

# UN ESTUDIO LEXICOMÉTRICO DE UN TEXTO LITERARIO CON PROGRAMAS INFORMÁTICOS

ÁNGELES SÁNCHEZ HERNÁNDEZ  
Universidad de Las Palmas de Gran Canaria

## 1. INTRODUCCIÓN

Hablar, a comienzos del siglo XXI, de nuevas tecnologías refiriéndonos al ordenador podría parecer ya un poco trasnochado cuando una máquina de estas características cuenta con más de medio siglo de antigüedad. Estudios que datan de diez años atrás ya señalaban que el 60% de los universitarios especialistas en Humanidades utilizaban un tratamiento de textos para sus trabajos (Meyer, 1991:134). Su utilidad es, por tanto, muy clara para investigadores cuya tarea profesional se relaciona directamente con la palabra como forma de expresión de sus ideas. Los programas de tratamiento de textos han tenido una gran repercusión para la elaboración y presentación de trabajos. No obstante, su empleo es aún reducido cuando se trata de aplicar la herramienta informática en el proceso de investigación literaria, a pesar de poder ofrecer beneficios importantes para su desarrollo.

En la actualidad, existen numerosas bases de datos en las que se pueden encontrar los textos ya *digitalizados*. En concreto, en lengua francesa encontramos *FRANTEXT* que es, según M. Bernard (1999:21), el banco de datos textuales más importante del mundo con una colección de unos 3.600 textos íntegros que ha sido elaborado a lo largo de más de treinta años por el *Institut National de Langue Française*. Los textos son literarios en su mayoría aunque también constan algunos otros técnicos, administrativos o filosóficos, y el interesado debe estar abonado a esta base de datos para poder acceder a ellos. Puede hallarse también una versión abreviada que se ofrece en *CD-ROM* denominada *DISCOTEXT* pero está limitada a unos 534 textos literarios publicados entre 1827 y 1923.

Sin embargo, para nuestra investigación no hemos podido servirnos de bases de datos ya establecidas, puesto que el texto que hemos elegido es *Citadelle* de Antoine de Saint-Exupéry que no se encontraba digitalizado sino que ha sido tarea nuestra ese traslado de información de un tipo de soporte a otro. El trabajo que presentamos a continuación trata de mostrar un modelo de investigación lexicométrica fundamentada en el recuento de palabras, y cómo ésta puede verse facilitada enormemente por el uso de un programa informático. Este estudio numérico de los vocablos hace extraer, posteriormente, unos datos de los que se deducen características estilísticas, método que ya utilizaban los clásicos. El libro elegido para ello guarda una buena cantidad de dificultades por tratarse de una obra póstuma, inacabada, de gran extensión y de características, aparentemente, muy distintas de las que se atribuyen al resto de la producción literaria del mismo autor.

Este trabajo abarca varias partes. En una primera fase, se trata de trasladar el texto con el que vamos a trabajar del soporte papel en el que está editado originariamente a un soporte numérico, es decir, debemos trasformarlo en una serie de códigos legibles para la máquina. Así se pueden llevar a cabo todas las fases posteriores. Empleamos el programa

FRECON.UIB<sup>1</sup> con el que se elaboran los llamados *documentos lexicométricos*. Las posibilidades son variadas: glosarios, concordancias, inventarios, tablas léxicas y tablas de segmentos repetidos. Nos interesan particularmente los dos tipos citados en primer lugar. Los *glosarios* que consisten en una reorganización del vocabulario del texto, en orden lexicográfico, es decir alfabético, o en orden lexicométrico según frecuencias decrecientes, y, a igual frecuencia, recurriremos al orden alfabético. Estos listados léxicos son necesarios para poder operar sobre ellos con los análisis estadísticos pertinentes que conducen a la obtención de resultados objetivos para lograr unas conclusiones significativas desde el punto de vista interpretativo.

Las palabras sobre las que se trabaja habitualmente en estas investigaciones son las consideradas *llenas*, o sea, las que llevan la mayor carga semántica de la frase como son los nombres, los verbos y los adjetivos, terminología que se emplea como oposición a palabras *gramaticales* o funcionales. Las frecuencias más altas suelen corresponder a vocablos de menor carga semántica que no siempre son ignorados, puesto que son útiles cuando se busca provocar una reacción específica en el lector como es el caso de los discursos políticos (Bécue Bertaut, 1999:86).

Posteriormente, se elaboran las *concordancias* que presentan los contextos discursivos en los que aparece una forma, denominada *forma-polo*.<sup>2</sup> Este contexto es facilitado por el soporte informático utilizado, proporcionándonos las líneas del texto anteriores y posteriores a la línea en que se encuentra dicha forma en la extensión deseada por nosotros. La elaboración de concordancias es imprescindible para el estudio de los centros temáticos, puesto que es la manera de determinar el sentido de las formas-polo seleccionadas de los glosarios siguiendo un criterio de frecuencia y de distribución, o sea, si aparecen en determinadas partes de la obra y en otras no. Tenemos que precisar que esta investigación ha trabajado a partir de las llamadas *formas gráficas*<sup>3</sup> porque son más fácilmente reconocibles por el ordenador y, por tanto, más propicias para la segmentación automática. No es la única unidad posible de análisis puesto que otros investigadores prefieren el *lema*, nombre dado a una forma que recoge todos los vocablos que cuentan con una misma raíz y semejante significación sin tener en cuenta las variaciones morfológicas de las palabras.

