

<https://artnodes.uoc.edu>

ARTÍCULO

La transformación de la ciencia ficción: un análisis cuantitativo del patrón emocional en los premios Hugo

José Luis Arroyo-Barrigüete

Universidad Pontificia Comillas

Fecha de presentación: abril de 2021

Fecha de aceptación: mayo de 2021

Fecha de publicación: julio de 2021

Cita recomendada

Arroyo-Barrigüete, José Luis. 2021. «La transformación de la ciencia ficción: un análisis cuantitativo del patrón emocional en los premios Hugo». *Artnodes*, núm. 28: 1-10. [Consulta: dd/mm/aa]. <http://doi.org/10.7238/a.v0i28.385684>



Los textos publicados en esta revista están sujetos –si no se indica lo contrario– a una licencia de Reconocimiento 4.0 Internacional de Creative Commons. La licencia completa se puede consultar en https://creativecommons.org/licenses/by/4.0/deed.es_ES.

Resumen

La ciencia ficción como género literario ha estado en constante evolución desde sus orígenes, a principios del siglo pasado. Sin embargo, en las últimas décadas esta evolución se ha acelerado, pasando de ser un género de corte conservador a adoptar enfoques más próximos al progresismo político (Oleszczuk, 2017). El objetivo de esta investigación es determinar si esta transformación se aprecia también en la construcción estilística de las obras y, más concretamente, en su patrón emocional, desarrollando un estudio sincrónico dual de las décadas de 1960 y 2010. Concretamente se han analizado veinticuatro obras galardonadas con el premio Hugo, probablemente el más relevante en el ámbito de la ciencia ficción y, en menor medida, de la fantasía, doce de la década de 1960 y otras doce de la década de 2010. Sobre este corpus de documentos se han aplicado tres técnicas de lingüística computacional: análisis de sentimientos, clustering y componentes principales, a fin de identificar grupos homogéneos de novelas. Los resultados muestran la agrupación en tres clústeres claramente

diferenciados, con patrones emocionales disímiles, y que correlacionan fuertemente con la década en que fueron escritas. Esto confirma que, adicionalmente a la evolución temática, se ha producido un cambio sustancial desde el punto de vista estilístico.

Palabras clave

ciencia ficción, literatura, lingüística computacional, análisis de sentimientos.

The transformation of science fiction: a quantitative analysis of the emotional pattern at the Hugo Awards

Abstract

Science fiction as a literary genre has been in constant evolution since its origins, at the beginning of the last century. However, in recent decades this evolution has accelerated from being a conservative genre to adopt approaches closer to political progressivism (Oleszczuk, 2017). The aim of this research is to determine whether this transformation is also seen in the stylistic construction of the novels, and more specifically, in their emotional pattern, conducting a dual synchronic study of the 1960s and 2010s. Specifically, it has been analysed 24 novels that received the Hugo Award, probably the most relevant in the field of science fiction and, to a lesser extent, fantasy, 12 from the 1960s and another 12 from the 2010s. Three computational linguistic techniques were applied to this corpus of documents: sentiment analysis, clustering and principal components, in order to identify homogeneous groups of novels. The results show the grouping into three clearly differentiated clusters, with dissimilar emotional patterns, and which correlate strongly with the decade in which they were written. This confirms that, in addition to the thematic evolution, there has been a substantial change from a stylistic point of view.

Keywords

Science fiction, literature, computational linguistic, sentiment analysis

1. Introducción

La popularidad del término *Science-Fiction* se debe al ínclito escritor y editor Hugo Gernsback, pese a no ser él quien acuñó el término. De hecho, inicialmente utilizó la expresión *scientifiction*, que debió abandonar tras perder el control de su revista *Amazing Stories* en 1929. Esto le llevó a buscar un nuevo vocablo para referirse al tipo de relatos que comenzó a publicar en tres nuevas revistas que fundó pocos meses después: *Air Wonder Stories*, *Science Wonder Stories* y *Science Wonder Quarterly*. Es en esta época, los inicios de la ciencia ficción, cuando surge el término *Pulp science fiction*¹, con unas connotaciones claramente negativas, para referirse a obras con una baja calidad literaria, muy centradas en la aventura y la acción, principalmente física, y que en aquellos momentos domi-

naban claramente el género. También es en este periodo en el que surge el concepto de *space opera*, probablemente el subgénero de la ciencia ficción más conocido en la actualidad, término acuñado por Wilson Tucker en 1941 para describir «[historias superficiales, gastadas, nauseabundas y manidas de naves espaciales]» (Pringle, 2000: 35)². Así pues, los orígenes de este género parecen marcados, y probablemente con cierto sentido, por una percepción negativa sobre su calidad literaria. En aquellos momentos el objetivo era el mero entretenimiento, quizá algo infantil, que demandaba un creciente número de aficionados.

