

**KÖLNER BEITRÄGE ZUR LATEINAMERIKA-FORSCHUNG**

Herausgegeben von Christian Wentzlaff-Eggebert und Martín Traine

La voz del pueblo en el espacio cultural  
europeo: El pueblo y su identidad

editado por Christian Wentzlaff-Eggebert

Universidad de Colonia

Centro de Estudios sobre España, Portugal y América Latina

Universität zu Köln

Arbeitskreis Spanien – Portugal – Lateinamerika

La voz del pueblo en el espacio cultural europeo: El pueblo y su identidad

Contribuciones de Christian Wentzlaff-Eggebert, Alfredo Crespo Borrallo, Barbara Hagg-Huglo, Cándido Martín, Antonio Martínez González, Mario Garvín, Gloria Chicote, Antonio Frías Delgado, Antonio Tordera, David Porcel Bueno, Sofía Barrón, Javier Lluch-Prats, R. Sergio Balches Arenas, Marina Bianchi, Enrico Lodi, Gonzalo Aguila, Ewa Stala, Daniela Marcheschi y Facundo Tomás.

El presente proyecto ha sido financiado con el apoyo de la Comisión Europea. Esta publicación es responsabilidad exclusiva de los autores. La Comisión no es responsable del uso que pueda hacerse de la información aquí difundida.



Programm für  
lebenslanges  
Lernen

**DAAD**

Deutscher Akademischer Austausch Dienst  
German Academic Exchange Service

Köln / Colonia 2015

Arbeitskreis Spanien – Portugal – Lateinamerika  
Centro de Estudios sobre España, Portugal y América Latina  
Albertus-Magnus-Platz  
50923 Köln

ISSN 1438-6887

Redacción: Marja Nalesinski

# ANTONIO FRÍAS DELGADO: LAS VOCES DE LOS QUIJOTES

## **Abstract:**

In this work we present a detailed analysis of the character profile in the three Quijotes by Cervantes and Avellaneda. Several properties of the languages and sub-languages are explored.

## **Introducción**

*Las voces del "Quijote"* es el título del Estudio Preliminar que aparece en la edición del Instituto Cervantes dirigida por Francisco Rico. Su autor, Fernando Lázaro Carreter, argumenta que

La mutación fundamental que introduce el Renacimiento en la literatura de ficción consiste, esencialmente, en la independencia creciente de los personajes.<sup>1</sup>

La independencia ha de entenderse no tanto en lo que dicen (al fin y al cabo, Antígona y Creonte piensan y dicen cosas opuestas, ninguna de las cuales tiene por qué coincidir con lo que pensaba Sófocles), sino en cómo lo dicen, en sus voces, en el lenguaje. Son numerosos los estudios dedicados al análisis del lenguaje del Quijote desde un punto de vista cualitativo y/o histórico.<sup>2</sup> Cualquier lector atento, sin necesidad de ser especialista, repara en los grandes rasgos de los perfiles lingüísticos, sobre todo de Sancho y de don Quijote: el uso de refranes, la tendencia a trabucar vocablos, el uso de arcaísmos, la emulación del lenguaje de los libros de caballerías, etc. La talla gigantesca de estos personajes -don Quijote y Sancho- casi que entierra el interés por otros personajes que tienen una cierta importancia en las novelas. En este trabajo presentamos algunos resultados de análisis cuantitativos de las distintas voces que perfila Cervantes. Los análisis cuantitativos tienen más interés y son más adecuados cuando se comparan lenguajes distintos. Por ello hemos añadido a las voces de los Quijotes cervantinos las del de Avellaneda.

Además del interés propio del análisis cuantitativo de los textos mismos, nos proponemos indirectamente dos tareas adicionales. Una de ellas tiene que ver con la estilometría. Son numerosos los estudios en el campo de la

---

<sup>1</sup> Lázaro Carreter, Fernando. Las voces del "Quijote". En Cervantes Saavedra, Miguel de. Don Quijote de la Mancha. Edición dirigida por Francisco Rico. Instituto Cervantes-Crítica. Barcelona, 1998., págs. XXI-XXXVII.

