

## **LAS PREGUNTAS DE RESPUESTA ABIERTA Y CERRADA EN LOS CUESTIONARIOS. ANÁLISIS ESTADÍSTICO DE LA INFORMACIÓN**

Ramón Álvarez Esteban  
*Universidad de León*

### **RESUMEN**

Frecuentemente podemos encontrar en los cuestionarios anotaciones al margen de las preguntas que los entrevistados realizan para poder explicar o matizar una determinada respuesta. A medida que el nivel de cualificación del encuestado es mayor, el nivel de detalle de las preguntas aumenta, se interroga por conceptos abstractos o la propia naturaleza del tema es más compleja, entonces la extracción de información con preguntas de respuesta cerrada plantea importantes restricciones.

La utilización de preguntas de respuesta abierta en cuestionarios supone un mayor esfuerzo de transcripción que cuando se trabaja con respuestas cerradas, a lo que hay que añadir un cierto grado de subjetividad. Este coste de tiempo, económico y de personal, solamente es justificable —dentro de un análisis coste-beneficio— si los resultados alcanzados son significativamente distintos.

Los análisis estadísticos de correspondencias y de agrupaciones son dos instrumentos fundamentales en el análisis estadístico de datos textuales, con ciertos matices en su aplicación e interpretación que difieren del análisis de datos no textual y que, por tanto, deben ser considerados.

Palabras clave: *Preguntas de respuesta abierta. Análisis estadístico de datos textuales. Análisis de Correspondencias. Análisis Cluster.*

## Introducción

Las preguntas abiertas en cuestionarios, o dicho de manera más precisa, las preguntas con respuesta abierta, constituyen una posibilidad dentro del análisis de encuestas que es necesario considerar a la hora de plantearnos la recogida de información.

Abascal y Franco, en un reciente artículo (Abascal y Franco, 2002), sintetizan una estructurada exposición del enfoque con el que podemos enfrentar el procesamiento o tratamiento estadístico de preguntas con respuesta abierta. La claridad en el planteamiento de los objetivos y el desarrollo seguido por los autores nos permitirá intentar aportar algunos aspectos que no se encontraban, por razones obvias, dentro de los fines del trabajo de los mismos.

Comenzaremos, pues, analizando la propia esencia de las preguntas con respuesta abierta en los cuestionarios, para después referirnos al análisis textual de estas preguntas, y finalizaremos haciendo mención de diversos análisis estadísticos de datos aplicables al estudio de las mismas.

### Las preguntas con respuesta abierta en los cuestionarios

Tradicionalmente, se ha venido asumiendo que las preguntas con respuesta abierta en los cuestionarios proporcionan una información de gran riqueza que es útil sobre todo en las fases iniciales de un estudio (por ejemplo en cuestionarios piloto). Esta visión podría sintetizarse en la utilización de la pregunta abierta como el instrumento más adecuado para determinar las modalidades de una variable que deben explicitarse en el cuestionario final. En otras palabras, abrir inicialmente la pregunta para cerrarla finalmente.

Un segundo enfoque sobre la utilización de este tipo de preguntas estaría relacionado con el control y depuración de la información recogida tradicionalmente en un cuestionario. Correspondería al *¿por qué?* que frecuentemente encontramos detrás de una pregunta con respuesta cerrada. Es conocido que una de las más claras fuentes de error en la recogida de información está en la respuesta del entrevistado cuando no se tiene opinión, o se desconoce el tema sobre el que se debe opinar. En este caso, una pregunta cerrada intentaría construir un experimento de laboratorio que se aleja en gran medida de las condiciones de actuación del encuestado en situaciones normales. En términos generales, un decisor que no tiene un conocimiento suficiente ante la elección de diferentes alternativas podrá escoger entre alguna de las modalidades que se le presenten si el coste no es elevado o si los resultados de la elección no tienen demasiada importancia, como puede ser el momento de elegir un plato de comida atendiendo a lo sofisticado del nombre que el cocinero haya querido darle.

