



---

## ESTUDIOS / RESEARCH STUDIES

---

### **Análisis documental de noticias de prensa en sistemas de información factual**

Antonio García Gutiérrez\*

\*Universidad de Sevilla

Correo-e: [algarcia@us.es](mailto:algarcia@us.es)

Recibido: 06-06-2013; 2ª versión: 19-09-2013; Aceptado: 20-09-2013.

**Cómo citar este artículo/Citation:** García Gutierrez, A. (2014). Análisis documental de noticias de prensa en sistemas de información factual. *Revista Española de Documentación Científica*, 37(2):e046. doi: <http://dx.doi.org/10.3989/redc.2014.2.1094>

**Resumen:** Los archivos digitales de noticias proporcionan a sus usuarios una gran cantidad de información en forma de texto completo o referencias hemerográficas que habitualmente hacen necesaria una gran inversión de tiempo adicional dedicado a la búsqueda de datos factuales: sujetos, consecuencias, fechas, lugares, causas... En esta investigación se propone un procedimiento pragmatista de análisis y recuperación que se centra en las empíricamente reconocibles "noticias de prensa" y en los modos de formulación pronominal de los usuarios finales y con claras extrapolaciones a otros géneros periodísticos informativos, como entrevistas o reportajes, que han de ser estudiados separada y específicamente. El estudio parte de la base de la construcción de macroproposiciones sobre un corpus de noticias, siguiendo reglas prácticas y sencillas de reproducir, para obtener un modo de lectura estandarizado y automático en los archivos de prensa y hemerotecas.

**Palabras clave:** Documentación periodística; hemerotecas digitales; noticias de prensa; información factual; análisis documental; búsqueda pronominal.

#### **Document Analysis of News Items for Factual Information Systems**

**Abstract:** Digital news libraries provide users with great amounts of information both in full text and references, that usually require a great amount of additional time for searching for factual data: subjects, consequences, dates, places, causes, backgrounds...The current research proposes a pragmatic procedure of analysis and retrieval focusing on press news items and based on pronominal request modes formulated by end-users. This has a clear potential for extrapolation to other journalistic genres, such as interviews or news reports, which would have to be separately and specifically addressed. This study begins with the construction of macro-propositions based on a corpus of news items and following rules that are both practical and easy to reproduce. The goal is to obtain a means for accessing press libraries and archives that is both standardized and automatic.

**Keywords:** Journalistic documentation; digital news libraries; news; factual data; document analysis; pronominal search.

**Copyright:** © 2014 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0.

## 1. INTRODUCCIÓN

Los archivos digitales de prensa han supuesto una gran innovación en la recuperación de información escrita en los medios de comunicación, tanto por la disponibilidad y aumento de resultados en los procesos de búsqueda como por la potenciación de la propia memoria mediática, ya sea para el uso de la empresa informativa que la genera o bien para terceros. Estudios recientes sobre las hemerotecas digitales de prensa de los principales medios españoles indican el crecimiento y expansión de este sector (Rubio Lacoba y Blanco García, 2010; Guallar y otros, 2012). Son relevantes, desde otros puntos de vista que también desvelan el auge de la investigación en documentación periodística, trabajos como los desarrollados sobre medios regionales y locales por Fernández Gil (2010) o Bonet y Fernández Quijada (2006).

Sin embargo, se observa que la deseada automatización de los fondos escritos por los célebres softwares de gestión documental (sgd), cuya presencia y diversidad han proliferado de manera notoria en los medios de comunicación en los últimos veinte años, ha venido acompañada por un nuevo tipo de problema, a saber, la obtención masiva de respuestas con altos índices de ruido y silencio, en cualquier caso a menor escala que los millones de ocurrencias ofrecidas desde los motores de búsqueda de *Google* o *Yahoo!* pero que, a pesar de ello, pueden implicar un colapso de la memoria por saturación de acuerdo con el teórico de la memoria social Huyssen (2000).

Los sistemas de documentación periodística, apoyados en potentes y costosas aplicaciones, albergan y gestionan buena parte de la memoria reciente y sirven a sus usuarios cantidades ingentes de datos mezclados en millares de textos e imágenes o sugeridos en referencias y metadatos que requieren un laborioso trabajo de criba posterior para el que los propios periodistas no disponen de tiempo. La cantidad de información suministrada por los buscadores electrónicos eclipsa en muchas ocasiones la información relevante para el usuario. Esta falta de precisión responde a dos causas derivadas del incremento exponencial de textos digitalmente archivados: por un lado, el análisis documental del contenido de las noticias (la entrada de información en el sistema) es, paradójicamente, cada vez más genérico cuando no inexistente (en ocasiones consiste en un mero escaneado), al mismo tiempo que aumenta la precisión de los propios buscadores digitales en la salida de información, sin lograr paliar el problema. En parte, tal abandono ha sido inducido por la confianza excesiva depositada en las herramientas de recuperación habilitadas por las mismas aplicaciones si bien la reducción de plantillas en los archivos de prensa sería una razón previa a la aludida.

Por otro lado, y más específicamente, la saturación y distorsión en las respuestas de los sistemas de documentación periodística puede atribuirse

a la ausencia de adjudicación de funciones o roles a los conceptos extraídos de la noticias como palabras-clave y *links*, y naturalmente, a la falta de desarrollo de esta posibilidad por parte de las aplicaciones de bases de datos periodísticas, lo que ha generado históricamente la identificación o confusión conceptual. Veamos algunos ejemplos: el rol del Ayuntamiento de Zaragoza como firmante de un contrato (sujeto colectivo) poco tiene que ver con el rol del Ayuntamiento de Zaragoza como lugar en el que se ha celebrado una conferencia o se han manifestado unos vecinos (locativo). En la ecuación de búsqueda "España vende aceite a Portugal" la respuesta habitual de un sistema suele intercambiar y confundir al sujeto de la venta, en este caso España, con su receptor, Portugal, pues al estar exentos de funciones los conceptos involucrados (y ser ambos países productores de aceite) no es posible identificar quién vende a quién. España y Portugal, que en el ejemplo aparecen como sujetos colectivos (los gobiernos, las empresas, los consumidores) también serán identificados como lugares, añadiendo aún más ruido a la respuesta. Finalmente, y como pequeña muestra inicial del problema, no sería posible preguntar al sistema quién, cuándo, cómo vende o qué consecuencias tiene la venta de aceite a Portugal o viceversa y ni siquiera que confirme si se vende aceite a Portugal (verificación).

Y un último apunte introductorio: las propias aplicaciones informáticas (desde el intérprete *Ariety* y las redes sociales a las tecnologías *Ajax* y la *web 3.0*) favorecen la presencia de los usuarios en los sistemas digitales y, por lo tanto, también la potenciación de la figura del usuario modelo (Eco, 1993) en documentación periodística como parte de la construcción de la memoria digital mundial. Si hay una disciplina pragmatista (o pragmática, pero en el sentido clásico de James (2007)) por excelencia, bien pudiera ser la documentación, pues nada en ella tiene sentido ni significado si el conocimiento que promueve no tiene consecuencias. De ahí que coloquemos la dimensión pragmática de la documentación periodística como eje privilegiado, esto es, como posición epistemológica dominante de la que partirán las observaciones y propuestas de este artículo y de todo nuestro programa de investigación.

Esta investigación, centrada en el análisis documental del género informativo por excelencia - la noticia- y fundamentada en varios años de observación empírica, abre un programa de trabajo sobre otros géneros y estilos periodísticos persiguiendo los mismos objetivos a partir de las variables que impone cada tipología textual particular en la recuperación de datos factuales. De modo paralelo y teniendo en cuenta tales variables discursivas en cada caso, ha sido ya publicado un trabajo sobre análisis de entrevistas y textos declarativos (2013) y, ya abordado, el reportaje y los principales géneros de opinión (editorial, análisis, columna y crítica) con el fin de obtener productos

integrados, precisos y fiables en los sistemas de documentación periodística organizados desde una perspectiva pragmática pronominal. Todos estos procedimientos y resultados pertenecen a una dimensión textual y discursiva de la documentación periodística que habrá de ser completada con aplicaciones informáticas *ad hoc* en una fase ulterior.

## 2. OBJETIVOS ESPECÍFICOS

La presente investigación orienta su aparato teórico conceptual y su metodología a la consecución de dos objetivos específicos:

1. proponer un procedimiento de análisis de noticias que, partiendo de una simulación sobre cómo pregunta un usuario-modelo, regule y explicita el modo a través del cual los documentalistas extraen los ejes significativos de las noticias con el fin de ser convertidos en ángulos de búsqueda (abandonando los habituales criterios no parametrizables como la intuición, el "oficio" o la experiencia) y

2. propiciar una organización y búsqueda de la información basadas en roles conceptuales a partir de preguntas pronominales de modo que se reduzcan, o incluso anulen en un alto porcentaje, las distorsiones de resultados que reciben los usuarios, adecuándolos a la información que solicitan.

Los usuarios de sistemas de documentación periodística, sean periodistas, historiadores, investigadores sociales o un público más genérico, no siempre buscan información en los formatos en los que los mismos servicios suelen proporcionarlas (texto íntegro, referencias, resúmenes o metadatos). El procedimiento permitiría al sistema ofrecer los datos factuales en formatos mucho más personalizados y adecuados a demandas y perfiles específicos a desarrollar en estudios posteriores.

