

El análisis estadístico para el estudio de los campos estilísticos en una obra literaria

Angel Igelmo, Gabriel M. Jordà, Carlota Vicens

Universitat de les Illes Balears
Cra. de Valldemossa, Km 7,5 – 07071 – Palma de Mallorca – España
angel.igelmo@uib.es, gmjorda@uib.es, cvicens@uib.es

Abstract

The purpose of this page is to outline some of the possibilities offered by using the FRECONWIN tool in the textual analysis. Vocabulary can be seized as a dynamic whole ; we consider textuality as a network of thin mutual relations between recurrent units, a vocable is rather a field, a discontinuous cotext, wich rules its own endogenous norm for the distribution of its cocurrents. The object is a statistical and stylistic investigation of 'Citadelle' by Antoine de Saint-Exupéry.

Resumen

El objetivo de esta comunicación es el de mostrar la fecundidad de los análisis estadísticos efectuados con la aplicación FRECONWIN en el estudio de los campos estilísticos de una obra literaria. Siguiendo a P. Guiraud, consideramos "campos estilísticos" los nudos de interrelaciones que se establecen entre los diferentes elementos del léxico de una obra. Hemos elegido la que Saint-Exupéry consideraba como la más importante de su vida, *Citadelle*, y con la finalidad de descubrir su estructura léxica y poder conocer mejor el pensamiento y el universo simbólico del autor hemos estudiado los campos estilísticos de los nombres, de los más específicos en alguna de las etapas de la vida del autor y también de los más homogéneos, de los que sistemáticamente aparecen a lo largo de la obra.

Keywords : statistique textuelle, homogénéité et spécificité, champs stylistiques, Saint-Exupéry.

1. Valores estilísticos del vocabulario

A menudo se ha utilizado la estadística léxica como rama auxiliar de una estilística de normativa exógena ; es decir, se tenía como representativo del estilo de un autor el conjunto de desviaciones en la frecuencia del empleo de determinados vocablos respecto a un corpus externo considerado como característico del uso de la lengua en un tiempo-espacio determinados (Guiraud, 1954). De este modo se esperaba de los métodos estadísticos, no sólo la identificación y comparación de las desviaciones, sino también su interpretación estilística, considerando como elemento esencial, como *mots-clés*, la lista de palabras "sobrerepresentadas". Sin embargo el mismo Guiraud insistió, unos años más tarde, en "la extrema complejidad del problema", constatando con sinceridad que "la mayor parte de los estudios ... de los *mots-clés* y de los *écarts* en el uso de las formas y construcciones son, por regla general, simples inventarios pasivos de los que sólo pueden extraerse conclusiones vanas o tautológicas" (Guiraud, 1969 : 16). Frente a este problema se produjo un cambio de perspectiva al considerar que los vocablos se caracterizan por su colocación a lo largo del texto (Peytard, 1970), por su "cotexto discontinuo" (Massonnie, 1986 ; Muller, 1973 y 1977 ; Brunet, 1985) ya trabajaron desde esta perspectiva comparando cada obra de Corneille o cada novela de Zola con el conjunto del teatro dramático o

con el de los Rougon-Macquard ; por su parte Viprey estudia la dinámica del vocabulario en las relaciones de afinidad que se observan a través de su distribución a lo largo del texto, la “surreprésentation d’un vocable dans le cotexte d’un autre vocable du même ensemble” (Viprey, 1977 y 1998).

Ya dentro de esta perspectiva, nosotros hemos seguido el punto de vista de Guiraud cuando considera la lengua de una obra literaria como un sistema de valores en el cual los signos funcionan oponiéndose unos a otros, recibiendo un sentido a medida que se van forjando sus relaciones recíprocas dentro del conjunto de la obra. Surge así la noción de “campo estilístico” para definir los nudos de interrelaciones que se establecen entre los diferentes elementos del léxico de una obra. Nuestro trabajo se centra por consiguiente en el estudio de cada vocablo dentro del conjunto de la obra y, para ello, el análisis estadístico se inicia no sólo con la consideración de los elementos “sobrerepresentados”, “específicos”, sino también de los más “homogéneos”, los que aparecen de forma sistemática a lo largo de toda la obra que, desde este momento, consideramos como un mensaje del que pretendemos reconstruir el código (Guiraud, 1969).

