

# Modelos y métodos computacionales para la crítica literaria: el estado del arte\*

Fabio Ciotti

Universidad de Roma “Tor Vergata”, Roma, Italia

[fabio.ciotti@uniroma2.it](mailto:fabio.ciotti@uniroma2.it)

**E**N EL CURSO DE LOS últimos quince años, el multiforme campo de estudios bajo el nombre de Humanidades Digitales se ha convertido en un fenómeno de importancia primordial tanto en el mundo de la investigación y la enseñanza universitaria (si bien la diferencia entre los distintos contextos nacionales sea relevante y el reconocimiento institucional inadecuado, en especial en nuestro país) como en el debate cultural.<sup>1</sup>

En la historia de la relación entre tecnologías informáticas digitales y ciencias humanas algunas áreas disciplinares han jugado tradicionalmente un papel estimulante. Aunque hoy el campo de las ciencias históricas parezca estar a la vanguardia, uno de los pilares de este vínculo ha sido el dominio de las ciencias literarias, hasta el punto de que la etiqueta disciplinar de Estudios Literarios Digitales (*Digital Literary Studies*) ha adquirido autonomía propia, sancionada hace algunos años por el valioso y voluminoso *Companion to Digital Literary Studies*, que siguió al más conocido *Companion to Digital Humanities*.

\* Ciotti, Fabio. “Modelli e metodi computazionali per la critica letteraria: lo stato dell’arte.” *L’Italianistica oggi: ricerca e didattica*, Atti del xix Congresso dell’ADI–Associazione degli Italianisti, 9, 10, 11 y 12 de septiembre de 2015, Roma, Italia. Editado por Beatrice Alfonzetti et al. Roma, Adi editore, 2017, págs. 1-11. Traducido por Elena Pérez Palacio y revisado por Anastasia Belousova.

1 En el resto del artículo, por brevedad, usaremos el acrónimo HD en lugar del nombre completo. La literatura sobre la definición disciplinar y la evolución de este campo es ahora sobreabundante: nos limitamos a señalar *Defining Digital Humanities* y, en particular en el mismo volumen, la excelente reconstrucción histórica de Vanhoutte, “The Gates of Hell. History and Definition of Digital | Humanities | Computing” (págs. 119-156). Para una panorámica histórica italiana ver Orlandi y Ciotti, “A landscape”. Fundamental para una visión de conjunto de la pluralidad de temas y enfoques que componen las HD es el monumental *A Companion to Digital Humanities* y su reciente segunda edición *A New Companion to Digital Humanities*.

Esta centralidad parece natural si se considera cómo la edición y la codificación digital de textos y documentos (entendidos ora como fuentes primarias ora como literatura secundaria) representan un componente fundamental para todas las HD y no solo para los estudios digitales de orientación literaria.<sup>2</sup> Pero, en este artículo, no me detendré sobre cuestiones concernientes a la edición digital, temas y técnicas ya ampliamente tratadas y experimentadas en el contexto italiano. En su lugar, pretendo proponer algunas reflexiones sobre el estado del arte de los métodos computacionales para el análisis textual y la crítica literaria, ámbito que, en el panorama de los estudios literarios italianos, parece hoy todavía restringido.

## La tradición del análisis computacional de los textos

Los métodos y los instrumentos de análisis textual (*text analysis*) han contribuido a la construcción y evolución de las HD desde sus orígenes que, como es conocido, se remontan a los primeros experimentos de automatización del análisis lexical del vocabulario de Santo Tomás por parte del Padre Roberto Busa.<sup>3</sup> Como uno de los pioneros y más acreditados estudiosos en este campo, el anglista David Hoover ha observado que “el análisis textual asistido por computadores tiene una larga y rica historia a pesar del hecho de que, como ha sido frecuentemente anotado, no haya sido ampliamente adoptado por los estudios literarios contemporáneos”<sup>4</sup> (“Textual Analysis” s. p.).

Claro está que la estrecha relación con la lingüística computacional, disciplina muy correlacionada con las HD y que desde hace mucho tiempo goza de una dignidad institucional consolidada, ha jugado un papel importante en esta presencia histórica. El análisis textual computacional, justamente, tiene fundamentos comunes e instrumentos operativos compartidos con la lingüística de los *corpus*; basta pensar en la tradición de los léxicos y

- 
- 2 La bibliografía sobre el tema no tiene fin. Para una introducción sobre estos temas referimos a Ciotti, *Il manuale*; una discusión teórica y técnica sobre los límites del XML para las ciencias humanas se encuentra en Ciotti, “La rappresentazione”. Sobre la *Text Encoding Initiative*, se pueden consultar obviamente los numerosos documentos disponibles en el sitio del TEI Consortium.
  - 3 Para una biografía del padre Busa y una reconstrucción de su papel intelectual en la historia de la informática humanista, ver el reciente *Roberto Busa, S.J. and the Emergence of Humanities Computing: The Priest and the Punched Cards* de Steven E. Jones.
  - 4 La traducción de las citas del inglés es de Elena Pérez Palacio.

concordancias. Pero tales raíces son al mismo tiempo (al menos en parte) la causa de la limitada recepción en el ámbito crítico literario de la que Hoover se lamenta.

La imposibilidad de superar la barrera del significante en ámbito informático, por así decirlo, ha restringido las capacidades de las técnicas de análisis textual tradicional a los métodos puramente cuantitativos (conteo y análisis de la distribución de *token* sublexicales, lexicales y morfemáticos), dando origen a la estilometría computacional clásica. Ciertamente, en este ámbito no han faltado contribuciones interesantes. Para limitarnos a los pioneros, basta pensar en los trabajos de Paul Fortier sobre tematología en la literatura francesa, basados en el análisis estadístico de la distribución de *clusters* lexicales,<sup>5</sup> o los de John Burrows sobre el lenguaje de los personajes en la narrativa de Jane Austen, basados en el uso de métodos estadísticos como el análisis de grupos (*cluster analysis*) y el análisis de componentes principales (*principal component analysis*).<sup>6</sup> Y, permítaseme recordar, como eminente ejemplo de estudio crítico basado en métodos de computación tradicionales, el trabajo sobre *Memoriale* de Paolo Volponi publicado en *Letteratura Italiana Einaudi* por Giuseppe Gigliozi, una de las figuras más prominentes de la informática humanista de la escuela romana, lamentablemente fallecido de manera prematura. Finalmente, no olvidamos la gran relevancia y los interesantes resultados obtenidos por un área de investigación estrechamente conectada a la estilometría clásica, la de los estudios de atribución, que han tenido una importancia notable en los estudios anglísticos (como consecuencia de las conocidas problemáticas de atribución shakespeareanas) y de la Antigüedad clásica.<sup>7</sup>

El hecho es que, si para los estudios lingüísticos o atributivos la cuantificación de las formas es un dato suficiente para suministrar evidencia de conclusiones científicamente relevantes, en los estudios literarios, el problema del significado y de la interpretación es ineludible. Por esto, si bien no han faltado resultados

<sup>5</sup> Ver Fortier y McConnell; Fortier.