Sin embargo, en la década de 1940 se produjo un importante cambio de la mano de John W. Campbell, editor de la revista *Astounding Science Fiction*. Rodeado de jóvenes autores que posteriormente adquirieron una enorme notoriedad, como Isaac Asimov, Alfred Elton

1. Realmente el término *pulp* hace referencia al desecho de pulpa de madera con el que se fabricaba el papel, de muy baja calidad, en el que se imprimían estas revistas.

2. Una definición más moderna y algo menos peyorativa es la que ofrece Wolfe (1986: 120) en su glosario de términos de la ciencia ficción: «*space operas* [suelen ser aventuras intergalácticas a gran escala de ritmo rápido]». Por su parte, Moreno (2013: cap. 7) las caracteriza como «los equivalentes espaciales a los cantares épicos medievales o a los libros de caballerías renacentistas, contemplados desde el punto de vista de las líneas argumentales».

van Vogt, Philip K. Dick o Robert A. Heinlein, comenzó a ofrecer a sus lectores relatos atractivos, al tiempo que de una mayor calidad literaria. Comenzó así la llamada *edad de oro de la ciencia ficción*, comprendida entre los años cuarenta y cincuenta (Moreno, 2013), en la que surgieron y se consolidaron algunos de los autores icónicos. Podemos afirmar que desde ese momento el género ha gozado de buena salud, no solo por el apoyo de los aficionados, sino por un creciente interés entre la comunidad académica. Sin embargo, incluso ahora, siguen existiendo aspectos no resueltos y quizá el más chocante de ellos es la ausencia de una delimitación clara y universalmente aceptada, en especial por la existencia de características comunes con el llamado género de *fantasía*. Una de las más conocidas definiciones de la ciencia ficción es la que ofrece Suvin (1972: cap. 2)³: «[La CF es, pues, un género literario cuyas condiciones necesarias y suficientes son la presencia e interacción del extrañamiento y la cognición]», que en trabajos más recientes modifica para incorporar el concepto de cronotopo: «[un género literario (etc.) definido por la interacción del extrañamiento y la cognición histórica, y cuyo principal elemento formal es un cronotopo narrativo y/o agentes alternativos al mundo empírico del autor]» (2008: 116). Los escritores de ciencia ficción también plantean sus propias definiciones, así Isaac Asimov la describe como «la rama de la literatura que trata de la respuesta humana a los cambios científicos y tecnológicos» (Barceló, 2015: 29). Una de las más recientes definiciones del género, y por otra parte de las más sólidamente construidas desde un punto de vista teórico, es la que ofrece Moreno (2013: cap. 2): «género de ficción proyectiva basado en elementos no sobrenaturales». Son solo algunos ejemplos de visiones alternativas, que, si bien comparten elementos comunes, no son plenamente coincidentes. Hemos de mencionar aquí que, al margen de este debate académico, un reciente trabajo empírico desarrollado por Menadue, Giselsson y Guez (2020: 1) sobre una muestra de 232 aficionados al género demostró que la ciencia ficción «[puede definirse categóricamente. Los datos de la encuesta empírica son consistentes y demuestran una distinción inequívoca entre las definiciones populares de ciencia ficción y fantasía]». Es decir, parece que pese a las dificultades teóricas que existen a la hora de definir con nitidez el género, los aficionados sí son capaces de identificar claramente este tipo de obras. Se trata de un resultado aún muy reciente como para que haya sido replicado en estudios similares, que verifique, de este modo, su solidez, pero aun así resulta sorprendente: no está claro que podamos definir con límites precisos la ciencia ficción, pero sí reconocerla cuando la leemos. Y para añadir

todavía más confusión, sigue abierto el interesante debate entre la denominada *ciencia ficción dura* (*Hard SF*) y *ciencia ficción blanda* (*Soft SF*), distinción que, de nuevo, dista de estar resuelta⁴.

Ciertamente, definir unos límites nítidos para la ciencia ficción parece una tarea harto complicada, cuando no imposible, especialmente cuando estamos ante un género en constante evolución, que ha ido experimentando enormes cambios a lo largo de los años. Cambios que no siempre han sido bien recibidos por algunos aficionados. Como señala Oleszczuk (2017: 127), siendo tradicionalmente un género con un corte conservador, «[la ciencia ficción se ha convertido en un discurso cuya representación progresista de prácticamente todos los temas políticos controvertidos y el cuestionamiento crítico del *statu quo* político, se ha mantenido como ningún otro en la llamada cultura popular]». Esto ha llevado a que grupos de aficionados, organizados en torno a los movimientos autodenominados *Sad Puppies* y *Rabid Puppies*, se hayan manifestado en contra de lo que perciben como la represión de la visión más tradicional de la ciencia ficción.

Como apunta Oleszczuk, las temáticas efectivamente parecen haber cambiado y resulta de interés evaluar hasta qué punto los aspectos estilísticos también lo han hecho. ¿Es posible establecer una distinción nítida entre las obras de ciencia ficción clásica y las novelas actuales? Esta es, de hecho, la pregunta de investigación a la que se pretende dar respuesta en este trabajo. La hipótesis es que, efectivamente, desde los años sesenta se ha producido un cambio sustancial en el estilo narrativo de la literatura de ciencia ficción, más concretamente en el patrón emocional de las obras que conforman este género literario.

