

LOS LENGUAJES DOCUMENTALES EN LAS BASES DE INFORMACION POLITICA Y DE ACTUALIDAD

Cristina Rodríguez Vela*

Resumen: La vieja discusión entre lenguaje libre y lenguaje controlado cobra especial trascendencia cuando se trata de bases de información política y de actualidad. La conveniencia o no de aplicar un tesoro a las tareas de análisis documental de la prensa diaria y semanarios de actualidad es objeto de este trabajo. La riqueza conceptual de esta documentación desborda la adaptación de un tesoro a esa realidad cambiante. Sin embargo, la combinación de búsqueda en texto libre y mediante descriptores garantiza el éxito de la misma. La adopción de ambos procedimientos asegura una estrategia que, por su carácter complementario, parece la aconsejable para el tratamiento de este tipo de documentación.

Palabras clave: Análisis documental, búsqueda documental, tesoro, lenguaje libre, lenguaje controlado.

Abstract: The old controversy between free and controlled language becomes specially bitter when dealing with political information and current events bases. This study questions the suitability of applying a thesaurus to the task of document analysis of daily and weekly current events press. The conceptual richness of these documents goes well beyond the appropriateness of applying a thesaurus to such changing reality. Nevertheless, success is guaranteed with combined search on free-text and by means of descriptors. The adoption of both procedures assures strategy which, due to its complementary characteristic, seems to be most advisable when treating this sort of documentation.

Keywords: Document analysis, document retrieval, thesaurus, free language, controlled language.

1. Introducción

Abordar el análisis del lenguaje documental más apropiado a un dominio temático tan amplio exige una reflexión previa, no exenta de complejidad, a la luz de los materiales documentales concretos y teniendo en cuenta el uso que de ellos se requiere.

Cuando se habla de información política y de actualidad se utiliza la coordinación de dos términos intencionadamente, sin olvidar que la información de actualidad en gran medida se refiere a la información política, o, dicho de otro modo, la información política ocupa un lugar privilegiado en la información de actualidad.

* Ministerio del Portavoz del Gobierno. Documentación
Recibido 24-9-91

Sin embargo, utilizamos la expresión información política para lo que se entiende por tal y se reserva el término actualidad para denominar de una forma amplia, general e imprecisa un campo de significación donde cabe toda clase de información que cumpla el requisito de ser actual.

Presentada la cobertura temática de estas bases en este marco de generalidad, se consideran adecuadas a estos propósitos todas aquellas informaciones que tengan un interés especial para la vida política del país o especial relevancia por su valor de opinión en temas concretos. Todo aquello que sucede, sin que importe tanto su naturaleza como su incidencia y repercusión, interesa a estas Bases. Tratándose de este tipo de información no constituye ninguna originalidad señalar que en la diversidad de fuentes tengan un lugar destacado los periódicos, sean estos escritos o no, de forma que el periodismo de radio y TV puede proporcionar también materiales documentales. Adentrándonos en la línea de los contenidos políticos en soporte impreso llegaríamos a materiales tan apropiados y específicos como el Diario de Sesiones de las Cortes, que recoge las intervenciones parlamentarias en general y discursos de los Ministros y del Presidente del Gobierno.

Pero vamos a centrar nuestra atención en la prensa periódica. Así pues, los materiales impresos periódicos, diarios o semanales, nacionales o de cualquier otro ámbito, así como cualquier otra fuente impresa con contenidos políticos constituyen las fuentes de alimentación de estas Bases.

2. El lenguaje periodístico y el lenguaje documental

Las especiales características de los materiales documentales, periodísticos preferentemente, tanto informativos como de opinión, aconsejan hacer algún apunte sobre las particularidades del lenguaje periodístico, para analizar después el lenguaje documental más idóneo.

Nos interesa el estudio de este lenguaje, aún someramente, para centrar las notas que lo definen y hacer la opción del lenguaje documental más adecuado a ese lenguaje de los periódicos que constituye, en cierto modo, una lengua especial, «que —hoy como nunca— recoge de manera viva y puntual todos los registros de la lengua hablada» (1).

Las controversias sobre el lenguaje periodístico desembocan en una valoración desigual del panorama de la expresión lingüística en los medios. «Cuando alguien habla del lenguaje periodístico es muy de temer que esté hablando de un lenguaje frívolo, fugaz, sin consistencia. Aunque, eso sí, más o menos agradable» (1).

