

Exploración textual en el contexto del Modelo de Valores en Competencia. Aplicación al tipo de cultura de la UPV-EHU(1)

por

K. FERNÁNDEZ AGUIRRE,
J. I. MODROÑO HERRÁN

Departamento de Economía Aplicada III. (Econometría y Estadística)
Universidad del País Vasco - Euskal Herriko Unibertsitatea

M. I. LANDALUCE CALVO

Departamento de Economía Aplicada
Universidad de Burgos

RESUMEN

Este trabajo explora la utilización de dos técnicas de análisis de datos textuales: por un lado, el análisis de correspondencias, que permite una representación gráfica de una tabla de contingencias, para la investigación de asociaciones; por otro lado, la clasificación automática, que permite identificar grupos a partir de observaciones multivariantes. El estudio empírico se centra en una pregunta abierta incluida en un cuestionario sobre las características dominantes de la cultura organizacional en la Universidad del País Vasco–Euskal Herriko Unibertsitatea (UPV-EHU). Este estudio presenta la particularidad de su planteamiento en el marco del Modelo de Valores en Compe-

(1) Este trabajo ha sido financiado por el Grupo de Investigación Consolidado UPV 038.321-13631/2001 y por el Proyecto de Investigación FIT-080000-2001-0063.

tencia de Cameron y Quinn. Los resultados muestran una tipología de personal docente que se ajusta parcialmente al modelo propuesto.

Palabras clave: estadística textual, AC, clasificación

Clasificación AMS: 62H25

1. INTRODUCCIÓN

Cuando se plantea un estudio en el que la información de partida se va a obtener a través de encuestas, surge la duda sobre la naturaleza de las preguntas a incluir en el cuestionario a utilizar. La utilización de preguntas cerradas tiene la ventaja de una relativa facilidad de tratamiento e interpretación de los resultados, y es la razón principal de su empleo generalizado. Sin embargo, por construcción, estas preguntas condicionan y limitan las respuestas del encuestado. El menor uso de preguntas abiertas tiene relación directa con la mayor dificultad de tratamiento y, sobre todo, de interpretación de los resultados.

El análisis estadístico de respuestas a preguntas abiertas es uno de los objetivos de la *Estadística Textual* que se centra en el análisis de las frecuencias de los términos o vocablos que aparecen en un texto y de su relación con características del sujeto o sujetos autores del mismo. Está relacionada con métodos estadísticos factoriales y de clasificación como puede verse en, por ejemplo, Etxeberria et al. (1995), Lebart y Salem (1988 y 1994), Lebart, Salem y Berri (1998) o Lebart, Salem y Bécue (2000). Estas técnicas han sido aplicadas con éxito a datos textuales procedentes de cuestiones abiertas incluidas en encuestas o de textos políticos, históricos, narraciones, discusiones de grupo, entrevistas y otros documentos (véase, por ejemplo, Fernández Aguirre (2003)). Su importancia no deja de crecer en la llamada Sociedad de la Información, en la que los volúmenes de información son ingentes y, en particular, lo son las grandes bases de datos de textos como las que manejan, por ejemplo, los buscadores de términos de Internet como Google o Yahoo. A este área, relacionada con la Inteligencia Artificial, ya se la conoce como *Text (o Textual) Data Mining* o Minería de Datos Textual (véase, por ejemplo, Hearst (1999)).

La principal ventaja de las técnicas englobadas dentro de la Estadística Textual aplicada sobre una pregunta abierta es que son más objetivas en la obtención y proceso de los datos que en el caso de respuestas a preguntas cerradas, cuya confección requiere una fuerte intervención del investigador. En el caso de una pregunta abierta, correctamente planteada, el investigador interviene de forma decisiva sólo en la fase de interpretación de los resultados lo que hace sencilla su revisión. En el caso de preguntas cerradas, existen otro tipo de condicionantes

sobre su formulación, opciones de respuesta... y sería necesario volver a realizar la encuesta si se dudase sobre la corrección del cuestionario.

En este estudio, el interés se centra en el análisis estadístico de una cuestión abierta, centrada en cuatro aspectos diferentes relativos a la mejora de la cultura organizacional de una universidad pública, en el ámbito del Modelo de Valores en Competencia de Cameron y Quinn (1999). Pueden encontrarse otros estudios con el mismo objetivo pero usando preguntas cerradas, como Sporn (1996) o Berrio (2003). A diferencia de ellos, en este trabajo se utiliza la Estadística Textual para extraer los términos más relevantes de las respuestas, así como su caracterización basada en las características de los encuestados.

La estructura de este artículo se describe a continuación. En la sección segunda se introduce la metodología estadística empleada. En el tercer punto se describe el modelo organizativo, como marco teórico previo a los análisis realizados. En el cuarto se presentan los datos y en la sección quinta se analizan detalladamente los resultados. Finalmente, se exponen las conclusiones en la sección sexta.

2. METODOLOGÍA

Estadística Textual

Cuando el investigador en ciencias sociales se encuentra ante datos de naturaleza textual, es decir, datos expresados en forma de cadenas verbales de cierta extensión y que generalmente acaban siendo registrados como textos escritos, puede recurrir a procedimientos de análisis de tipo cualitativo y/o cuantitativo.

El modo más habitual y conocido de analizar cuantitativamente los datos textuales consiste en la aplicación de técnicas estadísticas a los valores numéricos resultantes del recuento del número de veces (frecuencia) que una palabra aparece en el conjunto de los datos textuales, correspondientes a cada individuo estudiado.

Estos valores numéricos son capaces de conducir a una descripción objetiva y sistemática del contenido de los textos. En esta línea se inscriben los denominados enfoques lexicométricos o de la estadística textual.

La primera operación realizada sobre los textos objeto de análisis consiste en la *segmentación del corpus textual*, que implica diferenciar sus unidades elementales. Siguiendo diferentes criterios, es posible diferenciar diversas unidades que han sido tomadas en consideración por los investigadores: forma gráfica, segmentos repetidos, etc.

La *forma* es una unidad física, fácilmente reconocida por un ordenador y que propicia, por tanto, una segmentación automática. Es la unidad más frecuentemen-

te empleada y generalmente coincide con una palabra. Lebart y Salem (1988) definen *forma gráfica* como una secuencia de caracteres no delimitadores (en general, letras) comprendida entre dos caracteres delimitadores (espacios o signos de puntuación).

Considerando que las palabras aisladas resultan descontextualizadas y su análisis puede conducir a interpretaciones equívocas, también se pueden utilizar como unidades los *segmentos repetidos*, definidos como una secuencia de dos o más formas, no separadas por un delimitador de secuencia, que aparecen más de una vez en un corpus de datos textuales.

Las tablas léxicas constituyen formas de reorganizar la información que presentan interés de cara a la descripción. Una *tabla léxica* es una tabla de doble entrada en la que las filas (o columnas) corresponden a las unidades resultantes de la segmentación del texto, generalmente palabras, y las columnas (o filas) a los textos elementales considerados; como por ejemplo, la respuesta de cada individuo. Considerando un corpus dividido en n partes ($E_1, E_2 \dots E_j \dots E_n$) en el que han sido diferenciadas r formas ($F_1, F_2 \dots F_i \dots F_r$), la tabla léxica recoge la frecuencia de aparición de la forma situada en la fila i dentro del subcorpus correspondiente a la columna j (k_{ij}).

Cuando cada columna de esta tabla se construye a partir del agrupamiento de varios textos elementales, por ejemplo, todas las respuestas de los individuos de una determinada edad, recibe el nombre de *tabla léxica agregada*. Una tabla de este tipo permite al investigador, siempre que los textos tengan una extensión similar, realizar comparaciones entre las frecuencias alcanzadas en cada uno de ellos por determinadas formas relevantes para el estudio.

Es habitual establecer unos niveles mínimos de frecuencia que deben alcanzar las *formas o segmentos repetidos* para ser incluidos en una tabla como la anterior, sin que los resultados varíen significativamente (Salem (1981)). También lo es excluir de las tablas léxicas las *palabras funcionales*, tales como artículos, preposiciones, conjunciones, demostrativos, algunos adverbios, etc., por su escasa significatividad en relación al contenido de las ideas que transmite un texto.

Las posibles técnicas de la Estadística Textual son diversas; nuestro interés se centra en los métodos de análisis de correspondencias (en adelante AC) y clasificación automática, dos métodos exploratorios multivariantes complementarios adecuados al tratamiento de datos cualitativos.

