

Una aproximación a la escritura de Miguel Delibes desde la estilometría

María Pilar Celma Valero

Universidad de Valladolid. Facultad de Filosofía y Letras. Departamento de
Literatura española y Teoría de la Literatura. Valladolid. España

pilar@fyl.uva.es

Cristina Ruiz Urbón

Universidad de Valladolid. Facultad de Filosofía y Letras. Departamento de
Literatura española y Teoría de la Literatura. Valladolid. España

cristina.ruiz.urbon@uva.es

An approach to Miguel Delibes's writing from stylometry

Fecha de recepción: 8.03.2021 / Fecha de aceptación: 11.06.2021

Tonos Digital, 41, 2021 (II)

RESUMEN: De un tiempo a esta parte, la estilometría ha ofrecido grandes resultados en el establecimiento del perfil lingüístico de una determinada persona, la atribución de autoría de un texto anónimo o dubitado, la detección de plagio o el análisis de sentimientos, gracias a la posibilidad de cuantificar de manera rápida y objetiva rasgos mensurables como la frecuencia léxica, la media de palabras por frase, la ratio *type/token* o la utilización de vocabulario relacionado con ciertos estados afectivos. En el presente artículo pretendemos profundizar en otra de las aplicaciones de los análisis cuantitativos, menos estudiada pero enormemente interesante, que guarda relación con la posibilidad de determinar la evolución estilística de un autor a lo largo del tiempo (variación intrapersonal), tomando como referencia la obra de Miguel Delibes. Así, en las siguientes páginas estudiaremos cómo el estilo delibesiano va cambiando de manera paulatina con el paso de los años y cómo esa variación se manifiesta de manera especialmente significativa en *Los santos inocentes*, una novela publicada en 1981 pero cuya génesis compositiva es de finales de los años 60.

Palabras clave: Delibes; estilometría; análisis cuantitativos; periodización; evolución estilística.

ABSTRACT: For some time now, stylometry has offered great results in establishing the linguistic profile of a given individual, the authorship attribution of an anonymous text or a text with a dubious author, the detection of plagiarism, or the analysis of feelings. This is possible thanks to the possibility of quickly and objectively quantifying measurable traits such as lexical frequency, average number of words per sentence, type/token ratio or verbal expression of emotions. In this article we intend to delve into another one of the applications of quantitative analysis, less studied but enormously interesting, which is related to the possibility of determining the stylistic evolution of an author over time (intrapersonal variation), taking the work of Miguel Delibes as a reference. Thus, in the following pages we will study how the delibesian style gradually changes over the years and how this variation manifests itself in an especially significant way in *Los santos inocentes*, a novel published in 1981 but whose compositional genesis dates back to the end of the 60s.

Keywords: Delibes; stylometry; quantitative analysis; periodization; stylistic evolution.

Al ocuparnos de autores con una larga trayectoria literaria, como es el caso de Miguel Delibes, hay cuestiones que era imposible plantear antes de que la estilometría abriera caminos para el tratamiento de "big data". Los análisis cuantitativos, como el que aquí presentamos, han dado ya grandes resultados en el campo de la psicología (Cohn, Mehl, y Pennebaker, 2004), de la política (Hart, 1984) y también están empezando a cobrar una dimensión relevante en los estudios literarios en general (Stone, Dunphy, Smith y Ogilvie, 1966).

Es habitual en los estudios literarios la periodización de la obra de los autores longevos por etapas o la clasificación de su producción por géneros. Sin embargo, era realmente difícil traducir a cifras exactas las variables lingüístico-literarias sobre las que sustentar las susodichas periodizaciones o clasificaciones (Alpers, Winzelberg, Classen, Roberts, Dev, Koopman y Taylor, 2005). Ciertas preguntas podían resultar inimaginables hasta hace poco tiempo. Hoy, sin

embargo, podemos cuantificar las diferencias estilísticas entre dos periodos o entre dos géneros diferentes dentro de un mismo autor.

