

# Estilometría computacional, algunas experiencias en el marco del proyecto TRACE.

Bia Alejandro.

Cita:

Bia Alejandro (2016). *Estilometría computacional, algunas experiencias en el marco del proyecto TRACE. Humanidades Digitales: Construcciones locales en contextos globales. Asociación Argentina de Humanidades Digitales, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/aahd2016/8>

ARK: <https://n2t.net/ark:/13683/ey3x/xUo>



Esta obra está bajo una licencia de Creative Commons.  
Para ver una copia de esta licencia, visite  
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>.

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.*



ASOCIACIÓN ARGENTINA DE HUMANIDADES DIGITALES

# Humanidades Digitales: construcciones locales en contextos globales

Actas del I Congreso Internacional  
de la Asociación Argentina de  
Humanidades Digitales (AAHD)



Humanidades Digitales : Construcciones locales en contextos globales : Actas del I Congreso Internacional de la Asociación Argentina de Humanidades Digitales - AAHD / Agustín Berti ... [et al.] ; editado por Gimena del Rio Riande, Gabriel Calarco, Gabriela Striker y Romina De León - 1a ed. - Ciudad Autónoma de Buenos Aires : Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires, 2018.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-987-4019-97-4

1. Actas de Congresos. 2. Humanidades. 3. Digitalización. I. Berti, Agustín II. del Rio Riande, Gimena, ed.

CDD 301

ISBN 978-987-4019-97-4



# Humanidades Digitales. Construcciones locales en contextos globales

Gimena del Rio Riande, Gabriel Calarco, Gabriela Striker y Romina De León (Eds.)

ISBN: 978-987-4019-97-4

---

## > Índice

### *I. Preliminares*

**FUNES**, Leonardo. *Palabras Preliminares*

del **RIO RIANDE**, Gimena. *Cuando lo local es global*

**FIORMONTE**, Domenico. *¿Por qué las Humanidades Digitales necesitan al Sur?*

### *II. Métodos y herramientas de las Humanidades Digitales*

**BIA**, Alejandro. *Estilometría computacional, algunas experiencias en el marco del proyecto TRACE*

**SALERNO**, Melisa; **HEREÑÚ**, Daniel y **RIGONE**, Romina. *Modelado 3D del cementerio de la antigua Misión Salesiana de Río Grande: tareas efectuadas y potenciales usos*

**VÁZQUEZ CRUZ**, Adam Alberto y **TAYLOR**, Tristan. *Adnoto: un etiquetador de textos para facilitar la creación de ediciones digitales*

**BRACCO**, Christian; **CORREA**, Facundo; **CUEVAS**, Lucas; **CEPEDA**, Virginia; **DELLEDONNE**, Francisco; **VOSKUIL**, Anne Karin; **PAPARAZZO**, Nicolás y **TORRES**, Diego. *Una wiki semántica para las artes escénicas. Conceptos e implementación de la plataforma colaborativa Nodos*

**IZETA**, Andrés Darío y **CATTÁNEO**, Roxana. *¿Es posible una arqueología digital en Argentina? Un acercamiento desde la práctica*

**LACALLE**, Juan Manuel y **VILAR**, Mariano. *Una lectura distante de la investigación actual en*

*Letras en Argentina*

**MARTIN**, Jonathan y **TORRES**, Diego. *Análisis de patrones en la evolución de wikis*

**MARTÍNEZ CANTÓN**, Clara Isabel; **DEL RIO RIANDE**, Gimena y **GONZÁLEZ-BLANCO GARCÍA**, Elena. *Poetriae. Una colección de poéticas medievales basada en conceptos métricos únicos y referenciables*

**SUED**, Gabriela. *Ciudades visibles: estética y temática de tres ciudades iberoamericanas en la red social Instagram. Un estudio exploratorio desde las Humanidades Digitales*

### ***III. Educación, políticas públicas, Humanidades Digitales en el aula***

**DAVICO**, María Luz; **LINEARES**, Gabriel y **PEZZUTTI**, Luciana. *Literacidad electrónica en la enseñanza universitaria: cómo, cuándo y dónde*

**MUÑOZ**, Patricia Alejandra. *Valoración de un proyecto de desarrollo tecnológico y social en la enseñanza de Inglés como lengua extranjera*