## 3. APROXIMACIÓN PRAGMATISTA E HIPÓTESIS DE TRABAJO

El uso determina el significado, solía advertir Wittgenstein (2004). En documentación periodística, los usuarios no solo necesitan textos completos o referencias hemerográficas si no también datos puntuales (fechas, lugares, sujetos, procesos...) o derivados de ellos (comparaciones y verificaciones: más que, menos que, sí, no o porcentajes). Para realizar este "giro pragmatista" es necesario, en primer lugar, identificar qué y cuántas preguntas de tipo pronominal de primer orden podría formular un usuario modelo y serían contestables por el sistema. En segundo lugar, hay que observar si los datos que solicita son manipulables y organizables de acuerdo a los parámetros de búsqueda establecidos en la fase anterior. Para ello, habrán de tenerse en cuenta: a) las tipologías textuales, las funciones conceptuales, los conectores y otros elementos implicados en las estructuras lógico-semánticas de las unidades que constituyen un archivo periodístico, b) correlacionar los modos de interrogación

del usuario-modelo con los criterios explícitos de lectura del documentalista de prensa y con las estrategias de la propia producción textual elegidas por los periodistas. En suma, para cada modo de interrogación debe generarse una categoría consistente (Andréu, 2000) que segmente el texto registrado, proporcionando una respuesta factual acorde a la pregunta formulada.

Antes de proceder a inventariar los modos de interrogación de un teórico usuario, es necesario restringir su formulación a un tipo de texto particular con el fin de asegurar unas bases estructurales que impidan la dispersión de preguntas posibles (repárese que nos interesa repertoriar "cómo" se solicita la información factual -una estructura categorial- y no "qué" información factual se solicita -una estructura temática-). Dado el conocimiento empírico generalizado sobre cómo se solicita la información factual, ha sido suficiente contar con la colaboración desinteresada de varios profesores, periodistas y alumnos de periodismo para acceder casi de forma inmediata a un umbral de repetición satisfactorio de doce preguntas elementales.

Debemos excluir aquí las demandas que buscan datos formales de la producción o difusión de la noticia, como autor, medio, fecha de publicación, cuya resolución corresponde a técnicas catalográficas y no propiamente al análisis estructural de contenido que nos ocupa, pues pertenecen a otro plano de descripción. Restringiendo, en consecuencia, las búsquedas a preguntas que persiguen respuestas factuales que teóricamente se prevén en un perfil de búsqueda sobre noticias escritas en español estándar, de extensión media y estructura canónica, obtenemos el siguiente inventario exhaustivo, una vez reducida y controlada la sinonimia y polisemia de formulaciones similares:

1. ¿Quién/Qué cosa ataca, gana, cae, se desborda, vende, destruye, se reúne, firma...ejecuta una acción determinada?
2. ¿A/Contra quién ataca, gana, vende...?
3. ¿Qué cosa es atacada, ganada, vendida...?
4. ¿Por qué ataca, gana, vende, cae...?
5. ¿Cómo ataca, gana, cae, vende...?
6. ¿A través de qué medio ataca, gana, vende ...?
7. ¿Para qué ataca, gana, vende...?
8. ¿Con qué consecuencias ataca, gana, vende...?
9. ¿Con quién se ejecuta?
10. ¿En qué estado se ejecuta?
11. ¿Cuándo ataca, gana, reúne, firma, cae...?
12. ¿Dónde ataca, gana, reúne, firma, cae...?

Parafraseando a Lasswell y tomando la teoría de casos universales como horizonte, podríamos hablar de las "12Ws" o preguntas de orden pron-

minal de primer nivel (efectuadas a la acción) que todo usuario plantea o debe poder plantear a un sistema de documentación de noticias esperando obtener una información concisa, precisa y adecuada a su petición. Las 12Ws, a falta del análisis y depuración que será realizado más adelante, abarcan una tipología de demandas de datos factuales (un sujeto concreto, una fecha, una consecuencia, una herramienta...) y no textos completos o párrafos de noticias o reportajes en los que todavía habría que buscar y desbrozar esos datos específicos.

Una vez determinados los objetivos y establecidos los límites y conjeturas sobre el objeto de investigación, procede enunciar la siguiente hipótesis de trabajo: en los sistemas de documentación periodística es posible elaborar un procedimiento de análisis y búsqueda de noticias periodísticas escritas basado en un enfoque pragmático que parte de los modos de interrogación pronominal de un usuario modelo, basados en la teoría de casos universales, de forma que el sistema pueda proporcionar datos de tipo factual (y derivados como comparaciones y verificaciones) adecuados a los términos y precisión exigidas en la demanda, sin perjuicio de otros productos convencionales de los sistemas de información (referencias y textos completos). Tal procedimiento coadyuva a dos mejoras inmediatas: a) precisión y adecuación de las respuestas a las preguntas formuladas, y b) homologación, explicitación y verificación del procedimiento de análisis seguido por el documentalista de modo que pueda ser revisado por terceros (lo que dota de mayor fiabilidad y científicidad a un proceso, hasta el momento, altamente intuitivo y precario).

A diferencia de otros estudios previos, emprendidos por mí mismo (1999), que han partido de la posición opuesta, esto es, desde la perspectiva de la producción textual, esta investigación -sin desecher los avances alcanzados en aquella línea y de los que se sirve como referencia crítica- prioriza qué necesidades de información tiene el usuario de la documentación periodística de prensa en términos factuales, cómo expresa esas necesidades en forma de preguntas pronominales y cómo el documentalista y el sistema deben operar con ellas.

Quedan varios asuntos de extrema relevancia que deben ser advertidos y, en su caso, resueltos a un cierto nivel, en función de los objetivos e hipótesis de trabajo:

- a. Construcción de registros empíricos sobre los cuales serán hechas las observaciones y propuestas que compondrán el corpus de experimentación como una de las aproximaciones de la investigación para obtener resultados parciales. El elemento fundamental de estos registros será la elaboración de macroposiciones -MPs- (Van Dijk, 1990) a partir de las noticias seleccionadas. Las MPs permiten formalizar, homologar y regular los procedimientos y estruc-

turas con el fin de facilitar inicialmente las observaciones.

- b. Elaboración y depuración de un inventario de conectores vinculados a modos de interrogación pronominal de usuario.
- c. Discusión sobre los roles de la acción como eje central del procedimiento buscado, si bien su inventario temático, no relevante a efectos de los objetivos de este trabajo, es una tarea urgente de la documentación periodística.
- d. Depuración de los modos de interrogación en bruto y asociación a categorías y reglas de análisis con las que el documentalista y, eventualmente, un software inteligente de análisis textual pudieran ulteriormente operar.
- e. Elaboración de MPs en laboratorio que recojan exhaustivamente respuestas a todos los modos posibles de interrogación en primera instancia. Esta aproximación está orientada a completar los resultados derivados de la vía empírica expuesta en el punto a).
- f. Señalar extrapolaciones y aplicaciones de las formulaciones parciales o totales alcanzadas con el fin de abrir líneas de investigación pragmática en otros géneros periodísticos (como la entrevista y los géneros de opinión, investigaciones ya finalizadas) y otras áreas de acción de la documentación periodística.

#### 4. APARATO TEÓRICO-CONCEPTUAL

Para demostrar las hipótesis y comprobar las conjeturas que las sostienen, deben conjugarse los intereses de búsqueda factual de los usuarios (las 12Ws) con el aparato teórico-conceptual que varias disciplinas han solido prestar históricamente a la documentación periodística. Éste es el caso de la lingüística teórica y del análisis del discurso, del análisis de contenido y de los estudios de usuario así como del pragmatismo (como filosofía del conocimiento) en el sentido del ya mencionado James (2007). En esa línea, conceptos propios de la semántica y de la morfosintaxis, fundamentalmente procedentes de la lingüística estructural de Pottier (1976), de la teoría de los casos universales enunciada por Fillmore (1968) y de las aportaciones de autores como Cintra (1987), Amaro (1991), Cunha (1990) o Lara (2011) en el ámbito de la lingüística aplicada a la organización del conocimiento, del seminal análisis automático de Pêcheux (1978), de la propuesta de análisis del discurso periodístico de Van Dijk (1990) o de los análisis de contenido de Krippendorff (1990) y Andréu (2000), entrarán en diálogo con el aparato conceptual tradicional de la documentación periodística (Galdón (2002), Fuentes (1997) o García Gutiérrez y Lucas Fernández (1987).

## 5. CORPUS DE EXPERIMENTACIÓN

La primera medida, a efectos de aplicar el método y demostrar los distintos componentes de la hipótesis de trabajo, es delimitar aleatoriamente un corpus suficiente de noticias periodísticas escritas y organizar éstas en unidades de información o registros. De acuerdo con Krippendorff, el método aleatorio garantiza, para muestras simples como la presente, que cualquier registro del universo total tenga igualdad de oportunidades de estar incluido o excluido del corpus, por lo que apela al sorteo para determinar qué unidades lo compondrán a través de cualquier procedimiento que asigne la misma probabilidad (Krippendorff, 1990). Se trataba de observar, desde la perspectiva de una búsqueda pronominal a partir de las 12Ws, si las noticias, con las acotaciones formales más abajo descritas, atendían a estructuras conceptuales, generales y estables a través de un cierto periodo de tiempo de modo que el procedimiento matriz no tuviese que ser sometido a modificaciones constantes. A pesar de comprobarse la existencia de estas macroestructuras profundas, hubo de acudirse a la construcción de observables (las macroproposiciones, *vid.* más abajo) dadas las dificultades para operar con el lenguaje y las retóricas de la superficie de los textos e incluso de los títulos.

