

Piper, Andrew. *Enumerations: Data and Literary Study*. Chicago, University of Chicago Press, 2018, 258 págs.

El análisis literario computacional y digital se suele relacionar con conceptos como velocidad, objetividad, mecanicidad, lectura distante y productividad, términos polémicos para la crítica literaria reciente. Investigadores como Daniel Allington, Sarah Brouillette y David Columbia ven en las humanidades digitales una forma de trabajo académico que produce resultados inmediatos para la industria y que, además, forma egresados entrenados para satisfacer las necesidades de un mercado laboral neoliberal. Si bien reconocen que las humanidades digitales no son en esencia un enfoque metodológico exclusivo del neoliberalismo, sí afirman que así es como están siendo empleadas actualmente en las universidades norteamericanas. En *Enumerations*, Andrew Piper intenta mostrar una faceta distinta de los métodos computacionales digitales.

Como punto de inicio, para comprender aquello que llamamos humanidades digitales, es necesario distinguir tres conceptos fundamentales a menudo confundidos: lo cuantitativo, lo computacional y lo digital. Los primeros dos conceptos se refieren a formas de análisis, mientras que el otro está relacionado con los medios empleados para ello. El análisis cuantitativo parte de identificar iteraciones y cantidades en la información de un texto literario. Este procedimiento está implícito en muchas de las prácticas de lectura tradicionales de un crítico literario como lo son la identificación de recurrencias de palabras, ideas o motivos, la descripción de un patrón, o el reconocimiento de una tendencia. El análisis computacional implica diseñar un modelo o algoritmo que permita dar sentido a tales cantidades. Por ejemplo, la idea de romanticismo en la literatura puede ser un modelo para explicar la repetición o coincidencia de ciertos rasgos ideológicos, formales y temáticos en un conjunto de obras. Por su parte, lo digital tiene que ver con el uso de herramientas informáticas para facilitar la obtención y el procesamiento de los datos, cuantitativos o no.



En *Enumerations*, Piper discute la necesidad de incorporar las aproximaciones metodológicas de las humanidades digitales al amplio conjunto de herramientas de análisis del crítico literario. Tal integración resulta en una nueva noción de pensamiento humanista que cuestiona la oposición entre lo numérico y lo literario. Si bien se suele asociar lo literario con el carácter subjetivo, individual y creativo del pensamiento y lo numérico con lo objetivo, exacto y universal, el nuevo humanismo debe partir de comprender holísticamente ambos componentes del pensamiento humano. Su visión busca eliminar la idea de que el uso de herramientas tecnológicas aniquilará lo subjetivo en busca de un positivismo ciego. Todo lo contrario: enfatiza en el rol central del sujeto al momento de formular las preguntas que orientan el análisis de datos y su responsabilidad en la posterior interpretación.

Ahora, el primer obstáculo para tal integración parece ser la gran distancia entre las formas de análisis de la crítica tradicional y los métodos de las humanidades digitales. Sin embargo, Piper sostiene que tal distancia no es tan grande, pues, con recursos digitales o sin ellos, los procedimientos de lectura, como la *close reading*, y prácticas como la transcripción, parafraseo, traducción, antologización, canonización y memorización son formas de repetición y lectura computacional que desde tiempos remotos se han empleado como mecanismos de estabilización cultural.

Los medios digitales ofrecen un conjunto de herramientas y aproximaciones metodológicas para poner a prueba, confirmar, corregir o reorientar aquellas afirmaciones que la crítica tradicional ha construido sobre la literatura. Un ejemplo de aquellas prácticas de la crítica literaria que podrían mejorar a partir de la incorporación de métodos computacionales digitales es la tendencia a establecer generalizaciones sobre la historia de la literatura desde una evidencia que no es suficientemente representativa. Particularmente, Piper describe el procedimiento que Erich Auerbach emplea en *Mimesis* de la siguiente manera: se presenta un pasaje ejemplar y, a partir de este, se propone una generalización sobre una gran obra que, a su vez, le permite construir parte de una generalización más amplia sobre la historia de la literatura occidental. El problema en su metodología es el fallo en la representatividad de los ejemplos. Un fragmento o una obra, por destacables que sean, son insuficientes para describir toda la literatura de una época.

La crítica tradicional, con Auerbach como ejemplo, falla en mostrar la tensa relación entre las partes y el todo, entre un documento particular y un contexto desbordantemente más amplio (Piper 7).