Análisis de Correspondencias y Clasificación

Los datos textuales son susceptibles de ser analizados mediante técnicas de reducción de dimensionalidad como el AC o los métodos de clasificación. Ambos métodos están ampliamente tratados en la bibliografía de Estadística Multivariante (véase Greenacre (1984) o Lebart, Morineau y Piron (2000)), por lo que aquí nos ocuparemos únicamente de destacar la posibilidad de su aplicación a datos de naturaleza textual.

El AC es aplicado, en el marco de la Estadística Textual, a tablas de contingencia en las que disponemos de información sobre la frecuencia en que aparecen determinadas unidades textuales en distintos textos o partes de un corpus textual considerado. Si, por ejemplo, construimos una tabla (*tabla léxica*) en la que indicamos la frecuencia en que cada forma gráfica (forma lematizada o segmento repetido) aparece en el texto emitido por un sujeto, es posible representar a los individuos en el espacio determinado por las formas gráficas. Se trata, por tanto, de comparar los perfiles léxicos de los textos correspondientes a cada individuo, permitiendo localizar en el espacio los individuos que resulten más cercanos (o más alejados) entre sí, al tiempo que las formas causantes de esa proximidad.

Si las tablas de contingencia son construidas agrupando en cada columna a los individuos (*tabla léxica agregada*) para los cuales una determinada variable adopta el mismo valor (edad, nivel social...), podemos identificar la posición relativa de unos grupos respecto a otros de acuerdo con el vocabulario empleado, y caracterizar la posición de los mismos por las formas próximas en el espacio. Por ejemplo, a veces es interesante conocer si distintos grupos de edad responden ante una pregunta abierta empleando el mismo vocabulario; o tal vez pretendemos comparar las opiniones sobre un tema expresadas por sujetos procedentes de diferentes ámbitos.

Los métodos de clasificación permiten la obtención de una partición en clases de las formas (u otro tipo de unidades textuales consideradas) en función de su empleo por parte de los sujetos o la clasificación de los individuos en función del vocabulario empleado por los mismos. Los métodos de clasificación conducirán a la construcción de clases de individuos homogéneas en cuanto al vocabulario empleado, o a la determinación de grupos de palabras que suelen ser empleadas por los mismos individuos y que delimitan, por tanto, campos semánticos o temáticas conectadas entre sí.

La clasificación de los individuos o unidades textuales puede realizarse a partir de las coordenadas de estos tras llevar a cabo un AC. Ambos métodos de análisis pueden ser utilizados conjuntamente, permitiendo una mejor descripción de los datos (véase Lebart (1994) o Greenacre (1987)).

Valores-test

Los valores-test (véase Lebart et al. (1984, p. 95), Morineau (1984, p. 24) o Etxebarria (1995)) conforman una herramienta de caracterización de un eje factorial de un AC a partir de las modalidades de una variable suplementaria o de un grupo o cluster obtenido a través de un método de clasificación.

En el primer caso, los valores-test son mayores cuanto mayor sea la coordenada de una categoría suplementaria sobre el factor y cuanto mayor sea el número de individuos que escogen esa modalidad. Bajo la hipótesis nula de que la categoría no esté representada mejor que otra u otras sobre el eje factorial, el valor-test sigue una distribución normal tipificada.

Para caracterizar un grupo de una clasificación, el valor-test, bajo la hipótesis nula de que la proporción de individuos que escogen una modalidad dentro de una clase es la misma que la proporción de individuos que la escogen en el total de la muestra, sigue una distribución hipergeométrica. La interpretación del contraste es similar a la del primer caso.

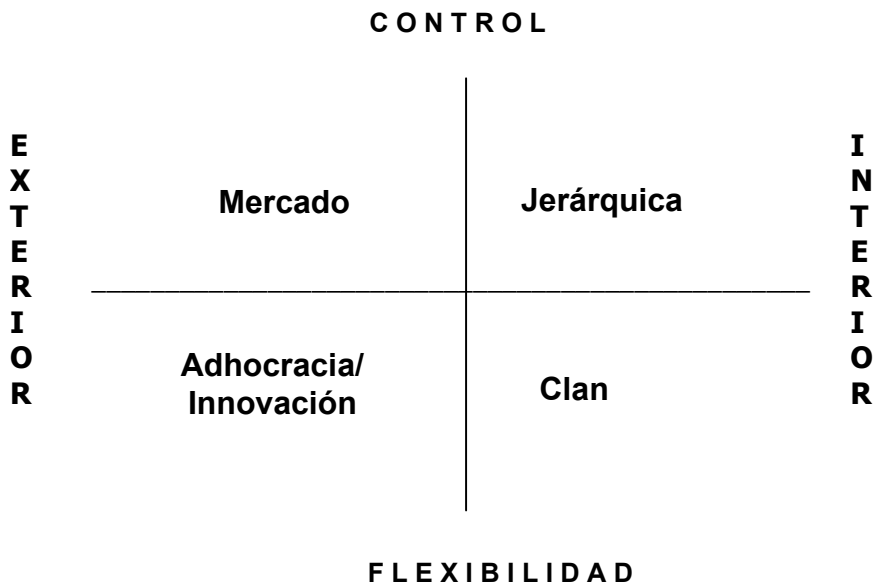
3. EL ANÁLISIS DE LA CULTURA ORGANIZACIONAL: EL MODELO DE VALORES EN COMPETENCIA

Existen diversas aproximaciones al concepto de cultura organizacional, uno de los modelos más aceptados es el adoptado en el diseño de la encuesta: el Modelo de Valores en Competencia de Cameron y Quinn (1999). Este modelo se basa en dos dimensiones bipolares que, a modo de ejes cartesianos, se cruzan produciendo cuatro cuadrantes con sus respectivos tipos de cultura.

La primera dimensión, en el eje horizontal, contrapone la estrategia de orientación interna que tiene la organización (*interna, a corto plazo y actividades simples*) a la estrategia de orientación externa (*externa, a largo plazo y de actividades orientadas al logro de objetivos*). La segunda dimensión, en el eje vertical, contrapone, según las características del control ejercido por la organización sobre sus miembros, *flexibilidad, individualidad y espontaneidad*, frente a *estabilidad, control y predecibilidad*.

Estos dos ejes definen cuatro tipos de cultura, como se representa en la figura 2.

Figura 2
 TIPOS DE CULTURA ORGANIZACIONAL.
 MODELO DE CAMERON Y QUINN



Los cuatro tipos de cultura son:

Clan: con orientación interna y control flexible. Este tipo de cultura se caracteriza por los valores y objetivos compartidos, la cohesión, la participación, desarrollo de los individuos, el sentido de grupo, funcionamiento mediante trabajo en equipo, normas de lealtad y consenso. Su nombre se debe a su similitud con una organización familiar.

Jerárquica: con orientación interna y control estable. Este tipo de cultura se basa en las reglas, la especialización, impersonalidad, responsabilidad y jerarquía. Las metas de la organización a largo plazo son la estabilidad, la previsibilidad y la eficiencia.

De Mercado: con orientación externa y control estable. Es un tipo de organización que funciona como un mercado en sí misma. Los valores centrales son la competitividad y la productividad. Una cultura de mercado es un espacio de trabajo orientado hacia los resultados.

Adhocrática: con orientación externa y control flexible. Es una cultura de iniciativas pioneras e innovadoras como llaves del éxito. La tarea de gestión más importante es estimular el conocimiento, el riesgo y la creatividad para poder estar siempre en la vanguardia.

Como se explica en la sección siguiente, en la cuestión abierta se pregunta sobre las posibles mejoras que el profesor encuestado propondría para distintos aspectos mejorables en la institución. Como hipótesis de partida podemos imaginar la terminología empleada por un tipo de profesorado que concibe las mejoras en términos de cada uno de los tipos de cultura descritos. Así, si los encuestados en sus respuestas hablan de departamento (entorno más próximo de un profesor dentro del trabajo), cooperación, profesores, diálogo, ... podremos suponer que subyace un deseo de mejoras en términos de cultura de tipo **Clan**. En el caso de que se hable predominantemente de autoridad, poder, dirección, ... pensaremos en una cultura **Jerárquica**. Términos como recursos, gestión, competir, ... nos indicarán cultura de tipo **Mercado**. Por último, si la terminología incide en términos como investigación, grupos de investigación, universidad, creatividad, ... la cultura sugerida sería de tipo **Adhocracia**.

4. LOS DATOS DE LA ENCUESTA. EL ANÁLISIS TEXTUAL DE UNA PREGUNTA ABIERTA

Los datos que analizamos forman parte de una encuesta realizada, mediante muestreo aleatorio estratificado, a profesores de la Universidad del País Vasco (UPV-EHU), por un grupo de trabajo dentro del proyecto multidisciplinar titulado *Gestión del conocimiento en una universidad pública: El proceso de Investigación-Desarrollo-Transferencia del conocimiento científico técnico*(2).

