

**II**  
**REPÚBLICA**  
**Y LITERATURA**



## **OLLA, OBÚS, OLVIDO: ANÁLISIS COMPARATIVO DE LÉXICO EN LOS CUENTOS DE ARTURO BAREA / OLLA, OBÚS, OLVIDO: COMPARATIVE LEXICAL ANALYSIS IN ARTURO BAREA'S TALES**

**CARMEN TORRIJOS CARUDA**  
Universidad Autónoma de Madrid

Recibido: 15/04/2021

Aceptado: 25/06/2021

**Resumen:** Este artículo es una revisión de los cuentos de Arturo Barea desde un punto de vista de la analítica léxica. En primer lugar, se presenta el análisis de distintividad como herramienta informatizada que nos capacita para aplicar a los textos un análisis comparativo basado en las palabras que los construyen. Este análisis se enfoca en el recuento de las frecuencias de aparición de las palabras con el fin de caracterizar el texto a partir de datos objetivos, y de este modo hacer aflorar la motivación última del autor en el proceso creativo que conduce a la selección léxica consciente o inconsciente. En segundo lugar, se presentan los resultados de este análisis junto con un breve comentario, con el fin de comprobar la validez de la taxonomía elegida y la pertinencia del análisis realizado.

**Palabras clave:** cuento literario, análisis estilométrico, análisis comparativo de léxico, selección léxica.

**Abstract:** This article is a review of Arturo Barea's tales from a lexical analytical point of view. First, distinctiveness analysis is presented as a computerised tool that enables us to apply a comparative analysis to the texts based on their words. This analysis focuses on frequencies for words in a text, in order to characterise it on the basis of objective data, and thus bring to light the author's ultimate motivation in the creative process that leads to conscious or unconscious lexical selection. Lastly, the results of this analysis are presented together with a brief note in order to check the validity of the taxonomy design and the relevance of the analysis.

**Key words:** literary tales, stylometric analysis, comparative lexical analysis, lexical selection.

## Introducción

Arturo Barea, figura clave de la narrativa del exilio, es principalmente conocido por la trilogía *La forja de un rebelde*, que incluye tres novelas, *La forja*, *La ruta* y *La llama*. Principalmente caracterizadas por sus elementos autobiográficos, *La forja* se corresponde con su niñez y adolescencia en Madrid, *La ruta*, con su experiencia en la guerra del Riff y *La llama*, con la Guerra Civil, la última experiencia vivida por Barea antes de exiliarse a Inglaterra en 1939, con 42 años. *La forja de un rebelde* sería además objeto de una adaptación televisiva de seis capítulos emitida en 1990 por TVE.

Los cuentos de Arturo Barea conservan esta fuerte impronta autobiográfica, tan descriptiva de la situación histórica y social de su tiempo vital. Sin embargo, la habitual acogida irregular de la narrativa breve con respecto a la novela hizo que hubiera que esperar hasta el año 2001 para contar con una edición de sus cuentos completos. Históricamente y por motivos editoriales se han agrupado los cuentos de Barea en tres apartados, que corresponden en sus *Cuentos completos* a «Cuentos misceláneos» (16 cuentos), «Cuentos de valor y miedo» (20 cuentos) y «El centro de la pista» (14 cuentos). En el apartado «Cuentos misceláneos» aparecen mezclados cuentos de la guerra de Marruecos, cuentos de exilio en Inglaterra y dos descubrimientos sorprendentes: un cuento pseudofantástico de inspiración directa en los relatos fantásticos del siglo XIX, en el estilo de «El monte de las ánimas» o «La cruz del diablo», y dos cuentos futuristas, de ciencia ficción, que presentan una historia hilada. Los mismos personajes se repiten en estos dos cuentos y en un tercero que, por alguna razón desconocida, no aparece incluido aquí sino en «El centro de la pista». «Cuentos de valor y miedo» es el apartado más unitario, dado que todos son bien cuentos militares de batalla y trinchera, o bien cuentos de ciudad muy ligados a la guerra. Se publicó como un volumen en Barcelona en 1938, inmediatamente antes de la entrada de las tropas franquistas. Barea vivió su infancia con su madre en el barrio de Lavapiés, por lo que los cuentos de este apartado establecen un vínculo innegable con la ciudad de Madrid, con la vida del barrio, con las gentes y con el impacto generacional de la guerra, como afirma Julio Llamazares en su artículo de *El País*, «Barea»:

Español por encima de todo, republicano y rebelde hasta su final, el hijo de la lavandera que veía hincharse con el viento los pantalones de los soldados que su madre colgaba de los tendederos después de lavarlos en el Manzanares y que apenas pudo estudiar a causa de su pobreza es el autor de una trilogía que más que contar su vida cuenta la de su generación, una generación castigada y diezmada por una guerra y una dictadura cuyos efectos aún son perceptibles hoy (Llamazares, J., 6 de enero de 2018).

