

EVOLUCIÓN DE LA TEMÁTICA DE LA REVISTA METODOLOGÍA DE ENCUESTAS. ANÁLISIS DEL VOCABULARIO DE LOS TÍTULOS Y DE LAS PALABRAS CLAVE

Elena Abascal Fernández (*).

E-mail: eabascal@unavarra.es. Teléfono 948 169197

Vidal Díaz de Rada Igúzquiza ()**.

E-mail: vidal@unavarra.es Teléfono 948 168903

Ignacio García Lautre (*).

E-mail: nacho@unavarra.es Teléfono 948 169198

M. Isabel Landaluze Calvo (*)**.

E-mail: iland@ubu.es Teléfono 947 259388

(*) Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra

(**) Departamento de Sociología, Universidad Pública de Navarra

(***) Departamento de Economía Aplicada, Universidad de Burgos

RESUMEN. En este artículo se realiza un estudio sobre el contenido de la revista *Metodología de Encuestas* al cumplirse 11 años de la publicación del primer número; con el objetivo principal de comprobar el cumplimiento de la vocación multidisciplinar con la que inició su andadura. El estudio se focaliza en el análisis del vocabulario empleado por los autores tanto en los títulos como en las palabras clave de los artículos publicados. Dada la naturaleza de la información seleccionada se ha empleado el enfoque de la Estadística Textual basada en las técnicas factoriales del Análisis de Correspondencias y del Análisis de Clasificación, llevando a cabo el análisis y descripción completa de la riqueza de aspectos tratados en las publicaciones. Los resultados obtenidos se han completado clasificando previamente los artículos atendiendo a la estructura temática interna de la revista, considerando seis áreas: aspectos generales, cuestionario, muestreo, trabajo de campo, tratamiento de datos y otros.

PALABRAS CLAVE: Análisis de Datos Textual, Análisis de Correspondencias, Análisis de Clasificación.

ABSTRACT. In this paper we make a research on the subjects of the journal *Metodología de Encuestas* 11 years after first issue. The main purpose of verifying compliance about its multidisciplinary vocation. The study focuses on analyzing the vocabulary used by the authors in the titles as keywords in the articles published in the journal. Given the nature of the information gathered the approach to Statistical Textual factor based on the techniques of Correspondence Analysis and Classification Analysis has been used. The results have been completed previously classified items in response to the internal thematic structure of the magazine, considering six areas: general aspects, questionnaire, sampling, fieldwork, data processing and others.

KEYWORDS: Textual data, correspondence analysis, classification analysis.

Recibido: 3 de diciembre 2009

Revisado: 25 de febrero 2010

Aceptado: 15 de abril de 2010

Los autores desean agradecer las sugerencias recibidas por dos evaluadores anónimos.

Introducción y objetivos

En 1999 se publicó el primer número de la revista metodología de encuestas. En la presentación Vicente Manzano se preguntaba ¿qué se investiga acerca de las encuestas? Desde su punto de vista "las investigaciones en metodología de encuestas obedecen a dos perspectivas diferentes. Por un lado, existen trabajos cuya motivación es mejorar, innovar o proponer algo con respecto a algún elemento del proceso: modelos de muestreo, resolución de problemas de inferencia, tratamiento de la no respuesta,... Por otro lado, las encuestas deben responder a exigencias particulares según los contextos de aplicación concretos lo que lleva a plantear estudios y reflexiones orientadas a problemas de investigación como ocurre con las encuestas electorales, las estadísticas oficiales,..." En consecuencia, el ámbito de investigación en metodología de encuestas es muy amplio. La utilización de las encuestas se realiza en multitud de campos por investigadores formados en disciplinas diferentes como Sociología, Psicología, Estadística, Marketing, Ciencias de la información,... En este contexto, *Metodología de Encuestas* nace como una revista "cuya vocación es ser construida y consumida por quienes piensan y realizan encuestas". Pasados 11 años nos planteamos ahora si ha cumplido con esta vocación. Es hora de analizar el contenido de la revista en este periodo.

En el año 2004 se publicó un primer estudio titulado "un lustro de metodología de encuestas" (G. Guillera, J. Gómez Benito y J. L. Padilla, 2004). En este artículo se presenta un estudio bibliométrico de la revista durante sus primeros cinco años. Se realiza un análisis de su contenido con criterios bibliométricos y se presentan argumentos sobre la evolución esperable del campo de la metodología de encuestas y la deseable adaptación de la revista.

Transcurridos seis años, afrontamos un nuevo estudio de la revista contemplando todos los artículos publicados durante once años. En este caso no se parte de un punto de vista bibliométrico sino que se plantea un estudio de los temas publicados, el análisis del contenido de la revista a través del estudio del vocabulario de los títulos de los artículos, así como de las palabras clave utilizadas.

Este enfoque se basa en el empleo de la Estadística Textual para explorar el vocabulario utilizado en los títulos y en las palabras clave. Además de intentar responder a la pregunta de si la revista ha cumplido con su objetivo de ser multidisciplinar, detectará los temas que han suscitado más interés a los autores.

Este trabajo comienza con una breve presentación de la metodología y de los datos analizados. En el apartado siguiente se presentan los resultados del análisis de los títulos y de las palabras clave. En primer lugar se realiza un estudio descriptivo a través de las frecuencias y seguidamente se realiza el análisis por áreas temáticas.

Datos y Metodología

Los documentos incluidos en el estudio han sido todos los artículos aparecidos en la revista desde el año 1999 hasta el 2009, obteniendo un conjunto de 154

observaciones¹. Los artículos han sido clasificados atendiendo a la estructura interna de la revista considerando seis áreas temáticas: aspectos generales, cuestionario, muestreo, trabajo de campo, tratamiento de datos y otros.