<sup>6</sup> Ambas son técnicas de análisis lineales que permiten crear grupos de datos (en nuestro caso de textos) con base en la similitud de su distribución. Ver Burrows.

<sup>7</sup> La literatura sobre métodos cuantitativos para la atribución es bastísima. Recordamos solamente a Joula; Craig y Kinney; Love. Recientemente, en el debate literario italiano, estos métodos han salido a escena por su uso en la demostración de la naturaleza apócrifa del *Diario Postumo* de Montale, sobre el cual se puede consultar: Italia y Canettieri; pero, sobre todo, el amplio y documentado volumen de Condello.

críticos notables, los métodos computacionales no han tenido una influencia satisfactoria en los estudios literarios “tradicionales”, permaneciendo confinados en un enclave que crece en dimensiones, pero que no logra amalgamarse con el territorio que lo rodea.

Esta es una situación que ha estado bien presente en la conciencia de algunos de los estudiosos más sagaces teórica y epistemológicamente. Por ejemplo, Jerome McGann, en el prefacio a su influyente libro del 2001, *Radiant Textuality*, escribe:

La tecnología digital usada por académicos humanistas se ha enfocado casi exclusivamente en métodos para organizar, acceder y diseminar grandes *corpus* de materiales y en ciertos problemas especializados de estilística y lingüística computacional. Sobre este punto los estudios rara vez enfrentan aquellas preguntas sobre interpretación y reflexión autoconsciente que son preocupaciones centrales para la mayoría de académicos y educadores en las humanidades [...]. El campo general de la educación y la academia humanista no se tomará en serio el uso de la tecnología digital hasta que se demuestre cómo sus herramientas mejoran el modo en que exploramos y explicamos obras de arte —hasta que, precisamente, expandan nuestros procedimientos de interpretación. (xi-xii)

Pocos años después vuelve sobre este tema uno de los teóricos y metodólogos más acreditados del campo de las HD, Willard McCarty, quien en una serie de trabajos lamentó el papel puramente instrumental de los métodos computacionales para la crítica literaria:

La informática literaria está confinada a probar evidencia a favor o en contra de lo que ya conocemos o sospechamos. Está fuertemente inhibida en su capacidad de sorprender. Proveer evidencia parece justificación suficiente, pero la evidencia se torna cada vez más problemática en la medida en que el volumen de información excede la norma de las prácticas críticas formadas antes del crecimiento exponencial de recursos en la red. A medida que este volumen crece también lo hace la probabilidad de una escogencia arbitraria, así como la facilidad con la que una afirmación puede estar conectada con cualquier otra. Los buenos críticos pueden hacer mejores investigaciones al encontrar más de lo que necesitan; los críticos mediocres pueden prontamente

convertirse en peores con facilidad. El punto, no obstante, es que la informática literaria ha fungido de esta manera solo como una sirvienta callada y obediente, y no ha hecho mucho para rescatarse a sí misma de esta posición de debilidad desde la cual difícilmente puede realizar los beneficios que sus fieles le reconocen. Ha hecho poco por la educación metodológica de los académicos. (40-41)

McCarty sugiere que la causa principal de ese escaso impacto es la insuficiente definición teórica de la noción de texto en los estudios literarios digitales. Aun así, me parece posible decir que si un tema ha sido debatido desde hace tiempo, aunque sea solo en la reflexión teórica y metodológica sobre los procesos de codificación digital, es justamente aquel sobre la naturaleza del concepto de texto. Tal vez se podría argumentar con mayor razón que el gran peso de los enfoques posestructuralistas y neoidealistas en los estudios literarios de los últimos treinta años ha jugado un papel relevante en la profundización de la distancia entre ciencias literarias tradicionales y experimentaciones computacionales: la Teoría sin adjetivos no interactúa fácilmente con el formalismo de los métodos computacionales como ha señalado nuevamente David Hoover:

Mucha teoría de alto nivel está profundamente influenciada por ideas sobre la inestabilidad del signo y la tendencia de los textos a desintegrarse bajo presión crítica, ideas seguramente asociadas con el difunto Jacques Derrida. [...] Enfoques críticos como estos y una desconfianza más generalizada del vínculo entre el signo y el significado en la teoría literaria ayudó a crear un clima inhóspito al análisis textual y la estilística.<sup>8</sup> ("The End" s. p.)

También es cierto que los métodos y los instrumentos analíticos de la estilometría clásica eran demasiado rudimentarios para aspirar a una vasta adopción en la crítica literaria dominante. Debemos, en fin, admitir que

---

8 En realidad, en este ensayo, Hoover asume como objeto de polémica también (si no sobre todo) la tradición generativista chomskiana, acusándola de haber focalizado la atención de la lingüística a los aspectos sintácticos y gramaticales, en detrimento del interés por la semántica. Asimilar dos tradiciones tan distantes (si no opuestas) parece un poco arriesgado, pero en parte es justificado por la tan particular situación académica de los Estados Unidos, donde estas dos tradiciones han dominado *de facto* los estudios (y las oportunidades de carrera) en sus respectivos ámbitos.

estos han enfrentado insatisfactoriamente el problema de la especificidad y de la complejidad de los textos literarios y de los fenómenos culturales relacionados (géneros, estilo, períodos, temas y motivos, influencia, personaje, intertextualidad, isotopía).

## Lectura distante: un nuevo paradigma

En los últimos años, la emergencia de un nuevo paradigma metodológico en los estudios literarios parece abrir la posibilidad de superar la condición subalterna de los métodos computacionales en el análisis textual. Nos referimos a la noción de lectura distante (*distant reading*), teorizada por Franco Moretti por primera vez, en el ensayo “Conjectures for World Literature” y en el afortunado volumen *Graphs, Maps, Trees: Abstract Models for a Literary History*, que después se convirtió en marco de referencia metodológica para una cantidad considerable de estudios y experimentaciones. La acertada formulación de Moretti nace en manifiesta antítesis a la lectura atenta (*close reading*), elaborada en primera instancia por I. A. Richards y, sucesivamente, redefinida y elevada a método crítico de referencia de los formalistas americanos de la Nueva Crítica (después convertido en patrimonio común de los estudios crítico-literarios del siglo xx).<sup>9</sup>

La idea fundamental de la lectura distante es que existen hechos y fenómenos literarios y culturales, diacrónicos y sincrónicos, que no son accesibles a los métodos tradicionales de lectura profunda y de interpretación puntual de pocas (grandes o al menos notables) obras, sino que requieren el análisis masivo de centenares o miles de textos y documentos (y de millones de palabras). Como afirma Moretti en el ya citado “Conjectures”:

Pero el problema de la lectura directa (en todas sus encarnaciones, desde el nuevo criticismo a la deconstrucción) es que depende necesariamente de un canon muy reducido. Esto quizás se haya convertido ya en una premisa inconsciente e invisible, pero no deja de ser férrea: sólo invertimos tanto en los textos individuales si creemos que muy pocos de ellos valen realmente la pena. [...] En el fondo, es un ejercicio teológico —un tratamiento muy

---

<sup>9</sup> La noción de lectura atenta aparece por primera vez en el libro *Practical Criticism* de Richards. Después sucesivamente es adoptada y refuncionalizada por los New Critics; sobre esto ver North.

solemne de unos cuantos textos que se toman muy en serio—, mientras que lo que realmente necesitamos es un pequeño pacto con el diablo: sabemos leer textos, ahora aprendamos cómo no leerlos. La lectura distante, en la que la distancia, permítaseme repetirlo, es una condición para el conocimiento, nos permite centrarnos en unidades mucho menores o mucho mayores que el texto: recursos, temas, tropos; o géneros y sistemas. Y si entre lo muy pequeño y lo muy grande desaparece el texto en sí, bien, es uno de esos casos en los que es justificable decir que menos es más. (Moretti, “Conjeturas” 67-68)

Es evidente cómo esta actitud metodológica —que, nótese, Moretti elabora prescindiendo de mencionar específicas técnicas de análisis computacional, por demás ignoradas por él, como afirma en más de una ocasión<sup>10</sup> otorga una formidable base teórica para el renovado interés hacia los métodos cuantitativos y los análisis computacionales de los textos literarios que hemos presenciado en la última década. Una base teórica “interna” a los estudios literarios que interactúa con dos fenómenos, por así decirlo, “externos”: la disponibilidad de vastos archivos de textos digitales y de metadatos, producto de las campañas de digitalización masiva emprendidas en la última década del siglo pasado; y el desarrollo y difusión, también en el ámbito humanista, de técnicas e instrumentos de *big data analysis*, que hasta hace pocos años era prerrogativa de ámbitos disciplinarios asaz distantes como las ciencias físicas, biomédicas y parte de las económicas (Aiden y Michel).

Estas técnicas, aplicadas a grandes conjuntos de documentos digitales o de metadatos, permiten hacer *emergir* estructuras, regularidades y *patterns* de otra manera incognoscibles, que tienen un papel explicativo significativo en la comprensión de procesos literarios como la evolución de los géneros, la difusión de un estilo, la intertextualidad, la presencia de temas y contenidos recurrentes en un dado periodo histórico-literario. La crítica y la historiografía literaria, entonces, pueden finalmente disponer de una serie de aparatos instrumentales que permiten justificar sus explicaciones y generalizaciones sobre la base de la evidencia empírica, proporcionadas por el análisis de datos y no de la intuición y argumentación del estudioso. Y esto, además, se vuelve posible en una escala macro que hace irrelevante el estudio detallado de un

---

<sup>10</sup> Por ejemplo, ver la reciente entrevista que realizó con Melissa Dinsman.

único texto, como afirma Matthew Jockers en su *Macroanalysis*, otro texto de referencia para los cultivadores de la lectura distante:

Los masivos *corpus* digitales nos ofrecen un acceso no antes visto al registro literario e invitan, incluso demandan, un nuevo tipo de recolección de evidencia y de creación de sentido. El literato del siglo XXI ya no puede atenerse a evidencia anecdótica, con “cosas” casuales recogidas de algunos textos, incluso “representativos”. Debemos hacer un esfuerzo por entender estas cosas que encontramos interesantes en el contexto de todo lo demás, incluyendo la masa de textos posiblemente “no interesantes”. (8)

Afirmaciones tan radicales (también si, a decir verdad, vale la pena señalar que el extremismo de las formulaciones de manifiesto teórico no siempre corresponde a la verdadera práctica crítica consiguiente) obviamente determinaron reacciones críticas justificadas sobre las que volveremos más adelante. Antes, sin embargo, es oportuno hacer reseña de las técnicas y los métodos principales que han sido adoptados en los estudios que se inscriben en el paradigma de la lectura distante.

## **Los instrumentos de la lectura distante**

El conjunto de los métodos e instrumentos adoptados en las investigaciones y en los estudios inspirados por la lectura distante se puede subdividir en dos categorías: las técnicas de minería de textos (*text mining*) y aprendizaje automático (*machine learning*) y las de análisis de redes (*network analysis*; Jockers y Underwood).

En el primer grupo entran los métodos e instrumentos de análisis de datos en capacidad de determinar patrones, esquemas o tendencias emergentes dentro de cantidades enormes de datos textuales (y, por lo tanto, escasamente estructuradas) o de extrapolar de ellas características cuantificables (variables escondidas) útiles para subdividir el conjunto de textos en categorías significativas (*Clustering*). Estos métodos de análisis adoptan complejos algoritmos probabilísticos, de los cuales los más notorios derivan de la teoría probabilística bayesiana y del análisis multivariante.<sup>11</sup>

---

<sup>11</sup> Ver Han; Feldman y Sanger.

En particular en el ámbito de las HD y de los estudios literarios computacionales, en los últimos años, han gozado de un notable éxito los métodos de *topic modeling*. Desde el punto de vista técnico, el *topic modeling* es un procedimiento de *clustering* textual no supervisado (esto es, que no requiere de un conjunto de ejemplares clasificados manualmente para adiestrar el sistema), basado sobre dos tesis: un texto es una distribución de probabilidad de un conjunto de temas o argumentos (*topic*)<sup>12</sup> y un tópico es a su vez una distribución de probabilidad de un conjunto de términos y palabras. El aspecto particularmente relevante para fines de la aplicación en campo literario es que el procedimiento, aplicado a un conjunto de textos (o a segmentaciones arbitrarias de textos), produce de manera completamente automática una serie de listas de formas lexicales, que pueden ser interpretadas como grupos temáticos, y calcula la importancia en cada texto: en otras palabras, un análisis temático automatizado.<sup>13</sup>

Existen en efecto distintos algoritmos de *topic modelling*, pero el más difundido es aquel conocido como *Latent Dirichlet Allocation* (LDA; Blei). Intuitivamente, podemos decir que a la base de LDA se encuentra un modelo generativo del texto simplificado: cuando un autor escribe un texto, escoge el conjunto de los argumentos (*topic*) de los que quiere hablar y después determina la proporción con que cada argumento estará presente. Admitamos ahora que cada tópico posible pueda ser caracterizado como un conjunto no ordenado de palabras con una dada distribución: una especie de caja de palabras en la que una palabra puede repetirse  $n$  veces debido a su pertinencia respecto al argumento. El autor podrá “pescar” de manera casual en las distintas cajas que corresponden a los argumentos sobre los que pretende escribir y extraer de cada una un número de palabras proporcional al peso que pretende asignar a cada argumento. El texto resultante de la composición de estas palabras presentará, consecuentemente, una distribución lexical determinada por la relativa relevancia de los argumentos que trata y de la pertinencia de cada

---

<sup>12</sup> En el resto de la traducción utilice el término en castellano “tópico” [n. de la t.].

