Bibliography

Bilge, Sirma. “Théorisations féministes de l’intersectionnalité.” *Diogène*, vol. 225, no. 1, Presses Universitaires de France, 2009, pp. 70–88.

Brown, Susan. “Categorically Provisional.” *PMLA*, vol. 135, no. 1, Modern Language Association, 2020, pp. 165–74. doi:10.1632/pmla.2020.135.1.165.

Crenshaw, Kimberle. “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.” *University of Chicago Legal Forum*, vol. 1989, 1989, pp. 139–68.

D’Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT P, 2020.

Drabinski, Emily. “Queering the Catalog: Queer Theory and the Politics of Correction.” *The Library Quarterly: Information, Community, Policy*, vol. 83, no. 2, 2013, pp. 94–111. doi:10.1086/669547.

Gevisser, Mark. *The Pink Line: Journeys Across the World’s Queer Frontiers*. Farrar, Straus and Giroux, 2020.

Maxwell, John. “Text Processing Techniques and Traditions (or: Why the History of Computing Matters to DH).” *Doing More Digital Humanities: Open Approaches to Creation, Growth, and Development*, edited by Constance Crompton, Raymond Siemens, and Richard Lane., Routledge, 2019.

Roberto, K. R. *Radical Cataloging: Essays at the Front*. McFarland, 2008.

Sheffield, Rebecka T. “More than Acid-Free Folders: Extending the Concept of Preservation to Include the Stewardship of Unexplored Histories.” *Library Trends*, vol. 64, no. 3, Johns Hopkins UP, 2016, pp. 572–84. doi:10.1353/lib.2016.0001.

Tanguay, Christian. “Les archives LGBTI: le droit à un lieu pour sa culture.” *Les cahiers de la LCD*, vol.1, no. HS1, 2018, pp. 145–54.

Ware, Syrus Marcus. *“No One Like Me Seemed to Have Ever Existed”: A Trans of Colour Critique of Trans Scholarship and Policy Development in Post-Secondary Schools*. University of Toronto, 2011.

Una mirada digital a cuatro obras dramáticas de Federico García Lorca. Aportes de las técnicas de visualización a la interpretación de texto dramático

Dabrowska, Monika
monika.dabrowska@unir.net
Universidad Internacional de la Rioja (UNIR)

Santa Maria, Teresa
teresa.santamaria@unir.net
Universidad Internacional de la Rioja (UNIR)

Introducción

La propuesta parte del convencimiento del gran potencial de la visualización de la información para el estudio de los textos literarios, apoyándose en la teoría de las redes sociales y técnicas de la visualización de los datos. Si el texto literario es una red de fenómenos discursivos, lo es por excelencia una obra dramática, una red de parlamentos e interacciones entre los personajes. Aplicando los instrumentos de análisis y representación gráfica, se pueden ilustrar las relaciones discursivas entre los personajes. El tratamiento de los datos permite visualizar y “medir” estas relaciones, así como evidenciar fenómenos difícilmente perceptibles en la lectura tradicional, creando nuevos conocimientos e interpretaciones.

Objetivo

El propósito de esta presentación es mostrar ejemplos de visualización de texto y sus posibilidades interpretativas, basándose en los cuatro dramas rurales del autor granadino: *Bodas de sangre* (1933), *Yerma* (1934), *Doña Rosita la soltera* (1935) *La casa de Bernarda Alba* (1936). Se exploran las potencialidades de Gephi y RAWGraphs para sintetizar, sistematizar y representar los datos mediante diferentes grafos y otras representaciones infográficas (dendrogramas, diagramas aluviales, etc.). La exploración muestra los nuevos conocimientos que se pueden extraer a partir de transformación de los datos en imágenes, referentes a la estructura, composición y contenidos de las obras teatrales.