<sup>2</sup> Gutiérrez Cuadrado, Juan. La lengua del "Quijote": rasgos generales. En [3], Volumen Complementario, págs. 819-856.

estilometría que defienden la existencia de un estilo, cuantificable, en el lenguaje que es propio de cada autor. Cervantes tendría su estilo diferenciado del de Avellaneda, aunque escribiesen sobre el mismo tema. Sin embargo, si un autor logra tal maestría en el lenguaje de sus personajes que puede diferenciarse, digamos, el de hombres y mujeres y los pertenecientes a una clase u otra, etc. ¿dónde queda espacio para su estilo propio que lo diferencie de otro autor? La otra tarea de investigación cuantitativa es sólo tentativa. Según algunos psicolingüistas (los trabajos de Pennebaker y colaboradores<sup>3</sup> son unas de las referencias más importantes en este campo), en el lenguaje, en las palabras escritas o dichas, vamos dejando un rastro inconsciente de nosotros mismos. Lo importante de esta tesis es (i) que el rastro es inconsciente; (ii) aquello de lo que dejamos el rastro es de aspectos como nuestro género, nuestra edad, nuestra posición social, nuestras tendencias depresivas, etc. ¿Aparecen en el lenguaje de los personajes de los Quijotes trazas de estos rasgos?

## Los textos

### ***Los textos utilizados son principalmente:***

*El ingenioso hidalgo don Quijote de la Mancha* (1605, CI en adelante)

*Segunda parte del ingenioso caballero don Quijote de la Mancha* (1615, CII en adelante)

Versión electrónica del Proyecto Gutenberg:

<http://www.gutenberg.org/ebooks/2000>

*Segundo tomo del ingenioso hidalgo don Quijote de la Mancha* (1614)

Alonso Fernández de Avellaneda (AV en adelante)

Versión electrónica de la Biblioteca Cervantes Virtual.

<http://www.cervantesvirtual.com/obras/autor/528/Fernández>  
de Avellaneda, Alonso

### ***Aspectos globales***

i.- Las tres obras son de desigual longitud. El total de palabras, considerando las formadas sólo de caracteres alfabéticos es el siguiente: AV: 136.484; CI: 180.050; CII: 190.216. El vocabulario, las palabras diferentes calculadas sobre tokens sin lematizar, es el siguiente: AV: 12.675; CI: 14.349; CII: 15.959. Puede observarse que la riqueza léxica – considerando que ésta es mayor cuanto menos veces se repite el

<sup>3</sup> Chung, Cindy and Pennebaker, James. The psychological function of function words. En K. Fiedler, editor: Social Communications. Psychology Press. New York. Pgs. 343-359. y <http://homepage.psy.utexas.edu/HomePage/Faculty/Pennebaker/Home2000/JWPhoto.htm>

vocabulario en el texto completo- es mayor en AV (10,76) que en CII (11,91) y ésta mayor que en CI (12,54).

#### *ii.- Aspectos compositivos: capítulos*

Las obras están divididas en capítulos (AV: 36, CI: 52, CII: 74), no siempre resultado del proyecto compositivo del autor sino, a veces, debido a necesidades de impresión. Aun así, los capítulos pueden darnos pistas sobre la planificación global del autor.

Consideremos el número de palabras de cada capítulo y compáremoslo con la media (el número de palabras si todos los capítulos fuesen homogéneos). CI se mantiene por debajo de la media hasta el capítulo 20 y en los 10 últimos. Los relatos incrustados alargan hasta 4 veces los capítulos iniciales más cortos.

CII ofrece una figura bien diferente; hay oscilaciones en torno a la media desde el comienzo al fin. Téngase en cuenta que en el capítulo 59 Cervantes da noticia de la aparición del Quijote de Avellaneda.

AV ofrece una imagen similar a CII, con los picos de los relatos incrustados.

Las diferencias en los proyectos compositivos se aprecian mejor observando los capítulos acumulados. La siguiente figura ofrece una comparativa del desarrollo compositivo por capítulos.

La siguiente figura muestra datos de frecuencias de los personajes principales.