**PACHECO DE OLIVEIRA**, Maria Livia y **SÁ DE PINHO NETO**, Júlio Afonso. *Brecha digital e o acesso à informação: projetos de inclusão digital*

**CASASOLA**, Laura. *Experiencia educativa con TIC: Celulares en acción*

**DÍAZ**, Aída Alejandra y **HUALPA**, Mariela. *Una experiencia de aprendizaje en educación superior mediada por TIC*

**FRESCURA TOLOZA**, Claudio Daniel. *Computación en la nube en la enseñanza de escritura académica*

**LEÁNEZ**, Nancy; **LECETA**, Andrea; **MARTÍN**, Marcela y **MORCHIO**, Marcela. *Hacia una reconfiguración del aula de lengua extranjera*

**OLAIZOLA**, Andrés. *Los escritores vernáculos digitales y el concepto de valor en las escrituras digitales*

**CHECHELE**, Patricia; **LURO**, Vanesa y **PINTOS ANDRADE**, Esteban. *Afiliarse en la distancia. El ingreso a la educación superior en un entorno virtual de aprendizaje*

**ALLÉS TORRENT**, Susanna y **DEL RIO RIANDE**, Gimena. *Enseñar edición digital con TEI en español. Aprendizaje situado y transculturación*

#### ***IV. Medios, re-mediación, redes sociales***

**RODRIGUEZ KEDIKIAN**, Martín. #100DiasdeMacri. *Analítica cultural en la construcción de los primeros cien días de la presidencia de Mauricio Macri en conversaciones en Twitter*

**ALONSO**, Julio; **ALAMO**, Sofía; **GONZALEZ OCAMPO**, María Eugenia; **GIAMBARTOLOMEI**, Guido; **MANCHINI**, Lucas y **TOSCANO**, Ayelén. *¿Hacia una algoritmización de los sentimientos?*

**DE MIRANDA**, Jair Martins. *Samba Global– Do mundo do samba ao samba no mundo*

**ORTIZ**, María. *Las migraciones en los tiempos del software*

**SANTOS**, Laura. *Arte urbano, de la calle a las redes*

**ALAMO**, Sofía; **BORDOY**, Giselle; **CHETTO**, Melisa; **IBAÑEZ**, Fernanda, **MIGLIORINI**, Agustina y **GONZALEZ OCAMPO**, María Eugenia. #NiUnaMenos: *Big Data para la comprensión de una problemática de género*

**KLIMOVSKY**, Pedro. *El documental digital y la representación de lo real*

**BERTI**, Agustín. *Fotogramas autorizados: La crisis de la noción de obra cinematográfica ante las remasterizaciones*

**BORDOY**, Giselle. *El disco como obra abierta en interacción con las audiencias*

**COELHO**, Cidarley. *Forma Material Digital: livro e leitura na sociedade contemporânea*

#### ***V. Reflexiones sobre/desde/hacia lo digital***

**VISCARDI**, Ricardo. *Actuvirtualidad e inter-roogación: un lugar entre-otros*

**ÁLVAREZ GANDOLFI**, Federico y **DEL VIGO**, Gerardo Ariel. *Hatsune Miku, una idol digital: entre el otakismo y el waifuismo*

**SAÁ**, Guido. *Reflexiones sobre música y narración: Recursos retóricos y exegéticos musicales en la línea narrativa y el pathos en BioShock 2 y BioShock Infinite*

**GLUZMAN**, Georgina Gabriela. *Algunas reflexiones sobre la Base de datos de mujeres artistas en Buenos Aires (1924-1939)*

**DOMINGUEZ HALPERN**, Estela; **ALAMO**, Sofía; **ALONSO**, Julio. *Entramados y ciudades. Visibilizando Baldosas por la Memoria*

**GÓMEZ**, Verónica Paula. *Territorios nacionales, territorialidades ciberespaciales: disputas discursivas sobre la soberanía en la circulación de literatura digital*

**RIGAL COLLADO**, Pablo Alonso; **MAESTIGUE PRIETO**, Nancy y **GARCÍA VÁZQUEZ**, Mayté. *La narración hipertextual. El reto cubano*

## ***VI. La publicación científica y el Acceso Abierto desde las Humanidades Digitales***

**TSUJI**, Teresa y **CANELLA**, Rubén. *Lenguajes y recursos multimediales para la difusión de la ciencia. Desafíos y oportunidades digitales*