Las *tablas léxicas* o las *tablas de segmentos repetidos* constituyen otras formas de organizar la información para el posterior análisis estadístico ya que pueden ser útiles en una comparación de textos de autores diferentes, como ejemplos tenemos la realizada para comparar el vocabulario de *La Condition Humaine* de Malraux, o la realizada con las obras publicadas en Francia entre 1930 y 1936, que figuran en la base de datos FRANTEXT (Phalèse, 1995:41-44). También es posible estudiar la evolución del vocabulario de un autor concreto, según la época en la que escribió las diferentes obras. Esa verificación la podremos realizar con una simple lectura de las tablas elaboradas, por ser una información que se ofrece de forma precisa y clara.

---

(1) Ha Programa desarrollado por Ángel Igelmo (Análisis Matemático), Gabriel M.<sup>a</sup> Jordà (Filología Francésa) y José L. Guerrero (Informática), de la Universitat de Les Illes Balears.

(2) Término empleado por Etxeberria et al. (1995:154) para hablar de la palabra que se quiere estudiar a través de distintas concordancias.

(3) Definidas por Lebart y Salem (Etxeberria, 1995:147) como una secuencia de caracteres no delimitadores (letras) comprendida entre dos caracteres delimitadores (espacios o signos de puntuación).

El presente estudio se fundamenta en la selección de los *verbos-clave* de *Citadelle* para lo que emplearemos la Estadística Textual. Los datos que obtenemos estadísticamente son útiles para la constitución de lo que P. Guiraud denomina *campos estilísticos*, que se establecen sobre los *campos semánticos* que la lingüista Gardes-Tamine (1988:102) define como: “l’association d’un champ notionnel et d’un champ lexical”. En España, la profesora Bobes Naves (1989:23) explica este mismo concepto en términos similares. Las dos investigadoras coinciden en la necesidad de establecer listas cerradas de palabras para los estudios sobre ellas, estableciendo el investigador los criterios de selección.

Para Guiraud, está claro que el sentido de una palabra en la obra depende de su relación con las demás, dicha palabra tendrá sentido no por sí misma sino en función del lugar que ocupa en la comunicación. Es un estudio sintagmático. En nuestro caso, la significación de los verbos seleccionados está estrechamente relacionada con el resto de los vocablos que la rodean en cada contexto preciso. El sentido de cada una de las palabras elegidas se irá revisitando de otros matices, irá evolucionando y transformándose con la misma escritura del texto, y constituirá una significación diferente a la que podría aportarnos un diccionario. Los temas sugeridos por esas frecuencias de uso de determinados verbos constituyen la continuación de la investigación en una fase interpretativa posterior. Este tipo de trabajos sobre frecuencias de vocabulario no ha sido muy experimentado con autores contemporáneos ni con obras completas de una extensión como la que nos ocupa.

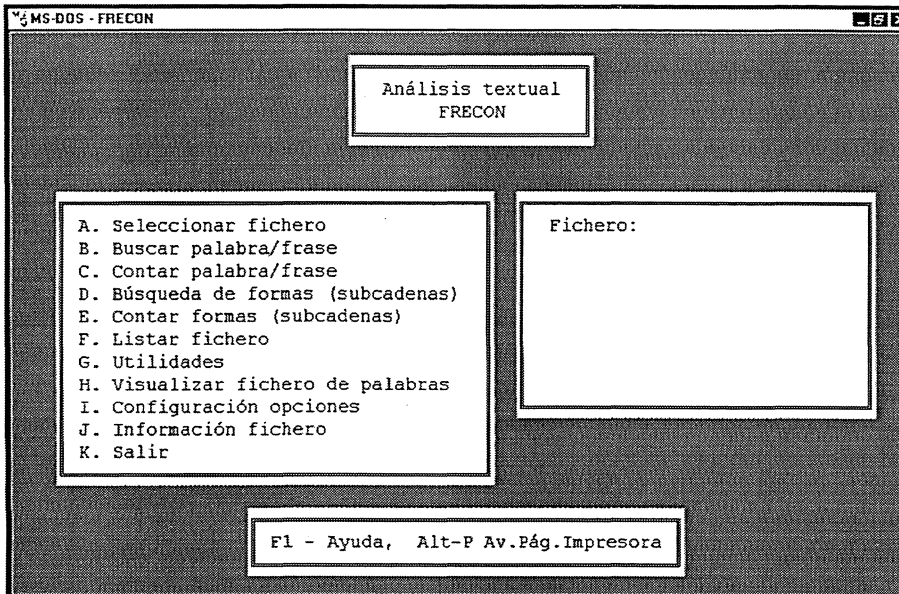
La edición de *Citadelle* empleada para poner la obra en soporte informático es la publicada por Gallimard en 1971, en su colección *Le Livre de Poche*, n.º 1.532, 1.533, 1.534. El manuscrito, que su autor no pudo corregir, llegó a la editorial lleno de tachaduras, incorrecciones y líneas en gran parte ilegibles. Esta edición presenta el texto considerado hoy como definitivo gracias a la puesta a punto de Simone Lamblin, Pierre Chevrier y Léon Wencelius.

