

# Filología digital: transcripción automática de mss, colación asistida y creación de corpus textuales

Mónica Castillo Lluch  
Andrea Escobar Castillo  
José Manuel Frajedas Rueda  
Francisco Gago Jover

Curso de verano, 2024

**UIMP** Universidad Internacional  
Menéndez Pelayo

## Curso: Filología digital: transcripción automática de manuscritos, colación asistida y creación de corpus textuales

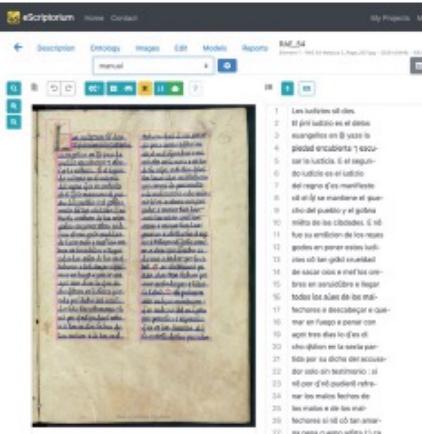
HORA*	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES
9h00 - 10h30	Introducción al curso (Mónica Castillo Lluch)	Transcripción automática de manuscritos medievales (Mónica Castillo Lluch)	Colación asistida de diversas versiones textuales con CollateX (Mónica Castillo Lluch)	7PartidasDigital: edición de mss e impresos con Transkribus (José Manuel Frajedas Rueda)	Introducción a la creación y aprovechamiento de los corpus textuales (Francisco Gago Jover)
10h30 - 12h	Filología tradicional y digital (Mónica Castillo Lluch)	Transcripción automática de manuscritos modernos (Mónica Castillo Lluch)	Colación asistida de diversas versiones textuales con CollateX (Andrea Escobar Castillo)	7PartidasDigital: análisis estilométrico (José Manuel Frajedas Rueda)	Creación de corpus textuales (Francisco Gago Jover)
12h - 12h30		Descanso			
12h30 - 14h	Preparación técnica de los talleres del curso (Andrea Escobar Castillo)	Taller eScriptorium (Andrea Escobar Castillo)	Análisis de datos tras la colación con CollateX (Mónica Castillo Lluch)	Taller Transkribus (José Manuel Frajedas Rueda)	Aprovechamiento de los corpus textuales (Francisco Gago Jover)
14h - 15h		Comida			
15h - 16h30		Taller eScriptorium (Andrea Escobar Castillo)	Taller CollateX (Andrea Escobar Castillo)	Taller Transkribus (José Manuel Frajedas Rueda)	
16h30 - 18h.		Taller eScriptorium (Andrea Escobar Castillo)		Sesión de discusión: proyectos y problemas (Francisco Gago Jover)	

# Transcripción automática de mss medievales

# FUERO JUZGO digital

dirección: Mónica Castillo Lluch  
apoyo informático: Andrea Escobar Castillo  
asistentes: Orléane Chioccola y Domitille Gaillard

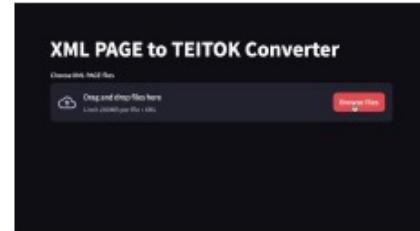
## 1 Uso de eScriptorium para transcripción semiautomática



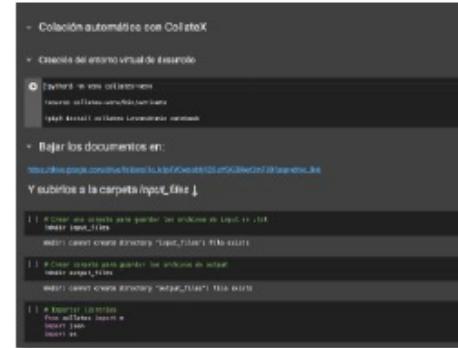
## 2 Corrección humana



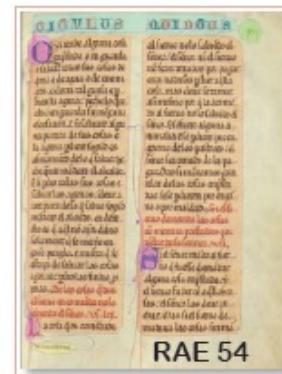
## 3 Volcado en TEI:TOK (script)



## 4 Colación asistida para comparar versiones (CollateX-Notebook)



El proyecto Fuero Juzgo digital pretende superar la falta de conocimiento de este texto importante para la historia de la lengua y del derecho, mediante la producción y puesta en línea de ediciones de sus testimonios antiguos más interesantes. La comparación de los testimonios permitirá explorar los contrastes entre las dos recensiones, fernandina y alfonso, y mostrará las diferentes formas de la recensión alfonso, conservadas en manuscritos de los siglos XIII al XV.



RAE 54

En junio de 2024 hemos finalizado la transcripción de RAE 54 (ca. 1260, recensión alfonso) con el programa eScriptorium y revisión humana. Su subida a TEI:TOK está prevista antes de finales de este año.



## Problemas de transcripción de mss medievales y modernos:

- Segmentación
- Las abreviaturas, ¿cómo transcribirlas? ¿desarrollarlas?
- ¿Modelo grafemático o alográfico?

mss medievales

Existencia de modelos entrenados  
con gran volumen de datos y de  
lenguas diversas

mss modernos

inexistencia de modelos



¿Ventajas e inconvenientes de transcribir en detalle las abreviaturas?

Versión del 10.05.24

## Abreviaturas

**⚠️ No deben copiar y pegar ningún signo de este documento en [eScriptorium](#)**  
En verde las abreviaturas muy convencionales (ya desde el latín y de palabras muy frecuentes o del ámbito religioso)

	Facsímil	Abreviatura	Significado	Combinación + Unicode (NFC)
1.		<u>ñ</u>	a + nasal (an/ am)	a + \u0303 (nasal o tilde horizontal)
2.		<u>ñ<sup>s</sup></u>		\u01DE4 (s superpuesta)
3.		b	<u>bra</u>	
4.		b	<u>ber</u>	\u0180
5.		b	<u>bre</u>	b + \u0364 (e superpuesta)
6.		<u>p</u>	<u>per</u>	\u0127 (barra horizontal)
7.		<u>p</u>	<u>san&gt;c&lt;b&gt;a/o</u>	c + \u0303 (nasal o tilde horizontal)
8.		<u>c</u>	<u>cri</u>	c + \u0365 (i superpuesta)
9.		<u>c</u>	<u>cro</u>	
10.		d	<u>der</u>	
11.	 	d	<u>der/ don</u>	\u0111
12.		<u>e</u>	<u>e + nasal (en/ern)</u>	e + \u0303 (nasal o tilde horizontal)
13.		<u>e<sup>s</sup></u>	<u>es</u>	\u01DE4 (s superpuesta)
14.	 	<u>g</u>	<u>gua / gca (falta ejemplo)</u>	g + \u0363 (a superpuesta)

Line #27



tel iuez. fueras end si el sē-

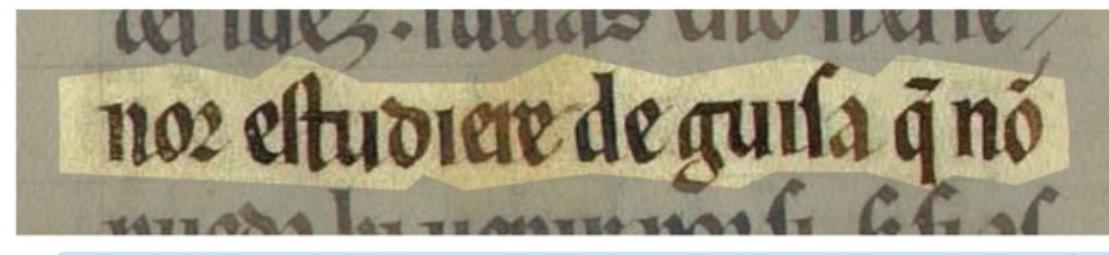
by Orléane (eScriptorium) on Wed Dec 06 2023 21:26:06 GMT+0100

+Toggle history

?

Line #28

Line #28



nor estudiere de guisa q nō

by Orléane (eScriptorium) on Wed Dec 06 2023 21:26:06 GMT+0100

+Toggle history

?

Folio 79, líneas 27-28: corte de la palabra *señor*



¿Ventajas e inconvenientes de transcribir en detalle las abreviaturas?

Versión del 10.05.24

## Abreviaturas

**⚠️ No deben copiar y pegar ningún signo de este documento en [eScriptorium](#)**  
En verde las abreviaturas muy convencionales (ya desde el latín y de palabras muy frecuentes o del ámbito religioso)

	Facsímil	Abreviatura	Significado	Combinación + Unicode (NFC)
1.		<u>ñ</u>	a + nasal (an/ am)	a + \u0303 (nasal o tilde horizontal)
2.		<u>ñ<sup>s</sup></u>		\u01DE4 (s superpuesta)
3.		b	<u>bra</u>	
4.		b	<u>ber</u>	\u0180
5.		b	<u>bre</u>	b + \u0364 (e superpuesta)
6.		<u>p</u>	<u>per</u>	\u0127 (barra horizontal)
7.		<u>p</u>	<u>san&gt;c&lt;b&gt;a/o</u>	c + \u0303 (nasal o tilde horizontal)
8.		<u>c</u>	<u>cri</u>	c + \u0365 (i superpuesta)
9.		<u>c</u>	<u>cro</u>	
10.		d	<u>der</u>	
11.	 	d	<u>der/ don</u>	\u0111
12.		<u>e</u>	<u>e + nasal (en/ern)</u>	e + \u0303 (nasal o tilde horizontal)
13.		<u>e<sup>s</sup></u>	<u>es</u>	\u01DE4 (s superpuesta)
14.	 	<u>g</u>	<u>gua / gca (falta ejemplo)</u>	g + \u0363 (a superpuesta)

# Modelos de transcripción de 2 tipos:

Grafemático

The screenshot shows a digital transcription interface. At the top, there are navigation buttons and a title "Line #44". Below this is a photograph of a medieval manuscript page with text in a Gothic script. Overlaid on the bottom half of the image is a modern transcription of the text: "ga lo el obispo o el alcalde". Below this transcription is a table with numbered entries (42 to 46) and their corresponding text in Spanish. At the very bottom of the interface is a keyboard.

dadera miétre d' seran tenu- dos por falsos. fueras en- de si lo dexeré por engaño dotorarre. o por manado do del Rey q'lo no pudie- ron fazer saber fasta aquella dia. o por otra coytá. De-	42	cado. E lo q'dixieré estos si- eruos por su iuraméto: fa- ga lo el obispo o el alcalde
mado por ellos e	43	escriuir después. esa cōfir- mado por ellos e por el Rey.
Que la manda e	44	
deue seer mosti- cado. E lo q'dixiere estos	45	
Obispo e ante l	46	
mas. e compri- mido por ellos e		

Alográfico

The screenshot shows a digital transcription interface. At the top, there are navigation buttons and a title "Line #44". Below this is a photograph of a medieval manuscript page with text in a Gothic script. Overlaid on the bottom half of the image is a modern transcription of the text: "ga lo el obispo o el alcalde". Below this transcription is a keyboard labeled "RAE54" and a "Manage keyboards" button. To the right of the transcription is a column of numbered entries (12 to 22) with their corresponding text in Spanish. The text is written in a style that matches the original manuscript.

12	ecleros q'el muerto. E q' los deftigos cleu lo fazer la
13	bera aq'llos hereleros q' lo
14	en la manda fafta vi. me
15	fes. E fi lo nō fizieré fafta
16	vi. mefes ono cupieren
17	lo q' es clicho enefta ley fâ
18	ca aq'l tiempo. l'epanuer
19	clabera miétre q'feran denu
20	bos por falfos fueras en
21	de fi lo glerare por engano
22	

¿Ventajas e inconvenientes?

# Modelo alográfico

## Ventajas:

- se conserva información desechada normalmente (p. ej. posición de los alógrafos) que:
  - permitiría la identificación de perfiles de escribas individuales o colectivos
  - puede ayudar a datar el ms.

(Camps 2016, Gille Levenson 2023)

## Inconvenientes:

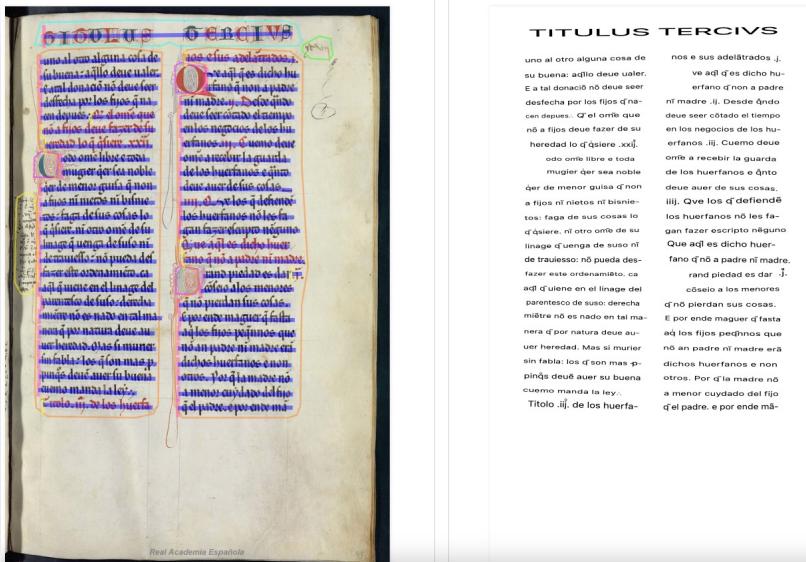
- Riesgo de perder la interoperabilidad de los datos
- Supone más trabajo tras la transcripción y tasa de error superior (más clases → más errores)

(Gille Levenson 2023)



# ¿Qué modelo elegir: grafemático o alográfico?

- tensión entre editar (mejor modelo grafemático) y datar (mejor modelo alográfico)



# Transcripción grafemática completa de RAE 54

Alográfico								Documents	Edit	Reports	<a href="#">Create new Document</a>			
Select	Preview	Name	Owner	Last modified	Image count	Tags	Actions							
<input type="checkbox"/>		RAE293	Orléane	June 27, 2024	4 images.									
<input type="checkbox"/>		Murcia	Orléane	June 27, 2024	6 images.									
<input type="checkbox"/>		Z-III-6	Orléane	June 27, 2024	7 images.									
<input type="checkbox"/>		RAE 54	Orléane	June 27, 2024	8 images.									

## Transcripción alográfica de algunos folios

**C**orenmas: - las él pueblo xpuies. iii.

**D**ios q fiso todas las cosas. creó  
y confecho la cibeg del dñ de jeso. q  
yo nacer dela cibeg todas las oídas para  
dios todo meublos del mundo el dñ. Once  
qz ello el dñha cateca. pr q los oíos me-  
mos comunes. qz nacer della. y forma en  
la cibeg la lúbar dlo oíos. pr q pudieles  
oír nacer las oídas qle pudieles emprecar.  
y forma en ella la memoria q entender. y  
pr q pudieles gobernlar tozcaner los oíos  
miembras dlo cuerpo. qson sometidas. pr  
ello lo maestros q son fabres aní han cu-  
ta cel mal dela cibeg. q todos oíos miembros  
del cuerpo. y por entz la melenina el ante.  
pr q entente el maestro q qz mayor penglo.  
q si la cibeg es sana ancianos ensi pr  
q poden sanar todos lo oíos miembros. mas  
si la cibeg fuer enferma ní poden sanar sa-  
lud alos oíos miembros q nlo a ensi. y por  
entz tenemos pma memoria entaner lo que  
ches los énques. por qson más categorias.  
y dentear su uña y su salutre. q despeza  
to oxigar las colas del pueblo. q en mi-  
ent q el el Rey el consalte. q por qno fiz  
me menen defenzer lo pueblos. **D**e toll  
la cobrida dlo énques. y como éne secr  
ches los énques ensi nobr.  
**E**n la éna. o el senor estonc semiea q  
ama lass das oídas celestiales. qmz apri-  
gar lo gremios. y quellles tener puelo.  
One auenur suele q mayor pl gana dela  
salz dlo oíos q dela suuu. ca quanto lo oíos son  
mas tanto mayor ganancia suele ave-  
nir dlos. q quanto el espia d'auer pl de

Si mismo. sola menen no semiea mucho  
de ganar bñ hñ de si mismo q es un oí  
solo. Et por esto queremos aguantar la sa-  
luz dlo oíos el pueblo. mas qz uno sola  
menen. Ens q el énque semiea q ama la  
saluz dlo pueblo sola menen pr palabra.  
y no por fecho. que entre lo q rueda el pu-  
eblo. q estone aní el puelo el pueblo.  
qnto entiendes qlo oíos dlo qz al remando  
y q gela oxigena. One como lo pmares aní  
estones mas colgadotes te rebas el pueblo  
en lo qz q son passados. y de crescentar  
el su dlo leño. y nos entremos agarra la melqui-  
dad dlo sometidas q la grá de dlo spu. pu-  
es q dimes las leyes al sometidas q tomen  
sen. qmno poner freno. q dimes ala obedi-  
cia dlo énques. Et por ello establecemos  
alli pr nos como pr todos nros successio-  
nes. q nñq Rey no constriega pr freca a  
nñq dñ q laga escrito dlo roda q de-  
ne a entz. ni mane q gela laga bas por  
freno. ni por q nñqmo pienla q dñ rero.  
ni qz su voluntad las oídas qz qz reue. Et  
si algñ dñ qster par algunas oíadas al éna  
y pr sua voluntad. o el énque ganar al  
guna oíada pr algñ algo qfiso. tea enel  
escrito puelo q gela da y pr su voluntad y  
como gela da. y q gela da. y pr q nñq  
pueda q entende si gela da pr freca.  
o qz engano del énque. y si pudier q  
entender q gela da q su voluntad. o gela  
diese el énque. desfaga lo q fiso mal.  
xpuies q su muerte convine las oíadas.  
ad q gelas pierda. o dlo heredero. q aquella  
oíadas q fueron todas al énque sin nñqda

Z-III-6\_def.txt

Qv di 19 < > Fermer Remplace

Delas mādas daqllos q̄ uan en romeria o en hueste: y los alcança la muerte. cō mo deuen feer firmadas. .xij.

q̄l q̄ muere en romia o en hueste. si quiere oms libres cōfígo eſcūa fu māda cō fu mano aſtelloſ. o fi nō ſopiere eſcūir. o nō aſiere o nō pidiſre por enfermedat faga fu māda aſte fg fieruos q̄ ſepa el obispo q̄ fon de buena fe. y q̄ nō fueſſen aſte falladoſ II[II]

en peccado. y lo q̄ dixiera eftos fieruos por fu iuramēto antel alc ſagalo el obispo. o el alc eſcūir deſpues y ſea oſfirmado por ellos y por el rey.