Frecuencias de ‘Sancho’ y ‘Quijote’ en CI, AV, CII (en porcentajes relativos sobre el total de palabras de las obras, normalizado)

Es de notar el bajo porcentaje relativo de Sancho y de don Quijote en CI. Ello se debe a que CI es un proyecto en el que a las aventuras de ambos personajes se le añaden los relatos incrustados y otros de Arcadía pastoril en el que son meros espectadores. Publicado CI, debió ser evidente que el personaje de Sancho ofrecía muchas más posibilidades narrativas de las explotadas en CI.

### **Las voces de los *Quijotes***

¿Cuántas voces hay en los *Quijotes*? La pregunta es sencilla y parecería que hubiera de haber una respuesta simple e inmediata. No es así. El número de personajes e intervenciones depende del criterio que se adopte.

Puede variar si consideramos una misma voz o tres diferenciadas, por ejemplo, a Sansón Carrasco, el Caballero del Bosque y el Caballero de la Blanca Luna; si diferenciamos los personajes que hablan en estilo directo o indirecto en las narraciones, etc. A veces el autor no especifica, cuando se trata de un grupo, quién interviene. En CII, el juego de espejos llega a la complicación de que el narrador se presente como copista de una traducción del relato de Cid Hamete Benengeli ('en este punto dice el traductor que Benengeli dice...'). Hemos optado por diferenciar los personajes cuando actúan de narradores de historias autocontenidas; por ejemplo, Bracamonte o el ermitaño en sus intervenciones como personajes y como narradores. Nuestro método aporta estas cifras:

	CI	AV	CII
<b>Personajes</b>	73	93	136
<b>Intervenciones</b>	1744	1814	2588
<b>Media de palabr/ interv.</b>	103	75	73

El mayor número de palabras corresponde al autor/narrador, a don Quijote y a Sancho. Las intervenciones del autor, de don Quijote y Sancho, en porcentajes sobre el total de las obras, muestran diferencias de interés. Se ha diferenciado en AV y CI al narrador principal y a los narradores de los relatos incrustados. Al desaparecer estos en CII, el porcentaje del autor es mayor. Es llamativa la baja intervención de Sancho en CI. Que el personaje podía dar más juego narrativo se explota en AV y en CII. Cervantes se encuentra más cómodo, o ha encontrado un perfil del personaje que le permite más registros, con don Quijote que Avellaneda. Lo contrario ocurre con Sancho. Avellaneda abordó el proyecto desde una óptica más cercana a la picaresca o a la comedia de bobo.

Algunos personajes tienen una intervención muy reducida. Los personajes que intervienen con más del 1% del texto global son los siguientes.

Personaje	AV	CI	CII
Autor	---	1,364	---
Narrador P	27,977	30,093	33,825
Narrador RD	3,45	---	---
Narrador FA	6,926	---	---
Narrador CI	---	4,797	---
Narrador CA	---	7,321	---
Narrador CE	---	1,295	----
Narrador CP	---	1,058	---
<b>Narraciones</b>	<b>38,353</b>	<b>44,564</b>	<b>33,825</b>
Álvaro Tarfe	2,773	---	---
Bárbara	2,62	---	---
Luisa (priora)	1,773	---	---
Gregorio	1,497	---	---
Dorotea	---	4,106	---
Cardenio	---	3,92	---
Cura	---	2,914	---
Lotario	---	2,027	---
Dueña(dolorida)	---	---	1,529
Duquesa	---	---	1,512
Carrasco	---	---	1,495
Duque	---	---	1,076
Ricote	---	---	1,025
<b>Quijote</b>	<b>17,866</b>	<b>18,718</b>	<b>21,795</b>
<b>Sancho</b>	<b>18,943</b>	<b>8,067</b>	<b>17,343</b>
<b>Total</b>	<b>83,825</b>	<b>84,316</b>	<b>79,6</b>

Autor: el autor del prólogo de la obra. P: principal; RD: rico desesperado; FA: felices amantes; CI: curioso impertinente; CA: cautivo.