2. El corpus analizado

El objeto del análisis ha sido *Citadelle*, obra póstuma de Antoine de Saint-Exupéry. Apreciado en todo el mundo por su producción literaria y mal interpretado a veces por sus compatriotas, Saint-Exupéry es una figura literaria clave en el período de entreguerras. La aplicación utilizada, FRECONWIN, ha sido desarrollada en la Universidad de les Illes Balears con el objetivo fundamental de ofrecer los cálculos de homogeneidad y especificidad de formas y segmentos léxicos, tanto en un corpus general como en una clase de vocablos de dicho corpus. La personalidad del escritor y la del ser humano están íntimamente ligadas ; coherente con su vida y con su obra, en una carta dirigida a André Breton que nunca llegó a enviar, afirma : “Il n’est pas une ligne écrite au cours de toute mon existence qu’il me soit nécessaire de justifier, de taire ou démentir” (1982 : 136). El manuscrito de *Citadelle* se encontró en el interior de un maletín que él confió a su compañero Gavaille la víspera de su muerte. *Citadelle* ha sido la obra más controvertida de Saint-Exupéry y la que ha provocado las valoraciones más polémicas, pero era la obra en la el autor que tenía depositadas todas sus esperanzas, “c’est l’œuvre de toute ma vie”, afirmó al confiar el manuscrito a su amigo (Richelmy, 2000 : 129) ; fue comenzada hacia 1936, aunque algunos estudiosos sitúan el inicio de su redacción a 1933 y otros, asociándola a su experiencia del desierto en Cap Juby, a 1927. De lo que no cabe duda es de que fue escrita simultáneamente a *Terre des hommes*, *Pilote de guerre*, *Lettre à un otage* y *Le petit prince*, y que la parte substancial de ella se escribió durante su exilio en New York, es decir, cuando Europa se enfrentaba al totalitarismo, a la guerra y a la destrucción, no sólo física, sino también de los valores de la sociedad humana, por los que él siempre luchó. Según propia afirmación pensaba consagrar a la redacción del libro todavía diez años más para pulirlo y revisarlo. Estas características convierten el manuscrito de *Citadelle* en un corpus ideal para el estudio de su estructura léxica mediante el análisis estadístico y de los campos estilísticos ; Saint-Exupéry, contra lo que algunos piensan, era extremadamente minucioso en la corrección de sus obras, manuscritos de 400 páginas quedaban a menudo reducidas a unas 150 ; éste no es el caso de *Citadelle* ; esta obra póstuma es un compendio de notas manuscritas del autor o grabadas durante la noche para ser más tarde escritas, por lo que no puede en absoluto ser considerada desde el punto de vista formal como una obra acabada y, en consecuencia, sometida a comparación estilístico-formal con las otras obras del autor ; por el contrario, al ser un compendio de relatos y reflexiones surgidos en un primer momento de elaboración, efectuados a lo largo de casi toda su vida literaria y con los que Saint-Exupéry pensaba crear su “Biblia”, según propias palabras, tienen para

nosotros el valor de la inmediatez y la espontaneidad, constituyen un testimonio inapreciable para, desde el análisis estadístico, poder descubrir la estructura léxica y penetrar en el pensamiento y universo simbólicos del autor ; el estudio de los campos estilísticos de los nombres más específicos y homogéneos nos permite conocer el proceso en que se van creando los dominios y subdominios conceptuales, un proceso del que Saint-Exupéry era muy consciente : “Et quand j’ai dit rempart il faut aussi remplir le mot. Et les géomètres y ajoutent quelque chose, et les poètes, et les conquérants, et l’enfant pâle et la mère [...]” (1971 : 107). La edición de *Citadelle* empleada es la publicada por Gallimard (1971), en su colección “Le Livre de Poche”, nº 1.532, 1.533 y 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.

