

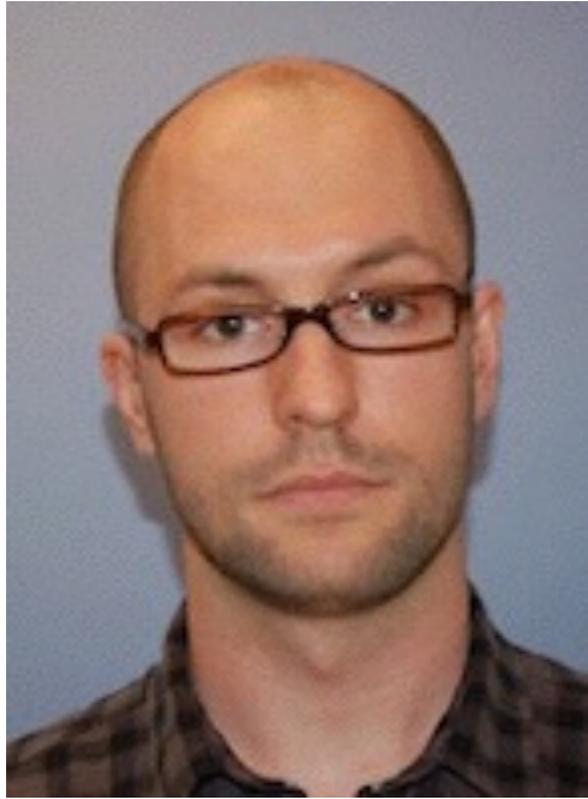
Unsupervised Modeling of Code-Switching and Orthographic Variation for Historical Document Transcription

Dan Garrette

UW CSE

<http://dhg.ai>

Collaborators



Taylor Berg-Kirkpatrick

Dan Klein

Hannah Alpert-Abrams

Computer Science
UC-Berkeley

Comparative Literature
UT-Austin

Document Transcription

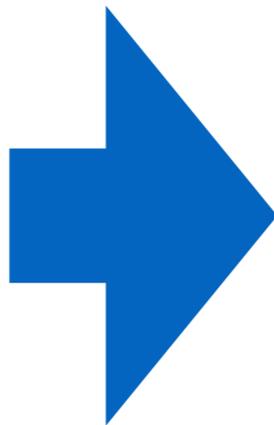
Aduertencias para

como parece manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimonia dāt in cælo Pater, Verbum, & Spiritus sanctus : & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas personas son vn Dios verdadero: o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzin, con que se quita toda amphibologia y duda diziendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzin, 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 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] çace Dios tlahtohuani, imme reihttōtica, y a algūos de sus ministros les ha parecido el meteihhtōtica, vn vocablo en si d

Document Transcription

Aduertencias para

como parece manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniā dāt in cælo Pater, Verbum, & Spiritus sanctus : & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas 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 diziendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzitzin, 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 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] çace Dios tlahtohuani, imme teihttotica, y a algũos de sus ministros les ha parefcido el meteihttotica, vn vocablo en si di uino



Aduertencias para

como parece manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniū dāt in cælo Pater, Verbum, & Spiritus sanctus: & hi tres vnum sunt. 1. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas 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 diziendo. 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 estas 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] çace Dios tlahtohuani, imme teihttotica, y a algũos de sus ministros les ha parefcido el meteihttotica, vn vocablo en si di uino

Linguistic Noise

Code switching (multilinguality)

Orthographic variation (cross-linguality)

Primeros Libros

Books printed in Mexico in the 1500s



Primeros Libros

LICENCIA.
Antonio del Rincon para que haga ympri-
mir a Pedro Balli ympressor la arte que de
sufo se haze mencion con que sea confor-



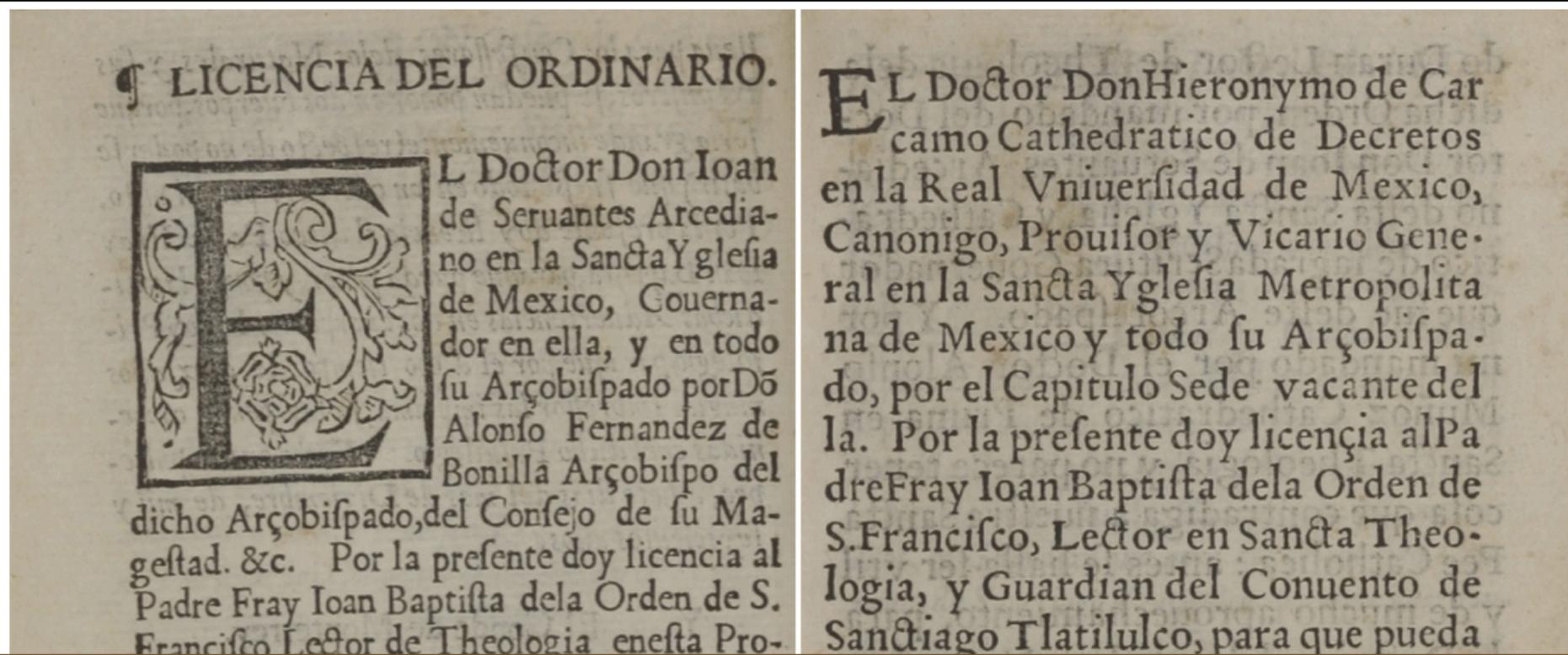
O Estevan Paez Pro-
cial de la Compañia
Iesus en esta nueva Eipa

PRIMEROS LIBROS DE LAS AMERICAS

Advertencias para los confesores de los naturales (Primera parte)
Juan Bautista, fray, 1555-approximately 1613 - Published: 1600 Benson Latin American Collection (COPY 1) Detalles | Descargar PDF

1 1 2 p q r d gr ver todo del texto

Páginas 10-11 de 370 Ir a la página 10 IR

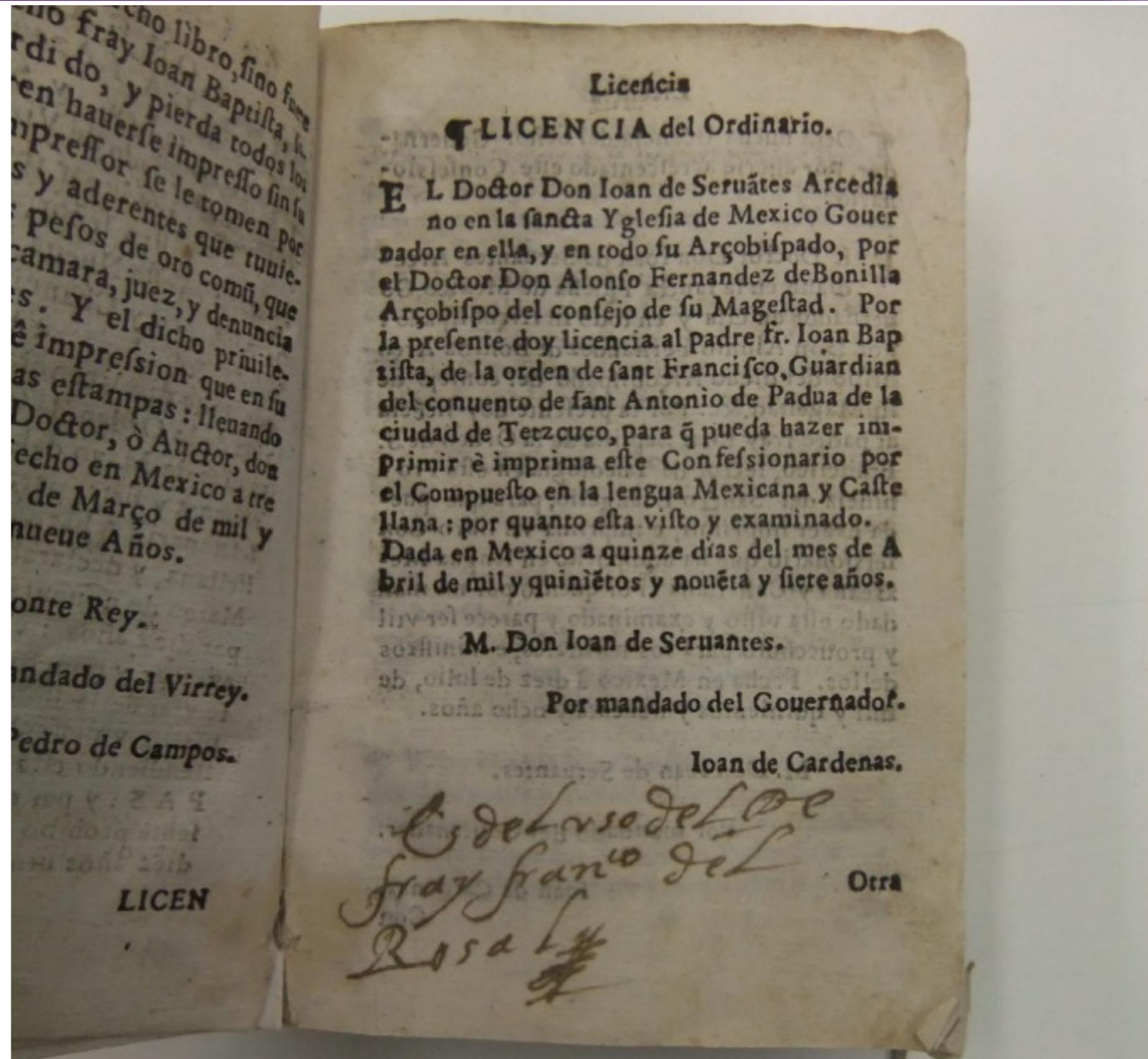


This Talk

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

role along in silent **room**

Irregular alignment

the Death of the Deceased,

Unknown Fonts



g g g g g g

Unknown Fonts

Gante	(1553)	motlacatilia: ynica sacramento Baptis
Anunciación	(1565)	¶ Rincóneltoquitia yndios
Sahagún	(1583)	Yoan oquihui in Emperador, in taca
Rincón	(1595)	etion.v.g.tetlaçotlaliztli.amatio, vel,
Bautista	(1600)	¶imo, hæc supra dictus doctor Medina.

