



Facultad de Ciencias Económicas y Empresariales
Universidad de León

Grado en Marketing e Investigación de Mercados
Curso 2016 / 2017

ANÁLISIS DE DATOS TEXTUALES MULTILINGÜE EN EL MERCADO
PERIODISTICO INTERNACIONAL: LA INFLUENCIA DE DONALD
TRUMP Y LAS ELECCIONES ESTADOUNIDENSES.

(MULTILINGUAL TEXTUAL DATA ANALYSIS IN THE INTERNATIONAL
JOURNALISTIC MARKET: THE INFLUENCE OF DONALD TRUMP AND
THE USA ELECTIONS)

Realizado por la alumna D^a. Laura Abou-Jamous Andrés

Tutelado por el Profesor D. Ramón Álvarez Esteban

León, a 14 de Septiembre de 2017

ÍNDICE DE CONTENIDO

RESUMEN	6
ABSTRACT	7
1. INTRODUCCIÓN	8
2. OBJETIVOS	10
3. METODOLOGÍA	12
3.1. FUENTES DE INFORMACIÓN	12
3.2. EXTRACCIÓN DE DATOS	12
3.3. CARACTERÍSTICAS DE LA MUESTRA	13
3.4. PROGRAMAS UTILIZADOS Y TIPOS DE ANÁLISIS	14
4. EL MARKETING PERIODÍSTICO Y LA POLÍTICA	15
4.1. EL CONSUMIDOR DEL MERCADO PERIODÍSTICO	16
4.2. EL CONTENIDO COMO ELEMENTO DE DIFERENCIACIÓN	17
4.3. LA PRENSA COMO ELEMENTO FUNDAMENTAL DEL MARKETING POLÍTICO	18
5. EL ANÁLISIS DE DATOS TEXTUALES Y SU EVOLUCIÓN	20
6. CORPUS	24
7. ANÁLISIS INDIVIDUALES	27
7.1. ANÁLISIS ESPAÑA	27
7.2. ANÁLISIS FRANCIA	38
7.3. ANÁLISIS REINO UNIDO	47
7.4. ANÁLISIS REGIÓN ÁRABE	54
7.5 ANÁLISIS ESTADOS UNIDOS	62
8. ANÁLISIS CONJUNTO	69
9. CONCLUSIONES	79

REFERENCIAS	82
-------------------	----

ÍNDICE DE CUADROS

Cuadro 6.1: División del periodo de tiempo estudiado en semanas.....	24
Cuadro 6.2.: Clasificación de temáticas.	26

ÍNDICE DE FIGURAS

Figura 6.1.: Periódicos utilizados para la elaboración de cada base de datos individual.....	25
--	----

ÍNDICE DE GRÁFICOS

Gráfico 4.1. Evolución de la inversión publicitaria en España.....	15
Gráfico 7.1.1.: Mapa de las palabras sobre las dimensiones 1 y 2 españolas.....	31
Gráfico 7.1.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 españolas.....	32
Gráfico 7.1.3.: Mapa de las palabras sobre las dimensiones 1 y 3 españolas.....	33
Gráfico 7.1.4.; Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 españolas.....	34
Gráfico 7.1.5.: Mapa de grupos de semanas similares de España.	37
Gráfico 7.1.6.: Dendrograma España.....	37
Gráfico 7.2.1.: Mapa de las palabras sobre las dimensiones 1 y 2 francesas.....	40
Gráfico 7.2.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 francesas.	40
Gráfico 7.2.3.: Mapa de las palabras sobre las dimensiones 1 y 3 francesas.....	42

Gráfico 7.2.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 francesas.....	43
Gráfico 7.2.5.: Mapa de grupos de semanas similares de Francia.....	44
Gráfico 7.2.6.: Dendrograma Francia.....	45
Gráfico 7.3.1.: Mapa de las palabras sobre las dimensiones 1 y 2 británicas.....	49
Gráfico 7.3.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 británicas.....	50
Gráfico 7.3.3.: Mapa de las palabras sobre las dimensiones 1 y 3 británicas.....	51
Gráfico 7.3.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 británicas.....	51
Gráfico 7.3.5.: Mapa de grupos de semanas similares de Reino Unido.....	53
Gráfico 7.3.6.: Dendrograma Reino Unido.....	53
Gráfico 7.4.1.: Mapa de las palabras sobre las dimensiones 1 y 2 árabes.....	55
Gráfico 7.4.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 árabes....	56
Gráfico 7.4.3.: Mapa de las palabras sobre las dimensiones 1 y 3 árabes.....	58
Gráfico 7.4.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 árabes....	58
Gráfico 7.4.5.: Mapa de grupos de semanas similares de la región árabe.....	60
Gráfico 7.4.6.: Dendrograma región árabe.....	61
Gráfico 7.5.1.: Mapa de las palabras sobre las dimensiones 1 y 2 de EEUU.....	64
Gráfico 7.5.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 de EEUU.....	64
Gráfico 7.5.3.: Mapa de las palabras sobre las dimensiones 1 y 3 de EEUU.....	65
Gráfico 7.5.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 de EEUU.....	66
Gráfico 7.5.5.: Mapa de grupos de semanas similares de Estados Unidos.....	67

Gráfico 7.5.6.: Dendrograma Estados Unidos.	68
Gráfico 8.1.: Mapa de las semanas sobre las dimensiones 1 y 2.	72
Gráfico 8.2.: Mapa de las palabras sobre las dimensiones 1 y 2.	72
Gráfico 8.3.: Mapa de las palabras específicas seleccionadas sobre las elecciones ...74 estadounidenses y Donald Trump en las dimensiones 1 y 2.....	74
Gráfico 8.4.: Representación de la distancia entre países/regiones por semanas.	76
Gráfico 8.5.: Representación de las dimensiones parciales sobre los factores del AFM.	77
Gráfico 8.6.: Representación de los países/regiones.....	78

ÍNDICE DE TABLAS

Tabla 7.1.1.: Las 50 palabras más características del vocabulario global de la base de datos española.....	28
Tabla 7.2.1.: Las 50 palabras más características del vocabulario global de la base de datos francesa.	38
Tabla 7.3.1.: Las 50 palabras más características del vocabulario global de la base de datos británica.....	47
Tabla 7.4.1.: Las 50 palabras más características del vocabulario global de la base de datos árabe.	54
Tabla 7.5.1.: Las 50 palabras más características del vocabulario global de la base de datos estadounidense.	62
Tabla 8.1.: Valores propios por dimensión y contribuciones por país.	70

RESUMEN

La celebración de las elecciones presidenciales estadounidenses es un acontecimiento que genera gran repercusión en todos los países del mundo. En el caso de las celebradas en noviembre de 2016 la repercusión fue aún mayor, debido a la candidatura por el partido republicano del magnate Donald Trump, famoso por sus polémicas declaraciones y programa electoral. Como es lógico, esta repercusión se vio reflejada en toda la prensa internacional, ocupando cada vez más portadas y titulares.

A lo largo de este trabajo, se estudiarán en primer lugar, dentro de un marco teórico, los aspectos más relevantes del mercado periodístico así como su relación con el marketing político. En segundo lugar, la evolución del análisis de datos textuales gracias a los avances tecnológicos, que ha permitido la realización de este tipo de estudios.

Por último, desde un marco práctico, se han elaborado cinco bases de datos correspondientes a España, Francia, Reino Unido, región árabe y Estados Unidos, en sus idiomas de origen, a partir de los artículos de los periódicos digitales más leídos de cada región. A continuación, se llevaron a cabo análisis tanto individuales como de manera conjunta, para obtener las características de cada país y las diferencias entre ellos.

Palabras clave: Análisis estadístico de datos textuales, multilingüe, marketing periodístico, política, Trump, elecciones estadounidenses.

ABSTRACT

The holding of the USA elections always generates a huge impact in every single country across the globe. With regards to the elections that took place in November 2016, this impact was even greater, due to the republican's party candidature of Mr. Donald Trump, notorious for his controversial statements and election programme. Logically, this impact was reflected in the whole international press, taking up more front covers and headlines day by day.

Throughout this study, will be considered first of all, within a theoretical framework, are journalism marketing's most important issues, as well as it's relationship to the political marketing. Secondly, the data textual analysis evolution through technological development, which has enabled this type of studies.

Finally, from a practical framework, five databases related to Spain, France, Great Britain, Arab countries, and the United States have been developed, in their own original languages, from the most read digital press articles of each areas. Thereafter, the individual and the pooled analysis were carried out in order to obtain the features of each country and the differences between all of them.

Keywords: Statistical textual data analysis, multilingual, journalism marketing, politics, Trump, U.S.A. elections.

1. INTRODUCCIÓN

El desarrollo de nuevas tecnologías ha provocado una revolución en la sociedad. Los mercados, y empresas que compiten en ellos, han tenido que adaptarse a los nuevos cambios con el objetivo de poder competir y sobrevivir en un entorno en el que la rivalidad aumenta día a día.

Los medios de comunicación, concretamente, se han visto realmente afectados por los avances tecnológicos, apareciendo nuevas formas de transmitir la información e incrementando su contenido por el mayor acceso a esta, si comparamos con épocas pasadas.

Respecto a la prensa, la aparición de los periódicos digitales ha ido en detrimento del formato en papel tradicional, llegando estos a un mercado mucho más amplio y heterogéneo y alcanzando un contenido mucho más abundante, variado y actualizado. Consecuentemente, la inversión publicitaria, que constituía gran parte de los ingresos del sector, siguió a los lectores en el proceso de cambio de formato, posicionando a los diarios digitales como líderes de los medios de comunicación escritos (Montero Rodríguez, 2005)

Como se verá a lo largo del estudio, el contenido que proporcione cada marca será el elemento diferenciador de las que compiten en el mismo mercado. Generalmente, la temática líder sobre la que versa la prensa internacional corresponde a la política, ya sea nacional o internacional, pero las publicaciones sobre este tema se multiplican si se trata de nueva información de la política estadounidense, y aún más cuando se acercan elecciones a presidente del gobierno con un candidato, y posterior ganador de ellas, muy polémico internacionalmente por sus declaraciones y programa político, el magnate Donald Trump.

Por otro lado, el sector de la estadística ha dado un enorme salto gracias a los avances informáticos, ya que han permitido la realización de análisis antes impensables de manera manual y han reducido la dificultad y el tiempo de ejecución de otros.

Concretamente, un análisis textual de la envergadura de este trabajo sería prácticamente impensable manualmente, debido a la gran cantidad de datos a analizar.

Lo que se pretende con este trabajo es, en primer lugar, aportar un estudio más al escaso número de publicaciones acerca del análisis de datos textuales multilingüe, analizándose conjuntamente los idiomas español, inglés, francés y árabe.

En segundo lugar, analizar la repercusión que Donald Trump y las elecciones estadounidenses de noviembre de 2016 han tenido en la prensa internacional.

La razón por la que considero este estudio interesante es la diferencia cultural existente entre los países analizados, latente también en la prensa y su contenido, sobre todo entre los países occidentales, como España, Reino Unido, Estados Unidos y Francia, y los países árabes.

2. OBJETIVOS

Existen dos objetivos principales: Realizar un análisis textual a partir de datos en cuatro idiomas diferentes, y analizar la repercusión de las elecciones estadounidenses y del magnate Donald Trump como candidato en ellas en la prensa internacional, en el periodo comprendido entre el día 3 de octubre de 2016 y el día 20 de febrero de 2017, es decir, desde un mes previo a las elecciones estadounidenses hasta un mes después de su subida al poder. Para ello, establecemos una serie de subobjetivos:

- Conocer de manera teórica los aspectos más relevantes del mercado periodístico y su relación con el marketing político, así como la evolución del análisis de datos textuales en el tiempo.
- Obtener las palabras características de cada base de datos para aproximar a partir de ellas, los temas más característicos de la prensa de cada país en ese periodo de tiempo.
- Agrupar, mediante los temas más relevantes en cada nación/región, las diferentes semanas analizadas en el estudio, para así determinar en cuantas y en cuales la política internacional (política nacional en el caso de Estados Unidos) es la temática principal.
- Siguiendo con el objetivo anterior, determinar si existe una evolución temática en el tiempo.
- Determinar si existe similitud entre el vocabulario utilizado por los distintos países/regiones en las mismas semanas.

- Partiendo de vocabulario concreto sobre Donald Trump y las elecciones estadounidenses, observar la utilización del mismo de cada país/región respecto a los demás.
- Comprobar la similitud o disimilitud general de la prensa de los distintos países/regiones en este periodo.

3. METODOLOGÍA

El estudio se ha elaborado fundamentalmente desde un marco práctico, con el objetivo de realizar un trabajo novedoso sobre la actualidad periodístico-política a partir del análisis de datos textuales en cuatro diferentes lenguas, de lo cual no existen muchos artículos publicados.

3.1. FUENTES DE INFORMACIÓN

Se han creado cinco bases de datos de: España, Francia, Reino Unido, la región árabe y, Estados Unidos. Respecto a la región árabe es importante añadir que no se refiere tan solo a un país, sino a todos los países en los que se habla este idioma, y que la razón por la que, para la elaboración de este trabajo, se han juntado en la misma región es que los dos periódicos seleccionados son los dos periódicos más leídos en todos estos países, criterio que se utilizó a la hora de elegir la prensa para la elaboración de cada corpus¹. Los países/regiones seleccionados para la realización del estudio se deben al conocimiento de las diferentes lenguas habladas en estos países/regiones, ya que la recopilación de los diferentes artículos se ha basado en periódicos propios de cada uno de los países y, por ende, en sus respectivos idiomas. La información para crear las bases de datos ha sido extraída a partir de los periódicos digitales que fueron seleccionados.

En la interpretación de dichas bases de datos se ha utilizado información de fuentes secundarias como libros, estudios publicados de temas relacionados y apuntes de estadística avanzada.

3.2. EXTRACCIÓN DE DATOS

Se seleccionaron 150 artículos de cada uno de los dos periódicos elegidos para la elaboración de cada corpus, 300 por cada base de datos. Para la extracción de estos artículos se utilizó la plataforma de noticias “*Google News*”, la cual mediante su algoritmo de “ordenar por importancia”, devuelve las noticias más relevantes del periodo de tiempo seleccionado sobre el criterio de búsqueda.

¹ “Se considera corpus a un texto o conjunto de textos con dominios semánticos similares, es decir, aquellos sobre los que se va a efectuar el estudio estadístico” (Álvarez Esteban, 2015 a, p.209)

El periodo de tiempo que comprende el estudio es desde el día 3 de octubre de 2016 hasta el 20 de febrero de 2017, esto es, desde un mes antes de las votaciones para la elección del nuevo presidente del gobierno de Estados Unidos, hasta un mes después de la subida al poder del actual presidente de los Estados Unidos, Donald Trump.

En este caso, el criterio de búsqueda era el nombre del periódico del que se querían recopilar las noticias y configuraba en “herramientas de búsqueda” el periodo semanal deseado. Tras esto, recogía las noticias en un archivo de *excel*, a la vez que leía cada una de ellas para completar datos, como su temática, el periódico que publicó ese artículo o la semana de publicación, hasta llegar a las 300 noticias por cada una de las bases de datos.

3.3. CARACTERÍSTICAS DE LA MUESTRA

Dado que el trabajo se basa en un análisis *crosscultural*, se han elaborado una base de datos o corpus sobre Francia en francés, otra sobre España en español, dos bases de datos diferentes en lengua inglesa pertenecientes a Estados Unidos y a Reino Unido, respectivamente; y por último, una en idioma árabe perteneciente al conjunto de los países árabes, ya que, estos poseen “*Al Jazeera*” y “*Al Arabiya*” como los dos periódicos comunes más leídos. Los motivos de realizar estas diferentes bases de datos se encuentran en los objetivos perseguidos por el trabajo, como desvelar las particularidades de la prensa de cada uno de los países por separado y poder comparar y encontrar diferencias al realizar un análisis conjunto.

La muestra se compone, por tanto, de cinco bases de datos con 300 individuos o artículos en cada una de ellas, correspondientes al periodo de tiempo de octubre de 2016-febrero de 2017. Las razones de seleccionar 300 individuos por muestra son dos, en primer lugar tener un número lo suficientemente grande como para poder analizar y que salgan resultados significativos y, en segundo lugar, para poder comparar los diferentes corpus entre sí en el posterior análisis conjunto.

Además de las noticias, las cuales son las variables principales del análisis, se encuentran otras variables como la “Semana”, formada a partir de la división del periodo de tiempo estudiado en 20 semanas y que constituirá la variable de agregación del análisis, es decir la variable a partir de cuyas modalidades se clasificarán los

resultados. Por otro lado se encuentra “Temática”, la cual es una variable construida por mi a partir del establecimiento de 8 modalidades (guerra y asuntos bélicos, política internacional, política nacional, economía internacional, economía nacional, sucesos y atentados terroristas, ciencia y tecnología, y cultura) y la posterior identificación y clasificación de los artículos en ellas. La razón de la creación de esta variable es su introducción en el análisis como variable suplementaria², con el objetivo de que ayude a la interpretación de los resultados. Por último, la variable “Periódico” que contiene los dos periódicos elegidos en cada base de datos. La elección de estos fue a partir de la identificación de los más representativos de cada país o región, atendiendo a los *rankings* de los más leídos de cada lugar.

3.4. PROGRAMAS UTILIZADOS Y TIPOS DE ANÁLISIS

Para elaborar las diferentes bases de datos se ha utilizado el programa *excel* debido a su facilidad de utilización, y de la posibilidad de posterior exportación al programa con el que se ha realizado el análisis, el cual es el programa estadístico R (Hornik, 2017).

Tras la importación de los archivos *excel* a formato *.RData*, se procedió al análisis individual de cada base de datos a partir del paquete *XplorText* cuya función es el análisis de datos textuales (Bécue-Bertaut, Álvarez-Esteban y Sánchez-Espigares, 2017).

El mayor problema con el programa fue el idioma árabe, debido a su diferente alfabeto, ya que no lo leía y por tanto no se podía realizar el análisis, pero se consiguió tras probar con distintas versiones del programa y la búsqueda de *scripts* que permitiesen su lectura.

Se han utilizado diversos análisis estadísticos con el fin de conseguir los objetivos planteados para los distintos países/regiones por separado, como el análisis de vocabulario específico, el análisis de correspondencias (AC) y el análisis clúster.

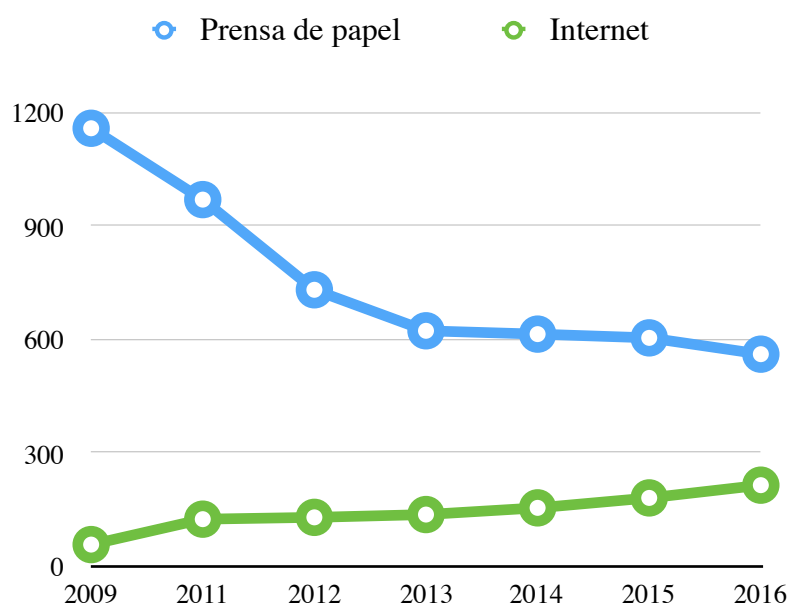
Por otro lado, el análisis realizado a las bases de manera conjunta fue el análisis factorial múltiple (AFM).

² “Aquellas variables que pueden aparecer representadas y contribuir a explicar de forma significativa los primeros ejes, aunque su contribución absoluta siempre será 0, porque no participan activamente en el análisis.” (Álvarez Esteban, 2015)

4. EL MARKETING PERIODÍSTICO Y LA POLÍTICA

El marketing periodístico es un tipo de marketing sectorial orientado a la prensa diaria, la cual se divide en dos grandes grupos: periódicos y revistas. Aunque en la práctica sean mercados muy diferenciados, poseen particularidades comunes y están doblegados a influencias del entorno similares. Entre sus particularidades comunes, se destaca una en concreto, y es que es un producto que se vende dos veces, y esto es debido a que los ingresos procedentes de los productos vienen, en primer lugar, de la venta de los ejemplares a los lectores y, en segundo lugar, de la publicidad por la que pagan las marcas. La inversión publicitaria es fundamental en este mercado y ha experimentado importantes cambios en los últimos años a causa del aumento de la utilización de la prensa digital en detrimento de la prensa escrita, con la consecuente caída de la inversión publicitaria en esta última, a favor de la primera (Montero Rodríguez, 2005). A partir de los datos recogidos en el “Índice de Inversión Publicitaria 2016” de Media Hotline y Arce Media (como se cita en Zarzalejos, 2017), se ha creado una gráfica para mostrar esta evolución en el caso de España:

Gráfico 4.1. Evolución de la inversión publicitaria en España
(Cifras en millones de euros)



Fuente: Elaboración propia a partir de datos de “Medios de comunicación: La prensa de papel, en el chasis: la publicidad también caerá este año víctima de internet” El Confidencial (2017)

Entre las influencias del entorno comunes destacamos la coyuntura económica del país, los cambios legales, los cambios sociales y políticos y la competencia de otros medios de comunicación, sobre todo con la llegada de internet y las nuevas tendencias digitales.

4.1. EL CONSUMIDOR DEL MERCADO PERIODÍSTICO

La primera cuestión a tratar en el marketing editorial debe centrarse en el análisis del consumidor y las motivaciones de su compra.