3. Análisis estadístico, campos estilísticos y estructura léxica

En la preparación del texto hemos escogido la opción de los “no lematizadores”, “formalistas” (Lafon, 1984) en vez de la de los “lematizadores”, “lemmatiseurs” (Muller, 1984) y, una vez realizada la desambiguación, dadas las características de las seiscientas páginas de *Citadelle*, un conjunto de pensamientos, reflexiones y notas que no fueron revisadas ni estructuradas por el autor, hemos escogido la “División automática” del FRECONWIN, optando por dividir el documento en cinco partes de tamaño idéntico. La elección de cinco partes es arbitraria en el sentido de que, dadas las características de *Citadelle*, no existe ninguna razón estilística, temática o biográfica para justificar una división determinada, la única razón es la estadística por lo que hemos optado por un número que sin ser demasiado bajo para un análisis estadístico aceptable, dividiera la obra en partes de una extensión significativa (39.621, 39.475, 39.476, 39.475 y 39.467 formas).

Una vez dividido el documento hemos iniciado nuestro análisis lexicométrico ; éste se ha centrado en el estudio de los “mots pleins”, es decir, el nombre, el verbo, el adjetivo y el adverbio. Nuestro objetivo ha sido obtener las formas y segmentos más específicos y homogéneos para, a continuación y utilizando los medios que nos proporciona la aplicación FRECONWIN, encontrar sus campos estilísticos y elaborar la estructura léxica de la obra. Como se ha señalado, esta comunicación expone el estudio de los nombres.

3.1. Análisis estadístico

Para el estudio de la homogeneidad la aplicación sigue el método del Chi-2. Este contraste o comparación lo equiparamos a un contraste de homogeneidad de muestras. Se trata de contrastar la hipótesis nula :

$$H_0 : p_1 = p_2 = \dots = p_k$$

con un nivel de significación α , frente a la hipótesis alternativa, H_1 , de que alguna igualdad no se cumple, con nivel de significación α . El nivel de significación es la probabilidad de rechazar la hipótesis nula, H_0 , siendo cierta. En este trabajo emplearemos $\alpha = 0'05$.

En este caso se trata de verificar si la clase gramatical analizada es homogénea en los k textos, lo cual significa, bajo la hipótesis de que H_0 sea cierta, que las diferencias entre las proporciones de los textos son debidas al azar y no representan un cambio de estilo.

Si n_1, n_2, \dots, n_k son las frecuencias absolutas de la clase gramatical analizada en cada texto y N_1, N_2, \dots, N_k son los tamaños de dichos textos, entonces el estadístico de contraste es :

$$W = \frac{1}{p(1-p)} \sum_{i=1}^k \frac{(n_i - pN_i)^2}{N_i}$$

siendo :

$$p = \frac{n_1 + n_2 + \dots + n_k}{N_1 + N_2 + \dots + N_k}$$

El estadístico de contraste tiene una distribución Chi-cuadrado con $k-1$ grados de libertad :

$$W \sim \chi_{k-1}^2$$

lo cual permite determinar la región crítica, RC, o región de rechazo de tamaño α .

El criterio decisorio será por tanto : rechazar H_0 si $W \in RC$. El *nivel de significación* nos muestra la posibilidad de que se rechace la hipótesis nula aún siendo ésta cierta (según el estadístico, según la hipótesis desarrollada) ; dicho de otra forma, nos muestra el índice de error : una significancia de 0'05 quiere decir que tenemos una probabilidad de acierto de un 95%.

Las aproximaciones estadísticas para realizar el estudio de la especificidad han sido múltiples, basándose en teorías tales como *Chi-cuadrado*, *la ley normal* o *ley de Poisson*. Sin embargo, es la *ley hipergeométrica* aplicada al análisis de textos por Lafon (1980) la que se adapta mejor a las ocurrencias del vocabulario. La especificidad puede ser positiva, y por tanto haber una sobreutilización, o ser negativa lo que señalaría su escasez. Lafon desarrolló el cálculo de las especificidades siguiendo la ley de la distribución hipergeométrica y demostrando que ésta se adapta a la perfección al campo actual de la lexicometría. Éste es pues el método utilizado en el *FRECONWIN*.

