Building Early Colonial Corpora for Digital Scholarship

Hannah Alpert-Abrams
CLIR Postdoctoral Fellow
University of Texas at Austin

LLILASBENSON

https://goo.gl/atVK2Q

Hannah Alpert-Abrams

with

Dan Garrette
Taylor Berg-Kirkpatrick
Bryan Tarpley
Maria Victoria Fernández
Albert Palacios
Stephanie Wood
Kent Norsworthy
...

Corpus Creation Challenges

Curation

Creation

Encoding

Colonial Corpus Challenges

Language
Orthography
Epistemology

+ more?

How do we draw on our training as humanists to build better colonial corpora?

How do we draw on our training as humanists to build better colonial corpora?

Methods

Tools

Evaluation

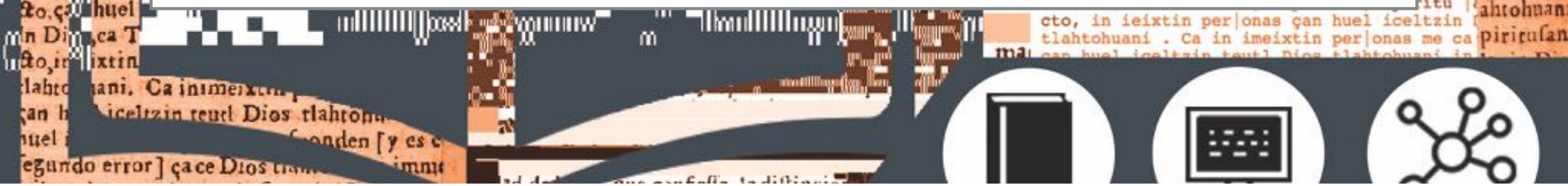
Reading the First Books

Multilingual, early-modern OCR for the Primeros Libros

sites.utexas.edu/firstbooks

primeroslibros.org

emop.tamu.edu



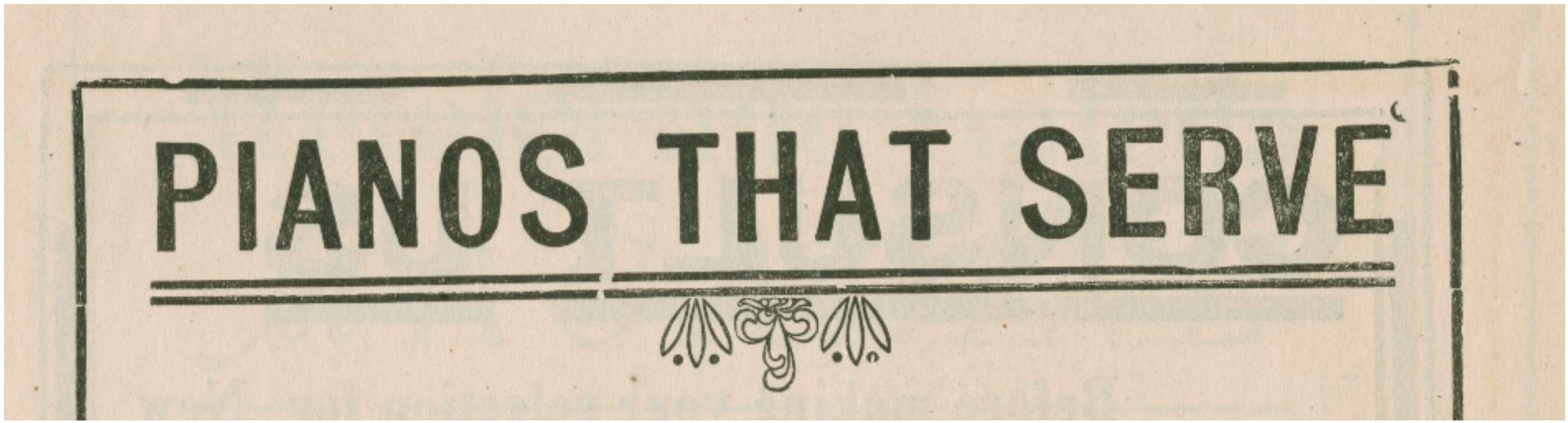
Hannah Alpert-Abrams

https://goo.gl/atVK2Q

Digitization methods shape corpora.

Hannah Alpert-Abrams

https://goo.gl/atVK2Q



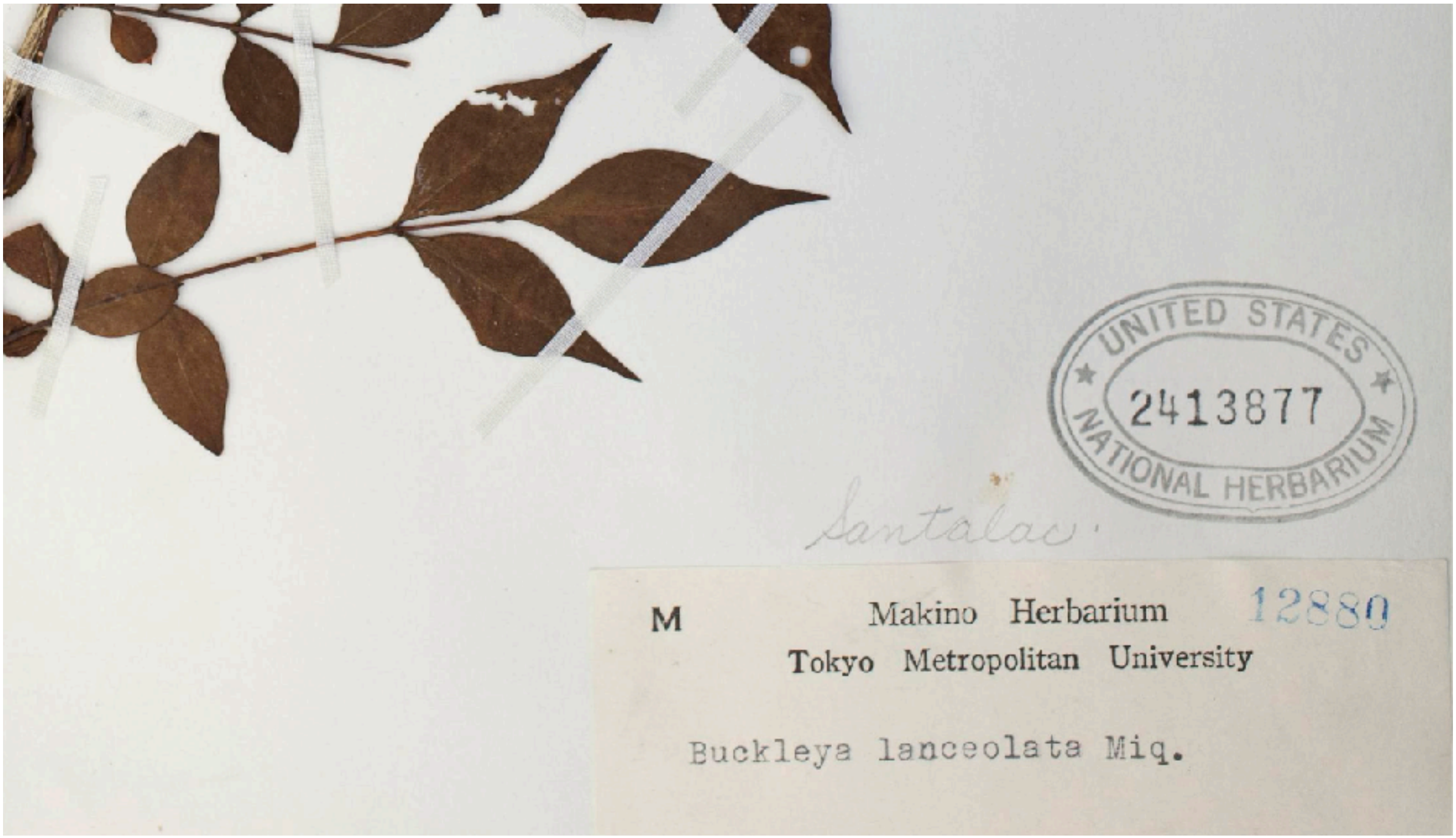
Smithsonian Transcription Center:

