

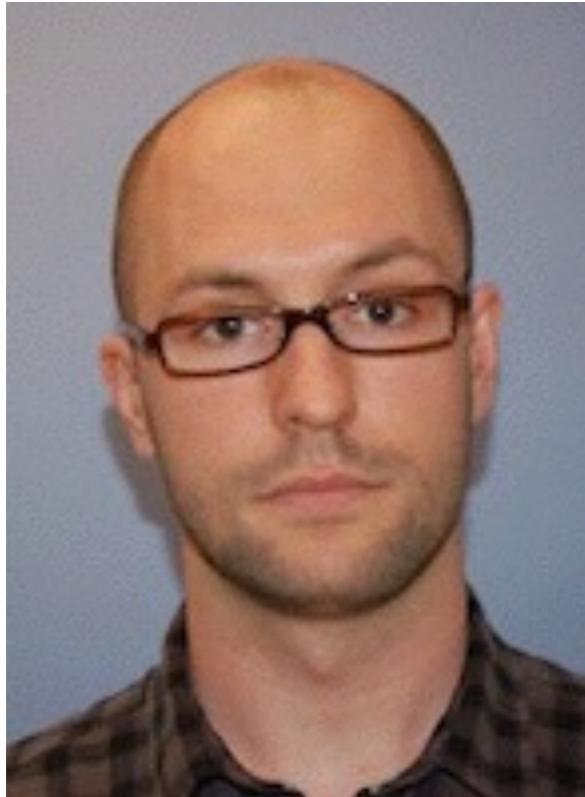
Unsupervised Modeling of Code-Switching and Orthographic Variation for Historical Document Transcription

Dan Garrette

UW CSE

<http://dhg.ai>

Collaborators



Taylor Berg-Kirkpatrick

Computer Science
UC-Berkeley

Dan Klein

Hannah Alpert-Abrams

Comparative Literature
UT-Austin

Document Transcription

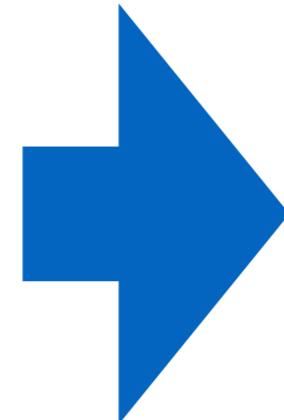
Advertencias para

como parece manifiesto en las palabras de S.
Ioan, que dize. Tres sunt qui testimonium dant
in cœlo Pater, Verbum, & Spiritus sanctus :
& hi tres vnum sunt. i. Ioann. ultimo. Por
loqual deuen ser instruydos y enseñados, que
todas tres diuinæ personas son vn Dios ver-
dadero: o reformando la sobre dicha proposi-
cion, y añadiendo esta palabra. In huelime
ixtintzitzin, con que se quita toda amphibio-
logia y duda diciendo. In Deus, ca Tettatzin
Tepiltzin, Spiritu sancto, ei personas, çan ce
huelnelli teutl Dios in huelimeixtintzitzin,
q. d. Deus es Padre, Hijo, y Spū sancto tres
personas, vn solo Deus verdadero todas tres,
cō la qual reduplicacion se quita toda dubda.
Tambien se quita con estas propoſiciones.
In D I O S, ca Tettatzin, Tepiltzin, Spiritus san-
cto, çan huel iceltzin teutl Deus tlahtohuani.
In Deus, ca Tettatzin, Tepiltzin, Spiritus san-
cto, imēixtin personas çan huel iceltzin Dio-
tlahtohuani. Ca inimēixtin personas me ca-
çan huel iceltzin teutl Deus tlahtohuani in-
huel imēixtin. ¶ Otros responden [y es el
segundo error] çace Deus tlahtohuani, imē
teihttotica, y a algios de sus ministros les ha
parecido el meteihhtotica, vn vocablo en si di-

Document Transcription

Aduertencias para

como parece manifiesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniu dāt in cœlo Pater, Verbum, & Spiritus sanctus: & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinias personas son vn Dios verdadero: o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzitzin, con que se quita toda amphibologia y duda diciendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzitzin, q. d. Dios es Padre, Hijo, y Spu sancto tres personas, vn solo Dios verdadero todas tres, cō la qual reduplicacion se quita toda dubda. Tambien se quita con estas propoſiciones. In DIOS, ca Tettatzin, Tepiltzin, Spiritus sancto, çan huel iceltzin teutl Dios tlahtohuani. In Dios, ca Tettatzin, Tepiltzin, Spiritus sancto, imeixtin personas çan huel iceltzin Dios tlahtohuani. Ca inimeixtin personas me caçan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] çace Dios tlahtohuani, imme teihttotica, y a algios de sus ministros les ha parecido el meteihhtotica, vn vocablo en si di-



Aduertencias para

como parece manifiesto en las palabras de S. Ioan, que dize. Tres sunt qui teſtimoniu dāt in cœlo Pater, Verbum, & Spiritus sanctus: & hi tres vnum sunt. 1. Ioann. vltimo. Por lo qual deuen ser inftruydos y enfeñados, que todas tres diuinias personas son vn Dios verdadero: o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzitzin, con que se quita toda amphibologia y duda diciendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli tuetl Dios, in huel imeixtintzitzin, q. d. Dios es Padre, Hijo, y Spu sancto tres personas, vn solo Dios verdadero todas tres, cō la qual reduplicacion se quita toda dubda. Tambien se quita con eftas propoſiciones. In DIOS, ca Tettatzin, Tepiltzin, Spiritu sancto, çan huel iceltzin teutl Dios tlahtohuani. In Dios, ca Tettatzin, Tepiltzin, Spiritu sancto, imeixtin personas çan huel iceltzin Dios tlahtohuani. Ca inimeixtin personas me caçan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] çace Dios tlahtohuani, imme teihttotica, y a alguos de sus ministros les ha parecido el meteihhtotica, vn vocablo en si di uino

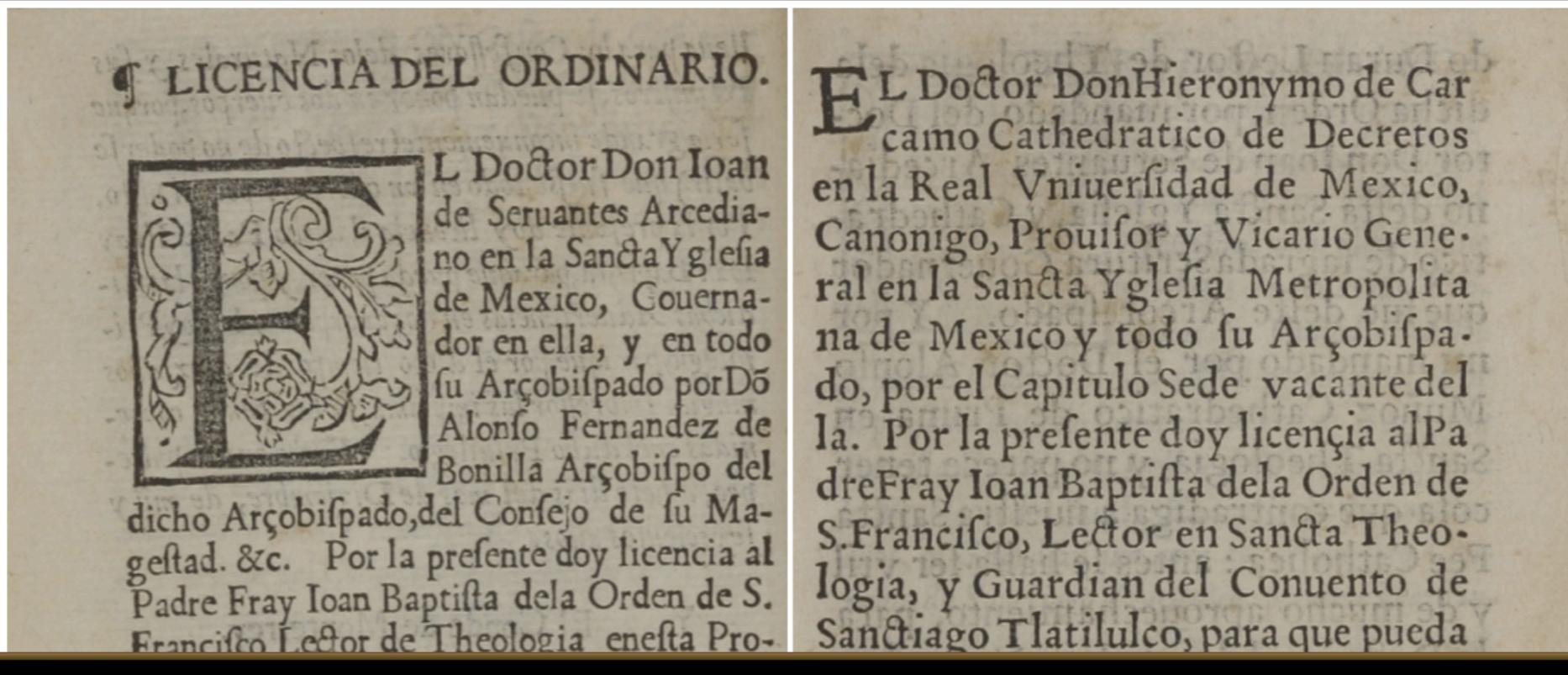
Linguistic Noise

Code switching (multilinguality)

Orthographic variation (cross-linguality)

Primeros Libros

Books printed in Mexico in the 1500s

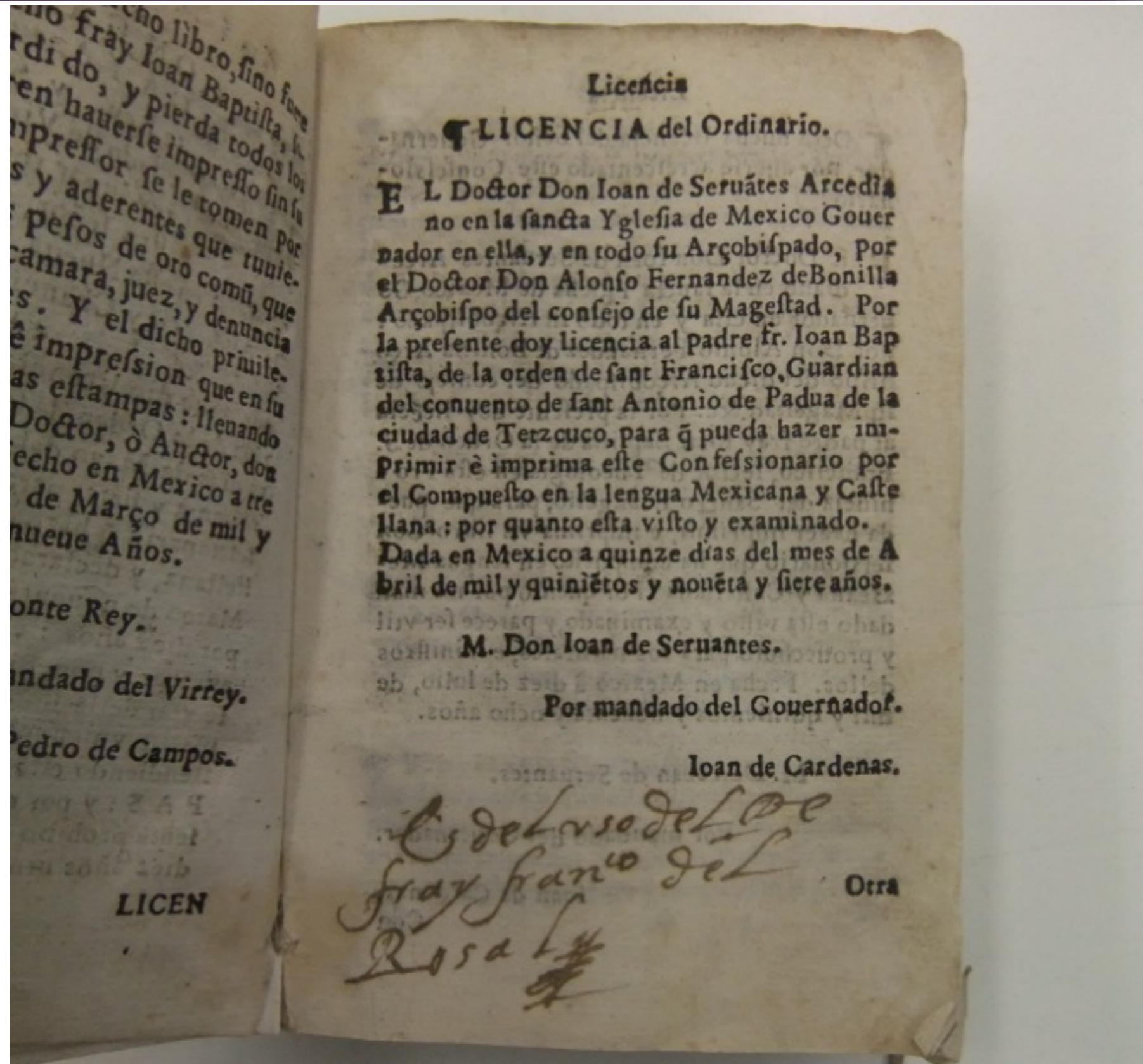


This Talk

1. What makes the problem hard
2. How we model the problem
3. Implications for the humanities

Transcription Challenges

Historical Document Transcription



Visual Noise

Too little or too much ink

roule along in silence

Irregular alignment

the Death of the Decimated,

Unknown Fonts



g g g g g g g

Unknown Fonts

Gante	(1553)	motlacatilía:ýnica sacramento Baptis-
Anunciación	(1565)	¶ Tlön non el toqui tíá yndíos
Sahagún	(1583)	Yoan oquilhui in Emperador, intlaca,
Rincón	(1595)	etion.v.g.tetlaçotlaliztli.amatio, vel,
Bautista	(1600)	Rimog bæc supra dictus docto Medina.

