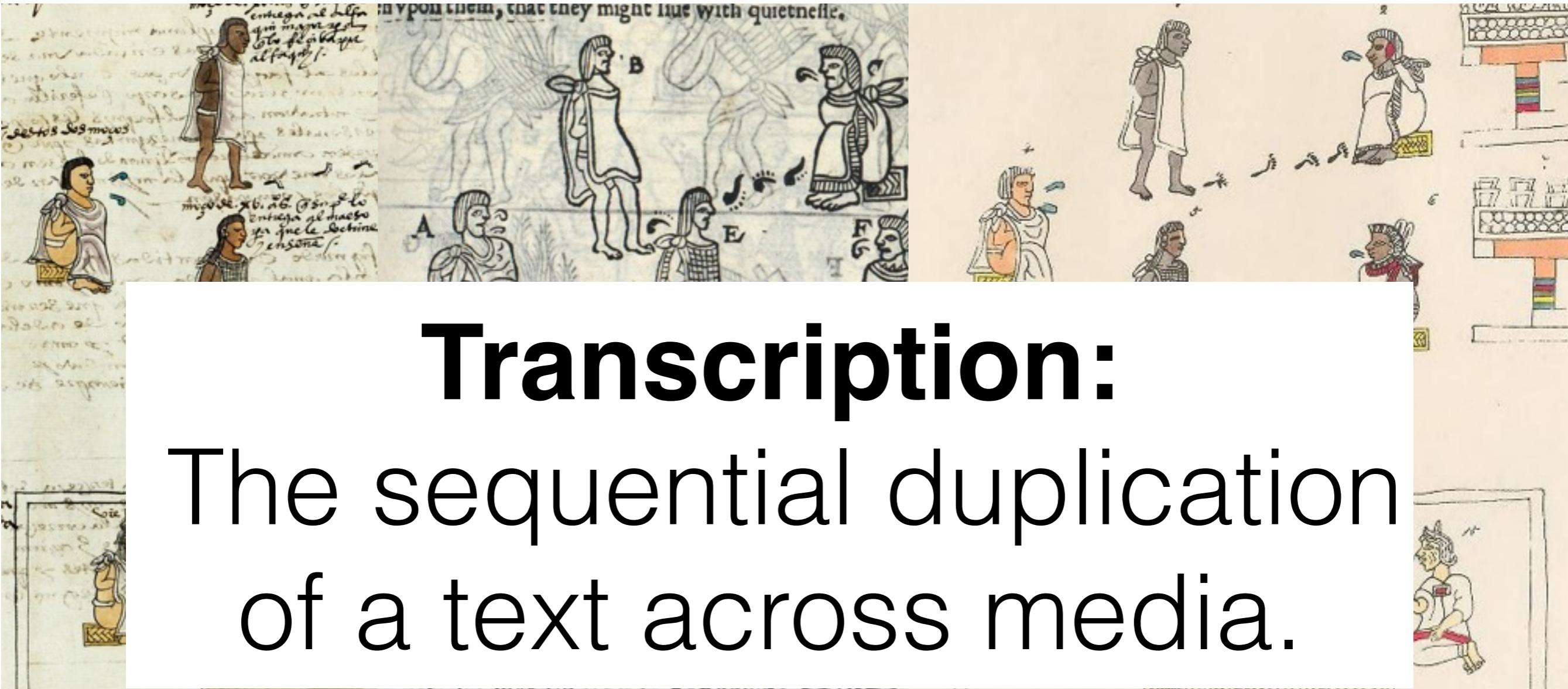


Esta āphibologia no ay ē latin

Machine reading linguistic
hybridity in the *Primeros Libros*

Hannah Alpert-Abrams
University of Texas at Austin

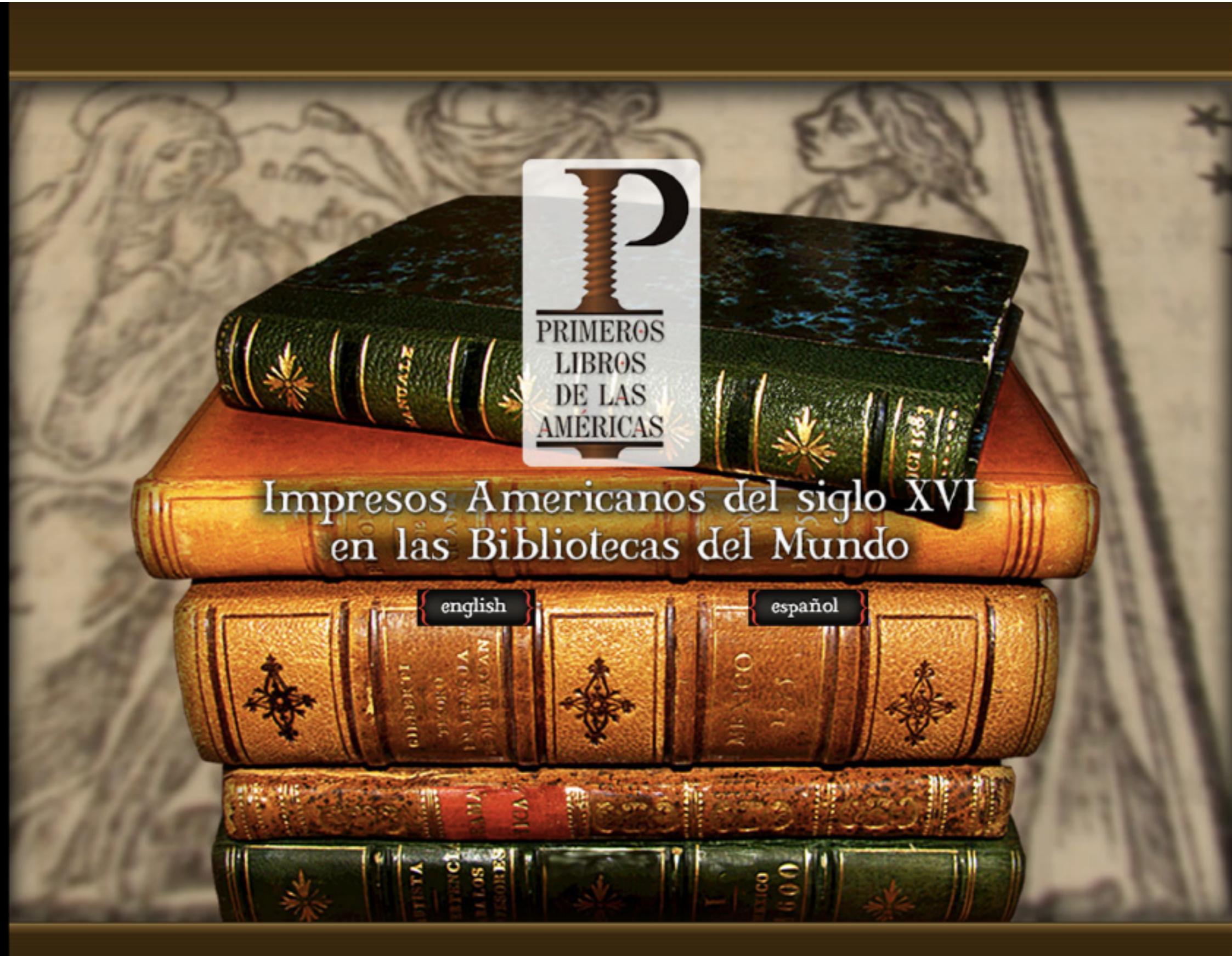