Las variables utilizadas han sido el área temática del artículo, el título completo y las palabras clave. Las dos últimas variables son de naturaleza textual y como tal se han tratado, teniendo en cuenta cada palabra o forma, así como el contexto en que se sitúan. La otra variable es cualitativa y se ha utilizado fundamentalmente para definir grupos. La información así estructurada se ha analizado mediante el programa estadístico SPAD, más concretamente con el módulo de Estadística Textual.

El análisis exhaustivo del vocabulario de los artículos se ha realizado a partir de las dos variables textuales señaladas. En primer lugar se ha procedido al análisis de las palabras clave y en segundo lugar se ha repetido el análisis para el título completo del artículo. Ambos estudios tienen un procedimiento común. Inicialmente se construye una lista de palabras² tras eliminar las preposiciones, artículos, conjunciones, etc. Esto es, el análisis se realiza a partir de las llamadas “palabras plenas”. Se retiene el léxico de las formas que cuentan cada una al menos con 3 ocurrencias. Con esta forma de proceder desaparecen aquellas palabras que durante los 11 años solamente habrán aparecido 1 o 2 veces, que pueden ser palabras de un tema muy específico o, simplemente, sin interés.

A partir de las palabras retenidas y de las variables categóricas se realizan análisis de las frecuencias y se construyen diferentes tablas de contingencia o tablas léxicas agregadas que cruzan las palabras del título o palabras clave con las categorías asociadas a las áreas temáticas. Estas tablas se analizan mediante dos métodos multivariantes complementarios, el Análisis de Correspondencias y Análisis de Clasificación. El primero proporciona una visión gráfica simplificada de las asociaciones entre las palabras y las áreas temáticas Lebart, L (1994), Lebart, L., Salen, A. y Bécue, M. (2000), Lebart, L., Salem, A. y Berri, E. (1998). El Análisis de clasificación realizado sobre los factores obtenidos en el Análisis de Correspondencias lo completa proporcionando grupos de palabras que se utilizan en las mismas áreas temáticas, Lebart, L. (1994). Los principales resultados numéricos y gráficos se presentan en la siguiente sección.

Resultados

En el cuadro 1 se refleja la distribución de los 154 artículos según áreas temáticas. Exceptuando la categoría “otros”, que incluye gran variedad de artículos, las áreas a las que se ha dedicado mayor número de artículos han sido “trabajo de campo” (35, un 23%) y “tratamiento de datos” (29 artículos, un 19%).

¹ Conviene precisar que no se incluyen los artículos de opinión ni comentarios de artículos y libros.

² Más exactamente formas de palabras: el singular, se considera una palabra diferente del plural.

Cuadro 1.

Reparto de los artículos por áreas temáticas

Área temática	Número de artículos	% de artículos
Aspectos generales	18	11,7
Cuestionario	18	11,7
Muestreo	17	11,0
Trabajo de campo	35	22,7
Tratamiento de datos	29	18,8
Otros	37	24,0
Total	154	100,0

*i) Análisis del vocabulario**a) Análisis del vocabulario de los títulos*

El estudio comienza con la construcción de las palabras y los segmentos repetidos (secuencia de palabras) incluidos en los títulos de los artículos. El número de palabras –o formas– que aparecen en el recuento inicial de los 154 títulos es de 1.840 palabras, siendo 599 de ellas diferentes, lo que nos da una idea de la riqueza del vocabulario utilizado por los autores en los títulos de los artículos; así como de la variedad de aspectos considerados en la revista. El cuadro 2 presenta una lista

Cuadro 2.

Palabras no-herramienta más frecuentes en los títulos

Palabras empleadas	Frecuencia	Palabras empleadas	Frecuencia	Palabras empleadas	Frecuencia
encuestas	78	estudio	5	resultados	3
especial	25	social	5	efecto	3
análisis	23	campo	5	calculo	3
encuesta	18	población	5	revisión	3
investigación	13	información	5	múltiples	3
respuesta	13	estadístico	4	España	3
eurobarómetro	12	medición	4	cuestionarios	3
internet	10	presentación	4	pregunta	3
comisión	10	no	4	salud	3
europea	10	electoral	4	telefónica	3
flash	10	mercados	4	comportamiento	3
mediante	9	diseño	4	nacional	3
calidad	9	metodología	4	México	3
representación	9	estadística	4	modelo	3
ciudadana	8	incidencias	4	respuestas	3
muestreo	8	escalas	4	representatividad	3
datos	7	móvil	4	preguntas	3
estimación	7	técnicas	4	formatos	3
aplicación	7	electorales	4	uso	3
trabajo	6	opinión	4	entrevista	3
telefónicas	5	estudios	3	diferentes	3
cuestionario	5	investigaciones	3	género	2

depurada, sin las llamadas palabras herramienta (artículos, preposiciones,...) de las formas gráficas utilizadas y su frecuencia.

Si se tiene en cuenta que solo hay 6 artículos que en su título repiten la palabra encuesta o encuestas, se tiene que casi la mitad incluyen en el título la palabra *encuestas*, llegando a un 58% si consideramos también este término en singular. El 15% incluye la palabra *análisis*.

