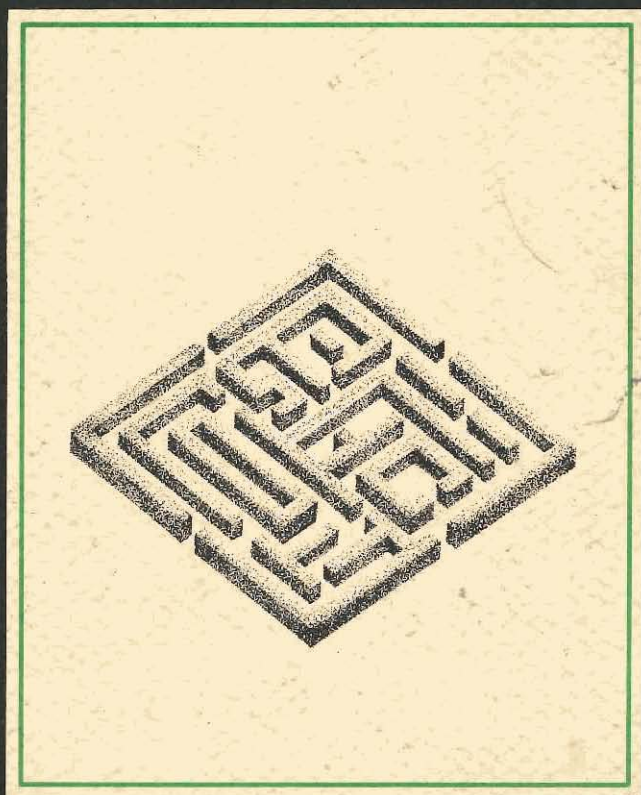


ANALISIS DOCUMENTAL DEL DISCURSO PERIODISTICO

ANTONIO GARCIA GUTIERREZ



025
GAR
9

R. 13 931

1000
1903

ANTONIO GARCÍA GUTIÉRREZ

ANALISIS DOCUMENTAL DEL DISCURSO PERIODISTICO

490013



Madrid 1992

GARCIA GUTIERREZ, Antonio
Análisis documental del discurso periodístico /Antonio García Gutiérrez.
Madrid: CTD, S. A. - 1992.

160 p.

Bibliografía

Análisis documental. Periodismo. Análisis del Discurso

Diseño de cubierta: Teresa del Valle

Copyright: Antonio García Gutiérrez

Edita: CTD - Centro de Tratamiento de la Documentación, S. A.
San Romualdo, 26 - 28037 Madrid

Imprime: Gráficas Iglesias
San Romualdo, 26 - 28037 Madrid

Depósito Legal: M-5.107-1992.

ISBN: 84-604-1875-8

Sin duda, porque no se ha sometido la connotación a una tipología de los textos, no tiene buena prensa.

(R. Barthes: S/C. p.4)

TABLA DE MATERIAS

INTRODUCCION	9
0. OBJETO, METODO Y CORPUS	15
0.1 PRINCIPIOS E HIPÓTESIS DE INVESTIGACION	15
0.2 CORPUS TEÓRICO Y CORPUS DE APLICACION	17
0.3 METODOLOGÍA GENERAL DE INVESTIGACIÓN	18
1. CUADRO TEORICO: OBJETO Y METODO DEL ANALISIS DOCUMENTAL (AD)	21
1.1 MÉTODO Y NORMA DE ANÁLISIS	24
1.2 DELIMITACION TEÓRICA: EL PARAGIGMA DEL PROCESO DOCUMENTAL	26
1.2.1 ANÁLISIS	28
1.2.2 SÍNTESIS	29
1.2.3 REPRESENTACION	29
1.2.4 RECUPERACION	30
1.3 AD: STATUS QUAESTIONIS. PROBLEMAS GENERALES	31
1.4 CRISIS DE LA INVESTIGACIÓN ACTUAL EN AD	34
1.5 CONDICIONES PREVIAS A LA FORMULACION DE UN MÉTODO ANALÍTICO	37
1.6 USUARIO MODELO	42
1.7 APLICACIONES DEL AD: LAS RESTRICCIONES DISCURSIVAS	44
2. CUADRO REFERENCIAL TECNOLOGICO	47
2.1 INTELIGENCIAS ARTIFICIALES (IA)	48
2.1.1 EL ACCESO AL SIGNIFICADO EN IA	51

2.1.2	PRIMITIVAS, GUIONES Y ESCENARIOS	53
2.2	SISTEMAS EXPERTOS	55
3.	EL AD DEL DISCURSO PERIODISTICO: CONSIDERACIONES PREVIAS	67
3.1	EL OBJETO PRAGMÁTICO DE OBSERVACION (ACOTACIONES)	71
3.1.1	REFLEXIONES EN TORNO A LA EXPRESION Y AL CONCEPTO DE INFORMACION DE ACTUALIDAD	72
3.1.2	LA UNIDAD DISCRETA DE TRABAJO: ELEMENTOS PARA SU FORMULACION EN INFORMACIÓN DE ACTUALIDAD	76
3.1.2.1	UNIDAD DE CONTENIDO: DESCRIPCION Y EVALUACIÓN	79
3.1.2.2	RECHAZO DE ALGUNAS CONJETURAS	85
3.1.2.3	ALTERNATIVAS Y TIPOLOGIA DE UNIDADES POSIBLES: LA UNIDAD DE INFORMACION	87
3.1.3	ESTRUCTURA DE LA INFORMACION DE ACTUALIDAD	96
3.1.4	APROXIMACIONES DESDE EL ANÁLISIS DEL DISCURSO	98
3.1.5	GRAMÁTICA Y SEMÁNTICA DE LA NOTICIA	103
3.1.5.1	CONTENIDO DE LA NOTICIA	103
3.1.5.2	TÍTULOS Y MACROESTRUCTURAS	105
3.1.5.3	EL CORPUS DE LA NOTICIA	106
3.1.5.4	ESQUEMAS FORMALES	108
3.1.5.5	PROPOSICIONES	109
3.1.6	PALABRA, ENUNCIADO Y TEXTO: RELEVANCIA DOCUMENTAL	111
3.1.7	TIPOLOGÍA DE OMISIONES EN EL DISCURSO PERIODÍSTICO	113
3.2	LOS AGENTES DEL AD EN EL DISCURSO PERIODÍSTICO	117
3.3	OBJETIVOS ESPECÍFICOS DEL AD EN INFORMACIÓN DE ACTUALIDAD	120
3.3.1	NUEVOS PRODUCTOS PARA NUEVOS MÉTODOS	121
3.3.2	OBJETIVOS DEL ANÁLISIS: TIPOS DE SEGMENTOS RELEVANTES	123
4.	EVALUACION DE METODOS DE ANALISIS: CONDICIONES GENERALES	127
4.1	MÉTODOS DE ANÁLISIS BASADOS EN LA REPRESENTACION FORMAL	130
4.1.1	INDIZACIÓN POR UNITÉRMINOS Y DESCRIPTORES EN LENGUAJE LIBRE	131
4.1.2	CÁLCULO FRECUENCIAL AUTOMÁTICO	133
4.1.3	INDIZACIÓN MEDIANTE LÉXICOS	134