Definition



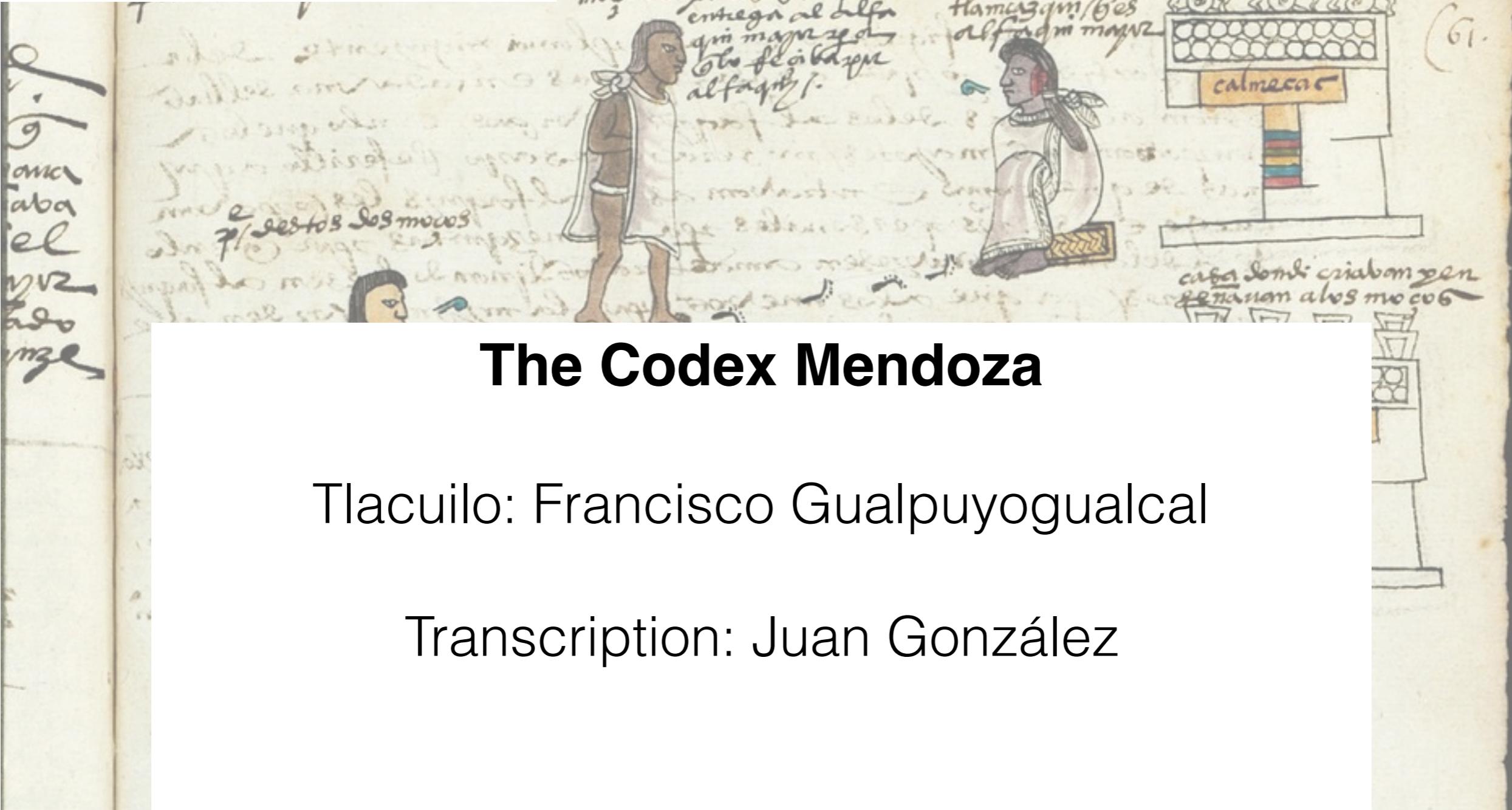
Reading the First Books

Automatic transcription for multilingual, early modern printed documents

sites.utexas.edu/firstbooks/



Colonial Transcription





Capítulo desisiete, del razonamiento, lleno de muy buena doctrina en lo moral, que el señor hacia a sus hijos, quando ya auja llegado a los a nos de discencion: exhortandolos ahuyr, los vi cios, ya que se diesen, a los exercicios de nobleza, y de virtud.