La encuesta se planteó con el objetivo de analizar la cultura organizacional de una universidad pública, en concreto de la UPV-EHU. Por una parte, se trataba de definir los valores dominantes de la organización universitaria y, por otra, de describir las prácticas derivadas de los valores dominantes. Los 60 ítems de la encuesta fueron diseñados, por el equipo de trabajo, tomando como punto de partida el Modelo de Valores en Competencia presentado en la sección tercera y sus respuestas en una escala de Likert de 5 puntos fueron tratadas mediante una metodología clásica de búsqueda de escalas aditivas (Aritzeta, Ayestarán y Balluerka

(2) El proyecto fue dirigido por el Dr. A. Rodríguez Castellanos. La encuesta fue diseñada por el grupo de trabajo dirigido por el Dr. S. Ayestarán Etxeberria e integrado por K. Fernández Aguirre, J. Landeta, J. Larrea, P. Mariel, A. Martín y F. Panera.

(2003), pag. 248)(3) mediante análisis de componentes principales. Además, se utilizaron otros análisis estadísticos complementarios: análisis de varianza, regresión probit, análisis discriminante de Fisher, AC múltiples y de clasificación (véase Fernández, Modroño y Landaluce (2004) y Fernández, Mariel y Martín (2004)). En la última parte de la encuesta se introdujo una cuestión abierta, con la presunción de que el análisis textual podría constituir un instrumento válido para colaborar en la posible definición de los valores dominantes de la UPV-EHU. Dicha cuestión abierta fue: ***Escriba de forma esquemática cómo se podrían mejorar:***

1. La forma en que se ejerce la autoridad,
2. Los sistemas de comunicación,
3. El desarrollo de los recursos humanos,
4. Las conductas de manejo de conflictos.

La cuestión abierta planteada se analizó agrupando los cuatro aspectos, de modo que en el análisis lexical el corpus analizado, así como la frecuencia de las formas empleadas fuera lo suficientemente grande. Las comparaciones de perfiles lexicales serán significativas desde un punto de vista estadístico si las formas textuales tienen un mínimo de frecuencia (véase Lebart, Salem y Berri (1998), pag. 104) (4).

El problema que motiva un análisis estadístico puede tratar de formalizar un modelo probabilístico a priori; o bien, puede estar planteado en términos generales con el objetivo de explorar un universo tomado en consideración. En el análisis textual de una cuestión abierta, por ejemplo, se trata de explorar los motivos que llevan a futuros alumnos de universidad a elegir una carrera u otra (Abascal y Franco, (2002)). La diferencia del análisis textual, en nuestro caso, estriba en que se trata de explorar la cultura dominante y su posible ajuste al Modelo de Valores en Competencia de Cameron y Quinn.

Mediante el análisis se trata de identificar la cultura dominante en relación, fundamentalmente, con los distintos grupos definidos en la encuesta por las categorías o modalidades de las variables de caracterización de los profesores, así como en

(3) En el artículo se presenta el Modelo Aditivo, entre otros modelos de búsqueda de variables latentes en diferentes niveles.

(4) Por una parte, al establecer un umbral de frecuencia el tamaño del vocabulario empleado se reduce drásticamente. Por otra, en la mayoría de los casos la precisión de un modelo obtenido es proporcional a la raíz cuadrada del tamaño de la muestra. La precisión de la localización de un elemento en el plano formado por las dos primeras dimensiones de un AC crece con la frecuencia.

relación con nueve cuestiones de tipo diferencial semántico planteadas en la encuesta.

Las variables sobre las características personales de los profesores encuestados, son las siguientes: Sexo, Edad, Rango (Catedrático, Titular, Asociado), Participación en Proyectos de investigación y Áreas de la Ciencia y la Tecnología (UNESCO).

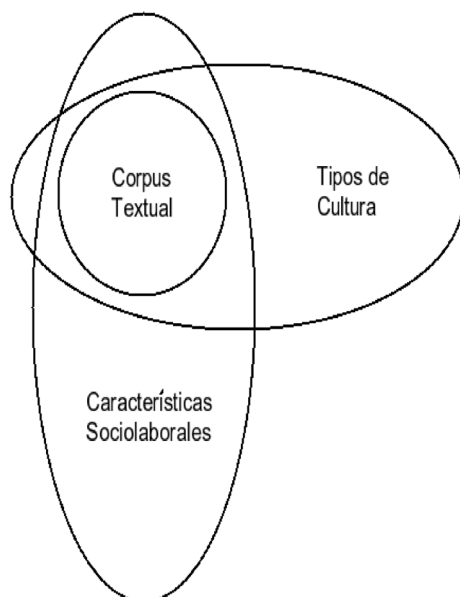
Las nueve variables de tipo diferencial semántico consideradas, con respuestas desde 1 (totalmente de acuerdo con la opción señalada) hasta 5 (totalmente de acuerdo con la opción opuesta)(5) en el otro extremo son las reflejadas en la tabla 1:

Tabla 1

CULTU1:	<i>1=Rígida</i>	<i>5= Flexible</i>
CULTU2:	<i>1=Abierta hacia el exterior</i>	<i>5=Centrada en la propia Universidad</i>
CULTU3:	<i>1=Individualista</i>	<i>5=Trabajo en equipo</i>
CULTU4:	<i>1=Innovadora</i>	<i>5=Tradicional</i>
CULTU5:	<i>1=Planificada</i>	<i>5=Imaginativa</i>
CULTU6:	<i>1=Cooperación</i>	<i>5=Rivalidad</i>
CULTU7:	<i>1=Resultados a largo plazo</i>	<i>5=Resultados a corto plazo</i>
CULTU8:	<i>1=Valoración de la investigación</i>	<i>5=Indiferencia hacia la investigación</i>
CULTU9:	<i>1=Investigación básica</i>	<i>5=Investigación aplicada</i>

La información procesada es esquemáticamente la siguiente:

(5) En el caso de CULTU9, en realidad, no se trata de opciones opuestas.



En los análisis realizados buscamos clasificar el vocabulario empleado en las respuestas de los profesores. Para ello, utilizamos el AC, análisis de gran robustez que proporciona una clasificación de formas textuales en un espacio continuo, seguido de un análisis de clasificación automática que proporciona clases disjuntas de dichas formas.

Las clases son a priori desconocidas y el análisis realizado en esta investigación presenta la particularidad de una búsqueda en el marco del Modelo de Valores en Competencia de Cameron y Quinn, que, en sí mismo, supone una clasificación teórica de partida.

4. RESULTADOS DEL ANÁLISIS TEXTUAL

A continuación, presentamos diversos análisis multivariantes precedidos de un análisis del vocabulario empleado por los encuestados, como una primera aproximación que proporciona el conocimiento de las formas y segmentos más frecuentes.

Análisis del vocabulario

Para realizar este análisis se ha aplicado el módulo textual del programa Spad para Windows (Decisia (2003)). En el conjunto de las respuestas tenemos 11316 formas textuales de las que 2333 son formas diferentes. Esto da idea de la riqueza de vocabulario empleado por los 306 profesores que responden a la pregunta planteada para explicar cómo se pueden mejorar los cuatro aspectos de la cultura organizacional de la universidad considerados. En la tabla 2 se presenta una lista de las formas más frecuentes, después de prescindir de las palabras funcionales.

Tabla 2

<i>Formas empleadas</i>	<i>Frecuencia</i>	<i>Formas empleadas</i>	<i>Frecuencia</i>	<i>Formas empleadas</i>	<i>Frecuencia</i>
diálogo	83	consenso	32	sistemas	16
comunicación	67	gestión	31	profesores	16
autoridad	63	entre	29	dirección	16
mayor	58	conflictos	27	ser	16
todos	47	grupos	27	canales	15
formación	45	reuniones	23	respeto	15
investigación	42	departamentos	21	profesorado	15
información	41	participación	20	cursos	15
personas	40	forma	20	hay	15
personal	38	dialogando	19	transparencia	15
menos	37	personales	19	poder	15
recursos	34	todo	18	necesidades	15
capacidad	33	profesional	18	grupo	15
universidad	33	debe	17	aspectos	14
son	33	real	16	dando	14

Un examen de dicha tabla permite comprobar que, claramente, aparecen formas que evocan los distintos tipos de cultura. Las formas *diálogo*, *comunicación*, *consenso*, *personas*, *personal*, *profesores*, *grupo* sugieren una cultura de tipo **Clan**. Las formas *autoridad*, *dirección*, *respeto*, *poder* podemos asociarlas a la cultura de tipo **Jerárquica**. Por otra parte, las formas *gestión*, *recursos*, *capacidad* nos sugieren una cultura de **Mercado**. Por último, formas como *formación*, *investigación*, *información* y *universidad* corresponderían a una cultura **Adhocrática** o innovadora.

