

#23

EL TEXTO LITERARIO HECHO DATOS: F. M. DOSTOIEVSKI EN EL MARCO DE LAS HUMANIDADES DIGITALES Y LOS ENFOQUES CUANTITATIVOS

Benamí Barros García

Universidad de Granada



Resumen || El discurso en las obras de F. M. Dostoievski lleva más de 150 años atrayendo la atención de numerosos especialistas, cada vez procedentes de ámbitos más dispares. Con la consolidación de las Humanidades digitales (HD) como espacio necesariamente interdisciplinar se han conseguido interesantes avances y no menos sugestivas propuestas que parecen arrojar luz sobre la interpretación de ciertas cuestiones narratológicas de las obras de este autor. En el presente estudio abordaremos el estado actual de la imbricación de las HD y los estudios literarios y mostraremos el potencial del análisis cuantitativo, las técnicas de visualización y el análisis de redes para el estudio de textos literarios y, en particular, de la obra literaria de Dostoievski.

Palabras clave || Dostoievski | Visualización de datos | Análisis del discurso | Estilometría | Análisis de redes

Abstract || Discourse in Dostoevsky's works has drawn the attention of numerous specialists in increasingly diverse fields for more than 150 years. As Digital Humanities (DH) become consolidated as an interdisciplinary space of research, interesting advances and proposals have been made that seem to shed new light on the interpretation of certain narratological questions in the works of this author. This contribution studies the current state of intersection of DH with Literary Studies and shows the potential of quantitative analysis, data visualization techniques and network analysis as applied to the study of literary texts and, particularly, the literary works Dostoevsky.

Keywords || Dostoevsky | Information visualization | Discourse analysis | Stylometry | Network analysis

Resum || El discurs en les obres de F. M. Dostoievski porta més de 150 anys atraient l'atenció de nombrosos especialistes, cada vegada procedents d'àmbits més dispars. Amb la consolidació de les Humanitats digitals (HD) com a espai necessàriament interdisciplinari s'han aconseguit interessants avanços i no menys suggestives propostes que semblen llançar llum sobre la interpretació d'unes certes qüestions narratològiques de les obres d'aquest autor. En el present estudi abordarem l'estat actual de la imbricació de les HD i els Estudis literaris i mostrarem el potencial de l'anàlisi quantitativa, les tècniques de visualització i l'anàlisi de xarxes per a l'estudi de textos literaris i, en particular, de l'obra literària de Dostoievski.

Paraules clau || Dostoievski | Visualització de dades | Anàlisi del discurs | Estilometria | Anàlisi de xarxes

0. Introducción

Quien hoy en día decide dedicarse al análisis del texto literario puede beneficiarse del enorme potencial de las herramientas y enfoques basados en las Tecnologías de la información y la comunicación (TIC), con los que podrá acceder a una información y generar un conocimiento difícilmente alcanzable mediante otras perspectivas y métodos. Encontrará, además, un ámbito o marco de escurridiza y difusa definición que lo ampara (Terras et al., 2013; Kaplan, 2015; Sula y Hill, 2019), las Humanidades Digitales (HD), pero que también le obliga a ser cauto y riguroso a nivel metodológico (cfr. Gold, 2012).

Todavía más difusas en su definición tras un proceso de profunda transformación (Gold y Klein, 2012), las HD parecen querer ser claras en sus objetivos y valores (Spiro, 2012), muestra de lo cual es el continuo ejercicio de diferenciación con respecto a otras áreas y disciplinas que fueron, o son, afines (para una revisión: Kaplan, 2015). Las HD son un claro ejemplo de la necesaria interdisciplinariedad en la creación y difusión del conocimiento, de un conocimiento digital más complejo (Romero Frías y Bocanegra Barbecho, 2018: 12) y de que la comunicación de la ciencia es más eficiente si esta se acerca y ajusta a la realidad en la que se construye y se comparte, es decir, si el conocimiento transferido se puede integrar fácilmente en el conocimiento del receptor (Burkhard, 2005).

A pesar de los avances en el análisis del texto en general (Egan, 2019) y a diferencia de lo ocurrido en otros ámbitos, continúa la lucha por encontrar un respaldo amplio en los estudios literarios que sirva para consolidar los conocimientos que han derivado de las HD:

The digital humanities community has produced some beautiful work and made many advances in technology, design, and standards. Now is the moment to consolidate that knowledge and to articulate the community's requirements and goals. Go forth and do great things... (Borgman, 2009).

Una de las problemáticas fundamentales que supone la imbricación de estas nuevas herramientas y los estudios literarios es el miedo a que la automatización de los análisis, de alguna manera, pueda desvirtuar el discurso literario mediante su descontextualización. La cuestión quizá radique en que la finalidad de la imbricación de las HD y los estudios literarios no sea tanto el análisis del sentido de los textos y fenómenos literarios, sino posibilitar que ese análisis sea más fundamentado, que pueda ser más amplio y profundo, llegando a tener en cuenta elementos y comportamientos discursivos difícilmente advertibles desde otros enfoques. De esta forma, el análisis computarizado de textos literarios vendría no tanto a hablarnos de sentidos últimos, de los sentimientos o intenciones del autor, sino, más bien, a enseñarnos cómo este nuevo método puede ayudar a contextualizar una obra dentro de una ingente colección

de obras literarias (Jockers, 2013: 32) y, por tanto, a facilitar y perfeccionar su interpretación y su conocimiento, superando las limitaciones inherentes a los enfoques basados en estrategias de lo que llamaron *close reading* (Schöch, 2013). Siguiendo la clásica advertencia, estaríamos hablando de abandonar esa cuestionable posición del investigador que escribe sobre literatura sin poder observar parte de su material o, dicho de otro modo, haciendo como si la mitad de su material no existiera (Burrows, 1987: 1).

En este sentido, es llamativo que muchos investigadores dentro de las HD consideren que su único o prioritario foco de atención sea precisamente el análisis textual (cfr. Sula y Hill, 2019). Parece, por tanto, que las reticencias están del lado de algunos estudiosos de la literatura que no aceptan las HD como método/s y no tanto de los investigadores en HD, que sí tienen claro que uno de sus objetos de estudio es o podría ser el texto literario.

En el presente artículo nos centraremos en las aportaciones y potencial de la visualización de datos a la hora de analizar literatura. En particular, plantearemos y mostraremos las posibilidades de análisis que brindan diferentes técnicas y herramientas de visualización a la hora de estudiar las obras de F. M. Dostoievski, tanto individualmente como en su conjunto o en comparación con otras obras y autores. Para ello partiremos de una cuestión que suele despertar suspicacias entre algunos críticos, que conciben el hecho de hablar de datos de los textos o del análisis computarizado de textos como una práctica arriesgada y a veces poco consistente: ¿qué son y cómo deben ser los datos en Humanidades? Sin entrar en mayores debates (cfr. Allés Torrent, 2019), aquí entenderemos que los datos en Humanidades, ya sean definidos como textos, artefactos o información procesable, no son una evidencia por sí mismos, sino que son objetos polifacéticos que pueden usarse como evidencia para apoyar un argumento (Owens, 2011). En los estudios literarios estos datos, recordemos los *smart big data* de Schöch (2013), tienen la capacidad de funcionar como *fuentes potenciales de información*, haciendo evidente aquello a lo que nos referíamos anteriormente con la necesidad de comprender cuáles son los objetivos reales de la incorporación de los métodos digitales a los estudios literarios: la *distant reading* (Moretti, 2013; 2011; 2000; 2005) siempre sería capaz de producir nuevos artefactos que pueden ser abordados mediante una *close reading* (Owens, 2011). Gavin (2019) lleva un poco más lejos la reflexión, al rebajar la importancia de si el análisis computacional de textos consiste o requiere una *close* o *distant reading*, para plantear la cuestión de qué son los datos (palabras) en los que descompone los textos que analiza, sugiriendo que no son ya textos, sino una nueva forma de textualidad (Gavin, 2019: 21). No hay que olvidar en este mismo sentido que las expectativas que un lector o investigador en la actualidad ponen en un texto no son las

mismas que los de otra época o, incluso, de otro nicho sociocultural, y que esta no coincidencia cambia la naturaleza del propio texto (Jones Nulty, 2019).

En nuestra experiencia con la implementación de técnicas de visualización de la información para el análisis de los textos literarios de F. M. Dostoievski hemos constatado que el mayor potencial de las visualizaciones no es la corroboración de hipótesis previas vertidas por la crítica y literatura científica, sino la exploración de nuevos elementos, patrones, sucesos del texto tanto a nivel particular de la obra o del segmento analizado como a nivel global de la contextualización de la obra en la colección o repertorio literario en cuestión.