Q̄ la māda del muerto duee feer moſtrada antel obispo y ante las teſtimoiias y cōplida faſta .vj. meſes. .xiij.

a māda del muerto q̄ es fecha por eſc̄ptro. faſta .vj. meſes duee feer manifeſtada aſtel obispo. y ſi algū ōe la aſcōdiere por engaño. y la nō aſiere moſtrar. peche otro tāto deſto fuyo a aqllos q̄ deuié auer aqlla māda. q̄nto les diera el muerto enella.

Delos eſc̄ptos q̄ fon dubdofos cōmo deue feer cōcerrađos cō otros eſc̄ptos. .xiiiij.

os pleytos y los eſc̄ptos q̄ fon fechos. y aql q̄ los fiſo faž. y las teſtimoiias fon muertas fi pareſſer fu ſenal dellos o del muerto en el eſc̄ptro. deue el alc catar las ots ſenales. y los ots eſc̄ptos q̄ ellos fizieroſ cōcerraſ aql eſc̄ptro cō los ots eſc̄ptos. y deue abonðar tres eſc̄ptos. o q̄tro q̄ ſeā ſemeiableſ. A aql por puar aql. fueras ende filos tiēpos en q̄ fuerō fechadas las leyes dižē q̄ aqllos eſc̄ptos nō deue ualer.

Delos eſc̄ptos oſtracioſ y fon las mādas q̄ eſ- cuē los oms cuyas fon por fus manos. .xv.

or q̄ los oſ an coyta alas uezes. y nō pueđe cōplir las leyes. Por ende enlos logares o ōe pueđe non L

fallar tātos oſ cōmo māda la ley. cada un ōe duee eſcūir fu māda cō fu mano. y diga eſpecial miétre q̄ māda faž de fg cofas. o aden māda ordeñar fg cofas. o aden las māda. y notar y el dia y el año. en q̄ fažē la māda. Et pues q̄ todo esto ouiere eſc̄ptro. eſcūa enfondō dela carta q̄ lo ſirma cō fu mano. Et deſpues q̄los hedos o fus fijos fallare esta māda fecha. faſta .xxv. años mueſtrē la al obispo dela T̄ra. o al alc faſta .vj. meſes. y el obpo. o el alc tomē tales .iii. eſc̄ptos q̄ fuerō fechos por fu mano daql q̄ fiſo la māda. Et por aqllos iii. eſc̄ptos ſiſfe ſemeiare la leſ dela māda ſea cōfirmada la māda. Et pues q̄ todo esto ouiere conoſciđo el obispo y el alc. o ots teſtimoiias oſfirmē el eſc̄ptro dela māda oſ uez.

## Búsqueda a ojo

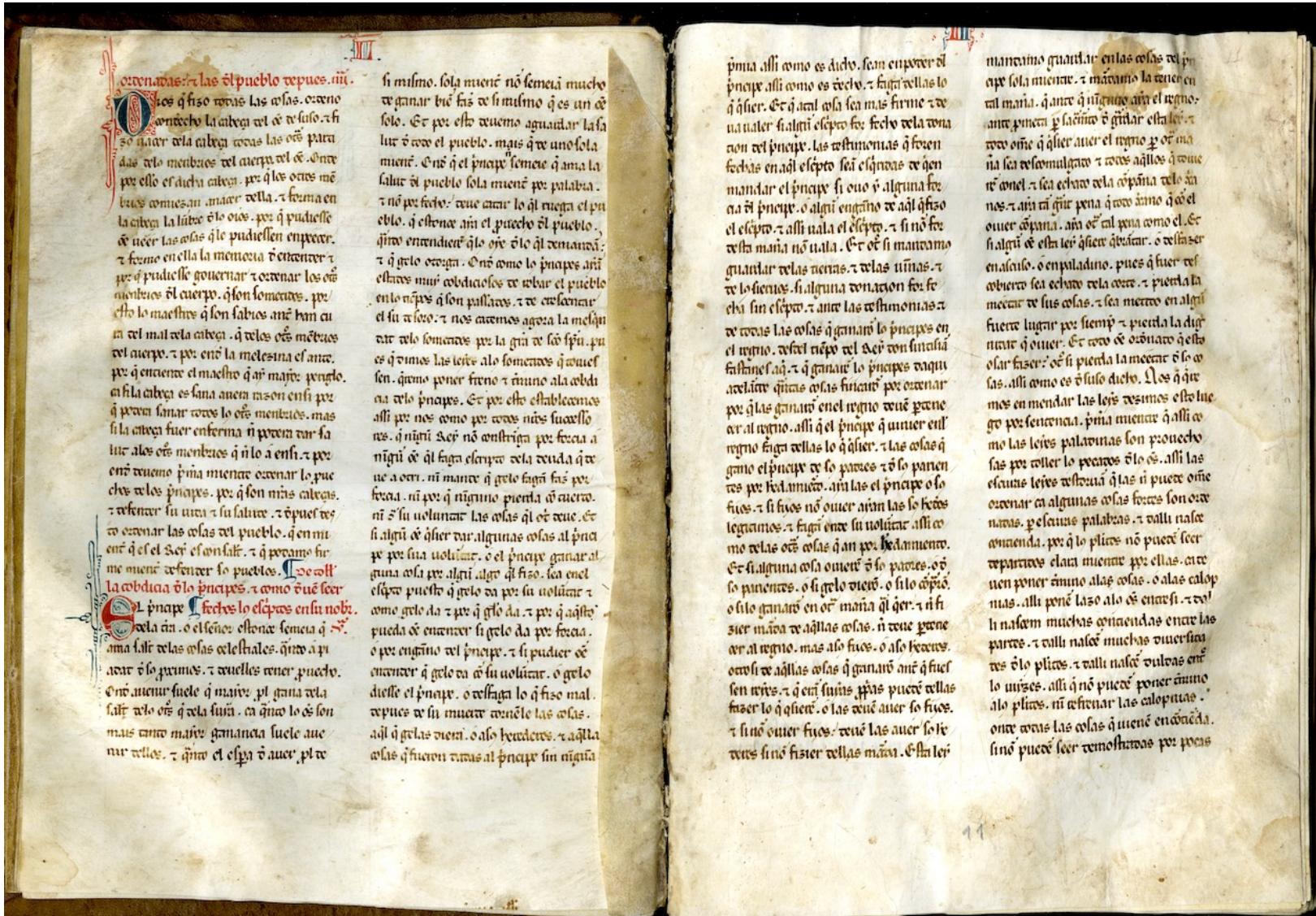
## Búsqueda automática

# Problema de la datación de los mss del FJ

Propuestas de datación

Documento de trabajo MCL

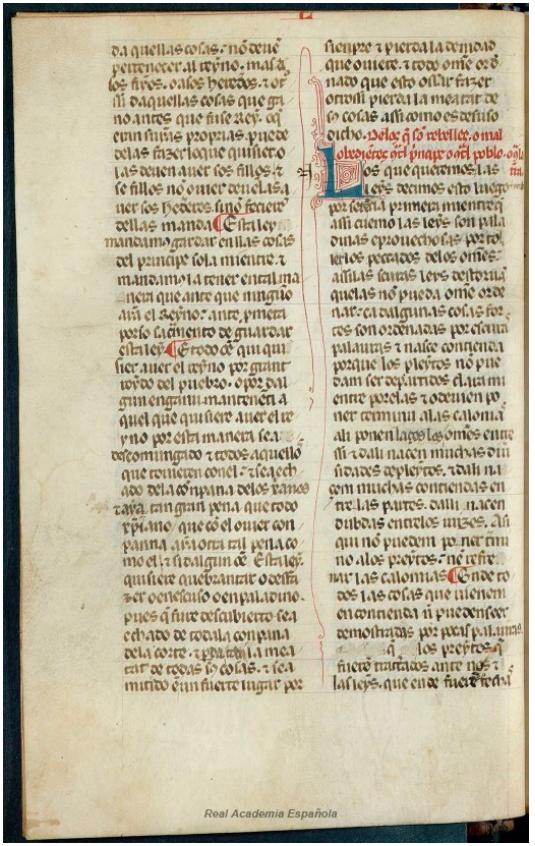
1	RAE 49 (avant Campomanes) Madrid En ligne sur la <i>Colección digital</i> de la RAE	<b>PhB</b> (2021): 4, manid 1345, cnum 451 Transmitir a PhB: <ul style="list-style-type: none"> <li>denominación antigua: Campomanes (RAE 1815)</li> <li>corregir:</li> </ul> <p><b>Copied</b> 1260 ca. a quo - 1300 ca. (Simón)</p> <ul style="list-style-type: none"> <li>Galindo y de Vera (1863) le croit antérieur au ms. de Murcie</li> </ul>	Occidental	[PhB : 1260 ca. a quo – 1300] Galindo y de Vera (1863) le croit antérieur au ms. de Murcie <u>ERD: 1243 post quem</u> (1260- 1270) RRP : 1250s
2	RAE 51 (avant Béjar) Madrid En ligne sur la <i>Colección digital</i> de la RAE		Occidental	[PhB : 1260 ca. a quo – 1300] <u>ERD: 1260-1270</u>
3	RAE 53 (avant Malpica 1 = Gondomar 2) Madrid En ligne sur la <i>Colección digital</i> de la RAE CC2021: copista Pedro Mtnez <u>Gallego</u> (en colofón "Pedro Martiz Gualego me scripsit" parte antigua a la que le faltan folios / parte moderna s. XVI a. p. de Z-III-6 --> rama beta en su segunda parte (moderna, que copia Z-III-6) MCL 23.6.23: Los folios modernos de la Hispanic (ms. B2713) copian este ms.			[PhB : 1260 ca. a quo – 1300] <u>ERD : 1246-1260</u> RRP : 1240s
4	RAE 54 (avant Malpica 2 = Gondomar 1) Madrid En ligne sur la <i>Colección digital</i> de la RAE <a href="https://bibliotecavirtualmadrid.comunidad.madrid/bv_madrid_publicacion/es/catalogo_imagenes/grupo.do?">https://bibliotecavirtualmadrid.comunidad.madrid/bv_madrid_publicacion/es/catalogo_imagenes/grupo.do?</a>			[PhB : 1260 ca. a quo – 1300] <u>CC2021: ad quem 1260</u> <u>ERD: 1260 ad quem</u> RRP: ¿1260-1280?



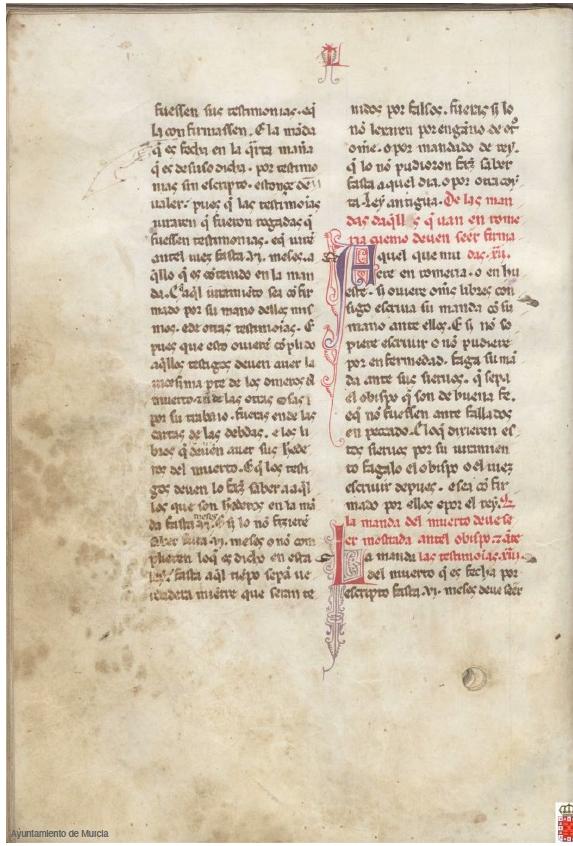
# Rasgos tempranos

- *above top line*
  - di (inicial e interior)
  - ðo, ðe + do, de
  - fusión irregular de curvas (ca, ðe, oe, pc)
  - Morfología avanzada pero no la cadena gráfica

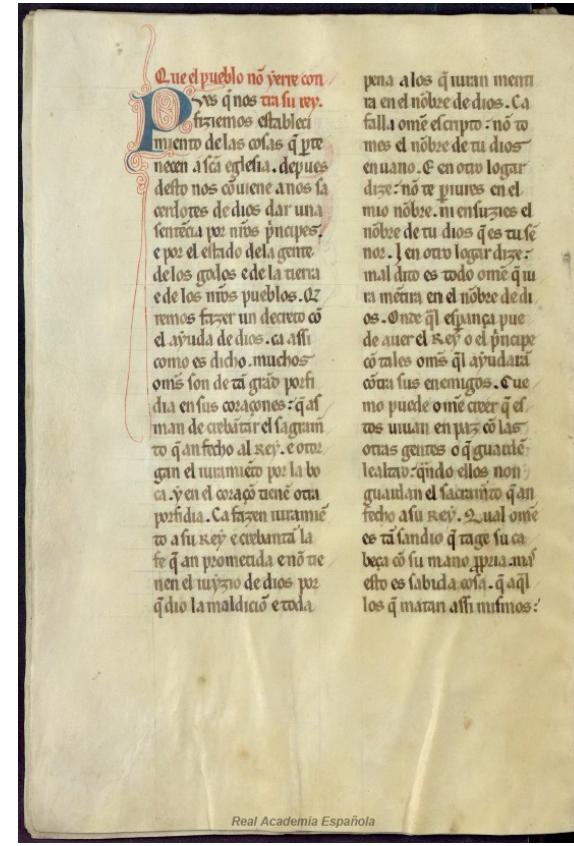
→ ca 1260 (ERD)



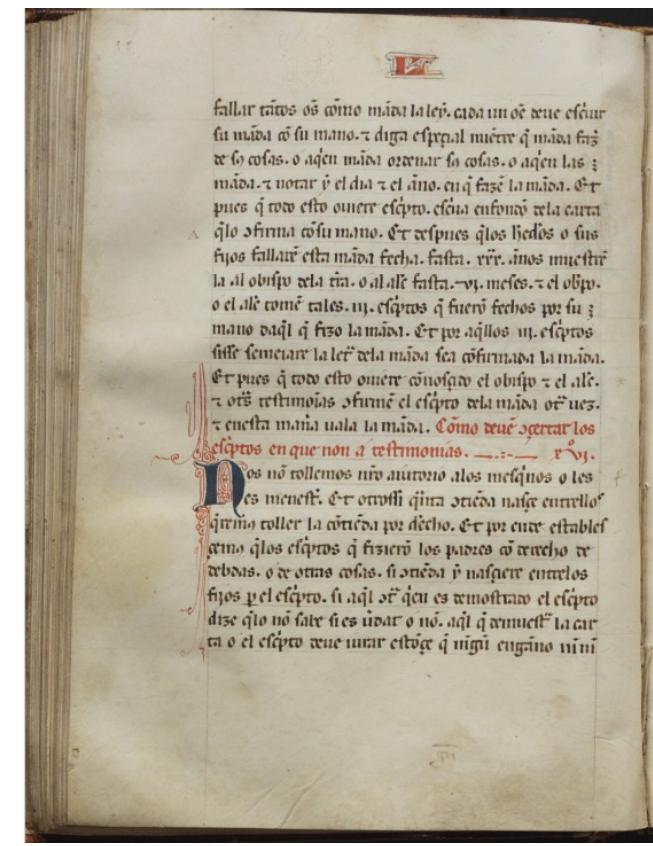
RAE 293  
datado 1289



Murcia  
dataable 1288  
(ERD: datado)



RAE 54  
ca 1260



Z-III-6  
1290-1300

## Investigaciones recientes sobre datación de mss castellanos y leoneses ss. XII y XIII

Rodríguez Díaz, Elena E. (2024), «Elementos para fechar los códices leoneses y castellanos según los manuscritos datados (ss. XII y XIII)», en Ángeles Romero Cambrón (ed.): *La ley de los godos. Estudios selectos*. Bern, Peter Lang.