**CATALDI**, Marcela; **DI CÉSARE**, Victoria; **FERNÁNDEZ**, Néstor; **HERNÁNDEZ**, Alicia; **LIBERATORE**, Gustavo y **VOUTTO**, Andrés. *Sistema taxonómico de organización de los recursos de información autoarchivados en el Repositorio Institucional de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata*

**ÁLVAREZ**, Leonardo Javier y **CORDA**, María Cecilia. *FLACSOAndes Tesis: comunicación científica de investigaciones realizadas en maestrías y doctorados del sistema FLACSO*

## ***VII. Digitalización, políticas y prácticas, archivo y memoria***

**AUTHIER**, Carlos; **GIORDANINO**, Eduardo y **LUIRETTE**, Carlos. *La preservación de la memoria audiovisual en Argentina*

**GAMBA**, Guido; **HEIDEL**, Evelin; **RAIA**, Matías; **ACUÑA**, Ezequiel; **ACTIS CAPORALE**, Carla; **DE LA HERA**, Diego y **ACEVEDO**, Melisa. *Recursos digitales para el acceso a los bienes culturales en dominio público*

**FLORES MUTIGLIENGO**, Jennifer. *Arte y Archivo*

**BUGNONE**, Ana y **SANTAMARÍA**, Mariana. *La política de democratización del archivo: el caso del Centro de Arte Experimental Vigo*

**GAMBA**, Guido; **HEIDEL**, Evelin; **RAIA**, Matías; **ACUÑA**, Ezequiel; **ACTIS CAPORALE**, Carla; **DE LA HERA**, Diego y **ACEVEDO**, Melisa. *Digitalización: Una experiencia de campo*

# Estilometría computacional, algunas experiencias en el marco del proyecto TRACE<sup>1</sup>

BIA, Alejandro / Centro de Investigación Operativa. Universidad Miguel Hernández (UMH) – [abia@umh.es](mailto:abia@umh.es)

---

» *Palabras clave: estilometría, clasificación, reconocimiento de autoría.*

## > **Resumen**

En este trabajo veremos el estado del arte de la estilometría computacional y algunos experimentos realizados con un corpus en español, en el marco del proyecto TRACE. Describiremos los resultados preliminares sobre clasificación y agrupación estilométrica. Hemos realizado varios experimentos en un corpus de textos en español para probar la clasificación por período de tiempo y por género del autor, así como el reconocimiento de la autoría por agrupamiento. Como no se encontró literatura relevante sobre estilometría aplicada a textos en español, el propósito principal de estos experimentos fue probar que la estilometría computacional funciona de la misma manera que cuando se aplica a textos en inglés, de lo que sí hay muchas publicaciones.

Algunos de los problemas que hemos resuelto satisfactoriamente son los siguientes:

- Clasificación por época (método oposición de textos, *Zeta de Craig*)
- Clasificación por género del autor (método oposición de textos, *Zeta de Craig*)
- Agrupamiento por autor: (método *bootstrap concensus tree*)
- Agrupamiento por autor (método *cluster analysis*)

Estos experimentos se han realizado utilizando el paquete *Stylo*, basado en el programa de análisis estadístico R. Los ensayos realizados han demostrado que, haciendo los ajustes necesarios y utilizando los parámetros correctos, la herramienta es capaz de realizar la clasificación y el agrupamiento de forma fiable. También aprendimos de nuestros propios

---

<sup>1</sup> Este trabajo ha sido desarrollado dentro del proyecto TRACE: *Herramientas de Software para el Análisis Contrastivo de Textos en Corporaciones Bilingües Paralelas*, y ha sido financiado con la ayuda FFI2012-39012-C04-02 del VI Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica del MINECO (Ministerio de Economía y Competitividad de España).

errores, particularmente en lo referente a la preparación del texto. Finalmente, se propone un flujo de trabajo para realizar pruebas estilométricas de manera metódica.

## > **Introducción**

El ahora popular *meta-campo* o *campo paraguas* de la ciencia de los datos o *Data Science* abarca una amplia gama de disciplinas, técnicas y teorías, como se muestra claramente en la visualización tipo mapa de metro de Chandrasekaran (2013).