El objetivo principal del presente artículo es plantear y/o mostrar una serie de análisis llevados a cabo en torno a la obra de F. M. Dostoievski desde la perspectiva del estudio computacional del texto literario y, en concreto, de la visualización de la información, con el fin de evidenciar algunas de las posibilidades que ofrece el hecho de poder trabajar con cantidades ingentes de datos derivados de las obras literarias. Estas visualizaciones, en su gran mayoría, reflejarán la tendencia clásica en los análisis cuantitativos de la literatura hacia la construcción de visualizaciones poco interactivas y orientadas hacia la representación estática de contenidos, sucesos o elementos: nubes de palabras, histogramas, grafos, dendrogramas, etc. En los últimos años se ha abogado por el uso de visualizaciones que puedan facilitar la creación de nuevo conocimiento de manera colaborativa precisamente acentuando el carácter interactivo de las mismas, lo que, a su vez, permitiría sumergirse en otros niveles del texto literario:

But despite the recent rise of digital methods, most visualization techniques for literary studies still exhibit an epistemological imbalance to communicating results, but seldom generating new knowledge. One facet of this reduction to presentation is the use of visualization as unquestioned data representations denying its interactive potential (Schwan et al., 2019).

Esta aparente falta de eficiencia a la hora de comunicar los resultados, si bien nos parece algo discutible, puede tener que ver con el foco en el que se ha puesto la atención, siendo el estudio del léxico y, en concreto de las palabras más frecuentes (MFW), la piedra angular del análisis cuantitativo de textos. Entre otros factores, la irrupción del análisis de redes cambió el paradigma de representación y puso de relieve que el léxico no era el único dominio en el que las visualizaciones podían hacer aportaciones. Se comenzó de esta manera a plantear la posibilidad de representar visualmente los niveles de narración, las interrelaciones entre personajes, las conexiones entre diferentes pasajes, obras, autores, etc.

1. La obra de F. M. Dostoievski como datos

Una de las formas en la que pueden ser comprendidas las HD es como *digitized humanities*, es decir, como la creación, tratamiento y uso de grandes bancos de datos (Roth, 2019). En este sentido, es indispensable disponer de los materiales digitalizados siguiendo una serie de preceptos y recomendaciones bien estipulados, entre otros, por el TEI Guidelines for Electronic Text Encoding (para una reciente revisión: Cummings, 2019). Probablemente de aquí surge el todavía actual debate acerca de la diferenciación entre las HD y la Biblioteconomía y Documentación (entre otros, Conteh, A. y Wilms, 2013), que ha dado lugar a grupos como DARIAH Bibliodata (<https://www.dariah.eu/activities/working-groups/bibliographical-data-bibliodata/>). Es cierto que muchos de los primeros centros universitarios y laboratorios de HD surgieron como proyectos albergados en las bibliotecas de las universidades (Terras et al., 2013), precisamente porque su justificación solía estar asociada a la creación de catálogos de obras literarias, autores, corpus de textos, etc.

El caso de la literatura rusa es algo singular en este sentido, ya que la digitalización de la literatura en Rusia fue una práctica masiva desde los comienzos de la web 2.0. Mucho antes de que las bibliotecas se plantearan cuestiones acerca de la digitalización de sus fondos, ya existían páginas web en las que se podía tener acceso a prácticamente toda la literatura rusa y gran parte de la universal traducida al ruso. Entre otros proyectos, en 1994 apareció la Biblioteca Maksím Moshkov (www.lib.ru), con apoyo de instituciones estatales, en la que se reunía una enorme colección de obras literarias, política, humor, historia, etc., en lengua rusa o traducidas al ruso. Se trata de una biblioteca digital que los usuarios y administradores podían ir ampliando y modificando. La gran mayoría de los textos incluyen metadatos, se pueden descargar como texto plano y están organizados por autor, tipología, género, fecha de publicación, e incluso se aportan comentarios críticos a cada una de las obras. Así, cuando irrumpen las HD en los estudios literarios en Rusia, una de las primeras cuestiones que se planteó no fue tanto la necesidad de digitalizar, sino la de determinar qué ediciones debían estar disponibles y si las colecciones de textos digitalizados existentes eran o son realmente las adecuadas para trabajar con ellos. Sobre esta cuestión en el caso de Dostoievski volveremos un poco más adelante.

La digitalización y conformación del corpus de la literatura rusa experimentaría un punto de inflexión con el lanzamiento y posteriores mejoras de uno de los corpus nacionales de lengua más completos del planeta (www.ruscorpora.ru), que, por supuesto, incluía un gran número de textos literarios del repertorio ruso. Además, ya desde

fases tempranas de desarrollo permitía crear subcorpus por autor, género, época, tipología, etc., así como descargar algunos datos obtenidos en ficheros XML y visualizar los resultados mediante una gráfica de ocurrencias a lo largo del tiempo. Hoy las posibilidades de trabajo con este corpus son muy amplias, pudiendo desempeñar algunas tareas básicas de análisis cuantitativo de textos y visualización de corpus literarios.

En el caso concreto de los textos del autor que nos ocupa, el Corpus Nacional de la Lengua Rusa (CNR) contiene 44 documentos con un total de casi 2 millones de palabras, significativamente inferior al corpus de textos (todavía en proceso de corrección) elaborado por el grupo de textología de la Universidad de Petrozavodsk, dirigido por Vladimir Zakharov, y que actualmente contiene unos 2,7 millones de palabras solo de obras literarias (en corpus aparte disponen de cerca de 5 millones de palabras de textos manuscritos, cartas, textos periodísticos, etc.). Este corpus, todavía no disponible como tal *online*, forma parte del relevante proyecto Digital Dostoevsky, también dirigido por Zakharov (2019; Кибальник, 2018), que, a lo largo de más de veinte años, ha generado diferentes contenidos digitales de gran interés accesibles en philolog.petsu.ru/fmdost/index.htm. En colaboración con el Instituto de Literatura rusa de la Academia de Ciencias de Rusia se creó la web «El archivo de F. M. Dostoevski» (dostoevsky-archive.ru) de enorme valor para el estudio de los manuscritos del autor. Algunos textos originales y manuscritos pueden consultarse en la web del Archivo estatal ruso de literatura y arte (РГАЛИ, rgali.ru).

Por otro lado, conviene mencionar el Diccionario de la lengua de Dostoevski (Караулов, 2008-2010, 2017) como uno de los primeros recursos específicamente cuantitativos para el estudio de sus textos. Se trata de un proyecto que continúa en edición hasta completar todas las palabras que caracterizan el idiolecto del autor, su «idioglosario». La obra contiene diferentes tipos de diccionarios, entre los que hay un interesante diccionario de frecuencias.

Quizá algo más ecléctica sea la propuesta del Laboratorio de Literatura Fantástica (fantlab.ru), fundado en 2004 por Alexei Lvov. En esta web se recopila información acerca de obras de literatura fantástica y hay una sección muy completa dedicada a Dostoevski. Mediante un algoritmo implementado y descrito por el autor (fantlab.ru/article374) desarrollan, entre otras funciones, un análisis «lingüístico» del texto, en el que se obtienen datos acerca del repertorio léxico, palabras únicas, longitud de frase, curvas de crecimiento y disminución de la distribución léxica, tablas de frecuencia por palabras, partes de la oración, etc., así como un muy interesante gráfico e indicador de la carga de diálogos en el texto. El problema, claro, es parecido al que tienen otras herramientas de

visualización de corpus automatizadas: no tenemos control sobre los datos ni sobre las diferentes conversiones que se realizan. Adicionalmente, en esta web aportan los resultados de un leve análisis estilométrico para determinar qué autores podrían haber escrito cierta obra o el grado de parecido entre varios autores. Para algunos de los parámetros arriba reseñados se establece un *ranking* dentro del corpus de textos y/o autores que contiene la web.

Por otro lado, en 2013 se inició la ingente labor de publicar las nuevas obras completas de F. M. Dostoievski en 35 tomos. De las candidaturas presentadas para encargarse del proyecto, finalmente fue escogida la del Instituto de Literatura Rusa (Casa Pushkin) de la Academia de Ciencias de Rusia, siendo Vsevolod Bagnó el editor principal (Bagnó, 2013-2019). Los primeros tomos están disponibles *online* en la web del propio Instituto (russian-literature.org/author/Dostoyevsky), donde también se pueden consultar la edición anterior, hasta ahora canónica, de las obras completas del autor (Bazanov, 1972-1990).

Todos estos recursos no son aún suficientes y esperamos que, en 2021, con motivo del 200 aniversario del escritor, puedan aparecer corpus específicos y herramientas que faciliten el estudio cuantitativo de los datos de Dostoievski. No obstante, creemos que queda patente el interés por la digitalización de sus textos, así como por hacer accesibles los datos, intención que mucho tiene que ver con la atracción que despierta este autor entre los investigadores en HD.