Inicialmente, se estableció un corpus provisional, disperso y abierto en torno a cien noticias de prensa a fin de elaborar las primeras conjeturas y servir de base a un test provisional destinado a averiguar si las noticias contienen datos que satisfacen las demandas factuales establecidas a partir de las 12Ws y bajo qué categorías conceptuales y conectores podrían ser acogidos. Los primeros ensayos ya sugerían que a partir de ese primer centenar de noticias se alcanzaba un umbral de repetición notable en cuanto a conectores y ciertas categorías conceptuales pero todavía insuficiente en cuanto a otras así como respecto al número de acciones y temas generales y específicos tratados, una labor relacionada con la representación que habría de emprenderse mediante la formalización de verbos, sustantivos y adjetivos, pero que excede los límites de la presente investigación y no tiene implicaciones en el objetivo de detección de categorías conceptuales buscado.

El criterio que determinó la extensión definitiva del corpus, y el número total de registros que lo componía, provino de la metodología de validación adoptada, de modo que tanto el tamaño de las muestras derivadas del mismo, así como las propiedades de sus unidades constituyentes pudieran satisfacer los criterios de validación y, de ese modo, estar en disposición de verificar los resultados. La técnica de validación elegida fue la denominada "técnica de la mitad". La división en mitades se hace en un estadio temprano de la construcción del corpus. De acuerdo con Krippendorff (1990), la técnica de la mitad ofrece un muestreo satisfactorio. La muestra inicial es dividida en dos muestras

de igual tamaño. Si cada parte aporta el mismo nivel de confianza, la totalidad puede ser considerada como adecuada en sus dimensiones. De lo contrario, el analista debe incrementar el tamaño de la muestra hasta que las condiciones deseadas se cumplan. De este modo, se procede a operar con dos muestras, A y B, de forma que las inferencias elaboradas a partir de ensayos con registros de la muestra A pudieran dar lugar a unos resultados que, de acuerdo a lo estipulado por el procedimiento de validación, permitieran su verificación en una muestra ajena a la utilizada como base del primer ensayo.

Se procedió, en consecuencia, a generar dos muestras, A y B, mediante la ubicación en cada una de ellas de noticias extraídas por sorteo de varios medios impresos españoles, de tirada nacional y local (El País, El Mundo, ABC de Sevilla, entre otros) entre 1991 y 2011, amplio periodo al que se recurre con el fin de verificar que las estructuras textuales profundas de las noticias permanecían estables con independencia de sus tendencias ideológicas y libros de estilo (lo que no ocurre con la "prensa histórica" a la que habría de dedicar estudios específicos y cautelosas extrapolaciones, en su caso). La muestra A tuvo que ser ampliada hasta alcanzar los 205 registros, umbral en el que comenzaron a repetirse las estructuras de menor frecuencia y, por tanto, cada una de las 12 formulaciones presentaban ya correspondencia factual en una mitad del corpus. Así mismo, la primera muestra ya servía con suficiencia conectores y acciones (en bruto) como para realizar un inventario no temático sino relativo a la conectividad de ambos elementos léxicos. A continuación, se procedió a realizar la misma operación, con el fin de sostener los mismos extremos, enriqueciendo la muestra B hasta el mismo límite, comprobándose similares propiedades categoriales. Como consecuencia de estas operaciones, se obtuvo un corpus total de 410 registros para el que fueron construidas otras tantas MPs de procedencia empírica subordinadas a acciones dominantes y basadas en el cuadro de las 12Ws aún en bruto (la depuración, estructuración, organización y normalización final de las MPs se presenta en los ejemplos exhaustivos del epígrafe 6).

Los registros de observación, cuya estructura se describe más abajo, se restringieron a noticias de estructura canónica, de extensión media (con un mínimo de dos párrafos lo que excluía, por ejemplo, necrológicas y algunos "despachos" breves), escritas en español contemporáneo de España con el fin de evitar una casuística legítima, pero incontrolable a efectos de esta primera aproximación al problema, de otros usos periodísticos del español que deben ser objeto de investigación específica al igual que otras lenguas próximas derivadas del latín. Se observó, del mismo modo, que la inclusión de variables temáticas o ideológicas, elementos decisivos en una fase ulterior de representación en medios específicos y en la construcción

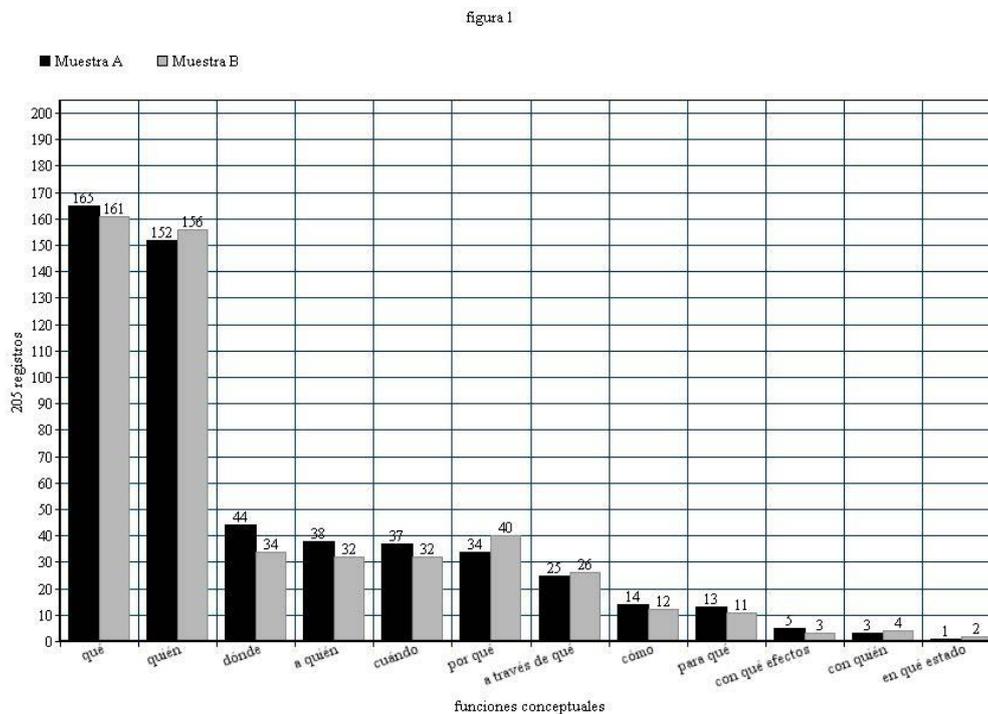
de lenguajes documentales, no era determinante en esta fase inicial del programa de investigación centrado en la detección de funciones conceptuales universales subyacentes abordadas por géneros periodísticos.

A nivel estructural, y ya desde las primeras noticias, se detectaron categorías que superaban y completaban las 5Ws clásicas, a la vez que sugerían la posibilidad de ser transformadas en preguntas pronominales de usuario para obtener, en primera instancia, datos factuales organizados predicativamente en torno a acciones dominantes. El umbral de repetición se obtuvo (y sostuvo) rápidamente en el caso de algunas categorías en tanto que otras precisaron la selección de muchos más registros manteniendo una baja ocurrencia en otros. Registros consultados al azar, más allá de esos límites, no aportaron ya nuevas estructuras conceptuales de primer orden (subordinadas a una acción dominante). Como observación colateral se desprendía que si obviamente ninguna noticia real cubría simultáneamente las 12Ws (pues se trataría, en ese caso utópico, de un relato total) la mayoría de ellas tampoco satisfacía simultáneamente las clásicas 5Ws y solían aportar de forma azarosa datos sobre las siete Ws restantes de forma desigual y dispersa. Pero, en definitiva, el corpus manejado sugería que incluso las formulaciones pronominales de menor frecuencia podrían verse satisfechas con algunas respuestas en los registros controlados y, por

tanto, en mucha mayor medida en una base de datos con millones de registros.

Las estructuras conceptuales -protocategorías- obtenidas respondían mayoritariamente a las preguntas ¿qué cosa? y ¿quién?, cayendo drásticamente las ocurrencias, como se refleja en la figura 1, a una cuarta parte de las anteriores en las preguntas ¿a quién?, ¿por qué?, ¿dónde?, ¿cuándo? e incluso, en menor medida, ¿a través de qué? En cuanto a las restantes, las formulaciones modales y finales ¿cómo o mediante qué? y ¿para qué?, su frecuencia se reduce aproximadamente a la mitad de las anteriores. Por último, ¿con qué consecuencias?, ¿con quién? y ¿en qué estado? ocupan el rango menor de ocurrencias. En la misma figura se aprecia la frecuencia de aparición de estructuras conceptuales de hecho, asociables a interrogaciones pronominales potenciales. Debe señalarse, no obstante, que las de mayor frecuencia no necesariamente portan más relevancia desde la perspectiva pragmática (aunque sí lo hicieran desde la perspectiva opuesta de la producción textual) pues, en nuestra óptica, la relevancia es determinada por el usuario. Ahora bien, aunque la frecuencia no sea una señal de relevancia, sí es un indicador de saturación, ruido y otros accidentes lógico-semánticos en el sistema que habrían de ser tenidos en cuenta y resueltos en otros estudios mediante reglas de representación y recursos informáticos.