Nuestra línea de trabajo abarca tres subcampos relacionados con el análisis de textos: extracción de información de textos (*text mining*), estilometría computacional y técnicas de visualización de datos. Estos campos utilizan métodos similares o cooperan entre sí.

Tanto la minería de textos como la estilometría computacional utilizan técnicas de clasificación y agrupación (Jockers & Witten, 2010; Oakes, 2004), así como métodos de procesamiento del lenguaje natural o PLN (Nieto, 2004) y, en algunos casos, redes neuronales (Merriam & Matthews, 1994; Ramyaa *et al.*, 2004). De hecho, la mayoría de las hipótesis que queremos demostrar pueden ser probadas usando métodos de clasificación y agrupación, primero con muestras de entrenamiento, y luego verificando el método con muestras de prueba, lo mismo que en la minería de datos. Las técnicas de visualización, por su parte, facilitan la buena comprensión del problema y permiten un mejor análisis de los resultados experimentales.

## > **Clasificación estilométrica y agrupación**

Hay varios problemas que queremos resolver usando estilometría computacional, como, por ejemplo:

- Identificar a los escritores por su estilo de escritura.
- Estudiar cómo el bilingüismo afecta el estilo de un escritor.
- Clasificar textos por periodo de tiempo.
- Separar las traducciones afectadas por la censura de las traducciones sin censura.
- Clasificar textos por género del escritor.
- Clasificar textos por autor (reconocimiento de autoría).

Hemos realizado experimentos con un corpus de textos españoles donde trabajamos:

- Clasificación por tiempo (método de oposición, *Zeta de Craig*) (véase figura 1)
- Clasificación por género del autor (método de oposición, *Zeta de Craig*) (véase figura 2)
- *Clustering* de autores (método *bootstrap*, árbol de consenso y análisis de

conglomerados)

Existen varias herramientas para el análisis estilométrico. Entre los más recientes se distingue el Programa de Atribución de Autoría Gráfica Java (JGAAP) (Juola, 2009) y el paquete *Stylo* basado en R (Eder *et al.*, 2016). Los experimentos descritos aquí se han realizado utilizando el paquete *Stylo*, basado en el programa de análisis estadístico R. *Stylo* proporciona funciones para el análisis estilométrico, una interfaz gráfica de usuario y diagramas de calidad de impresión, y está siendo desarrollado y mantenido por el grupo de estilística computacional<sup>2</sup>.

En algunos de los experimentos mencionados en la lista anterior, como la agrupación por período de tiempo y por género de autor, el método aplicado fue el mencionado *Zeta de Craig* (Craig & Kinney, 2009) para una clasificación binaria según características estilométricas discriminativas. El *Zeta de Craig* es una extensión de la métrica *Zeta* originalmente propuesta por Burrows (2007).

En otros experimentos se aplicó el método *Bootstrap Consensus Tree* (BCT) y un diagrama tipo dendrograma como visualización para el reconocimiento de la autoría. Según Eder *et al.* (2016), el método BCT se basa en la idea de que los resultados se estabilizan cuando dividimos la lista de palabras más frecuentes (MFWs) en bandas de frecuencia no idénticas, pero potencialmente superpuestas y luego analizamos estas bandas independientemente entre sí (Eder, 2012). El método BCT fue originalmente tomado prestado por Eder del campo de la evolución del lenguaje y la genética, ya que en la literatura se han descrito varias aplicaciones exitosas de esa técnica (Rybicki & Heydel, 2013; Dalen-Oskam, 2014; Stover *et al.*, 2016). Para el reconocimiento de la autoría también utilizamos la característica de análisis de clúster del paquete *Stylo*, que agrupa las muestras por ramas de una estructura de árbol jerárquica.

El corpus utilizado para el experimento consistió en 83 obras literarias completas de 36 escritores españoles de siglos diferentes (del XVII al XX) y de género diferente, incluyendo algunos ejemplos falsos para probar la consistencia de los métodos utilizados. Como ejemplo de estos últimos, hemos utilizado algunos textos en catalán y en portugués para verificar que no se tomaran por el español antiguo.

---

<sup>2</sup> Sitio web del *Computational Stylistics Group*: <https://sites.google.com/site/computationalstylistics/>.