Es necesario dividir el corpus textual en varias partes y suprimir la ambigüedad de algunos vocablos para proceder a la segmentación de la obra. La fragmentación del texto debe atender a un equilibrio en su longitud como explica la siguiente cita de Lafon (1985:59): “Les résultats des tests s’avèrent pertinents surtout si les ensembles sont de longueur analogue”. En este caso, la división se estableció en cinco partes de igual longitud con similar número de palabras. Algunas de ellas necesitaron de un marcaje especial para solucionar los problemas de homonimia que han surgido.

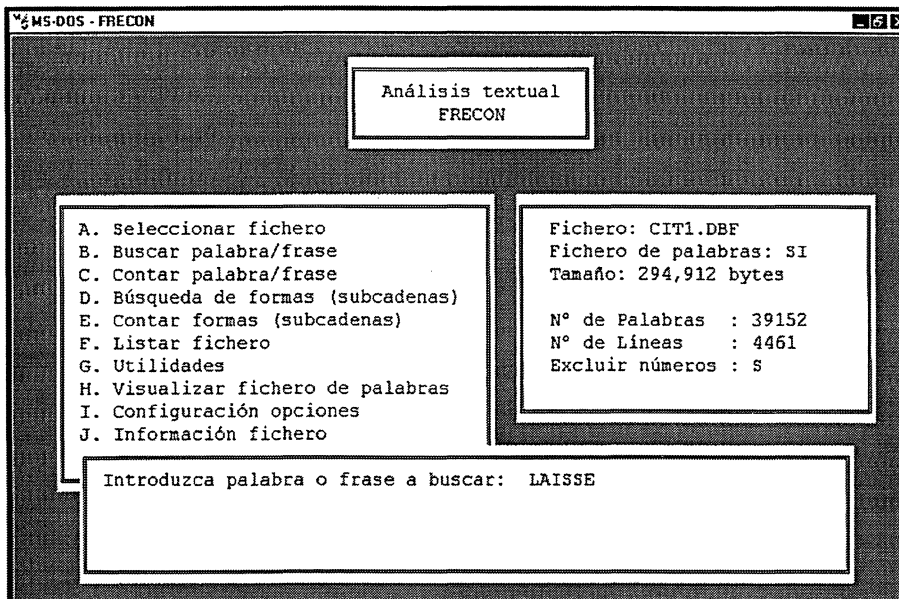
El traslado del libro a soporte informático conlleva la creación de archivos de texto en ASCII y, posteriormente, su pase a archivos en formato DBF, para así poder trabajar con el programa FRECON.UIB.

## **2. CREACIÓN DE DOCUMENTOS LÉXICOS CON EL PROGRAMA FRECON.UIB**

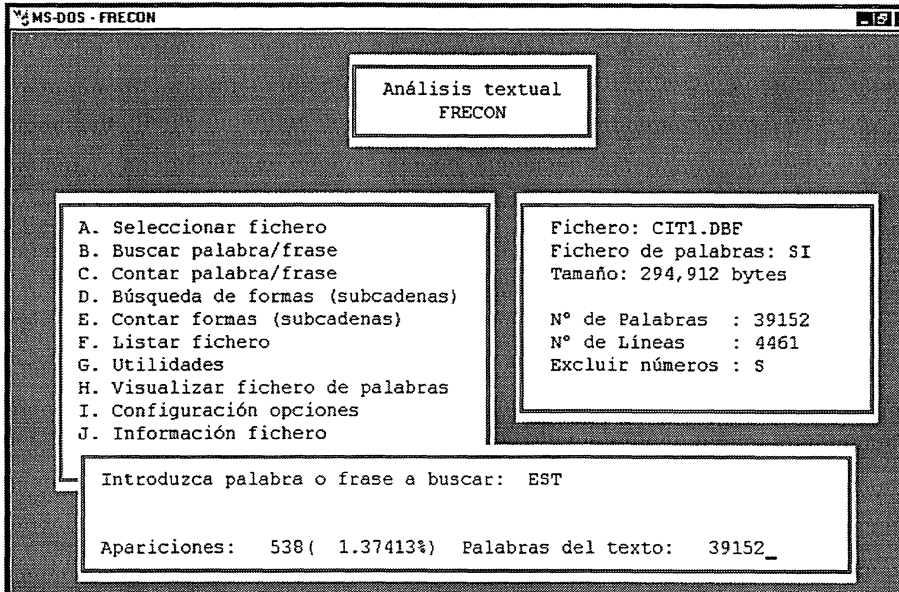
El programa se compone de un fichero de base de datos que comprende un campo por registro, llamado línea, de tipo alfanumérico, con 65 caracteres de longitud. En cada registro se introduce una frase del libro, o parte de ella, secuencialmente. Dicho programa permite la búsqueda y clasificación de palabras, frases, formas, etc. Junto a la identificación de las unidades que deseemos, va a realizar un tratamiento estadístico sencillo de las mismas, básicamente cálculo de frecuencias relativas y absolutas. Al abrir el programa, tras la pantalla de presentación, accedemos al menú principal:



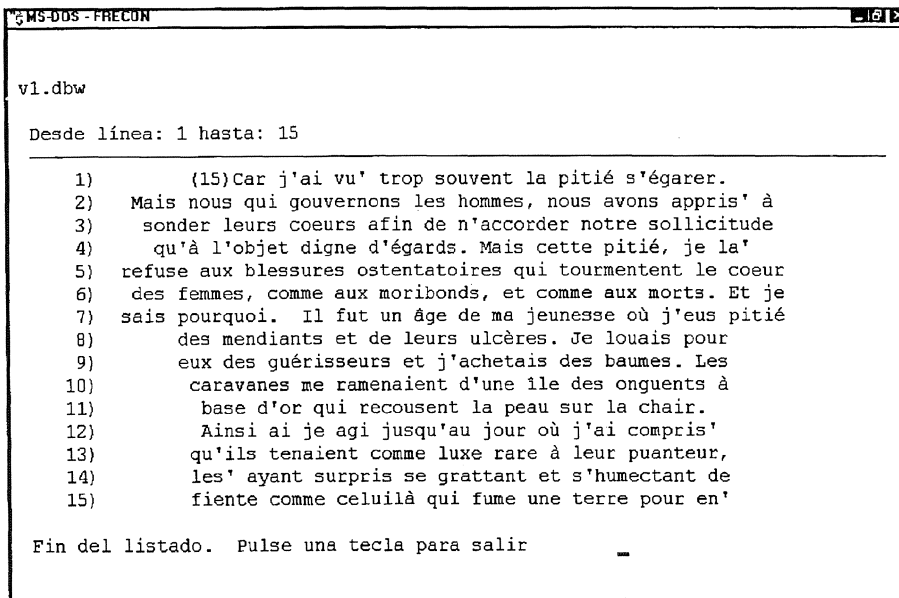
Consta de once subrutinas como hemos visto en la pantalla anterior. Antes de cualquier otra opción hay que abrir el fichero donde están los datos que vamos a estudiar. El archivo que seleccionemos será siempre uno del tipo DBF. Así, por ejemplo, el primero que utilizamos en nuestro trabajo es CIT1.DBF y, sucesivamente, hacemos lo mismo con las otras cuatro partes. Una vez elegido el archivo, el ordenador nos ofrece el número de palabras y el número de líneas que contiene. La opción siguiente nos va a servir para buscar la palabra o la frase como vemos a continuación:



La tercera opción, letra C del menú, nos cuenta las apariciones de una frase o palabra en el texto, su frecuencia absoluta y relativa, así como el número total de palabras con que cuenta el texto, tal como aparece en la figura siguiente:



Los menús D y E son análogos a B y C pero se refieren a formas gramaticales o subcadenas de texto. El menú F permite ver el contenido de todo el fichero y, por tanto, el texto íntegro, seleccionando aquella parte que nos interesa examinar en ese momento, tal como muestra esta pantalla:





```
MS-DOS - FRECON
v1.dbw
Desde línea: 1592 hasta: 1598
-----
1592)      leur ferai trancher la tête dans un grand cirque de
1593)                      clairons d'or."
1594)
1595)      "Quiconque      , disait mon père, c'est qu'il
1596)                      est bas." "Jamais un chef, disait mon
1597)                      père, ne sera jugé par ses subalternes."
1598)
Fin del listado.  Pulse una tecla para salir
```

Así se obtuvieron los inventarios lexicográficos y lexicométricos. Estas tablas nos han permitido acceder a los términos de mayor presencia en el corpus, y nos han dado una primera información sobre la importancia cuantitativa de los significados representados en ellos.

### 3. ANÁLISIS DE ESPECIFICIDADES

Los archivos resultantes de la fase anterior nos sirven de texto base sobre el que se realiza el *cálculo de las especificidades*, por medio del programa SPAD.T de CISIA (Centre International de Statistique et Informatique Appliquées). Este método estadístico es de gran utilidad por ser el que resulta más exacto en las frecuencias débiles de los grandes corpus. Consiste en determinar los elementos que se revelan más característicos en un texto cuando llevamos a cabo un estudio comparativo de varios de ellos. El carácter específico de una palabra en una de las partes se determina por comparación con el corpus formado por el total de las partes en que dividiremos la obra de *Citadelle*.