Figura 1. Representatividad de las categorías en las muestras A y B



Por otra parte, un espacio muy rico en información lo constituyen los titulares de la noticia, justamente el resultado de una síntesis que realiza su productor con el fin de atraer la curiosidad del lector. Sin embargo, como se comprueba a primera vista, los títulos de las noticias no siempre proporcionan toda la información relevante que la misma noticia desarrolla e incluso suelen responder a criterios comerciales, políticos, estilísticos o de índole no netamente textual (esto es, no representan adecuadamente el contenido de la noticia que preceden ni sus estructuras conceptuales) y están adornados de una implacable retórica dentro de una tipología textual muy extensa de la que destacamos indicativamente los casos más recurrentes:

- oraciones atributivas del tipo "A es " (se transforman en "enunciados de acción").
- oraciones impersonales del tipo "se construye, se produce..."
- oraciones intransitivas como "X muere..."
- oraciones subordinadas introducidas por "que" del tipo "A hace que B..."
- oraciones transitivas interrumpidas "X hace"
- oraciones simples en cualquier orden hiperbático: "tanto monta, monta tanto".
- oraciones con omisión verbal: "X, en TV".
- oraciones en pasiva: "X es acusado por Y de..."
- oraciones coordinadas: "Se va X, se queda Y"
- yuxtaposiciones opositivas: "juicio verdadero, juicio falso"
- encabezamientos con gerundios y participios.
- construcciones heterogéneas como nombres de personajes seguidos de frases lapidarias, sintagmas nominales, frases extraídas de declaraciones, etc. Esta diversidad justifica la elaboración de MPs a partir de los títulos enriquecidos por elementos del texto de la noticia.

En el caso de títulos descriptivos, estos son ampliados (*expanded titles*) con las estructuras conceptuales detectadas en el texto tras su interrogatorio mediante las 12Ws y organizadas en torno a la acción dominante. Sin embargo, en caso de metáfora, metonimia, elipsis, eufemismo o cualquier otra figura retórica en el título, éste es sustituido completamente por una construcción descriptiva similar al caso anteriormente descrito. La ventaja de trabajar con MPs normalizadas es evidente aunque requiere un trabajo que, por parte del investigador o del documentalista de medios, se vislumbra como ineludible para no perder la precisión requerida en las búsquedas: la "desambiguación" y "desretorización" del titular de la noticia, además de su complementación con elementos descritos y estratégicamente ubicados en su contenido (este problema aumentaría enormemente en el caso de los textos de opinión).

Las MPs permiten una mejor observación de las tipologías textuales, de acciones y, especialmente, de conectores que albergan las noticias y, por lo tanto, facilitan la detección y normalización de las funciones conceptuales asociadas a preguntas de usuarios, de cuántas y cuáles son, lo que nos daría la clave para elaborar las categorías elementales de organización textual de la noticia, desde el punto de vista de su recuperación documental y, por tanto, también los protocolos de análisis que habrían de ser encomendados a los documentalistas (y posteriormente a los sistemas inteligentes) para organizar datos procedentes de textos periodísticos en función de demandas pronominales.

### 5.1 Constitución del registro

El registro es una unidad de información (Andréu, 2000) procedente de una transformación de cada una de las noticias que constituyen el corpus. Tal transformación es efectuada con objeto de homologar estructuralmente los datos observables, detectar fácilmente lagunas, solapamientos, erratas y omisiones con el fin de aplicar de manera más precisa los distintos instrumentos metodológico. Se compone de los siguientes campos y subcampos básicos:

- a) descripción externa: identificación de autor (apellidos, nombre), título (todos los elementos del titular por orden de importancia, medio de publicación (nombre del periódico o revista), fecha (ocho dígitos invertidos en el formato aaaammdd) y páginas (o url) en la que se encuentra la noticia.
- b) macroproposición: síntesis textual elaborada a partir de la formulación de las 12Ws a cada una de las noticias de las muestras que consiste en "reescribir" (y, en su caso, sustituir) y enriquecer el título en un mismo enunciado canónico del tipo: sujeto-acción-predicado con los datos que el propio texto original explícitamente proporcione. Veamos en primer lugar la acción, por ser el elemento articulador de toda la estructura:
  - Acción: verbo o nombre de proceso que presenta la acción dominante en el texto (no necesariamente en los titulares). En un texto puede haber muchas acciones secundarias y, muy plausiblemente, se encuentre explícita una acción que subsume a todas ellas si bien, en caso de hallarse en modo tácito, es necesario crearla para articular un predicado en la MP. Por ejemplo, en una noticia se menciona que un ministro español llega a México en un vuelo oficial, baja del avión, saluda a su homólogo que ha ido a recibirlo, pasa revista a las tropas, entra en la sala de prensa del aeropuerto, dirige unas palabras a los periodistas y a algunos compatriotas residentes, le recoge un coche oficial y toma rumbo a la embajada. Todas estas acciones particulares se subsumen en la acción dominante "viaje" o "visita" (oficial). En determinadas noticias ocurre que una parte del

texto no presenta subordinación temática alguna respecto a otra (por ejemplo, mientras se produce y describe la visita, ha entrado en erupción el volcán Popocateptl). La única razón de que /visita/ y /erupción/ sean abordados bajo la misma noticia, titular y periodista, se debe a una coincidencia espacial o temporal (ha ocurrido cerca del lugar o el mismo día de la visita). Esos casos serán tratados como dobles textos o "subtextos" y se procederá a elaborar dos macroproposiciones (una por subtexto), físicamente bien delimitadas en el registro mediante puntos y sangrado, una dedicada a la visita y otra a la erupción.

Por otra parte, y aunque debe tenderse a conservar la terminología original siempre que sea posible, se acudirá a vocablos alternativos que eviten los usos retóricos verbales (disminuir o mantener en lugar de "frenar" o "congelar los precios", etc.). Instalar desde el inicio esta cautela antirretórica, aunque no tiene repercusiones categoriales relevantes (salvo en lo que afecta a los propios roles verbales transitivos, intransitivos, impersonales, etc.), facilita la homologación del vocabulario y, por ende, las comparaciones estructurales aunque no es necesaria su resolución a efectos inmediatos de este trabajo. Finalmente, se opta, a efectos de formalización de las MPs, por emplear el presente de indicativo sobre hechos ocurridos en el pasado (el ejército ataca, el gobierno se reúne...) y, en el caso de futuro (los sindicatos se manifestarán..., un astronauta viajará a Marte...), se acude a ciertos verbos como anunciar o programar, en presente de indicativo, que implícitamente proyectan futuro (se anuncia que los sindicatos..., se programa un viaje a Marte...).

- Sujeto: el sujeto abarca, con independencia de las precisiones que habrán de ser hechas más adelante, a todo ser vivo, colectivo o cosa que ejecuta (por sí misma) la acción dominante, previamente seleccionada. Entran, dentro de este segmento, las personas, las instituciones, las bacterias y los ríos, volcanes, huracanes o cualquier otro tipo de fenómeno o instancia natural que cumpla la condición de ejecutar acciones por sí mismo (hablar, reunirse, invadir, desbordar, entrar en erupción, destruir...). Se obtiene trasladando la pregunta más básica que suelen hacer los usuarios: ¿Quién o qué cosa ejecuta la acción?

- Predicado: última parte de la macroproposición que se compone de elementos derivados de la ejecución de la acción seleccionada: en ella hay cabida para el resto de interrogantes del perfil que el usuario traslada al documentalista: a quién o qué cosa beneficia o perjudica, cómo se ejecuta y a través de qué medios, por qué motivos y con qué consecuencias... En la elaboración de esta parte de la MP es importante llevar a cabo dos operaciones: por un lado, encadenar los distintos conceptos, de acuerdo a la función que desempeñan, mediante conectores ( epígrafe 5.2) que comporten una ca-

pacidad de funcionalidad primaria. A pesar de la polisemia de los conectores habituales, el propio contexto de la noticia puede actuar como instancia mitigante de tal dispersión (por ejemplo, reservando "a través de" a la instrumentalidad: TV por cable = a través de cable, o reservando "por" a la causalidad: muerte por hambre, por enfermedad, por accidente...). Naturalmente, se llama la atención sobre estas cautelas a fin de ir normalizando y depurando un inventario de conectores que nos darán la clave de unas estructuras artificiales que puedan ser "entendidas" por las aplicaciones inteligentes. Por otra parte, la mera y exacta ubicación física de las funciones conceptuales -desprovista de conectores- dentro del predicado es también muy relevante, especialmente porque facilita la comparación entre funciones precedida de conectores y la detección de otras funciones que carecen de ellos (en ciertos casos, la mera posición sería indicativa de función, como es el caso de sujetos o complementos directos, por ejemplo). El orden predicativo convencionalmente establecido es el siguiente: tras la acción, precedida en su caso por sujetos u objetos ejecutantes, se sitúan los conceptos relativos a la pregunta /a quién/ y /a qué cosa afecta la acción/, seguido de /por qué/ y /para qué/, /cómo/ y /a través de qué/, /con qué resultados/, /dónde/ y /cuándo/, /con quién/ y /en qué estado/. Se presenta, a continuación, la configuración general de los registros:

Registro de análisis de noticias:

Signatura:

Descripción externa: Autor: título. "medio", fecha, página/url

Macroproposición: sujeto-verbo-predicado

Observaciones:

- conectores:
- acciones: (reservado para investigaciones posteriores)
- descriptores: (reservado para investigaciones posteriores)

Las operaciones llevadas a cabo, simultáneamente al enriquecimiento de los enunciados de los títulos son: eliminación de predicados atributivos, eliminación de datos secundarios o de segundo nivel de subordinación, eliminación de conceptos retóricos, eliminación de todos los tiempos verbales, y de la voz pasiva, salvo el presente de indicativo. Estas operaciones de eliminación llevan aparejadas operaciones reconstructivas y sustitutivas: generalización de verbos y conceptos secundarios, elaboración de enunciados descriptivos, transformación en presente de indicativo, unificación de acciones mediante la sustitución de formas verbales comunes e incorporación de lugares o fechas en la macroproposición para explicitar espacio y temporalidad.