Al tratar de contextualizar estas formas pasamos al análisis de los segmentos repetidos. En la tabla 3 se presentan los 50 segmentos de mayor frecuencia en el

corpus textual. Previamente se obtuvieron y analizaron los segmentos de longitud menor o igual a 3 y de frecuencia mayor o igual a 5.

Tabla 3

<i>Frecuencia</i>	<i>Segmento</i>	<i>Frecuencia</i>	<i>Segmento</i>
40	la autoridad	9	los cargos
26	de investigación	9	las partes
26	las personas	9	más recursos
24	la universidad	9	de personas
23	todos los	9	el diálogo
22	de comunicación	9	se ejerce
20	de gestión	8	a todos los niveles
16	grupos de	8	mayor comunicación
16	la comunicación	8	la participación
15	los conflictos	8	el dialogo
15	capacidad de	8	acceso a
13	la información	8	más dialogo
13	personas que	8	los profesores
12	grupos de investigación	7	correo electrónico
11	a través de	7	necesidades de
11	las personas que	7	teniendo en
11	de la universidad	7	de información
11	en cuenta	7	a las personas
11	los departamentos	7	formación y
10	la investigación	7	de las personas
10	las necesidades	7	la gestión
10	de conflictos	7	y consenso
9	de recursos	7	comunicación bidireccional
9	la calidad	7	participación de

Nuevamente se observan segmentos(6) que evocan los cuatro tipos de cultura. Casi el 50% de ellos evocan una cultura de tipo **Clan**: *las personas, de comunica-*

(6) En los segmentos repetidos la frecuencia dentro del corpus textual de cada forma lexical se ve reducida. Por ejemplo el segmento "la autoridad" tiene una frecuencia de 40 mientras que la forma "autoridad" tiene una frecuencia de 63, esto quiere decir que en 23 ocasiones esta forma no va precedida del artículo la.

ción, todos los, grupos de, la comunicación, personas que... Aproximadamente el 25% evocan una cultura de tipo **Adhocrática: de investigación, la universidad, la información, de formación...** El 12% evocan una cultura de tipo **Jerárquica: la autoridad, los conflictos, se ejerce, de conflictos...** Finalmente, encontramos algunos segmentos que evocan cultura de **Mercado: de gestión, de recursos, más recursos...**

Lo mismo que en el caso de las formas textuales, la adscripción a un determinado tipo de cultura resulta discutible y es, sin duda, subjetiva. Este hecho se engloba en un fenómeno más general que ya en 1910 señalaba el lingüista Ferdinand de Saussure(7) que decía:

“En cualquier lugar hay objetos que cada uno puede considerar desde diferentes puntos de vista. El punto de vista surge en primer lugar, correcto o equivocado, es sólo un punto de vista mediante el cual creamos objetos secundarios. Estas creaciones se corresponden con la realidad cuando el punto de partida es correcto y no cuando el punto de partida es equivocado, pero en ningún caso existe un objeto correcto en sí mismo. Ni tan siquiera cuando consideramos hechos materiales, hechos que aparentemente tienen una definición obvia, como unas series de vocales sonoras.”

Nuestro punto de partida, aunque basado en el modelo de Cameron y Quinn, es inevitablemente subjetivo, lo mismo que la interpretación dada a las formas y segmentos más frecuentes; sin embargo, la exploración textual confirma el punto de vista de partida y las conclusiones a las que llegamos no deben estar muy alejadas de la realidad por su coherencia. En concreto, el análisis frecuencial efectuado confirma el resultado obtenido en estudios anteriores basados en el análisis de los datos, no textuales, de esta misma encuesta (véase Fernández, Modroño y Landaluze (2004) y Fernández, Mariel y Martín (2004)), así como en los primeros análisis basados en el Modelo Aditivo (Rodríguez, (2002). Este análisis frecuencial preliminar sugiere que la cultura de la UPV-EHU es sobre todo flexible y predominantemente de tipo **Clan**, seguida de una cultura de tipo tipo **Adhocrática** o innovadora, con importancia relativa. El objetivo perseguido en este trabajo, sin embargo, es más ambicioso, tratamos de identificar distintos grupos asociados, con mayor o menor intensidad, a diversos tipos de cultura y para ello realizamos el análisis de distintas tablas de contingencia lexicales.

A continuación, presentamos varios análisis exploratorios secuenciales que pueden englobarse en el concepto de minería de datos textuales.

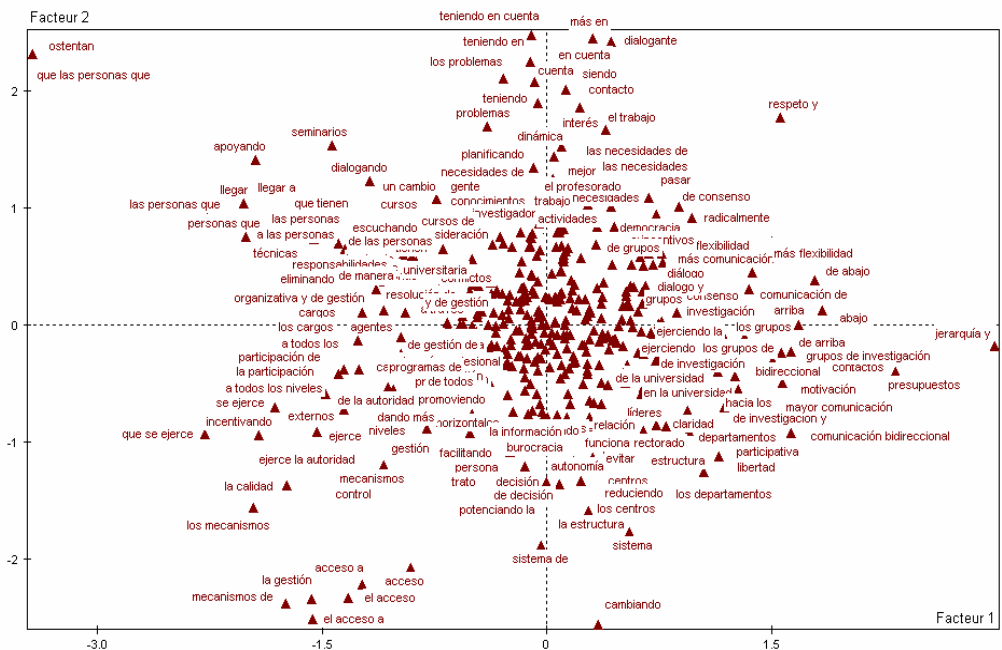
(7) Tomado de Lebart *et al.* (1998)

Análisis de Correspondencias de la tabla léxica directa

En este caso, partimos del AC de la tabla de contingencia que cruza 301 individuos (profesores) en filas(8) con las 441 formas y segmentos repetidos de mayor frecuencia en columnas. La tabla contiene, asimismo, las columnas suplementarias correspondientes a las 9 variables de cultura y a las 5 variables de caracterización de los encuestados. Las categorías de estas variables suplementarias no contribuyen a la creación de los ejes y planos factoriales pero se proyectan sobre los obtenidos en el AC de la tabla activa. Los valores-test que nos proporciona el programa nos informan de su posición más o menos significativa sobre los ejes factoriales. La figura 3 muestra el plano principal, formado por los dos primeros ejes, de la tabla activa.

Figura 3

PLANO PRINCIPAL DEL AC DE LA TABLA LÉXICA DIRECTA. FORMAS Y SEGMENTOS REPETIDOS



(8) Tras los primeros análisis se ha prescindido de cinco individuos excesivamente específicos y que impedían un análisis claro de la tabla de datos.

En este plano(9) podemos observar, si recorremos el mismo partiendo del primer cuadrante y en sentido directo, en primer lugar, formas y segmentos como *comunicación, comunicación de, flexibilidad, más flexibilidad, de consenso, el trabajo, dialogante...* En el segundo cuadrante se proyectan *planificando, gente, dialogando, seminarios*, y una variada serie de segmentos que incluyen la forma *personas*. Este vocabulario indica una percepción de las mejoras en términos de cultura flexible, y de tipo **Clan**.

En el tercer cuadrante se habla de participación pero en su mayoría tenemos formas y segmentos que evocan una cultura de tipo **Jerárquica**: *que se ejerce, ejerce, ejerce la autoridad, de la autoridad, control* y de tipo **Mercado**: *gestión, decisión, la decisión, la gestión*.