Unknown Fonts

These are all ‘a’s from one page of one book:

a a a a a a a a a a

Starting point: *Ocular*

Taylor Berg-Kirkpatrick, Greg Durrett, and Dan Klein
ACL 2013

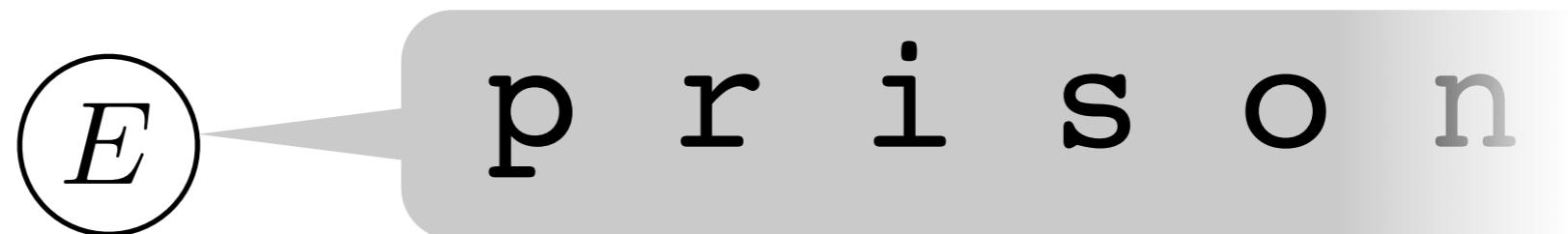
Starting Point: *Ocular*

Generative Model in 3 parts:

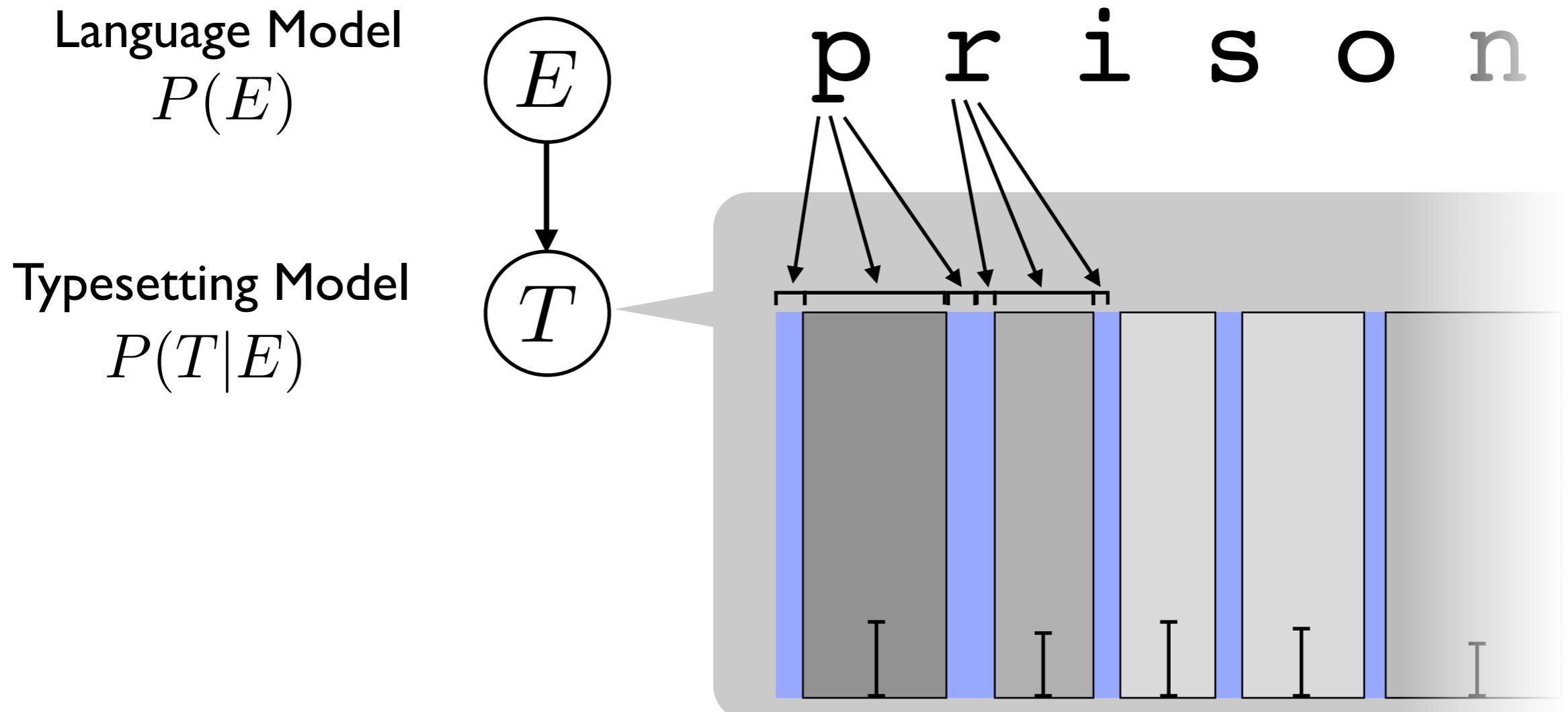
1. Language model
2. Typesetting model
3. Rendering model

Ocular's Generative Model

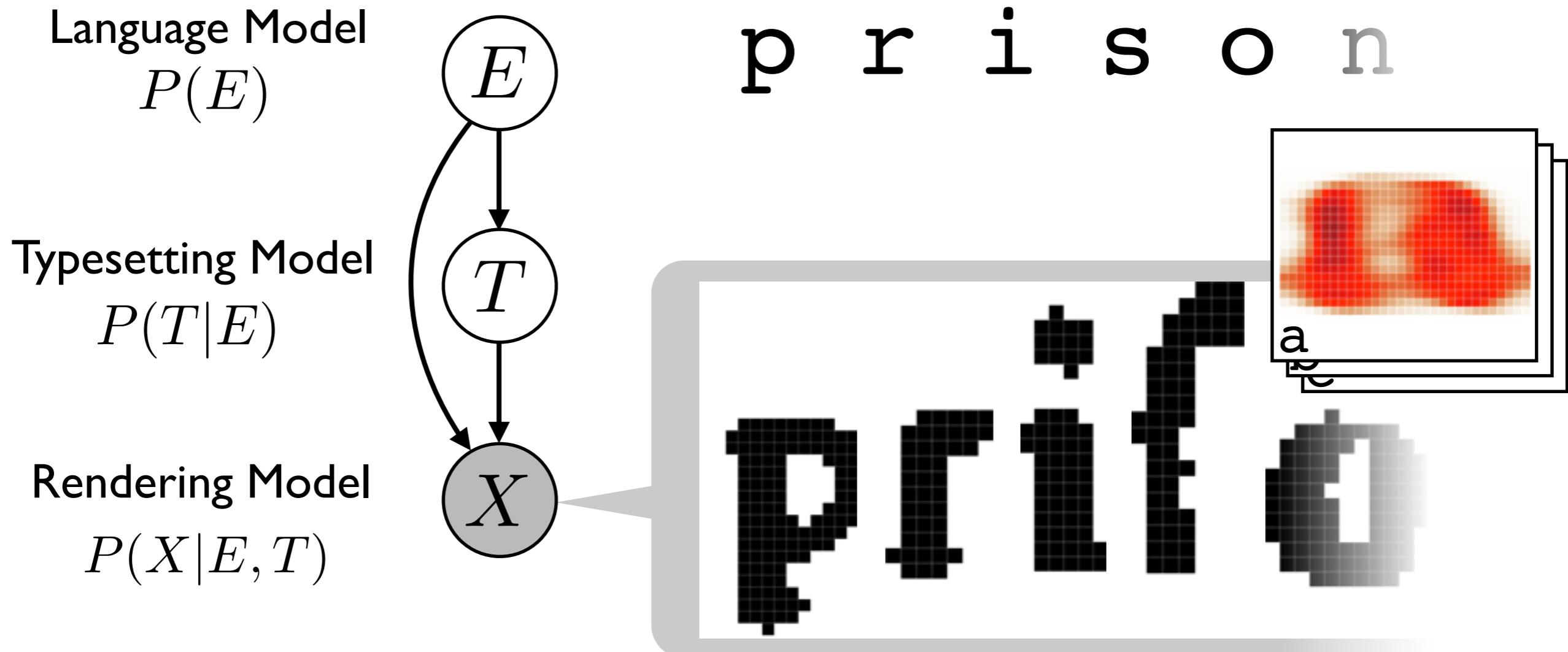
Language Model
 $P(E)$



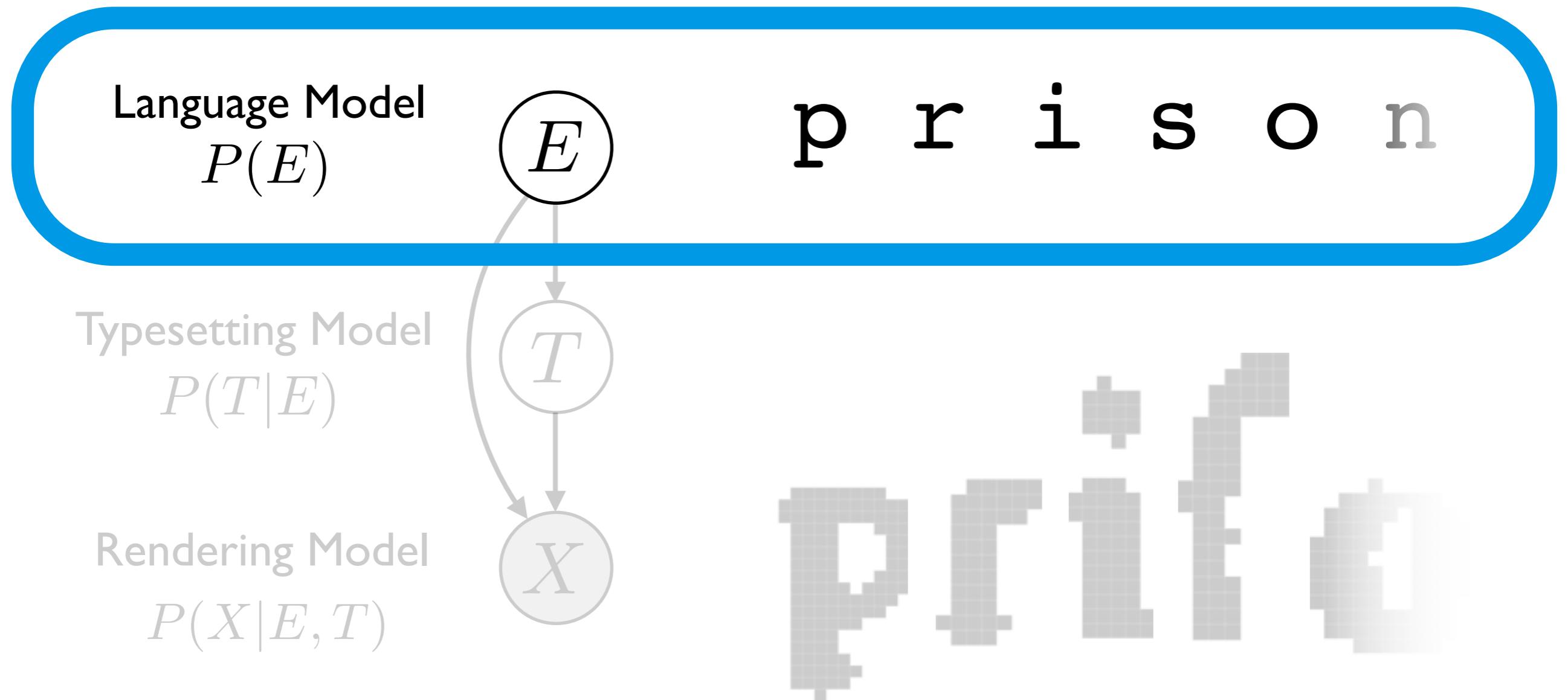
Ocular's Generative Model



Ocular's Generative Model



Our Focus



Our Focus

1. Multilingual code-switching
2. Orthographic variation

Multilingual Texts

Multilingual Texts

Many historical documents are written in, and switch readily between, multiple languages.

Aduertencias para

como parece manifiesto en las palabras de S.
Ioan, que dize. *Tres sunt qui testimoniū dāt
in cœlo Pater, Verbum, & Spiritus sanctus :*
& hi tres vnum sunt. i. Ioann. ultimo. Por
loqual deuen ser instruydos y enseñados, que
todas tres diuinias personas son vn Dios ver-
dadero; o reformando la sobre dicha proposi-
cion, y añadiendo esta palabra. *In huelime
ixtintzitzin,* con que se quita toda amphibo-
logia y duda diciendo. *In Dios, ca Tettatzin
Tepiltzin, Spiritu sancto, ei personas, çan ce
huelnelli teutl Dios in huelime ixtintzitzin,*
*q. d. Dios es Padre, Hijo, y Spū sancto tres
personas, vn solo Dios verdadero todas tres,
cō la qual reduplicacion se quita toda dubda.*
Tambien se quita con estas propoſiciones.
*In D I O S, ca Tettatzin, Tepiltzin, Spiritus san-
cto, çan huel iceltzin teutl Dios tlahtohuani.*
*In Dios, ca Tettatzin, Tepiltzin, Spiritus san-
cto, imēixtin personas çan huel iceltzin Dio-
s tlahtohuani.* Ca inimēixtin personas me ca-
çan huel iceltzin teutl Dios tlahtohuani in-
huel imēixtin. ¶ Otros responden [y es el
segundo error] çace Dios tlahtohuani, immē-
teihttotica, y a algños de sus ministros les ha
parecido el meteihttotica, vn vocablo en si di-

Spanish

Latin

Nahuatl

Aduertencias para
como parece manifiesto en las palabras de S.
Ioan, que dice.

[REDACTED] Por
loqual deuen ser instruydos y enseñados, que
todas tres diuinias personas son vn Dios ver-
dadero; o reformando la sobre dicha proposi-
cion, y añadiendo esta palabra. [REDACTED]
[REDACTED] con que se quita toda amphibio-
logia y duda diciendo. In Dios, [REDACTED]
[REDACTED] personas,

Dios
q. d. Dios es Padre, Hijo, y [REDACTED] tres
personas, vn solo Dios verdadero todas tres,
cō la qual reduplicacion se quita toda dubda.
Tambien se quita con estas propoſiciones.