Las aproximaciones estadísticas para realizar este estudio han sido múltiples, basándose en teorías tales como Chi-cuadrado, la ley normal o ley de Poisson. Sin embargo, es *la ley hipergeométrica* desarrollada por P. Lafon la que se adapta mejor a las ocurrencias del vocabulario. La especificidad puede ser positiva, y por tanto existir una sobreutilización, o ser negativa lo que señalaría su escasez. Esa probabilidad de una palabra de aparecer en un texto en mayor o menor medida no vamos a explicarla ahora matemáticamente, pero sí diremos que su finalidad es la de tener en cuenta que las palabras no son lanzadas al azar en las obras literarias, para ello daremos una cita que nos parece aclaratoria de la función que cumple (Bernard, 1999:44):

Imaginez un casier disposé ainsi:

		A	B	

Si je disperse au-dessus une poignée de confettis, on sent bien qu'il serait étonnant qu'il en tombe la moitié dans la case B, mais que l'on pourrait trouver normal qu'il en soit ainsi pour la case A. Quelle est la probabilité de ces deux événements? C'est justement ce qu'expriment leurs spécificités, sous la forme d'une valeur comprise entre 0 (l'impossible) et 1 (l'inévitable). Si la probabilité est très faible (on n'utilise généralement que les probabilités inférieures à 0,01, soit une chance sur 100), c'est que l'accumulation ou la rareté de la forme mérite l'attention, et relève d'une explication littéraire (thématique, stylistique, historique, ...)

Todo esto para reseñar que el fenómeno de sobreuso o de infrautilización de una forma obedece a la manera de pensar del autor. Este tipo de cálculo se ha usado con éxito en textos cerrados como el nuestro, ejemplo de ello son los de Lamartine, Gautier, Verlaine y Baudelaire que se incluyen en el directorio EXPOEME ofrecido por SPAD.T, a pesar de ser un método más usado para textos de cuestiones abiertas.

Para proceder a este tipo de análisis estadístico hubo que transcribir todos los verbos en documento ASCII según la frecuencia de aparición. Este texto base nos permitió obtener, por medio de la ley hipergeométrica, las palabras infrautilizadas y aquellas otras que gozaban de una sobreutilización como ya hemos expuesto. Estas palabras son las que alcanzan los índices más altos para el valor-test medio, calculado a partir del valor-test de cada una de las formas que la componen. Los resultados se ofrecen a continuación siguiendo las cinco partes en que hemos dividido la obra:

PRIMERA PARTE. Los verbos de *Citadelle* (Páginas 15-133)

Formas sobreutilizadas	Valor test	Formas infrautilizadas	Valor test
venaient	4.106	est	-10.040
vide	3.481	es	-7.247
faites	3.476	soit	-5.542
disait	3.398	as	-4.491
sommes	3.247	sont	-4.340
enterre	3.195	faire	-4.268
chantait	3.195	dire	-3.768
enseignerez	3.195	sois	-3.683
édifie	3.195	fait	-3.659
forme	3.181	juger	-3.544



SEGUNDA PARTE. Los verbos de *Citadelle* (Páginas 133-253 )

Formas sobreutilizadas	Valor test	Formas infrautilizadas	Valor test
renie	3.737	ont	-5.253
laisse	3.179	parle	-3.795
réponds	2.894	prétends	-3.066
abordent	2.850	irai	-2.917
collabore	2.850	es	-2.478
fane	2.850	sert	-2.433
connaissais	2.830	voir	-2.412
donné	2.740	vide	-2.337
donnes	2.738	importe	-2.292
pouvoir	2.687	changer	-2.277

TERCERA PARTE . Los verbos de *Citadelle* (Páginas 253-373)

Formas sobreutilizadas	Valor test	Formas infrautilizadas	Valor test
est	7.937	suis	-3.409
es	6.729	change	-3.065
voir	4.882	été	-3.045
noue	4.282	saurais	-2.316
trouve	4.185	seront	-2.257
soit	4.069	porte	-2.147
communiquer	3.862	cherche	-2.040
rit	3.822	était	-1.983
pouvoir	3.716	venir	-1.767
sourire	3.437	enferme	-1.967

CUARTA PARTE. Los verbos de *Citadelle* (Páginas 374-496)

Formas sobreutilizadas	Valor test	Formas infrautilizadas	Valor test
ait	6.204	vient	-6.472
es	5.045	vois	-5.095
agir	4.855	vu	-5.095
étant	4.175	vivre	-4.164
ira	4.006	voir	-4.021
allé	3.952	vint	-3.444
dévore	3.658	pouvoir	-3.332
ressemblent	3.287	refuse	-2.820
est	3.272	vont	-2.782
dirai	3.220	croire	-2.433

QUINTA PARTE. Los verbos de *Citadelle* (Páginas 496-617)

Formas sobreutilizadas	Valor test	Formas infrautilizadas	Valor test
été	5.477	es	-5.939
plaît	4.971	a	-5.735
marche	4.708	trouve	-4.281
suis	4.676	étant	-3.350
viens	4.422	étaient	-3.347
vois	4.053	saisir	-2.886
prétends	3.851	créer	-2.375
sauverai	3.813	sourire	-2.343
irai	3.699	laisse	-2.127
chantes	3.630	sachant	-1.995

Lo interesante de estos resultados es la comparación de las subpartes de un texto como *Citadelle*, escrito a lo largo de varios años, para constatar las fluctuaciones o la persistencia de algunas ideas que sus palabras expresan.

