

ANÁLISIS TEXTUAL DE ENCUESTAS: APLICACIÓN AL ESTUDIO DE LAS MOTIVACIONES DE LOS ESTUDIANTES EN LA ELECCIÓN DE SU TITULACIÓN

Elena Abascal Fernández
María de los Ángeles Franco Manero
Universidad Pública de Navarra

RESUMEN

El objetivo principal de este trabajo es mostrar cómo los métodos de estadística textual permiten analizar de forma sistemática una encuesta realizada a 898 estudiantes, en la que se incluye una pregunta de respuesta libre sobre las motivaciones que les han conducido a la elección de la titulación que están estudiando, así como preguntas cerradas que recogen las características del estudiante.

La recogida de información para una investigación de estas características, en la que se pueden encontrar gran variedad de motivaciones y, además, múltiples, con gran riqueza de vocabulario, sólo puede obtenerse a partir de una encuesta con una pregunta de respuesta libre. De esta forma, no se limita al estudiante a elegir entre unas pocas opciones excluyentes, sino que se le deja libertad para expresar todas sus motivaciones con sus propios matices.

La Estadística Textual proporciona métodos de análisis que permiten realizar un estudio de la encuesta, desde una perspectiva multivariante. Además, considera las respuestas completas, teniendo en cuenta las palabras y sus contextos.

El estudio realizado analiza las motivaciones de los estudiantes en la elección de la titulación, y sirve para ilustrar las posibilidades que las técnicas estadísticas del Análisis Textual, poseen en la investigación a través de encuestas con preguntas de respuesta libre en muestras de gran tamaño.

Palabras clave: *Análisis de Correspondencias, Análisis de Clasificación, Estadística Textual, Análisis de Especificidades.*

Introducción

Las preguntas de respuesta libre son muy ricas en información y sin embargo, no son muy frecuentes en las encuestas, debido a que las respuestas obtenidas presentan gran dificultad para su tratamiento estadístico posterior. Habitualmente se utilizan en encuestas pequeñas o en la fase preparatoria, principalmente con el objetivo de obtener información que permita formular adecuadamente preguntas cerradas.

El desarrollo actual de la estadística textual permite el tratamiento sistemático de la información obtenida en una encuesta en la que se incluyen preguntas de respuesta libre, incluso en muestras de gran tamaño. La estadística textual considera las respuestas libres en su forma original, es decir, las respuestas completas emitidas por los interrogados a la cuestión abierta. De esta manera tiene en cuenta las palabras y sus contextos. Además utiliza la información sobre las características de los encuestados, recogidas en preguntas cerradas.

La estadística textual proporciona métodos de análisis que permiten realizar un estudio de la encuesta desde una perspectiva multivariante. No parte de una reducción de la información a priori sino que utiliza toda la información disponible sobre el encuestado. Utilizando estas técnicas de estudio se evitan las deformaciones, sesgos y pérdidas de información que se pueden ocasionar a través de la post-codificación y permiten además analizar un elevado número de respuestas.

Por otra parte, la universidad está sometida actualmente a un proceso de creación y modificación de las titulaciones implantadas. En este contexto, interesa conocer los motivos por los cuales un estudiante elige la carrera en la que se encuentra y averiguar si las causas de éxito o fracaso pueden relacionarse con los motivos en los que se basó la elección.

La recogida de información para un estudio de estas características, en el que se puede encontrar gran variedad de motivaciones y además múltiples, con gran riqueza de vocabulario, sólo puede obtenerse a partir de una encuesta con una pregunta de respuesta libre, en la que no se limite al estudiante a elegir entre unas pocas opciones excluyentes, sino que le deje libertad para expresar todas sus motivaciones con sus propios matices.

El objetivo principal de este trabajo es mostrar cómo los métodos de estadística textual permiten analizar de forma sistemática una encuesta realizada a 898 estudiantes, en la que se incluye una pregunta de respuesta libre sobre las motivaciones que les han conducido a la elección de la titulación que están estudiando, así como preguntas cerradas que recogen las características del estudiante.

Los objetivos específicos de este trabajo son:

- Estudiar las distintas motivaciones que llevaron a los estudiantes a elegir la titulación que están cursando en la actualidad.
- Analizar si existen diferencias según las distintas titulaciones, o bien, si éstas tienen su origen en otras características como sexo, nivel de notas obtenidas en educación secundaria, selectividad, etc.
- Mostrar la adecuación de las técnicas de estadística textual para el análisis de preguntas abiertas en muestras de gran tamaño.

El artículo se organiza del siguiente modo: en primer lugar, se presentan las características de la metodología del análisis textual de encuestas. Seguidamente, se entra en el proceso de análisis y resultados obtenidos en el estudio de las motivaciones de los estudiantes en la elección de su titulación. Por último, se recogen algunas conclusiones relativas a la metodología y a las motivaciones.

El análisis textual de preguntas libres

Las preguntas de respuesta libre —llamadas también “cuestiones abiertas”—, constituyen una herramienta imprescindible en todas aquellas encuestas que tratan de estudiar y profundizar en un tema complejo o poco conocido. En estos casos, se hace necesario trabajar con datos de naturaleza textual, es decir, datos expresados en forma de cadenas verbales de extensión variable.