Lo problemático en la generalización

El problema de la generalización no es la operación mental en sí, sino la forma en que se realiza. El estudio literario computacional no abandona la generalización, sino que pone en evidencia la carencia de una ciencia de la generalización que permita reunir el conocimiento, que la crítica ha elaborado a partir de obras singulares, y reflexionar sobre su representatividad con respecto a toda la producción literaria conocida para determinado momento histórico. La *close reading*, como herramienta de cómputo de información, permite al crítico dar cuenta de las relaciones de sentido de una obra a partir del análisis focalizado de sus componentes. No obstante, ¿con qué herramientas se cuenta al momento de construir ideas más amplias sobre un texto o un contexto a partir de tal análisis? (9). Una posible respuesta está en la aplicación de métodos cuantitativos en el estudio de la literatura. Piper propone que, a través de herramientas como el modelado de datos, el uso de diagramas y el análisis de la distribución de la información, es posible construir generalizaciones más acertadas y comprobables respecto a la literatura. Esto se fundamenta en la hipótesis de que la literatura no parte de lo raro y lo singular, sino de lo común y lo colectivo, pues el pensamiento y el lenguaje como material primario funcionan, en gran medida, por medio de la repetición de estructuras. Es decir, que prestar atención a las iteraciones y las cantidades puede revelar diversos canales y variantes de la expresión cultural, así como conexiones profundas entre palabras, ideas y formas.

Cada uno de los apartados de *Enumerations* funciona como un ejemplo concreto de cierto tipo de aproximación cuantitativa y computacional a problemas ya conocidos. El capítulo uno versa sobre el sentido de la puntuación, su distribución y sobreacumulación. Aquí se emplea la utilidad *grep* como herramienta de análisis. En los capítulos dos y tres, se aproxima a asuntos como la trama y el modelado de tópicos a través de espacios vectoriales y circuitos sociales (*social networks*). En el capítulo cuarto (ficcionalidad), se aborda el problema de la distinción entre ficción y no

ficción por medio del aprendizaje automatizado (*machine learning*). En el capítulo quinto, se emplea el análisis de dependencias y la resolución de correferencias para estudiar cómo empleamos el lenguaje para crear eso que llamamos personaje. Y el capítulo final se vale de modelos anidados para tratar de medir el cambio en el estilo de un autor a lo largo de su obra.

Es llamativo que Piper enfoque sus aproximaciones en aquellos problemas que la crítica tradicional no ha abordado por la ausencia de una metodología para tal fin. Fenómenos como la puntuación han sido ampliamente estudiados por la lingüística, pero sobre su funcionamiento en la literatura y su relación con los géneros poco se ha elaborado. Piper retoma el llamado de Georges Bataille a la construcción de una historia de la economía de la puntuación para entender de qué manera las tácticas de interrupción, retardamiento, ritmo y periodicidad construyen sentido y constituyen una forma esencial de comunicación (23). En la poesía en lengua inglesa del siglo xx encontraremos fenómenos como la sobrepuntuación, práctica poética que manifiesta a un ser incapaz de adherirse a sí mismo o una comunidad. Los escritores de la época encontraron en la coma y las cesuras una forma de expresar un carácter reflexivo, a veces contradictorio, mientras que el punto resultó útil para expresar lo concreto y enfático. En este ejemplo, convergen la lectura distante de Franco Moretti y la cercana de Ivor Richards y William Empson, pues se emplean técnicas computacionales para el análisis global de expresiones regulares en múltiples obras, pero tales datos tienen un sentido particular para cada pieza analizada, reconociendo lo distintivo de los poemas y lo general de la práctica poética. Es decir que se emplea la *close reading* para validar prácticas de lectura distante y proponer así nuevas generalizaciones o modelos de interpretación. Este caso, además, es interesante dado que, en contra de la visión de la computación como una herramienta al servicio de la velocidad industrial, el procedimiento empleado aquí obliga a detenerse para reflexionar sobre aquello que constituye sentido en la literatura.

El modelado de datos como proceso de pensamiento

Una idea central en la obra de Piper es que la modelación de algoritmos, digitales o no, es una herramienta de análisis literario que permite al investigador tomar una posición crítica ante su propio proceso de lectura. Se parte de que un modelo es una construcción teórica que media entre el

investigador y sus observaciones, una forma de dar sentido a fenómenos complejos como la literatura. Es por ello por lo que examinar la manera en la cual construimos modelos es comprender que el conocimiento siempre está mediado por una construcción artificiosa. Así, la literatura siempre se ha cimentado sobre modelos como lo es la teoría de Auerbach. La diferencia entre los modelos teóricos tradicionales y los computacionales es que estos últimos tienen la virtud de ser explícitos y exhibir sus limitaciones. En la crítica tradicional, los grandes académicos hablan de pasajes y obras que encajan en sus hipótesis, pero no es habitual que se hable de todos aquellos otros fragmentos u obras que contradecían sus afirmaciones o que simplemente no fueron tomados en consideración. En resumidas cuentas, la teoría suele desconocer la prevalencia de los fenómenos observados (11). Por supuesto, Piper reconoce que siempre hay y habrá dimensiones tácitas en los modelos, pero, precisamente, por manifestar su naturaleza, es posible acercarse a ellos con conciencia de su limitación.