Sin embargo, cuando tenemos que elegir entre alternativas con mayor coste, o que afectan de forma importante a otros aspectos de nuestra vida, generalmente optamos por retrasar la toma de decisiones hasta recoger la información que juzgamos necesaria. En otras palabras, la mayor implicación o trascendencia de la decisión que debe realizarse está en relación directa con el tiempo que utilizamos para realizar la elección. La consideración de las preguntas con respuesta abierta, desde este punto de vista de control, no

plantea mayores dificultades que la de optar por eliminar o no las respuestas sin justificación o insuficientemente justificadas.

El tercer enfoque está relacionado con el grado de detalle en las preguntas y respuestas que se debe utilizar en un cuestionario. El investigador debe ser consciente de que la determinación de un número grande de categorías puede proporcionar al encuestado una información que no poseía a priori, pero también que si entre los encuestados existen diferentes niveles de experiencia, los que posean un mayor conocimiento sobre el tema pueden sentirse incómodos con las modalidades sugeridas al considerar que representan conceptos demasiado genéricos, con lo que habría que matizar, e incluso dividir cada modalidad, para poder recoger respuestas más precisas. Las preguntas con respuestas libres permiten al encuestado, en este caso, fijar su propio nivel jerárquico de detalle de las modalidades o establecer el nivel de decisión en función de su experiencia.

De Groot (Libby, 1985: 79) realizó un experimento con jugadores de ajedrez distribuidos en tres grupos de individuos según su nivel de experiencia. En un primer análisis se mostró un tablero con las piezas en un momento determinado de una partida; mientras que en el segundo, la posición de las piezas era aleatoria. Al tratar de reconstruir la posición sugerida, los resultados mostraron que en el caso aleatorio el nivel de recuerdo era igual en todos los grupos, mientras que si la posición correspondía a una partida interrumpida, un mayor nivel de experiencia mostraba claramente un mayor acierto, ya que eran capaces de reconstruir la posición de cada una de las piezas relacionándola con la posición lógica de las restantes. Es decir, mientras que en el primer caso trabajaba la memoria a corto plazo, en el segundo lo hacía la experiencia.

Un cuarto enfoque ha sido el de las motivaciones. Frecuentemente oímos o pronunciamos la expresión de que "cada persona es un mundo" para indicar que las motivaciones de cada decisor pueden hacer que las alternativas elegidas sean variadas y todas ellas asumibles desde el punto de vista de la propia racionalidad. Parece justificable, desde este punto de vista, que al preguntar sobre las motivaciones de los entrevistados se deje la posibilidad de que cada individuo exponga su propia motivación, aunque también cabe la postura contraria: es posible establecer un grupo de motivaciones y dejar una última modalidad bajo el nombre de "otras". El evidente peligro que entraña esta última alternativa es que el entrevistado pueda evaluar en ese momento algunas posibilidades de elección que nunca hubiera considerado por sí mismo, habiendo contaminado la muestra y reproduciendo entonces una situación difícilmente extrapolable al resto de la población.

Esta última decisión debiera llevarnos a considerar la necesidad de realizar una reflexión más profunda, y genérica a la vez, sobre la utilización de las preguntas con respuesta cerrada y con respuesta abierta. Un planteamiento más genérico en cuanto a si la utilización de unas puede ser preferible a la de otras en todos los casos, y un planteamiento más profundo en cuanto a si realmente miden lo mismo o, dicho de otra forma, si los resultados que alcanzamos con ambas son equivalentes o muestran diferencias.

Los estudios realizados sobre la comparación de los resultados obtenidos a partir de la utilización en cuestionarios de preguntas con respuesta abierta y con respuesta cerrada han mostrado que los resultados no son iguales (Shuman y Presser, 1981; Juan, 1986), lo que nos lleva inmediatamente a reflexionar sobre si la diferencia se debe al instrumento

de medida utilizado o a que realmente estamos midiendo fenómenos con perspectivas u objetivos distintos.

En un estudio sobre los candidatos presentados a las elecciones municipales de la provincia de León se les planteó la pregunta abierta *¿Puede indicarnos, en su opinión, los mayores problemas existentes en su municipio?*, así como la pregunta cerrada *¿Puede indicarnos si aumentaría o disminuiría el dinero utilizado en su municipio para...?*, exponiendo a continuación una serie de 21 partidas presupuestarias con cinco modalidades (Álvarez Esteban, 1996).

