

Una nueva visión de la supuesta influencia de Madame Bovary en La Regenta a través de la estilometría y el análisis de sentimientos basados en lenguaje R

María LOZANO ZAHONERO
Università degli Studi di Roma "Tor Vergata"

Resumen

La supuesta influencia de *Madame Bovary* en *La Regenta*, rodeada desde el inicio de polémicas y enfrentamientos, ha sido objeto de numerosos estudios críticos. El enfoque tradicionalmente adoptado ha sido de tipo cualitativo y se ha fundado en datos parciales, no siempre objetivos. Es más, en ocasiones, se han tomado como base de las distintas hipótesis tan solo impresiones meramente anecdóticas y, en consecuencia, los resultados obtenidos han sido discordantes. El objetivo principal de este trabajo es aportar datos cuantitativos que contribuyan a dar respuesta a esta cuestión aún abierta. Con este fin, llevaremos a cabo un análisis computacional de los patrones estilísticos y la dimensión emotiva que subyacen en ambas novelas utilizando para ello el lenguaje de programación R. Además de este objetivo primario se abordará también secundariamente la comparación de la versión original de *Madame Bovary* con su traducción al español, a fin de someter a experimentación un nuevo modelo de acercamiento a la equivalencia traductora.

A pesar de que, dada su novedad, este enfoque presenta aún limitaciones, puede constituir un primer paso para explorar nuevas vías de investigación de fenómenos como la asimilación, la imitación, la intertextualidad o el plagio en textos literarios, así como de la equivalencia en traducción.

Palabras clave: estilometría, análisis de sentimientos, lenguaje R, análisis computacional de textos literarios, afinidad estilística.

Abstract

Madame Bovary's supposed influence on *La Regenta* has been the subject of numerous critical studies although, since the beginning, it has been surrounded by controversy and debate. The traditionally adopted approach has been qualitative and based on partial, and not always objective, data. Furthermore, only merely anecdotal impressions have been sometimes the basis of different hypotheses and, consequently, the results obtained have been discordant. The main goal of this work is to provide quantitative data that allow to answer this still open question. To this end, a computational analysis of both the stylistic patterns and the emotional dimension, which underlie both novels, will be carried out by using the programming language R. In addition, the comparison between the original version of *Madame Bovary* and its translation into Spanish will also be addressed to test a new model for identifying equivalence in translation.

Despite its limitations due its novelty, this approach can be a first step to examine new ways for investigating phenomena such as assimilation, imitation, intertextuality or plagiarism in literary texts, as well as equivalence in translation.

Key words: stylometry, sentiment analysis, R language, literary texts' computational analysis, stylistic influence.

1. INTRODUCCIÓN

La supuesta influencia de *Madame Bovary* en *La Regenta* ha sido objeto de numerosos estudios críticos y también de diversas polémicas y críticas¹. Se ha hablado de asimilación, inspiración, homenaje, deuda, coincidencias, ecos, intertextualidad, imitación e incluso de plagio. Se han destacado las semejanzas en la trama, en la ambientación, en determinadas escenas y, en particular, en la caracterización de las dos protagonistas, Emma Rouault y Ana Ozores, y en sus sentimientos, emociones y deseos, tan distantes en ambos casos de la realidad provinciana y burguesa en la que viven. Los dos extremos de la cuestión quedaron plasmados en la encendida polémica que enfrentó en su tiempo a Luis Bonafoux con el mismo Clarín. Bonafoux acusó a Clarín de plagio, llegando a afirmar que *La Regenta* no era “más que una mala traducción de *Madame Bovary*” (Clarín, 1888: 11), acusación que Clarín rechazó decididamente, negando cualquier tipo de influencia directa de *Madame Bovary* en su novela:

Bonafoux asegura que cierta novela mía, titulada *La Regenta*, es plagio de *Madame Bovary*, y para ello se funda en que madame Bovary va una noche á un teatro con su marido y allí se encuentra con su amante, y no pasa en el teatro nada de particular; y en *La Regenta* también va la protagonista al teatro, y allí está un señor que la quiere decir que la adora, pero que todavía no se lo ha dicho. Tenemos como prueba de plagio, un teatro: teatro en *Madame Bovary*, teatro en *La Regenta*. Un marido: marido en *Madame Bovary*, marido en *La Regenta*; una esposa (id. ídv. id.); un amante en *Madame Bovary*, un pretendiente inconfeso en *La Regenta*. Ese es el plagio, esa es la mala traducción de la novela de Flaubert (Clarín, 1888: 23).

En todos los estudios críticos sobre los supuestos vínculos de filiación entre *Madame Bovary* y *La Regenta* se ha adoptado hasta ahora un enfoque cualitativo, necesariamente parcial, que ha arrojado resultados discordantes. El objetivo principal de este trabajo es aportar nuevos datos que puedan contribuir a dar respuesta a esta cuestión aún abierta, utilizando para ello instrumentos capaces de procesar grandes volúmenes de información. Con este fin, llevaremos a cabo un análisis computacional de ambas novelas mediante el lenguaje de programación R y su entorno asociado, que nos permitirá detectar los patrones estilísticos, la dimensión emotiva y la trayectoria de la narración subyacentes. Además de este objetivo primario, se abordará también paralelamente la comparación de la versión original de *Madame Bovary* con su traducción

¹ Hacer una revisión de los principales estudios y controversias críticas sobre la influencia de *Madame Bovary* en *La Regenta* queda fuera del ámbito de nuestro estudio. Para mayores detalles, el lector interesado puede ver Bonafoux (1888), Botrel (1968), Clarín (1888), Clavería (1942), Dupont (2007), Eoff (1965), Hatzfeld (1977), Lafitte (1943), Lissorgues (1987), Martínez Cachero (1953), Melón Ruiz de Gordejuela (1952), Richmond (1977), Sobejano (1981, 1985), Toledano García (1989) y Ventura Agudiez (1963).

al español, a fin de explorar las posibilidades de un nuevo modelo de determinación de la equivalencia traductora.

Dado el estado de desarrollo de estos recursos tecnológicos y la novedad del enfoque, ni los resultados ni las conclusiones de este trabajo pretenden ser concluyentes ni excluyentes. Consideramos, sin embargo, que el enfoque adoptado proporciona una nueva visión de los estudios literarios y constituye un primer paso hacia la superación de métodos tradicionales basados en datos fragmentarios e incluso, en ocasiones, en meras impresiones más o menos fundadas o anecdóticas, fundamentalmente por cuatro razones. La primera es que el análisis computacional de textos es un análisis lingüístico en profundidad de la materia del texto; la segunda es que este enfoque “distante” funciona a la vez como un telescopio y un microscopio y permite contemplar de cerca aspectos y detalles del texto que pasarían desapercibidos al ojo humano; la tercera es que las hipótesis formuladas a partir de los datos obtenidos son hipótesis falsables, que pueden confirmarse o refutarse, lo que “[...] provide us with some opportunity for the kind of falsification that Karl Popper and post-positivism in general offer as a compromise between strict positivism and strict relativism” (Jockers, 2014: x); la cuarta es que, cuando se combina con los métodos tradicionales, el enfoque distante tiene el potencial de desafiar los viejos supuestos y descubrir nuevos patrones (Underwood, 2017). A pesar de ello, somos conscientes de la necesidad de mejorar los instrumentos de extracción, análisis y visualización de datos de los que disponemos en la actualidad para el español, así como de validar los resultados obtenidos mediante análisis posteriores específicos.

A continuación, y antes de entrar en el análisis, expondremos, para mayor claridad, el marco teórico y metodológico en el que se encuadra nuestro trabajo y describiremos los principales instrumentos utilizados.

2. MARCO TEÓRICO Y METODOLÓGICO

El análisis computacional del texto literario puede encuadrarse dentro del marco teórico y metodológico de las humanidades digitales establecido en los estudios de teoría y crítica literaria de Moretti (2000a, 2000b, 2005, 2013a, 2013b, 2013c, 2016a, 2016b, 2017). La noción subyacente es la “lectura a distancia”², “lectura distante”, “lectura lejana” o “no lectura” frente a la “lectura cercana” (*distant reading / close reading*):

Distant reading: where distance, let me repeat it, *is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something (Moretti 2000a: 57-58).

² Para una visión histórica del concepto de lectura distante no asociado a las humanidades digitales, puede verse Underwood (2017).

Desde un punto de vista teórico, Moretti adopta sobre la lectura lejana una postura radical, limitada y fuertemente provocadora (el texto como objeto de estudio tiene que desaparecer), que ha sido objeto en los últimos años de fuertes críticas³. La mayor parte de los teóricos de este enfoque asumen, sin embargo, la necesidad de teorizar una especie de compromiso entre lectura cercana y lejana (Ascari, 2014; Bode, 2017; Hammond, 2017; Jin, 2017; Khadem, 2012; Serlen, 2010). El presente trabajo se encuadra en este marco mixto y considera que la diferencia fundamental entre ambos enfoques es la relevancia concedida en la lectura distante a las unidades “*appartenenti sia all’ “infinitamente grande” delle macro-categorie in cui il testo è inquadrato all’interno del mondo, sia all’ “infinitamente piccolo” di tutte le componenti formali di cui il testo è composto*” (Gerolamo Mossa, 2016: 4). Esta nueva escala nos permite acceder a datos sobre los textos a los que no es posible llegar (solo) mediante la lectura cercana. En los estudios de crítica literaria se ha utilizado tradicionalmente “*the middle of the scale: a text, a scene, a stanza, an episode, an excerpt ... An anthropocentric scale, where readers are truly “the measure of things”*”. But the digital humanities, Alan Liu has written, have changed these coordinates, by focusing “*on microlevel linguistic features [...] that map directly over macrolevel phenomena [Liu, 2012]*” (Algee-Hewitt *et al.*, 2015: 4).

Los datos obtenidos mediante la lectura lejana servirán para confirmar o refutar hipótesis generales previas al análisis cuantitativo y constituirán además la base para la construcción de interpretaciones críticas del hecho y del texto literarios a partir de un tipo de lectura cercana distinto del tradicional:

In it, the text does not so much disappear as reappear newly reconstituted through an encounter with forms of knowledge that are not saturated with the textual innards that conventional close reading would throw up, but with the details of a richly recovered contextual history (Joshi 2006: 496).

Según lo anterior, desde el punto de vista metodológico, la lectura distante se ocupará de “*fatti e fenomeni letterari, sia sincronici sia diacronici, che non sono accessibili ai tradizionali metodi di close reading [...] ma che richiedono l’analisi di massa di centinaia o migliaia di testi*” (Ciotti 2015: 13) con el objetivo de extraer patrones recurrentes comunes.

En nuestro trabajo aplicamos la lectura distante al texto literario con objeto de sacar a la luz parte de lo infinitamente grande e infinitamente pequeño que permanecía oculto en *Madame Bovary* y *La Regenta*. Este tipo de lectura implica no solo la extracción de grandes cantidades de datos de ambas novelas, sino también la determinación, a partir del análisis de los datos obtenidos, de los rasgos y patrones recurrentes de las dos obras y su visualización gráfica. Para la minería de datos y metadatos, su análisis y su

³ Para una revisión crítica del planteamiento distante de Moretti, puede consultarse Ciotti (2015), Gerolamo Mossa (2016), Goodwin y Holbo (2011), Khadem (2012), Kristal (2002), Ferguson (2008), Orsini (2002), Parla (2004) y Prendergast (2001, 2005).

representación gráfica, se utilizarán algoritmos escritos en lenguaje R, así como los paquetes y librerías de este entorno que se describen a continuación.

2.1. EL LENGUAJE R

El lenguaje R es un lenguaje de programación de alto nivel creado en 1993 como una implementación de código abierto del lenguaje de programación S (Becker y Chambers 1984; Ihaka y Gentleman 1996). Para escribir y ejecutar código en R, puede trabajarse, para mayor comodidad, en un entorno de desarrollo integrado como RStudio (RStudio Team 2017). El código se escribe y guarda en una ventana y se ejecuta en otra, mientras que el resultado se muestra en una tercera, como puede verse en la Figura 1.