En cuanto a la aproximación al fenómeno estudiado, los humanistas por tradición muchas veces han sido opacos con respecto a sus metodologías, a las evidencias que tuvieron en cuenta y a lo que consideraron digno de ser valorado como evidencia. En este sentido, el modelado resulta provechoso en cuanto hace innecesario el argumento de autoridad. Los modelos demuestran su validez al definir previamente cada paso del proceso intelectual que lleva de las partes al todo. Claro está que las instrucciones del algoritmo pueden ser erróneas y sesgadas. Por ello, Piper comparte al final de su libro todos los códigos de programación que empleó de modo que el procedimiento de cómputo queda abierto a todos para ser evaluado, criticado y mejorado. Esto implica que el modelado computacional no debe representar una práctica de deshumanización de la lectura, sino de colectividad. Tampoco se presenta como una forma única de conocimiento ni la única válida, pero sí una útil. Es un acto de reflexión sobre el propio pensamiento, pues a través de ella hacemos explícitos los pasos que estamos tomando para llegar a una afirmación sobre la literatura (en palabras de Walter Benjamin: una conciencia del “inconsciente léxico” del texto). En resumen, el modelado no pretende establecer un valor de verdad objetiva y absoluta, más bien afirma que, bajo determinadas condiciones específicas establecidas por el investigador, uno podría creer que algún aspecto de la literatura puede ser cierto.

La idea que más me interesa de Piper es su planteamiento sobre la posibilidad de autoanalizar nuestro pensamiento, nuestra forma de leer e, incluso, de escribir a través de los datos. Piper no solo aplica el análisis de datos a otras obras literarias, sino que también lo aplica a su libro, siendo este mismo la prueba de que los métodos cuantitativos son una poderosa herramienta de introspección. De datos como la cantidad de palabras empleada en cada capítulo, cuántas y cuáles eliminó de versiones anteriores, Piper obtiene información valiosa sobre sus propios procesos de pensamiento (qué posibles aproximaciones ha contemplado y descartado a lo largo de la composición del libro). Así mismo, compara sus términos más frecuentes y construye un campo semántico de su propia obra para analizar qué es lo que realmente ha comunicado.

La lectura de *Enumerations* deja claro que los métodos computacionales no son una forma de mecanizar el análisis literario ni de eliminar su carácter subjetivo. En todo modelo está implicado un artífice que dirige *a priori* qué información vale la pena estudiar. Él mismo es quien lee e interpreta los datos para formular cierta afirmación sobre la obra o la literatura en general.

Por medio de los datos, el investigador puede retomar una posición crítica no solo ante su obra sino ante todo un campo de estudio o disciplina. Es importante estudiar en torno a qué ha girado la crítica que emplea métodos cuantitativos, qué ideas son recurrentes, qué se ha ignorado, qué información se cita y qué impacto ha tenido. Todo ello para entender cómo se ha desarrollado el campo como un todo y evaluar aquello que se ha logrado y aquello en lo que se ha fallado. Un ejemplo de ello es la confirmación de un canon a través de la labor investigativa. Shakespeare aparecía en artículos y trabajos 4.1 veces más que John Milton en 1870. Hoy en día Shakespeare aparece 4.5 veces más que otros autores: la dominación del canon continua.

Otra forma de adoptar una postura crítica a partir del análisis cuantitativo es evaluar a quiénes leemos y la incidencia en ello de ciertas prácticas académicas de la crítica literaria moderna. Piper encuentra información valiosa sobre la inequidad institucional. Luego de revisar la cantidad de publicaciones académicas, identifica un sesgo hacia ciertas instituciones de la élite que son responsables de la producción de gran parte de la investigación en el campo de los estudios literarios. Más de la mitad de las publicaciones académicas está centrada en diez universidades de las cuales las primeras dos, Harvard y Yale, tienen la mayor parte. De aquí surge la necesidad

de construir un sistema académico más inclusivo. Nuestros métodos de selección, contratación, citación, promoción y publicación están sesgados y refuerzan desproporciones institucionales en el campo académico. En otras palabras, los datos además permiten ver estos desbalances, irregularidades y sesgos ideológicos que no siempre son evidentes, son una herramienta que problematiza nuestra perspectiva crítica, pero también a nosotros mismos como investigadores dentro de un contexto.

Andrew Piper es el director del laboratorio de análisis cultural.*txtlab* y editor de la revista *Journal of Cultural Analytics* en la universidad de McGill, donde también es profesor. Su investigación se ha enfocado en el estudio histórico del impacto de la tecnología en los procesos de lectura y en el estudio de la narrativa a través de herramientas de la ciencia de datos, el aprendizaje automático y el procesamiento del lenguaje natural. Entre sus publicaciones se destaca *Dreaming in Books: The Making of the Bibliographic Imagination in the Romantic Age* (Chicago 2009), obra que recibió el premio de la Asociación de lenguas modernas y una mención honorífica en el Premio Harry Levin de la Asociación americana de literatura comparada.

John Alexander Contreras Caicedo
Universidad Nacional de Colombia, Bogotá, Colombia

Obras citadas

Allington, Daniel. “Neoliberal Tools (and Archives): A Political History of Digital Humanities.” *Los Angeles Review of Books*. Web. 1 mayo de 2016.