Si bien es cierto que la existencia de una enorme literatura científica dedicada al análisis del léxico, idiolecto y estilo de este autor podría ya ser suficiente para pensar que está justificado el interés por él de los enfoques cuantitativos, quizá haya que tener en cuenta también algunas particularidades del estilo de Dostoievski a la hora de abordar la cuestión. No es el objetivo aquí profundizar en estas cuestiones, pero recordaremos que desde hace más de un siglo los críticos comenzaron a prestar atención a las singularidades de la obra de Dostoievski a nivel léxico, simbólico, estilístico y, en general, narratológico. Puede que muchas de ellas deriven de una muy interesante relación con la palabra (cfr. Barros García, 2015a). Además, se trata de un autor en el que las redacciones previas pueden llegar a ser significativamente distintas a las obras finales, un autor tendente a los cambios profundos incluso en el tipo de narrador en el último momento, que dictó algunas de sus obras, que quiso corregir sus obras en las diferentes ediciones que se publicaron en vida, que sufrió una más que probable censura de capítulos y escenas completas. Estamos, por tanto, ante un autor cuyo estudio, sin miedo a equivocarnos, se puede ver enormemente beneficiado por los enfoques cuantitativos y basados en técnicas de visualización, modelos que permiten observar los datos en su

conjunto para poder interpretar de forma más fundamentada las obras y fenómenos particulares.

2. Aportaciones de las HD al estudio del discurso en la obra de Dostoievski

2.1. En busca de patrones léxicos y discursivos

El estudio del léxico en Dostoievski encuentra en los enfoques cuantitativos y herramientas de análisis de contenido un gran aliado a la hora de validar las hipótesis teóricas sobre la existencia de patrones léxicos y discursivos en este autor a la hora de decir el mundo (cfr. Barros García, 2017). La arquitectura de la mayoría de las herramientas de HD para el análisis del texto literario está basada en el estudio de las palabras más frecuentes (MFW), por lo que se ajustan y pueden contribuir al logro de objetivos relativos a la localización de estructuras recurrentes en la obra de este autor.

Desde esta perspectiva y con este objetivo, comienzan las HD a postularse como un enfoque dentro de la crítica dostoievskiana, como más tarde lo harán las técnicas de visualización en cuanto herramientas corroborativas de observaciones previas no computarizadas. Uno de los principales focos de atención será el estudio de los elementos léxicos característicos de este autor (cfr. Barros García, 2012): entre otros, el uso excesivo de *вдруг*, traducido habitualmente por «de repente» al español y tema de debate en la literatura especializada desde principios del siglo XX, bien como rasgo particular de un estilo en el que todo sucede de repente (Terras, 2002: 134), bien como consecuencia de cierta impulsividad en la creación (Карякин, 1989: 645), o bien como reflejo de la visión del mundo de la lengua rusa (Арутюнова, 1999). De igual forma, se buscarán otros dos elementos léxicos constantes y especialmente relevantes en el discurso de este autor: *как бы* y *как будто*. Sin entrar aquí a discernir las diferencias a nivel cognitivo y de uso entre ambas, diremos que para Arutiunova (Арутюнова, 1999) las dos forman parte del conjunto de términos culturalmente específicos de la visión de la lengua rusa que sirven para marcar cierta distancia entre el sujeto de enunciación y el enunciado, razón por la que ambos, en un proceso de simplificación cuestionable, se suelen traducir al español por «como si».

El análisis cuantitativo de los datos corroboraba la excesiva presencia de este tipo de elementos en el corpus de textos de Dostoievski, y particularmente en los textos literarios frente a otros géneros cultivados por el autor (Barros García, 2012; Young, 2018). Asimismo, permite visualizar la distribución de estos elementos a lo

largo de cada obra o del corpus global de las obras completas del autor. Igualmente, revelaba información acerca de las colocaciones y otros datos centrados en las agrupaciones léxicas, todo ello desde la perspectiva de empleo de las visualizaciones como herramientas corroborativas o estadios finales de cierto análisis basado en hipótesis previas. No obstante, y como hemos defendido, los datos más interesantes derivaban del carácter exploratorio de las visualizaciones, es decir, eran resultado de ver aquello que de otra manera no se podía observar:

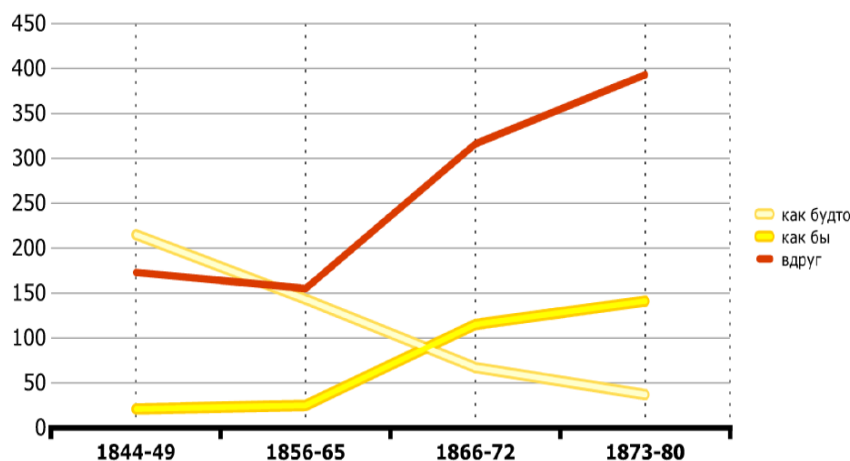


Fig. 1: Frecuencia de ocurrencia (en tantos por 100000) de *kak by* (amarillo intenso), *kak budto* (amarillo claro) y *vdrug* (rojo) en el corpus completo de textos literarios de Dostoievski. Véase Barros García (2012) para una interpretación.

De esta manera, la visualización permitía revelar un comportamiento especialmente atractivo de similitud entre *kak by* y *vdrug*, mientras que ponía de manifiesto un descenso significativo del uso de *kak budto* en la obra tardía del autor. A partir de estos datos, se trabajó con corpus literarios de la época y de otros autores para comprender si se trataba de un rasgo particular de Dostoievski o si existió realmente una tendencia hacia el sobreuso de unos elementos frente al paulatino desuso de otro (cfr. Barros García, 2012).

La visualización del peso de Dostoievski sobre el corpus general de textos literarios de la época en lo relativo al uso de estos elementos se puede realizar a nivel superficial y de forma sencilla incluso mediante herramientas propias del CNR o de Google n-gram viewer (Michel et al, 2010):

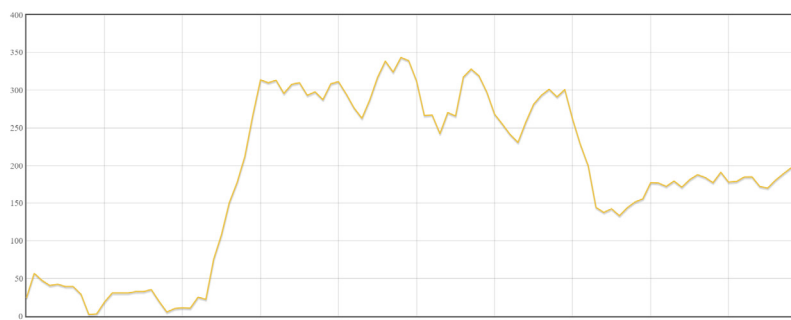


Fig. 2. Distribución de ocurrencias de *kak budto* en el corpus general de textos literarios rusos del CNR (1800-1899).



Fig. 3. Distribución de ocurrencias de *kak by* en el corpus general de textos literarios rusos del CNR (1800-1899).

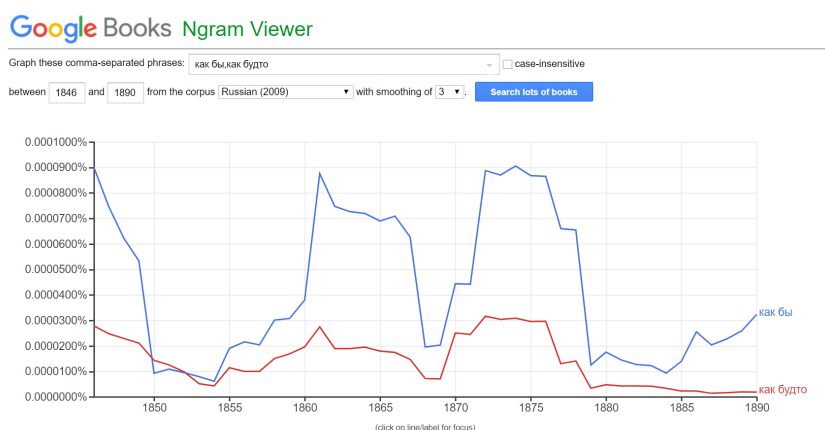


Fig. 4. Distribución de ocurrencias de *kak by* (azul) y *kak budto* (rojo) en Google Ngram Viewer (1800-1899)

Conviene advertir que aquí nos enfrentamos a un problema metodológico importante: si bien el hecho de usar este tipo de *software* (para una recapitulación de algunas herramientas, Neuendorf, 2017) o programas automatizados facilita el acceso y manejo de este tipo de herramientas, debemos ser conscientes de que con ellas se pierde el control de las transformaciones que sufren los datos y decisiones importantes que se deben tomar para su tratamiento (Goldstone, 2012), así como no olvidar los problemas específicos de diferente índole de ciertas herramientas como Google Ngram Viewer para el análisis cuantitativo de datos textuales (Jockers, 2013). Veremos más adelante otras formas de abordar los análisis basados en frecuencias o MFW.