Unknown Fonts

These are all 'a's from one page of one book:

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

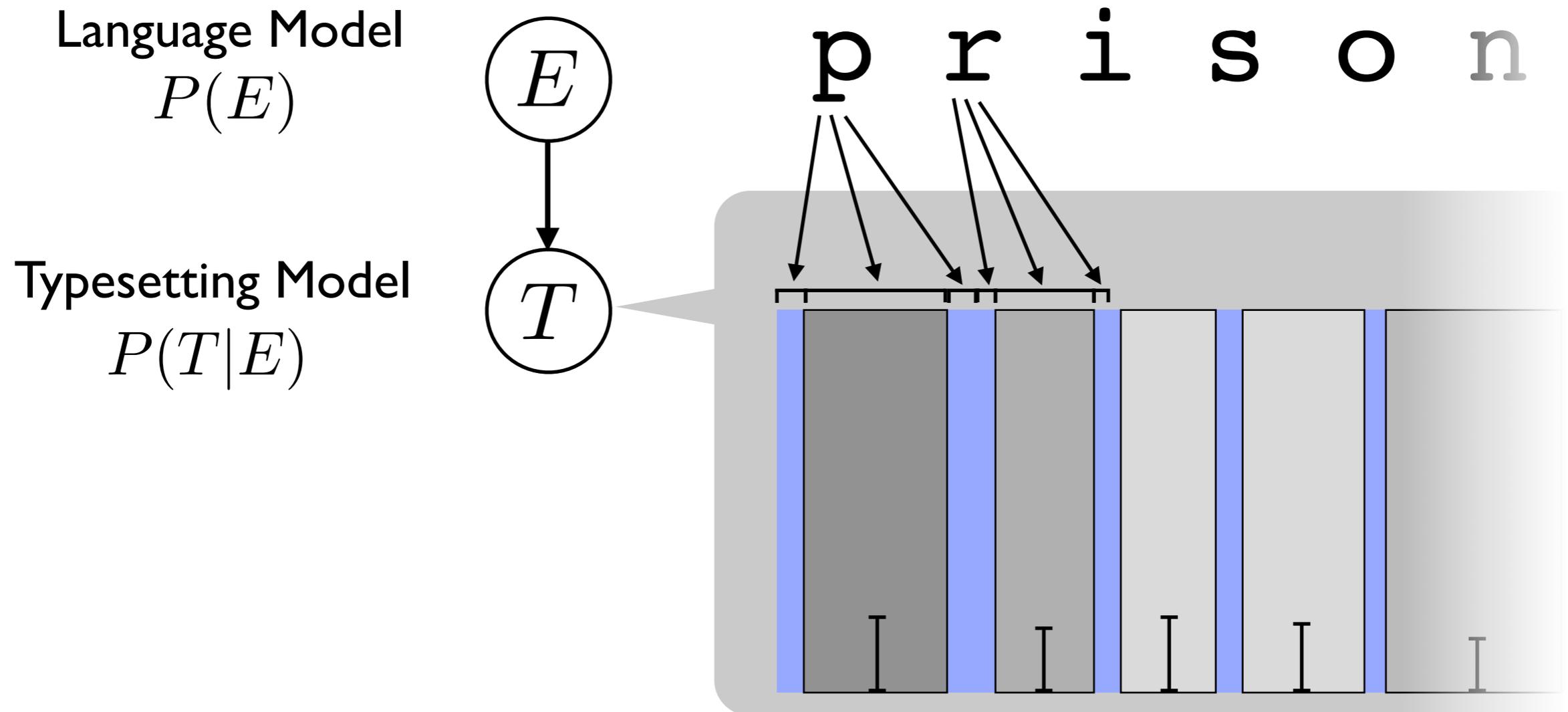
Ocular's Generative Model

Language Model
 $P(E)$

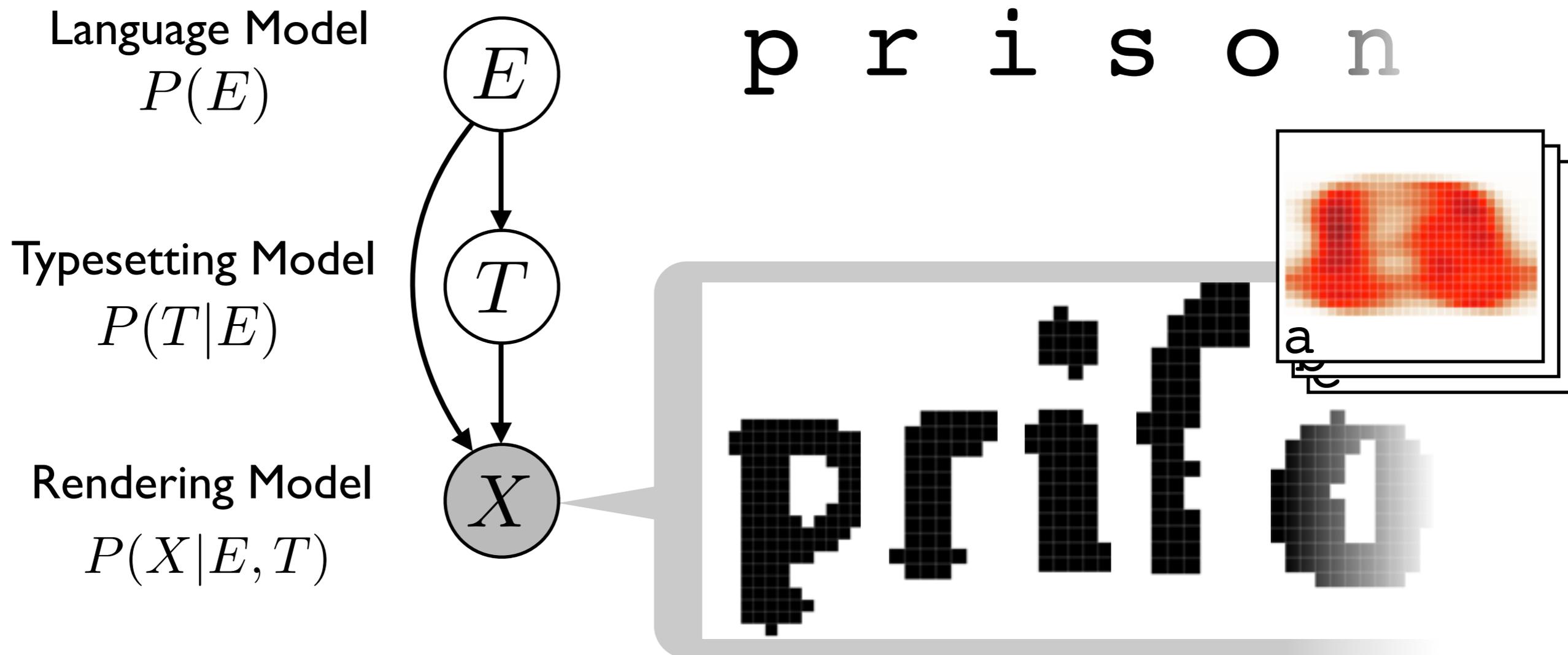
E

p r i s o n

Ocular's Generative Model



Ocular's Generative Model



Our Focus

Language Model
 $P(E)$

E

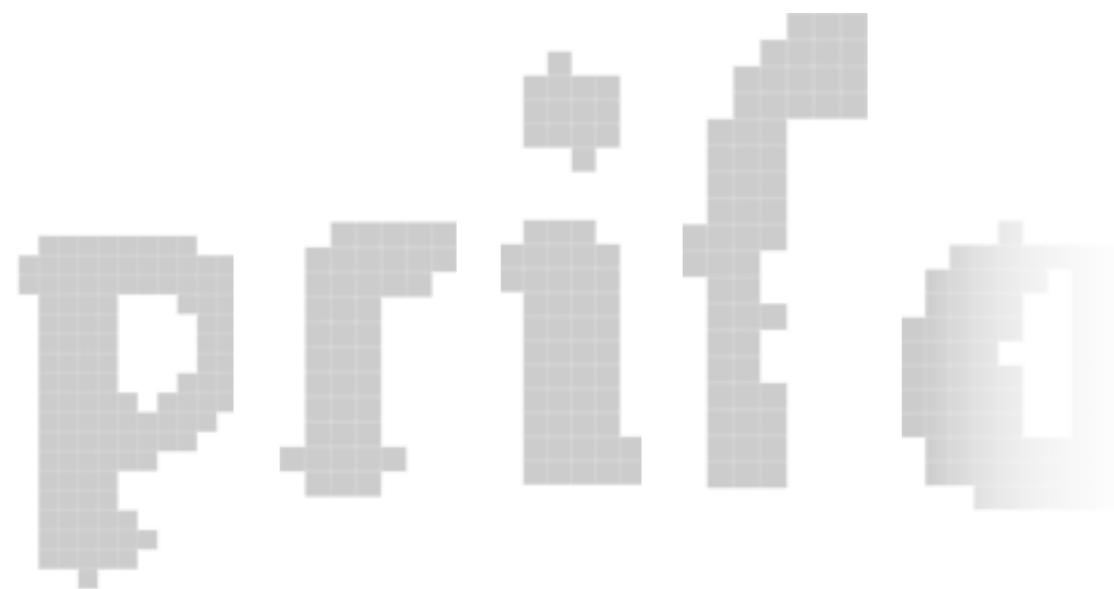
p r i s o n

Typesetting Model
 $P(T|E)$

T

Rendering Model
 $P(X|E, T)$

X



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 manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniū dāt in cælo Pater, Verbum, & Spiritus sanctus : & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas personas son vn Dios verdadero; o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzin, con que se quita toda amphibologia y duda diziendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzin, 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 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es segundo error] ça ce Dios tlahtohuani, immetihttotica, y a algūos de sus ministros les ha parecido el metehttotica, vn vocablo en si d

Aduertencias para

como parece manifesto en las palabras de S.
Ioan, que dize.

[Redacted]

Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas personas son vn Dios verdadero; o reformando la sobre dicha proposicion, y añadiendo esta palabra.

[Redacted] con que se quita toda amphibologia y duda diziendo. In Dios,

[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.

DIOS, [Redacted] Dios

Dios, [Redacted] personas Dios

[Redacted] personas me Dios

Otros responden [y es e segundo error Dios

[Redacted] y a algños de sus ministros les ha parecido el [Redacted] vn vocablo en si d

[Redacted] uino

Spanish

Latin

Nahuatl

Tres sunt qui testimoniā dāt
in cælo Pater, Verbum, & Spiritus sanctus :
& hi tres vnum sunt. i. Ioann. vltimo.