Pretender atacar el problema mediante el cálculo directo de la distribución hipergeométrica (mediante números combinatorios), no es el método más adecuado para un ordenador, ya que los valores numéricos son demasiado elevados. El método que seguimos, aplica la escala logarítmica para intentar disminuir los números procesados en el ordenador.

Traduzcamos ahora el modelo mediante fórmulas matemáticas : convenimos en llamar a T a la longitud total del *corpus* y t a la longitud de una parte. El computo de las posibles muestras de longitud t es elemental : es el número de combinaciones que pueden ser formadas eligiendo t elementos entre T . Este número se representa por :

$$\binom{T}{t} = \frac{T(T-1)\dots(T-t+1)}{t(t-1)\dots 2 \cdot 1}, \text{ o expresado de otra forma : } \frac{T!}{t!(T-t)!}$$

Sea ahora f la frecuencia total de una forma dada F . Del conjunto extraemos una muestra de longitud t . El número de ocurrencias de F en la muestra es una variable aleatoria X que puede tomar los valores $0, 1, 2, \dots, k, \dots, f$. Podemos calcular la probabilidad para que $X=k$, es decir para que F figure exactamente k veces en la muestra.

Para esto calculamos la muestra que contiene k veces F . Proviene de todas las combinaciones posibles de k elementos entre f , es decir $\binom{f}{k}$, y para cada una de ellas de todas las combinaciones de $(t-k)$ elementos entre $(T-f)$; es decir :

$$\binom{T-f}{t-k}$$

En la hipótesis de la equiprobabilidad de las muestras la probabilidad buscada es pues :

$$\text{Pr ob}(X = k) = \frac{\binom{f}{k} \binom{T-f}{t-k}}{\binom{T}{t}}$$

3.2. Resultados del análisis estadístico

3.2.1. Especificidad

La aplicación informática nos ha permitido la obtención de las formas específicas, sobreutilizadas e infrautilizadas, de cada una de las partes.

Ofrecemos a modo de ilustración una de las pantallas del cálculo de la especificidad con el FRECONWIN, aunque hay que indicar que éste nos permite también agrupar las formas por categorías, en el caso que nos ocupa, la de los nombres, y llevar a cabo los cálculos estadísticos dentro de cada una de ellas. En el análisis de los nombres en *Citadelle* respetamos la diferencia entre mayúsculas / minúsculas y singular / plural porque a menudo el autor utiliza estas marcas para otorgar distintos sentidos a las palabras, p. e. : *Dieu / dieu / dieux ; demeure / demeures ; Homme / homme / hommes ; Intelligence ; Esprit.*

A continuación ofrecemos los resultados de los diez nombres sobreutilizados e infrautilizados en cada una de las cinco partes en las que hemos dividido la obra de Saint-Exupéry.

The screenshot shows the 'FRECONWin - CITADELLE' window. The main title is 'ESPECIFICIDAD DE FORMAS LÉXICAS'. The interface includes input fields for 'Número de formas' (set to 10) and 'Umbral de probabilidad' (set to 0.05). Below these is a text area displaying the results of the analysis. The results are organized by part ('PARTE 1: Parte 1 (UNO)') and list various words with their corresponding specificity values in scientific notation.

Forma	Valor
ils	5,2103E-028
père	1,8009E-011
on	5,8303E-011
généraux	2,3570E-010
leurs	7,8265E-010
vous	1,8076E-007
sa	2,2986E-007
nous	2,4405E-007
moutons	1,2697E-006

At the bottom of the window, there are three buttons: 'Calcula especificidad', 'Ver Especificidad', and 'Salir'. The taskbar at the bottom shows the Start button, 'IGELMO - Microsoft Word', and 'FRECONWin 1.0'.

