

Redes semánticas en textos periodísticos: propuestas técnicas para su representación¹

MODESTO ESCOBAR

Universidad de Salamanca
modesto@usal.es

Recibido: 10.06.2008

Aceptado: 26.05.2009

En este artículo se presenta una serie de técnicas basadas en los residuos ajustados para representar en modo de red los contenidos de personajes, instituciones, territorios y temáticas, a partir de un conjunto de textos periodísticos seleccionados de un determinado periodo histórico con el propósito de documentar los acontecimientos más relevantes de la transición política española de la dictadura a la democracia.

Ante el análisis de cualquier conjunto numeroso de textos es útil que quien se enfrente a este reto disponga de herramientas que le faciliten la labor. Si estos textos están acompañados de índices, como es el caso de esta base, resulta muy provechoso emplearlos para la selección de aquellos artículos que despierte el interés del analista. Pero, si de lo que se trata, es de estudiar el contenido global de la información obtenida, deberían hacerse análisis más relevantes que el mero recuento de la aparición de los temas indexados. Por ello, aquí se propone una herramienta para detectar la conexión existente entre ellos, al tiempo que se resaltan los más abundantes empleando la metodología de redes, cuyo cometido original, bajo el nombre de sociograma, era encontrar las personas más populares en un pequeño grupo. Pues bien, ahora con la llegada de los ordenadores, que son capaces de analizar matrices millonarias de relaciones, la misma técnica puede emplearse para detectar de modo entrelazado los temas más frecuentes e interconectados en un amplio conjunto de textos.

¹ Parte de las ideas de este artículo se están desarrollando también en el proyecto *La identidad social; nuevos enfoques y metodologías*, financiado por el Ministerio de Educación y Ciencia (SEJ2007-66242). Versiones previas de este texto han sido discutidas en el Departamento de Sociología IV de la Universidad Complutense y en el IX Congreso Español de Sociología de la FES; por tanto el autor agradece los comentarios realizados por diversas personas, entre las que cabe citar a Francisco Alvira, Ignacio Sánchez-Cuenca, Carlos Lozares, Julio Carabaña y Miguel Ángel Quintanilla. Obviamente, ninguno de los errores aquí expuestos debe atribuirse a ellos.

Los antecedentes teóricos del análisis sociológico de redes pueden encontrarse en la sociología formal de Simmel (2002), quien concibe la sociedad como el efecto recíproco de la acción de los individuos. Sin embargo, los primeros estudios empíricos aplicados provienen de los estudios de Moreno (1932) sobre las relaciones personales en grupos y, específicamente, de la elaboración de sus sociogramas, en los que los actores son representados por puntos o nodos y las relaciones entre ellos por líneas que los conectan, es decir, empleó grafos para la representación de las relaciones sociales, lo que dio pie a la construcción de la metáfora de red para ser aplicada a la organización y estructura sociales. Hoy en día, al menos en la comunidad de sociólogos, nadie habla de los sociogramas, pero todos conocen, cuando no emplean recursivamente, el denominado análisis de redes, e incluso hay quienes caracterizan a la sociedad post-moderna, como sociedad-red (Castells 1996).

La idea de construir redes temáticas con las palabras no es original de este trabajo. Verd (2005) denomina redes semánticas a «una representación gráfica de [conceptos con] nodos y arcos interconectados que tiene como objetivo representar un determinado conocimiento» y distingue varias aproximaciones clasificadas principalmente en dos grupos: las de quienes se centran en las *redes de palabras* basadas en la presencia conjunta de los vocablos y la de quienes optan por *redes textuales*, que ponen su énfasis en el modo en el que se relacionan las palabras en los textos. Los segundos basan la conexión de las palabras por su significado; los primeros lo hacen por la regularidad con la que aparecen conjuntamente en un texto. Dentro de la primera tradición pueden encontrarse los trabajos de Doerfel 1998, Van Meter (1999), Danowski (1988 y 1993), Svenja (2006), el equipo de Corman (2001, 2002 y 2003), que propugna el CRA (*Centering Resonance Analysis*)², y Van Atteveldt (2008); mientras que entre los partidarios de análisis más textuales, en los que se representan las conexiones semánticas de naturaleza reticular entre distintos conceptos, se sitúan Carley (1993), Popping (2000, 2003, 2005 y 2007) y Lozares (2000 y 2003).

Las páginas que siguen a continuación muestran una serie de procedimientos para elaborar redes semánticas o textuales, enmarcados dentro de la primera de las dos tradiciones apuntadas.³ Los datos que aquí se exponen para presentar y defender la técnica señalada proceden del *Archivo de la Transición Democrática Española*, constituido por unos 76.000 artículos de prensa recogidos por Juan J. Linz en el periodo comprendido entre 1958 y 1987. En dicha base de datos⁴ se ha aplicado un sistema de categorización de las noticias basado en un te-

² El CRA es un conjunto de técnicas que transforman la estructura de la aparición natural de un texto en una red semántica (Corman *et al.* 2002)

³ Para ello se han utilizado cuatro herramientas informáticas; SQL para la catalogación y disposición de las noticias; SPSS para la tabulación de los descriptores; Excel para la elaboración de los residuos estandarizados y Ucinet/Netdraw para la construcción de las redes.

⁴ Es una base de datos abierta a disposición del público investigador en <http://www.march.es/ceacs/linz>

sauro con más de 10.000 entradas, cuyas líneas fundamentales se han presentado en otro artículo (Knecht, Peach y Fernández, 2006)⁵.

Para tratar de explicar los procedimientos de análisis, se presentan tres secciones. En la primera, se esbozan los resultados globales del archivo de textos periodísticos sobre política en la transición democrática española (1975-1978) recopilados por Linz durante una extensión de más de veinte años y se realiza una breve presentación de los descriptores más relevantes de la base de datos en uso; en la segunda se explica el fundamento estadístico para las representaciones que se exponen en este artículo, y finalmente en la tercera se realiza una propuesta de distintas modalidades de representación con un breve comentario de los resultados obtenidos.

No debo dejar de mencionar que esta herramienta metodológica fue generada en un estudio⁶ sobre la presencia de la cultura científica, tecnológica y de la innovación empresarial en los libros de texto de la Educación Secundaria Obligatoria, enmarcada en los trabajos llevados a cabo por el grupo de Cultura Tecnológica de la Fundación COTEC (1994, 1998, 1999, 2004 y 2005)⁷. El objetivo general de este colectivo es analizar el concepto de cultura tecnológica y explorar cómo los factores culturales afectan a los procesos de innovación y en qué medida este tipo de factores puede ayudar a entender mejor las carencias y las posibilidades de la ciencia en España.

⁵ Este archivo se ha elaborado, a partir del material aportado por el profesor Juan José Linz, por un equipo dirigido por J. Ramón Montero, compuesto por académicos de distintas Universidades como Kerman Calvo (Universidad de Essex), Elisa Chuliá (UNED), Rafael Durán (Universidad de Málaga), Modesto Escobar (Universidad de Salamanca), Ignacio Lago (Universitat Pompeu Fabra) y Víctor Sampredo (Universidad Rey Juan Carlos I), así como por las bibliotecarias asociadas al Instituto Juan March de Estudios e Investigaciones, Martha Peach, Almudena Knecht, Paz Fernández, y por más de una docena de otros documentalistas asociados: Berta del Águila, Ana Aldea Alonso, Cristina Estévez Sánchez, Javier Fernández Fernández, Carolina Fiaño Valverde, Margherita Frigheni, Alicia Hernández Martín, Carmelo Martín de La Calle, Nuria Nieto Mate, Juan Antonio Sánchez Jiménez, Juan Antonio Simón Sanjurjo, Luca Tedde, Nicola Liguori, Raquel Centeno, José Luis Salado, Sandra Aguado, Clara Grau, Dolores Montero e Inés Frade. Ha contado además con subvenciones del Ministerio de Cultura (Dirección General de Cooperación Cultural, 3584C/2002 y 4970C/2003) y del Ministerio de Educación y Ciencia (Plan Nacional para la Promoción General del Conocimiento BSO2001-5350, BSO2002-12733 y SEJ2005-25742-E).