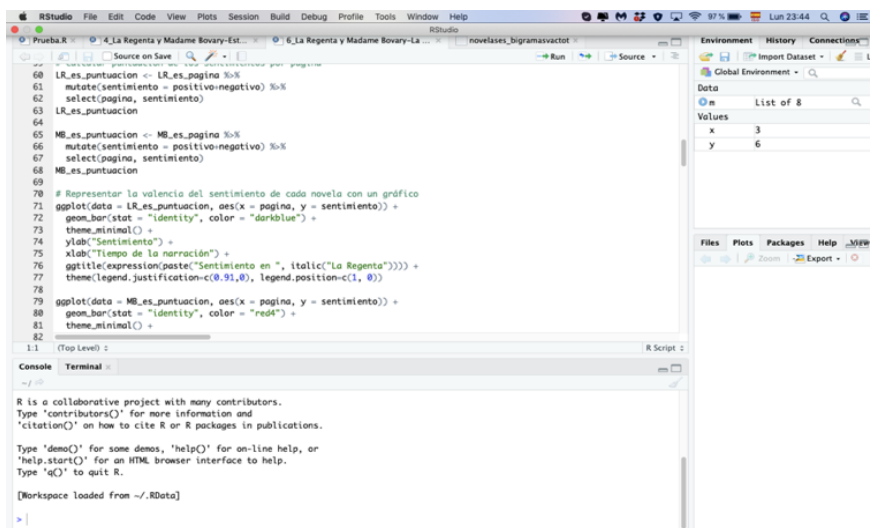


Figura 1. Entorno RStudio

R está orientado al análisis estadístico de datos masivos en todas sus fases (extracción, preparación, presentación, aplicación). Por esta razón, ha adquirido recientemente una notable popularidad en las humanidades digitales, ámbito en el que una de las tareas principales es el análisis de fuentes textuales, en particular el análisis de datos no estructurados en formatos múltiples o sin formato.

Este lenguaje forma parte de un proyecto abierto y colaborativo⁴ y cuenta con un amplio conjunto de utilidades para la visualización gráfica de alta calidad de los resultados, así como con una serie de paquetes adicionales que extienden la configuración básica. Estos paquetes son conjuntos de datos y funciones de código abierto con funciones específicas que los usuarios pueden publicar y descargar gratuitamente. De las herramientas basadas en R orientadas al análisis computacional de

⁴ El entorno R puede descargarse gratuitamente en <https://www.r-project.org/>.

textos literarios⁵, en nuestro trabajo hemos utilizado dos de las de mayor desarrollo⁶: el paquete *Stylo* para el análisis estilométrico y el paquete *Syuzhet* para el análisis de sentimientos. A continuación definimos brevemente estos dos campos de estudio, así como las herramientas asociadas.

2.2. LA ESTILOMETRÍA Y EL PAQUETE *STYLO*

El inicio de la estilometría, también denominada *análisis estilístico computacional* o *análisis computacional de estilo*, puede situarse a finales del siglo XIX con los estudios pioneros de Morgan, Mendenhall (cf. Holmes, 1998) y Lutoslawski (1897), que introduce el término, aunque no se constituirá como disciplina hasta la aparición de herramientas informáticas capaces de llevar a cabo el procesamiento de datos masivos.

Si bien la estilometría suele asociarse al concepto de atribución de autoría⁷, tiene un alcance más amplio. Puede definirse como una técnica de análisis lingüístico-estadístico de corpus textuales basada en la lectura distante, cuya finalidad consiste en determinar las relaciones existentes entre el estilo de los textos y los metadatos incluidos en ellos –en particular la identidad y sexo del autor, la datación y el género textual (Boto Bravo, 2017: 100)–, contribuyendo así a establecer patrones de similitudes y diferencias recurrentes que no pueden percibirse mediante la lectura cercana. De esta manera, la estilometría permite abordar no solo problemas de autoría, sino también, entre otras, cuestiones relativas a la datación de una obra, la ordenación cronológica de los textos, el género literario o, como veremos aquí, los vínculos estilísticos entre distintas obras o autores, ayudando a confirmar o refutar las interpretaciones cualitativas tradicionales (Eder *et al.*, 2015).

La estilometría se basa en la noción de *huella lingüística* o conjunto de rasgos lingüísticos que caracterizan el estilo de escritura de un autor o una obra y permiten identificarlo como un estilo individual y único que lo diferencia de otros estilos de otros autores u otras obras. A la hora de determinar esta marca de identidad estilística, los rasgos lingüísticos que tienen mayor peso son, contrariamente a lo que podría pensarse, los rasgos con un alto índice de frecuencia, en particular, las palabras funcionales o gramaticales frente a las palabras con significado léxico. Esto es debido a que las primeras no dependen del entorno (tema, género, época) y, sobre todo, a que se usan

⁵ Para mayores detalles, se aconseja la lectura de Jockers (2013) y Jockers (2014), dos libros pioneros que han inspirado la mayor parte de los sucesivos trabajos en este campo.

⁶ Además de las herramientas citadas, se han usado también las librerías gráficas (*ggplot*) para la visualización de los resultados y el ecosistema *tidy*, con los paquetes *tidyverse* y *tidytext*, para el manejo de datos textuales mediante tablas (*data frames* o *tibbles*) (Silge y Robinson, 2019).

⁷ Recientemente, la estilometría se ha aplicado a la atribución de autoría en campos como “criminal law, e.g., identifying the writers of harassing letters or ransom notes [...]; intelligence agencies work, e.g., linking intercepted messages to known terrorists or enemies [...]; civil law, e.g., solving estate disputes or copyright issues [...]; plagiarism detection, e.g., determining whether work submitted by a student was written by someone else [...]. Authorship identification has also become a major part of other identification technologies including intrusion detection systems, cryptography and signatures [...]” (Sarwar *et al.*, 2018: 324).

de manera involuntaria e inconsciente (Stamatatos, 2009: 540), configurando de esta manera, sin ningún tipo de filtros racionales, una clave estilística idiosincrásica:

The correlation between particular language features and stylistic idiosyncrasy is quite a delicate and uneasy balance between what is common and what is unique in the language. Extracting as many unique elements of style as possible is the goal of authorship attribution. These identified and extracted stylistic features are referred to as *style-markers*. Usually, the most desired style-markers are those undiscoverable with the naked eye and thus beyond authorial control; this is because the real “uniqueness” in style should be resistant to imitation, plagiarism and parody (Eder, 2011: 103).

Entre los marcadores de estilo utilizados en la mayor parte de los enfoques actuales, además del ya mencionado vector de las “palabras más frecuentes” (MFW, por su sigla en inglés), constituido por palabras funcionales, se encuentran también los signos de puntuación, la longitud de las palabras, la longitud de las oraciones, los patrones rítmicos sílaba tónica-átona, el número de sílabas por palabra, la frecuencia de determinadas secuencias de letras, la distribución de *n-grams*, la riqueza léxica, las colocaciones o la distribución de las partes de la oración (Baayen, 2001; Baayen *et al.* 2002; Grieve, 2007; Hirst y Feiguina, 2007; Holmes, 1998; Hoover, 2002; Juola, 2006).

El paquete *Stylo*⁸ de R (Eder y Rybicki, 2011; Eder *et al.*, 2013, 2016) es el programa de análisis estilométrico de mayor desarrollo en la actualidad. Centrado primordialmente en el análisis de estilo literario y dotado de una interfaz gráfica que facilita su uso (Figura 2), integra los procedimientos estilométricos con procedimientos estadísticos.

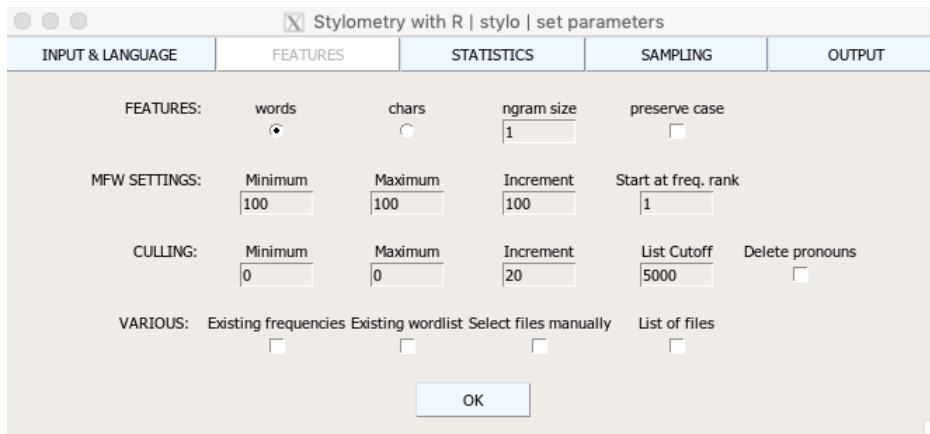


Figura 2. Interfaz de *Stylo*

Dado el objetivo del presente estudio, de entre las distintas funcionalidades que ofrece este paquete⁹, se ha seleccionado `stylo()`, que es una funcionalidad orientada

⁸ Para más detalles, puede verse el manual de referencia del paquete *Stylo* en <https://cran.r-project.org/web/packages/stylo/stylo.pdf>.

⁹ Para más detalles, véase Eder *et al.* (2019).

específicamente a la evaluación y representación gráfica de las similitudes estilísticas entre textos por medio de técnicas estadísticas. En nuestro estudio se ha utilizado el análisis de grupos, también denominado *análisis de conglomerados* o *agrupamiento* (*cluster analysis*), una técnica de análisis estadístico multivariante que calcula, a partir de las MFW y, si procede, otros marcadores de estilo, la distancia entre los distintos archivos textuales de un corpus agrupándolos de manera homogénea según dicho grado de distancia. Cuanto menor sea la distancia entre los archivos, mayor será la proximidad estilística, mientras que a medida que aumente la distancia, disminuirá la afinidad de estilo. La funcionalidad `style()` genera automáticamente representaciones gráficas de redes nodales a partir de la distancia entre los textos. En nuestro caso, se ha seleccionado la visualización mediante dendrogramas, un tipo de diagramas de árbol en los que los archivos del corpus aparecen agrupados según su grado de cercanía estilística. Para completar la representación, se han usado también los árboles de consenso, que son el resultado gráfico de trazar y comparar varios dendrogramas subyacentes (§ 3.2).

Según lo anterior, la hipótesis de influencia de *Madame Bovary* en *La Regenta* desde el punto de vista estilístico formal quedará confirmada o refutada en función de la distancia que medie entre ambas obras, tal y como reflejarán los relativos dendrogramas y árboles de consenso.

2.3. EL ANÁLISIS DE SENTIMIENTOS, LOS LEXICONES AFECTIVOS Y EL PAQUETE SYUZHET

El análisis de sentimientos (*sentiment analysis*), también conocido como minería de opinión (*opinion mining*), es un tipo de análisis semántico computacional que se ocupa de la detección, extracción, clasificación, análisis y representación gráfica de la dimensión subjetiva asociada a los temas o tópicos presentes en un corpus de textos¹⁰ a distintos niveles (documento, párrafo, oración, palabra), entendiéndose por dimensión subjetiva todo tipo de sentimientos, emociones, estados afectivos y anímicos, creencias, percepciones, puntos de vista, actitudes, opiniones, juicios y valoraciones. El análisis de sentimientos asigna a esta dimensión subjetiva una polaridad, que puede ser positiva, negativa o neutra (Mohammad *et al.*, 2009; Turney y Littman, 2003). En algunos casos, como en nuestro estudio, el análisis se refina hasta llegar a detectar las emociones principales (Alm y Sproat, 2005; Alm *et al.*, 2005; Mohammad, 2012; Mohammad y Turney, 2010).

El análisis de sentimientos registra a partir del año 2000 un notable desarrollo, que corre paralelo al crecimiento exponencial de textos en formato digital como reseñas, chats, foros, blogs o microblogs en Internet y las redes sociales, donde predominan la expresión de opiniones y puntos de vista personales (Liu, 2015). Este tipo de análisis encontró su primera aplicación inmediata en la investigación de mercado y se extendió después a numerosos ámbitos, en particular a aquellos en los que predomina la

¹⁰ Si bien el análisis de sentimientos se aplicó en un principio exclusivamente a textos escritos, recientemente se ha utilizado también con productos multimedia y productos multimodales. Al respecto puede verse Poria *et al.* (2018).

comunicación persuasiva. En la actualidad, prácticamente en todas las esferas e instancias, desde las empresariales, comerciales, financieras y publicitarias hasta las políticas, sociales y gubernamentales, se llevan a cabo actividades de minería de las opiniones expresadas en las redes sociales, lo que plantea, en muchos casos, cuestiones éticas de relieve. Conocer estas opiniones concede una ventaja comparativa respecto a los competidores, ya que puede ayudar a mejorar los productos y servicios y a construir campañas y estrategias que se ajusten en mayor medida a las necesidades y expectativas de los usuarios o clientes, pero abre también las puertas a un peligroso control ideológico y de comportamientos. Recientemente, el análisis de sentimientos ha empezado a utilizarse con fines predictivos en campos tan variados como, por ejemplo, el campo financiero, las elecciones políticas o el cine y la literatura, en donde ha llegado a pronosticar con un elevado índice de fiabilidad cuáles son las empresas (Bollen *et al.*, 2011; Zhang y Skiena, 2010; Zhang *et al.*, 2010), los políticos (O'Connor *et al.*, 2010; Tumasjan *et al.*, 2010), las películas (Mishne y Glance, 2006; Sadikov *et al.*, 2009) o las obras literarias (Archer y Jockers, 2016) que obtendrán un mayor rendimiento, mejores resultados o más votos o ventas.