National Baptist Metoka and Galeda Bible Class Magazine, September 1917

Hannah Alpert-Abrams https://goo.gl/atVK2Q

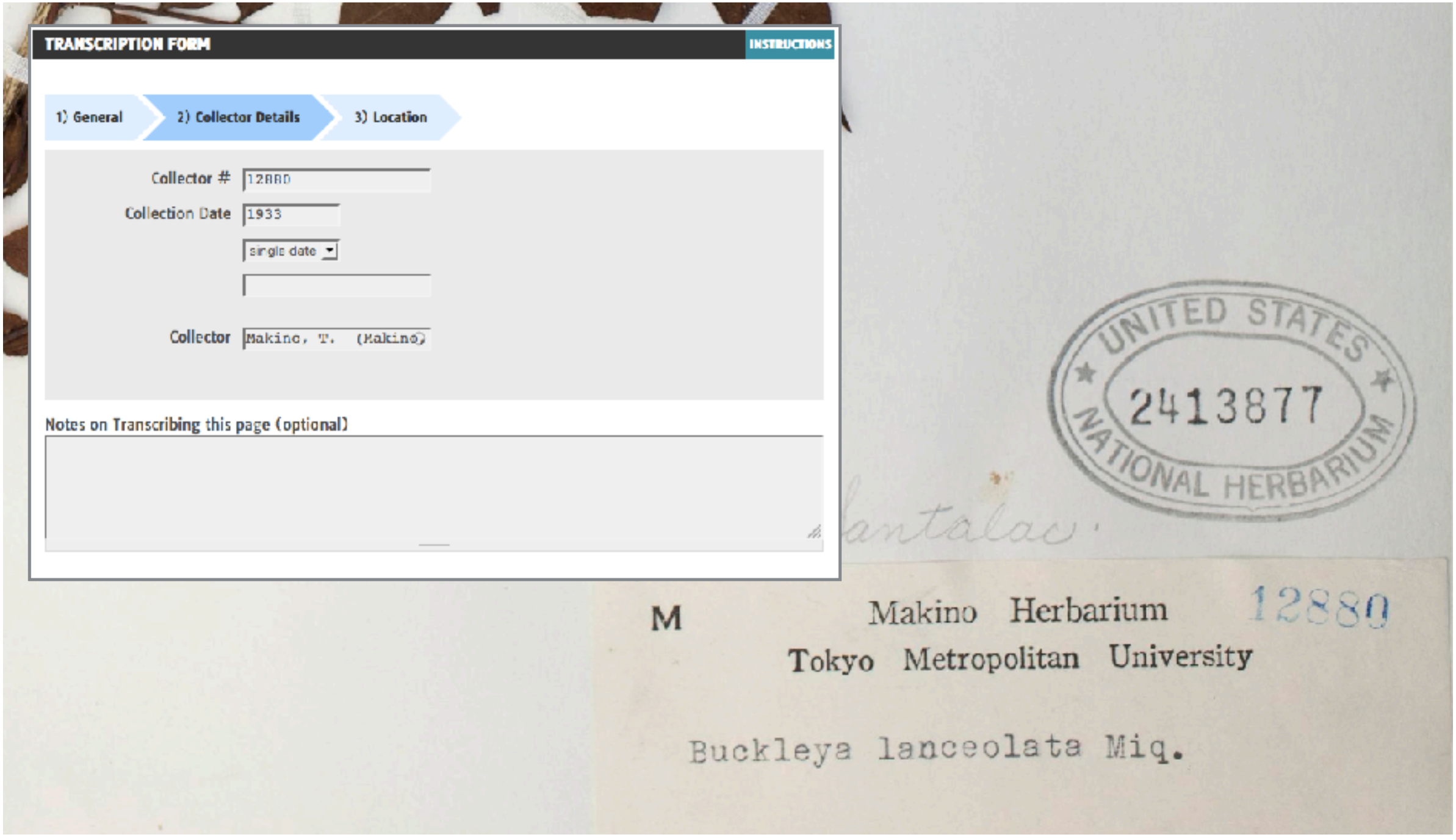
[[boxed advertisement]]
PIANOS THAT SERVE
[[double line]]
[[image - small floral sketch]]

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



Hannah Alpert-Abrams

Smithsonian Transcription Center:
Herbarium Project: Santalaceae



Hannah Alpert-Abrams

Smithsonian Transcription Center:
Herbarium Project: Santalaceae

Elizabeth Hill Boone, The Red and the Black

Viceroy Mendoza
(maguey + tozan [gopher])

Elizabeth Hill Boone, The Red and the Black

Transcription is interpretation.