«Cuentos de valor y miedo» es también el apartado que da lugar a la valoración que leemos en la introducción de Nigel Townson a los *Cuentos completos*, donde el investigador afirma que los cuentos de Barea estaban fuertemente orientados a ser un instrumento más en la defensa del proyecto republicano frente al golpe militar:

En otras palabras, él consideraba sus relatos, al igual que sus programas radiofónicos, como un medio más de lucha contra el enemigo. En consecuencia, muchos de los cuentos son en gran parte sketches propagandísticos que exaltan la causa republicana en conflicto con las fuerzas «fascistas» de Franco. Por tanto, las historias no hablan de las limitaciones y divisiones dentro del campo republicano. En concreto, Barea no hace referencia a su relación cada vez más difícil con el PCE, la cual desembocó en su decisión de abandonar España en febrero de 1938 (Barea, A., 2001:7).

Para comprender esta afirmación hay que contar con datos biográficos esenciales: Barea trabajó durante la Guerra Civil como responsable del servicio de prensa extranjera desde la Telefónica de Gran Vía, a las órdenes del gobierno republicano. Él mismo realizaba, además, emisiones radiofónicas clandestinas conocidas como *La voz incógnita de Madrid*, en las que narraba los sucesos y el avance de la irrupción fascista. Un impulso periodístico de tal intensidad no le permitía descansar de la labor de reporte ni siquiera cuando se encontraba en los territorios de la ficción, por lo que muchos cuentos narran instantes de la lucha, escenas cotidianas y acontecimientos al modo de un reportero en las calles. Cabe, por otra parte, la posibilidad de que esa fuera la auténtica voluntad de Barea tras los cuentos: ilustrar instantes de realidades propias y ajenas marcadas exclusivamente por la visión republicana. Si bien en el exilio inglés continúa trabajando en radio, nada menos que en su propio espacio de la BBC, todo apunta a que había convertido la escritura de cuentos en una actividad terapéutica. Su matrimonio con la periodista Ilse Kulcsar es también un hecho a tener en cuenta en este impulso protector de la imagen de la República:

Tras un primer matrimonio fallido, Arturo se topó con Ilse Kulcsar, una periodista austriaca que también trabajaba en el edificio. Ilse era judía, activista y comunista. Había llegado a España con Leopold, su marido, para apoyar la causa republicana. Ilse colaboró con Arturo en que la imagen de la República de cara al extranjero fuera intachable. Y como pasa en las historias de celuloide, mientras el mundo se derrumbaba los dos se enamoraron (Campoy, A., marzo de 2017).

Por su parte, «El centro de la pista» se publicó póstumamente en 1960, también como un solo volumen. Recoge cuentos que no tienen entre sí ninguna similitud temática, pero que obedecen a una faceta de Barea más lírica, ligeramente más contemplativa de una realidad que, con todo, tampoco se permite desvincular de la guerra.

Más allá de esta clasificación de sus cuentos, que como hemos dicho obedece a motivos editoriales y de publicación, el denominador común es que todas las historias hablan de España en los años de la guerra o de españoles exiliados, estos últimos cuentos muy marcados por el exilio inglés del autor. El objetivo de este estudio es confirmar, tras la lectura de los *Cuentos completos*, la intuición en el lector de que otra clasificación es posible para los cuentos de Arturo Barea, más ceñida a criterios temáticos y de ambientación.

## Metodología

El análisis estilométrico se ocupa de la relación entre el estilo de escritura de un texto y los metadatos que lo acompañan, como la fecha, el género o la autoría (Eder, Rybicki y Kestemont, 2016:1). No obstante, la principal aplicación de este tipo de análisis ha sido, históricamente, la de identificar o confirmar la autoría de un texto (Calvo Tello y Cerezo Soler, 2018:1) mediante la búsqueda de patrones estilísticos similares en otras obras del autor o los autores candidatos. A medida que los avances en las tecnologías informáticas lo han permitido, esta búsqueda de patrones ha podido optimizarse y alcanzar una mayor sofisticación, sirviéndose de técnicas de ciencia de datos y procesamiento de lenguaje natural para hacer búsquedas más eficientes, facilitando así el trabajo de los investigadores. Por ejemplo, en estudios avanzados, la atribución autorial puede realizarse también mediante el modelado de tópicos (Seroussi, Zukerman y Bohnert, 2014:270). Esta técnica consiste en someter los textos a un procesamiento automático para extraer y contabilizar frecuencias de palabras que se agrupan después en diferentes tópicos o temas. Esto permite caracterizar el texto a un nivel más profundo y analítico del que es capaz el ojo humano, descubrir cuáles son los temas principales tratados de forma no evidente, qué palabras los reflejan y en qué medida estos pueden coincidir con los temas de otras obras confirmadas del mismo autor. Otro ejemplo de ciencia de datos aplicada a la estilometría es el estudio de Vélez y Marmolejo-Ramos sobre *Cien años de soledad*, de Gabriel García Márquez, en el cual los autores se proponen ratificar de manera empírica las repetidas afirmaciones de la crítica literaria en torno a los principales rasgos estilísticos de la novela, como la alusión constante a los personajes, la narración de eventos en tono neutro o el nivel de concreción de las palabras (Vélez y Marmolejo-Ramos, 2016:271-273). Con diferente propósito, en los últimos años se han aplicado también técnicas estilométricas a la poesía de la llamada generación *millennial* con el fin de calcular diferentes índices de

