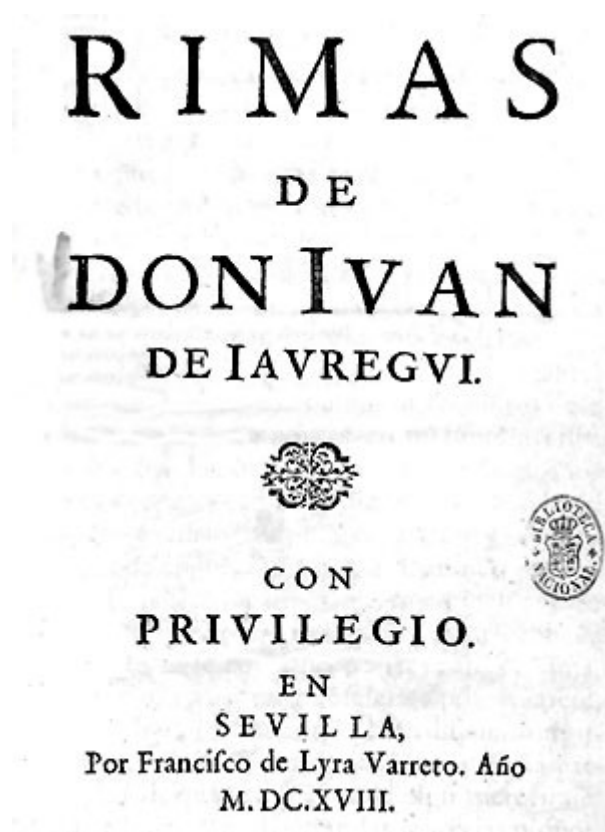




Enrique García Arias

# Humanidades Digitales: Analizando los Sonetos de Juan de Jáuregui



*Portada de la primera edición de rimas (Sevilla 1618)*



*Este documento constituye una aplicación de los métodos y técnicas de las Humanidades Digitales al análisis literario y se deriva del Trabajo Final que he presentado para la asignatura Introducción a las Humanidades Digitales incluida en el Máster Universitario en Humanidades Digitales: Métodos y Buenas Prácticas de la UNED.*

## ÍNDICE

1. Corpus escogido y justificación.....	3
2. Objetivos de investigación.....	3
3. Análisis.....	4
3.1. Preparación del Corpus.....	4
3.2. Expresiones regulares.....	6
3.3. Voyant Tools.....	8
3.4. Recogito.....	20
4. Resultados.....	23
5. Conclusiones.....	25
6. Bibliografía.....	26

## 1. Introducción y Objetivos

El objetivo de esta investigación consiste en identificar **¿Cómo los métodos y prácticas de Humanidades Digitales, en particular el uso de Expresiones Regulares, Voyant Tools y Recogido, posibilitan la identificación y representación de los principales temas de análisis aplicables a los sonetos de Juan de Jáuregui?**.

Para fundamentar esta investigación, se ha llevado a cabo una exhaustiva revisión previa sobre el autor y su obra, identificando elementos clave para este estudio, como la influencia italianizante, el empleo de la mitología, la dualidad de facetas como pintor y poeta, y el uso del cultismo. A partir de esta información, se formulan las siguientes hipótesis de investigación:

H1. Las Humanidades Digitales permiten destacar y facilitar el análisis de las tendencias literarias presentes en la obra de Juan de Jáuregui.

H2. Las Humanidades Digitales contribuyen de manera significativa a una comprensión más profunda de la forma y el contenido de la obra al presentar la información de manera efectiva.

H3. Las Humanidades Digitales tienen el potencial de estimular la participación activa y la colaboración en esta actividad de investigación.

## 2. Corpus escogido y justificación

Se ha realizado una primera selección centrada en el periodo del Siglo de Oro, incluyendo a los autores Juan de Arguijo (Sevilla 1567 – 1622) y Juan de Jáuregui (Sevilla 1583 – Madrid 1641).

La elección de este período se fundamenta, en parte, en la lectura del artículo "Humanidades Digitales y Literatura española: 50 años de repaso histórico y panorámica de proyectos representativos"(Hernández Lorenzo, 2020). Este artículo, con especial atención al Siglo de Oro, destaca el predominio de estudios relacionados con esta época, indicando referencias de Antonio Rojas Castro o el proyecto ADSO (Análisis distante del soneto castellano de los Siglos de Oro) (Herrero, 2018) entre otros, por lo que he considerado que el disponer de suficientes fuentes de información puede orientar mi introducción a un área hasta ahora desconocida para mí, como es el análisis de la literatura.

Otra motivación surge del contrapunto que representa el Siglo de Oro entre la brillante creación literaria y la decadencia del Imperio español, específicamente durante el reinado de Felipe IV y su valido, el Gran Duque de Olivares, un contexto histórico que me interesa y que desde un punto de vista literario supone la introducción de elementos italianos, como el soneto, y la recuperación de las fuentes clásicas.

En este contexto, considero que Juan de Jáuregui se erige como un exponente particularmente relevante pues exhibe dos tendencias de interés. Por un lado, se percibe una influencia italianizante, el petrarquismo y un tratamiento distintivo del amor. Por otro lado, se manifiesta el culteranismo a través del uso de mitología, metáforas, hipérboles y antítesis. En este sentido, resulta interesante analizar la progresión que se produce en su obra desde la vertiente italianizante hacia la culterana. Dicha evolución también encuentra eco en las disputas que Jáuregui sostuvo con los principales autores contemporáneos; se opuso a Góngora, pero también se enfrentó a Lope y atacó a Quevedo.

Un hito en la polémica gongorina lo supuso la difusión del Antídoto contra la pestilente poesía de las “Soledades”. Jáuregui se aparta pero también se acerca a Góngora, no se opone a los cultismos sino a su abuso pero es interesante analizar la coherencia entre la teoría y la práctica.

Otro aspecto de interés del autor es su faceta como pintor y cómo pudo influir o tuvo reflejo en su obra. Todos estos aspectos me han motivado a tomar como base de partida el corpus de los sonetos de Juan de Jáuregui.

### 3. Análisis

#### 3.1. Preparación del Corpus

Esta tarea reviste gran relevancia, ya que a partir de la información base constituida por el corpus se generarán los resultados finales. Por ello, la verificación de las fuentes y el enriquecimiento semántico de estructuras de datos estándar, como el formato TEI-XML, se consideran prioritarios para posibilitar una mayor interoperabilidad y explotación de los resultados.

EL resultado de esta tarea está disponible en el repositorio GitHub en la dirección: “[https://github.com/enriuegarciaarias/introHD\\_tarea\\_final](https://github.com/enriuegarciaarias/introHD_tarea_final)”

La información de partida para proceder a la preparación del Corpus ha consistido en:

- Archivo jauregui\_todoslossonetos.txt un único fichero que encadena todos los sonetos sin diferenciación.
- Archivo jauregui\_cadasoneto.zip fichero comprimido con la carpeta disco089g incluyendo un archivo por soneto. Estos archivos individuales se nombran con la convención “disco089g\_xxxx.txt” donde “xxxx” abarca el rango de 161 a 183

Las tareas a realizar incluyen, en primer lugar, la verificación de las fuentes para confirmar el origen de los archivos y realizar posibles correcciones o ampliaciones de esta información recibida como datos iniciales. En segundo lugar, se llevarán a cabo los procesos de tratamiento de datos necesarios para generar ficheros de salida elaborados.