## 5.2 Tipologías de conectores

Los conectores constituyen uno de los elementos más relevantes de asociación entre conceptos y de indicación de función conceptual en los textos. Para Pêcheux (1978), uno de los autores más influyente en esta materia, particularmente en relación a las lenguas derivadas del latín, los conectores son nexos carentes de significado propio y suficiente que tienen por objeto la articulación de las categorías léxicas provistas de significación dominante (básicamente sustantivos, verbos y adjetivos). Por así decirlo, sustituyen los viejos cometidos de las declinaciones en las lenguas no declinadas. Son conectores: las preposiciones, las conjunciones y todo tipo de partículas léxicas que conectan elementos semánticamente relevantes del enunciado.

En muchos casos, el conector adquiere una presencia tácita recogida en la posición del elemento en el enunciado, bajo una adjetivación o en otros tipo de variadas elipsis y, en otros, viene dotado de una perjudicial polisemia para nuestro objetivo investigador (por ejemplo, como se ha visto, los diferentes sentidos de /por/: causal /muerte por accidente/, instrumental /tv por cable/; /en/ tiene valor temporal /en 2012/, instrumental /en tv/ y locativo /en Madrid/, etc.). La dimensión más relevante del conector, a nuestros efectos, es la relación de asociación que puede ser construida con los modos de interrogación pronominal (12Ws), por un lado, y con categorías de análisis con las que el documentalista pueda operar, de modo que se erigen como puente entre los modos de buscar de un usuario modelo y las formas de segmentación textual con las que el documentalista organiza las noticias para que puedan ser recuperadas.

Pêcheux (1978) realiza un amplio estudio sobre los conectores en su metodología para analizar automáticamente los discursos, pero la exhaustiva tipología que presenta se muestra poco útil y excesiva dadas las restricciones de nuestros registros, no obstante, nos servimos en gran medida de sus aportaciones teóricas para establecer el modo de detección y reducción pragmática de conectores.

El conector introduce una función conceptual trasladable a un categoría práctica de análisis, de ahí su alto interés para este estudio. En las muestras se localizan varias decenas de conectores de distinto orden y con un alto grado de sinonimia y polisemia (mediante, a través de, en, por, a causa de, por motivo de, en respuesta a, sobre, tras, entre, contra, para, con objeto de, desde, durante, de, respecto a, relativo a, en relación a, según, de acuerdo a, en función de, con el fin de, por tanto, en consecuencia...) Se realizan operaciones de eliminación de polisemia y de control de sinonimia (de acuerdo a contextos dados) de modo que se obtiene la siguiente lista depurada, sometida a ciertas convenciones formales (como posición y siglas) y asociada a los modos de interrogación pronominal (12Ws) de manera unívoca y reordenada de un modo más próximo al lenguaje natural:

1. ¿Quién/Qué?: omitido, el conector es la posición en la MP. El sujeto precede a la acción.
2. ¿A/Contra quién?: a, dativo. Receptores directos de la acción, seres vivo, sucede a la acción.
3. ¿Qué cosa es objeto de la acción?: omitido, el conector es la posición. Sucede a la acción, tras el dativo.
4. ¿Mediante qué?: mediante/-mente, modal. Procesos no tangibles especializados u ordinarios.
5. ¿A través de qué?: a través de, instrumental. Objetos tangibles a través de los cuales se ejecuta la acción, precisan la mediación o manipulación por sujetos.
6. ¿Por qué?: a causa de, causal. Extensible a motivos y antecedentes.
7. ¿Para qué?: con el fin de, final. Intenciones y objetivos no alcanzados.
8. ¿Con qué consecuencias?: en consecuencia. Resultados alcanzados y constatados.
9. ¿Cuándo?: desde, hasta, durante, fecha, tiempo, periodo. Evitar "en" con valor locativo.
10. ¿Dónde?: en, lugar. Espacios genéricos o específicos.
11. ¿Con quién/qué?: Con, contra, asociativo. Sujetos colaboradores, no protagonistas.
12. ¿En qué estado?: omitido, el conector es la posición. Aparece junto al elemento que matiza.

Como se advierte, en ciertas lenguas no declinadas, como el castellano, la propia posición de un concepto en el enunciado es indicativa de función (habitualmente los sujetos y objetos directos), por lo que el listado de conectores genera indicios de un alto porcentaje de funciones pero no en todos los casos. A efectos de la construcción de MPs, siempre es necesario reservar unas posiciones determinadas en sus enunciados con valor de función unívoca. Por otro lado, es preciso insistir en el alto grado de polisemia de los conectores, lo que exige una formalización que puede resultar expresivamente tosca y señalar, además, que ciertos conectores no tienen ninguna función conectiva relevante a efectos de esta investigación, por ejemplo los que introducen negación: sin, no, pues la negatividad, como elemento de búsqueda o representación en una base de datos, debe reflejarse en conceptos que la contengan -rechazar, negar, desautorizar- o mediante operadores booleanos u otros recursos de búsqueda de las aplicaciones informáticas. En cuanto a otras partículas o elementos no abordados, debe resaltarse /de/ con función partitiva o clasemática no relevante como conector ya que su función permanece en el interior de los descriptores sintagmáticos, es decir, dentro de parámetros de la representación.

Restarían por clasificar y resolver, en una próxima investigación, pues son casos peculiares que excederían los objetivos de la presente, la problemática que introduce /entre/ como conector de orden participativo (cooperativo), casi siempre asociado a operaciones de sujetos y objetos (esto es, de segundo orden o nivel intercategorial). Finalmente quedarían también para resolución informática, las funciones de los importantes conectores de cuantificación como /más que/, /menos que/, de momento, integrables para ciertas acciones (vender, gastar y otras que permiten cuantificar) en derivaciones de preguntas sobre qué y con qué consecuencias (se gasta, vende, etc.) y se remite a otra investigación de distinto rango el estudio de las funciones valorativas de conceptos (mejor, peor, bueno, malo, etc), dimensión axiológica que no pertenece en puridad a la conectividad pero que requiere reflexión y solución por parte de la documentación periodística.

### 5.3 Acciones

En relación a los verbos que se han obtenido de la muestra de noticias, se advierte que son subsumibles, por generalización o sinonimia, en no más de un centenar de acciones prácticas cuyo objetivo es articular las MPs (ejemplo anterior en el que se subsumen, bajo visita o viaje, varias decenas de acciones propias o derivadas de una visita/viaje oficial). Inventariar todos los verbos, organizarlos en jerarquías y formalizarlos como descriptores no es una labor necesaria a efectos de los objetivos de este trabajo pero sí debe abordarse en una fase posterior para solventar la problemática de su representación (junto a descriptores de otra procedencia).

Ya en el nivel de registros, se observan acciones que, independientemente de su significado, interesan por sus implicaciones funcionales (transitivas o intransitivas, impersonales y reflexivas, con conectores explícitos o tácitos, etc.) y tienen repercusión directa en la construcción de las MPs. Una gran cantidad de verbos utilizados en las noticias, forman parte del campo semántico de "declarar" (decir, hablar, manifestar, dirigirse al auditorio, advertir, indicar, señalar, describir, disertar, explicar, exponer, afirmar...) y han sido objeto de estudio publicado en otro lugar (García Gutiérrez, 2013) por constituir una casuística peculiar, lo que se suma al hecho de que muchas noticias que basculan en torno a verbos de hacer (reunión, ataque, erupción...) en realidad son declaraciones, relatos del acontecimiento, generando MPs completamente diferentes. Debe ponerse especial cuidado en distinguir cuándo una noticia parte de un hecho o de un "dicho" sobre un hecho (¿se describe el ataque o la narración la hace un protagonista del ataque?), pues de la elección del verbo dependerá el resto del análisis y también la recuperación de la noticia (la acción declarar "arrastra" una radiografía conceptual de la noticia y "atacar" otra bien distinta). En ocasiones, hay que esperar al

último párrafo de la noticia para descubrir esta circunstancia. Por otra parte, hay que tomar en consideración, y eliminarla en su caso, la polisemia inherente a ciertos verbos (analizar, manifestar, dirigirse, juzgar, condenar...la misma acción de "declarar" en una entrevista o ante un juez) que pueden ser usados tanto en enunciados declarativos como de acción. Este artículo se centra en noticias con enunciados dominantes típicamente de acción, si bien la aplicación conjunta y homologada de los resultados de ambas investigaciones (sobre enunciados de acción y "de decir") será esencial para obtener un procedimiento eficaz (un ejemplo indicativo de declaración como subtexto de la MP2 al final del epígrafe 6).