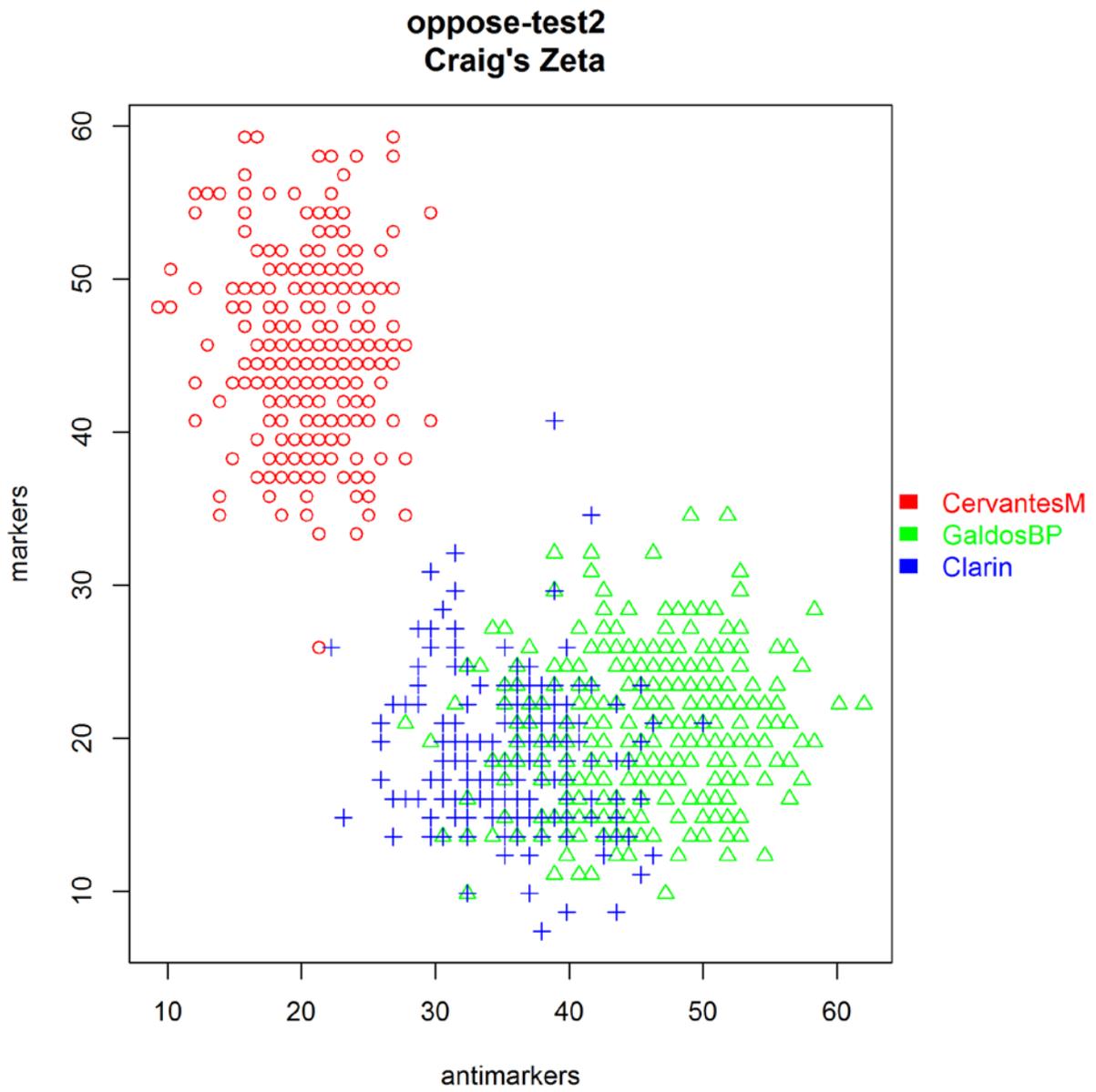


Figura 1. Test de oposición (Craig). Agrupación por período de tiempo: Cervantes, Galdós y Clarín.