#### Verificación de fuentes

Se identifican varias fuentes:

- Corpus DISCO The Diachronic Spanish Sonnet Corpus que contiene los sonetos en TEI-XML entre las siglos XV y XX (4530 sonetos de 1216 autores de 22 países)(*corpus DISCO*, s. f.)
- Corpus del proyecto ADSO desarrollado por la universidad de Alicante con fondos de la fundación BBVA. Incluye entre otros autores los 23 Sonetos del corpus facilitado de Juan de Jáuregui, en formato TEI-XML(Navarro-Colorado, 2020).

- Poesía Edición de Juan Matas Caballero(Jáuregui y Aguilar & Matas Caballero, 1993)

Se identifican otros sonetos principalmente en la fuente de Juan Matas Caballero, pero que no se incluyen en el corpus.

- Soneto para la justa en honor a la Inmaculada Concepción de 1616
- Dos Sonetos dedicados a D. Rodrigo Calderón de 1621
- Soneto para la Justa en honor a los santos Ignacio de Loyola y Francisco Javier
- Soneto inspirado en la hazaña del príncipe Baltasar Carlos al matar un toro bravo de un arcabuzazo en 1631.
- Soneto al santo fray Luis de Granada.
- Soneto de título “A una mujer a quien un hombre no pudiendo gozarla viva mató a puñaladas para gozarla muerta”
- Soneto de título “Al mismo intento” continuación del anterior.

### Proceso de los textos

La segunda fase implica la complementación del corpus proporcionado con los títulos de los sonetos. Dado que es necesario incorporar cierta estructura TEI-XML a los mismos, se ha decidido desarrollar un programa en Python con las siguientes funcionalidades:

- Leer los sonetos individuales del corpus proporcionado.
- Asociar el título correspondiente a cada soneto.
- Identificar y extraer la información sobre las rimas presentes.
- Modificar la estructura de las líneas de soneto según los requisitos establecidos.
- Agregar elementos TEI-XML necesarios, completando un fichero con la estructura estándar de TEI-XML.
- Guardar un único fichero de salida que contenga toda la información en formato TEI-XML.
- Generar un fichero de salida adicional con las palabras que riman.

Toda la información se ha publicado en github en la dirección:

[https://github.com/enriquegarciaarias/introHD\\_tarea\\_final](https://github.com/enriquegarciaarias/introHD_tarea_final)

Su contenido resumido es:

- Programa python

[https://github.com/enriquegarciaarias/introHD\\_tarea\\_final/blob/main/CorpusToTEI.py](https://github.com/enriquegarciaarias/introHD_tarea_final/blob/main/CorpusToTEI.py)

- Branch Resultados contiene los dos ficheros resultado del proceso.
- Branch Sonetos con el corpus facilitado.
- Branch Work con archivos de trabajo necesarios para la ejecución del proceso.

## 3.2. Expresiones regulares

La utilización de expresiones regulares en el programa Python se resume en:

### 1. Separación del texto y la última palabra en cada línea de los sonetos

Con el objetivo de extraer las palabras que riman y facilitar la composición de la nueva línea en el archivo de salida en formato TEI-XML, se implementan varios comandos que incluyen el uso de expresiones regulares.

```
def proc_regex(d):
    d = re.sub(r"\r?\n?", "", d) # reemplaza carriage return y line feed
    d = re.sub("^[ ]+|[ ]+$", "", d) # reemplaza espacios al inicio y al final
    pattern = r"^(.*)\b(\w+)([\^\\w]+)?$" # localiza la última palabra
    match = re.search(pattern, d)
    return match
```

### 2. Extracción de títulos del fichero de trabajo

Para integrar los títulos en el archivo TEI-XML generado, el programa se basa en un archivo de trabajo que incluye la asociación de títulos con archivos, siguiendo el modelo:

```
<file>disco089g_0161</file><title>A San Ignacio de Loyola</title>
```

Para extraer el título correspondiente a cada archivo (soneto) leído del corpus proporcionado, la expresión regular utilizada en la línea 29 del código es:

```
pattern = r"^<file>([^\<]+)</file><title>([^\<]+)</title>$"
```

### 3. Personalización del fichero de cabeceras TEI-XML

El propósito es finalizar la creación del archivo TEI-XML con una estructura estándar, incluyendo las cabeceras. Para hacer este proceso modular y reutilizable, se emplea un archivo TEI-XML "plantilla de cabecera" con el formato:

```
<edition>Edición digital de <date>**</date></edition>
```

Esto permite personalizar las cabeceras TEI-XML mediante el uso de las funciones re.search y re.sub en el código:

```
def fill_teihead():
    ct = datetime.datetime.now().isoformat()
    cust = [[r"<title>|*|<", "Sonetos"],
            [r"<author>|*|<", "Jáuregui, Juan de"],
            [r"<date>|*|<", ct],
            [r"<sourceDesc>|*|<,"
```

```
'<p>Jáuregui, Juan de. <title>Poesía</title>. Ed. De Juan Matas Caballero.  
Madrid: Cátedra, 1993. Impreso.</p>']]  
with open('tei_headers.txt', 'r') as h:  
    ls = h.readlines()  
    headers = []  
    for li in ls:  
        for c in custom:  
            if re.search(c[0], li):  
                print(c[0])  
                li = re.sub(r"\\*", c[1], li)  
            headers.append(li)  
    return headers
```

Al ejecutar el programa, se visualizan en la consola los resultados, que incluyen el número de sonetos (23), la cantidad de registros TEI grabados (601) y la cantidad de registros de rimas grabados (322). Posteriormente, se lleva a cabo la grabación de dos archivos, los cuales se almacenan en la rama "Resultados" (un subdirectorío cuando el programa se ejecuta localmente). Estos archivos son jauregui\_rimas.txt y jauregui\_tei.xml.

### 3.3. Voyant Tools

En esta sección, se describen de manera concisa los procesos llevados a cabo utilizando la herramienta web Voyant Tools. En el primer paso, se procedió a incorporar el corpus y crear una lista de "palabras vacías", la cual se utilizó en la interfaz de la herramienta. La dirección pública del interfaz de Voyant Tools, con estos datos específicos, se encuentra disponible en el siguiente enlace:

<https://voyant-tools.org/?corpus=a3f64b3fec2d36bfl8d6997a966430>

En la parte inferior izquierda de la interfaz, el módulo de Resumen (Summary) proporciona información cuantitativa sobre las características del corpus:

El corpus está conformado por 22 textos, que en su conjunto reúnen un total de 1.982 palabras o tokens y 944 palabras únicas o types.