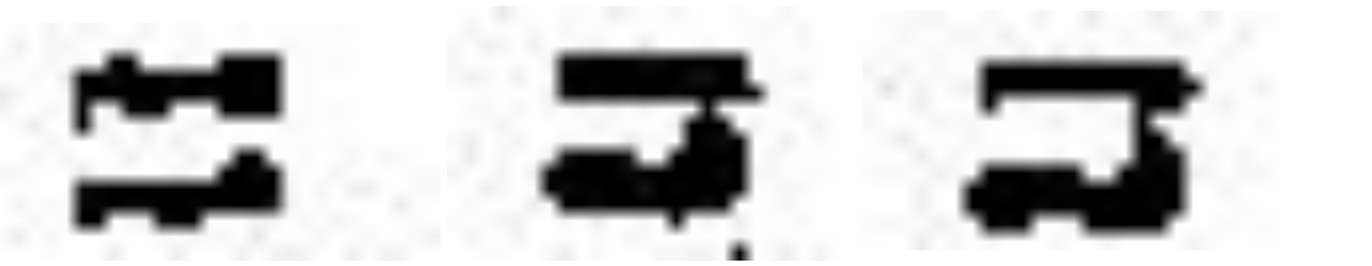
Transcription is specialized interpretive labor.

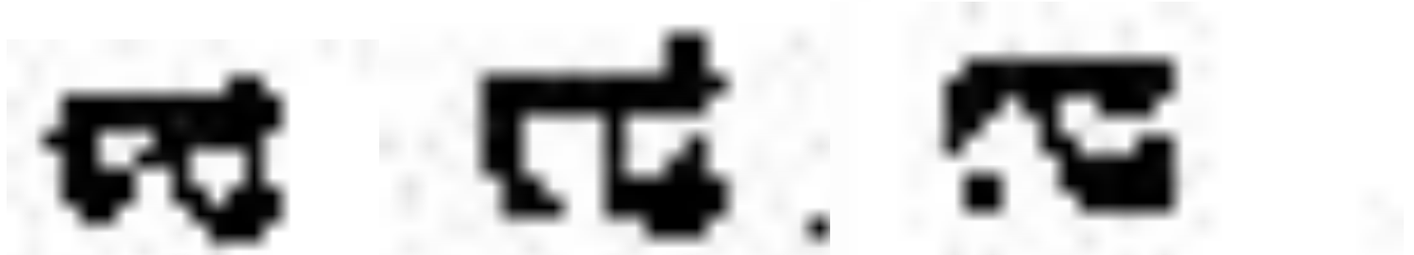
Digitization methods shape corpora.

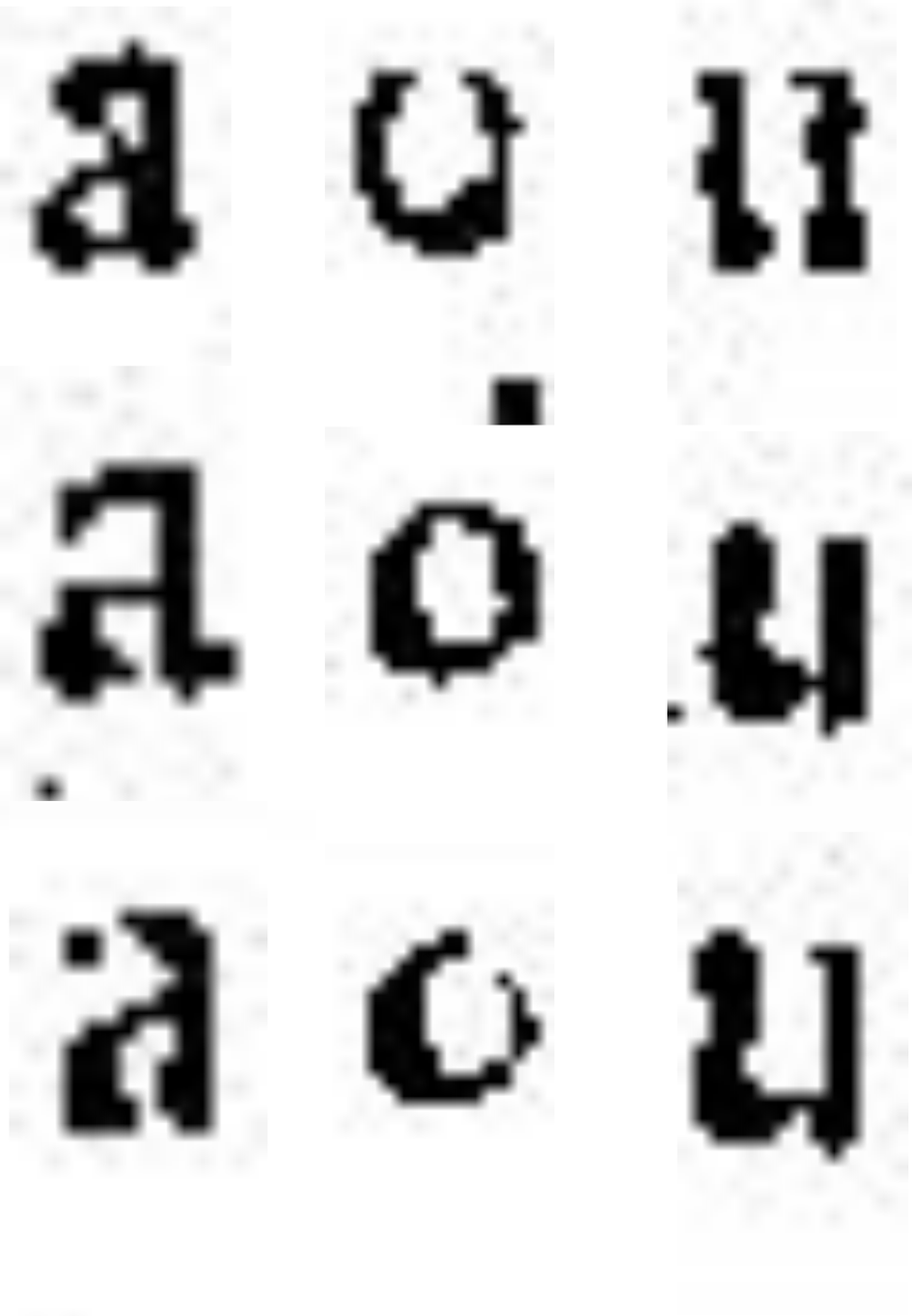
We need to think about the ways we employ labor in colonial corpus creation.

Transcription is specialized interpretive labor even when it is done by a machine.