3.2.1.1. *Los nombres de 'Citadelle' : especificidad positiva y negativa*

Parte 1 (páginas 15-133)

Positiva :

père ; généraux ; demeure ; moutons ; demeures ; temple ; or ; sédentaires ; collaboration ; sel.

Negativa :

Cérémonial ; clous ; poème ; ami ; lignes ; planches ; sentinelle ; roi ; heure ; étage.

Parte 2 (páginas 134-253) :

Positiva :

vie ; paysage ; crête ; loisir ; prière ; plan ; création ; ordre ; Dieu ; montagne.

Negativa :

choses ; cérémonial ; condition ; sentinelle ; remparts ; blé ; graine ; part ; besoin ; matériaux

Parte 3 (páginas 254-273) :

Positiva :

sens ; liberté ; sentinelle ; choses ; temple ; empire ; âmes ; réalité ; contrainte ; objets.

Negativa :

Soir ; instant ; signe ; tour ; condition ; mot ; façon ; morts ; part ; bonheur.

Parte 4 : (páginas 374-496) :

El análisis estadístico para el estudio de los campos estilísticos en una obra literaria

Positiva :

Graine ; part ; ElKsour ; champ ; roi ; géant ; qualité ; change ; sucs ; homme.

Negativa :

Vie ; cause ; sourire ; femmes ; besoin ; demeure ; corps ; or ; frère ; pouvoir.

Parte 5 : (páginas 497-617) :

Positiva :

Perle ; condition ; frère ; fête ; cérémonial ; Seigneur ; soif ; amour ; bijou ; échecs.

Negativa :

Sens ; langage ; homme ; liberté ; hommes ; forme ; porte ; Dieu ; armée ; part.

Ya de entrada es posible hacer las siguientes observaciones :

- Las formas más utilizadas para cada una de las partes son las siguientes : *père*, *vie*, *sens*, *graine* y *perle*, frente a las menos utilizadas : *cérémonial*, *choses*, *soir*, *vie* y *sens*.
- La forma *vie* es la más utilizada en la segunda parte y la menos utilizada en la cuarta. Lo mismo ocurre con *sens*, la más utilizada de la tercera parte y la menos de la quinta.
- Las formas sobreutilizadas ofrecen en general un índice de especificidad más homogéneo.
- Ninguno de los nombres sobreutilizados se repite en ninguna de las partes. De entre los nombres infrautilizados se repiten los siguientes : *cérémonial* en la 1ª y 2ª partes, *sentinelle* en la 1ª y 2ª partes, *condition* en la 2ª y 3ª, *besoin* en la 2ª y 4ª y “part” en la 2ª, 3ª y 5ª.
- Al agrupar los nombres según campos semánticos se observa que, a grandes rasgos, estos coinciden tanto para los nombres más utilizados como para los menos utilizados. Las refe-

rencias a la religión, con palabras como *Dieu, cérémonial, temple, prière, âme* o *Seigneur* (con mayor presencia en el grupo de las formas sobreutilizadas), así como las referencias al *habiter*, habitar la casa (*demeure, père, frère, amour, loisir, ami ...*) o habitar la *citadelle* (*sentinelle, empire, roi, remparts, tour, porte ...*) son las más frecuentes, aunque los nombres empleados no siempre son los mismos. Por otra parte en el grupo de las formas sobreutilizadas observamos abundantes referencias a la naturaleza (*paysage, montagne, graine, moutons, champ ...*), mientras que en el grupo de las infrautilizadas llaman la atención aquellas formas que se refieren a la comunicación (*lignes, poèmes, mot, signes, langage, sens ...*).