*Tabla 1: Valores de Riqueza Léxica y Extensión de los documentos*

Métrica	Documento y Valor	Título
Mayor densidad de vocabulario	Disco089g_0164 (0.835)*	A un navío destrozado
Menor densidad de vocabulario	Disco089g_0179 (0.633)*	A la invención de la cruz
Más largo	Disco089g_0166 (101)**	Al sol, amaneciendo
Más corto	Disco089g_0168 (83)**	A Mucio Cévola

\* Ratio de palabras únicas en este documento

\*\* Número de palabras en este documento

La densidad de vocabulario, que indica la riqueza léxica, se obtiene dividiendo el número de palabras únicas (U) entre el número total de palabras (T) mediante la fórmula:  $100 \times U / T$ . Un valor más elevado refleja una mayor diversidad léxica, mientras que un valor más bajo sugiere la repetición de palabras y una menor variabilidad en el uso del vocabulario. En el caso de los sonetos del corpus, se observa un rango de 0.63 a 0.83 para esta métrica. Estos valores son notables, ya que los sonetos, en general, buscan expresar ideas con concisión.

Es opinión generalizada que el soneto desarrolla un único asunto lo que obliga a una intensa concentración en el tema sin divagaciones, y así lo hace Jauregui (Matas Caballero, 1990).

Los índices de extensión de los documentos varían entre 83, el más corto en el soneto "A Mucio Cévola", y 101, presente en el soneto "Al sol amaneciendo". Dada la aplicación uniforme de la estructura del soneto en todos los documentos, se aprecia poca dispersión en estos valores.

En cuanto al índice de palabras por oración, se destaca la significativa superioridad en los sonetos "Epitafio a las ruinas de Roma", "En el túmulo que fabricó Sevilla a la reina Doña Margarita" y "A la hazaña de don Álvaro Pérez de Guzmán el Bueno, en la defensa de Tarifa", con un promedio de palabras superior a 88, en comparación con el resto que se sitúa por debajo de 51.



## Supresión de “palabras vacías”



Figura 1: Nube de palabras "Cirrus" tras excluir las palabras vacías

Tras la exclusión de las palabras vacías, carentes de significado semántico, surgen patrones destacados tanto en la nube de palabras "Cirrus" como en la tabla de términos. La palabra "cielo" destaca con 9 apariciones, seguida por "vida" con 8, y luego "pecho", "muerte", "hoy" y "amor", cada una con 6 apariciones. Esto sugiere una singularidad y relevancia particulares de "cielo" y "vida", alineándose con las motivaciones barrocas vinculadas a temas religiosos, la efímera naturaleza del tiempo y el sentido de la existencia.

Las palabras que siguen con 6 apariciones también reflejan estos mismos ideales, enfatizando especialmente "amor" y "pecho", donde se pueden identificar influencias petrarquistas. Se fortalece el sentido barroco de la muerte y el transcurso del tiempo a través de las relaciones "ayer", "hoy" y "mañana". Para obtener una comprensión más detallada de estas métricas, se proporciona información adicional a continuación.

Tabla 2: Palabras más utilizadas tras excluir las palabras vacías

Términos	Contar	Relativo	Coeficiente	Oblicuidad
cielo	9	0,0045	-2,0222	0,4032
vida	8	0,0040	1,0383	1,5218
pecho	6	0,0030	-0,8159	1,1139
muerte	6	0,0030	2,3136	1,8061
hoy	6	0,0030	3,6594	2,0180
amor	6	0,0030	3,7016	2,0244
virgen	5	0,0025	14,0585	3,6252
siempre	5	0,0025	-0,0310	1,4041
pudo	5	0,0025	14,1445	3,6362
mano	5	0,0025	6,5183	2,5451
fuego	5	0,0025	-0,0028	1,4097
cruz	5	0,0025	22,0000	4,6904
voz	4	0,0020	1,2987	1,7816
luz	4	0,0020	8,0574	2,8652
fuerte	4	0,0020	8,4254	2,9103
feliz	4	0,0020	10,2394	3,1409
culpa	4	0,0020	8,1702	3,0677
ausencia	4	0,0020	8,7298	2,9502

La asimetría estadística, nos proporciona información sobre cómo se distribuyen las palabras en un conjunto de texto, una asimetría por encima de 1 nos indica una presencia irregular, mas acusada en pocos documentos, cuanto más cercana a cero nos indica que la presencia de éste término es más regular. Destaca en este sentido la asimetría de estos términos, todos por encima de 1 excepto el término cielo que tiene una presencia más regular.

### Sin suprimir las palabras vacías

En este caso no se aplica la lista de palabras vacías o “stopwords”, los procesos se aplican sobre el conjunto del texto original.



*Figura 2: Nube de palabras sin excluír palabras vacías*

Los 7 términos más frecuentemente empleados (y, el, la, de, que, en, su) exhiben una asimetría estadística consistentemente inferior a 1, lo que sugiere una notable regularidad en la utilización de estos términos a lo largo de los sonetos. Destaca la asimetría estadística en la utilización de los términos "no", "tu" y "mi", la cual podría estar relacionada con una preferencia por un lenguaje menos directo, favoreciendo un estilo retórico y afirmativo. Este énfasis en la afirmación se manifiesta también en la representación intensificada del dolor y el sufrimiento emocional, así como en la exploración de temas de moralidad y religiosidad.

*Tabla 3: Palabras más utilizadas sin excluir las palabras vacías*

Términos	Contar	Relativo	Coeficiente	Oblicuidad
y	105	0,0530	-0,1338	-0,8564
el	92	0,0464	-0,7516	0,3807
la	71	0,0358	-0,6727	0,3825
de	58	0,0293	1,4431	0,9327
que	56	0,0283	-0,3524	0,4293
en	46	0,0232	0,3685	0,3227
su	39	0,0197	-0,1862	0,5864

al	37	0,0187	3,4515	1,3273
a	32	0,0161	-0,9700	0,3323
del	30	0,0151	0,6743	1,2043
no	14	0,0071	2,4342	1,6234
con	13	0,0066	1,9686	1,5193
se	12	0,0061	-0,4406	0,9523
tu	11	0,0055	2,9501	1,7532
las	11	0,0055	0,0934	1,0478
ya	10	0,0050	0,0426	1,2445
mi	10	0,0050	3,6322	2,0035
un	9	0,0045	0,6811	1,4767
si	9	0,0045	0,8462	1,3908
cielo	9	0,0045	-2,0222	0,4032
vida	8	0,0040	1,0383	1,5218

La conjunción copulativa “y” puede indicar el uso de un enfoque narrativo así como la acumulación de ideas similares, polisemias y contrastes creando sensación de riqueza, profusión, ornamentación y densidad semántica. El valor de la asimetría estadística nos indica una presencia regular de estos primeros 7 términos en el conjunto de sonetos.