Además de sus numerosas aplicaciones prácticas en la vida real¹¹, el análisis de sentimientos es también un campo de investigación abierto y en rápida expansión que despierta un fuerte interés entre los estudiosos de distintas disciplinas que trabajan con datos textuales, si bien, dada su novedad, no está exento de problemas¹². La literatura no se ha quedado al margen. Los textos literarios son un canal privilegiado para la expresión de emociones y sentimientos, de ahí que en distintos trabajos de investigación se haya empezado a aplicar este tipo de análisis computacional para examinar las tendencias en la expresión y distribución emotiva en distintos géneros literarios, autores o épocas (Alm y Sproat, 2005; Alm *et al.*, 2005; Jockers y Mimno, 2013; Kakkonen y Kakkonen, 2011; Kim *et al.*, 2017; Klinger *et al.*, 2016; Mohammad, 2011; Wilhelm *et al.*, 2013). Estas tendencias centrales se han utilizado a su vez para trazar y visualizar lo que se conoce como *forma de la historia* o *arco emotivo de la narración*, es decir, el flujo o trayectoria de los sentimientos de la obra literaria que subyace al flujo o trayectoria de la trama (Archer y Jockers, 2016; Booker, 2004; Jockers, 2015; Reagan *et al.*, 2016).

Con el objetivo de detectar posibles vínculos de filiación de *La Regenta* con *Madame Bovary*, en nuestro estudio hemos llevado a cabo, en primer lugar, un análisis comparativo de los sentimientos y emociones presentes en ambas obras y, en segundo lugar, hemos trazado y comparado la forma de la historia de las dos novelas. A continuación presentamos brevemente estas dos líneas de investigación y los recursos utilizados, es decir, los lexicones para el análisis de sentimientos y el paquete *Syuzhet* para la forma de la historia.

¹¹ Para el análisis de sentimientos en español, puede verse Henríquez Miranda y Guzmán (2016).

¹² Para mayores detalles, puede consultarse Liu (2015) y Pang y Lee (2008).

2.3.1. ANÁLISIS DE SENTIMIENTOS Y LEXICONES AFECTIVOS

Para el análisis de sentimientos de *Madame Bovary* y *La Regenta* hemos utilizado el enfoque semántico¹³ basado en lexicones afectivos, también denominados *diccionarios de sentimientos* o simplemente *lexicones*. Una de las principales limitaciones de este trabajo es, sin duda, la carencia de lexicones diseñados de manera específica para lenguas distintas del inglés (en nuestro caso, el español y el francés) o que permitan llevar a cabo comparaciones interlingüísticas (Balahur y Turchi, 2012; Schmidt y Burghardt, 2018), así como de lexicones que contemplen la dimensión diacrónica (Buechel *et al.*, 2016). Habida cuenta de la necesidad de una mayor investigación sobre estos aspectos, los resultados no han de considerarse concluyentes, sino meramente ejemplificativos de la aplicación de un nuevo modelo de estudio. No obstante, a pesar de las limitaciones observadas, el margen de fiabilidad de este enfoque es elevado cuando no se aplica solo a una oración, sino a textos amplios que contienen una variedad de expresiones emotivas, como en el caso del presente estudio (Mohammad, 2011: 107).

En este trabajo se ha utilizado el lexicon NRC de Saif M. Mohammad¹⁴ (Mohammad *et al.*, 2009; Mohammad y Turney, 2010; Mohammad y Yang, 2011) en las versiones para el español y francés incluidas en el paquete *Syuzhet*, que describimos brevemente en la siguiente sección, así como en la versión de Fradejas Rueda¹⁵ para el español. El lexicon NRC contiene una lista de palabras asociadas con ocho emociones básicas (*anger, anticipation, disgust, fear, joy, sadness, surprise, trust*) y dos extremos de polaridad (*positivo y negativo*).

Como veremos (§ 3.3), el análisis semántico computacional mediante lexicones permitirá no solo detectar los temas o núcleos significativos principales de las dos novelas, sino también cuantificar su carga afectiva y ponerlos en relación con las emociones básicas principales. Este aspecto es especialmente relevante a la hora de analizar comparativamente textos literarios, ya que elementos de la “escala media” observables mediante una lectura cercana, como una escena, una ambientación o un episodio, pueden ser comunes a distintos textos, pero tener en cada uno de ellos una polaridad o una valencia emocional distinta y, por tanto, un peso narrativo diferente. Según lo anterior, hipotetizamos que para hablar de influencia de *Madame Bovary* en *La Regenta* no es suficiente la existencia de núcleos significativos o elementos de la escala media comunes a ambas obras: es necesario, además, que estos puntos comunes tengan el mismo peso y la misma polaridad o valencia.

¹³ Existen otros métodos de análisis de sentimientos como los basados en el aprendizaje automático (*machine learning*) (Pang *et al.*, 2002) o en el aprendizaje profundo (*deep learning*), en particular en redes neuronales (Socher *et al.*, 2010, 2013; Stuart y Majewski, 2015), así como métodos híbridos, pero no resultan aplicables a nuestro caso. Para un análisis comparativo de varios métodos de análisis de sentimientos, véase Gonçalves *et al.* (2013).

¹⁴ Puede accederse al lexicon NRC original en inglés y a sus versiones en distintas lenguas desde el enlace: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.

¹⁵ Este lexicon forma parte de una base de datos facilitada en el marco del curso de *Análisis de textos y estilometría con R* de la UNED por el profesor Fradejas Rueda, a quien agradecemos desde aquí su aportación.

2.3.2. EL PAQUETE SYUZHET

Para la comparación de la forma de la historia de *Madame Bovary* y *La Regenta* se ha utilizado el paquete *Syuzhet* del entorno R, uno de los sistemas de mayor desarrollo en la actualidad para el análisis de obras literarias en prosa (Jockers, 2013, 2015). La fundamentación teórica del sistema es una idea de Propp:

The fundamental idea, of course, was nothing new. What I was after is what the Russian formalist Vladimir Propp had defined as the narrative's *syuzhet* (the organization of the narrative) as opposed to its *fabula* (raw elements of the story). *Syuzhet* is concerned with the linear progression of narrative from beginning (first page) to the end (last page), whereas *fabula* is concerned with the specific events of a story, events which may or may not be related in chronological order (Jockers, 2015: 2-3).

Además de la distinción de Propp entre *syuzhet* y *fabula*, hay otras tres ideas que constituyen la base teórica, metodológica y aplicativa del sistema de Jockers. La primera es que las fluctuaciones de los sentimientos y emociones a lo largo de una obra son un indicador altamente fiable de los movimientos de la trama (Jockers, 2015: 2). La segunda es que la trama y la forma de la historia de las obras literarias se ajustan a un número limitado de patrones arquetípicos básicos (tres según Foster-Harris, 1959; seis según Reagan *et al.*, 2016; siete según Booker, 2004; veinte según Tobias, 1993; treinta y seis según Polti, 1921). La tercera es que, como afirmó Vonnegut¹⁶ en su conferencia sobre la forma de las historias, “[...] there is no reason why the simple shapes of stories can't be fed into computers, they are beautiful shapes”. Estos tres supuestos implican que los sentimientos, la trama y la forma de la historia están interrelacionados y pueden analizarse computacionalmente. El paquete *Syuzhet* obedece a este fin.

Según lo anterior, consideramos que para hablar de influencia de *Madame Bovary* en *La Regenta* no es suficiente que haya semejanzas puntuales en la trama, ya que las tramas de obras del mismo género se ajustan a modelos arquetípicos. Es necesario, además, que exista una correspondencia entre la forma de la historia de *Madame Bovary* y la de *La Regenta* y, en consecuencia, entre las fluctuaciones de los sentimientos y emociones a lo largo de la narración en ambas obras.

2.4. EL PROCEDIMIENTO DE ANÁLISIS Y LA EQUIVALENCIA TRADUCTORA

Un objetivo secundario de este estudio es proponer un acercamiento experimental a la equivalencia traductora a través de la estilometría y el análisis de sentimientos. Este objetivo deriva del procedimiento seguido para analizar *Madame Bovary* y *La Regenta*. Para comparar ambos textos, el primer problema que surge es que están escritos en dos lenguas diferentes, mientras que los recursos de los paquetes *Stylo* y *Syuzhet* no pueden aplicarse a corpus multilingües debido a su dependencia de cada idioma específico. De entre los enfoques que se han propuesto para realizar análisis de corpus multilingües (Sarwar *et al.*, 2018), en este estudio se aplica el consistente en traducir todos los textos

¹⁶ Kurt Vonnegut, *Shapes of stories*: <https://www.youtube.com/watch?v=oP3c1h8v2ZQ>.

a una única lengua y llevar a cabo después un análisis comparado monolingüe (Bogdanova y Lazaridou, 2014). En consecuencia, *La Regenta* se comparará con la traducción al español de *Madame Bovary*¹⁷ y no con el original francés.

El límite principal de este enfoque es que la traducción puede contener elementos estilísticos propios que enmascaren el estilo del texto original (Sarwar *et al.*, 2018). Por esta razón, se propone como medida de corrección analizar la equivalencia entre el texto de origen (TO) y el texto meta (TM), es decir entre el original francés de *Madame Bovary* y su traducción al español. Como es sabido, la descripción de una traducción y la evaluación de su calidad se basan en el análisis comparativo del TO y el TM para determinar la equivalencia entre ambos. Esta equivalencia ha de ser primordialmente una equivalencia *semántica*, pero en las obras literarias, y en general en todos los textos con una función estética, tienen también una importancia fundamental la equivalencia *estilística* y la equivalencia que podríamos denominar *emotiva*.

La equivalencia estilística, ampliamente estudiada desde distintas perspectivas teóricas, se basa en la correspondencia de los rasgos de estilo del TO y el TM. A este respecto la aportación novedosa de este estudio consiste en utilizar la estilometría para analizarla. Se considera que la equivalencia estilística entre el TO y el TM será mayor cuanto menor sea la diferencia entre las agrupaciones de los capítulos y las distancias que medien entre ellos, tal y como aparecen representadas en los dendrogramas¹⁸.

Por su parte, la equivalencia emotiva consistiría en la correspondencia entre la forma de la historia del TO y el TM y, en consecuencia, entre las valencias expresivo-afectivas, positivas y negativas, de ambas obras y sus fluctuaciones a lo largo de la narración. Este tipo de equivalencia tiene, por tanto, como raíz el estilo de la narración, es decir, la manera en la que el autor presenta los acontecimientos a los lectores (*synzhet*). Se considerará que la equivalencia emotiva entre el TO y el TM es mayor cuanto menor sea la diferencia entre la forma de sus historias, tal y como aparecen representadas en el sistema de coordenadas.

3. ANÁLISIS DE MADAME BOVARY Y LA REGENTA

3.1. EL CORPUS: PREPARACIÓN Y OPERACIONES PRELIMINARES

Dados los objetivos del presente trabajo, y visto que no es posible comparar textos en distintas lenguas (§ 2.4), el primer paso ha sido constituir un corpus inicial con tres obras: *La Regenta* (LR), *Madame Bovary* en la versión original francesa (MB_fr) y *Madame Bovary* en la versión traducida al español (MB_es). Para poder aplicar las

¹⁷ Sería interesante completar el análisis que aquí proponemos con un análisis especular en el que se comparara *Madame Bovary* con la traducción al francés de *La Regenta*.