Muchos personajes intervienen muy poco; algunos lo hacen con una sola frase de 5 palabras. Agrupar los personajes por ‘clases’ o por estratos es tarea controvertida. A modo de ejemplo, hemos hecho la siguiente agrupación de los personajes menores de AV. Se excluyen los narradores, don Quijote y Sancho. Se indica el número de personajes, el porcentaje sobre el total de personajes y el porcentaje de sus intervenciones sobre las palabras totales de la obra. (Para obtener el porcentaje de las intervenciones sobre el total de las intervenciones de los personajes menores, ha de multiplicarse por 4,26. Así, las palabras de las intervenciones de las mujeres religiosas suponen el 2,27% del total del texto y el 9,67% del total de las intervenciones de los personajes menores.

Sexo	Grupo	Nº	% Per.	%P.
	Religiosos	3*	3,65	2,27
	Alta	3	3,65	0,64
<b>Mujeres</b>	Admon.			
	Oficios	2	2,43	0,04
	Domésticos	2	2,43	0,18
	Pueblo	4*	4,87	2,97
<b>Total</b>		<b>14</b>	<b>17,07</b>	<b>6,1</b>
	Religiosos	8	9,75	2,15
	Alta	14	17,07	7,66
<b>Hombres</b>	Admon.	10	12,19	1,32
	Oficios	9	10,97	2,18
	Domésticos	12	14,63	2,79
	Pueblo	15	18,29	0,95
<b>Total</b>		<b>68</b>	<b>82,92</b>	<b>17,05</b>

El grupo de mujeres religiosas incorpora las intervenciones de la priora D<sup>a</sup> Luisa, personaje del relato de los felices amantes. Esto aumenta su porcentaje. El grupo de mujeres de pueblo incorpora la figura de Bárbara, con gran participación en la novela.

### Algunos aspectos cuantitativos

Un texto no es sólo un conjunto de decenas o centenares de miles de palabras al azar. Sigue unas reglas y presenta regularidades. Algunas de estas pautas pueden estudiarse construyendo modelos de lenguaje, averiguando y comparando distribuciones de probabilidad subyacentes. En este apartado analizaremos algunos aspectos especialmente relevantes.

La divergencia de Kullback-Leibler (KL) es una medida no simétrica muy usada que permite establecer la distancia que hay entre dos textos. La divergencia KL del texto A al B puede ser diferente de la divergencia KL del B al A. Tiene sentido averiguar divergencias relativas: cuáles de los subconjuntos de lenguajes están más próximos de otros. La tabla siguiente muestra los textos de AV, CI, CII y en cada caso los que corresponden al autor (A), don Quijote y Sancho. Los valores máximos están en negrita y los mínimos en cursiva. Forzando un poco lo que puede inferirse de los valores de las divergencias, presentamos valores de medias y desviaciones típicas para tener una idea más global de los resultados.



	AVA	AQ	AS	CIA	CIQ	CIS	CIIA	CIIQ	CIIS
<b>AVA</b>		16.3345	16.273	<i>13.5244</i>	<b>16.3274</b>	<b>17.6905</b>	13.2314	<b>16.0558</b>	<b>16.7597</b>
<b>AVQ</b>	17.047		16.4386	<b>16.807</b>	15.8029	17.6144	<b>16.6717</b>	15.7682	16.4957
<b>AVS</b>	16696	<i>15.5142</i>		16.3532	15.6121	16.2433	16.1842	15.123	14.8746
<b>CIA</b>	<i>14.374</i>	16.7922	16.7005		15.6943	17.0157	<i>12.9889</i>	15.6017	16.2325
<b>CIQ</b>	16.99	15.7933	16.331	15.5555		16.4324	15.6565	14.4678	15.39
<b>CIS</b>	16.7896	15.58	<i>15.1103</i>	15.068	<i>14.0354</i>		15.1271	<i>13.8815</i>	<i>13.7134</i>
<b>CIIA</b>	14.7735	<b>17.0279</b>	<b>17.0358</b>	13.566	16.1889	17.5679		15.4816	16.2096
<b>CIIQ</b>	<b>17.3013</b>	16.1568	16.5115	16.0726	15.0425	16.7863	15.6719		15.2516
<b>CIIS</b>	17.1607	15.9589	15.4658	15.9541	14.8373	<i>15.696</i>	15.5383	14.072	
<b>Total</b>	131.132	129.157	129.866	122.9	123.54	135.046	121.07	120.451	124.927
<b>Media</b>	16.3915	16.1447	16.2333	15.3626	15.4426	<b>16.8808</b>	15.1337	<i>15.0564</i>	15.6158
<b>D.T.</b>	0.27194	<i>0.06569</i>	0.27135	0.20912	0.214	<b>0.41889</b>	0.14303	0.34805	0.12879