En primer lugar, en cuanto al perfil sociodemográfico del lector medio, en el caso de España, según datos del CIS (Centro de Investigaciones Sociológicas) de 2014, destaca la prevalencia de los hombres lectores de prensa escrita sobre las mujeres, con un 75% de estos como lectores habituales frente a un 46,5% en el caso de lectoras habituales. Respecto al grupo de edad predominante, destacan dos: personas entre 45 y 54 años (21%) y personas de más de 65 años (20,9%), según una estadística realizada por statista.com en 2016. Siguiendo con los grupos de edades que predominan como lectores de prensa, se observa un claro envejecimiento del perfil de los usuarios, algo que es preocupante para el sector. Según esta misma estadística los grupos de edad jóvenes como el grupo de 14-19 años, el de 20-24 y el de 25-34, constituyen el 4%, 4,2% y 12,7%, respectivamente, de los lectores totales en este año. Según estudios sociológicos, se cree que una de las causas principales de este bajo porcentaje de jóvenes entre los lectores es el gran peso de la información política en los diarios, algo que interesa poco a este grupo de población. Quizás, llegados a este punto y teniendo en cuenta que el tema a tratar se basa en el marketing de este sector, la pregunta ahora sería: ¿Por qué si este contenido no interesa a los jóvenes no se cambia el contenido? La respuesta a esta pregunta se encuentra en dos razones: La primera, a la rutina y hábitos de trabajo establecidos entre responsables y directores editoriales, gran parte de ellos formados en la cultura informativa de hace 30 o 40 años, en la que la política sí se consideraba tema de interés entre los jóvenes de la época: Y, la segunda, al claro posicionamiento de ciertos diarios a determinados partidos políticos y opiniones, que fomentan cada día con el contenido que publican (Montero Rodríguez, 2005).

Respecto a las motivaciones de compra, el motivo principal es la *necesidad de conocimiento e información* a partir de noticias de mayor profundidad, que solo puede

proporcionarle la prensa. El segundo motivo es el *posicionamiento ideológico* de cada periódico que provoca que el lector, en la mayoría de los casos, compre uno u otro en base a la coherencia con sus propias ideas, ya que este percibirá la información como mejor, respecto a otros diarios. Por último, otras motivaciones a la compra pueden ser el formato, la edición, o incluso el suplemento que se publica con el periódico en cuestión.

4.2. EL CONTENIDO COMO ELEMENTO DE DIFERENCIACIÓN

La diferenciación es “una estrategia de marketing que trata de resaltar las características del producto, sustanciales o simplemente accesorias, que pueden contribuir a que sean percibidas como único.” (Montero Rodriguez, 2005, p.112). En base a esta definición, se determina por tanto que la estrategia de diferenciación de este mercado es el contenido, la información que se proporciona en los artículos publicados por cada periódico.

En base al contenido de un diario, una tradicional clasificación es la que diferencia la *prensa de calidad* de la *prensa popular*. La primera basa su contenido en temas de profundidad y análisis como la política, la economía, la ciencia, etc. , en otras palabras, temas serios. La segunda está basada en contenidos de poca repercusión a nivel social pero con un alto grado de entretenimiento, como los deportes o el corazón, a lo que una gran parte de la población da un gran valor.

La elección de un tipo de contenido u otro radica en el propio contenido o el target³ al que se dirige el periódico. Generalmente, el nivel sociocultural de los consumidores de *prensa de calidad* es superior al de los que consumen *prensa popular*.

La credibilidad es el valor principal que posee la *prensa de calidad*, lo que provoca que genere gran influencia sobre la opinión pública (Montero Rodriguez, 2005).

Según una estadística realizada por statista.com en 2016, el periódico más leído de España es “Marca” cuyo contenido se relaciona con la *prensa popular* por dedicarse exclusivamente a la temática deportiva. Los periódicos pertenecientes a la *prensa de calidad* más leídos son “El País” y “El Mundo” los cuales serán utilizados para realizar la base de datos de España y su posterior análisis.

³Público al que va dirigido.

4.3. LA PRENSA COMO ELEMENTO FUNDAMENTAL DEL MARKETING POLÍTICO.

“El marketing político o electoral es una especialidad del marketing cuyo objetivo final es <<vender>> una ideología, un programa o un candidato.” (Barranco Saiz, 2003, p. 20)

La prensa es el medio más tradicional de comunicación política. Esta posee gran aceptación por parte del público por la profundidad de análisis del contenido que se publica, además de permitir una gran cobertura distinguiendo el mensaje conforme al entorno de difusión. Por otro lado, hay que tener en cuenta que la publicidad en la prensa no es rechazada por los lectores, sino que es una norma admitida por estos que el periódico en cuestión posea un espacio para ella entre los diferentes artículos. Por tanto, la introducción de anuncios políticos no produce efecto de rechazo alguno y, sin embargo, el recuerdo que genera es mucho más alto que cualquier campaña política publicitada en televisión, lo que deja clara la razón por la que la prensa es el medio de comunicación más tradicional de este contenido. (Luque, 1996; Barranco Saiz, 2003).

En cuanto al contenido del mensaje político, en 1944 Lazarsfeld, Berelson y Gaudet en su artículo *“The People’s choice: How the Voter Makes Up His Mind in a Presidential Campaign”* (La elección de la gente: Como los votantes componen su mente en una campaña presidencial) (como se cita en Luque, 1996), señalaban que los efectos de los medios de comunicación en las campañas políticas eran muy acotados y que solo afectaban a los votantes ya comprometidos. A partir de esta teoría y con el paso del tiempo, se fueron formando otras. En 1988 aparece la teoría de Crain y Goff, que indicaba que los votantes sí reaccionaban ante los cambios de información. Además, se señalaba que “el tipo de información que los votantes buscan y usan, directa e indirectamente, depende del coste de obtener la información, por lo que un cambio tecnológico que afecte a la habilidad para recoger información de los votantes altera las variables políticas.” (Luque, 1996, p.169).

Entre las cuestiones de más interés de los mensajes políticos para este estudio se encuentran las siguientes (Luque, 1996):

- Como indicaba Rudd en 1989, el hecho de que el votante coincida con la posición política del candidato no tiene efectos significativos en términos de imagen, pero sí que los temas se enfoquen de una manera específica.
- Según Patterson-McClure, la publicidad de los temas es un potencial real para cambiar las ideas públicas. Sostenía que a menor nivel de información del votante y con una cobertura de noticias limitada, se crea un entorno muy favorable para el máximo impacto de la publicidad.
- Según un estudio realizado en 1984 de Rust, Bajaj y Haley, “los lectores de periódicos y revistas están más relacionados positivamente con la participación, mientras que no se encontró una relación fuerte entre los telespectadores y la participación electoral, con la excepción de que los asiduos telespectadores eran menos propensos a votar” (Luque, 1996, p.172)
- Atendiendo, desde la perspectiva gramatical, al discurso político, Patzold y Verlinden en 1984 (como se cita en Luque, 1996) determinan que la variedad de vocabulario entre los candidatos a presidente del gobierno francés en las elecciones de 1974 y 1981 era diferente, y los candidatos ganadores utilizaron un “vocabulario cuantitativamente menor”. Esto hizo replantearse si la buena comprensión pudiera verse obstaculizada por la riqueza de vocabulario.
Además, indicaron también la presencia de cambios en aquellos discursos en los que aparecía vocabulario “personalizado”, desaparición o cambio en ciertos términos e hincapié en otros, incremento de los verbos de autoridad y referencias o citas a sus contrincantes en los discursos.

5. EL ANÁLISIS DE DATOS TEXTUALES Y SU EVOLUCIÓN

En primer lugar y antes de comenzar a hablar sobre la evolución de este tipo de análisis, se comenzará por su definición:

“Bajo el término “análisis de datos” se engloban antiguos métodos de lexicología, a los que la utilización del computador ha dado una nueva vitalidad, y más recientes métodos estadísticos descriptivos de grandes tablas de datos. Por tanto, el análisis estadístico de datos es una metodología nacida de las posibilidades de cálculo que ofrece el ordenador, en la que se parte de los datos y se trata un gran número de dimensiones, simultáneamente. Permite tratar una información compleja de forma global, resumirla y representarla de forma sintética. Estos métodos consideran los individuos representados en el espacio referenciado por las variables” (Bécue-Bertaut, 1991, p.8-9).

En cuanto a su evolución, según Laurance Bardin (1977), el interés por el análisis de textos no comenzó en el siglo XX con los avances tecnológicos, el ejemplo más claro de esto era la práctica antigua de intentar analizar textos sagrados o misteriosos con el objetivo de interpretarlos y buscar una explicación, y esto no surge de la mera observación o la intuición, ya que detrás del discurso se esconde un sentido que, generalmente, es simbólico y polisémico⁴, y el objetivo es que sea desvelado.

Dentro de las prácticas de observación de un discurso, previo al análisis de contenido, se encuentran la retórica, que estudia las maneras de expresión más adecuadas para la “*declamación persuasiva*”, y la lógica, que define las reglas formales del razonamiento a través del estudio de las declaraciones de un discurso y su consecuente estallido (Bardin, 1977).

El análisis de contenido, que se define como “un conjunto de instrumentos metodológicos cada vez más refinados y en constante mejora que se aplican a los “discursos” extremadamente diversificado, cuyo factor común de estas técnicas es una hermenéutica⁵ controlada, basada en la deducción: la inferencia.” (Bardin, 1977, p.12) ,

⁴ Que tiene más de un significado.

⁵ Técnica de interpretar un texto para descubrir su sentido.

tuvo su despegue a principios del siglo XX en Estados Unidos. Durante esta época y hasta cuarenta años después este análisis fue fundamentalmente periodístico. La primera institución que realizó estos estudios fue la Escuela de Periodismo de Columbia y gracias a ella, los estudios cuantitativos periodísticos se multiplicaron. Entre los análisis realizados por la institución destacan el seguimiento de la evolución de un órgano de prensa, la medición del grado de “sensacionalismo” de sus artículos y la comparación de los periódicos semanales y diarios de la ciudad. La Primera Guerra Mundial dio pie al análisis de la propaganda, el cual fue amplificado durante la Segunda.

Entre los años 1940 y 1950, el departamento de ciencias políticas de los Estados Unidos tuvo un importante papel en el desarrollo del análisis de contenido. Como se mencionó anteriormente, los problemas de este periodo hicieron que se amplificara el número de análisis de este tipo durante la Segunda Guerra Mundial, constituyendo la investigación política, el 25% de los estudios empíricos de análisis de contenido de la época.

Durante estos años de guerra, el gobierno de EEUU utilizó analistas para desvelar periódicos y publicaciones periódicas sospechosas de ser propaganda conspiradora (generalmente propaganda nazi). Los métodos de criba que se utilizaron fueron:

- Comparación entre el contenido de periódicos ofensores y el de los emisores de índole nazi contra los Estados Unidos.
- La identificación de temas favorables al enemigo.
- Comparación de dos publicaciones sospechosas con dos publicaciones de evidente patriotismo.
- Estudio de libros y publicaciones periódicas a favor o en contra de los temas “El éxito de la Unión Soviética” y “Las doctrinas comunistas son verdaderas” y otros temas relacionados.
- Análisis léxico a partir de una lista de palabras claves de la política y propaganda nazi.

Entre 1960 y 1975, el uso del ordenador revolucionó las investigaciones de análisis de contenido. Gracias a él, se permitió realizar pruebas estadísticas impensables de practicar anteriormente y a digerir rápidamente masas de datos que no se podían manipular de manera manual, como el análisis de frecuencias.

A partir de 1975, el desarrollo de los micro-ordenadores y las experiencias en inteligencia artificial acrecentaron la esperanza en las posibilidades informáticas. Gracias a los estudios emprendidos de ciertos tipos de análisis del discurso y bajo la influencia de una lingüística más abierta, la experimentación informática superó los problemas de programación y controló relativamente los conteos léxicos, problemas que experimentaba en el periodo temporal anterior.

A partir de este punto, el análisis de contenido multiplica sus aplicaciones, centrándose en las alteraciones de la tecnología de la información, en el tema de la innovación metodológica.

Desde la llegada de los avances informáticos y de internet y su expansión, hasta llegar a que millones de personas usen este recurso diariamente, los análisis estadísticos así como la prensa han dado un giro de 180°. El desarrollo de programas estadísticos potentes, como Lexico, Hyperbase, ALCESTE o SPAD, ha provocado un importante avance en este sector, ya que ha permitido la realización de análisis estadísticos que antes eran impensables, por el hecho de tener que realizarlos a mano. Lógicamente, estos avances también han afectado mucho al tiempo que conlleva la ejecución de un análisis, pasando de los meses o semanas antes de la aparición del ordenador, a días u horas tras la aparición de este, a segundos con la invención de estos programas informáticos que se comentaron anteriormente.

Atendiendo al análisis de datos textuales en concreto, la aparición de estos programas ha sido una verdadera revolución, ya que al manejarse una gran cantidad de datos, hay análisis que son imposibles de realizar sin ordenador.

- El análisis de contenido y la lingüística.

Supuestamente la lingüística y el análisis de contenido tienen el mismo objeto formal: el lenguaje, pero este no es el caso. La diferencia está marcada en la distinción entre lenguaje y habla, según el planteamiento de F. de Saussure.

La lingüística trabaja sobre el “lenguaje teórico, concebido como un conjunto de sistemas, que permite combinaciones controladas y sustituciones de funcionamiento del

lenguaje, a parte de las variaciones individuales o sociales tratadas por la psicolingüística y la sociolingüística.” (Bardin, 1977, p.48)

Por el otro lado, el análisis de contenido trabaja sobre el habla, o dicho de otra manera, sobre la práctica que se lleva a cabo por transmisores reconocibles de la lengua.

Es decir, la lingüística estudia un lenguaje para describir su funcionamiento, mientras que el análisis de contenido explora para conocer qué hay detrás de las palabras a las que la lingüística se refiere.

Por último, para finalizar este subapartado, se exponen algunas definiciones que serán de utilidad para situar el análisis práctico sobre el que se basa este trabajo (Bardin, 1977) :

- *Semántica*: Estudio del significado de las unidades lingüísticas.
- *Sociolingüística*: Se desplaza desde el lenguaje hasta las palabras para implantar de forma sistemática las correlaciones (covarianza) entre las estructuras lingüísticas y sociales.
- *Lexicología*: Es el estudio científico del vocabulario.
- *Estadística léxica*: Es la aplicación de los métodos estadísticos a la descripción del vocabulario. Estudia unidades simples de significado (palabras) y utilizan clasificaciones de frecuencia y conteo.
- *Análisis de discurso*: Estudia unidades lingüísticas superiores a la oración, como enunciados, textos, etc.

6. CORPUS

Como ya se ha comentado anteriormente, el corpus está formado por cinco bases de datos de los países España, Francia, Reino Unido, región árabe y Estados Unidos, con 300 individuos o artículos en cada una de ellas, que comprenden el periodo de tiempo 3 de octubre de 2016 - 20 de febrero de 2017. Este periodo fue elegido específicamente por considerarse una época en la que la prensa de las diferentes partes del mundo enfocaron gran parte de su contenido en la situación electoral de Estados Unidos, por lo que es interesante ver la evolución del mismo con el paso del tiempo. Por tanto, para ver dicha evolución, se realizó la división del periodo de tiempo que comprende el estudio en semanas de la siguiente manera:

Cuadro 6.1: División del periodo de tiempo estudiado en semanas.

Semana 1	3 - 9 de octubre de 2016	Semana 11	12 - 18 de diciembre de 2016
Semana 2	10 - 16 de octubre de 2016	Semana 12	19 - 25 de diciembre de 2016
Semana 3	17 - 23 de octubre de 2016	Semana 13	26 de diciembre de 2016 al 1 de enero de 2017
Semana 4	24 - 30 de octubre de 2016	Semana 14	2 - 8 de enero de 2017
Semana 5	31 de octubre al 6 de noviembre de 2016	Semana 15	9 - 15 de enero de 2017
Semana 6	7 - 13 de noviembre de 2016	Semana 16	16 - 22 de enero de 2017
Semana 7	14 - 20 de noviembre de 2016	Semana 17	23 - 29 de enero de 2017
Semana 8	21 -27 de noviembre de 2016	Semana 18	30 de enero al 5 de febrero de 2017
Semana 9	28 de noviembre - 4 de diciembre de 2016	Semana 19	6 - 12 de febrero de 2017
Semana 10	5 - 11 de diciembre de 2016	Semana 20	13 - 20 de febrero de 2017

Fuente: Elaboración propia

Una vez determinado el periodo de tiempo, se procedió a la elección de los periódicos representativos de cada país o región atendiendo a los *rankings* de los más leídos de cada lugar, que se exponen en la siguiente figura:

Figura 6.1.: Periódicos utilizados para la elaboración de cada base de datos individual.



Fuente: Elaboración propia

A continuación se realizó el proceso de selección de los artículos a partir de la plataforma de búsqueda “Google News”, introduciendo cada periódico digital en la barra de búsqueda y determinando en la configuración cada periodo semanal correspondiente. De cada semana se seleccionaron las 15 noticias de entre los dos

periódicos seleccionados, que según el algoritmo de la misma plataforma, son las más relevantes de ese periodo de tiempo.

De cada artículo se recopiló información para crear las siguientes variables:

- Periódico: Variable que indica el diario de procedencia de cada artículo. Esta variable contiene dos modalidades diferentes en cada base de datos, como se comentó anteriormente.
- Semana: Número de semana en la que el artículo fue publicado. Esta variable contiene 20 modalidades como se expuso en el cuadro 6.2.
- Titular: Título de cada noticia.
- Noticia: Texto del artículo. Introducido en el idioma de origen (español, inglés, francés o árabe)
- Temática: Clasificación de las materias que recogen los artículos que componen la base de datos. La siguiente tabla recoge las diferentes temáticas que se contemplaron en las bases de datos con sus correspondientes asignaciones numéricas para simplificar su identificación:

Cuadro 6.2.: Clasificación de temáticas.

<i>No.</i>	<i>Temática</i>
1	Conflictos bélicos, guerras, asuntos de guerra.
2	Política y asuntos internacionales.
3	Política y asuntos nacionales.
4	Economía y asuntos económicos internacionales.
5	Economía y asuntos económicos nacionales.
6	Sucesos.
7	Ciencia y tecnología.
8	Cultura.

Fuente: Elaboración propia

7. ANÁLISIS INDIVIDUALES

A continuación se presentan los análisis, con sus respectivas interpretaciones, que se han realizado a cada una de las bases de datos por separado, para determinar las características específicas de las mismas.

7.1. ANÁLISIS ESPAÑA

La base de datos correspondiente a España se realizó a partir de los artículos publicados por los dos periódicos digitales más leídos del país, dentro de lo que llamábamos anteriormente *prensa de calidad*, los cuales son “El País” y “El Mundo”.

El primer análisis realizado a dicha base de datos fue un análisis de datos textuales para poder obtener las palabras más características del corpus. Este análisis fue llevado a cabo a partir de la función “*TextData*” del paquete “*XplorText*” y, como se puede ver en la sintaxis presentada a continuación, para la realización del análisis se han eliminado los números, se han pasado las palabras a minúsculas, se han seleccionado las palabras con una longitud mínima de 2 que aparecen al menos 10 veces en al menos 10 noticias de las 300 totales, y se han eliminado las palabras herramienta españolas establecidas en el paquete “*tm*” (preposiciones, conjunciones, conectores textuales, etc.) (Bécue-Bertaut, M. et al., 2017).

```
TDES <- TextData(ESPANOLA, var.text = "Noticia", var.agg = "Semana", context.quali =  
c("Periodico", "Tematica"), lower = TRUE, remov.number = TRUE, Fmin = 10, Dmin = 10,  
Fmax = Inf, stop.word.tm = TRUE, idiom = "es", lminword = 2)
```

En la siguiente tabla se muestran las 50 primeras palabras con sus respectivas frecuencias de aparición:

Tabla 7.1.1.: Las 50 palabras más características del vocabulario global de la base de datos española.

No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.
1.	trump	735	18.	españa	215	35.	así	176
2.	presidente	580	19.	parte	213	36.	casa	173
3.	si	419	20.	política	211	37.	elecciones	173
4.	país	415	21.	podemos	202	38.	eeuu	170
5.	gobierno	407	22.	pp	201	39.	hacer	169
6.	dos	364	23.	unidos	199	40.	rajoy	167
7.	según	337	24.	año	197	41.	vez	166
8.	años	331	25.	países	197	42.	caso	164
9.	ser	281	26.	hace	194	43.	general	159
10.	partido	279	27.	dijo	193	44.	solo	159
11.	ahora	271	28.	nuevo	193	45.	hecho	157
12.	millones	256	29.	campana	188	46.	primera	156
13.	después	241	30.	además	186	47.	méxico	155
14.	puede	232	31.	pasado	183	48.	acuerdo	154
15.	personas	226	32.	aunque	181	49.	tiempo	153
16.	tras	219	33.	congreso	181	50.	ex	152
17.	psoe	218	34.	mundo	178			

Fuente: Elaboración propia

Como se puede comprobar en la tabla anterior las palabras más frecuentes del corpus son “trump” con una frecuencia de 735 y “presidente” con 580. además entre las más frecuentes también aparecen: “país” (415), “gobierno” (407), “partido” (337), “psoe” (218), “españa” (215), “política” (211), “podemos” (202), “pp” (201), “campana” (188), “elecciones” (173), “eeuu” (170) y “rajoy” (167).

Según las palabras citadas anteriormente, por su relevancia, se esclarecen los dos temas que predominan sobre el resto entre la prensa española durante el periodo Octubre 2016 - Febrero 2017:

- Situación de la política nacional y de las relaciones entre los partidos que componen la misma.

Especial mención en este punto a la alta puntuación de frecuencia de la palabra “PSOE”, lo cual es debido a la inestabilidad interna que sufrió este partido político durante los últimos meses del año 2016 y su consecuente convocatoria de elecciones primarias, lo que ocupó las páginas de muchos periódicos nacionales durante este periodo tiempo.

- Elecciones generales de Estados Unidos, las cuales tuvieron en esta ocasión una especial repercusión debido al candidato del partido republicano, Donald Trump, conocido mundialmente por sus controvertidas declaraciones desde sus años de magnate hasta su actual puesto en el despacho oval.

Tras la obtención de estas palabras características del corpus, se procedió a la caracterización del objeto “*TDES*”, que se expuso anteriormente, para obtener el vocabulario específico de cada semana, el cual será de gran ayuda a la hora de realizar el análisis clúster para determinar el motivo por el que se juntan en un mismo grupo distintas semanas, lo cual se explicará a continuación del AC. De este análisis se obtienen las palabras más utilizadas y las más infrautilizadas de estos pequeños periodos de tiempo. Por cuestiones de relevancia a la hora de conseguir los objetivos y de extensión, este análisis no va a ser particularmente analizado en los corpus, si no que se mencionará en la interpretación de los clústers. Para obtener el vocabulario específico semanal, se parte de la función “*LexChar*” perteneciente al mismo paquete que la anterior función. La sintaxis utilizada es la siguiente:

```
LDES <- LexChar(TDES, maxDocs=300, maxCharDoc=300, maxPrnDoc=Inf)
```

A continuación, se realizó un Análisis de Correspondencias (AC) con el propósito de simplificar la información y poder analizar los resultados de manera más fácil.

En primer lugar y dado que es la primera vez que se desarrolla este tipo de análisis en el estudio, se define el AC como “una técnica factorial, cualitativa y multidimensional que permite la representación conjunta de las modalidades de distintas variables en un número reducido de sub-espacios factoriales y la proyección sobre éstos de las diferentes categorías o modalidades, e incluso de otras que no han intervenido

activamente en el cálculo y que se denominan ilustrativas o suplementarias” (Álvarez-Esteban, 2015 b).