¹⁸ Cabe pensar que, dado que los lexicones utilizados para el español y el francés no son idénticos, obtener dendrogramas iguales o con un elevado grado de similitud sea prácticamente imposible. Sin embargo, esta es una cuestión sobre la que sería necesario realizar una labor de investigación para poder definir con precisión las nociones de “diferencia” y “distancia”.

herramientas R, es necesario que los archivos estén en texto plano (*.txt). LR¹⁹ y MB_fr²⁰ se han descargado directamente en formato .txt, mientras que MB_es²¹ se ha descargado en formato EPUB y se ha convertido después en texto plano mediante el programa Calibre²². Este último archivo es el que presenta un mayor grado de suciedad textual.

Para preparar el corpus para su procesamiento, se han editado los textos y se han efectuado una serie de operaciones de limpieza que, en términos generales, han consistido en: i) eliminación de metadatos y paratextos que pudieran interferir en el análisis; ii) cambio de mayúsculas por minúsculas; iii) remoción de líneas y espacios en blanco innecesarios.

Además, se han llevado a cabo una serie de transformaciones para facilitar las operaciones de segmentación previas al análisis estilométrico. En particular, se ha decidido uniformar en las tres obras la designación y numeración de los capítulos. *Madame Bovary* se divide en tres partes: la primera consta de 9 capítulos, la segunda de 15 y la tercera de 11. En cada una de ellas, la numeración de los capítulos comienza desde 1. En el archivo de MB_fr cada capítulo se abre simplemente con un número romano, mientras que en MB_es los números romanos van precedidos de la palabra “CAPÍTULO”, en versales. Por su parte, *La Regenta* consta de dos tomos y treinta capítulos. En el archivo LR los capítulos están numerados correlativamente en romanos precedidos de la palabra “Capítulo”. En el corpus listo para procesar, los capítulos de las tres obras han numerado siempre correlativamente utilizando numeración arábiga a doble cifra (01, 02, 03... 10, 20, 30) precedida por la palabra “Capítulo”. De esta manera, en los archivos MB_fr y MB_es listos para procesar la primera parte de la obra comprende los capítulos 01-09, la segunda parte, los capítulos 10-24 y la tercera parte, los capítulos 25-35. En LR, el Tomo I comprende los capítulos 01-15 y el Tomo 2, los capítulos 16-30. Cada uno de los capítulos de las obras constituye un archivo del corpus.

Una vez preparado el corpus textual para su procesamiento, y tras las operaciones de carga de ficheros, librerías, paquetes y herramientas de R, se pasó a la fase previa al análisis consistente en la segmentación de los textos y la obtención de estadísticas preliminares. Con un *script* de R se procesaron los cien ficheros del corpus, se segmentaron los textos en *palabras-token* y en *palabras-tipo* y se procedió al recuento de las palabras más frecuentes (con y sin palabras vacías) y a la comparación de estos datos preliminares (frecuencia absoluta y frecuencia relativa). Por su representatividad, se incluyen a continuación los gráficos de las palabras léxicas con mayor índice de frecuencia absoluta y relativa (gráficos 1 y 2) obtenidos después de haber aplicado como filtro un catálogo de palabras vacías generales (765 palabras funcionales) y otro de

¹⁹ LR se ha descargado del siguiente enlace facilitado por el profesor Fradejas Rueda en el marco del curso sobre *Análisis de textos y estilometría con R* de la UNED:

<<https://raw.githubusercontent.com/7PartidasDigital/AnaText/master/datos/textos/LaRegenta.txt>>.

²⁰ MB_fr se ha descargado del sitio del proyecto *Gutenberg*: <<https://www.gutenberg.org/>>.

²¹ MB_es se ha descargado de la biblioteca virtual gratuita *One More Library*:

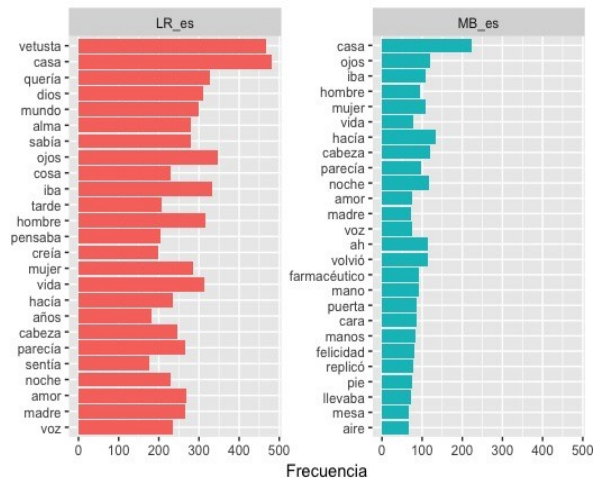
<<https://onemorelibrary.com/index.php/es/>>.

²² Véase, <https://calibre-ebook.com>.

palabras vacías específicas²³ de cada obra (nombres propios, palabras sin peso semántico y palabras-error), con el fin de analizar cómo se configuran en una primera aproximación los campos semánticos de ambas novelas.

No obstante se trate de resultados preliminares con un valor meramente indicativo, se perfilan dos mundos en cierta medida paralelos, pero también distantes. En *La Regenta*, la ciudad, Vetusta, parece tener un papel protagonista, mientras que el entorno exterior no está presente en esta escala en *Madame Bovary*. La dimensión religiosa o espiritual (*dios, alma*), muy marcada en *La Regenta*, está ausente en *Madame Bovary*, mientras que la *felicidad*, tan presente en *Madame Bovary*, no es uno de los nodos centrales de *La Regenta*. En ambas obras hay menciones a partes del cuerpo (fundamentalmente de la parte superior, *ojos, cabeza, cara; mano(s)* y *pie* aparecen solo en *Madame Bovary*). La dimensión temporal está marcada en ambas obras (*tarde, día, noche, años*) y los verbos, siempre en pretérito imperfecto en *La Regenta*, sugieren un tiempo suspendido, que en *Madame Bovary* se rompe solo esporádicamente (*volvió, replicó*). Veremos cuál es la valencia emotiva de algunos de estos núcleos significativos al tratar del análisis de sentimientos (§ 3.3).

Gráfico 1. Palabras léxicas más frecuentes en LR y MB_es
(frecuencia absoluta)

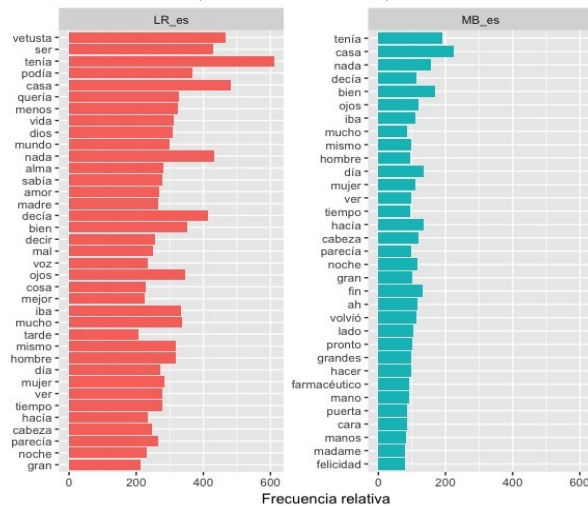


²³ El catálogo de palabras vacías adhoc es el siguiente:

```
vacias_adhoc <- tibble(palabra = c("don", "ana", "magistral", "marqués",
"álvaro", "victor", "regenta", "señor", "doña", "mesía", "fermín",
"quintanar", "petra", "señora", "frígilis", "obdulia", "anita", "paula",
"paco", "marquesa", "provisor", "pompeyo", "bermúdez", "glocester",
"guimarán", "petronila", "madame", "monsieur", "emma", "carlos", "bovary",
"león", "homais", "rodolfo", "ho mais", "lheureux", "yonville", "justino",
"rouault", "binet", "hipólito", "bournisien", "11 ", "í", "allí", "ho
mais"))
```

A pesar de ser una palabra de las que hemos denominado 'propias', hemos decidido no incluir *Vetusta* en el catálogo debido al valor simbólico y metafórico que, según algunos estudiosos, tiene la ciudad en la novela.

Gráfico 2. Palabras léxicas más frecuentes en LR y MB_es
(frecuencia relativa)



3.2. ANÁLISIS ESTILOMÉTRICO

El análisis estilométrico²⁴ con el paquete *Stylo* nos permitirá:

1) Comparar el original francés de *Madame Bovary* (TO) con su traducción al español (TM) (gráficos 3 y 4) para analizar si la agrupación por capítulos del TO es congruente con la del TM y comprobar, por tanto, si puede hablarse o no de equivalencia estilística.

2) Comparar MB_es y LR (gráficos 5 y 6). Las hipótesis son:

H1. Si el sistema no agrupa los conglomerados de MB_es (1) y de LR (2) por separado, es decir, si no media ninguna distancia entre (1) y (2), no puede descartarse la hipótesis de vinculación estilística entre ambas novelas (plagio, influencia, imitación o afinidad estilística).

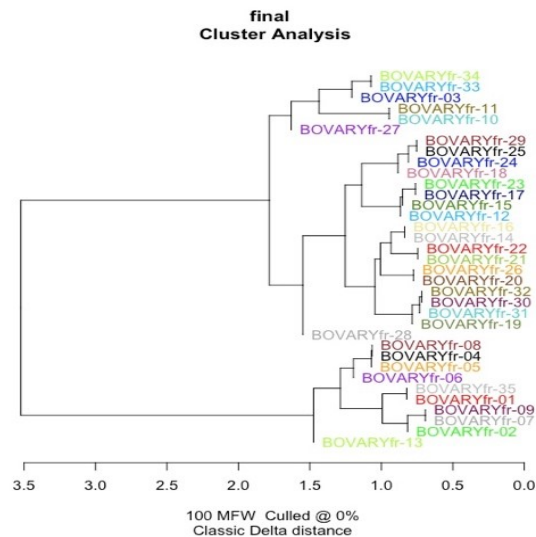
H2. Si el sistema agrupa los conglomerados de MB_es (1) y los de LR (2) por separado, pero la distancia entre (1) y (2) es menor o igual que la distancia interna existente entre los archivos de (1) entre sí o entre los archivos de (2) entre sí, no puede descartarse la hipótesis de vinculación estilística entre ambas novelas (plagio, influencia, imitación o afinidad estilística).

H3. Si el sistema agrupa los conglomerados de MB_es (1) y los de LR (2) por separado y la distancia entre (1) y (2) es mayor que la distancia interna existente entre los archivos de (1) entre sí o entre los archivos de (2) entre sí, puede descartarse la hipótesis de vinculación estilística entre ambas novelas (plagio, influencia, imitación o afinidad estilística).

²⁴ Todos los análisis se han realizado con R mediante desarrollo de códigos propios.

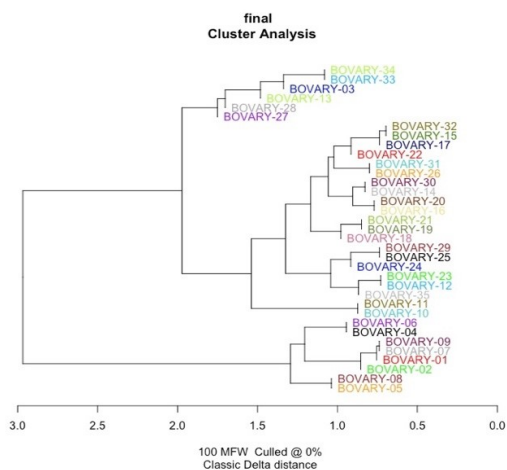
No hay acuerdo sobre cuáles son los valores mínimos y máximos de MFW y *n*-grams y la medida estadística de `stylo` () que deben seleccionarse para que los resultados obtenidos tengan el mayor índice de fiabilidad posible²⁵. En consecuencia, hemos realizado análisis de agrupaciones aplicando distintos parámetros –en particular, número de MFW entre 100 y 5000, *1-gram*, *2-grams* y *3-grams*, distancia Euclídea y distancia Delta de Burrows (Classic)–, sin que se haya podido apreciar una variación significativa en los datos necesarios para confirmar o refutar las hipótesis propuestas. Para ilustrar los resultados principales, a continuación figuran los dendrogramas y árboles de consenso obtenidos aplicando los parámetros 100 MFW y 100-1000 MFW, respectivamente, y la medida Delta clásica (Argamon, 2008; Burrows, 2002; Fradejas Rueda, 2016).

Gráfico 3. Análisis de agrupaciones: capítulos MB_fr



²⁵ Al respecto puede consultarse Boto Bravo (2017), Burrows (2002), Craig y Kinney (2009), Eder (2011), Fradejas Rueda (2016), Smith y Aldridge (2011).