La estadística textual se apoya en técnicas estadísticas desarrolladas por la escuela francesa de análisis de datos. La base de este método consiste, a *grosso modo*, en utilizar procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y realizar algún tipo de análisis estadístico a partir de los resultados de dichos recuentos.

En su origen, el análisis o estadística textual, se enfocó hacia el estudio de textos (estadística lingüística). Posteriormente su campo de aplicación se ha ampliado y actualmente se analizan datos textuales procedentes de cuestiones abiertas incluidas en encuestas, datos textuales de discusiones de grupo, entrevistas o documentos diversos.

A continuación se describen los elementos fundamentales del análisis textual

Las unidades de la estadística textual

La *forma gráfica* es una unidad que se define como una secuencia de caracteres no delimitadores (en general letras) comprendida entre dos caracteres delimitadores (espacios en blanco o signos de puntuación) (Lebart y Salem 1988). Ejemplos de formas gráficas en la encuesta realizada son “gusta”, “salidas” y “muchas”. El conjunto de formas de un texto constituye su *vocabulario*.

El *perfil léxico* es un vector cuyos componentes son las frecuencias de cada una de las formas utilizadas por un individuo o un grupo de individuos.

Una vez conocido el vocabulario de texto, interesa conocer los perfiles de las frecuencias de las formas gráficas y especialmente las diferencias entre ellos. Veremos que los estudiantes de Economía y de Ingeniería Industrial tienen perfiles más próximos que éstos últimos y los de Magisterio.

Los *segmentos repetidos* se definen como una secuencia de dos o más palabras, no separadas por un delimitador de secuencia, que aparecen más de una vez en un corpus de datos textuales. Ejemplos de segmentos repetidos en la encuesta realizada son “salidas profesionales” y “tiene muchas salidas”. Son unidades más largas, compuestas por varias formas, dotadas de un sentido propio y que no siempre se deduce de las formas que entran en su composición, si éstas se consideran aisladamente.

Todo el tratamiento estadístico aplicable a las formas gráficas, puede generalizarse a los segmentos. Los resultados obtenidos son enriquecedores por la introducción de contexto en las formas, que hace desaparecer gran parte de las ambigüedades.

La *lematización* consiste en declarar equivalentes las formas gráficas que tienen la misma raíz y un significado equivalente. Una vez creado el vocabulario y antes de proceder a su tratamiento, existe una etapa opcional en la que éste se depura, lematizando y eliminando las palabras utensilio o herramienta como artículos, conjunciones, preposiciones, etc.

Se ha comprobado que esta etapa no es indispensable y puede presentar problemas (Lebart 1986). Por ejemplo, una palabra puede estar ligada a un contexto y a un contenido particular, y ciertas preposiciones, artículos, etc. (denominados palabras herramientas), pueden caracterizar de manera efectiva las actitudes u opiniones. En este estudio encontramos que el empleo del verbo gustar en formas del pasado se asocia con las asinaturas, mientras que la utilización en formas del presente se utiliza más para referirse a la profesión. Por ello, es conveniente analizar bien los contextos de utilización de las palabras, antes de proceder a la lematización.

Para el tratamiento estadístico se asigna a cada forma gráfica diferente un número de orden, el mismo siempre que aparece. La numeración da lugar a la creación de un diccionario de formas propio de cada análisis. En el caso en que se procede a lematizar, las palabras que se declaran equivalentes les corresponde el mismo número.

Las tablas léxicas

Las respuestas libres pueden ser numeradas de manera completamente automática por el analista y, a partir del diccionario de formas, construir tablas rectangulares que serán analizadas mediante las técnicas multivariantes descriptivas, como el Análisis Factorial de Correspondencias y el Análisis de Clasificación (Benzécri y otros, 1973, 1981; Abascal y Grande, 1989; Escofier y Pagès, 1992; Lebart, Morineau y Piron, 1995).

Estas tablas recogen la distribución de palabras entre las distintas respuestas individuales o entre grupos de individuos. Los grupos se definen de acuerdo con alguna característica de los entrevistados, recogida en sus respuestas a una pregunta cerrada. Las tablas utilizadas son:

- La *tabla léxica T* (individuos x palabras). Es una matriz de k filas, una por cada respuesta, y un número de columnas igual al de formas gráficas utilizadas por el conjunto de individuos. En la intersección de la fila i y la columna j , figura la frecuencia con que aparece la forma j en la respuesta del individuo i . Esta tabla recoge los perfiles léxicos de los individuos.
- La *tabla léxica agregada* (palabras x grupos de sujetos). Es la tabla de contingencia cuyo término general recoge el número de veces que la forma i ha sido utilizada en la respuesta libre, por el conjunto de individuos que eligen la modalidad j en la respuesta cerrada. A partir de esta tabla se pueden comparar los perfiles léxicos de los segmentos de población definidos.

Se puede obtener una tabla léxica agregada para cada pregunta cerrada, y comparar los perfiles léxicos de las diferentes categorías de la población.

Técnicas descriptivas de análisis de datos multivariantes.

El análisis de datos textuales se apoya, como ya hemos dicho anteriormente, en los métodos de análisis estadístico de grandes tablas de datos, concretamente el análisis de correspondencias y el análisis de clasificación.