Para realizar este análisis se utilizó la función “LexCA”, del mismo paquete de análisis también:

```
LCAES <- LexCA(TDES)
```

El análisis se realizará en base a las distintas semanas que componen el periodo de tiempo de los artículos de la muestra, de esta manera, se estudiarán las modalidades de esta variable y sus respectivas palabras conjuntamente en base a su disposición sobre las dos dimensiones mencionadas. Para esto, se utilizarán cuatro gráficos, el gráfico de la dimensión 1 y 2, el gráfico de la dimensión 1 y 3 y, por último, los gráficos de las variables cualitativas suplementarias en dichas dimensiones, las cuales ayudarán a interpretar mejor los resultados. Para realizar los gráficos se han utilizado las siguientes instrucciones:

```
plot(LCAES,selDoc=NULL,selWord=NULL,quali.sup="ALL")
```

```
plot.LexCA(LCAES, axes = c(1,2), quanti.sup = NULL, quali.sup = NULL, selWord = "meta II")
```

```
plot.LexCA(LCAES, axes = c(1,3), quanti.sup = NULL, quali.sup = NULL, selWord = "meta 8")
```

Analizando conjuntamente los dos gráficos que se exponen a continuación, se comenzará por definir la **dimensión o factor 1**, dispuesto en el eje X. Este factor representa un porcentaje de la inercia⁶ total del 9,21%.

La modalidad de la variable semana que caracteriza a esta primera dimensión se sitúa en el lado positivo, y es la semana 4 con una coordenada de 22,26. Aparecen en el mismo lado las palabras “pp” (1,45), “rajoy” (1,37), “diputados” (1,25), “maduro” (0,98), “psoe” (2,28), “abstención” (1,21), “sanchez” (0,95), “gestora” (1,26) y “partido” (0,73).

Por otro lado, en el negativo, no hay ninguna semana que sobresalga por su elevada coordenada, por lo que no hay ninguna que caracterice al factor por su parte negativa.

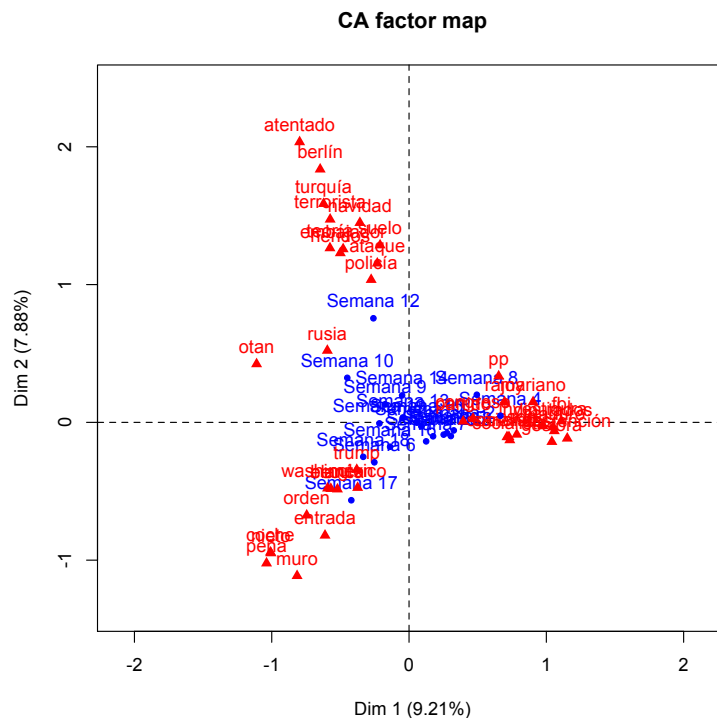
⁶ La inercia o varianza es una medida de dispersión definida como el cuadrado de las desviaciones en relación a la media de una distribución.

Las palabras que están relacionadas con el lado negativo de la dimensión 1 son: “trump” (-1,82), “eeuu” (-0,93) , “rusia” (-0,89) , “otan” (-0,85) y “coche” (-0,74), entre otras.

Con el gráfico en el que aparecen representadas las variables ilustrativas, comprobamos que las palabras que estaban situadas en la parte positiva del factor coinciden con la temática que se posiciona en este gráfico en el mismo lugar, política y asuntos nacionales. Misma situación que en la parte negativa con la temática de asuntos y política internacional. Por otro lado, aparecen en el mismo gráfico las modalidades de la variable suplementaria “Periodico” pero ambas están muy cercanas al valor 0, por lo que no ayudan a la definición de ninguno de los factores de manera significativa.

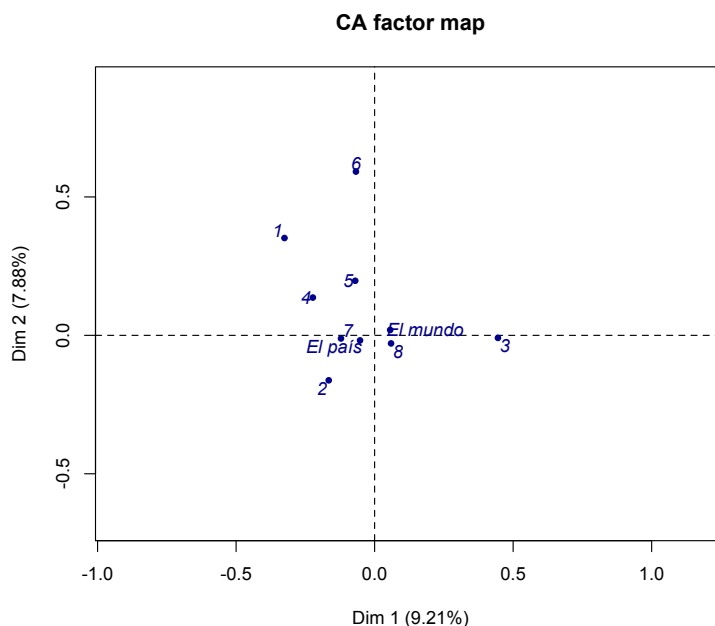
En base a estos resultados el factor 1 se caracteriza de manera clara por las semanas cuyo contenido se centra en la política y los asuntos nacionales españoles, debido a que las palabras relacionadas con esta temática son las que poseen coordenadas más elevadas en términos absolutos.

Gráfico 7.1.1.: Mapa de las palabras sobre las dimensiones 1 y 2 españolas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.1.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 españolas.



Fuente: Elaboración propia a partir de Xplortext.

Una vez analizado el factor 1, se presenta a continuación el estudio del **factor 2**, situado en el eje Y, y que representa el 7,87 % de la varianza total y, que junto al primer factor representa el 17,08%.

En el lado positivo del eje se encuentra la semana 12 con una coordenada de 46,45, por lo que claramente es esta la que define al factor por su lado positivo. junto a ella se sitúan las palabras “atentado” (3,20), “berlín” (2,07), “policía” (1,70), “turquía” (1,60), “terrorista” (1,20) y “ataque” (1,05), entre otras.

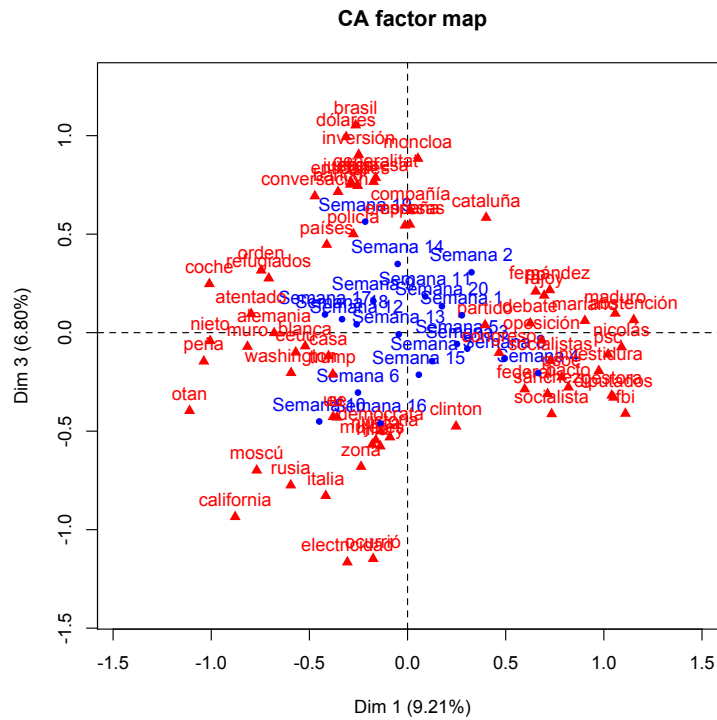
La semana 17 con una coordenada de -24,51 es la que caracteriza al segundo factor por su lado negativo y aparece junto a las palabras “trump” (-1,73), “muro” (-1,23), “eeuu” (-0,78), “coche” (-0,76), “peña” (-0,72), “entrada” (-0,69) y “méxico” (-0,69), entre otras.

A partir del gráfico de las variables ilustrativas se confirma que las palabras relacionadas con la parte positiva del factor pertenecen a la temática 6, sucesos y atentados, mientras que aquellas posicionadas en la parte negativa se relacionan con la temática 2, política y asuntos internacionales. En esta dimensión, ambos periódicos se distancian de la

temática sobre sucesos y atentados, acercándose mucho más a temáticas como la política internacional, la ciencia y la cultura.

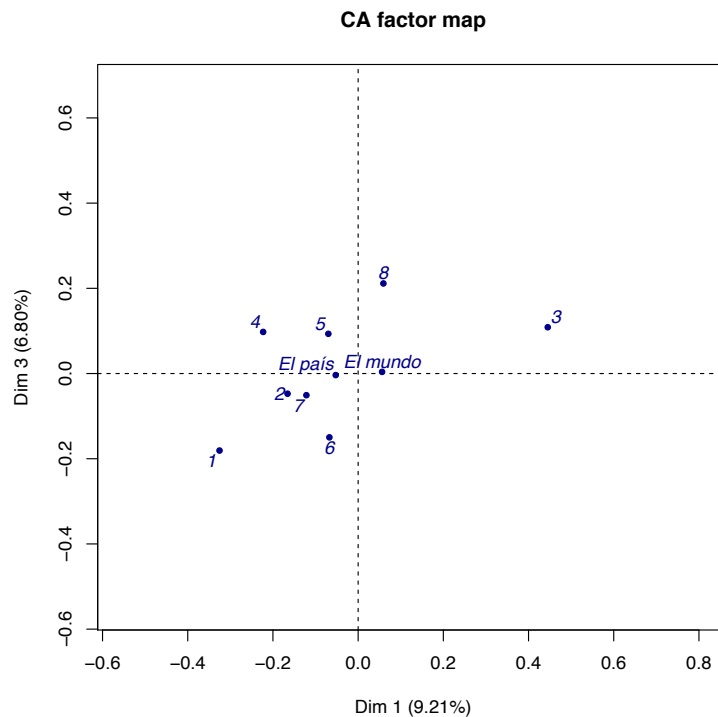
Por tanto la dimensión dos, en este caso, se caracteriza por las semanas cuyo contenido está relacionado con la temática seis, es decir, sucesos y atentados, ya que las palabras relacionadas con esta temática son las que poseen coordenadas más elevadas en términos absolutos.

Gráfico 7.1.3.: Mapa de las palabras sobre las dimensiones 1 y 3 españolas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.1.4.; Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 españolas.



Fuente: Elaboración propia a partir de Xplortext.

Por último, la dimensión 3, la cual se puede estudiar a partir del gráfico 7.1.3., explica un 6,8% de la varianza total y, junto a las dos anteriores, el 23,8%. En la parte positiva se sitúa la semana 19 con una coordenada del 27,30, convirtiéndose en la semana que define a este factor, y las palabras “dólares” (1,77), “españa” (1,49), “empresa” (0,93), “países” (0,90), “justicia” (0,86), “brasil” (0,79) e “inversión” (0,69), entre otras.

En el lado contrario es la semana 10 la que caracteriza a esta dimensión, con una coordenada de -17,92. Las palabras situadas en esta parte del eje Y son rusia (-2,04), “clinton” (-0,77), “trump” (-0,75), “zona” (-0,71) y “california” (-0,62), entre otras.

El gráfico de variables ilustrativas (gráfico 7.1.4) no presenta resultados significativos para ninguna de las dos variables representadas en esta dimensión.

Por tanto, la diferencia entre las semanas marcadas por contenido sobre política internacional y aquellas cuyo contenido destacado se basa en la economía nacional son las que definen el factor tres, ya que poseen las coordenadas más altas.

Una vez realizado el AC, se realiza un análisis clúster para identificar de manera significativa grupos de semanas que, según su contenido, poseen similitud. La sintaxis utilizada para realizar el análisis es la siguiente:

```
HCCAES <- LexHCca(LCAES, graph = TRUE, max=5, nb.clust = 0, order = TRUE)
```

Como se podrá ver en el *dendrograma*⁷ y mapa de factores, que aparecen en el final de este apartado, se diferencian cinco clusters o grupos similares:

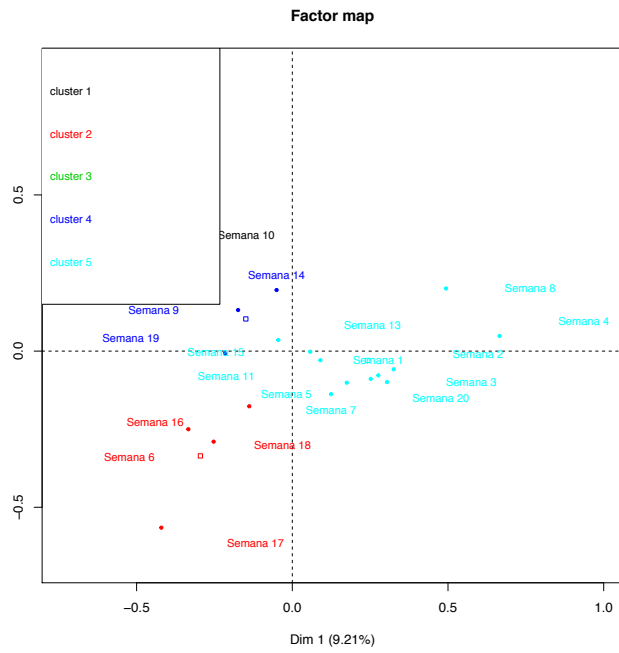
- El clúster número 1 está formado tan solo por la semana 10, que comprende la semana del 5 al 11 de diciembre de 2016. A partir del vocabulario específico se puede determinar el vocabulario más frecuente de esta semana, el cual se caracteriza por las palabras “rusia” con una frecuencia de 55 repeticiones en el contenido de este periodo, “otan” con una frecuencia de 19, “california” con 16 y “rusa” con 13. No es casualidad que estas palabras sean las que definan el contenido de la semana, ya que los artículos de prensa de esa semana se centraron en la sospecha de la ayuda rusa a la campaña de Donald Trump para ganar las elecciones estadounidenses.
- El clúster número 2, está compuesto por la semana 6 (7-13 de noviembre de 2016), semana 17 (23-29 de enero de 2017) y semana 18 (30 de enero-5 de febrero de 2017). No es casualidad que estas tres semanas formen un solo grupo, ya que como se vio en el AC estas semanas se sitúan en las partes negativas de las dos dimensiones, cuyo tema principal es la política internacional. Atendiendo al vocabulario específico, estas tres semanas centran su vocabulario en este tema, siendo la palabra “trump” la más utilizada con una frecuencia de 302 veces durante estos periodos. El motivo de esto es claro, ya que estas semanas son claves en la situación electoral estadounidense; en la semana 6 tiene lugar el día 8 de noviembre las elecciones presidenciales; en la semana 17, el presidente electo, Donald Trump, fue recién instalado en la Casa Blanca y asumía definitivamente el poder del país; y, por último, la semana 18 fue una semana marcada por la puesta en marcha de las

⁷ Gráfico que representa los datos en forma de árbol que sintetiza el proceso de agrupación de los clusters.

controvertidas propuestas sobre asuntos exteriores de su campaña electoral, lo cual tuvo un gran impacto mundial, y por ende, en la prensa internacional.

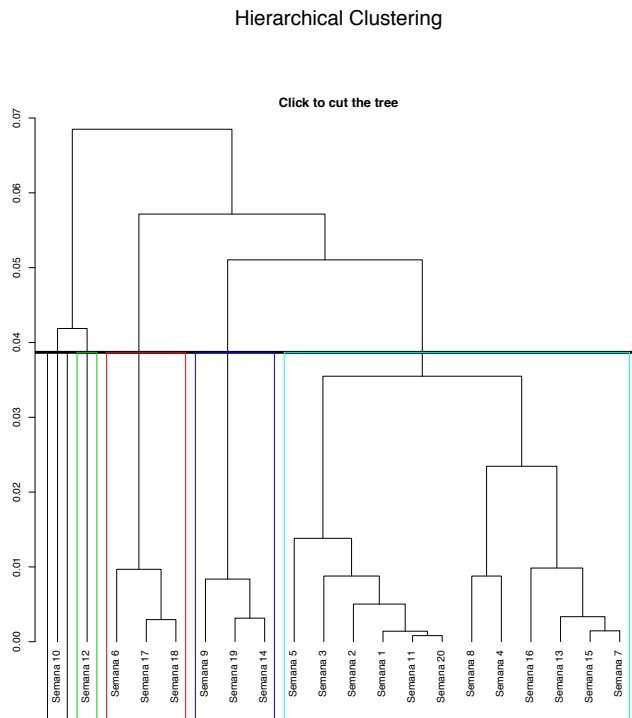
- El clúster número 3 está compuesto por la semana 12, que comprende el periodo del 19 al 25 de diciembre de 2016. El vocabulario específico de esta semana define a la perfección la temática que ocupó los titulares de la prensa, con tan solo las primeras palabras con más frecuencia se esclarece que fueron días marcados por los atentados terroristas (temática 6) de Alemania y Turquía: “atentado” (26), “berlín” (19) y “turquía” (17).
- El clúster número 4 se compone de las semanas 9 (28 de noviembre-4 de diciembre de 2016), 14 (2-8 de enero de 2017) y 19 (6-12 de febrero de 2017). El vocabulario específico de estas semanas define el tema por el que se han juntado en un mismo grupo. Las palabras más frecuentes son “españa” con 86 repeticiones, “dólares” con 51 repeticiones y “millones” con 71, por lo que dejan claro que el contenido de este periodo fue marcado por la economía y los asuntos económicos nacionales (temática 5)
- Por último, el clúster 5 agrupa un gran número de modalidades de la variable “Semana”, y atendiendo al vocabulario más frecuente entre el que destaca “gobierno” con 261 repeticiones, “psoe” con 188, “partido” con 186, “pp” con 147 y “podemos” con 143, se puede determinar que estas semanas se caracterizan por un contenido relacionado con la política nacional.

Gráfico 7.1.5.: Mapa de grupos de semanas similares de España.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.1.6.: Dendrograma España.



Fuente: Elaboración propia a partir de Xplortext.

7.2. ANÁLISIS FRANCIA

La base de datos francesa se realizó a partir de los artículos de los dos periódicos más relevantes de la nación, “Le Figaro” y “Le Monde”.

Como en el caso anterior, el primer análisis realizado a esta base de datos fue un análisis de datos textuales para obtener las palabras más características de la muestra, a partir de la función “*TextData*”. Los resultados se muestran en la siguiente tabla:

Tabla 7.2.1.: Las 50 palabras más características del vocabulario global de la base de datos francesa.

No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.
1.	plus (<i>más</i>)	826	19.	euros (<i>euros</i>)	227	36.	ainsi (<i>así</i>)	168
2.	comme (<i>como</i>)	370	20.	monde (<i>mundo</i>)	225	37.	paris	168
3.	être (<i>ser</i>)	362	21.	françois	224	38.	autres (<i>otro</i>)	162
4.	france (<i>francia</i>)	350	22.	avoir (<i>tener</i>)	217	39.	candidat (<i>candidato</i>)	162
5.	deux (<i>dos</i>)	340	23.	très (<i>muy</i>)	216	40.	mois (<i>mes</i>)	161
6.	fait (<i>hecho</i>)	299	24.	ministre (<i>ministro</i>)	213	41.	primaire (<i>primaria</i>)	156
7.	aussi (<i>también</i>)	299	25.	trump	212	42.	encore (<i>aún</i>)	155
8.	président (<i>presidente</i>)	292	26.	français (<i>francés</i>)	206	43.	plusieurs (<i>varias</i>)	153
9.	si (<i>tan</i>)	275	27.	dont (<i>cuyo</i>)	202	44.	moins (<i>menos</i>)	153
10.	depuis (<i>desde</i>)	274	28.	premier (<i>primero</i>)	201	45.	gauche (<i>izquierda</i>)	152
11.	pays (<i>país</i>)	267	29.	bien (<i>bien</i>)	197	46.	etat (<i>estado</i>)	149
12.	entre (<i>entre</i>)	261	30.	où (<i>donde</i>)	192	47.	tous (<i>todos</i>)	148
13.	tout (<i>todo</i>)	250	31.	politique (<i>política</i>)	191	48.	personnes (<i>personas</i>)	147
14.	après (<i>después</i>)	249	32.	alors (<i>entonces</i>)	179	49.	non (<i>no</i>)	145
15.	faire (<i>hacer</i>)	246	33.	figaro	178	50.	également (<i>igualmente</i>)	144
16.	ans (<i>años</i>)	241	34.	avant (<i>antes</i>)	177			
17.	selon (<i>según</i>)	236	35.	peut (<i>puede</i>)	176			
18.	contre (<i>contra</i>)	233						

Fuente: Elaboración propia

Tal y como muestra la tabla anterior, se pueden destacar las palabras más relevantes para relacionarlas con los temas que destacan en la prensa francesa en el periodo estudiado. Entre las palabras más características se pueden distinguir: “france” (*Francia*) con una frecuencia de 350, “président” (*presidente*) con 292, “euros” (*euros*) con 227, “monde” (*mundo*) con 225, “françois” con 224, “ministre” (*ministro*) con 213, “trump” con 212, “politique” (política) con 191 y “candidat” (*candidato*) con 162.