<sup>13</sup> Los dos autores de los trabajos más significativos en este ámbito son Matthew Jockers y Ted Underwood: ver Jockers, “Computing”; Jockers y Mimmo; Underwood, *Why; Goldstone y Underwood*. De ambos, además, se consultan sus respectivos blogs, en los que están publicados numerosos e interesantes artículos e intervenciones: ver Jockers, *Matthew L. Jockers*; Underwood, *The Stone*.

palabra en el contexto de cada argumento. Este modelo generativo, bajo ciertas presunciones, puede ser invertido, lo que es justamente el objetivo de LDA.<sup>14</sup>

Otra técnica de análisis nacida en el ámbito de la sociología cuantitativa, y que recientemente ha encontrado aplicación en lo literario, es el análisis de sentimientos (*sentiment analysis*). Esta técnica consiste en la atribución de un valor cuantitativo de la valencia emotiva de frases o de textos por medio de una métrica emocional que atribuye pesos a un dado conjunto de palabras consideradas significativas (Liu). Recientemente, Matthew Jockers ha experimentado con estas técnicas para desarrollar una aplicación llamada *syuzhet*, utilizada para un arriesgado y controversial análisis de textos narrativos decimonónicos en inglés. La idea teórica (inspirada en una sugerencia de Kurt Vonnegut)<sup>15</sup> consiste en que por medio del análisis de sentimientos es posible trazar la trayectoria emotiva de los textos, lo cual, a su vez, está correlacionado con la estructura profunda de la trama (*plot*) y, finalmente, en última instancia sería posible derivar un conjunto restringido de estructuras fundamentales del relato (Jockers, “A Novel Method”; “Revealing Sentiment”).<sup>16</sup>

En cuanto al análisis de redes, se trata de un método formal desarrollado para estudiar la estructura de las relaciones que se pueden instituir entre grupos de individuos (en los que los individuos pueden ser personas, instituciones, objetos o conceptos abstractos), basada en la teoría matemática de los grafos (Kadushin; Caldarelli y Catanzaro). Cada individuo constituye un nodo (o vértice) y cada relación un arco (o arista). La red resultante es un modelo formal

<sup>14</sup> Para una introducción no técnica a LDA ver Underwood, “Topic modeling”.

<sup>15</sup> Se trata en realidad de una sugerencia elaborada inicialmente por el escritor estadounidense para su trabajo de grado (por demás rechazado), muchos años después retomada en una conferencia (“Kurt Vonnegut”) cuyo texto está incluso en la autobiografía *A Man without a Country*, así resumida: “La idea fundamental es que las historias tienen formas que pueden ser graficadas en papel, y que la forma de las historias de una dada sociedad es al menos tan interesante como la forma de sus materas o puntas de lanza” (*Palm Sunday* 285).

<sup>16</sup> Consiguió un vivaz debate que involucró al mismo Jockers y a numerosos otros estudios muy activos en este campo (como Ted Underwood, Andrew Piper, Scott Weingart y Annie Swafford) que evidenció numerosos límites, técnicos y teóricos, en la propuesta. Para una reconstrucción ver Clancy. De cualquier manera, Jockers continuó con su trabajo y las sucesivas elaboraciones son descritas en los distintos artículos publicados en su blog. Me permito señalar las fuertes dudas que nutro acerca de la utilidad y los fundamentos teóricos de estos cuanto menos curiosos experimentos. Estas dudas han sido confirmadas luego de que un grupo de matemáticos haya publicado, sin citar a Jockers por cierto, un artículo que documenta una experimentación muy similar, cuyo nivel de (in)conciencia teórica es vergonzoso: ver Reagan et al.

y altamente abstracto de la estructura relacional interna del grupo. Algunas propiedades matemáticas de la red pueden ser adoptadas como sustitutos de las características cualitativas del dominio: la centralidad de un nodo, la distancia de (un subconjunto de) nodos respecto a otros, la individuación de subconjuntos de nodos proximales, entre otros. Estas técnicas deben su encanto también (si no, sobre todo) al hecho de poder ser naturalmente transformadas en cautivadoras gráficas dotadas de un elevado valor explicativo. Falta demostrar si y en qué medida los métodos de visualización puedan entregar una contribución genuina al conocimiento que de otra manera no podría conseguirse mediante el simple escrutinio de datos numéricos o de sus elaboraciones formales.

Las aplicaciones del análisis de redes en el ámbito de las HD son numerosas, sobre todo en ámbito sociológico e historiográfico. Pero también son relevantes las aplicaciones en el campo literario, en el que han sido utilizadas, ya sea para estudios de la sociología de la literatura, inherentes a la vida social y a la recepción de textos y las redes de relaciones entre escritores y las dinámicas de las instituciones literarias y editoriales, o para analizar las dinámicas relacionales de los personajes al interior de un texto (Moretti, “Network Theory”; Trickle; Prado et al.; So y Long).

## Límites y críticas de la lectura distante

Las técnicas eminentemente cuantitativas de *data analysis* adoptadas en los trabajos inspirados en la lectura distante parecen abrir horizontes ilimitados para los estudios literarios computacionales, sin ni siquiera requerir el dispendioso trabajo presupuesto por la investigación informática humanista clásica, tanto para la formalización de conceptos, modelos e hipótesis críticas como para la codificación rigurosa de los datos y de los documentos (en efecto, basándose en algoritmos estadísticos y probabilísticos, estas tienen una intrínseca tolerancia a la imprecisión de datos, por lo menos dentro de ciertos límites). Pero, como anticipado, si en una dirección el nuevo paradigma metodológico, teorizado por Moretti y fuertemente promovido por académicos como Matthew Jockers y Ted Underwood, ha encontrado un notable consenso y ha producido una gran cantidad de aplicaciones, análisis y estudios, es también verdad que ha suscitado numerosas reacciones críticas.