El análisis de correspondencias de una tabla proporciona una visión gráfica, simplificada, de la información que contiene y pone en evidencia las diferencias entre los distintos perfiles léxicos. Las representaciones permiten situar a los individuos en un espacio determinado por las formas, permitiendo comparar los perfiles de sus respuestas, localizando aquellos más semejantes y las formas causantes de estas semejanzas. Del mismo modo, se detectan los sujetos más diferentes explicando sus causas. La representación de las formas permite estudiar las proximidades entre aquellas formas que son utilizadas simultáneamente por los mismos sujetos, es decir, estudiar los contextos, ya que la proximidad entre dos formas es mayor cuando aparecen en la misma respuesta.

Si se analiza la tabla léxica agregada se pueden estudiar las posiciones relativas de diferentes grupos de individuos en relación al vocabulario empleado y caracterizarlos por las palabras utilizadas.

Con este tipo de análisis, la presencia de palabras herramienta (preposiciones, artículos, conjunciones, pronombres...), está en principio perfectamente justificada si estas palabras caracterizan a algunas categorías, o su posición próxima puede ser interesante para interpretarlas. Si por el contrario, su reparto es aleatorio y se aproximan o se sitúan en la parte central del gráfico, interesa eliminarlas. Por otra parte, la presencia próxima de varias formas de un mismo verbo, puede constituir una herramienta de validación y justificar la lematización.

Un posterior análisis de clasificación enriquece los resultados (Lebart y Salem 1988, Bécue 1991). La clasificación proporciona grupos de individuos homogéneos en cuanto a las palabras utilizadas en sus respuestas o bien, grupos de palabras frecuentemente utilizadas por los mismos sujetos.

El análisis de datos textuales se complementa con otros métodos léxicométricos y técnicas clásicas como el estudio del vocabulario, las concordancias, las formas características, etc. (Lebart y Salem 1988; Bécue 1991; Etxeberria, García y Rodríguez 1995).

El sistema informático SPAD.T está especialmente diseñado para el análisis de datos textuales. Es compatible con SPAD constituyendo en conjunto un instrumento muy útil para el análisis global de una encuesta.

Las motivaciones de los estudiantes en la elección de la titulación

El objetivo de este trabajo es estudiar las distintas motivaciones que llevaron a los estudiantes a elegir la titulación que están cursando en la actualidad y analizar si existen diferencias significativas entre las distintas titulaciones, o bien si éstas tienen su origen en otras características, como sexo, nivel de notas obtenido en educación secundaria, selectividad, etc. Además, se busca obtener un perfil de las respuestas de cada grupo y seleccionar de entre ellas, las más representativas y las más frecuentes.

El estudio parte de la información obtenida en una encuesta realizada en la Universidad Pública de Navarra a 898 estudiantes pertenecientes a seis titulaciones diferentes: Economía (249), Ingeniería Industrial (206), Enfermería (102), Magisterio (152), Relaciones Laborales (104) y Derecho (85).

En este trabajo se analizan las respuestas a la siguiente pregunta “¿*Cuáles son los motivos principales por los que has elegido esta titulación?*”.

Además, se utilizan otras cuestiones o *preguntas cerradas* que permiten caracterizar al estudiante. Son preguntas relativas al sexo, titulación, notas de selectividad, notas de secundaria, tipo de centro de secundaria, opción en que eligió la carrera, curso actual, tipo de acceso, situación laboral, tipo de domicilio (familiar, residencia,...), asistencia a clase, opción en que eligió la carrera, asistencia a tutorías, cursos y asignaturas repetidas, años en la titulación y utilización de la biblioteca.

Análisis del vocabulario

El estudio comienza con la construcción de las formas gráficas y segmentos repetidos contenidos en las respuestas de los estudiantes. El número de formas que aparecen en el recuento inicial es de 12750, siendo 1507 de ellas diferentes, lo que nos da una idea de la gran riqueza del vocabulario utilizado por los estudiantes para explicar los motivos por los que eligieron una determinada titulación, así como de la amplia gama de motivaciones que encuentran. La tabla 1 presenta una lista depurada, sin las llamadas *palabras herramienta* de las formas gráficas utilizadas.

Formas empleadas	Frecuencia	Formas empleadas	Frecuencia	Formas empleadas	Frecuencia
carrera	232	profesionales	31	gustaban	23
gusta	220	derecho	31	tener	22
salidas	118	interés	30	vida	22
gustaba	78	trabajar	30	buenas	21
futuro	65	ser	29	elegido	21
trabajo	61	vocación	29	educación	20
interesante	51	profesional	29	salida	20
niños	47	elegí	27	profesión	20
asignaturas	46	estudiar	27	parecía	19
gustan	44	estudios	26	relacionado	19
economía	43	ciencias	26	también	19
parece	42	conocimientos	26	ayudar	18
gustado	42	ofrece	25	opción	18
mundo	42	gustaría	25	atraía	17
carreras	40	posibilidades	25	elegir	17
titulación	39	hacer	25	podía	17
quería	38	laborales	25	música	17
tenía	34	poder	24	puede	16
laboral	34	universidad	24	personas	16

Tabla 1. *palabras no-herramienta más frecuentes en las respuestas.*

Aparecen con gran frecuencia las formas *salidas*, *futuro* y *trabajo*, que pueden hacer referencia a que el estudiante piensa en las futuras salidas o trabajos a los que pueda acceder a través de la titulación elegida. También hay formas que hacen referencia a los gustos o preferencias del estudiante, como factores importantes en la elección (formas del verbo gustar, *interesante*). Otras formas parecen a priori relacionadas con algunas titulaciones: *niños* (Magisterio), *economía* (Economía).