Dentro del análisis de contenido basado en visualizaciones, encontramos otras búsquedas de elementos característicos del discurso de este autor, como pueden ser los fenómenos de corrección y reformulación (Barros García, 2017):

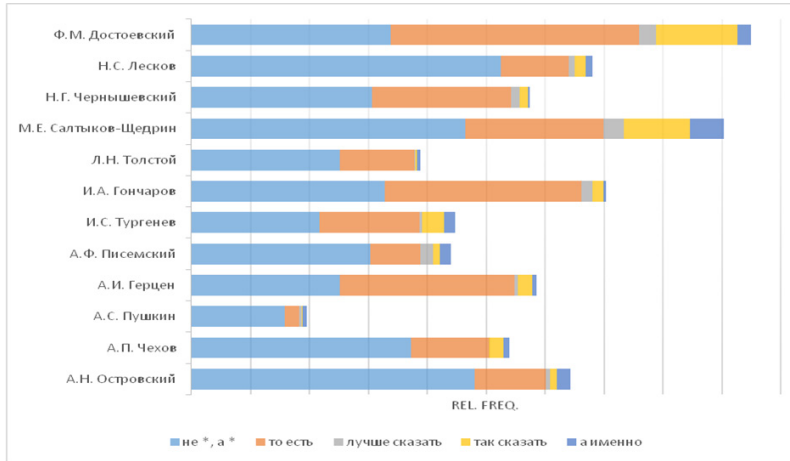


Fig. 5. Diferentes elementos de la reformulación y corrección (en la leyenda, de izquierda a derecha: no X, sino Y; es decir; mejor dicho; por así decirlo; sino precisamente) en la obra literaria de autores rusos: Dostoievski, Leskov, Chernyshevski, Saltykov-Schedrin, Tolstoy, Goncharov, Turgueniev, Pisemski, Hertzen, Pushkin, Chejov, Ostrovski (de arriba abajo). Datos basados en el CNR.

Dentro del ámbito léxico, no dejan de ser sugestivas las visualizaciones en forma de nube de palabras, si bien hay que tener en cuenta sus limitaciones y habituales sesgos asociados (Sinclair y CardewHall, 2008; McNaught, 2010; Barros García, 2015b). Aunque en este punto conviene matizar que si se trabaja con un corpus propio (en este caso basado en los textos literarios de Dostoievski), el uso de *software* o herramientas automatizadas pierde potencial frente a otras opciones como R o Python, bien porque ese *software* no esté optimizado para el análisis del alfabeto cirílico (por ejemplo, en la lematización), bien porque suele implicar un control mucho más limitado de los datos de origen.

Más allá de querer descubrir cuáles son las palabras más o menos frecuentes en las obras de Dostoievski, podemos llevar el análisis a dimensiones difícilmente observables mediante enfoques no cuantitativos y visuales. Por ejemplo, podemos preguntarnos acerca de las palabras exclusivas de cada uno de los personajes, cuya representación en R es bastante sencilla. En particular y por motivos obvios a la trama, parece interesante indagar qué palabras dirá Sonia que nunca pronunciará Raskolnikov, a pesar de que el peso a nivel de discurso en el texto de este es marcadamente superior al de ella:



Fig. 6. Nube de las palabras que usa Sonia y no usa Raskolnikov a lo largo de toda la obra *Crimen y castigo*. Elaborada con www.wordle.net.

En este caso hemos optado por no tokenizar las palabras porque nos interesa precisamente ver cada una de las formas exclusivas de Sonia. El etiquetado del texto para extraer los discursos de los diferentes personajes se realizó con Atlas.ti (para una descripción: Barros García, 2012). Llama, entre otras, la atención la palabra *жалко*, concepto relacionado con el campo semántico de la *toska* (normalmente traducido como *anguish* en inglés), siendo este rasgo distintivo a nivel discursivo de ciertos personajes capitales (Martinsen, 2018). Las visualizaciones revelan, de nuevo, un patrón discursivo de gran relevancia. Aparte de plantearse qué personajes verbalizaban este concepto en toda la obra del autor, cabría preguntarse con qué elementos suele coocurrir en los diferentes textos:

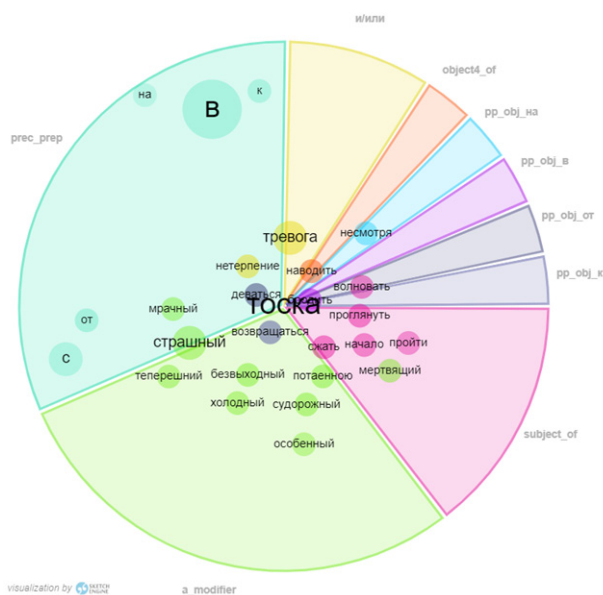


Fig. 7. Colocaciones de *toska* en la obra *Crimen y castigo*, etiquetadas a su vez por partes de la oración. Elaborada con Sketch Engine.

En este caso hemos usado otra de las herramientas más difundidas y fáciles de manejar para manejo de corpus y visualización, Sketch Engine (Kilgarriff et al., 2014; Kilgarriff et al., 2004), disponible en www.sketchengine.eu. No tenemos la intención aquí de dilatar esta línea de análisis, pero creemos que queda claro su potencial tanto a nivel contrastivo intratextual como intertextual. Esta misma herramienta nos puede ayudar a generar una visualización de los verbos que suelen coocurrir con los elementos anteriormente mencionados *kak budto* y *kak by*, que aparecerán escritos como una sola palabra precisamente para evitar que la herramienta incluya coocurrencias (por ejemplo, inducidas por la eliminación de los signos de puntuación o palabras frecuentes) de los dos elementos de cada una de esas partículas en otros usos o funciones:



Fig. 8. Colocaciones de *kak budto* y *kak by* con verbos en *Crimen y castigo*. Elaborada con Sketch Engine. Se permite observar con qué verbos suelen coocurrir estos elementos, así como el hecho de que solo comparten un verbo en las colocaciones, algo que llama la atención al ser elementos tan cercanos a nivel semántico.

Esta tipología de análisis cuantitativo de datos literarios puede llevarse al ámbito de los estudios comparados, por ejemplo, con la intención de ver cómo se comportan diferentes elementos léxicos en varias obras. En esta línea es frecuente recurrir al Term Frequency and Inverse Document Frequency (tf-idf) como método para cuantificar la importancia que tienen diferentes elementos dentro de un documento que es parte de una colección y para el que en R existen paquetes de gran eficacia como tidytext (Silge y Robinson, 2016):

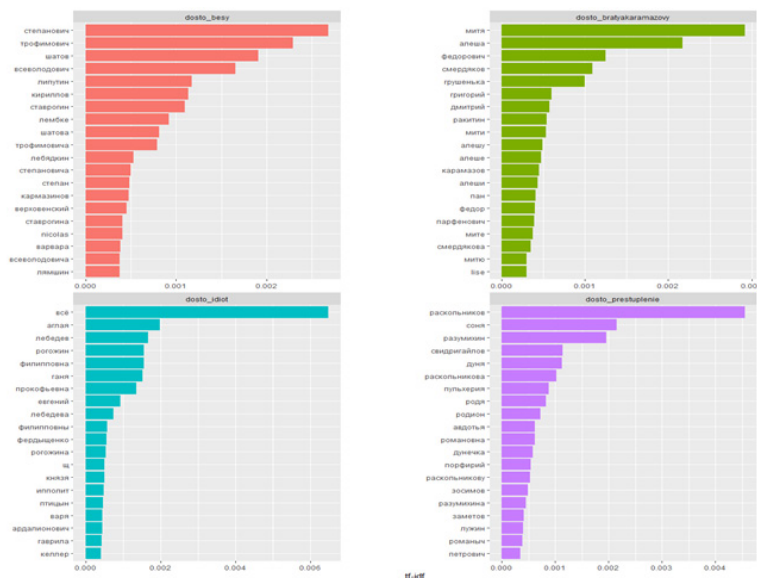


Fig. 10. Valores más altos de tf-idf para las obras *Los demonios*, *Los hermanos Karamazov*, *El idiota* y *Crimen y castigo* (de izquierda a derecha y de arriba abajo). Elaborados en R.