Los resultados de ambas preguntas fueron sensiblemente diferentes. Para la pregunta abierta los problemas obtenidos según el orden de importancia (palabras que más se repitieron, con la frecuencia entre paréntesis) fueron los siguientes: agua (56), calle (54), comunicaciones (40), infraestructuras (37), deporte (35), pavimentación (35), trabajo (32), carreteras (28), despoblación (27), población (26) y envejecimiento (16). En cambio, cuando la pregunta se formuló de manera cerrada se obtuvo como mayor problema la reforestación (el 85% de los candidatos aumentaría o aumentaría mucho el dinero gastado en reforestación), cultura (84%), instalaciones deportivas (80%), viviendas sociales/suelo (76%), jardines (75%), mejora de carreteras (75%), depuradoras de agua (74%), etc. Por otro lado, algunos factores que no se incluyeron en la pregunta cerrada aparecieron en la abierta: despoblación, envejecimiento de la población, trabajo, paro y polución.

De estas diferencias, aunque algunas puedan atribuirse al instrumento de medida, la mayor parte son debidas al hecho de que frecuentemente la decisión o motivación del encuestado se intenta medir con el prisma del investigador (expresado a través de las modalidades de la pregunta con respuesta cerrada).

A la hora de plantearnos la utilización de una pregunta con respuesta abierta o cerrada debiéramos reflexionar en primer lugar sobre si el enfoque de investigación que deseamos está basado en atributos (preguntas con respuestas cerradas) o bien sobre si dicho enfoque está basado en conceptos o relaciones conceptuales (preguntas con respuestas abiertas). Las preguntas abiertas y cerradas serían dos formas diferentes de obtención de información que pueden considerarse complementarias y en ocasiones alternativas, pero en ningún caso equivalentes.

Así, por ejemplo, el ya mencionado trabajo de Abascal y Franco (Abascal y Franco, 2002) se sitúa bajo la perspectiva de las motivaciones, por lo tanto, dentro del estudio de las relaciones conceptuales de los entrevistados, tratando de explicar el motivo por el que cada uno de los estudiantes ha elegido una determinada titulación.

Como resumen de lo dicho hasta el momento, podríamos indicar que las preguntas abiertas suelen utilizarse cuando se desea investigar antecedentes causales, condiciones para que ocurra un hecho, permitirlo o autorizarlo, consecuencias causales, verificación de ideas, elección entre diversas posibilidades, procedimientos asociados al pronombre interrogativo "cómo", qué se pretende con la elección, expectativas, juicios, peticiones o solicitudes, etc.

La siguiente figura intenta sintetizar el proceso de toma de decisiones y su relación con las preguntas con respuesta abierta y cerrada:

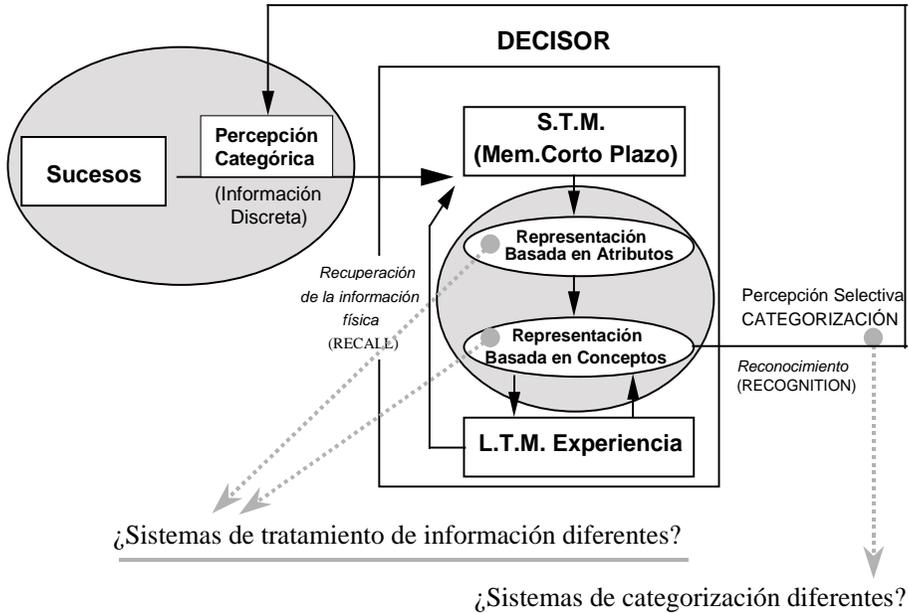


Figura 1. *Proceso de toma de decisiones y preguntas abiertas en cuestionarios.*

Si se desea realizar un cuestionario que pretenda recuperar información almacenada en la memoria a corto plazo la representación puede considerarse fundamentalmente basada en atributos y, por lo tanto, en términos generales, las preguntas cerradas pueden ser un instrumento útil (con las limitaciones que se han expuesto con anterioridad). Si una encuesta pretende recuperar la información almacenada en la memoria a largo plazo (experiencia), ésta se encuentra en su mayor parte estructurada en forma de conceptos o representaciones conceptuales y, ante una nueva información la comparación no se realizará en base a los atributos, sino en base a conceptos (un ejemplo puede ser el de las motivaciones). Cabe pues, considerar de mayor interés la utilización de una pregunta de respuesta abierta en este caso.

### **El análisis textual de respuestas abiertas**

El primer paso en el análisis textual de las respuestas abiertas en los cuestionarios es la determinación de la unidad de tratamiento de información estadística, frecuentemente la *forma gráfica*. Un juego tradicional como el de las "palabras prohibidas" nos puede servir como ejemplo para mostrar cómo cuando nos encontramos con una pregunta con dominio estrecho (una pregunta específica dentro de un contexto determinado) la utilización de ciertas palabras clave indica claramente el tema tratado, la respuesta que el encuestado ha elegido. Éste es un hecho que se puede constatar fácilmente con la lectura de las respuestas dadas en una pregunta abierta: la presencia o ausencia de una palabra o

forma gráfica indica si se considera o no una determinada alternativa de elección, una motivación, etc.

Este primer paso a la hora de analizar un texto supone la división del mismo en las unidades mínimas que deseamos estudiar (palabras, segmentos o frases, etc.), así como fijar los delimitadores que separan cada unidad mínima (retorno de carro, punto, coma, etc.). Este proceso se denomina *segmentación* del texto, y *ocurrencia* al conjunto de caracteres que se encuentran entre dos delimitadores.

A continuación (aunque esta fase puede ser previa a la segmentación) se debe proceder a una *normalización* del texto, igualando las graffías en mayúsculas o minúsculas, abreviaturas, diversas maneras de expresar los números, fechas, quitando las mayúsculas después de un punto, etc. Por ejemplo: SA, S.A., Sociedad Anónima, sociedad anónima; CC.AA., CCAA, Comunidades Autónomas; Estado, estado, Estado de las Autonomías, estado de las Autonomías.

Un problema que se plantea en el análisis de datos textuales es el de la *lematización*, con el fin de quitar la ambigüedad que frecuentemente existe en las distintas formas (por ejemplo palabras homógrafas). Es posible transformar las formas verbales al infinitivo, adjetivos al masculino singular, sustantivos al singular, obtener familias morfológicas, etc. La lematización manual en un cuestionario es poco factible si el número de respuestas es elevado y plantea, en términos de un análisis coste-beneficio, si los resultados obtenidos distan sustancialmente de los resultados sin lematizar. Generalmente se alcanzan diferencias que no parecen compensar en la mayor parte de los casos este esfuerzo (Lebart, 1994).

Por otra parte, el análisis de las formas verbales, tal y como se presentan en las respuestas, puede proporcionar resultados interesantes dependiendo de que los verbos se encuentren en pasado o en futuro (por ejemplo, en la política los partidos en el poder y en la oposición toman un discurso de presente, pasado o futuro según sus respectivos intereses; del mismo modo, el verbo "gustar" en relación con las motivaciones tiene un significado distinto cuando las respuestas hablan de "gustaba", "gusta" o "gustará"), pudiendo constatarse entonces que también las motivaciones en ocasiones cambian con el tiempo.