Este artículo se enmarca en el contexto de una serie de trabajos que abordan el análisis literario desde la óptica del *sentiment analysis*. Se trata de un área que, si bien es relativamente reciente debido a la disponibilidad de herramientas computacionales de análisis de texto, está despertando un considerable interés entre la comunidad académica. El conocido trabajo de Reagan et al. (2016) abrió una fructífera línea de investigación al concluir, tras analizar 1.327 relatos, que apenas un conjunto de seis arcos emocionales constituía los bloques esenciales de patrones emocionales más complejos. Siguiendo este enfoque, Kim, Padó y Klinger (2017: 24-25), tras un completo análisis de 2.019 obras pertenecientes a cinco géneros distintos, entre ellos la ciencia ficción, concluyeron que «[la clasificación basada en las emociones asociadas a palabras tiene un rendimiento similar al basado en las características tradicionales de los géneros [...]]. Nuestra primera conclusión es, por tanto, que las emociones contienen información muy relevante para

3. Realmente este ensayo es la versión revisada de una ponencia que impartió en la Universidad de Yale en 1968. Junto con otros ensayos similares, fue publicado en forma de libro en 1979, bajo el título de *Metamorphoses of Science Fiction: on the Poetics and History of a Literary Genre* (Suvin, 2016 [1979]).

4. Higgins y Duncan (2014: cap. 7) en *The Science Fiction Handbook* definen la *Hard SF* como «[un término utilizado para designar la ficción que se centra en conocimientos y principios científicos establecidos o en una extrapolación cuidadosamente estudiada de dichos conocimientos y principios. La distinción entre ciencia ficción 'dura' y 'blanda' suele ser paralela a la misma distinción en las ciencias: la mayoría de la ciencia ficción dura se identifica por su énfasis en la física, la astronomía, la química, la ingeniería y la biología]».

distinguir los géneros]». Otros trabajos como los de Wang, Zhang y Smetannikov (2020) han confirmado que, efectivamente, el análisis emocional de las novelas de ficción no solo permite identificar a qué género pertenece, sino incluso predecir con una notable precisión (73,4%) la popularidad de una determinada obra. Parece, por tanto, que el patrón emocional de una obra o conjunto de obras proporciona mucha información útil desde el punto de vista del análisis literario, razón por la que se ha optado por este enfoque para tratar de verificar la hipótesis de investigación antes mencionada.

El presente trabajo se estructura en cuatro partes diferentes. En primer lugar, se detalla la metodología de análisis empleada, que combina tres técnicas de lingüística computacional: análisis de sentimientos, clústering y componentes principales. A continuación, se muestran los resultados obtenidos, así como su interpretación. Por último, se presentan las principales conclusiones del trabajo, así como sus limitaciones y futuras líneas de investigación.

2. Metodología

Se ha seleccionado un conjunto de veinticuatro novelas de ciencia ficción, doce de la década de los sesenta del siglo xx, como representantes de la ciencia ficción más clásica, y otras doce de esta última década, representantes de la última ola en ciencia ficción. Concretamente se ha optado por todas las novelas ganadoras del prestigioso premio Hugo, probablemente el más relevante en el ámbito de la ciencia ficción⁵, en ambos periodos, pues fue en la década de 1960 cuando se concedieron los primeros galardones. En la tabla 1 se muestra el título y autor de cada novela, así como el año en que fue galardonada.

Para cada novela se ha llevado a cabo un análisis de sentimientos (*sentiment analysis*), según el esquema habitual en algoritmos del tipo *Bag-of-Words*. Como primera etapa se han procesado los datos, trabajando sobre todas las palabras que componen cada obra: transformación de letras mayúsculas a minúsculas, eliminación de signos de puntuación, eliminación de las denominadas *stop words* o palabras

vacías⁶ y *tokenización* (división del texto en piezas más pequeñas o *tokens*, que en el caso de este trabajo son palabras). Posteriormente, se ha empleado el lexicón NRC-VAD (Mohammad, 2018) para asociar cada token a las correspondientes emociones. Este lexicón es un listado de 20.007 palabras en lengua inglesa que asigna a cada una de ellas una puntuación según tres dimensiones: *valence* (nivel de positividad o negatividad de una palabra), *arousal* (dimensión que evalúa la excitación o calma) y *dominance* (dimensión correspondiente al grado de fuerza o debilidad). Para cada palabra, la puntuación toma un valor de 0 (nivel más bajo de V/A/D) a 1 (nivel más alto V/A/D). Así, por ejemplo, la palabra *punish* toma valores de 0,061/0,808/0,610, lo que implica nivel muy bajo de *valence* (palabra negativa), bastante alto de *arousal* (alta excitación) y medio-alto de *dominance* (nivel medio-alto de fuerza). Por el contrario, la palabra *harmony* (0,949/0,230/0,509) es clasificada como muy positiva, de baja excitación y fuerza media. Tras llevar a cabo este cálculo para todas las palabras de cada novela, se ha procedido a calcular su valor medio, realizando la suma y dividiendo por el total de palabras que la componen; así se ha obtenido una métrica que podríamos denominar *densidad emocional*. De este modo, se ha asignado a cada obra tres valores (*valence/arousal/dominance*) que corresponden a la intensidad media con que cada una de dichas dimensiones se manifiesta.