*Tabla 4: Contextos de uso de la conjunción copulativa "y"*

Izquierda	Términos	Derecha
al pueblo amado, el fuego	y	el granizo ved mezclado, porque
ángel fuerte, de pureza armado,	y	en procelosas ondas arrojado, a
Pasó la primavera	y	el verano de mi esperanza
el verano de mi esperanza,	y	el agravio mío en la
turbado pecho pierde el brío	y	se entorpece la cobarde mano
reliquias mueran de memorias mías.	y	el desengaño, como fénix rara
acosado Antonio, al fuerte Augusto	y	a Cleopatra mira; una al
su favor conspira del medo	y	persa, egipcio y macedonio Puede
del medo y persa, egipcio	y	macedonio Puede triunfar de Augusto
puede vencer astuto su malicia	y	arte Trueca la acción; y

Los artículos el y la pueden relacionarse con las descripciones detalladas y enriquecidas presentes en la literatura barroca y donde también se presta una mayor atención a la forma y el estilo refinado y en ocasiones sofisticado.

Tabla 5: Contextos de uso de los artículos "el" y "la"

Izquierda	Términos	Derecha
columna que descubre senda en	el	piélago rubro al pueblo amado
piélago rubro al pueblo amado,	el	fuego y el granizo ved
pueblo amado, el fuego y	el	granizo ved mezclado, porque el
el granizo ved mezclado, porque	el	egipcio su amenaza entienda. Ved
su amenaza entienda. Ved sobre	el	mar, porque su golfo encienda
mar, porque su golfo encienda,	el	ángel fuerte, de pureza armado
Pasó	la	primavera y el verano de
y el agravio mío en	la	estéril sazón del seco estío
el brío y se entorpece	la	cobarde mano Mas la razón
entorpece la cobarde mano Mas	la	razón, que mi derecho ampara
su malicia y arte Trueca	la	acción; y del contrario huyendo
índice remoto. Ausente yace de	la	selva cara, do el verde

## Palabras diferenciadas

Las palabras diferenciadas son los términos que hacen aparecen con mayor frecuencia en un documento en relación con el resto y que por tanto nos permiten identificar temas o patrones.

Archivo	Palabras y valores									
disco089g_0161:	veréis	1	tantos	1	santos	1	rubro	1	pureza	1
disco089g_0162:	vulcano	1	verano	1	turbado	1	sagaz	1	renazca	1
disco089g_0163:	huyendo	2	cleopatra	2	augusto	2	vil	1	victorias	1
disco089g_0164:	vio	1	verde	1	vasto	1	trajo	1	selva	1
disco089g_0165:	yerto	1	vivaz	1	vistió	1	vientre	1	verdes	1
disco089g_0166:	sólo	2	sombras	2	punto	2	noche	2	rubio	1
disco089g_0167:	pena	2	vuestra	1	trueco	1	tormento	1	todos	1
disco089g_0168:	diestra	2	pudo	3	feliz	2	mano	2	varón	1
disco089g_0169:	hijo	2	fuerte	2	voces	1	vea	1	ti	1
disco089g_0170:	aquí	2	vulto	1	vuelo	1	tierras	1	tesoros	1
disco089g_0171:	cerca	2	verá	1	tonante	1	términos	1	tenaro	1
disco089g_0172:	envidia	2	culpa	2	vicio	1	venganza	1	veneno	1
disco089g_0173:	lisipo	2	escultor	2	marte	2	arte	2	ves	1
disco089g_0174:	alma	2	vandalia	1	universo	1	tristeza	1	pretende	1

disco089g_0175:	guarecida	1	engañóme	1	ecceso	1	viví	1	siguiendo	1
disco089g_0176:	olvido	2	ausencia	2	voluntad	1	tormentos	1	teme	1
disco089g_0177:	fe	2	venus	1	sentir	1	remota	1	quisiera	1
disco089g_0178:	luz	2	verbo	1	ve	1	sonoro	1	simeón	1
disco089g_0179:	cruz	5	cristo	3	resucita	2	leño	2	halla	2
disco089g_0180:	eterna	2	doce	2	zona	1	signos	1	sacra	1
disco089g_0181:	esposo	2	yelo	1	vuela	1	vivas	1	unida	1
disco089g_0182:	sacerdocio	2	limpieza	2	virgen	3	culpa	2	yaces	1
disco089g_0183:	guarnecida	1	extiende	1	engañome	1	viví	1	siguiendo	1

### Palabras más utilizadas

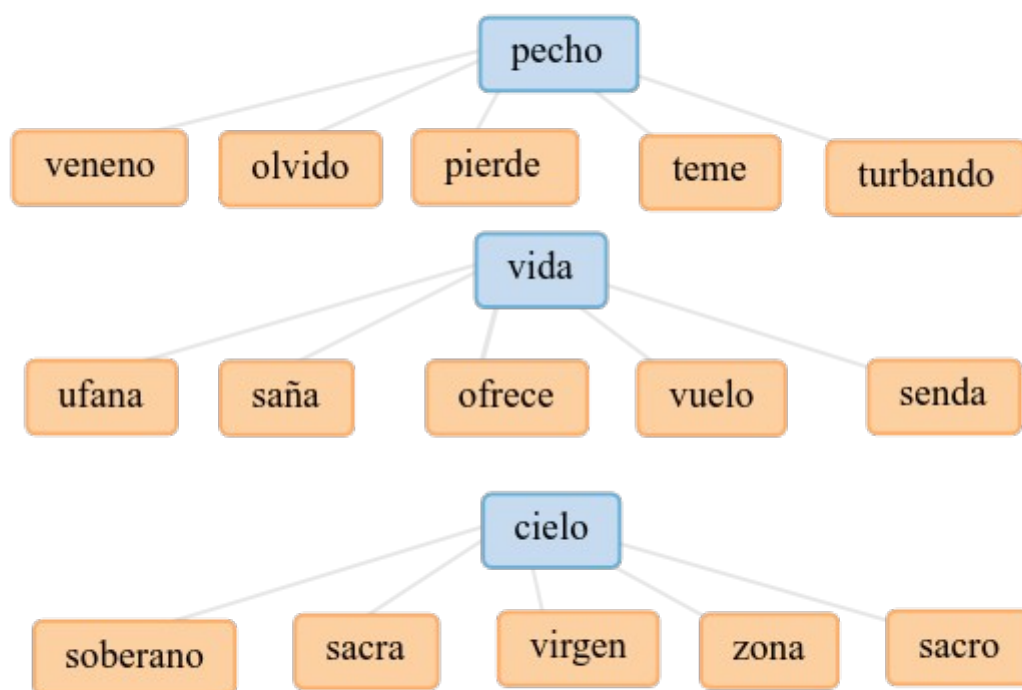


Figura 3: Vista de la pestaña Enlaces

Esta visualización nos presenta los términos más recurrentes, mostrando su relación con aquellos que aparecen en mayor proximidad. Esto amplía el contexto y el significado de los términos predominantes, destacando especialmente "pecho", "vida" y "cielo", los cuales revelan los temas centrales en la poesía de Juan de Jáuregui, desarrollados con más detalle posteriormente.

Es interesante observar que los términos relacionados con estos tres "principales" no se comparten entre sí, reflejando así la presencia de tres temáticas distintas, cada una con un fuerte protagonismo y contenido propio.

En el análisis detallado de la distribución, que utilizaremos para examinar estos tres términos principales, se incluyen los "Collocates", términos que tienden a aparecer junto al término principal, y los "Correlations", términos con una correlación contextual elevada respecto al término principal.