Spanish

Latin

Nahuatl

Spiritu sancto, ei

Spū sancto

Spiritusan

āo,

Spiritusan.

āo,

Spanish

Latin

Nahuatl

[redacted]

In huel ime
ixtintzitzin,
[redacted] ca Tettatzin
Tepiltzin, [redacted] çan ce
huelnelli teutl in huel imeixtintzitzin,

In [redacted] ca Tettatzin, Tepiltzin,
[redacted] çan huel iceltzin teutl [redacted] tlahtohuani,
In [redacted] ca Tettatzin, Tepiltzin, [redacted]
[redacted] imeixtin [redacted] çan huel iceltzin
tlahtohuani. Ca inimeixtin [redacted] ca-
çan huel iceltzin teutl [redacted] tlahtohuani in
huel imeixtin. ¶ [redacted]
[redacted] çace [redacted] tlahtohuani, imne
teihhtotica, [redacted]
[redacted] meteihhtotica, [redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Spanish

[Redacted]

Latin

[Redacted]

Nahuatl

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

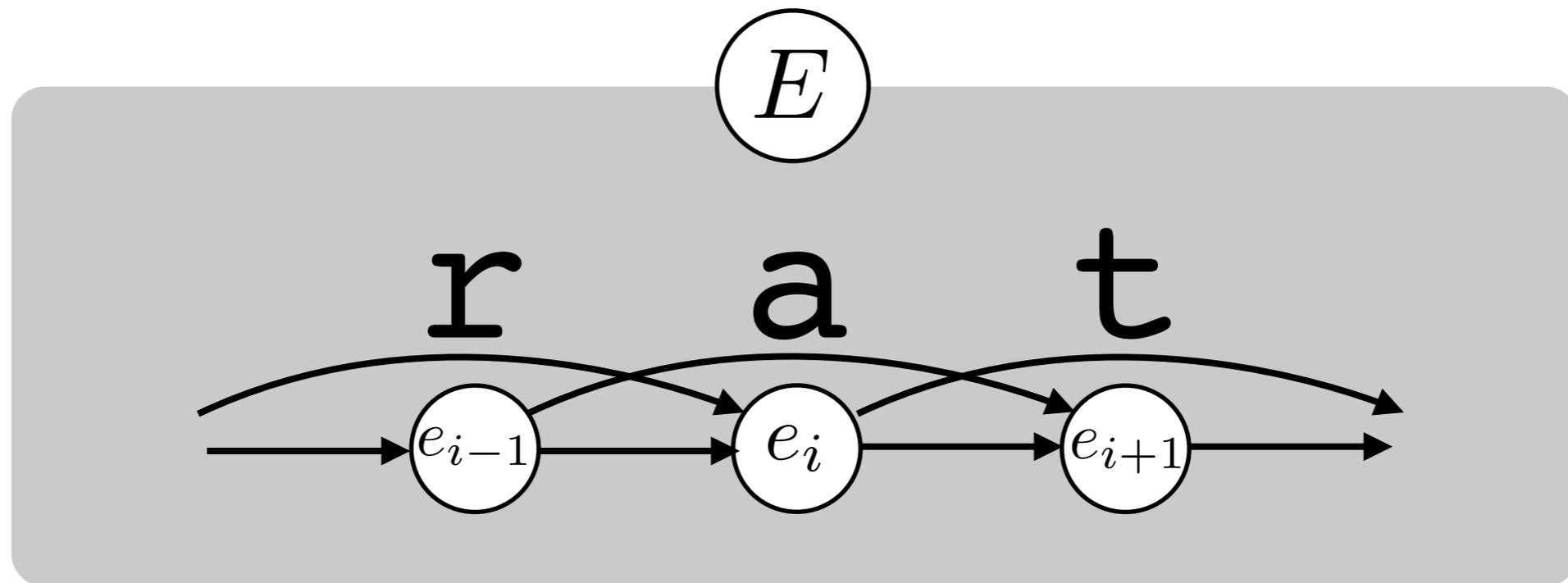
[Redacted]

Why Code Switching?

ficion, y añadiendo esta palabra. **In huelime
ixtintzitzin**, con que se quita toda amphibo-
logia y duda diziendo. In Dios, ca Tetzatzin

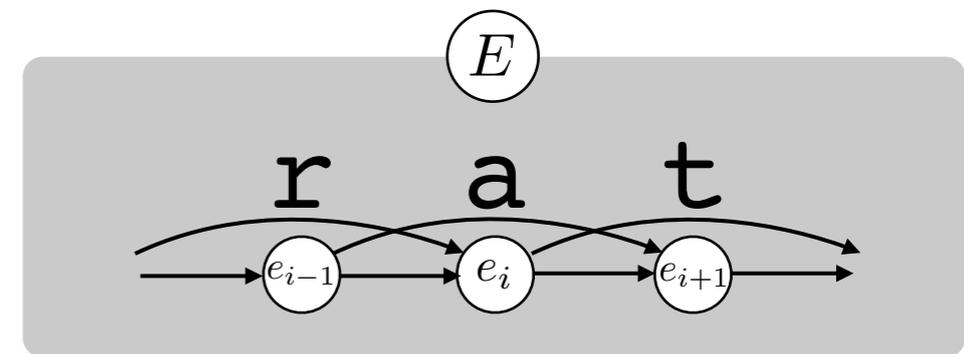
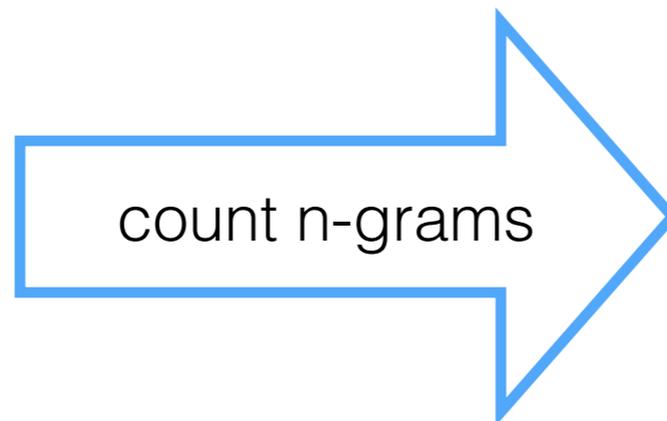
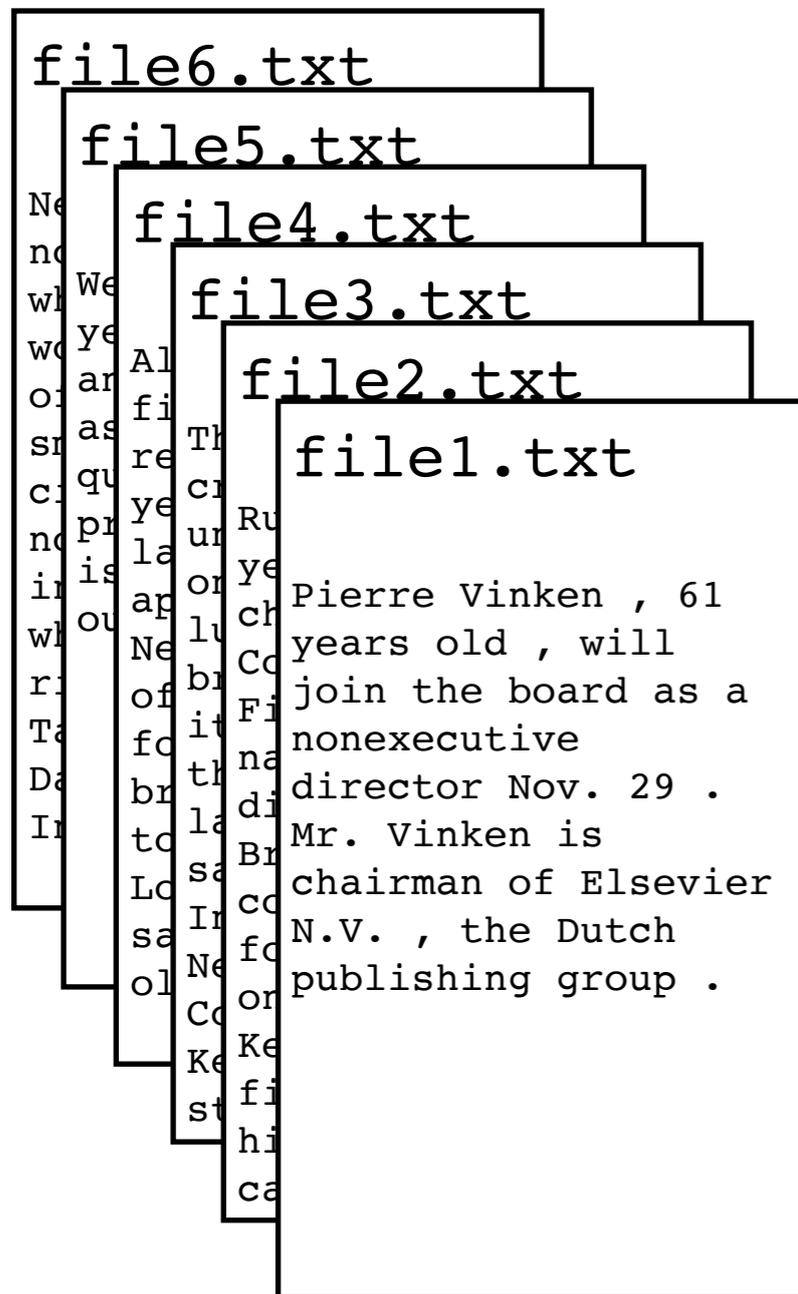
...and adding these words. **[Nahuatl phrase]**,
with which all ambiguity and doubt is removed...