Inés Fernández-Ordóñez (2023), Comunicación sobre variación gráfica en los códices alfonsíes e inmediatamente anteriores y posteriores en el *Encontro Internacional Forais medievais peninsulares: questões históricas e linguísticas*, Lisboa, 29/11/2023.

Inés Fernández-Ordóñez (2024), «Los fueros como documentos. La primera transmisión del Fuero real», conferencia en el VIII Congreso Internacional de la Red CHARTA *La historia del español a través de sus documentos: coordenadas geográficas y sociales de la variación lingüística*, Palma, 21/06/2024.

# Estudio de las alternancias entre d/ð y r/ɹ

ERD (2024):

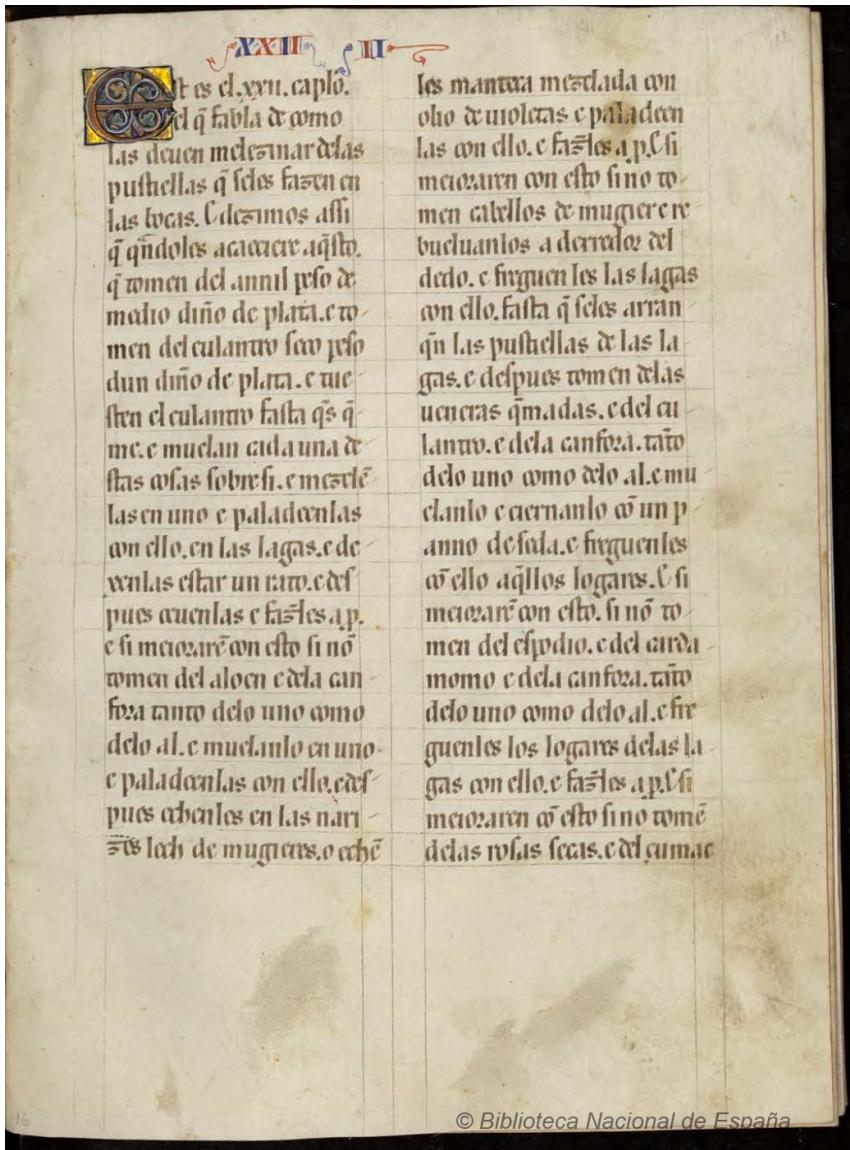
- *d* “se usó de forma cada vez más restringida a partir de 1256 [...] En los casos más tardíos el uso de la *d* recta es más frecuente en los grupos *di/u/n/r*, es decir, seguida de letras cuyo primer trazo es vertical”.
- Se usa con frecuencia *d* hasta 1246 y se usa delante de letras curvas hasta 1253 y 1255
- *d* disminuye en la década de 1260 para casi desaparecer a principios de la década de 1270
- *ð* se usó sistemáticamente en posición inicial desde 1259
- *ðe / ðo* con nexo desde mediados del s. XIII
- *ð*: 1250-1259 empieza a predominar y desde 1259 es claramente dominante
- 1256 → *ð*, con restos de *d* + letras de palo hasta principios de los 1270
- *or* desde 1165
- *pr, br, ðr* hasta 1256-1259, *desde 1260 pɹ, bɹ, ðɹ*

# Características arcaizantes en los códices prealfonsíes

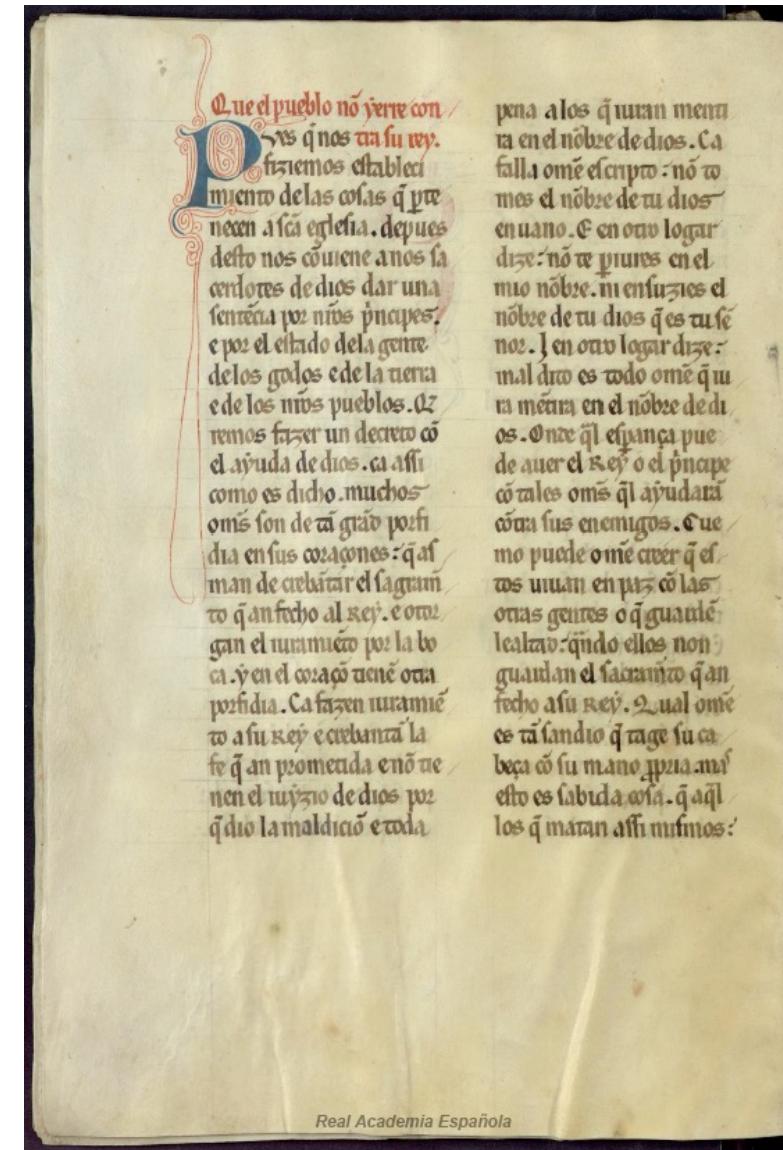
Título	di	ði	du	ðu	da	ða	do	ðo	de	ðe	o+r	o+r	d+r	ð+r	b+r	b+r	p+r	p+r	Número de palabras
<i>Animalias</i> 1250	34 = 100%	-	9 = 100%	-	25 = 92%	2 = 7%	13 = 61%	8 = 38%	112 = 67%	54 = 32%	X	X			X		X		1223
RAE54																			
Z-III-6...																			

*Libro de las animalias:* La uncial comienza ante <o> y <e>. Los contornos más resistentes son ante <i> e <u> con <a> en posición cercana.

La <r> en forma de 2 parece comenzar tras <o>, mientras que la <r> recta se mantiene en <dr, pr, br>. Se escribe encima de la primera línea del pautado.



Libro de los animales que cazan  
1250



RAE 54  
ca 1260

Título	di	ði	du	ðu	da	ða	do	ðo	de	ðe	o+r	o+r	d+r	ð+r	b+r	b+r	p+r	p+r	Número de palabras
<i>Animalias</i> 1250	34 = 100%	-	9 = 100%	-	25 = 92%	2 = 7%	13 = 61%	8 = 38%	112 = 67%	54 = 32%	X	X		X		X			1223
RAE54	13 = 52%	12 = 48%	5 = 100%	-	13 = 20,63%	50 = 79,37%	4 = 17,39%	19 = 82,61%	76 = 79,17%	20 = 20,83%	-	46 = 100%	4 = 44,44%	5 = 55,56%	-	4 = 100%	-	4 = 100 %	1337
Z-III-6	19 = 76%	6 = 24%	-	3 = 100%	-	63 = 100%	-	22 = 100%	-	86 = 100%	-	40 = 100%	-	9 = 100%	-	3 = 100%	-	-	1232
Murcia	-	27 = 100%	-	4 = 100%	-	59 = 100%	-	23 = 100%	-	91 = 100%	-	46 = 100%	-	5 = 100%	-	4 = 100%	-	1 = 100%	1230
RAE293	-	25 = 100%	-	5 = 100%	1= 2%	49 = 98%	-	20 = 100%	-	88 = 100%	-	29 = 100%	-	9 = 100%	-	4 = 100%	-	3 = 100%	1036

- RAE54 presenta aún usos antiguos (en verde) de *d*, pero en menores proporciones que *Animalias*; pierde r y generaliza r̥, excepto en 4 casos *dr*) → datación ca 1260
- Z-III-6 presenta usos modernos, pero *di* es rasgo antiguo
- M y RAE293 presentan perfil más moderno, coherente con las fechas de 1288 y 1289.

Título	di	ði			du			ðu			da			ða			do			ðo			de			ðe			d+r							
RAE 54	13			12			7			-			13			50			4			19			76			20			4			5		
	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I	PP	PL	I						
	11	-	2	2	1	9	3	-	4	-	-	-	7	3	3	1	3	46	3	-	1	0	0	19	69	1	6	1	2	17	-	4	-	0	0	5
	84,62%	1,5, 2,8 %		25%	75 %		42,86%	57,1 4 %		-	-		76,92%	23,0 8 %		8%	92 %	75%	25 %	0%	10,0 %	92,11%	7,89 %		15%	85 %	100%	0 %	0 %	100 %						
Z-III-6	19			6			No hay alternancia																													
	P P	PL	I	PP	PL	I																														
	9	-	10	5	-	1																														
	47,37%		52,6 3%	83,33%		16,6 7 %																														
RAE 293	Apenas hay alternancia. La única d recta aparece en la secuencia da en interior de palabra (Representa 2% del total de los 49 casos de da y 2,56% del total de los 39 casos de da interior)																																			

PP: Principio de palabra

PL: Principio de línea con cesura

I: Interior de palabra

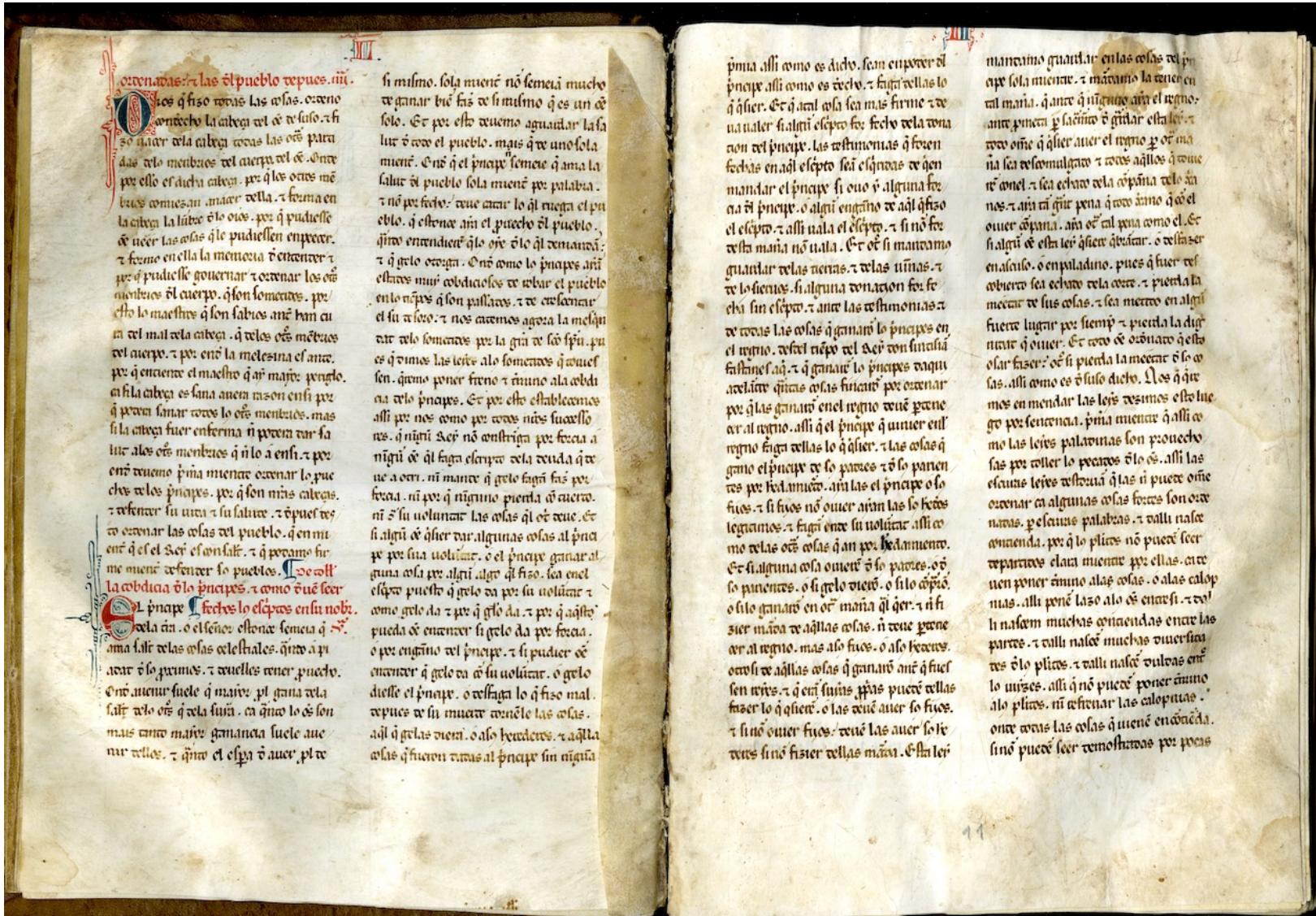
# Detalles del contexto gráfico: RAE54 y Z-III-6

## RAE54

- *Di* en interior precedido de *u* (*udi* 2 veces VS *uði* 3 veces)
- *Du* en interior precedido de 1 *a*, 1 *e*
- *Da* en interior precedido de 1 *r*, 1 *e*, 1 *o*
- *Do* en interior precedido de *b* (*bdo* 1 vez VS *bðo* ninguna vez)
- *De* en interior precedido de 1 *a*, 1 *p*, 4 veces *e* (*ede* 4 veces VS *eðe* 4 veces)

## Z-II-6

- *Di* en interior precedido de *u* (*udi* 4 veces VS *uði* ninguna vez)
  - ✓ (*fdi* 3 veces VS *fði* ninguna vez)
  - a* (*adi* 2 veces VS *aði* 1 vez)
  - o* (*odi* 1 vez VS *oði* ninguna vez)



# Rasgos tempranos

- *above top line*
  - di (inicial e interior)
  - ðo, ðe + do, de
  - fusión irregular de curvas (ca, ðe, oe, pc)
  - Morfología avanzada pero no la cadena gráfica

→ ca 1260 (ERD)

# HSA B2567: ¿varias manos? ¿si una mano, razón de esta variación?

Título	di	ði	du	ðu	da	ða	do	ðo	de	ðe	o+r	o+r	d+r	ð+r	b+r	b+r	p+r	p+r	Número de palabras
B2567: 11	10 = 100%	-	-	-	8 = 36,36%	14 = 63,64%	-	8 = 100%	1 = 1,92%	51 = 98,08%	-	30 = 100%	-	-	-	8 = 100%	11 = 100%	-	1105
B2567: 21	12 = 50 %	12 = 50 %	1 = 50%	1 = 50%	9 = 27,27%	24 = 72,73%	-	25 = 100%	7 = 7,45%	87 = 92,55%	1 = 2,63%	37 = 97,37%	-	1 = 100%	-	7 = 100%	-	1 = 100%	962
B2567: 25-26	-	25 = 100%	-	3 = 100%	1 = 1,72%	57 = 98,28 %	-	28 = 100%	-	97 = 100%	-	49 = 100%	-	-	-	3 = 100%	-	-	1166



Share

## PhiloBiblon 2024 n. 4 (mayo). Los manuscritos del *Fuero Juzgo*: denominaciones antiguas y signaturas modernas

Posted on May 8, 2024 by Charles Faulhaber

Mónica Castillo Lluch  
(Université de Lausanne)Carmen del Camino Martínez  
(Universidad de Sevilla)

Tan nocivos como el fuego y otros desastres naturales pueden resultar para el conocimiento de los textos antiguos los cambios de denominación de los manuscritos. La confusión que produce un nuevo nombre o firma en un testimonio puede obstaculizar, o incluso impedir, el avance de la investigación, por lo que a filólogos e historiadores se nos impone, como primer paso para el estudio de una tradición manuscrita, la aclaración de las concordancias entre las denominaciones de cada pieza, mutantes a lo largo del tiempo. Descifrar esos códigos puede ser una tarea más o menos exigente en función de la edad de los manuscritos: cuanto más antiguo es un testimonio, mayor cantidad de mudanzas ha podido sufrir de una biblioteca a otra, y más cambios posibles de firmas. Si además tratamos con tradiciones complejas, compuestas de abundantes testimonios, la dificultad se multiplica.