Pondremos un ejemplo: sabemos que Benito Pérez Galdós publicó sus *Episodios nacionales* en cinco series, con una distancia de casi 40 años entre la primera y la última serie. Sometidos los textos de los *Episodios* a un algoritmo estadístico de componentes principales (PCA), realizado con RStudio, obtenemos un resultado tan explícito como el siguiente:

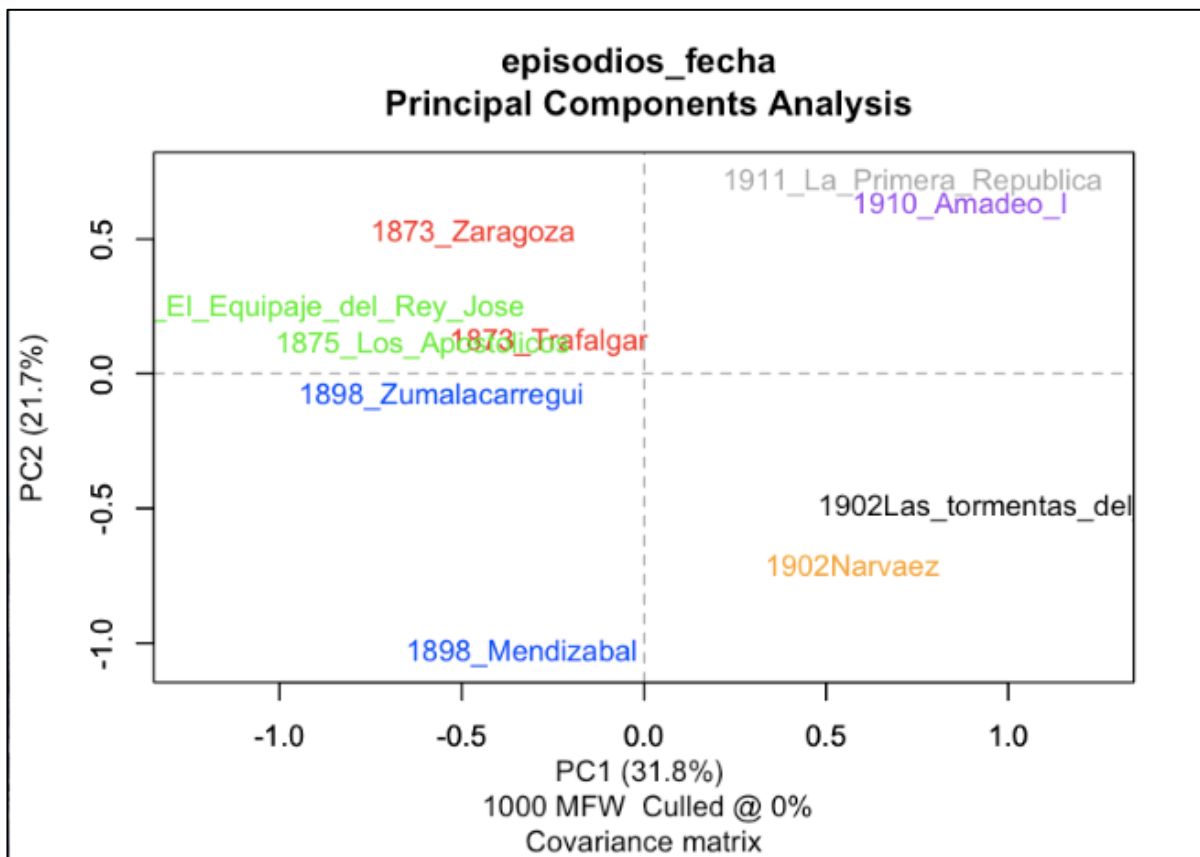


FIGURA 1. Componentes principales de los *Episodios nacionales* de Galdós (RStudio).
Fuente: Elaboración propia.

La imagen que refleja la Figura 1 es la de una evidente evolución estilística desde los primeros episodios a los últimos, con un punto de inflexión muy marcado que se localiza entre 1898 y 1902. ¿Es posible hacer lo mismo con la obra de Miguel Delibes? O, planteado de otro modo, ¿la obra de Delibes responderá de una manera tan clara y ofrecerá un gráfico de evolución tan claramente discriminatorio? Y, suponiendo que los análisis cuantitativos sean capaces de detectar la evolución estilística de un autor a lo largo del tiempo,

¿qué podemos esperar de estos análisis al ocuparse de obras cuya génesis – como es el caso de *Los santos inocentes*– se ha prolongado varios lustros?