Ocular's Language Model



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

Ocular's Language Model



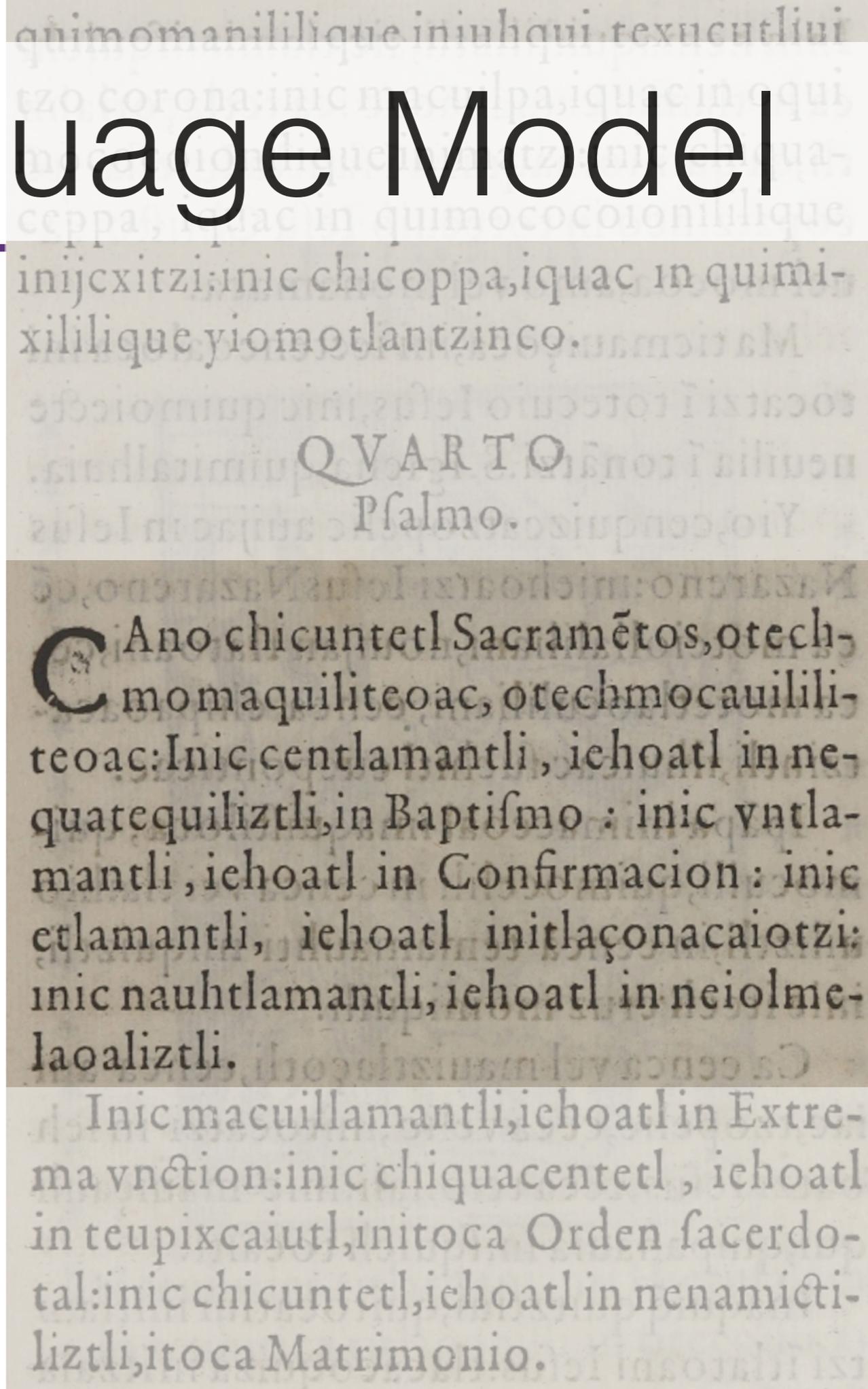
Ocular's Language Model

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** climrocaruisiliteorage Intergentilamandis, **tie board isn't** quate quilted it in Raptisme's snig Vadam and **swig broad** in Confirmacion's intergirlamandis **wife broad**. Trinidagon arcators internault dramandi, is board in **petrol** mess **oasis** di-

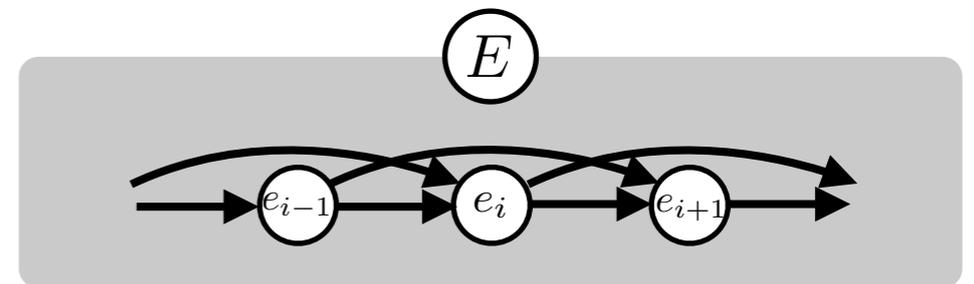


Baseline Multilingual Model

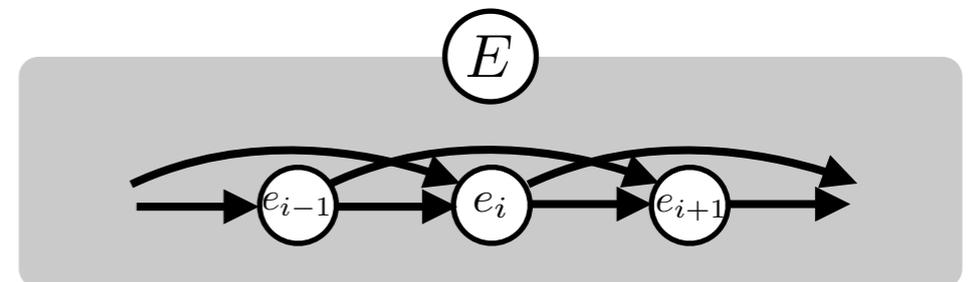
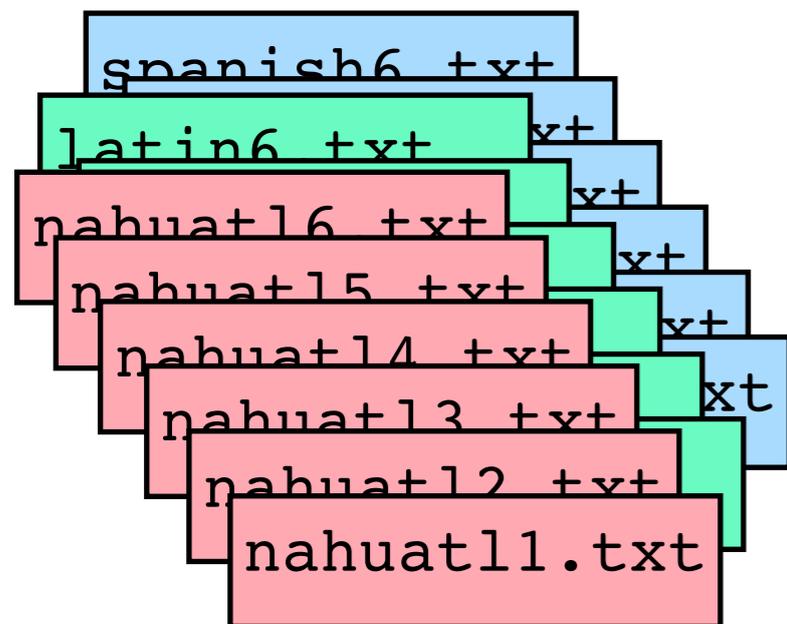
spanish6.txt
spanish5.txt
spanish4.txt
spanish3.txt
spanish2.txt
spanish1.txt