4.1.3.1	TÉRMINOS STOP	134
4.1.3.2	ANTIDICCIONARIOS	134
4.1.3.3	LENGUAJES COMBINATORIOS	135
4.1.3.4	PONDERACIÓN CONCEPTUAL	136
4.2	MÉTODOS DE ANÁLISIS BASADOS EN LA SIGNIFICACIÓN	136
4.2.1	EL ANÁLISIS LOGICISTA (GARDIN)	137
4.2.2	EL MODELO DE LASSWELL	140
4.2.3	EL MODELO DE CASOS (FILLMORE/POTTIER)	140
4.2.4	EL MODELO FACETADO: EL LENGUAJE DOCUMENTAL COMO SISTEMA EXPERTO	144
4.2.5	EL ANÁLISIS DEL DISCURSO (VAN DIJK)	145
4.2.6	MÉTODO CUANTITATIVO-SEMÁNTICO (GUIRAUD)	148
4.2.7	MÉTODO DE LAS MODALIDADES (GREIMAS)	149
4.3	OTRAS APORTACIONES METODOLÓGICAS	151
4.3.1	CONECTIVAS LÓGICAS	151
4.3.2	ANÁLISIS SINTÁCTICO	153
4.3.3	OTROS MÉTODOS	154
5.	CONCLUSIONES	155
	BIBLIOGRAFIA	157

INTRODUCCION

Este estudio persigue, a través de la reflexión crítica, la detección de las insuficiencias del Analisis Documental (AD) efectuado sobre contenidos y, en particular, sobre el discurso periodístico. Se construye un modelo para la investigación teórica, conceptual y tecnológica del AD, a partir de otras disciplinas, aprovechando los objetivos pragmáticos de la Documentación, con el fin de realizar propuestas para la simulación de sistemas de información inteligentes que escapen al control de los métodos de análisis impuestos por los equipos tecnológicos suministrados por las empresas comerciales. La aproximación desconfiada del documentalista o, en su defecto, de cualquier programa inteligente de análisis automático que lo emule, debe primar sobre el trabajo "inocuo" de los sistemas de computadoras actuales.

Trás la reflexión genérica, se realiza una revisión de métodos de análisis describiendo y evaluando sus resultados en lo que concierne a un determinado discurso social proponiéndose, finalmente, un punto de partida para la simulación de un método integrador en una tecnología inteligente. En este capítulo introductorio se recogen las principales hipótesis de trabajo, algunas formadas mediante aproximación especulativa, otras mediante observación empírica. Los resultados de la demostración de ellas han de clarificar uno de los núcleos más confusos de la Teoría de la Documentación.

El mito de la objetividad del análisis efectuado por documentalistas y bibliotecarios, que trabajan en el ámbito de las disciplinas sociales, perjudica la elaboración de un método adecuado de interpretación dentro del marco de segmentación del AD. Este mito es explotado por fabricantes y distribuidores de paquetes de indización automática con el fin de colocar sus productos en los procesos y momentos fundamentales de la trasmisión de conocimiento y de cultura con dos resultados: Por una parte, la justificación de la renovación de unos softwares con alto grado de obsolescencia y bajas prestaciones debido a

los errores de fundamentación desde su concepción y, por otra, a la obtención de una estabilidad favorable al *statu quo* imperante dada la inactividad de los transmisores de información documental. Por ello, no podemos separar la investigación documentológica de los patrones socio-económicos imperantes que afectan, de igual modo, a la investigación de la comunicación desde la perspectiva de industria cultural.

Trás casi un siglo de la implantación de la moderna concepción de la Documentación, a raíz de los esfuerzos realizados dentro del más puro positivismo por Paul Otlet desde su Instituto Internacional de Bibliografía, y después de más de veinticinco años de existencia de paquetes automáticos para la lectura de textos con fines recuperacionales, no se ha producido una reforma, ni siquiera leve, de los modos de captación del documento a pesar del potente desarrollo de otras técnicas aledañas facilitadas por la tecnología de la información o de la profundización adquirida por el análisis de contenido en otras aplicaciones. A su vez, las publicaciones y Congresos en los que se produce la prometedora interacción Documentación-lenguaje-tecnología refuerzan el papel tecnológico y la representación en detrimento de la modelización de los procesos cognitivos de interés documental.

Parece obvio que, sin un estudio riguroso de los objetivos del documentalista en los procesos de información, de las condiciones de producción de los documentos leídos o de las posibles interrogaciones a que un sistema puede ser sometido por el usuario (U), es inviable establecer los parámetros de un método del AD humano que supere el mito de la objetividad y alerte, por el contrario, a U de las posibles omisiones y manipulaciones del original.

Es, también, evidente que la ausencia del método de captación aludido impide cualquier propuesta de extrapolación a un ordenador clásico o experto. Entonces ¿Cómo es posible que la industria tecnológica esté produciendo sofisticados instrumentos de recuperación de información sin que haya constancia de una propuesta metodológica surgida de la reflexión teórica de los documentólogos u otros investigadores sociales? En mi opinión, la industria informática y su aplastante vinculación a la denominada industria cultural no necesita esos métodos de AD elaborados por los teóricos de la Documentación porque, en el mejor de los casos, no son rentables para el vigente sistema de producción y consumo o, en una visión pesimista, el desconocimiento coadyuva a la expansión de las mercancías y a su rápida obsolescencia.

Los presupuestos millonarios que destinan las multinacionales a la investigación puramente tecnológica en busca, por ejemplo, de la compatibili-

dad de sistemas operativos e, incluso, de equipos, como el conocido acuerdo IBM-Apple para acercar sus productos, lo que representa una encubierta concentración de mercado, crean la falacia, entre los consumidores, de recibir una ayuda desinteresada de estas empresas en favor de la conexión y de la comunicación de diferentes sistemas informáticos cuando en realidad es el reflejo de una política empresarial para la homogeneización de productos. Es cierto que la compatibilidad en equipos y programas beneficia al usuario hasta un punto pero, es más cierto aún, que los principales beneficiarios son las propias multinacionales que suscriben los pactos de compatibilidad. Podemos vincularnos con esta reflexión y aceptar, en términos generales, los postulados de la denominada Teoría de la dependencia (centro-periferia) desarrollada por Schiller o Varis, entre otros.

Una mínima intervención de los usuarios de los softwares de indización automática en los procesos de producción de los mismos redefiniendo los objetivos o provocando la derivación de una parte de los presupuestos de investigación de las empresas hacia la mejora cualitativa al servicio del conocimiento, y no viceversa, supondría un avance considerable en los métodos de análisis por ordenador. Pero, este deseo, lejos de ser realidad contribuye a una mayor pasividad por parte de bibliotecarios y documentalistas, cada día más abrumados y conscientes de la responsabilidad del manejo de cantidades y no de calidades de datos. El usuario final es la víctima real de esta situación. Salvo en los casos de investigadores o personas de alta capacidad crítica o de reacción, la mayoría de los usuarios de un sistema documental, como nuevo refuerzo de los sistemas de generación de opiniones y hábitos en una sociedad determinada, consumen los datos obtenidos gracias a la "buena voluntad" del ordenador y de los documentalistas que lo nutren. Este estado de cosas es ya un hecho en varios países entre los que destacamos, por su cercanía y afinidad, Francia y su plan telemático nacional de información al hogar. La documentación no sólo debe formular métodos sino también describir procesos de deterioro cultural en los que intervengan modos y medios documentales con el fin de evitar profundas sustituciones de los hábitos y de las culturas locales en función de lo establecido en el paradigma de la dependencia.