riqueza léxica y estructural, en un intento por cuantificar la calidad de los poemas y la naturaleza de su relación con el éxito de público (Sánchez, 2018:176). En concreto, el análisis de variedad en el vocabulario se ha aplicado con frecuencia en estudios que comparan la disponibilidad léxica de estudiantes de español como lengua extranjera y como lengua materna (Sifrar Kajan, 2012:1), o como herramienta de evaluación del nivel de competencia lingüística del profesorado en lengua extranjera (Santos Díaz, 2015:1).

El análisis comparativo de léxico puede utilizarse también como herramienta estilométrica. En un nivel general, se trata de un procedimiento informatizado que trabaja con el texto en formato digital y que tiene como base la frecuencia de aparición de las palabras. No obstante, como afirman Moreno y Guirao, es clave la diferenciación entre análisis de frecuencias, cuántas veces aparece una palabra en un texto, y análisis de distintividad, cuántas veces aparece una palabra en ese texto en comparación con otros, por lo que no son necesariamente equivalentes:

En contraposición a la frecuencia oponemos el concepto de distintividad, entendido como unidad léxica que es característica y distintiva de una variedad porque aparece típicamente en dicho registro y no tanto en otros. Informativamente, una unidad distintiva destaca porque proporciona más información sobre el contenido y contexto. Igualmente, la probabilidad de que dicho lema aparezca en un determinado registro es más alta que en otros contextos (Moreno y Guirao, 2009:196).

El trabajo de estos autores cuenta entre sus bases con la teoría expuesta por Ted Dunning en 1993, según la cual la manera de saber si una palabra es característica o distintiva de un texto no es comparar su frecuencia de aparición en ese texto con su frecuencia habitual en el corpus general del idioma:

Statistics based on the assumption of normal distribution are invalid in most cases of statistical text analysis unless either enormous corpora are used, or the analysis is restricted to only the very most common words (that is, the ones least likely to be of interest). This fact is typically ignored in much of the work in this field. Using such invalid methods may seriously overestimate the significance of relatively rare events (Dunning, 1993:71).

Lo que Dunning sugiere es que, por una parte, para establecer esta frecuencia supuestamente habitual o normal en la lengua es necesario partir de grandes volúmenes de texto que además sean ampliamente representativos, y por otra parte, este análisis puede no ser justo con palabras que son altamente distintivas de ese texto, pero que aparecen en él en una baja frecuencia. El análisis de distintividad por frecuencias contrastadas necesita, en cambio, únicamente dos textos que comparar, y no actúa en perjuicio de las palabras menos frecuentes:

The practical effect of this improvement is that statistical textual analysis can be done effectively with very much smaller volumes of text than is necessary for conventional tests

based on assumed normal distributions, and it allows comparisons to be made between the significance of the occurrences of both rare and common phenomena (Dunning, 1993:65).

La principal aplicación del análisis de distintividad suele ser la contraposición de textos, que puede darse en diferentes modalidades: un texto aislado frente a otro, un texto aislado frente a un conjunto de textos, o un conjunto de textos frente a otro. El objetivo es conocer cuáles son las palabras más distintivas del texto o conjunto de textos base en contraposición al resto, y por tanto cuál es su caracterización, qué temas distinguen su discurso. Al enfrentar dos textos, el recuento contrastado de frecuencias de aparición en un texto y en otro puede arrojar luz sobre el grado de representatividad o distintividad de la palabra en el texto en el que se encuentra, es decir, en qué medida la aparición de esta palabra caracteriza y localiza el texto, y operaría del mismo modo con conjuntos de textos. Llegamos así a la conclusión de que una palabra no es distintiva de un documento porque aparezca muchas veces en él, sino porque aparece en él y no en el resto de documentos que participan de la comparación. Del mismo modo, esto aumenta la probabilidad matemática para esa palabra de aparecer en un contexto similar, lo que conlleva un análisis estadístico del tipo: basándonos en sus frecuencias de aparición en determinados contextos temáticos, la probabilidad para esta palabra de aparecer en este texto es mayor que en otros textos. Y así, si una palabra tiene la misma probabilidad de aparecer en todos los textos objeto de estudio, entonces su grado de distintividad para con estos textos es muy bajo o ninguno. La herramienta ACL, desarrollada por el Instituto de Ingeniería del Conocimiento y disponible bajo demanda<sup>1</sup>, nos permite automatizar este análisis de distintividad procesando los textos mediante técnicas avanzadas de procesamiento de lenguaje natural. Como ejemplo de uso, se plantea una interesante comparación de la polémica entre Quevedo y Góngora a partir del análisis comparativo de léxico de sus textos (Moreno Sandoval, 2012).