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

nahuatl6.txt
nahuatl5.txt
nahuatl4.txt
nahuatl3.txt
nahuatl2.txt
nahuatl1.txt



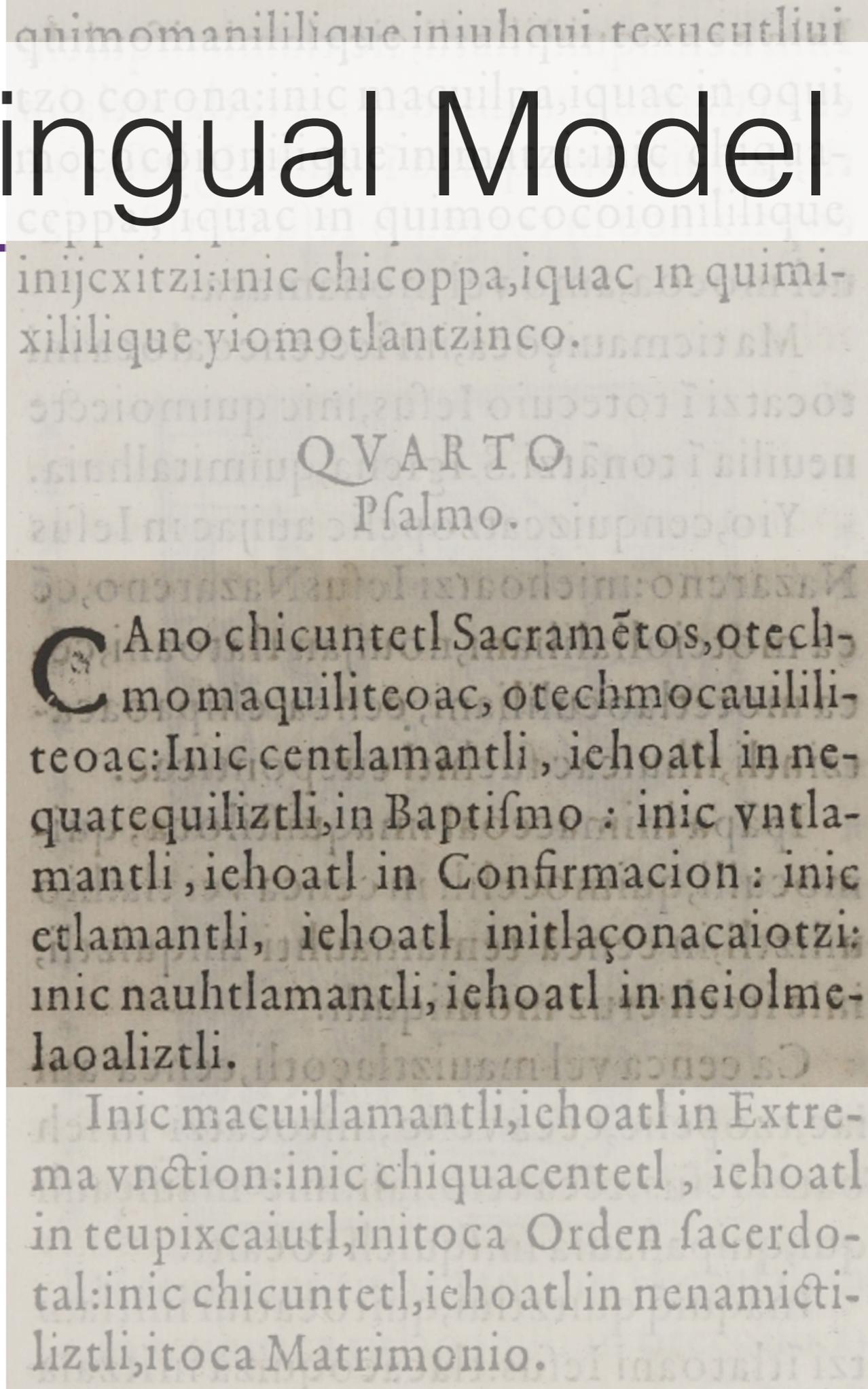
Baseline Multilingual Model



Baseline Multilingual Model

Multilingual Blur

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

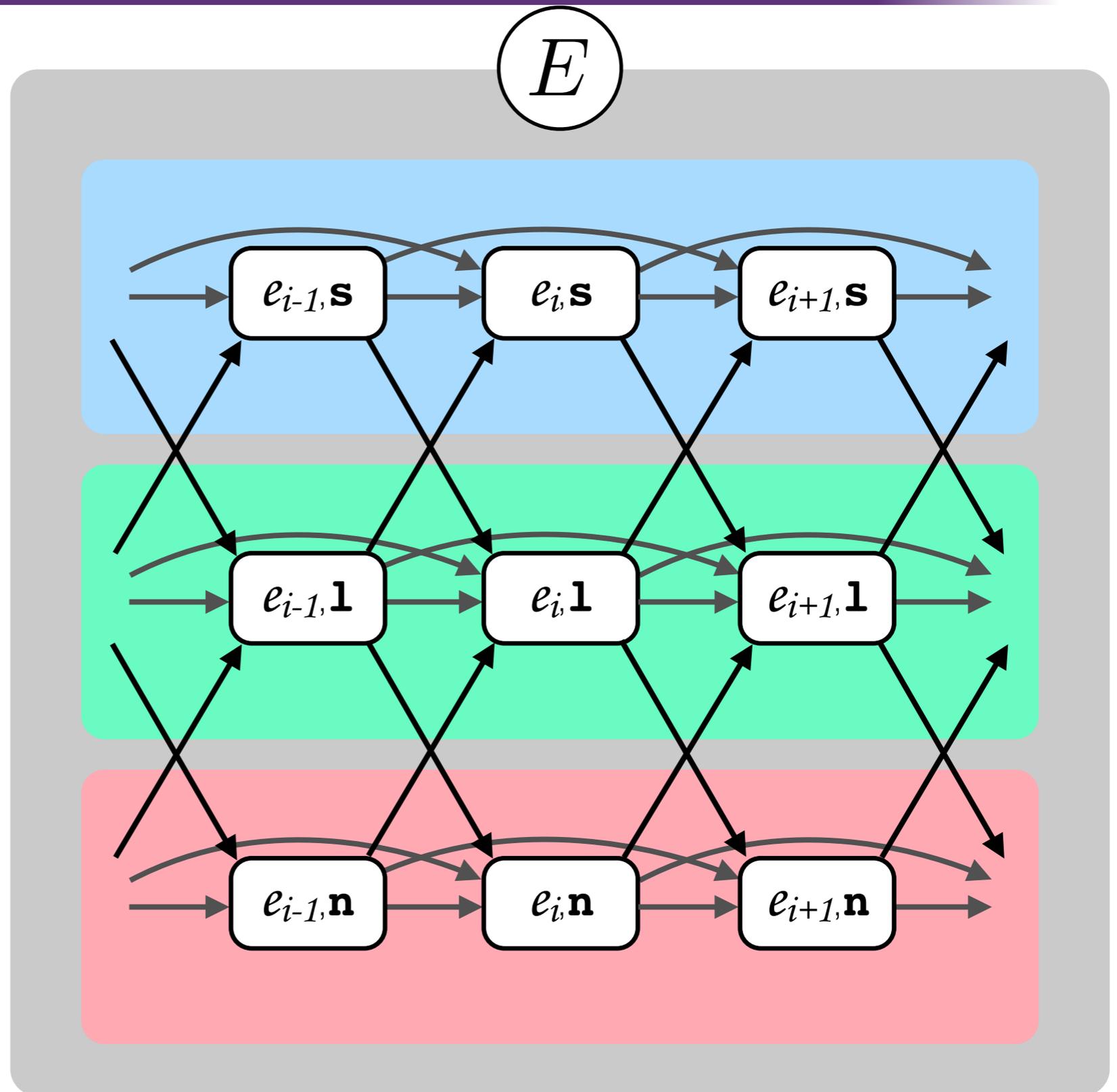


Code-Switching Language Model

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

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

nahuatl6.txt
nahuatl5.txt
nahuatl4.txt
nahuatl3.txt
nahuatl2.txt
nahuatl1.txt



Aduertencias para

como parece manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniū dāt in cælo Pater, Verbum, & Spiritus sanctus : & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas personas son vn Dios verdadero; o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzin, con que se quita toda amphibologia y duda diziendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzin, 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 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es segundo error] ça ce Dios tlahtohuani, immetihttotica, y a algūos de sus ministros les ha parecido el metehttotica, vn vocablo en si d

ixtintzitzin, con que

Spanish

AÁBCDÉFGHIÍJKLMÑOÓPQRSTUVWXYZ
aábcdeéfgghiíjklmñoópqrstuúvwxyz
01234567890.,/\"' : ; -

Latin

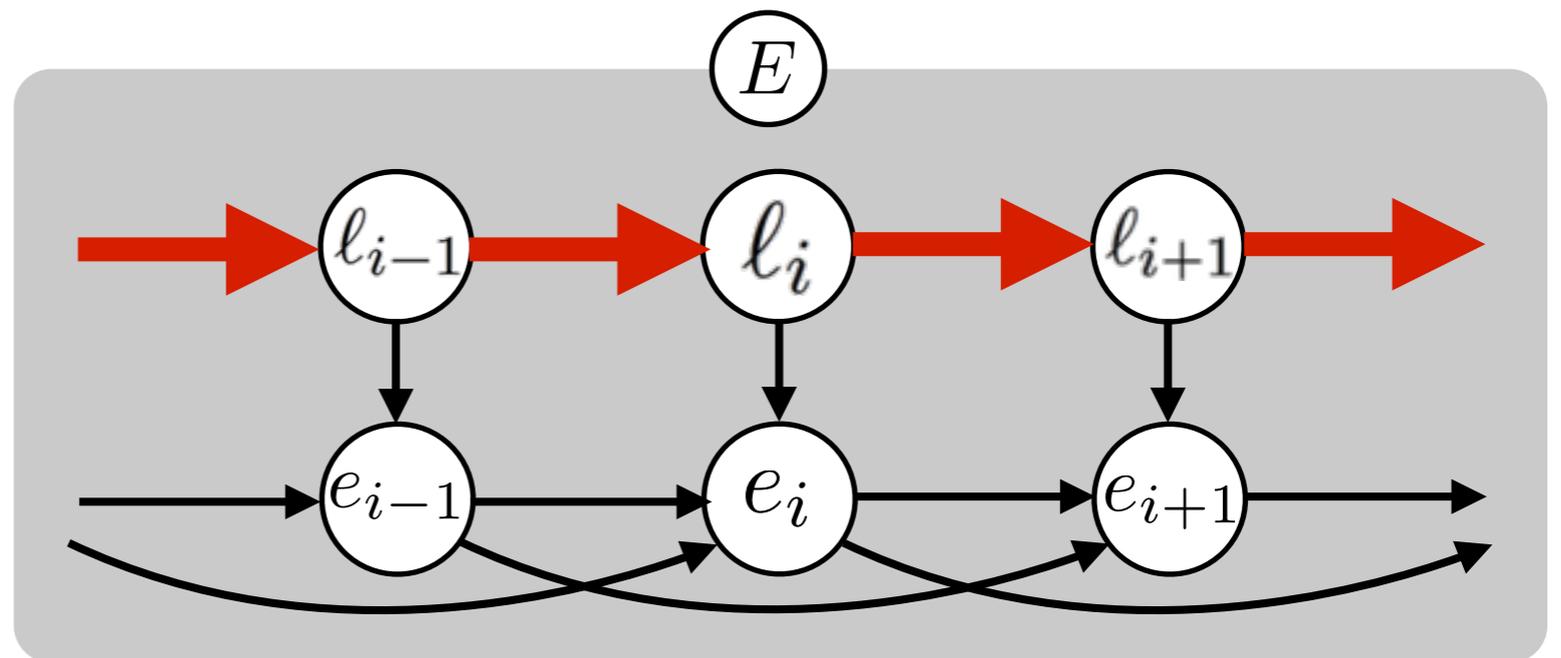
ABCDFGHIJKLMOPQRSTUVWXYZ
abcdfghijklmopqrstuvwxyz
01234567890.,/\"' : ; -

Nahuatl

ABCDFGHIJKLMOPQRSTUVWXYZ
abcdfghijklmopqrstuvwxyz
01234567890.,/\"' : ; -

Code-Switching Language Model

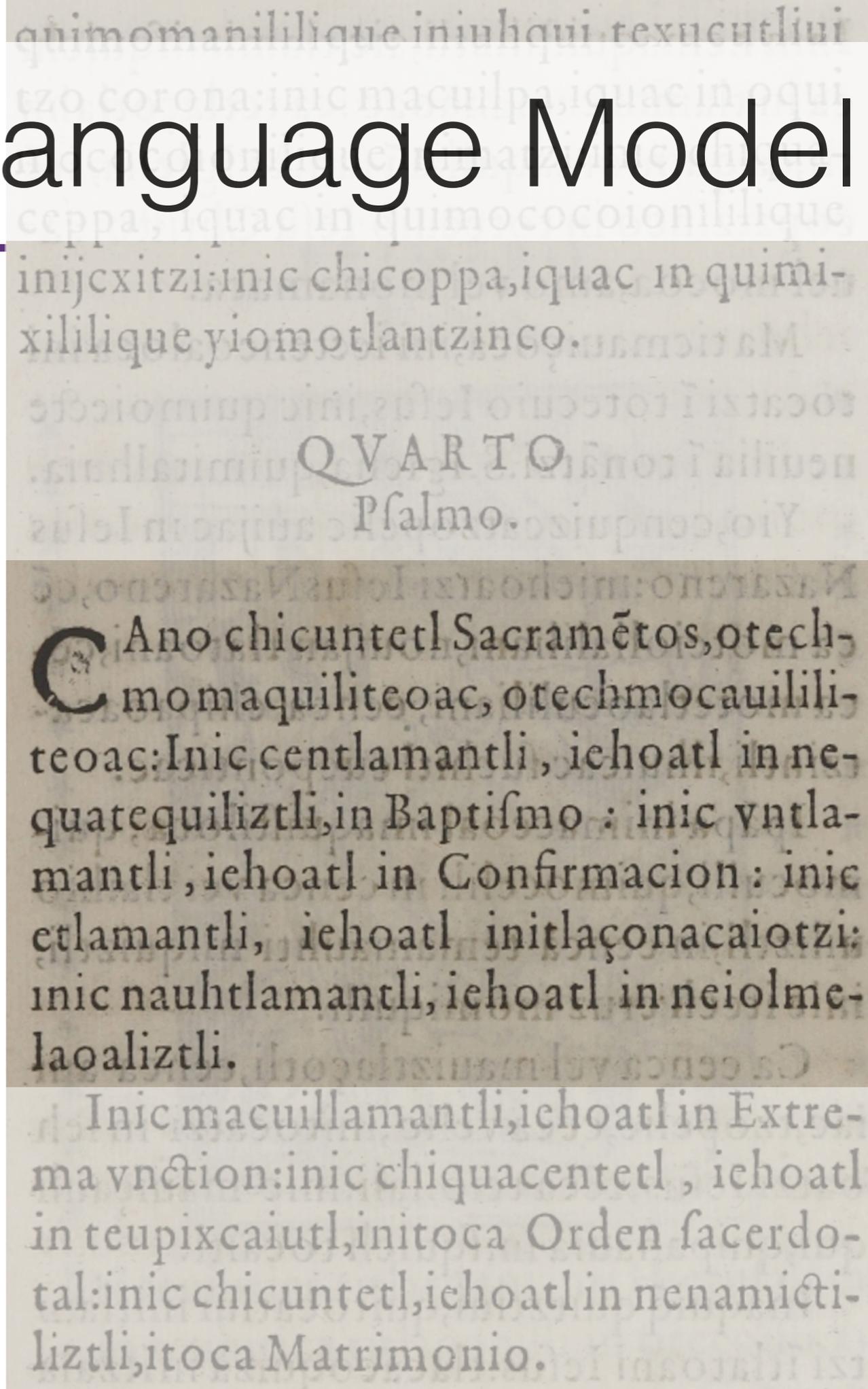
$P(l_i | l_{i-1})$ is learned unsupervised via EM, with a hyperparameter biasing the model toward **not** switching (long language spans)