DIO S, [REDACTED] Dios
Dios, [REDACTED] personas [REDACTED] Dioſ
[REDACTED] personasme [REDACTED]
Dios [REDACTED]
Otros responden [y es el
segundo error] [REDACTED] Dios [REDACTED]
[REDACTED] y a alḡos de sus ministros les ha
parecido el [REDACTED] yn vocablo en si di-
uiño

Spanish

Spiritu sancto, ei

Spū sancto

Latin

Nahuatl

āto,

āto,

Spiritusan

Spiritusan.

qui

Tres sunt qui testimoniam dāt
in cœlo Pater, Verbum, & Spiritus sanctus :
& hi tres unum sunt. i. Ioann. vltimo.

Spanish

Latin

Nahuatl

In huelime
ixtintzitzin,
ca Tettazin
Tepiltzin,
huelnelli teutl
çan ce
in huel imieixtintzitzin.

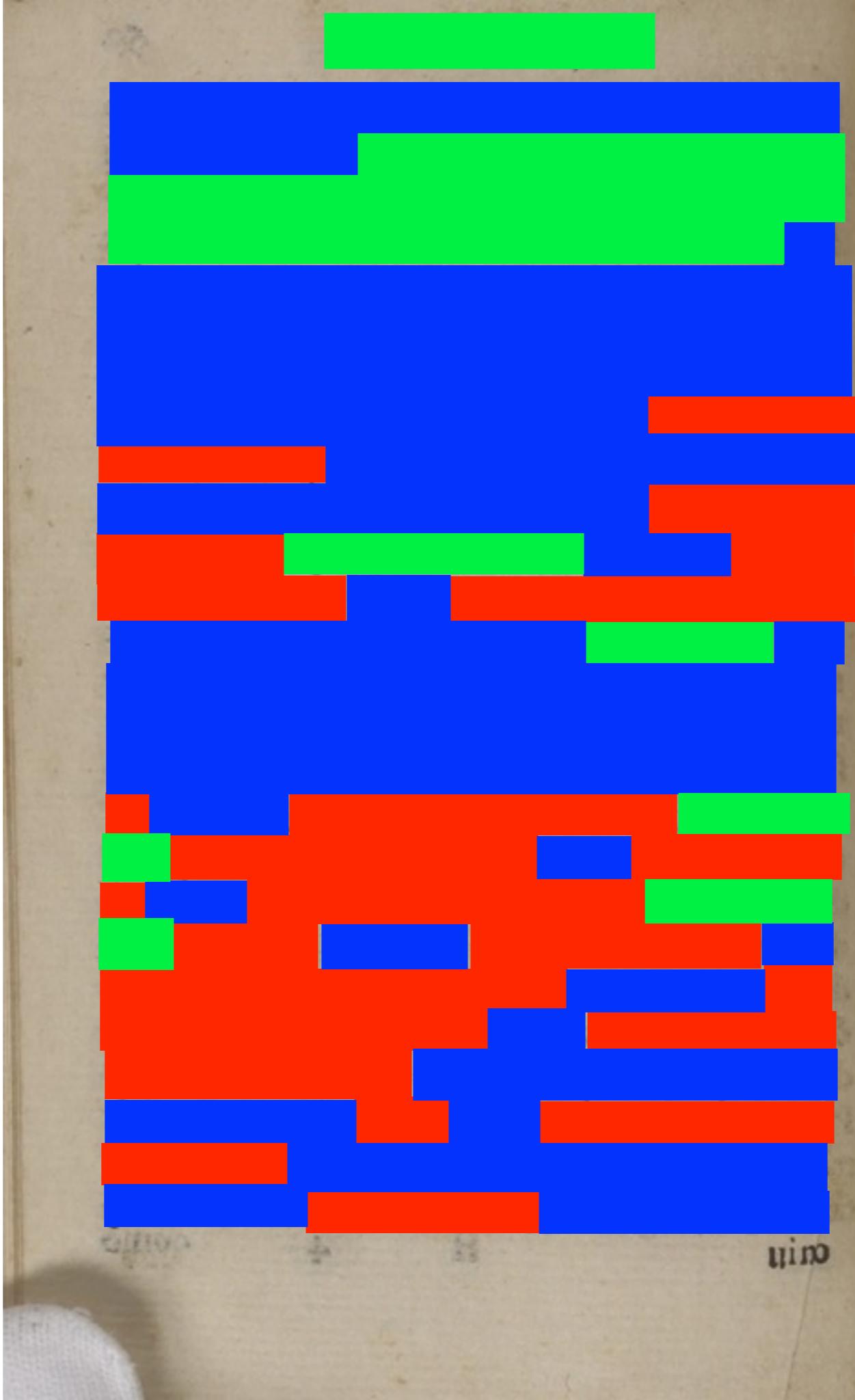
In ca Tettazin, Tepiltzin,
çan huel iceltzin teutl tlahitohuani.
In ca Tettazin, Tepiltzin,
imeixtin çan huel iceltzin
tlahitohuani. Ca inimeixtin ca-
çan huel iceltzin teutl tlahitohuani in-
huel imieixtin. ¶ çace tlahitohuani, immie
teihttotica,
meteihttotica,

qui

Spanish

Latin

Nahuatl

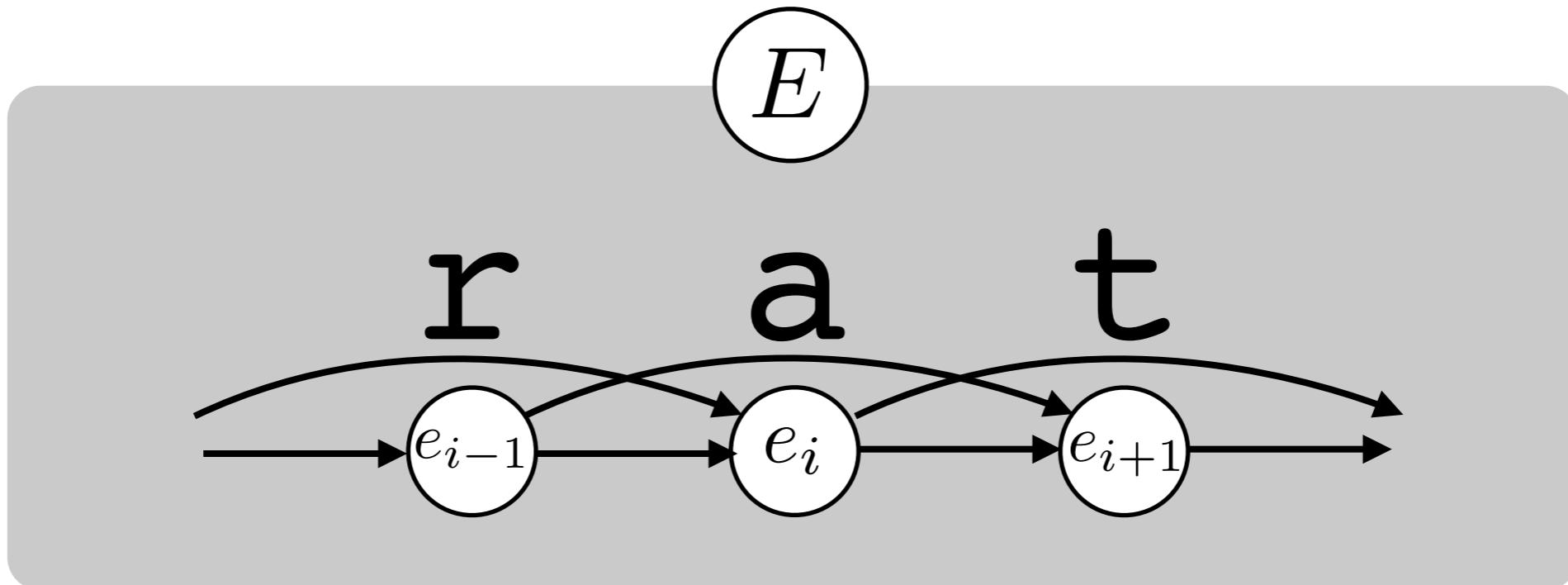


Why Code Switching?

ficion, y añadiendo esta palabra. **In huelime**
ixtintzitzin, con que se quita toda amphibología y duda diciendo. In Dios, ca Tettatzin

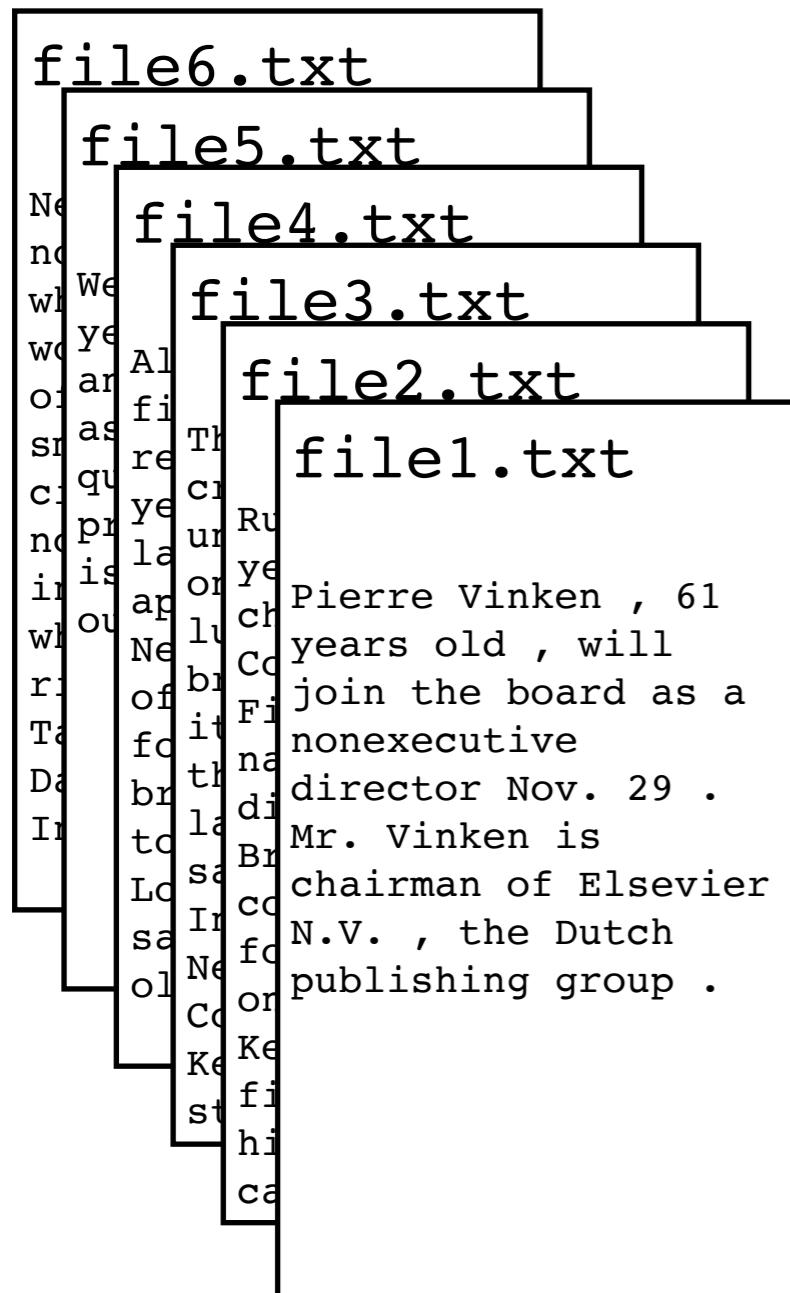
...and adding these words. [Nahuatl phrase],
with which all ambiguity and doubt is removed...

Ocular's Language Model

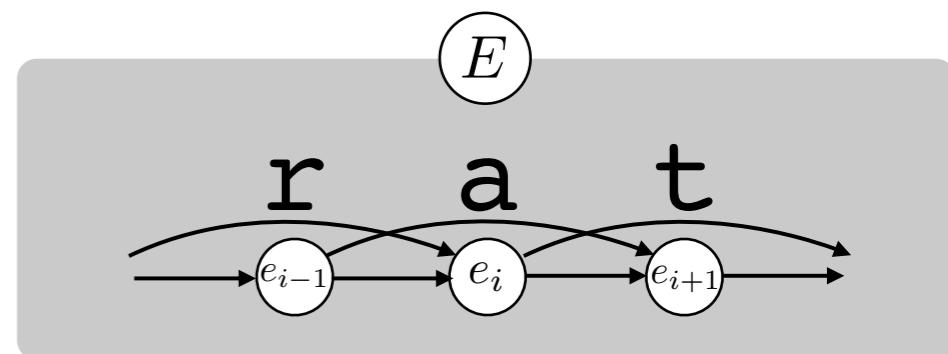


Kneser-Ney smoothed **character** 6-gram

Ocular's Language Model



count n-grams



Ocular's Language Model

Ocular was originally evaluated on court documents printed in England in the 1700-1800s.

Ocular's Language Model

English

An orchic unred Sacrameters **were else** momaquiliteo **are were** climrocaruisili-teorage Intergentilamandis, **tie board isn't** quate quilted it in Raptisme's snig Vadam and **swig broad** in Confirmacion's inter girlamandis **wife broad**. Trinidagon arcators internault dramandi, is board in **petrol** mess **oasis** di-

inijcxitzi;inic chicoppa,iquac in quimixililique yiomotlantzinco.