Desde una visión purista y normativa de la lengua, el lenguaje de la prensa levanta indignación con frecuencia aunque, por otra parte, se le reconozca un papel activo y de largo alcance. Valoración positiva es la de Carlos Seco cuando dice que «los medios de comunicación de masas son los principales maestros del lenguaje», o la de Alarcos, que reconoce que «hay una unidad fundamental patentemente mantenida por la lengua de la literatura... y por la lengua menos cuidada de los medios de comunicación» (1).

Lingüistas y académicos se han ocupado especialmente del lenguaje periodístico y se han detenido a analizar el estado de la lengua en este lenguaje. Aunque

sigue prevalenciando una cierta inclinación negativa, se le reconoce sin embargo riqueza y variedad, notas éstas que serán determinantes de la opción del lenguaje documental más apropiado.

Pero si «el periódico es un nido de lenguajes, una especie de permanente diccionario de actualidad donde se cobija toda habla por efímera y fugaz que sea» (1), no hay que olvidar que, además de esa riqueza, se vale de lenguajes sectoriales, especializados. Estos lenguajes especializados pertenecientes a diversas disciplinas configuran un panorama lingüístico propio de un dominio universal del conocimiento.

Hay que resaltar con especial énfasis, parafraseando a Brian Steel, que la característica más importante reside en «la variedad de vocabulario, de los giros, de las construcciones gramaticales» y, sobre todo, hay que señalar de una manera muy especial que esta lengua, además de trabajar con palabras, es una «lengua representada... trabaja también con fotografías y procedimientos gráficos», en palabras de René Pucheu.

La relación entre el lenguaje periodístico y el lenguaje de intercambio verbal nos traslada al núcleo del problema —desde la perspectiva de los lenguajes documentales—: el texto de un periódico no se puede leer como el texto de un libro porque aquél presenta un lenguaje complejo donde los acompañamientos paralingüísticos y no lingüísticos pueden cambiar y transformar el sentido. El periodismo dispone de técnicas capaces de manipular los mensajes verbales. Mediante la retórica tipográfica inclina al lector a formarse un juicio de un suceso aparentando una correcta objetividad en la descripción.

Interesa esa combinación de «distorsión» y de «representación» por la dificultad que añade a nuestro estudio que, tratando del lenguaje documental, no puede más que ceñirse al estudio de la palabra como herramienta de comunicación.

Por otra parte el lenguaje periodístico se caracteriza por la rapidez del cambio de su catálogo léxico. Podría discutirse si tiene o no un léxico amplio, pero parece generalmente admitido que, sea como sea, es cambiante. «El catálogo léxico de las ofertas del consumo es muy amplio, cada vez mayor, y ésta es quizás su singularidad, pero al mismo tiempo es un catálogo muy limitado y en continuo cambio... la prensa, la radio y la TV imponen... pautas de pensamiento y de expresión que se renuevan con rapidez» (1).

Se atribuye, pues, a la prensa, radio y TV un doble efecto. Si por una parte contribuyen a la estandarización de la lengua, y por tanto a cierta igualdad lingüística, este mismo efecto democratizador es perverso desde el punto de vista lingüístico por desarrollar una lengua común estandarizada, artificialmente creada. Es lo que Garelli, en su polémica con Mauro, llama burolingua o lengua burocratizada, empleada por el periodismo. Sin embargo Duneton, que comparte esta posición, salva al periodismo español de «esta forzada igualación idiomática ya que se encuentra en los diarios españoles mayor vigor lingüístico» (1).

En cualquier caso, aunque la prensa española sea más vigorosa, está de igual forma sometida a esa tensión de fuerzas contrapuestas: *riqueza* frente a *estandarización*. Queda así planteado el panorama de discusión en torno al lenguaje periodístico, definido por notas que en principio pueden parecer contradictorias, riqueza y variedad frente a igualación y estandarización, y en ambos casos cambio y renovación de expresiones. En cualquier caso, el lenguaje periodístico supone un

elemento de enriquecimiento «ya que... no es un producto de laboratorio... sino, casi siempre, reflejo de los hábitos verbales vivos que circulan por la calle» (1).

3. Definición de la base y análisis documental

3.1. Definición de la Base

El tratamiento de estos materiales documentales requiere una estrategia adecuada en las *fases de análisis y recuperación* de la documentación.