Estas visualizaciones permiten, además, observar la preocupante tendencia de algunos traductores de eliminar elementos repetidos en el texto origen, algo discutible sobre todo si tenemos en cuenta que estas repeticiones parecen no ser fruto de deslices o inconsistencias

en el proceso creador ni tampoco de hábitos en la lengua origen, por lo que respetarlas podría ser interesante desde un punto de vista filológico, sobre todo en lo que concierne a elementos característicos del estilo de este autor (cfr. Barros García, 2012; Barros García 2017). A continuación, se muestran las palabras con valores más altos de tf-idf en el texto original de *El idiota* y en una de sus traducciones más citadas al inglés:

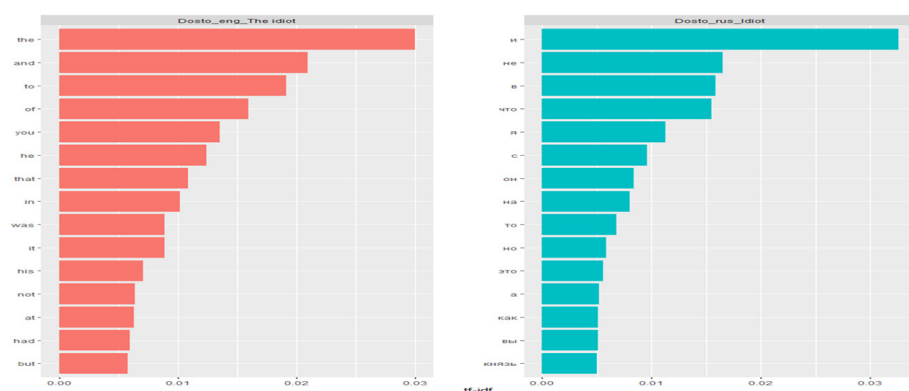


Fig. 11. Valores más altos de tf-idf para *El idiota* en traducción al inglés de Eva Martin y la obra original en ruso. Elaborados en R.

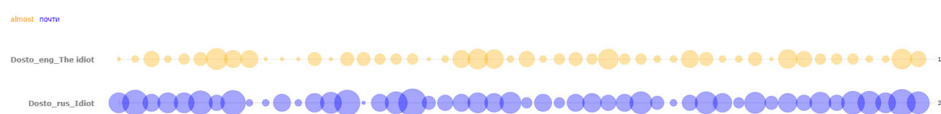


Fig. 12. Distribución de ocurrencias por orden de lectura en la traducción al inglés de *El idiota* (arriba) y en el original (abajo). Elaborada con Voyant Tools (Sinclair y Rockwell, 2012).

Si bien es cierto que cabe cuestionarse cuán natural es la excesiva repetición de ciertos elementos en la cultura meta, también se debe hacer lo mismo con respecto a la cultura de origen y, en caso de que en esta también sea poco habitual, entonces tal vez convenga mantener la repetición como rasgo identificativo y característico del autor.

En este apartado hemos pretendido resumir algunas de las líneas básicas de análisis cuantitativo y basado en visualizaciones que se pueden llevar a cabo en textos literarios. Para las visualizaciones hemos empleado herramientas populares y fácilmente disponibles, a pesar de que con ellas se pierde cierto control sobre los datos, con la intención de concienciar de que es posible acercarse a estos enfoques sin necesidad de tener conocimientos de programación, aunque estos, sin duda, ayudan a poder profundizar en el control y análisis de los datos.

2.2. En busca del autor

Siguiendo la línea marcada por Fradejas Rueda (2016, 2019) y Boto Bravo (2017), cabe preguntarse cuáles son las aportaciones que la estilometría puede hacer al análisis del texto literario. Si bien la estilometría con frecuencia se centra en los estudios de autoría, el hecho de que se base en la detección automatizada de similitudes estilísticas entre textos, puede llevarla a trascender esa función. En los ejemplos que siguen se ha usado el paquete de R *stylo* (Eder, Rybicki y Kestemont, 2016) precisamente para observar, entre otras cuestiones, similitudes y diferencias entre obras de Dostoievski o entre capítulos de una misma obra.

En los estudios sobre Dostoievski encontramos enconados debates en torno al punto de inflexión que parece suponer su condena a trabajos forzados a nivel ideológico y narratológico. Además, tenemos otro factor importante de que algunas obras fuesen dictadas, otras escritas con una presión muy intensa para llegar a la entrega a tiempo, otras que pudieran haber sido censuradas por los editores de las diferentes revistas donde iban a ser publicadas, y un largo etcétera de sucesos que hacen especialmente atractivo el hecho de comparar hasta qué punto guardan o no similitudes los textos de Dostoievski entre sí.

Partiremos de un análisis de *clusters* (Fig. 13) con el fin de agrupar los textos literarios de Dostoievski «de manera que se dé la máxima homogeneidad posible dentro de cada grupo y, a la vez, la mayor diferencia entre los diversos grupos» (Fradejas Rueda, 2016).

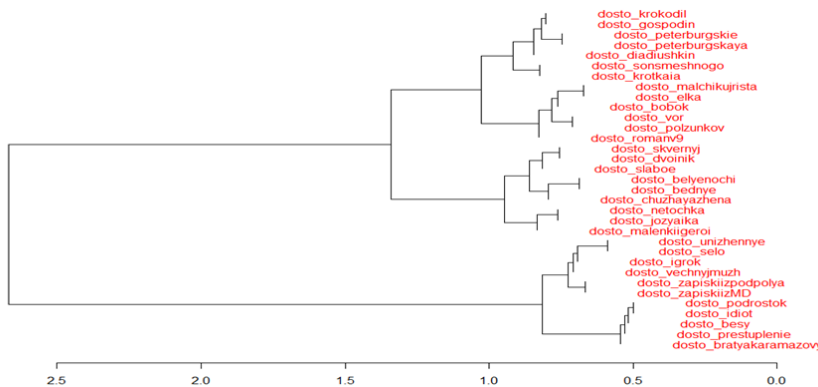


Fig. 13. Dendrograma con 5000 MFW y distancia Classic Delta, tomando palabras completas por N-grams. Elaborado con *stylo*.

El resultado no deja de ser llamativo, pues se han agrupado las denominadas cinco grandes novelas de Dostoievski (*El adolescente*, *El idiota*, *Los demonios*, *Crimen y castigo* y *Los hermanos Karamazov*) en el subgrupo inferior, al mismo tiempo que parece acertar en el reconocimiento de similitudes entre obras tempranas, con carácter

epistolar, relatos fantásticos, etc. No entraremos aquí a profundizar sobre las implicaciones de este agrupamiento, pero sí queremos comentar que tiene bastante sentido tanto a nivel narratológico y estilístico como temático y cronológico.

Es cierto que hemos optado por hacer el análisis sobre un valor muy alto de palabras frecuentes (5000), pero la agrupación de las cinco grandes novelas se observa incluso en valores de MFW de 100-300:

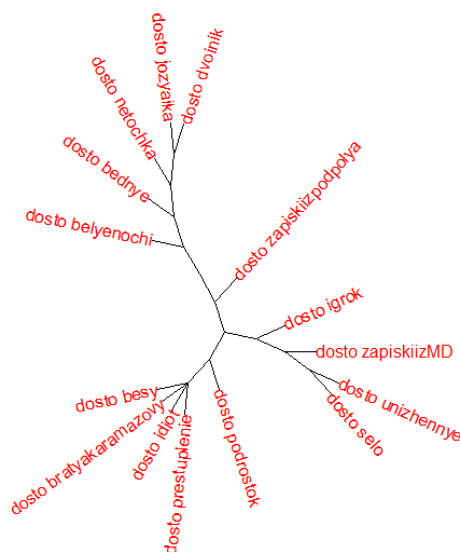


Fig. 14. Árbol de consenso con 100-300 3-grams MFW y distancia Classic delta.