El panorama investigador y, por ende, bibliográfico sobre este aspecto en documentación es desolador. Con la excepción de unos pocos libros y esporádicos artículos aparecidos en fuentes dispersas, además de irrecuperables tesis y otros trabajos inéditos, la documentación obtenida sobre métodos alternativos de análisis de documentos u otras ópticas que consideren este campo es esencialmente de tipo evaluativo y basada en criterios de rentabilidad de los datos, satisfacción del usuario y relación de las tasas de precisión y exhaustividad,

posibles mejoras en los módulos de recuperación de determinados softwares por encargo de las empresas, o comparación entre ellos para facilitar su venta. El problema radica en que este campo no tiene soluciones mecánicas, visión aséptica de la Documentación que se trata de implantar a nivel informático, puesto que hablamos de instrumentos que contribuyen a la mediación en la cultura y en la conciencia.

Estas constataciones parecen suficientes para proponer, como prioridad de investigación, la búsqueda de un método de AD factible, desde el punto de vista pragmático, en la labor cotidiana del documentalista el cual, a través de simplificaciones y codificaciones, debe ser transportable a un sistema de análisis automático con la fiabilidad de que el ordenador es capaz de captar las omisiones significativas, las manipulaciones o, al menos en principio, lo relevante de un texto, objetivos que por el momento están lejos de ser alcanzados. El paradigma propuesto o el aprovechamiento de los métodos considerados habrían de tomar parte en la formación de documentalistas.

La complejidad del estudio global de un método de AD es evidente. Por ello, se aborda aquí su primera fase, el análisis o lectura, considerando la presencia y la influencia constantes de las tres fases restantes. Con ello, en resumen, se pretende lo siguiente: crear las condiciones para proponer o investigar sobre métodos de análisis cuya aplicación sea factible en un sistema inteligente, auxiliado por el documentalista en la primera fase, con unos resultados verificables.

No obstante y a pesar de la especificidad del objeto que voy a tratar, mantengo en un sustrato superior un objetivo más ambicioso: poder pensar la Documentación desde perspectivas distintas a las operativas y tecnológicas predominantes, replantear el rol de los agentes que la ejecutan, investigan o reciben, considerar los procesos documentales como fuentes generadoras de nuevas formas de entender los procesos culturales.

En cualquier caso, la contribución de este trabajo al desarrollo de la Teoría del Análisis documental se orienta más a la formulación de problemas que a la corrección de procedimientos inadecuados. Y ello se debe al vacío existente en torno a un corpus teórico necesario, a la falta de un modelo general de referencia y a la necesidad de estudiar con urgencia los parámetros para construir una teoría del uso de la información que modelice los conceptos de necesidad, satisfacción y utilidad/utilización de tal forma que podamos abordar nuevos métodos de lectura y representación en función de las tipologías de la demanda.

A pesar de la contundencia de las últimas afirmaciones, no quiero ocultar una paradoja que planea sobre el presente texto: si debemos pensar los procesos culturales masivos exclusivamente en marcos tecnológicos y prescindimos de la intervención humana en las fases de lectura y representación, tenemos que plantear la investigación en dos posibles opciones: documentación artesanal o documentación industrial. Si negamos la primera vía, como ocurre en este libro ¿No estaremos contribuyendo al refuerzo de la rechazada?

Madrid, Diciembre de 1991.

CAPITULO 0

01. PRINCIPIOS E HIPOTESIS GENERALES

Se proponen, a continuación, los principios e hipótesis generales, que determinan los elementos y métodos utilizados en el texto:

- La investigación del AD aplicado debe realizarse dentro de un paradigma teórico que globaliza e interrelaciona cada uno de sus componentes. La investigación sobre uno o varios elementos del AD debe considerar la presencia simultánea de todo el paradigma.
- La investigación sobre AD debe realizarse interdisciplinariamente mediante aproximaciones que toman por objeto el contenido documental.
- En Documentación y, en consecuencia, en AD, la vía experimental es la más útil en investigación teórica y metodológica. El AD debe considerar, además, una aplicación determinada en la que puedan validarse empíricamente los resultados. No es posible la universalidad de métodos de AD para todo discurso. Las extrapolaciones de reglas y elementos deben realizarse en casos determinados y afines. La validación sólo es posible mediante muestras de unidades discretas. Como ocurre en otras disciplinas similares, la validación en AD no solamente es un método de evaluación sino también objeto de estudio en sí misma.
- La estructura y tipología particulares de un discurso condicionan, en el estado actual, las investigaciones y los resultados sobre métodos de

AD. En el discurso periodístico esto es válido, además, para géneros, extensión y fragmentación considerados como variables. Es necesario inventariar todas las variables discursivas posibles que afecten a la generalización.

- No son posibles las propuestas de métodos de AD fuera de marcos tecnológicos hipotéticos, potenciales y reales. Por lo tanto, la investigación dirigida a procesos documentales artesanales es poco práctica puesto que la explicitación del raciocinio humano (intuición, sentido común y otras variables) no es simulable por máquina usando los mismos criterios y procedimientos.
- Para poder simular el AD, o algunos de sus componentes, es preciso, primeramente, objetivar y modelizar los procesos de lectura pragmática realizados por el ser humano puesto que se pretenden los mismos resultados mediante máquina (a través de un diferente procedimiento de acuerdo con la hipótesis anterior). Un método integrado de interrogación textual mediante analizadores es el paso previo a la elaboración de los árboles de decisión.
- Las unidades discretas de las muestras deben basarse en categorías temáticas, a partir de un inventario de éstas, aunque deben establecerse otros criterios para la construcción de tales unidades.
- Las propuestas de métodos de AD desconectados de lenguajes documentales sólo funcionan en laboratorio. Cualquier simulación de AD debe ser compatible y necesita un sistema de referencia de macrodescripciones y macrosituaciones además de un cuadro de demandas y modos de demandas. Para lo primero es preciso definir los parámetros extensionales e intensionales así como la relación entre elementos. Para lo segundo es preciso construir la figura del usuario modelo, por vía empírica (fusión de perfiles de usuarios).
- La representación, tanto en el nivel teórico como práctico, debe ser determinada por los resultados del análisis y no viceversa (hipótesis contraria a la imposición de formas previas de análisis).
- Es necesaria la descripción de nuevos productos como resultados de los nuevos métodos aplicados en las tecnologías inteligentes aunque éstas deben conservar la capacidad de generar productos tradicionales.

- Es necesario definir la estructura del texto para hacer posible la investigación de modelos de análisis aplicados. Es necesario formular la estructura repetible y mínima de todos los textos en los que se cumplen las condiciones analíticas.

02. CORPUS TEORICO Y CORPUS DE APLICACION

La presencia de un corpus de referencia teórica explicitado es necesaria para confirmar la hipótesis sobre la importancia de la intervención de otras disciplinas en la teoría del análisis documental y sustraer de éstas los instrumentos conceptuales y metodológicos válidos para nuestros objetivos. En efecto, el AD no ha descubierto el texto como objeto, ni tan siquiera el documento. Podemos atribuir al AD la innovación de adscribir su indagación científica en unos objetivos que parecen serle exclusivos aunque para obtenerlos debe compartir un amplio campo común con otras disciplinas.