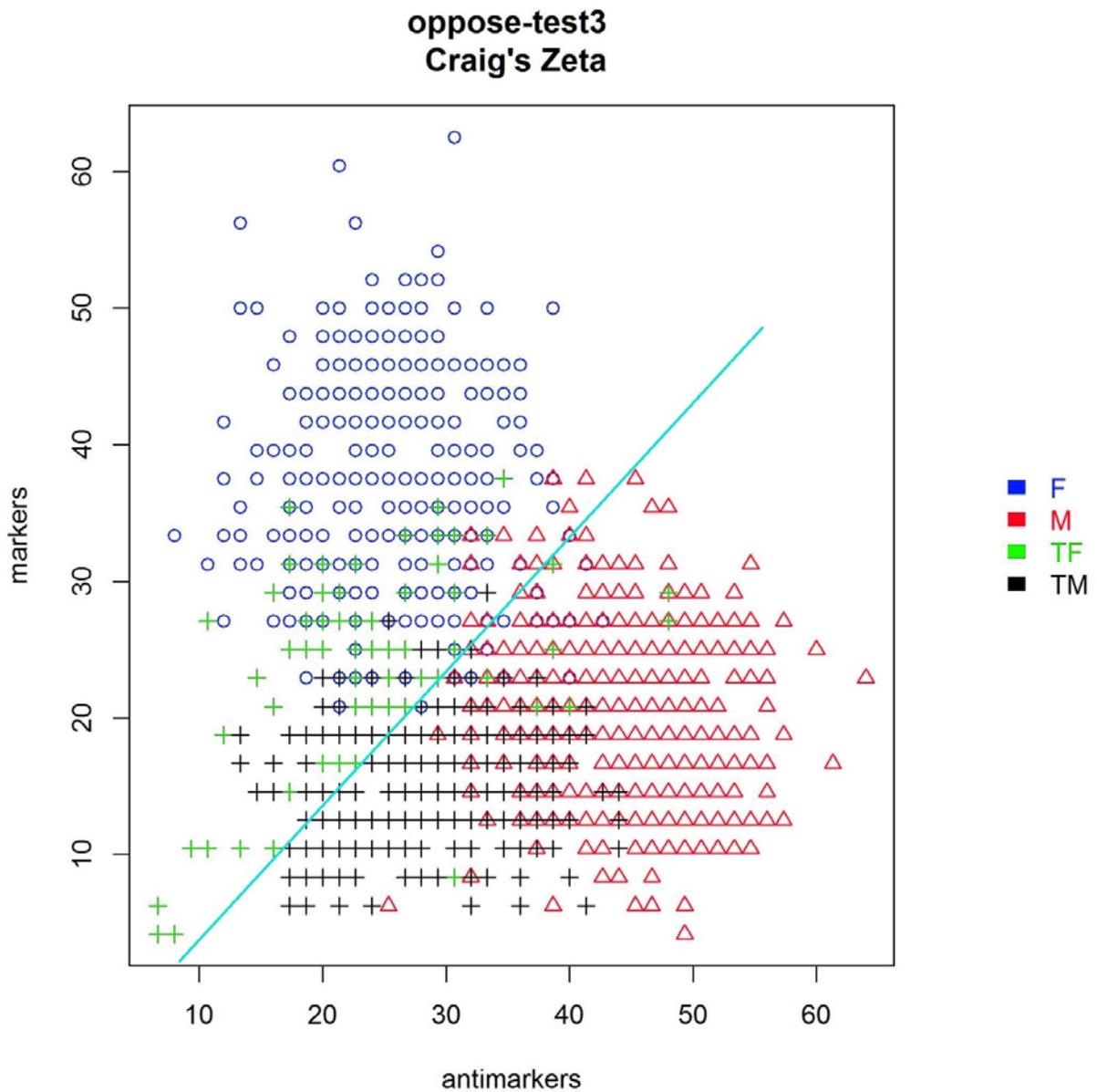


Figura 2. Test de oposición (Zeta de Craig). Agrupación por género: las muestras de prueba están casi perfectamente separadas por el género del autor a cada lado de la línea verde.

## > Conclusiones

Las pruebas realizadas con *Stylo* han demostrado que, haciendo los ajustes necesarios y utilizando los parámetros correctos, la herramienta es capaz de realizar la clasificación por período, el análisis de género y el reconocimiento de autoría de manera confiable en un

corpus español de obras literarias, del mismo modo que cuando se aplica a textos en inglés.

Por experiencia, aprendimos lo obvio: que los textos deben estar limpios de comentarios y anotaciones (que no sean los del propio autor), y hay que tener cuidado de no utilizar textos modernizados, modificados o traducidos, ya que estas fueron las causas de algunos resultados fallidos en la clasificación o la agrupación durante nuestros primeros experimentos. En otras palabras, los textos utilizados para la estilometría deben ser muestras puras no contaminadas del estilo de escritura de los autores correspondientes.