De nuevo vemos en este caso, realizado sobre MFW 3-grams, que las cinco grandes novelas siguen estando agrupadas. Además, se atisba una distinción entre las obras anteriores y posteriores a *Memorias del subsuelo*, lo que corroboraría que es precisamente en esta obra donde sucede cierto punto de inflexión estilístico e ideológico en el autor en parte motivado por la condena a trabajos forzados.

Este tipo de enfoque se podría aplicar, aceptando las limitaciones que implica manejar un corpus de extensión muy limitada (Hoover, 2019) al trabajar con y sobre una sola obra, a una cuestión de enorme interés dentro de la crítica sobre *Crimen y castigo*: la función, pertinencia y calidad del Epílogo (cfr. en español: Barros García, 2012). Este es el resultado obtenido mediante análisis de componentes con el fin de observar las similitudes entre los diferentes capítulos de la obra *Crimen y castigo*:

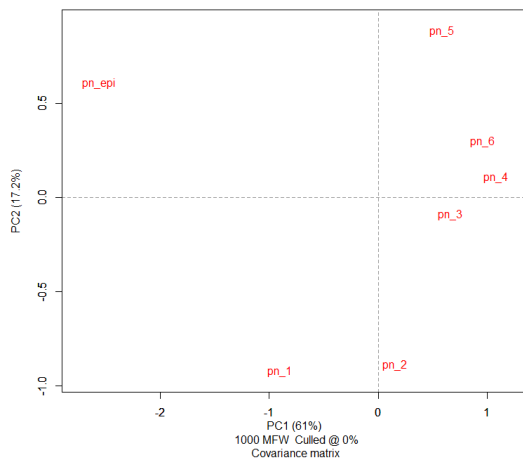


Fig. 15. Dendrograma de análisis de componentes principales (PCA) con distancia matriz de covarianza con 1000 MFW. Elaborado con R.

Si bien asumimos las limitaciones comentadas, sobre todo relativas al volumen de palabras en cada parte y al bajo número de palabras totales del corpus analizado (solo una novela), creemos que en esta línea sería muy interesante continuar y completar la interpretación del aparente *aislamiento* obtenido mediante este PCA del Epílogo con respecto al resto de la obra.

Cabe advertir aquí que todos estos análisis estilométricos se pueden beneficiar enormemente de las herramientas y técnicas de análisis de redes, como veremos a continuación.

2.3. En busca de conexiones y redes

Si bien con anterioridad al presente siglo encontramos algunas propuestas basadas en el análisis del texto literario mediante diagramas (cfr. Paranyushkin, 2011), es en estos últimos 15 años, y probablemente acentuado con la sistematización de Moretti (2000, 2005, 2011, 2013) y la profusión de *software* de visualización de corpus, cuando se aprecia un mayor número de publicaciones que intentan, sobre todo, establecer las redes de personajes basadas en su interacción comunicativa mediante, entre otros, gráficos *who-talks-to-whom* (Cherven, 2015: 9).

Como hemos apuntado, Moretti (2011) sistematiza un modelo de análisis literario basado en redes que tendrá enorme influencia. Paulatinamente hemos atendido a la consolidación del uso de Gephi (Bastian et al., 2009) para el análisis de redes en los estudios cuantitativos del texto literario, y a la posterior aparición del algoritmo implementado ForceAtlas2 (Jacomy et al., 2014), en el que se basará la mayoría de estudios sobre diferentes escritores y corpus de textos literarios (Levine, 2015). Asimismo, surge el interesante debate en torno a la función y sentido de las redes en la cultura contemporánea (Jagoda, 2016).

En el caso concreto de Dostoievski, uno de los primeros análisis que se hicieron en esta línea mostraba las interrelaciones discursivas

entre los personajes de la novela *Crimen y castigo* (Barros García, 2012, 2015b):

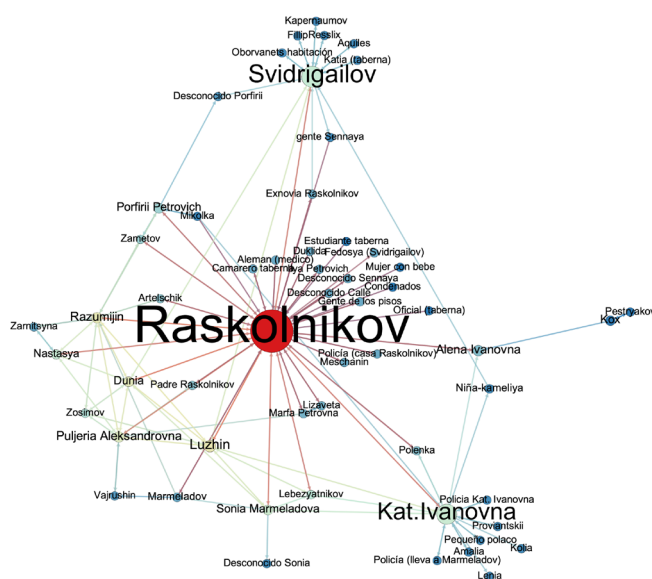


Fig. 16. Red de relaciones discursivas de los personajes de *Crimen y castigo*, incluidas las no bidireccionales (un personaje le habla a otro, pero no obtiene respuesta). Adaptado de Barros García (2015b)

Estos datos ofrecían posibilidades impensables desde otros enfoques como, por ejemplo, analizar los índices de centralidad e intermediación para descubrir que la vieja usurera, Aliona Ivánovna, a pesar de su ínfima presencia discursiva en el texto, muestra un índice de intermediación que la convierte en un eje fundamental de la trama (Barros García, 2012). Recientemente Chloë Kitzinger (2020) ha implementado el análisis y la visualización de la red de los personajes de esta misma obra, consiguiendo mapear de forma más certera las relaciones entre personajes, llegando a incluir incluso las que se sobreentienden (rumores, por ejemplo). Análisis parecidos se han realizado con otros autores de la literatura rusa (Ilchuk, 2017).

Otro objeto de estudio recurrente es la visualización de redes de relaciones entre obras o autores. En este sentido, debemos mencionar la propuesta de Kelley et al. (2009) para el estudio de la influencia de unos autores sobre otros, comentando, solo tangencialmente, que Dostoievski es uno de los autores más influyentes de todo el corpus manejado de literatura universal. En este caso, las relaciones se establecían mediante la declaración o testimonio explícito de un autor de haber sido influido por otro autor.

Pero volvamos ahora un poco hacia atrás para mostrar el potencial obtenido en esta línea de investigación al combinar dos métodos descritos en este trabajo: la estilometría y el análisis de redes:

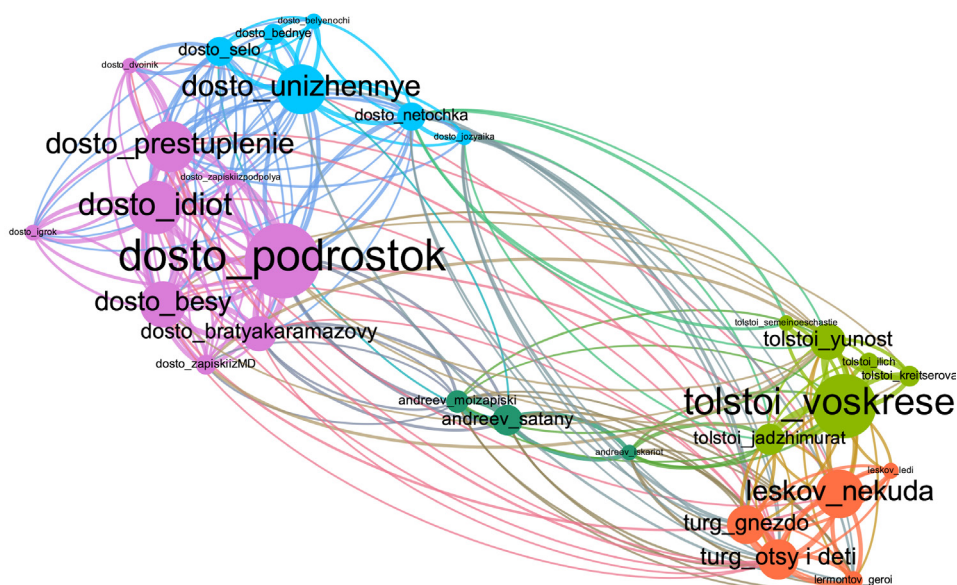


Fig. 17. Comunidades localizadas para un corpus de 27 obras de Dostoievski, Tolstoi, Turgueniev, Leskov, Lermontov y Andreiev.