Gráfico 4. Análisis de agrupaciones: capítulos MB_es



Si comparamos los gráficos 3 y 4 anteriores, observamos que la distribución de los conglomerados, así como la distancia entre ellos es, en términos generales, congruente excepto en el caso de: i) algunos capítulos sueltos que se agrupan de manera diversa en MB_fr y MB_es (en particular, los capítulos 10, 11, 13, 28); ii) la distancia entre el conglomerado de la parte superior y el resto de conglomerados (capítulos 3, 27, 33, 34, además de 10, 11 en MB_fr y 13, 28 en MB_es), la cual es decididamente mayor en MB_es que en MB_fr. Tomando como base estos resultados, puede afirmarse que entre el TO y el TM hay equivalencia estilística, si bien sería necesario investigar más a fondo la disimetría de agrupación y/o distancia que se observa en los capítulos 3, 10, 11, 13, 28, 33 y 34. En particular, sería interesante aplicar a MB_es las funcionalidades `rolling.delta ()` y `rolling.classify ()` del paquete *Stylo*, que permiten analizar las idiosincrasias estilísticas internas de un texto, a fin de comprobar si la diferencia de estilo de los capítulos del conglomerado superior puede imputarse a que son fruto de un segundo traductor, distinto del que ha traducido los restantes capítulos (Eder, 2016).

Si observamos ahora los gráficos 5 y 6, vemos que tanto el dendrograma como el árbol de consenso confirman plenamente la hipótesis H3, según la cual puede descartarse cualquier tipo de vinculación estilística (plagio, influencia, imitación o afinidad) entre *Madame Bovary* en *La Regenta*, ya que la agrupación por separado de los respectivos conglomerados y la marcada distancia que media entre las novelas revela el uso de patrones estilísticos claramente diferenciados. En este punto la estilometría da plenamente razón a Clarín (§ 1).

Gráfico 5. Comparación estilo: capítulos LR y MB_es

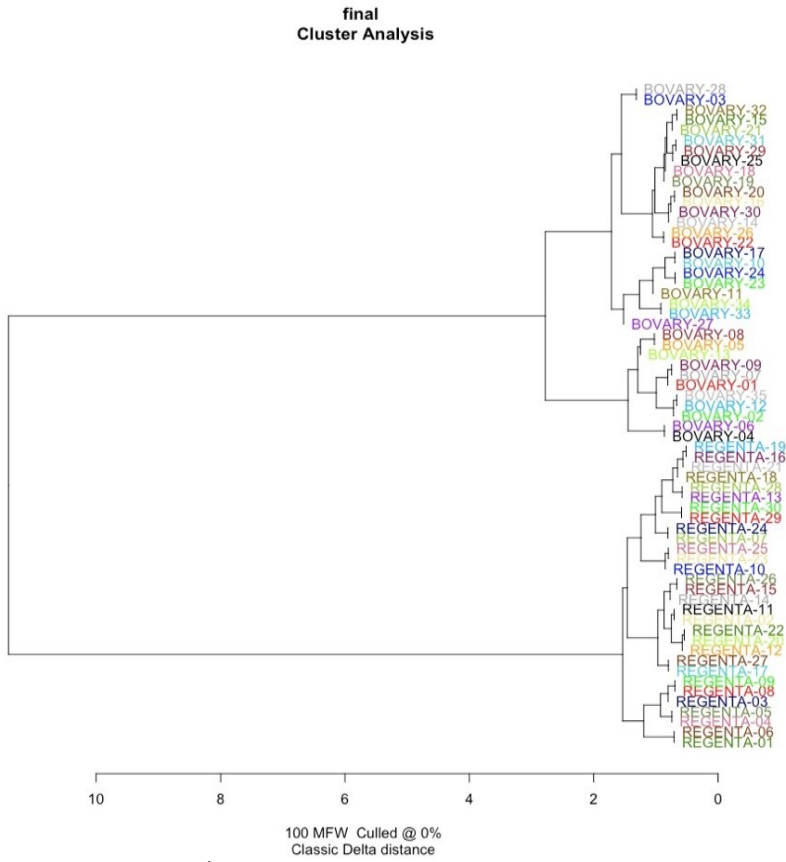
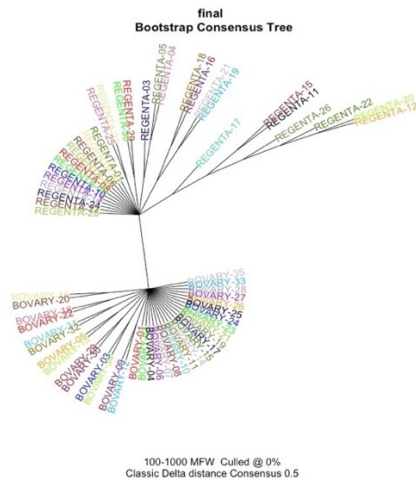


Gráfico 6. Árbol de consenso comparativo de LR y MB_es



3.3. ANÁLISIS DE SENTIMIENTOS

Como hemos mencionado antes, para analizar la dimensión temporal y afectiva de *Madame Bovary* y *La Regenta*, uno de los ámbitos en los que los críticos han señalado la presencia de elementos comunes, se ha aplicado un enfoque basado en lexicones afectivos (§ 2.3) con objeto de determinar los temas principales de ambas obras, cuantificar su polaridad o valencia y asociarlos con las emociones básicas principales. Este es un aspecto particularmente significativo ya que para poder hablar de influencia de *Madame Bovary* en *La Regenta* no solo es necesaria la presencia de elementos comunes a las dos novelas, sino también que estos elementos tengan la misma polaridad y peso emotivo y, por tanto, el mismo valor narrativo.

Los resultados del análisis semántico computacional se han representado en los gráficos 7 a 12, que figuran a continuación. En el gráfico 7 pueden observarse las palabras asociadas a polaridad positiva o negativa en ambas novelas y su peso relativo. En el gráfico 8 se completa esta visión positiva/negativa con las ocho emociones básicas del lexicon NRC. Esta información se representa en forma de nubes de palabras en los gráficos 9 a 12. En la interpretación de los gráficos hay que tener en cuenta lo siguiente: i) en los gráficos 7 y 8, las emociones con polaridad positiva se sitúan a la derecha de la línea central y las negativas a la izquierda, mientras que en los gráficos 9 y 10 la parte positiva se encuentra situada en la mitad inferior y la negativa en la superior, ii) en las nubes de palabras, el tamaño de las palabras refleja su peso relativo en la novela; ii) se ha respetado la denominación de las emociones básicas contenidas en la versión española del lexicon NRC de Fradejas Rueda en todos los gráficos correspondientes, aun considerando que no es siempre la más adecuada. Las ocho emociones básicas que abarca este diccionario en su versión original en inglés (Mohammad *et al.*, 2009; Mohammad y Turney, 2010; Mohammad y Yang, 2011) son *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, *trust*, que aparecen traducidas al español como *ira*, *premonición*, *disgusto*, *miedo*, *alegría*, *tristeza*, *asombro* y *confianza*, si bien la traducción de *disgust* como *disgusto* en lugar de como *repugnancia* o *asco* es claramente un calco y *anticipation* debería traducirse en este contexto como *expectación* y no como *premonición*. A lo largo del texto, sin embargo, se utiliza *repugnancia* y *expectación*.

Gráfico 7. Sentimientos positivos y negativos en *La Regenta* y *Madame Bovary* (MB_es)

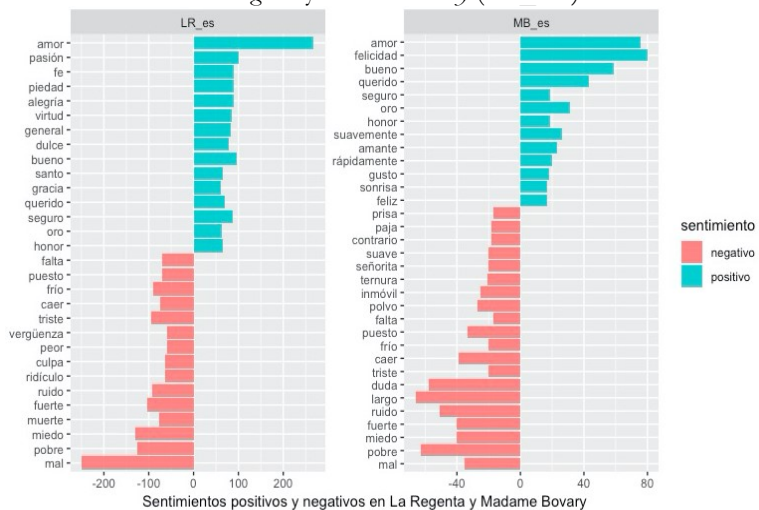


Gráfico 8. Sentimientos positivos y negativos y emociones básicas en *La Regenta* y *Madame Bovary* (MB_es)

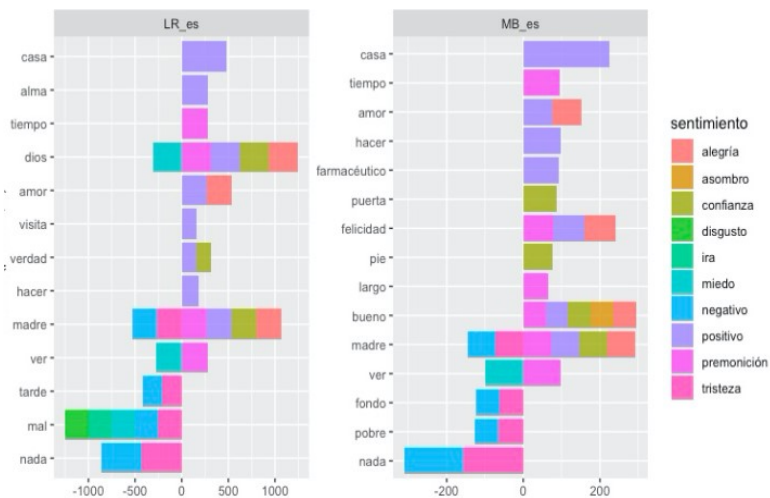


Gráfico 9. Valencia positiva y negativa en *La Regenta*

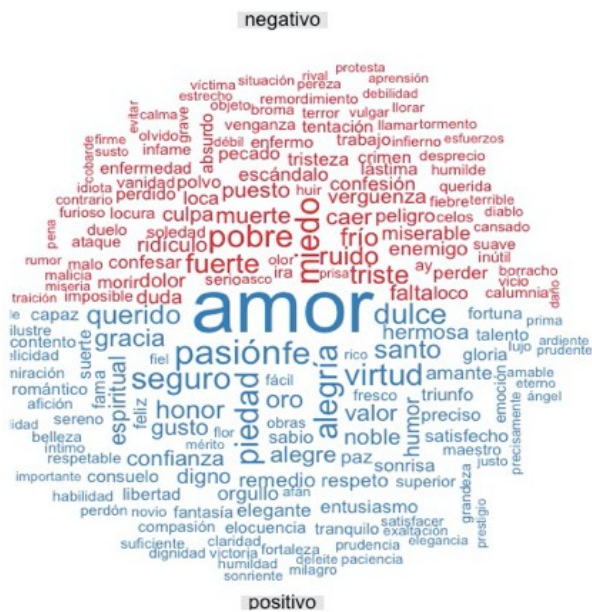


Gráfico 10. Valencia positiva y negativa en *Madame Bovary* (MB_e_s)



En términos generales, coinciden las dos obras en el fuerte peso positivo del *amor*, eje central de ambas como queda reflejado en todos los gráficos, o de la *casa* (gráficos 8, 11 y 12), así como en el fuerte peso negativo del *miedo*, *la tristeza*, *el frío* (todos los gráficos), *la nada* (gráfico 8) –ligada en los dos casos a la tristeza– y *el mal* (gráfico 7), si bien en torno a este último nodo hay una diferencia significativa: la visión del mal en *La Regenta* tiene mayor peso y complejidad emotiva que en *Madame Bovary*, articulándose en torno a la tristeza, el miedo, la ira y la repugnancia (gráfico 8). El análisis de sentimientos confirma también la idea común del *tiempo* como tiempo de espera que se perfilaba ya en el uso predominante de formas verbales en pretérito imperfecto (§ 3.1). Es un tiempo suspendido, premonitorio, ligado a la expectación (gráfico 8). El nodo semántico *madre* es un nodo complejo en las dos novelas (gráfico 8). Evoca sentimientos positivos y negativos, conjugando en sí alegría, confianza, expectación y tristeza en ambas obras. Sería interesante investigar con mayor profundidad si cada una de estas diferentes valencias emotivas está asociada a la figura de la madre o a la maternidad. Merecerían también mayor investigación los nodos *fuerte*, *pobre* y *ruido*, que tienen una polaridad negativa con un peso importante en ambas obras, pero son difícilmente interpretables solo a partir de estos datos.