El examen de las palabras y sus frecuencias se queda muy limitado pues su significado depende mucho de la frase o contexto en que se encuentra ubicada, lo que impide hacer afirmaciones sobre relaciones. Los segmentos repetidos completan la información que proporcionan las palabras. Se definen como una secuencia de dos o más palabras no separadas por un delimitador de secuencia que aparecen más de una vez en el texto. Por ejemplo “trabajo de campo”, “análisis de datos”, “investigación de mercados”. Están compuestos por varias palabras dotadas de sentido propio y que no siempre se deduce de las palabras que entran en su composición, si éstas se consideran aisladamente.

En el cuadro 3 se recoge la lista de segmentos repetidos con algún significado (se elimina, por ejemplo, los formados por una conjunción y un artículo) obtenidos tras eliminar del vocabulario inicial las formas que aparecen menos de 3 veces.

Cuadro 3:
Lista de segmentos repetidos por orden de frecuencia

Frecuencia	Texto del segmento	Frecuencia	Texto del segmento
21	las encuestas	5	trabajo de campo
14	especial encuestas	5	de respuesta en
14	análisis de	5	encuestas telefónicas
13	de encuestas	5	en la investigación
12	encuestas en	5	sobre la
11	de las encuestas	5	estimación de
10	encuestas y	5	encuestas de
10	la encuesta	5	en las encuestas
9	de respuesta	5	de datos
9	encuestas en internet	4	del trabajo
8	encuestas y representación ciudadana	4	análisis de datos
8	de la encuesta	4	muestreo en
7	especial encuestas en internet	4	investigación d mercados
7	especial encuestas y representación ciudadana	4	las encuestas telefónicas
7	en encuestas	4	diseño de
7	la investigación	4	el análisis
6	respuesta en	3	calidad de las encuestas

La palabra predominante en los títulos es *encuesta*, tanto en singular como en plural. No obstante, puede observarse que algunos títulos hacen referencia a *especial encuestas*, otros a *encuesta en*, o bien *de encuestas*, lo que significa que no todos tratan el mismo tema. La destacada frecuencia en *especial encuestas* y *representación ciudadana* tiene su explicación en la publicación de un monográfico dedicado a este tema.

La información proporcionada por el cuadro 4 pone de manifiesto que los títulos más largos corresponden a los artículos sobre trabajo de campo y de tratamiento de datos. Estos artículos son, a su vez, los que utilizan mayor número de palabras diferentes (junto a los artículos incluidos en la categoría “otros”), lo que podría ser indicativo de que en estas áreas hay mayor variedad de aspectos tratados.

Cuadro 4.
Reparto de las palabras por áreas

Áreas	Nº de palabras en el área	% del total de palabras	Media de palabras por artículo	Nº de palabras diferentes en el área	% palabras del área	Número de artículos	% del artículos
Aspectos generales	198	8,50	11,0	99	50,00	18	11,7
Cuestionario	283	12,15	15,7	140	49,47	18	11,7
Muestreo	236	10,13	13,9	142	60,17	17	11,0
Trabajo de campo	560	24,03	16,0	215	38,39	35	22,7
Tratamiento de datos	470	20,17	16,2	228	48,51	29	18,8
Otros	583	25,02	15,8	216	37,05	37	24,0
Total	2330	100,00	15,1			154	100,0

b) *Análisis del vocabulario de las palabras clave*

En el estudio focalizado en las palabras clave hay que señalar la reducción de las observaciones a 96 artículos por la inexistencia de estas palabras en 48 de los artículos seleccionados inicialmente. El número total de palabras clave es de 888, siendo 387 de ellas diferentes, lo que confirma la gran variedad de temas considerados en la revista.

Cuadro 5.

Palabras no-herramienta más frecuentes en las palabras clave

Palabras	Frec.	Palabras	Frec.	Palabras	Frec.	Palabras	Frec.
análisis	27	entrevista	5	sucesivo	3	representatividad	3
encuestas	22	tablas	5	telefonía	3	control	3
muestreo	20	metodología	5	móvil	3	formatos	3
datos	16	telefónica	5	objetos	3	precisión	2
encuesta	15	tasa	4	deseabilidad	3	relativa	2
respuesta	13	estadístico	4	apareamiento	3	superpoblación	2
trabajo	12	cuestionarios	4	preferencias	3	mercados	2
cuestionario	12	modelo	4	software	3	poblacional	2
investigación	11	telefónicas	4	múltiples	3	simple	2
campo	10	ítems	4	estadística	3	finitas	2
incidencias	8	correspondencias	4	recogida	3	artificiales	2
escalas	8	conjunto	4	segmentación	3	proximidad	2
estimador	7	error	3	población	3	vida	2
no-respuesta	7	empresas	3	modelos	3	cati	2
social	6	índice	3	no	3	predicción	2
clasificación	6	entrevistas	3	variable	3	ganancia	2
diseño	6	muestral	3	imputación	3	estabilidad	2
calidad	6	errores	3	fracción	3	múltiple	2
redes	2	preguntas	2	significación	2	medida	2
efectos	2	construcción	2	procedimientos	2	electoral	2
razón	2	cobertura	2	likert	2	política	2
aplicaciones	2	ponderación	2	aleatorio	2	cluster	2
prueba	2	muestra	2	satisfacción	2	administración	2
personales	2	textual	2	tamaño	2	grupo	2
participación	2	validez	2	programas	2	ordenación	2
factorial	2	poblaciones	2	ordenador	2	respuestas	2
voto	2						

La palabra clave más frecuente es *análisis*, lo que podría hacer pensar que en muchos casos se trata de artículos centrados en aplicaciones empíricas. Sin embargo los segmentos repetidos recogidos en el cuadro 6 permiten vislumbrar que, en ocasiones, es solo una parte del nombre de una metodología estadística como Análisis Conjunto o Análisis de Correspondencias.