Explorando los contextos de cada término, podemos estudiar cómo se emplea en diferentes situaciones. La herramienta proporciona información sobre el texto adyacente a la izquierda y a la derecha de cada término en cada documento (soneto). También es evidente cuando el término forma parte del principio o del final de la frase. Toda esta información nos permite identificar patrones y significados en profundidad.

## CIELO

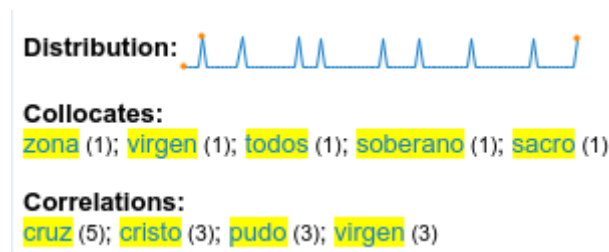


Figura 4: Detalle de la distribución y correlaciones del término "cielo"

Tabla 6: Contextos del término "cielo"

Soneto	Izquierda	Térm.	Derecha
A la edad del año	en plácidos amores gozando,	cielo	, su amador constante, de las
Soneto 167	amarrado, a eterno giro el	cielo	te condena, y llevad todos
Epitafio a las ruinas de Roma	sacro imperio, y al empíreo	cielo	
Condena al fabricar soberbios palacios	se levante máquina excelsa, al	cielo	convecina, tanto más cerca atiende
En una estatua del rey Filipo III, esculpida por insigne artífice toscano	soberano reino te aumenta el	cielo	, gran Filipo, así te da
En el túmulo que fabricó Sevilla a la Reina Doña Margarita	cenizas envolver pretende; y el	cielo	puro, que atesora el alma
A la coronación de Nuestra Señora	de luciente zona, que el	cielo	os ciñan de la sacra

A la paloma que salió de la boca desta Santa en su muerte	se encumbra; ya penetra el	cielo	¡Oh, cuánto la remira el
Soneto 182	y fiera, hasta que el	cielo	resolvió, aplacado, que una Virgen

Esta información nos permite resaltar los principales temas que abarcan lo Sacro, Laudatorio y Ético-moral. Se observa que numerosos poemas están dedicados a San Ignacio y a la Virgen, y muchas de estas composiciones de índole religiosa fueron concebidas para participar en certámenes, adoptando un tono más circunstancial y menos íntimo.

En el tema laudatorio, se aprecia una variante elegíaca donde se manifiesta el sentimiento de frustración característico del barroco frente al inexorable paso del tiempo, junto con un mensaje cristiano de liberación del cuerpo.

*Tabla 7: Temáticas de los sonetos con el término "cielo"*

Soneto	Temáticas
A la edad del año	Destaca el paso del tiempo y la lección moral
Soneto 167	Temática petrarquista de énfasis del dolor
Epitafio a las ruinas de Roma	Paso del tiempo y desengaño barroco
Condena al fabricar soberbios palacios	Pensamiento ético y contraste temporal
En una estatua del rey Filipo III, esculpida por insigne artífice toscano	Laudatorio al rey Felipe III
En el túmulo que fabricó Sevilla a la Reina Doña Margarita	Laudatorio a la muerte de la reina Margarita
A la coronación de Nuestra Señora	Rima sacra
A la paloma que salió de la boca desta Santa en su muerte	Rima sacra
Soneto 182	Rima sacra en honor a la inmaculada concepción

## VIDA



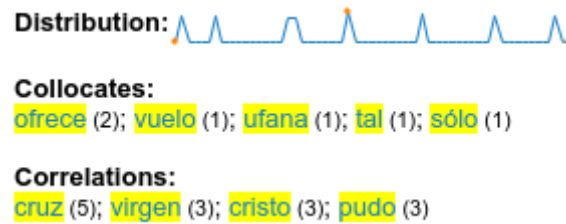


Figura 5: Detalle de la distribución y correlaciones del término "vida"

Tabla 8: Contextos del término "vida"

Soneto	Izquierda	Térm.	Derecha
A San Ignacio de Loyola	ved el arca ufana, que	vida	ofrece en la borrasca a
Al sol, amaneciendo	el día, tal que de	vida	un punto no me debas
Al sol, amaneciendo	del ausencia mía: que es	vida	sólo el tiempo que me
A Mucio Cévola	admiración cambia la saña, la	vida	al enemigo reservando, que para
Epitafio a las ruinas de Roma	en encendidas guerras, porque su	vida	renovase el vuelo Y si
Soneto 176	mis males, me reserva en	vida	: que es mayor mal que
En el día de la presentación	ofrece, al tiempo que la	vida	tiene, y el mismo Espíritu
En el día de la presentación	senda nos muestras a la	vida	ausente!) llegue, y en paz

Destacamos los principales temas, como lo Sacro, el inexorable paso del Tiempo y la marcada influencia Petrarquista.

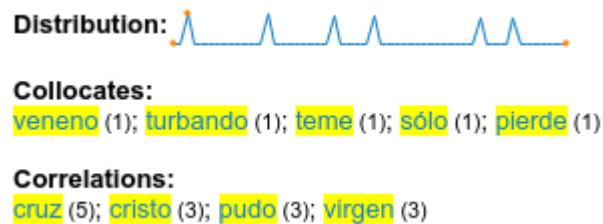
En "Al sol, amaneciendo", se evidencia claramente una influencia petrarquista-herreriana, explorando el sufrimiento asociado al implacable transcurrir del tiempo, tema que también se refleja en "Epitafio a las ruinas de Roma". Asimismo, resalta el Soneto 176, donde se aborda de manera clásica el motivo petrarquista, con la ausencia y el desdén cruel de la amada como causa del sufrimiento.

Tabla 9: Temáticas de los sonetos con el término "vida"

Soneto	Temáticas
A San Ignacio de Loyola	Rima sacra
Al sol, amaneciendo	Rima petrarquista, enfrentamiento amor y tiempo Paso del tiempo y desengaño barroco
A Mucio Cévola	lectura moral, con valor sacrifica su vida por un ideal
Epitafio a las ruinas de Roma	Desengaño del hombre barroco, paso del tiempo reflejado en las piedras convertidas en polvo

Soneto 176	Petrarquista de la ausencia y el desden de la amada
En el día de la presentación	Rima sacra el hijo engendrado 2ª persona de la santísima trinidad

## PECHO



*Figura 6: Detalle de la distribución y correlaciones del término "pecho"*

*Tabla 10: Contextos del término "pecho"*

Soneto	Izquierda	Térm.	Derecha
Un amante, abrazando las prendas de su dama	hielo frío, donde el turbado	pecho	pierde el brío y se
Al sol, amaneciendo	el bello objeto que mi	pecho	adora sólo le gozo entre
Soneto 167	águila y cadena, herido el	pecho	, y al Caúcaso atado. Dame
A la hazaña de don Álvaro Pérez de Guzmán el Bueno, en la defensa de Tarifa	mas el alto valor del	pecho	fuerte se aventajó con espantable
La virtud a la envidia	tu veneno empleo turbando el	pecho	que mis obras culpa; que
Soneto 176	Si en el amado	pecho	más constante teme el olvido

Resaltamos la marcada influencia Petrarquista en el tema amoroso, así como el mensaje Ético-moral.