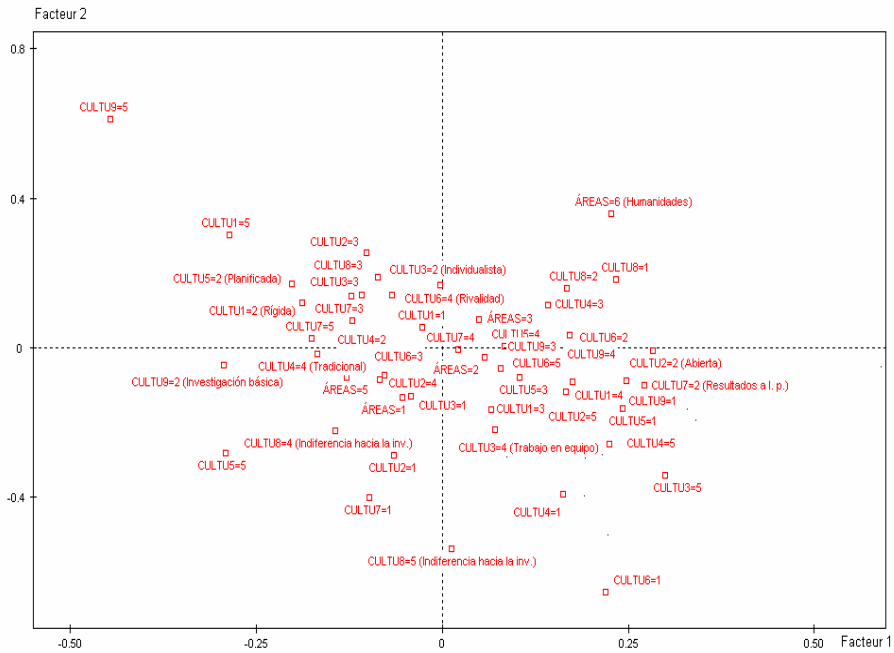
Por último, en el cuarto cuadrante se sitúan las formas y segmentos propias de una cultura de corte innovador o **Adhocracia**: grupos de investigación, de investigación, comunicación bidireccional, mayor comunicación, de investigación y, de la universidad, en la universidad, departamentos, claridad, líderes, centros, los centros, los departamentos, participativa, libertad, sistema. A diferencia de lo que ocurre en la parte superior del plano y sobre todo en el segundo cuadrante, la percepción de las posibles mejoras ya no se centra en el profesorado y las personas, sino en términos de institución en sus diversos aspectos, de libertad, de claridad y de investigación.

El análisis de los valores-test de las categorías de las variables suplementarias, permite vislumbrar que únicamente hay un valor-test reseñable en el caso de las características de los profesores encuestados, se trata de la categoría Humanidades de la variable área de conocimiento. Esta categoría se proyecta en el primer cuadrante del plano factorial analizado. Sin embargo, tenemos varios valores-test significativos asociados a categorías de las variables sobre tipos de cultura. Estas categorías se han etiquetado de forma expresa en su proyección en el plano principal.

En la figura 4 se ha prescindido de las formas y segmentos repetidos para facilitar la interpretación.

(9) En el plano principal surgen términos que no podemos tener en cuenta, bien por no estar correctamente representados, en general los que están cerca del origen de coordenadas, o bien, porque no evocan ningún tipo de cultura concreto, como por ejemplo, "radicalmente" en el primer cuadrante de la Figura 3.

Figura 4
PLANO PRINCIPAL DEL AC DE LA TABLA LÉXICA DIRECTA.
CATEGORÍAS DE VARIABLES SUPLEMENTARIAS



El análisis de este plano, teniendo en cuenta que cada categoría se sitúa en el centro de gravedad corregido de los individuos que la han escogido, pone de manifiesto, por otra parte, que los que piensan que la cultura en la UPV-EHU es abierta, de resultados a largo plazo y de trabajo en equipo conciben las mejoras con terminología de corte innovador o **Adhocracia**. Por otra parte, que los que piensan que predomina la investigación básica, que es rígida, planificada, tradicional, de indiferencia hacia la investigación y de rivalidad conciben la mejora con terminología de los otros tres tipos de cultura.

El AC de la tabla directa, ha permitido detectar la relación entre ciertas características individuales, como creencias del tipo de cultura existente o bien la pertenencia al área de humanidades con la visión sobre las mejoras que pueden establecerse en términos de vocabulario empleado. Sin embargo, debemos profundizar en el análisis y para ello procederemos al análisis de la tabla léxica agregada y posteriormente a la obtención de un análisis de clasificación sobre los primeros factores del análisis efectuado.

Análisis de Correspondencias de la tabla léxica agregada

La tabla léxica agregada es una tabla de contingencia que cruza 144 formas y segmentos repetidos en filas con las 45 categorías de respuesta correspondientes a las 9 variables de cultura en columnas (véase Tabla 1). Esta tabla contiene, además, como columnas suplementarias, las correspondientes a las categorías de respuesta de las características personales de los encuestados. El análisis de este plano, teniendo en cuenta que cada categoría se sitúa en el centro de gravedad corregido de los individuos que la han escogido, pone de manifiesto, por una parte, que los que piensan que la cultura en la UPV-EHU es abierta, de resultados a largo plazo y de trabajo en equipo conciben las mejoras objetivo que perseguimos al analizar esta tabla agregada consiste en la profundización en la relación entre el vocabulario empleado y las modalidades de cultura percibida por los encuestados, así como con las características de los mismos. Para ello, realizamos el AC de la tabla activa y proyectaremos en los ejes y planos obtenidos las características de los encuestados, como elementos suplementarios o ilustrativos.

La propiedad de doble representación simultánea que posee el AC proporciona una imagen del vocabulario y de las categorías de cultura en el plano formado por los dos primeros ejes factoriales. Así, cada elemento del vocabulario se encuentra proyectado en el centro de gravedad corregido de las categorías de cultura y cada una de estas, a su vez, se proyecta en el centro de gravedad corregido de las formas y segmentos repetidos. Las figuras 5 y 6 presentan ambas imágenes en dos planos separados para obtener una mayor claridad en la representación; no obstante, no debemos olvidar que cada categoría de cultura se relaciona con las formas lexicales y viceversa, por lo que tenemos que interpretar conjuntamente ambos planos.