A continuación, se ha llevado a cabo un análisis clúster, es decir, un procedimiento para identificar grupos homogéneos según un conjunto de variables, en este caso: *valence*, *arousal* y *dominance*. El objetivo es agrupar todas las novelas en clústeres o grupos relativamente homogéneos de acuerdo a estas tres características, de manera que las novelas que conforman un grupo sean similares entre sí y diferentes a las novelas de otros grupos. Se ha optado por un algoritmo jerárquico⁷ aglomerativo⁸, empleando la distancia euclídea⁹ y el método de enlace de Ward (de varianza mínima)¹⁰, pues dicho método tiende a producir clústeres compactos. A fin de determinar el número óptimo de grupos, se han comparado veinticuatro índices diferentes, seleccionando el recomendado por la regla de la mayoría.

Por último, se han representado los tres grupos de novelas en función de las dos primeras componentes principales¹¹, a fin de vi-

5. Y en menor medida, de la *fantasía*, pues en ocasiones obras de este género también han sido galardonadas.

6. Palabras que carecen de significado en sí mismas, pues modifican o acompañan a otras, como artículos, pronombres o preposiciones.

7. Dentro de las dos grandes categorías de análisis clúster encontramos los métodos jerárquicos y los no jerárquicos. Los primeros están especialmente indicados cuando no se conoce a priori el número de grupos y la cantidad de objetos no es muy grande, mientras que los segundos requieren que se estime de antemano el número de grupos, y adicionalmente suelen emplearse en grandes muestras. En este caso, dado que la muestra es reducida (veinticuatro novelas) y preferimos no realizar hipótesis a priori, los métodos jerárquicos parecen más adecuados.

8. Comienza con tantos clústeres como objetos que clasificar y va uniendo progresivamente los elementos en grupos de mayor tamaño.

9. Calculada en este caso según la siguiente expresión: distancia entre las novelas A y B = $d(A,B) = \sqrt{\sum_{i=1}^3 |A_i - B_i|^2}$

10. Este método parte de la idea de que los grupos deben formarse de modo que, al fusionar dos clústeres, la pérdida de información sea mínima (la información se cuantifica como la suma de las distancias al cuadrado de cada elemento respecto al centroide del grupo al que pertenece). De este modo, en cada paso se unen aquellos clústeres que generan una menor pérdida de información, es decir, una menor suma de cuadrados de las distancias intra conglomerado.

11. El análisis de *componentes* principales permite describir un conjunto de datos (en este caso las veinticuatro novelas) en términos de nuevas variables que denominamos componentes, y que se generan como una combinación lineal de las variables originales. Estos componentes se ordenan por la cantidad de varianza original que describen, es decir, por la cantidad de información que recogen. En nuestro caso, los dos primeros componentes explican el 99% de la variabilidad, de modo que la representación basada en dichos componentes proporciona una imagen bastante fidedigna.

Año	Título	Autor	Observaciones
1960	Starship Troopers	Robert A. Heinlein	
1961	A Canticle for Leibowitz	Walter M. Miller, Jr.	
1962	Stranger in a Strange Land	Robert A. Heinlein	Se ha analizado la versión ampliada de esta obra
1963	The Man in the High Castle	Philip K. Dick	
1964	Here Gather the Stars (Way Station)	Clifford D. Simak	
1965	The Wanderer	Fritz Leiber	
1966	(empatado) Dune	Frank Herbert	
1966	(empatado) ...And Call Me Conrad	Roger Zelazny	Se ha analizado la versión ampliada de esta obra (This Immortal)
1967	The Moon Is a Harsh Mistress	Robert A. Heinlein	
1968	Lord of Light	Roger Zelazny	
1969	Stand on Zanzibar	John Brunner	
1970	The Left Hand of Darkness	Ursula K. Le Guin	
2010	(empatado) The Windup Girl	Paolo Bacigalupi	
2010	(empatado) The City & the City	China Miéville	
2011	Blackout/All Clear	Connie Willis	
2012	Among Others	Jo Walton	
2013	Redshirts: A Novel with Three Cudas	John Scalzi	
2014	Ancillary Justice	Ann Leckie	
2015	The Three-Body Problem	Cixin Liu	Se ha analizado la versión traducida al inglés por Ken Liu
2016	The Fifth Season	N. K. Jemisin	
2017	The Obelisk Gate	N. K. Jemisin	
2018	The Stone Sky	N. K. Jemisin	
2019	The Calculating Stars	Mary Robinette Kowal	
2020	A Memory Called Empire	Arkady Martine	

Tabla 1. Novelas seleccionadas (ganadoras del premio Hugo en las décadas de 1960 y 2010)

sualizar gráficamente hasta qué punto se trata de clústeres realmente homogéneos y claramente separados.