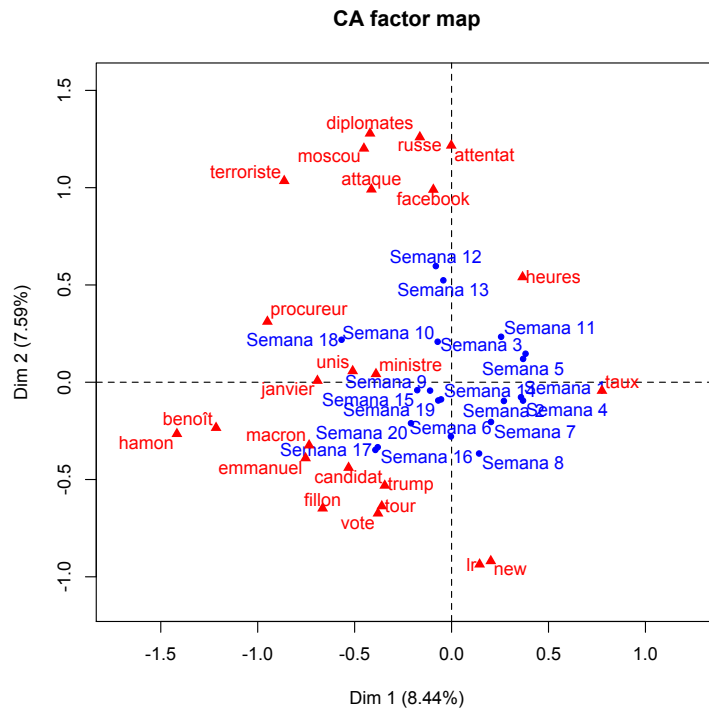
Nuevamente, según las palabras que hemos señalado como relevantes, destacamos dos temas principales entre los artículos recogidos en la base de datos:

- La situación política nacional francesa, lo cual era algo a esperar dado que las elecciones presidenciales francesas tuvieron fecha el 23 de abril de 2017 (primera vuelta) y el 7 de mayo de este mismo año (segunda vuelta). Por esto, dado que el análisis está basado en los meses previos a las elecciones se puede comprobar como la prensa francesa centró mucho su contenido en la política nacional y en las campañas de los respectivos partidos.
- La situación política estadounidense. Al igual que en el resto del mundo la prensa también centró su contenido en las elecciones estadounidenses y su posterior situación política.

A continuación, como en el caso anterior, se caracterizó el nuevo objeto creado a partir de “*TextData*” para obtener el vocabulario específico de cada semana que servirá para el posterior análisis clúster, y un AC agregado por semanas, para simplificar la información y facilitar la interpretación de los resultados.

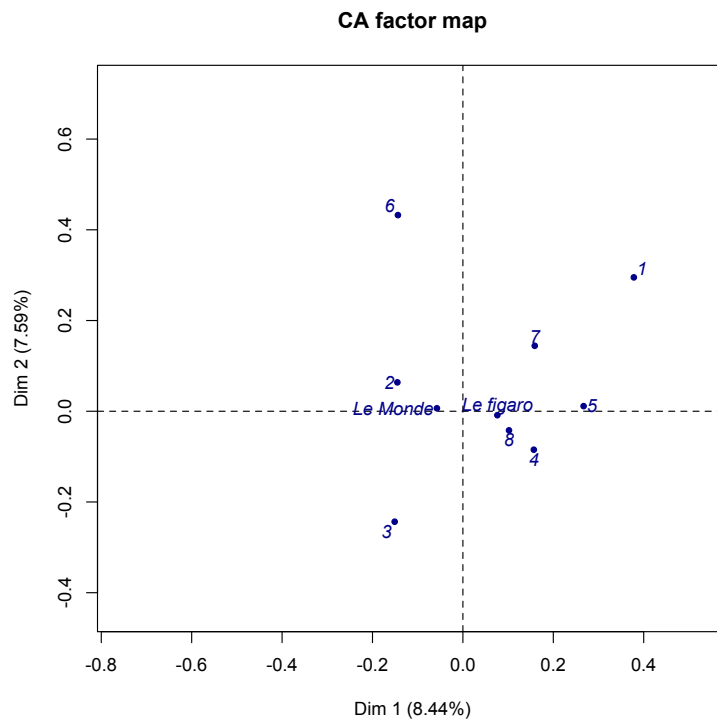
Tras ejecutar las instrucciones para el análisis en el programa estadístico R, la salida nos proporciona los gráficos, que de manera similar al caso español, se analizarán conjuntamente.

Gráfico 7.2.1.: Mapa de las palabras sobre las dimensiones 1 y 2 francesas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.2.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 francesas.



Fuente: Elaboración propia a partir de Xplortext.

Comenzando por la **primera dimensión**, que representa el 8,44% de la varianza total, no hay ninguna modalidad de la variable semana que defina la parte positiva del factor. Las palabras que aparecen en este lado del factor son: “taux” (*tasa*) (1,06), “sarkozy” (0,55), “milliards” (*millones*) (0,52) , “étude” (*estudio*) (0,51), “octobre” (*octubre*) (0,43), “euros” (*euros*) (0,35), “argent” (*dinero*) (0,29), “etat” (*estado*) (0,26), “défense” (*defensa*) (0,25), “mossoul” (*Mousul*) (0,233), “prix” (*precio*) (0,22), entre otras.

En el lado negativo de la primera dimensión aparece la semana 18 con una coordenada de -30,03, por lo que es la que define a este factor. Junto a esta semana se posicionan también las palabras características “hamon” (-2,59) , “janvier” (*enero*) (-1,51) , “fillon” (-1,41) , “macron” (-1,30) , “benoît” (-1,03) y “candidat” (*candidato*) (-1,03) , entre otras.

Por otro lado, en cuanto a las variables ilustrativas, las temáticas de las que se habló durante el análisis del eje coinciden con el vocabulario situado en su misma posición. Relacionado con la variable “Periódico” ambas modalidades se sitúan cerca del 0, por lo que, aún estando “Le Monde” más cerca de la política nacional francesa, la interpretación no es significativa.

En base a los resultados obtenidos se puede ya dar nombre a esta dimensión, dado que la parte positiva corresponde a vocabulario económico en sus coordenadas positivas más altas, y la parte negativa a vocabulario relacionado con la política nacional, por lo que el factor uno se caracteriza por el vocabulario relacionado con la política nacional francesa, ya que posee mayores valores de coordenadas.

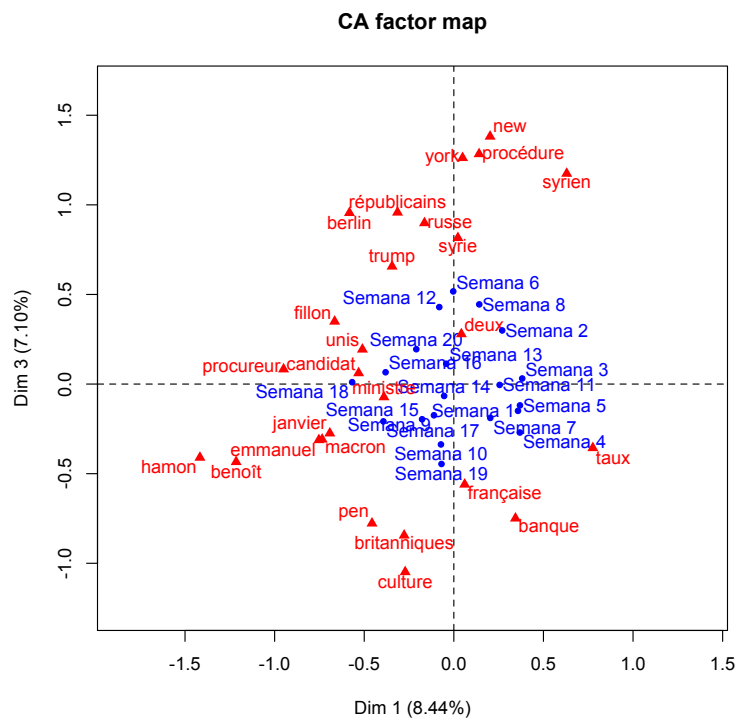
En cuanto a la **dimensión dos**, que representa el 7,5% de la inercia total y junto al factor 1 el 16,02%, se encuentran en su lado positivo las semanas que caracterizan al factor que son la 13 y la 12 con coordenadas de 23,65 y 21,92, respectivamente.

En cuanto al vocabulario, destacan en el lado positivo las palabras “attaque” (*ataque*) (1,70), “russe” (*ruso*) (1,24), “attentat” (*atentado*) (1,19), “terroriste” (*terrorista*) (0,83), “diplomates” (*diplomáticos*) (0,82), “heures” (*horas*) (0,74) y “moscou” (*Moscú*) (0,72), entre otras, mientras que por el lado negativo se señalan por su relevancia:

“trump” (-1,50), “fillon” (-1,49), “tour” (*vuelta*) (-0,80), “candidat” (*candidato*) (-0,78), “vote” (*voto*) (-0,76) y “donald” (-0,57).

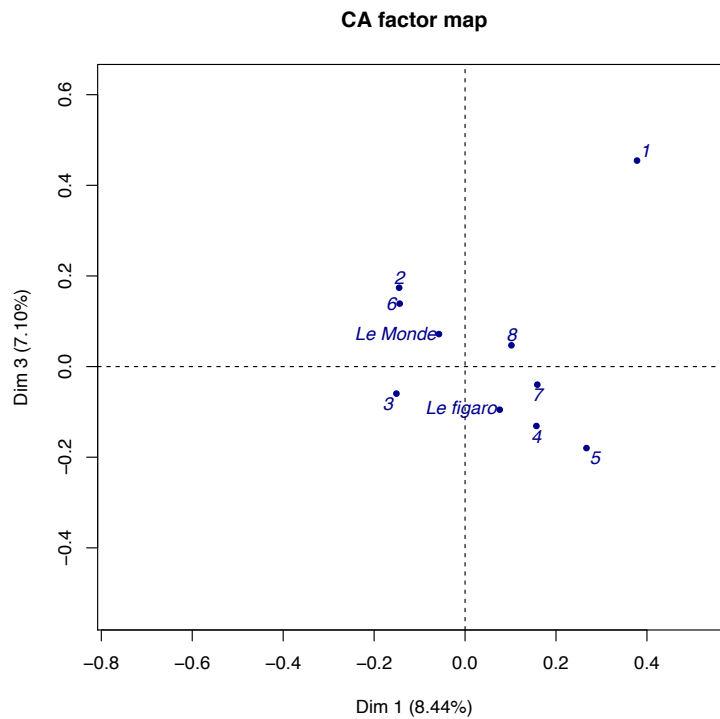
Analizando el gráfico de las variables suplementarias, el vocabulario que acaba de ser expuesto coincide con la temática que se posiciona sobre el eje en el mismo lugar, relacionándose la parte positiva del factor con la temática 6, sucesos y atentados, y la parte negativa con la política nacional.

Gráfico 7.2.3.: Mapa de las palabras sobre las dimensiones 1 y 3 francesas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.2.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 francesas.



Fuente: Elaboración propia a partir de Xplortext.

Por último, en la parte positiva de la **dimensión tres**, que representa el 7,09% de la varianza y junto a las anteriores el 23,12% del total, se sitúan las semanas que caracterizan a este factor, que son las semanas 6 y 8 con coordenadas de 17,80 y 17,69, respectivamente. En su misma posición destacamos vocabulario como: “trump” (2,46), “procédure” (*procedimiento*) (1,51), “berlin” (0,93), “syrie” (*Siria*) (0,93) y “republicains” (*republicanos*) (0,71), entre otras.

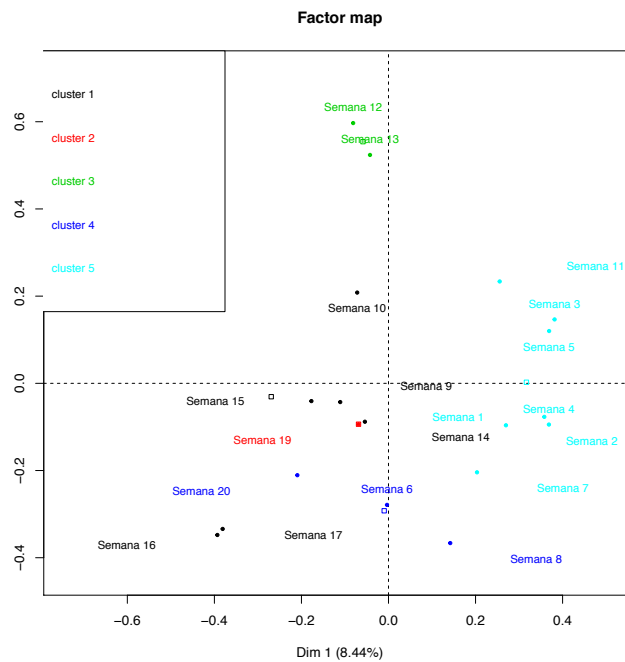
Por el lado negativo de la dimensión no hay semanas concretas que definan a la dimensión. Las palabras situadas en esta parte son: “culture” (*cultura*) (-1,18), “française” (*francesa*) (-0,74), “pen” (-0,73), “banque” (*banco*) (-0,69), “britanniques” (*británicos*) (-0,69), “dollars” (*dólares*) (-0,57).

Según el vocabulario descrito y las coordenadas que aparecen entre paréntesis no se puede interpretar de manera clara la temática que define al factor, por lo que se recurre a la gráfica de las variables suplementarias, que ayudan a la interpretación del mismo. Gracias a este gráfico se comprueba que es la temática 1 (guerra y conflictos bélicos) la

que caracteriza a esta dimensión, aunque comprobamos también que la palabra “trump” posee coordenadas altas y la temática 2, según el gráfico de variables suplementarias, también posee un peso importante en esta dimensión. Por tanto, serían estas dos temáticas las que caracterizarían al factor. Además, señalar que el periódico “Le Monde” se sitúa en la parte positiva de la dimensión, por lo que se relaciona más con la publicación sobre este tipo de temáticas que “Le Figaro”.

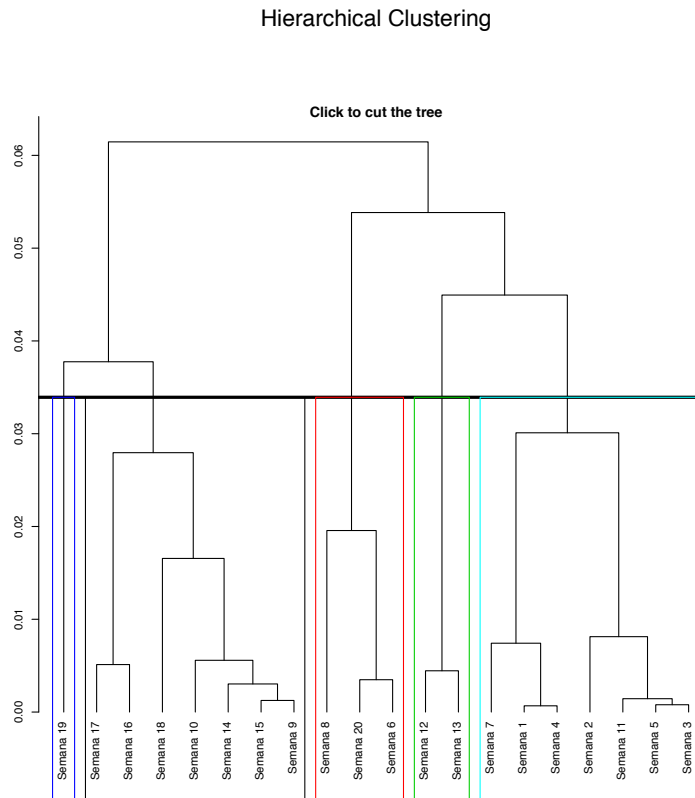
Tras el AC, se realizó un análisis clúster, como en el caso anterior, para crear grupos de semanas que coinciden en temática de contenido. La exposición de la interpretación de los resultados se realizará nuevamente a partir de los gráficos proporcionados por el programa estadístico:

Gráfico 7.2.5.: Mapa de grupos de semanas similares de Francia.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.2.6.: Dendrograma Francia.



Fuente: Elaboración propia a partir de Xplortext.

Tomando como base los gráficos, se pueden diferenciar cinco clusters o grupos similares:

- El cluster número 1 está compuesto por las semanas 9, 10, 14, 15, 16, 17 y 18. Dentro del vocabulario específico de estas semanas se destacan palabras como: “ministre” (*ministro*) con 119 repeticiones, “premier” (*primero*) con 104, “janvier” (*enero*) con 101, “candidat” (*candidato*) con 92, “fillon” con 78, “macron” con 66, “hamon” con 54 y “benoît” con 28.

Por tanto, la razón de similitud es clara, el contenido principal de todas estas semanas fue la temática relacionada con la política nacional francesa, debido a sus, entonces próximas, elecciones a presidente del gobierno.

- El clúster número dos es la semana 19, entre el vocabulario común específico destacamos “culture” (*cultura*) con 24 repeticiones, “pen” con 21 y “française” con 24. Teniendo en cuenta el conjunto de las palabras específicas de este periodo y el lugar donde el grupo está situado en los ejes, se esclarece el motivo por el que esta semana aparece como un solo grupo, y es que su contenido principal es la cultura francesa.
- El clúster tres se compone de las semanas 12 y 13. En el vocabulario específico común de estas semanas destacan “russe” (*ruso*), “attentat” (*atentado*), “attaque” (*ataque*), “russie” (*Rusia*), “décembre” (*diciembre*) e “israël” (*Israel*). Estas palabras esclarecen la temática que une en el mismo grupo a estos tres periodos de tiempo, la cual es la número 6, sucesos y atentados.
- El clúster cuatro está formado por las semanas 6, 8 y 20, y entre el vocabulario específico se destacan las palabras “trump”, “donald”, “procédure” (*procedimiento*), “républicains” (*republicanos*) y “président” (*presidente*). Por tanto estas semana se centraron su contenido en el nuevo presidente de los Estados Unidos, Donald Trump.
- Por último, el clúster número 5 está formado por las semanas 1, 2, 3, 4, 5, 7 y 11. Del vocabulario común específico se extraen palabras como “taux” (*tasa*), “impôt” (*impuesto*), “étude” (*estudio*), “euros” (*euros*), “milliards” (*millones*), y “octobre” (*octubre*), lo cual deja ver que este grupo está formado por las semanas que han centrado su contenido en economía y asuntos económicos.

7.3. ANÁLISIS REINO UNIDO

La base de datos británica fue elaborada a partir de las noticias publicadas digitalmente por los dos diarios más leídos de este país, “The Guardian” y “The Sun”.

En primer lugar se realizó un análisis de datos textuales para la obtención de las palabras más frecuentes del corpus con el mismo procedimiento que en los casos anteriores. En la siguiente tabla se exponen las 50 primeras de estas palabras:

Tabla 7.3.1.: Las 50 palabras más características del vocabulario global de la base de datos británica.

No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.
1.	said (<i>dijo</i>)	1580	20.	president (<i>presidente</i>)	204	37.	way (<i>manera</i>)	150
2.	will (<i>voluntad</i>)	825	21.	first (<i>primero</i>)	194	38.	security (<i>seguridad</i>)	150
3.	trump	532	22.	russia (<i>rusia</i>)	193	39.	make (<i>hacer</i>)	149
4.	us (<i>nosotros</i>)	463	23.	years (<i>años</i>)	191	40.	say (<i>decir</i>)	147
5.	people (<i>gente</i>)	460	24.	russian (<i>ruso</i>)	191	41.	since (<i>desde</i>)	145
6.	government (<i>gobierno</i>)	432	25.	two (<i>dos</i>)	187	42.	european (<i>européo</i>)	142
7.	uk (<i>reino unido</i>)	403	26.	just (<i>solo</i>)	186	43.	added (<i>añadido</i>)	139
8.	new (<i>nuevo</i>)	398	27.	guardian (<i>guardián</i>)	178	44.	party (<i>partido</i>)	139
9.	also (<i>también</i>)	361	28.	many (<i>muchos</i>)	176	45.	get (<i>obtener</i>)	138
10.	one (<i>uno</i>)	342	29.	britain (<i>gran britaña</i>)	170	46.	even (<i>incluso</i>)	138
11.	can (<i>poder</i>)	319	30.	minister (<i>ministro</i>)	166	47.	including (<i>incluyendo</i>)	138
12.	year (<i>año</i>)	315	31.	world (<i>mundo</i>)	159	48.	take (<i>tomar</i>)	137
13.	eu (<i>unión europea</i>)	309	32.	country (<i>país</i>)	154	49.	week (<i>semana</i>)	137
14.	may	271	33.	like (<i>gustar</i>)	154	50.	law (<i>ley</i>)	133
15.	last (<i>último</i>)	246	34.	work (<i>trabajo</i>)	151			
16.	brexit	233	35.	labour (<i>trabajo</i>)	151			
17.	now (<i>nuevo</i>)	221	36.	public (<i>público</i>)	150			
18.	time (<i>tiempo</i>)	219						
19.	told (<i>contó</i>)	206						

Fuente: Elaboración propia

Entre las palabras características que se pueden ver en la tabla anterior, sobresalen por su contexto “trump”, “government” (*gobierno*), “uk” (*Reino Unido*), “eu” (*Unión Europea*), “*brexit*”⁸, y “president” (*presidente*), ya que dejan ver desde un primer momento los temas más característicos de la prensa británica durante estos cinco meses:

- La situación de inestabilidad política que atraviesa Reino Unido por el reciente referéndum a favor del “*Brexit*” y las próximas elecciones presidenciales previstas para junio de 2017, provocaron la poca variación de contenido de la prensa durante estos meses estudiados.
- Por otro lado, la política internacional vuelve a ser protagonista también en este caso debido a las elecciones a presidente de Estados Unidos, y la posterior victoria del magnate Donald Trump.

Tras la realización del análisis de correspondencias con la variable “semana” como variable de agregación, se presentan los siguientes resultados:

En la parte positiva de la dimensión 1 (eje X), las semanas que por sus coordenadas más elevadas contribuyen a la caracterización de la dimensión son las semanas 14 y 12. En el mismo lado de este eje aparecen palabras como “basic” (*básico*) (0,89), “health” (*salud*) (0,85), “nhs”⁹ (0,75), “water” (*agua*) (0,69), “funding” (*financiación*) (0,69), “search” (*buscar*) (0,63), “drivers” (*conductores*) (0,63), “energy” (*energía*) (0,52), “investment” (*inversión*) (0,46) y “money” (*dinero*) (0,46), entre otras.

Por el lado negativo, las modalidades de semana que más caracterizan al factor son la número 15 y la número 6 y junto a ellas, las palabras “trump” (-3,52), “moscow” (*Moscú*) (-1,17), “president” (*presidente*) (-0,97), “supreme” (*supremo*) (-0,90), “russian” (*ruso*) (-0,89), “article” (*artículo*) (-0,79), “troops” (*tropas*) (-0,77), “us” (*Estados Unidos*) (-0,76) y “obama” (-0,73), entre otras.

⁸ Es el acrónimo de las palabras “Britain” y “exit” utilizada para referirse al proceso político en curso que tiene como objetivo la desanexión de Reino Unido de la Unión Europea.