Una parte de estas reacciones provienen del mundo de los estudios literarios “tradicionales” y, en general, fundan su crítica a los métodos computacionales

sobre la (afirmada) irreductibilidad de lo literario a cualquier enfoque formal o cuantitativo por encima de su naturaleza “cualitativa” (argumentos no disímiles fueron usados contra el estructuralismo y la semiótica del texto en los años setenta).<sup>17</sup> Por ejemplo, Stanley Fish, aunque conceda que los análisis cuantitativos puedan, en algunos casos, producir resultados notables, recuerda cómo aquello que cuenta verdaderamente es la interpretación, que procede de un complejo proceso de construcción de hipótesis, contextualizaciones y formulaciones de juicios por parte del crítico, y concluye:

[...] Sea cual sea la visión de las humanidades digitales proclamada, tendrá poco lugar para críticos como yo y para el tipo de crítica que practico: una crítica que limita el sentido a los significados diseñados por un autor, una crítica que generaliza de un texto tan breve como media línea, una crítica que insiste en la distinción entre verdadero y falso, entre lo que es pertinente y el ruido, entre lo que es serio y lo que es simplemente un juego. (“Mind your P’s” s. p.)

Del mismo calibre, si no más radicales, son las críticas formuladas por exponentes de las distintas escuelas posestructuralistas que recuerdan cómo cada lectura es una operación de desarrollo crítico y que, por lo tanto, no puede intrínsecamente basarse sobre datos objetivos (*a fortiori* no sobre datos empíricos de carácter cuantitativo).<sup>18</sup>

Estas polémicas asumen en modo metonímico la lectura distante (o métodos cuantitativos) para expresar en realidad un rechazo, o cuanto menos un profundo escepticismo, frente a la totalidad del campo de las HD. Pero tanto los fundamentos teóricos como los éxitos aplicativos del nuevo paradigma han sido objeto de un acérrimo debate crítico también al interior de las HD. Numerosos trabajos producidos por estudios expertos en *data analytics* han puesto en duda la aplicabilidad, en el contexto humanista y literario, de métodos nacidos para finalidades y dominios en todo sentido diferentes o han evidenciado cómo un uso “poco consciente” pueda llevar a falacias o malentendidos. Por ejemplo, Sculley y Pasanek señalan algunas presunciones o principios *a priori* que están a la base de sistemas de *machine*

<sup>17</sup> Ver un ejemplar de esta actitud intelectual neoimpresionista: Marche.

<sup>18</sup> Un brillante ejemplo de este género de críticas a las HD, y en particular al giro cuantitativo de la lectura distante, es la de Eyers.

learning y que no son aceptables o tienen consecuencias problemáticas en ámbito humanista:

- la presunción de una naturaleza indiferenciada de la distribución de los datos;
- la presunción según la cual el espacio de las hipótesis derivables de los resultados sea restringido;
- la relevancia a los fines del análisis de los diversos métodos de cuantificación y codificación de los datos;
- las consecuencias del así llamado “*No free lunch theorem*”: ningún algoritmo es óptimo y no es posible construir un metaalgoritmo que minimice la criticidad y los límites. Como consecuencia, cada método produce resultados que son inevitablemente determinados por el método mismo y esto tiene consecuencias relevantes, *in primis* el riesgo de introducir circularidad en la argumentación cuando tales resultados son utilizados en contextos explicativos (Sculley y Pasanek).

Más recientemente Schmidt ha demostrado cómo una aplicación no consciente de los algoritmos de *topic modeling* probabilísticos como LDA pueda llevar a graves errores interpretativos, debido a dos presunciones teóricas tácitas que se revelan falaces: la de la coherencia semántica de la lista de palabras recogida en un tópico y aquella relativa a la estabilidad semántica de un tópico cuando aparece en textos distintos.

Sustancialmente, considero posible decir que me parece asaz controversial, si no para descartar del todo, la posibilidad de usar los resultados de un algoritmo probabilístico (como los de *topic model*) como evidencias empíricas en un contexto de justificación o explicación, como ya concluía John Unsworth hace algunos años:

Puedo estar inclinado a añadir una más, que es evitar acercarse a la aplicación de esta tecnología como un asunto de probar la veracidad de una hipótesis. Para fines literarios, [...] creo que tiene más sentido pensar que las herramientas de la minería de textos ofrecen provocaciones, traen a la superficie evidencia, sugieren patrones y estructuras, anuncian vagamente tendencias. Aunque la minería de textos es usualmente sobre predicción, precisión y verdades

fundamentales, en los estudios literarios, creo que es más sobre sorpresa, sugerión y capacidad negativa.<sup>19</sup> (s. p.)

Y esto no solo porque, como ha sido observado, en un contexto explicativo es necesario ser plenamente consciente del modo en que los datos han sido generados y de qué papel juegan como evidencia, pero, también, porque una explicación en ámbito literario debe dar cuenta de nociones y conceptos intencionales.<sup>20</sup>

En comparación con las criticidades teóricas y técnicas avanzadas de otras partes y de distintos contextos teóricos y disciplinarios (de los cuales hemos suministrado una limitada ejemplificación), algunos investigadores han elaborado enfoques de “mediación” metodológica (o, como se suele decir en ciencias sociales, *mixed methods*;<sup>21</sup> Johnson, Onwuegbuzie y Turner). Me refiero a la idea de lectura graduable (*scalable reading*) propuesta por Martin Mueller, acreditado anglista e historiador exponente de las HD, o la de hermenéutica computacional (*computational hermeneutics*) avanzada por Andrew Piper. En general, ellos proponen adoptar una síntesis dialéctica tanto de métodos cuantitativos, a nivel molar, para explorar la plausibilidad de hipótesis analíticas definidas, como de métodos cualitativos, a nivel molecular, para construir y justificar interpretaciones textuales:

Mi objetivo en este ensayo es ofrecer una polémica metodológica contra las posiciones exclusivistas en los debates de lectura atenta versus distante o profunda versus superficial que han hecho metástasis dentro de nuestro discurso crítico actual. Quiero que veamos cuán imposible es no moverse entre estos dos polos al momento de intentar construir argumentos literarios que operan a cierta escala (aunque cuándo ocurre este punto de giro permanezca siendo poco claro). En particular, quiero que veamos la integración necesaria del razonamiento cualitativo y cuantitativo que, como intentaré demostrar, tiene una naturaleza fundamentalmente circular y, por lo tanto, hermenéutica. (Piper 59)

<sup>19</sup> Observación hecha durante la serie de conferencias *How Not To Read A Million Books*, sostenidas entre el 2008 y 2009.

<sup>20</sup> Sobre este aspecto he insistido en Ciotti, “Sul distant reading”.

<sup>21</sup> En el resto de la traducción utilizo el término en castellano “métodos mixtos” [n. de la t.].

Hecha la excepción para los estudios que se ubican estrechamente en el ámbito socioliterario, podemos decir que tales enfoques metodológicos híbridos aparecen sin duda más apropiados cuando se les mide con el análisis de aquellas particulares tipologías textuales que son los textos literarios. Por otra parte, el mismo Moretti, en sus ensayos de carácter metodológico, afirma expresamente la necesidad de mantener el papel central de la interpretación crítica, si no de los textos, al menos de los datos producto del análisis (“Operationalizing”). Aun así, resta preguntarnos si tal interpretación deba entenderse en absoluto como un residuo subjetivo no analizable o si, en cambio, no es posible encontrar estrategias formales y computacionales adecuadas para modelizarla.