El análisis se ha llevado a cabo íntegramente en R (R Core Team, 2013), un entorno de programación orientado al análisis estadístico, empleando diversas librerías para elaborar el código: *NbClust* (Charrad et al., 2014), *tidytext* (Silge y Robinson, 2016), *ggplot2* (Wickham, 2016), *fmsb* (Nakazawa, 2018), *tidyr* (Wickham y Henry, 2019), *scales* (Wickham, 2018), *stringr* (Wickham, 2019), *dplyr* (Wickham et al., 2020), *factoextra* (Kassambara y Mundt, 2020) y *readtext* (Benoit y Obeng, 2020).

3. Resultados

El análisis preliminar (véase tabla 2) ya parece indicar que existen diferencias entre ambas décadas, pues en las tres dimensiones consideradas, los niveles son menores en las obras escritas en el periodo 2010-2020. Sin embargo, esto podría deberse al hecho de que, dada una cierta tendencia a novelas más largas en las últimas décadas, la densidad emocional ha ido disminuyendo. Sin duda, esto en sí mismo es un resultado interesante, pero el objetivo de este trabajo es profundizar en los patrones, más allá de que,

	VALENCE	AROUSAL	DOMINANCE
Década de 1960	0,36	0,29	0,33
Década de 2010	0,33	0,27	0,30

Tabla 2. Valores medios de *valence/arousal/dominance* de las obras en las dos décadas

efectivamente, parece observarse una disminución en la densidad emocional.

Mucho más relevante resulta el mapa de calor de las veinticuatro novelas (fig. 1). Los colores indican el nivel de cada una de las tres dimensiones emocionales para cada novela, correspondiendo el rojo a niveles más altos y el azul a niveles más bajos. La interpretación vertical muestra las agrupaciones de las tres dimensiones emocionales, e indica que *valence* y *dominance* están más relacionadas entre sí que *arousal*. No obstante, no es el objetivo de esta investigación entrar a analizar las posibles relaciones entre dichas dimensiones, de modo que es la interpretación horizontal la que presenta un mayor interés a efectos del presente trabajo. Como puede comprobarse, encontramos un primer grupo (clúster 1) formado por nueve novelas, de las que ocho corresponden a la década de 1960, siendo *The Three-Body Problem* de Cixin Liu la única excepción (2015). En términos comparativos, estas novelas se caracterizan por niveles elevados de *valence* (orientación positiva), pues, de hecho, este conjunto de obras son precisamente las que manifiestan una orientación más positiva de todas las novelas consideradas. Del mismo modo, su *arousal* y *dominance* también presentan niveles superiores a los del resto de clústeres. Expresándolo en términos comparativos, estamos ante un conjunto de obras que tienen una orientación clara hacia emociones positivas y están caracterizadas por niveles de fuerza y acción elevados. El tercer grupo, antagónico al primero, está formado exclusivamente por novelas de la década de 2010 y presenta los niveles más bajos

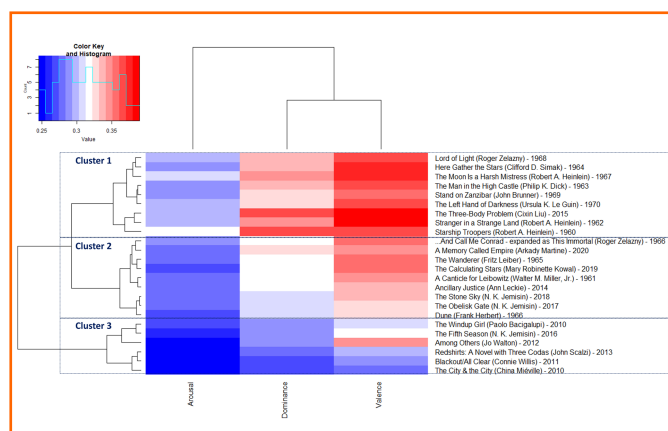


Figura 1. Mapa de calor y dendrogramas para las veinticuatro novelas y las tres dimensiones emocionales

	VALENCE	AROUSAL	DOMINANCE
Clúster 1	0,37	0,30	0,34
Clúster 2	0,34	0,28	0,32
Clúster 3	0,31	0,25	0,28

Tabla 3. Valores medios de *valence, arousal* y *dominance* en cada uno de los clústeres

en las tres dimensiones. Es decir, una orientación más negativa con menos acción y fuerza. Por último, el clúster 2 incluye cuatro novelas de la década de 1960, y cinco de la década de 2010, y presenta niveles medios en las tres dimensiones. En la tabla 3 se muestran los valores medios por clúster, observándose importantes diferencias entre ellos, especialmente entre el 1 (novelas mayoritariamente de la década de 1960) y el 3 (novelas de la década de 2010), del 18%, 15% y 19% respectivamente para *valence, arousal* y *dominance*.