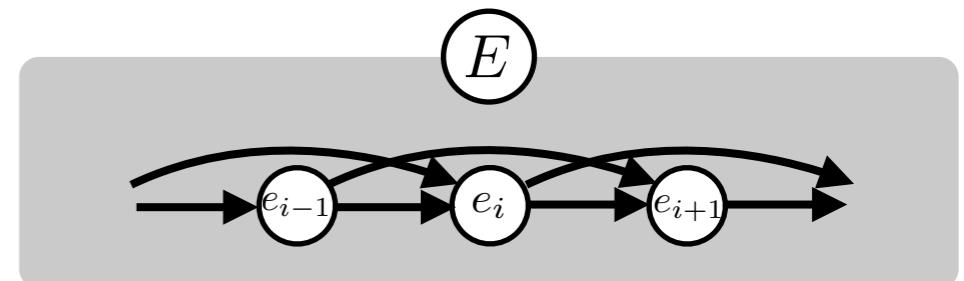
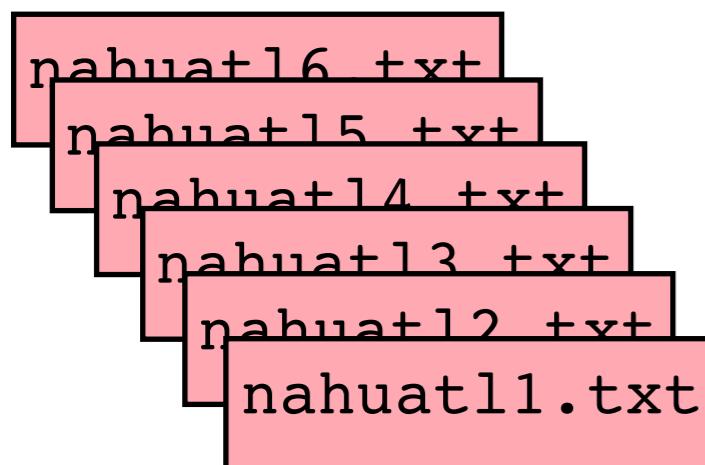
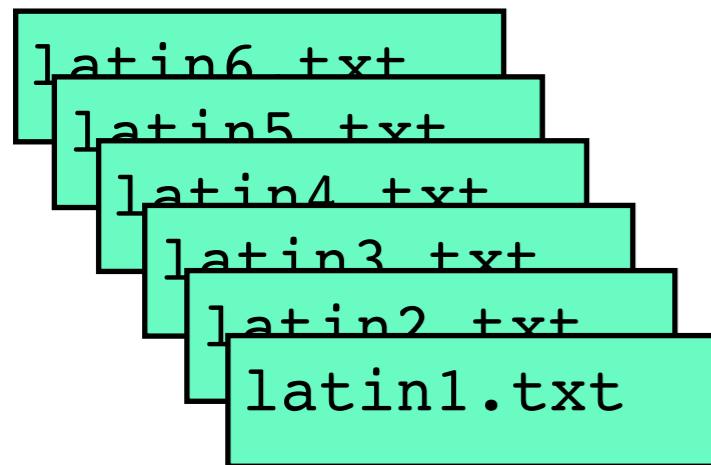
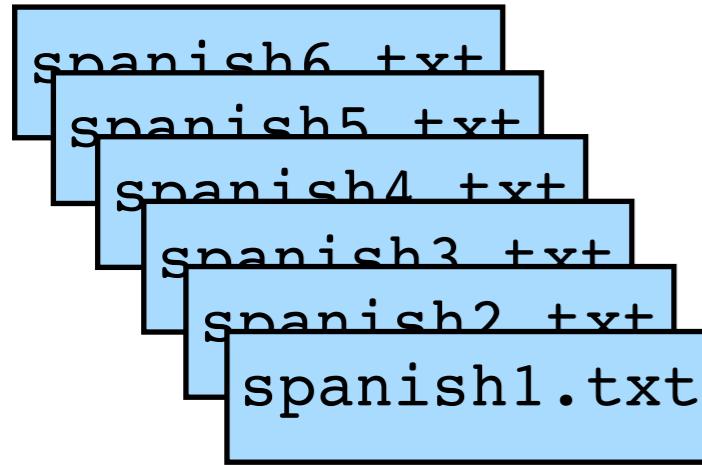
Q V A R T O

Psalmo.

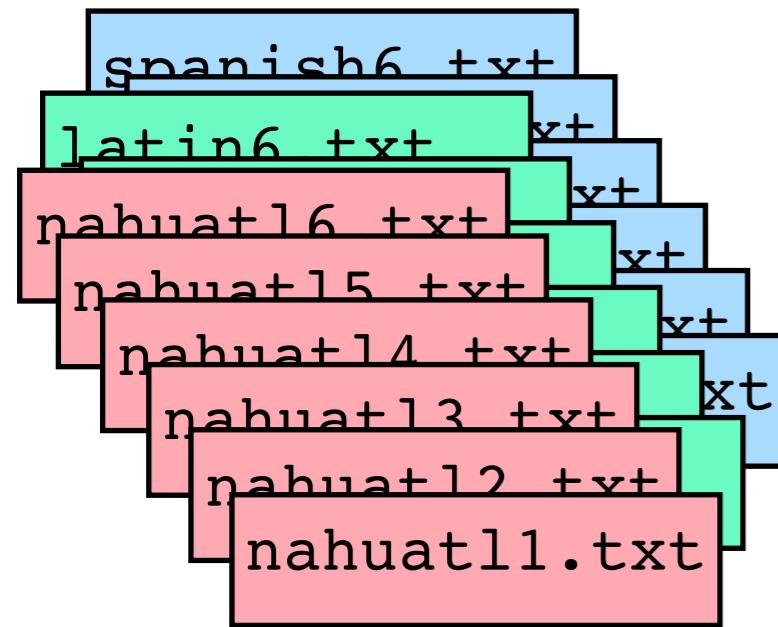
C Ano chicuntetl Sacramētos, otechmomaquiliteoac, otechmocauilileteoac: Inic centlaimantli, iechoatl in nequatequiliztli, in Baptismo : inic vntlamantli, iechoatl in Confirmation : inic etlamantli, iechoatl initlaçonacaiotzi: inic nauhtlamantli, iechoatl in neiolme-laoaliztli.

Inic macuillamantli, iechoatl in Extrema vñction: inic chiquacentel, iechoatl in teupixcautl, initoca Orden sacerdotal: inic chicuntetl, iechoatl in nenamictiliztli, itoca Matrimonio.

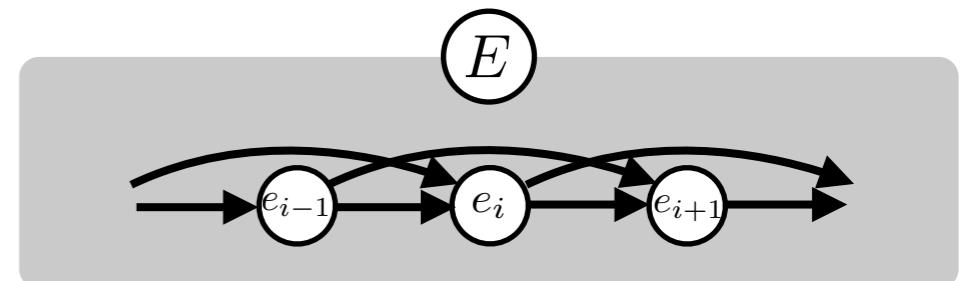
Baseline Multilingual Model



Baseline Multilingual Model



count n-grams



Baseline Multilingual Model

Multilingual Blur

A nō chicunt **et** l. Sacramētos, ó **trecho** momaquili texoac, otechmocavilili - texoac. Inic centlamantli, iehoatl in **ne** - quatequiliztli in Baptismo . inic vntlamantli, iehoatl in Confirmación . imic etlamantli, iehoatl, in itlaçònacaiotzi, inic nauhtlamantli, iehoatl in **netolme** laoaliztli.

inijcxitzi;inic chicoppa,iquac in quimi- xililique yiомotlantzinco.

Q V A R T O

Psalmo.

CAno chicuntetl Sacramētos, otechmomaquiliteoac, otechmocauiliteoac: Inic centlamantli, iehoatl in nequatequiliztli, in Baptismo : inic vntlamantli, iehoatl in Confirmacion : inic etlamantli, iehoatl initlaçonacaiotzi: inic nauhtlamantli, iehoatl in neiolme laoaliztli.

Inic macuillamantli, iehoatl in Extrema vnction: inic chiquacentetl , iehoatl in teupixcaiutl, initoca Orden sacerdotal: inic chicuntetl, iehoatl in nenamictiliztli, itoca Matrimonio.

Code-Switching Language Model

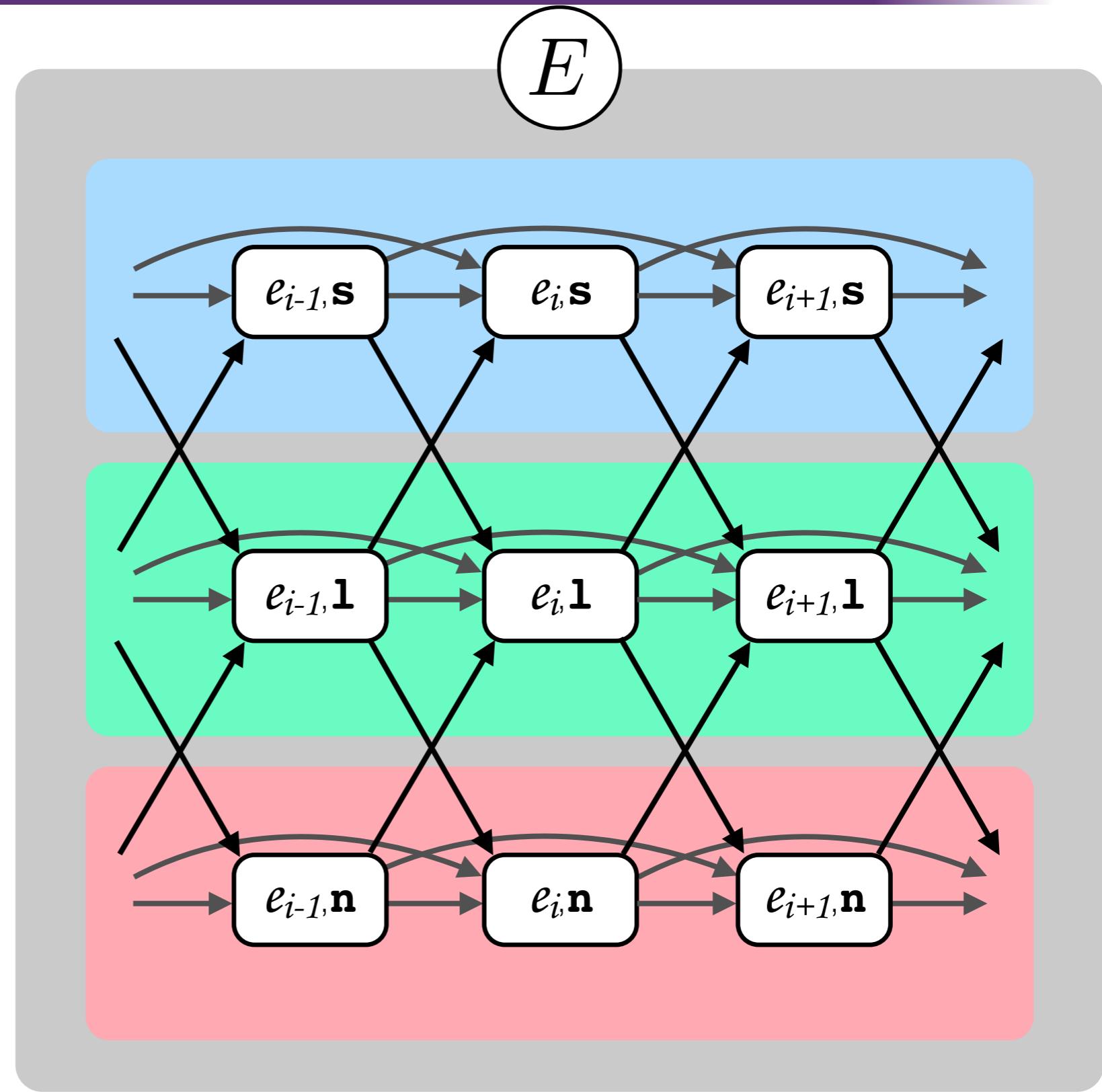
spanish6 +xt
spanish5 +xt
spanish4 +xt
spanish3 +xt
spanish2 +xt
spanish1.txt



latin6 +xt
latin5 +xt
latin4 +xt
latin3 +xt
latin2 +xt
latin1.txt



nahuatl16 +xt
nahuatl15 +xt
nahuatl14 +xt
nahuatl13 +xt
nahuatl12 +xt
nahuatl11.txt



Aduertencias para

como parece manifiesto en las palabras de S.
Ioan, que dize. *Tres sunt qui testimoniū dāt
in cœlo Pater, Verbum, & Spiritus sanctus :*
& hi tres vnum sunt. i. Ioann. ultimo. Por
loqual deuen ser instruydos y enseñados, que
todas tres diuinias personas son vn Dios ver-
dadero; o reformando la sobre dicha proposi-
cion, y añadiendo esta palabra. *In huelime
ixtintzitzin,* con que se quita toda amphibo-
logia y duda diciendo. *In Dios, ca Tettatzin
Tepiltzin, Spiritu sancto, ei personas, çan ce
huelnelli teutl Dios in huelime ixtintzitzin,*
*q. d. Dios es Padre, Hijo, y Spū sancto tres
personas, vn solo Dios verdadero todas tres,
cō la qual reduplicacion se quita toda dubda.*
Tambien se quita con estas propoſiciones.
*In D I O S, ca Tettatzin, Tepiltzin, Spiritus san-
cto, çan huel iceltzin teutl Dios tlahtohuani.*
*In Dios, ca Tettatzin, Tepiltzin, Spiritus san-
cto, imēixtin personas çan huel iceltzin Dio-
s tlahtohuani.* Ca inimēixtin personas me ca-
çan huel iceltzin teutl Dios tlahtohuani in-
huel imēixtin. ¶ Otros responden [y es el
segundo error] çace Dios tlahtohuani, immē-
teihttotica, y a algños de sus ministros les ha
parecido el meteihttotica, vn vocablo en si di-

íxtintzitzin, con que

Spanish

AÁBCDÉFGHIÍJKLMÑOÓPQRSTUÚVWXYZ
aábcdéfghiíjklmññoópqrstuúvwxyz
01234567890., /\()?"':;-