## 6. DESARROLLO DEL PROCEDIMIENTO Y DISCUSIÓN DE RESULTADOS

Una vez observado el corpus inicial, cuyo fin es la provisión de inferencias de procedencia empírica, se procedió a la "fabricación" de las anunciadas MPs de laboratorio (y temáticamente ficticias), exclusivamente atendiendo al objetivo de la inclusión exhaustiva de las 12Ws en una estructura ideal y puramente teórica que simulase a un usuario-modelo que construye con sus preguntas una síntesis "factualmente exhaustiva". Este segundo procedimiento, que parte desde un lugar de enunciación distinto al corpus, permite además la detección de lagunas y duplicidades estructurales en las MPs obtenidas de los registros empíricos, obligando a un regreso y revisión de los textos de partida, y generando una mayor formalización de las MPs de las muestras al ser más aproximadas estructuralmente a los requisitos del modelo de demanda establecido. Veamos un ejemplo de texto simple MP1 para posteriormente, una vez que se hayan afianzado las denominaciones y alcances de las categorías asociadas a interrogaciones pronominales, proceder con un segundo ejemplo complejo MP2 (dos subtextos). La posición de las funciones conceptuales (temporalidad, objetos, lugares, etc.) ha sido sometida aquí a los objetivos de una mayor comprensión del texto, alteración que no comporta incidencia alguna para el análisis por parte de documentalistas pero que sería inviable en caso de una lectura automática en la que la posición implica función:

MP1 "El ejército ruso, apoyado por el gobierno georgiano, ataca en marzo de 2001 a los ya diezmados guerrilleros chechenos, bombardeando el mercado de Grozny con aviones Mig y artillería pesada y causando la destrucción del lugar y cien muertos civiles, con el fin de controlar la capital chechena y en respuesta a dos atentados ocurridos en Moscú semanas antes atribuidos a los guerrilleros, todo ello bajo una fuerte presión internacional en contra de la intervención militar rusa."

Segmentación de las funciones conceptuales asociadas a preguntas pronominales prototipo del modelo de demanda:

0. ataque: acción dominante.
  1. ejército ruso: ¿quién ejecuta la acción? sujeto colectivo.
  2. gobierno georgiano: ¿con quién? sujeto asociado
  3. marzo de 2001: ¿cuándo? fecha de la acción.
  4. diezmados (derrotados): ¿en qué estado se encuentra el sujeto pasivo? estado de quién sufre la acción.
  5. guerrilleros chechenos: ¿a quién perjudica/beneficia la acción? sujeto beneficiado/perjudicado.
  6. bombardeo: ¿cómo se produce la acción? procedimiento de la acción
  7. mercado de Grodzny: ¿qué cosa se beneficia o perjudica de la acción? objeto directo
  8. con aviones Mig y artillería pesada: ¿a través de qué se ejecuta la acción? medios tangibles
  9. causando la destrucción del lugar y cien muertos civiles: ¿qué consecuencias tiene la acción? resultados, efectos palpable, independientemente de las intenciones.
  10. obtener el control de la capital chechena: ¿qué fines tiene la acción? intenciones, deseos, independientemente de los resultados.
  11. en respuesta a dos atentados en Moscú atribuidos a los guerrilleros: ¿por qué se ejecuta la acción? motivo de la acción, también con un valor de antecedente.
  12. bajo fuerte presión internacional contra la intervención rusa: ¿en qué estado se encuentra el sujeto? Nuevamente aparece esta función (campo 4) asociada ahora al sujeto colectivo 1.

Con el fin de dotar de una denominación a cada una de las preguntas del formulario de usuario, se procedió a comparar el alcance de las mismas con el ámbito semántico de las categorías de organización textual, ya ensayadas por este investigador en otro lugar (García Gutiérrez, 1999) a partir de los trabajos de Cunha (1990) a su vez basados en la teoría de los casos universales de Fillmore (1968), coincidiéndose en gran medida en las categorías de análisis documental propuestas aun siendo necesario modificarlas y adaptarlas al entrar en juego los formatos de interrogación de los usuarios como horizonte director. Los casos universales originales de los que partía la discusión de Isabel Cunha con el objetivo de aplicarlos a un lenguaje documental sobre historia colonial portuguesa, fundamentalmente relacionada con Brasil, fueron: Agentivo, Dativo, Nominativo, Ergativo, Acusativo, Causativo, Instrumental, Sociativo, Final, Beneficiativo y Locativo, que la misma autora redujo en: Agente, Objeto, Instrumento, Lugar, Tiempo, Producto y Finalidad.

Una vez adaptadas estas categorías esenciales, tanto conceptual como terminológicamente, al mo-

delo de las 12Ws, fueron organizadas sobre ellas todas las MPs de la muestra empírica en caso de estar explícitas las oportunas respuestas en algún lugar de las noticias. Las respuestas, además, aparecen dispersas y camufladas bajo enunciados en ocasiones retóricos, lo que hace pensar en una elaboración permanente de MPs por parte de los documentalistas, incluso en el caso de plena automatización del sistema. La macroproposición sería la clave del éxito o fracaso de un supuesto análisis automático de los textos y de todas las rutinas derivadas por lo que debe estar reservada a construcción humana. Por otra parte, se manifiesta la necesidad de elaborar definiciones simples y denominaciones intuitivas para dotar de mayor formalización a las preguntas del protocolo que han de seguir los analistas.

A continuación, se ofrece una propuesta desarrollada del modelo de interrogación pronominal junto a una breve discusión normativa en cuanto a su alcance, límites e interacciones, encabezadas por la acción y agrupadas por afinidad e incompatibilidad manifiestas, junto a una posible sigla mnemotécnica que aligere su reconocimiento y algunas claves para superar la confusión ocasional. Tras ello, se ofrece la resolución de la MP2, de mayor complejidad que la anterior, con dos subtextos independientes. Formalizando las denominaciones de las funciones conceptuales implicadas en el texto, obtenemos las siguientes categorías operativas para el documentalista, a partir de una simple pregunta pronominal del usuario y ordenadas en función de su proximidad y mayor comprensión:

- Acciones: la acción (a) se obtiene en primer lugar, pues constituye el ángulo vertebrador de todas las interrogaciones a que se somete el texto, y es representada habitualmente por un verbo en la MP (aunque su representación final habría de ser un sustantivo que indique proceso, como ataque, reunión, prohibición, erupción...). Se trata de elegir el verbo/acción que destaque como dominante respecto a las múltiples acciones secundarias dispersas en la noticia y de ahí la necesidad imperiosa de abordar el inventario y modelización de acciones procedentes de noticias en otro trabajo. En el caso de evidenciarse dos o más subtextos en una misma noticia, es decir, dos o más relatos no subordinados a la acción dominante primera (posiblemente manifestados en los subtítulos de la noticia), es necesario establecer nuevas acciones dominantes, tantas como subtextos se hayan detectado (más adelante se ofrece un ejemplo de dos subtextos), y proceder a análisis paralelos como si se tratara de noticias independientes (se produce una /reunión/ de ministros de petróleo en El Cairo y se /desborda/ el Nilo el mismo día: las acciones dominantes serían reunión y desbordamiento). La subsunción de todas las formas verbales de un texto bajo una rúbrica dominante es la operación más importante de todo el procedimiento pues de ella depende el desarrollo de las demás categorías. La acción dominante es como la columna vertebral de la MP y las categorías

restantes constituyen los distintos elementos del esqueleto que se observa en la radiografía textual que hacemos a partir de la óptica de la acción central. Es preciso subrayar la necesidad de formalizar los tiempos y voces verbales de la noticia a presente de indicativo, especialmente para evitar futuribles que pueden no tener lugar (los trabajadores irán a la huelga o se manifestarán...). En el caso de futuro, como se expuso en 5.3, tal tiempo debe ser sustituido por un verbo que lo contenga en presente de indicativo (los trabajadores anuncian o convocan una huelga...) de este modo se evita una complicada revisión retrospectiva, habida cuenta de que, a la vista de la muestra, un alto porcentaje de las noticias son "futuribles".

- ¿Quién/Qué ejecuta la acción? Sujetos y objetos ejecutantes de acciones: por un lado, se trata de sujetos individuales o colectivos (seres vivos) que ejecutan acciones y serán denominados agentes (ag). Los objetos que ejecutan acciones por sí mismos, como fenómenos o elementos inorgánicos de la naturaleza (el río Nilo se desborda, un meteorito cae, el volcán entra en erupción...) sugieren una tipología aparte, y deben ser considerados como objetos ejecutantes o activos (obc), si bien podrían estar integrados en una doble categoría con los agentes "orgánicos" o seres vivos, humanos, animales o virus ejecutores de acciones, siempre que vayan debidamente diferenciados y mantengan incompatibilidad (la misma acción no puede ser ejecutada por un ag y un obc simultáneamente). Al efectuar el usuario del sistema una pregunta del tipo ¿quién ataca a los chechenos...?, ¿quiénes se reúnen en El Cairo...?, ¿qué cae en Rusia central...?, ¿qué entra en erupción?, el sistema de búsqueda debe relacionar quién/qué al campo de ag/obc a partir de las acciones ejecutadas con los campos de los registros en los que aparecen ocurrencias como ataque, El Cairo, el meteorito M, el volcán V...siempre asociadas a sus funciones conceptuales y de ese modo proporcionar las respuestas adecuadas.

- ¿A quién/A qué cosa va dirigida la acción? Seres vivos/objetos receptores: los seres vivos que son directamente objetivos de una acción se considerarán pacientes (p) y los objetos en la misma circunstancia se considerarán objetos pasivos (op). Determinados verbos, como los intransitivos y también, por su naturaleza semántica, muchos transitivos, no admiten la presencia de pacientes u objetos por lo que puede ser necesario modificar los verbos asociados a las acciones si en las macroproposiciones existen las funciones conceptuales incompatibles con los verbos inicialmente seleccionados. Por ejemplo, el verbo juzgar admite paciente (A juzga a P) y también determinados objetos pasivos: no se puede juzgar una casa pero sí una declaración (A juzga la declaración del ministro sobre...).