A continuación, se ha llevado a cabo la representación de los clústeres en sus dos primeras componentes principales. El objetivo es verificar hasta qué punto existen solapamientos entre ellos, lo que sería una indicación de que la agrupación no es excesivamente buena, o, por el contrario, los clústeres no se solapan, lo que indicaría que la agrupación sí parece adecuada. En este caso, se observa que, efectivamente, no aparecen solapamientos, de modo que podemos concluir que los grupos están claramente diferenciados, considerando que la varianza explicada es del 99%.

Estamos, por tanto, ante un resultado que pone de manifiesto una clara evolución en el patrón emocional a lo largo del tiempo

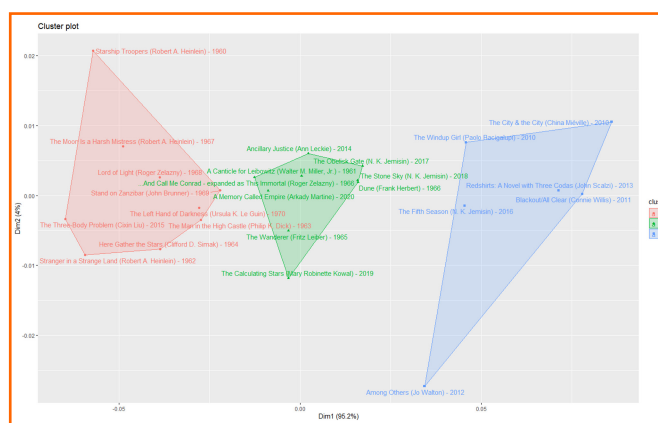


Figura 2. Representación de los clústeres en sus dos primeras componentes principales

	CLÚSTER 1: NIVELES ALTOS EN LAS 3 DIMENSIONES	CLÚSTER 2: NIVELES MEDIOS EN LAS 3 DIMENSIONES	CLÚSTER 3: NIVELES BAJOS EN LAS 3 DIMENSIONES
Porcentaje de obras (década de los 1960)	67%	33%	0%
Porcentaje de obras (década de los 2010)	8%	42%	50%

Tabla 4. Valores medios de *valence*, *arousal* y *dominance* en cada uno de los clústeres

(véase tabla 4). La gran mayoría de las novelas escritas en la década de 1960 (el 67%) pertenecen al clúster 1 y están caracterizadas por las emociones positivas, la fuerza y la acción. Es decir, encajan en la concepción más clásica de la ciencia ficción, enlazando así con el concepto de *space opera*, tan popular en los orígenes de este género: aventuras intergalácticas a gran escala. Apenas un 33% de las obras galardonadas en este periodo quedan englobadas en el clúster 2, y ninguna en el clúster 3, que, como ya hemos indicado, presenta niveles sensiblemente inferiores en las tres dimensiones emocionales consideradas. Con las obras premiadas en la década de 2010 sucede todo lo contrario. Apenas un 8% (una única novela, *The Three-Body Problem*) pertenece al clúster 1, estando el 92% restante en los grupos 2 y 3. Es decir, se trata de novelas que transmiten una menor positividad y presentan niveles de acción y fuerza inferiores a las escritas en la década de 1960. De hecho, resulta muy interesante el hecho de que la única novela atípica en la década de 2010 sea *The Three-Body Problem* de Cixin Liu, pues es la única obra escrita en otro idioma y posteriormente traducida al inglés por Ken Liu. La cuestión que surge inmediatamente es hasta qué punto el anómalo comportamiento de esta obra se debe, precisamente, a la traducción. En otras palabras, probablemente el hecho de haber analizado la novela traducida haya distorsionado el análisis y quizá la obra original muestre un patrón emocional que podría encuadrarse mejor en los clústeres 2 o 3. Se trata de un tema que dejamos abierto como futura línea de investigación, en tanto que requeriría un análisis comparativo muy pormenorizado de la obra original y su traducción.

4. Conclusiones

A lo largo de este trabajo se ha analizado el patrón emocional de veinticuatro novelas de referencia en el género de la ciencia ficción. Concretamente todas las ganadoras del premio Hugo en las décadas de 1960 y 2010. Basándonos en las dimensiones de *valence*, *dominance* y *arousal*, todas las obras pueden clasificarse