El análisis semántico, en el más amplio sentido dado por Gardin, comprende el Análisis del discurso y de contenidos, el análisis gramatical, el análisis de casos, etc., procedimientos correspondientes a disciplinas como Semiótica, Sintaxis y Sociología, entre otras, y Análisis documental, sin duda, para el que debe realizarse un preciso recorte teórico-conceptual. El criterio para tal recorte nos viene marcado, nuevamente, por los objetivos específicos: recuperación del conocimiento a través de descripciones y representaciones textuales. En consecuencia, partimos de algunas experiencias de modelización con otros objetivos para avanzar en el establecimiento de un método integrador para el AD.

Los objetivos documentológicos son, como se ve, de orden pragmático. El pragmatismo impone, para confirmar las hipótesis relativas a las aplicaciones, validar los resultados y los elementos de los métodos en corpus organizados o delimitados en discursos específicos. En este trabajo, se ha optado por el discurso periodístico, con las restricciones señaladas en la tercera parte, como campo de aplicación del AD con los evidentes problemas que comporta: enciclopedismo y dispersión, entre otras características particulares profundizadas aquí desde la óptica estructural.

No podemos pretender, en esta fase de las investigaciones, dar respuesta total a todos los elementos del paradigma general bajo el que queremos situarnos, conscientes de la necesidad de un abordaje

epistemológico global y de la confirmación pendiente de no pocas hipótesis de trabajo. Deseamos comprobar, en un nivel teórico intermedio, una serie de postulados e intuiciones que ayuden al AD a salir del estancamiento en el que se halla sumido debido, fundamentalmente, a la perspectiva normativista con la que, en general, es tratado bibliográficamente, salvo contada excepciones, entre otras el grupo de Gardin en París con unas aplicaciones muy concretas y el grupo de Smit en Sao Paulo, con preocupaciones más amplias y cercanas a nuestro interés.

El objetivo inmediato de este trabajo no es, por las razones anteriores, menos ambicioso: establecer un aparato conceptual, las condiciones teóricas, prácticas y tecnológicas y la evaluación de algunos modelos de análisis. Por lo tanto, la puerta queda abierta, y ya proponemos nuevos trabajos para aplicar las consecuencias de este estudio en el que observamos, a priori, los siguientes problemas pendientes (todos ellos sobre la misma aplicación):

1. Desarrollo simulado de una base de reglas sobre una muestra de discurso periodístico, mediante la propuesta de analizadores inteligentes que emulen el proceso humano de lectura empírica.
2. Condiciones para la construcción de síntesis mediante la aplicación de generadores inteligentes.
3. Compatibilidad de los métodos de análisis con las estructuras de representación construidas como universo de referencia.
4. Compatibilidad con sistemas sintácticos naturales de recuperación inteligente.
5. Análisis del comportamiento de los conceptos en el enunciado para permitir el reconocimiento artificial de categorías funcionales y la atribución de roles.

03. METODOLOGIA DE INVESTIGACION

A veces hablamos de una teoría del AD aunque en realidad nos referimos a un corpus teórico en construcción todavía lejos de su consolidación. Como he expuesto anteriormente, la investigación documentológica está sometida a las aplicaciones y, en consecuencia, se declara radicalmente empírica siendo

necesario introducir criterios de validación en cada caso. La construcción de una teoría general del AD, por tanto, tendría un desarrollo lento y presentaría los primeros visos de maduración cuando hayamos conseguido explicar mediante leyes y reglas la sistemática general y reiterada de los procedimientos documentales en función de todo discurso posible. Entre tanto, nos encontramos con la ardua tarea de describir las estructuras de los discursos objeto y de como influyen éstas en la metodología del AD como paso previo a la modelización y simulación automática.

En principio necesitamos un corpus conceptual y teórico para dar cuenta de cómo se organiza el conocimiento en el discurso periodístico y cómo es representado mediante la lectura efectuada por el documentalista de prensa dentro de sus objetivos cotidianos. En ese complejo tránsito influyen, no sólo los bagajes ideológico-culturales de lector y productor del texto, sino también la institución (en sentido saussureano) que impone, desde la ausencia, estrictos códigos de lectura, la propia naturaleza y objetivos del discurso en cuestión y los usuarios modelo que dominan el imaginario de los documentalistas. Para ello se aborda estructuralmente cómo se organiza la noticia periodística, materia prima de este trabajo, apoyándonos en otras disciplinas, esencialmente, en el Análisis del Discurso.

Ya que se trata de modelizar un proceso analítico para evitar rutinas en la *gestión de las noticias, los resultados y consecuencias finales y parciales* del estudio se contrastan con un marco referencial tecnológico, concretamente los sistemas expertos, en el pueden simularse los procedimientos aunque con reducciones importantes en la actualidad ya que la mayoría de los consultados se limitan a la inferencia lógica, recurso útil pero no único en un sistema documental.

La simulación inteligente necesita ciertos requisitos como: formular el cuadro empírico de categorías de macrodescripción y demanda en las que opera el método y proponer productos de salida del sistema. Téngase en cuenta que mis objetivos van más allá de una aplicación en un sistema concreto. Pretendo realizar propuestas que sean, posteriormente, integradas y parametrizables (esto es, descritas y sistematizadas) por los ingenieros del conocimiento en un programa experto para el que este estudio es una contribución en el plano operativo que debe ser completada con los otros planos del paradigma sobre el que trabajamos.

Al procurar la máxima generalización dentro de las condiciones discursivas imperantes he abordado diversos modelos de análisis provistos por otras

disciplinas y con otros objetivos. Sin embargo, se han utilizado idénticos criterios para la evaluación en orden de homogeneizar el dictámen y los resultados aplicables en el AD. La validación hubiera sido deseable para testar cada método pero también inviable al no disponer de instrumentos, tanto de procedimiento como mecánicos, para llevarla a cabo. En cualquier caso, mi objetivo es extraer consecuencias del análisis comparativo.

He insistido, a lo largo de muchas páginas, en las condiciones para la construcción de una muestra como único elemento garante de la validación empírica. En este sentido y consciente de la complejidad de la aplicación que me ocupa, discrepo de mis colegas de Sao Paulo, a pesar de estar inmerso en la misma línea de trabajo, que utilizan la vía de ejemplarización liberando en exceso, pienso, la validación de unas propuestas que arriesgan ser difícilmente extrapolables o reducirse a meros análisis individualizados de textos negando el carácter social o colectivo del AD y provocando un alejamiento del enfoque pragmático de nuestra disciplina.

Por último, para la estructura general de este estudio he realizado recortes instrumentales en distintas disciplinas que podríamos agrupar en torno a las llamadas Ciencias cognitivas. Así, me he servido del Análisis del Discurso para reconocer la estructura de la información de actualidad con la ayuda, además, de la Semántica estructural y de la Semiótica del texto. Determinados conceptos y procedimientos tomados tímidamente de la Psicología cognitiva en su entroncamiento con la Inteligencia Artificial y de la Lógica formal, en especial aplicada a enunciados, me ayudaron a entender las posibilidades de la alta tecnología respecto al análisis de textos. La Morfosintaxis, la Gramática de casos, el Análisis de contenido de tendencia sociológica y la Teoría general de la Comunicación me facilitaron la comprensión de las estructuras de los conceptos y mensajes y la construcción de modelos parciales de interrogación textual como paso inicial hacia la elaboración de un árbol de reglas de decisión para la lectura automática del texto. Finalmente, la Documentología me aportó un escaso aunque esencial bagaje teórico para encuadrar este estudio dentro de unos objetivos bien delimitados y la Documentación, un universo de problemas prácticos.