Hemos realizado esta red como ejemplo de lo que podría suponer una investigación de mayor envergadura. En la Fig. 17 se han visualizado las similitudes localizadas mediante análisis de agrupamiento (100-3000 MFW, distancia Classic delta). Los resultados obtenidos se han exportado a Gephi y han sido distribuidos por ForceAtlas2. Los colores marcan las comunidades formadas, siendo todas significativas estadísticamente, y estando mediadas por la comunidad intermediaria que suponen los textos de Andreiev. Los tamaños de los nodos se corresponden al grado ponderado con los pesos de las aristas. Vemos que el agrupamiento detectado coincide plenamente con la autoría y apunta a la potencial validación de las hipótesis sobre el estilo e influencia entre estos autores, al mismo tiempo que incita a la realización de sucesivas implementaciones del ejemplo aquí propuesto.

3. Conclusiones

Es probable que el concepto de dato en los estudios literarios siga despertando algunas reacciones adversas. Pero es igualmente probable que los enfoques cuantitativos y basados en visualizaciones generen un conocimiento de enorme relevancia en el análisis de textos literarios.

A lo largo de estas páginas hemos partido del debate existente en torno a esta y otras cuestiones relativas al posicionamiento de las HD en el análisis del discurso y el estudio de la literatura (cfr. Allés

Torrent, 2019) con el fin de ejemplificar la pertinencia de enfoques basados en métodos de visualización de la información y análisis de redes en el estudio de la obra de F. M. Dostoievski. El objetivo principal ha sido mostrar que las visualizaciones son mucho más que formas estáticas y estadios últimos de representación de datos, pudiendo llegar a ser concebidas como procesos de investigación en sí mismas (Hinrichs, Forlini y Moyiniham, 2019) y que muestran un enorme potencial para el estudio de textos literarios concretos, así como grandes corpus de obras. En particular, hemos comprobado algunas posibilidades que ofrecen para el estudio de las frecuencias, desde la caracterización discursiva y léxica de personajes, obras, autores, hasta los análisis de similitudes en el estilo entre partes de una misma obra. También hemos mostrado algún ejemplo de los interesantes resultados que se obtienen al combinar análisis estilométricos con el análisis de redes, en particular para detectar similitudes o distancias entre autores y obras de un corpus de 27 obras de Dostoievski, Tolstoi, Turgueniev, Leskov, Lermontov y Andreiev.

Algunas de las limitaciones de la metodología seguida han sido valoradas a lo largo del texto, como el tamaño y equilibrio del corpus, la simplificación de ciertos procesos de minería de los datos y análisis (sobre todo con las herramientas automatizadas) o haber trabajado en los ejemplos únicamente con las palabras como unidades de análisis. En este sentido, conviene destacar que las HD están demostrando una virtud interesante que define su compromiso con la construcción de la ciencia: se tienen en cuenta las carencias detectadas. Quizá por su juventud, quizá por su naturaleza, parece haber y fomentarse una mayor tendencia hacia el continuo debate y revisión que en otros ámbitos. Probablemente esto tenga también que ver con que los resultados normalmente son concebidos como posibles puntos de partida para otras futuras investigaciones.

En el caso concreto de la imbricación de las visualizaciones y análisis cuantitativos y los estudios literarios, es esperable que pronto se consoliden algunas propuestas dentro de lo que supone uno de los grandes retos de las HD en este ámbito: la representación de los niveles de narración, estructura, etc. (Schwan et al., 2019). En el caso de Dostoievski, en cuyas obras la complejidad estructural y narratológica ha sido puesta de manifiesto en innumerables ocasiones, a buen seguro ayudarán a arrojar luz sobre, entre otras cuestiones, cómo se construyen las múltiples dimensiones o niveles (Neuhäuser, 1986) que observamos en sus obras y, en suma, sobre cómo se construye la dinámica de la experiencia lectora (Belknap, 2016). De momento conocemos la propuesta de Culpepper (2019), probablemente basada en la línea ya apuntada por Jockers (2013) acerca de la representación de sucesos en el tiempo narrativo con el paquete de R Syuzhet.

Otro tipo de análisis cuantitativos de la obra de Dostoievski conllevaría implementar las herramientas, paquetes y recursos de los que disponemos actualmente, algo que nos parece especialmente necesario dado el potencial demostrado por este tipo de enfoques. Algunas de estas implementaciones tienen que ver con las características intrínsecas de la lengua rusa (lematización, análisis de sentimientos, etc.), actualmente solo abordables con aprendizaje automático, lo que dificulta que puedan ser utilizadas sin conocimientos de programación.

Bibliografía citada

- ALLÉS TORRENT, S. (2019): Sobre la complejidad de los datos en Humanidades, o cómo traducir las ideas a datos, *Revista de Humanidades Digitales*, 4, 1-28, <<https://doi.org/10.5944/rhd.vol.4.2019.24679>>, [23/05/2020].
- BAGNO, Vs. E. (2013-2019): *Достоевский Ф. М. Полное собрание сочинений и писем: В 35 томах*, СПб.: Наука.
- BARROS GARCÍA, B. (2012): «Análisis y visualización del discurso en *Crimen y castigo* de F. M. Dostoevski» (tesis doctoral), Granada: Universidad de Granada, <<https://digibug.ugr.es/handle/10481/20265>>, [01/12/2019].
- BARROS GARCÍA, B. (2015a): «En torno a la función del personaje en la ficción literaria», *452ºF. Revista de Teoría de la Literatura y Literatura Comparada*, 13, 187-203, <https://www.452f.com/pdf/numero13/13_452f_Barros_orgnl.pdf>, [11/12/2019].
- BARROS GARCÍA, B. (2015b): «Las aportaciones de la visualización del texto al estudio del texto literario», *Tonos Digital*, 28, <<http://www.tonosdigital.es/ojs/index.php/tonos/article/viewFile/1210/738>>, [01/12/2019].
- BARROS GARCÍA, B. (2017): «In Other Words: Reformulation Strategies in Dostoevsky's works», *Russian Literature*, 91, 1-25, <<https://doi.org/10.1016/j.ruslit.2017.09.001>>, [11/12/2019].
- BASTIAN, M., HEYMANN, S. y JACOMY, M. (2009): «Gephi: an open source software for exploring and manipulating networks», International AAAI Conference on Weblogs and Social Media, <<http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>>, [12/12/2019].
- BAZANOV, V. G. (ed.) (1972-1990): *Достоевский Ф. М. Полное собрание сочинений в тридцати томах*, Ленинград: Наука.
- BELKNAP, R. (2016): *Plots*, Nueva York: Columbia University Press.
- BORGMAN, C. L. (2009): «The Digital Future is Now: A Call to Action for the Humanities», *Digital Humanities Quarterly*, 3, <<http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html%20/000077.html>>, [12/12/2019].
- BOTO BRAVO, M. A. (2017): «Mapa estilométrico de la narrativa de Eduardo Mendoza: aproximación a un análisis estilístico computacional de textos literarios», *EPOS*, XXXIII, 99-114.
- BURKHARD, R. A. (2005): «Towards a Framework and a Model for Knowledge Visualization: Synergies Between Information and Knowledge Visualization» en Tergan, S. O. y Keller, T. (eds.), *Knowledge and Information Visualization. Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer, 238-255.
- BURROWS, J. F. (1987): *Computation into criticism: a study of Jane Austen's novels and an experiment in method*, Oxford: Clarendon Press.
- CHERVEN, K. (2015): *Mastering Gephi network visualization*, Birmingham: Packt Publishing Ltd.
- CONTEH, A., y WILMS, L. (2013): «“Where do you need us?” — The National Library in the Digital Humanities», *Journal of Digital Humanities*, 3, 1, <<http://journalofdigitalhumanities.org/3-1/where-do-you-need-us-the-national-library-in-the-digital-humanities/>>, [10/11/2019].
- CULPEPPER, A. J. (2019): «Visually Mapping the Narrative System of Dostoevsky's *The Idiot*». <https://trace.tennessee.edu/utk_eureca/2019/artsscience/14/>, [22/01/2020].
- CUMMINGS, J. (2019): «A world of difference: Myths and misconceptions about the TEI», *Digital Scholarship in the Humanities*, 34, 1, <<https://doi.org/10.1093/llc/fqy071>> [30/01/20].
- EDER, M., RYBICKI, J. y KESTEMONT, M. (2016): «Stylometry with R: a package for computational text analysis», *R Journal*, 8, 1, 107-121.
- EGAN, G. (2019): «Introduction to a special section on “Computational Methods for Literary–Historical Textual Scholarship”», *Digital Scholarship in the Humanities*, 34 (4), 818–824. <<https://doi.org/10.1093/llc/fqz071>>, [01/01/20].
- FRADEJAS RUEDA, J. M. (2016): «El análisis estilométrico aplicado a la literatura española: las novelas policíacas e históricas», *Caracteres*, <<http://revistacaracteres.net/revista/vol5n2noviembre2016/analisis-estilometrico/#footnote-20>>, [01/12/2019].
- FRADEJAS RUEDA, J. M. (2019): «Alfonso X y la estilometría», <7partidas.hypotheses.org/1072>, [21/12/2020].
- GAVIN, M. (2019): «Is there a text in my data? (Part 1): on counting words», *Journal of Cultural Analytics*, <<https://culturalanalytics.org/article/11830-is-there-a-text-in-my-data-part-1-on-counting-words>>, [12/01/2020].