#### 4. BÚSQUEDA DE LA HOMOGENEIDAD

En la siguiente fase estadística, hemos obtenido las palabras más homogéneas aplicando el método Chi-cuadrado. El análisis matemático se ha llevado a cabo por medio del programa ESTADI.<sup>4</sup> Diremos que han resultado homogéneas aquellas palabras cuya frecuencia de aparición ha sido similar en las cinco partes en que hemos dividido la obra. Por tanto, no presentaban diferencias estadísticamente significativas.

La metodología *Chi-cuadrado* es la técnica estadística que se aplica al análisis de frecuencias absolutas y relativas a fin de calcular la homogeneidad de una unidad o conjunto de unidades en uno o varios textos. Se habla de hipótesis nula ( $H_0: P_1 = P_2$ ), cuando esta utilización responde a un cálculo aleatorio. Se establece matemáticamente la región crítica (RC), es decir, la zona en la que los cálculos, al ser la estadística la ciencia de lo probable, pueden ser considerados como hipótesis nulas de acuerdo con un índice de significancia, y se rechaza la hipótesis nula en caso contrario, o sea, cuando el empleo de la unidad o del conjunto de unidades no responde a una utilización aleatoria, que, trasladado al caso del estilo de un autor, nos lleva a una intención consciente o inconsciente de preferencia de un término sobre otro. La filología debe analizar, interpretar y, en su caso, obtener conclusiones de este dato. El *índice de significancia* nos muestra la posibilidad de que se rechace la hipótesis nula aun siendo ésta cierta (según el estadístico, según la hipótesis desarrollada). En otras palabras, nos muestra el índice de error: una significancia de 0'05 quiere decir que tenemos una probabilidad de acierto de un 95%.

(4) Programa del Dr. Igelmo Ganzo de la UIB.

Todas las palabras que se seleccionan se introducen una a una en el ordenador. El programa pide el número del documento, que corresponde a cada una de las partes en las que se divide *Citadelle*. Dado el número, la pantalla siguiente requiere la introducción, en cada una de las partes, del tamaño del documento, es decir, el número total de formas gráficas incluidas en cada parte, para pasar luego a anotar la frecuencia absoluta de las formas verbales en cada una de ellas. Y, con estos datos, obtenemos automáticamente en pantalla el resultado (sí/no) de su homogeneidad y su estadístico numérico. Previamente, se había elaborado una lista con los verbos cuyo umbral de frecuencia era superior a 99, es decir las formas que en todo el texto de *Citadelle* tenían una ocurrencia de 100 veces o de un número superior. El programa ha aplicado el estadístico Chi-cuadrado sobre estas palabras con un nivel de significancia  $\alpha=0.05$  y 4 grados de libertad (9.488), y nos ha ido dando el estadístico numérico. Vemos los resultados en el listado que ofrecemos a continuación, ordenado de mayor a menor según grado de homogeneidad:

sais	2.93	sí
sera	3.51	sí
dit	4.39	sí
peut	5.74	sí
faut	5.97	sí
dis	6.15	sí
ai	7.08	sí
dire	8.39	sí
veux	9.48	sí
sont	9.71	no
agit	11.43	no
être	13.47	no
fait	13.97	no
faire	16.54	no
était	17.68	no
suis	25.15	no
soit	28.01	no
a	29.83	no
vient	32.82	no
es	73.72	no
est	79.32	no

Comprobamos que se acepta la hipótesis nula en nueve de estas palabras.

## 5. SELECCIÓN DE VERBOS PARA EL ESTUDIO DE LA ESTRUCTURA LÉXICA

Obtenidos todos los datos, tenemos que proceder a seleccionar un número determinado y restringido que nos permita profundizar en ellos. Está claro que, en los listados precedentes, debemos optar por un número restringido de verbos para establecer una lista cerrada que facilite el acceso a una investigación más exhaustiva de cada uno de ellos, puesto que la totalidad haría el trabajo inabarcable. Hemos considerado conveniente elegir los tres verbos más sobreutilizados de cada una de las cinco partes, siempre que su valor test sea superior a 3.000, y los nueve que presentan homogeneidad, quedándonos pues con un total de veintitrés para adentrarnos en el estudio de los campos estilísticos y en la estructura léxica de *Citadelle*. Pensamos que son un número suficiente para establecer conclusiones puesto que pertenecen a campos semánticos variados, dándose además un equilibrio en el número de palabras homogéneas y específicas. Los verbos-clave seleccionados son los que siguen:

CARACTERÍSTICAS	1° parte	2° parte	3° parte	4° parte	5° parte
	venaient	renie	est	ait	été
	vide	laisse	es	es	plaît
	faites		voir	agir	marche
HOMOGÉNEAS:	sais, sera, dit, peut, faut, dis, ai, dire, veux.				