macion mattinemj, maco, nijtijchuenj injnemj, inj qijcahats, auh injuetzta: a uihmaic val chocas, maic oal mellaquaoas inoiquac ita ipanxholoto, in nomalauh, in nomotecujnj. A motzonterecsin antelki

qujuhtzin i, qujmomaduile, maxime moxlamatca halili int: auh maxi motequ, qujmomaduile, maxime nanamqijucan intloq mraoq que, mioai, qujmomaduile, maxime nonotzaliztlatolli, neremj listilonj, injc qujn nonotzia, ipilhoan hatoanj: inj quacie ixhamati, ietlaca quj, qujntlaquauihma raia, injc qujtlacahujs que, injxquich in aqualli, innaiectli. Auh injxquijtl quauhtzitzqujzque, impilte qujtl, intlatocatequjtl: auh injxquich qualli iectli.

The Florentine Codex

Misrendered images reflect imperfect literacy

El Título de Santa María Ixhuatán (17th c.)

Teohuanhuaco:

*Tacuilo pochot tacuilo
teuopixqui*

*(dibujo de una ceiba, dibujo
de un sacerdote)*

Teowakwawnawako

lugar al lado del árbol divino

Teokwawko

lugar del árbol divino

19th Century Transcription

bieron³ S. principales, que
gobernaban⁴ por capitanes. Los de
Culiba aparecieron⁵ gente de mas cuantia y
S. principales. Los unos y los otros vini-
eron á la Laguna de Mexico. Los de Culiba entraron por la parte de Oriente, y
edificaron un pueblo que se dice Culancion⁶ Pueblo
^{Culaniungs} Tlanco⁷ Tlancio
tollan (hoy Tula)⁸ Tlancio
fueron a Tula, doce leguas de Mexico, á la
parte del Norte, y vinieron poblando hacia
Texcoco Texcoco⁹ Texcoco
Tlancio, que es la orilla del agua de la La-
guna de Mexico, cinco leguas de Tlancio
y ocho de Tlancio. Texcoco está la parte de
Oriente, y Mexico al Occidente, la laguna
en medio. Algunos quieren decir que Te-
xcoco se dice Culiba por respeto de estos
que allí poblaron. Despues el Señorío de
Texcoco Texcoco¹⁰ Texcoco
Texcoco fue tan grande como el de Mexico.

Transcribing Historical Documents:

information loss
changes in signification
changes in value

Reading the First Books

Automatic transcription for multilingual, early modern printed documents

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

Multiple Languages

"¶ Quid dicendum de Ocnamacaque. f. de
las pulqueras, o taberneros, qui vendunt vi-
num indorum, indis, quod dicitur Octli .&
etiam de iis qui eis venduntur non verum His

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

Bautista, *Advertencias*, 1601

Normalization

Norm.: illustres. E logo os Reitores **que** taõ juntos, derão o mesmo
Hist.: illustres. E logo os Reitores \~qe taõ juntos, derão o mefmo

Norm.: aviso a seus Collegios, **sendo** em todos, & cada hum delles
Hist.: avifo a feus Collegios, fendo em todos, & cada hum delles

Norm.: recebida a nova com as **mesmas demonstrações**, & al**voroço**
Hist.: recebida a nova com as **mefmas demôstrações**, & aluoroço

Norm.: de alegria começando cada **huma** tratar do que con**vinha** pa-
Hist.: de alegria começando cada hū tratar do que conuinha pa-

illustres. E logo os Reitores q̄ eraõ juntos, derão o mesmo
auiso a seus Collegios, sendõ em todos, & cada hum delles
recebida a noua com as **mesmas demôstraçõeſ**, & aluoroço
de alegria, começando cada hū a tratar do que conuinha pa-
ra a festejar, como fizeraõ, & em seus lugares se dirà.