Pero se plantea una cuestión previa, qué tipo de base interesa: referencial o, por el contrario, de texto íntegro. Aun disponiendo de un sistema de gestión documental capaz de almacenamiento del texto completo y recuperación en texto libre, se definen estas bases como bibliográficas o referenciales, si bien la referencia va acompañada de un lead o resumen de la noticia, elemento válido de información, que sustituye adecuadamente al texto.

Es cierto que las bases de texto completo, cada vez más numerosas, se multiplican con la expansión de la edición electrónica y con la disponibilidad de documentos escritos sobre soportes electrónicos legibles por ordenador. La tendencia, pues, apunta a que nazcan como bases de texto completo aquéllas resultantes de la edición electrónica.

Cuando no se parte de este supuesto, sino que la base es producida por un centro diferente al que genera la noticia, se opta por caminos más económicos en tiempo y en equipo y se definen las bases como referenciales, tras las razones que a continuación pasamos a enunciar.

Partiendo de que la alimentación de la base es externa, la entrada de datos supone una costosa tarea que puede ser resuelta por métodos automáticos mediante periféricos —scanner o digitalizador OCR— o, en su defecto, según procedimientos convencionales de grabación.

La imagen del scanner —constituida, en este caso, por texto— presenta el gran inconveniente de que ocupa mucha más memoria que un texto procesado, además de no admitir modificaciones, ya que el scanner no interpreta palabras.

Los digitalizadores OCR, sin embargo, reconocen cada carácter presentado y lo almacenan como texto, únicos ficheros sobre los que operan los programas de recuperación documental. Las tecnologías OCR —reconocimiento óptico de caracteres— presentan la ventaja de no necesitar más memoria que cualquier texto procesado, sin embargo, tienen el inconveniente de no reconocer cualquier tipo de letra. Es ahí donde radica el factor de dificultad.

Recordemos que se trata de una *lengua representada* y que la «representación» descansa en gran medida en la diversidad tipográfica, ante la que los procedimientos OCR todavía no responden con éxito absoluto.

Se desestima, de esta suerte, la entrada masiva del texto íntegro mediante procedimientos automatizados de digitalización, único procedimiento viable para la alimentación de este tipo de bases. Esta dificultad de tipo técnico apoya la elección de una *base referencial* en la que la introducción de datos de los campos textuales no suponga un trabajo de grabación convencional inabarcable.

3.2. Análisis Documental

En este marco de base referencial nos encontramos con una situación paradójica de entrada. El campo temático de nuestro trabajo documental abarca, en principio, una temática de tipo general, enciclopédica, donde, además, caben los universos o dominios de especialización, aunque no de manera rigurosa. Sobre estos dominios enciclopédicos/especializados se abre una gran interrogación: qué técnicas de análisis aplicar y qué procedimientos son los más recomendables para un adecuado uso de la base, es decir, para una recuperación exitosa del documento, salvando la siempre existente amenaza del ruido y del silencio.

En estos dominios tan universales se enmarca el debate del *análisis documental*: conviene la aplicación de un *lenguaje controlado*, a pesar de la amplitud de la cobertura del mismo, o es preferible la búsqueda en *texto libre*, sin ningún control de los términos. Ambas posibilidades van unidas a un planteamiento de análisis sustancialmente contrapuesto: se somete el texto a un *análisis inteligente*, manual, o por el contrario se confía el texto a los procedimientos de *indización automática*; ¿se aplican técnicas de sistematización o de análisis? En resumen, la misma pregunta desde perspectivas diferentes: qué lenguaje documental adoptar en las tareas de indización.

3.2.1. Sistematización versus análisis

Desde un punto de vista riguroso lo analítico se contrapone a lo sistemático, son operaciones en dirección contraria. Sin embargo es sabido que en el análisis inteligente se tiende a sistematizar. «La clasificación y la indización son aspectos interdependientes del análisis temático... todo lenguaje de indización se beneficia de la clasificación previa de los conceptos a indizar... (2)» El manejo de los tesauros es un recorrido por las clasificaciones o agrupamientos temáticos de los descriptores.

Sin embargo, a pesar de estas inevitables operaciones de clasificación en la actividad del análisis, es conveniente observar las recomendaciones de la norma ISO 5963-1986 sobre indización. Estas recomendaciones son claras en cuanto a la identificación de conceptos, recomendando exhaustividad y especificidad, que se pierde cuando un concepto se expresa con un término más general, con independencia de que su uso pueda ser conveniente en algunos casos. Análisis y sistematización se dan conjuntamente en una actividad de indización inteligente.