El examen de las formas y sus frecuencias se queda muy limitado, pues su significado depende mucho de la frase o contexto en que se encuentra ubicada y esto impide hacer afirmaciones sobre relaciones que se puedan establecer.

En la tabla 2 se recoge la lista de 40 segmentos repetidos (ordenados según sus frecuencias), obtenidos tras eliminar del vocabulario inicial las formas que aparecen menos de 5 veces, las de longitud menor de 3 y algunas palabras herramienta de longitud mayor.

Texto del segmento	Frecuencia	Texto del segmento	Frecuencia
salidas profesionales	26	tiene más	7
muchas salidas	21	quería hacer	6
más salidas	17	única carrera	6
desde siempre	15	gustaba desde	6
salidas laborales	14	bastantes salidas	6
sobre todo	13	relaciones laborales	6
tiene muchas salidas	11	futuro profesional	6
mercado laboral	11	gustado desde	6
parece interesante	11	han gustado	6
gusta mucho	11	principal motivo	6
salida profesional	9	tienen más salidas	6
buenas salidas	9	tiene salidas	6
buenas perspectivas	8	querido estudiar	5
gustaría trabajar	8	motivos principales	5
carrera interesante	8	otras titulaciones	5
mundo laboral	7	tiene buenas salidas	5
más adecuada	7	quería estudiar	5
parecía interesante	7	muchas salidas profesionales	5
tiene buenas	7	poder ayudar	5
otras carreras	7	algo relacionado	5

Tabla 2: lista de segmentos repetidos por orden de frecuencia

Los temas que predominan son los referentes al futuro laboral, los gustos y deseos. Se aprecia también que aparecen diversas formas de un mismo verbo, lo cual nos puede inducir a lematizar y convertirlos en una única forma. Sin embargo, existe la posibilidad de que no se incluyan en los mismos contextos y tengan diferentes significados, por lo cual se realiza un análisis previo a esta lematización.

El objetivo del estudio va más allá de una mera descripción de los distintos motivos que tienen los estudiantes para elegir titulación. Se pretende detectar también los moti-

vos principales de los diferentes grupos de estudiantes, de ahí que sea necesario realizar diversos análisis de las tablas léxicas y léxicas agregadas que permitan caracterizar los distintos grupos.

Análisis de la tabla directa

En este apartado se analiza la tabla léxica directa que cruza individuos con formas gráficas. Contiene 898 filas de individuos y 143 columnas, de las cuales 120 se corresponden con las formas gráficas de mayor frecuencia y las otras 23 columnas con las variables cualitativas que caracterizan a los estudiantes. Estas últimas se analizan como ilustrativas.

El objetivo del análisis es doble. En primer lugar, obtener una tipología de los individuos que utilizan formas iguales o parecidas en sus respuestas, caracterizándolos por las formas utilizadas y por las categorías de las variables cualitativas, es decir, por las características de los estudiantes que forman el grupo. El segundo objetivo es representar las proximidades entre aquellas formas que son utilizadas simultáneamente por los mismos estudiantes, es decir, estudiar los contextos, ya que la proximidad entre dos formas es mayor cuando aparecen en la misma respuesta.

La metodología del análisis consiste en realizar un Análisis Factorial de Correspondencias de la tabla directa. A través de este análisis se obtiene una visión gráfica de las asociaciones entre las formas y los grupos definidos por categorías de respuestas.

El análisis de correspondencias de la tabla léxica directa proporciona 119 ejes. Sobre el gráfico 1, plano formado por los dos primeros ejes, se pueden extraer algunas informaciones interesantes.

Las formas del pasado *gustaba* y *gustaban* (en el segundo cuadrante) se asocian con formas como *ciencias*, *letras*, *asignaturas*, *estudios* y *carreras*. Significa que son utilizadas por los estudiantes que eligen su titulación en función del gusto por el tipo de asignaturas (de ciencias o de letras) presentes en la carrera. Sin embargo, se aprecia que la forma *mucho* así como las formas *gusta*, *gustan* y *gustaría* del verbo gustar están asociadas a *trabajar*, *profesión*, *mundo*, *vida...*(cuarto cuadrante). Son formas utilizadas por aquellos que eligen la titulación pensando en sus preferencias profesionales.

Sobre el tercer cuadrante se detecta una asociación entre *salidas*, *muchas*, *buenas*, *laborales*, *profesionales*, *perspectivas* y *tiene*. Esto se puede interpretar en el sentido de que cuando utilizan las formas *muchas* o *buenas*, generalmente se están refiriendo a un futuro profesional o laboral.

Por último, las formas *gustado* o *querido* se asocian con la forma *vocación* situada también en el primer cuadrante, pero muy alejada del centro por lo que no aparece en el gráfico.