Figura 5
PLANO PRINCIPAL DEL AC DE LA TABLA LÉXICA AGREGADA.
CATEGORÍAS DE CULTURA

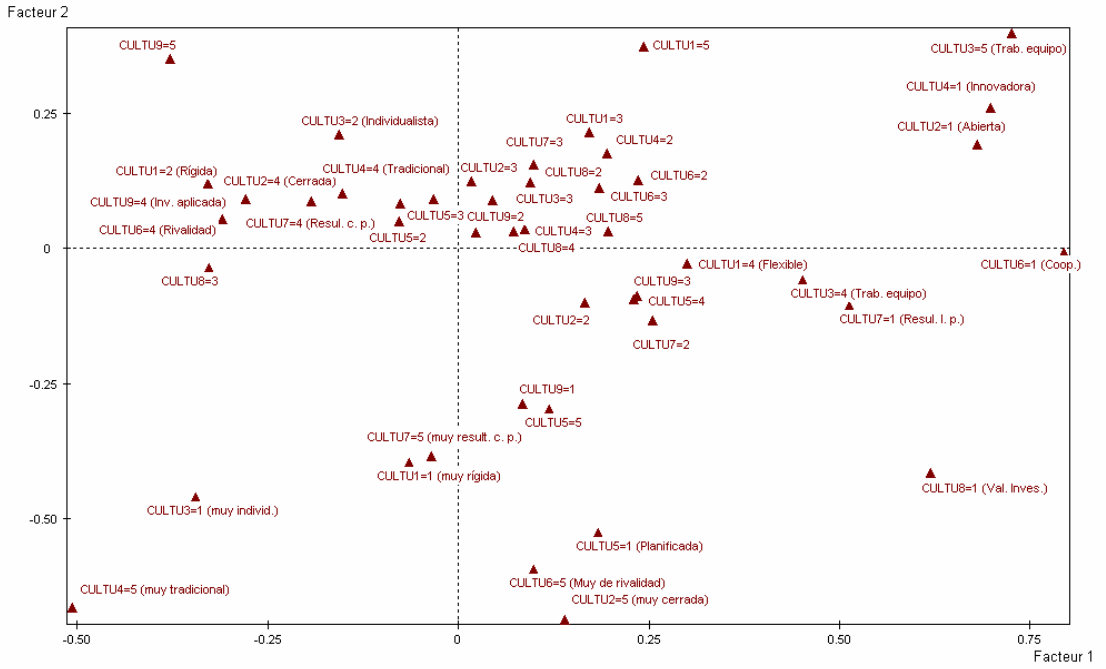
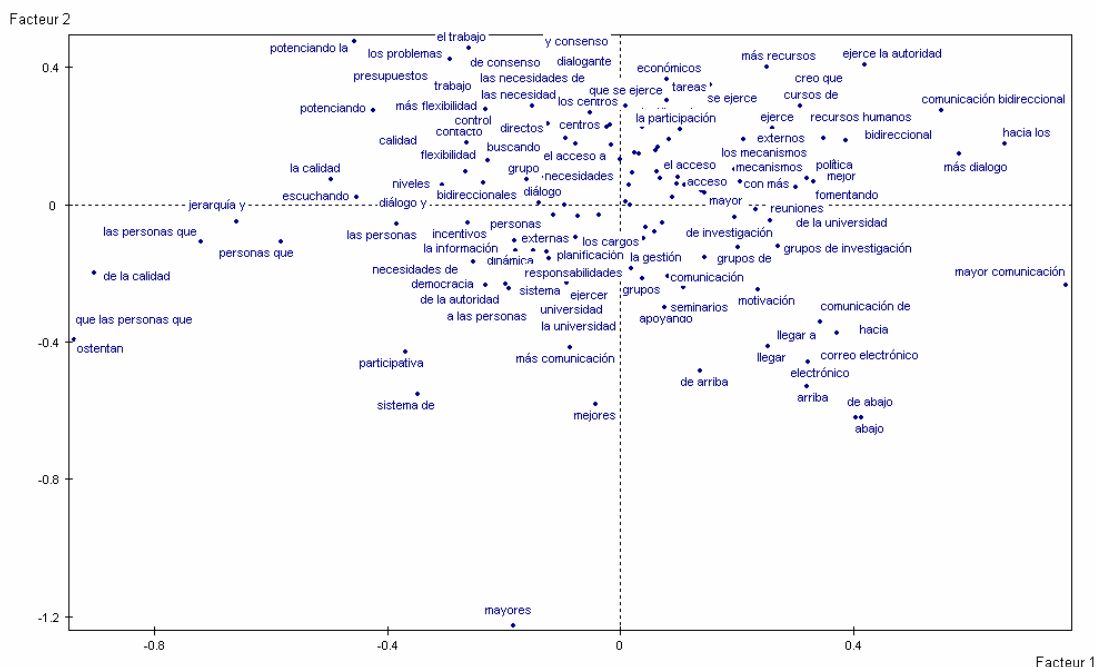


Figura 6
PLANO PRINCIPAL DEL AC DE LA TABLA LÉXICA AGREGADA.
FORMAS Y SEGMENTOS REPETIDOS



En cuanto a las categorías de cultura, en el plano factorial (Figura 5) podemos observar las categorías más relevantes, con su significado entre paréntesis. Se trata de las categorías que más contribuyen a la formación de los ejes y que están bien representadas en el plano. Así, en el primer cuadrante tenemos representados a los profesores que perciben la cultura de la universidad como de *trabajo en equipo*, *innovadora*, *abierta* y en cuanto a los términos empleados (Figura 6) vemos: *comunicación bidireccional*, *ejerce la autoridad*, *más recursos*, *recursos humanos*, *cursos de*, *bidireccional*, *más diálogo*, *política*, *fomentando*. Estos profesores sugieren una mejora en términos de cultura abierta pero enfatizando la colaboración entre los miembros de los grupos, con necesidad de recursos. Es decir, en cuanto al léxico empleado en sus respuestas sugieren una cultura de tipo **Clan**.

En el segundo cuadrante tenemos representados a los profesores que perciben la cultura como de *investigación aplicada*, *rígida*, *de rivalidad*, *cerrada*, *de resultados a corto plazo*, *tradicional e individualista* y en cuanto a los términos empleados observamos: *escuchando*, *la calidad*, *potenciando*, *flexibilidad*, *potenciando la y*

control. Estos profesores opinan que existe una cultura de tipo **Jerárquico**, y al mismo tiempo sugieren la necesidad de una cultura más flexible, potenciando la calidad y la cultura, quizás intermedia entre **Adhocracia** y de **Clan** que huye de una cultura tradicional e individualista.

En el tercer cuadrante sobresalen las categorías *muy individualista, muy tradicional, muy de resultados a corto plazo, muy rígida*, y asociadas principalmente al segundo eje. En cuanto a las formas y segmentos tenemos: *ostentan, que las personas que, de la calidad, las personas que, personas que, participativa, sistema, autoridad*. En este caso perciben todavía con mucha más nitidez las mejoras en términos de cultura de tipo **Clan**, aunque aparecen algunos términos asociables a una cultura **Jerárquica**.

Por último, en el cuarto cuadrante destacan las categorías *muy de rivalidad, muy cerrada y planificada*, asociadas al segundo eje y próximas en el plano a las citadas anteriormente al hacer referencia al tercer cuadrante. En el extremo derecho del plano y hacia el primer cuadrante tenemos los tipos de cultura más representativas de este cuadrante, destacando las categorías: *de valoración de la investigación, de resultados a largo plazo, de trabajo en equipo, flexible y de cooperación*. Representan a los profesores que tienen una visión más positiva de la universidad, piensan que es innovadora y sus respuestas a la cuestión planteada son en términos de: *mayor comunicación, grupos de investigación, de la universidad, comunicación de, hacía, correo electrónico, arriba, abajo, de abajo, electrónico, llegar, llegar a*. Podemos decir que este sector responde sobre las mejoras en términos de cultura innovadora o de **Adhocracia**.

Todos estos resultados permiten concluir que la cultura organizacional de la UPV-EHU no se refleja en cuatro cuadrantes similares a los del modelo teórico de Cameron y Quinn, pero los dos primeros factores sí que distinguen claramente entre dos corrientes de opinión, la primera de una universidad abierta e innovadora y la segunda de una entidad rígida, cerrada, tradicional e individualista.

Análisis de clasificación automática

La clasificación de los términos lexicales, que presentamos en la tabla 4, se ha realizado sobre los seis primeros factores del AC de la tabla léxica agregada, tomados como nuevas variables que recogen la estructura fundamental de los datos. El método utilizado ha sido el de la clasificación ascendente jerárquica de acuerdo con el criterio de Ward generalizado (véase Lebart, Morineau y Piron (2000)).

La clasificación en seis grupos de términos lexicales, que engloban distintos porcentajes del corpus textual, ha resultado ser la más adecuada. El valor-test indica las categorías de las variables suplementarias más características de cada

clase. La interpretación de cada uno de estos grupos requiere tener en cuenta que una categoría es característica de una clase si su presencia es significativamente mayor (valor-test positivo) o significativamente menor (valor-test negativo) que en el conjunto de los datos sin distinción de clases.

Tabla 4

(Continúa)

Formas y segmentos	Modelo Cultura	Características
Clase 1 (5,55 %)		
motivación	Valoración investigación	Menor proporción de profesores del área de Ingenierías y Tecnología: Áreas=2 (valor test (v.t.) = -3,24)
comunicación de arriba, de arriba	Centrada en la propia Universidad	
abajo, de abajo	Rivalidad	
correo electrónico	Planificada	
	Rígida	
Clase 2 (43,87 %)		
departamentos	Trabajo en equipo	Mayor proporción de profesores del área de Ciencias Exactas y Naturales: Áreas=1 (v.t. = 2,98)
diálogo	Flexible	
investigación	Imaginativa	Edad entre 41 y 50 (v.t. = 2,43)
facilitando	Resultados a largo plazo	
reuniones	Investigación básica	Menor proporción de profesores de Edad entre 51 y 60 (v.t. = -2,44)
gestión	Escasa rigidez	
información		
Clase 3 (8,34 %)		
política	Trabajo en equipo	Menor proporción de catedráticos: Rango=1 (v.t. = -2,59)
cursos	Abierta	
ejerce	Cooperación	
acceso	Innovadora	
consenso	Poco tradicional	
individual		
recursos		

Tabla 4

(Conclusión)