El caso de la tradición manuscrita del *Fuero Juzgo* ([BETA texid 1191](#)) reúne todas las características que lo hacen complejo en cuanto a las denominaciones de los testimonios: está constituida por un número elevado de testimonios (50), de los cuales 36 (más 3 fragmentos) son medievales, que han cambiado de poseedores y que se han editado y estudiado haciendo uso de designaciones que ya no corresponden a las de hoy. En consecuencia, muchas preguntas importantes han estado en el aire durante décadas, incluso durante siglos, sobre la tradición manuscrita y las ediciones existentes del *Fuero Juzgo*: ¿qué manuscrito editó Villadiego en 1600? ¿A qué manuscritos corresponden exactamente las denominaciones que utilizó la RAE en su edición de 1815? ¿Están en el aparato de la edición académica de 1815 los manuscritos que editaron Mencé (1996) u Orazi (1997)? Por muy sorprendente que pueda parecer, nadie hasta la fecha ha intentado establecer una lista completa de correspondencias entre las denominaciones antiguas y las firmas modernas de los testimonios del *Fuero juzgo*, como tampoco se ha identificado el manuscrito editado por Villadiego en 1600, y ni siquiera en las ediciones de Mencé y de Orazi se llega a indicar a qué denominaciones académicas (RAE 1815) correspondían los manuscritos escurialenses editados por ambas autoras (Z-III-6 y Z-III-21 respectivamente).

En PhiloBiblon hasta ahora [abril/mayo de 2024] sólo se presentaban tres concordancias entre denominaciones antiguas y firmas modernas, extraídas del catálogo de Zarco Cuevas (1924, 1926, 1929). A estas referencias se limitaba casi el debate de la cuestión, al que debemos añadir otras tres informaciones. En el caso de los otros dos códices estudiados

Signatura actual	PhiloBiblon BETA manid	Denominación RAE	Denominación Burriel	Denominación Floranes	Equivalencia entre ediciones y códices de la RAE (1815)
Archivo Municipal de Murcia, Serie 3, ms. 53	manid <a href="#">1350</a>	Murcia			
RAE 49	manid <a href="#">1345</a>	Campomanes			
RAE 50	manid <a href="#">1358</a>	San Bartolomé			
RAE 51	manid <a href="#">1346</a>	Béjar			
RAE 53	manid <a href="#">1347</a>	Malpica 1		Gondomar 2	cotejado por Floranes
RAE 54	manid <a href="#">1348</a>	Malpica 2		Gondomar 1	editado por Floranes (y antes por Diego Covarrubias: ms. A 331/15 de la Universidad de Sevilla)
BNE Vitr. 17-10	manid <a href="#">4458</a>	Toledo	Toledo 4		editado por Burriel
BNE 244	manid <a href="#">1356</a>	(BR5)			
BNE 2978	manid <a href="#">1364</a>	(BR6)			
BNE 5774	manid <a href="#">3684</a>	(BR4)			
BNE 5814	manid <a href="#">1365</a>	BR2			editado por Villadiego
BNE 5975	manid <a href="#">1359</a>	BR3			
RBME d-III-18	manid <a href="#">1360</a>	Esc. 3			
RBME M-II-18	manid <a href="#">1341</a>	Esc. 5			
RBME M-III-5	manid <a href="#">1354</a>	Esc. 4			
RBME P-II-17	manid <a href="#">1353</a>	Esc. 6			
RBME Z-II-9	manid <a href="#">1361</a>	(Esc. 5 en prólogo, pero no recogido en aparato)			
RBME Z-III-6	manid <a href="#">1355</a>	Esc. 1			editado por Mencé
RBME Z-III-18	manid <a href="#">1163</a>	Esc. 2			
RBME Z-III-21	manid <a href="#">1337</a>	BR1			editado por Orazi
?		Estudios Reales			
T 43-9	manid <a href="#">1363</a>		Toledo 5		cotejado por Burriel
T 43-10	manid <a href="#">1357</a>		Toledo 6		cotejado por Burriel

**Hacia un *stemma codicum* del Fuero Juzgo desde el Humanismo hasta hoy**  
MÓNICA CASTILLO LLUCH y CHARLES MABILLE

Université de Lausanne  
ORCID: 0000-0002-8460-2492  
0000-0002-6815-7608

**Resumen:** En este estudio se expone el estado de la cuestión sobre la transmisión manuscrita y, en particular, las relaciones de parentesco entre los manuscritos del Fuero Juzgo desde el siglo XVI hasta hoy, y se presenta una propuesta de *stemma codicum* a partir de la *collatio* de cinco leyes hecha con todos los testimonios conocidos en la actualidad.

**Palabras clave:** Fuero Juzgo, historiografía, crítica textual, *stemma codicum*, *collatio codicum*.

**Abstract:** This study presents the state of the art on manuscript transmission and, in particular, the genetic relationships between the manuscripts of the Fuero Juzgo from the 16th century to the present day and offers a proposal for a *stemma codicum* based on the *collatio* of five laws made with all the currently known witnesses.

**Key words:** Fuero Juzgo, historiography, textual criticism, *stemma codicum*, *collatio codicum*.

## 1. INTRODUCCIÓN

La tradición manuscrita del Fuero Juzgo se desconoce ampliamente. Las respuestas disponibles a cuestiones tan centrales como ¿de qué testimonios se compone y a qué firmas modernas corresponden las denominaciones antiguas?, ¿cuáles fueron sus fechas y lugares de producción?, ¿en qué variedades romances están redactados?, ¿transmiten esos manuscritos una o varias versiones del texto? son parciales y a menudo inexactas. A la pregunta de ¿qué vínculo tienen estos códices con los redactados en latín que fueron su modelo? apenas aún se ha formulado respuesta<sup>1</sup>.

Este trabajo contesta sintéticamente a estos interrogantes y se plantea como objetivos primordiales exponer el estado de la cuestión sobre las filiations entre los testimonios de la ley visigótica en romance desde el Humanismo hasta la actualidad y ofrecer un primer esbozo de *stemma codicum* de toda esa tradición, basado en la *collatio* de una ley y el examen de otras cuatro teniendo en cuenta todos los manuscritos antiguos y modernos conocidos. Los contenidos se presentan en dos apartados: el primero (§2) se abre con la información esencial

<sup>1</sup> Es de notar, a este respecto, que la tradición manuscrita del Liber Iudiciorum se conoce bastante mejor que la del Fuero Juzgo, gracias al estudio fundamental de Yolanda García López (1996).

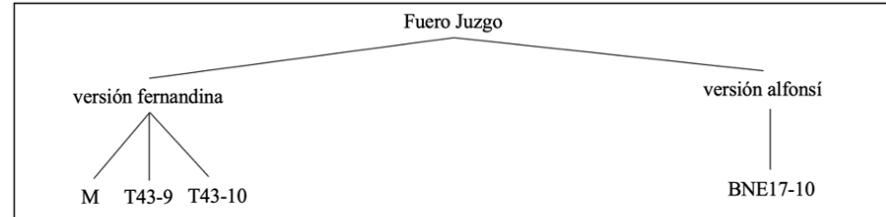


FIGURA 1  
*Stemma codicum del Fuero Juzgo según Burriel (1758).*

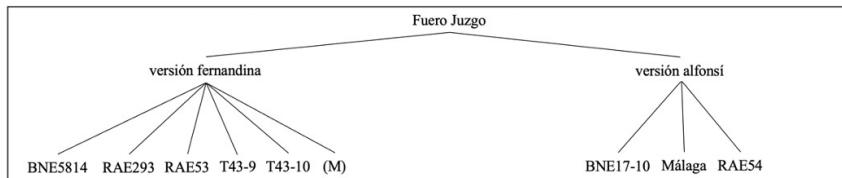


FIGURA 2  
*Stemma codicum del Fuero Juzgo según Floranes (1784), basado en la división en dos ramas propuesta por Burriel (1758).*

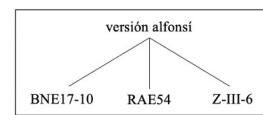


FIGURA 3  
Códices con versión alfonsí del Fuero Juzgo según RAE (1815), basándose en la división en dos ramas propuesta por Burriel (1758).

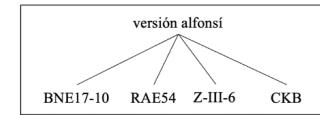


FIGURA 5  
Códices con versión alfonsí del Fuero Juzgo según Castillo Lluch (2016, 2018, 2020) y Castillo Lluch y Mabille (2021).

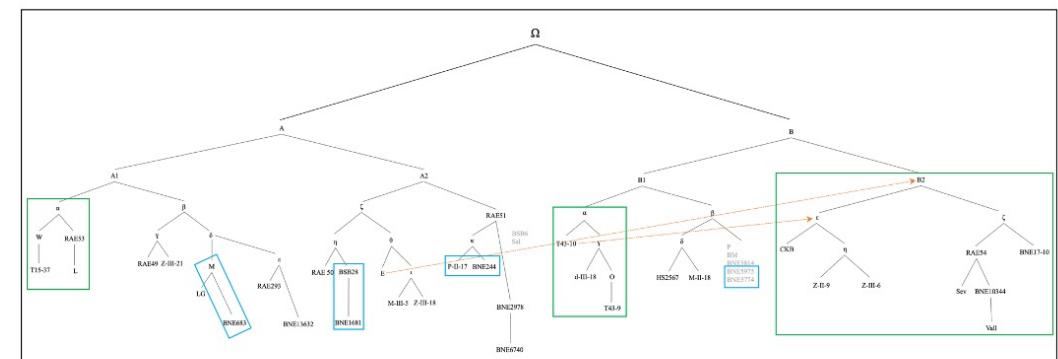
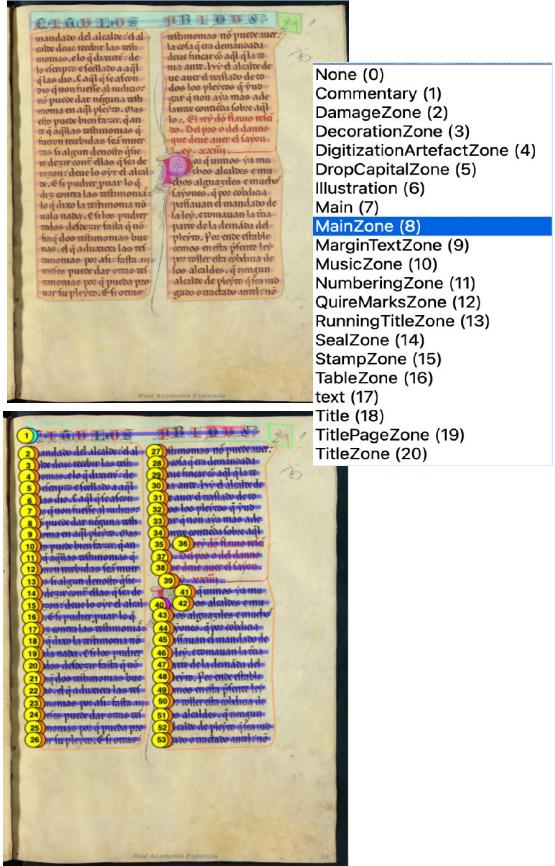


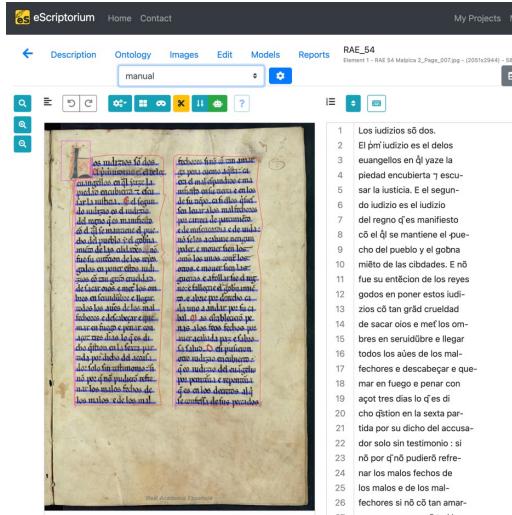
FIGURA 7  
*Esbozo de stemma codicum a partir de los lugares críticos presentes en 8.2.1, con las familias de mss que se pueden confirmar con 1.1.2, 3.5.7, 8.4.1 y 11.1.2.*

# Transcripción de RAE54 con *eScriptorium*

# Segmentación



# Transcripción automática



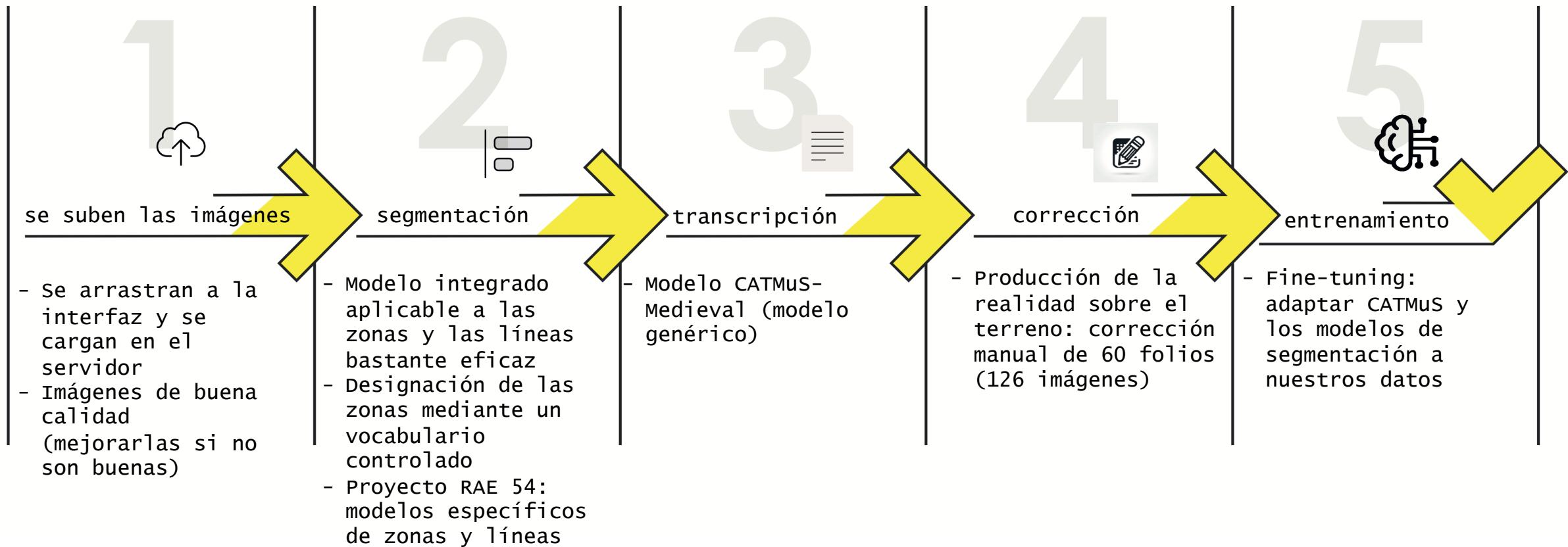
CATMuS Medieval

Modelo grafemático en acceso abierto que se puede aplicar a una variedad grande de escrituras medievales