En los temas de "amor", se observa una coincidencia entre el personaje del amante y el poeta, aunque sin la descripción de una experiencia real. Se destaca el sufrimiento barroco, que incluso supera el de los mitos, y la relación intrínseca entre el amor y el tiempo, que inevitablemente conduce a la muerte.

Tabla 11: Temáticas de los sonetos con el término "pecho"

Soneto	Temáticas
Un amante, abrasando las prendas de su dama	Desengaño amoroso petrarquista
Al sol, amaneciendo	Petrarquista enfrentamiento amor y tiempo
Soneto 167	Celos y amor, petrarquista el dolor es mayor al de los mitos
A la hazaña de don Álvaro Pérez de Guzmán el Bueno, en la defensa de Tarifa	Ejemplaridad moral
La virtud a la envidia	Moralizante
Soneto 176	Motivo petrarquista de la ausencia y también del desdén de la ama

### 3.4. Recogito

Recogito nos brinda la posibilidad de añadir anotaciones a nuestro corpus, enriqueciendo semánticamente su información con datos sobre personas, ubicaciones georeferenciadas y eventos. Estas anotaciones se estructuran mediante etiquetas.

Hemos utilizado el documento generado en la fase de "Preparación del corpus", estructurado en formato TEI-XML bajo el nombre jauregui\_tei.xml en la "branch" de resultados en el repositorio GitHub de este proyecto. El proyecto en Recogito está públicamente disponible en la siguiente dirección: <https://recogito.pelagios.org/document/hczu0toq2teqjf/part/1/edit>

Después de llevar a cabo las operaciones de anotación con la herramienta, hemos creado un nuevo documento TEI-XML que incorpora la información adicional en metadatos. Este documento está disponible en la "branch" de Resultados con el nombre jauregui\_recogito.xml.

#### Proceso de anotación semántica



Figura 7: Ejemplo de marcado de una palabra en Recogito

Una vez que nuestro archivo base en formato TEI-XML está disponible en la herramienta, accedemos a la funcionalidad "New +" ubicada en la esquina superior izquierda de la página

principal para llevar a cabo las anotaciones. Estas anotaciones se centran en términos identificativos de personas y lugares, a los cuales añadimos breves descripciones complementarias generalmente extraídas de Wikipedia. Clasificamos estas anotaciones mediante etiquetas "tags" en tres categorías que consideramos relevantes: Mitología, Histórico y Cristiano. Esta elección se fundamenta en nuestro interés por identificar la frecuencia y la forma en que Jáuregui emplea referencias clásicas y mitológicas, así como la importancia de los temas religiosos. Estos conceptos se asocian de manera significativa con referencias geográficas, y su representación visual en Recogito puede agregar un mayor significado.

A continuación, presentamos la visualización en el mapa resultante de este proceso. En el caso de los puntos, un tamaño mayor indica un mayor número de referencias, y los colores indican las clasificaciones para los elementos identificados como "places" (lugares).

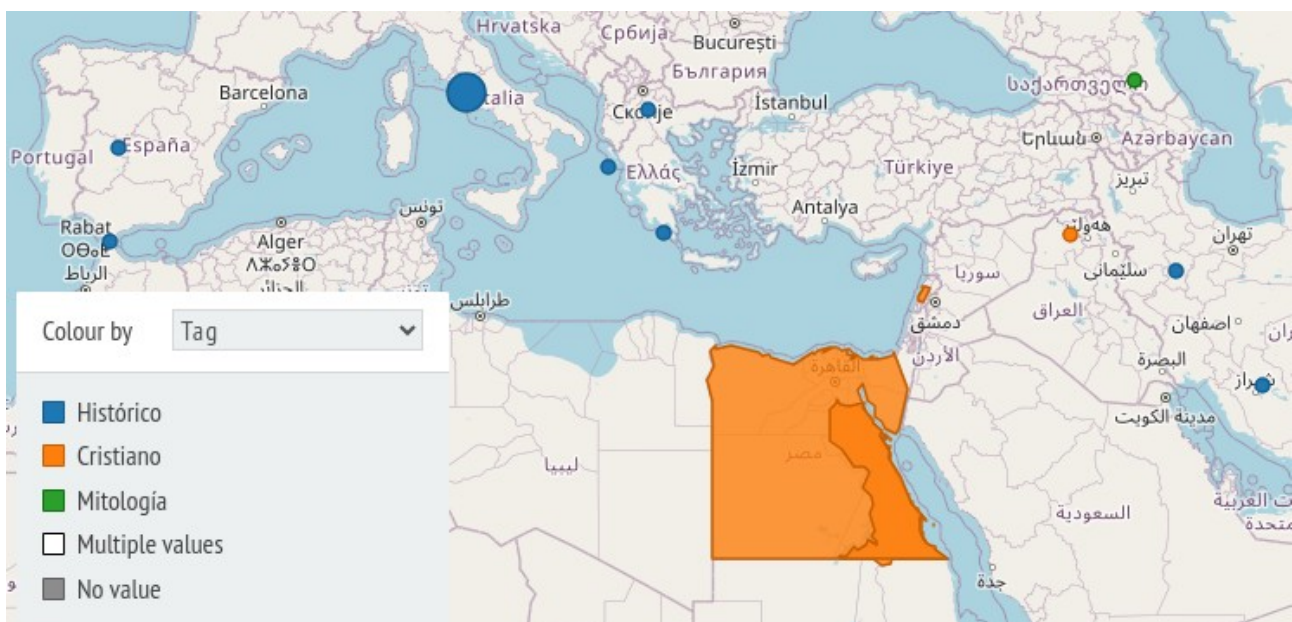


Figura 8: Representación en mapa de palabras georeferenciadas

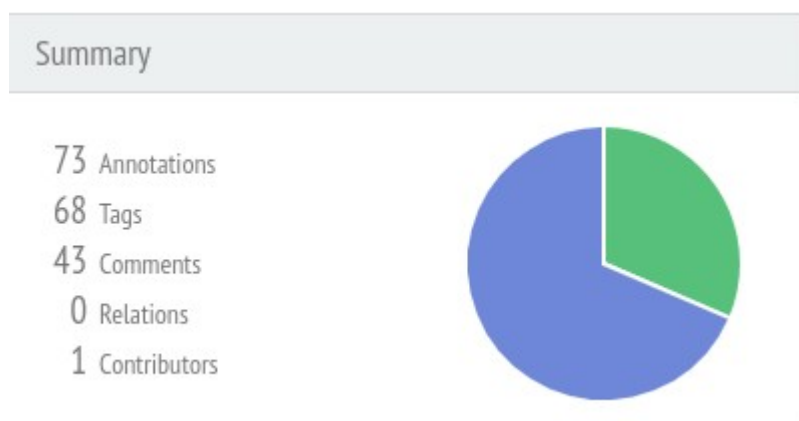


Figura 9: Estadísticas de uso en el Corpus

Es muy interesante la funcionalidad dinámica de mostrar el contexto asociado a la ubicación mediante una extracción del texto y el enlace al texto fuente.