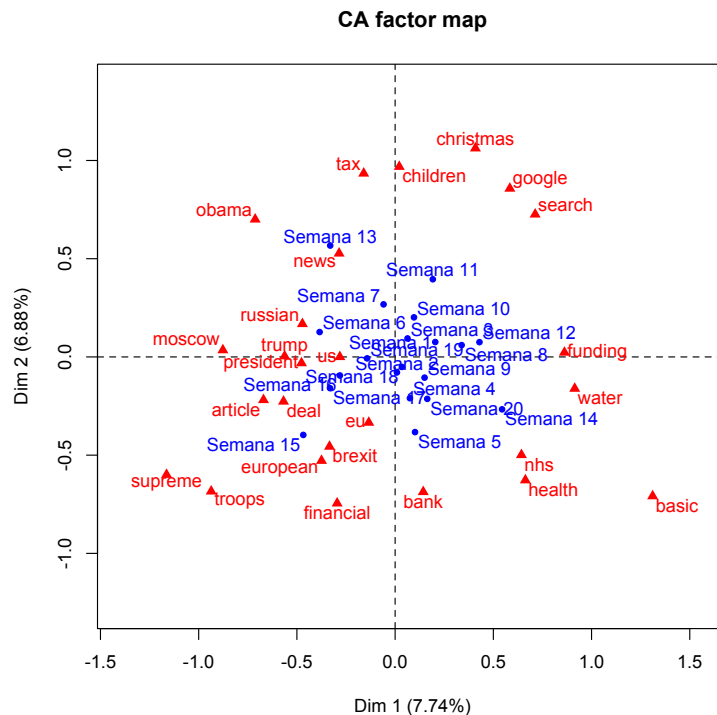
⁹ NHS es el acrónimo que utilizan los británicos para referirse a su Servicio Nacional de Salud.

A partir del gráfico de las variables suplementarias se comprueba que las palabras examinadas en el párrafo previo coinciden en temática con la modalidad de esta variable, siendo la temática 5 (economía nacional) y la temática 2 (política internacional) las que definen al factor desde sus lados opuestos.

Las modalidades de la variable “Periódico” están posicionadas en el 0, por lo que no ayudan a explicar ninguno de estos dos factores (dim 1 y dim 2).

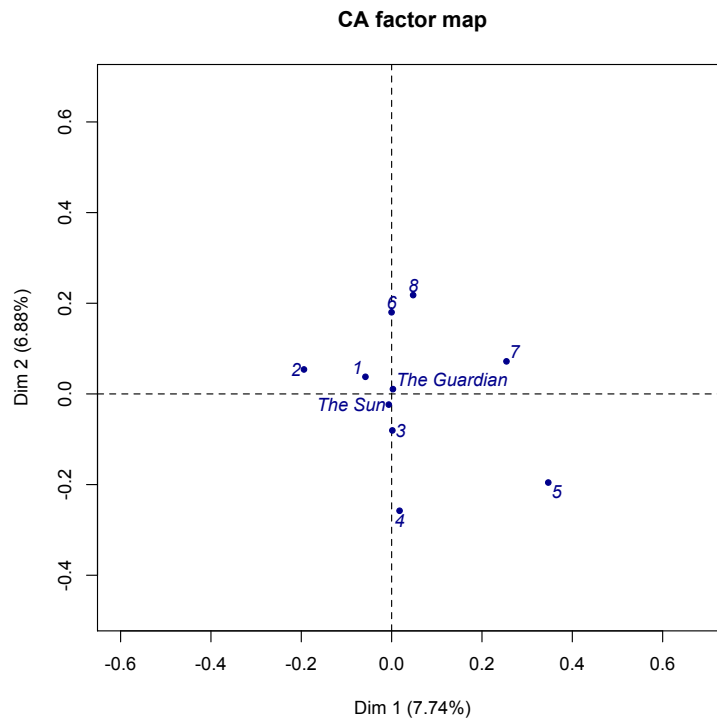
Por tanto, en base al análisis anterior, la dimensión 1 la definen los contenidos de la semana 14 (16,93) cuyas temáticas principales son las temática de ciencia y tecnología y la economía nacional, y por otro lado, los contenidos de la semana 15 (-20,30) que basa su contenido en la temática de la política internacional.

Gráfico 7.3.1.: Mapa de las palabras sobre las dimensiones 1 y 2 británicas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.3.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 británicas.



Fuente: Elaboración propia a partir de Xplortext.

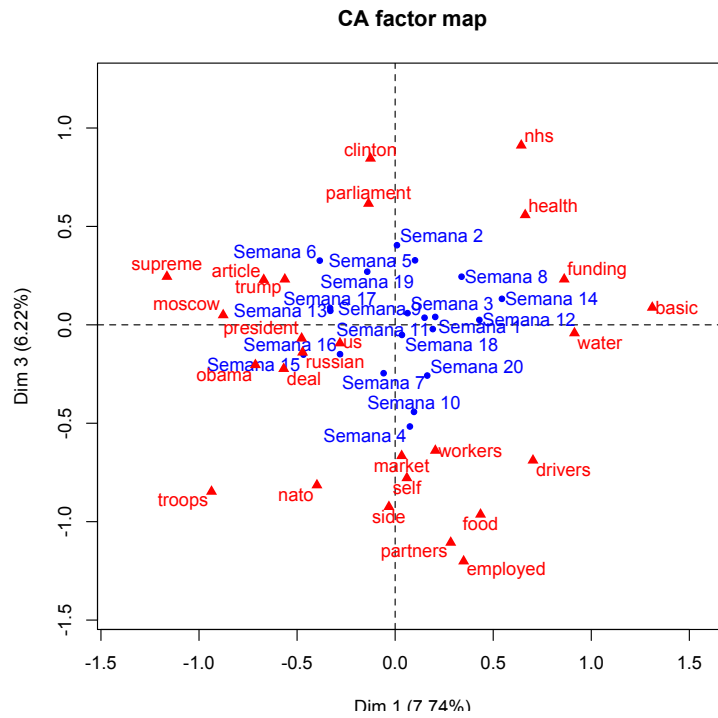
En cuanto a la dimensión 2, las semanas que más contribuyen a la explicación del factor son la 13 y 11 por el lado positivo y la 15 por el lado negativo.

Respecto a las palabras asociadas a cada parte del eje, comenzando por la parte positiva se encuentran: “children” (*niños*) (2,49), “tax” (*tasa*) (1,97), “google” (1,10), “christmas” (*Navidad*) (0,93) y “obama” (0,79), entre otras. Por el lado negativo se encuentran: “brexit” (-1,14), “bank” (*banco*) (-1,06), “european” (*europeo*) (-0,93), y “financiamiento” (*financiero*) (-0,88), entre otras.

En el gráfico de las variables ilustrativas se puede señalar que las modalidades de la variable temática que ayudan a explicar el factor son la 4 y 5 por la parte negativa, economía internacional y economía nacional respectivamente, y por el lado positivo, la temática 6, sucesos y atentados, y la 8, cultura.

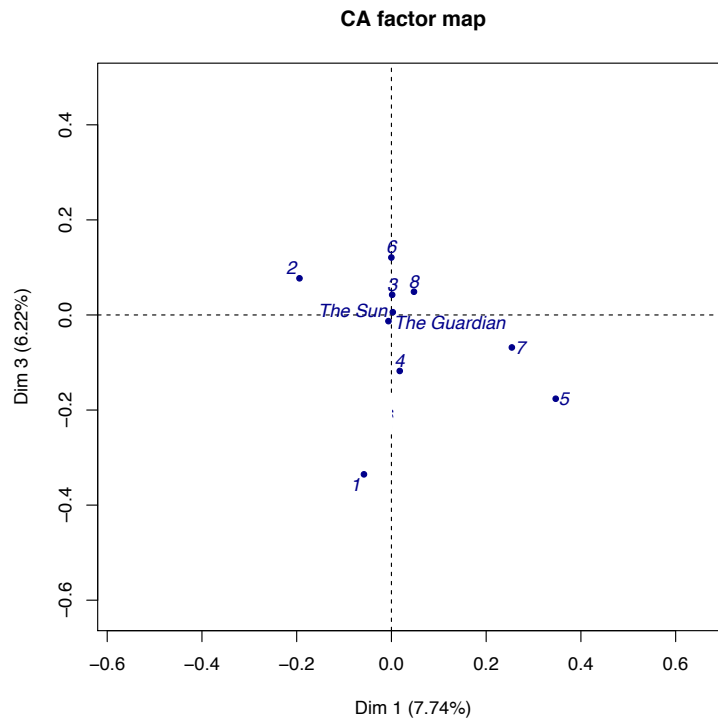
Siguiendo lo anterior, se concluye que la dimensión dos se caracteriza por las temáticas de economía.

Gráfico 7.3.3.: Mapa de las palabras sobre las dimensiones 1 y 3 británicas.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.3.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 británicas.



Fuente: Elaboración propia a partir de Xplortext.

Por último, la dimensión tres se define a partir de las semanas 4 (-24,9) y 10 (-16,8).

Por el lado positivo se sitúan palabras como “nhs” (1,88), “clinton” (1,67), “parliament” (*parlamento*) (0,98), “health” (*salud*) (0,75) y “trump” (0,74), entre otras; mientras que, en el lado negativo aparecen: “nato” (*OTAN*) (-1,38), “employed” (*empleado*) (-0,97), “food” (*comida*) (-0,91), “side” (*lado*) (-0,91) y “partners” (*asociados*) (-0,82), entre otras.

De nuevo, a partir del vocabulario que se posiciona sobre la dimensión no se puede ver claramente la temática que define a esta dimensión, por lo que se vuelve a recurrir a la ayuda de las variables suplementarias. En el gráfico de estas, se comprueba que la temática 1 se aleja significativamente del 0, por lo que es la que caracteriza a este factor.

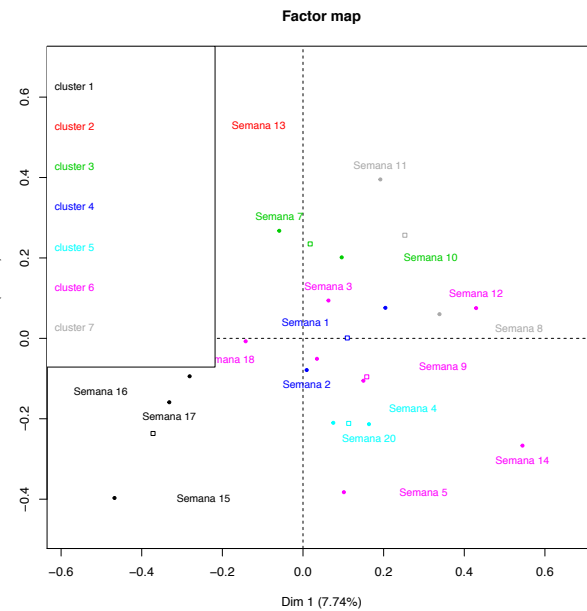
Tras el análisis de correspondencias se realizó, de nuevo, un análisis clúster. En este caso se presentan 7 grupos en vez de 5 como en los casos presentados con anterioridad, el motivo es la interpretación de los resultados, ya que en el caso de Reino Unido la agrupación de todas las semanas en tan solo cinco clústers provocó que se juntaran semanas cuyas temáticas eran diferentes, creando con ello grupos con temas muy variados en los que identificar la razón de similitud entre las semanas agrupadas era complicado.

Siguiendo el procedimiento anterior e identificando el vocabulario específico de los grupos, se pueden identificar las temáticas por las que han sido agrupadas las distintas semanas. En este caso y los sucesivos, tras haber sido expuesto en anteriores ocasiones el procedimiento de identificación de los distintos clústers, el análisis se centrará en la definición de aquellos grupos cuya temática principal tenga que ver con el caso a estudiar, es decir, con la política internacional.

Por tanto, en base a lo anterior, son los clusters 1 y 2 aquellos cuyas temáticas principales son la política internacional. Nuevamente, la razón se encuentra en los acontecimientos que sucedieron a lo largo de las semanas que agrupan estos clusters. El grupo 1 está compuesto por las semanas 15, 16 y 17, periodo que estuvo marcado por la instalación de Donald Trump en la Casa Blanca y su definitiva toma del control del gobierno de los Estados Unidos. En cuanto al grupo 2, se forma de las semanas 6 y 13

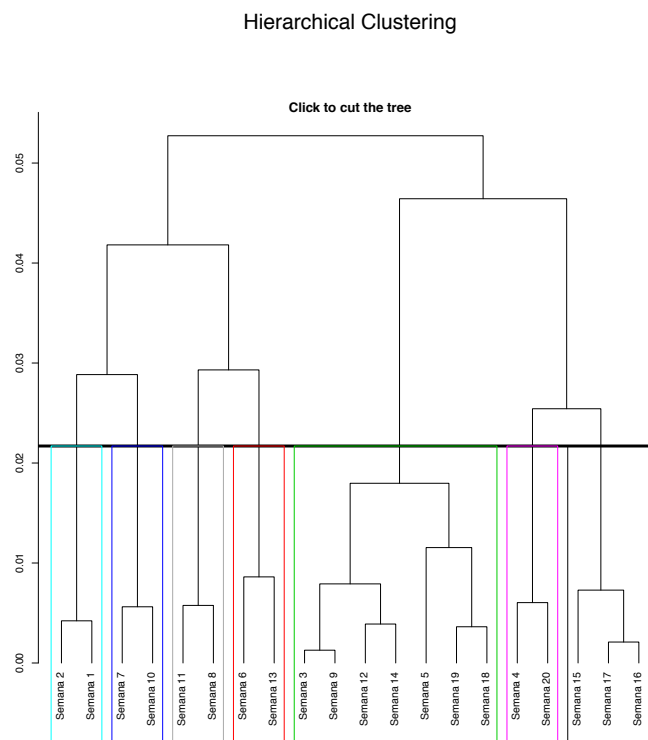
donde tuvieron lugar las elecciones de EEUU y el enfrentamiento entre este país y Rusia por el ataque informático realizado por los hackers rusos, respectivamente.

Gráfico 7.3.5.: Mapa de grupos de semanas similares de Reino Unido.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.3.6.: Dendrograma Reino Unido.



Fuente: Elaboración propia a partir de Xplortext.

7.4. ANÁLISIS REGIÓN ÁRABE

En el caso de la base de datos árabe, no se refiere tan solo a un país, si no a todos los países en los que se habla este idioma, ya que poseen en común dos periódicos que predominan en número de lectores sobre el resto, tal es así, que ambos periódicos trabajan en la publicación de sus artículos también en lengua inglesa debido a su gran alcance. Estos dos periódicos son “Al Jazeera” y “Al Arabiya”.

Tras su análisis textual, las 50 palabras más frecuentes del corpus se muestran en la siguiente tabla:

Tabla 7.4.1.: Las 50 palabras más características del vocabulario global de la base de datos árabe.

No .	Palabras más frecuentes	Frec.	No .	Palabras más frecuentes	Frec.	No .	Palabras más frecuentes	Frec .
1.	(el presidente) الرئيس	310	18.	(oposición) المعارضة	157	35.	(el año) العام	114
2.	(unida) المتحدة	303	19.	(consejo) مجلس	156	36.	(ministro) وزير	113
3.	(entre) بين	291	20.	(aquello) ذلك	148	37.	(Mousul) الموصل	113
4.	(Trump) ترمب	277	21.	(la isla) الجزيرة	144	38.	(Alepo) حلب	112
5.	(y ha dicho) وقال	276	22.	(añadido) وأضاف	142	39.	(internacional) الدولي	110
6.	(después) بعد	252	23.	(exterior) الخارجية	136	40.	(millar) مليار	110
7.	(americana) الأميركية	241	24.	(dolar) دولار	135	41.	(las fuerzas) القوات	109
8.	(durante) خلال	226	25.	(presidente) رئيس	133	42.	(ciudad) مدينة	108
9.	(esta) هذه	209	26.	(Siria) سوريا	129	43.	(seguridad) الأمن	106
10.	(régimen) النظام	190	27.	(el pasado) الماضي	128	44.	(Egipto) مصر	104
11.	(arábica) العربية	186	28.	(todo) كل	124	45.	(el gobierno) الحكومة	103
12.	(americano) الأميركي	185	29.	(Rusia) روسيا	121	46.	(corresponsal) مراسل	101
13.	(estados) الولايات	183	30.	(distinto) غير	118	47.	(Donald) دونالد	100
14.	(hoy) اليوم	176	31.	(Clinton) كلينتون	118	48.	(organizar) تنظيم	100
15.	(dijo) قال	161	32.	(estaba) كان	117	49.	(número) عدد	92
16.	(fuerzas) قوات	158	33.	(el sirio) السوري	117	50.	(el ejército) الجيش	91
17.	(antes) قبل	157	34.	(el estado) الدولة	115			

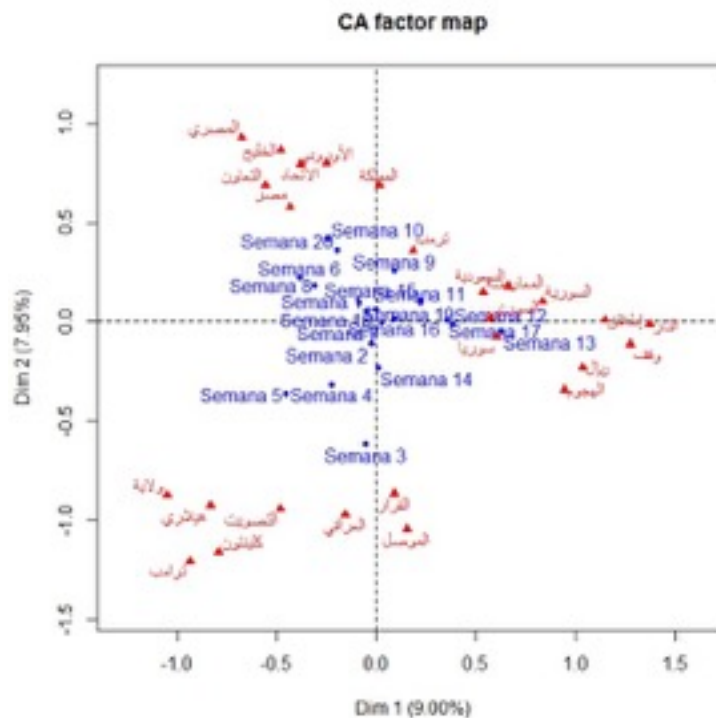
Fuente: Elaboración propia

Entre estas palabras expuestas anteriormente destacamos, por similitud temática, (*el presidente*) “الرئيس”, (*Trump*) “ترمب”, (*americana*) “الأميركية”, (*fuerzas*) “قوات”, (*oposición*) “المعارضة”, (*Siria*) “سوريا”, (*Clinton*) “كلينتون”, (*Mousul*) “الموصل”, (*Alepo*) “حلب” y (*el ejército*) “الجيش”. Por ello, a partir de estas palabras destacamos dos temas principales:

- La situación política de Estados Unidos, debido a la elección del nuevo presidente del gobierno.
- La guerra de Siria y otros conflictos bélicos y enfrentamientos ocurridos en países de Asia y África.

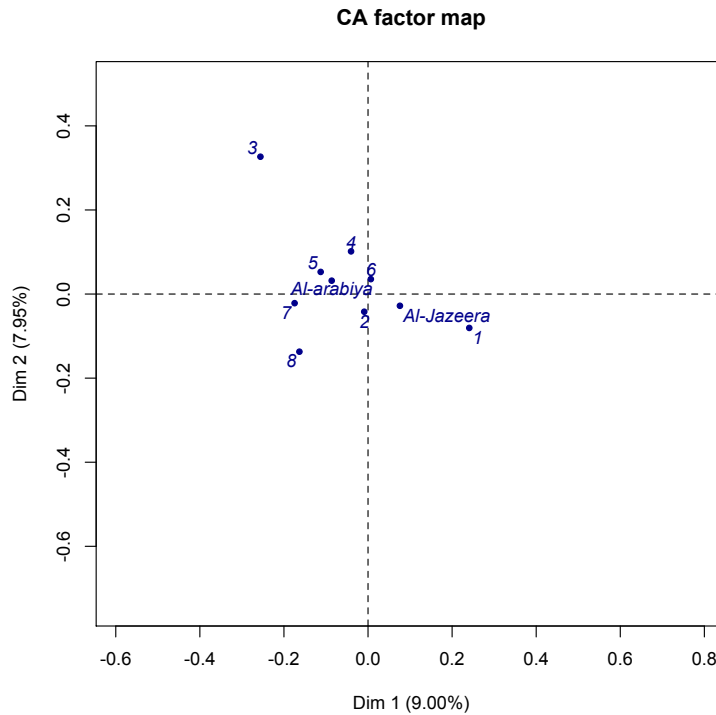
El análisis de correspondencias realizado con el programa estadístico nos proporcionan los siguientes resultados:

Gráfico 7.4.1.: Mapa de las palabras sobre las dimensiones 1 y 2 árabes.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.4.2.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 árabes.



Fuente: Elaboración propia a partir de Xplortext.

La dimensión 1 es definida por las semanas 13 (31,15) y 5 (-15,88) ya que son las que poseen las coordenadas más elevadas en términos absolutos.

En la parte positiva se encuentran palabras como:

- | | |
|--------------------------------|--------------------------------------|
| (3,56) (fuego) “النار” , | (2,79) (detener) “وقف” |
| (2,33) (lanzamiento) “إطلاق” , | (1,95) (conflicto sirio) “السورية” , |
| (1,31) (Siria) “سوريا” , | (1,25) (la oposición) “المعارضة” , |
| (1,01) (ataque) “الهجوم” , | (0,93) (Arabia Saudita) “السعودية” , |

Mientras que en la parte negativa se posicionan las siguientes palabras:

- | | |
|---------------------------------|---------------------------------|
| (-2,55) (Trump) “ترامب” , | (-2,05) (Clinton) “كلينتون” , |
| (-1,19) (estado) “ولاية” , | (-0,73) (el jeque) “الشيخ” , |
| (-0,65) (el egipcio) “المصري” , | (-0,64) (americana) “الاقتصادي” |

En base a esto, y coincidiendo con el gráfico de las variables ilustrativas, se confirma que la dimensión 1 está definida por el vocabulario relacionado con la temática 1, es decir, guerra y conflictos bélicos. Además, a partir de este último gráfico mencionado, se comprueba que el periódico “Al Jazeera” está más próximo a este contenido que “Al Arabiya”.