Integrating with eMOP



IDHMC
TEXAS A & M UNIVERSITY

Initiative for Digital Humanities, Media, and Culture

FirstBooks Dashboard

Results Filter

Ground Truth: All | Language: All | Add Conditions

Works: All | Print Font: All

Collection: All | OCR completed date:

OCR Batch: All | From: [] To: []

Reset Filter | Apply Filter

Job Queue

Scheduled	In-Progress	Await Postprocess	Failed	Ingested	Ingest Failed	TOTAL
545	0	0	9	33	0	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_pfx_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_pfx_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

300 exemplars, eight languages

Reading the First Books

NATIONAL ENDOWMENT FOR THE
Humanities

LLILAS BENSON
LATIN AMERICAN STUDIES AND COLLECTIONS

EMOP
EARLY MODERN OCR PROJECT

14

Transcription

christiana.	Fo	10.
chriltana	chrilt	
¶ Nic. ca y	¶ Nic. ca y	
lozque ini	lozque ini	
nemitico i	nemitico i	
quimochik	quimochik	
inictehuar	inictehuar	
nitemacht	nitemacht	
quicuiloqu	quicuiloqu	
tame.	tame.	
¶ Nic. Ca y	¶ Nic. Ca y	
yn Euan	yn Euange	
totech	totech mo	
chihuaz\~	chihuaz\~	
íntictelchi	íntictelchi	
huaz\~q y	huaz\~q y	
yntlacamo	yntlacamo	
cuiilotiaqu	cuiilotiaqu	
tizqui.	tizqui.	
¶ No. nicm	¶ No. nicm	
huicac cu	huicac cu	
oquicuilo	oquicuilo	
¶ Nic. ca \	¶ Nic. ca \	
miliztic	miliztic	
tin inqu	tin inqu	
que yn ilh	que yn ilh	

christiana.	Fo 10.
chriltiana.	Fo 10.
¶ Nic. ca ypampa inic chrltiana.	¶ Nic. ca ypampa inic yehuantin quic Fo 10.
lozque	lozque ininemilztzin yn ictlalticpac
nemitico	nemitico in totecuyo Jeļu x\~po: in
quimoc	quimochihuilico in quimotemachtil
inictehu	inictehuantin tictotepotztoquilizque
nitemac	nitemachtitzin. ¶ No. Catlehuatl in
quicuilo	quicuiloque in nahuintin in Euange
tame.	tame.
¶ Nic. C	¶ Nic. Ca yehuatl yn auhtlamantli
yn Euar	yn Euangelio yn ipan mopia yn ixo
totech	totech monequi in ticneltocazque
chihuaz	chihuaz\~q. Yhu\~a intleyn tictlalc
íntictel	íntictelchihuazque. Yhuan intley ti
huazq	huaz\~q yye yx\~qch totech mone
yntlac	yntlacamo ticneltocazque in tlein o
cuiotia	cuiotiaque çan nim\~a ahueltitom
tizqui.	tizqui.
¶ No. ni	¶ No. nicmatiznequi yn ic tiazque y
huicac	huicac cuix ça yeyyo ticneltocazq
oquicu	oquicuilotiaque.
¶ Nic. c	¶ Nic. ca \~itlacamo tictopielic\~a
miliztic	miliztican initeotenahuatitzin, yye
tin inqu	tin inquicaughtiaque caniman ahue
que yn	que yn ilhuicac.

christiana. Fo 10.

ca ypampa inic chrltiana. quic Fo 10.

¶ Nic. ca ypampa inic yehuantin quicui
lozque ininemilztzin yn ictlalticpacmo
nemitico in totecuyo Jeļu x\~po: intley
quimochihuilico in quimotemachtilico:
inictehuantin tictotepotztoquilizque y
nitemachtitzin. ¶ 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 intley tictlalcahuiz\~q
íntictelchihuazque. Yhuan intley tichi
huaz\~q yye yx\~qch totech monequi. Tluh
yntlacamo ticneltocazque in tlein qui-
cuitotiaque çan nim\~a ahueltitomaquiax-
tizqui.

¶ No. nicmatiznequi yn ic tiazque yn il-
huicac cuix ça yeoyo ticneltocazque yn
oquicuitotiaque.