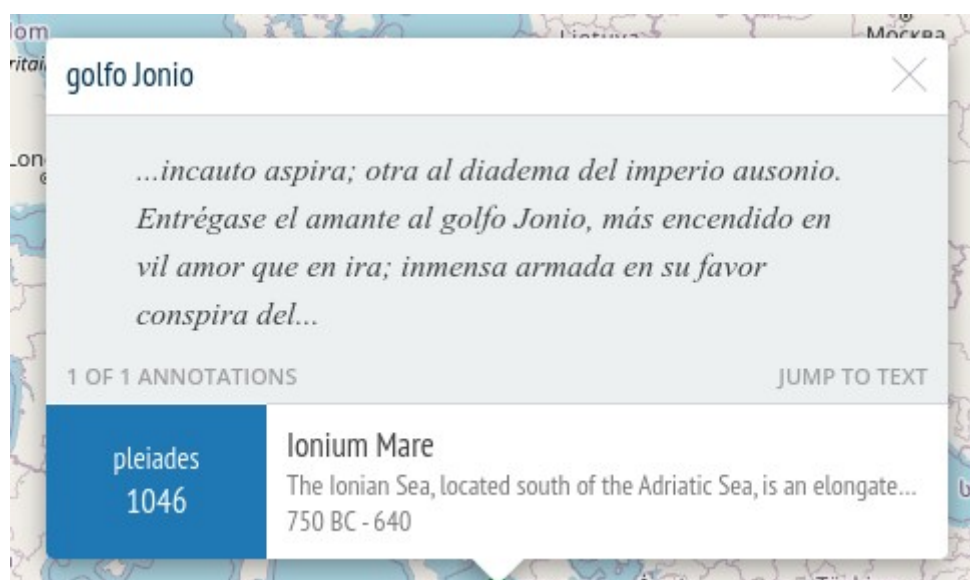


Figura 10: Información contextual asociada al elemento georeferenciado

Recogito no solo posibilita la georreferenciación del corpus, como podría hacerse con Google Maps, sino que también enriquece semánticamente el contenido. Esto permite abordar preguntas de investigación más vinculadas al contenido y los temas tratados en el corpus. Esta información se amplía mediante la identificación de personas y sus descripciones, además de la funcionalidad de relaciones, que no se ha utilizado en este caso. Adicionalmente, la capacidad de exportar en varios formatos, específicamente en TEI-XML en nuestro caso, amplía el alcance de uso de esta información.

## 4. Resultados

### **H1. Las Humanidades Digitales permiten destacar y facilitar el análisis de las tendencias literarias presentes en la obra de Juan de Jáuregui.**

Mediante el análisis de las palabras frecuentes y su contexto, hemos logrado identificar las temáticas más relevantes y analizar el enfoque utilizado por Juan de Jáuregui en sus sonetos. Al asociar los títulos, hemos incorporado contextualización y clarificación del sentido del texto, utilizando herramientas de Expresiones Regulares que permiten este proceso de forma automática, escalable a un mayor número de textos sin incremento de esfuerzo.

El análisis de las palabras frecuentes se lleva a cabo a través de las interfaces "Cirrus" y "Términos" de Voyant Tools. La selección de estas palabras nos permite profundizar en el contexto mediante los "Enlaces" y los "Contextos". Así, hemos explorado cómo Jáuregui aborda los temas sacros, morales, relacionados con el paso del tiempo y de influencia petrarquista principalmente. Aunque no hemos abordado inmediatamente el análisis del uso de la mitología, los cultismos o la influencia de la pintura, se reconoce que requerirían una elaboración previa y una mayor disponibilidad de tiempo para su realización con Voyant Tools.

En la herramienta Recógito, hemos clasificado los términos relacionados con mitología, historia o religión mediante "tags". Esto enriquece los textos con contenidos estructurados en metadatos en el fichero TEI-XML resultante, facilitando procesos ulteriores.

### **H2. Las Humanidades Digitales contribuyen de manera significativa a una comprensión más profunda de la forma y el contenido de la obra al presentar la información de manera efectiva.**

Las capacidades de visualización de la información, ilustradas con capturas de pantalla en este documento, se potencian mediante el uso de herramientas en línea que posibilitan una interacción dinámica, proporcionando respuestas visuales, cuantificadas y contextualizadas a las preguntas de investigación planteadas. Además, las capacidades de exportación de datos en varios formatos permiten que el investigador integre esta información con sus propias herramientas de análisis o publicación de manera efectiva.

### **H3. Las Humanidades Digitales tienen el potencial de estimular la participación activa y la colaboración en esta actividad de investigación.**

La accesibilidad y la intuitividad de las herramientas evaluadas (Voyant Tools y Recogito), junto con su dinamismo e interactividad, posibilitan que individuos con diversos niveles de conocimiento literario se acerquen al análisis y comprensión de las temáticas abordadas por el autor. Al explorar visualizaciones específicas, se generan nuevas preguntas y se despierta el interés por profundizar en significados e interpretaciones, haciendo que estas herramientas resulten atractivas para un público más amplio.

La disponibilidad en línea y las funciones de gestión de usuarios y colaboradores permiten la formación de equipos de trabajo multidisciplinarios, enriqueciendo y ampliando las investigaciones. La adopción de estrategias open source y la utilización de sistemas basados en estándares como TEI-XML favorecen la interoperabilidad y la aceptación dentro de la comunidad de investigación.

Estas características claramente fomentan y respaldan la participación activa y la colaboración en actividades de investigación.

Podemos ahora dar respuesta a nuestra pregunta de investigación: **¿Cómo los métodos y prácticas de Humanidades Digitales, en particular el uso de Expresiones Regulares, Voyant Tools y Recogido, posibilitan la identificación y representación de los principales temas de análisis aplicables a los sonetos de Juan de Jáuregui?**

Hemos realizado un repaso al procedimiento y los resultados ofrecidos por estas herramientas y su análisis nos ha evidenciado la riqueza y variedad temática del corpus del que extraemos 3 grandes temas de análisis como son la poesía religiosa, la poesía ético-moral y la amorosa. Otros grupos temáticos se entrecruzan pues consideramos que el tratamiento del tiempo es mas horizontal y el tratamiento laudatorio tiene cabida en el ejemplo ético-moral. Los 3 términos principales (Cielo, vida y pecho) y sus contextos han sido una guía importante para llegar a esta clasificación y los elementos de representación han permitido comprender cómo utiliza el poeta estos conceptos a lo largo de la obra.

Por tanto se ha evidenciado que las herramientas son de fácil utilización y proporcionan resultados que mejoran la comprensión de la obra literaria guiando al investigador en líneas temáticas, líneas de análisis que no son evidentes con la lectura superficial y que acercan al lector inexperto a la comprensión de la literatura activando su interés y al investigador más versado permiten bucear en los textos apoyando sus investigaciones y descubriendo nuevas interpretaciones o preguntas de investigación.