La divergencia mínima es entre el narrador de CI y el narrador de CII. En general el lenguaje de los narradores, tanto Avellaneda como Cervantes, es bastante homogéneo, presenta las mínimas divergencias: media 13.7432 y desviación típica 0.07234. Los personajes de los Quijotes presentan una media de 15.50525 y una desviación típica de 0.188916. Los personajes de Sancho presentan una media de 15.1839 y una desviación típica de 0.20906394.

Si agrupamos los tres textos de cada autor, obtenemos los resultados que refleja la siguiente tabla

	AV	CI	CII
<b>AV</b>		13.7125	13.3111
<b>CI</b>	13.7112		12.496
<b>CII</b>	14.1346	13.2921	

Puede observarse que las divergencias medias son menores cuando se consideran los textos globales que cuando estos se dividen en sublenguajes diferenciados.

## Etiquetado

Los textos han sido sometidos a un etiquetado de categorías gramaticales con las herramientas de FreeLing<sup>4</sup>. Aunque el proceso de etiquetado no es

<sup>4</sup> Lluís Padró and Evgeny Stanilovsky. FreeLing 3.0: Towards Wider Multilinguality.

totalmente preciso, suponemos que los errores se distribuyen por igual en todos los textos y pueden darnos unos valores aceptables para compararlos.

Las categorías principales que utiliza son: adjetivos (A), adverbios (R), determinantes (D), nombres (N), verbos (V), pronombres (P), conjunciones (C), interjecciones (I), preposiciones (S), signos de puntuación (F), numerales (Z), fechas y horas (W).

Los porcentajes de cada categoría en Avellaneda y Cervantes se presentan en la siguiente tabla. Para establecer referencias externas presentamos también las de TSO y FyJ.

Categoría	AV	CI y CII	TSO	FyJ
<b>N</b>	21.538	20.1967	20.7646	20.6858
<b>V</b>	18.6788	19.7549	20.0633	19.6728
<b>S</b>	14.861	<b>13.5332</b>	13.7848	<b>13.2392</b>
<b>D</b>	13.0253	<b>13.62</b>	12.9292	<b>14.4698</b>
<b>P</b>	11.5782	11.7165	11.5256	11.7985
<b>C</b>	9.3299	10.6829	9.968	8.2751
<b>R</b>	5.1621	5.5215	5.8315	6.3303
<b>A</b>	4.9769	4.3254	4.5483	4.734
<b>Z</b>	0.5474	0.4682	0.4776	0.4752
<b>I</b>	0.2792	0.1688	0.0241	0.2814

Las categorías más frecuentes, más de 5.000 ocurrencias, en los textos cervantinos son:

Categoría	Frecuencia	Lectura
<b>SPS00</b>	51302	Preposición
<b>NCMS000</b>	24038	Nombre común masculino singular
<b>NCFS000</b>	23756	Nombre común femenino singular
<b>CC</b>	23434	Conjunción coordinada
<b>CS</b>	17063	Conjunción subordinada
<b>RG</b>	14662	Adverbio general
<b>DA0MS0</b>	12133	Artículo masculino singular
<b>VMN0000</b>	11914	Verbo principal infinitivo
<b>NP00000</b>	11847	Nombre propio
<b>PR0CN00</b>	10569	Pronombre relativo
<b>DA0FS0</b>	9143	Artículo femenino singular
<b>VMIS3S0</b>	8927	Verbo principal indicativo pasado 3ª persona del singular
<b>NCMP000</b>	8177	Nombre común masculino plural
<b>NCFP000</b>	7548	Nombre común femenino plural

<b>VMIP3S0</b>	6280	Verbo principal indicativo presente 3ª persona del singular
<b>RN</b>	6269	Adverbio negativo
<b>VMP00SM</b>	5901	Verbo principal participio masculino singular