Digitization tools shape corpora.

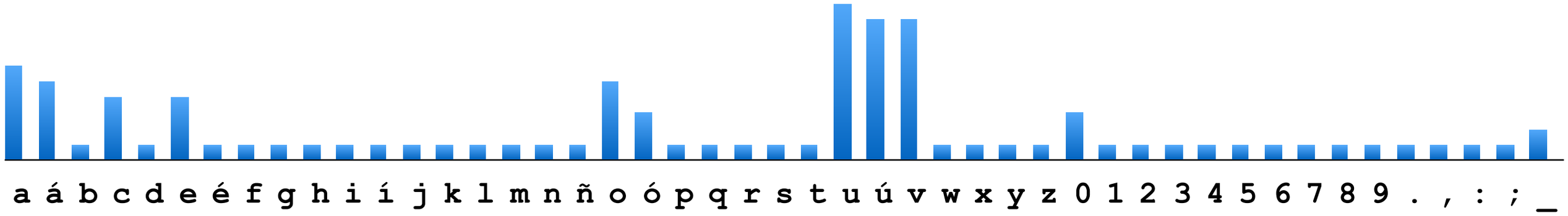


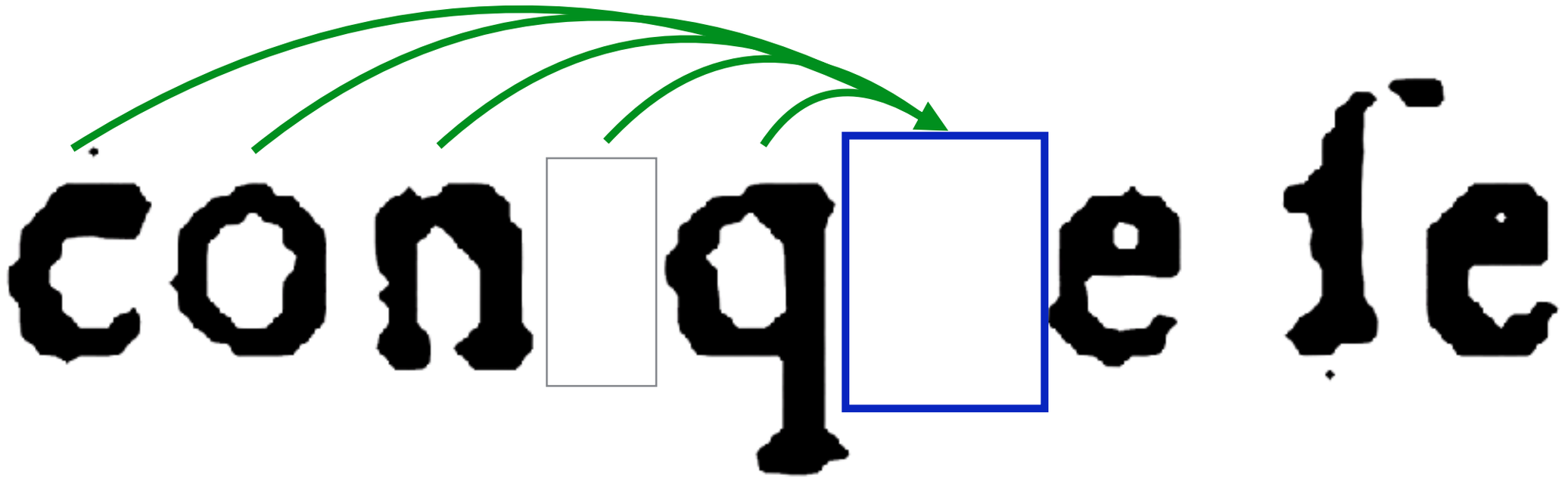


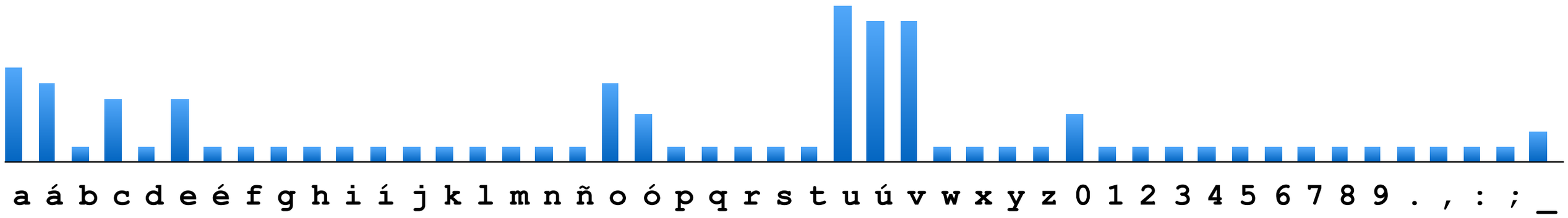


a á b c d e é f g h i í j k l m n ñ o ó p q r s t u ú v w x y z 0 1 2 3 4 5 6 7 8 9 . , : ;

Font Model







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 . , : ; _

con que fe

con que fe

reſp. por la mayor parte. Ca yehuatzin in

reſp pool a may or parte. Ca yeſtnat win ſit

responsortia mayor Paris Chairman analysts

reſpi por la mayor parte. Ca yehuatzin in

Digitization tools shape corpora.

We need tools that are sensitive to the needs of our collections.

Elvia Arroyo Ramírez, Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files.