Las herramientas estilométricas demostraron ser muy útiles, pero el análisis humano y la interpretación de los resultados son esenciales para obtener conclusiones interesantes. En esto, las técnicas de visualización juegan un papel muy importante. Por otra parte, es deseable tener hipótesis de partida en primer lugar (a menos que se pretenda obtener *descubrimientos* inesperados). Esto apunta a la necesidad de una práctica metódica (véase figura 3): tener una hipótesis de partida o problema que resolver, reunir y preparar los textos, probar/ajustar/entrenar la herramienta, entrenar con una colección de muestras, probar con el conjunto de investigación objetivo, visualizar y analizar los resultados, y finalmente, sacar conclusiones.

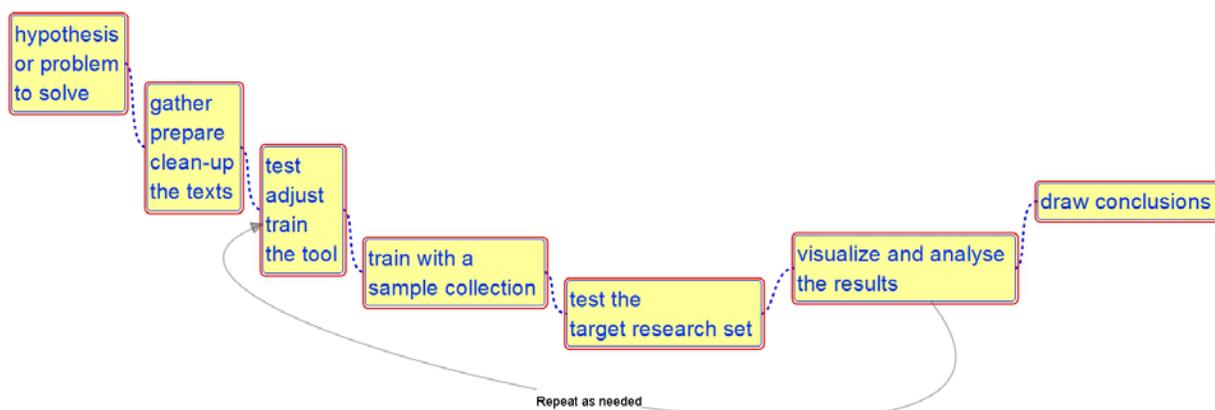


Figura 3. Propuesta de flujo de trabajo para el análisis estilométrico.

Como trabajo futuro, se pretende explorar el uso de estas herramientas en informática forense, para determinar la autoría y el género en mensajes cortos, como se muestra en la literatura (Brocardo *et al.*, 2013; Calix *et al.*, 2008; Burrows, 2002; Iqbal *et al.*, 2010).

Aunque la estilometría no es una ciencia exacta, y depende mucho de las destrezas y el esfuerzo del investigador, escogiendo los métodos apropiados y ajustando adecuadamente los parámetros de la prueba, la investigación sobre corpus de ofuscación ha demostrado que los métodos más robustos y precisos pueden ser efectivos incluso en casos de ofuscación engañosa (Juola & Vescovi, 2011).