- ¿Con quién/Quién acompaña (y no protagonista)? Asociativos (asc): sujetos asociados a agentes

o pacientes como cómplices o acompañantes (el apoyo de Georgia al ejército ruso). Los asociativos son preguntados a los roles actanciales y no a la acción misma: estos implican a los agentes, pacientes (quien presenció o acompañaba a la víctima de un atentado podría ser su asociativo). En caso de objeto activo (un huracán (obc) puede venir acompañado de fuertes lluvias (asc)), o incluso un objeto pasivo podría tener la incidencia de un asociativo (edificio anexo al mercado, con daños menores, que no fue objeto directo del ataque). En consecuencia, aun no siendo un ángulo de búsqueda esencial, la función-conceptual asociativa podría ofrecer matizaciones y detalles muy útiles para algunos usuarios, por lo que se sugiere mantener su presencia como categoría secundaria adscrita a los cuatro roles actanciales con los que podría estar potencialmente vinculada.

- ¿En qué estado? Situación (st): se trata de otra categoría que puede matizar la situación de sujetos y objetos. La situación se refiere al estado involuntario en el que pueden encontrarse las categorías que ejecutan o sufren la acción, por tanto es necesario efectuar la pregunta a todas ellas: ¿en qué estado o situación se encuentra ag, obc, p, op? Posibles respuestas serían: presión internacional (presidente sirio, ag), latente (volcán, obc), cercado (atracador, p), contaminado (cuartel, op). La categoría más próxima con la que la situación puede entrar en confusión es el modo. Para evitarlo, debe aplicarse el criterio de involuntariedad de la situación frente a la voluntariedad del modo (además de la condición secundaria de la situación, dependiente de los actantes, frente a la "primeridad" del modo, dependiente de la acción).

- ¿A causa de qué/por qué? Causa (c): la categoría causal debe ser entendida de un modo amplio pues por eficacia operativa, en estas primeras instancias de investigación, se le ha hecho abarcar tanto las motivaciones directas (las causas propiamente dichas), los antecedentes (atc) que aparecen en la noticia y su contexto (ctx) cuando es necesario invocarlo (por ejemplo, un bombardeo determinado se inserta en el contexto "guerra de Irak", si fuera necesario, a efectos de no perder el registro por descontextualización excesiva). Aunque parece obvio el campo que restringe la pregunta ¿por qué?, es necesario advertir que en numerosas experiencias de análisis se ha observado que hay una tendencia (derivada de la comunicación ordinaria) a responder a "por qué" con "para" en lugar de "porque", lo que provoca una confusión de la categoría causal con la final. Por tal razón, se sugiere efectuar la pregunta ¿a causa de qué ocurre...? en lugar del lapidario ¿por qué ocurre...? de manera que pueda asegurarse la dimensión causal de la respuesta.

- ¿Para qué? Finalidad (f): deseos, objetivos, propósitos no alcanzados. Una vez mencionadas las cautelas que han de tomarse en relación a la causa, también debe señalarse la identificación, más

común de lo que sería esperable, de la finalidad con la consecuencia (cc). La diferencia, evidente, entre una y otra, no solo reside en la pregunta sino fundamentalmente en el carácter no consumado de la respuesta en el caso de la finalidad y consumado, con datos y efectos comprobables en la consecuencia (por ejemplo; se promociona la película para que tenga éxito (f) vs. se promociona la película y tuvo un verificado éxito de taquilla (cc)).

- ¿Con qué consecuencias? Consecuencia (cc): son los resultados, efectos, conclusiones en relación a una acción. La categoría que puede interactuar en algunas ocasiones con la consecuencia, además de la finalidad -ya aclarada-, es el objeto pasivo además, en menor medida, del paciente. Pero la consecuencia en ningún caso es el objeto directo de la acción (se atenta contra el ministerio (op)/ o contra el ministro (p)) sino las consecuencias reales y constatadas de la ejecución de la acción (destrucción del edificio (cc), ileso o herido o muerto (cc)).

- ¿Mediante qué/Cómo? Modo (m): procedimiento, técnica, manera, género. La principal confusión del modo procede de sus cruces con la categoría instrumento (i) y se disipa acudiendo al carácter tangible del medio o instrumento frente a la intangibilidad de la categoría modal. Los adverbios modales relevantes y asociados a la acción dominante, pertenecen a esta categoría (provisionalmente, urgentemente, cautelosamente...).

- ¿A través de qué? Instrumento (i): el instrumento es el medio físico, la tecnología, la empresa a través de la cual se desarrolla o tiene lugar la acción. La respuesta debe ser un objeto tangible. X (ag) atraca (a) a Y (p) mediante golpes (m) con/a través de un bate de béisbol (i).

- ¿Dónde y Cuándo? Lugar (l) y Tiempo (t): ambos locativos constituyen las coordenadas espacio-temporales de los datos. En principio, no se han observado interferencias entre los locativos y las categorías establecidas previamente salvo rara vez con el op (lo que se ataca y dónde se ataca).

Finalmente, se presentan esos mismos resultados en las MP1 y MP2, en las que se recoge toda su tipología de modo integrado:

MP1 (a) ataque

¿Quién ataca?<-->(ag)<-->ejército ruso

¿A quién ataca?<-->(p)<-->guerrilleros chechenos

¿Qué se ataca?<-->(op)<-->mercado de Grodzny

¿Quién apoya a ag, obc, p, op?<-->(asc)<-->apoyo político/ Georgia (ag)

¿En qué estado se encuentran ag, obc, p, op?<-->(st)<-->presión internacional (ag)/ derrotados (p)

¿Por qué ataca?<-->(c/atc)<-->atentados Moscú

¿Cómo/mediante qué se ataca?<-->(m)<-->bombardeo

¿A través de qué se ataca?<-->(i)<-->aviones Mig/ artillería pesada

¿Para qué se ataca?<-->(f)<-->control militar/ Chechenia

¿Qué consecuencias tiene el ataque?<-->(cc)<-->destrucción/ mercado/muertos civiles

¿Dónde se ataca?<-->(l)<-->Grodzny/ Chechenia

¿Cuándo se ataca?<-->(t)<-->20010300

MP2: Subtexto a: El volcán V, considerado ya en reposo tras las erupciones catastróficas de 1880, entra en erupción ayer (22 de noviembre de 2010) lanzando al exterior, mediante explosiones irregulares, rocas y cenizas que han causado pánico entre la población del Estado mexicano de EEE, además de numerosos destrozos en las aldeas cercanas. Los ecologistas habían advertido ya del peligro por la proximidad de las aldeas al volcán. A la erupción del volcán se suma una intensa tormenta tropical. // Subtexto b: Por otro lado, el presidente mexicano PM, contando también con la presencia del presidente de la Comisión Europea PCE, declaró en primicia que asiste mañana (24 de noviembre de 2010) en Oaxaca a la inauguración del X Congreso internacional de Vulcanología, según dijo en entrevista con la periodista Lourdes Pérez emitida con urgencia por el canal 10TV y a petición de la Sociedad mexicana de Vulcanología, con el fin de conseguir mayor participación en el evento y hacer un llamamiento a la sensibilidad de la comunidad científica ante las constantes catástrofes humanitarias.

MP2 (a) 1. erupción 2. declaraciones

¿Quién?<-->(ag)<-->2. PM/ presidente/ México

¿Qué?<-->(obc)<-->1. volcán V

¿A quién?<-->(p)<-->2. Pérez, Lourdes/ periodista

¿A/sobre qué?<-->(op)<-->2. asistencia/ inauguración/ X Congreso internacional de Vulcanología/ Oaxaca/ 20101124

¿Con quién/con qué?<-->(asc)<-->1. tormenta tropical (obc) 2. PCE/ presidente/ Comisión Europea (ag)

¿En qué?<-->(st)<--> 1. reposo (obc)

¿A causa de qué?<-->(c)<-->1. (atc.) catástrofe/ 18800000 / (atc.) advertencias/ ecologistas/ peligro/ proximidad/ aldeas 2. (motivo) petición/ Sociedad mexicana de Vulcanología

¿Cómo/mediante qué?<-->(m)<-->1. explosiones irregulares 2. entrevista/urgencia/primicia

¿A través de qué?<-->(i)<-->1. cenizas volcánicas/ rocas volcánicas 2.Canal 10TV

¿Para qué?<-->(f)<--> 2. aumento/ participación/ congreso/llamamiento/ sensibilidad/ comunidad científica/ catástrofe humanitaria