CAPITULO I

1. CUADRO TEORICO: OBJETO Y METODO DEL ANALISIS DOCUMENTAL (AD)

No pretendo, en este trabajo, revisar las polémicas sobre el espectro epistemológico del AD o la idoneidad de la expresión. Me remito a la bibliografía de la última década que trata de situarlo como operación dentro del proceso documental para dejar zanjada la cuestión. Emplearé decididamente la denominación AD para abarcar exclusivamente el conjunto de procedimientos que utilizan tanto el documentalista en la captación de las proposiciones del texto, su segmentación y representación formal como el usuario en el proceso recuperador, interrogador e interactivo con la memoria documental. Todos y cada uno de los elementos citados merecen un mayor detenimiento dentro del marco teórico que proponemos. Excluyo, sin embargo, las operaciones formales vinculadas a las normas de descripción bibliográfica del tipo ISBD o AACR dentro de esta denominación: tales operaciones comienzan afortunadamente a ser consideradas independientes, a pesar de ser descriptivas, del AD en algunos planes de estudios documentales en España.

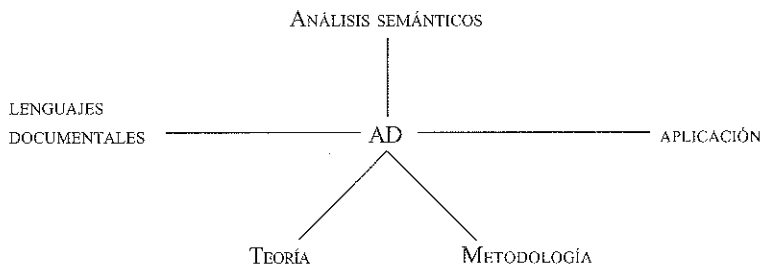
El AD opera sobre conocimiento: es un objeto cuya fenomenología podemos situar, con Gardin, en el campo de la Epistemología práctica¹: disciplina que tiene por objeto el estudio del conocimiento científico, su delimitación, articulaciones y construcción en áreas específicas. El AD se dirige

¹ Denominación propuesta por Gardin para designar una disciplina cuyo fin es clarificar los fundamentos conceptuales de las construcciones en Ciencias humanas tal como se presentan en la práctica, tarea que emprende sobre corpus textuales específicos con sus colaboradores en : Gardin, J.C. et al.: *La logique du plausible: essais d'épistémologie pratique en Sciences humaines*. Paris: Maison des Sciences de l'Homme, 1987, p. 27-100.

estrictamente hacia los contenidos de los documentos o, por extensión, hacia el discurso fragmentado en ellos con el fin de optimizar el proceso de recuperación de información. Se destacan, de ahí, dos principios: el pragmatismo como elemento que impregna cualquier propuesta y la adscripción a los procesos comunicativos o, por extensión, sociales.

El AD, por otra parte, participa del universo teórico y metodológico de otros métodos: análisis semiótico, análisis sintáctico, análisis de contenidos y del discurso, análisis estadístico y de otras disciplinas: Semántica, Lógica formal, Teoría de la Comunicación, Terminología y Normalización, cuyas relaciones ya he descrito en otro lugar². La vinculación del AD con la denominada Lingüística documental (LD) es obvia: es uno de sus principales objetos de observación y descripción y sobre el que la LD debe realizar sus mayores esfuerzos en lo que concierne a propuestas metodológicas. Como ha demostrado Cunha, a pesar de no utilizar la denominación citada, es en la Lingüística donde el análisis documental realiza su recorte epistemológico más importante³.

SITUACION DEL AD Y OTROS TIPOS DE ANALISIS



(podemos considerar que los análisis del discurso, de contenido y semióticos pertenecen a la categoría general de los análisis semánticos).

² El marco interdisciplinar que disfruta la Documentología y algunos recortes epistemológicos necesarios, se recogen en García Gutiérrez, Antonio: Estructura lingüística de la Documentación: teoría y método. Murcia: Servicio de publicaciones de la Universidad de Murcia, 1990, cap. 1

³ El objeto de investigación de la denominada en España "Lingüística documental" coincide con la línea de investigación en Análisis documental abierta en la Universidad de Sao Paulo (USP). Véanse su presupuestos teóricos en Cunha, Isabel: Do mito à análise documentária. Sao Paulo: EDUSP, 1990, 163 p.

La mayor gravitación, que he defendido en los últimos cinco años, del AD sobre las disciplinas relacionadas con el texto/discurso y su desestructuración se ha debido a un deseo de equilibrar la fuerte inclinación de los últimos 25 años hacia la Informática, hasta tal punto, que los ingenieros proponían, y lo siguen haciendo, los modelos a seguir. Una vez encontrado mayor equilibrio en algunas contribuciones, muy pocas, españolas y extranjeras, es preciso señalar el necesario papel de la tecnología, bien dirigida, en la consecución de los procesos documentales en los que interviene el AD.

Por ello, tras la exposición de las triviales confusiones detectadas en lo que respecta a método y norma documentales, propondré a la par un cuadro teórico adecuado a las necesidades de la investigación actual sobre AD y un marco referencial tecnológico en el que se presupone que deben cuajar pragmáticamente las propuestas teóricas y los métodos. Debemos pasar, por tanto, de un positivismo a ultranza que ha dominado la Documentación desde la época de los fundadores⁴, a la necesidad de ciertas dosis de especulación teórica, sin llegar a la ficción futurista, sobre las potencialidades del AD. En definitiva, debemos imaginar, dentro de una tecnología factible y en constante renovación, la posición de la Documentación dentro de un década y realizar previsiones metodológicas que sean asumibles por las computadoras.

Esto supone dar al traste con las tendencias tradicionales de unas disciplinas que asumen, en su propio discurso, el estatismo de sus objetos orgánico-pragmáticos, archivos, bibliotecas o centros de documentación, a las que este trabajo pretende aportar una pequeña contribución. No obstante, ha de resaltarse la validación empírica como culminación de la propuesta de métodos documentales.

No debemos olvidar, aunque no se realicen menciones constantes sobre ello, la articulación del AD con los lenguajes documentales, el otro gran objeto de estudio de la LD. Pero a ellos hemos dedicado, muy recientemente, un estudio conceptual, instrumental y metodológico que ahora pretendo equilibrar en el área de las condiciones operacionales. En cualquier caso los lenguajes documentales afectan al AD en la metodología de la representación y son contruidos a partir de los modos estructurantes del discurso⁵. Esos modos han

⁴ La perspectiva positivista de Otlet es recogida por José López Yepes, en el marco de la evolución y las polémicas en torno a la disciplina, en *Teoría de la Documentación*. Pamplona: Eunsa, 1978, XXXI+337 p. y por Félix Sagredo y José María Izquierdo, con un enfoque etimológico y lógico-semántico, en *Concepción lógico-lingüística de la Documentación*. Madrid: Ibercom- Red Comnet, 1983, 440 p.