3.2.2. Lenguaje controlado, tesauros, versus lenguaje libre

Es sabido que la existencia de lenguajes documentales estructurados —tesauros— es directamente proporcional a la normalización de las disciplinas de que traten (3). El conjunto de noticias, escritos de opinión, que participan de dominios muy variados del conocimiento no permiten hablar de una disciplina única y por tanto no se puede hablar de normalización desde el punto de vista terminológico,

aunque sí caben en esa política y actualidad ciertos campos de especialización con sus respectivos lenguajes especiales, como ya se ha adelantado anteriormente.

En este sentido los tesauros escasean, dada la universalidad que se exige a su cobertura. Los existentes hacen un recorrido sistemático, ordenando la realidad en términos amplios. Esta cobertura universal hace que se denominen macrotesauros. Por otra parte, abandonando los lenguajes combinatorios y entrando en los lenguajes jerarquizados, esa realidad total ha sido recogida históricamente por la CDU, sin embargo el sistema de relaciones y polijerarquías a que se presta la actualidad y la política aconseja el trabajo de análisis y de indización con otro tipo de herramientas.

3.2.3. *Indización automática versus indización inteligente*

La controversia indización automática/indización manual va asociada en esta exposición a lenguaje libre/lenguaje controlado, respectivamente. Es cierto que se puede utilizar lenguaje libre en la indización inteligente, tomando términos del lenguaje natural sin ninguna operación de control, pero esta posibilidad está descartada por las múltiples entradas a que daría origen y el alto riesgo de silencio en las búsquedas.

Así pues, la *indización automática* va asociada al análisis automático que acomete el programa informático sobre el texto que se le somete. El programa actúa sobre el texto y organiza sus índices término a término, descartando aquellos que han sido previamente definidos en un tesoro negativo.

Esta indización es la base de la búsqueda en texto libre, que opera sobre unitérminos, como veremos más adelante, con los consiguientes resultados ruidosos. La simplicidad de las búsquedas se complica por el ruido que conllevan.

No hay que olvidar que para obtener resultados satisfactorios los programas informáticos documentales de tratamiento del texto íntegro son capaces de mejoras que consisten en el «... reforzamiento de las cualidades de los operadores de proximidad: adyacencia o distancia... y mejora de la eficacia de la indización automática por la introducción de algoritmos lingüísticos en los lógicos documentales» (4).

Mediante la búsqueda con adyacencias o con «links» se mejora notablemente la búsqueda en texto libre. Viene al caso citar textualmente a Bares cuando dice: «Es necesario que la extracción automática funcione: Con el *menor silencio* posible, ningún documento que esté relacionado con el tema debe ser olvidado. Con el *menor ruido* posible, debe aparecer el menor número de documentos que estén fuera del tema» (4).

En la *indización inteligente* el indizador somete el texto a una operación de conversión al lenguaje documental adoptado, en este sentido hablábamos antes de que se realiza involuntariamente una operación de sistematización en el intento de buscar el término adecuado. La tendencia y conveniencia de la indización manual es utilizar descriptores precoordinaados que proporcionan mayor especificidad frente a los unitérminos que conducen con facilidad a la falta de pertinencia y con frecuencia al excesivo «ruido».

Las recomendaciones de las normas ISO o UNE sobre construcción de

tesauros, y concretamente a propósito de la morfología de los descriptores, aconsejan la desagregación de ciertos términos y la poscoordinación en las búsquedas. El peligro de que las combinaciones booleanas de unitérminos conduzcan a resultados poco exitosos en la poscoordinación de los mismos es de sobra conocido. Afortunadamente algunos softwares de recuperación permiten un alto grado de pertinencia al introducir algún efecto corrector, de modo que todas las combinaciones posibles no son igualmente válidas. La posibilidad de búsqueda con «links» permite una certera combinación de unitérminos y sale al paso del tradicional ruido que se produce con las combinaciones indiscriminadas. Para ello los indizadores pueden establecer nexos entre los descriptores que formen parte del mismo subcampo, de forma que no haya más que un modo de combinación posible para una correcta recuperación.