## Análisis

La motivación de utilizar el análisis de distintividad sobre los *Cuentos completos* de Arturo Barea es proponer una nueva clasificación de sus cuentos basada exclusivamente en criterios temáticos. Para ello, se ha seguido la metodología de comparación de frecuencias de palabras expuesta anteriormente: el número de

---

<sup>1</sup> <http://innova.iic.uam.es/acl/index.php>.

apariciones en un conjunto de textos va a determinar el poder distintivo de una palabra para ese conjunto, siempre en comparación con el número de apariciones en el conjunto opuesto. Este estudio precisa, por tanto, de una herramienta informatizada capaz de calcular y comparar de manera automática la capacidad distintiva de las palabras en los textos, para lo cual utilizaremos la herramienta ACL del Instituto de Ingeniería del Conocimiento, descrita en el apartado anterior.

Con el fin de realizar el análisis de léxico distintivo, se llevó a cabo una segmentación manual previa de los textos con un criterio estrictamente temático que pudiera después ser respaldado por el resultado del análisis, esto es: comprobar si las palabras más distintivas de cada categoría de cuentos, mostradas aisladamente, corresponden efectivamente al reflejo de su contenido semántico. Tras la lectura, los cuentos fueron reclasificados de manera intuitiva siguiendo una taxonomía basada en un primer nivel con cuatro categorías:

1. Cuentos de guerra,
2. cuentos de ciudad,
3. cuentos de exilio y
4. cuentos futuristas.

El cuento pseudofantástico, por su ambientación, cumple los criterios para ser incluido en cuentos de ciudad. Como consideraciones preliminares, hay que tener en cuenta que un alto porcentaje de los cuentos respira la Guerra Civil y las causas del exilio a pesar de estar ambientado en otro tiempo y otro lugar, incluso los más costumbristas o los cuentos de infancia. Hay que señalar también que en buena parte de los cuentos estamos trabajando con traducciones de los textos originales escritos en inglés, e incluso con traducciones al español de los textos originales previamente traducidos al inglés por Ilse Kulcsar.

En este caso vamos a aplicar el análisis de distintividad a los cuentos de Arturo Barea utilizando la modalidad que permite oponer dos conjuntos de textos, comparando cada una de las cuatro categorías temáticas por separado frente al conjunto completo de los cuentos. Esto es, por ejemplo, el grupo de cuentos clasificados como cuentos de guerra frente a todos los cuentos completos, y los primeros incluidos en los segundos. Efectuado el análisis, se obtiene una lista de las palabras más distintivas de cada categoría. Hay que señalar que de la lista resultante de palabras distintivas se han excluido prácticamente todas las palabras gramaticales: preposiciones, determinantes, conjunciones, pronombres y cuantificadores han quedado fuera del estudio al no poder aportar significado



léxico para contribuir a la caracterización de los textos. Se han excluido también casi todos los nombres propios, que son exclusivos y por tanto distintivos por naturaleza de cada cuento, no pudiendo aportar tampoco información que nos permitiera diferenciar un texto de otro en el aspecto temático.

Trasladamos por tanto a los resultados únicamente las palabras léxicas, incluyendo el resultado en crudo de la herramienta y un resumen a continuación (en **negrita**) de las palabras más relevantes para el análisis cualitativo. Podemos ver así que las palabras más distintivas de los cuentos de guerra frente a los cuentos completos constituyen una lista en cierto modo predecible:

ángel	37	41
trinchera	35	37
manzanares	27	27
nalguitas	23	23
sargento	35	54
moros	22	25
capitán	31	47
soldado	23	29
soldados	18	19
miguel	18	19
parapeto	17	17
rosa	18	20
sacos	14	14
meses	21	31
posición	16	20
casas	20	30
burro	14	16
cazadores	13	14
allí	62	157
sancho	12	12
balas	12	12
mosca	14	18
hombres	35	76
romero	10	10
pollo	10	10
máquina	16	24
camino	20	35
carabanchel	11	13
terrereros	9	9
rata	9	9