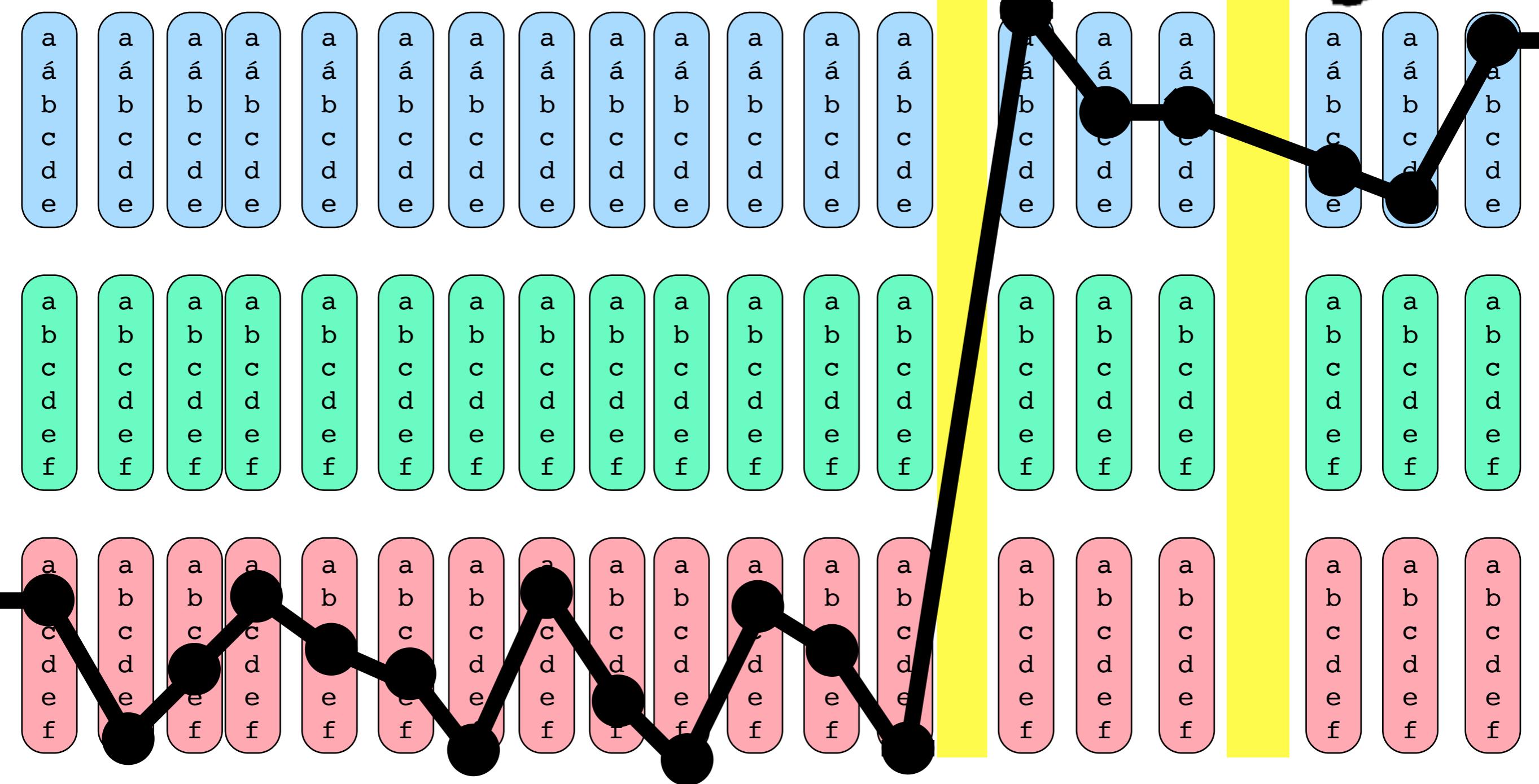
Latin

ABCDEFGHIJKLMOPQRSTUVWXYZ
abcdefghijklmopqrstuvwxyz
01234567890., /\()?"':;-

Nahuatl

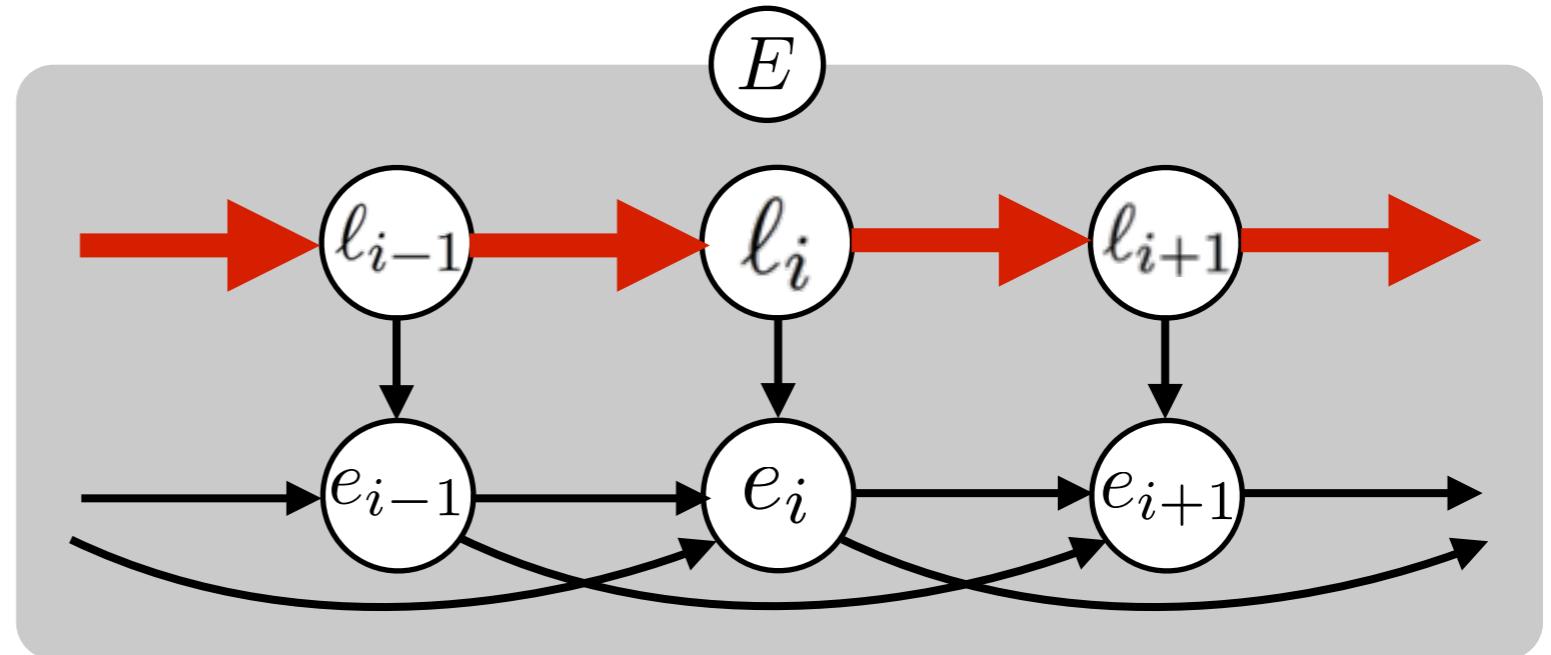
ABCDEFGHIJKLMOPQRSTUVWXYZ
abcdefghijklmopqrstuvwxyz
01234567890., /\()?"':;-

íxtinzitzin, con que



Code-Switching Language Model

$P(\ell_i | \ell_{i-1})$ is learned unsupervised via EM, with a hyperparameter biasing the model toward **not** switching (long language spans)



Code-Switching Language Model

Code-Switching

Ano chicuntetl Sacramētos, otech momaquili teoac, otechmocavililī - teoac. Inic centlamantli . iehoatl in nè- quatequiliztli, in **Baptismo** . inic vntlamantli, iehoatl in **Confirmación** : inic etlamantli, iehoatl, in itlaçonacaiotzi; inic nauhtlamantli, iehoatl in neiolme- laoaliztli.

quimomanililique inihqui texnichtlii
tzo corona:inic macuilpa, iquac in oqui-
ceppa, iquac in quimococoionilique
inijcxitzi:inic chicoppa, iquac in quimi-
xililique yiomotlantzinco.

Q V A R T O

Psalmo.

C Ano chicuntetl Sacramētos, otech- momaquiliteoac, otechmocauilili- teoac: Inic centlamantli , iechoatl in ne- quatequiliztli, in Baptismo : inic vntla- mantli , iechoatl in Confirmacion : inic etlamantli, iechoatl initlaçonacaiotzi; inic nauhtlamantli, iechoatl in neiolme- laoaliztli.

Inic macuillamantli, iechoatl in Extre- ma vnction: inic chiquacentel , iechoatl in teupixcaiutl, initoca Orden sacerdo- tal: inic chicuntetl, iechoatl in nenamicti- liztli, itoca Matrimonio.

Orthographic Variability

Orthographic Variability

	Original form	Modern form
dize	dize	dice
número	numero	número
Dõde	Dõde	Donde

Orthographic Variability

Spelling, accent usage, shorthand, etc vary by:

- region
- time period
- author
- printer
- document
- page
- etc

Orthography-specific supervision would be impossible.

Orthographic Variability

- Latinized orthographies of indigenous languages (such as Nahuatl) were being developed concurrently.

Orthographic Variability

- We train our language models from available text (e.g. Project Gutenberg)
- Modern transcribers use modern spellings, which often do not match the printed documents

Orthographic Variability

- Scholarly editing tells us that it's hard to determine what “correct” output looks like.
 - Literal transcription (aka **Diplomatic**)
 - **Normalized** transcription
- Long history in determining how to prepare texts for publication.
- Different scholars want different outputs (e.g. search engines or historical linguistics).

Orthographic Variability

Two spellings of ‘mentira’ from one page of one book

mentira

niçira

Orthographic Variability

merita

Without handling orth. variation: merita

Correct diplomatic transcription: mētira

Correct normalized form: mentira

Unsupervised Orthography

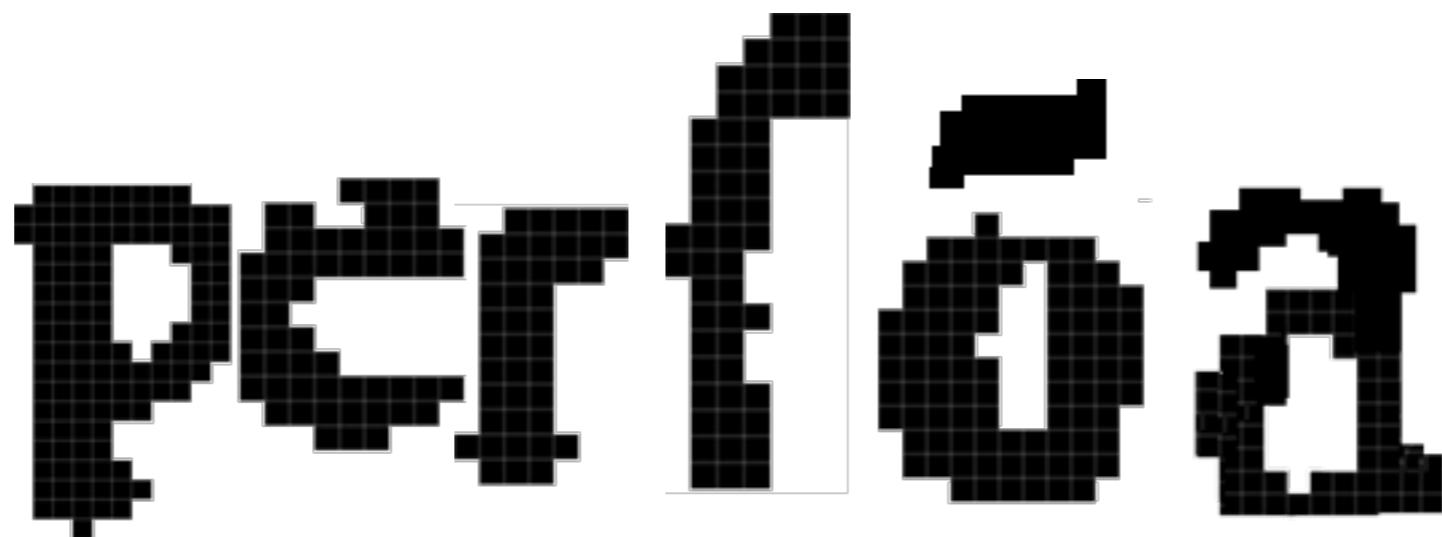
Our approach:

Induce a probabilistic mapping between
normalized writing and the orthography of the
document.