La posibilidad de combinación indiscriminada de términos llevó a Austin a desarrollar un sistema de indización sintáctica, el Preserved Context Indexing System —(*PRECIS*)— con motivo de la elaboración del índice impreso de la British National Bibliography. Este sistema requiere una operación inteligente y es el ordenador el que se hace cargo de la rotación de descriptores y la creación de reenvíos. Según Austin los encabezamientos originales y las entradas suplementarias obtenidos automáticamente son «tan completos como el tema tal como es visto por el indizador», es decir, cada entrada presenta un «resumen completo del tema» del que trata el documento.

Este sistema de indización sintáctica opera mediante codificación rigurosa de las funciones gramaticales y sintácticas de los descriptores elegidos para designar las nociones del enunciado de un tema. Permite así evitar alteraciones de sentido y salvaguardar el contexto en el que aparecen estos conceptos. De este modo, los mismos principios que regulan el buen encadenamiento de las palabras de una frase en la lengua natural aseguran la coherencia sintáctica de los encabezamientos *PRECIS*, sistema que concede gran importancia a la acción, así como a los objetos y agentes de esa acción (2).

Nos permitimos traer a colación este sistema empleado en la elaboración de los encabezamientos de materias del índice de una Bibliografía Nacional por tratarse de un tipo de indización que evita el ruido mediante la aplicación de una lógica sintáctica, pilar sobre el que descansa la salvaguarda del sentido.

Se recomienda que a falta de normativa se creen los mecanismos propios y claves que eliminen la ambigüedad de la frase documental. Son pues, válidos los métodos de gramaticalización, como los indicadores de rol o los nexos. El método posicional o el topográfico son también adecuados para diferenciar las funciones de cada descriptor en la frase documental. De esta manera, mediante la lógica booleana y los operadores semánticos se dota al sistema de recuperación de un instrumento preciso que elimine la posibilidad de malentendidos (5).

3.2.4. *Nombres propios y geográficos*

Para evitar los silencios conviene normalizar en el proceso de análisis los nombres propios y geográficos que, excepto los nacionales, con frecuencia aparecen bajo formas diferentes. La falta de normativa nacional con respecto a los

nombres extranjeros nos hace asomarnos a la normativa del país vecino, Francia, donde este tipo de problema está solucionado en gran medida. Los nombres de personas y geográficos se afrancesan según su notoriedad en Francia. Serán dados en la forma francesa cuando ésta exista en el diccionario enciclopédico Larousse. En su defecto, se emplea la lengua original si pertenece a las lenguas románicas o germánicas (incluido el inglés), si no se utiliza el equivalente en lengua francesa.

Tampoco están normalizados los nombres geográficos, y de nuevo volvemos la vista a Francia donde los nombres de Estados y sus capitales están recogidos en una lista oficial dada en una orden conjunta del Ministerio de Asuntos Exteriores y de Educación en 1985 (6).

Se derivan también problemas de la utilización de siglas, que conviene desarrollar, excepto aquellas que se han impuesto con el uso como verdaderos sustantivos y son universalmente conocidas bajo esa forma.

La falta de normativa para los nombres propios de personas y geográficos obliga a la adopción de convenciones internas en el equipo de análisis, siguiendo alguna fuente de autoridad reconocida. Así pues, más allá del lenguaje controlado en los nombres comunes, se abre un terreno de problemas referentes a los nombres propios, bien sean estos geográficos, de personas o instituciones, sobre los que hay que ejercer un control, ya sea bajo fórmulas adoptadas oficialmente, ya sea bajo convenciones de carácter más doméstico.

3.2.5. *Indización automática - Tesoro negativo*

Frente a las operaciones de indización manual y de precoordinación inteligente, la indización automática trabaja con unitérminos. Todas las palabras que componen un texto, una a una, son objeto de indización excepto las expresamente contenidas en el tesoro negativo o de palabras vacías, que en una base de Política y de actualidad cobra características muy particulares.

La confección de un tesoro negativo en una base cuyo lenguaje es tan variado y al tiempo tan estandarizado, salpicado de fórmulas acuñadas, es una labor singular no coincidente con cualquier otro tesoro negativo.

Se procede en este caso a recoger grandes grupos de palabras que desde el punto de vista gramatical son vacías per se: preposiciones, conjunciones, pronombres personales, relativos, adjetivos y pronombres posesivos, demostrativos e indefinidos. También algunos adverbios utilizados con más frecuencia.