Para responder a estas preguntas procedemos a crear un corpus con las siguientes obras de ficción narrativa: *La sombra del ciprés es alargada* (1948), *Aún es de día* (1949), *El camino* (1950), *Mi idolatrado hijo Sisí* (1953), *Las ratas* (1962), *El disputado voto del señor Cayo* (1978), *Los santos inocentes* (1981), *El hereje* (1998); y los tres diarios: *Diario de un cazador* (1955), *Diario de un emigrante* (1958), *Diario de un jubilado* (1995).

Analizamos, en primer lugar, las obras narrativas de ficción y la respuesta a la cuestión de la detección de una posible evolución de estilo es bastante clara:

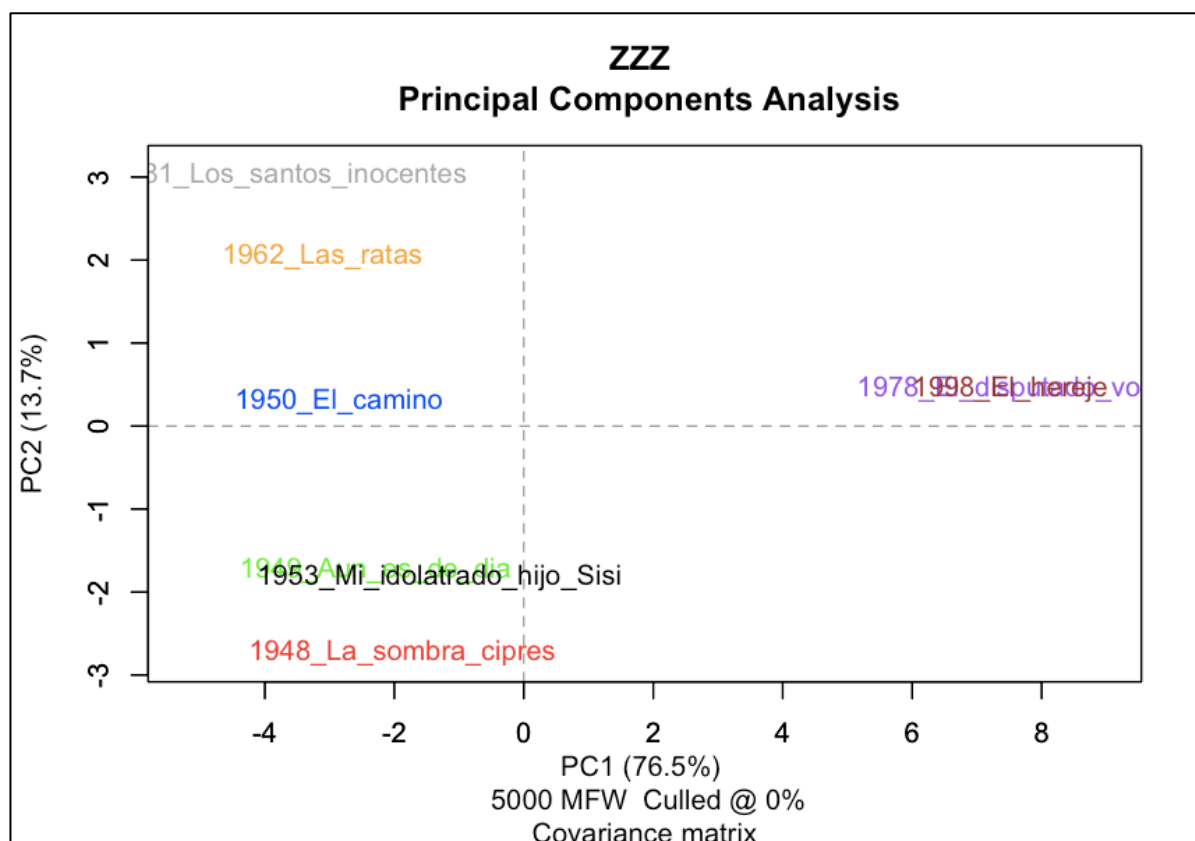


FIGURA 2. Componentes principales del corpus de obras narrativas de Delibes (RStudio). Fuente: Elaboración propia.

Como muestra la Figura 2, estas obras se agrupan de la siguiente manera: en el cuadrante inferior izquierdo aparecen las tres primeras obras, hasta 1953 (la confluencia estilística entre *Aún es de día* y *Mi idolatrado hijo Sisí* es tal que incluso se superponen); en el cuadrante superior izquierdo aparecen *El camino*

(1950), *Las ratas* (1962) y *Los santos inocentes* (1981); y en el cuadrante superior derecho, se superponen *El disputado voto del señor Cayo* (1978) y *El hereje* (1998), lo que refleja una absoluta identidad estilística entre ambas.