Hannah Alpert-Abrams

https://goo.gl/atVK2Q

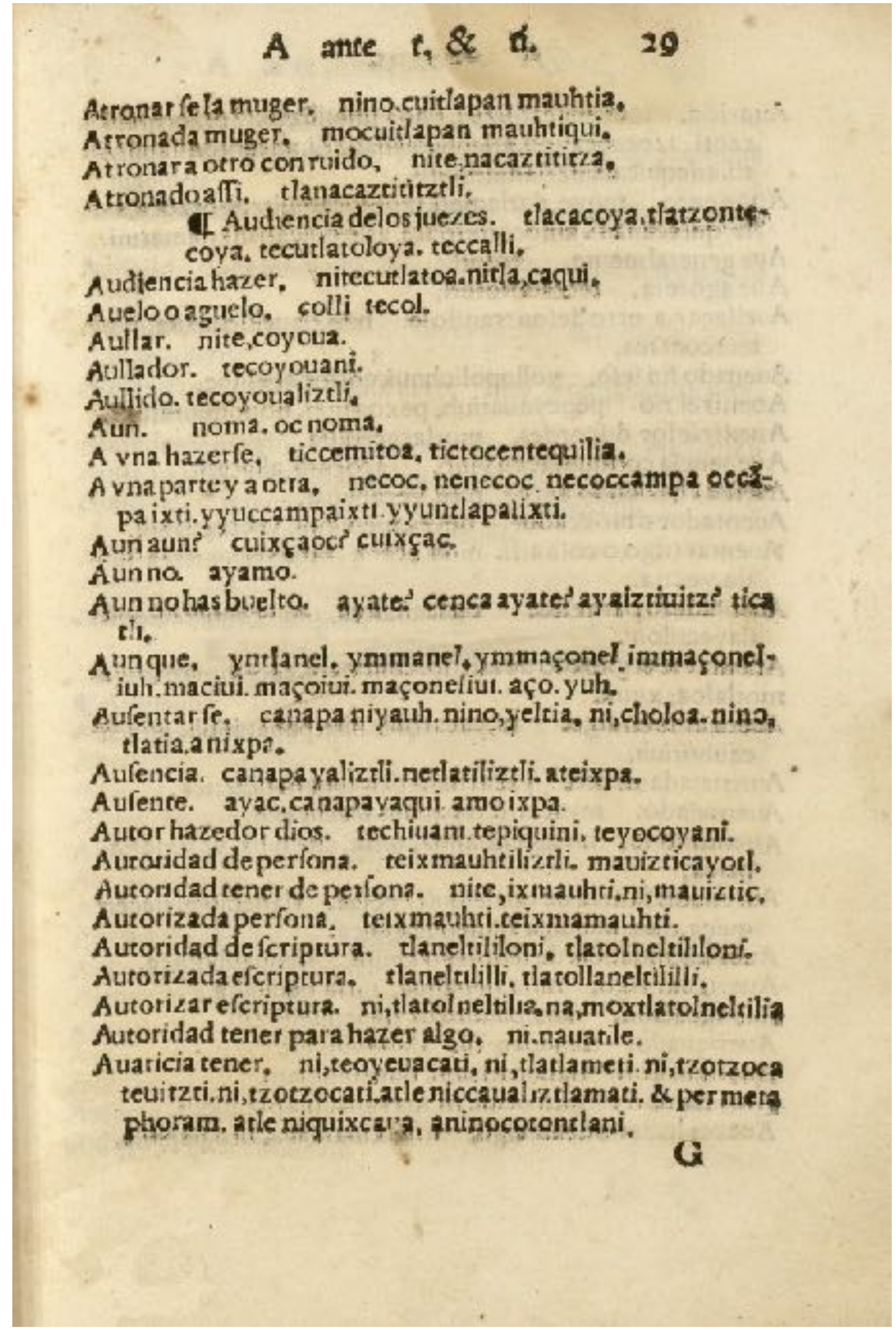
Evaluation systems must be context-specific

Evaluating Colonial Corpora

Aqui comiença vn vocabulario enla lengua Castellana y Mexicana, Compuesto por el muy reuerendo padre fray Alonso de Molina: Guardiã d̃l couẽto d̃ sant Antonio d̃ Tetzcuco d̃la ordẽ delos frayles Menores.

Signasti domine seruum tuũ Franciscum signis redemptionis nostre.

dorum nimia te fecit prole parentem. qui genuit moriens, quos pater alme foues. Confixus viuis, languens: cum mente reuoluis. vulnera, cum spectas, stigmata carne geris.



A ante t, & u. 29

Atronar se la muger. nino.cuitlapan mauhtia.
Atronada muger. mocuitlapan mauhtiqui.
Atronar a otro con ruido. nite.nacaztititza.
Atronado assi. tlanacaztititztli.
¶ Audiencia delos juezes. tlacacoya.tlatzontecoya. tecutlatoloya. teccalli.
Audiencia hazer. nitecutlatoa.nitla,caqui.
Auelo o aguelo. colli tecol.
Aullar. nite,coyoua.
Aullador. tecoyouani.
Aullido. tecoyoualiztli.
Aun. noma. oc noma.
A vna hazerse. ticcemitoa. tictocentequilia.
A vna parte y a otra. necoc. nenecoc necoccampa occãpa ixti.yyuccampaixti.yyuntlapalixti.
Aun aun? cuixçaoc? cuixçac.
Aun no. ayamo.
Aun no has buelto. ayate? cenca ayate? ayaiztiuitz? ticatli.
Aunque. yntlanel. ymmanel. ymmaçonel. immaçoneliuh. maciui. maçoiui. maçoneliui. aço. yuh.
Ausentar se. canapa niyauh. nino,yeltia. ni,choloa. nino,tlatia. anixpa.
Ausencia. canapa yaliztli. netlatiliztli. ateixpa.
Ausente. ayac, canapayaqui. amo ixpa.
Autor hazedor dios. techiuani.tepiquini. teyocoyani.
Autoridad de persona. teixmauhtiliztli. mauizticayotl.
Autoridad tener de persona. nite,ixmauhti.ni,mauiztic.
Autorizada persona. teixmauhti.teixmamauhti.
Autoridad de scriptura. tlaneltililoni. tlatolneltililoni.
Autorizada escriptura. tlaneltililli. tlatollaneltililli.
Autorizar escriptura. ni,tlatolneltilia.na,moxtlatolneltilia
Autoridad tener para hazer algo. ni.nauatile.
Auaricia tener. ni,teoyeuacati. ni,tlatlameti. ni,tzotzocateuitzti.ni,tzotzocati.atle niccaualiztlamati. & per metaphoram. atle niquixcaua. aninocotontlani.