Método

En el trabajo se aplican métodos de análisis cuantitativo y cualitativo. Los textos digitalizados en formato TEI, GEFX y CVS provienen de Drama Corpora Proyect (DraCor, disponible en:). Para representar la red de conexiones entre los personajes se crearán los grafos, que representan la red de protagonistas y el peso cuantitativo de las intervenciones discursivas de cada una. Para su procesamiento y generación de representaciones gráficas se utiliza el software libre Gephi 0.9.2-beta (Gephi.org 2008-2017) y RAWGraphs (https://rawgraphs.io/). Se comparan los grafos generados por ambas herramientas con los creados por la API de DraCor llamada Shiny DraCor (). Para visualizar las relaciones entre las palabras y unidades léxicas más significativas se recurre a los grafos léxicos. Para mapear las estadísticas textuales se rastrean otras 29 herramientas de representación gráfica incluidas en la interfaz de RAWGraphs.

Avance de conclusiones

El interés principal de la propuesta radica en la exploración de las visualizaciones y detección de las singularidades de los elementos léxicos en el tejido del texto lorquiano. Los resultados de análisis informático se confrontan con los enfoques críticos tradicionales. Se evalúa críticamente la eficacia de las técnicas de visualización para la interpretación de la obra literaria, su valor epistemológico, las posibilidades y limitaciones de las herramientas utilizadas (algunas de ellas automatizadas) y los retos que planean a los estudios literarios. Sin duda, es una vía que puede revelar tendencias y patrones inesperados y aportar claves interpretativas de interés.

Bibliografía

Barros García, B. B. (2020). El texto literario hecho datos: F. M. Dostoievski en el marco de las Humanidades digitales y los enfoques cuantitativos. *452ºF. Revista De Teoría De La Literatura Y Literatura Comparada*, (23), 53–77.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361-362.

Burley, D., Ashburn, V. (2010). Information Visualization As A Knowledge Integration Tool. *Journal of Knowledge Management Practice*, 11(4), 18.

Cherven, K. (2015). *Mastering Gephi Network Visualization*. Packt Publishing Ltd

Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5 (1)

Martínez Carro, Elena. (2018) Los personajes femeninos lorquianos desde una interpretación digital, en: Álvarez Ramos, Eva y Blasco Pascual, Javier (eds.). *Humanidades Digitales. Retos, Recursos y Nuevas Propuestas,* Agilice Digital, Valladolid.

Santa María, M.T., Calvo Tello, J., Jiménez, C.M. (2021). ¿Existe correlación entre importancia y centralidad? Evaluación de personajes con redes sociales en obras teatrales de la Edad de Plata, Digital Scholarship in the Humanities, Volume 36 (1), 81–88. https://doi.org/10.1093/llc/fqaa015

Word frequencies in authorship attribution: A simple tweak to improve performance

Eder, Maciej

maciej.eder@ijp.pan.pl
Institute of Polish Language (Polish Academy of Sciences)

Abstract

In this paper, I introduce a very simple method of computing relative word frequencies for authorship attribution and similar stylometric applications. The proposed method outperforms classical most-frequent-word approaches by a few percentage points.

Introduction

In a vast majority of stylometric studies, relative frequencies of the most frequent words (MFWs) are used as the language features to betray the authorial "fingerprint". A vector of such relative word frequencies is then passed to one of the multidimensional machine-learning classification techniques, ranging from simple distance-based lazy learners, such as Delta (Burrows, 2002; Evert et al., 2017), to sophisticated neural network setups (Gómez-Adorno et al., 2018).

Even if alternative types of features have been introduced (Peng et al., 2003; Hirst and Feiguina, 2007; Lučić and Blake, 2013) and tested in controlled experiments (Eder, 2011), the standard approach relying of word frequencies continues to be predominant in the field (Grieve, 2007; Stamatatos, 2009). In this paper, I propose to count the relative frequencies in a slightly different way, in order to better capture the authorial choice of words.

Words that (might) matter

The notion of relative word frequencies is fairly simple. We count all the tokens belonging to particular types (e.g. all the attestations of the word "the", followed by the attestations of "in", "for", "of" etc.), and for each word, we divide the number of types by the total number of words in a document. Consequently, each word frequency is equal to its percentage within the document (e.g. "the" = 0.0382), and all the frequencies sum up to 1. The reason of converting occurrences to relative frequencies is obvious: by doing so, one is able to reliably compare texts that differ in length.

For the sake of this paper, it is important to note that such frequencies are relative to *all the other words* in a document in question. Convenient as they are, these values are at the same time very small and – importantly – can be affected by other word frequencies. Now, what if we disregard thousands of other words in a text, and compute