⁶ En el mencionado estudio, aún sin publicar, se analizaron 81 libros de texto correspondientes a las materias obligatorias y optativas que estudian los escolares en los cursos de primero y cuarto de la ESO, publicados por cuatro grandes editoriales de textos escolares de España, para detectar el peso relativo de la cultura científica y tecnológica en el conjunto de contenidos que se transmiten a los estudiantes de la Enseñanza Secundaria Obligatoria.

⁷ Estos trabajos, dirigidos por M. Ángel Quintanilla, versaban sobre la cultura científica y tecnológica en los libros de textos de la educación secundaria obligatoria. En su elaboración participaron además J. Antonio Montero Becerra y Rubén Palacios Sánchez. Otros colaboradores de este proyecto que deben citarse son Ana Cuevas, Joaquín G. Carrasco, Óscar Montañés, Arlyn Orellana, M. Cruz Sánchez y el Centro de Investigaciones Lingüísticas de la Universidad de Salamanca (CILUS), cuya directora es la Dra. María Jesús Mancho Duque.

1. LA BASE DE ARTÍCULOS

Para la presentación de las técnicas que se van a proponer a continuación, se han utilizado los artículos seleccionados de la prensa Española por J.J. Linz durante el periodo 1958-1987. Como quiera que no se trata de una selección aleatoria de textos, conviene en primer lugar caracterizarlos con el fin de que el lector tenga una apreciación certera del conjunto de noticias con el que se está trabajando.

Ya ha quedado dicho que se está trabajando con más de 76.000 recortes procedentes en su inmensa mayoría de la prensa española. Más del 90% de estas piezas de información fueron extraídas de 8 periódicos enumerados a continuación por orden del número de artículos recogidos en cada uno de ellos: ABC, Diario 16, Informaciones, Ya, El País, Pueblo, Arriba y El Alcázar, como se ve todos ellos editados en Madrid. También hay incluidos periódicos de otra localidades, de Barcelona principalmente, pero en bastante menor proporción, como es el caso de La Vanguardia (403 recortes) y El Periódico de Cataluña (106).

Tabla 1. *Contenido de la base de datos por fuente*

	Frecuencia	%
ABC	20124	26,4
Diario 16	12223	16,0
Informaciones	12209	16,0
Ya	8375	11,0
El País	5141	6,7
Pueblo	4887	6,4
Arriba	4795	6,3
El Alcázar	2702	3,5
Resto	5781	7,6
Total	76237	100,0

La mayor parte de los recortes que se realizaron son crónicas o noticias; pero también se han incluido bastantes artículos de opinión redactados tanto por plumas habituales del periódico como por colaboradores esporádicos. Otros estilos periodísticos recogidos con frecuencia mayor que el millar son las breves noticias, los editoriales, las entrevistas y los reportajes.

Todos estos recortes fueron clasificados por su productor en 300 cajas temáticas, algunas de las cuales, 32 en concreto, fueron divididas temporalmente por la gran cantidad de artículos incluidos. Realizando un recuento de los recortes de estas cajas múltiples, los cinco grandes temas son: las elecciones, el País Vasco, el orden público, artículos de fondo y vida política, todos ellos con más de 3000 unidades recortadas. El resto de cajas incluyen, además de Cataluña, temáticas, personalidades e instituciones, entre las que también se encuentran los partidos políticos más importantes de la transición (Alianza Popular, PCE, PSOE y UCD).

Tabla 2. *Contenido de la base de datos por tipo de recorte*

	Frecuencia	%
Crónica, noticias	53345	70,0
Opinión (colaboradores habituales)	6998	9,2
Opinión (firmas invitadas)	1959	2,6
Nota simple, Breve	6183	8,1
Editorial	2776	3,6
Entrevista	1773	2,3
Reportaje	1123	1,5
Anuncio, Esquela, Fotografía	789	0,8
Cartas al Director	577	0,8
Crítica	108	0,1
Total	76237	100,0

Tabla3. *Contenido de la base de datos por carpetas***Carpetas de Linz**

Elecciones	8895	11,7	Alianza Popular	1375	1,8
Vascos	7210	9,5	Franco	1358	1,8
Orden público	4503	5,9	UCD	1208	1,6
Artículos	3189	4,2	Economía	1175	1,5
Vida política	3180	4,2	Agricultura	1060	1,4
Asuntos exteriores	2576	3,4	Juan Carlos I	888	1,2
Sindicatos	2215	2,9	Monarquía	533	0,7
Iglesia	2180	2,9	Chistes	387	0,5
Educación	1948	2,6	Guerra Civil	262	0,3
PSOE	1821	2,4	Mujeres	189	0,2
Trabajo	1766	2,3	Justicia	146	0,2
Catalunya	1742	2,3	Sanidad	142	0,2
Cortes	1723	2,3	Colegios Profesionales	74	0,1
PCE	1639	2,1	Consejo de Ministros	74	0,1
Gobierno	1622	2,1	Juventud	51	0,1
Extrema derecha	1545	2,0	Resto	18164	23,8
Regional	1397	1,8			
			Total	76237	100,0

Para una total descripción del material que se analiza en este artículo, queda por referir la clasificación por descriptores realizada por un equipo de personas que se encargaron de acompañar a los artículos de su información contextual, incluyendo datos como el periódico, la fecha, el tipo de noticia, así como el contenido. Para esto último, se emplearon unos 10.000 descriptores distintos, entre los cuales 442 aparecieron en más de 200 artículos diferentes. Con objeto de simplificar la descripción de tanta información, se restringió el análisis a aquellos con una frecuencia por encima del mencionado límite, teniendo en

cuenta que tal cantidad implica una frecuencia de aparición superior al 2,5 por mil. En la tabla 4, aparecen los veinte descriptores más frecuentes, en tanto que en el apéndice se listan por orden alfabético aquellos con más de doscientas apariciones.

Tabla 4. *Descriptor de aparición más frecuente*

20 descriptores de aparición más frecuente

Descriptor	Frecuencia	%	Descriptor	Frecuencia	%
Posición política	5119	6,7	Manifestaciones	2913	3,8
Elecciones de 1977	4775	6,3	Terrorismo	2797	3,7
Partido Socialista Obrero Español	4184	5,5	Partidos políticos	2761	3,6
Gobierno Suárez	3925	5,1	Atentados	2757	3,6
Crítica política	3762	4,9	Resultados electorales	2611	3,4
Declaraciones	3677	4,8	Unión de Centro Democrático	2594	3,4
ETA	3330	4,4	Estrategia política	2592	3,4
Reforma política	3228	4,2	Campaña electoral	2571	3,4
País Vasco	3092	4,1	Negociación política	2563	3,4
Madrid	2927	3,8	Partido Comunista de España	2499	3,3

Ahora bien, a diferencia de las anteriores tablas de frecuencia, los descriptores se caracterizan por no ser mutuamente excluyentes, esto es, en un mismo artículo pueden coexistir varios de ellos. Esto no era así en el caso de las variables ya descritas, puesto que un artículo sólo aparece en un periódico, sólo se clasifica en un género y sólo se clasifica en una carpeta. Por ello, con el fin de estudiar cómo se conjugan en las noticias los grandes temas de la transición, se va a analizar el hecho de que dos determinados descriptores se ubiquen en un mismo artículo. Sin embargo, no han sido desarrolladas hasta la actualidad herramientas que permitan realizar análisis automáticos de apariciones conjuntas de valores múltiples, esto es, de atributos no excluyentes de casos⁸.

2. LA TÉCNICA DEL ANÁLISIS DE COAPARICIONES

La construcción de una red de descriptores en un corpus periodístico tiene como objetivo la representación automática tanto del conjunto de actores y temáticas importantes por frecuencia y sentido (nodos o vértices), como la de la proximidad de sus significados a través de su aparición conjunta en las unidades periodísticas, que son los artículos. Es decir, en la práctica, se parte de la suposición de que existe una especial conexión entre los descriptores que aparecen en el mismo artículo.