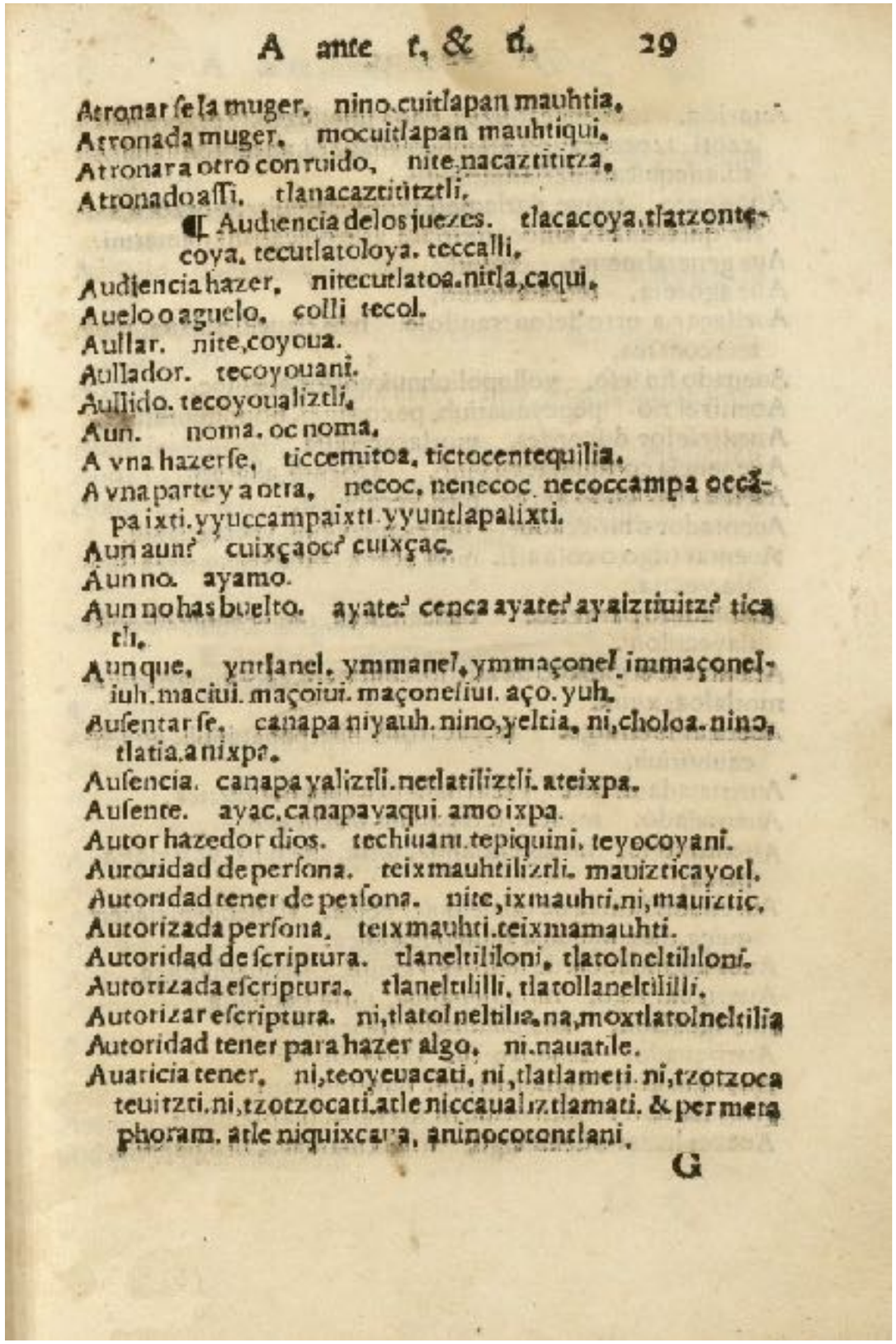
G

Hannah Alpert-Abrams

https://goo.gl/atVK2Q

Evaluating Colonial Corpora

?? ??
Arronarſe la muger, ni no, cuitlapan mauhtia.
Atró nada muger, mocuitlapan mauhtiqui.
Atronar á otro con ruido, ni te nacaztititza,
á tronado allí, tlanacaztitĩztli,
ã[Audiencia de los Juezes. tlacaco ya
coya. tecutlatoloya, 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 aunc̃cuix ça occc97xçac.
Aun no. ayamo.
Aun no ha abuelto, ayatec̃cen casayatec̃ayaíz
tli.
Aun que, yntla nel, ymmanel, ymmaçonel, ii
iuh maciui, maçoiui, maçoneltin, aço yuh.
Aufentat fe. canapa niyauh. nino-yeltia, ni-cl
tlatia. anĩxpa,
Aufencia. canapa yaliztli, netlatiliztli, atẽĩxpa,
Aufente. ayac. canapa yaqui amo ixpa.
Autor hazedor dios. techiuani. tepi quini, te yc
Autoridad de perſona, teixmauhtiliztli. mau
Autoridad tener de perſona, ni tesix mauhti, ni
Autorizada perſona, teixmauhti, teixmamauj
Autoridad de feriptura. tlaneltililoni, tlatolti
Autorizada eſcriptura, tlaneltililli, 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. anĩnocotontlani.



A ante t, & d. 29

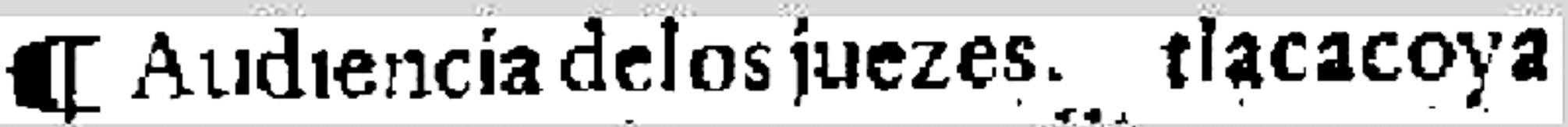
Atronarſe la muger, nino.cuitlapan mauhtia.
Atronada muger, mocuitlapan mauhtiqui.
Atronara otro con ruido, nite.nacaztititza.
Atronadoaſſi, tlanacaztititztli.
¶ Audiencia delos juezes. tlacacoya.tlatzonte-
coya. tecutlatoloya. teccalli.
Audiencia hazer, nitecutlatoa.nitla.caqui.
Auelo o aguelo, colli tecol.
Aullar. nite.coyoua.
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 aunc̃ cuixçaocc̃ cuixçac.
Aun no. ayamo.
Aun no has buelto. ayatec̃ cenca ayatec̃ ayaizticuitzc̃ tica
tli.
Aunque, yntlanel. ymmanel. ymmaçonel. immaçonel-
iuh.maciui. maçoiui. maçoneltiui. aço. yuh.
Auſentarſe. canapa niyauh. nino.yeltia. ni.choloa. nino.
tlatia.anixpa.
Auſencia. canapa yaliztli. netlatiliztli. ateixpa.
Auſente. ayac.canapayaqui amo ixpa.
Autor hazedor dios. techiuani.tepiquini. teyocoyani.
Autoridad de perſona. teixmauhtilizrli. mauizticayotl.
Autoridad tener de perſona. nite.ixmauhti.ni.mauiztic.
Autorizada perſona. teixmauhti.teixmamauhti.
Autoridad de ſcriptura. tlaneltililoni. tlatolneltililoni.
Autorizada eſcriptura. tlaneltililli. tlatollaneltililli.
Autorizar eſcriptura. ni.tlatolneltilia.na.moxtlatolneltilia
Autoridad tener para hazer algo. ni.nauatile.
Auaricia tener. ni.teoyeuacati. ni.tlatlameti. ni.tzotzoca
teuitzti.ni.tzotzocati.atle niccaualiztlamati. & per me-
phoram. atle niquixcatia. anĩnocotontlani.
G

Evaluating Colonial Corpora

Atró nada muger, mocuitlapan mauhtiqui.

