

Estudios literarios computacionales en la Universidad de Alicante

Computational literary studies at the University of Alicante

Borja Navarro Colorado | borja@dlsi.ua.es

¡Busco
colaboradores!

Análisis distante del soneto del Siglo de Oro (ADSO)

Objetivos / aportaciones

- Análisis de la métrica recurrente («gustos métricos») en los sonetos del Siglo de Oro.
- Análisis de temas y motivos recurrentes mediante *Topic Modeling*.
- Financiación: Fundación BBVA 2016-2018.

Recursos desarrollados

Corpus de sonetos del Siglo de Oro:

<https://github.com/bncolorado/CorpusSonetosSigloDeOro>

- Más de 5000 sonetos (de la Biblioteca de Soneto de Ramón García González, publicado en la Biblioteca Virtual Miguel de Cervantes)
- Anotación métrica automática con el sistema de escansión ADSO:
 - <https://github.com/bncolorado/CorpusSonetosSigloDeOro>
 - 10% sonetos revisados y corregidos a mano

Corpus general de poesía del Siglo de Oro:

<https://github.com/bncolorado/CorpusGeneralPoesiaLiricaCastellanaDelSigloDeOro>

- Muestra de 475 poemas con diferentes tipos métricos y estróficos: canción, lira, églogas, coplas, etc.
- Anotación categorial (Freeling) y métrica (sistema de escansión ADSO)
- 10% revisado y corregido a mano.

Herramientas desarrolladas

Sistema de escansión ADSO

<https://github.com/bncolorado/adsoScansionSystem>

- Basado en reglas más frecuencias para resolución de ambigüedades.
- Endecasílabos. Solo métrica: combinación de tónicas y átonas.

Bibliografía

NAVARRO COLORADO, Borja (2019) "Por un análisis distante y profundo: un corpus piloto de la poesía lírica castellana del Siglo de Oro" *Revista de poética medieval*, 33 (2019)

NAVARRO COLORADO, Borja (2018) "On Poetic Topic Modeling: extracting themes and motifs from a corpus of Spanish poetry", *Frontiers in Digital Humanities. Computational Linguistics and Literature*, doi 10.3389/fdigh.2018.00015

NAVARRO COLORADO, Borja (2018): "A Metrical Scansion System for Fixed-Metre Spanish Poetry". *Digital Scholarship in the Humanities* (doi)

NAVARRO COLORADO, Borja (2016): "Hacia un análisis distante del endecasílabo áureo: patrones métricos, frecuencias y evolución histórica". *Rhythmica. Revista española de métrica comparada*, 14.

NAVARRO COLORADO, Borja; RIBES LAFOZ, María y SÁNCHEZ, Noelia (2016): "Metrical annotation of a large corpus of Spanish sonnets: representation, scansion and evaluation". *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*, Portorož (Eslovenia).

NAVARRO COLORADO, Borja (2015): "A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects". *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver (EEUU).

Distant Reading for European Literary History (COST Action)

Objetivos y financiación

<https://www.distant-reading.net/>

- Crear un corpus de novela europea (*European Literary Text Collection: ELTeC*) con al menos 2500 novelas en 10 idiomas diferentes.
- Plantear, desarrollar y compartir métodos computacionales para el análisis multilingüe de la novela europea.
- Reflexionar y replantear conceptos básicos de la teoría e historia literaria a partir de estos métodos y recursos.

Financiación: Unión Europea. Programa *COST Action CA16204* 2017 – 2021 (ampliado a 2022)

Grupos de trabajo

WG1. Desarrollo de la European Literary Text Collection (ELTEC)

WG2: Desarrollo de métodos y herramientas computacionales para el análisis de la novela

WG3: Implicaciones en la teoría y crítica literaria de los métodos computacionales (*distant reading*)

Recursos: Corpus ELTeC

<https://github.com/COST-ELTeC>

Criterios de selección:

- 100 novelas por idioma.*
- Novelas publicadas entre 1840-1920 en Europa (25 novelas por cada 20 años).
- Compensado por género (si es posible). Mínimo 10% mujeres.
- Una novela por autor. Solo 11 autores con tres novelas.
- Compensado por «canonicidad»: mínimo 30% novela no canónica.

Anotación XML-TEI metadatos, estructura, palabras. Información lingüística en proceso: estilo directo, UD, etc.

Situación actual: <https://distantreading.github.io/ELTeC/>

- Español/Castellano: 80 novelas (faltan 20)

* Necesito colaboración para seleccionar/anotar novelas en catalán/valenciano, euskera y gallego. Si te interesa participar, ¡ponte en contacto conmigo!



Universitat d'Alacant
Universidad de Alicante

Departament de Llenguatges i Sistemes Informàtics
Departamento de Lenguajes y Sistemas Informáticos