¶ Nic. ca \~itlacamo tictopielic\~a tictone
miliztican initeotenahuatitzin, yyehu\~a
tin inquicaughtiaque caniman ahuel tiaz
que yn ilhuicac.

20th Century Transcription

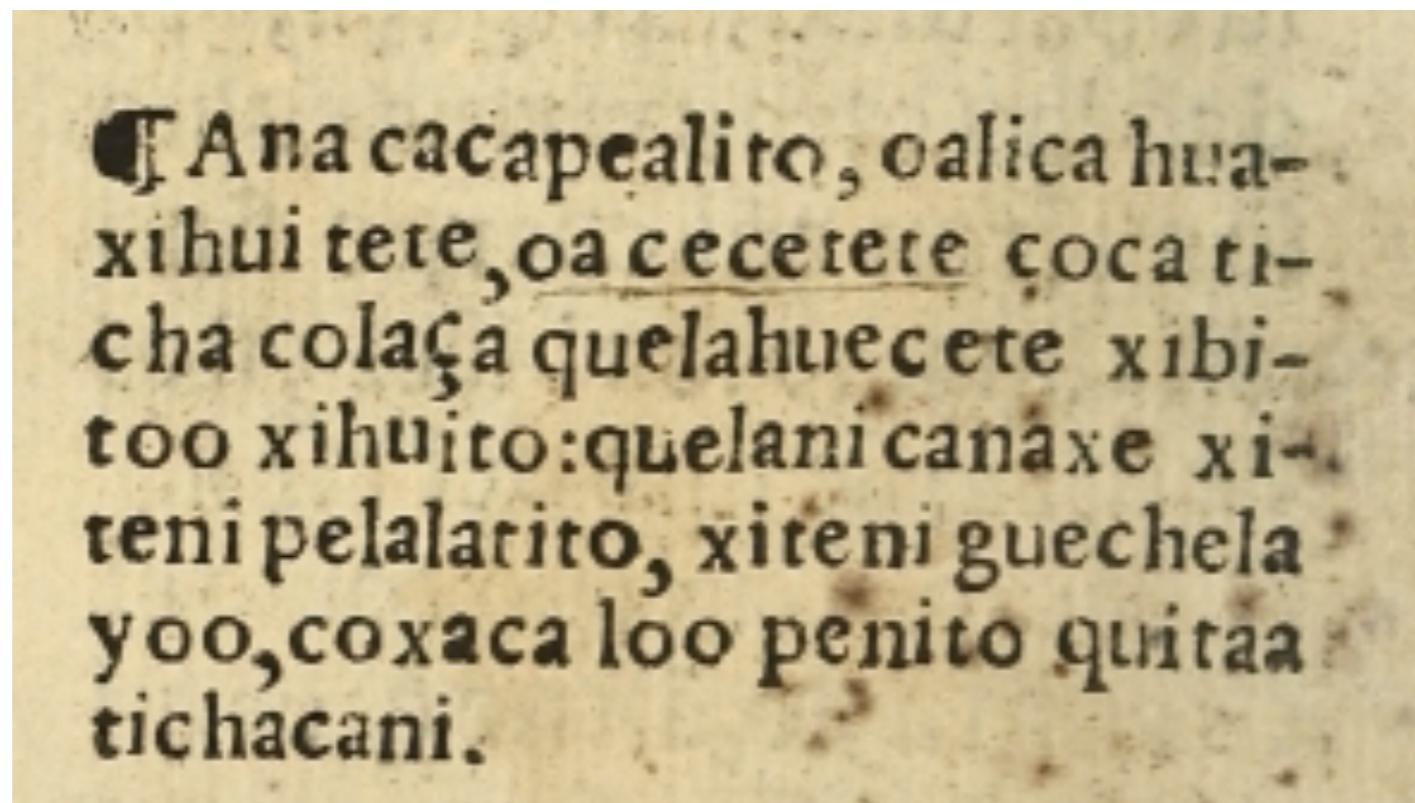
spacing

æ Ana **caca pea lito**, oalica hua-xihui teresoa cgçetete goca ti-cha colaça quelahue cete Xibitoo xihui torquelani canâ xe xi-teni pelala rito, xiteni Sue chela yoo, coxaca loo pẽniso guitia