En el gráfico 1 se representa también la proyección, sobre el plano factorial, de las variables cualitativas que han resultado más significativas, la titulación y el sexo.

El análisis de clasificación de los individuos, es decir, de los perfiles léxicos de sus respuestas, muestra nuevamente que las variables más relacionadas con la clasificación, las que obtienen mayores diferencias en las respuestas de los estudiantes, son la titulación a la que pertenecen y el sexo. Estas diferencias son menos claras en el caso del nivel de notas previas a los estudios universitarios, es decir, de secundaria y selectividad. En

consecuencia, titulación y sexo son las variables que pueden resultar interesantes al definir las tablas léxicas agregadas que serán analizadas más adelante

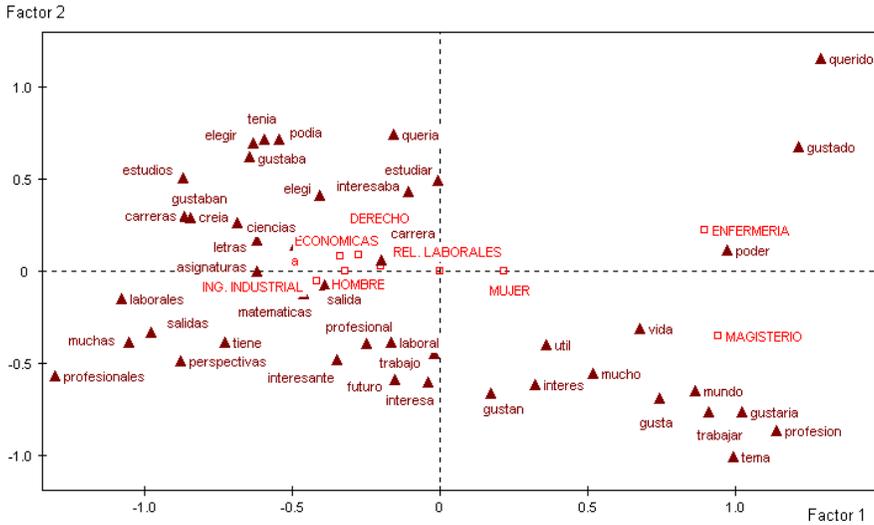


Gráfico 1: *representación sobre plano (1,2) de las formas y de las categorías más significativas de las variables cualitativas .*

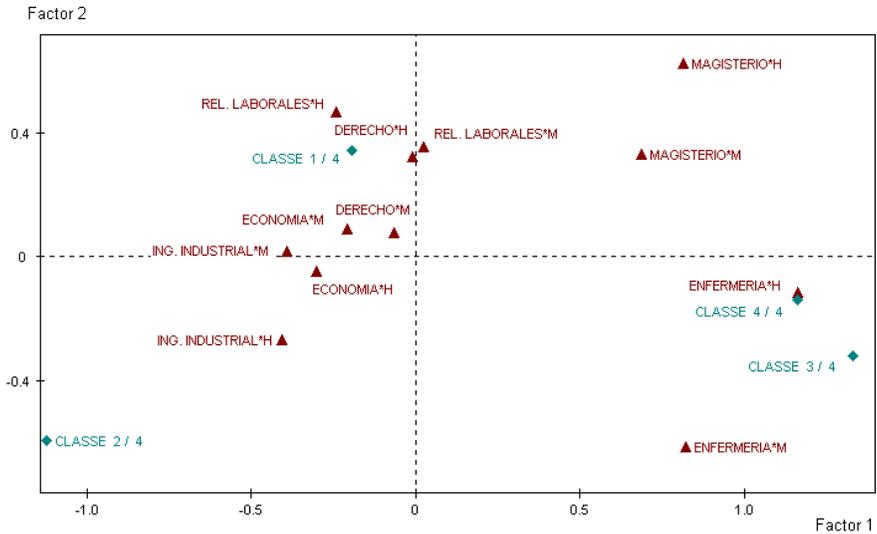


Gráfico 2: *representación, sobre el plano (1, 2), de las categorías de la variable titulación*sexo y de las clases.*

Análisis de la tabla léxica agregada

El siguiente objetivo es estudiar cómo utilizan las formas los estudiantes de las diferentes titulaciones y sexos. Para ello se construye la tabla léxica agregada de formas gráficas (120 filas) y categorías de la variable cruzada titulación*sexo (12 columnas). Se adjuntan como ilustrativos los segmentos repetidos. Se realiza, de nuevo, un análisis de correspondencias completado con el análisis de clasificación. Esta técnica permite obtener una representación gráfica de las asociaciones entre las formas (o segmentos) y los grupos o categorías de estudiantes definidas, en este caso por la titulación y sexo. La proximidad entre palabras y categorías ayuda a comprender las respuestas de cada uno de los grupos.

Sobre el plano que determinan los dos primeros factores se representan las categorías y las clases obtenidas (gráfico 2) y las formas (gráfico 3). Se han separado en dos gráficos para obtener mayor claridad evitando la duplicidad de puntos, pero ambos gráficos se pueden superponer (utilizando la misma escala) e interpretar las asociaciones entre palabras y categorías, lo que permite comprender las respuestas de cada grupo. En los gráficos se observan muy claramente las diferencias entre las titulaciones. Por otro lado, se puede afirmar que, en general, no se aprecian grandes diferencias por sexo sino que éstas radican más en la titulación, ya que los puntos relativos a los sexos de una misma titulación están próximos en el plano.