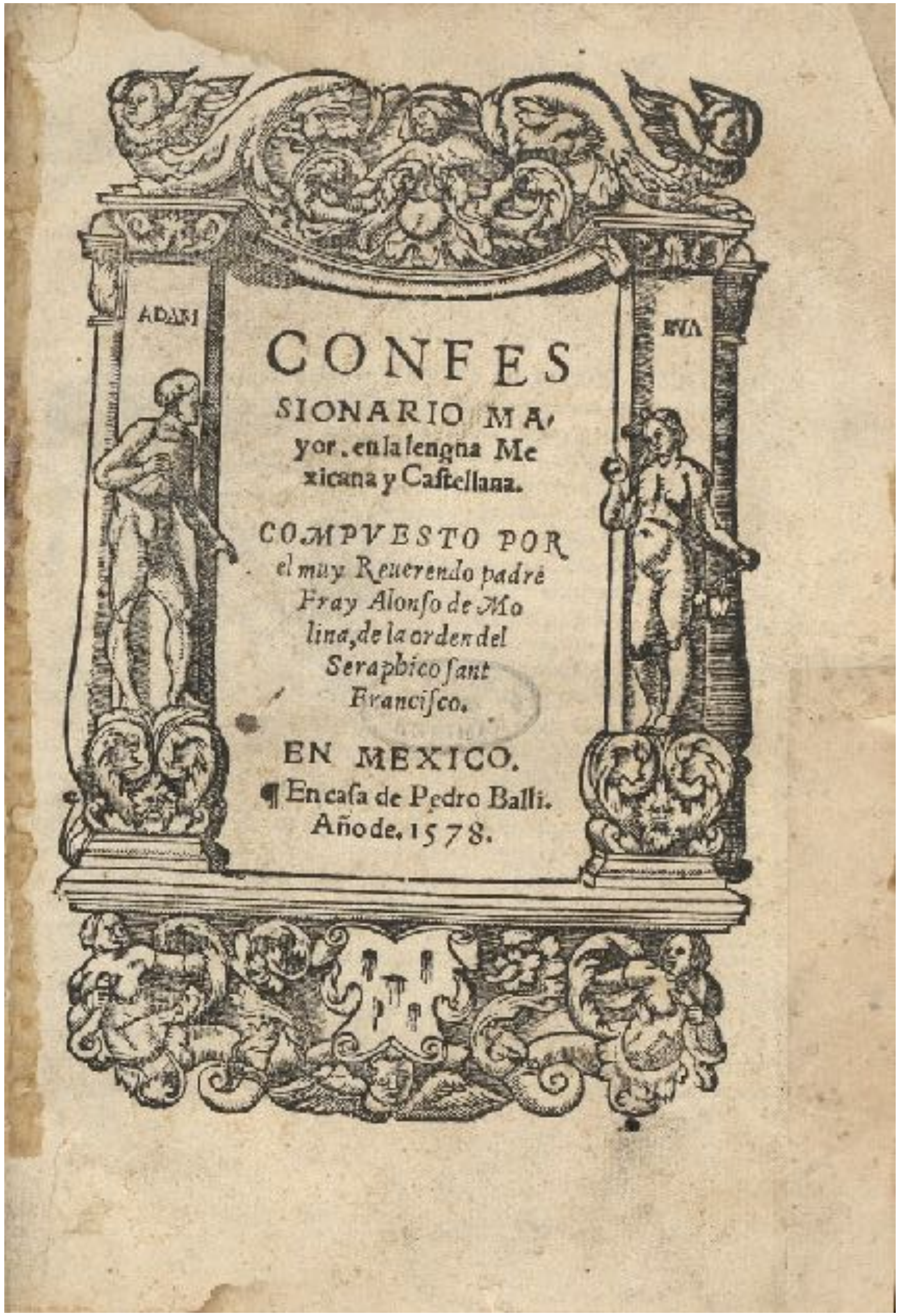
Atronar á otro con ruido, ni te nacaztititza,