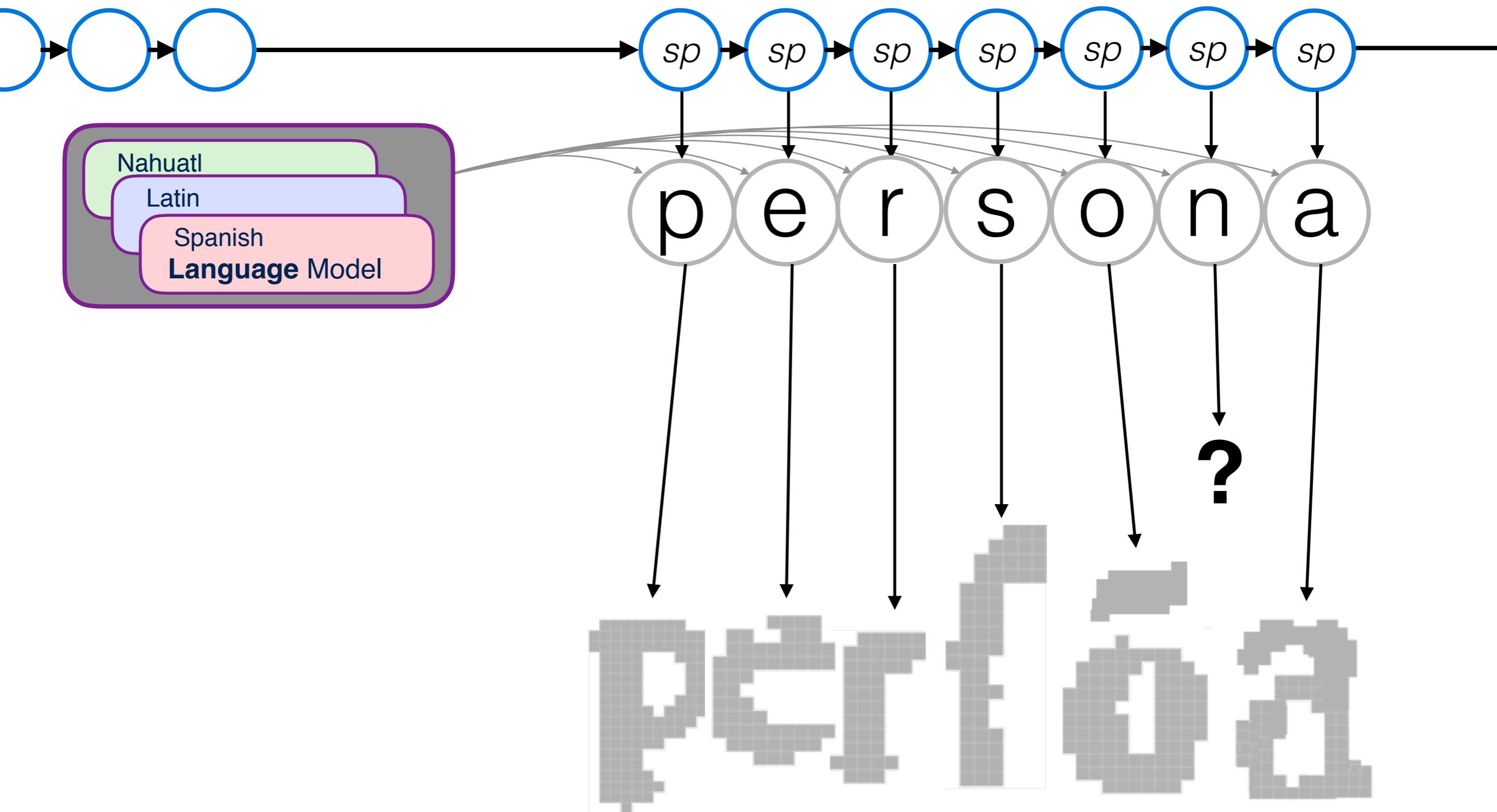
Unsupervised Orthography

Our statistical model attempts to answer the question:

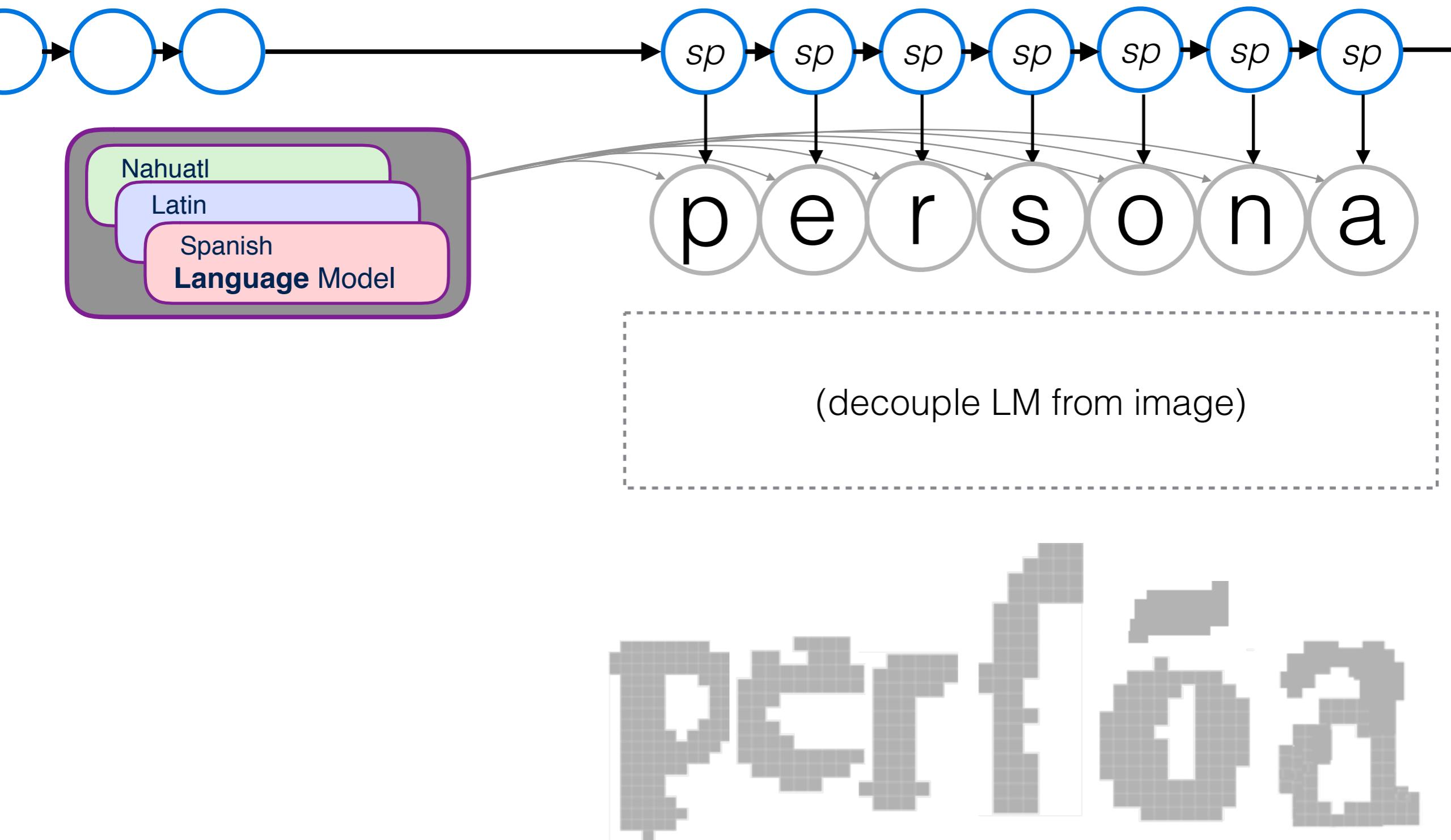
“How did these pixels end up on the page?”