Cuadro 6.
Lista de segmentos repetidos por orden alfabético

Frec.	Texto del segmento	Frec.	Texto del segmento
4	análisis conjunto	4	encuesta telefónica
9	análisis de	3	encuestas telefónicas
4	análisis de correspondencias	6	escalas de
3	análisis de datos	4	escalas de clasificación
5	calidad de	5	estimador de
5	de clasificación	3	formatos de
14	de datos	3	fracción de apareamiento
3	de encuesta	4	incidencias no-respuesta
7	de encuestas	3	modelo de
3	de escalas	3	muestreo sucesivo
10	de respuesta	3	recogida de datos
3	de trabajo	4	tasa de respuesta
4	del cuestionario	3	telefonía móvil
3	deseabilidad social	10	trabajo de campo
3	diseño de	3	trabajo de campo incidencias

Con el fin de profundizar en los aspectos estudiados en cada área temática, objetivo principal de este estudio, será preciso realizar varios análisis de las tablas léxicas agregadas. Ahora bien, para la obtención de las tablas léxicas definitivas debe tenerse en cuenta la existencia de diversas formas que aparecen tanto en singular como en plural. Asimismo, se observan diferentes conjugaciones de un mismo verbo. Esta situación recomienda, antes de llevar a cabo los análisis factoriales, a lematizar y convertirlas en una única forma.

ii) Estudio por áreas temáticas

a) Estudio de las palabras clave

El estudio de las palabras clave por área temática revela que los artículos relativos a *aspectos generales* emplean una media de 18,2 palabras clave por artículo (Cuadro 7), un número notablemente superior a la media de 10,8. También se observa un valor alto en *cuestionario*.

Cuadro 7.
Reparto de las palabras clave por áreas

Área temática.	Nº de palabras Clave en el área	% del total de palabras clave	Media de palabras clave por artículo con palabras clave	Nº de palabras clave diferentes en el área	% palabras clave del área	Número artículos	Número artículos con palabras clave
Aspectos generales	91	8,78	18,2	57	62,64	18	5
Cuestionario	223	21,50	12,4	119	53,36	18	18
Muestreo	164	15,81	10,25	94	57,32	17	16
Trabajo de campo	222	21,41	10,6	95	42,79	35	21
Tratamiento de datos	248	23,92	9,5	152	61,29	29	26
Otros	89	8,58	8,9	66	74,16	37	10
Total	1037	100,00	10,8			154	96

El número de artículos que no disponen de palabras clave es importante, algo más del 30%, pero no se distribuye igualmente en todas las áreas temáticas. Como se observa en el cuadro 7 esta ausencia afecta principalmente a los bloques temáticos relativos a *aspectos generales* y a *otros*.

La lematización constituye la etapa previa a la construcción de la tabla léxica agregada, tabla que cruza las palabras clave con las categorías asociadas a las áreas temáticas. Se ha procedido a declarar semejantes las palabras que aparecen en singular y en plural (por ejemplo encuesta y encuestas), y también en algunos casos cuando la misma palabra aparece en distinto género, como conjunta y conjunto. Se siguen considerando únicamente aquellas palabras plenas de sentido, que aparecen un mínimo de 3 veces.

Con el fin de detectar diferencias entre las áreas temáticas en el uso de palabras clave se realiza un Análisis de Correspondencias de la tabla léxica agregada, completado con un Análisis de Clasificación sobre todos los ejes factoriales obtenidos (5), adjuntando como ilustrativos los segmentos repetidos. Al ser ilustrativos no intervienen en la obtención de los factores pero enriquecen la interpretación, ya que se pueden representar sobre los planos y ver las palabras en su contexto.

El gráfico 1 corresponde al plano principal que determinan los dos primeros factores. En él se representan las áreas temáticas, las palabras clave y las clases o grupos de palabras que se utilizan en las mismas áreas temáticas. Se interpretan las

asociaciones entre palabras y áreas, lo que permite detectar cuales son las principales palabras clave asociadas a cada área y por lo tanto los temas más tratados.

Entre los 3 primeros factores recogen el 78,36% de la información. El primer factor de diferenciación representa la oposición entre las palabras clave utilizadas en los artículos asociados al área de *muestreo* y las utilizadas en el resto de las áreas temáticas. El segundo factor refleja la contraposición entre las palabras clave utilizadas en los artículos relativos al *trabajo de campo* frente a los asociados al *tratamiento de datos*. Por su parte, el tercer factor opone las palabras clave utilizadas en los artículos relacionados con el *questionario* frente a los relativos a *otros temas* y a *tratamiento de datos*.

Gráfico 1.

Representación, sobre el plano (1, 2), de las áreas temáticas, palabras clave y los grupos de palabras