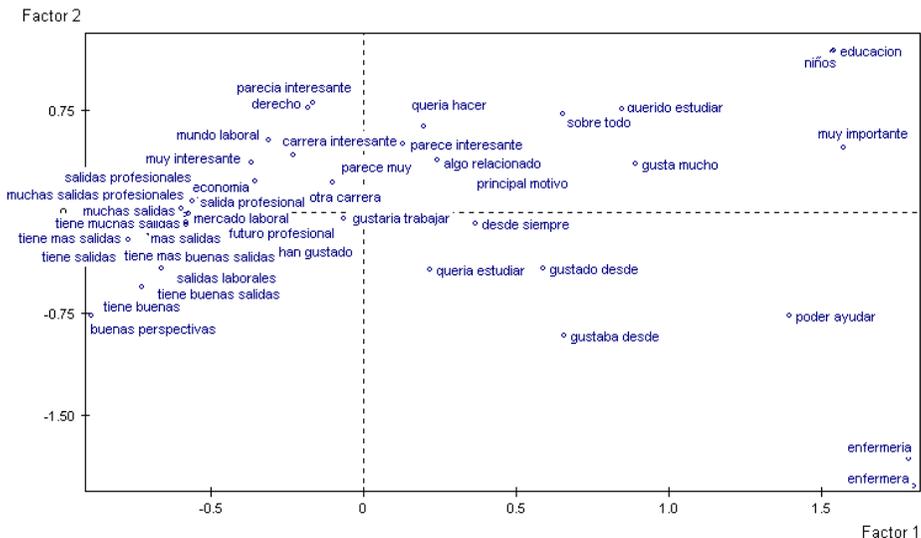


Gráfico 3: *representación de las palabras más significativas sobre el plano (1, 2)*

Los estudiantes de Magisterio y Enfermería comparten las formas situadas a la derecha del plano. En general, se asocian con *gustaría, profesión, vocación, siempre* y por no utilizar, o hacerlo en menor medida, las formas relativas a los temas laborales o salidas

profesionales y aquellas que puedan estar relacionadas con las asignaturas: *física, matemáticas o letras*. Todo lo contrario ocurre en las titulaciones de Ingeniería y Economía, que utilizan con mayor frecuencia las formas relativas al futuro profesional y el gusto o disgusto por algunas asignaturas.

Las diferencias entre Magisterio y Enfermería se observan sobre el segundo eje. Los estudiantes de Magisterio utilizan más las formas *niños, enseñanza y música*; mientras que los de Enfermería están más asociados a *ayudar, personas, demás*, etc. El vocabulario que hace la distinción entre estas dos titulaciones es el específico de cada carrera o bien, el que se asocia con su papel en la sociedad.

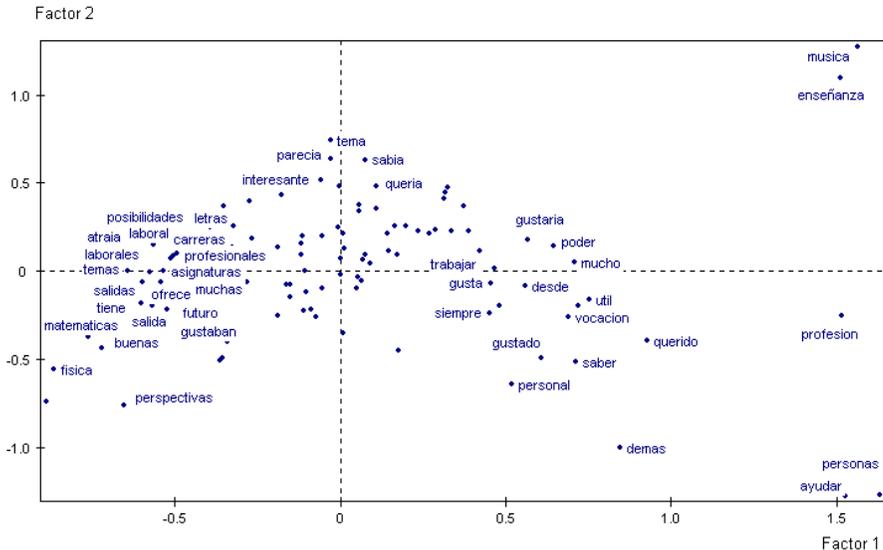


Gráfico 4: representación de los segmentos repetidos más significativos sobre el plano (1, 2)

Sobre el mismo plano factorial se han representado algunos de los segmentos repetidos (gráfico 4). En él se observan mejor las asociaciones entre las formas. Por ejemplo, respecto a Enfermería, se aprecia que *poder* se utiliza junto con *ayudar*. En Magisterio, la asociación es entre *querido* y *estudiar, sobre y todo, gusta y mucho*. Mientras que en el lado negativo del primer eje se sitúan los segmentos repetidos *salidas profesionales, salidas laborales, tiene más salidas, buenas perspectivas*, etc.; asociados principalmente a Ingeniería y Economía.