dirty OCR

96956894-

overcorrection



Feria, *Doctrina Cristiana*, 1567

20th Century Transcription

morpheme
parsing

Norm.: In **teu** iutica **ta** ueuetl, in titlatzca, into-

Hist: In **teu** iutica **ta** ueuetl, in titlatzca, into-

teoyotica
+
tawewetl

Norm: ca metl, in toiauhquauitl, in tepiltzi sancta

Hist: ia metl, in taiauhquauitl, in tipiltzi fancta

Norm: yglesia. in toquichtli xipapaqui, ximotla-

Hist: Iglesia. in toquichtli xipapaqui, ximotla

In teuiütica taucuetl,in titlatzca,into
iametl,in taiauhquauitl,in tipiltzi sancta
Iglesia,in toquichtli xipapaqui , ximotla
machti.

Sahagún, *Psalmodia*, 1583

20th Century Transcription

over -
correction

Norm: l̄ yollo, oquimopanatili in cemana-
Hist: l̄ yollo, oquimmopanauili in cemana-

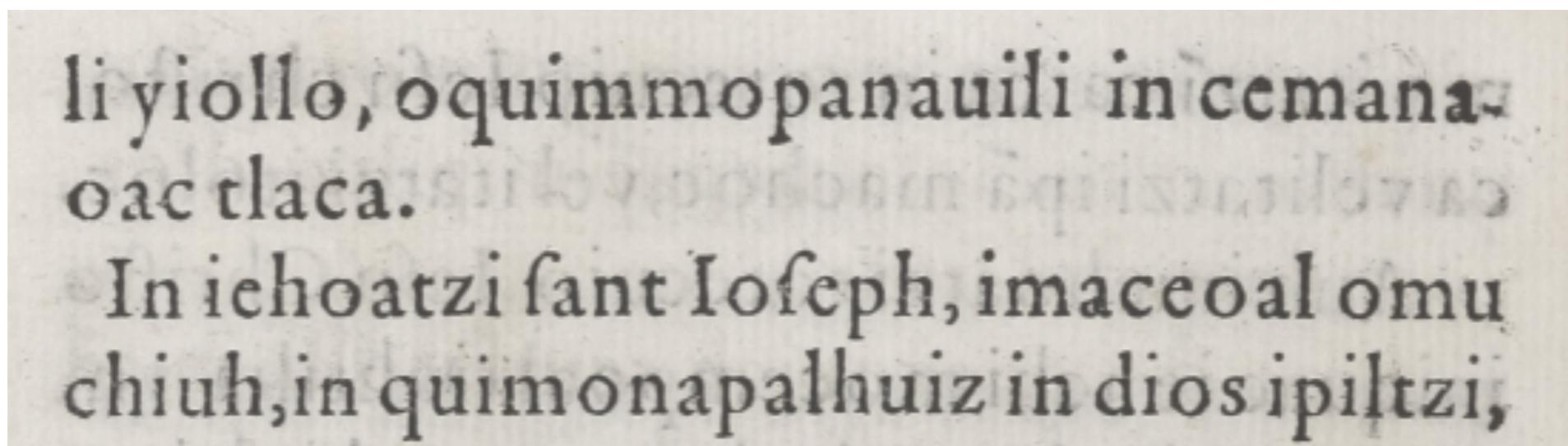
false
substitution

Mod: oac tlaca.
Hist: oac tlaca.

Norm: In iehoatzi sant joseph. ymaceoal omo- —
Hist: In iehoatzi fant fofeph. imaceoal omu

correct
substitution!

Norm: chiuh, in quimonapalhuiz in dios ypiltzi,
Hist: chiuh, in quimonapalhuiz in dios ipiltzi,



Sahagún, *Psalmodia*, 1583

“Reading the First Books”

sites.utexas.edu/firstbooks

Hannah Alpert-Abrams

halperta@gmail.com