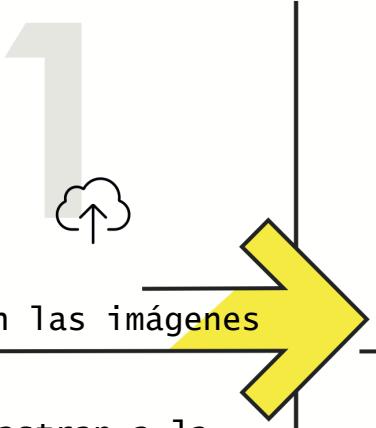
# Corrección manual



29.		p	pre	p + \u0303 (barra horizontal)
30.		p̄	pri	p + \u0365 (superpuesta)
31.		p̄	pro	\u03a753
32.		\u0304	qua	q + \u0363 (a superpuesta)
33.		\u0304/\u0304	que	q + \u0303 (nasal o tilde horizontal)
34.		\u0304	qui	q + \u0365 (superpuesta)
35.		\u0304	nifo (n<uest>ro) gras (gr<a>cas)	r + \u0303 (nasal o tilde horizontal)
36.		\u0304	ter	t + \u0309 (coma superpuesta)
37.		\u0304	tra	t + \u0363 (a superpuesta)
38.		\u0304	tri	t + \u0365 (i superpuesta)



1



se suben las imágenes

- Se arrastran a la interfaz y se cargan en el servidor
- Imágenes de buena calidad  
(mejorarlas si no son buenas)

**Description**   **Ontology**   **Images**   **Edit**   **Models**   **Reports**   ground\_truth\_RAE54

Projects   Models   Tasks

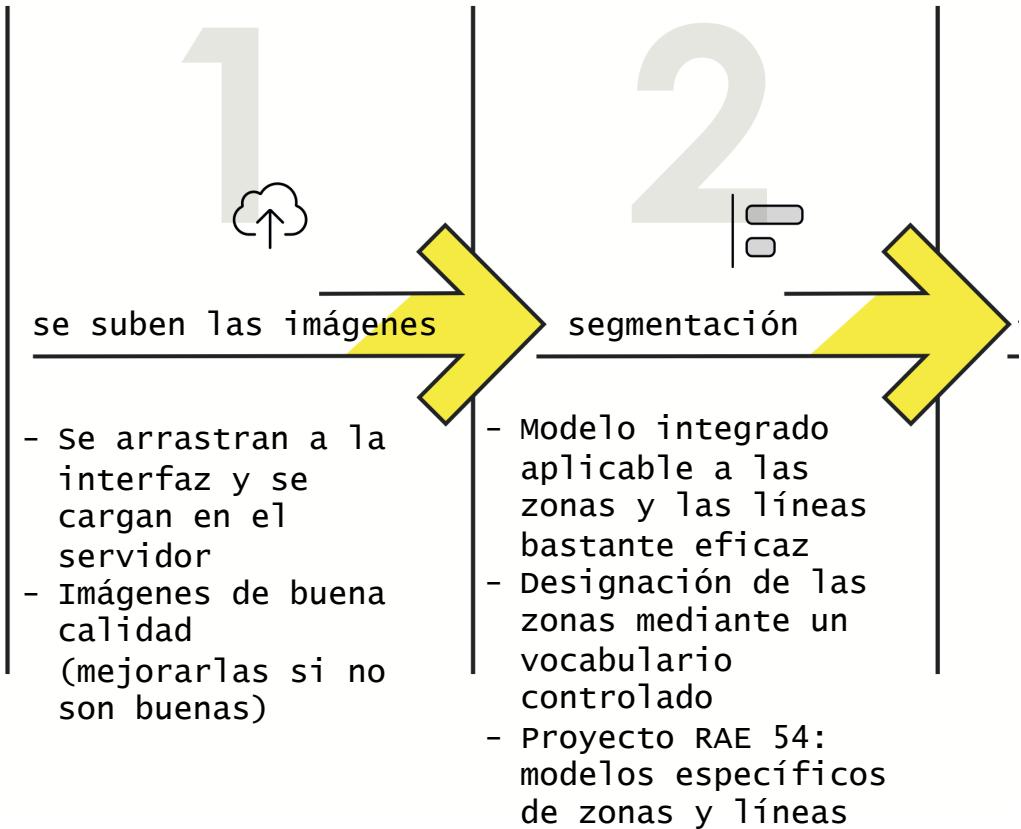
Drop images here or click to upload.

Select all   Unselect all   Selected 0/126   Import   Export   Train   Binarize   Segment   Transcribe

1   2   3   4   5   6   7

Profile

Light mode



Description    Ontology    Images    Edit    Models    Reports    METS Import | export   

Projects    Models    Tasks   

ImageRegion

Real Academia Española

on cuidado del amo de x e co grād diligencia de dñ mandado muy gñfo Rey despanha t de fracia: todos Obispnos aueramos en nōbre de nro señor en uno en la ciudad de toledo. q pue el mādado del Rey epor

su ensēnamēto fiziessemos todos comunal mientre un tctado de las cosas de s̄a eglēsia e de sus establecimētos. q pue mera miétre nos todos diemos grās al nō saluador dios. q pue fazer todas las cosas. Ede pue esto al de uandicho Rey el q es fructo muy sobre pulant. e mu-

HeadingLine:rubric



Real Academia Española

1



se suben las imágenes

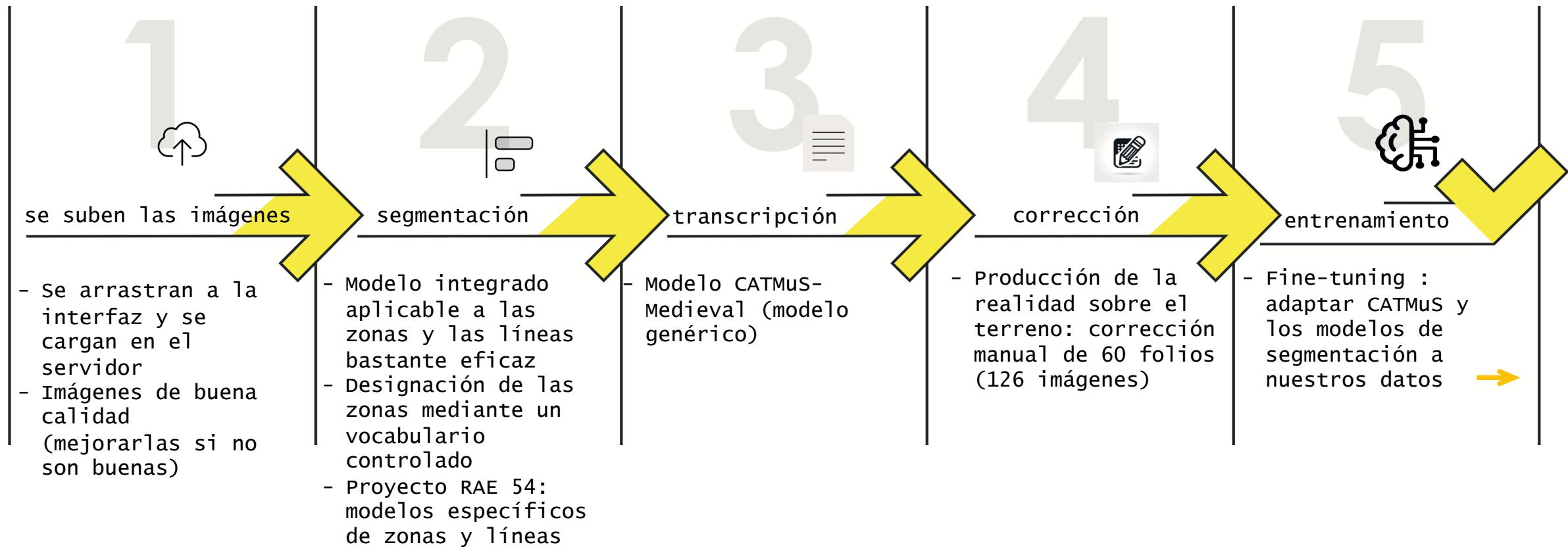
2



segmentación

- Se arrastran a la interfaz y se cargan en el servidor
- Imágenes de buena calidad  
(mejorarlas si no son buenas)
- Modelo integrado aplicable a las zonas y las líneas bastante eficaz
- Designación de las zonas mediante un vocabulario controlado
- Proyecto RAE 54: modelos específicos de zonas y líneas

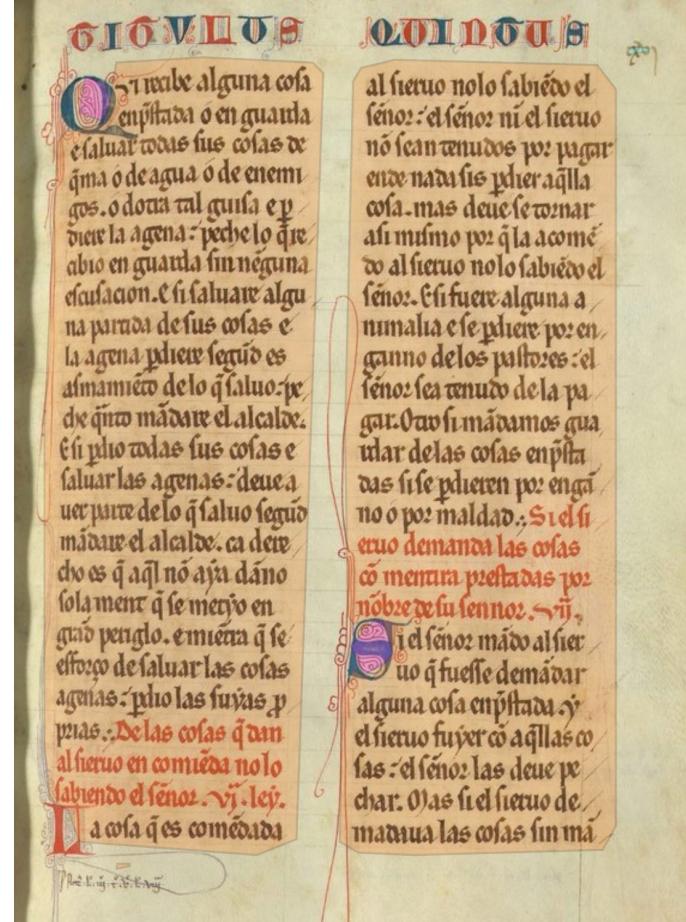
a coytar se deuē de sacar los coytados delas qbrantancas. q'daq adelâtre nō ayan poder los malos de enpecer los. E todo ome coytado aya entendemiento q'aya remedio desu coyta por esta nr̄a ley. Onde cuelmo en los tiépos q'son pasados. los señores fuessen muy graues los señores nō gouernass̄ los pueblos q'les erā dados por derecho mas por uicta. Entendemēto de bien fecho: dar firmedū bre a lo q'faze. E fazer leuatar la cosa q'era cayda. assi si los om̄s menores deuen cuydar de toller la carga ela coyta de sobre los pueblos esobre si. Muy mas graue culpa sera en los plados de sc̄a eglesia q'só mayores si ellos nō q̄sierē acorrer a los pueblos q'les son dados de dios en adlā manera q' ellos puedē. Onde



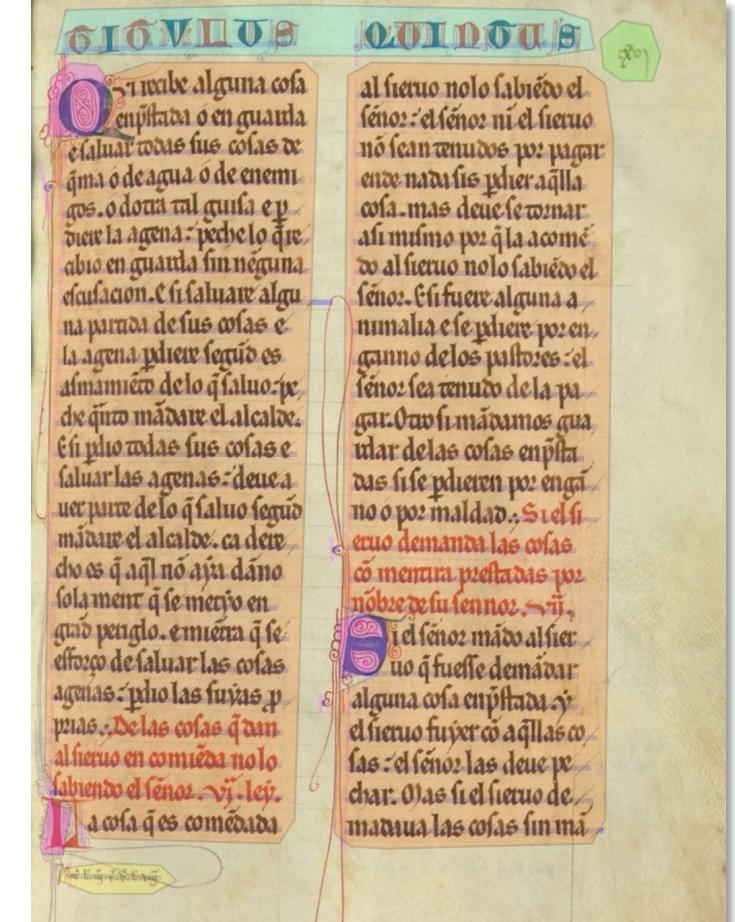
## Segmentación con diversos modelos



Modelo eSCRIPTORIUM por defecto



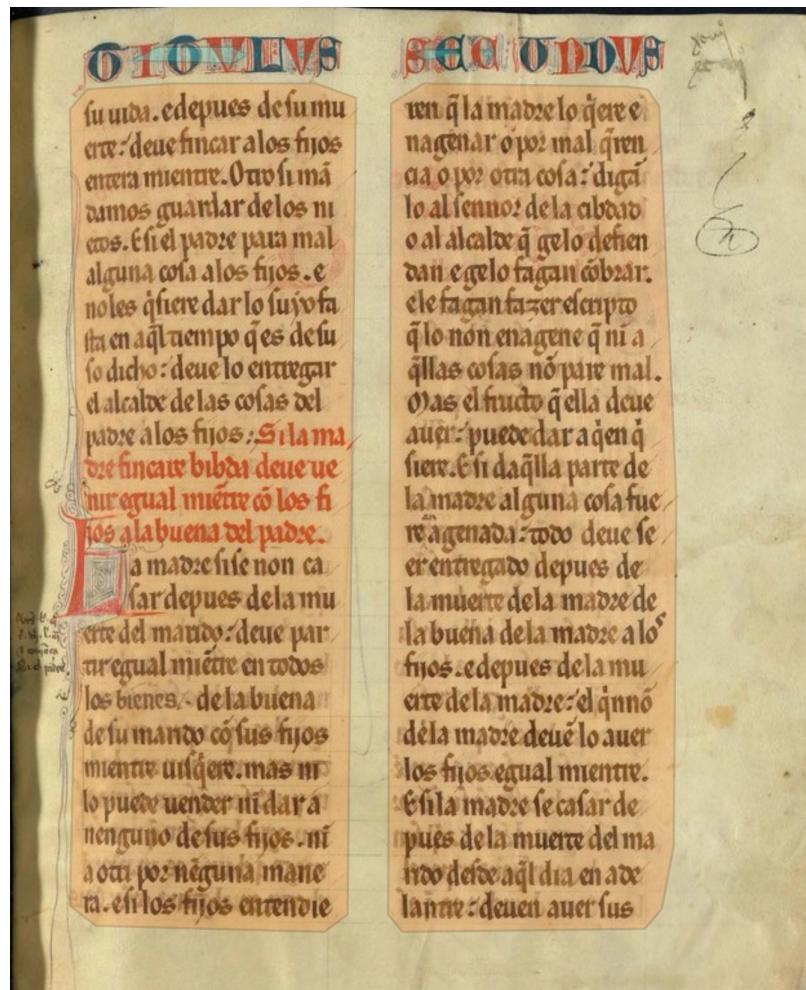
Modelo segmOnto v2 (zonas)