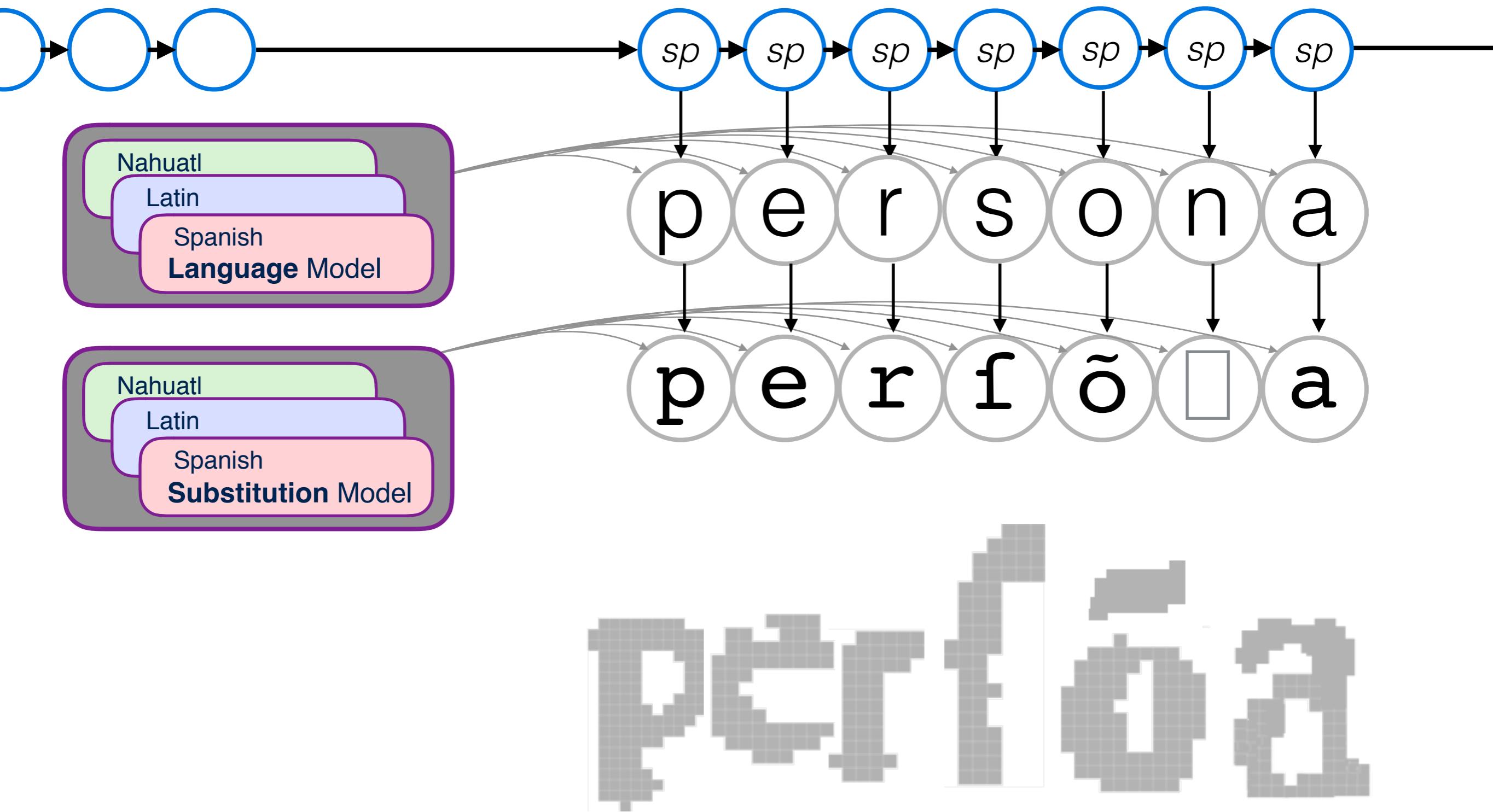
Unsupervised Orthography



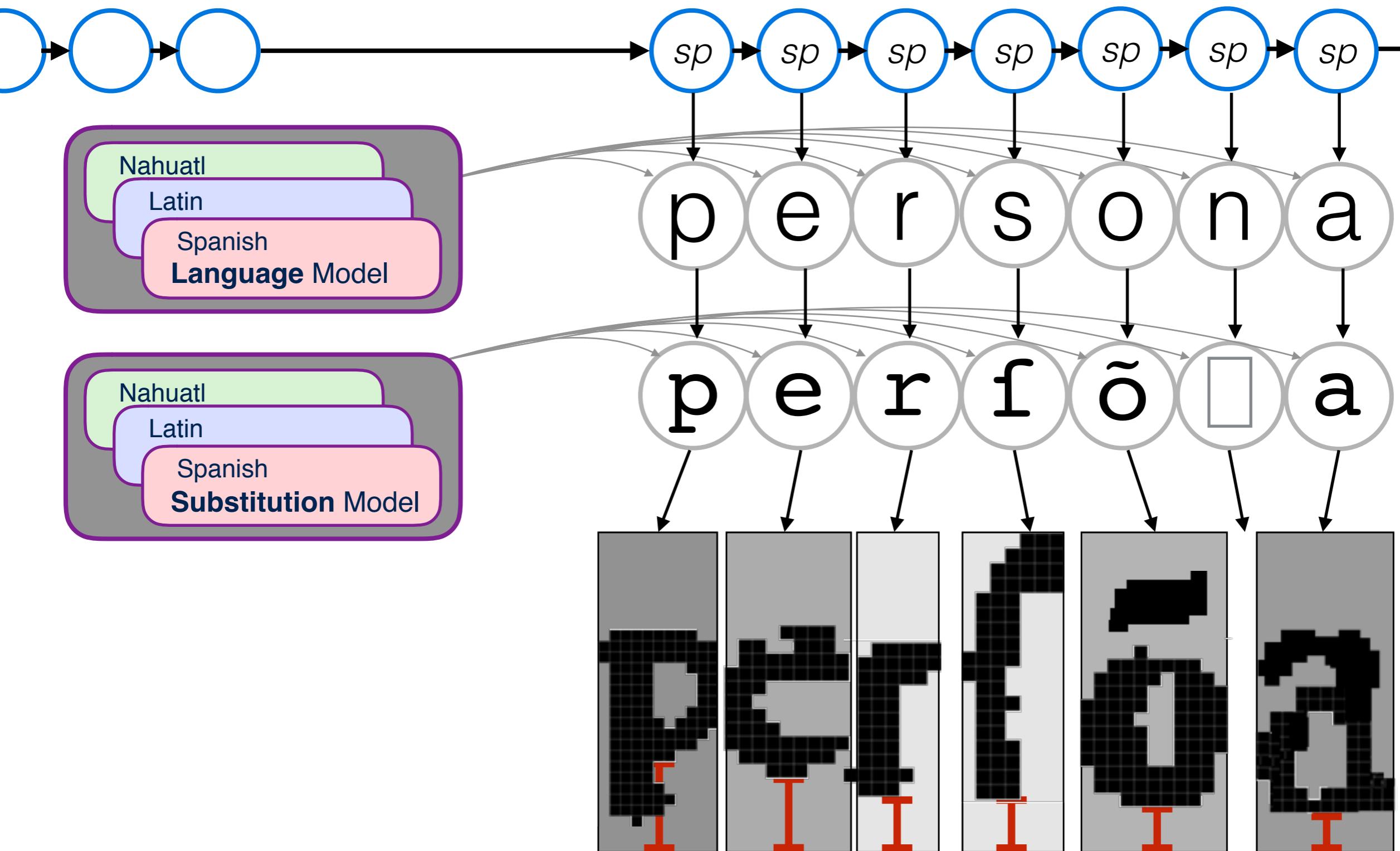
Unsupervised Orthography



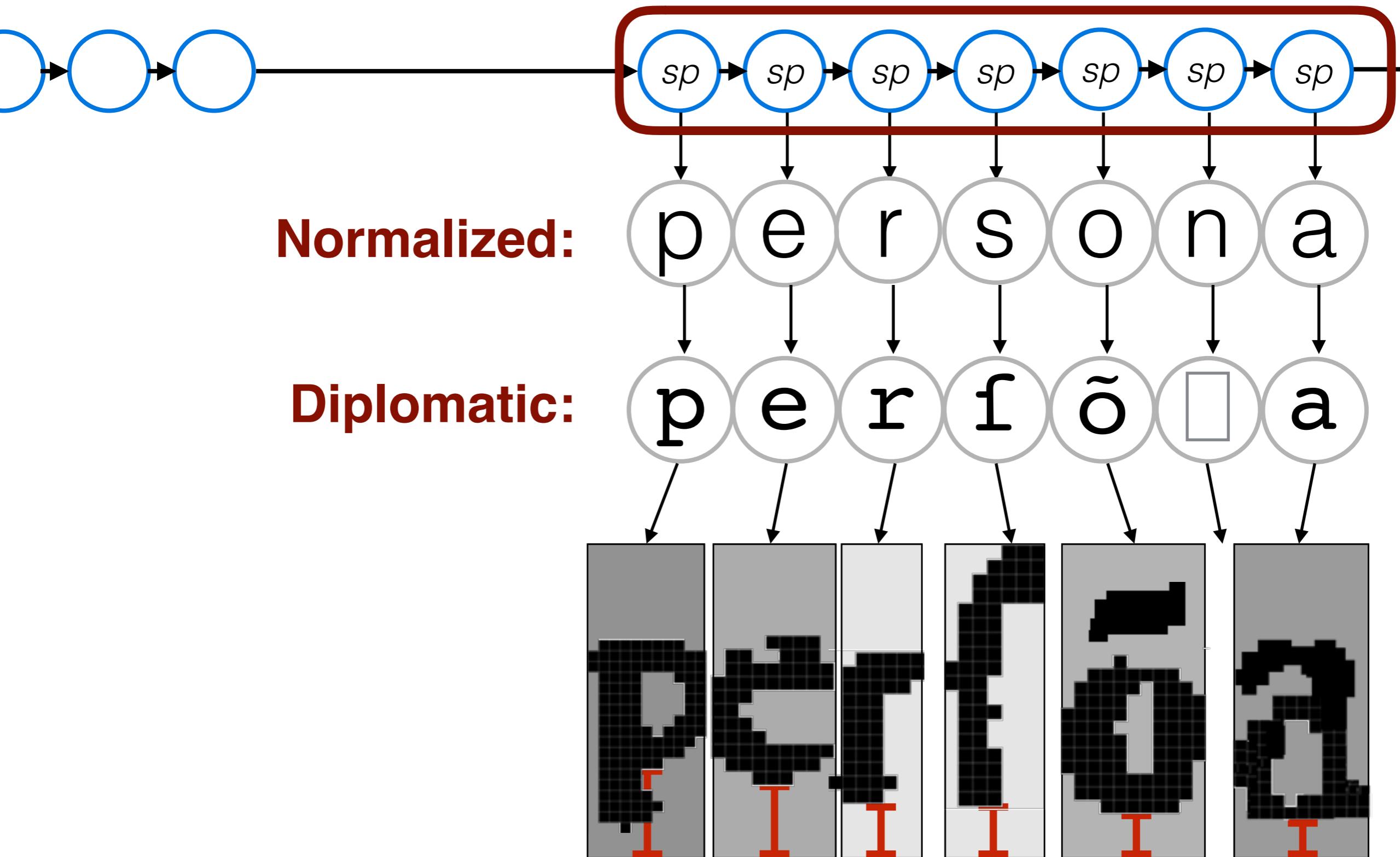
Unsupervised Orthography



Unsupervised Orthography



Unsupervised Orthography



Unsupervised Orthography

- We produce simultaneous diplomatic and normalized transcriptions.
- Standard approach is a pipeline: diplomatic transcription, then rule-based normalization.
- Joint modeling works better.

Joint Dual Transcription

For Nahuatl texts

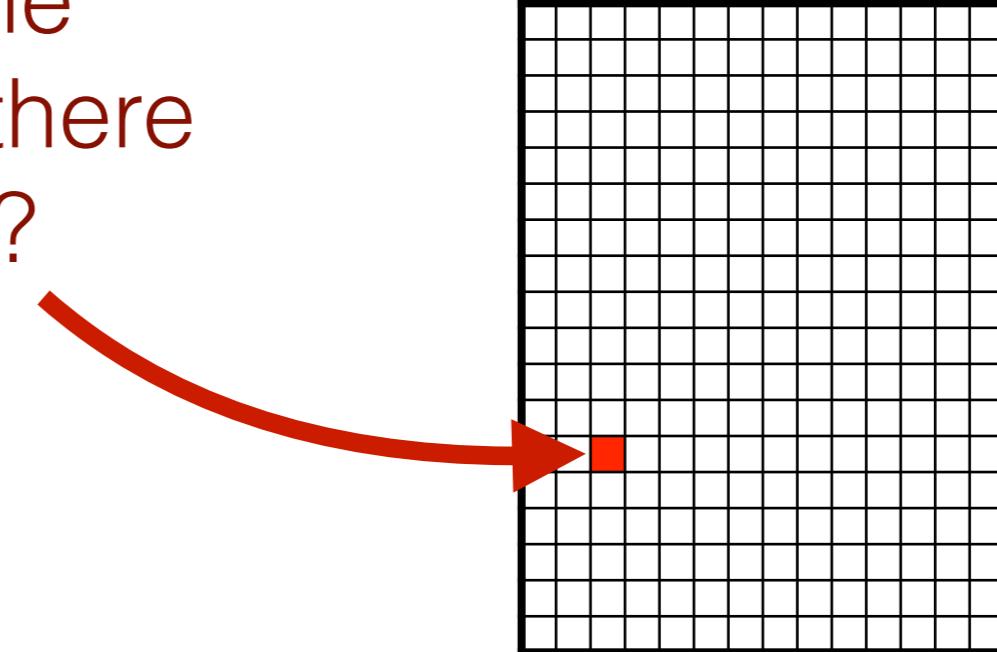
- Latinized orthography was developed by people who referred to Nahuatl as “barbaric” and “deficient” in certain sounds.
- Recently proposed orthographies seek to *decolonize*.
 - e.g. with tonal markings
 - and co-authoring with native speakers

Learning the Model Parameters

Parameters

- How do we get the “statistics” for this statistical model?

For the letter ‘a’, what’s the probability in this font of there being a **black pixel** here?



Problem

Train an OCR system from only:

1. Images of pages
2. Modern monolingual text

(no manually-transcribed pages,
no orthographically-variant example text)

Parameter Learning

- We want the model to learn good parameters on its own (unsupervised).
- We use the Expectation-Maximization algorithm

Parameter Learning

Expectation-Maximization

- Initial:
 - Estimate language model n-gram counts.
 - Strong prior on “don’t substitute”, otherwise uniform.
 - Average all fonts installed on the computer.
- Transcribe pages, update parameters, repeat.

The EM Algorithm

0. Initialize your expectations.

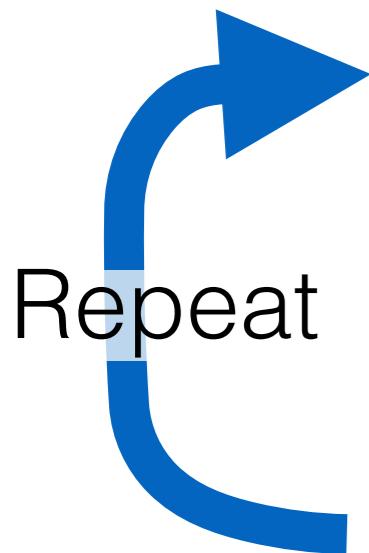
“I think this is what an ‘a’ looks like.”

1. Transcribe data based on your expectations.

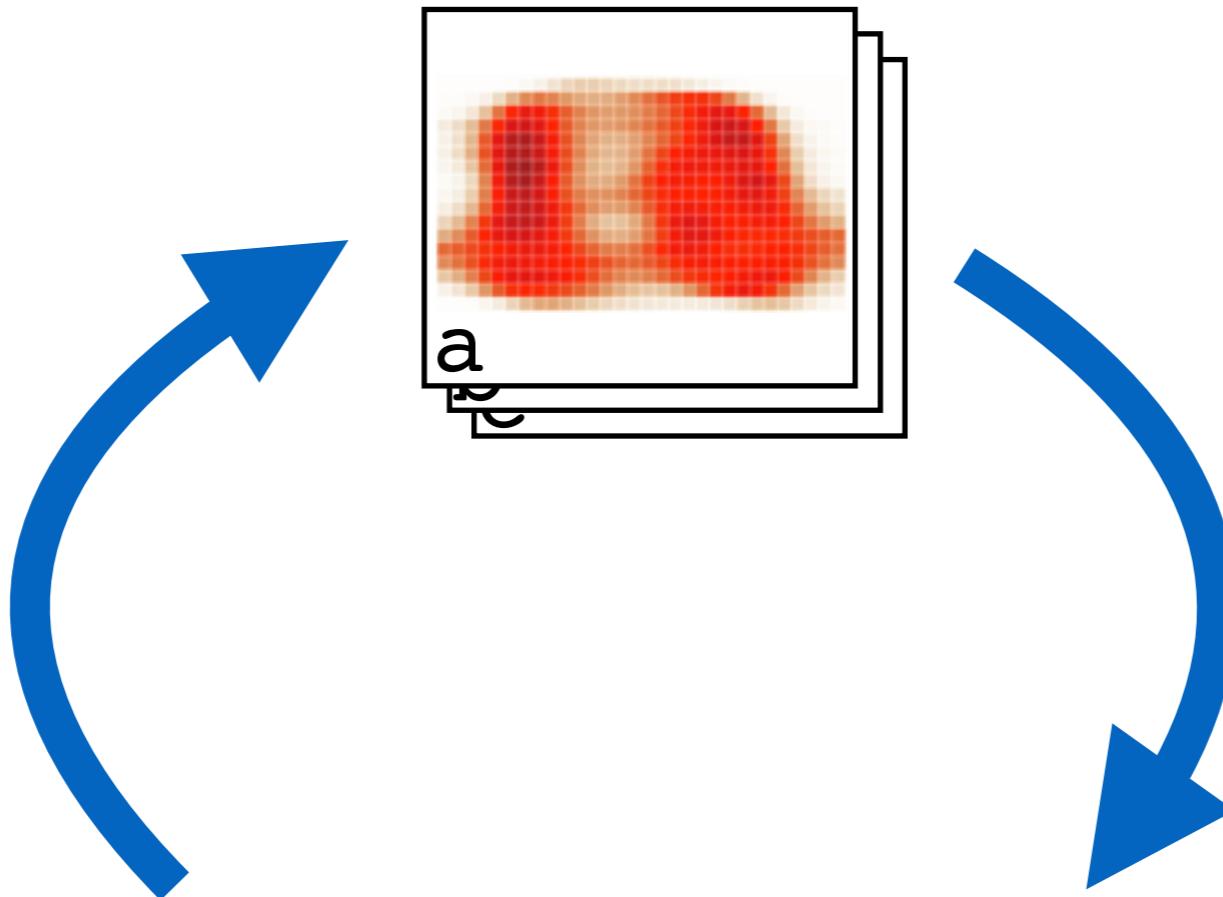
“I think these are all the ‘a’s.”

2. Update your expectations.

“Oh, I guess *that’s* what an ‘a’ looks like.”



The EM Algorithm



a coftúbrar el anima a

acoftúbrar el **anima** a

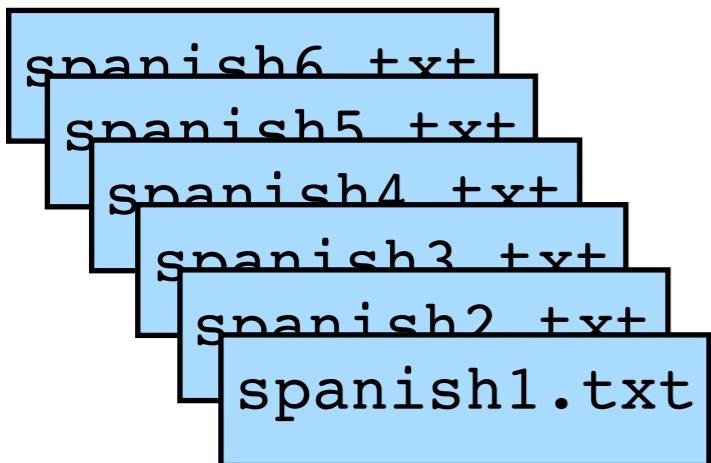
Experiments

Experiments

- Evaluated on seven different *Primeros Libros* books
- Years 1553 to 1600
- Differing fonts, language proportions, clarity, etc

Orthography Baseline

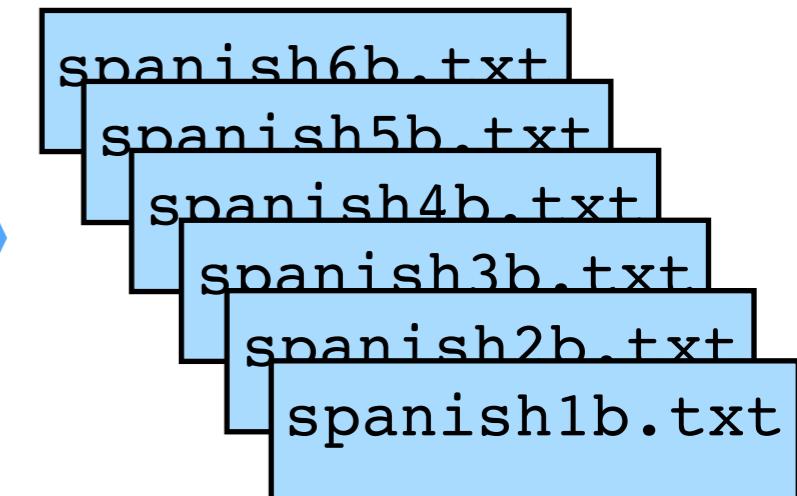
Modern Spanish



Replacement Rules

u → v
c → z
ú → u
on → õ
que → q̃

“Old” Spanish



Experiments

<i>Orthographic variation strategy</i>	Diplomatic		Normalized	
	CER	WER	CER	WER
No handling	13.2	45.7	17.4	47.6
Hand-written rules	8.5	30.8	13.1	37.9
Unsupv. joint model	8.6	32.7	9.5	27.6