### **Una síntesis entre métodos semánticos y métodos cuantitativos**

Como he tenido ocasión de señalar, para aplicar de manera eficaz los métodos computacionales en el análisis crítico e histórico de la literatura es necesario definir su estatuto teórico de manera formal (Ciotti, “Toward”; “Sul distant reading”). Un método de análisis computacional guarda interés en la medida en que constituye un modelo formal de términos y nociones teóricas relevantes en una teoría del texto literario o en una metodología crítica. Desde este punto de vista, la modelización computacional es una variante de la noción de interpretación semiótica como explicación del texto propuesta por Umberto Eco:

La interpretación semántica o semiótica es el resultado del proceso por el cual el destinatario, frente a la manifestación lineal del texto, la llena de significado. La interpretación crítica o semiótica es en cambio aquella por medio de la cual se busca explicar por qué razones estructurales el texto puede producir esas (u otras alternativas) interpretaciones semánticas. (29)

La convergencia entre las definiciones de ontologías formales de dominio literario y la marcación digital de los textos mediante lenguajes formales permite, a mi parecer, operacionalizar la noción de interpretación semiótica de Eco (Gruber; Guarino, Oberle y Staab). La elaboración de interpretaciones y explicaciones, por parte de un estudioso, depende de presunciones que no son puramente individuales: metodologías, prácticas disciplinares y

criterios de aceptabilidad racional de las explicaciones son compartidas por una comunidad interpretativa (concepto que entendemos en buena parte en el sentido propuesto por Stanley Fish; *Is There a Text*), incluso si sus miembros pueden estar en desacuerdo sobre la interpretación puntual de hechos o fenómenos. Esta comunidad de intérpretes produce y valora las interpretaciones textuales y generalizaciones teóricas sobre la base de un *framework* conceptual y normativo tácito. La modelación ontológica permite no solo objetivar y explicitar (al menos en parte) dicho marco, sino también disponer de potentes métodos de elaboración y análisis:

- el conocimiento compartido, explicitado en una ontología, se vuelve operable con métodos de investigación semántica e inferencia automática;
- distintas ontologías pueden ser conectadas, por medio de la marcación semántica al mismo pasaje textual, creando múltiples niveles de marcación que abran la complejidad semántica;
- textos digitales con marcaciones de múltiples capas (*multi-layer*) pueden ser re-usados en diversos contextos de fruición y por distintos tipos de usuarios: expertos y estudiosos profesionales y no expertos pueden, entonces, cooperar, leer, visualizar y analizar los recursos en distintos niveles de complejidad y, por lo tanto, contribuir a su ulterior enriquecimiento.

Sobre estas bases se puede proceder a la construcción de un ambiente virtual de investigación que incluya vastos conjuntos de datos textuales en formatos de codificación estandarizada, instrumentos de marcación cooperativa ontológicamente orientada (*ontology driven*) e instrumentos de análisis cuantitativo y semántico y de visualización.<sup>22</sup> En este contexto, el enfoque estadístico cuantitativo y el formal cualitativo pueden cooperar, como sugieren los abanderados de los métodos mixtos, para analizar los fenómenos literarios en distintos niveles de abstracción y complejidad.

Puesto que tal ambiente tiene el objetivo de formalizar y externalizar los procesos interpretativos de los estudiosos y de los relativos resultados críticos,

---

<sup>22</sup> Naturalmente, la máquina hermenéutica no podría ser *sic et simpliciter* un proyecto basado en el *crowdsourcing*, que tanto atrae muchos exponentes de las HD, así como la marcación científica no es indexación social (*social tagging*). Podremos, sin embargo, definirla como un proyecto basado en una comunidad competente y motivada.

propongo llamarlo *Hermeneutic Machine*. Su construcción es obviamente una tarea bastante compleja, pero muchos de sus posibles componentes están ya disponibles y requieren en definitiva so lo una oportunidad de integración.<sup>23</sup> Para concluir, ya sea que adopten enfoques estadísticos cuantitativos o que tengan predilección por enfoques semánticos, basados en la computación enfocada a lo humano (*human centered computing*; Jaimes et al.; Jaimes, Sebe y Gatica-Pérez), un requisito parece claro: las infraestructuras de investigación colaborativa en la era digital son estratégicas, también, para la investigación humanista y digital. Esta es lamentablemente una debilidad estructural de la investigación humanista italiana, ya sea por mentalidad o por falta de recursos. Los investigadores de muchos otros países y áreas lingüísticas y culturales pueden usufructuar en este momento de grandes archivos abiertos y compartidos con decenas de miles de recursos primarios y secundarios, de ambientes digitales de investigación, de instituciones de apoyo a la investigación digital. Hay que desear que el mundo de los estudios humanistas italianos (de los cuales la Italianística es, por obvios motivos, uno de los pilares) sepa asumir esta oportunidad e invertir en esta dirección recursos e inteligencias.

## Obras citadas

- Aiden, Erez, y Jean-Baptiste Michel. *Uncharted: Big Data as a Lens on Human Culture*. Nueva York, Riverhead Books, 2013.
- Blei, David M. “Topic Modeling and Digital Humanities”. *Journal of Digital Humanities*, vol. 2, núm. 1, 2013, s. p.
- Burrows, John. *Computation into Criticism*. Clarendon, Oxford, 1987.
- Busa, Roberto. “The Annals of Humanities Computing: The Index Thomisticus”. *Computers and the Humanities*, vol. 14, núm. 2, 1980, págs. 83-90. DOI: <https://doi.org/10.1007/BF02403798>

<sup>23</sup> Después de este artículo, Fabio Ciotti ha optado por los métodos cuantitativos y sostiene que existe una cesura entre la lectura distante y los estudios literarios tradicionales con su paradigma hermenéutico. En un artículo del 2021, Ciotti argumenta que “la teoría biocognitiva de la literatura y la evolución cultural puede ofrecer la mejor opción para potenciar los métodos de la literatura distante y para comprender un conjunto completo de fenómenos relacionados con la literatura y la cultura que se encuentran más allá de los acercamientos tradicionales” (“Distant Reading” 210) [n. de la t.].