Una de las observaciones –reproches- que hace Cervantes sobre el texto de Avellaneda es que omite artículos. La tabla siguiente muestra el porcentaje de artículos y nombres de ambos autores:

	<b>AV</b>	<b>CI y CII</b>
<b>Artículos</b>	7.619686	8.097763
<b>Nombres comunes</b>	17.968812	17.071594

Empleo de las principales categorías por los narradores de Avellaneda

Empleo de las principales categorías por los personajes con más texto de Avellaneda

La siguiente figura permite ver los perfiles de los personajes principales de Avellaneda: (1) Álvaro, (2) Bárbara, (3) Gregorio, (4) Luisa, (5) don Quijote, (6) Sancho

Empleo de las principales categorías por los narradores de CI.

CI. Perfil de los personajes: (1) Cardenio, (2) Cura, (3) Dorotea, (4) Lotario, (5) Quijote, (6) Sancho

CII. Perfil de los personajes (1) Narrador, (2) Dueña dolorida, (3) Duquesa, (4) Duque, (5) don Quijote, (6) Ricote, (7) Sansón Carrasco, (8) Sancho, (9) Teresa

Hemos tratado de obtener un “perfil lingüístico” de cada personaje dependiendo del porcentaje en que usa cada categoría distinta. Un número único para cada uno se ha obtenido como la media armónica de los porcentajes de uso. Tal número, con independencia de otras consideraciones, no nos interesa en sí mismo, sino como modo de agrupar perfiles cercanos y tratar de obtener qué grupos están usando el lenguaje de forma similar.

Grpos de uso similar en AV: 1.- Autor/Narrador (2.1) 2.- Narrador del rico desesperado (2.3) 3.- Narrador de los felices amantes (2.4) 4.- Álvaro Tarfe, Gregorio, don Quijote (2.7) 5.- Bárbara, Luisa, Sancho (2.8)

Grupos similares en CI: 1.- Autor/prólogo (2.5) 2.- Cardenio, Narrador (2.6) 3.- Cabrero (Pedro), Narrador del cautivo (2.7) 4.- Camila, Dorotea

(2.8) 5.- Cura, Lotario, Narrador del curioso impertinente, don Quijote, Sancho (2.9)

Grupos similares en CII: 1.- Narrador (2.3) 2.- Dueña dolorida, Teresa (2.7) 3.- don Quijote, Ricote, Sancho (2.9) 4.- Duque, duquesa, Sansón Carrasco (3)

Un uso de categorías muy superior a lo esperado por azar lo encontramos en los personajes populares de Avellaneda (Bárbara y Sancho) con CS, las conjunciones subordinadas. Este mismo rasgo caracteriza en CI y CII a los lenguajes de don Quijote y de Sancho. Los lenguajes del Sancho de Avellaneda y de CI se caracterizan también por el elevado porcentaje de adverbios.

### **Conclusiones**

1.- Los sublenguajes más diferenciados del resto son los del narrador del cautivo y el narrador de CII, muy próximos estos entre sí.

2.- Los sublenguajes de los que un mayor número de sublenguajes cervantinos está lejano son el del narrador de rico desesperado y el de la priora Luisa de Avellaneda.

3.- De los sublenguajes cervantinos, el que más lejano resulta de Avellaneda es el de Camila. El de Camila es también el sublenguaje más lejano para el resto de los lenguajes de Cervantes.

4.- El sublenguaje de don Quijote es en los 3 casos neutro, ni se aproxima ni se separa mucho de ningún otro.

5.- Ningún rasgo de los que según Pennebaker y colaboradores correlacionan con tipos de etiquetado resalta claramente. En concreto, ni los rasgos asociados a género (masculino, femenino) ni a estatus social (como la longitud de las palabras, el mayor o menor uso de primera persona, pronombres, artículos, verbos, preposiciones) agrupan claramente a sublenguajes específicos. Parece haber una tendencia en Avellaneda a que sean más homogéneos los sublenguajes por estatus (Sancho, Bárbara) que por género (Bárbara, Luisa); en Cervantes ocurre lo contrario: tienden a ser más homogéneos los sublenguajes por género que por estatus.

6.- En general, el lenguaje de Cervantes es más homogéneo que el de Avellaneda.