### **Análisis de correspondencias textual**

Una de las primeras elecciones que deben realizarse a la hora de efectuar un análisis de correspondencias textual aplicado al estudio de las preguntas con respuesta abierta en cuestionarios es el establecimiento de un umbral de frecuencia mínima que una forma debe tener para que pueda ser tratada, eliminando de esta manera las formas poco frecuentes (que corresponderían de alguna manera a las respuestas poco representativas). El problema es, en principio, similar al que nos encontramos en un análisis de correspondencias no textual cuando el número de respuestas que señalan una determinada modalidad es muy pequeño. El análisis, por la ponderación que se da a cada modalidad, puede incorporar en los primeros ejes modalidades que tengan una frecuencia baja, encontrándonos así con unos ejes que tan sólo explicarían lo que ocurre con un porcentaje muy pequeño de la encuesta realizada y, por lo tanto, con representatividad bastante dudosa.

La solución más habitual, en el caso de preguntas cerradas, suele ser la agregación entre sí de estas modalidades poco representativas.

La selección del umbral de frecuencia mínima en las preguntas con respuesta abierta no deja de ser una decisión subjetiva, siendo habitual analizar solamente las palabras que aparecen al menos cinco veces.

Aún así, el aspecto de una tabla de contingencia textual (tabla léxica agregada) dista mucho de lo que podemos considerar como habitual entre las tablas de contingencia no textuales. Al tener tantas filas como palabras o formas, y tantas columnas como modalidades de la pregunta cerrada, el número de filas es muy superior al de columnas. En estos casos, el resultado en términos de varianza total se muestra sensible a las dimensiones de la tabla.

Por otro lado, la presencia de valores nulos (ceros) en la tabla de contingencia (ausencia de una forma o palabra en una respuesta) suele ser muy alta. Por este motivo, mientras en un análisis de correspondencias no textual las distancias entre modalidades-fila o modalidades-columna se establecen en términos de diferencias de frecuencia, en un análisis de correspondencias con una tabla léxica las distancias suelen estudiarse en términos de presencia y ausencia.

En definitiva, tendremos una matriz esparsa (también denominada "mal condicionada") sobre la que aplicaríamos un análisis de correspondencias. Los valores propios obtenidos, interpretados en términos de la importancia de cada uno de ellos en la solución final, no suelen tener la misma significación que en un análisis no textual, ya que la inercia suele aparecer muy repartida entre todos los factores.

Una cuestión importante que creemos debiera ser considerada en todo análisis de correspondencias textual es la de las contribuciones en su doble vertiente: las contribuciones absolutas y las contribuciones relativas.

La representación gráfica de las palabras sobre los planos factoriales aparece en muchos casos como una verdadera nube, en la que las palabras se solapan entre sí e impiden la interpretación. Este problema se hace más grave a medida que tenemos un mayor número de palabras retenidas (mayor número de filas).

Por otro lado, la representación de las palabras sobre el plano factorial tan sólo nos da una idea intuitiva de la importancia que tiene cada una de ellas a la hora de explicar un factor, así como la importancia del factor a la hora de explicar cada variable.

La contribución absoluta permite cuantificar el porcentaje de la varianza del factor que es explicado por cada modalidad, encontrándonos en el análisis de respuestas abiertas con que al existir un gran número de palabras o de formas expresadas en filas, la importancia de cada factor suele aparecer muy repartida entre todas ellas. No solemos tener (a diferencia del análisis no textual) una modalidad o un número reducido de éstas que sean capaces de caracterizar totalmente un factor, sino que poseeremos un considerable número de palabras que conjuntamente explican el factor, a pesar de que individualmente cada una de ellas tenga contribuciones absolutas pequeñas.

La contribución relativa expresa el porcentaje de la varianza de la variable explicado por el factor. En el análisis de correspondencias textual esta información suele ser un instrumento de extrema utilidad para poder descubrir los conceptos que forman las palabras.