- Caldarelli, Guido, y Michele Catanzaro. *Networks: A Very Short Introduction*. Oxford, Oxford University Press, 2012.
- Ciotti, Fabio. “A landscape of Digital Humanities in Italy: backgrounds, institutions and infrastructures”. *Testo e senso*, vol. 16, 2015, s. p.
- . “Distant Reading in Literary Studies: A Methodology in Quest of Theory”. *Testo e Senso*, núm. 23, 2021, págs. 195-213.
- . *Il manuale TEI Lite*. Milán, Bonnard, 2004.
- . “La rappresentazione digitale del testo: il paradigma del *markup* e i suoi sviluppi”. *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlando*. Editado por Lorenzo Perilli y Domenico Fiornante. Florencia, Le lettere, 2011.
- . “Sul distant reading: una visione critica”. *Semicerchio*, vol. LIII núm. 2, 2015, págs. 12-20.
- “Toward a Formal Ontology for Narrative”. *MATLIT: Materialidades Da Literatura*, vol. 4, núm. 1, 2016, págs. 29-44. doi: [https://doi.org/10.14195/2182-8830\\_4-1\\_2](https://doi.org/10.14195/2182-8830_4-1_2)
- Clancy, Eileen. “A Fabula of Syuzhet: A Contretemps of Digital Humanities (with Tweets)”. *Storify*. Web.
- Condello, Federico. *I filologi e gli angeli. È di Eugenio Montale il “Diario postumo”?* Boloña, BUP, 2014.
- Craig, Hugh, y Arthur F. Kinney, eds. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge, Cambridge University Press, 2009.
- Dinsman, Melissa. “The Digital in the Humanities: An Interview with Franco Moretti”. *Los Angeles Review of Books*. Web. 2 de marzo de 2016.
- Eco, Umberto. *I limiti dell'interpretazione*. Milán, Bompiani, 1990.
- Eyers, Tom. “The Perils of the ‘Digital Humanities’: New Positivisms and the Fate of Literary Theory.” *Postmodern Culture*, vol. 23 núm. 2, 2013, s. p. doi: <https://doi.org/10.1353/pmc.2013.0038>
- Feldman, Ronen, y James Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge/Nueva York, Cambridge University Press, 2007.
- Fish, Stanley. *Is There a Text in This Class? The Authority of Interpretive Communities*. Cambridge (Mass.)/Londres, Harvard University Press, 1980.
- . “Mind your P's and B's: The Digital Humanities and Interpretation”. *Opinionator Blog, The New York Times*. Web. 23 de enero de 2012.

- Fortier, Paul A. "Analysis of Twentieth-Century Prose Fiction: Theoretical Context, Results, Perspective". *Literary Computing and Literary Criticism. Theoretical and Practical Essays on Theme and Rhetoric*. Editado por Rosanne G. Potter. Filadelfia, University of Pennsylvania Press, 1989, págs. 77-95.
- Fortier, Paul A., y J. Colin McConnell. *THEME: A System for Computer-Aided Theme Searches of French Texts*. Winnipeg, Department of French and Spanish/University of Manitoba, 1975.
- Gigliozzi, Giuseppe. "Memoriale di Paolo Volponi". *Letteratura italiana. Le opere*. Vol. iv. Turín, Einaudi, 1995, págs. 729-69
- Goldstone, Andrew, y Ted Underwood. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us". *New Literary History*, vol. 45, núm. 3, 2014, págs 359-384. doi: <https://doi.org/10.1353/nlh.2014.0025>
- Gruber, Tom. "Ontology". *Encyclopedia of Database Systems*. Editado por Ling Liu y M. Tamer Özsu. Nueva York, Springer, 2009.
- Guarino, Nicola, Daniel Oberle y Steffen Staab. "What is an Ontology?". *Handbook on ontologies*. Berlin/Heidelberg, Springer, 2009, págs. 1-17.
- Han, Jiawei. *Data Mining: Concepts and Techniques*. Burlington, Elsevier, 2011.
- Hoover, David L. "Textual Analysis". *Literary Studies in the Digital Age: An Evolving Anthology*. Editado por Kenneth M. Price y Ray Siemens. Nueva York, Modern Language Association, 2013, s. p. doi: <https://doi.org/10.1632/llda.2013.3>
- . "The End of the Irrelevant Text: Electronic Texts, Linguistics, and Literary Theory". *Digital Humanities Quarterly*, vol. 1, núm. 2, 2007, s. p.
- Italia, Paola y Paolo Canettieri. "Un caso di attribuzionismo novecentesco: il 'Diario Postumo' di Montale". *Cognitive Philology*, núm. 6, 2013, s. p.
- Jaimes Alejandro, Nicu Sebe y Daniel Gatica-Pérez. "Human-centered computing: a multimedial perspective". *Proceedings of the 14<sup>th</sup> annual ACM international conference on Multimedia*. Nueva York, ACM Press, 2006, págs. 855-864. doi: <https://doi.org/10.1145/1180639.1180829>
- Jaimes, Alejandro et al. "Human-Centered Computing: Toward a Human Revolution". *IEEE Computer*, vol. 40, núm. 5, 2007, págs. 30-34. doi: <https://doi.org/10.1109/mc.2007.169>
- Jockers, Matthew L. "A Novel Method for Detecting Plot". *Matthew L. Jockers*. Web. 5 de junio de 2014.

- . “Computing Ireland’s Place in the 19<sup>th</sup> Century Novel: A Macroanalysis”. *Breac: A Digital Journal of Irish Studies*. Web. 7 de octubre de 2015.
- . *Macroanalysis: Digital Methods and Literary History*. Urbana/Chicago/Springfield, University of Illinois Press, 2013.
- . *Matthew L. Jockers*. Web. <https://www.matthewjockers.net/>
- . “Revealing Sentiment and Plot Arcs with the Syuzhet Package”. *Matthew L. Jockers*. Web. 2 de febrero del 2015.
- Jockers, Matthew L. y David Mimno “Significant themes in 19th-century literature”. *Poetics*, núm. 41, 2013, págs. 750-769. DOI: <https://doi.org/10.1016/j.poetic.2013.08.005>
- Jockers, Matthew L. y Ted Underwood. “Text-Mining in the Humanities”. *A New Companion to Digital Humanities*. Editado por Susana Schreibman, Ray Siemens y John Unsworth Malden, Wiley Blackwell, 2016, págs. 291-306.
- Johnson, R. Burke, Anthony J. Onwuegbuzie y Lisa A. Turner. “Toward a Definition of Mixed Methods Research”. *Journal of Mixed Methods Research*, vol. 1, núm. 2, 2007, págs. 112-133. DOI: <https://doi.org/10.1177/1558689806298224>
- Jones, Steve E. *Roberto Busa, S.J. and the Emergence of Humanities Computing: The Priest and the Punched Cards*. Nueva York, Routledge, 2016.
- Joula, Patrick. “Authorship Attribution”. *Foundations and Trends® in Information Retrieval* vol. 1, núm. 3, págs. 233-334. DOI: <http://doi.org/10.1561/1500000005>
- Kadushin, Charles. *Understanding Social Networks: Theories, Concepts, and Findings*. Nueva York, Oxford University Press, 2012.
- “Kurt Vonnegut on the Shapes of Stories”. *Youtube*, publicado por David Comberg. Web. 30 de octubre de 2010.
- Liu, Bing. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Nueva York, Cambridge University Press, 2015.
- Love, Harold. *Attributing Authorship: An Introduction*. Cambridge, Cambridge University Press, 2002.
- Marche, Stephen. “Literature is not Data: Against Digital Humanities”. *Los Angeles Review of Books*. Web. 28 de octubre del 2012.
- McCarty, Willard. “Literary enquiry and experimental method: What has happened? What might?”. *Storia della Scienza e Linguistica Computazionale: Sconfinamenti Possibili*. Editado por Liborio Dibattista. Milán, Franco Angelli, 2009, págs. 32-54