La matriz de covarianza aplicada al análisis de componente principales nos devuelve una figura en el que claramente se perciben tres momentos y una incuestionable secuenciación cronológica, con una frontera entre las dos primeras etapas, marcada por *El Camino*, y una segunda frontera hacia finales de los años 70 y principios de los 80, con la sorprendente salvedad de *Los santos inocentes*, que debería haber ido al cuadrante superior derecho en vez de asociarse con *Las ratas*, lo que deja abierta una interrogante. Si repetimos el análisis incorporando ahora los *Diarios*, el resultado obtenido se refleja así:

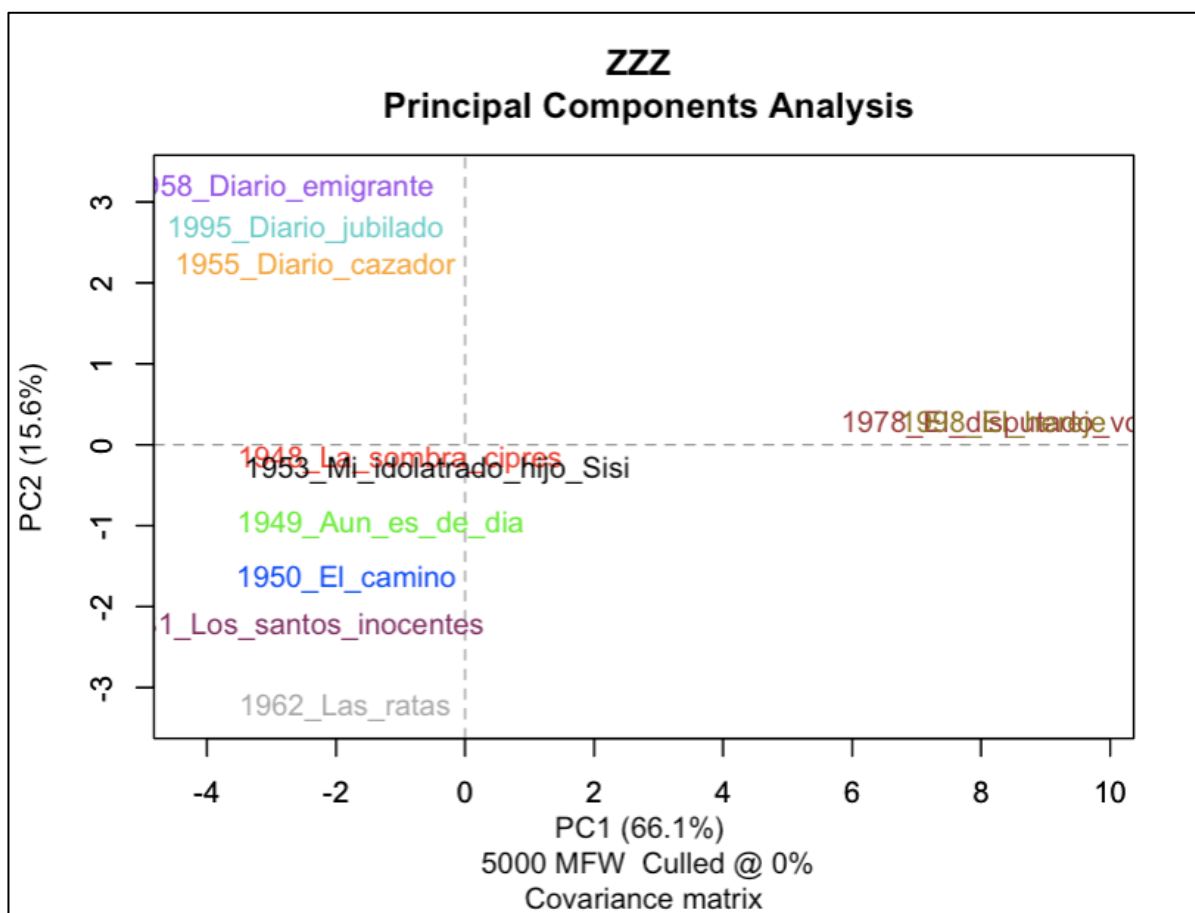


FIGURA 3. Componentes principales del corpus completo de Delibes (RStudio). Fuente: Elaboración propia.