Como hemos indicado, el tipo de tabla de contingencia más habitual sobre la que se aplica el análisis de correspondencias es la denominada *tabla léxica agregada*, formada por tantas filas como palabras retenidas y tantas columnas como modalidades de la variable correspondiente a la pregunta cerrada o característica estudiada del individuo.

Otra posibilidad es el análisis de una pregunta abierta de forma independiente, construyendo para ello una tabla que contiene las palabras retenidas en filas y en columnas. El valor de una celda indica el número de veces que en una respuesta aparecen conjuntamente (a la vez) la palabra fila y la palabra columna.

También es posible construir *tablas yuxtapuestas*, para ello se unen horizontalmente tantas tablas como variables de las preguntas cerradas quieran analizarse.

El posicionamiento de variables cuantitativas o cualitativas en forma de elementos ilustrativos sobre los planos factoriales adquiere una importancia fundamental a la hora de la interpretación de todos estos resultados.

El análisis de correspondencias sobre la información textual de preguntas con respuesta abierta en cuestionarios no deja de aportar, a pesar de ser uno de los análisis con más tradición en la escuela francesa, nuevas visiones para intentar comprender la información que recogemos mediante cuestionarios. Así, por ejemplo, los problemas o características propias señaladas del análisis de correspondencias textual hacen que se planteen dudas sobre su estabilidad, y por lo tanto de la representatividad de las conclusiones que obtengamos (Lebart, 1996; Álvarez y otros, 2002). Cuando se considera que las filas (formas) actúan como variable dependiente mientras que las columnas (pregunta cerrada) lo hacen como variable independiente (o viceversa), pudiera ser más adecuada la utilización del denominado NSCA (análisis de correspondencias no simétrico). Por su parte, Mónica Bécue y Jérôme Pagès están proporcionando interesantes soluciones a la hora de comparar dos preguntas abiertas entre sí (Bécue y Pagès, 2002).

## **Análisis cluster sobre información textual**

El análisis cluster o de agrupaciones tiene en el análisis de datos textual una especial relevancia cuando utiliza como input de información los resultados de los factores obtenidos con un análisis de correspondencias previo. El objetivo puede ser la obtención de agrupaciones de modalidades de las preguntas cerradas, la obtención de grupos de palabras (buscando descubrir conceptos o palabras relacionadas para cada factor), e incluso el posicionamiento de las agrupaciones sobre los planos factoriales resultado del análisis de correspondencias.

Las técnicas utilizadas son las de agrupamiento jerárquico en el caso de que se obtengan los clusters de las modalidades de las preguntas cerradas, y el agrupamiento a partir de técnicas de clasificación directa o mixta cuando lo que se desea agrupar son las formas o palabras, ya que el gran número de éstas no permite utilizar técnicas jerárquicas. El hecho de aplicar el análisis cluster con la misma métrica chi-dos del análisis de correspondencias sobre los factores que hemos obtenido de éste, tiene un doble objetivo: eliminar la influencia de los ejes que contienen la parte aleatoria (los que son menos importantes) y el de obtener representaciones conceptuales a partir de palabras o formas que no son fácilmente observables en un plano factorial con una gran nube de palabras. De esta forma, es posible encontrar en una agrupación términos como "espacios, verdes,

jardines" o "basura, recogida, vertederos", que realmente hacen referencia al mismo concepto.

## Otros análisis estadísticos aplicados a textos

En el campo lingüístico es frecuente analizar los discursos de políticos que han sido pronunciados en diferentes momentos de tiempo, siendo la evolución la característica que se desea estudiar. De la misma forma, es factible realizar comparaciones sobre tablas de contingencia en diferentes periodos de tiempo, en particular mediante el análisis de correspondencias de más de dos vías. Un caso particular es el análisis de series textuales cronológicas.

El análisis textual discriminante se enfrenta con el problema de la existencia de un número de variables generalmente mucho mayor que el de individuos o, al menos, en una proporción nada habitual. Hasta el momento, la mayor parte de los estudios realizados en este campo se han llevado a cabo con respecto a la clasificación automática de documentos.