Las preposiciones son objeto de un análisis detenido. Aunque su papel es principalmente nexivo o relacionante no son totalmente vacías. Algún caso es objeto de excepción, como la preposición *contra* por el uso en el lenguaje periodístico para denominar el movimiento de oposición al régimen sandinista en Nicaragua. En este caso la preposición ha adquirido un valor sustantivo «la *contra*». Nunca se podrían encontrar documentos relativos a este asunto de no excluirla del listado de palabras vacías.

De igual forma el adverbio negativo no es excluido de esta lista por semejantes razones, ya que es susceptible de ser usado en el lenguaje periodístico, y en el común, tantas cuantas veces se quiera designar lo contrario de lo expresado por sustantivos y sobre todo adjetivos. Contamos con fórmulas acuñadas en el

lenguaje de política internacional como no-alineados, fórmula muy usual referente a países.

Por lo que respecta a otros grupos de palabras, se han incluido las formas conjugadas de los verbos auxiliares haber, ser y estar, de tan alto uso en cualquier discurso, con alguna excepción. La tercera persona del futuro del verbo ser, será, se ha eliminado por la confusión a que podría dar pie si por error no fuera acentuada ortográficamente y se confundiera con el nombre del importante periódico italiano *Corriere della Sera*.

Por último, se respetan los sustantivos por su especial capacidad significativa, sobre todo adjetivados con calificativos. Sobre estas formas gramaticales descansa lo sustantivo, valga la redundancia, desde el punto de vista de la significación. A pesar de esta rotunda afirmación, se ha optado por incluir en este tesoro negativo aquellos que por su alta frecuencia y su absoluta vaciedad no desempeñan ningún papel significativo, antes bien ocupan demasiado espacio en los índices. En este grupo de sustantivos se encontraría «caso», sustantivo muy usado, que queda excluido por el uso que se hace para denominar algunos asuntos de transcendencia política. De la misma forma, nos consta que otros sustantivos son inoperantes por su falta de significado, pero dada su baja aparición en el discurso cotidiano de la prensa de actualidad no viene al caso suprimirlos.

Todas estas consideraciones y la propia limitación impuesta por el sistema de recuperación condicionan el tesoro negativo resultante, producto especial dada la idiosincrasia de este lenguaje.

4. Conclusiones/recomendaciones

Los estudios comparativos de los dos procedimientos de indización muestran que la *indización inteligente* utiliza un *vocabulario más restringido* además de una tendencia sistemática a la utilización de *macrotérminos*.

La *indización inteligente* tiende a la búsqueda de descriptores genéricos, si no existe el término apropiado, lo que hace que esta operación resulte teñida de una nota de generalización. Las operaciones de síntesis de que hablábamos antes inducen a los indizadores a introducir ciertos términos generales de forma sistemática, de suerte que pudiera probarse, por tan alta frecuencia de estos términos, que existe un hábito de aplicar el término general aunque se disponga de un término con mayor grado de especificidad.

Además la indización inteligente se ve limitada por el uso del tesoro, que inevitablemente va por detrás de las necesidades terminológicas, lo que obliga a la utilización de un vocabulario más restringido. Ese desfase del vocabulario controlado con respecto al léxico periodístico exige puestas al día en función de las necesidades experimentadas, sobre todo en ámbitos que evolucionan muy rápidamente.

Por el contrario, hay que resaltar que la especificidad que proporciona la *indización automática* va acompañada, en gran parte de los casos, de la ambigüedad inevitable cuando se descontextualizan los términos. La polisemia de ciertos términos, sin embargo, es neutralizada en la indización inteligente al contar con una herramienta estructurada jerárquica y semánticamente —el tesoro—.

La *indización automática*, operando sin estas interferencias inteligentes, refleja de forma clara «la utilización de una mayor *variedad de términos* que traducen simultáneamente un efecto de *dispersión y especificidad* en el vocabulario».

Estas observaciones sobre la indización manual y la automática aconsejan no enfrentarlas como contrapuestas sino considerarlas como complementarias.