⁸ La excepción a esta afirmación puede ser determinadas variantes del análisis factorial de correspondencias. Consúltese a este respecto el artículo XX Metodología de Encuestas XX.

El procedimiento que se propone a continuación para elaborar una red conceptual automática consta de los siguientes pasos:

En primer lugar se selecciona un conjunto de artículos, a partir del cual se construirán esas redes. Es importante esta selección, porque en función de ella el resultado puede ser muy diferente. Si recopilamos temas de política interior, como éste es el caso, los descriptores que se relacionen entre sí tendrán que ver con las vicisitudes de las instituciones de la esfera política. Si se seleccionan las noticias culturales, la constelación de descriptores tendrá que ver con los distintos géneros, sus autores, críticos y divulgadores. Se pueden realizar muchas hipótesis de los distintos resultados a los que puede dar lugar este procedimiento; pero en este artículo me limitaré a presentar una sola prueba aplicada a la base de datos de los artículos políticos de la transición seleccionados por Linz. En otros trabajos, como los mencionados en la introducción, se ha aplicado una técnica semejante con libros de texto que los estudiantes españoles han de cursar en la Enseñanza Secundaria Obligatoria, pero en lugar de emplear descriptores como material de análisis, se emplearon conceptos del vocabulario científico.

El segundo paso es el de la selección de los descriptores. En este caso, se puede proceder básicamente de dos maneras: Una consiste en disponerlos de antemano, cuando se sabe muy bien lo que se va buscando y se posee un corpus de vocablos del que queremos conocer sus conexiones. La otra pasa por su constitución automática. El modo más simple de hacer esto último es mediante un recuento de las palabras en el texto, previa omisión claro está de las partículas idiomáticas (artículos, preposiciones, conjunciones, adverbios, verbos auxiliares y adjetivos determinativos). Evidentemente, se podría hacer estrategias mixtas consistentes tanto en una selección y reorganización de los obtenidos por el segundo de los procedimientos; lo que puede implicar bien la eliminación de aquellos que no tengan que ver con la materia que se desea investigar, bien la agrupación de conceptos con similares significados o palabras con raíces semejantes; como en confeccionar una lista provisional de descriptores que se va ampliando a medida que se encuentran artículos o temáticas no recogidas en el repertorio inicial.

Una vez que se dispone de una colección de los descriptores centrales del análisis, en el tercer paso debe procederse a su aplicación a los artículos. Para ello sería muy útil hacer uso de procedimientos automáticos presentes en la mayor parte de programas de análisis de contenido.

Ahora bien, hay que tener en cuenta que para proceder con la codificación cada descriptor debe estar referido a una unidad de aparición. Hay para ello varias alternativas:

- 1) Una unidad amplia podría ser un *ejemplar de diario*. Esta no es recomendable en la medida en que implicaría una gran cantidad de coocurrencias que debilitaría el análisis de proximidad de los conceptos.
- 2) Una unidad óptima sería el *artículo*. Para estos menesteres, se considera la más oportuna en cuanto que este conjunto tipográfico ha de reflejar una unidad de sentido para el escritor.

- 3) Otras posibilidades serían el *párrafo* y la *oración*, en tanto que unidades mínima de sentido y comunicación. Al utilizar estas opciones se limitarían muchísimo las coocurrencias, a cambio de aumentar la aparición de los descriptores y el trabajo de asignarlos. Téngase en cuenta que un determinado tema que apareciera en un artículo tres veces en distintas oraciones contaría el triple según este proceder.

Dispuesto para cada ejemplar, artículo, párrafo u oración un vector con variables dicotómicas (presencia o ausencia) de todos los descriptores asignados, es relativamente fácil construir una matriz de coocurrencias, obteniendo el cruce de estas variables consigo mismas (véase la Tabla 5 A con un ejemplo referido sólo a tres descriptores).

Tabla 5. Ejemplo de matriz de coocurrencias

A. Matriz de frecuencias					B. Matriz de frecuencias teóricas				
	Actos elect.				Actos elect.				
	A.P.	Madrid		A.P.	Madrid		A.P.	Madrid	
Actos electorales	1325	108	139	1325	23	30	51		
AP	108	1753	75	1753	30	40	67		
Madrid	139	75	2927	2927	51	67	112		
	1325	1753	2927	76237					

C. Matriz de residuos estandarizados					D. Matriz de residuos ajustados				
	Actos elect.				Actos elect.				
	A.P.	Madrid		A.P.	Madrid		A.P.	Madrid	
Actos electorales	271,3	14,0	12,4	276,1	14,3	12,7			
AP	14,0	269,8	0,9	14,3	276,1	1,0			
Madrid	12,4	0,9	265,5	12,7	1,0	276,1			

Tras la obtención de este cruce, las diagonales de la tabla muestran el número de unidades en las que se encuentra cada concepto y el resto de casillas son simétricas con respecto a la diagonal, puesto que el número de veces que aparece un descriptor A junto con otro B es el mismo que el de veces que ocurre el B en compañía del A.

Sobre esta matriz de ocurrencias, cuyos marginales no se obtienen sumando las frecuencias, sino reflejando sólo la casilla diagonal, no se pueden calcular las frecuencias esperadas con la típica fórmula empleada para el cálculo de χ^2 . Lo que debe hacerse es que en el cálculo de los marginales de la tabla dejen de contarse los recuentos no diagonales, ya que éstos en realidad no implican una aparición distinta, sino la misma contada en la diagonal, pero con otro elemento. Por ello, las fórmulas para obtener los nuevos marginales (c_i y c_j) son las siguientes:

$$c_i = f_{ii}$$

$$c_j = f_{jj}$$

De este modo, lo único que se realiza es la eliminación de las coapariciones en los marginales, ya que en la diagonal (f_{ii} o f_{jj}) figuran las veces que aparece un descriptor, aunque esté sin acompañar por otro término similar. Y la suma respectiva de estos específicos marginales de fila o columna totaliza el conjunto de apariciones de las categorías en estudio presentes en el texto analizado. Como puede apreciarse, no es éste el caso de la tabla 5 A, porque sólo es un fragmento, ya que se han omitido de ella los casi 10.000 otros descriptores usados en la investigación. En cualquier caso, el marginal total (p) deber reflejar el total de unidades analizadas y no la suma de las apariciones de las categorías. Dicho más específicamente, debería expresar el conjunto de unidades a las que se les ha asignado al menos un descriptor.

A partir de estas frecuencias marginales específicas, que coinciden con la frecuencia diagonal, podría calcularse la frecuencia esperada de cada coaparición de una pareja de conceptos dividiendo sus marginales por el número total de unidades (p), artículos en este caso, en las que aparezca al menos un descriptor⁹. El marginal partido por esa última cantidad es la probabilidad de que una categoría aparezca en un artículo codificado. Por definición, si los dos descriptores son independientes, la probabilidad de que ocurra una coaparición es igual al producto de sus dos correspondientes probabilidades:

$$\frac{c_i}{p} * \frac{c_j}{p} = p_{ij}^{\bullet}$$

De ahí que para obtener la frecuencia esperada de una coaparición en el caso de que los descriptores fueran independientes, es necesario multiplicar dicha probabilidad por el número de párrafos en los que haya presencia de categorías:

$$f_{ij}^{\bullet} = p_{ij}^{\bullet} * p = \frac{c_i * c_j}{p}$$

A partir de la obtención de esas coapariciones teóricas (Tabla 5 B), pueden calcularse los residuos, restando las frecuencias teóricas a las empíricas. Si éstas son mayores que aquellas, indicaría que el par de descriptores en cuestión tiene una tendencia a aparecer conjuntamente; mientras que, si es menor, implicaría es

⁹ También se podría hacer que p fuera variable para cada par de descriptores, estableciendo que su tamaño sea las veces que aparezca en uno u otro, en lugar del número de unidades en las que hay al menos un descriptor. Ello produciría consistentemente valores esperados de coaparición mayores. El autor prefiere la opción expuesta en el texto porque es constante a lo largo de toda la matriz y porque no genera una tabla con una casilla de frecuencia igual a 0, en el caso de que un par de descriptores no se hallen juntos en ningún artículo.

que las calificaciones en cuestión no suelen presentarse juntas, es decir, lo hacen por debajo de la probabilidad de que fuera al azar.