Code-Switching Language Model

Code-Switching

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



Orthographic Variability

Orthographic Variability

	<u>Original form</u>	<u>Modern form</u>
dize	dize	dice
numero	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

miēcira

Orthographic Variability

merit̃ira

Without handling orth. variation: `merita`

Correct diplomatic transcription: `mẽtira`

Correct normalized form: `mentira`

Unsupervised Orthography

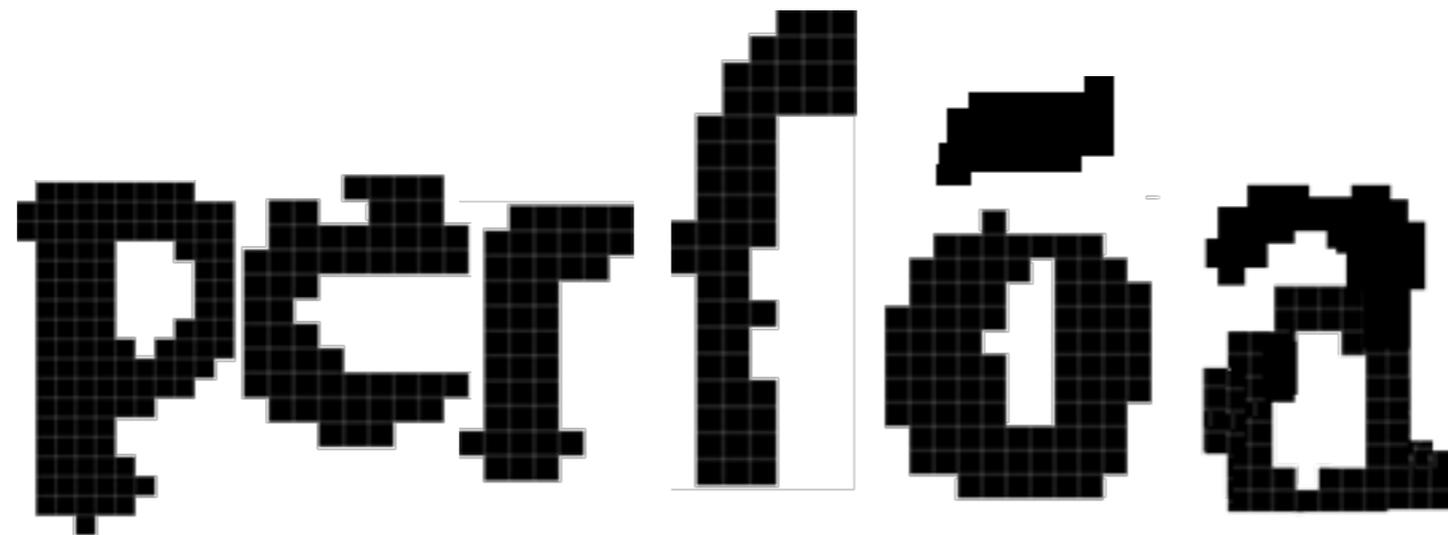
Our approach:

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

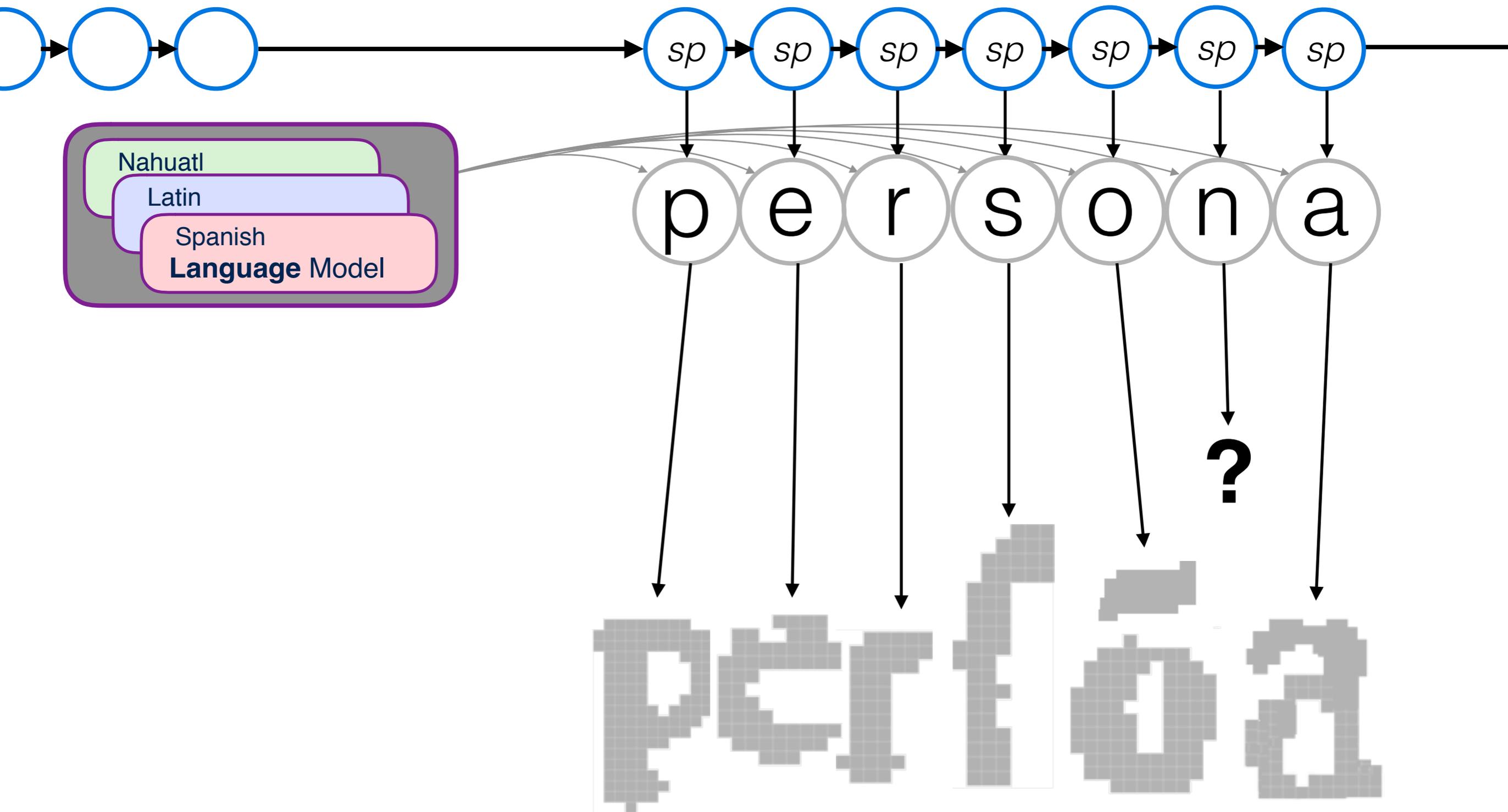
Unsupervised Orthography

Our statistical model attempts to answer the question:

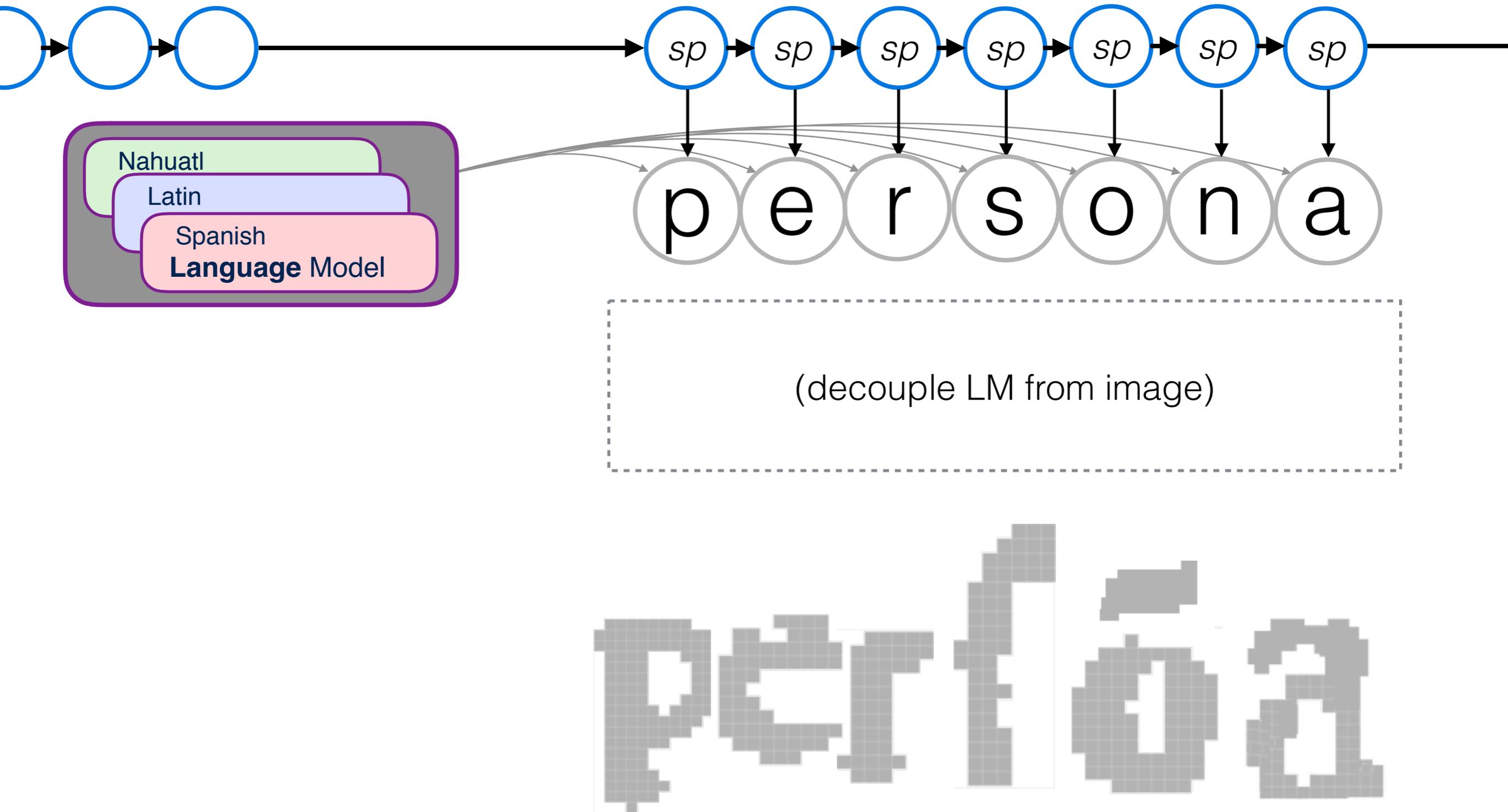
“How did these pixels end up on the page?”