Figura 1: Tabla de frecuencias de palabras en los cuentos de guerra

**Trinchera - sargento - moros - capitán - soldado - parapeto - sacos - meses - posición - casas - burro - cazadores - balas - mosca - hombres - pollo - máquina - camino - terreros - rata**

En la imagen podemos observar tres columnas: la primera columna es la palabra, la segunda columna es el número de apariciones en la categoría específica de cuentos de guerra y la tercera columna es el número total de apariciones en los

*Cuentos completos*. Podemos ver así que algunas palabras únicamente aparecen en los cuentos de guerra, como *balas*, que aparece 12 veces en los cuentos de guerra y ese mismo es su total de apariciones en los cuentos completos, por lo que su grado de distintividad para caracterizar los cuentos de guerra es muy alto. Más allá de los elementos previsible, como los grados militares (*sargento, capitán, soldado*), destaca el papel de los animales (*burro, mosca, pollo, rata*) en la vida de la trinchera, y también la palabra *meses*, la única en este grupo relacionada con el paso del tiempo. Durante la guerra, es posible que el tiempo pasara muy rápido para contar en semanas y muy despacio para contar en años, por lo que los meses se convirtieron en la unidad de medida, en la expectativa inmediata y en la cuenta retroactiva de la duración de las penurias al mirar hacia atrás.

Ahora que conocemos el funcionamiento del análisis, podemos pensar durante un momento, antes de leer el resultado, qué palabras esperaríamos como más distintivas de los cuentos de ciudad.

encarnita	52	52
agua	107	135
madre	106	135
don	85	105
antonio	52	56
mamá	45	47
curro	42	44
señora	57	66
conchi	34	34
dinero	55	65
abuela	32	32
puerta	63	80
tío	37	41
le	446	741
tinaja	25	25
johnny	25	25
conde	25	25
rufo	27	28
tía	28	30
paula	22	22
pesetas	31	35
payaso	24	25
número	24	25
manuel	21	21
lupe	21	21
médico	28	31
toñín	20	20
juana	20	20
chicos	28	32
circo	21	22
hermano	23	25
luis	18	18
juez	18	18
señores	22	24
céntimos	22	24
charito	17	17
diez	30	36
mi	247	407
triana	21	23
mister	21	23
serafín	16	16
fuentes	18	19

Figura 2: Tabla de frecuencias de palabras en los cuentos de ciudad

**Agua - madre - mamá - señora - dinero - abuela - puerta - tinaja - tío - tía - pesetas - número - médico - chicos - monedas - hermano - juez - señores - céntimos - fuente**

En este grupo de cuentos, las personas emergen como parte importante de la vida en la ciudad, muestra de ello es el alto número de nombres propios. Muchos son cuentos de infancia, en los que los protagonistas son niños de corta edad que nombran y describen en su discurso el universo que les rodea (*señora, señores, chicos*), con especial foco en la familia (*hermano, tío*) y con un ligero predominio de los roles femeninos, propios de los cuidados en tiempo de guerra (*madre, mamá, abuela, tía*). Todas las palabras reflejan cotidianidad: *puerta, fuente, tinaja*. Por otro lado, mientras que en la batalla pierde importancia el dinero, en la ciudad supone uno de los temas más distintivos de los cuentos (*dinero, pesetas, monedas, céntimos*).

Podemos pensar ahora qué palabras esperaríamos de los cuentos de exilio, si bien resultan mucho menos predecibles:

teresa	27	32
padre	48	115
peter	20	20
carlos	21	26
había	133	589
iglesia	24	42
tony	18	29
su	190	1069
verde	15	30
dios	22	62
azul	16	37
fe	12	23
taylor	7	7
demasiado	15	35
personas	14	31
nunca	31	112
mi	81	407
tabernero	9	14
era	101	547
tortilla	6	6
servando	6	6
higado	6	6
columpios	6	6
nuestra	18	53
sartén	10	19
quizá	9	16
ella	45	204
todo	50	234
jóvenes	6	7
saunders	5	5
j.	5	5
crueldad	5	5
cobardía	5	5
rojos	7	11
jardín	7	11
temía	6	8
quizás	6	8
supuesto	7	12
que	501	3452
inglaterra	9	20
flores	12	33
probablemente	5	6
bicicleta	5	6
mujer	23	89
demás	12	34
país	8	17
negro	21	80
niños	10	26
vicario	4	4

Figura 3: Tabla de frecuencias de palabras en los cuentos de exilio

**Padre - iglesia - verde - dios - azul - fe - personas - tabernero - tortilla - hígado - columpios - sartén - jóvenes - crueldad - cobardía - rojos - jardín - Inglaterra - flores - bicicleta**