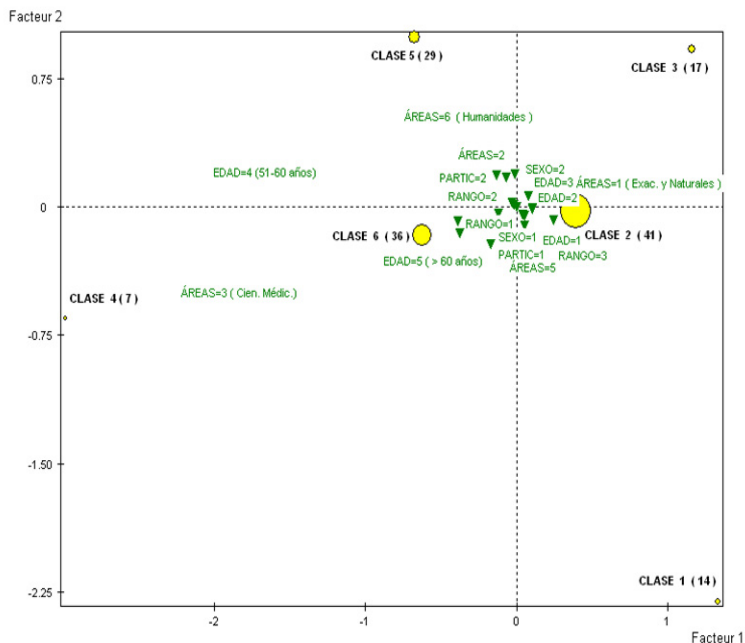
Formas y segmentos	Modelo Cultura	Características
Clase 4 (3,07 %)		
las personas que	Tradicional	Mayor proporción de profesores de Ciencias Médicas: Áreas = 3 (v.t. = 5,40)
jerarquía y ostentan la calidad de la calidad	Individualista Investigación aplicada Rígida Poco flexible	Edad entre 51 y 60 (v.t. = 5,12) Menor proporción de profesores de Edad entre 41 y 50 (v.t. = -4,18)
Clase 5 (12,86 %)		
Control grupo los centros, centros niveles las necesidades flexibilidad diálogo contactos	Individualista Investigación aplicada Resultados a corto plazo Indiferencia hacia la investigación	Edad entre 51 y 60 (v.t. = 4,10) Mayor proporción de profesores de Humanidades: Áreas=6 (v.t. = 2,9) y de Ingeniería y Tecnología: Áreas=2 (v.t. = 2,41) Menor proporción de profesores del área de Ciencias Sociales: Área=5 (v.t. = -2,61)
Clase 6 (25 %)		
Autoridad Personas, la autoridad diálogo necesidades bidireccionales incentivos	Planificada Rígida Investigación Básica/Aplicada Resultados a corto plazo	Edad más de 60 (v.t. = 2,47) Menor proporción de profesores de Edad entre 51 y 60 (v.t. = -2,46) Menor proporción de profesores del área de Ciencias Exactas y Naturales: Áreas=1 (v.t. = -4,16)

Resulta interesante observar en la figura 7 el plano principal del AC de la tabla léxica agregada. En él se han proyectado las clases obtenidas, así como las categorías de las características de los profesores encuestados(10). Aunque la clasifi-

(10) Se obtuvieron respuestas de 306 profesores, de los 600 incluidos en la muestra, a la parte textual que ocupaba el último lugar de un cuestionario exhaustivo. Sin embargo, análisis de homogeneidad realizados nos confirmaron que podemos estimar que se mantiene la estratificación establecida de partida en la muestra.

cación se ha efectuado sobre los seis primeros factores, este plano constituye un vínculo gráfico de unión entre el AC y el análisis de clasificación automática efectuado.

Figura 7
PLANO PRINCIPAL DEL AC DE LA TABLA LÉXICA AGREGADA.
CLASES Y CARACTERÍSTICAS DE LOS PROFESORES



En dicho gráfico se puede apreciar fácilmente que las clases, representadas por círculos de superficie proporcional a su tamaño, son bastante desequilibradas en cuanto al porcentaje del léxico que contienen. En este trabajo, dicho tamaño solamente representa el número de formas textuales próximas entre sí, dentro de cada clase, que utilizan los profesores encuestados y no el número de profesores que las utilizan. En ese sentido, la descripción de las clases, en lo que respecta a la identificación de sus variables características, es lo que relaciona dichos términos con los posibles tipos de profesores encuestados. La interpretación del plano factorial y la posición de estas clases en el mismo, así como los tipos de cultura organizacional diferenciados en el modelo de Cameron y Quinn, permiten realizar la siguiente interpretación:

- Las clases 4 y 6 corresponden a una cultura intermedia entre tipo **Jerárquico** y **Clan**, en la que predomina la percepción de una cultura rígida y de corte tradicional, en general. La clase 4 es un subgrupo más extremo que el 6 y está caracterizado por una mayor proporción de profesores del área de Ciencias Médicas. La clase 6 contiene una proporción baja de profesores de Ciencias Exactas y Naturales. Ambas contienen mayor proporción de profesores de edad superior que la media.
- La clase 5 representa una cultura de tipo **Clan**, aunque con algunas características de cultura rígida, como las referencias al vocablo *control*, y refleja una percepción bastante negativa de la institución. Está caracterizada particularmente por profesores de Humanidades y de Ingeniería y Tecnología, por profesores de edad elevada y por tener menos profesores que la media del área de Ciencias Sociales.
- La clase 2 es la clase más numerosa en cuanto a la parte del corpus que engloba y parece estar, por su posición en el plano y por sus características principales, en sintonía con elementos definitorios de una organización de cultura innovadora o **Adhocrática** con algunos elementos que evocan una cultura de tipo **Clan** en algunas de sus formas textuales características como *reuniones departamentos* o *diálogo*. Está caracterizada particularmente por profesores de Ciencias Exactas y Naturales, de mediana edad y por profesores con una percepción de la cultura organizacional de la universidad totalmente positiva.
- La clase 3 puede interpretarse como un grupo de cultura innovadora o **Adhocrática** donde son característicos términos como *recursos*, *política* o *cursos*. Es un grupo de profesores con una menor proporción de catedráticos y donde únicamente el término consenso evoca una cultura de tipo clan.
- La clase 1 contiene formas textuales como *arriba*, *abajo* y *comunicación* que están asociadas a una cultura rígida, planificada, rival y centrada en la propia institución. A pesar de la posición que ocupa en el plano con respecto al primer eje, esta clase está mucho mejor representada en el segundo y debe interpretarse como un grupo de cultura de tipo **Jerárquico**. Está caracterizado principalmente por estar completamente en desacuerdo con una cultura de clan (que sería la clase 5, en una posición diametralmente opuesta en el plano factorial) y también con una cultura innovadora. Es una clase poco representativa de profesores del área de Ingeniería y Tecnología.

En resumen, los análisis realizados muestran fundamentalmente tres tipos de cultura. La primera, en cuanto al porcentaje del léxico empleado que supone, es tipo **Adhocrática** con elementos de cultura de tipo **Clan** y abarca a las clases 2 y 3, formadas particularmente por profesores de edad media, del área de Ciencias Exactas y Naturales y entre los cuales hay pocos catedráticos. La segunda es una cultura de tipo **Clan**, representada en la clase 5 y particularmente caracterizada por

profesores de Humanidades, Ingenierías y Tecnología y de mayor edad. La tercera, de tipo **Jerárquico**, corresponde a la clase 1 que se entiende como completamente opuesta a la clase 5 y en la que hay pocos profesores del área de Ingenierías y Tecnología. Existen, además, dos clases intermedias (las 4 y 6) de profesores aparentemente descontentos, que piensan que la cultura es poco o nada innovadora, y donde se entremezclan trazas de cultura de tipo **Jerárquico** y **Clan**; estas clases tienen más profesores que la media de Ciencias de la Salud y menos que la de Ciencias Exactas y Naturales, siendo, además profesores de mayor edad (modalidades 4 y 5).

Para finalizar podemos decir que, en cuanto al léxico empleado al proponer mejoras en la universidad, predomina claramente una cultura flexible, que engloba las culturas de tipo **Adhocrática** y de tipo **Clan**, opuestas a una cultura de tipo **Jerárquico** que aparece en mucha menor medida. Asimismo, la cultura de tipo **Mercado** es prácticamente inexistente. Este resultado ha podido ser relacionado, aunque con las debidas cautelas, con distintos tipos de profesorado en cuanto a su edad y áreas de conocimiento en el análisis de la tabla léxica agregada.