## > **Bibliografía**

- Brocardo, M., Traore, I., Saad, S. y Woungang, I. (2015). Authorship Verification for Short Messages Using Stylometry. *Proceedings of the International Conference on Computer, Information and Telecommunication Systems, Piraeus-Athens, Greece, May 7-8, 2013* (1-6). Recuperado de <https://dl.acm.org/citation.cfm?id=2827309> el 25/01/2017.
- Burrows, J. (2002). 'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing Journal*, 17(3), 267-287.
- Burrows, J. (2007). All the Way through: Testing for Authorship in Different Frequency Strata. *Literary and Linguistic Computing Journal*, 22(1), 27-48.
- Calix, K., Connors, M., Levy, D., Manzar, H., McCabe, G. y Westcott, S. (2008). Stylometry for E-mail Author Identification and Authentication. *Proceedings of CSIS Research Day, Pace University, May 2008*.
- Chandrasekaran, S. (2013). Becoming a Data Scientist-Curriculum via Metromap. *Pragmatic Perspectives*. Recuperado de <http://nirvacana.com/thoughts/becoming-a-data-scientist/> el 27/11/2016.
- Craig, H. y Kinney, A. (Eds.) (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- Dalen-Oskam, K. van (2014). Epistolary voices: The Case of Elisabeth Wolff and Agatha Deken. *Literary and Linguistic Computing*, 29(3), 443-451.
- De Vel, O., Corney, M., Anderson, A. y Mohay, G. (2002). Language and Gender Author Cohort Analysis of E-mail for Computer Forensics. *Proceedings of The Digital Forensic Research Conference, DFRWS 2002, Syracuse, NY (Aug 6<sup>th</sup>-9<sup>th</sup>), USA*. Recuperado de [https://www.dfrws.org/sites/default/files/session-files/paper-language\\_and\\_gender\\_author\\_cohort\\_analysis\\_of\\_e-mail\\_for\\_computer\\_forensics.pdf](https://www.dfrws.org/sites/default/files/session-files/paper-language_and_gender_author_cohort_analysis_of_e-mail_for_computer_forensics.pdf) el 27/11/2016.
- Eder, M. (2012). Computational Stylistics and Biblical Translation: How Reliable can a Dendrogram Be? En T. Piotrowski y L. Grabowski (Eds.), *The Translator and the Computer* (155-170). Wrocław: WSF Press.
- Eder, M., Rybicki, J. y Kestemont, M. (2016). Stylometry with R: a Package for Computational Text Analysis. *R Journal*, 8(1): 107-121. Recuperado de <http://journal.r-project.org/archive/2016-1/eder-rybicki-kestemont.pdf> el 27/11/2016.
- Iqbal, F., Khan, L., Fung, B. y Debbabi, M. (2010). E-mail Authorship Verification for Forensic Investigation. *Proceedings of the 2010 ACM Symposium on Applied Computing, SAC '10* (1591-1598). New York: ACM.
- Jockers, M. y Witten, D. (2010). A Comparative Study of Machine Learning Methods for Authorship Attribution. *Literary and Linguistic Computing*, 25 (2), 215-223.
- Juola, P. (2009). JGAAP: A System for Comparative Evaluation of Authorship Attribution. *Journal Of*

- The Chicago Colloquium on Digital Humanities and Computer Science*, 1(1). Recuperado de <http://jdhcs.uchicago.edu/> el 27/11/2016.
- Juola, P. y Vescovi, D. (2011). Analyzing Stylometric Approaches to Author Obfuscation. En G. Peterson y S. Sheno (Eds.), *IFIP International Conference on Digital Forensics. DigitalForensics 2011: Advances in Digital Forensics VII* (115-125). Berlin-Heidelberg: Springer.
- Merriam, T. y Matthews, R. (1994). Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe. *Literary and Linguistic Computing*, 9(1). Recuperado de <http://robertmatthews.org/wp-content/uploads/2016/03/Neural-Computing-II-Marlowe.pdf> el 27/11/2016.
- Nieto, V. (2004). Authorship Attribution with Help of Language Engineering. Homework Assignment for Language Engineering, 2D1418, Academic Year 2004/05 (October 16, 2004). Recuperado de <https://www.nada.kth.se/kurser/kth/2D1418/uppsatser04/victor.pdf> el 27/11/16 el 27/11/2016.
- Ramyaa, C. H., Rasheed, K. (2004). Using Machine Learning Techniques for Stylometry. *Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications (MLMTA'2004)* (897-903).
- Oakes, M. (2004). Ant Colony Optimization for Stylometry: The Federalists Papers. *RASC2004, School of Computing and Technology*. Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=102ED7B13FD2F84FE8416E0980FB9F22?doi=10.1.1.108.799&rep=rep1&type=pdf> el 27/11/2016.
- Rybicki, J. y Heydel, M. (2013). The Stylistics and Stylometry of Collaborative Translation: Woolf's 'Night and Day' in Polish. *Literary and Linguistic Computing*, 28(4),708-717.
- Stover, J., Winter, Y., Koppel, M. y Kestemont, M. (2016). Computational Authorship Verification Method Attributes New Work to Major 2<sup>nd</sup> Century African Author. *Journal of the American Society for Information Science and Technology*, 67(1), 239-242.