

Herramientas digitales para la literatura: el diccionario de lemas y formas del Quijote

Antonio Moreno Sandoval

I Jornadas Humanidades Digitales

9 de enero de 2024



Esquema

1. Orígenes del proyecto, problemas y equipo
2. El método de compilación del diccionario
3. La conversión a base de datos
4. La revisión y la herramienta de consulta (LYNEAL)
5. La publicación

Orígenes, problemas y equipo

- La idea surge en 2015, en el cuarto centenario:
- **¿por qué no existe un diccionario de lemas y sus formas?**
 - hay listados de formas con frecuencias
 - no hay agrupación de formas y variantes bajo un lema
 - no hay clasificación por categorías sintácticas
- **¿cómo hacerlo con una obra de casi 400000 palabras?**

--> solo se puede realizar mediante anotación automática y revisión manual. Filología digital: un etiquetador morfosintáctico y lingüistas

El equipo del diccionario

Coordinadores: Hiroto Ueda (Tokio) y Antonio Moreno (UAM)

Colaboradores:

- María Agujetas
- Tomás Albaladejo
- Carmen Chacón
- Francisco Jesús Fernández
- José María Guirao
- Inmaculada Martínez
- Pedro Sánchez-Prieto

- ¿Qué versión del Quijote elegir?
 - En formato electrónico
 - Con autorización de los autores de la edición crítica

--> Florencio Sevilla y Antonio Rey, ed. 1996. UAM.

La presente copia en disquete (3,5;HD) contiene el texto de la novela, tal y como se ha fijado en la edición impresa, archivado en formato MS-DOS (ASCII) [...]
En definitiva, sin pretender sustituir al libro impreso, la copia en disco está pensada como una herramienta de trabajo complementaria que potencie y agilice la labor intelectual del lector actual.

Otros problemas (estudio preliminar con GRAMPAL)

-
- La variación formal: distintas variantes de la misma forma:
 - *malencolía*
 - *malenconía*
 - *melancolía*
 - *melanconía*
 - Palabras “desconocidas” para el lexicon de GRAMPAL:
 - 3915 formas de 220000 palabras (2ª parte DQ)
 - nombres propios
 - amalgamas (*desta, della*)
 - arcaísmos (*agora*)
 - tiempos obsoletos (*quisiere, antojare*)
 - uso de clíticos con diferentes tiempos verbales (*sírvate, siguióle*)
 - léxico de la época (*harbar, alcatifa*)

El método de compilación del diccionario

- Especialistas en lingüística computacional:
 - conversión del texto a base de datos
 - análisis morfosintáctico automático
- Especialistas en español del XVI y XVII:
 - variantes formales
- Especialistas en la obra de Cervantes:
 - rasgos sociolingüísticos de los personajes
 - nomenclatura de los capítulos

Fases de la compilación

1. Conversión del texto ANSI a UTF8
2. Transformación del texto narrativo en intervenciones por personaje y narrador, características sociolingüísticas (fichero excel)
3. Indexación del texto por intervenciones en un buscador (LYNEAL)
4. Análisis morfosintáctico automático (lematización, categoría y rasgos)
5. Revisión manual de las formas analizadas
6. Generación automática del borrador del diccionario
7. Revisión manual del diccionario, consultando fuentes

Conversión a base de datos por personaje

Identificación de los personajes

Personaje	Abr.	Cap.	Sexo	Edad	Clase	Interlocutor	
.	.	.	*	*	*		
Andrés	a	1,04	hombre	joven	llano/a		
Amiga-1	aa1	2,62	mujer	mayor	noble		
Amiga-2	aa2	2,62	mujer	mayor	noble		
Ambrosio	ab	1,13	hombre	mayor	llano/a		
Arcaláus	ac	2,34	hombre	mayor	llano/a		
Anciano-1	ac1	2,45	hombre	mayor	llano/a		
Anciano-2	ac2	2,45	hombre	mayor	llano/a		
Amadis	ad	1,00	hombre	mayor	noble		
Altisidora	al	2,44	mujer	mayor	noble		
Alifanfuón	alf	1,18	hombre	mayor	noble	int	
Ama	am	1,05	mujer	mayor	llano/a		