Example Outputs

	char sub. (c → q)	char sub. (s → long s)	elision (que → q̄)	accent drop (ó → o)	doubled (c → cc)
Original image	qual	esta	aql	consideracion	peccados
No orth. handling	qual	eña	á ól	confideracion	peccados
Our diplomatic	qual	efsta	aql	confideracion	peccados
Our normalized	cual	esta	aquel	consideración	pecados

Example Outputs

Original image

de las dos que se siguen en las cuales apro
uecha mucho acostúbrar el anima á se le-

Our diplomatic

de las dos que se siguen en las cuales apro
uecha mucho acostúbrar el anima á se le-

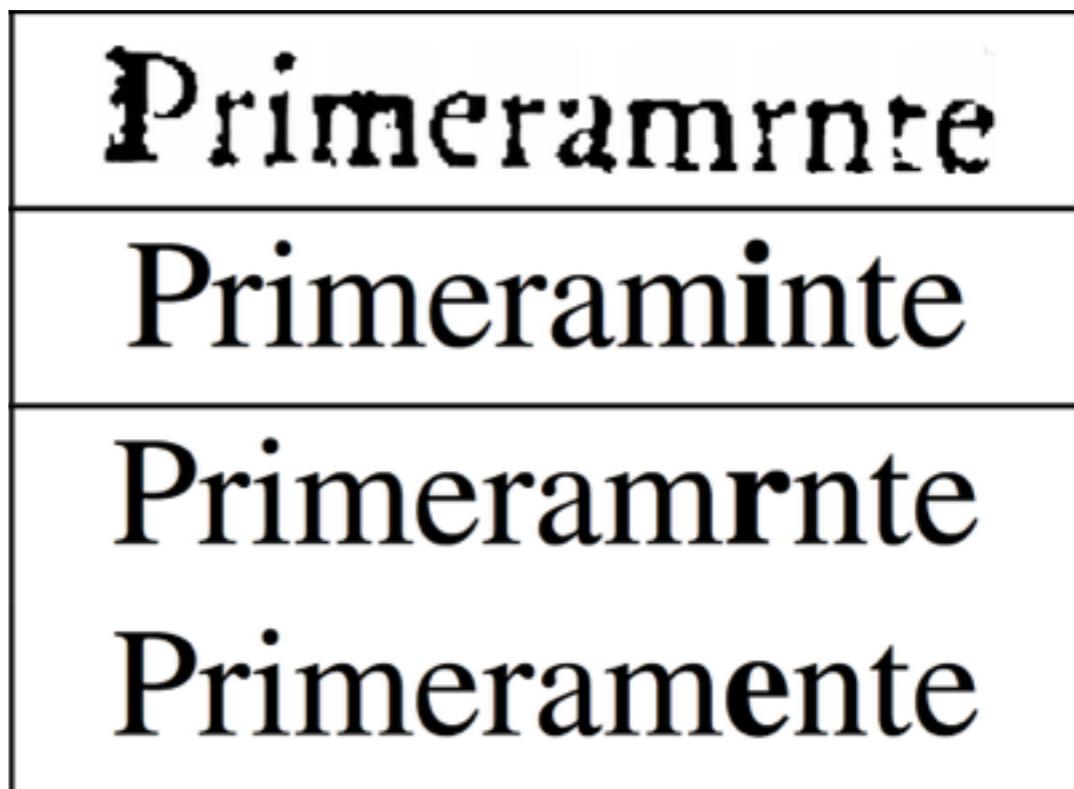
Our normalized

de las dos que se siguen en las cuales apro
uecha mucho acostumbrar el ánima á se le-

Example Outputs

- Original image
- No orth. handling
- Our diplomatic
- Our normalized

Typographical Error



Example Outputs

Original image

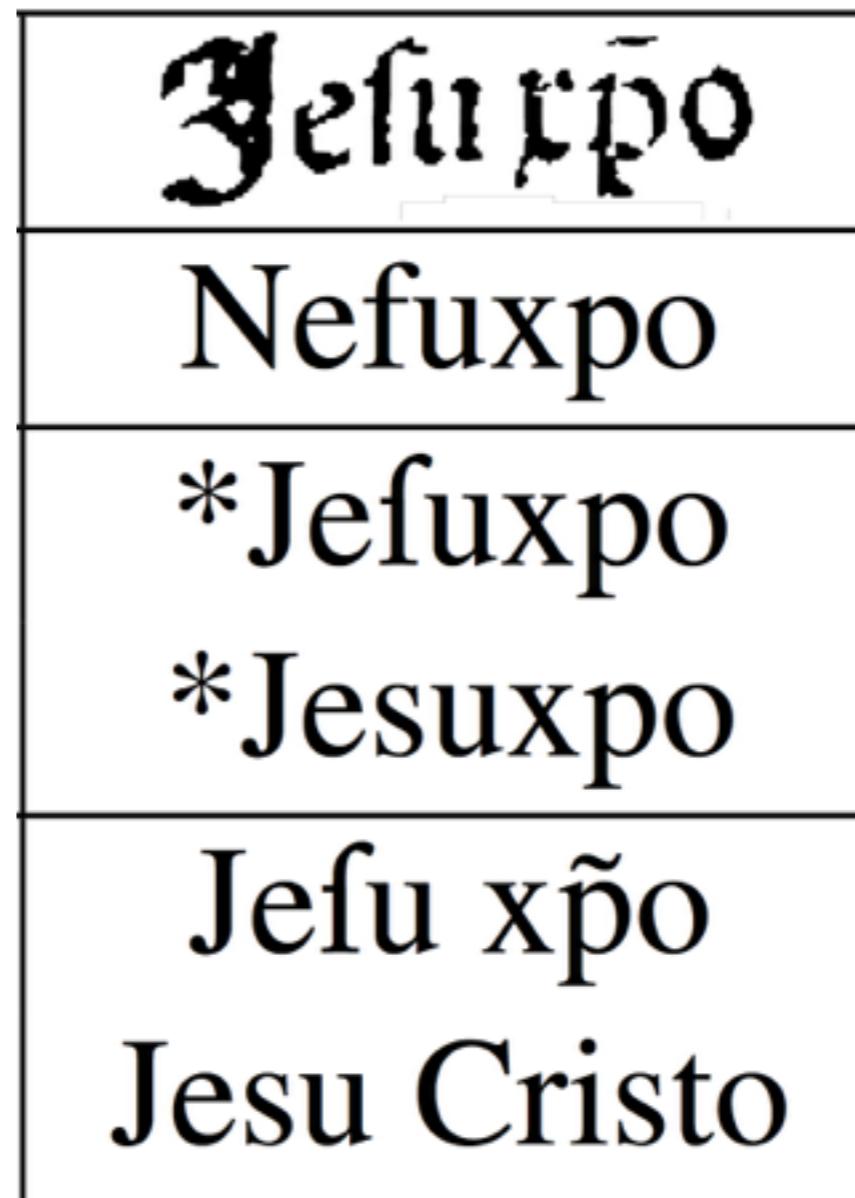
Baseline

Our diplomatic

Our normalized

Gold diplomatic

Gold Normalized



Example Outputs

Los verbales en li, o en tli, antepo-

Model output:

sí

tīi

← Spanish

Gold transcription:

li

tli

← Nahuatl

Model avoids switching languages, but this is actually from a description of Nahuatl grammar.

Example Outputs

tetechtlamieccâquixtiliztli

tetechtla miec caquixtiliztli

Spacing in Nahuatl writing was not standardized.

Broader Implications

Joint Dual Transcription

- Normalization provides access for:
 - Students and non-native speakers for whom historical orthography is hard to read.
 - Screen-readers for the visually impaired.

Joint Dual Transcription

Particularly relevant for transcribing texts in indigenous languages.

- More dramatic changes in orthography.
 - 1570: Nahuatl becomes official language of New Spain.
 - Nahuatl spreads across Mesoamerica.
 - 1696: All indigenous languages are banned.
- Reviving and standardizing written forms of these languages is an important effort today.

Joint Dual Transcription

For Nahuatl texts

- Latinized orthography was developed by people who referred to Nahuatl as “barbaric” and “deficient” in certain sounds.
- Recently proposed orthographies seek to *decolonize*.
 - e.g. with tonal markings
 - and co-authoring with native speakers

Joint Dual Transcription

- Our approach is able to target any “normal” orthography, by giving different LM training data.
 - Many scholars have their own preferences about Nahuatl orthographic standards.
- We can map from the original, colonial, orthography to an actively decolonial form.
- *At the same time*, preserving the historical/societal context in which the book was originally printed.

Output as Metadata

- Language-tagged output provides additional information that would be difficult or impossible to determine by hand.
 - Languages appearing in a book
 - Orthographic patterns

Output as Metadata

c	<i>replacement</i>	$freq(sp., c, g)$	$P_{spanish}^{\text{GLYPH}}(g \mid c)$
-	ELIDED	52	0.0881
ó	o	31	0.0526
s	f (long s)	325	0.0352
q	ñ	9	0.0222
n	ELIDED	57	0.0136
v	u	55	0.0129
ó	ö	20	0.0091
c	cc	23	0.0028

Output as Metadata

Allows inquiry into further questions:

- Can patterns of orthography and code-switching tell us the author, printer, timeframe, region, etc?
 - Work has been done on this to identify the typesetters of Shakespeare texts.
- Changes in orthography over time can give insight into historical linguistics.

Documentary Linguistics

- Linguists often need to aggregate small resources acquired from various sources.
- Many texts produced by naïve authors who don't conform to any orthographic standards.
- Good opportunity for low-resource text normalization (with or without OCR component)

Conclusion

Conclusion

Aduertencias para

como parece manifiesto en las palabras de S.
Ioan, que dize. Tres sunt qui testimonium dāt
in cœlo Pater, Verbum, & Spiritus sanctus :
& hi tres vnum sunt. i. Ioann. ultimo. Por
loqual deuen ser instruydos y enseñados, que
todas tres diuinæ personas son vn Dios ver-
dadero; o reformando la sobre dicha proposi-
cion, y añadiendo esta palabra. In huelime
ixtintzitzin, con que se quita toda amphibio-
logia y duda diciendo. In Dios, ca Tettatzin
Tepiltzin, Spiritu sancto, ei personas, çan ce
huelnelli teutl Dios in huelimeixtintzitzin,
q. d. Dios es Padre, Hijo, y Spū sancto tres
personas, vn solo Dios verdadero todas tres,
cō la qual reduplicacion se quita toda dubda.
Tambien se quita con estas propoſiciones.
In D I O S, ca Tettatzin, Tepiltzin, Spiritus an-
to, çan huel iceltzin teutl Dios tlahuohuani.
In Dios, ca Tettatzin, Tepiltzin, Spiritus an-
to, imēixtin personas çan huel iceltzin Dio-
tlahuohuani. Ca inimēixtin personas me ca-
çan huel iceltzin teutl Dios tlahuohuani in-
huel imēixtin. ¶ Otros responden [y es el
segundo error] çace Dios tlahuohuani, imme-
teihtotica, y a algunos de sus ministros les ha-
parecido el meteihhtotica, vn vocablo en si di-

Conclusion

Diplomatic Transcription

como parece manifiesto en las palabras de S. joan, que dize. Tres sunt qui testimonium dā: in cælo Pater, Verbum, & Spiritus sanctus; g. hi tres vnum sunt. 1. Joann. ultimo por lo qual deuen fer inftruydos y enfeñados, que todas tres diuinas perfonas son vn Dios verdadero; o reformando la sobre dicha proposicion, y añadiendo ella palabra. in huel ini: ixtintzitzin, con que sé quita toda amphilogia y duda diciendo. In Dios. ca Tettatzin Tepiltzin, Spiritu sancto, et perfonas,) an (: huel nelli teutl Dios in huel imeixtintzitzin. q. d. Dios es Padre, Hijo, y Spū sancto tres perfonas, vn solo Dios verdadero todas tres, cō la qual reduplicacion se quita toda dubda. También se quitá con eftas proposiciones.

In D;f O S.ca Tettatzin, Tepiltzin, Spirituf aa cfo, can huel iceltzin teutl Dios tlahhuani. in Dios. ca Tettatzin, Tepiltzin, Spiritu sancto, in ieixtin. perfonas can huel iceltzin Dios tlahhuani. Ca in imeixtin perfonasme ca- can huel iceltzin teutl Dios tlahhuani in- huel imeixtin. ¶ Otros respondenj y es é segundo error j ca ce Dios tlahhuani, im iut tītfittoti cá, y á algunos de fus ministros les fu parescido et merei tutorica, vn vocablo enjid ui in

Normalized Transcription

como parece manifiesto en las palabras de S. joan, que dice. Tres sunt qui testimonium dā: in cælo Pater, Verbum, & Spiritus sanctus; g. hi tres unum sunt. 1. Joann. ultimo por lo cual deben ser instruidos y enseñados, que todas tres divinas personas son un Dios verdadero; y reformando la sobre dicha proposicion, y añadiendo ella palabra. in huel ini: ixtintzitzin, con que sé quita toda amphilogia y duda diciendo. In Dios. ca Tettatzin Tepiltzin, Spiritu sancto, et personas,) an (: huel nelli teutl Dios in huel imeyxtintzitzin. q. d. Dios es Padre, Hijo, y Spū sancto tres personas, un solo Dios verdadero todas tres, cō la cual reduplicación se quita toda dubda. También se quizá con estas proposiciones.

In D;s O S.ta Tettatzin, Tepiltzin, Spiritus ad cro, can huel icaltzin teutl Dios tlahhuani. in Dios. ca Tettatzin, Tepiltzin, Spiritu sancto, in ieixtin. personas can huel icaltzin Dios tlahhuani. Ca in imeyxtin personasme ca- can huel icaltzin teutl Dios tlahhuani in- huel ineixtin. ¶ Otros respondena y es é- segundo error y ca ce Dios tlahhuani, in iuh ticteittiti cá, y á algunos de sus ministros les su parescido et merei tutorica, in vocablo evid- ui in

Conclusion

- By accounting for **multilingual text** and **obsolete orthography**, we can improve the state-of-the-art for historical OCR.
- High-quality transcription tools can open up vast new resources for scholars to study.
- Code and papers available at <http://dhg.ai>