Con el uso frecuente de estas palabras, confirmamos ideas recurrentes en los cuentos que responden a los pensamientos, también recurrentes, del Arturo Barea exiliado: la culpabilización de la iglesia, la añoranza de la gastronomía y la idealización de Inglaterra. Este fue su lugar de exilio último, retratado en los cuentos como lugar de jardines y flores, bicicletas y columpios, en contraste con la España gris y pobre vivida en los años de guerra. Dos sustantivos abstractos coronan el análisis del exilio desde el exilio: *crueldad* y *cobardía*. En cuanto a la palabra *rojos*, curiosamente y en contra de lo que podríamos esperar, sólo en dos de las siete apariciones en los cuentos de exilio se utiliza como sustantivo, mientras que en el resto de los casos alude efectivamente al color.

Y, por último, las palabras distintivas de los cuentos futuristas:

smilton	32	32
miria	30	30
dr.	18	18
laboratorio	18	19
blanca	25	47
negra	24	45
ken	16	16
consejo	17	23
negros	19	32
color	19	32
miedo	33	101
blancos	18	29
raza	15	19
bakuko	13	13
piel	23	54
blanco	23	56
smojden	12	12
razas	12	12
dakar	12	12
may	11	11
negro	25	80
tu	21	59
caso	17	42
doctor	9	9
cies	9	9
supremo	9	10
hospital	12	22
sáhara	8	8
mary	8	8
productos	8	9
genética	8	9
director	11	21
sector	9	13
tí	14	36
rebotson	7	7
leyes	9	14
orden	12	27
enfermedad	8	11
secreto	10	19
epidemia	7	8
casos	8	12
mundo	23	97
fue	29	141
químicos	6	6
biotic	6	6
miles	8	13
experimento	6	7

Figura 4: Tabla de frecuencias de palabras en los cuentos futuristas

**Doctor - laboratorio - blanca - negro - consejo - color - miedo - raza - piel - casa - supremo - hospital - productos - genética - director - sector - leyes - orden - enfermedad - secreto - epidemia**

Sabemos por la lectura que los cuentos están ambientados en el futuro, pero no podríamos afirmarlo únicamente leyendo sus palabras distintivas ya que estas reflejan principalmente el argumento de las obras, más que un contexto general de ciencia-ficción. En este caso, quizá la intención última del autor no es escribir sobre el futuro o las posibilidades de la ciencia. Los cuentos hablan de la alteración genética, del cambio de color en la piel de las personas a través de un experimento cuidadosamente diseñado y extendido en la sociedad con el fin de acabar con la supremacía de la raza blanca, en un esfuerzo mundial de concienciación antirracista. Barea utiliza la ciencia ficción y el futuro, como ocurre a menudo, como un medio para alertar sobre los conflictos racistas en el presente del siglo XX o sobre los peligros de la incipiente ingeniería genética.

Hemos analizado ya el primer nivel de la taxonomía. Los cuentos de ciudad son, tras la lectura, los que presentan una mayor diversidad interna, por lo que podemos ampliar el análisis estableciendo un segundo nivel dentro de esta categoría. El objetivo en este caso es distinguir dentro de los cuentos de ciudad entre aquellos que podemos llamar de inspiración obrera y aquellos que podemos llamar de ambientación burguesa. La principal y única consideración es que esta distinción es osada, y debe tenerse en cuenta que tanto el concepto de clase obrera como el concepto de burguesía están aún en debate, dependiendo de la posición de la frontera, en gran medida, del esquema ideológico seguido.

Estableciendo como texto de base para la comparación los cuentos obreros, hemos extraído sus palabras distintivas pero esta vez cambiando el esquema del análisis: en lugar de enfrentar los cuentos clasificados en esta categoría al total de los *Cuentos completos*, en este caso enfrentamos exclusivamente los textos obreros con los textos burgueses, con el fin de determinar qué elementos léxicos tienen mayor probabilidad de aparecer en un grupo y no en otro, y por tanto cuáles son los que distinguen a una categoría de otra. Las palabras distintivas de los cuentos obreros son las siguientes:

la	1572	882
agua	89	18
abuela	31	1
madre	86	20
pueblo	30	1
señora	50	7
antonio	45	7
hermano	22	1
papá	15	1
mano	45	13
cura	23	4
pobres	13	1
fuentes	16	2
Madrid	50	17
muy	82	34
abajo	26	6
treinta	12	1
monedas	12	1
donde	41	13
chicos	23	5
pan	20	4
medio	20	4
soy	14	2
ha	70	30
veinte	16	3
zapatero	10	1
estalló	10	1
pues	18	4
corriendo	18	4
sentía	15	3
juez	15	3
iglesia	15	3
dado	15	3
médico	22	6
suelo	24	7
portal	12	2
manantial	12	2
patio	9	1