Puede verse mucho más claro mediante un ejemplo. Imaginemos un conjunto con 100 artículos analizados. En 20 de ellos se toca el tema de las *elecciones* y en 15 de ellos aparece el descriptor *parlamento*. Ocurre además que en 10 ocasiones se encuentran en el mismo artículo ambas palabras. En consecuencia con estos datos, el valor esperado de apariciones conjuntas es de sólo 3: $(20 \times 15 / 100)$. El que haya 10 indica que ambos conceptos ocurren por encima de lo esperado. El residuo resultante es positivo con un valor de 7.

Los residuos se pueden estandarizar e incluso ajustar mediante su división por la raíz cuadrada de la frecuencia teórica y, en el segundo caso, también por la desviación típica, tal como propusiera Haberman (1979)¹⁰. Siguiendo con el ejemplo antes expuesto, a partir del residuo bruto, el estandarizado sería de 4,0 y el ajustado de 4,9. Pueden verse ejemplos de estas dos medidas sobre los tres descriptores seleccionados de la Tabla 5 en sus secciones C y D.

Trabajando con los residuos ajustados, pueden aplicarse diversos análisis en forma de red, empleándolos como medida de la relación entre los conceptos. Dependiendo del tamaño del conjunto de categorías disponibles habrá que tomar decisiones sobre qué procedimiento emplear. De modo general, se propone omitir todo residuo que no sea positivo significativo con hipótesis alternativa unidireccional (residuo inferior a 1,64 o 2,33 según se trabaje con un nivel de significación del 5% o del 1%)¹¹ y eliminar también los residuos de una categoría consigo misma, ya que con un tamaño suficiente de la muestra siempre han de ser significativos¹².

Finalmente, también cabría trabajar con medidas más próximas a las correlaciones. En este ámbito podrían utilizarse todas aquellas medidas de similitud empleadas en el análisis de conglomerados para variables dicotómicas¹³. Sin embargo, en línea con las anteriores medidas, aquí se propone dividir el residuo

¹⁰ La fórmula del residuo estandarizado es $\frac{(f_{ij} - f_{ij}^*)}{\sqrt{f_{ij}^*}}$ y la del ajustado, $\frac{(f_{ij} - f_{ij}^*)}{\sqrt{f_{ij}^*(1 - c_i/p)(1 - c_j/p)}}$.

¹¹ Esta decisión estadística parte del falso supuesto de que los artículos han sido extraídos al azar, puesto que en este caso concreto no es así, ya que han sido seleccionados «intencionalmente» por un determinado investigador. Alternativamente, puede concebirse este conjunto de más de 76.000 artículos analizados como un conjunto poblacional que se está comparando con un modelo teórico, en cuyo caso las pruebas estadísticas lo que están indicando es el grado de discrepancia de la realidad con el supuesto de independencia de aparición de los descriptores.

¹² Otra operación posible es direccionar los residuos, de modo tal que sólo se consideren los que vayan de la categoría menos frecuente a la más numerosa en apariciones. De este modo, cuando se realice un grafo de red, en lugar de aparecer una flecha bidireccional, aparecerá una sola, que va desde la categoría menor a la mayor en importancia de aparición.

¹³ Entre ellas se encuentran los índices de Russell-Rao, Jaccard, Rogers y Tanimoto, Q de Yule, Ochiai, Sokal y Sneath y Kulczynski. Véase algunas de ellas en Albenderfer y Blashfield (1984) o en Hernández (2001).

ajustado por su valor máximo, que es la raíz cuadrada del número de unidades analizadas.

De este modo, a partir del conjunto de residuos ajustados expuestos en la Tabla 5 D, se podría dibujar un gráfico, que indicara una asociación entre *Actos Electorales* y *Alianza Popular*, esperable en la medida en que es un partido político que compite en las elecciones; a la vez que entre el primer descriptor y el denominado *Madrid*, probablemente debido a que la prensa analizada es fundamentalmente madrileña y a que la noticia de un acontecimiento electoral debe acompañarse del lugar donde se celebra. Sin embargo no se observan conexiones significativas entre *Alianza Popular* y *Madrid*, al tratarse de un partido de alcance nacional, por muy centralista que se caracterizara su ideario. El tamaño de los nodos puede dibujarse proporcional al número de apariciones de cada descriptor, con lo que se obtiene una apreciación de su importancia en la base de artículos.

3. RESULTADOS Y COMENTARIOS DEL ANÁLISIS DE LA BASE DE DATOS

Aplicando este sistema a todos los descriptores de la transición española con más de 400 apariciones en esta base de datos, que como puede comprobarse en el listado de la Tabla I del apéndice son 232, se ha construido la matriz de residuos ajustados.

Este estadístico, obtenido para cada par de descriptores, es la base para construir un mapa reticular de las conexiones estadísticamente significativas entre ellos. En este mapa, cada nodo (punto) es un descriptor y cada línea, una relación significativa entre ellos. De este modo, dos temas muy vinculados entre sí aparecen conectados. Aplicando además las técnicas de redes se pueden dibujar espacialmente los nodos, representando en el centro aquellos con más vinculaciones, o dicho con la terminología de las redes sociales, con más elecciones o conexiones, dejando en la periferia aquellos temas, que por específicos, tienen muy bajo número de conexiones significativas con otros descriptores, e incluso dejando de representar aquellos que no están vinculados de este modo con ningún otro. Complementariamente, los tamaños de los nodos se dibujan proporcionalmente a la aparición de cada nodo, para que pueda realizarse una evaluación de qué importancia o frecuencia poseen en la base de datos. De este modo, se distinguen los temas *importantes* por su frecuencia, de los temas *transversales*, que son aquellos conectados con muchos otros¹⁴.

Sin embargo, siguiendo este procedimiento con 232 elementos, se obtendría un mapa completamente ilegible si se dibujaran todas las relaciones significati-

¹⁴ También debe evitarse interpretaciones de conexión por el hecho de estar cerca dos categorías a menos que estén unidas por una línea. Sobre todo en los extremos, dos nodos cercanos no indican proximidad de significados

Secundariamente, se encuentra un conjunto de tres descriptores: a) *Cataluña* con su *Gobierno Autónomo* y su primer presidente, *Josep Tarradellas*; b) el *Partido Comunista de España* conectado con *Santiago Carrillo* y su *Legalización* c) las *Relaciones bilaterales*, vinculadas con los *Estados Unidos* y el *Gobierno de los Tecnócratas*, y d) *Coalición Popular* con otros dos descriptores: el *Centro Democrático y Social*, junto con las *Elecciones de 1986*.

Finalmente, aparecen un conjunto de relaciones binarias que vienen a indicar los grandes temas políticos de la transición democrática española reflejados en la prensa. Temáticamente pueden agruparse en a) Internacionales: *Negociación internacional* y *Comunidad Económica Europea*, por un lado, *Eurocomunismo* y *Unión Soviética*, además de la *Opinión internacional* y su *Repercusión internacional*, por el otro; b) Estrictamente políticos como *Política gubernamental* junto con *Consejo de Ministros*, *Unidad política* vinculada con *Socialistas*, y *Enrique Tierno Galván* con el *Partido Socialista Popular*; c) Sociales: *Huelgas* con *Reivindicaciones políticas*, *Amnistía* con *Presos políticos*, *Unión General de Trabajadores* con *Comisiones Obreras*; d) Electorales: *Elecciones de 1977* con *Candidatura electoral*, por un lado, y *Encuestas* con *Previsiones electorales*, por el otro; e) Violencia política: *Terrorismo* con *ETA* y *Secuestro* con *GRAPO*. Por último, queda por reseñar la pareja de descriptores que vincula el *Humor gráfico* con la *Sátira política*.