La posibilidad de buscar en texto libre en los campos título y resumen, es decir, la facilidad que proporciona la indización automática del título —que en la mayoría de los casos tiene un gran valor informativo— y de los resúmenes o cabeceras, de indiscutible valor, nos permite hacer una recomendación para evitar lo que resultaría una *sobreindización*. Dicho de otra manera, evitar en la indización inteligente la utilización de los mismos términos que los utilizados por el texto de la noticia, inevitable contaminación a la que el indizador le resulta difícil sustraerse. Es esta contaminación la que permite, por otra parte, una continua puesta al día del lenguaje controlado, que se ve invadido por la aparición de nuevas palabras relativas a esa realidad viva y cambiante.

En la perspectiva del efecto de complementariedad hay que recordar que el grado de especificidad en la indización automática es muy alto, ya que todas las palabras significativas e informativas de títulos y resúmenes pueden constituir palabras de acceso al documento. Las ventajas de estos términos en las búsquedas reside en su valor informativo intrínseco, se trata a menudo de términos muy específicos y recientes que permiten seguir la evolución terminológica de una disciplina. La desventaja se ve paliada por la labor complementaria del vocabulario controlado.

Pierre Albert piensa que la renuncia a la indización inteligente y la adopción de la automática puede acarrear consecuencias perversas, pues es un procedimiento que exige conjuntos homogéneos de texto y no es el caso de los artículos de prensa... En su opinión, si se utiliza el texto completo e indización automática hay riesgo de ser contestado con una avalancha desalentadora de respuestas y se podrá también obtener silencios debido a homonimias y a grafías defectuosas. Según Albert «la generalización del texto completo supone una indización previa o incluso un resumen de los contenidos de los documentos. Piensa que cuanto más aumente el volumen de los textos almacenados más se necesita una fina indización. A medio plazo la coexistencia de sistemas... le parece la mejor solución» (8).

En esa complementariedad hay que destacar una nota propia de la indización inteligente: permite un fácil y gradual acceso a la información. Frente a esa rigurosa especificidad, tan de apreciar en búsquedas muy concretas, las operaciones inteligentes segregan estadios intermedios de descripción que permiten organizar la información en niveles intermedios. Esa tendencia a la generalización, al empleo de macrotérminos, que recuerda un plan de clasificación, proporciona una progresividad en el acceso muy apreciable en las operaciones de recuperación. De suerte que nos permitiríamos concluir que la variabilidad que presenta la indización inteligente —muy difícil que coincidan al cien por cien todos los indizadores— junto a la tendencia a ampliar el nivel de descripción y a generalizar, en un intento de sistematización y clasificación, constituyen un marco perfecto de encuadramiento de esa otra descripción término a término, fruto de la indización automática, tan susceptible de ruido a pesar de los recursos de los programas informáticos documentales.

En una base de información política y de actualidad, donde el lenguaje constituye un elemento tan a tener en cuenta y la cobertura de su campo temático es tan universal, la complementariedad de ambos procedimientos de análisis se justifica especialmente para asir esos contenidos de la realidad cambiante, única forma de controlarla y poder recuperarla.

Este trabajo corresponde a una ponencia anunciada en el programa de las Terceras Jornadas Españolas de Documentación Automatizada, que no llegó a presentarse.

Bibliografía

1. HERNANDO, Bernardino M. *El Lenguaje de la Prensa*. Madrid, Eudema, 1990.
2. NEET, Hanna E. *A la recherche de la mot clé: Analyse documentaire et Indexation alphabétique*. Genève, Les Editions IES, 1989.
3. ALVARO, C.; VILLAGRA, A.; SORLI, A. Desarrollo de lenguajes documentales. *Revista Española de Documentación Científica*. (Madrid) 12 (2-3), 1989.
4. BARES, Michel. *La búsqueda documental en el contexto telemático y utilización de bases de datos*. Madrid. Díaz de Santos; 1989.
5. GARCIA GUTIERREZ, A.; LUCAS FERNANDEZ, R. *Documentación automatizada en los medios informativos*. Madrid. Paraninfo; 1987.
6. RICHTER, Noë. *Les langages documentaires encyclopédiques: Guide pratique d'Indexation*. Marigné. Editions de la Queu du Chat; 1990.
7. CHARTRONG, G.; DALBIN, S.; MONTEIL, M.G.; VERILLON, M. *Documentaliste*. (Paris) 26 (4-5), 181-187, 1989.
8. ALBERT, P.; LETEINTURIER, C. Presse et banques de données: les enjeux. *Bulletin des Bibliothèques de France*. (Paris) 30 (2), 118-123, 1985.