á tronado allí, tlanacaztitĩtztli,



ã[Audiencia de los Juezes. tlacaco ya

Evaluating Colonial Corpora



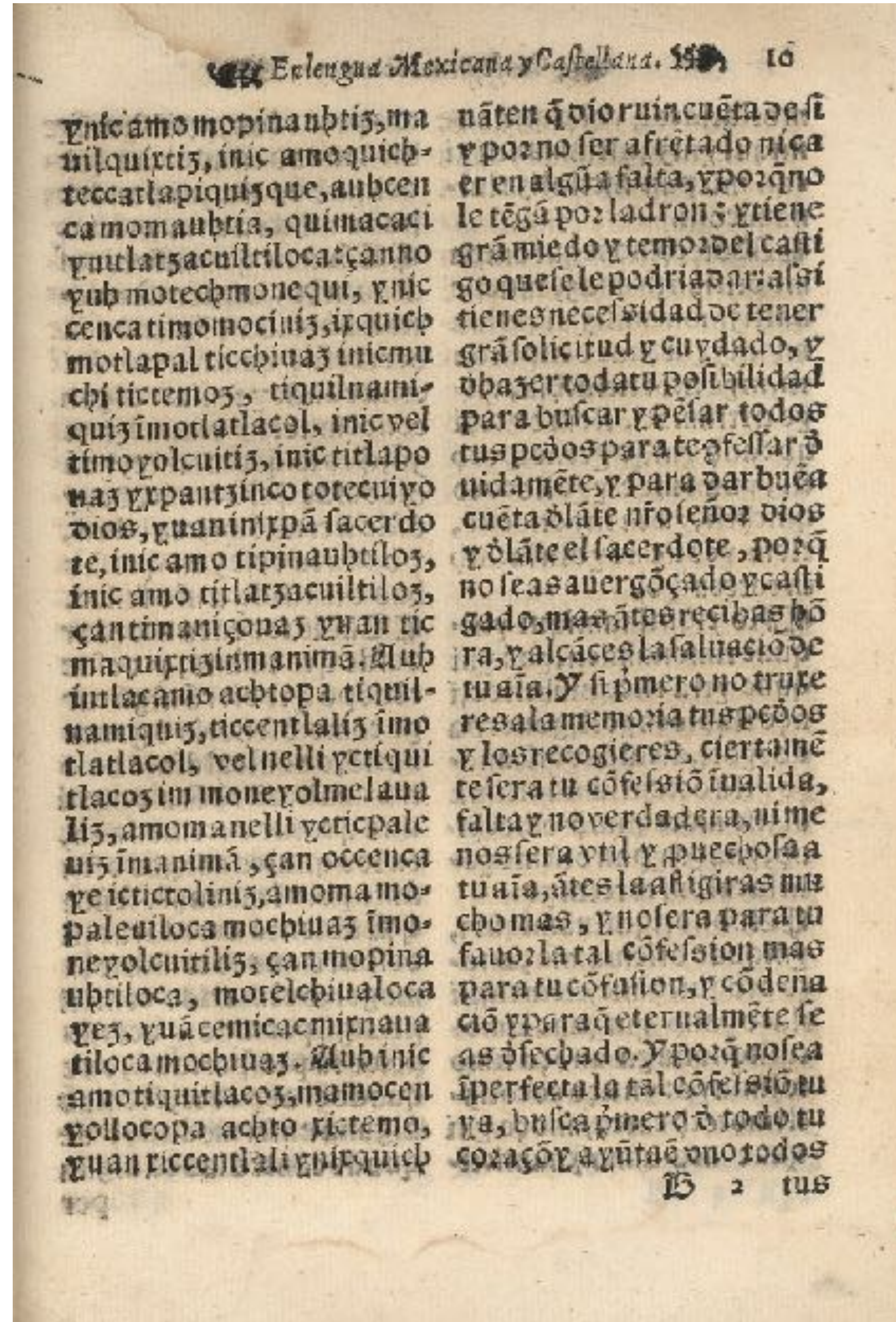
ADAM

EVA

CONFES
SIONARIO MA-
yor, en la lengua Me
xicana y Caſtellana.

COMPVESTO POR
el muy Reuerendo padre
Fray Alonſo de Mo
lina, de la orden del
Seraphico ſant
Franciſco.

EN MEXICO.
¶ En caſa de Pedro Balli.
Año de. 1578.



ynic amo mopinauhtiz, ma
nilquixtiz, inic amo quich-
teccatlapiquizque, auhcen
ca momauhtia, quimacaci
ynictlatzacuiltilocatçanno
yuh motechmonequi, ynic
cenca timomociuiz, ixquich
motlapal ticchiuaz inicmu
chi tictemoz, tiquilnami-
quiz imotlatlacol, inic vel
timoyolcuitiz, inic titlapo
uaz yxpantzinco toteuiyo
dios, yuaninixpã ſacerdo
te, inic amo tipinauhtiloz,
inic amo titlatzacuiltiloz,
çan timaniçouaz yuan tic
maquixtiz im manimã. Auh
intlacamo achtopa tiquil-
namiquiz, ticcentlaliz imo
tlatlacol, vel nelli yctiqui
tlacoz im moneyolmelaua
liz, amo ma nelli yeticpale
uiz imanimã, çan occenca
ye ictictoliniz, amo ma mo-
paleuiloca mochiuaz imo-
neyolcuitiliz, çan mopina
uhtiloca, motelchiualoca
yez, yuã cemicacmixnaua
tiloca mochiuaz. Auh inic
amo tiquitlacoz, mamocen
yollocopa achto xictemo,
yuan xiccentlali ynixquich

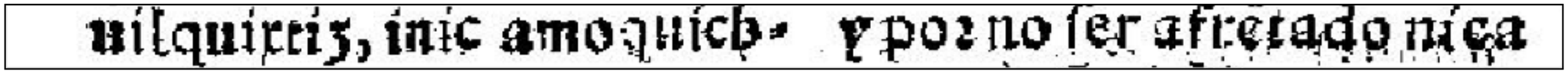
nãten q̃ dio ruin cuẽta de ſi
y por no ſer afrẽtado nĩ ca
er en algũa falta, y porq̃ no
le tẽgã por ladron; y tiene
grã miedo y temor del caſti
go que ſe le podria dar: aſsi
tienes neceſsidad de tener
grã ſolicitud y cuydado, y
d hazer toda tu poſibilidad
para buſcar y pẽſar todos
tus pcdos para te cõfeſſar d
uidamẽte, y para dar buẽa
cuẽta dlãte nr̃o ſeñor dios
y dlãte el ſacerdote, porq̃
no ſeas auergõçado y caſti
gado, mas ãtes recibas hõ
ra, y alcãces la ſaluaciõ de
tu aĩa. Y ſi pmero no truxe
res a la memoria tus pcdos
y los recogieres, ciertamẽ
te ſera tu cõfeſsiõ inualida,
falta y no verdadera, ni me
nos ſera util y puechoſa a
tu aĩa, ãtes la afligiras mu
cho mas, y no ſera para tu
fauor la tal cõfeſsion mas
para tu cõfuſion, y cõdena
ciõ y para q̃ eternalmẽte ſe
as deſechado. Y porq̃ no ſea
ĩperfecta la tal cõfeſsiõ tu
ya, buſca pmero d todo tu
coraçõ y ayũtaẽ vno todos

Hannah Alpert-Abrams

https://goo.gl/atVK2Q

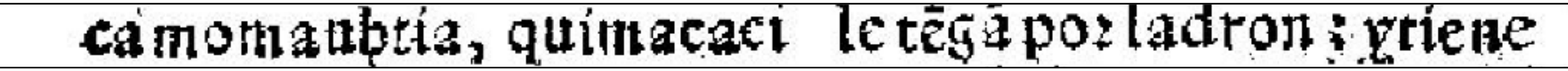
Evaluating Colonial Corpora

ynic amo mopĩnãuhtij, ma nãten. q̃Dior uincuẽtaoe-lĩ



uilquirtij, inic amo quicb - y poz no fer atrétado nica

teccatlapiquijque, auh cen eren algua falta, y poz q̃n.o



ca momaubtia, quimacaci le tẽgã poz ladron ; ytiene

Evaluation systems must be context-specific

We need standards for evaluating our colonial corpora.