en tres grupos bien definidos, que, en términos generales, están ligados a la mayor o menor densidad emocional¹². De este modo, el primer grupo presenta los mayores niveles en las tres dimensiones; el tercero, los valores más bajos; y el segundo, unos niveles intermedios. La gran mayoría de las novelas escritas en la década de los sesenta, concretamente el 67%, pertenecen al primer clúster, es decir, siguen un patrón emocional que encaja con la visión más típica de la ciencia ficción. Por el contrario, la práctica totalidad de las obras escritas en la década de 2010, el 92%, se reparte entre los clústeres 2 y 3, es decir, presentan un patrón emocional claramente diferente, con una densidad emocional significativamente menor. Si asumimos que los premios Hugo son un buen reflejo de la ciencia ficción en general, algo que parece razonable dado el prestigio de este galardón, la conclusión es que se aprecia una clara evolución en el tiempo. Este género no solo ha ido modificando sus temáticas e incluso su orientación política (Oleszczuk, 2017), sino que desde un punto de vista estilístico parece haber mutado hacia obras con una menor densidad emocional en general, de modo que las emociones que se transmiten son menos positivas y con menor fuerza y acción. Ciertamente, este resultado es coherente con la creciente tendencia a que la ciencia ficción esté cada vez más ligada a problemas sociales, y menos a la acción y la aventura tan propias de las primeras obras de este género. O quizá, como sostienen Bould y Vint (2011: 1, cursiva añadida) «[no existe tal cosa como la ciencia ficción, sino múltiples y cambiantes formas de producir, comercializar, distribuir, consumir y entender los textos como ciencia ficción. La experiencia de un texto cambia de varias maneras una vez que se lo califica o no como ciencia ficción; y diferentes características del texto dominarán *en relación con la situación social e histórica del lector*]».

Una segunda aportación de este trabajo, más general y de carácter metodológico, es el haber comprobado que el *sentiment analysis* no solo puede resultar útil a efectos de identificación de géneros, como hacen una buena parte de los trabajos que optan por esta técnica, sino que resulta de utilidad para estudiar

12. Una interesante línea de investigación que queda abierta para futuros trabajos es evaluar si se observa un esquema de agrupación similar cuando se consideran las ocho emociones básicas de Plutchik (2009), esto es: alegría, confianza, miedo, sorpresa, tristeza, asco, ira y anticipación.

en profundidad y de un modo cuantitativo distintas obras de un mismo género.

4. Limitaciones del estudio

Este trabajo presenta dos limitaciones principales. La primera de ellas está relacionada precisamente con uno de los problemas más importantes de los algoritmos tipo *Bag-of-Words*. Al establecer una asignación unívoca de sentidos, la interpretación semántica del texto podría resultar imprecisa en pasajes en los que se niegue la palabra procesada, lo que sugeriría un sentido opuesto al identificado por el algoritmo. Este problema, a día de hoy, no está adecuadamente resuelto, pues presenta una enorme complejidad desde el punto de vista computacional. La aproximación más sencilla consiste en capturar bigramas (dos palabras) para identificar negaciones, pero esto requiere a su vez una limpieza de términos, como los verbos («do not be afraid» no es un bigrama: es necesario eliminar el verbo para que el algoritmo capture el bigrama con la negación «not afraid»). Además, no permite el tratamiento de las negaciones implícitas asociadas al uso de palabras como «forbid». Partiendo de este enfoque, el más sencillo posible y no demasiado efectivo, en la actualidad se está desarrollando una muy intensa actividad investigadora para encontrar soluciones satisfactorias. De hecho, existen paquetes de R como *SentimentR* (Rinker, 2019) que son capaces de calcular la polaridad considerando los *valence shifters*, es decir, los negadores, los amplificadores o intensificadores, los desamplificadores y las conjunciones adversativas. Este tipo de aproximaciones solventan parcialmente el problema, pero cuando el análisis no se limita a la polaridad, como sucede en este trabajo, sino que tienen en cuenta otras variables como *arousal* y *dominance*, la dificultad es mucho mayor. Hasta donde llega nuestro conocimiento, en la actualidad no existe ninguna solución realmente satisfactoria para lidiar con esta problemática, debido a la juventud de esta área de investigación.

La segunda limitación es que realmente no se ha realizado un estudio diacrónico desde 1960 hasta 2019, sino un estudio sincrónico dual de las décadas de 1960 y de 2010. Este es un matiz importante a la hora de afirmar que ha habido una evolución en el género. Como futura línea de investigación resultaría de interés llevar a cabo un estudio similar considerando todas las obras del periodo, a fin de verificar la robustez del resultado aquí alcanzado.

Agradecimientos

El autor agradece sinceramente a los dos evaluadores anónimos su minuciosa revisión del trabajo, así como sus comentarios, que sin duda han contribuido a mejorar sustancialmente el artículo.