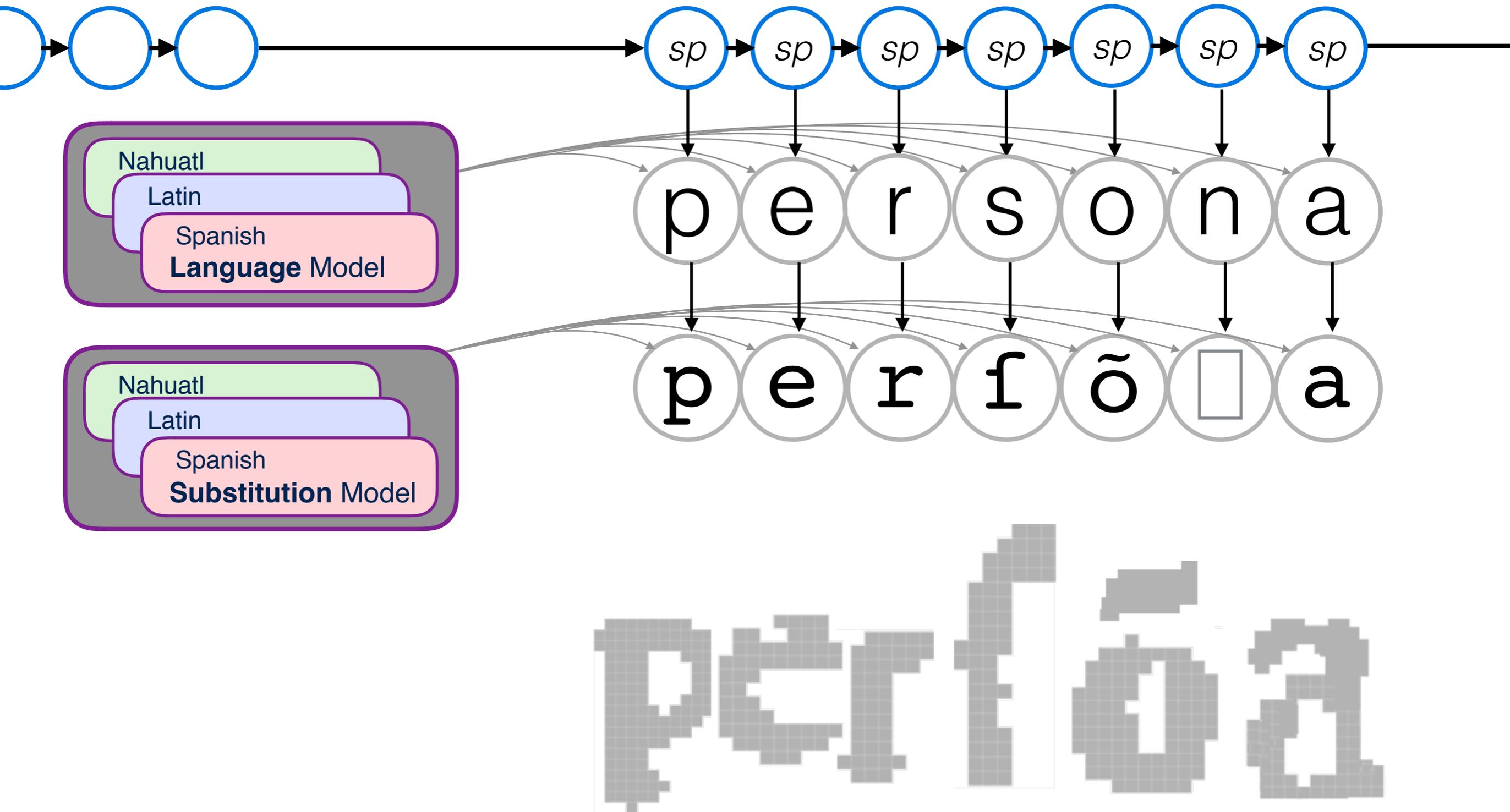
Unsupervised Orthography



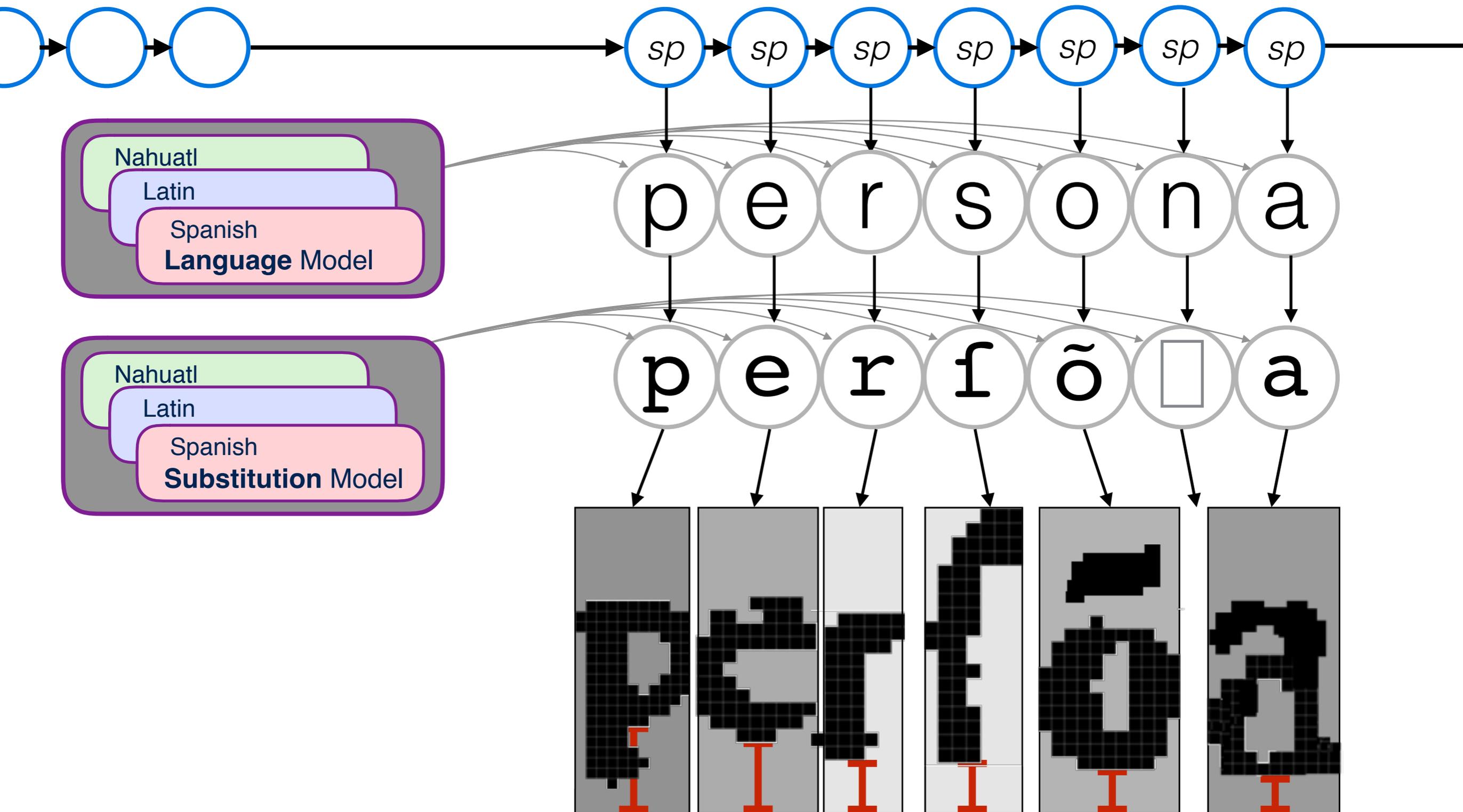
Unsupervised Orthography



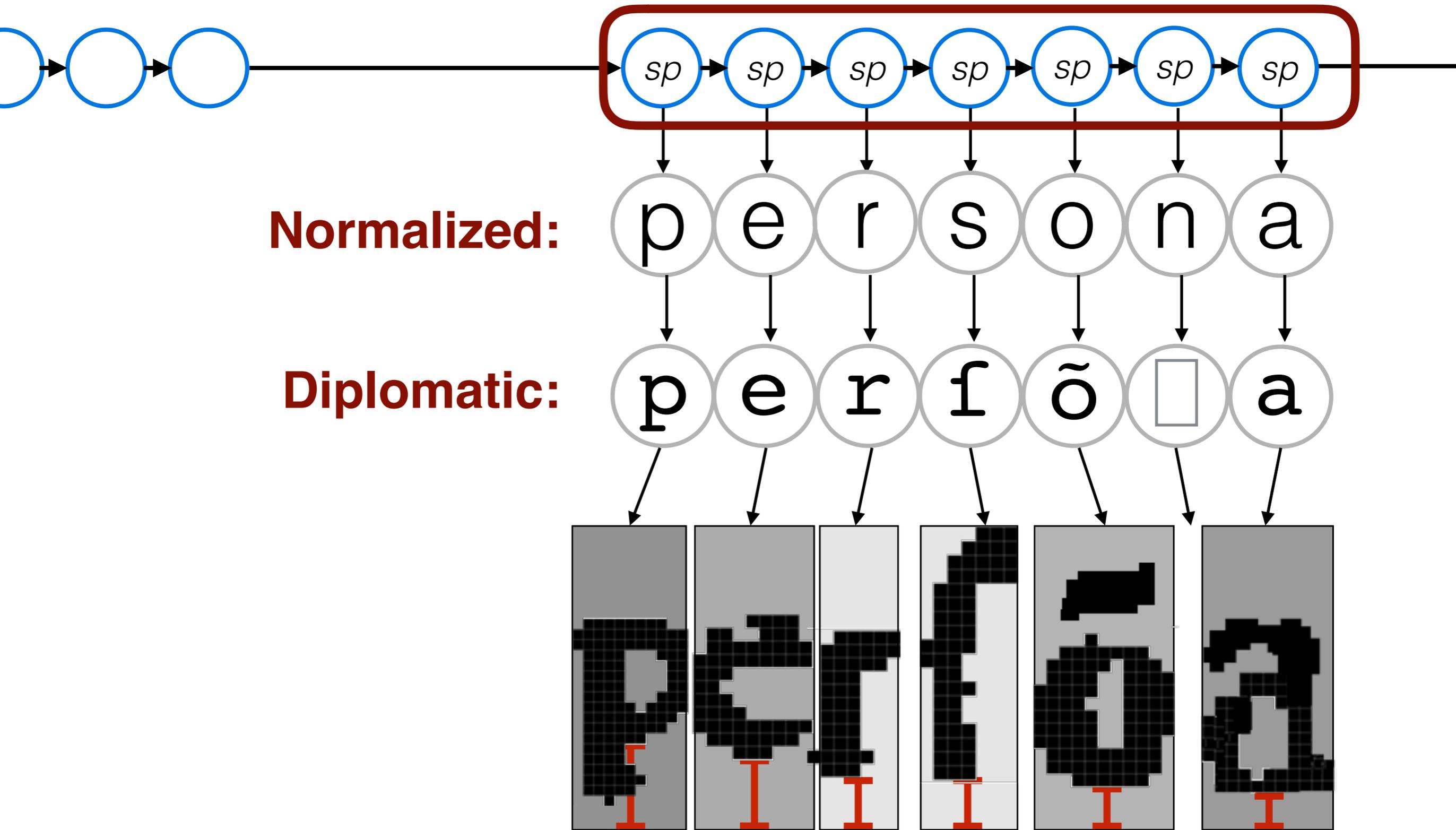
Unsupervised Orthography



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

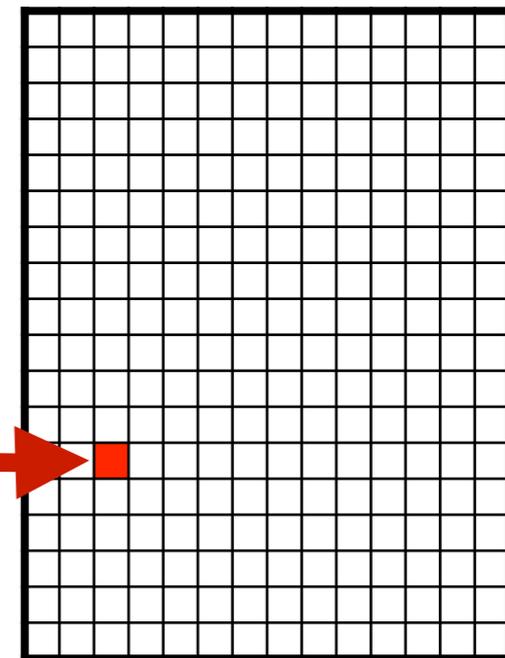
- Latinized orthography was developed by people who referred to Nahuatl as “barbaric” and “deficient” in certain sounds.
- Recently proposed orthographies seek to *decolonize*.
 - *e.g.* with tonal markings
 - and co-authoring with native speakers

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.”

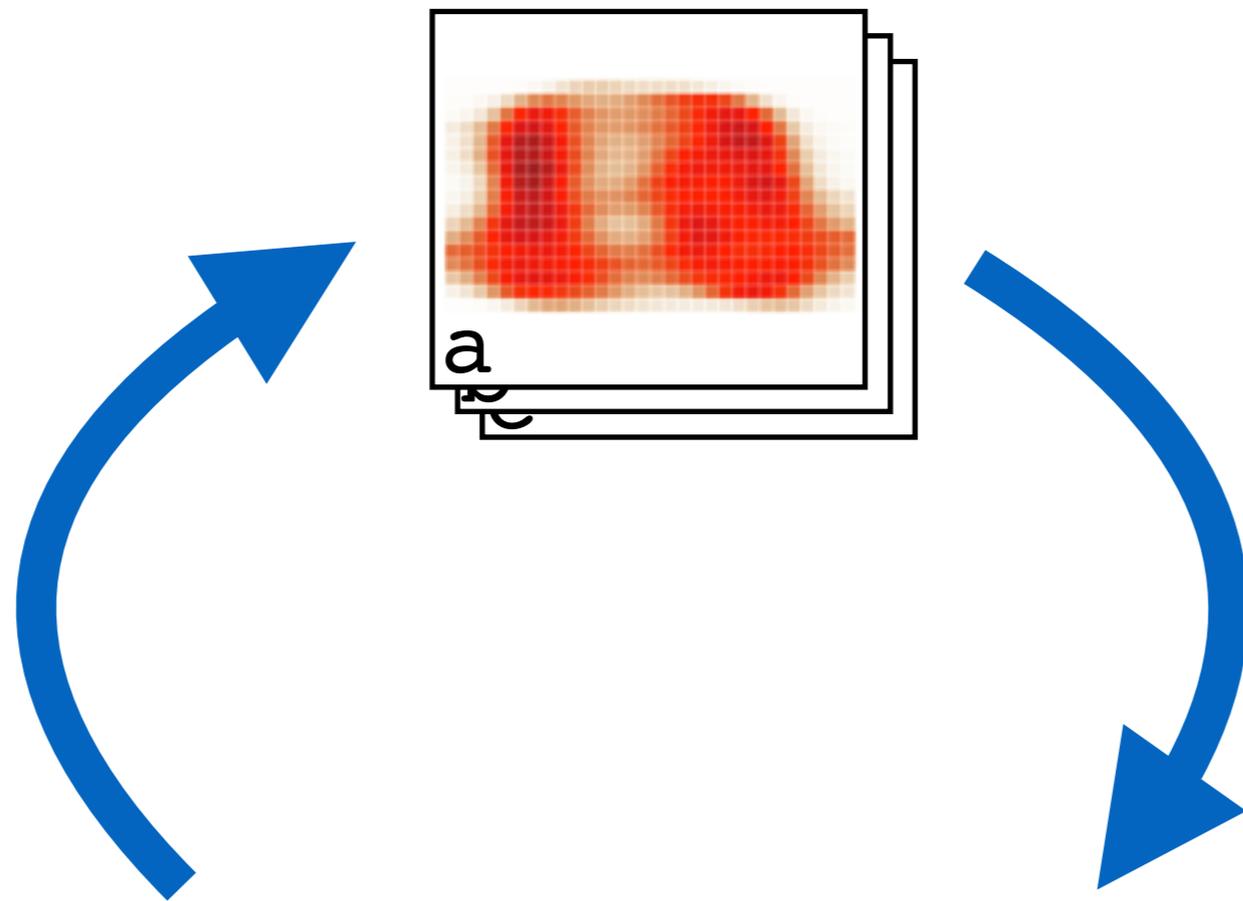
2. Update your expectations.

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

Repeat



The EM Algorithm



acoftũbrar el anima a

acoftũbr**a**r el **a**nima **a**

Experiments

Experiments

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

Orthography Baseline

Modern Spanish

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



Replacement Rules

u → v
c → z
ú → u
on → ñ
que → ñ



“Old” Spanish

spanish6b.txt
spanish5b.txt
spanish4b.txt
spanish3b.txt
spanish2b.txt
spanish1b.txt

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

Original image
No orth. handling
Our diplomatic
Our normalized

char sub. (c → q)	char sub. (s → long s)	elision (que → q̃)	accent drop (ó → o)	doubled (c → cc)
qual	eſta	aq̃l	confideracion	peccados
qual	eña	á ól	confideracion	peccados
qual	eſta	aq̃l	confideracion	peccados
cual	esta	aquel	consideración	pecados

Example Outputs

Original image

de las dos que se figuen en las quales apro-
uecha mucho acostũbrar el anima a se le-

Our diplomatic

de las dos que se figuen en las quales apro-
uecha mucho acostũbrar el anima á se le-

Our normalized

de las dos que se siguen en las quales apro-
vecha mucho acostumbrar el ánima á se le-

Example Outputs

Typographical Error

Original image

No orth. handling

Our diplomatic

Our normalized

Primeramrnte

Primeraminte

Primeramrnte

Primeramente

Example Outputs

Original image

Jefu xp̄o

Baseline

Nefuxpo

Our diplomatic

*Jefuxpo

Our normalized

*Jesuxpo

Gold diplomatic

Jefu xp̄o

Gold Normalized

Jesu Cristo

Example Outputs

Los verbales en, li, o en tli, antepo-

Model output:

sí

tĩi

← Spanish

Gold transcription:

li

tli

← Nahuatl

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

Example Outputs

tetechtlam̄ieccaquixtiliztli

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

<i>c</i>	<i>replacement</i>	<i>freq(sp., c, g)</i>	$P_{spanish}^{GLYPH}(g c)$
-	ELIDED	52	0.0881
ó	o	31	0.0526
s	f (long s)	325	0.0352
q	q̃	9	0.0222
n	ELIDED	57	0.0136
v	u	55	0.0129