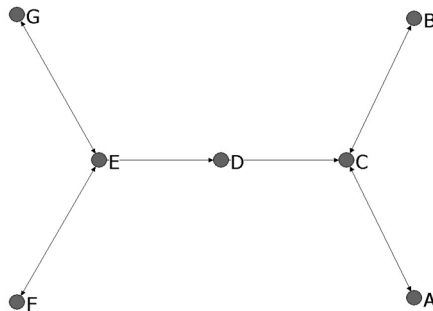
Podríamos denominar a este gráfico de *significación*, en la medida en que resalta las conexiones más significativas del conjunto. El resultado esperable es la aparición de aquellos descriptores o temas más estrechamente relacionados entre sí: bien por una propia coyuntura histórica (*Referendum* con *Reforma política*, con *Constitución de 1978* y con *OTAN*), bien porque sean sinónimos (*Encuestas* y *Previsiones electorales*).

2) La segunda propuesta implica otra selección de los nodos. Además de emplear el criterio de significatividad, éste se mantiene en el 5% o el 1%, se realiza una selección de sólo aquellos descriptores con un alto grado de intermediación (*Betweenness*). Este es uno de los conceptos más utilizados en la literatura que analiza redes de palabras. Se trata de una medida que indica para cada sujeto (en este caso, descriptor) su capacidad de conectar grupos de sujetos (descriptores). Se comprende mejor este índice, cuya medida propuso Freeman (1977), mediante una ilustración. En el Gráfico 2, los puntos C y E son los más centrales, es decir, aquellos con más conexiones con el resto. Sin embargo el poder de intermediación de D es tan grande como el de los anteriores conceptos, ya que es un concepto necesario para vincular A, B y C, por un lado, con E, F y G. El *Centering Resonance Analysis* considera esta medida como un indicador del

bría que multiplicar la significación del residuo por el número de pruebas realizadas ($221 \times 220 / 2 = 24310$). De esta forma el valor crítico de la distribución normal (una sola cola) sería de 4,61 y 4,93 para significaciones respectivas de 0,05 y 0,01. Sin embargo, en este contexto se ha sobredimensionado el umbral para obtener sólo las coaparaciones más sobresalientes.

valor estructural de las palabras, de los descriptores en el presente caso. (Corman et al. 2002).

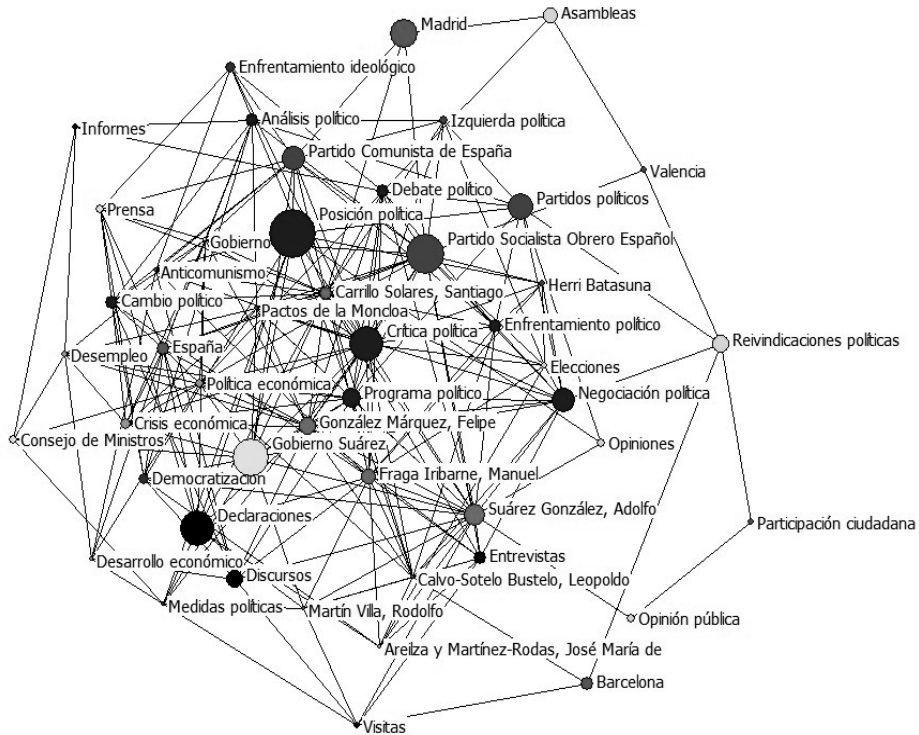
Gráfico 2. Representación de la intermediación



Por estas características de la intermediación, se propone un gráfico en el que sólo se representen los nodos (descriptores) con mayor valor estructural. Se proponen aquellos que tengan un valor normalizado superior a aquella cantidad que haga legible el gráfico, 0,28 con los datos comentados. De acuerdo con este criterio y, siguiendo representando los nodos con un tamaño proporcional a su frecuencia, se obtiene en el Gráfico 3 la representación de los más de 76000 artículos analizados.

El principal resultado de aplicar este nuevo criterio es que se representan sólo aquellos conceptos con más capacidad intermediadora. Desaparecen los dos conceptos anteriormente más conectados (*Referendum* y *Audiencias*) por no poseer esta propiedad. Sin embargo, aparecen conceptos centrales en el estudio de la transición, como son los líderes y partidos políticos, como se deduce de un examen atento de los nodos representados. Incluso, puede realizarse una agrupación temática en el gráfico, aplicando el mismo color a aquellos descriptores que, estando cercanos espacialmente, se consideren en la misma categoría a juicio del investigador. De este modo, el dibujo de la red del Gráfico 3 parece haber aglutinado conceptos en las siguientes categorías:

Además de las dos acabadas de mencionar: a) líderes políticos (*Adolfo Suárez, Felipe González, Manuel Fraga, Santiago Carillo, José M de Areilza y Rodolfo Martín Villa*), y b) partidos políticos, curiosamente de izquierda, (*PSOE, PCE y HB*); también aparecen descriptores referidos a: c) procesos políticos (*Posición política, Crítica política, Negociación política, Crítica política, Programa político, Debate político, Análisis político, Enfrentamiento ideológico, Enfrentamiento político y Democratización*), d) actividades políticas (*Declaraciones, Discursos, Entrevistas, Medidas políticas, Visitas, Informes*), e) instituciones (*Gobierno Suárez, Consejo de Ministros, Prensa*), f) ciudades (*Madrid, Barcelona y Valencia*), y las esferas f) ciudadana (*Asambleas, Reivindicaciones políticas, Opinión pública, Participación ciudadana*) y g) económica (*Política económica, Crisis económica, Desempleo, Pactos de la Moncloa*).

Gráfico 3. Red de descriptores intermediadores

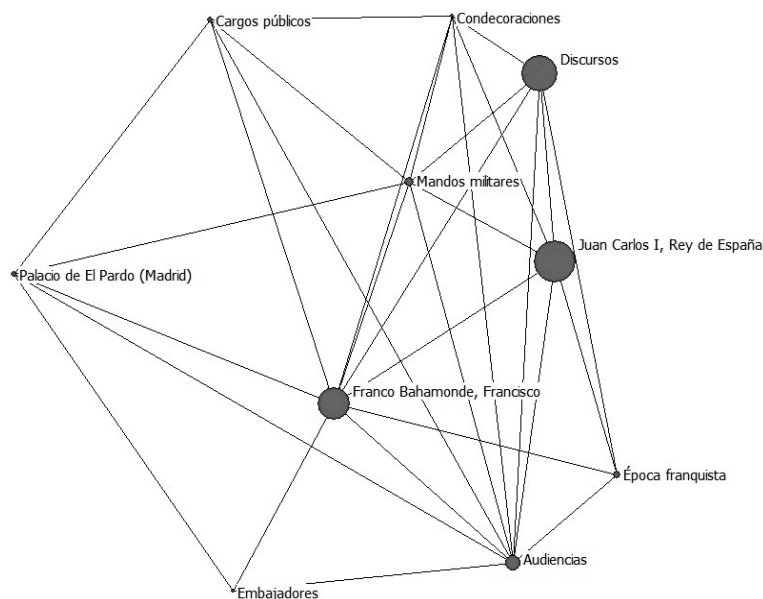
A este segundo tipo de gráfico lógicamente puede denominársele de *intermediación*, en la medida en que resalta aquellos descriptores o temáticas que poseen mayor capacidad de conexión en el conjunto analizado. Sin duda, el proceso de selección de puntos es mucho más rico temáticamente que en el caso de los gráficos de significación.