Título para los capítulos

Cap.	Título	Revisado
1,00	Prólogo de la primera parte	
1,01	Condición y ejercicio	Condición y ejercicio de la caballería andante
1,02	Venta-1. Castillo	Primera salida
1,03	Don Quijote armado	Don Quijote armado caballero
1,04	Andrés, mercaderes de Toledo	
1,05	Vuelta a casa	
1,06	Libros de caballería	
1,07	Salida con Sancho	Segunda salida, con Sancho
1,08	Molinos de viento	
1,09	Vizcaíno	Aventura del vizcaíno
1,10	Conversación de don Quijote y Sancho	Conversación sobre la profesión caballeresca

Lema y forma

n	3n	p	en	en	1	de la condición y ejercicio del famoso hidalgo don Quijote de la Mancha \$ ^	En	.	.
n	3n	ar	un	un	1	la condición y ejercicio del famoso hidalgo don Quijote de la Mancha \$ ^ En	un	.	m.s
n	3n	n	lugar	lugar	1	condición y ejercicio del famoso hidalgo don Quijote de la Mancha \$ ^ En un	lugar	.	m.s
n	3n	p	de	de	1	y ejercicio del famoso hidalgo don Quijote de la Mancha \$ ^ En un lugar	de	.	.
n	3n	ar	el	la	1	ejercicio del famoso hidalgo don Quijote de la Mancha \$ ^ En un lugar de	la	.	f.s
n	3n	tp	Mancha	Mancha	1	del famoso hidalgo don Quijote de la Mancha \$ ^ En un lugar de la	Mancha	.	.
n	3n	s	,	,	1	famoso hidalgo don Quijote de la Mancha \$ ^ En un lugar de la Mancha	,	.	.
n	3n	p	de	de	1	hidalgo don Quijote de la Mancha \$ ^ En un lugar de la Mancha,	de	.	.
n	3n	r	cuyo	cuyo	1	don Quijote de la Mancha \$ ^ En un lugar de la Mancha, de	cuyo	.	m.s
n	3n	n	nombre	nombre	1	Quijote de la Mancha \$ ^ En un lugar de la Mancha, de cuyo	nombre	.	m.s
n	3n	ng	no	no	1	de la Mancha \$ ^ En un lugar de la Mancha, de cuyo nombre	no	.	.
n	3n	v	querer	quiero	1	la Mancha \$ ^ En un lugar de la Mancha, de cuyo nombre no	quiero	.	i.pr.1
n	3n	v	acordar	acordarme	1	Mancha \$ ^ En un lugar de la Mancha, de cuyo nombre no quiero	acordarme	.	inf

La revisión y la herramienta de consulta: LYNEAL

Generación automática del borrador de lemas y formas

acercar (v) acercando [g] 3, acercado [pp.m.s] 1, acercó [i.pi.3s] 2 <6>.

acero (n) acero [m.s] 9, aceros [m.p] 1 <10>.

acertadísimo (a) acertadísimo [m.s] 1.

acertar (v) acertar [if] 11, acertando [g] 1, acertado [pp.m.s] 22, acertados [pp.m.p] 1, acertada [pp.f.s] 3, acertadas [pp.f.p] 1, acierta [i.pr.3s] 3, acertáis [i.pr.2p] 1, aciertan [i.pr.3p] 2, acertaba [i.im.1s] 1, acertaba [i.im.3s] 2, acertaban [i.im.3p] 1, acerté [i.pi.1s] 1, acertó [i.pi.3s] 14, acertastes [i.pi.2p] 1, acertaron [i.pi.3p] 2, acertaré [i.ft.1s] 2, acertaréis [i.ft.2p] 1, acierte [s.pr.1s] 1, aciertes [s.pr.2s] 1, acierte [s.pr.3s] 1, acertase [s.ims.3s] 1, acertasen [s.ims.3p] 1, acertara [s.imr.1s] 3, acertara [s.imr.3s] 2, acertáramos [s.imr.1p] 1, acertare [s.ft.1s] 1, acertare [s.ft.3s] 4 <86>.

acogimiento (n) acogimiento [m.s] 10, acogimientos [m.p] 1 <11>.

acometedor (a) acometedor [m.s] 2, — (n) acometedores [m.p] 1 <3>.

acometer (v) acometer [if] 33, acometiendo [g] 3, acometido [pp.m.s] 3, acometidos [pp.m.p] 1, acometida [pp.f.s] 1, acomete [i.pr.3s] 5, acometemos [i.pr.1p] 1, acometen [i.pr.3p] 3, acometía [i.im.3s] 3, acometían [i.im.3p] 1, acometí [i.pi.1s] 1, acometió [i.pi.3s] 2, acometieron [i.pi.3p] 3, acometería [i.cd.3s] 1, acometa [s.pr.1s] 1, acometa [s.pr.3s] 1, acometiese [s.ims.3s] 1, acometiera [s.imr.1s] 1, acometieran [s.imr.3p] 1, acometa [ip.3s] 1 <67>.