¿Qué consecuencias?<-->(cc)<-->1. pánico/ población/ destrucción/ aldeas

¿Dónde?<-->(l)<-->1. estado EEE/ México

¿Cuándo?<-->(t)<--> 1. 20101122 2. 20101124

## 7. CONCLUSIONES

Una vez culminado el análisis, procede resumir las conclusiones obtenidas en relación a la hipótesis de trabajo, que queda confirmada en todos sus términos generales, y explicar las fases acometidas y los resultados particulares alcanzados: 1) se parte de un inventario pragmático de los modos de preguntar de los usuarios, desde la posición más exhaustiva posible, pero con el objetivo de obtener respuestas factuales (personajes, fechas, lugares, hechos, modos, causas, consecuencias...) de un sistema de documentación periodística, en esta investigación inicialmente restringida a noticias escritas en medios españoles, y se procede a formalizar las preguntas de orden pronominal en 12 formulaciones (12Ws), 2) se elabora, por otro lado, un muestra de noticias de acuerdo a la metodología de validación denominada "técnica de la mitad" para asegurar la exhaustividad del corpus con objeto de estudiar las tipologías textuales y los conectores, 3) se revela como una operación fundamental elaborar macroproposiciones a partir de la muestra empírica de noticias para realizar las observaciones y extraer resultados homologables, 4) se elaboran varias macroproposiciones teóricas a partir de los modos de formulación con el fin de ser cotejadas con los registros empíricos y favorecer su formalización estructural retroactivamente, 5) se construye un inventario de conectores a partir de la muestra y tras su depuración (reducción de sinónimos y polisemias) se asocia cada conector a un modo de interrogación pronominal determinado, 6) se comprueba que las 12Ws encuentran respuesta bajo tipologías textuales muy heterogéneas en los registros de la muestra y que no aparece ninguna otra función conceptual básica que pueda satisfacer preguntas adicionales (salvo cuantificaciones, valoraciones, y asociaciones intercategoriales o de segundo orden remitidas a una investigación ulterior), 7) se asocia cada conector y posición funcional, determinados por los modos de formulación, a categorías de interrogación y segmentación textual con objeto de establecer un procedimiento de análisis para el documentalista, 8) se comprueba que a cada formulación pronominal específica corresponde una respuesta factual precisa y adecuada mediada por una categoría de análisis y segmentación textual, 9) se confirma el alto nivel de homologación, explicitación y verificación del procedimiento de análisis seguido por el documentalista de modo que pueda ser revisado por terceros (lo que dota de mayor fiabilidad y cientificidad a un proceso hasta el momento altamente subjetivo y precario), 10) el procedimiento de análisis y representación propuesto no impide obtener los productos

convencionales de las bases de datos periodísticas (texto completo y referencias) toda vez que estos se obtienen mediante la coincidencia de ocurrencias en el sistema entre los mismos descriptores resultantes de la formulación pronominal. Además, el procedimiento permite obtener otras respuestas derivadas de las factuales como son las verificaciones y comparaciones de datos y 11) se infiere la utilidad del enfoque pragmático para abordar problemas de análisis, representación y recuperación en sistemas de documentación periodística.

Debe advertirse, a efectos prácticos, que si bien no es necesario el auxilio de software en la fase de análisis de la distintas categorías y ejes de significación inmediatamente convertidos en ángulos de búsqueda, sí es imprescindible una aplicación informática con potente motor de búsqueda relacional que enlace cada función conceptual a una categoría y cada categoría a un modo de interrogación pronominal de usuario. No es en absoluto descartable la posibilidad de una casi completa automatización del sistema de recuperación y análisis por parte de una aplicación inteligente, si bien la construcción de las macroproposiciones a partir de una inmensidad de enunciados libres, heterogéneos y cargados de retórica, procedentes de los textos de las noticias, habría de permanecer, al menos a largo plazo, bajo control de los documentalistas de prensa.

## 8. CAMPOS DE INVESTIGACIÓN INMEDIATOS Y COLATERALES

Se indican, a continuación, asuntos que deben ser abordados en otras investigaciones y algunas extrapolaciones posibles a raíz de las conclusiones obtenidas en este estudio:

- deben ser objeto de una investigación específica otros géneros informativos (entrevistas, reportajes) y de opinión en medios impresos que busque, no obstante, una homologación de categorías generales con el procedimiento propuesto para las noticias con el fin de formar parte de sistemas integrados de documentación periodística.
- debe resolverse la formalización de los operadores de cuantificación y valoración en el interior de las categorías primarias establecidas o como categorías secundarias independientes.
- el idioma para el que está pensado el procedimiento es el español estándar contemporáneo usado en medios españoles y no han sido considerados otros medios hispanófonos no españoles aunque, inicialmente, no se observan obstáculos importantes para aplicarlo sobre textos periodísticos enunciativos en los diversos usos legítimos del español y, ya con mayor grado de dificultad, si bien manteniendo la fundamentación del procedimiento, en otras lenguas derivadas del latín como catalán, gallego, portugués o italiano.

- el procedimiento propuesto abre un vasto campo de investigación teórica y práctica en materia de lenguajes de representación, construcción de thesauri, ontologías y epistemografías en el campo de la información de actualidad.
- las propuestas de este estudio van dirigidas a la optimización de los procedimientos de análisis y recuperación en bases de datos periodísticas profesionales (especialmente con la asistencia de aplicaciones inteligentes, reservando el trabajo humano a la elaboración de MPs y atención al usuario) pero también plantean beneficios evidentes para la dimensión docente de la documentación periodística cuya eficacia está restringida por la libre interpretación en el análisis documental de noticias.
- el procedimiento es particularmente útil para los sistemas de información escrita de cualquier medio informativo (prensa, radio, tv o internet) pero su extrapolación podría ser inmediata a las hemerotecas de prensa contemporánea y, aplicando las modificaciones oportunas, a las hemerotecas de prensa antigua.

## BIBLIOGRAFÍA

- Amaro, R. (1991). *Contribuição da análise do discurso para a análise documental: o caso da documentação jornalística*. Sao Paulo; Eca/Usp, 87 p.
- Andréu Abela, Jaime (2000). *Las técnicas de Análisis de Contenido: Una revisión actualizada*. <http://public.centrodeestudiosandaluces.es/pdfs/S200103.pdf> (consult. sept. 2013).
- Bonet, M.; Fernández Quijada, D. (2006). El reto de la digitalización del archivo sonoro en los servicios públicos de radiodifusión. El caso de Catalunya Ràdio. *El Profesional de la Información*, vol. 15 (5), septiembre-octubre, 390-396. <http://dx.doi.org/10.3145/epi.2006.sep.08>
- Cebrián, Bernardino (1994). El índice del periódico: una sugerencia para mejorar la información local. *Comunicación y Sociedad*, v.VII, n.1 [http://www.unav.es/fcom/comunicacionsociedad/es/articulo.php?art\\_id=223](http://www.unav.es/fcom/comunicacionsociedad/es/articulo.php?art_id=223) (consult. sept 2013).
- Cintra, A.M.M. (1987). Estrategias de lectura en documentación. En J. Smit (org.), *Análise documental*, Brasilia; Ibict, 27-35.
- Cunha, Isabel (1990). *Do mito a análise documental*. Sao Paulo; Edusp, 163 p.
- Eco, Umberto (1993). *Lector in fabula. La cooperación interpretativa en el texto narrativo*. Barcelona; Lumen, 330 p.
- Fernández Gil, J.R. Fuentes de análisis para el estudio de la prensa diaria. *Anales de Documentación*, 2010, vol. 13, 135-158.
- Fillmore, C.J. (1968). The Case for case. En E. Bach y R.T. Harms (eds.) *Universals in Linguistic Theory*, New York; Holt Rinehart and Winston, 1-88.
- Fuentes i Pujol, M. Eulalia (1997). *Documentación y Periodismo*. Pamplona; Eunsa, 130 p.
- Galdón, Gabriel, (coord.) (2002). *Teoría y práctica de la documentación informativa*. Barcelona: Ariel, 2002, 67-89.
- García Gutiérrez, Antonio y Lucas Fernández, Ricardo (1987). *Documentación automatizada en los medios informativos*. Madrid; Paraninfo, 264 p.
- García Gutiérrez, Antonio (coord.) (1999). *Introducción a la documentación informativa y periodística*. Sevilla; Mad, 510 p.
- García Gutiérrez, Antonio (2013). Entrevistas periodísticas y textos declarativos: un procedimiento de análisis y recuperación documentales basado en modos de búsqueda pronominal. *El Profesional de la Información*, julio-agosto, v. 22, (4), 315-325. <http://dx.doi.org/10.3145/epi.2013.jul.07>
- Gómez Vázquez, Miguel. Servicio de archivo y documentación de un periódico: El Mundo. *Cuadernos de documentación multimedia* nº 6-7 <http://www.ucm.es/info/multidoc/multidoc/revista/cuad6-7/gomez.htm> [consultado en noviembre 2012].
- Guallar, Javier; Abadal, Ernest y Codina, Lluís (2012). Hemerotecas de prensa digital: evolución y tendencias. *El Profesional de la Información*, noviembre-diciembre 2012, vol. 21, núm. 6.
- Huysen, Andreas (2000). *Seduzidos pela memoria*. Rio de Janeiro; Ed. Aeroplano; Universidade Candido Mendes; Museo de Arte Moderna, 2000, 116 p.
- James, William (2007). *Pragmatismo*. Madrid; Alianza ed., 258 p.
- Krippendorff, K. (1990). *Metodología de análisis de contenido*. Barcelona; Paidós, 270p.
- Lara, M. L. G. (2011). Conceitos de organização e representação do conhecimento na ótica do Grupo Temma. *Informação & informação*, vol. 16, 92-121.
- Pêcheux, Michel (1978). *Hacia el análisis automático del discurso*. Madrid; Gredos, 269 p.
- Pottier, Bernard (1976). *Linguística general*. Madrid; Gredos, 426 p.
- Rubio Lacoba, María y Blanco García, Juan Carlos (2010). Mejor un buscador que un encontrador. *Documentación de las Ciencias de la Información*, vol. 33, 273-287.
- Sanz Casado, Elías (1994). *Manual de estudios de usuarios*. Madrid; Fundación Germán Sánchez Ruipérez, 290 p.
- Van Dijk, Teun (1990). *La noticia como discurso*. Barcelona; Paidós, 284 p.
- Wittgenstein, Ludwig (2004). *Investigaciones filosóficas*. Barcelona; Crítica, 552 p.