Figura 5: Tabla de frecuencias de palabras en los cuentos de ciudad de inspiración obrera

**Agua - abuela - madre - pueblo - señora - hermano - papá - mano - cura - pobres - fuentes - Madrid - monedas - pan - zapatero - juez - iglesia - médico - suelo - portal**

Recordemos que la primera columna corresponde a la palabra, la segunda columna al número de apariciones de esa palabra en los cuentos de inspiración obrera y la tercera columna al número de apariciones en los cuentos de

ambientación burguesa, con el fin de poder establecer la comparación. No obstante, si invertimos el experimento y buscamos qué palabras distinguen a los cuentos de burguesía de los cuentos obreros, utilizando esta vez los cuentos burgueses como base de la comparación, la lista cambia considerablemente:

tío	33	4
toro	14	1
toros	13	2
mil	13	2
hotel	13	2
él	81	62
cuenta	22	8
nariz	10	1
después	72	54
señores	17	5
cabo	15	4
hijo	19	7
mío	9	1
sudor	11	2
pista	11	2
mundo	25	12
eso	21	9
mi	127	120
público	10	2
oficio	8	1
gloria	8	1
fuera	29	17
dijo	53	41
ser	38	26
gran	20	10
tiempos	9	2
oro	9	2
trabajar	7	1
marido	7	1
línea	7	1
corridas	7	1
comenzaron	7	1
banco	7	1
capitán	12	4
así	48	37
joven	16	7
qué	67	58
luis	13	5
viejo	27	17
sólo	42	32
le	214	232
ella	57	48
historia	21	12
buen	21	12
libros	8	2
borracho	8	2
amor	8	2
eres	11	4
mujer	25	16
no	377	439
última	6	1
secretario	6	1
quieren	6	1
dejaron	6	1
flores	12	5
traje	9	3

Figura 6: Tabla de frecuencias de palabras en los cuentos de ciudad de ambientación burguesa

**Toros - hotel - nariz - señores - mundo - oficio - gloria - oro - marido - corridas - banco - capitán - joven - libros - borracho - amor - mujer - secretario - flores - traje**

En primer lugar, el segundo nivel de la taxonomía es respaldado por el hecho de que no hay apenas palabras coincidentes entre ambas listas, lo cual confirma la diferenciación semántica en la inspiración de los cuentos. En los cuentos obreros toman importancia la familia, el dinero, la comida, el agua, la iglesia, el médico: la supervivencia, la vida básica. En la segunda lista, aparecen ante nosotros palabras, y por tanto conceptos, no vistos hasta el momento en ninguna de las listas anteriores: los toros, las corridas, un hotel, la gloria, el oro, el banco, un borracho, el amor, la mujer, un traje. Otra esfera de la realidad que parece preservada de la guerra o la pobreza, conservada en la normalidad con preocupaciones a distintos niveles. Como ocurría ya en el primer nivel, el análisis comparativo de léxico nos muestra de nuevo la pertinencia de la taxonomía establecida y nos confirma la posibilidad de otra clasificación para estos cuentos. Nos proporciona, además, una base más sólida para los argumentos a favor de una lectura extrínseca de los textos, como por ejemplo en la conclusión de que efectivamente Barea buscaba, con su actividad cuentística, no sólo calmar el dolor del exilio sino dejar una marca social.

## Conclusiones

Sabíamos ya que la brevedad del cuento tiene un poder intensional. Aplicar este mismo análisis a una novela, donde por la extensión del texto y el contexto las mismas palabras pueden aparecer un alto número de veces, puede ser traicionero, no arrojar ninguna luz o confundir la interpretación. Sin embargo, la aplicación de este análisis aporta especial valor en el formato del cuento, donde la extensión está fuertemente restringida y la voluntad creativa cobra una importancia capital teniendo en cuenta, además, que la selección léxica en un autor puede ser tanto un proceso creativo consciente como inconsciente, difícil de revelar en una lectura lineal. En todo caso, el análisis comparativo de léxico nos proporciona una visión más amplia de la obra literaria, despegada de la subjetividad, que puede aproximarnos a sus claves desde un ángulo insospechado. En definitiva, una herramienta de extracción de distintividad léxica nos permite, en el peor de los casos, confirmar intuiciones, y en el mejor de los casos, la extracción de



conclusiones que enriquezcan el análisis cualitativo. Qué mejor que tres palabras, *olla*, *obús* y *olvido*, para resumir un volumen de cuentos, o más bien la vida, de un autor que nos habla sobre la batalla, la ciudad y el exilio.