3) La tercera alternativa para analizar las conexiones entre descriptores consiste en elaborar *ego-redes*, es decir, representaciones que sólo incluyan los descriptores vinculados con uno elegido como relevante, más los vínculos entre ellos. Aquí se propone el uso de la palabra *autorede*, puesto que no se supone un ego en los descriptores. Para ver el comportamiento de estas redes, se ponen los ejemplos de tres descriptores estelares de los anteriores procedimientos: *Audiencias*, *Referendum*, y *Posición política*. De este modo, lo que se logra es aumentar la densidad sobre los descriptores más entrelazados obtenidos con los mecanismos precedentes.

La *autorede* correspondiente a las *Audiencias* muestra una estrecha asociación con las dos figuras de Jefe de Estado durante la transición, *Francisco Franco* y *Juan Carlos I*. Como apareció en la primera red representada, el descriptor correspondiente a los *Mandos militares* también es importante. La novedad de este

tipo de gráfico es que permite mostrar asociaciones también significativas entre descriptores que no fueron mostradas antes para evitar la saturación del gráfico. De este modo, además de los ya tres mencionados descriptores presentan conexiones significativas con las *Audiencias*, los *Discursos*, las *Condecoraciones*, el *Palacio de El Pardo*, los *Embajadores* y la *Época Franquista*. También deber resaltarse que el descriptor correspondiente al dictador tiene tantas relaciones con los otros descriptores como el descriptor (*Audiencias*) que sirve para la confección de la selección de los nodos presentes en el gráfico.

Gráfico 4. *Autored del descriptor audiencias*



Frente al mapa correspondiente a las *Audiencias*, puede contrastarse el de *Referendum*, que ostenta muchas más relaciones significativas con descriptores de distinto signo.

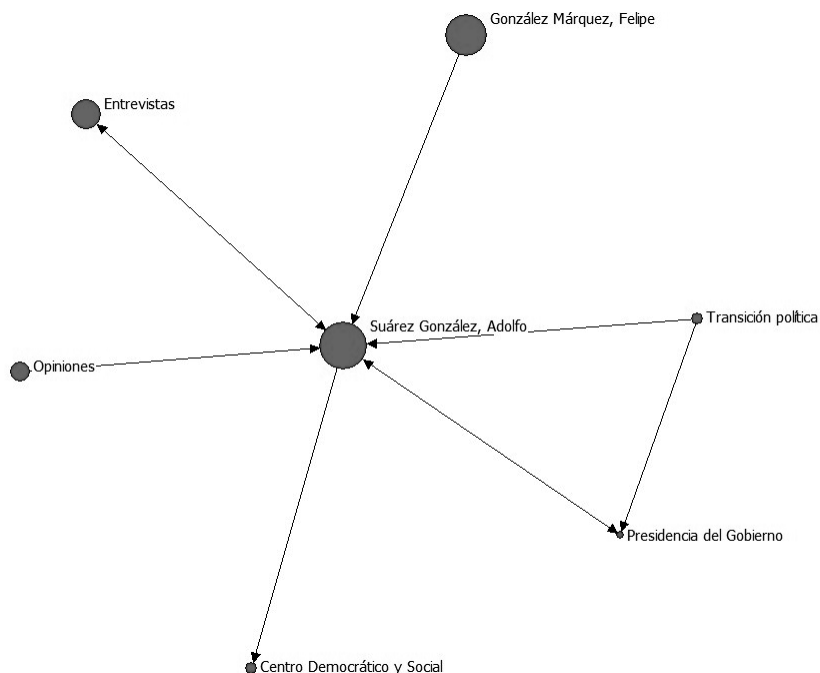
Recuérdese que en el primer gráfico se conectaba la *Ley de la Reforma Política* con la *Constitución de 1978* y con la *OTAN*, además de aparecer la propia *Reforma Política*. En este nuevo gráfico también aparecen los *Gobiernos de Suárez* y del *PSOE*, *Cataluña*, el *País Vasco* y *Andalucía*, los *Estatutos de Autonomía*, y la *Autonomía política*, así como procesos como la *Posición política*, la *Oposición política*, la *Participación política*, la *Participación ciudadana*, el *Sistema electoral*, el *Comportamiento electoral*, la *Campaña electoral*, la *Propaganda política*, la *Propaganda electoral*, *Radio Televisión Española*, las *Previsiones electorales*, las *Encuestas*, los *Resultados electorales*... Aparecen además algunos descriptores de algo más difícil justificación como *Rodolfo Martín*

política. Por último, queda un par de descriptores de la esfera internacional (*Opinión internacional* y *Repercusión internacional*) y otro par más difícil de aglutinar, pero no por ellos menos importantes, como el *Gobierno Suárez* y *Radio Televisión Española*.

4) La autoredes pueden, si es necesario, simplificarse con la introducción de una limitación que haga disminuir las conexiones en el gráfico, a fin de que queden diáfanas las relaciones más importantes entre los descriptores. Hasta ahora se han seleccionado aquellas relaciones que salen significativas. Un procedimiento distinto, que daría lugar a una *red de preferencias*, emulando la metodología de los sociogramas, sería la selección para cada descriptor de las tres relaciones más importantes, siempre y cuando sean significativas¹⁶.

De este modo, del nodo central seleccionado saldrán tres flechas hacia los tres nodos más significativos, si los hay; pero hasta él pueden llegar menos o más, según la importancia relativa con que aparezcan con otros descriptores. Como el atento lector habrá advertido, éste es el único gráfico en el que las flechas pueden ser uni o bidireccionales.

Gráfico 8. *Autored simplificada del descriptor Adolfo Suárez González*



¹⁶ Con ello se trata de realizar el mismo artificio que se logra en el sociograma cuando se pide a los sujetos que seleccione a las tres personas preferidas para una relación de tipo instrumental o afectivo.

Como ejemplo, se expone en el Gráfico 8 la autorede simplificada correspondiente a Suárez. En ella se advierten las tres relaciones que salen del descriptor seleccionado: *Presidencia del Gobierno*, *Centro Democrático y Social* y *Entrevistas*. De modo complementario, se ven otros descriptores para los que este personaje tiene la conexión más fuerte. Se trata de *Felipe González*, *Transición Política* y *Opiniones*.

4. CONCLUSIONES

Las técnicas de análisis de redes pueden emplearse para la representación de las relaciones y conexiones entre una serie de ideas o conceptos presentes en un conjunto de textos. Sin embargo, como en la mayor parte de análisis exploratorios estadísticos de búsqueda de estructuras, hay modos muy diversos de realizar el proceso selectivo que destaque las relaciones que se producen efectivamente en un conjunto de textos.