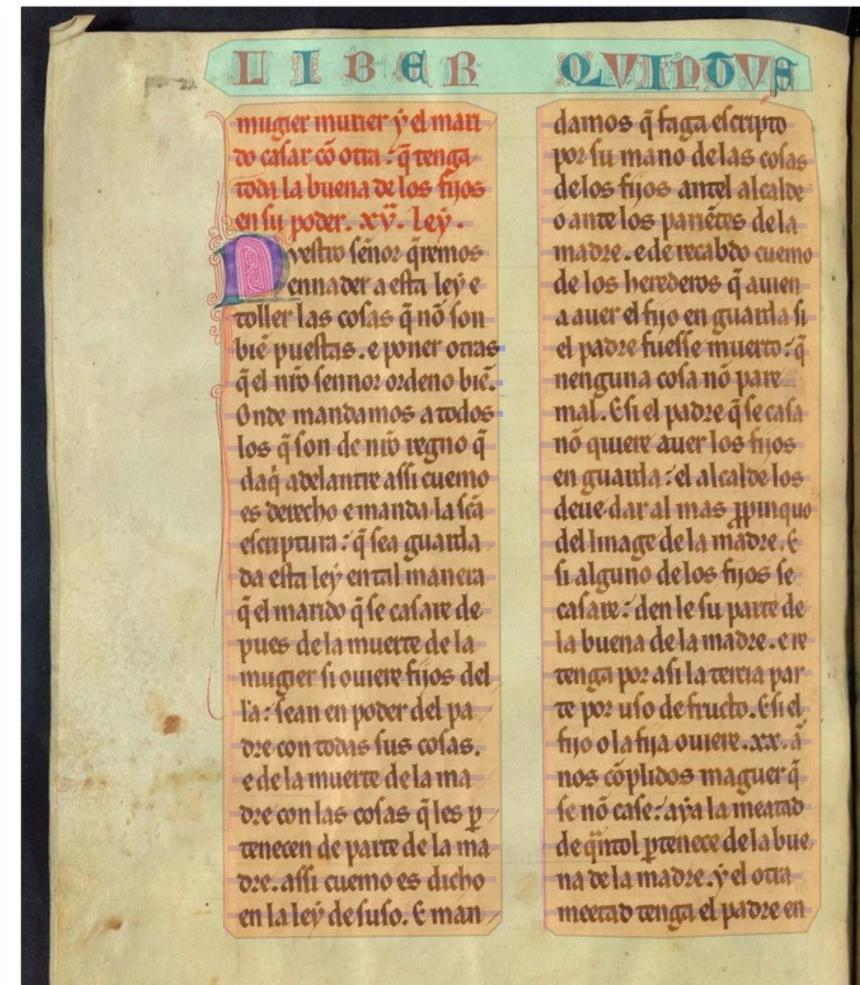
Fine-tuning

# Resultados de los primeros entrenamientos (modelos de segmentación)

*From scratch* con 126 images



Fine-tuning a partir de segmOnto v2



## Resultado de los primeros entrenamientos (modelos de transcripción)

A partir de CATMuS Medieval

Fine-tuning con 40 imágenes

Fine-tuning con 12 imágenes

```
== report ==
5287 Characters
445 Errors
91.58% Accuracy

126 Insertions
119 Deletions
200 Substitutions
```

```
== report ==
5287 Characters
280 Errors
94.70% Accuracy

47 Insertions
95 Deletions
138 Substitutions
```

```
== report ==
5287 Characters
147 Errors
97.22% Accuracy

54 Insertions
24 Deletions
69 Substitutions
```

```
Errors Correct-Generated
93 { } - { COMBINING TILDE }
59 { - } - { }
57 { õ } - { o }
36 { SPACE } - { }
32 { ē } - { e }
21 { t } - { d }
14 { ã } - { a }
11 { COMBINING TILDE } - { }
```

Errors	Correct-Generated
75	{ } - { COMBINING TILDE }
38	{ õ } - { o }
30	{ ē } - { e }
19	{ - } - { }
14	{ SPACE } - { }
12	{ ã } - { a }
8	{ ī } - { i }
7	{ } - { - }
5	{ COMBINING TILDE } - { }

Errors	Correct-Generated
30	{ - } - { }
10	{ SPACE } - { }
7	{ } - { SPACE }
7	{ ē } - { e }
6	{ } - { COMBINING TILDE }
4	{ } - { i }
4	{ COMBINING TILDE } - { }
4	{ ã } - { a }
3	{ u } - { i }

Transcripción automática de mss modernos

- Calidad de las imágenes dispar
- Escritores no profesionales = mayor irregularidad en la escritura
- Ningún modelo de aplicación general disponible para este tipo de escritura (coloso español no sobresale)



El tiempo empleado en obtener una transcripción satisfactoria aumenta



# Las imágenes

## Preprocesar o no preprocesar

early and taking effective corrective actions. If a project gets too far out of control, it may be difficult to achieve the project objective without sacrificing the scope, budget, schedule, or quality.

Throughout a project, some activities will be completed on time, some will be finished ahead of schedule, and others will be finished later than scheduled. Actual progress—whether faster or slower than planned—will have an effect on the schedule of the remaining, uncompleted activities of the project. Specifically, the actual finish times (AFs) of completed activities will determine the earliest start and earliest finish times for the remaining activities in the network diagram, as well as the total slack.

Throughout a project, changes may occur that have an impact on the schedule. These changes might be initiated by the customer or the project team, or they might be the result of an unanticipated occurrence. Any type of change—whether initiated by the customer, the contractor, the project manager, a team member, or an unanticipated event—will require a modification to the plan in terms of scope, budget, and/or schedule. When such changes are agreed upon, a new baseline plan is established and used as the benchmark against which actual project performance will be compared.

Once data have been collected on the actual finish times of completed activities and the effects of any project changes, an updated project schedule

early and taking effective corrective actions. If a project gets too far out of control, it may be difficult to achieve the project objective without sacrificing the scope, budget, schedule, or quality.

Throughout a project, some activities will be completed on time, some will be finished ahead of schedule, and others will be finished later than scheduled. Actual progress—whether faster or slower than planned—will have an effect on the schedule of the remaining, uncompleted activities of the project. Specifically, the actual finish times (AFs) of completed activities will determine the earliest start and earliest finish times for the remaining activities in the network diagram, as well as the total slack.

Throughout a project, changes may occur that have an impact on the schedule. These changes might be initiated by the customer or the project team, or they might be the result of an unanticipated occurrence. Any type of change—whether initiated by the customer, the contractor, the project manager, a team member, or an unanticipated event—will require a modification to the plan in terms of scope, budget, and/or schedule. When such changes are agreed upon, a new baseline plan is established and used as the benchmark against which actual project performance will be compared.

Once data have been collected on the actual finish times of completed activities and the effects of any project changes, an updated project schedule

Modificar los facsímiles es principalmente una cuestión estética, ya que preprocesar las imágenes no afecta necesariamente los resultados de la segmentación.

# Las imágenes

## Preprocesar o no preprocesar

early and taking effective corrective actions. If a project gets too far out of control, it may be difficult to achieve the project objective without sacrificing the scope, budget, schedule, or quality.

Throughout a project, some activities will be completed on time, some will be finished ahead of schedule, and others will be finished later than scheduled. Actual progress—whether faster or slower than planned—will have an effect on the schedule of the remaining, uncompleted activities of the project. Specifically, the actual finish times (AFs) of completed activities will determine the earliest start and earliest finish times for the remaining activities in the network diagram, as well as the total slack.

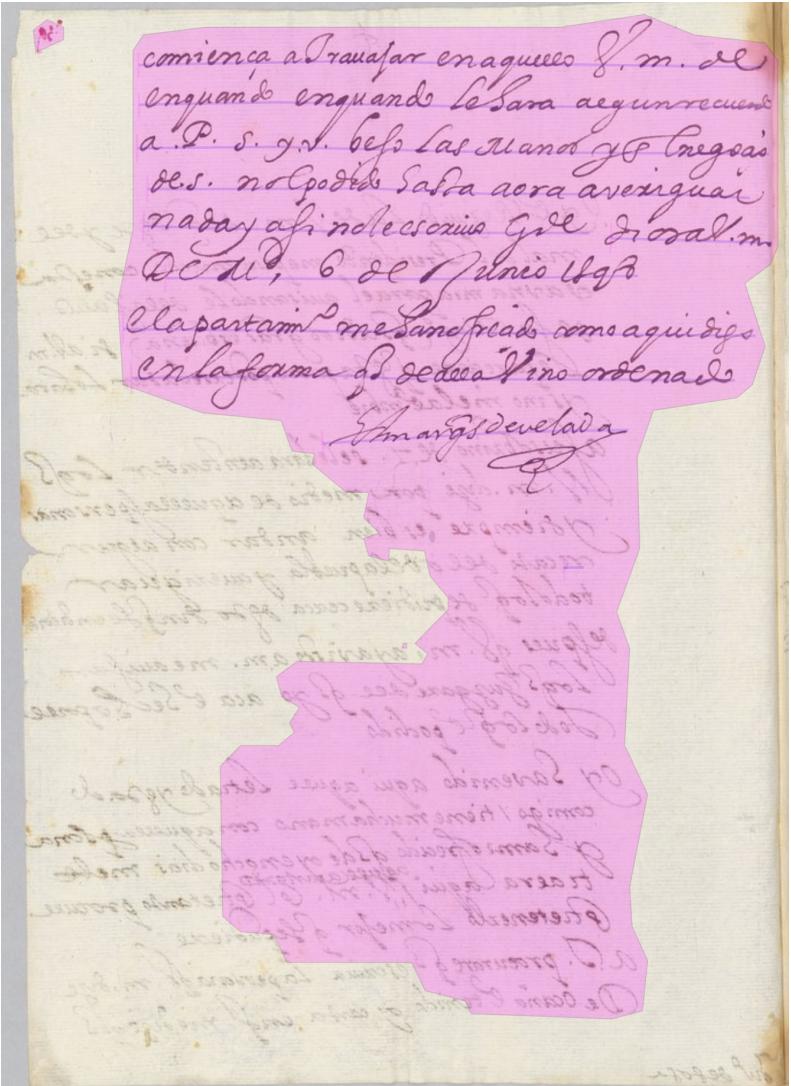
Throughout a project, changes may occur that have an impact on the schedule. These changes might be initiated by the customer or the project team, or they might be the result of an unanticipated occurrence. Any type of change—whether initiated by the customer, the contractor, the project manager, a team member, or an unanticipated event—will require a modification to the plan in terms of scope, budget, and/or schedule. When such changes are agreed upon, a new baseline plan is established and used as the benchmark against which actual project performance will be compared.

Once data have been collected on the actual finish times of completed activities and the effects of any project changes, an updated project schedule

La binarización tampoco mejora los resultados en el estado actual de la tecnología

# Las imágenes

Preprocesar o no preprocesar

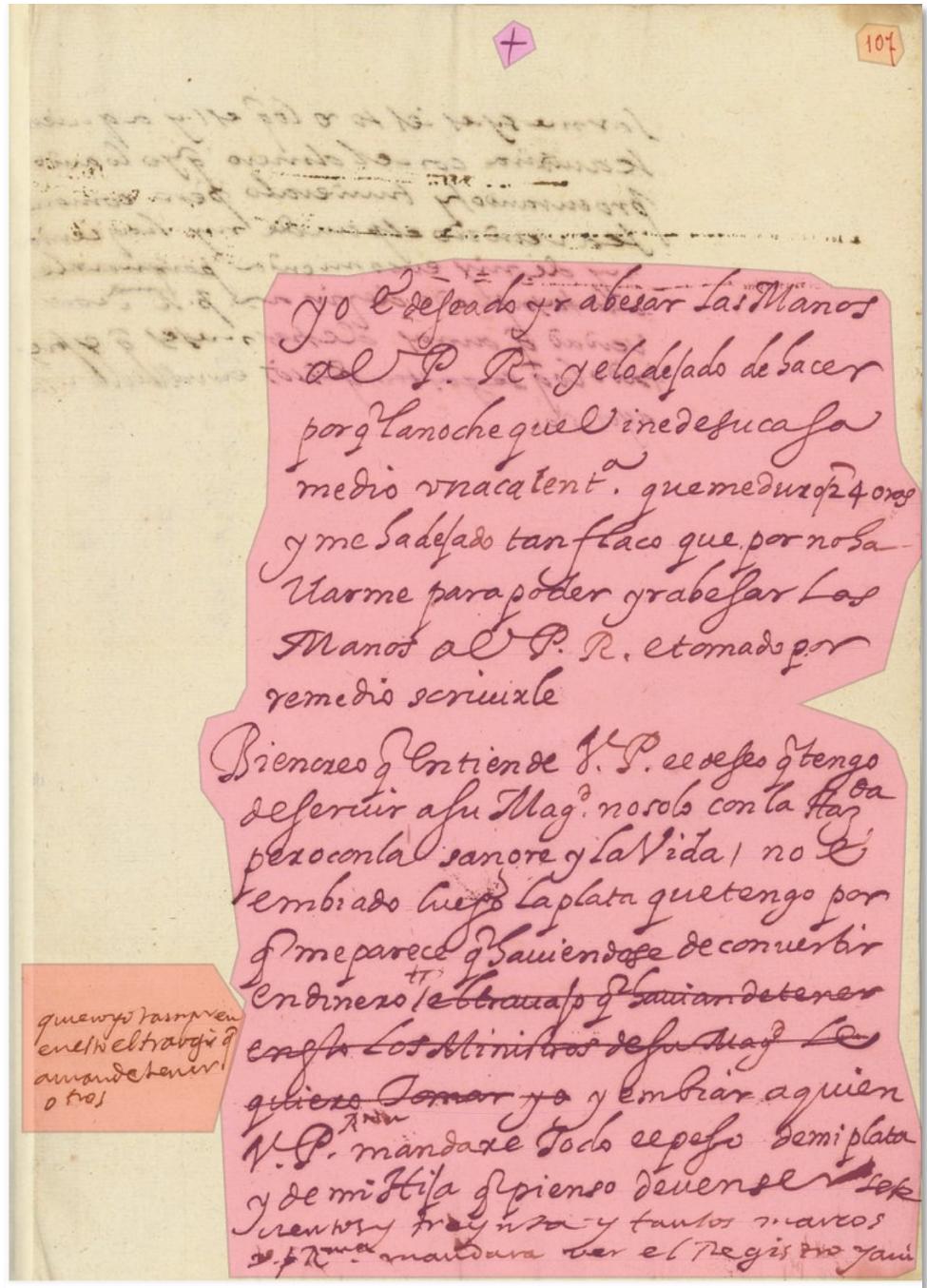


JAKE-CLARK.TUMBLR

# La segmentación

- La segmentación es tan importante como la transcripción para el resultado final.
- Conocer sus archivos para determinar el grado de detalle en la identificación de zonas.

Identificación de zonas (texto, columnas, figuras, numeración)



# La segmentación

104

Yo lo dejado y abeñ la Manos  
tal P. R. y el dejado de hacer  
porq Yanoche que elline de su casa  
medio vna calent<sup>a</sup> que medura 24000  
y me salido tan flaco que por no ha  
tarne para poder grabafar los  
Manos a D. P. R. etomado por  
un medio scriurle

Bienveo y Contiene V.P. ese se pofeno  
seforio ahi Mag<sup>o</sup> nrolo con la Roca  
y diconta sanore y la Vida / no Es  
embiado ahi la plata que tengo por  
que paseo y huiendo de convertir  
en dineros le lleva qd suian detener  
en qd los lleva de la Mag<sup>o</sup> 12  
qd ibero somar ya y embiar aquien  
V.P. manda de todo cepejo semiplata  
y demitila qd piense devuelo /  
cremto / qd una y tanto manos  
de puma mandara ver el regisro yani

que uno qd  
que el qd  
que uno qd  
que uno qd

104

Yo lo dejado y abeñ la Manos  
tal P. R. y el dejado de hacer  
porq Yanoche que elline de su casa  
medio vna calent<sup>a</sup> que medura 24000  
y me salido tan flaco que por no ha  
tarne para poder grabafar los  
Manos a D. P. R. etomado por  
un medio scriurle

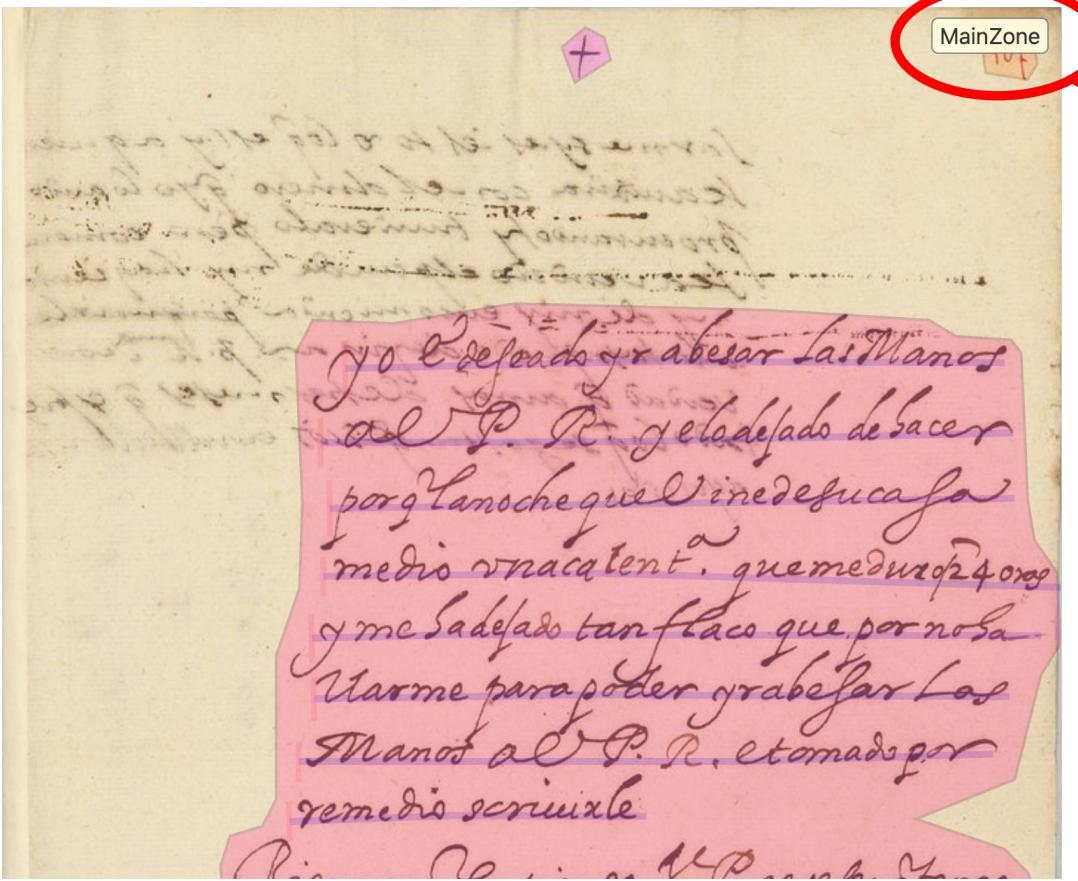
Bienveo y Contiene V.P. ese se pofeno  
seforio ahi Mag<sup>o</sup> nrolo con la Roca  
y diconta sanore y la Vida / no Es  
embiado ahi la plata que tengo por  
que paseo y huiendo de convertir  
en dineros le lleva qd suian detener  
en qd los lleva de la Mag<sup>o</sup> 12  
qd ibero somar ya y embiar aquien  
V.P. manda de todo cepejo semiplata  
y demitila qd piense devuelo /  
cremto / qd una y tanto manos  
de puma mandara ver el regisro yani

que uno qd  
que el qd  
que uno qd  
que uno qd

Identificación de líneas de texto

# La segmentación

- En *eScriptorium* se pueden descargar las transcripciones en formato ALTO (XML), PAGE (XML) y texto plano (.txt)
  - Denominación de zonas para asegurar el control de calidad de los datos y con vistas a su explotación