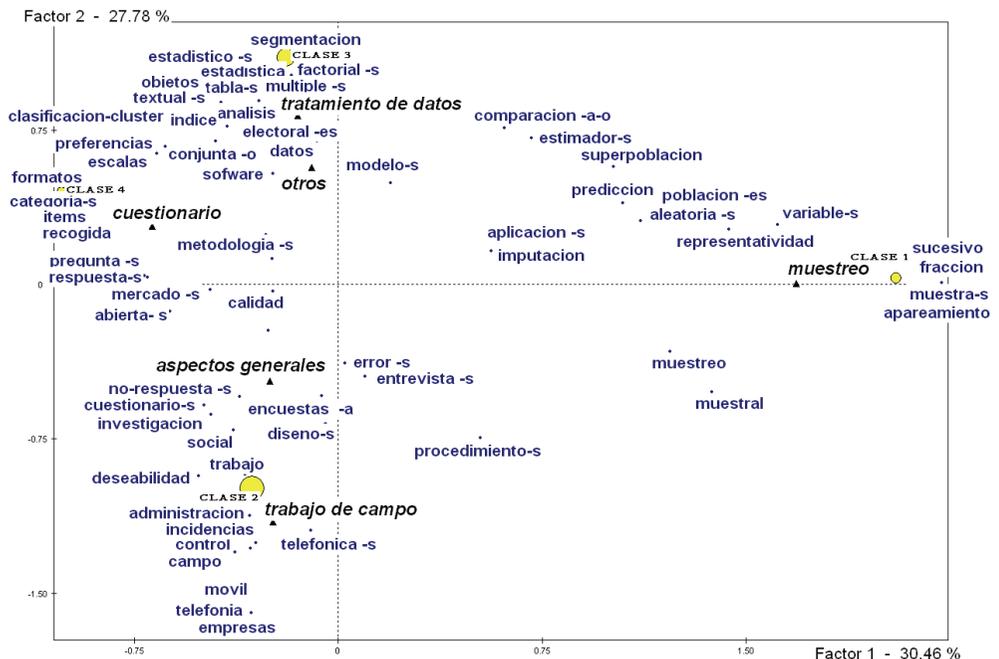
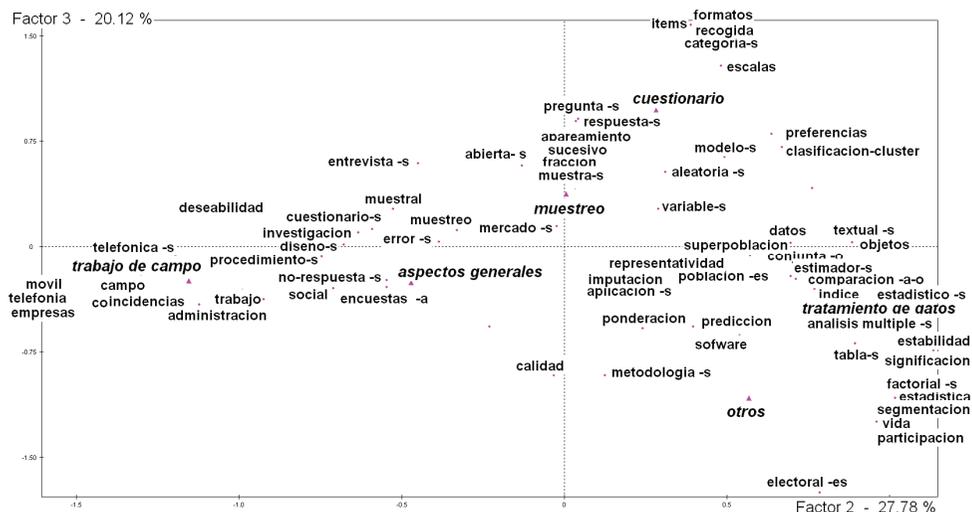


Gráfico 2.
 Representación, sobre el plano (2, 3), de las áreas temáticas y palabras clave



En los gráficos 3 y 4 se proyectan, con carácter ilustrativo, los segmentos repetidos asociados a las palabras clave. Al tratarse de una *proyección* sobre los tres primeros factores, es una representación clarificadora puesto que sitúa cada palabra en su contexto. Así, por ejemplo, se observa que los temas más citados en análisis de datos son *análisis de correspondencias*, *análisis conjunto* y *de clasificación*. La expresión de *encuestas* se utiliza en todo tipo de artículos, al igual que *de encuesta*, pero esta última con mayor frecuencia en los dedicados al muestreo. Los principales temas tratados en los artículos relativos al cuestionario son los que tienen que ver con las *escalas de*, *formatos* y también los que incluyen *recogida de datos*. Señalar que esta expresión es usada más frecuentemente en esta área que en la de muestreo. En cuanto al área temática trabajo de campo los aspectos más considerados son los de la *no-respuesta*, *tasa de respuesta*, *encuesta telefónica* y *diseño de*. La palabra *respuesta* se emplea fundamentalmente en dos áreas: en el trabajo de campo, asociada con la *tasa de respuesta*, y en los artículos sobre el cuestionario.

Cuadro 8.
Descripción de la clasificación de las palabras clave

Áreas temáticas	% de palabras del área en la publicación	% de palabras del área en la clase	% de la clase en el área	p-valor	valor-test (*)
Clase 1 / 4					
Muestreo	15,82	73,24	72,22	0,000	12,4
Clase 2 / 4					
Trabajo de campo	22,64	51,35	92,23	0,000	12,44
Aspectos generales	8,57	15,14	71,79	0,000	3,95
Clase 3 / 4					
Tratamiento de datos	24,18	57,14	72,73	0,000	10,56
Otros	7,25	13,57	57,58	0,001	3,14
Clase 4 / 4					
Cuestionario	21,54	79,66	47,96	0,000	10,35

(*) Valor-test es un indicador proporcionado por el programa SPAD. Mide la diferencia entre las dos proporciones de palabras, en la clase y en el total, y se expresa en número de desviaciones de la ley normal de manera que las diferencias son más significativas cuanto mayor es el valor-test en valor absoluto.