## BIBLIOGRAFÍA

ACL: Herramienta de análisis comparativo de léxico. Desarrollada por el Instituto de Ingeniería del Conocimiento y la Universidad Autónoma de Madrid. Disponible bajo demanda en: <http://innova.iic.uam.es/acl/> [Consultado 02/06/2021].

Barea, A. (2001). *Cuentos completos*. Madrid: Debate.

Calvo Tello, J. y Cerezo Soler, J. (2018). «La conquista de Jerusalén ¿de Cervantes? Análisis estilométrico sobre autoría en el teatro del Siglo de Oro español» en *Digital Humanities Quarterly*, Tomo 12, N.º 1. Providence. Recuperado de: <https://www.proquest.com/openview/98767f47677420f01f4902beb8b238cd/1?pq-origsite=gscholar&cbl=5124193> [Consultado 06/11/2021].

Campoy, A. (Marzo de 2017). «Arturo Barea, la forja de una memoria». *Jot Down*. Recuperado de: <https://www.jotdown.es/2017/03/arturo-barea-la-forja-una-memoria/> [Consultado 09/11/2021].

Dunning, T. (1993): «Accurate methods for the statistics of surprise and coincidence». *Computational Linguistics*, 19 (1): 61-74. Recuperado de: <https://aclanthology.org/J93-1003.pdf> [Consultado 13/11/2021].

Eder, M., Rybicki, J., y Kestemont, M. (2016) “Stylometry with R: A Package for Computational Text Analysis”. *The R Journal* Vol. 8/1. Recuperado de: <https://journal.r-project.org/archive/2016-1/eder-rybicki-kestemont.pdf> [Consultado 06/11/2021].

Llamazares, J. (5 de enero de 2018). «Barea». *El País*. Recuperado de: [https://elpais.com/elpais/2018/01/05/opinion/1515152559\\_203725.html](https://elpais.com/elpais/2018/01/05/opinion/1515152559_203725.html) [Consultado 09/11/2021].

Moreno, A. (2012). «Análisis de la polémica entre Góngora y Quevedo, parte 1». En *Blog*. Instituto de Ingeniería del Conocimiento. Recuperado de: <https://www.iic.uam.es/noticias/analisis-de-la-polemica-entre-gongora-y-quevedo-parte-1/> [Consultado 09/11/2021].

- Moreno, A. y Guirao, J. M. (2009). «Frecuencia y distintividad en el uso lingüístico: casos tomados de la lematización verbal de corpus de distintos registros». *A survey of corpus-based research*, 195-210. Madrid: Universidad Autónoma de Madrid. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=4805635> [Consultado 09/11/2021].
- Sánchez, F. J. (2018). «Análisis de la riqueza léxica de los poetas “millennial”. Primera aproximación». En Sánchez García, M.R. (Coord.), *Nuevas poéticas y redes sociales: Joven poesía española en la era digital*, 175-186. Madrid: Siglo XXI.
- Santos Díaz, I. (2015). «Análisis comparativo del léxico en español y en lengua extranjera del futuro profesorado». En *e-AESLA. Revista Digital*, Universidad de Málaga. Recuperado de: [https://www.researchgate.net/publication/299470676\\_Analisis\\_comparativo\\_del\\_lexico\\_en\\_espanol\\_y\\_en\\_lengua\\_extranjera\\_del\\_futuro\\_profesorado](https://www.researchgate.net/publication/299470676_Analisis_comparativo_del_lexico_en_espanol_y_en_lengua_extranjera_del_futuro_profesorado) [Consultado 09/11/2021].
- Seroussi, Y., Zukerman, I., y Bohnert, F. (2014). “Authorship Attribution with Topic Models”. En *ACL Anthology* 40 (1), 269-310. Recuperado de: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.234&rep=rep1&type=pdf> [Consultado 06/11/2021].
- Sifrar Kajan, M. (2012). «Análisis comparativo de la disponibilidad léxica en español para estudiantes de español como lengua extranjera (ELE) y como lengua materna (ELM)». En *Revista de didáctica ELE*, nº 15. Recuperado de: [https://marcoele.com/descargas/15/sifrar-disponibilidad\\_lexica.pdf](https://marcoele.com/descargas/15/sifrar-disponibilidad_lexica.pdf) [Consultado 09/11/2021].
- Vélez, J. I. y Marmolejo-Ramos, F. (2016). «Los secretos de *Cien años de soledad*: Una aproximación estilométrica para la investigación en psicolingüística». En *Revista Colombiana de Psicología*, 25(2), 265-288. Recuperado de: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0121-54692016000200006&lng=es&nrm=iso&tlng=es](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-54692016000200006&lng=es&nrm=iso&tlng=es) [Consultado 06/11/2021].