acometimiento (n) acometimientos [m.p] 3.

acomodar (v) acomodar [if] 16, acomodando [g] 7, acomodado [pp.m.s] 15, acomodados

Proceso de revisión manual del diccionario

Consulta simultánea de varias fuentes:

- Texto indexado en LYNEAL
- Diccionario de Lengua Española (RAE) en línea
- Versión crítica del CVC

Ejemplos: errores en la categoría

liso (a) liso [m.s] 4, lisa [f.s] 2, lisas [f.p] 3 4 < 9
10 >. **liso** (n) lisas [f.p] 1. Es adjetivo.

Ejemplos: errores en el lema

posadero**oas** (n) posaderas [f.p] 5. Sinónimo de *nalgas*. @no tiene nada que ver con el posadero. El lema es en fp, como sale en el DLE.

Ejemplos: errores en variante arcaica

----- \ / ----- LETU 1 2 ----- LETU 1 2 -----
ledanías (n) ledanías [f.p] 1. Es variante arcaica
de “letanía”. Mover

Ejemplos: ampliación de significado

mohatra (n) mohatra [f.s] 1. Caballero de moa-
tra, hombre que aparentaba sin serlo (DLE).

Ejemplos: corrección de la crítica

invenerable (a) invenerable [c.s] 1. La crítica
corrige el posible error como “invulnerable”.
DLE no reconoce esta palabra.

Don Quijote en LYNEAL

[MENÚ] [Datos] [Letras] [Números] [Figuras]

■ MENÚ

Patrón: Ejecutar >> [Ctrl]+[Enter]

#Antonio Moreno

● Datos

● Letras

● Números

● Figuras

● Información

Inicio ● Language: Español ▼

LYNEAL

ver. 2023-02-10

■ Letras: Output: 8 / Límite de output:100000

Ventana / Copiar

2: Con esto se acabaron las preguntas y las respuestas, pero no se acabó la admiración en que todos quedaron, excepto los dos amigos de don Antonio, que el caso sabían. El cual quiso Cide Hamete Benengeli describirnos para que sepamos el mundo que vivió en él.

Núm.	Contexto anterior (A).	Forma	Contexto posterior (P)
1	*s niñerías que no pueden dejar de contarse \$ Don	Antonio Moreno	se llamaba el huésped de don Quijote, caballero r*
2	*n la tal cabeza se encerraba; y así, dice que don	Antonio Moreno	, a imitación de otra cabeza que vio en Madrid, fa*
3	*ea de burlas. \$ En resolución, aquella tarde don	Antonio Moreno	, su huésped, y sus dos amigos, con don Quijote y *
4	*, en este parecer, se desembarcó el virrey, y don	Antonio Moreno	se llevó consigo a la morisca y a su padre, encar*
5	*ta entonces le habían sucedido \$ La mujer de don	Antonio Moreno	cuenta la historia que recibió grandísimo content*
6	*ndo sería alguna nueva aventura fabricada por don	Antonio Moreno	, o por otro algún caballero de la ciudad, salió l*
7	*de Don Gregorio, y de otros sucesos \$ Siguió don	Antonio Moreno	al Caballero de la Blanca Luna, y siguiéronle tam*
8	*anda como en las bodas del rico Camacho, y de don	Antonio Moreno	; pero consideraba no ser posible ser siempre de d*

■ Números: 1. FA: Frecuencia absoluta (*) - 1. Puntuación

Número de filas: 1 / Número de columnas: 1

Ventana / Copiar con formato / Copiar sin formato

→ FA Todos *Reiniciar*

1 #Antonio Moreno 8 #Antonio Moreno

Reiniciar Todos

<http://shimoda.llf.uam.es/ueda/lyneal/quijote.htm>

La publicación

Ventajas de la edición en pdf

Versión **gratuita y descargable** en PDF:

- No ha limitación de páginas ni de tamaño de las fuentes
- Se pueden generar versiones corregidas de manera periódica
- Comodidad para usar en combinación con LYNEAL

Aprovechamiento de la base de datos

Diferentes versiones del diccionario:

- Lemas y formas por personajes principales
- Topónimos y Antónimos
- Listados por categorías léxicas: nombres, verbos...