En este artículo se ha propuesto un procedimiento empírico, que se sustenta en un mínimo conjunto de presupuestos objetivos sobre la relevancia de uno u otro concepto. Por ello, se han empleado los siguientes criterios:

- a) Una aproximación cuantitativa de la codificación de los textos.
- b) Una codificación automática, sólo en aquellos casos en los que se hayan seleccionado previamente los descriptores de trabajo.
- c) La selección de un criterio estadístico (los residuos ajustados) para determinar qué descriptores coaparecen en los artículos.
- d) La utilización del artículo como unidad de coaparición, aunque en otros contextos pueda ser diferente.
- e) Por último, se han sugerido cuatro modos de representación: el primero opta por representar mediante la *red de significaciones* todas las relaciones significativas; el segundo propone una representación, llamada *red de intermediación*, en la que sólo queden reflejados las relaciones significativas de aquellos conceptos con un alto grado de intermediación; el tercero consiste en una reducción de la maraña de relaciones empleando las *autoredes*, esto es, la representación de un solo concepto con todos los que se relacionan con él, junto con sus interrelaciones, y finalmente, se pueden simplificar aún más las autoredes, seleccionando solamente los tres vínculos más fuertes para cada descriptor, siempre y cuando sean significativos (*red de preferencias*).

Las ventajas de estos procedimientos, de naturaleza principalmente inductiva, residen principalmente en que son capaces de proporcionar una representación visual del conjunto de relaciones existentes entre una serie de categorías. En este sentido, en su finalidad, podrían semejarse al análisis factorial de correspondencias. Sin embargo, se diferencia de éste en que no emplea la métrica de χ^2

para el conjunto de categorías, sino a pares, por un lado; y en que las categorías con más asociación con el resto no son las que aparecen en el extremo del gráfico, sino en el centro. Las categorías que aparecen en el centro del gráfico del análisis de correspondencias son aquellas sin ninguna asociación; mientras que estas categorías desconectadas desaparecen en los gráficos de redes por la simple razón de estar aisladas. Ambas diferencias le otorgan a este proceder una simplicidad interpretativa de la que carece el ya consagrado método.

Obviamente, todos estos procedimientos gráficos que se han propuesto aquí no sólo sirven para el análisis de artículos de prensa. Ya se han aplicado con cierto éxito en el análisis de contenido de textos escolares y para la representación de las respuestas a las preguntas múltiples, como puede ser el caso, de la prensa leída por una población. Pero también podrían utilizarse para el análisis de manifiestos políticos, para la representación de autoimágenes que tienen las personas de sí mismas o para un sinfín de análisis de un conjunto múltiple de categorías de naturaleza no excluyente.

BIBLIOGRAFÍA

- AIBAR, E. y M. A. QUINTANILLA (2002): *Cultura Tecnológica. Estudios de Ciencia, Tecnología y Sociedad*. Barcelona, Horsori Editorial.
- ALDENDERFER, M. S. y R. K. BULASHFIELD (1984): *Cluster Analysis*. Newbury Park, Sage.
- BORGATTI, S. P. (2002): *Net Draw: Graph Visualization Software*. Harvard, Analytic Technologies.
- CARLEY, K. (1993): «Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis», en *Sociological Methodology*, 23, 75-126.
- CARLEY, K. y M. E. PALMQUIST (1992): «Extracting, Representing, and Analyzing Mental Models», en *Social Forces*, 70(3), 601-636.
- CASTELLS, M. (1996): *The Rise of the Network Society*. Malden, Mass., Blackwell Publishers.
- CORMAN, S. R. et al. (2002): «Studying Complex Discursive Systems: Centering Resonance Analysis of Communication», en *Human Communication Research*, 28(2), 157-206.
- COTEC (1994): *Un Análisis del Tratamiento de la Innovación Tecnológica en la Prensa Española*. (1992-1993). Madrid, COTEC.
- (1998): *Segundo Análisis del Tratamiento de la Innovación Tecnológica en la Prensa Española* (1996-1997). Madrid, COTEC.
- (1999): *Informar sobre Innovación*. Madrid, COTEC.
- (2004): *Documento para Debate: Situación en el 2003 del Sistema Español de Innovación*. Madrid, COTEC.
- (2005): *Creatividad e Innovación en la Práctica Empresarial*. Madrid, COTEC.
- DANOWSKI, J. A. (1988): «Network Analysis of Message Content», en *Progress in Communication Sciences*, 12, 198-221.
- (1988): «Organizational Infographics and Automated Auditing: Using Computers to Unobstrusively Gather as Well Analyze Communication», en G.M. Goldhaber et al. (eds.), *Handbook of Organizational Communication*. Norwood, NJ, Ablex.

- FECYT (2004): *Percepción Social de la Ciencia y la Tecnología en España*. Madrid, FECYT.
- FREEMAN, C. A. y G. A. BARNETT (1994): «An Alternative Approach to Using Interpretative Theory to Examine Corporate Messages and Organizational Culture», en L. Thayer et al. (eds.), *Organizational Communication. Emerging Perspectives*. Norwood, NJ, Ablex.
- FREEMAN, L. C. (1977): «A Set of Measures of Centrality Based on Betweenness», en *Sociometric*, 40, 35-41.
- (1979): «Centrality in Social Networks: Conceptual Clarification», en *Social Networks*, 1, 215-239.
- (1992): «Social Networks and the Structure Experiment», en L.C. Freeman et al. (eds.), *Research Methods in Social Network Analysis*. London, Transaction Publishers.
- (2000): «Visualizing Social Networks», en *Journal of Social Structure*, 1(1).
- HERNÁNDEZ, L. (2001): *Técnicas de Taxonomía Numérica*. Madrid, La Muralla-Hespérides.
- JANG, H. y G. A. BARNETT (1994): «Cultural Differences in Organizational Communication: A Semantic Network Analysis», en *Bulletin de Methodologie Sociologique*, 44, 31-59.
- JOHNSON, J. C. y L. KREMPEL (2007): «Network Visualization: The «Bush Team» In Reuters News Ticker 9/11-11/15/01», en *Journal of Social Structure*, 8(2).
- LOZARES, C. (2000): «La Actividad Situada y/o el Conocimiento Socialmente Distribuido», en *Papers. Revista de Sociología*, 62, 97-131.
- LOZARES, C. et al. (2003): «La Simulación Social, ¿una Nueva Manera de Investigar en Ciencia Social?», en *Papers. Revista de Sociología*, 72, 165-188.
- MCPHEE, R. D. et al. (2002): «Organizational Knowledge Expression and Management: Centering Resonance Analysis of Organizational Discourse», en *Management Communication Quarterly*, 16(2), 274-281.
- MORENO, J. L. (1932): *Application of the Group Method to Classification*. New York, National Committee on Prisons and Prison Labor.
- POPPING, R. (2000): *Computer-Assisted Text Analysis*. London, Sage.
- (2003): «Knowledge Graphs and Network Text Analysis», en *Social Science Information*, 42, 91-106.
- (2005): «Representation of Developments in Labour Market Research», en *Quality & Quantity*, 39, 241-251.
- (2005): «Text Analysis for Knowledge Graphs», en *Quality & Quantity*, 41, 691-709.
- QUINTANILLA, M. Á. (2005): *Tecnología: Un Enfoque Filosófico y Otros Ensayos de la Tecnología*. México, Fondo de Cultura Económica.
- SCHNEGG, M. y H. R. BERNARD (1996): «Words as Actors: A Method for Doing Semantic Network Analysis», en *Cultural Anthropology Methods Journal*, 8(2), 7-10.
- SCOTT, J. (2000): *Social Network Analysis*. London, Sage.
- SIMMEL, G. (2002): *Cuestiones Fundamentales de Sociología*. Barcelona, Gedisa.
- SOWA, J. F. (ed.). (1991): *Principles of Semantic Networks. Explorations in the Representation of Knowledge*. San Mateo, CA, Morgan Kaufman.
- SVENJA, A. (2006): *Introducing Electronic Text Analysis: a Practical Guide for Language and Literary Studies*. London, Routledge.
- VAN ATTENVELD, W. (2008): *Semantic Network Analysis: Techniques for Extracting, Representing, and Querying Media Content*. London, Routledge.

- VERD, J. M. (2005): «El Uso de la Teoría de Redes Sociales en la Representación y Análisis de Textos. De las Redes Semánticas al Análisis de Redes Textuales», en *Empiria. Revista de Metodología de Ciencias Sociales*, 10, 129-150.
- VERD, J. M. y C. LOZARES (2000): «La Teoría de Xarxes Socials Aplicada a la Interpretació D'entrevistes Narratives. Una Proposta», en *Revista Catalana de Sociologia*, 11, 191-196.