Como se puede apreciar en la tabla anterior, hay un verbo que se repite en la 3ª y en la 4ª parte por lo que aparecerán en los análisis posteriores 22 formas y no 23. Según el cuadro anterior, algunos verbos, como por ejemplo *renie* y *vide* dan la sensación de pasar página sobre elementos conocidos por el narrador. También se observa una constatación de gentes que se acercan hacia algo o hacia alguien con el verbo *venaient*. Las diferentes formas del verbo *être* deben analizarse en sus contextos de la misma forma que las dos del verbo *avoir* y, viendo su entorno, poder clarificar su sentido. Y otro tanto cabe decir de ese más explícito *faites* que se abre a un público plural, cara a cara, pero del que ignoramos las consignas dadas tras el verbo.

## 6. CAMPOS ESTILÍSTICOS

Con estos datos cuantitativos de objetividad difícilmente cuestionable podemos realizar el estudio cualitativo del texto para detenernos en los componentes literarios y estéticos partiendo de la base de que toda creación literaria es un universo autónomo y de que el significado de las palabras depende de la relación de éstas con las demás en el seno de la obra. El rey beréber que representa el eje central de *Citadelle* lo explica así: "J'ai rempli le mot d'un sens plus lourd". La fase que prosigue el presente estudio cuantitativo es la encargada de esclarecer cuál es ese significado.

Saint-Exupéry se apropia de cada término empleado para hacerlo suyo, con sentido personal y único, como ocurre con frecuencia con otros grandes autores. Por este motivo, la estilística nos resulta de gran provecho ya que se ocupa de la desviación que supone la obra literaria con respecto a la norma general, la lengua, por ser la manifestación de la individualidad a través de una forma precisa.

La extensión de un artículo no permite presentar exhaustivamente el contenido revelado por cada campo estilístico, pero repasaremos algunos verbos obtenidos en la estadística para ofrecer una visión global de los temas que se apuntan en la estructura léxica final, puesta de relieve por ese entramado verbal que los cálculos matemáticos han señalado como relevantes.

A través del estudio de estos verbos en sus concordancias discursivas la temática que se manifiesta como central muestra claramente que este libro es la síntesis del pensamiento de Saint-Exupéry. Si el autor no hubiera muerto antes de finalizar la obra está claro que hubiera suprimido numerosos párrafos y hubiera sintetizado sus ideas de forma más precisa, lo que facilitaría la tarea de comprensión del lector. Pero, a pesar de todas las dificultades que entraña su lectura, es necesario el estudio de esta obra para cualquiera que quiera profundizar en el conocimiento de este autor.

Los temas que han surgido en el análisis pueden dividirse en tres apartados. Las grandes preocupaciones que manifiesta *Citadelle* se centran en la construcción del individuo para transformarse en *hombre*. El camino que recorre cada persona en ese devenir personal es de vital importancia. Paralelamente y con igual intensidad, nos presenta a ese individuo como ser social; la sociedad constituye un macrocosmo con un campo de influencia amplio sobre los microcosmos particular dejando una huella necesaria para el progreso conveniente de cada individualidad. Y, por último, un eje intermedio que se ocupa de la comunicación, particularmente, del lenguaje. El fundamento de todo esto reside en el hombre como ya había puesto de relieve en toda su producción literaria anterior.

Cada uno de los tres apartados anteriores presenta múltiples matices. El primero, el mundo individual, se relaciona con su preocupación acerca de la significación de la verdad: “Et je **dis** vérité cela seul qui te exalte” (III, 3281-82),<sup>2</sup> con la muerte: “Le jardin s’offre à moi. Il **est**, derrière le petit mur, une patrie de mandariniers et de citronniers où **sera** reçue ta promenade” (IV, 4352-54), con la búsqueda del conocimiento: “Trouver c’**est** **voir**” (III, 3856), con los valores éticos que sostienen a cada hombre: “Je **veux** **voir** ton Dieu pour croire en toi” (V, 1105-06), y con la amistad: “Je t’accepte tel que tu **es**” (V, 3433), o con la felicidad: “Car l’homme, je te le **dis**, cherche sa propre densité et non pas son bonheur” (II, 3181-83).