Publicación electrónica del IULCE-UAM

Nueva colección en formato electrónico, con DOI.

No habrá coste de descarga

Uso en combinación con la herramienta de búsqueda, DQ en
LYNEAL

¡Gracias por su atención!

Antonio Moreno Sandoval
antonio.msandoval@uam.es

<http://www.llf.uam.es/es/>



<http://www.llf.uam.es/es/>



[Información](#)

[Recursos](#)

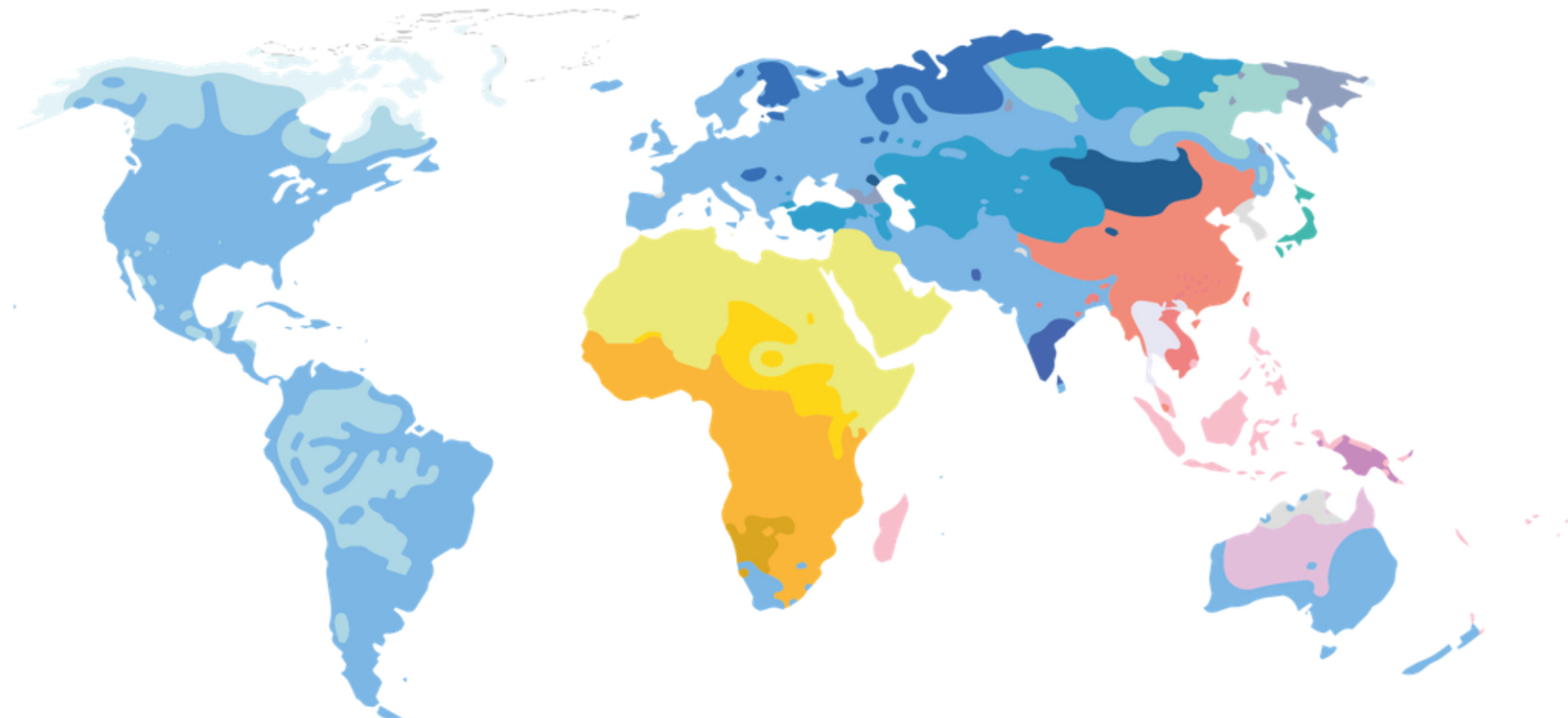
[Proyectos](#)

[Publicaciones](#)

[Personal](#)

[Actualidad](#)

Laboratorio de Lingüística Informática



Fuente: Wikipedia "Familia de Lenguas"