En cuanto a la dimensión número dos, por su elevada coordenada de -36,25, es la semana 3 la que caracteriza a esta dimensión. El vocabulario que se sitúa sobre el eje Y es, por la parte positiva:

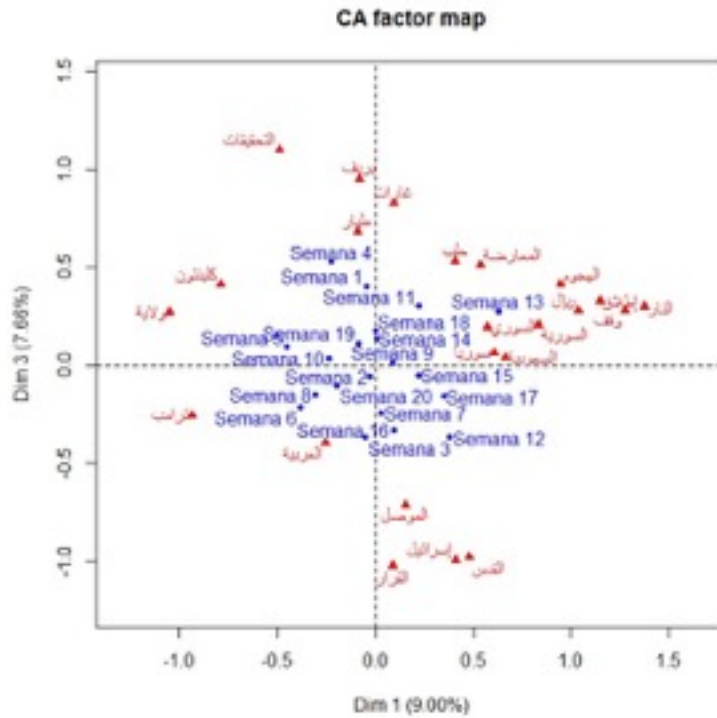
(1,39) (<i>el egipcio</i>) “المصري” ,	(1,08) (<i>Egipto</i>) “مصر” ,
(1,07) (<i>cooperación</i>) “التعاون” ,	(0,91) (<i>reino</i>) “المملكة” ,
(0,90) (<i>Unión Europea</i>) “الأوروبي” ,	(0,89) (<i>golfo</i>) “الخليج” ,
(0,88) (<i>unión</i>) “الاتحاد” ,	(0,72) (<i>promoción</i>) “تعزيز”

Por otro lado, en la negativa:

(-4,97) (<i>Clinton</i>) “كلينتون” ,	(-4,77) (<i>Trump</i>) “ترامب” ,
(-3,85) (<i>Mousul</i>) “الموصل” ,	(-2,06) (<i>resolución</i>) “القرار” ,
(-1,53) (<i>votación</i>) “التصويت” ,	(-0,93) (<i>estado</i>) “ولاية” ,
(-0,91) (<i>irakî</i>) “العراقي” ,	(-0,88) (<i>Hillary</i>) “هيلاري”

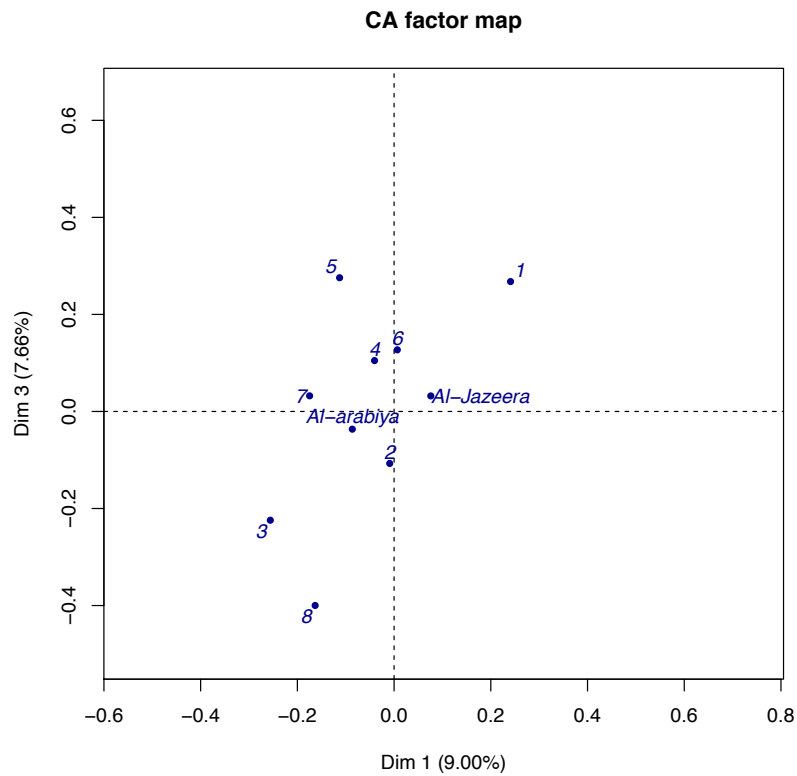
Nuevamente, para dar nombre al factor recurriremos a la ayuda de las variables suplementarias ya que, en algunos casos, el vocabulario puede pertenecer a varias temáticas al mismo tiempo. La modalidad 3 posee las coordenadas más altas en comparación con el resto de modalidades, que permanecen muy cercanas en su disposición sobre el eje, por ello, se determina que la dimensión 2 es definida por la política nacional, que como se pudo entrever en el comienzo del epígrafe, corresponde a la política nacional interna de uno de los países en concreto que conforma la región árabe o a los asuntos políticos que pongan en relación a dos o más países de la misma.

Gráfico 7.4.3.: Mapa de las palabras sobre las dimensiones 1 y 3 árabes.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.4.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 árabes.



Fuente: Elaboración propia a partir de Xplortext.

Por último en cuanto a la dimensión tres, la define la semana 4 con una coordenada de 22,43 junto a las palabras en la parte positiva:

(1,66) (<i>mil millones</i>) “مليار”,	(1,33) (<i>la oposición</i>) “المعارضة”,
(1,03) (<i>Alepo</i>) “حلب”,	(1,03) (<i>las afueras de la ciudad</i>) “بريف”,
(0,82) (<i>investigaciones</i>) “التحقيقات”,	(0,82) (<i>incursiones</i>) “غارات”,
(0,65) (<i>Clinton</i>) “كلينتون”,	(0,65) (<i>candidato</i>) “المرشحة” ;

y en la parte negativa:

(-2,99) (<i>resolución</i>) “القرار”,	(-2,88) (<i>Israel</i>) “إسرائيل” ,
(-2,06) (<i>Jerusalén</i>) “القدس”,	(-1,87) (<i>Mousul</i>) “الموصل”,
(-0,95) (<i>árabe</i>) “العربية”,	(-0,72) (<i>decisión</i>) “قرار”
(-0,67) (<i>irakí</i>) “العراقي”	(-0,64) (<i>potencia</i>) “السلطة”

Para definir la dimensión número tres, al igual que en la dos, se observarán las variables suplementarias para definirla con más precisión. Se observa que claramente es la temática 8 (cultura) y 3 (política nacional) las que definen a la dimensión tres, situadas en la parte negativa.

A continuación se realizó un análisis clúster de la misma manera que en los casos anteriores.

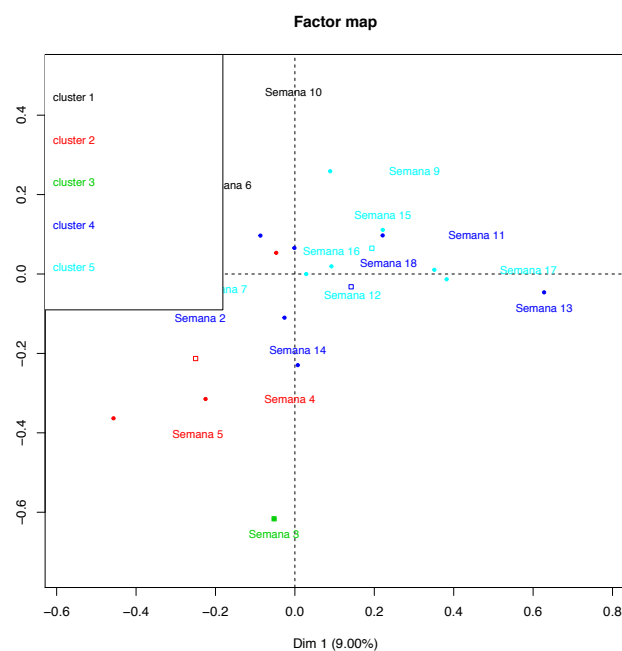
En este caso, los clusters que se centran en la política internacional son el 2 y el 5. Aunque ambos tengan esta como temática principal, a partir del vocabulario específico se encuentran los subtemas que abarca esta temática y, con ello, la razón por la que no conforman un mismo grupo perteneciendo a la misma temática.

En primer lugar, respecto al grupo dos, el vocabulario específico se centra en los nombres de los políticos estadounidenses junto a palabras como elecciones, votos o candidato. Teniendo en cuenta las semanas abarcadas por este cluster, se observa que entre ellas aparece la semana 6, periodo clave por acontecerse en ella las elecciones estadounidenses.

Y, en segundo lugar, el vocabulario específico del grupo cinco se centra en Israel y el conflicto israelí, aunque nuevamente aparece Trump entre las palabras más destacadas. Observando las semanas que componen el grupo, se comprueba que una gran parte de ellas pertenecen al periodo de enero de 2017, por tanto, en este grupo sobresale el conflicto palestino-israelí que durante el mes de enero fue reavivado por Donald Trump al iniciar conversaciones con Netanyahu, primer ministro de Israel. Las conversaciones fueron iniciadas para construir una relación de cooperación entre los países y para recibir el beneplácito del nuevo presidente estadounidense para construir 566 casas en la parte oriental de la ciudad de Jerusalén, ya que durante el mandato de Obama EEUU decidió intervenir para detener la nueva construcción al considerarse esta zona un territorio ocupado y no una parte del nuevo país. A todo lo anterior se le sumó la preocupación palestina por que Trump trasladase la embajada estadounidense a esta ciudad, lo que significaría una alineación completa por parte de EEUU con las posiciones de Israel.

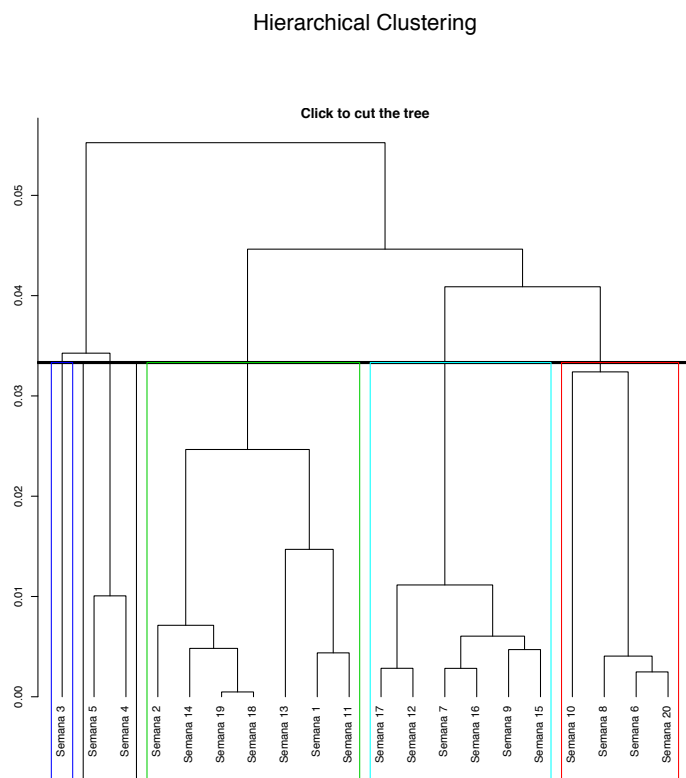
Los gráficos que se obtienen de la salida del programa y que se han utilizado para el análisis son los siguientes:

Gráfico 7.4.5.: Mapa de grupos de semanas similares de la región árabe.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.4.6.: Dendrograma región árabe.



Fuente: Elaboración propia a partir de Xplortext.

7.5 ANÁLISIS ESTADOS UNIDOS

Esta última base de datos se elaboró a partir de dos de los periódicos más prestigiosos, no solo de Estados Unidos, si no del mundo entero, por ello están posicionados en los dos primeros puestos dentro de la clasificación de número de lectores. Estos dos noticieros son “The New York Times” y “The Wall Street Journal”.

Tras el análisis textual del corpus se obtuvieron las 50 primeras palabras más frecuentes del mismo, que se presentan en la siguiente tabla:

Tabla 7.5.1.: Las 50 palabras más características del vocabulario global de la base de datos estadounidense.

No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.	No.	Palabras más frecuentes	Frec.
1.	mr (<i>señor</i>)	3523	18.	years (<i>años</i>)	394	35.	business (<i>negocio</i>)	302
2.	said (<i>dijo</i>)	2673	19.	administration (<i>administración</i>)	394	36.	including (<i>incluyendo</i>)	291
3.	trump	2205	20.	two (<i>dos</i>)	381	37.	white (<i>blanca</i>)	290
4.	new (<i>nuevo</i>)	1007	21.	million (<i>millón</i>)	371	38.	executive (<i>ejecutivo</i>)	282
5.	president (<i>presidente</i>)	741	22.	states (<i>estados</i>)	370	39.	now (<i>ahora</i>)	280
6.	will (<i>voluntad</i>)	652	23.	officials (<i>oficiales</i>)	363	40.	former (<i>antiguo</i>)	274
7.	year (<i>año</i>)	649	24.	campaign (<i>campana</i>)	360	41.	group (<i>grupo</i>)	273
8.	one (<i>uno</i>)	623	25.	united (<i>unidos</i>)	354	42.	financial (<i>financiero</i>)	272
9.	people (<i>gente</i>)	621	26.	many (<i>muchos</i>)	347	43.	clinton	271
10.	company (<i>compañía</i>)	608	27.	house (<i>casa</i>)	338	44.	just (<i>solo</i>)	271
11.	also (<i>también</i>)	550	28.	time (<i>tiempo</i>)	331	45.	since (<i>desde</i>)	268
12.	ms (<i>señora</i>)	460	29.	like (<i>como</i>)	323	46.	made (<i>hecho</i>)	268
13.	york	437	30.	companies (<i>compañías</i>)	321	47.	even (<i>incluso</i>)	265
14.	last (<i>último</i>)	421	31.	can (<i>poder</i>)	312	48.	election (<i>votación</i>)	255
15.	according (<i>según</i>)	416	32.	first (<i>primero</i>)	309	49.	country (<i>país</i>)	255
16.	state (<i>estado</i>)	413	33.	american (<i>americano</i>)	309	50.	federal (<i>federal</i>)	253
17.	government (<i>gobierno</i>)	395	34.	billion (<i>billón</i>)	307			

Fuente: Elaboración propia

Entre las palabras mencionadas en la tabla anterior destacamos las palabras “trump” con una frecuencia de 2205, una frecuencia mucho mayor respecto a la frecuencia que el apellido del nuevo presidente de los Estados Unidos posee en el resto de bases de datos; “president” (*presidente*) con una frecuencia de 741; “government” (*gobierno*) con 395; “administration” (*administración*) con 394; “million” (*millón*) con 371; “companies” (*empresas*) con 321; “american” (*americano*) con 309; business (*negocio*) con 302; “financial” (*financiero*) con 272; “clinton” con 271; y, por último, “election” (*votación*) con 255. A partir de las palabras que hemos subrayado de entre las 50 más características de toda la muestra, se pueden observar dos temas claros:

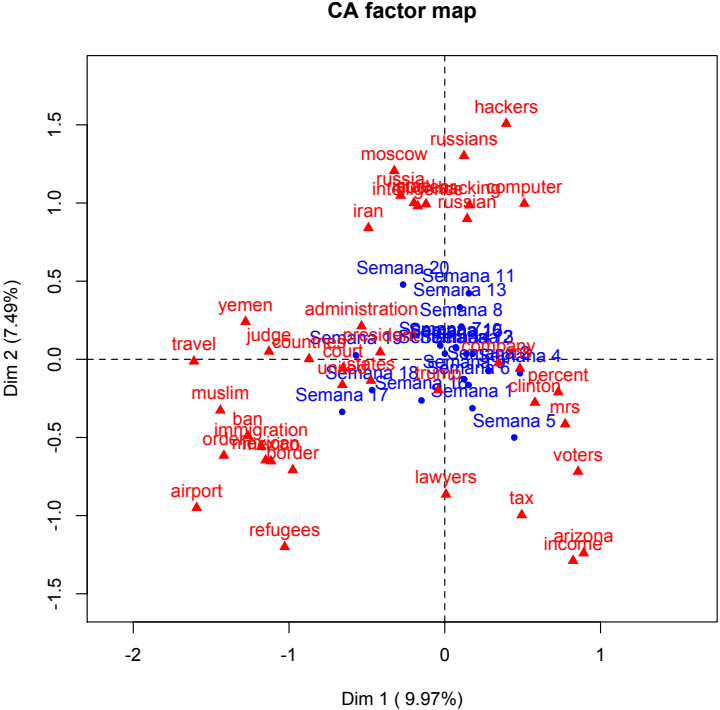
- La situación política nacional debido a las elecciones al nuevo presidente del gobierno y la consecuente victoria del candidato al partido republicano Donald Trump.
- Economía y asuntos económicos.

Los resultados del análisis de correspondencias fueron los siguientes:

Las semanas 17 y 19 son las que poseen mayor coordenada, por lo que son las que caracterizan al factor 1. Atendiendo al vocabulario que se dispone sobre los ejes, en la parte positiva del factor aparecen palabras como: “ms” (*señora*) (0,98), “clinton” (0,83), “percent” (*porcentaje*) (0,82), “voters” (*votantes*) (0,66), “company” (*empresa*) (0,66), “income” (*ingreso*) (0,50), “year” (*año*) (0,50), “tax” (*tasa*) (0,46) y “debt” (*deuda*) (0,39), entre otras. En la parte negativa aparecen “order” (*decreto*) (-4,08), “judge” (*juez*) (-1,47), “united” (*unidos*) (-1,41), “immigration” (*inmigración*) (-1,40), “muslim” (*musulman*) (-1,37), “president” (*presidente*) (-1,17), “ban” (*prohibición*) (-1,14) y “mexico” (-1,10), entre otras.

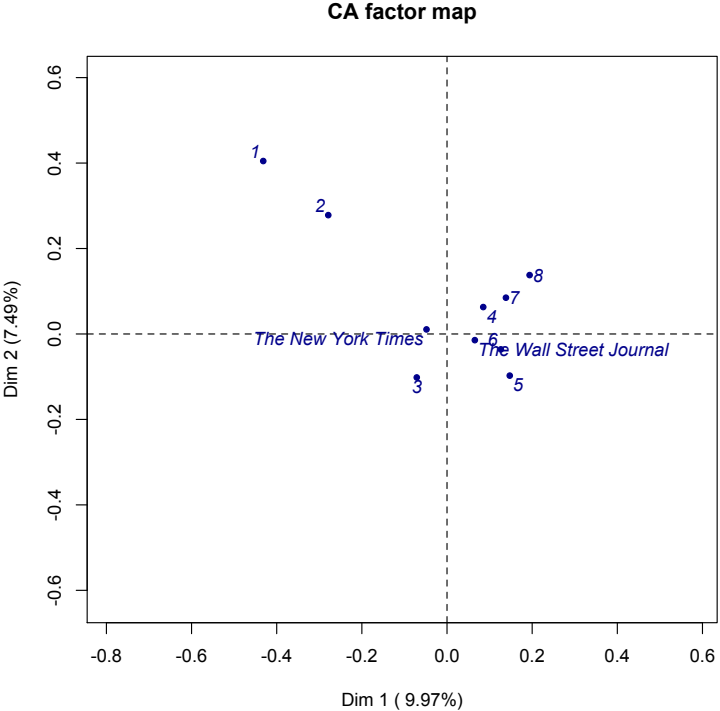
El vocabulario deja claro la caracterización de la dimensión en la parte positiva por la temática económica, mientras que en el lado negativo la política nacional es la que define al factor. Destaca en la dimensión 1 la cercanía del periódico “The Wall Street Journal” a la temática económica (lado positivo del eje) y “The New York Times” a la temática política.

Gráfico 7.5.1.: Mapa de las palabras sobre las dimensiones 1 y 2 de EEUU.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.5.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 2 de EEUU.

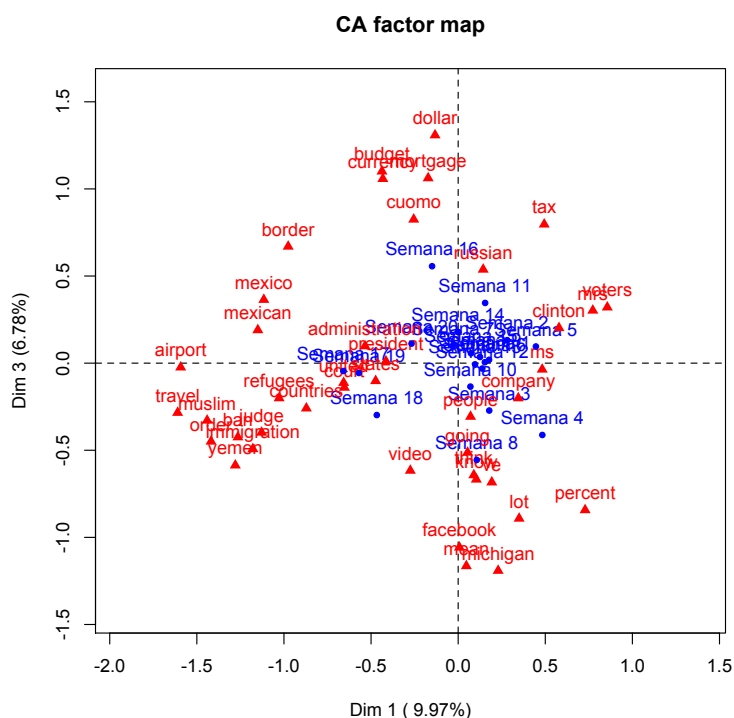


Fuente: Elaboración propia a partir de Xplortext.

Respecto a la dimensión dos, son las semanas 20 (21,10) y 11 (15,8) las que la determinan por la parte positiva, mientras que por la negativa es la semana 5 (-18,7). En la parte positiva se sitúa el vocabulario “russia” (*Rusia*) (2,50), “russian” (*ruso*) (2,08), “intelligence” (*inteligencia*) (1,99), “hackers” (*piratas informáticos*) (1,25), “moscow” (*Moscú*) (0,94), “nuclear” (*nuclear*) (0,83), “iran” (0,77) e “israel” (0,72). Por el lado negativo se encuentran “tax” (*tasa*) (-2,49), “income” (*ingreso*) (-1,64), “trump” (-1,04), “order” (*decreto*) (-1,02), “refugees” (*refugiados*) (-0,97), “voters” (*votantes*) (-0,62), “lawyers” (*abogados*) (-0,57), “arizona” (-0,56) y “mexico” (-0,50), entre otras.