La nueva figura nos sirve para confirmar que quedan detectadas en nuestro análisis las diferencias de estilo provocadas tanto por la supuesta evolución estilística a lo largo del tiempo, como las diferencias determinadas por el género. Y, sobre todo, que las marcas de género se imponen, en el estilo, por encima de las marcas temporales.

Con todo, observamos en ambos casos la misma anomalía en la ubicación de la novela *Los santos inocentes* (1981). Según la tendencia que muestran las Figuras 2 y 3, lo esperable era que esta obra se situase con las que vendrán después de 1970, en el cuadrante superior derecho; pero no ocurre así. De todos modos, tampoco nos sorprende mucho dicha anomalía, pues es natural que el carácter experimental de *Los santos inocentes* se manifieste de alguna manera en la lengua de los textos porque, como hemos podido observar en los gráficos anteriores, el "modo" literario (y quizá también el impulso literario) consigue imponer cambios estilísticos que llegan a tener más peso que los que se derivan de la evolución de la propia escritura. Sin embargo, en este caso la modalidad narrativa podría no ser el único elemento que explicase la anomalía. Recordemos, en relación con la génesis de *Los santos inocentes*, que el núcleo originario, como han señalado algunos estudiosos de Delibes (Rey, 2019: 192 y ss.), se localiza hacia 1967, con el relato "La milana". Allí ya aparecen los personajes de Azarías, Régula, Iván, así como un retrasado mental y se observa, igualmente, el protagonismo del pájaro. En 1970, nos recuerda Rey, Umbral dio a conocer un fragmento que coincide en gran medida con el inicio del libro segundo de *Los santos inocentes* y que el amigo de Delibes presenta así:

Este fragmento pertenece a un original inédito de una novela que aún no tiene título, y que Delibes no sabe si llegará a terminar... Se trata, evidentemente, de una novela rural... Una novela de señoritos feudales y paletos o servidores muy desgraciados (2019: 193).

De modo que el análisis de componentes principales –cuyos resultados quedan reflejados en la figura anterior– lo que detecta es no el momento puntual de la publicación del libro, esto es 1981, sino la extensa duración de su composición, un periodo que se prolonga desde 1967 hasta 1981. Por ello no debería extrañarnos el lugar en el que un texto como el que estudiamos ocupa en las Figuras 2 y 3.

Con esta información, procedemos ahora a estudiar algunas de las variables lingüísticas (léxicas, morfológicas, sintácticas, semánticas y pragmáticas) que, como marcas de estilo, están en la base de las figuras anteriores, a fin de confirmar nuestras suposiciones sobre el momento de redacción de *Los santos inocentes* y las características derivadas de su experimentalidad. El resultado del análisis de las clases de palabras, realizado con LIWC (Pennebaker y Graybeal, 2001; y Pennebaker, Mehl y Niederhoffer, 2003), es clarificador:

CLASES DE PALABRAS						
Filename	TotPron	PronPer	PronImp	Verbos	VerbAux	Adverb
1948_La_sombra_cipres.txt	16,39	11,38	5,01	10,22	1,21	4,69
1949_Aun_es_de_dia.txt	15,57	11,01	4,57	10,15	1,25	4,89
1950_El_camino.txt	14,95	10,70	4,25	10,34	1,05	5,05
1953_Mi_idolatrado_hijo_Sisi.txt	15,51	11,16	4,34	13,26	1,30	5,37
1962_Las_ratas.txt	12,53	9,57	2,95	9,79	0,65	4,29
1978_El_disputado_voto.txt	11,82	9,16	2,66	8,74	0,56	3,54
1981_Los_santos_inocentes.txt	14,76	10,87	3,90	9,85	0,99	5,11
1998_El_hereje.txt	12,58	9,28	3,30	6,10	0,44	3,10

TABLA I. Clases de palabras en el corpus de obras narrativas de Delibes (LIWC). Fuente: Elaboración propia.