⁵ Hipótesis sostenida en García Gutiérrez, A.: *Estructura lingüística... o.c.*, cap 5.

sido observados tanto para la construcción de vocabulario como para la desestructuración del texto con fines analíticos/sintéticos.

1.1 Método y normas

En las enseñanzas y aplicaciones del AD se suele identificar método y conjunto de normas. Unas cuantas normas que sólo afectan al nivel de superficie y a la representación de un documento han obtenido el título de método de AD en el quehacer profesional. En consecuencia, además de la confusión, observamos una reducción ya que se considera exclusivamente una de las fases del AD.

Si entendemos por método científico (del juego meta-hacia y odos-vía) el conjunto de elementos que conducen el espíritu a un procedimiento determinado⁶ y uno de los objetos de estudio de la Epistemología (junto a la delimitación y estructura del conocimiento científico) podemos comprobar la ausencia de ello en lo que se refiere al Análisis documental a pesar de declararse, fundamentalmente, como método. El método es lo que permite independizar el fenómeno observado de los intereses del observador. Sin embargo, el corpus actual del AD se adapta perfectamente a lo que podemos considerar normalización: codificar un proceso para resolver un problema que se repite con frecuencia ya que se reduce el método a una representación (codificación) en lengua natural o artificial.

Debemos considerar la normalización como parte del método para la ejecución del análisis (su regulación en la práctica), en consecuencia, y desarrollar los aspectos anteriores a la formalización textual no encontrados en la bibliografía sobre el tema o en los programas de enseñanza, dejando patentes las siguientes precisiones:

- 1) La denominación de método, para el AD tradicional, es insostenible; sencillamente no existe porque no afecta a todos los pasos del análisis excepto a la fase final en la que ya no son precisas normas de comportamiento lector, sino códigos convencionales que no hacen otra cosa que ocultar los objetivos de un verdadero análisis documental.

⁶ Dictionnaire de la Philosophie. París: Larousse, 1964, p. 182.

- 2) Las normas creadas para el análisis (en realidad, para la representación) no se preocupan de las características o la estructura del texto, sino tan sólo de la reducción de su espacio de almacenamiento y de un alto índice de recuperación cuando sea solicitado. Este último objetivo no planteó problemas durante décadas de contenido crecimiento de la información científica. En la actualidad, la saturación de datos científicos y periodísticos, en sí misma, exige la revisión de la normativa.
- 3) Los diferentes sistemas normativos propuestos desde los años 50, superándose unos a otros gracias a la tecnología (y en beneficio de la propia tecnología) ofrecían aparentes mejoras que, rápidamente, eran inmovilizadas por nuevas saturaciones. La lucha, en términos mercantiles, se planteó como busca de una óptima canalización de la información cualesquiera que fueran los mensajes.
- 4) El factor formal de coincidencia⁷ prioriza las técnicas de análisis automático y supeditan el análisis efectuado por el hombre. Los instrumentos ideológicos de representación documental, los lenguajes documentales, informan a los documentalistas, en última instancia, de la etiqueta aceptada o preferente. Los textos son almacenados con tales etiquetas formales en las memorias automáticas con claros síntomas de inventario de objetos y no de depósito cognitivo. Algunos organismos internacionales intentan, infructuosamente, construir lenguajes documentales interculturalmente aceptables en abierta contradicción con la excesiva formalización de significantes y estructuras arborescentes que requieren esos mismos instrumentos.
- 5) Las normas programadas en el software se basan en recuento frecuencial de palabras y, en el mejor de los casos, comparación y sustitución por términos preferenciales localizados en una memoria auxiliar. Es obvio que el sentido de un documento no se encuentra en el nivel de palabra, ni de enunciado, sino de texto y discurso. Por eso, considero que este campo de trabajo puede incluirse en la llamada "Tecnología del discurso" en su cruce con la Documentación como Ciencia cognitiva (en el plano de representación y recuperación del conocimiento).

⁷ La cuestión de la coincidencia formal y el fracaso de las investigaciones basadas en ese parámetro se encuentra en García Gutiérrez, A.: Nuevos parámetros para una Teoría de la Indización. In: López Yepes, J. (comp.): Fundamentos de Documentación e Información. 2ª ed. Madrid: Eudema, 1990, 485 p.

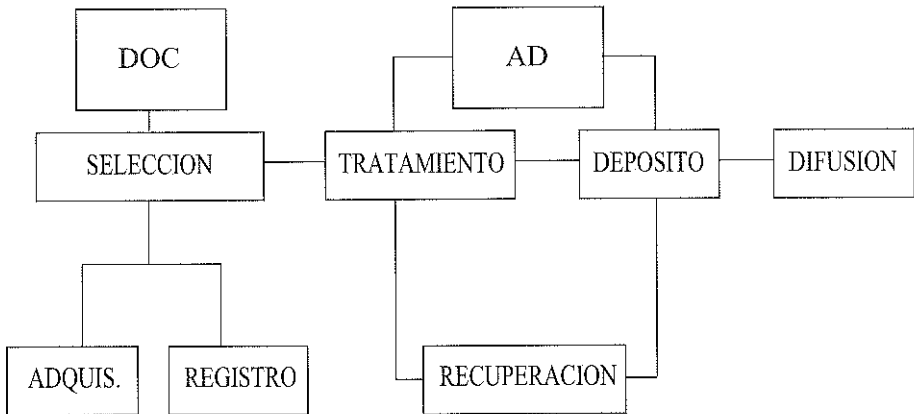
- 6) La captación lineal del texto por parte del ordenador desvirtua, asimismo la "interpretación". La alimentación de la memoria con fragmentos o síntesis previas efectuadas por los documentalistas ya hace innecesaria la presencia del ordenador en la fase de análisis, para la que se supone que debe trabajar. Las cuestiones relativas a entrada de datos mediante, teclado, scanner y Optical Recognition Character o cualquier otro dispositivo no afectan el sistema de lectura y, por tanto, las mejoras de esos aparatos no influyen en el estudio que efectuamos.
- 7) El análisis documental tradicional produce resultados insuficientes para responder a demandas lógicas planteadas por el usuario. No intuye, ni considera la presencia de un usuario ni sus posibles comportamientos en la recepción. Por el contrario, le impone un cuadro y estrategias que comportan grandes pérdidas de información y de libertad para pedir la información. El diálogo interactivo provisto por el AD automático encorseta al usuario en un diálogo manejado por y desde el ordenador y los productores de su soporte lógico. Esta situación puede agravarse enormemente a partir de la comercialización de bases de información en sistemas expertos e inteligentes.

1.2 Delimitación teórica: el paradigma del proceso documental

Debo detenerme en este punto para situar adecuadamente el marco en el que va a desenvolverse el método general del análisis documental, hasta ahora inscrito en el denominado proceso documental y, dentro de él, vinculado a la fase de entrada del organo centralizador llamado tratamiento⁸.

El proceso documental corresponde a las tareas o pasos de la cadena documental que comprende desde la selección física del documento hasta su integración en una memoria, manual o automática:

⁸ Jacques Chaumier representa la vinculación de la disciplina al organigrama empresarial, de gran influencia en la mayoría de los trabajos realizados en España en la pasada década. Véase, por ejemplo, *Les techniques documentaires*. Paris. PUF, 1974, 126 p. o cualquiera de sus obras sobre análisis y lenguajes. Traducida al español se encuentra *El análisis documental*. Barcelona: Mitre, 1986, 172 p.