En el mundo de la colectividad, se presenta a ese hombre como ser necesitado de una relación social. Leemos valoraciones rotundas sobre la jerarquización social y los principios cívicos que deben imperar en esa sociedad y, por ejemplo, nos describe las cualidades que debe poseer el jefe, la persona que debe ser guía para su pueblo, y ofrece un modelo patriarcal que evoca ideales bíblicos pero que no olvida preceptos fundamentales como la igualdad, entendida de forma muy personal: “La seule égalité qui **ait** une signification” (II, 2604-05), ni su oposición a todo totalitarismo: “Et de voir que les empires forts tranches les têtes tu **veux** créer ta force en les tranchant et tu n’**es** qu’un bouffon sanguinaire” (III, 1962-65). El ejemplo de jefe que ofrece persigue caminos diversos en los que cada ser pueda alcanzar su libertad particular: “Plus je trace des routes plus tu **es** libre” (III, 111-12). Además, presta atención a la trascendencia de la educación no sólo como tarea profesional de unos pocos,

---

(5) Esta numeración corresponde a la división llevada a cabo con *Citadelle*, los números romanos se refieren a una de las cinco partes que componen la obra en soporte informático y los números arábigos se refieren a las líneas correspondientes.

sino también como labor obligada de la colectividad para transmitir el legado cultural recibido de los antepasados a los herederos de esa civilización: “Le véritable enseignement n’est point de te parler mais de te conduire” (IV, 2178-79). Y no olvida la paz como elemento esencial de convivencia y como fruto de una preparación laboriosa: “La paix est un arbre lent à grandir” (I, 2840-41). La civilización que él ensalza es exigente con cada uno de los que la componen: “Une civilisation repose sur ce qui est exigé des hommes” (I, 1618-19).

Y, por último, el lenguaje como vía de comunicación entre el hombre y la sociedad, y como fuente de comprensión y de expresión de ese hombre preocupado por el mundo trascendente que ocupa gran parte de su vida anímica. Verifica las carencias del lenguaje verbal: “Ce qui est en moi il n’y a pas de mots pour le dire” (II, 482-83), y ofrece ejemplos de otros tipos de comunicación más completos para exteriorizar las preocupaciones humanas más íntimas: “C’est pourquoi je dis prière le labour, et labour, la méditation” (III, 2-3).

## 7. CONCLUSIONES

El tratamiento estadístico y el tratamiento informático dado se ha revelado como una herramienta indispensable para un texto con la complejidad de una obra como *Citadelle*. No sólo en la parte más técnica de extracción de datos, sino también en la segunda fase en la que la consulta de los contextos era necesaria, y a los que había que volver, una y otra vez, para contrastar las ideas que se manifestaban como primordiales. La agilidad y la precisión que se logran con el ordenador son muy superiores al antiguo sistema de fichas.

Al observar los campos estilísticos, la obra reúne en sí toda la compleja temática de Saint-Exupéry y añade una preocupación filosófica más difícil de ver en otras obras porque queda empañada por las hazañas relatadas. La dimensión de héroe que se ha querido dar a su figura basada en parte en sus escritos y en su biografía, encuentra explicación profunda en esta obra póstuma en la que podemos comprender que la razón que le movía a la acción no era la persecución de la gloria sino el deseo de apaciguar su espíritu y su necesidad de comprender y encontrar una razón a la existencia humana.

El verbo es la palabra que expresa la acción y que funciona como núcleo de la oración, sobre él se articulan el sujeto y el predicado. El estudio de los verbos, obtenidos de la manera más objetiva posible, nos ha llevado a entender en detalle las acciones de las que ese verbo es el eje fundamental. Como señala M. Grevisse, el verbo marca un proceso fundamentalmente, con él se designa y sintetiza las nociones de acción, existencia, estado y devenir en relación con el sujeto. Las formas verbales, analizadas en personas y tiempos distintos, desmenuzan las preocupaciones temáticas de su autor.

En la actualidad, los programas informáticos presentado han sido integrados en otro nuevo más completo y actualizado para un manejo más sencillo en el programa FRECON-WIN que se presenta en otro de los artículos de esta revista.

## BIBLIOGRAFÍA

- BÉCUE BERTAUT, M. (1999): “Análisis estadístico de textos” en BLECUA, J.M.; CLAVERÍA, G.; SÁNCHEZ, C. y J. TORRUELLA (eds.), *Filología e informática. Nuevas Tecnologías en los estudios filológicos*. Barcelona: Editorial Milenio i Universitat Autònoma de Barcelona, pp.81-109.
- BERNARD, M. (1999): *Introduction aux études littéraires assistées par ordinateur*. Paris: PUF.
- BOBES NAVES, M. C. (1989): *La semiología*. Madrid: Síntesis.

- ETXEBERRÍA, J., GARCÍA, E., GIL, J., RODRÍGUEZ, G. (1995): *Análisis de datos y textos*. Madrid: RA-MA.
- GARDES-TAMINE, J. (1988): *La Grammaire*. París: Armand Colin.
- LAFON, P. (1985): "Lexicométrie et associations syntagmatiques" en CHARPENTIER, L. et al. *La recherche française par ordinateur en langue et littérature*. París: Slatkine-Champion.
- MEYER, T. (1991): "Usage et représentation du traitement de texte chez les enseignants chercheurs et chercheurs", en ANIS, J. y J.-L. LEBRAVE (eds.), *Texte et ordinateur. Les mutations du Lire-Écrire*. Éditions de l'Espace européen.
- PHALÈSE, H. (1995): *Les voix de la 'Condition humaine'. La 'Condition Humaine' d'André Malraux à travers les nouvelles technologies*. París: Nizet.