Si se analizan ahora las divergencias, se observa que una de las principales diferencias temáticas, como ya habían puesto de relieve las palabras más frecuentes (§ 3.1), es la dimensión religiosa. Todos los nodos asociados a esta dimensión tienen en la obra de Clarín un peso importante, mientras que en *Madame Bovary* en su mayor parte ni siquiera están presentes. En *La Regenta*, el nodo principal de la dimensión religiosa es *dios*, un nodo complejo, asociado con emociones contrastantes como la alegría, la confianza y la expectación, pero también en muy amplia medida con el miedo (gráficos 8 y 11). Los nodos *fe*, *piedad*, *virtud*, *santo* (gráficos 7 y 9), *alma* y *espiritual / espíritu* (gráficos 8 y 11) tienen todos ellos un peso significativo y una marcada valencia positiva, y está también presente, con un peso menor y una valencia negativa asociada a la repugnancia, el nodo *pecado* (gráficos 9 y 11).

Siguiendo con las divergencias relevantes, mediante las palabras frecuentes se había detectado también la importancia de la *felicidad* en *Madame Bovary* frente a *La Regenta*, donde no es un nodo semántico. El análisis de sentimientos confirma esta marcada diferencia de peso: la felicidad es, junto con el amor, el núcleo principal de la obra de Flaubert, como puede apreciarse en todos los gráficos. Una felicidad que, a diferencia del amor, que es pura alegría positiva, es también expectación (gráfico 8). Por otro lado, dentro de los nodos con valencia negativa, en *La Regenta* *la culpa*, *la vergüenza*, *el ridículo* (gráficos 7 y 9) y *el escándalo* (gráfico 11) tienen un peso marcado, mientras que en *Madame Bovary* lo tiene la *duda*, que se asocia al miedo (gráficos 7, 10 y 12).

Para concluir el análisis semántico y antes de pasar a comparar la forma de la historia, es preciso mencionar, aunque solo sea brevemente, el elemento que ha sido el principal objeto de examen y debate en las investigaciones sobre la supuesta influencia de *Madame Bovary* en *La Regenta*: la escena del teatro. Sobre esta cuestión aún abierta, afirma Clarín (1888: 25):

En *Madame Bovary* la escena del teatro es un episodio insignificante, de los de menos relieve; en mi novela es un largo capítulo en que se estudia el alma de *La Regenta* por muchos lados, un capítulo de los principales para la acción interna del libro.

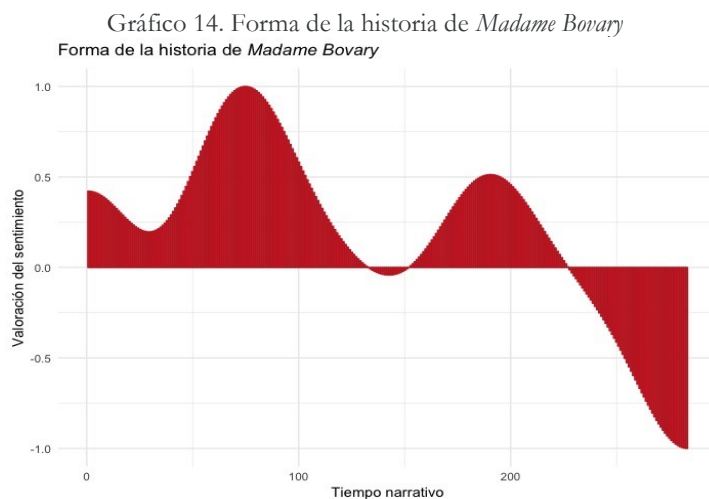
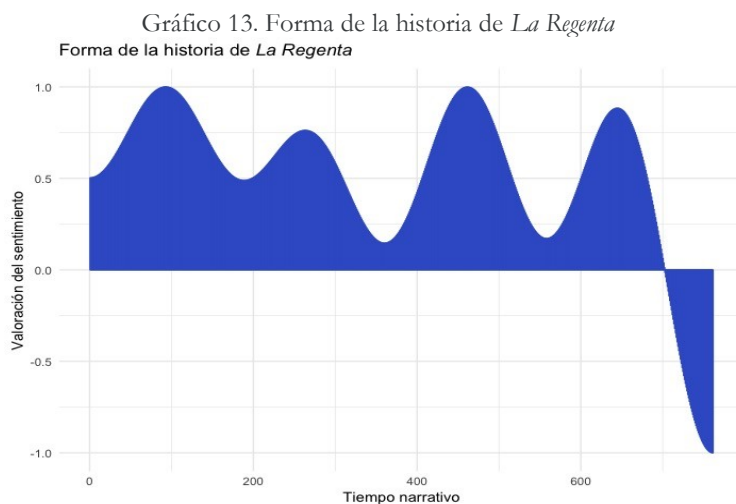
Aunque serían necesarias investigaciones con un mayor grado de granularidad para dar respuesta a esta cuestión, según los datos recabados (gráficos 11 y 12), el nodo *teatro* parece tener, tal como afirmaba Clarín, mayor peso en *La Regenta* que en *Madame Bovary*, si bien, dada su presencia entre los nodos temáticos, no cabe duda de que en esta última obra no es “un episodio insignificante”, sino de relieve. El *teatro* tiene, además, en ambas novelas una valencia positiva asociada a la alegría.

Para concluir, puede afirmarse que el análisis de sentimientos configura dos mundos cuyas valencias emotivas a escala “microscópica” no se corresponden o se corresponden solo tangencialmente, por lo que tampoco en este caso puede hablarse de influencia.

3.4. LA FORMA DE LA HISTORIA

En esta sección se explora la forma de la historia (§ 2.3.2). Para ello, se traza, en primer lugar, la trayectoria de la técnica utilizada por Clarín y Flaubert para dar forma a la narración (*syuzhet*) y se pasa después a comparar las dos trayectorias entre sí a fin de confirmar o refutar la existencia de correspondencia entre el arco emotivo de la *Madame Bovary* y el de *La Regenta* y, en consecuencia, entre las fluctuaciones de los sentimientos y emociones a lo largo de ambas obras. Para acabar se comparará también el arco emotivo del original francés de *Madame Bovary* con su traducción al español para analizar la equivalencia emotiva entre el texto origen y el texto meta (§ 2.4).

Se incluyen a continuación las visualizaciones globales de la forma de la historia que hemos obtenido aplicando el paquete *Syuzhet* a los datos del análisis de sentimientos obtenidos previamente (gráficos 13 y 14). Los resultados se representan mediante un sistema de coordenadas en el que en el eje de abscisas (x) se indica el avance del tiempo narrativo desde el principio hasta el final de la obra, mientras que en el eje de ordenadas (y) se indica la valencia positiva o negativa de los sentimientos y emociones presentes en cada momento de la narración.

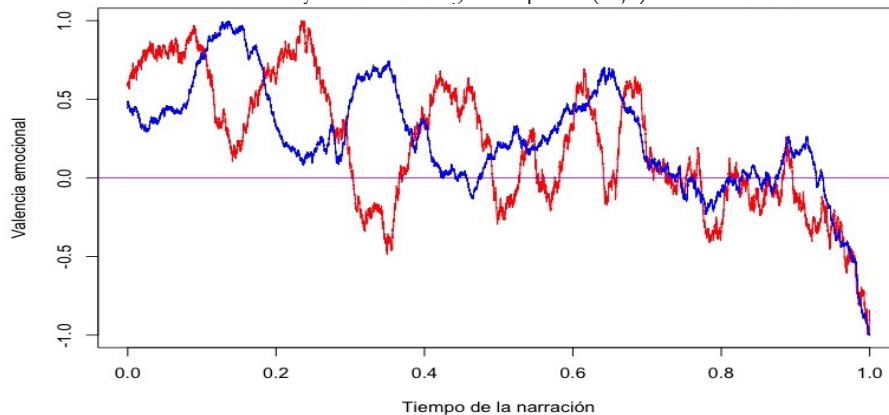


En los dos gráficos anteriores podemos reconocer una trayectoria común, característica de las tragedias (Booker, 2004): la historia se mantiene en general por encima de la línea media de valoración emotiva (0.0) para sufrir un drástico y brusco descenso al final. Sin embargo, la fluctuación de las dos curvas presenta diferencias. Dado que los textos tienen distinta longitud, para poder comparar los arcos emotivos con precisión, mediante las funcionalidades de *Synzhet* y el desarrollo de códigos en R, hemos calculado sus valencias, aplicado la función `rescale_x_2` y trazado con `plot` los gráficos comparativos que figuran a continuación.

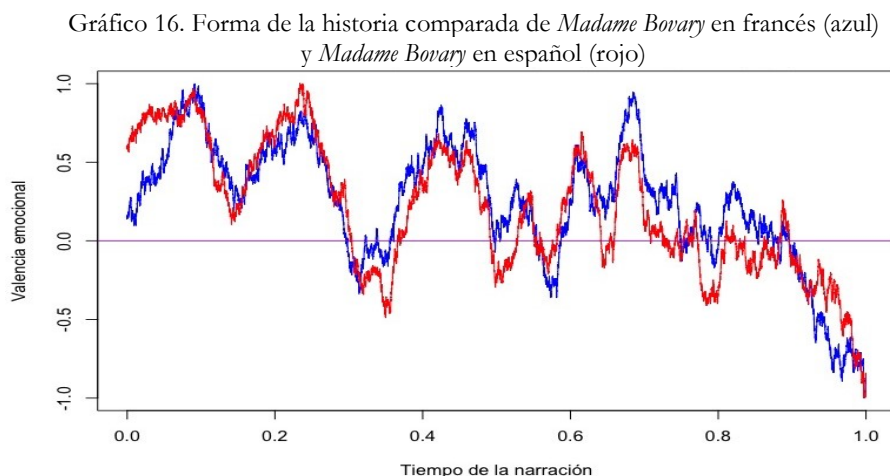
En el gráfico 15 puede observarse con claridad que la forma de la historia, como acabamos de apuntar, solo se corresponde en la última parte de las obras, caracterizadas por el marcado y rápido descenso de la línea de la valencia emocional por debajo de la media propio de las tragedias. Las fluctuaciones de las emociones a lo largo del tiempo narrativo son, sin embargo, diferentes y, en buena parte, contrapuestas (una línea sube

cuando la otra baja). Ambas novelas tienen una valencia emocional primordialmente positiva. Sin embargo, en *Madame Bovary* la curva de la historia presenta continuas fluctuaciones a lo largo de la narración, con una alternancia de drásticos ascensos y descensos, llegando a situarse antes del final de la obra hasta en siete ocasiones por debajo de la línea media para volver a subir siempre después por encima de ella de manera decidida. Esto no sucede en *La Regenta*, cuya curva emotiva presenta menos fluctuaciones y menos marcadas y se mantiene además, prácticamente siempre por encima de la línea media antes de llegar al final (solo en dos ocasiones se sitúa apenas por debajo de la media). En conclusión, también por lo que se refiere a la forma de narrar emociones y sentimientos, la hipótesis de la influencia de *Madame Bovary* en *La Regenta* queda refutada por los datos, en apoyo de cuanto afirmaba Clarín.

Gráfico 15. Forma de la historia comparada de *La Regenta* (azul) y *Madame Bovary* en español (rojo)



Por lo que se refiere, sin embargo, a la congruencia entre el TO en francés y el TM en español de *Madame Bovary*, las dos curvas emotivas presentan una perfecta correspondencia (gráfico 16). En consecuencia, puede afirmarse que hay equivalencia emotiva entre los dos textos.



4. CONCLUSIONES

En este trabajo hemos presentado una nueva forma de abordar el estudio de la influencia entre obras literarias mediante el análisis computacional, en particular mediante la estilometría y el análisis de sentimientos (nodos semánticos y forma de la historia) basados en lenguaje R. Pese a que este enfoque de lectura distante presente aún limitaciones debido principalmente a su novedad y a la falta de recursos específicos para lenguas distintas del inglés, abre futuras vías para la investigación de fenómenos relacionados con la influencia entre textos literarios (imitación, afinidad, plagio), así como con la equivalencia traductora.