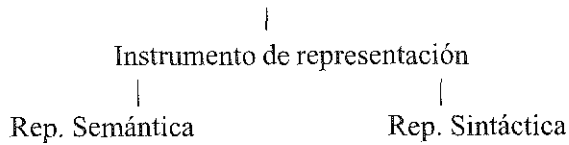
Veamos, a continuación, el modelo linguodocumental en el que he basado mis trabajos hasta 1987⁹, como reflejo de la fuerte presencia del Estructuralismo lingüístico, y sus vinculaciones disciplinarias esenciales:

DOCUMENTO	FORMA	FORMA	ANÁLISIS FORMAL
		CONTENIDO	
	CONTENIDO	FORMA	ANÁLISIS DE CONTENIDO
		CONTENIDO	

Propongo, a tenor de la línea investigadora abierta en las Universidades Complutense de Madrid y de São Paulo, un marco teórico para el AD basado en las operaciones que afectan estrictamente al contenido del documento al que llamaré paradigma del proceso documental (P/PD). Este modelo, y sus consecuencias generales, pretende ser universal para todo el AD independientemente de su aplicación y es un resultado de la fusión de los anteriores del que la teoría documental sale reforzada.

⁹ La excesiva influencia del estructuralismo fuerza el esquema modelo del documento, y su paralelismo al signo, para proponer un objeto propio de investigación documentológica en García Gutiérrez, Antonio: *Lingüística documental. Aplicación a la comunicación social*. Barcelona: Mitre, 1984, 279 p. El modelo aparece revisado en mis trabajos, con una mayor implicación y naturaleza documentológicas, a partir de 1986.

ANÁLISIS ↔ SÍNTESIS ↔ REPRESENTACIÓN ↔ RECUPERACIÓN



Estos cuatro segmentos -ASRR- mantienen una presencia simultánea en este estudio y deben tenerla en cualquier trabajo en torno al Análisis documental. ASRR se halla en fluctuación e interacción constante en ambos sentidos. Las propuestas planteadas aquí, a tenor de las pruebas empíricas y sus verificaciones, han de ser sometidas al P/PD influyendo todos los segmentos, como conjunto inseparable, en cualquiera de ellos individualmente aislado para observación. El concepto de AD abarca las operaciones de Análisis (lectura o reconocimiento) y síntesis (transformación), comparte la representación (con los instrumentos) y participa en la recuperación.

El P/PD sintetiza los principios y objetivos del AD como método operativo y de su disciplina matriz, la Lingüística documental, que también tiene, como objeto de estudio, los instrumentos denominados lenguajes documentales. Vemos, además, como el modelo se circunscribe dentro de los procesos culturales exactamente en el momento del paso de los discursos (sociales, científicos, etc.) a los metadiscursos (producidos a través de la acción documental).

1.2.1 Análisis: operación inicial equivalente a una lectura empírica¹⁰ que consiste en dos momentos: reconocimiento y desestructuración del texto de acuerdo a unos parámetros, por un lado, y selección de las proposiciones del productor, por otro. La metodología de análisis atañe, por tanto, al plano de la lectura dirigida de la segmentación del documento o unidad documental. Este es el aspecto fundamental (hasta ahora intuitivo o mecánico) y menos explotado del AD; el documentalista necesita un sistema referencial de ideas y reglas para interrogar el texto (en amplio sentido) que va a abordar. Revisaremos algunos métodos de análisis, sólo los más aprovechables para los fines del AD y propondremos algunos principios de procedimiento articulado que faciliten un proceso de explicitación de la lectura dirigida, teniendo siempre presentes los

¹⁰ Para Cintra, lectura metacognitiva en la que se realiza un esfuerzo consciente por parte del lector. Cintra, Ana M^a: Estratégias de leitura em Documentação. In: Smit, J. (org.): Análise documentária: A análise da síntese. Brasília: IBICT, 1987, p. 27-35

parámetros tecnológicos para una lectura artificial basada en el reconocimiento de la estructura textual desde la función y desde su organización por el productor para describir las condiciones de la construcción en un discurso específico. Sostenemos, por ello, la hipótesis que nos sugiere la imposibilidad de generalizar el AD.

1.2.2 Síntesis: la metodología de la síntesis se despliega una vez superada la primera fase de captación que produce segmentación virtual. Nuevamente el pragmatismo documental impone la condición de la economía en esta operación. El analista es sometido a restricciones, algunas extradocumentales, en su labor sintética de tal modo que la libertad de interpretación se convierte en una pérdida de información suplementaria apareciendo una nueva tipología de omisiones¹¹. La políticas de selección de un organismo se dirigen, sobre todo, a esta fase del AD, empero, descontextualizada. Ello trae como efecto posterior la síntesis parcial o incompleta de un texto en lugar de la síntesis contextual o globalizada que supone una condensación del texto en virtud de la relevancia de sus contenidos y la priorización de los temas esenciales detectados. La síntesis, punto de encuentro del texto/usuario, es una recomposición inductiva del discurso fragmentado previamente por el análisis deductivo. El conjunto de síntesis de un sistema refleja, a otro nivel, la esencia del discurso. La síntesis es una causa del extrañamiento del autor respecto a su obra: produce una desvinculación o abandono de la paternidad de la idea extractada. La propia naturaleza de la síntesis hace que el documentalista tampoco parezca haber participado como mediador o redactor de ella: parece salida espontáneamente del propio texto. Debo exponer, aquí, que hablar de síntesis exclusivamente como resultado del análisis es una reducción puesto que otros productos como datos, verificaciones o inferencias son también de interés documental (vid. 3.3.1). Preferimos, con algunas reservas, el concepto más genérico de transformación.

1.2.3 Representación: el método de representación se confunde con la norma habitual de los AD. No obstante, podemos realizar una crítica general a los sistemas de representación documentales: se basan exclusivamente en la

¹¹ La responsabilidad del documentalista en la lectura, hasta dónde debe llegar su labor de verificación (dentro del propio texto, intertextual o contextualmente), es debatida en García Gutiérrez, A.: Suficiencia estructural y tipología de la omisión en AD. "Documentación de las Ciencias de la Información", vol.13, 1990, p.73-86

coincidencia de significantes y este objetivo impone una fuerte codificación y, por tanto, anquilosamiento en la circulación de mensajes. Las condiciones de producción de los lenguajes documentales, o instrumentos que canalizan la representación, coadyuvan a la rigidez de la comunicación. Como tesis demostramos, en otro lugar, la necesidad de establecer el contacto en el nivel de significado: ello haría fluir la información haciendo inservibles muchos formalismos que afectan al significante. En definitiva, el tratamiento de la representación como parte del objeto teórico del AD debe tener unos objetivos claros: la naturalización del proceso, es decir, la adopción de recursos de todo orden de la lengua natural en el lenguaje documental liberalizando esta fase compartida con el AD: libertad de sinónimos, libertad controlada de polisemias, sustitución de operadores lógicos por operadores sintácticos, sustitución de "módulos de representación" por enunciados, entre otros. Identificamos dos niveles distintos en el plano de la representación: nivel semántico, que regula la forma de los conceptos, y nivel sintáctico, que regula su combinación en el plano funcional. Por ello, los estudios de representación deben cubrir tanto el corpus de significantes como de operadores gramaticales. La representación es un segmento que atañe tanto a los mensajes extraídos de los textos como a las demandas expresadas por los usuarios. En cuanto al primer punto, se verá más adelante la importancia de la representación al referirnos a la necesidad un cuadro de macrodescripción para que el análisis sea factible. Sobre el segundo punto, se justifica un inventario de interrogantes.