La clasificación de las formas que componen la tabla léxica agregada según la variable titulación*sexo, realizada sobre los ejes factoriales, complementa la interpretación de los planos factoriales. Se obtienen las cuatro clases siguientes:

La clase 1 está compuesta por formas asociadas principalmente a los estudiantes de Economía y Relaciones Laborales. Entre los elementos más característicos destacamos: *gustaba, mejor, carrera, profesionales, mundo, podía*. Se localiza en el segundo cuadrante del gráfico 2.

La clase 2 la definen formas con una presencia muy superior a la media entre los estudiantes de Ingeniería Industrial, tanto hombres como mujeres, así como los hombres de Economía. Las más características de esta clase son: *futuro, tiene, salidas, muchas, ciencias, gustaban, salida, interés, asignaturas*. Está situada en el tercer cuadrante del gráfico 2.

La clase 3 está compuesta por formas que emplean principalmente los estudiantes de Enfermería. Estas son: *vida, vocación, algo, principal, motivo, interesaba, cogieron, personas, ayudar*. Se localiza en el cuarto cuadrante del gráfico 2.

La clase 4 viene definida en gran medida por las formas habituales en el vocabulario de los estudiantes de Magisterio y las mujeres de Enfermería. Las formas más características son: *gusta, siempre, ser, desde, mucho, gustado, había, importante, elegido, gustaría*. Se sitúa próxima a la clase 3, en el cuarto cuadrante del gráfico 2.

Las respuestas modales o características

Los análisis multivariantes se pueden completar con otros procedimientos más clásicos de contrastes de hipótesis, en los que se compara la frecuencia con que aparece una forma en las respuestas de un grupo y en el conjunto de la encuesta, introduciendo dos nuevos elementos, las unidades características o especificidades y las respuestas características.

Las unidades características o especificidades son elementos (formas, segmentos) que destacan por su alta presencia o por su rareza, en las respuestas de un grupo de individuos.

Tabla 3: *especificidades de los estudiantes de Economía.*

Etiqueta de la forma gráfica	Porcentaje		Frecuencia		V. test	P.
	Interno	Global	Interna	Global		
1 economía	2,78	0,85	43	43	9,819	,000
2 carreras	1,75	0,79	27	40	4,674	,000
3 empresariales	0,39	0,12	6	6	3,154	,001
4 salidas	3,30	2,33	51	118	2,850	,002
5 económicos	0,32	0,10	5	5	2,789	,003
6 muchas salidas profesionales	0,32	0,10	5	5	2,785	,003
7 momento	0,39	0,14	6	7	2,638	,004
8 tener	0,84	0,43	13	22	2,572	,005
9 tiene	2,33	1,66	36	84	2,298	,011
6 ayudar	0,00	0,36	0	18	-2,990	,001
4 educación	0,00	0,39	0	20	-3,207	,001
3 profesión	0,00	0,39	0	20	-2,207	,001
2 gusta	2,71	4,34	42	220	-3,841	,000
1 niños	0,00	0,93	0	47	-5,401	,000

En la tabla 3 se recoge una muestra de las especificidades del grupo de estudiantes de Economía. Para cada forma que aparece en dicha tabla se especifica su frecuencia y

porcentaje en el grupo y en el conjunto total. Esta información se complementa con un valor de estadístico de contraste de igualdad de proporciones, en la titulación y en la muestra, normalizado (v.test).

Tabla 4: selección de las tres respuestas más características para ocho categorías de estudiantes (sin signos de puntuación).

Grupo	Distancia	Respuesta característica
<i>Economía Hombre</i>	0.188	me gusta y las salidas que tiene
	0.209	lo que más me gustaba era historia pero no tenía dinero para ir a la privada y no tiene salida elegí economía porque dentro de las carreras de la UPNA era de las que más me gustaba y puede tener salidas laborales
	0.234	parece que tiene salidas es de las que más me gustaban
<i>Economía Mujer</i>	0.139	porque es la que más me gusta
	0.142	porque es la carrera con más salidas y porque me gustan las salidas que tiene
	0.193	porque creo que tiene más salidas que otras y porque me gusta
<i>Ingeniero industrial Hombre</i>	0.166	es la que más me gusta
	0.240	por las salidas que ofrece y que es la que más interesante parecía
	0.244	porque me gusta
<i>Ingeniero industrial Mujer</i>	0.154	porque es la que más me gusta
	0.222	porque era la que me gustaba desde siempre y además tiene salidas
	0.233	porque me gusta esta carrera
<i>Magisterio Hombre</i>	0.543	me gusta mucho trabajar al lado de gente y sobre todo de niños y niñas creo que la profesion de maestro se adapta a mis características
	0.551	no lo sé el transcurso de mi vida académica nunca ha tenido un objetivo final claro me gusta aprender diferentes cosas y en diferentes momentos ayer técnico superior en salud ambiental hoy maestro y mañana
	0.552	porque me gustan tanto la musica como los niños
<i>Magisterio Mujer</i>	0.157	porque es una carrera que me gusta
	0.191	porque es la que más me gustaba
	0.229	porque me gusta
<i>Enfermería Hombre</i>	0.265	porque amo tanto la vida que quiero ayudar a las personas con alguna dificultad a que la vivan con alegría y la amen el amor a vivir y a la vida es mi principal motivo
	0.687	porque me interesaba y porque no estaba realmente decidido a hacer algo concreto
	0.815	vocación, mi novia y mi ilusión
<i>Enfermería Mujer</i>	0.137	porque es la carrera que más me gusta
	0.137	porque es la carrera que más me gusta
	0.167	porque me gusta la carrera

Las respuestas características son respuestas reales, completas, proporcionadas por los estudiantes. Se considera que una respuesta es modal o característica cuando su perfil es próximo al perfil léxico medio del grupo. Para calcular la proximidad entre perfiles se utiliza, en este caso, la distancia chi-cuadrado.