Como puede observarse en la tabla anterior (Tabla I), los porcentajes de incidencia descienden progresivamente en todas las clases contempladas (total de pronombres, pronombres personales e impersonales, verbos, verbos auxiliares, y adverbios) conforme pasan los años, con la salvedad de los índices correspondientes a *Los santos inocentes*, que se acercan mucho más a los de las obras anteriores a 1962, que a los de las posteriores.

A pesar del carácter experimental de *Los santos inocentes* –basta advertir que la longitud media de palabras por frase (27,55) es casi tres veces superior a la de la media de sus otras novelas–, los elementos cuantificables de su estilo denuncian (en gran medida) los rasgos dominantes de la época a la que

pertenece el núcleo originario de esta novela, esto es, a la década de los años 60, como se refleja en la tabla anterior. Sin embargo, un análisis de algunos componentes afectivos demuestra unos índices que contradicen la tendencia observada en las clases de palabras; una tendencia que empareja ahora *los santos inocentes* con las obras posteriores a 1962:

AFECTIVIDAD						
Filename	Afect	Emopos	Emoneg	Ansiedad	Enfado	Triste
1948_La_sombra_cipres.txt	4,48	2,25	2,23	0,45	0,62	0,71
1949_Aun_es_de_dia.txt	4,89	2,33	2,40	0,50	0,76	0,64
1950_El_camino.txt	4,20	1,95	2,14	0,41	0,68	0,66
1953_Mi_idolatrado_hijo_Sisi.txt	5,79	3,35	2,33	0,45	0,76	0,58
1962_Las_ratas.txt	3,41	1,45	1,89	0,31	0,58	0,48
1978_El_disputado_voto.txt	2,63	1,46	1,18	0,16	0,63	0,22
1981_Los_santos_inocentes.txt	3,07	1,68	1,33	0,22	0,47	0,33
1998_El_hereje.txt	3,40	1,83	1,51	0,27	0,53	0,38

**TABLA II. Afectividad en el corpus de obras narrativas de Delibes (LIWC).
Fuente: Elaboración propia.**

Es curioso observar cómo el avance de la cronología se corresponde con un sensible descenso de la emotividad, tanto en sus aspectos positivos como negativos, y, en este tránsito, *Los santos inocentes* sí que sigue la tendencia general de la escritura posterior a 1962; algo que Pennebaker y Stone (2003) ya habían observado, en general, para el inglés (específicamente, mostraron que a mayor edad los individuos tendían a utilizar menos palabras de emociones negativas).

En relación con los mecanismos cognitivos, se observan también índices menores de frecuencia en las obras más tardías, tanto en lo que se refiere al análisis de las "causas", como en las manifestaciones de "discrepancia", "tentativa" o "certeza" (Tabla III):

PROCESOS COGNITIVOS				
Filename	Causa	Discrep	Tentat	Certeza
1948_La_sombra_cipres.txt	1,12	0,87	2,02	1,25
1949_Aun_es_de_dia.txt	1,04	0,81	2,00	1,32
1950_El_camino.txt	1,09	0,93	2,05	1,19
1953_Mi_idolatrado_hijo_Sisi.txt	1,12	1,03	1,94	1,19
1962_Las_ratas.txt	1,04	0,62	1,76	0,73
1978_El_disputado_voto.txt	0,45	0,61	1,73	0,72
1981_Los_santos_inocentes.txt	1,05	0,72	1,80	0,72
1998_El_hereje.txt	0,60	0,47	1,81	0,84

TABLA III. Procesos cognitivos del corpus de obras narrativas de Delibes (LIWC).
Fuente: Elaboración propia

Por otro lado, se cumple perfectamente la ley enunciada por Pennebaker y Stone sobre la relación entre edad y el menor uso de palabras en tiempo pasado (2003), a la vez que se incrementa el uso de estructuras en presente, tal y como refleja la siguiente tabla:

TEMPORALIDAD		
Filename	Pasado	Presente
1948_La_sombra_cipres.txt	5,22	3,72
1949_Aun_es_de_dia.txt	4,75	4,06
1950_El_camino.txt	5,29	3,59
1953_Mi_idolatrado_hijo_Sisi.txt	6,36	5,48
1962_Las_ratas.txt	4,84	3,99
1978_El_disputado_voto.txt	2,57	5,65
1981_Los_santos_inocentes.txt	4,19	4,84
1998_El_hereje.txt	2,26	2,91

TABLA IV. Temporalidad del corpus de obras narrativas de Delibes (LIWC).
Fuente: Elaboración propia