En el caso de *Madame Bovary* y *La Regenta*, los datos sobre los patrones estilísticos, los nodos semánticos y la polaridad y emociones básicas correspondientes, y la forma de la historia permiten refutar la hipótesis de la influencia de la obra de Flaubert en la de Clarín en relación con estos aspectos.

Por lo que se refiere a la equivalencia traductora, se ha podido comprobar que, entre el texto del original francés de *Madame Bovary* y el texto de la traducción al español analizado, hay equivalencia tanto estilística como emotiva. Ambas merecerían, sin embargo, mayor investigación. En el primer caso, sería interesante analizar más a fondo las idiosincrasias estilísticas internas del texto meta, a fin de comprobar si las diferencias de estilo detectadas en él pueden deberse a la presencia de dos traductores. En cuanto a la equivalencia emotiva, sería importante investigar esta nueva forma de equivalencia traductora en literatura, así como su posible aplicación.

BIBLIOGRAFÍA

ALGEE-HEWITT, M; HEUSER, R.; MORETTI, F. (2015): "On Paragraphs. Scale, Themes, and Narrative Form", *Stanford Literary Lab Pamphlet 10*, <<https://litlab.stanford.edu/LiteraryLabPamphlet10.pdf>>.

- ALLISON, S.; HEUSER, R.; JOCKERS, M.; MORETTI, F.; WITMORE, M. (2013a): “Quantitative Formalism: an Experiment”, *Stanford Literary Lab Pamphlet 1*, <<https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>>.
- ALLISON, S.; GEMMA, M.; HEUSER, R.; MORETTI, F.; TEVEL, A.; YAMBOLIEV, I. (2013b): “Style at the Scale of the Sentence”, *Stanford Literary Lab Pamphlet 5*, <<https://litlab.stanford.edu/LiteraryLabPamphlet5.pdf>>.
- ALM, C. O.; ROTH, D.; SPROAT, R. (2005): “Emotions from text: Machine learning for text-based emotion prediction”, *Proceedings of the Joint Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMPNLP-2005)*, octubre de 2005, Vancouver (Canadá), pp. 579-586, <<https://www.aclweb.org/anthology/H05-1073.pdf>>.
- ALM, C. O.; SPROAT, R. (2005): “Emotional Sequencing and Development in Fairy Tales”, *Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction (ACII 2005)*, 22-24 de octubre de 2005, Beijing (China), Springer, pp. 668-674.
- ARCHER, J.; JOCKERS, M. L. (2016): *The Bestseller Code. Anatomy of the Blockbuster Novel*, St. Martin Press.
- ARGAMON, S. (2008): “Interpreting Burrows’s Delta: Geometric and Probabilistic Foundations”, *Literary and Linguistic Computing* 23, pp. 131-147.
- ASCARI, M. (2014): “The Dangers of Distant Reading”: Recessing Moretti’s Approach to Literary Genres”, *Genre* 47 (1), pp. 1-18.
- BAAYEN, H. (2001): *Word Frequency Distributions*, Dordrecht: Kluwer.
- BAAYEN, H. (2008): *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*, Cambridge: Cambridge University Press.
- BAAYEN, H.; VAN HALTEREN, H.; NEIJT, A.; TWEEDIE, F. (2002): “An Experiment in Authorship Attribution”, *Proceedings of JADT 2002*, Université de Rennes, St. Malo, pp. 29-37.
- BALAHUR, A.; TURCHI, M. (2012): “Multilingual Sentiment Analysis using Machine Translation?”, *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 52-60, Association for Computational Linguistics. <<https://www.aclweb.org/anthology/W12-3709>>.
- BECKER, R. A.; CHAMBERS, J. M. (1984): *S: An interactive environment for data analysis and graphics*. Boca Raton: CRC Press.
- BODE, K. (2017): “The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”, *Modern Language Quarterly* 78(1), pp. 77-106.
- BOGDANOVA, D.; LAZARIDOU, A. (2014): “Cross-language authorship attribution”, *LREC*, pp. 2015-2020.
- BOLLEN, J.; MAO, H.; ZENG, X. (2011): “Twitter Mood Predicts the Stock Market”. *Journal of Computational Science*, 2(1), pp. 1-8.
- BONAFoux, L. (1888): *Yo y el plagiaro Clarín. Tiquismiquis de Aramis*, Madrid: Administración.
- BOOKER, C. (2004): *The Seven Basic Plots: Why We Tell Stories*. Nueva York: Continuum.

- BOTO BRAVO, M. A. (2017): “Mapa estilométrico de la narrativa de Eduardo Mendoza: aproximación a un análisis estilístico computacional de textos literarios”, *EPOS*, XXXIII, pp. 99-114.
- BOTREL, J. F. (1968): “Últimos ataques de Bonafoux a Clarín”, *Archivum* 18, pp. 177-188.
- BUECHEL, S.; HELLRICH, J.; HAHN, U. (2016): “Feelings from the Past – Adapting Affective Lexicons for Historical Emotion Analysis”, *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, diciembre de 2016, Osaka (Japón), The COLING 2016 Organizing Committee, pp. 54-61, <<https://www.aclweb.org/anthology/W16-4008.pdf>>.
- BURROWS, J. (2002): “Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship”, *Literary and Linguistic Computing* 17, pp. 267–287.
- CIOTTI, F. (2015): “Sul distant reading: una visione critica”, en Stella, F.; Ciotti, F. (eds.), *The Mechanic Reader. Digital Methods for Literary Criticism*, *Semicerchio* 2, Pisa: Pacini Editore, pp. 12-20.
- CLARÍN, Leopoldo Alas (1888): “Mis plagios”, en *Folleto literarios IV*. Madrid: Librería de Fernando Fé.
<https://bibliotecadigital.jcyl.es/es/catalogo_imagenes/grupo.cmd?path=10074235>.
- CLAVERÍA, C. (1942): “Flaubert y *La Regenta* de Clarín”, *Hispanic Review*, 10 (2), pp. 116-125.
- CRAIG, H.; KINNEY, A. (Eds.) (2009): *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge: Cambridge University Press.
- DUPONT, D. (2007): “Accusation and Affirmation: Intertextuality in the Writings of Luis Bonafoux and Leopoldo Alas”, *Revista Canadiense de Estudios Hispánicos*, 31 (2), pp. 285-308.
- EDER, M. (2011): “Style-Markers in Authorship Attribution. A Cross-Language Study of the Authorial Fingerprint”, *Studies in Polish Linguistics*, 6, pp. 99-114.
- EDER, M.; KESTERMONT, M.; RYBICKI, J. (2013): “Stylometry with R: a suite of tools”, *Digital Humanities: Conference Abstracts*, University of Nebraska, Lincoln, NE, pp. 487-89.
- EDER, M.; RYBICKI, J. (2011): “Stylometry with R”, *Digital Humanities 2011: Conference Abstracts*, Stanford University, Stanford, CA, 308-311.
- EDER, M.; RYBICKI, J.; KESTERMONT, M. (2016): “Stylometry with R: a package for computational text analysis”, *R Journal* 8 (1), pp. 107-121. <<https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>>.
- EOFF, S. H. (1965): “En busca de un dios de amor: Gustave Flaubert, Leopoldo Alas”, en *El pensamiento moderno y la novela española, ensayos de literatura comparada*, traducción de Rosario Berdagué, Barcelona, Seix Barral, pp. 59-90.
- FERGUSON, F. (2008): “Planetary literary history: The place of the text”, *New Literary History* 393, pp. 657-684.

- FRADEJAS RUEDA, J. M. (2016): “El análisis estilométrico aplicado a la literatura española: las novelas policíacas e históricas”, *Caracteres. Estudios culturales y críticos de la esfera digital*, Vol. 5(2), pp. 196-245.
<<http://revistacaracteres.net/revista/vol5n2noviembre2016/analisis-estilometrico/>>.
- GEROLAMO MOSSA, M. (2016): “Il demone della distanza e l'angelo della vicinanza. *Riflessioni sulla teoria letteraria di Franco Moretti*”. Relazione per il corso “Seminario di cultura digitale” della professoressa Enrica Salvatori, Università di Pisa, a.a. 2015/16 <<http://www.labcd.unipi.it/wp-content/uploads/2018/05/Mario-Gerolamo-Mossa-Teoria-letteraria-di-Franco-Moretti-.pdf>>.
- GONÇALVES, P.; ARAÚJO, M.; BENEVENUTO, F.; CHA, M. (2013): “Comparing and combining sentiment analysis methods”, *Proceedings of the first ACM conference on Online social networks*, 7 y 8 de octubre de 2013, Boston (EE.UU.), pp. 27-38. <https://www.researchgate.net/publication/262356434_Comparing_and_combining_sentiment_analysis_methods>..
- GOODWIN, J.; HOLBO, J. (eds.) (2011): *Reading Graphs, Maps, Trees. Responses to Franco Moretti. A Valve book event*, Anderson: Parlor Press.
- GRIES, S. (2010): *Statistics for Linguistics with R: A Practical Introduction*, Berlín: De Gruyter Mouton.
- GRIEVE, J. (2007): “Quantitative Authorship Attribution: An Evaluation of Techniques”, *Literary and Linguistic Computing* 22, pp. 251-270.
- HAMMOND, A. (2017): “The double bind of validation: distant reading and the digital humanities’ through of disillusionment”, *Literature Compass*, 14: 8, pp. 1-13. <<http://www.adamhammond.com/wp-content/uploads/2015/05/double-bind-of-validation.pdf>>.
- HATZFELD, H. (1977): “La imitación estilística de *Madame Bovary* (1875) en *La Regenta* (1884)”, *Thesaurus*, 352 (1), pp. 40-53.
- HENRÍQUEZ, C.; GUZMÁN, J. (2017): “A review of sentiment analysis in Spanish”, *TECCIENCIA*, 12: 22, pp. 35-48.
<http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-36672017000100035>.
- HIRST, G.; FEIGUINA, O. (2007): “Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts”, *Literary and Linguistic Computing*, 22, pp. 405-417.
- HOLMES, D. (1998): “The Evolution of Stylemetry in Humanities Scholarship”, *Literary and Linguistic Computing*, 13, pp. 111-117.
- HOOVER, D. L. (2002): “Frequent Word Sequences and Statistical Stylistics”, *Literary and Linguistic Computing*, 17, pp. 157-180.
- HOOVER, D. L. (2004): “Testing Burrows’s Delta”, *Literary and Linguistic Computing*, 19, pp. 453-475.
- IHAKA, R.; GENTLEMAN, R. (1996): “R: A language for data analysis and graphics”, *Journal of Computational and Graphical Statistics*, 5(3), pp. 299-314.
- JIN, J. (2017): “Problems of Scale in “Close” and “Distant” Reading”, *Philological Quarterly*, 96: 1, pp. 105-129.

- JOCKERS, M. (2013): *Macroanalysis: Digital Methods and Literary History*, (Topics in the Digital Humanities), Chicago: University of Illinois Press.
- JOCKERS, M. (2014): *Text analysis with R for students of literature*. Nueva York: Springer.
- JOCKERS, M. (2015): “Reveling Sentiment and Plot Arcs with the Syuzhet Package”. <<http://www.matthewjockers.net/2015/02/02/syuzhet/>>
- JOCKERS, M. (2017): “Introduction to the Syuzhet Package”, en <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>.
- JOCKERS, M.; MIMNO, D. (2013): “Significant Themes in 19th-Century Literature”, *Poetics*, 41(6), pp. 750-769.
- JOCKERS, M.; THALKEN, R. (2020): *Macroanalysis: Digital Methods and Literary History*, (Topics in the Digital Humanities), Chicago: University of Illinois Press.
- JOSHI, P. (2006): “India, 1850-1900”, en Moretti, F. (ed.), *The Novel, Vol. I, History, Geography, and Culture*, Princeton - Oxford: Princeton University Press, pp. 495-508.
- JUOLA, P. (2006): “Authorship Attribution”, *Foundations and Trends in Information Retrieval* 1, pp. 233-334.
- KAKKONEN, T.; KAKKONEN, G. G. (2011): “SentiProfiler: creating comparable visual profiles of sentimental content in texts”, *Proceedings of Language Technologies for Digital Humanities and Cultural Heritage*, septiembre 2011, Hissar (Bulgaria), Association for Computational Linguistics, pp. 62-69. <<https://www.aclweb.org/anthology/W11-4110.pdf>>.
- KHADEM, A. (2012): “Annexing the unread: a close reading of «distant reading»”, *Neohelicon*, 39, pp. 409-421.
- KIM, E.; PADÓ, S.; KLINGER, R. (2017): “Prototypical Emotion Developments in Literary Genres”, *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, agosto 2017, Vancouver (Canadá), Association for Computational Linguistics, pp. 17-26. <<https://www.aclweb.org/anthology/W17-22.pdf>>.
- KLINGER, R.; SULIYA, S. S.; REITER, N. (2016): “Automatic Emotion Detection for Quantitative Literary Studies: A case study based on Franz Kafka’s *Das Schloss und Amerika*”, *Digital Humanities 2016: Conference Abstracts*, 11-16 de julio de 2016, Cracovia (Polonia), Jagiellonian University & Pedagogical University, pp. 826-828. <<https://dh2016.adho.org/abstracts/318>>.
- KRISTAL, E. (2002): “Considering coldly...?: A response to Franco Moretti”, *New Left Review*, 15, pp. 61-74.
- LAFITTE, G. (1943): “Madame Bovary et La Regenta”, *Bulletin Hispanique*, XLV, pp. 157-163.
- LERTNATTEE, V.; THEERAMUNKONG, T. (2004): “Effect of term distributions on centroid-based text categorization”, *Information Sciences*, 158, pp. 89-115.
- LISSORGUES, Y. (1987): “Introduction”, en Leopoldo Alas dit Clarín, *La Régente*, traduit de l'espagnol par Albert Belot, Claude Bleton, Jean-François Botrel, Robert Jammes, Yvan Lissorgues, Fayard, pp. 7-29.