CONCLUSIONES

El Análisis de Correspondencias y la posterior clasificación automática basada en los primeros factores del AC muestran la existencia de un tipo de cultura mixta en la universidad objeto de análisis. El Modelo de Valores en Competencia ha servido para establecer un marco de partida al que, sin duda, será difícil encontrar un ajuste claro en aplicaciones empíricas que traten la cultura organizacional de una universidad. Sin embargo, el plano principal del AC de la tabla léxica agregada recoge, bastante fielmente, las dos dimensiones bipolares del modelo de Cameron y Quinn. A la hora de concluir sobre las clases obtenidas debemos manifestar que, en primer lugar, un análisis de clasificación resulta siempre atractivo, aunque sus clases engloben únicamente el léxico empleado. Sin embargo, sabemos que sus resultados son menos robustos que los de un análisis factorial. Las clases obtenidas, por tanto, debemos tomarlas con cautela aunque no resultan contradictorias con otros tipos de análisis que docentes e investigadores experimentados de la cultura universitaria pudieran hacer. Estas clases reflejan en su seno, de nuevo, tipos de cultura mixta, existiendo, sobre todo, trazas de cultura de tipo **Clan** entremezcladas con cultura de tipo **Adhocrático** o innovador y también con culturas de tipo **Jerárquico**.

En relación con los tipos de docentes asociados a los distintos tipos de cultura, cabe señalar que la edad de los profesores y su percepción de cultura innovadora van en sentido inverso, hecho que ha podido verse favorecido por el impulso

investigador de la Universidad Española en las últimas décadas y que ha afectado sobre todo a profesores menores de 50 años. La variable edad señala una trayectoria temporal que marca posiciones tanto en el ámbito que nos ocupa como en otros muchos y este hecho se constata al realizar análisis exploratorios multivariantes (Fernández, Gallastegui, Modroño y Núñez, (2003)).

En lo relativo a la relación de las distintas áreas de conocimiento con los diversos tipos de cultura, podemos decir que la posición de las áreas de Humanidades y de Ciencias Exactas y Naturales no contrastan con los análisis derivados del carácter y trayectoria de estas áreas en la Universidad Española ligadas a la investigación referida, en ocasiones, a ámbitos más locales o más internacionales, respectivamente. En cuanto al área de Ciencias Médicas, podemos pensar que los docentes que han respondido a la encuesta no están, probablemente, implicados en la investigación más avanzada que se realiza en los grandes hospitales. Respecto del área de Ciencias Sociales poco podemos afirmar; no parece representativa de ningún tipo de cultura y únicamente podemos manifestar que el conocimiento adquirido de nuestra participación en otros trabajos, con los datos de esta misma encuesta, nos induce a pensar que se trata de un área heterogénea que engloba a todo tipo de docentes.

En este artículo hemos mostrado una aplicación de cómo las respuestas a preguntas abiertas pueden analizarse estadísticamente mediante el análisis textual para obtener tipologías descriptivas de individuos e incluso relacionarse con las respuestas a las clásicas preguntas de elección múltiple y, por ello, cuantificables. Como conclusión final podemos decir que el análisis textual de preguntas abiertas es un instrumento adecuado para explorar los datos de una encuesta y profundizar en el conocimiento proporcionado por otras técnicas cuantitativas. Al mismo tiempo, puede servir para sugerir futuras investigaciones de aspectos no suficientemente claros. La aplicación efectuada presenta la particularidad de haberse realizado, no como una exploración a ciegas, sino tomando como base el Modelo de Valores en Competencia y los resultados muestran una tipología de personal docente que se ajusta parcialmente al modelo de cultura organizacional propuesto, proporcionando, en todo caso, información sobre el tipo de cultura predominante y la importancia relativa o casi nula de los otros tipos.

REFERENCIAS

- ABASCAL FERNÁNDEZ, E., FRANCO MANERO, M. A. (2002), «Análisis textual de encuestas: aplicación al estudio de las motivaciones de los estudiantes en la elección de su titulación», *Metodología de Encuestas*, Vol. 4, N° 2.
- ARITZETA, A., AYESTARÁN, S. Y BALLUERKA, N. (2003), «Perspectiva sistémica y multinivel del análisis organizacional: una aproximación conceptual y metodológica», *Revista de Psicología Social*, Vol. 18, n° 3.
- BERRIO, M. A. (2003), «An organizational culture assessment using the competing values framework: a profile for Ohio State university extension», *Journal of Extension*, Vol. 41 No. 2, <http://www.joe.org>.
- CAMERON, K.S. Y QUINN, R.E. (1999), «Diagnosing and changing organizational culture: Based on the competing values framework», Addison Wesley, Reading MA.
- DECISIA (2003), SPAD Data Mining and Text Mining, <http://www.decisia.com>.
- ETXEBERRIA, J., GARCIA, E., GIL, E. Y RODRIGUEZ, G. (1995), «Análisis de Datos y Textos», RA-MA, Madrid.
- FERNÁNDEZ AGUIRRE, K. (2003), «Análisis textual: generación y aplicaciones», *Metodología de Encuestas*, Vol. 5, N° 1.
- FERNÁNDEZ AGUIRRE, K., GALLASTEGUI ZULAICA, I., MODROÑO HERRÁN, J. I., Y NÚÑEZ ANTÓN, V. (2003), «Clients' characteristic and marketing of products: some evidence from a financial institution», *The International Journal of Bank Marketing*, Vol. 21, N° 5.
- FERNÁNDEZ AGUIRRE, K., MODROÑO HERRÁN, J. I. Y LANDALUCE CALVO, M. I. (2004), «ACM y STATIS dual ponderado. Dos técnicas complementarias para analizar una visión de la cultura de la Universidad», *Estadística Española*, Vol. 46, No. 156.
- FERNÁNDEZ AGUIRRE, K., MARIEL, P. Y MARTÍN ARROYUELOS, A. (2004), «Analysis of the organizational culture at a public university», *Proceedings in Computational Statistics COMPSTAT-2004*. Ed. Jaromir Antoch, Physica-Verlag, pp. 1013-18.
- GREENACRE, M. J. (1984), «Theory and Applications of Correspondence Analysis» Academic Press, London.
- GREENACRE, M. J. (1987), «Graphical analysis of readership data, using correspondence analysis and clustering», *American Statistical Association, Proceedings of the Section on Statistical Graphics*, pp.40-8.

- HEARST, M. A. (1999), «Untangling Text Data Mining», *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*, University of Maryland, <http://www.sims.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html>.
- 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., MORINEAU, A., Y WARWICK, K. (1984), «Multivariate Descriptive Statistical Analysis», John Wiley & Sons, New York.
- LEBART, L., Y SALEM, A. (1988), «Analyse Statistique des Données Textuelles», Dunod. París.
- LEBART, L. Y SALEM, A. (1994), «Statistique Textuelle», Dunod, París.
- LEBART, L., SALEM, A. Y BERRI, E. (1998), «Exploring Textual Data» Dordrecht. Kluwer. Academic Publisher.
- LEBART, L., SALEM, A. Y BÉCUE, M. (2000), «Análisis estadístico de textos», Milenio
- LEBART, L., MORINEAU, A. Y PIRON, M. (2000), «Statistique exploratoire multidimensionnelle», 3ª edición, Dunod, París.
- MORINEAU, A. (1984), «Note sur la caractérisation statistique d'une classe et les valeurs test», *Bulletin technique du centre de statistique et informatique appliquées*, CISIA-CERESTA 2, pp. 20-28, París.
- RODRÍGUEZ CASTELLANOS, A. (2002), «Gestión del Conocimiento en una Universidad Pública: El Proceso de Investigación-Desarrollo-Transferencia del Conocimiento Científico-Técnico» Ponencia en la Conferencia «Gestión del Conocimiento científico y tecnológico en las Universidades: un reto y una oportunidad», Bilbao
- SALEM, A. (1981), «Signalement et inventaire lexical: textes politiques français de 1973» en Benzécri, J. P. et al.: *Pratique de l'Analyse des Données, T. 3, Linguistique et Lexicologie*, Paris, Dunod, pp. 183-197.
- SALEM, A. (1984), «La typologie des segments répétés dans un corpus, fondée sur l'analyse d'un tableau croisant mots et textes», *Les Cahiers de l'Analyse des Données*, (IX), 4, 489-500.
- SPORN, B. (1996), «Managing university culture: an analysis of the relationship between institutional culture and management approaches», *Higher Education*, 32, pp. 41-61.

EXPLORING TEXTUAL DATA IN THE COMPETING VALUES MODEL FRAMEWORK: AN APPLICATION TO THE KIND OF CULTURE AT THE UPV-EHU

ABSTRACT

This work explores the use of two statistical tools for the analysis of textual data: on one hand, correspondence analysis, which allows for a graphical representation of a contingency table for association research; on the other hand, automatic classification, which permits identifying groups from multivariate observations. The empirical application is about an open question included in a questionnaire over prevalent characteristics of the organizational culture at the University of the Basque Country. This work is particular in its placement in the Competing Values Model framework due to Cameron and Quinn. The results show a teaching staff typology partially adjusting to this proposed model.

Keywords: textual statistics, correspondence analysis, classification

AMS Classification: 62H25