Analizamos finalmente algunos aspectos relacionados con el contenido del corpus seleccionado:

CONTENIDOS				
Filename	Cuerpo	Salud	Sexual	Muerte
1948_La_sombra_cipres.txt	1,72	0,62	0,30	0,54
1949_Aun_es_de_dia.txt	1,86	0,46	0,34	0,40
1950_El_camino.txt	1,64	0,54	0,40	0,37
1953_Mi_idolatrado_hijo_Sisi.txt	1,77	0,55	0,59	0,34
1962_Las_ratas.txt	1,92	0,32	0,46	0,38
1978_El_disputado_voto.txt	1,78	0,26	0,46	0,24
1981_Los_santos_inocentes.txt	2,13	0,27	0,48	0,22
1998_El_hereje.txt	1,22	0,57	0,35	0,26

TABLA V. Contenidos del corpus de obras narrativas de Delibes (LIWC).
Fuente: Elaboración propia.

Aunque las referencias al cuerpo no sufren variaciones significativas, sí que se observa una menor atención a los temas relacionados con la salud y con la muerte conforme avanzan los años, a la vez que crece la atención e interés por la sexualidad. Las explicaciones de estas variaciones pueden ser de naturaleza diversa y, desde luego, no estrictamente literarias. Por ejemplo, la apertura que siguió a la caída de la dictadura pudiera haber favorecido la atención mayor a ciertos temas relacionados con la sexualidad.

Todavía nos queda una cuestión por dilucidar: ¿sería posible (desde un punto de vista meramente lingüístico, sin recurso documental a los manuscritos) determinar qué partes de *Los santos inocentes* pertenecen a las fechas de su concepción y cuáles a las fechas de su redacción definitiva? Para dar respuesta a esta pregunta, confrontamos con 'Stylo' de RStudio (función 'rolling.classify') el texto de esta novela con otros de diferentes momentos de la escritura de Delibes: *El camino* (1950), *Las ratas* (1962) y *El disputado voto del señor Cayo* (1978):