Referencias bibliográficas

- Barceló, Miquel. 2015. *Ciencia ficción. Nueva guía de lectura*. Barcelona: Nova.
- Benoit, Kenneth, y Adam Obeng. 2020. *readtext: Import and Handling for Plain and Formatted Text Files*. R package version 0.76. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=readtext>.
- Bould, Mark, y Sherryl Vint. 2011. *The Routledge Concise History of Science Fiction (Routledge Concise Histories of Literature)*. Routledge. DOI: <https://doi.org/10.4324/9780203830161>
- Charrad, Malika, Nadia Ghazzali, Veronique Boiteau, y Azam Niknafs. 2014. «NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set». *Journal of Statistical Software*, 61, n.o 6: 1-36. DOI: <https://doi.org/10.18637/jss.v061.i06>
- Higgins, David M., y Roby Duncan. 2014. «Key Critical Concepts, Topics and Critics». En *The Science Fiction Handbook*, editado por Nick Hubble y Aris Mousoutzanis. Londres: Bloomsbury Publishing.
- Kassambara, Alboukadel, y Fabian Mundt. 2020. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=factoextra>.
- Kim, Evgeny, Sebastian Padó, y Roman Klinger. 2017. «Investigating the relationship between literary genres and emotional plot development». En *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Vancouver: Association for Computational Linguistics, pp. 17-26. DOI: <https://doi.org/10.18653/v1/W17-2203>
- Menadue, Christopher Benjamin, Kristi Giselsson, y David Guez. 2020. «An Empirical Revision of the Definition of Science Fiction: It Is All in the Techne...». *SAGE Open*, 10, n.o 4. DOI: <https://doi.org/10.1177/2158244020963057>
- Mohammad, Saif. 2018. «Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words». En *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*: 174-184. Melbourne: Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/P18-1017>
- Moreno, Fernando Ángel. 2013. *Teoría de la literatura de ciencia ficción* (Spanish Edition). España: Sportula Ediciones.
- Oleszczuk, Anna. 2017. «Sad and rabid puppies: Politicization of the Hugo Award nomination procedure». *New Horizons in English Studies*, 2, n.o 1: 127-135. DOI: <https://doi.org/10.17951/nh.2017.2.127>
- Pringle, David. 2000. «What is this Thing called Space Opera?». En *Space and Beyond: The Frontier Theme in Science Fiction*: 35-47, editado por Gary Westfahl. Westport: Greenwood Press.

- Plutchik, Robert. 2009. «Emotions: A general psychoevolutionary theory». En *Approaches to emotion*: 197-219, editado por Klaus R. Scherer y Paul Ekman. Hillsdale: Psychology Press.
- Reagan, Andrew J., Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, y Peter Sheridan Dodds. 2016. «The emotional arcs of stories are dominated by six basic shapes». *EPJ Data Science*, 5, n.o 1: 31. DOI: <https://doi.org/10.1140/epjds/s13688-016-0093-1>
- Rinker, Tyler W. 2019. *SentimentR: Calculate Text Polarity Sentiment*. R package version 2.7.1. Acceso 12 de mayo de 2021. <http://github.com/trinker/sentimentr>.
- Silge, Julia, y David Robinson. 2016. «tidytext: Text Mining and Analysis Using Tidy Data Principles in R». *Journal of Open Source Software*, 1, n.o 3: 37. DOI: <https://doi.org/10.21105/joss.00037>
- Suvin, Darko. 2016[1979]. *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre*. Bern: Peter Lang.
- Suvin, Darko. 2008. «Of Starship Troopers and Refuseniks: War and militarism in U.S. Science fiction, Part 1 (1945-1974: Fordism)». En *New Boundaries in Political Science Fiction*: 115-144, editado por Donald M. Hassler y Clyde Wilcox. Columbia: University of South Carolina Press.
- Nakazawa, Minato. 2018. *fmsb: Functions for Medical Statistics Book with some Demographic Data*, R package version 0.6.3. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=fmsb>.
- Wang, Xing, Shouhua Zhang, e Ivan Smetannikov. 2020. «Fiction Popularity Prediction Based on Emotion Analysis». *2020 International Conference on Control, Robotics and Intelligent System*: 169-175. DOI: <https://doi.org/10.1145/3437802.3437831>
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Nueva York: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-319-24277-4_9
- Wickham, Hadley. 2018. *scales: Scale Functions for Visualization*. R package version 1.0.0. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=scales>.
- Wickham, Hadley. 2019. *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=stringr>
- Wickham, Hadley y Lionel Henry. 2019. *tidyr: Tidy Messy Data*. R package version 1.0.0. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=tidyr>
- Wickham, Hadley, Romain François, Lionel Henry, y Kirill Müller. 2020. *dplyr: A Grammar of Data Manipulation*. R package version 0.8.5. Acceso 25 de marzo de 2021. <https://CRAN.R-project.org/package=dplyr>
- Wolfe, Gary K. 1986. *Critical Terms for Science Fiction and Fantasy. A Glossary and Guide to Scholarship*. Westport: Greenwood Press.

CV

**José Luis Arroyo Barrigüete**

Universidad Pontificia Comillas

jlarroyo@comillas.edu

Doctor en Estudios Filológicos por la Universidad de Sevilla, doctor en Ciencias Económicas y Empresariales por la Universidad Complutense y doctor en Ingeniería por la Universidad Politécnica de Madrid. Actualmente es profesor en la Universidad Pontificia Comillas, donde imparte docencia en Analytics.