Hay que señalar que también ha habido algunos estudios aplicando el escalamiento multidimensional.

## Conclusiones

Las preguntas con respuesta abierta en los cuestionarios han demostrado proporcionar una información extremadamente rica que puede observarse fácilmente a partir de la lectura directa de las mismas.

A la hora de diseñar un cuestionario no debiéramos plantearnos la elección entre la formulación de una pregunta dejando la respuesta libre o con modalidades cerradas como una alternativa excluyente. Debemos ser conscientes de que una y otra proporcionan una información que, aunque pueda considerarse en principio alternativa, fundamentalmente es de carácter complementario. Los distintos estudios realizados en este sentido han demostrado claramente que estas preguntas no pueden ser consideradas como equivalentes.

La utilización de los ordenadores y programas estadísticos facilita la síntesis de la información textual para el tratamiento de encuestas realizadas sobre un gran número de individuos, el establecimiento de tipologías, etc., más allá del mero recuento de las palabras. No hay que olvidar que más del 90% de la información que se puede obtener en INTERNET es de tipo textual.

En definitiva, quedan abiertos muchos campos dentro del análisis estadístico de datos textuales que comprenderían, a título de ejemplo, modelos probabilísticos, caracterización de individuos/modalidades "fuertes", estudios de estabilidad, generalización al análisis de más de una pregunta abierta de forma conjunta, algunos problemas derivados del número elevado de datos o la especificidad de la composición de las matrices esparsas y algoritmos propios.

## Referencias

- Abascal, E. y Franco, M.A. (2003) Análisis textual de encuestas: aplicación al estudio de las motivaciones de los estudiantes en la elección de su titulación. *Metodología de Encuestas*, Vol 4 (2) 195-209.
- Álvarez Esteban, R. (1996) *Procesamiento humano de información y análisis estadístico de datos textuales para el tratamiento de preguntas abiertas en cuestionarios*. Tesis doctoral, Universidad de León.
- Álvarez Esteban, R.; Bécue, M; Lanero J.J. y Valencia, O. (2002) Results stability in textual analysis: its applications to the study of the Spanish investiture speeches (1979-2000). *JADT 2002 6<sup>th</sup> International Conference on Textual Data Statistical Analysis*, 1-12.
- Bécue, M. (1991) *Análisis de datos textuales. Métodos estadísticos y algoritmos*. París: CISIA.
- Bécue, M. y Pagès, J. (2002) Analyse conjointe de questions ouvertes et de questions fermées: méthodologie, exemple. *JADT 2002 6<sup>th</sup> International Conference on Textual Data Statistical Analysis*, 113-123.
- Dubois, D. (Ed). (1993) *Sémantique et cognition. Catégories, prototypes, typicalité*. Paris: CNRS EDITIONS.
- Juan, S. (1986) L'ouvert et le fermé dans la pratique du questionnaire. Analyse comparative et spécificités de l'enquête par correspondance. *Revue Française de Sociologie*. XXVII, 301-306.
- Lafon, P. (1984) *Dépouillements et statistiques en lexicométrie*. Gèneve: Slatkine-Champion.
- Lebart, L.; Salem, A. y Berry, L. (1991) Recents developments in the statistical processing of textual data. *Applied Stochastic Models and Data Analysis*, Vol 7, 47- 62.
- Lebart, L. y Salem, A. (1994) *Statistique textuelle*. Paris: Dunod.
- Lebart, L. (1996) Assessing sample variability and stability in the visualization techniques related to principal component analysis: Bootstrap and alternative simulation methods. *XII Symposium on Computational Statistics COMPSTAT*. Physica Verlag, Heidelberg, 205-210.
- Libby, R. (1985) *Accounting and Human Information Processing: Theory and applications*. Prentice Hall.
- Mechelen, I. y otros Eds. (1993) *Categories and concepts: theoretical views and inductive data analysis*. London: Academic Press.
- Shuman, H. et Presser, S. (1981) *Questions and answers in attitude surveys. Experiments on question form, wording and context*. New York: Academic Press.