Los resultados obtenidos a través del Análisis de Correspondencias se pueden corroborar y completar con los obtenidos a partir del Análisis de Clasificación. Este último detecta cuatro grupos o clases de palabras clave (gráfico 1):

- **La clase 1** está formada por palabras clave que se utilizan principalmente en los artículos dedicados al *muestreo*. Está compuesta por 20 palabras diferentes de las cuales las más representativas son: *variable(s)*, *muestreo*, *muestral*, *aleatoria(s)*, *población (es)*, *apareamiento*, *sucesivo*, *simple*, *razón*, *precisión*. Se localiza en el primer cuadrante del plano principal, casi sobre el primer factor.

- **La clase 2** está integrada por 28 palabras diferentes utilizadas principalmente en los artículos centrados en el *trabajo de campo*. Las más representativas son: *encuesta(s)*, *no respuesta*, *investigación*, *trabajo*, *social*, *error(es)*, *cuestionario(s)*, *campo*, *deseabilidad*, *control*. Está situada en el tercer cuadrante del plano principal.

- **La clase 3** contiene 33 formas diferentes que se emplean en los artículos asociados tanto al tratamiento de datos como a otros temas. Las más representativas son *análisis*, *múltiple(s)*, *datos*, *objetos*, *textual(s)*, *correspondencias*, *estadístico(s)*, *tabla(s)*, *estadística*, *factorial(s)*. Se sitúa en el segundo cuadrante del plano.

- **La clase 4** contiene 14 palabras clave diferentes que se utilizan en los artículos dedicados al cuestionario. Las más representativas son *escalas*, *respuesta(s)*, *categoría(s)*, *construcción*, *formatos*, *validez*, *likert*, *recogida*, *ítems* y *electrónica(o)*. Aparece localizado en el segundo cuadrante del plano principal.

b) Estudio de las palabras de los títulos por áreas temáticas

El análisis de las palabras clave es limitado ya que, como se ha comentado, hay un número importante de artículos que no las incluyen. No obstante, ha proporcionado una primera visión de la riqueza de los temas tratados en la revista Metodología de Encuestas a lo largo de sus 11 años de existencia.

El análisis de las palabras que constituyen los títulos de los artículos en relación con las áreas temáticas es más completo y, por tanto, va a proporcionar una visión más amplia de esta riqueza temática. En términos generales se puede destacar la existencia de 4 ejes que recogen el 87,23% de la información y 5 clases de palabras relevantes. Esto es, un eje y una clase más que en el análisis de las palabras clave.

El primer eje es un factor de *otros*. Está construido casi en un 70 % por el área temática *otros*, en oposición a las restantes áreas temáticas. Este eje surge en gran parte por el hecho de que se han dedicado dos números especiales en los que todos los artículos incluyen la misma frase como parte del título. Por un lado, se dedicó un número especial al "euro barómetro flash 151 de la comisión Europea", que se asocia con estas palabras y el área *otros*. Por otro lado, el "especial encuestas y representación ciudadana" se asocia con formas que aparecen también en el resto de áreas temáticas.

El segundo eje es un factor de *muestreo*. En su construcción interviene el área de *muestreo* en más del 70%, asociado a palabras como *cálculo*, *valores*, *modelos*, *estimadores* y, fundamentalmente a *muestreo(s)*. En los artículos de muestreo se utilizan expresiones como *aplicación de*, *de muestreo*, *de encuestas*, *estimación de la*, *de encuesta...* lo que indica que no son artículos de investigación teórica en muestreo sino que principalmente se aplican las técnicas de muestreo para estimaciones en las encuestas.

El tercer factor está creado en más del 90% por la diferenciación entre *cuestionario* y *trabajo de campo*. El cuestionario se asocia con las palabras *escalas*, *y formatos*, *electrónica(s)* y *cuestionario(s)*, mientras que las palabras más asociadas con trabajo de campo son *encuestadores*, *internet*, *móvil*, *telefónica(s)* *trabajo(s)* y *campo*. En los trabajos de cuestionario se utilizan expresiones como *de escalas*, *de respuesta*, *de respuesta en*, *de la información*, *del comportamiento*, *en las investigaciones*, *estadístico de*. En los artículos de trabajo de campo las más frecuentes son *análisis de la*, *del trabajo de campo*, *encuestas telefónicas*, *las encuestas telefónicas*, *encuestas en Internet*, *especial encuestas en Internet*, *incidencias*, *trabajo de campo*. Se abordan principalmente los problemas de los métodos más modernos de recogida trabajo de campo.

El cuarto factor refleja principalmente la oposición entre *cuestionario*, por un lado, y *tratamiento de datos* y *aspectos generales*, por otro. Por su parte, el cuestionario se asocia con las palabras *escalas*, *formatos*, *respuesta(s)*, *telefónicas* y *cuestionario(s)*. Las palabras más asociadas a las áreas relativas al tratamiento de datos y a los aspectos generales son *análisis*, *calidad*, *investigaciones*, *mercados*, *metodología*, *múltiples*, *revisión*, *usuario(s)* *técnica* y *uso*. Siendo las expresiones más utilizadas en estas áreas *análisis de datos*, *de las encuestas*, *el análisis*, *en la investigación*, *en el análisis de*, *encuesta de*, *estudio de*, *una aplicación*. Estos últimos son artículos en los que se emplean técnicas multivariantes (frecuentemente

Gráfico 7.
Representación, sobre el plano (1, 2), de las áreas temáticas y de los segmentos repetidos del título.