3.2.2. Homogeneidad

A continuación hemos obtenido el índice de homogeneidad de los nombres con un umbral de frecuencia cien o superior, con un nivel de significancia de 0,05 y 4 grados de libertad; el resultado ha sido que hemos rechazado la hipótesis nula en 18 de ellos. Aunque como cabía esperar, ninguna de las formas homogéneas coincide con ninguna de las sobre o infrautilizadas, es interesante constatar que los campos semánticos de unos y otros siguen siendo muy próximos. Así, seis de los 18 nombres se refieren a elementos de la naturaleza y un segundo grupo a la ciudadela que hay que habitar. Por el contrario no se observan palabras directamente relacionadas con la religión y ninguna referencia al mundo de la aviación.

HOMOGENEIDAD DE FORMAS LÉXICAS

Forma léxica	Núm.apa.	Valor homog.	¿Aplicable?	¿Homogéneo?
désert	124	2,269	S	S
déserts	10	1,006	S	S
désespoir	17	3,88	S	S
désir	64	4,718	S	S
désirais	6	0,669	S	S
désire	36	4,579	S	S
désirer	6	10,679	S	S
désires	17	8,019	S	S
désirs	16	1,502	S	S
désordre	16	6,506	S	S
désormais	43	3,378	S	S
dessin	7	3,698	S	S
destin	8	5,751	S	S
destruction	16	4,636	S	S

Nivel de significación: 0.05

N° de partes: 5

Tam. Parte : 1 : 39621
 Tam. Parte : 2 : 39475
 Tam. Parte : 3 : 39476
 Tam. Parte : 4 : 39475
 Tam. Parte : 5 : 39467

Mostrar sólo formas homogéneas

Descripción de la parte	Núm.apa.
Parte 1 (UNO)	28
Parte 2 (DOS)	23
Parte 3 (TRES)	27
Parte 4 (CUATRO)	19
Parte 5 (CINCO)	27

Cálculo de homogeneidad
 Exportación de resultados
 Impresión de resultados
 Salir

Los nombres de uso más homogéneo en 'Citadelle'

Forma léxica	Núm.apa.	Valor homog.	¿Aplicable ?	¿Homogéneo?
fois	175	001.71	S	S
désert	124	001.78	S	S
vent	124	002.23	S	S
maison	164	003.39	S	S
peuple	113	004.62	S	S
marche	107	004.79	S	S
monde	122	004.88	S	S
arbre	262	005.32	S	S
visage	196	005.50	S	S
enfant	132	006.04	S	S
jour	170	006.29	S	S
vérité	134	006.42	S	S
chose	186	007.14	S	S
mots	126	007.29	S	S
travail	106	007.49	S	S
mer	171	008.01	S	S
Pierre	108	009.07	S	S
yeux	102	009.10	S	S

3.3. Estudio de los campos estilísticos en 'Citadelle'

Finalmente hemos decidido adentrarnos en el estudio de los campos estilísticos y de la estructura léxica de *Citadelle* analizando los segmentos (específicos y homogéneos) junto con los respectivos contextos de 33 nombres, los tres más sobreutilizados de cada una de las cinco partes y de los dieciocho homogéneos: *père, généraux, demeure, vie, paysage, crête, sens, liberté, sentinelle, graine, part, El Ksour, perle, condition, frère, fois, désert, vent, maison, peuple, marche, monde, arbre, visage, enfant, jour, vérité, chose, mots, travail, mer, pierre, yeux*.

La aplicación FRECONWIN nos permite agrupar los segmentos que se articulan en torno a una "forma polo", descubrir cuáles son los que aparecen de una forma más recurrente y cuáles sólo lo hacen en un contexto léxico determinado, de este modo se articula la red que constituye el campo estilístico de la "forma polo" inicial y del que va brotando el significado de la palabra dentro de la estructura de *Citadelle*.

Del estudio de los campos estilísticos se deriva la absoluta interdependencia por una parte de los nombres de la obra en su construcción del centro; por otra de los diferentes símbolos en su caminar hacia este símbolo último que gobierna el paisaje citadeliano. Hacia el centro concurren efectivamente símbolos ascensionales, catamorfos, espectaculares, diurnos o nocturnos, cíclicos, diaréticos ..., que pueden reducirse a dos grupos según su relación con el mismo: los que se alejan de él, fuerzas centrífugas o de desintegración (*frère, généraux, sentinelle, vent*), y los que conducen a él, fuerzas centrípetas o de integración (*arbre, graine, désert, mer, perle, paysage, crête, yeux, visage, pierre, maison, demeure, peuple, père*). La llegada a este centro es inevitable para quien sigue los caminos de la ascensión, el *échange* y el *habiter*.