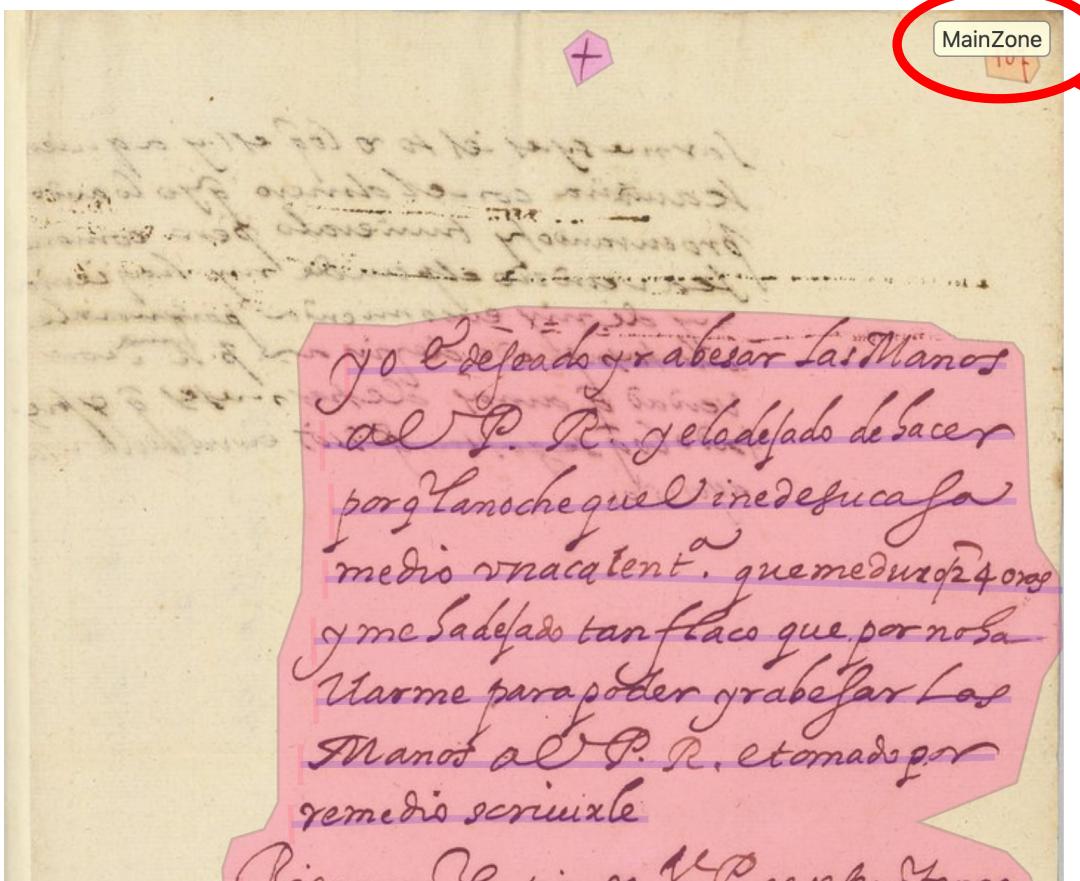


```
<Page imageFilename="coll_ed_favre_35_f042v.jpeg" imageWidth="2512" imageHeight="3689">
<TextRegion id="eSc_textblock_82bfc893" custom="structure {type:MainZone;}">
  <Coords points="2413,196 2487,229 2509,293 2479,368 2495,1014 2501,1077 2502,1276 2491,1374 2499,1467 2504,1519 2497,1698 2282,1
  <TextLine id="eSc_line_e8dac3dd" custom="structure {type:DetailLine;}">
    <Coords points="2444,189 2333,196 2296,215 2256,196 2250,196 2124,221 2069,196 2060,196 2013,215 1973,227 1721,209 1694,230 1
    <Baseline points="513,340 1960,282 2454,284"/>
    <TextEquiv>
      <Unicode>y sy parece ã elmoço es distraydo y ã la ha</Unicode>
    </TextEquiv>
  </TextLine>
</TextRegion>
```

# La segmentación

Pero no solo...

Las transcripciones obtenidas en un formato interoperable como ALTO pueden ser compartidas con otros investigadores. Para facilitar este intercambio, es ideal utilizar un estándar común.



Vocabulario controlado extensible: <https://segmonto.github.io/>

```
<Page imageFilename="coll_ed_favre_35_f042v.jpeg" imageWidth="2512" imageHeight="3689">  
<TextRegion id="eSc_textblock_82bfc893" custom="structure {type:MainZone;}">  
<Coords points="2413,196 2487,229 2509,293 2479,368 2495,1014 2501,1077 2502,1276 2491,1374 2499,1467 2504,1519 2497,1698 2282,1694 2301,209 2271,209 2209,230 2169,230 2137,227 2121,209 2060,196 2069,196 2013,215 1973,227 1721,209 1694,230 1691,230 1689,230 1687,230 1685,230 1683,230 1681,230 1679,230 1677,230 1675,230 1673,230 1671,230 1669,230 1667,230 1665,230 1663,230 1661,230 1659,230 1657,230 1655,230 1653,230 1651,230 1649,230 1647,230 1645,230 1643,230 1641,230 1639,230 1637,230 1635,230 1633,230 1631,230 1629,230 1627,230 1625,230 1623,230 1621,230 1619,230 1617,230 1615,230 1613,230 1611,230 1609,230 1607,230 1605,230 1603,230 1601,230 1599,230 1597,230 1595,230 1593,230 1591,230 1589,230 1587,230 1585,230 1583,230 1581,230 1579,230 1577,230 1575,230 1573,230 1571,230 1569,230 1567,230 1565,230 1563,230 1561,230 1559,230 1557,230 1555,230 1553,230 1551,230 1549,230 1547,230 1545,230 1543,230 1541,230 1539,230 1537,230 1535,230 1533,230 1531,230 1529,230 1527,230 1525,230 1523,230 1521,230 1519,230 1517,230 1515,230 1513,230 1511,230 1509,230 1507,230 1505,230 1503,230 1501,230 1499,230 1497,230 1495,230 1493,230 1491,230 1489,230 1487,230 1485,230 1483,230 1481,230 1479,230 1477,230 1475,230 1473,230 1471,230 1469,230 1467,230 1465,230 1463,230 1461,230 1459,230 1457,230 1455,230 1453,230 1451,230 1449,230 1447,230 1445,230 1443,230 1441,230 1439,230 1437,230 1435,230 1433,230 1431,230 1429,230 1427,230 1425,230 1423,230 1421,230 1419,230 1417,230 1415,230 1413,230 1411,230 1409,230 1407,230 1405,230 1403,230 1401,230 1399,230 1397,230 1395,230 1393,230 1391,230 1389,230 1387,230 1385,230 1383,230 1381,230 1379,230 1377,230 1375,230 1373,230 1371,230 1369,230 1367,230 1365,230 1363,230 1361,230 1359,230 1357,230 1355,230 1353,230 1351,230 1349,230 1347,230 1345,230 1343,230 1341,230 1339,230 1337,230 1335,230 1333,230 1331,230 1329,230 1327,230 1325,230 1323,230 1321,230 1319,230 1317,230 1315,230 1313,230 1311,230 1309,230 1307,230 1305,230 1303,230 1301,230 1299,230 1297,230 1295,230 1293,230 1291,230 1289,230 1287,230 1285,230 1283,230 1281,230 1279,230 1277,230 1275,230 1273,230 1271,230 1269,230 1267,230 1265,230 1263,230 1261,230 1259,230 1257,230 1255,230 1253,230 1251,230 1249,230 1247,230 1245,230 1243,230 1241,230 1239,230 1237,230 1235,230 1233,230 1231,230 1229,230 1227,230 1225,230 1223,230 1221,230 1219,230 1217,230 1215,230 1213,230 1211,230 1209,230 1207,230 1205,230 1203,230 1201,230 1199,230 1197,230 1195,230 1193,230 1191,230 1189,230 1187,230 1185,230 1183,230 1181,230 1179,230 1177,230 1175,230 1173,230 1171,230 1169,230 1167,230 1165,230 1163,230 1161,230 1159,230 1157,230 1155,230 1153,230 1151,230 1149,230 1147,230 1145,230 1143,230 1141,230 1139,230 1137,230 1135,230 1133,230 1131,230 1129,230 1127,230 1125,230 1123,230 1121,230 1119,230 1117,230 1115,230 1113,230 1111,230 1109,230 1107,230 1105,230 1103,230 1101,230 1099,230 1097,230 1095,230 1093,230 1091,230 1089,230 1087,230 1085,230 1083,230 1081,230 1079,230 1077,230 1075,230 1073,230 1071,230 1069,230 1067,230 1065,230 1063,230 1061,230 1059,230 1057,230 1055,230 1053,230 1051,230 1049,230 1047,230 1045,230 1043,230 1041,230 1039,230 1037,230 1035,230 1033,230 1031,230 1029,230 1027,230 1025,230 1023,230 1021,230 1019,230 1017,230 1015,230 1013,230 1011,230 1009,230 1007,230 1005,230 1003,230 1001,230 999,230 997,230 995,230 993,230 991,230 989,230 987,230 985,230 983,230 981,230 979,230 977,230 975,230 973,230 971,230 969,230 967,230 965,230 963,230 961,230 959,230 957,230 955,230 953,230 951,230 949,230 947,230 945,230 943,230 941,230 939,230 937,230 935,230 933,230 931,230 929,230 927,230 925,230 923,230 921,230 919,230 917,230 915,230 913,230 911,230 909,230 907,230 905,230 903,230 901,230 899,230 897,230 895,230 893,230 891,230 889,230 887,230 885,230 883,230 881,230 879,230 877,230 875,230 873,230 871,230 869,230 867,230 865,230 863,230 861,230 859,230 857,230 855,230 853,230 851,230 849,230 847,230 845,230 843,230 841,230 839,230 837,230 835,230 833,230 831,230 829,230 827,230 825,230 823,230 821,230 819,230 817,230 815,230 813,230 811,230 809,230 807,230 805,230 803,230 801,230 799,230 797,230 795,230 793,230 791,230 789,230 787,230 785,230 783,230 781,230 779,230 777,230 775,230 773,230 771,230 769,230 767,230 765,230 763,230 761,230 759,230 757,230 755,230 753,230 751,230 749,230 747,230 745,230 743,230 741,230 739,230 737,230 735,230 733,230 731,230 729,230 727,230 725,230 723,230 721,230 719,230 717,230 715,230 713,230 711,230 709,230 707,230 705,230 703,230 701,230 699,230 697,230 695,230 693,230 691,230 689,230 687,230 685,230 683,230 681,230 679,230 677,230 675,230 673,230 671,230 669,230 667,230 665,230 663,230 661,230 659,230 657,230 655,230 653,230 651,230 649,230 647,230 645,230 643,230 641,230 639,230 637,230 635,230 633,230 631,230 629,230 627,230 625,230 623,230 621,230 619,230 617,230 615,230 613,230 611,230 609,230 607,230 605,230 603,230 601,230 599,230 597,230 595,230 593,230 591,230 589,230 587,230 585,230 583,230 581,230 579,230 577,230 575,230 573,230 571,230 569,230 567,230 565,230 563,230 561,230 559,230 557,230 555,230 553,230 551,230 549,230 547,230 545,230 543,230 541,230 539,230 537,230 535,230 533,230 531,230 529,230 527,230 525,230 523,230 521,230 519,230 517,230 515,230 513,230 511,230 509,230 507,230 505,230 503,230 501,230 499,230 497,230 495,230 493,230 491,230 489,230 487,230 485,230 483,230 481,230 479,230 477,230 475,230 473,230 471,230 469,230 467,230 465,230 463,230 461,230 459,230 457,230 455,230 453,230 451,230 449,230 447,230 445,230 443,230 441,230 439,230 437,230 435,230 433,230 431,230 429,230 427,230 425,230 423,230 421,230 419,230 417,230 415,230 413,230 411,230 409,230 407,230 405,230 403,230 401,230 399,230 397,230 395,230 393,230 391,230 389,230 387,230 385,230 383,230 381,230 379,230 377,230 375,230 373,230 371,230 369,230 367,230 365,230 363,230 361,230 359,230 357,230 355,230 353,230 351,230 349,230 347,230 345,230 343,230 341,230 339,230 337,230 335,230 333,230 331,230 329,230 327,230 325,230 323,230 321,230 319,230 317,230 315,230 313,230 311,230 309,230 307,230 305,230 303,230 301,230 299,230 297,230 295,230 293,230 291,230 289,230 287,230 285,230 283,230 281,230 279,230 277,230 275,230 273,230 271,230 269,230 267,230 265,230 263,230 261,230 259,230 257,230 255,230 253,230 251,230 249,230 247,230 245,230 243,230 241,230 239,230 237,230 235,230 233,230 231,230 229,230 227,230 225,230 223,230 221,230 219,230 217,230 215,230 213,230 211,230 209,230 207,230 205,230 203,230 201,230 199,230 197,230 195,230 193,230 191,230 189,230 187,230 185,230 183,230 181,230 179,230 177,230 175,230 173,230 171,230 169,230 167,230 165,230 163,230 161,230 159,230 157,230 155,230 153,230 151,230 149,230 147,230 145,230 143,230 141,230 139,230 137,230 135,230 133,230 131,230 129,230 127,230 125,230 123,230 121,230 119,230 117,230 115,230 113,230 111,230 109,230 107,230 105,230 103,230 101,230 99,230 97,230 95,230 93,230 91,230 89,230 87,230 85,230 83,230 81,230 79,230 77,230 75,230 73,230 71,230 69,230 67,230 65,230 63,230 61,230 59,230 57,230 55,230 53,230 51,230 49,230 47,230 45,230 43,230 41,230 39,230 37,230 35,230 33,230 31,230 29,230 27,230 25,230 23,230 21,230 19,230 17,230 15,230 13,230 11,230 9,230 7,230 5,230 3,230 1,230 0,230"/>
```

# La segmentación

Denominación de zonas con el vocabulario SegmOnto (Gabay et al. 2023)