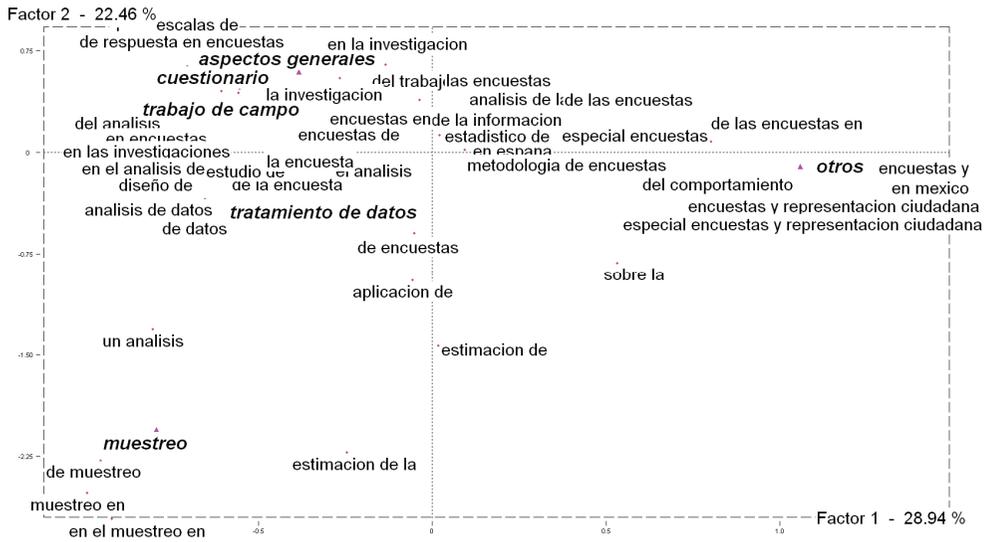
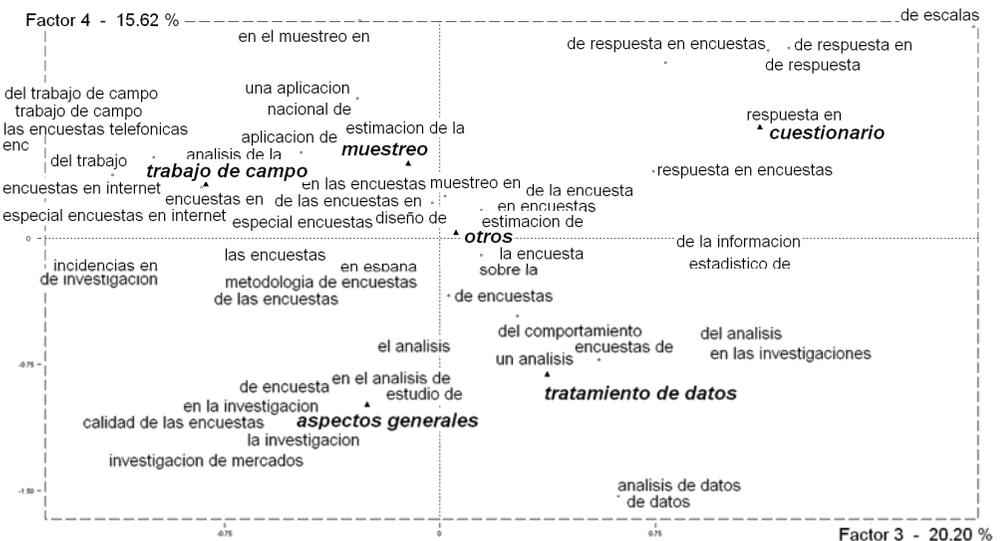


Gráfico 8.
Representación, sobre el plano (3, 4), de las áreas temáticas y de los segmentos repetidos del título.



denominadas técnicas de análisis de datos en la escuela francesa) para realizar estudios de encuestas, esto es, son aplicaciones de las técnicas. Sobre los gráficos anteriores se pueden apreciar las asociaciones comentadas.

Se detectan 5 grupos o clases o grupos de palabras:

- La **clase 1** contiene 27 palabras diferentes que emplean en tratamiento de datos y aspectos generales. Las más representativas son: *encuesta(s)*, *metodología(s)*, *aplicación(s)*, *índice(s)*, *información*, *estudio(s)*, *análisis*, *mediante*, *técnica*, *informatizada(s)*. Se localiza sobre el primer factor en el campo negativo.

- La **clase 2** contiene 11 palabras diferentes que se emplean en los títulos de artículos del área de cuestionario. Las más representativas son *respuesta(s)*, *cuestionario(s)*, *medición*, *modelo*, *electrónica*, *salud*, *efecto(s)*, *pregunta(s)*, *escalas* y *formatos*.

- La **clase 3** está formada 12 palabras diferentes utilizadas principalmente en el *trabajo de campo*. Las más representativas son: *telefónica(s)*, *internet*, *campo*, *entrevistadores*, *encuestadores(s)*, *móvil*, *trabajo*, *social*, *incidencias* y *nacional*.

- La **clase 4** está formada por palabras que se utilizan principalmente en los títulos de los artículos dedicados al *muestreo*. Está compuesta sólo por 4 palabras: *muestreo(s)*, *estimadores*, *cálculo* y *modelos*. Se localiza en el tercer cuadrante del plano (1,2).

- La **clase 5** está formada por palabras que se utilizan principalmente en los títulos de los artículos clasificados como *otros*. Está compuesta únicamente por 12 palabras. Las más representativas son: *europeo(s)*, *ciudadana*, *comisión*, *electoral(s)*, *flash*, *representación*, *eurobarómetro*, *especial*, *comportamiento*, *sesgo(s)*... Se localiza sobre el primer factor en el campo positivo.

Cuadro 9.