De esta forma el estudio lexicométrico de *Citadelle* nos ha permitido descubrir una estructura léxica en la que se manifiesta un universo simbólico que guarda gran coherencia con el resto de

la obra exuperiana, que se configura como una obra unitaria y circular. Así la imagen del *Espíritu* (con mayúscula) que impera en *Citadelle* como lo único capaz de construir al Hombre, se encuentra ya intuída o bocetada en obras anteriores (*Courrier sud*, *Terre des hommes*), lo mismo que la diferencia entre *Cuerpo/Inteligencia/Espíritu* (*Pilote de guerre*) o (*Petit prince*) el viaje entendido como camino iniciático. Es cierto que sorprende en *Citadelle* la ausencia del avión y del concepto de vuelo, pero el autor no renuncia sin embargo a la mirada desde un lugar elevado, al paisaje contemplado desde la altura, finalmente organizado y comprendido.

Formas léxicas

Forma léxica	Núm. apa.
dieu	70
Dieu	257
dieux	36
difféaient	1
diffère	5
différemment	5
différence	4
différences	1
différences	4

Búsquedas de formas léxicas -

Descripción de la parte	Núm. apa.
Parte 1 (UNO)	61
Parte 2 (DOS)	84
Parte 3 (TRES)	48
Parte 4 (CUATRO)	37
Parte 5 (CINCO)	27

Referencias

- Guerrero J.L., Igelmo A. y Jordà G.M. (2003). *Freconwin*. Palma de Mallorca. Universitat de les Illes Balears.
- Guiraud P. (1954). *Caractères statistiques du vocabulaire*. P.U.F.
- Guiraud P. (1969). *Essais de Stylistique*. Klincksieck.
- Massonnie J.-P. (1986). "Q-occurrences libres". In Brunet Ét., *Méthodes quantitatives et informatiques dans L'étude des textes*. Slatkine-Champion : 611-623.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Slatkine-Champion.
- Muller Ch. ([1977] 1992a). *Principes et méthodes de statistique lexicale*. Champion.
- Muller Ch. ([1977] 1992b). *Initiation aux méthodes de la statistique linguistique*. Champion.
- Muller Ch. (1984). Prologue. In Lafon P., *Dépouillements et statistiques en lexicométrie*. Slatkine-Champion.

- Peytard J. y Genouvrier É. (1970). *Linguistique et enseignement du français*. Larousse.
- Richelmy M. (2000). *Homme synthèse du siècle*. Éditions lyonnaises d'art et d'histoire.
- de Saint-Exupéry A. (1971). *Citadelle*. Gallimard. Folio.
- de Saint-Exupéry A. (1982). *Écrits de guerre 1939-1944*. Gallimard. Folio.
- Sánchez Hernández M^a A. (2001). *El verbo en 'Citadelle' (A. de Saint-Exupéry). Análisis estadístico, campos estilísticos y estructura léxica*. Tesis de doctorado. Las Palmas de Gran Canaria : Universidad de Las Palmas de Gran Canaria.
- Vicens Pujol C. (2000). *Los nombres de 'Citadelle' de Antoine de Saint-Exupéry : análisis estadístico, campos estilísticos y estructura léxica*. Tesis de doctorado. Barcelona : Universitat de Barcelona, División de ciencias humanas y sociales, Departamento de Filología Románica.
- Viprey J.-M. (1997). *Dynamique du vocabulaire des 'Fleurs du Mal'*. Slatkine-Champion.
- Viprey J.-M. (1998). "Une norme endogène pour le calcul stylistique du vocabulaire". In *Actes des JADT 1998*.