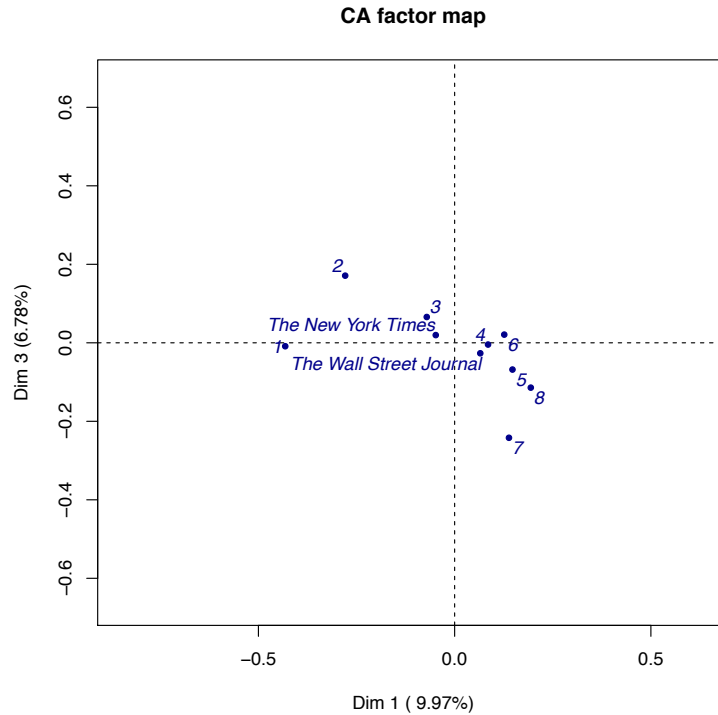
Analizando esto con el gráfico de las variables suplementarias se determina que la dimensión dos está caracterizada por las temáticas 1 y 2, guerra y política internacional, en la parte positiva, mientras que por la parte negativa se define a partir de las temáticas 3 y 5, es decir, política y economía nacional.

Gráfico 7.5.3.: Mapa de las palabras sobre las dimensiones 1 y 3 de EEUU.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.5.4.: Mapa de las variables ilustrativas sobre las dimensiones 1 y 3 de EEUU.



Fuente: Elaboración propia a partir de Xplortext.

En cuanto a la dimensión tres, la caracterizan por el lado positivo la semana 16 (24,6) y por el lado negativo la semana 8 (-30,91). Por el lado positivo aparecen las palabras “russian” (*ruso*) (0,82), “cuomo” (0,72), “currency” (*moneda*) (0,72), “mortgage” (*hipoteca*) (0,62) y “budget” (*presupuesto*) (0,57), entre otras.

Por la parte negativa aparecen palabras como: “facebook” (-2,29), “percent” (*porcentaje*) (-1,62), “people” (*gente*) (-0,79), “order” (*decreto*) (-0,60) y “michigan” (-0,55).

A partir del gráfico de las variables ilustrativas junto al vocabulario anterior se comprueba que es la temática 7 (ciencia y tecnología) , situada en la parte negativa, la que caracteriza esta dimensión.

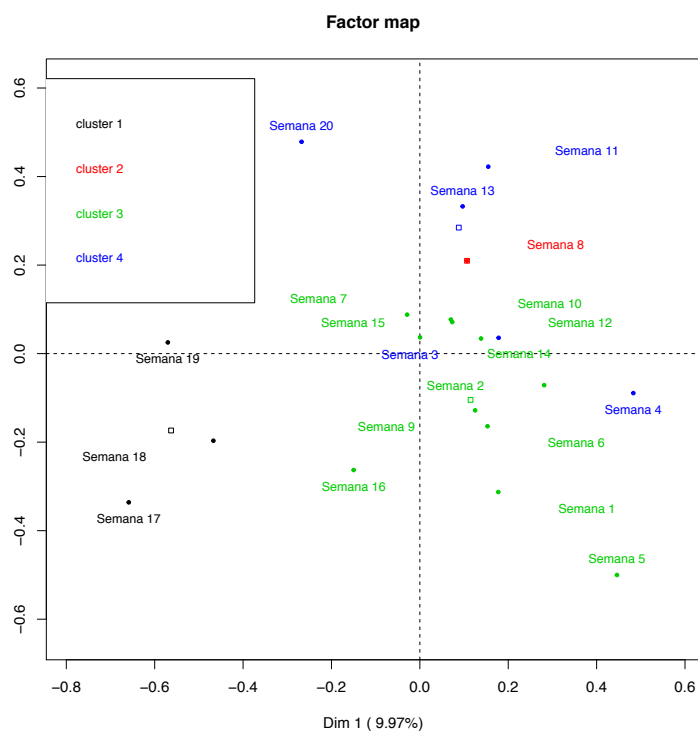
En cuanto al análisis clúster, de nuevo existen dos grupos cuyo tema principal gira alrededor de Donald Trump y las elecciones estadounidenses, lo cual en este caso pertenece a la política nacional.

El clúster número 1 está formado por las semanas 17, 18 y 19. El vocabulario específico relacionado con este periodo de tiempo es puramente político, y esto es debido a que este es el periodo en el que Donald Trump se mudó a la Casa Blanca y, con ello, asumió el poder del país.

Por último, el clúster número 3 agrupa un gran número de semanas, las cuales basan su temática en la economía y la política nacional, según el análisis de su vocabulario específico. Además, un dato más a señalar es que, según la disposición sobre los ejes, la mayoría de las semanas que son agrupadas por este clúster coinciden en la parte positiva de la dimensión 1 junto a “The Wall Street Journal” y la temática 5, el cual es el tema principal de este periódico. Por otro lado, las semanas restantes del grupo se posicionan cercanas a la temática 3, política nacional, y al periódico “The New York Times” en su parte negativa.

A continuación se muestran los gráficos a partir de los cuales se realizaron los análisis anteriores:

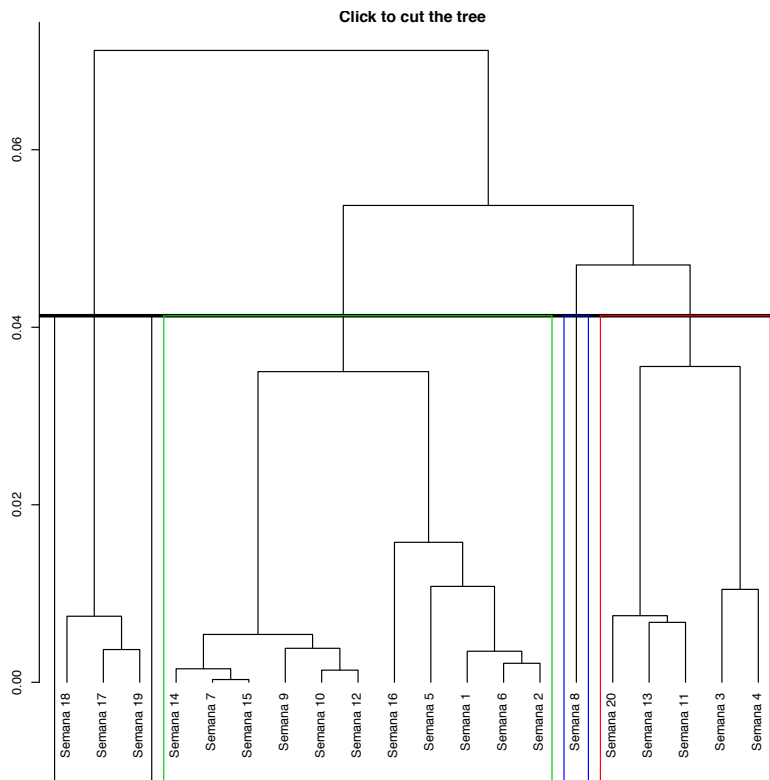
Gráfico 7.5.5.: Mapa de grupos de semanas similares de Estados Unidos.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 7.5.6.: Dendrograma Estados Unidos.

Hierarchical Clustering



Fuente: Elaboración propia a partir de Xplortext.

8. ANÁLISIS CONJUNTO

Una vez realizados los análisis a las bases de datos individuales con el objetivo de analizar la temática del contenido principal semanal de cada país/región, se realizará un análisis conjunto de todas las bases de datos a partir del estudio de las frecuencias de las palabras que forman el corpus para ver las similitudes y diferencias entre el vocabulario de los distintos países.

En primer lugar, se comenzó el análisis con los objetos “*TextData*” que se crearon en cada base de datos individual para construir una tabla de frecuencias múltiple. Esta tabla de frecuencias se forma con la función “*DocTerm*” de cada objeto creando una tabla léxica no agregada con tantas filas como documentos posee cada muestra, y tantas columnas como palabras seleccionadas (Bécue-Bertaut, M. et al., 2017). La función para la creación de la tabla de contingencia es la siguiente:

```
DocTerm_mul <- cbind(as.matrix(TDES$DocTerm), as.matrix(TDFR$DocTerm),  
as.matrix(TDUK$DocTerm), as.matrix(TDAR$DocTerm), as.matrix(TDEU$DocTerm))
```

Una vez creado el objeto con la tabla de contingencia se procede a la realización del análisis factorial múltiple. El objetivo de un análisis factorial es reducir un numeroso conjunto de variables cuantitativas en el menor número de dimensiones o factores comunes posibles sin perder información. Es similar al análisis de correspondencias, basando su diferencia en la procedencia de las variables, ya que el análisis de correspondencias trabaja con variables nominales y el análisis factorial con numéricas (Escofier, 1990).

Para la realización del análisis se utilizó la siguiente sintaxis:

```
res.mfact <- MFA(DocTerm_mul, group = c(ncol(TDES$DocTerm), ncol(TDFR$DocTerm),  
ncol(TDUK$DocTerm), ncol(TDAR$DocTerm), ncol(TDEU$DocTerm)), type = c("f", "f", "f",  
"f", "f"), name.group = c("ES", "FR", "UK", "AR", "EU"), graph = FALSE)
```

A continuación de crear el objeto que contiene el análisis factorial, sacamos las contribuciones de los distintos países a cada dimensión, así como los valores propios¹⁰ de cada dimensión y su contribución total a la inercia o varianza. Las instrucciones utilizadas, así como los resultados de las mismas son los siguientes:

```
round(res.mfact$group$contrib, 3)
round(res.mfact$eig,3)
```

Tabla 8.1.: Valores propios por dimensión y contribuciones por país.

	DIM 1	DIM 2	DIM 3	DIM 4	DIM 5
INERCIA TOTAL	3,96 (7,5%)	3,40 (6,4%)	3,26 (6,2%)	3,13 (5,9%)	3,10 (5,9%)
ESPAÑA	21.127	21.953	19.985	16.888	17.651
FRANCIA	21.458	20.707	19.848	20.297	20.281
R. UNIDO	16.792	21.244	19.165	20.831	18.118
R. ÁRABE	20.043	19.730	27.069	27.193	22.895
EEUU	20.581	16.366	13.933	14.790	21.055

Fuente: Elaboración propia

En la tabla anterior se exponen, en primer lugar, los valores propios de cada dimensión así como el total de varianza explicada por cada uno. Hay que señalar que en un AFM de tipo textual los valores propios no explican un porcentaje tan alto de la variante como en un AFM no textual.

El factor 1 es el que más valor propio presenta y, por tanto, mayor porcentaje de la inercia total explica, un 7,5%.

Los grupos que más contribuyen a la explicación de las dimensiones son Francia (Dim 1), España (Dim 2), y la región árabe (Dim 3, Dim 4 y Dim 5).

El porcentaje de inercia explicada acumulada con los cinco factores utilizados es de 32,12%.

¹⁰ El valor propio nos indica la aportación de un factor a la explicación de la varianza total.

Tras analizar la inercia y los factores, se procedió por un lado a la elaboración de los gráficos de los individuos activos sobre los ejes de coordenadas (semanas) y, por otro lado, a la creación del objeto llamado “*sel*” que contiene las contribuciones absolutas de las palabras de los diferentes grupos. La contribución absoluta, por definición, “de una variable *j* para un factor *alfa* indica el porcentaje del factor que es explicado por la variable” (Álvarez Esteban, 2015 a)

Una vez obtenido este objeto, procedemos a la elaboración de su gráfico para la facilidad de análisis. La sintaxis utilizada fue la siguiente:

```
plot.MFA(res.mfact, choix = "ind", invisible = c("col"), col.hab = rep("black", 6))
lines(res.mfact$ind$coord[1:3,1],res.mfact$ind$coord[1:3,2], lwd=2 )
lines(res.mfact$ind$coord[4:6,1],res.mfact$ind$coord[4:6,2], lwd=2 )

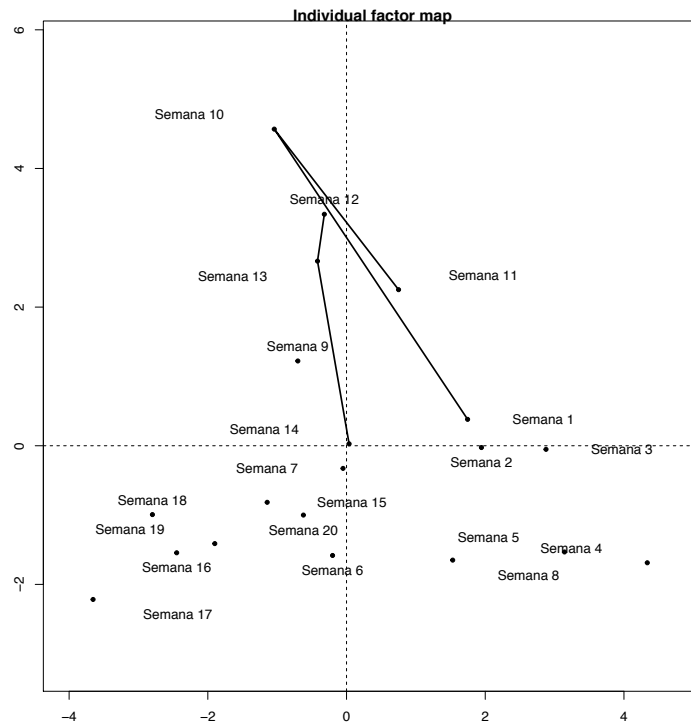
sel <- which(res.mfact$freq$contrib > 2*mean(res.mfact$freq$contrib))

plot.MFA(res.mfact, choix = "freq", invisible = c("row"), select = "cos2 0.5", unselect = 1,
autoLab = "yes", cex=0.6)
```

En el gráfico 8.1. se observan las semanas dispuestas sobre los ejes. Como se puede comprobar, existen tres bloques claros entre las distintas modalidades, uno posicionado en la parte positiva de la segunda dimensión y los otros dos dispuestos en su parte negativa, en las distintas partes del eje de la dimensión 1.

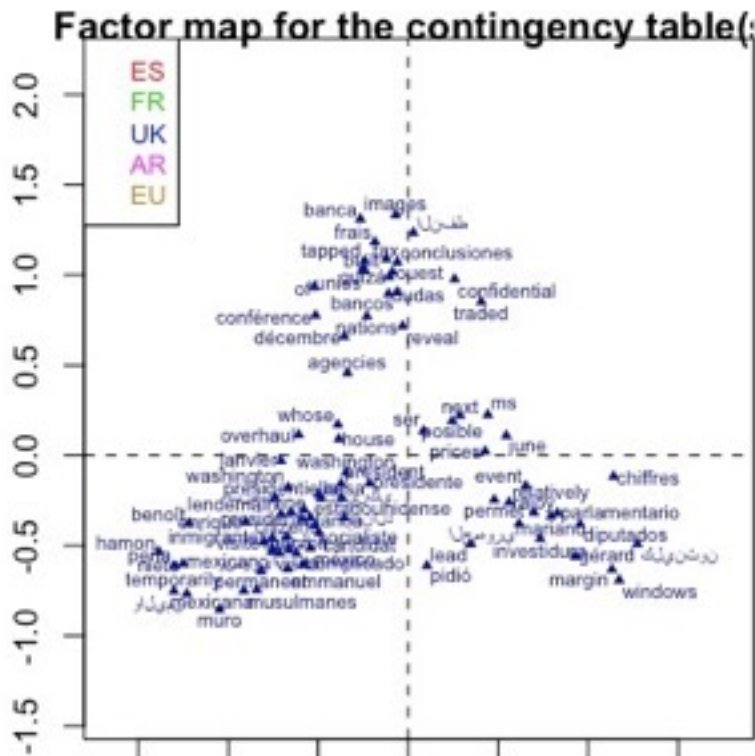
Se observa que no existe un frente temporal definido, si no que las semanas están dispuestas en el eje sin ningún tipo de agrupación cronológica. Esto es debido a que las temáticas de los contenidos de la prensa no son sucesivas, es decir, durante las distintas semanas las publicaciones son de temas muy variados y también puede haber temáticas de las que se publique en una semana y no se vuelva a hablar sobre ella hasta un mes después.

Gráfico 8.1.: Mapa de las semanas sobre las dimensiones 1 y 2.



Fuente: Elaboración propia a partir de Xplortext.

Gráfico 8.2.: Mapa de las palabras sobre las dimensiones 1 y 2.



Fuente: Elaboración propia a partir de Xplortext.

Con el análisis de las gráficas de manera conjunta se observa una temática muy clara en el grupo de semanas situadas en la parte negativa del factor 1, la política, ya que se sitúan en este lugar palabras como “presidente”, “estadounidense”, “washington” “benoît”, “hamon”, “nieto”, etc. Estas palabras están relacionadas con las semanas 16, 17, 18 y 19, periodo en el que Donald Trump tomó las riendas del gobierno de Estados Unidos y, por otro lado, se produjo la victoria de Benoît Hamon ante Manuel Valls en la segunda vuelta de las elecciones primarias del partido socialista francés.

Además, es interesante señalar la posición sobre los ejes de la semana 6, ya que como se vio en el análisis individual de las bases de datos, es la semana en la que se produjeron las elecciones a presidente de los Estados Unidos y este fue un tema muy marcado por la prensa internacional en este periodo. Sin embargo, en el análisis conjunto aparece situada muy cerca del valor 0 de la dimensión 2, prácticamente entre medias de los dos grupos formados en la parte negativa de esta dimensión y no donde se esperaría encontrarla en un principio, que sería en el grupo de la izquierda, por su temática. Esto es debido a que, aunque posee una fuerte relación con la temática de la política estadounidense, no fue el único tema que marco a esta semana, y tiene también cierta semejanza con las temáticas del grupo situado a la derecha.

Tras elaborar e interpretar los gráficos anteriores se creó un objeto llamado “c1” que recogía las coordenadas de las frecuencias de las palabras utilizadas para realizar el análisis factorial. A continuación se definieron, de acuerdo a los objetivos del estudio, siete palabras relacionadas con Trump y las elecciones estadounidenses y se tradujeron en los cuatro idiomas con los que se está trabajando, estas palabras fueron: “trump”, “eeuu”, “elecciones”, “voto”, “presidente”, “política” e “internacional”. Para identificar cada palabra con su país/región de procedencia, se procede a etiquetarlas con un prefijo, en el caso de España, por ejemplo, sería “ES_”. La sintaxis utilizada para todo lo expuesto y la gráfica con las palabras definidas se presenta a continuación. La gran longitud de esta sintaxis en comparación a las anteriores se debe a la complejidad del análisis, ya que se están juntando cinco estructuras diferentes.

```
c1 <- res.mfact$freq$coord
rownames(c1)[1:1763]<- paste0("ES_", rownames(c1)[1:1763])
rownames(c1)[1765:3248]<- paste0("FR_", rownames(c1)[1765:3248])
rownames(c1)[3249:4871]<- paste0("UK_", rownames(c1)[3249:4871])
rownames(c1)[4872:5971]<- paste0("AR_", rownames(c1)[4872:5971])
rownames(c1)[5974:8404]<- paste0("EU_", rownames(c1)[5974:8404])
```

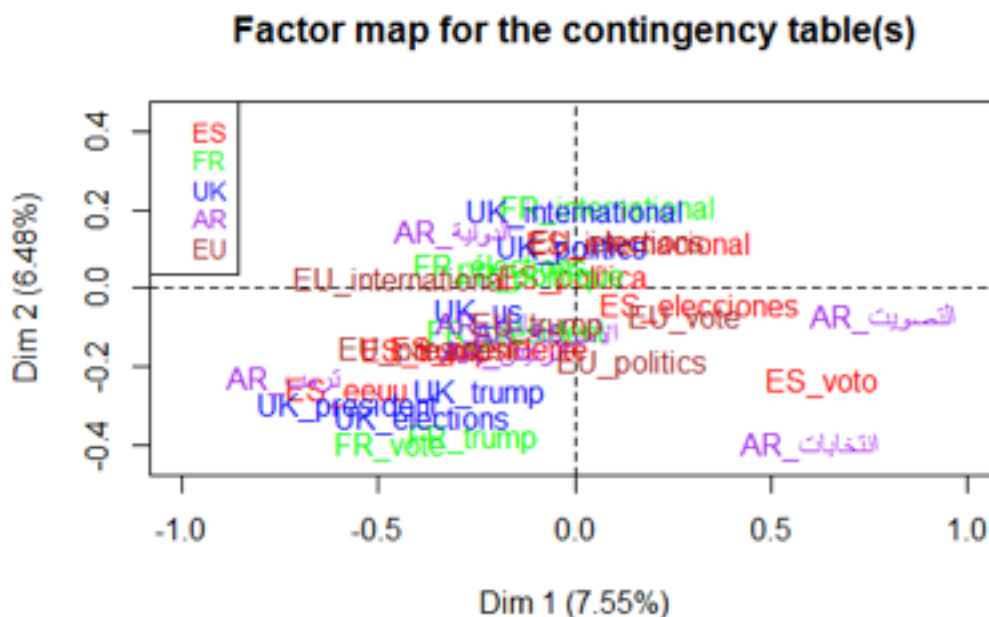
```

selfr <- which(rownames(c1) %in%c("FR_trump", "FR_eu", "FR_élections", "FR_vote",
"FR_président", "FR_politique", "FR_international"))
selar1 <- c("AR_الرئيس")
selar2 <- c("AR_ترمب")
selar3 <- c("AR_انتخابات")
selar4 <- c("AR_الأميركية")
selar5 <- c("AR_سياسة")
selar6 <- c("AR_تصويت")
selar <- which (rownames(c1) %in%c(selar1, selar2, selar3, selar4, selar5, selar6))
selUK <- which (rownames(c1) %in%c("UK_trump", "UK_us", "UK_elections", "UK_vote",
"UK_president", "UK_politics", "UK_international"))
seles <- which (rownames(c1) %in%c("ES_trump", "ES_eeuu", "ES_elecciones", "ES_voto",
"ES_presidente", "ES_política", "ES_internacional" ))
seleu <- which (rownames(c1) %in%c("EU_trump", "EU_usa", "EU_elections", "EU_vote",
"EU_president", "EU_politics", "EU_international"))

plot.MFA(res.mfact, choix = "freq", invisible = c("row", "col"), select = sel, unselect = 1,
cex=0.8, xlim = c(-0.8, 0.8), col.hab = c("red", "green", "blue", "purple", "brown"))
text(c1[seles,1], c1[seles,2], labels = rownames(c1)[seles], col = "red")
text(c1[selfr,1], c1[selfr,2], labels = rownames(c1)[selfr], col = "green")
text(c1[selUK,1], c1[selUK,2], labels = rownames(c1)[selUK], col = "blue")
text(c1[selar,1], c1[selar,2], labels = rownames(c1)[selar], col = "purple")
text(c1[seleu,1], c1[seleu,2], labels = rownames(c1)[seleu], col = "brown")

```

Gráfico 8.3.: Mapa de las palabras específicas seleccionadas sobre las elecciones estadounidenses y Donald Trump en las dimensiones 1 y 2



Fuente: Elaboración propia a partir de Xplortext.