Descripción de la clasificación de las palabras plenas de los títulos

Áreas temáticas	% de palabras del área en la publicación	% de palabras del área en la clase	% de la clase en el área	p-valor	Valor-Test
Clase 1 / 5					
Tratamiento de datos	16,46	27,76	85,00	0,000	6,92
Aspectos generales	7,61	11,84	78,38	0,000	3,45
Clase 2 / 5					
Cuestionario	13,37	60,34	53,85	0,000	9,18
Clase 3 / 5					
Trabajo de campo	24,69	83,05	40,83	0,000	10,11
Clase 4 / 5					
Muestreo	6,79	76,47	39,39	0,000	7,35
Clase 5 / 5					
Otros	31,07	86,92	61,59	0,000	13,83

Conclusiones

Uno de los objetivos que la revista “Metodología de Encuestas” se planteó desde sus inicios fue el de ser una plataforma de publicación de trabajos de y para todos los usuarios de encuestas. Dado que las encuestas tienen aplicación en prácticamente la totalidad de campos de investigación, la revista nació con una clara vocación de ser multidisciplinar.

Este trabajo, en el que se analiza el contenido de estos artículos, permite comprobar el cumplimiento de esta vocación. En concreto se ha considerado de interés el análisis de los títulos y de las palabras clave empleadas por los diferentes autores, ya que proporcionan información suficiente y complementaria sobre la riqueza temática de los diferentes aspectos tratados en las publicaciones. Dada la naturaleza de la información utilizada, se ha empleado el enfoque de la Estadística Textual, basado en las técnicas factoriales del Análisis de Correspondencias y del Análisis de Clasificación.

Los resultados obtenidos se han completado clasificando previamente los artículos atendiendo a la estructura temática interna de la revista, considerando seis áreas: aspectos generales, cuestionario, muestreo, trabajo de campo, tratamiento de datos y otros.

Entre los resultados obtenidos merecen ser destacados, por su contribución a la confirmación del cumplimiento de la vocación multidisciplinar de la revista “Metodología de encuestas”, los siguientes:

- La riqueza del vocabulario, medida en número de palabras distintas, utilizado por los autores, tanto en los títulos como en las palabras clave.
- Las áreas temáticas más proliferas son tratamiento de datos, trabajo de campo y otros.
- Muestreo es un área temática con un comportamiento peculiar frente al resto de áreas. En los artículos de muestreo se utiliza un vocabulario propio y no son, en general, trabajos de investigación teórica, sino aplicaciones de las técnicas de muestreo.
- Las áreas de tratamiento de datos y aspectos generales tienen ciertas similitudes en el vocabulario utilizado. Engloban, principalmente, trabajos que recogen aplicaciones de las técnicas multivariantes (también denominadas técnicas de análisis de datos).
- En trabajo de campo el vocabulario más frecuente pone de manifiesto que los trabajos abordan básicamente los problemas de los métodos más modernos de recogida de información.

A partir de estos resultados podría ser conveniente considerar en el futuro la recomendación de eliminar el término “encuesta” del título de los trabajos, en la medida que está presente en la propia denominación de la revista. A su vez se hace necesario el desarrollo de nuevas áreas temáticas relacionados con los aspectos más analizados en las principales publicaciones internacionales sobre el tema: no respuesta, efectos de los modos de recogida de datos, efectos de los ordenadores, pilotaje de cuestionarios, etc.

Bibliografía

- Abascal E. y Grande, I. (2005.). “*Análisis de encuestas*”. ESIC.
- Abascal, E. y Franco, G. (2002). Metodología de encuestas, vol6. nº 2 163-172. *Análisis textual de encuestas: aplicación al estudio de las motivaciones de los estudiantes en la elección de su titulación*” Metodología de encuestas, vol 4. nº 2 .
- Guillera, G ; Gómez Benito, J. y Padilla, J. L. (2004). *Un lustro de metodología de encuestas*. Metodología de encuestas, vol 6. nº 2 163-172.
- Lebart, L. (1992). *Traitements Statistiques de réponses libres en trois langues : Anglais, Français, Japonais (Etude de préférences alimentaires)* (1992). (avec Hiroshi Akuto) *Les Cahiers de l'analyse des Données*, Dunod, Paris, p 327-352.
- Lebart, L. (1994), “Complementary use of correspondence analysis and cluster analysis” in Greenacre, M. J. and Blasius, (Eds), *Correspondence Analysis in the Social Sciences*, Academic Press, London, pp. 162-78.
- Lebart, L., Salem, A. y Berri, E. (1998). *Exploring Textual Data*, Dordrecht. Kluwer. Academic Publisher.
- Lebart, L., Salen, A. y Bécue, M. (2000). Análisis estadístico de textos. Milenio
- Karl M. van Meter «2008). Analyses of quarter of century of publishing at the BMS. Bulletin de Méthodologie Sociologique (48), octubre 2008 nº 100 pp 6-15
- De Saint Leger, M. (2008). *Comment ont évolué les thématiques des 99 premiers numéros de BMS? Analyse avec un logiciel de fouille de texte*. Bulletin de Méthodologie Sociologique (48), octubre 2008 nº 100 pp 16-33
- Sagombaye Nodjiram y Thomassin M.M. (1992). *Analyse du vocabulaire des titres des publications du comité français de cartographie*. Les Cahiers de l'Analyse des Données vol XVII nº4 pp 471-480
- Sagombaye Nodjiram (1992). *Analyse du vocabulaire mathématique des titres des articles de deux périodiques: un siècle du bulletin de la S.M.F. et du journal de Liouville*. Les Cahiers de l'Analyse des Données vol XVII nº2 pp 133-158
- SPAD: *Système Portable pour l'Analyse des Données* (avec A. Morineau). (1982). CISIA Ed.