En la tabla 4 se recoge una selección de las tres respuestas más características para ocho categorías de estudiantes (sin signos de puntuación). En este caso las palabras están en su contexto original. En general, se observa que hay muchas respuestas que indican dos motivos al argumentar la elección. Entre las primeras respuestas características siempre se encuentra “me gusta”, motivo que está presente casi siempre y que no es propio de ningún grupo.

La lectura de las especificidades y de las respuestas más características corrobora la interpretación anterior, obtenida en el análisis de correspondencias y de clasificación.

Conclusiones

Los métodos de análisis de estadística textual se han manifestado como un instrumento muy útil para el análisis de una pregunta de respuesta libre, en la que se han obtenido respuestas con gran riqueza de información y de la que se dispone de un número elevado de observaciones, como ocurre en este caso.

En esta encuesta, el análisis clásico de codificación a posteriori, hubiera resultado prácticamente imposible, dado que la mayor parte de las respuestas incluyen dos o más motivos diferentes. Otros procedimientos clásicos más cualitativos se ven dificultados por el tamaño de la población analizada.

El Análisis Factorial de Correspondencias de la tabla directa ha permitido detectar cómo algunas palabras, que se podrían considerar a priori semejantes por tener la misma raíz o ser diferentes formas de un mismo verbo, no son empleadas de la misma forma ni en los mismos contextos, por lo que no deberían codificarse en una misma categoría.

Los distintos métodos utilizados: análisis de correspondencias de la tabla directa, y de la tabla léxica agregada, análisis de las especificidades, análisis de clasificación de las formas y análisis de las respuestas características, concluyen que existen motivos diferentes en la elección de las titulaciones, sin embargo, entre sus motivaciones casi siempre está presente *me gusta*.

Las diferencias detectadas inicialmente entre hombres y mujeres, están relacionadas con la elección de diferentes titulaciones y no se aprecian grandes diferencias entre éstos dentro de la misma titulación.

Los estudiantes de Enfermería y Magisterio, se distancian del resto de sus compañeros, pues el análisis de sus respuestas indica que son titulaciones que se suelen elegir por vocación. Expresan que la elección de su carrera no ha sido una decisión de última hora, sino que les gustaba *desde siempre*. Además, no suelen indicar ningún tipo de motivación económica.

Los estudiantes de Ingeniería Industrial y Economía eligen estas carreras pensando principalmente en las salidas profesionales y en el futuro laboral que puedan ofrecer. También indican el gusto por la titulación, estableciéndose aquí una distinción entre los dos grupos, pues los estudiantes de Ingeniería Industrial se refieren, además, al gusto por las asignaturas que se estudian en su carrera.

Los estudiantes de Relaciones Laborales y Derecho son los que utilizan un vocabulario menos específico y comparten características de los otros grupos, es decir, les gusta la titulación, piensan en las salidas profesionales. Por eso, sus respuestas tienen un perfil más próximo a la media.

Referencias

- Abascal, E. y Grande, I. (1989) *Métodos multivariantes para la Investigación Comercial*. Barcelona: Ariel.
- Bécue, M. (1991) *Análisis de datos textuales. Métodos estadísticos y algoritmos*. CISIA
- Benzécri J-P y col. (1973) *La taxonomie, Voll. L'Analyse des Correspondances, Vol II*. Paris: Dunod.
- Benzécri J-P y col. (1981) *Pratique de l'Analyse des Données, tome III, Linguistique et Lexicologie*. Paris: Dunod.
- Escofier, B. y Pagès, J. (1992) *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*. Universidad del País Vasco.
- Etzeberria, J.; García, E.; Gil J. Y Rodríguez G. (1995). *Análisis de datos y textos*. Madrid: Rama.
- Gragé, D. Y Lebart, L. (1994) *Traitements statistiques des enquêtes..* Paris : Dunod.
- Lebart, L. (1986) Analyse statistique des réponses libres dans les enquêtes par sondage. *Revue Française du Marketing*, 109.
- Lebart, L.; Salem, A. (1988) *Analyse statistique de données textuelles. Questions ouvertes et lexicométrie*. Bordas.
- Lebart, L. (1995). Recent Development in Textual Data Analysis. *Analyses Multidimensionnelles des Données. III Congrès International NGUS '95*.
- Lebart, L.; Morineau, A. y Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Lebart, L.; Morineau, A.; Bécue, M. y col. *SPAD. T. Système Portable pour l'Analyse de Données Textuelles. Manuel de référence*. CISIA.
- SPAD.T *Système Portable pour l'Analyse de Données Textuelles*. CISIA.
- SPAD.N *Système Portable pour l'Analyse de Données*. CISIA.