- LIU, A. (2012): “Where Is Cultural Criticism in the Digital Humanities?”, en Gold, M. K. (ed.), *Debates in the Digital Humanities*, University of Minnesota Press, cap. 29. <<http://dhdebates.gc.cuny.edu/debates/text/20>>
- LIU, B. (2015): *Sentiment Analysis. Mining Opinions, Sentiments and Emotions*, Nueva York: Cambridge University Press.
- LUTOLAWSKI, W. (1898): “Principes de stylométrie”, *Revue des Études Grecques*, 11 (41), pp. 61-81.
- MARTÍNEZ CACHERO, J.M. (1953): “Luis Bonafoux y Quintero, *Aramis*, contra *Clarín*. Historia de una enemistad literaria”, *Revista de Literatura* III, pp. 99-111.
- MELÓN RUIZ DE GORDEJUELA, S. (1952): “Clarín y el Bovarismo”, *Archivum*, II, pp. 69-87.
- MENDENHALL, T. C. (1901): “A Mechanical Solution of a Literary Problem”, *The Popular Science Monthly*, LX (7), pp. 97-105.
- MISHNE, G.; GLANCE, N. (2006): “Predicting Movie Sales from Blogger Sentiment”. *Proceedings of AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*, Stanford, California. <<https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-030.pdf>>.
- MOHAMMAD, S. M. (2012): “From once upon a time to happily ever after: Tracking emotions in mail and books”, *Decision Support Systems*, 53, pp. 730–741. doi:10.1016/j.dss.2012.05.030
- MOHAMMAD, S. M. (2011): “From once upon a time to happily ever after: Tracking emotions in novels and fairy tales”, *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 105-114.
- MOHAMMAD, S. M.; Turney, P. D. (2010): “Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon”, *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California, pp. 26-34. <<https://www.aclweb.org/anthology/W10-0204.pdf>>.
- MOHAMMAD, S. M.; DUNNE, C.; DORR, B. (2009): “Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus”, *Proceedings of Empirical Methods in Natural Language Processing (EMNLP-2009)*, Singapore, pp. 599-608.
- MOHAMMAD, S. M.; Yang, T. (2011): “Tracking sentiment in mail: how genders differ on emotional axes”, *Proceedings of the ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pp. 70-79 <<https://www.aclweb.org/anthology/W11-1709.pdf>>.
- MORETTI, F. (1998): *Atlas of the European Novel 1800-1900*, Londres - Nueva York: Verso.
- MORETTI, F. (2000a): “Conjectures on World Literature”, *New Left Review*, 1, pp. 54-68.
- MORETTI, F. (2000b): “The Slaughterhouse of Literature”, *Modern Language Quarterly*, 61: 1, pp. 54-67.

- MORETTI, F. (2005): *Graphs, Maps, Trees: Abstract Models for Literary History*, Londres - Nueva York: Verso; (tr. it., *La letteratura vista da lontano*, Turín: Einaudi).
- MORETTI, F. (2013a): *The Bourgeois - Between History and Literature*, Londres - Nueva York: Verso (tr. it., *Il Borghese. Tra storia e letteratura*, Turín: Einaudi, 2017).
- MORETTI, F. (2013b): *Distant Reading*, Londres - Nueva York: Verso.
- MORETTI, F. (2013c): “Operationalizing: or, the Function of Measurement in Modern Literary Theory”, *Stanford Literary Lab Pamphlet 6*.
<<https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>>.
- MORETTI, F. (2016a): “Literature, Measured”, *Stanford Literary Lab Pamphlet 12*.
<<https://litlab.stanford.edu/LiteraryLabPamphlet12.pdf>>.
- MORETTI, F. (2016b): “The Emotions of London”, *Stanford Literary Lab Pamphlet 13*.
<<https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>>.
- MORETTI, F. (2017): “Patterns and Interpretation”, *Stanford Literary Lab Pamphlet 15*.
<<https://litlab.stanford.edu/LiteraryLabPamphlet15>>.
- O’CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B. R.; SMITH, N. A. (2010): “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series”. *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, pp. 122-129.
<<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1536/1842>>.
- ORSINI, F. (2002): “India in the Mirror of World Fiction”, *New Left Review*, 13, pp. 75-88.
- PANG, B.; LEE, L. (2008): “Opinion mining and sentiment analysis”, *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.
- PANG, B.; LEE, L.; VAITHYANATHAN, S. (2002): “Thumbs up?: Sentiment classification using machine learning techniques”, *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Vol. 10, EMNLP ’02, pp. 79-86.
- PARLA, J. (2004): The object of comparison. *Comparative Literature Studies*, 41(1), pp. 116-125.
- POLTI, G. (1921): *The Thirty-Six Dramatic Situations*, James Knapp Reeve, Franklin OH.
- PORIA, S.; HUSSAIN, A.; CAMBRIA, E. (2018): *Multimodal Sentiment Analysis*, Socio-Affective Computing Series, Vol. 8, Springer.
- PRENDERGAST, C. (2001): “Negotiating World Literature”, *New Left Review*, 8, pp. 100-121.
- PRENDERGAST, C. (2005): “Evolution and Literary History: A Response to Franco Moretti”, *New Literary History*, 34, pp. 40-62.
- REAGAN, A. J.; MITCHELL, L.; KILEY, D.; DANFORTH, C. M.; DODDS, P. S. (2016): “The emotional arcs of stories are dominated by six basic shapes”, *EPJ Data Science*, 5(31), pp. 1-12.
- RICHMOND, C. (1977): “La polémica Clarín-Bonafoux y Flaubert”, *Ínsula* 365, pp. 1-12.
- RSTUDIO TEAM. (2017): *RStudio: Integrated development environment for*, R. Boston, MA: RStudio Inc.

- SADIKOV, E.; PARAMESWARAN, A.; VENETIS, P. (2009): “Blogs as Predictors of Movie Success”. *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM-2009)*, pp. 304-307.
<http://infolab.stanford.edu/~venetis/publications/2009/blog_predictions.pdf>.
- SARWAR, R.; Li, Q.; RAKTHANMANON, T.; NUTANONG, S. (2018): “A Scalable Framework for Cross-lingual Authorship Identification”, *Information Sciences*, pp. 323-339.
- SCHMIDT, T.; BURGHARDT, M. (2018): “An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing”, *Proceedings of Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 139-149
<<https://www.aclweb.org/anthology/W18-4516/>>.
- SERLEN, R. (2010): “The Distant Future? Reading Franco Moretti”, *Literature Compass* 7(3), pp. 214-225.
- SILGE, J.; ROBINSON, D. (2019): *Text Mining with R. A Tidy Approach*
<<https://www.tidytextmining.com>>.
- SMITH, P.; ALDRIDGE, W. (2011): “Improving Authorship Attribution: Optimizing Burrows’s Delta Method”, *Journal of Quantitative Linguistics*, 18, pp. 63-88.
- SOBEJANO, G. (1981): “*Madame Bovary en La Regenta*”, *Los Cuadernos del Norte. Revista Cultural de la Caja de Ahorros de Asturias*, núm. 7, año II, pp. 22-27. Oviedo, Caja de Ahorros de Asturias. Biblioteca Virtual Miguel de Cervantes:
<<http://www.cervantesvirtual.com/nd/ark:/59851/bmcm6216>>.
- SOBEJANO, G. (1985): *Clarín en su obra ejemplar*, Madrid: Castalia.
- SOCHER, R.; PERELYGIN, A.; WU, J. Y.; CHUANG, J.; MANNING, C. D.; NG, A. Y.; POTTS, C. (2013): “Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank”, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 18-21 octubre de 2013, Seattle (EE.UU.), Association for Computational Linguistics, pp. 1631-1342.
<<https://www.aclweb.org/anthology/D13-1170.pdf>>.
- SOCHER, R.; MANNING, C. D.; NG, A. Y. (2010): “Learning continuous phrase representations and syntactic parsing with recursive neural networks”, *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*, pp. 1-9.
<<http://www.robotics.stanford.edu/~ang/papers/nipsdluf10-LearningContinuousPhraseRepresentations.pdf>>.
- STAMATATOS, E. (2009): “A survey of modern authorship attribution methods”, *Journal of the American Society for Information Science and Technology*, Volume 60 (3), pp. 538-556. <<http://www.icsd.aegean.gr/lecturers/stamatatos/papers/survey.pdf>>.
- STUART, K. D.; MAJEWSKY, M. (2015): “Intelligent Opinion Mining and Sentiment Analysis Using Artificial Neural Networks”, en Arik, S; Huang, T.; Lai, W.K.; Q. Liu. Q. (eds): *Neural Information Processing, Proceedings of the International Conference on Neural Information Processing, ICONIP 2015, Parte IV*, Suiza: Springer, pp. 103-110.

- TOBIAS, R. B. (1993): *20 Master Plots: And How to Build Them*, Cincinnati, Ohio (EE.UU.): Writer's Digest Books, F&W Publications.
- TOLEDANO GARCÍA, M.C. (1989): "Presencia de Flaubert en Leopoldo Alas Clarín", en F. Lafarga (ed.), *Imágenes de Francia en las letras hispánicas*, Barcelona: Promociones y Publicaciones Universitarias, pp. 389-397.
- TUMASJAN, A.; SPRENGER, T. O.; SANDNER, P. G.; WELPE, I. M. (2010): "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", *Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010)*, pp. 178-185.
<<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1441/1852>>.
- TURNERY, P. D.; LITTMAN, M. (2003): "Measuring praise and criticism: Inference of semantic orientation from association", *ACM Transactions on Information Systems (TOIS)*, 21(4), pp. 315-346.
- UNDERWOOD, T. (2017): "A genealogy of distant reading", *Digital Humanities Quarterly*, 11 (2). <<http://digitalhumanities.org/dhq/vol/11/2/000317/000317.html>>.
- VENTURA AGUDIEZ, J. (1963): "Emma Bovary-Ana Ozores o el símbolo del amor", *The Romanic Review*, LIV, pp. 20-29.
- WILHELM, T.; BURGHARDT, M.; WOLFF, C. (2013): "To See or Not to See - An Interactive Tool for the Visualization and Analysis of Shakespeare Plays", en Franken-Wendelstorf, R.; Lindinger, E.; Sieck, J. (eds.): *Kultur und Informatik: Visual Worlds & Interactive Spaces*, Glückstadt: Verlag Werner Hülsbusch, pp. 175-185.
- ZHANG, W.; SKIENA, S. (2010): "Trading Strategies to Exploit Blog and News Sentiment", *Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010)*, pp. 375-378.
<<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1529/1904>>.
- ZHANG, X.; FUEHRES, H.; GLOOR, P. A. (2011): "Predicting Stock Market Indicators through Twitter 'I Hope It Is Not as Bad as I Fear'". *Procedia - Social and Behavioral Sciences*, 26, pp. 55-62.
<<https://www.sciencedirect.com/science/article/pii/S1877042811023895>>.