Al igual que en los análisis anteriores, la proximidad entre dos palabras es interpretada en términos de semejanza entre perfiles (Bécue, 2004; Lebart, Salem y Berry, 1998).

En el gráfico 8.3. se observa que palabras como “internacional” están situadas de manera muy próxima entre todos los países/regiones, aunque Estados Unidos se sitúa más lejos del grupo que forma el resto. Lo mismo ocurre con la palabra “presidente”, situada en el mismo lugar excepto para Reino Unido, que se encuentra más alejada. Esto quiere decir que cuando aparecen las palabras juntas, los países las interpretan de manera similar, sin embargo, EEUU y UK respectivamente, aparecen alejadas, por lo que estos dos países, en las respectivas palabras, las perciben de manera distinta.

Una situación parecida ocurre con las palabras política y voto, en el que las diferentes regiones las interpretan de la misma manera, excepto Estados Unidos y la región árabe en el primer caso y Francia en el segundo caso, que aparecen alejados del grupo formado por los que si poseen similitud. De esta manera, cuando los países como España, Francia y Reino Unido hablan de política, la relacionan con la palabra “internacional”, ya que se sitúan próximamente sobre los ejes, mientras que Estados Unidos la relaciona con “trump”, “presidente” o “voto”. Por otro lado, la palabra “voto” francesa se aleja tanto de la palabra “internacional” del resto de países como de las demás palabras francesas, por lo que interpreta esta palabra de forma diferente al resto de países.

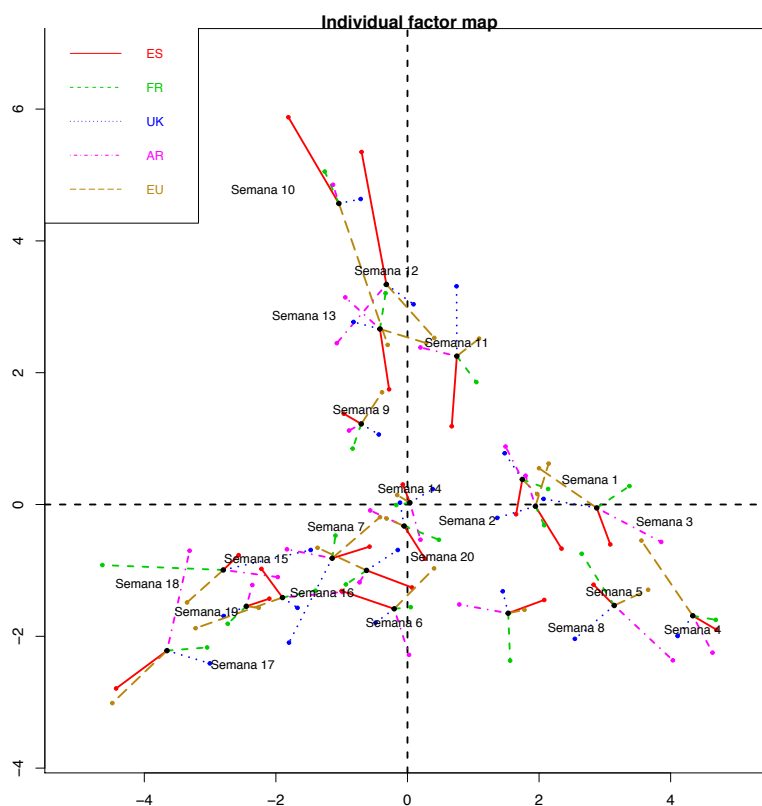
Respecto a la palabra “elecciones”, aparece próxima entre los países de Francia, Estados Unidos y España, aunque esta última algo más alejada, pero aparece muy lejos del resto en los casos de Reino Unido, situada junto a sus también propias palabras “trump” y “president” (*presidente*), y la región árabe situada solitaria en el gráfico.

Por último, la palabra “trump”, clave para conseguir los objetivos del trabajo, aparece para todos los casos en la parte negativa de la dimensión 1 y de la dimensión 2, aunque no están muy próximas entre sí. Las más alejadas son Francia, que la sitúa junto a su también propia palabra “vote” (*voto*); y la región árabe, que no la posiciona junto a ninguna otra palabra de su mismo idioma, por lo que no percibe de manera similar esta palabra ni respecto al resto de países, ni respecto al resto de palabras árabes (que sí

aparecen juntas) como “الرئيس” (*presidente*), “الأميركي” (*americana*) y “سياسة” (*política*), lo que quiere decir que interpreta la palabra “trump” de manera diferente.

Por último, con el objetivo de realizar un análisis completo de todas las bases conjuntamente, se crearon tres gráficos más que se presentarán e interpretarán a continuación.

Gráfico 8.4.: Representación de la distancia entre países/regiones por semanas.



Fuente: Elaboración propia a partir de Xplortext.

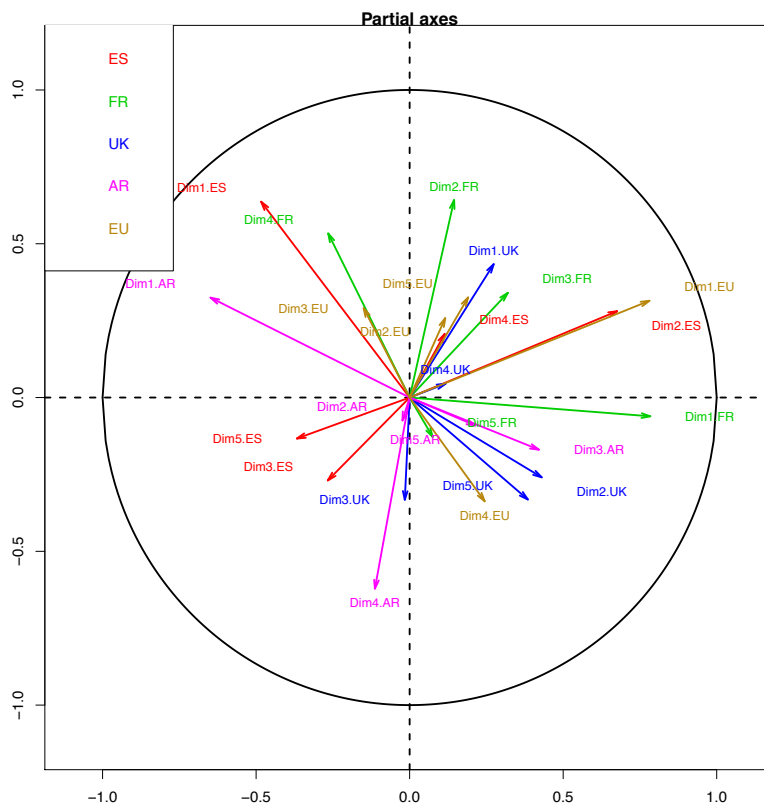
En este gráfico 8.4. se presentan las distintas semanas que componen el periodo en el que se realizó el estudio y se utiliza para comprobar las estructuras provocadas en ellas por las cinco nubes de palabras (cinco bases de datos). De esta manera, se podrán señalar las convergencias y divergencias entre países (Bécue, Pagès y Pardo, 2004) por semana.

En primer lugar, se observa que las semanas de la parte positiva del eje Y poseen mayor dispersión entre los distintos países que las semanas de la parte negativa.

Aunque son muchas las semanas cuyo contenido es interesante, se centrará en este caso el estudio en la semana 6 y la semana 16 ya que se consideran importantes dentro de la temática que se está analizando, Donald Trump y las elecciones estadounidenses, por las razones que hemos explicado a lo largo de todo el estudio. Atendiendo a las líneas que salen de cada punto, se observa que en el caso de Estados Unidos y España la línea es más larga, es decir, se separa más del centro, lo que quiere decir que el vocabulario de estos países de esta semana es muy diferente respecto al vocabulario característico del análisis conjunto del mismo periodo de tiempo. En la semana 6, la región árabe también se localiza en la misma situación junto a los otros dos países anteriores.

El siguiente gráfico que se analiza es el que representa las proyecciones de las distintas dimensiones parciales, es decir, las dimensiones de cada AC de los distintos países/regiones sobre los dos primeros factores del análisis factorial múltiple (AFM), para poder estudiar las relaciones entre sí (Escofier y Pagès, 1990).

Gráfico 8.5.: Representación de las dimensiones parciales sobre los factores del AFM.



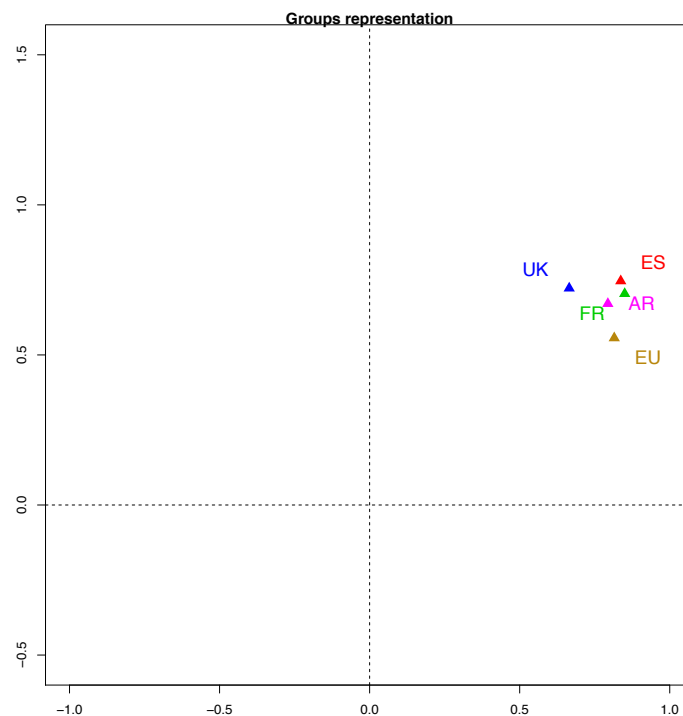
Fuente: Elaboración propia a partir de Xplortext.

Atendiendo al gráfico y la parte negativa de ambos ejes, que es donde se situaban las palabras relacionadas con Trump y las elecciones estadounidenses como se vio en el gráfico 8.2. las dimensiones que se relacionan con este vocabulario son la 3 y 5 españolas; la 2, 4 y 5 árabes; y, por último la dimensión 3 de UK, aunque esta última se sitúa muy cerca del valor 0 del eje Y, por lo que no está verdaderamente caracterizada por este vocabulario. Si volvemos a los análisis individuales y aún no siendo explicadas dimensiones como la 4 y la 5 en ninguno por cuestiones de relevancia, comprobamos *a posteriori* que, efectivamente, todas estas dimensiones se caracterizan por el vocabulario de la política internacional y por ello, tienen relación entre sí.

Por último se presenta el gráfico de grupos. Nuevamente, atendiendo al criterio de semejanza por proximidad, se observa similitud entre España, Francia, Estados Unidos y países árabes.

En cuanto a Reino Unido, que se muestra alejado del resto en el eje X, la diferencia reside en su destacado contenido sobre política nacional debido al candente tema del “*Brexit*” que ocupó un enorme número de titulares de la prensa británica.

Gráfico 8.6.: Representación de los países/regiones.



Fuente: Elaboración propia a partir de Xplortext.

9. CONCLUSIONES

Tras haber realizado el estudio del mercado periodístico y de la evolución del análisis textual, así como los análisis estadísticos, tanto individuales por país/región como conjuntamente, y teniendo en cuenta los objetivos expuestos al principio del trabajo se ha llegado a las siguientes conclusiones:

- La llegada de los periódicos digitales ha provocado un descenso en las ventas de ejemplares a papel y, con ello, una reducción considerable de la inversión publicitaria en el formato tradicional en favor del digital. Además, esto también ha producido un incremento de los lectores de este medio de comunicación, tanto en número como en variedad.
- Los avances informáticos han permitido un enorme progreso en el análisis de datos textuales, al poder trabajar con gran cantidad de datos simultáneamente y obtener los resultados en muy poco tiempo.
- Respecto a los análisis individuales se obtienen las siguientes conclusiones:
 - ➔ España: El vocabulario más característico del corpus indica la presencia de dos temas principales: la política nacional y la política internacional. En cuanto al análisis clúster los resultados indican que son cuatro las semanas cuyo tema principal es la política internacional: La semana 10 por la polémica sobre la ayuda rusa a la victoria de Donald Trump y las semanas 6, 17 y 18, semanas clave en Estados Unidos por tener lugar las elecciones presidenciales (semana 6) y el comienzo del gobierno de Donald Trump (semanas 17-18).
 - ➔ Francia: El vocabulario del corpus denota los dos mismos temas principales que España. Atendiendo al análisis clúster, son tres las semanas cuyo tema principal ha sido la política internacional, la 6, ya comentada anteriormente, la 8, y la 20, marcadas por declaraciones polémicas y visitas a otros países con resultado polémico.
 - ➔ Reino Unido: Los temas principales del corpus según el vocabulario característico son los mismos que en los casos anteriores. En el caso de

Reino Unido, son dos clústers los que se centran en la política internacional. El primero, agrupa las semanas 15, 16 y 17, semanas que rodean la instalación de Donald Trump en la Casa Blanca y asume las riendas del poder. El segundo está formado por las semanas 6 y 13, la semana 6 ya se ha comentado previamente y, en cuanto a la 13, fue una semana marcada por el enfrentamiento ruso-estadounidense por el ataque informático de hackers rusos a Estados Unidos.

- ➔ Región árabe: Según el vocabulario característico destaca la temática de guerra y conflictos bélicos junto a la política internacional. Teniendo en cuenta el análisis clúster existen dos grupos que corresponden a la política internacional, aunque solo uno posee el vocabulario correspondiente a las elecciones estadounidenses, el grupo formado por las semanas 6, 8, 10 y 20. Los motivos por los que estas semanas son importantes en este tema ya han sido expuestos en los casos anteriores.
- ➔ Estados Unidos: El vocabulario característico indica la relevancia de la política nacional y la economía nacional. Respecto a los clústers hay un grupo que por su vocabulario destaca en el tema a tratar, y es el que está compuesto por las semanas 17, 18 y 19. Las razones de este agrupamiento son claras, durante la semana 16 tiene lugar la toma del poder por parte del ganador de las elecciones, y es durante las semanas siguientes a esta fecha, cuando comienza la puesta en marcha de su programa político y comienzan las revueltas y manifestaciones en contra de ellas.
- No existe una evolución de temáticas relacionadas con el paso del tiempo, es decir, las temáticas no se van sucediendo a medida que el tiempo avanza, sino que coexisten a lo largo de los meses aunque haya temas más relevantes que otros dependiendo de la semana y de los acontecimientos que ocurran en ella.
- En las semanas en las que ocurre algo relacionado con Donald Trump o la política estadounidense, en toda la prensa internacional se marca este como tema principal, es decir, todas las semanas que en los análisis individuales han sido marcadas por este tema, aparecen situadas junto al vocabulario

relacionado con él en los diferentes idiomas.

- Eligiendo palabras concretas sobre el tema a estudiar se observa que es la región árabe la más diferente a la hora de utilizar el vocabulario, posicionando las palabras, en muchos casos, en los extremos de las dimensiones, lo que quiere decir que cuando utiliza, por ejemplo, la palabra “Trump” no utiliza la palabra “elecciones” o “voto”, como si hacen el resto de países.
- Por último, se quiso comprobar la similitud o disimilitud entre los distintos países/regiones, y se observó que, aunque todos se posicionan muy cerca entre sí por ser un periodo muy marcado por acontecimientos y polémica, Reino Unido es el que posee más diferencia. El motivo que probablemente provoque esto es la actual situación de este país con respecto a la Unión Europea, por la aprobación del *Brexit*, ya comentado a lo largo del trabajo, lo que hizo que la prensa británica dedicase gran parte del contenido a ello.

REFERENCIAS

Aaker, D. A., y Day, G. S. (1989). *Investigación de mercados* (2 ed.). Mexico: McGraw-Hill.

Álvarez-Esteban, R. (2015 a). *Apuntes de la asignatura Estadística avanzada I*. Grado Marketing e Investigación de Mercados, Facultad de Ciencias Económicas y Empresariales, Universidad de León. (Inédito)

Álvarez-Esteban, R. (2015 b). *Apuntes de la asignatura Estadística avanzada II*. Grado Marketing e Investigación de Mercados, Facultad de Ciencias Económicas y Empresariales, Universidad de León. (Inédito)

Bardin, L. (1997). *L'analyse de contenu*. Paris: Presses Universitaires de France.

Barranco Saiz, F. J. (2003). *Marketing político*. Madrid: Ediciones Pirámide.

Bécue Bertaut, M., Impr. Launay), A., y Guerrero, J. (1991 a). *Análisis de datos textuales: métodos estadísticos y algoritmos*. Paris: CISIA.

Bécue, M., Pagès, J., y Pardo, C.-E. (2004 b). Analysis of multilingual free responses. En *Journées internationales d'Analyse statistique des Données Textuelles* (pp. 119-127). Recuperado a partir de http://lexicometrica.univ-paris3.fr/jadt/jadt2004/pdf/JADT_010.pdf

Bécue-Bertaut, M., Álvarez-Esteban, R., y SánchezEspigares, J. A. (2017 c). Package 'Xplortext'

Careaga-Moya, J.-A., Medina-Urrea, A., y Torres-Moreno, J.-M. (2012). A New Cross-Lingua Automatic Summerization Approach Based on Textual Energy. En *Journées internationales d'Analyse statistique des Données Textuelles* (pp. 247-255). Recuperado a partir de <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Careaga-Moya>, Josue-Antonio et al. - A New Cross-Lingua Automatic Summerization Approach.pdf

Centro de Investigaciones Sociológicas. (2014). *Barómetro de diciembre de 2014*. Recuperado a partir de http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/3040_3059/3047/es3047mar.pdf

El Mundo. (2017). [Página web oficial]. Recuperado 10 de mayo de 2017, a partir de <http://www.elmundo.es/>

El País. (2017). [Página web oficial]. Recuperado 12 de mayo de 2017, a partir de <https://elpais.com/>

Escofier, B. y Pagès, J. (1990). *Análisis factoriales simples y múltiples: objetivos, métodos e interpretación*. Bilbao: Universidad del País Vasco.

Hồ Đình, O., y Valette, M. (2014). Analyse différentielle des discours de prévention du VIH: textes institutionnels et textes informels en français et en vietnamien. En *Journées internationales d'Analyse statistique des Données Textuelles* (Vol. 277-287). Recuperado a partir de <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/23-JADT2014.pdf>

Hornik, K. (2017). R FAQ. Versión 2017-09-05. Recuperado a partir de <https://cran.r-project.org/doc/FAQ/R-FAQ.html>

Icheboudene, Z., y Sjöblom, M. K. (2012). Exploration textométrique dans le paysage plurilingue de la presse algérienne francophone. En *Journées internationales d'Analyse statistique des Données Textuelles* (pp. 501-513). Recuperado a partir de <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Icheboudene>, Zina et al. - Exploration textométrique.pdf

Iglesias, F. (2001). *Marketing periodístico*. Barcelona: Ariel comunicación.

Janal, D. S., y De la Campa Pérez-Sevilla, M. A. G. (2000). *Marketing en Internet: cómo lograr que la gente visite, compre y regrese a su sitio Web*. España [etc.]: Pearson Educación de México.

Le Figaro. (2017). [Página web oficial]. Recuperado 15 de mayo de 2017, a partir de <http://www.lefigaro.fr/>

Le Monde. (2017). [Página web oficial]. Recuperado 17 de mayo de 2017, a partir de <http://www.lemonde.fr/>

Lebart, L., Salem, A., y Berry, L. (1998). *Exploring textual data*. Dordrecht: Kluwer Academic.

Luque, T. (1996). *Marketing Politico un analisis de intercambio politico*. Barcelona: Ariel.

Montero Rodríguez, F. (2005). *Marketing de periódicos y revistas*. Madrid: Esic.

Piqué, J. y Andreu-Besó, J.V. (1997). *Lingüística aplicada en su contexto académico*. Valencia: NAU Llibres.

Statista. (2017). Principales periódicos por lectores diarios España 2016. Recuperado a partir de <https://es.statista.com/estadisticas/476795/periodicos-diarios-mas-leidos-en-espana/>

Statista. (2017). Periódicos: distribución de lectores por edad España 2016. Recuperado a partir de <https://es.statista.com/estadisticas/476755/distribucion-de-lectores-de-periodicos-en-espana-por-edad/>

The Guardian. (2017). [Página web oficial]. Recuperado 22 de mayo de 2017, a partir de <https://www.theguardian.com/international>

The New York Times. (2017). [Página web oficial]. Recuperado 29 de mayo de 2017, a partir de <https://www.nytimes.com>

The Sun. (2017). [Página web oficial]. Recuperado 20 de mayo de 2017, a partir de <https://www.thesun.co.uk/>

The Wall Street Journal. (2017). [Página web oficial]. Recuperado 31 de mayo de 2017, a partir de <https://www.wsj.com/europe>

Zarzalejos, Á. G. (2017, febrero 19). Medios de comunicación: La prensa de papel, en el chasis: la publicidad también caerá este año víctima de internet. *El Confidencial*. Recuperado a partir de https://www.elconfidencial.com/comunicacion/2017-02-19/evolucion-sector-prensa-espana-papel-digital-publicidad-marcas-anunciantes_1333127/

(2017) الجزيرة.نت. [Página web oficial]. Recuperado 25 de mayo de 2017, a partir de <http://www.aljazeera.net/portal>

(2017) العربية.نت | الصفحة الرئيسية. [Página web oficial]. Recuperado 27 de mayo de 2017, a partir de <https://www.alarabiya.net/>