APENDICE

Listado alfabético de descriptores con frecuencia de aparición superior a las 400 noticias.

- Abstencionismo (474), Actividad parlamentaria (869), Actividad política (847), Actos electorales (1325), Actos oficiales (848), Alianza política (726), Alianza Popular (1753), Amnistía (1196), Análisis político (1293), Andalucía (900), Anticomunismo (452), Areilza y Martínez-Rodas, José María de (425), Asambleas (1609), Asesinato (1131), Asociacionismo político (948), Atentados (2757), Audiencias (758), Autonomía política (1839), Autonomía regional (440), Ayuntamientos (590).
- Barcelona (1290), Bilbao (574), Biografía (695).
- Calvo-Sotelo Bustelo, Leopoldo (432), Cambio político (1308), Cambio social (595), Campaña electoral (2571), Canarias (540), Candidatura electoral (2021), Carrillo Solares, Santiago (1457), Cataluña (1788), Centro Democrático y Social (500), Centro político (657), Clero (521), Coalición política (900), Coalición Popular (435), Comisiones Obreras (666), Comportamiento electoral (1071), Comunidad Económica Europea (932), Conferencias (1410), Conflicto del Sahara (494), Conflictos laborales (877), Congreso de los Diputados (720), Congresos (1500), Conmemoraciones (1168), Consejo de Ministros (762), Consejo Nacional del Movimiento (444), Constitución de 1978 (1000), Convocatoria electoral (799), Cortes constituyentes (519), Cortes españolas (1023), Crisis económica (1040), Crisis política (883), Crítica política (3762)
- Debate parlamentario (1664), Debate político (1304), Declaraciones (3677), Delitos (424), Democracia (764), Demócratas cristianos (678), Democratización (1178), Denuncia (454), Derecha política (496), Desarrollo económico (406), Descontento popular (596), Desempleo (635), Detención (1848), Dimisión (981), Diputados (441), Discursos (1780)
- Ejército (1160), Elección de cargos (2328), Elecciones (646), Elecciones autonómicas (1137), Elecciones de 1977 (4775), Elecciones de 1979 (1041), Elecciones de 1982 (689), Elecciones de 1986 (483), Elecciones municipales (1134), Elecciones sindicales (440), Empresa (452), Empresarios (656), Encuestas (703), Enfrentamiento ideológico (1102), Enfrentamiento político (1437), Entrevistas (1350), Época franquista (450), Escisión política (434),

- España (1368), Estados Unidos (1154), Estatutos de autonomía (600), Estrategia política (2592), ETA (3330), Eurocomunismo (420), Exequias (608), Extrema derecha (828).
- Falange Española de las JONS (401), Familia (402), Fraga Iribarne, Manuel (1606), Francia (938), Franco Bahamonde, Francisco (1572), Franquismo (705), Fuerza Nueva (524), Fuerzas de seguridad del Estado (883), Funcionarios (412).
- Galicia (487), Gobierno (471), Gobierno Arias (682), Gobierno autonómico (845), Gobierno Calvo-Sotelo (504), Gobierno de los tecnócratas (567), Gobierno PSOE (832), Gobierno Suárez (3925), Golpe del 23 de febrero (1215), González Márquez, Felipe (1936), GRAPO (531), Guardia Civil (859), Guerra González, Alfonso (464).
- Herri Batasuna (697), Homenajes (686), Huelgas (1311), Humor gráfico (523). Ideología política (939), Iglesia católica (1364), Informes (566), Investigación (633), Izquierda política (841).
- Juan Carlos I, Rey de España (2067).
- Labor legislativa (532), Legalización (1332), Ley 1/1977, de 4 de enero, para la Reforma Política (513), Libros (444), Líderes políticos (1268).
- Madrid (2927), Mandos militares (545), Manifestaciones (2913), Marruecos (426), Martín Villa, Rodolfo (485), Marxismo (445), Medidas económicas (695), Medidas políticas (464), Militancia política (643), Militares (688), Ministros (758), Monarquía (466), Movilización política (434), Movimiento Nacional (445).
- Nacionalismo vasco (683), Navarra (675), Negociación colectiva laboral (411), Negociación internacional (626), Negociación política (2563).
- Obispos (630), Opinión internacional (1232), Opinión pública (959), Opiniones (878), Oposición política (1292), Orden público (2429), Organización de partidos políticos (1542), Organización Sindical (521), OTAN (1246).
- Pactos de la Moncloa (518), País Vasco (3092), Participación ciudadana (666), Participación política (1281), Partido Comunista de España (2499), Partido Nacionalista Vasco (965), Partido Socialista Obrero Español (4184), Partido Socialista Popular (624), Partidos políticos (2761), Periodistas (404), Policía nacional (902), Política autonómica (929), Política económica (782), Política educativa (463), Política exterior (842), Política gubernamental (809), Posición política (5119), Prensa (920), Presos políticos (454), Previsiones electorales (699), Proceso judicial (1186), Procuradores en Cortes (489), Profesorado (501), Programa político (1996), Propaganda electoral (910), Propaganda política (897), Proyecto constitucional (551), Proyectos de ley (2131).
- Radio Televisión Española (582), Referéndum (2374), Reforma política (3228), Reino Unido (406), Reivindicaciones laborales (1255), Reivindicaciones políticas (1865), Relaciones bilaterales (1491), Relaciones exteriores (498), Relaciones Iglesia-Estado (704), Repercusión internacional (466), Representación política (599), Representación sindical (438), Represión policial (807), Resultados electorales (2611).

Salarios (489), San Sebastián (422), Sátira política (435), Secuestro (1171), Senado (449), Sentencias judiciales (521), Sindicatos obreros (872), Sistema electoral (516), Socialismo (488), Socialistas (760), Suárez González, Adolfo (2196).

Tarradellas i Joan, Josep (424), Tendencias políticas (956), Terrorismo (2797), Tierno Galván, Enrique (569), Trabajadores (1034), Transición política (557).

Unidad política (729), Unión de Centro Democrático (2594), Unión General de Trabajadores (664), Unión Soviética (576), Universidades (434).

Valencia (563), Valoración de la democracia (449), Viajes oficiales (586), Víctimas de atentados (596), Violencia política (1283), Visitas (533), Votos (529).

RESUMEN

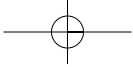
Este artículo presenta una serie de técnicas para la construcción de gráficos para la representación de la coaparición de conceptos o temas en un conjunto de textos. Con ello se pretende estudiar gráficamente las conexiones temáticas o semánticas presentes en un determinado discurso, a fin de descubrir no sólo los más frecuentes sino cómo están ligados entre sí. La idea clave está en la aplicación del análisis de redes a las apariciones conjuntas significativas de ideas en un determinado fragmento de texto. Para la explicación de estas técnicas se presenta un análisis de los descriptores más frecuentes correspondientes a la base de datos de la transición a la democracia española recopilada por Linz, compuesta por 76.000 artículos publicados por los principales periódicos de este país en el periodo comprendido entre 1958 y 1987.

PALABRAS CLAVE

Análisis de redes, Análisis de contenido, Transición española a la democracia.

ABSTRACT

This paper shows a set of techniques designed purpose of graph construction aimed at the apparition of either concepts or issues in a collection of texts. The aim is to analyze graphically the semantic connections that can be found in a selected textual corpus as a complement to the study of the statistical frequency of the different elements. The key contribution of this proposal is the use of network analysis for those items that are statistically connected. The use and range of these techniques are shown through the analysis of a database compiled



MODESTO ESCOBAR

REDES SEMÁNTICAS EN TEXTOS PERIODÍSTICOS... 39

by J.J. Linz composed by more than 76.000 Spanish newspapers articles on the transition to the democracy, published between 1958 and 1987.

KEYWORDS

Network analysis, Content analysis, Spanish democratic transition.