- GOLD, M. K. y KLEIN, L. F. (2012): «A DH that matters» en Gold, M. K. (ed.), *Debates in the Digital Humanities*, Minneapolis: University of Minnesota Press, <<https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/0cd11777-7d1b-4f2c-8fdf-4704e827c2c2>>, [17/01/20].
- GOLD, M. K. (ed.) (2012): *Debates in the Digital Humanities*, Minneapolis: University of Minnesota Press, <<https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities>>, [18/01/2020].
- GOLDSTONE, A. (2012): «Teaching Quantitative Methods: What Makes It Hard (in Literary Studies)» en Gold, M. K. (ed.), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, <<https://dhdebates.gc.cuny.edu/read/4805e692-0823-4073-b431-5a684250a82d/section/620caf9f-08a8-485e-a496-51400296ebcd#en164>>, [02/01/20].
- HINRICHS, U., FORLINI, S., MOYNIHAN, B. (2019): «In defense of sandcastles: Research thinking through visualization in digital humanities», *Digital Scholarship in the Humanities*, 34, 1, <doi.org/10.1093/llc/fqy051>, [10/01/2020].
- HOOVER, D. L. (2019): «Simulations and difficult problems», *Digital Scholarship in the Humanities*, 34, 4, <<https://doi.org/10.1093/llc/fqz034>>, [22/01/2020].
- ILCHUK, Yu. (2017): «*The Labyrinth of Linkages' Decoded: Character Networks in Anna Karenina*», 2017 AATSEEL Conference, San Francisco.
- JACOMY, M., VENTURINI, T., HEYMANN, S. y BASTIAN, M. (2014): «ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software», *PLoS ONE*, 9, 6, e98679, <<https://doi.org/10.1371/journal.pone.0098679>>, [17/01/2020].
- JAGODA, P. (2016). *Network Aesthetics*, Chicago/Londres: University of Chicago Press.
- JOCKERS, M. L. (2013): *Macroanalysis: Digital Methods and Literary History*, Urbana: University of Illinois Press.
- JONES, E. y NULTY, P. (2019): «Quantitative measures of lexical complexity in modern prose fiction», *Digital Scholarship in the Humanities*, 34, 4, 914-937.
- KAPLAN, F. (2015): «A map for big data research in digital humanities», *Front. Digit. Humanit.*, 2, 1, <<https://www.frontiersin.org/articles/10.3389/fdigh.2015.00001/full>>, [16/01/2020].
- KELLEY, S., GOLDBERG, M., MAGDON-ISMAIL, M, MERTSALOV, K., WALLACE, W. y ZAKI, M. (2009): «graphOnt: An ontology based library for conversion from semantic graphs to JUNG» en 2009 *IEEE International Conference on Intelligence and Security Informatics*, Dallas, <ieeexplore.ieee.org/document/5137292>, [20/01/2020].
- KILGARRIFF, A., BAISA, V., BUŠTA, J., JAKUBÍČEK, M., KOVÁŘ, V., MICHELFEIT, J., RYCHLÝ, P. y SUCHOMEL, V. (2014): «The Sketch Engine: ten years on». *Lexicography*, 1, 7-36.
- KILGARRIFF, A., RYCHLÝ, P., SMRŽ, P. y TUGWELL, D. (2004): «Itri-04-08 the sketch engine», *Information Technology*, <www.sketchengine.eu/wp-content/uploads/The_Sketch_Engine_2004.pdf>, [01/11/2020].
- KITZINGER, Ch. (2020, en prensa): «Mapping the Networks of Crime and Punishment» in Katz, M. y Burry, A. (eds.), *Approaches to Teaching Dostoevsky's Crime and Punishment*.
- LEVINE, C. (2015): *Forms: Whole, Rhythm, Hierarchy, Network*, Princeton: Princeton University Press.
- MARTINSEN, D. (2018): «*Anguish in Crime and Punishment*», *Достоевский и мировая культура. Филологический журнал*, 1, 59-68.
- MICHEL, J-B., LIEBERMAN-AIDEN, E. et al. (2010): «Quantitative Analysis of Culture Using Millions of Digitized Books», *Science*, 331, 6014, 176-182.
- MORETTI, F. (2000): «Conjectures on World Literature», *New Left Review*, 1, <newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>, [12/12/2019].
- MORETTI, F. (2005): *Graphs, Maps, Trees: Abstract Models for a Literary History*, Londres/Nueva York: Verso.
- MORETTI, F. (2011): «Network Theory, Plot Analysis», Stanford Literary Lab, <litlab.stanford.edu/LiteraryLabPamphlet2.pdf>, [12/12/2020].
- MORETTI, F. (2013): *Distant Reading*, Londres/Nueva York: Verso.
- NEUENDORF, K. A. (2017): *The Content Analysis Guidebook*, Los Angeles: SAGE.
- NEUHÄUSER, R. (1986): «A Contemporary Reading of Book VI, "A Russian Monk"», *Dostoevsky Studies*, 7, 135-51.
- OWENS, T. (2011): «Defining Data for Humanists: Text, Artifact, Information or Evidence?», *Journal of Digital Humanities*, 1, 1, <<http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/>>, [12/12/2020].

- PARANYUSHKIN, D. (2019): «Identifying the pathways for meaning circulation using text network analysis», *Nodus Labs*, <<https://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis>>, [10/01/2020].
- ROMERO FRÍAS, E. y BOCANEGRA BARBECHO, L. (eds.) (2018): *Ciencias Sociales y Humanidades Digitales Aplicadas: casos de estudio y perspectivas críticas*, Granada/Nueva York: Universidad de Granada/Downhill Publishing.
- ROTH, C. (2019): «Digital, digitized, and numerical humanities, Digital Scholarship in the Humanities», 34, 3, <<https://doi.org/10.1093/llc/fqy057>>, [11/01/2020].
- SCHÖCH, C. (2013): «Big? Smart? Clean? Messy? Data in the Humanities», *Journal of Digital Humanities*, 2, 3, <<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities>>, [15/11/2019].
- SILGE, J. y ROBINSON, D. (2019): *Text Mining with R. A Tidy Approach* <www.tidytextmining.com>, [26/01/2020].
- SINCLAIR, S. y ROCKWELL, G. (2012): *Voyant Tools*, <voyant-tools.org>, [30/01/2020].
- SPIRO, L. (2012): «“This is Why We Fight”: defining values of the digital humanities» en Gold, M. K. (ed.), *Debates in the Digital Humanities*, Minneapolis: University of Minnesota Press, <<https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbdde/section/9e014167-c688-43ab-8b12-0f6746095335>>, [30/01/2020].
- SULA, Ch, A. y HILL, H. V. (2019): «The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004)», *Digital Scholarship in the Humanities*, 34, 1, <<https://doi.org/10.1093/llc/fqz072>>, [02/01/2020].
- TERRAS, M, NYHAN, J. y VANHOUTTE, E. (eds.) (2013): *Defining Digital Humanities: A Reader*, Burlington, VT y Farnham: Ashgate Publishing Ltd.
- TERRAS, V. (2002): *A Karamazov companion: Commentary on the genesis, language and style of Dostoevsky’s novel*, Madison: University of Wisconsin.
- YOUNG, S. J. (2018): «Hesitation, Projection and Desire: The Fictionalizing “as if...” in Dostoevskii’s Early Works», *Modern Languages Open*, 1, 15, <<http://doi.org/10.3828/mlo.v0i0.183>>, [10/10/2020].
- ZAKHAROV, V. (2019): «Проблемы и перспективы развития проекта “Digital Dostoevsky”», *XVII Symposium of the International Dostoevsky Society*, Boston: Boston University.
- АРУТЮНОВА, Н. Д. (1999): *Язык и мир человек*, М.: Языки русской культуры.
- КАРАУЛОВ, Ю. Н. (2017): *Словарь языка Достоевского, Идиоглоссарий*. Н–По / Российская академия наук. Ин-т рус. яз. им. В. В. Виноградова, М.: Азбуковник.
- КАРЯКИН Ю. Ф. (1989): *Достоевский и канун XXI века*, М.: Советский писатель.
- КИБАЛЬНИК, С.А. (2018): «Международный научно-исследовательский семинар «Проблемы цифровой текстологии и атрибуции анонимных текстов»», *Русская литература*, 1, 259-261.