o	õ	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 manifesto en las palabras de S. Ioan, que dize. Tres sunt qui testimoniū dāt in cælo Pater, Verbum, & Spiritus sanctus : & hi tres vnum sunt. i. Ioann. vltimo. Por lo qual deuen ser instruydos y enseñados, que todas tres diuinas personas son vn Dios verdadero; o reformando la sobre dicha proposicion, y añadiendo esta palabra. In huel imeixtintzin, con que se quita toda amphibologia y duda diziendo. In Dios, ca Tettatzin Tepiltzin, Spiritu sancto, ei personas, çan ce huelnelli teutl Dios in huel imeixtintzin, 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 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 in imeixtin personas me ca çan huel iceltzin teutl Dios tlahtohuani in huel imeixtin. ¶ Otros responden [y es e segundo error] ça ce Dios tlahtohuani, immetihtotica, y a algños de sus ministros les ha parecido el metihtotica, vn vocablo en si d

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. vltimo por lo qual deuen fer instruydos y enseñados, que todas tres diuinas perfonas fon vn Dios verdadero; o reformando la fobre dicha propof; ficion, y añadiendo ella palabra. in huel ini: ixtintzitzin, con que fé quita toda amphibologia y duda diziendo. In Dios. ca Tettatzin Tepiltzin, Spiritu sancto, et perfonas,) an (: huel nelli teutl Dios in huel imeixtintzitzin. q. d. Dios es Padre, Hijo, y Spũ sancto tres perfonas, vn folo Dios verdadero todas tres, cõ la qual reduplicacion se quita toda dubda. También se quitá con estas propossiciones. In D;f O S.ca Tettatzin, Tepiltzin, Spirituf aa cfo, can huel iceltzin teutl Dios tlahtohuani. in Dios. ca Tettatzin, Tepiltzin, Spiritu sancto, in ieixtin. perfonas can huel iceltzin Dios tlahtohuani. Ca in imeixtin perfonasme can huel iceltzin teutl Dios tlahtohuani in-huel imeixtin. ¶ Otros respondenj y es é segundo error j ca ce Dios tlahtohuani, im iut tĩtfittoti cá, y á algunos de fus miniftros les fu parefcido 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 propos; sicion, y añadiendo ella palabra. in huel ini: ixtintzitzin, con que sé quita toda amphibología 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 Spũ sancto tres personas, un solo Dios verdadero todas tres, cõ 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 tlahtohuani. in Dios. ca Tettatzin, Tepiltzin, Spiritu sancto, in ieixtin. personas can huel icaltzin Dios tlahtohuani. Ca in imeyxtin personasme can huel icaltzin teutl Dios tlahtohuani in-huel ineixtin. ¶ Otros respondena y es é segundo error y ca ce Dios tlahtohuani, in iuh tictteittiti cá, y á algunos de sus ministros les su parescido et merei tutorica, in vocablo envid- 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>