- McGann, Jerome. *Radiant Textuality: Literature after the World Wide Web*. Nueva York, Palgrave, 2001.
- Moretti, Franco. “Conjecturas sobre la literatura mundial”. *The New Left Review*, vol. 3, 2000, págs. 65-76.
- . *Distant Reading*. Londres, Verso, 2015.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. Londres, Verso, 2005.
- . “Network Theory, Plot Analysis”. *The New Left Review*, vol. 2, núm. 68, 2011, págs. 80-102.
- . “Operationalizing: Or, the Function of Measurement in Literary Theory”. *The New Left Review*, vol. 2, núm. 84, 2013, págs. 103-119.
- Mueller, Martin. “Shakespeare His Contemporaries: collaborative curation and exploration of Early Modern drama in a digital environment”. *Digital Humanities Quarterly*, vol. 8 núm. 3, 2014, s. p.
- North, Joseph. “What’s ‘New Critical’ about ‘Close Reading’?: IA Richards and His New Critical Reception”. *New Literary History*, vol. 44, núm. 1, 2013, págs. 141-157. DOI: <http://doi.org/10.1353/nlh.2013.0002>
- Orlandi, Tito. “Per una storica dell’informatica umanistica”. *Dall’Informatica umanistica alle culture digitali, Atti del convegno di studi (Roma, 27-28 ottobre 2011) in memoria di Giuseppe GiglioSSI*. Editado por Fabio Ciotti y Gianfranco Crupi. Roma, Università Sapienza/Digilab, 2012, págs. 49-102.
- Piper, Adrew. “Novel Devotions: Conversational Reading, Computational Modeling, and the Modern Novel”. *New Literary History*, vol. 46, núm. 1, 2015, págs. 63-98. DOI: <https://doi.org/10.1353/nlh.2015.0008>
- Prado, Sandra D. et al. “Temporal Network Analysis of Literary Texts”. *Advances in Complex Systems*, vol. 19 núm. 3 y 4, 2016, s. p. DOI: <https://doi.org/10.1142/S0219525916500053>
- Reagan, Andrew J. et al. “The emotional arcs of stories are dominated by six basic shapes”. *EPJ Data Sci*, vol. 5, núm. 31, 2016, s. p. DOI: <https://doi.org/10.1140/epjds/s13688-016-0093-1>
- Richards, I. A. *Practical Criticism*. Londres, Kegan Paul, Trench, Trubner & Co, 1930.
- Schreibeman, Susan y Ray Siemens, editores. *Companion to Digital Literary Studies*. Malden, Blackwell, 2007.
- Schreibman, Susan, Ray Siemens y John Unsworth, eds. A *Companion to Digital Humanities*. Malden, Blackwell, 2004.
- . *A New Companion to Digital Humanities*. Malden, Wiley Blackwell, 2016.

- Schmidt, Benjamin M. "Words Alone: Dismantling Topic Models in the Humanities". *Journal of Digital Humanities*, vol. 2, núm. 1, 2012, s. p.
- Sculley, D. y Bradley M. Pasanek. "Meaning and Mining: The Impact of Assumptions in Data Mining for the Humanities". *Literary and Linguistic Computing*, vol. 23, núm. 4, 2008, págs. 409-424. DOI: <https://doi.org/10.1093/llc/fqno19>
- So, Richard Jean y Hoyt Long. "Network Analysis and the Sociology of Modernism". *Boundary*, vol. 2, núm. 1, 2013, págs. 147-182. DOI: <https://doi.org/10.1215/01903659-2151839>
- TEI Consortium. *Text Encoding Initiative*. Web. <https://tei-c.org/>
- Terras, Melissa, Julianne Nyhan y Edward Vanhoutte, eds. *Defining Digital Humanities. A reader*. Ashgate, Farnham, 2013.
- Trickle, Peer. "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft". *Empirie in der Literaturwissenschaft*. Editado por Philip Ajouri, Katja Mellmann y Christoph Rauen. Münster, Mentis, 2013, págs. 201-247.
- Underwood, Ted. *The Stone and the Shell*. Web. <https://tedunderwood.com/>
- . "Topic modeling made just simple enough". *The Stone and the Shell*. Web. 7 de abril de 2012.
- . *Why Literary Periods Mattered: Historical Contrast and the Prestige of English Studies*. Stanford, Stanford University Press, 2013.
- Unsworth, John et al. "How Not to Read a Million Books". *John M. Unsworth*. Web.
- Vonnegut, Kurt. *A Man without a Country*. Nueva York, Seven Stories Press, 2005.
- . *Palm Sunday*. Nueva York, Dial Press Trade Paperbacks, 2006.

### Sobre el autor

Fabio Ciotti es profesor asociado de la Universidad de Roma “Tor Vergara”, donde enseña “Teoría y crítica computacional de la literatura”. La actividad científica y didáctica de Fabio Ciotti se concentra en temas de las humanidades digitales y de los estudios literarios computacionales. En particular, se ha interesado por la aplicación de métodos computacionales en el análisis de textos literarios, por el uso de lenguajes de XML y TEI en el ámbito humanista, por la proyección e implementación de bibliotecas digitales, por la aplicación de modelos y tecnologías de la web semántica en las disciplinas humanistas. Sobre estos temas ha publicado distintas monografías y más de cuarenta artículos en revistas académicas y volúmenes misceláneos.

Fabio Ciotti ha participado en numerosos proyectos de investigación en Italia y en Europa. Ha organizado y ha sido miembro del comité científico de conferencias nacionales e internacionales (TEI Conference and Members Meeting; AIUCD conferences, IRCDL; DH), en particular, ha presidido el Comité de Programación de la conferencia DH2019 (Utrecht, 2019). Actualmente, preside el Comité Ejecutivo de la Asociación Europea de Humanidades Digitales (cuyas siglas en inglés son EADH) y ha sido fundador y presidente de la Asociación por la Informática Humanista y la Cultura Digital (cuyas siglas en italiano son AIUCD), de la cual dirige la revista *Umanistica Digitale*.

### Sobre la traductora

Elena Pérez Palacio es estudiante del pregrado en Estudios Literarios de la Universidad Nacional de Colombia. Actualmente, se desempeña como la coordinadora del grupo estudiantil de literaturas eslavas *Yasnaia Poliana*.