MainZone para el bloque de texto central

MarginTextZone para las anotaciones al margen

NumberingZone para el número de folio

DamageZone para todos los tipos de deterioros

GraphicZone:figure para las cruces u otras figuras

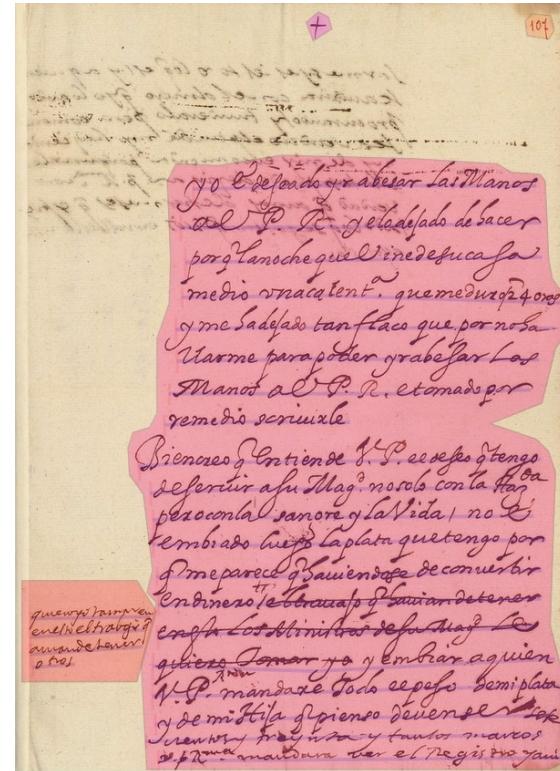
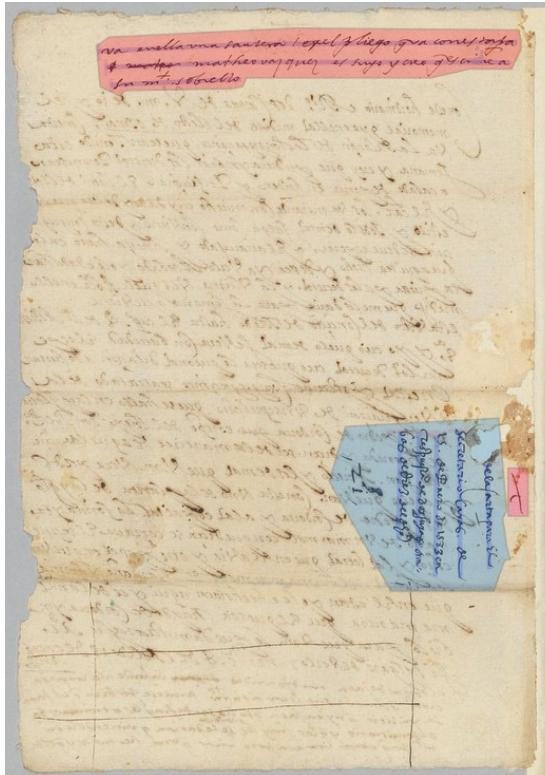
CustomZone:salute para el saludo

CustomZone:date para la fecha

CustomZone:postscript para los post scriptum

CustomZone:envelope para el sobrescrito

CustomZone:recipient para el destinatario cuando está escrito en alguna parte de la carta fuera de la zona principal



# La segmentación

## Líneas

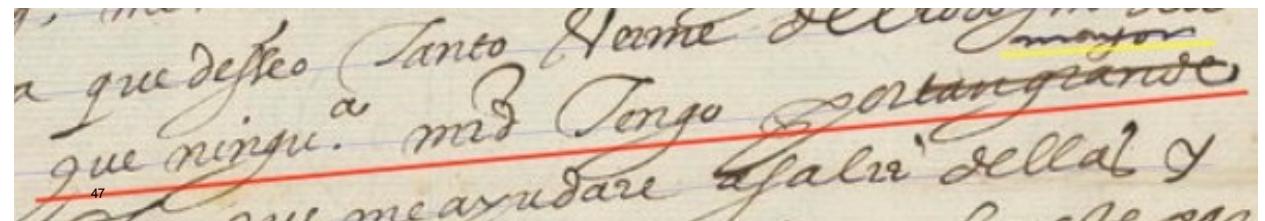
`DefaultLine` línea principal

`InterlinearLine:cross-out` interlíneas tachadas

`InterlinearLine:addition` para el texto añadido entre las líneas principales

`InterlinearLine:correction` para el texto que corrige el texto original

`InterlinearLine:continuation` para las líneas que continúan la línea principal por falta de espacio.



`InterlinearLine:correction` mayor

`DefaultLine` que ninguna merced tengo por [tangrande]

# Más decisiones: ¿Cómo transcribir?



- ¿Para qué transcribimos? ¿Queremos compartir nuestros datos? ¿Cuántos transcriptores y correctores hay en el proyecto?
- Es importante documentar todas las decisiones tomadas en el proyecto. Además, es más fácil eliminar información que añadirla una vez que el proyecto esté avanzado.
- Si hay más de un transcriptor, el riesgo de errores de interpretación aumenta.

A photograph of a page from a handwritten manuscript. The page is filled with dense cursive handwriting in black ink. Several sections of text are highlighted with colored paper: a large pink rectangle covers the upper right portion of the page, and two smaller orange rectangles are located at the bottom left and center. The text is in Spanish and discusses various topics, including health, wealth, and spiritual matters. Two specific words in the pink-highlighted section are circled in yellow: "Mag" and "Hija".

yo e dejado de abesar las Manos  
al D. P. R. y el dejado de besar  
por la noche que l'ine de su casa  
medio vna calent. quemadura 24 oreas  
y me haidado tan flaco que por no ha-  
tar me para poder grabesar las  
Manos al D. P. R. etomado p. r.  
remedio scriuixle

Bienares q entiendo d. V. P. ee sefio q tengo  
desfertir ahi Mag. nosolo con la Hija  
pero con la sanora y la Vida / no se  
embriado luego la plata que tengo por  
q me parecia q bauienda de convertir  
en dineros le bauia q bauian detener  
esta los administr de fa Mag. q  
glicero Tomar ya y embriar aquien  
V. P. mandare todo ee sefio semiplata  
y de mi Hija q pienso devuense a q  
trenor tray nra y tantos marcos  
q no maldura ver el Regis q no yani

# Más decisiones: ¿Cómo transcribir?



- Es fundamental conocer el corpus y definir criterios claros de transcripción, así como documentarlos adecuadamente. Además, es necesario precisar qué hacer en aquellos casos que no estén cubiertos por las normas.

yo e dejado de abesar las Manos  
al V. P. R. y el dejado de besar  
por la noche que l'ine de su casa  
medio vna calent. quemadura 24 oreas  
y me h'adado tan flaco que por no sa-  
tar me para poder grabesar las  
Manos al V. P. R. etomado p. r.  
remedio scriuixle

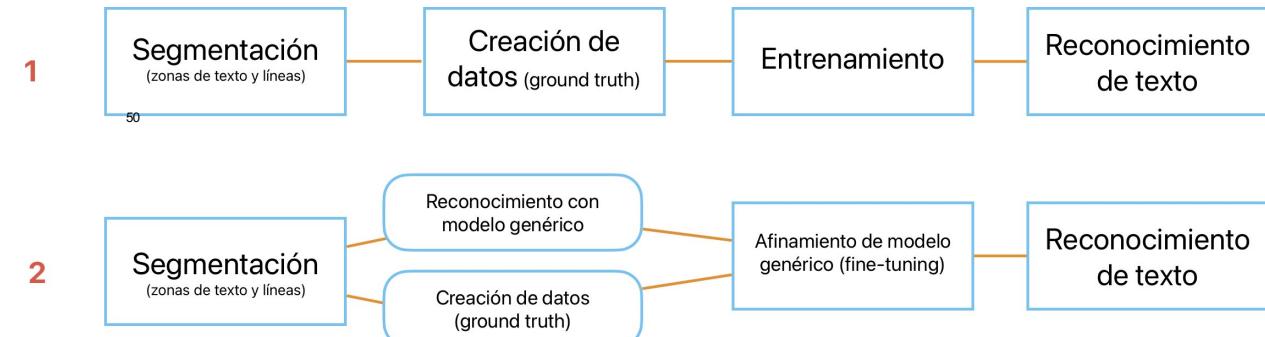
Bienres q' entiendo q' V. P. e epejo q' tempo  
desfertir ahi Mag. nosolo con la Hija  
pero con la sanora y la Vida / no se  
embriado luego la plata q' ten'go por  
q' me parecio q' bauiendo de convertir  
en dineros le lleva q' bauian detener  
esta los administr de fa Mag. q'  
glicero Tomar ya y embriar a quien  
V. P. mandare todo e epejo semiplata  
y de mi Hija q' pienso devuense q'  
trenor y frax nra y tantos marcos  
q' p' la maldura ver el Regis q' yani

que yo q' tamme  
en el ultrajos q'  
anando tener  
o los

# Para transcribir dos posibilidades:

**Workflow 1:** Transcripción manual y creación de un corpus (gold corpus) para entrenar un modelo adaptado a la escritura de los documentos.

**Workflow 2:** Transcripción utilizando un modelo existente y corrección de los resultados, seguido del afinamiento de un modelo genérico.



Para este tipo de letra no existe un modelo realmente eficaz según los criterios de algunos especialistas. 95 % de precisión sería el mínimo aceptable. **Esto es discutible.**

# Un experimento

- División del corpus según la mano: marqués de Velada (45 documentos), secretario (185 documentos), mixta (31 documentos) y regidor de Madrid (67 documentos).  
51
- Selección aleatoria de 8-10 documentos (aproximadamente 20 imágenes).
- Aproximadamente 550 líneas.



# Pero antes, ¿cómo se evalúa un modelo?

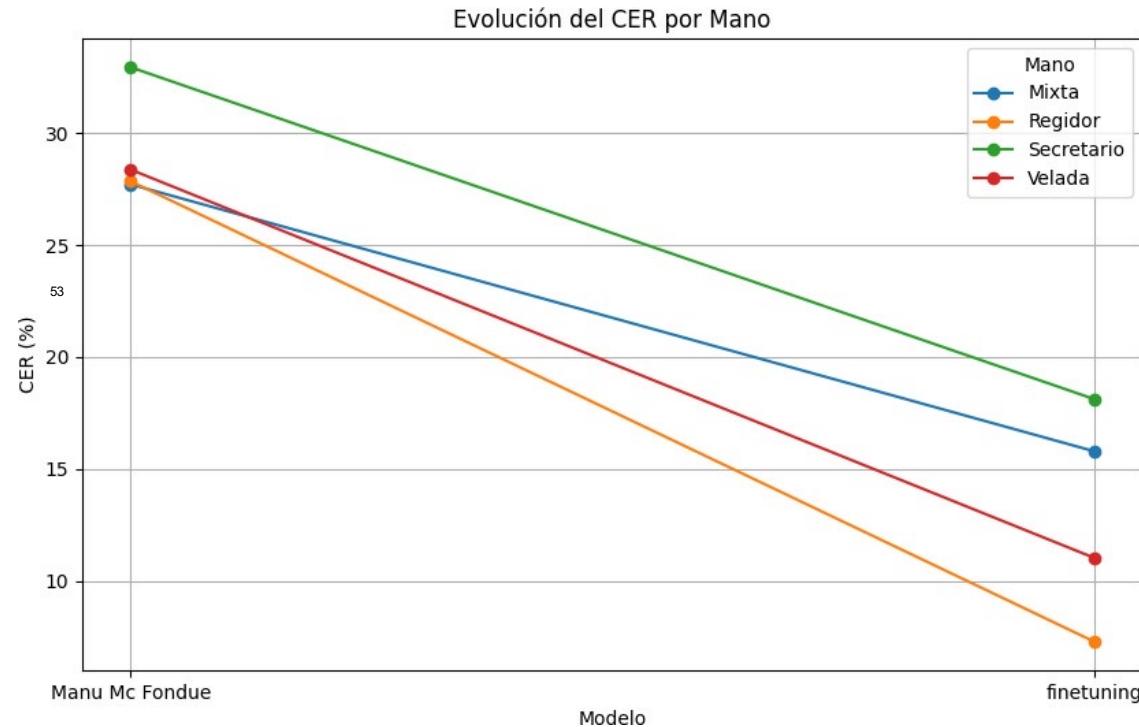
Con una medida de precisión llamada CER (Character Error Rate)



$$CER = \frac{(Sustitución + Eliminación + Inserción)}{N}$$

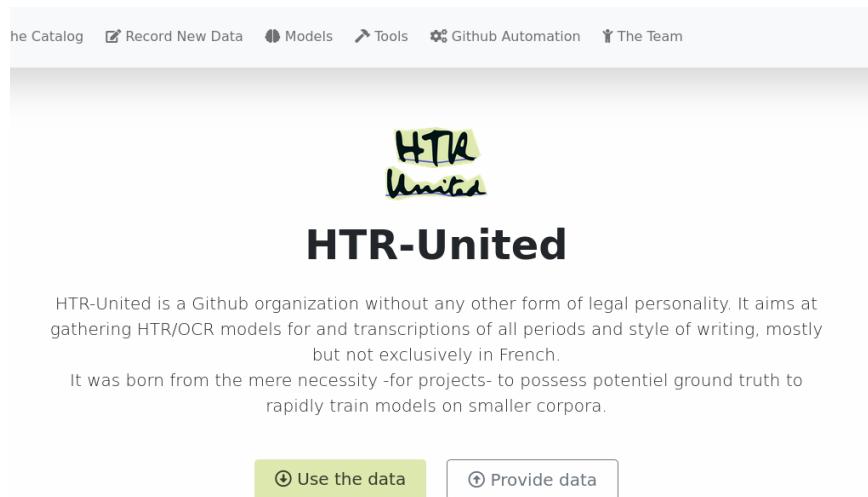
# Resultados

- Según la documentación de Transkribus, un margen de error inferior al 10% puede considerarse bueno (Bermúdez Carreño, 2023: 115).
- Tomar estas informaciones con cautela: corpus in-domain vs corpus out-of-domain



# Herramientas

- Para compartir y conseguir datos:



<https://htr-united.github.io/catalog.html>

# Herramientas

\*La evaluación de un modelo en eScriptorium se realiza mediante línea de comandos directamente en el motor de OCR.

Una solución: <https://huggingface.co/spaces/lterriel/kami-app>

The screenshot shows the 'Welcome to KaMI App!' screen. At the top is a logo featuring a red octopus-like creature inside a circular emblem. Below the logo, the text 'Welcome to KaMI App!' is displayed. A descriptive subtitle explains: 'KaMI stands for Kraken Model Inspector. It evaluates the success of a transcription task (HTR, OCR) comparing a correct transcription (reference) and a prediction.' A button labeled 'Do more with KaMI-lib colab' is present. The main area contains two input fields: 'Reference' (Maximum characters: 7000) and 'Prediction' (Maximum characters: 7000). Both fields have character count indicators (7000) and placeholder text: 'Drag & Drop or Paste the expected transcription...' and 'Drag & Drop or Paste the transcription produced by the model you wish to evaluate...'. The entire interface is contained within a light gray box.

## A modo de conclusión:

- En el estado actual de la cuestión, cada investigador se enfrenta a decisiones cruciales para el éxito de su etapa de transcripción: la precisión requerida de un modelo HTR subjetiva, cuántos datos son necesarios para afinar un modelo genérico también, qué criterios de transcripción emplear, que estándar para describir las zonas, etc.

# eScriptorium

Outil de reconnaissance de caractères applicable à des documents imprimés (*Optical Character Recognition* ou *OCR*) ou manuscrits (*Handwritten Text Recognition* ou *HTR*) à l'aide de techniques d'apprentissage automatique (*deep learning*).

Cette instance d'eScriptorium est mise à disposition par

**datumo**, le pôle < Services pour la recherche en sciences humaines >

de la

**Division calcul et soutien à la recherche (DCSR)**,

dans le cadre exclusif d'un cours. Si vous y participez, une invitation a dû vous être adressée par vos enseignantes.

Sinon, d'autres instances (instance de test et instance de production) suivront bientôt.

Si vous souhaitez être tenu-e au courant de la mise à disposition de ces nouvelles instances, vous pouvez nous contacter via [ce formulaire](#) ou directement à cette adresse [dscr-datumo@unil.ch](mailto:dscr-datumo@unil.ch).



## Automatic Transcription

Apply OCR/HTR to images of printed and handwritten documents using shared open models.



## Manual transcription

Make use of an ergonomic user interface leveraging modern browser technology to edit segmentations and transcriptions.



## Train Models

Create new models or finetune existing ones to improve automatic recognition.

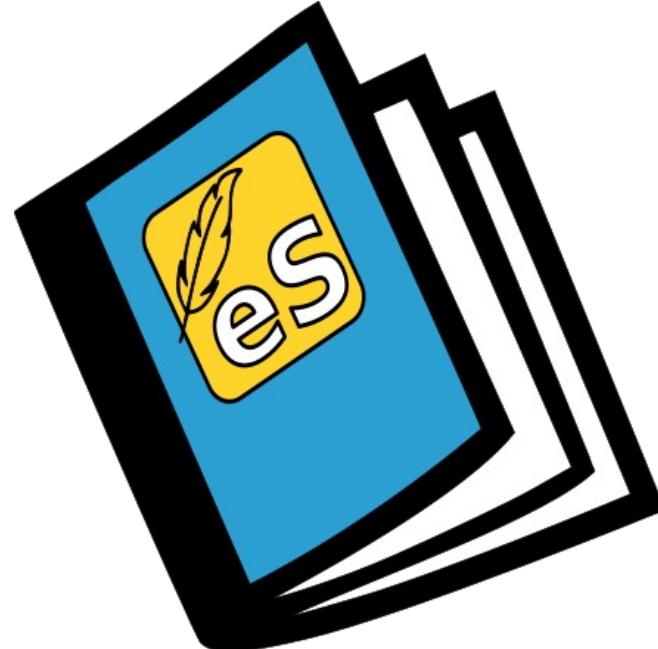


## Import/ Export

Import and export models and texts transcriptions in a variety of formats. Access data through a full REST API.

<https://escriptor.dcsr.unil.ch/>

Una chuleta ☺



# **ESCRITORIUM DOCUMENTATION**

<https://escriptorium.readthedocs.io/en/latest/>

## Otros recursos

Depósito abierto para modelos y datos: <https://zenodo.org/>

Chat eScriptorium: [https://app.gitter.im/#/room/#escripta\\_escriptorium:gitter.im](https://app.gitter.im/#/room/#escripta_escriptorium:gitter.im)