## 5. Conclusiones

Las herramientas analizadas, Voyant Tools y Recógito, destacan por su capacidad para representar información, permitiendo la visualización eficiente de extensos volúmenes de datos y reduciendo los tiempos de investigación. Facilitan la comprensión de estructuras complejas de forma intuitiva, revelando patrones y relaciones no siempre evidentes de manera directa. Estos aspectos potencian la investigación, fomentan el debate y promueven una participación más amplia.

Un factor crucial es la estrategia de código abierto adoptada por ambas herramientas, fortaleciendo la investigación y la difusión del conocimiento al posibilitar una mayor interacción con una comunidad amplia y diversa. Al adherirse a esta estrategia, se enriquecen y amplían las bases de conocimiento y las potencialidades de las herramientas.

La interactividad y dinamismo de estas herramientas son clave para estimular la exploración y motivar la formulación de nuevas preguntas ante la presentación de resultados que requieren nuevas interpretaciones.

Aunque estas herramientas ofrecen facilidades para aproximarnos a textos desconocidos, alentando la curiosidad y la iniciación en el análisis de textos, es crucial reconocer que por sí solas no transmiten conocimiento. Su eficacia depende de la tarea esencial de interpretación humana, una labor multidisciplinaria que combina el entendimiento del contexto en el que se desarrollan los contenidos con los datos cuantitativos generados. En este sentido, es esencial explorar la biografía del autor, sus motivaciones, su relación con contemporáneos y comprender las grandes tendencias interpretativas de la literatura del Siglo de Oro para aprovechar plenamente la información y las funcionalidades de estas herramientas y contribuir al valor de investigaciones más amplias.

Se abren nuevas vías de investigación, como las interconexiones entre la faceta de pintor y la obra literaria, el equilibrio entre la tendencia italianizante y la culterana, la utilización de cultismos, incluyendo el hipérbaton y los cultismos gongorianos, y la retroalimentación de las categorizaciones realizadas en Recógito para enriquecer el análisis en Voyant.

## 6. Bibliografía

- Casañas, J. T., & Blanch, R. C. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba: Anuario Galego de Filoloxía*, 44, 347-408.  
<https://doi.org/10.15304/verba.44.3155>
- Cervantes, B. V. M. de. (s. f.). *Sonetos del siglo XV al XVII*. Biblioteca Virtual Miguel de Cervantes. Recuperado 20 de enero de 2024, de [https://www.cervantesvirtual.com/obra-visor/sonetos-del-siglo-xv-al-xvii--0/html/001213e2-82b2-11df-acc7-002185ce6064\\_10.html](https://www.cervantesvirtual.com/obra-visor/sonetos-del-siglo-xv-al-xvii--0/html/001213e2-82b2-11df-acc7-002185ce6064_10.html)
- Corpus DISCO*. (s. f.). Recuperado 16 de enero de 2024, de <https://prf1.org/disco/>
- Del Río Riande, G., & Vitale, V. (2020). Recogito-in-a-Box: From Annotation to Digital Edition. *Modern Languages Open*, 1, 44. <https://doi.org/10.3828/mlo.v0i0.299>
- Hernández Lorenzo, L. (2020). Humanidades Digitales y Literatura española: 50 años de repaso histórico y panorámica de proyectos representativos. *Janus: estudios sobre el Siglo de Oro*, 9, 562-595.
- Herrero, M. Á. (2018). Reseña Proyecto ADSO. Análisis distante del soneto castellano de los Siglos de Oro (XVI y XVII). *Revista de Humanidades Digitales*, 2, 159-166.  
<https://doi.org/10.5944/rhd.vol.2.2018.22634>
- Jáuregui, J. de. (2020). *Discurso poético*. e-Spania Books.
- Jáuregui y Aguilar, J. de, & Matas Caballero, J. (1993). *Poesía*. Cátedra.
- Matas Caballero, J. (1990). Algunas notas sobre métrica: Sonetos y canciones de Juan de Jáuregui. *Estudios humanísticos. Filología*, 11, 11-26. <https://doi.org/10.18002/ehf.v0i11.4323>
- Matas Caballero, J. (2020). Juan de Jáuregui, perfil barroco de un poeta-pintor. *e-Spania. Revue interdisciplinaire d'études hispaniques médiévales et modernes*, 35.  
<https://doi.org/10.4000/e-spania.33732>
- Navarro-Colorado, B. (2020). *bncolorado/CorpusSonetosSigloDeOro: First release (v1.0.0)* [dataset]. Zenodo. <https://doi.org/10.5281/zenodo.3634442>
- Osuna Cabezas, M. J. (2014). *La polémica gongorina: Respuestas al «Antídoto» de Jáuregui*.  
<https://idus.us.es/handle/11441/42700>
- Voyant Tools*. (s. f.). [voyant-tools.org](http://voyant-tools.org)

## Índice de figuras

Figura 1: Nube de palabras "Cirrus" tras excluir las palabras vacías.....	9
Figura 2: Nube de palabras sin excluir palabras vacías.....	11
Figura 3: Vista de la pestaña Enlaces.....	14
Figura 4: Detalle de la distribución y correlaciones del término "cielo".....	15
Figura 5: Detalle de la distribución y correlaciones del término "vida".....	17
Figura 6: Detalle de la distribución y correlaciones del término "pecho".....	18
Figura 7: Ejemplo de marcado de una palabra en Recogito.....	20
Figura 8: Representación en mapa de palabras georeferenciadas.....	21
Figura 9: Estadísticas de uso en el Corpus.....	22
Figura 10: Información contextual asociada al elemento georeferenciado.....	22

## Índice de tablas

Tabla 1: Valores de Riqueza Léxica y Extensión de los documentos.....	8
Tabla 2: Palabras más utilizadas tras excluir las palabras vacías.....	10
Tabla 3: Palabras más utilizadas sin excluir las palabras vacías.....	11
Tabla 4: Contextos de uso de la conjunción copulativa "y".....	12
Tabla 5: Contextos de uso de los artículos "el" y "la".....	13
Tabla 6: Contextos del término "cielo".....	15
Tabla 7: Temáticas de los sonetos con el término "cielo".....	16
Tabla 8: Contextos del término "vida".....	17
Tabla 9: Temáticas de los sonetos con el término "vida".....	17
Tabla 10: Contextos del término "pecho".....	18
Tabla 11: Temáticas de los sonetos con el término "pecho".....	19

## Relación de referencias externas

Web	URL
Proyecto en gitHub	<a href="https://github.com/enriquegarciaarias/introHD_tarea_final">https://github.com/enriquegarciaarias/introHD_tarea_final</a>
Interfaz Voyant Tools	<a href="https://voyant-tools.org/?corpus=a3f64b3fec2d36bf1ed8d6997a966430">https://voyant-tools.org/?corpus=a3f64b3fec2d36bf1ed8d6997a966430</a>
Interfaz Recogito	<a href="https://recogito.pelagios.org/document/hczu0toq2teqjf/part/1/edit">https://recogito.pelagios.org/document/hczu0toq2teqjf/part/1/edit</a>
Zotero	<a href="https://www.zotero.org/egarias/items/JN3EVUQQ/item-details">https://www.zotero.org/egarias/items/JN3EVUQQ/item-details</a>