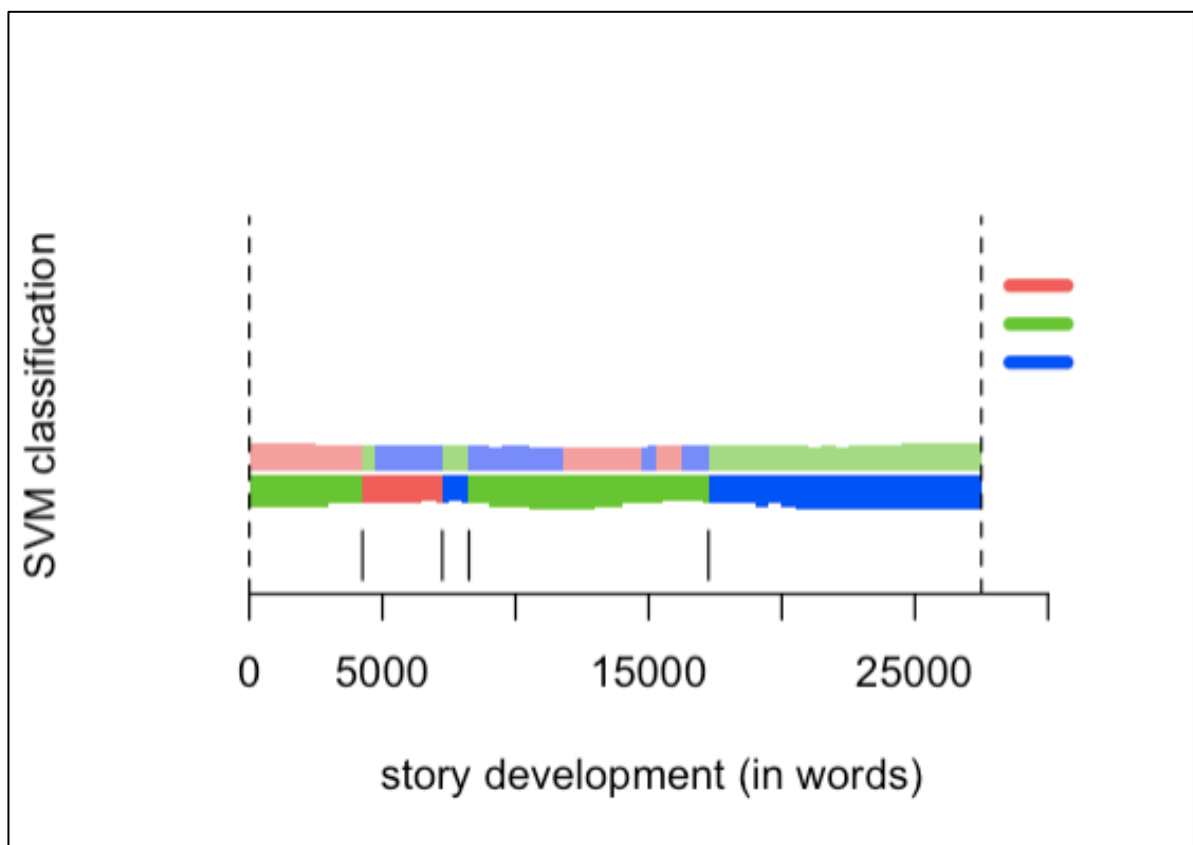


FIGURA 4. Cercanía lingüística entre *Los santos inocentes* y otras obras de Delibes representativas de tres periodos cronológicos distintos: *El camino* (en color rojo), *Las ratas* (en color verde) y *El disputado voto del señor Cayo* (en color azul) (RStudio).
Fuente: Elaboración propia.

El gráfico refleja estilos representativos de tres periodos cronológicos distintos de la escritura de Delibes. El color rojo representa *El camino* (1950); el verde, *Las ratas* (1962); y el azul, *El disputado voto del señor Cayo* (1978). La franja coloreada inferior marca los estadios o momentos creativos sobre el desarrollo de la historia de *Los santos inocentes*. Y, así, se muestra que, mientras que a partir de las 17 000 primeras palabras el estilo dominante es el de la última época, en las 17 000 primeras palabras el estilo remite a 1962, con huellas (entre las 4 900 y las 5 300 palabras) del estilo anterior e intervenciones esporádicas del estilo posterior a 1978. La franja superior de color más tenue ayuda a precisar intervenciones que bien pudieran tener que ver con las correcciones últimas de la obra, en fechas próximas a la de la edición.

CONCLUSIONES

Lingüísticamente *Los santos inocentes*, por lo dilatado de su génesis compositiva, se escapa de la tendencia observada para el resto de la narrativa

delibesiana al presentar huellas tanto de la narrativa representativa de finales de los 70 (de donde parece ser que procede el núcleo originario de esta novela), como de la narrativa representativa de los últimos años, aquellos en los que se inscribe la última redacción del texto y la edición (Figura 4). Sin embargo, en aquellos otros aspectos que hacen referencia a los contenidos temáticos, a los aspectos emocionales o a los mecanismos cognitivos, la mencionada novela sigue la tendencia general del conjunto de la obra de Delibes posterior a los años 70.

Y finalmente, por contestar a todas las preguntas que nos hacíamos al principio de este trabajo, el análisis estilométrico de *Los santos inocentes* parece ofrecer alta fiabilidad a la hora de recoger los cambios estilísticos que se producen a lo largo de los años, al menos cuando la trayectoria de un autor abarca varias décadas.

BIBLIOGRAFÍA

- Alpers, G. W., A. J. Winzelberg, C. Classen, H. Roberts, P. Dev, C. Koopman y C. B. Taylor (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*, 21 (2), 361-376.
- Cohn, M. A., M. R. Mehl y J. W. Pennebaker (2004). Linguistic indicators of psychological change after September 11, 2001. *Psychological Science*, 15, 687-693.
- Hart, R. P. (1984). *Verbal style and the presidency: A computer-based analysis*. New York: Academic Press.
- Pennebaker, J. W. y A. Graybeal (2011). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10, 90-93.
- Pennebaker, J. W., M. R. Mehl y K. G. Niederhoffer (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.
- Pennebaker, J. W. y L. D. Stone (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality and Social Psychology*, 85, 291-301.

Rey, A. (2019). *La narrativa de Delibes (1948-1998): cambio y traducción*. Santiago de Compostela: Universidad de Santiago de Compostela.

Stone, P. J., D. C. Dunphy, M. S. Smith y D. M. Ogilvie (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: M.I.T. Press.