1.2.4 Recuperación: La recuperación, o proceso de búsqueda y obtención de información, ha estado desterrada del ámbito teórico, metodológico y pragmático del AD. Sin embargo, es patente que la desvinculación es inviable: la recuperación es el objetivo de la primera fase, el análisis, condicionando desde ese momento, todas las operaciones. La recuperación ha sido, también, pasto investigador de las evaluaciones de sistemas que tienen como objetivo la rentabilidad. Este punto es comprobable en cientos de contribuciones publicadas en las revistas norteamericanas sobre Documentación de la última década. Esto ha llevado a la identificación de Documentología y técnicas de recuperación en muchos países.

Dentro del P/PD, la recuperación funciona como una fase final que retroalimenta el sistema, que lo condiciona desde su concepción inicial. Por ello debe cambiarse la orientación investigadora que prima sobre la recuperación por una óptica en la que se interrelacionan los documentalistas, como usuarios/mediadores, y los usuarios finales, que permite proponer cuadros universales de

interrogantes posibles a un sistema. Debe llegarse al punto de confluencia entre la estructuración del texto por parte del autor y la búsqueda de un texto, mediante interrogantes, por parte de un usuario. Existen casos universales de estructuración de mensajes que pueden ser transportados a computadoras y son coincidentes con la estructura mental de búsqueda (sus enfoques, objetivos y forma de plantear las demandas). En todo caso, las interrogaciones que efectúa un documentalista al texto en la fase de análisis deben reflejar todas las posibles demandas planteadas por el usuario final a ese mismo texto, como tal, y como fragmento de un entorno discursivo.

La dispersión de interrogantes obliga a inventariar y sistematizar todas las posibles. Por ello, en cualquier sistema de Documentación deben explicarse las condiciones de interrogación/consumo de información así como las reducciones o simplificaciones previstas en la arquitectura del sistema y los márgenes tolerables en la distorsión observada en la respuesta. La demanda tradicional de bases de datos automatizadas se funda en la coincidencia formal de significantes. El usuario es obligado a encorsetar su demanda, no solamente en el nivel de significado, con significados reductores, sino a un restrictivo planteamiento: este punto destaca el mito del llamado "diálogo hombre-sistema" ya que este proceso no es más que una parodia simbólica de la verdadera comunicación. La naturalización en la emisión de mensajes, por parte del usuario, debe ser uno de los objetivos de la metodología de la recuperación como corpus esencial de investigación del AD. Interesa en el análisis, por tanto, no solamente qué se va a buscar en el sistema (datos y relaciones de datos) sino también cómo va a ser buscada la información (estructura y gramática de la demanda). En definitiva, debemos interesarnos por los resultados de la recuperación para revisar los procedimientos pero el objetivo final es el uso del material obtenido o grado de satisfacción del usuario.

1.3 El AD: *status quaestionis*

El AD que tiene por objetivo la recuperación del documento mediante la representación de los trazos descriptivos de su contenido mediante términos, sintagmas u otras unidades discretas se denomina tradicionalmente indización. A pesar de que hay que superar la clásica y poco operativa división de productos en descriptores y resumen, debo indicar que el análisis automático se ha centrado exclusivamente, hasta la fecha, en la detección de conceptos a través de sus representaciones naturales y, en algunos casos, a la conversión de éstas

en otras alternativas. La transformación de originales en síntesis es algo reservado todavía al documentalista aunque los métodos de traducción automática disponibles para producir versiones en distintos idiomas son aprovechables durante un cierto recorrido común: al final, los traductores se interesan por la traducción lineal en una lengua terminal, los documentalistas por la traducción sintética en la misma lengua del original controlada desde un lenguaje de representación formalizado. Hasta ahora, los procesos de lectura en indización están determinados, no en función de un texto o de unos objetivos sino en función de un lenguaje formalizado que impone, en todo momento, sus condiciones. Este problema de influencia de la representación, no sólo en los resultados sino también en los modos de captación de las ideas, debe gozar de un amplio estudio documentológico que clarifique hasta dónde debe llegar el ld y de qué modo puede ser modificado por el proceso analítico cuando está operando.

En esta parte del trabajo me centraré en el AD como modo de indización textual revisando algunos aspectos que someten la labor de los documentalistas a los dictados del lenguaje. La indización humana (y utilizo este concepto, de ahora en adelante, como fase del AD que pretende sobre todo la recuperación entre otras funciones) está desprovista de objetividad y se basa en dos criterios asimilados por cualquier documentalista:

- * El sentido común del lector-documentalista, sumado a la memoria e intuición, determina lo relevante y rechaza lo insulso. El sentido común, en Documentación, no debe existir como tal ya que se comprueba que cada documentalista impone su aparato ideológico-cultural al segmentar el texto. En los Cursos de posgraduados he testado esta cuestión sobre resúmenes específicos leídos por licenciados en distintas especialidades ¹² llegando a distintos conceptos sobre el mismo texto, a pesar de haber sido establecidas unas condiciones que limitan la dispersión, en un porcentaje de variación que depende de la extensión y de la complejidad del documento pero nunca inferior al treinta por ciento sobre una media de diez descriptores. Obviamente sin condiciones y en textos más amplios el desajuste crecerá. El mismo documentalista, trabajando sobre el mismo texto en tiempos distintos, detecta o rechaza sentidos modificando los resultados anteriores. El sentido común, en estos casos, no es válido para la gestión de una memoria

¹² Durante algunos años, las prácticas de análisis de textos en el Curso de Documentación para posgraduados de la UCM, se basaban en resúmenes de autor. A pesar de esa facilidad del texto controlado, la dispersión de resultados provocaba debates en torno a lo relevante, a lo esencial, a la intuición y al sentido común enfrentados con la epistemología de la construcción conceptual del resumen.

institucional en la que suceden frecuentes cambios de personal poniéndose en peligro la continuidad del sistema. Esto se debe a que el documentalista no dispone de un modo de lectura que no eliminaría las diferentes opiniones pero reduciría distorsiones importantes. En consecuencia, el lector debe disponer de un conjunto de procedimientos (interrogantes y reglas) propios del AD que le permitan fijar su atención en determinadas estructuras del texto para resaltar lo que interesa, en general, al sistema y a un usuario-modelo.

- * Uso de un lenguaje documental que regula la representación al cien por cien y, en muchos casos, la selección de conceptos determinando su recuperación según su existencia o no en el lenguaje. En efecto, un Id perfecciona los resultados de la indización pero sólo en el nivel de la normalización y de la unificación de criterios entre los documentalistas, o entre estos y los usuarios. Sin embargo, no debería influir en la actividad analítica excepto para recordar cuales son los límites selectivos de una base de datos concreta. En consecuencia, el exceso de uso de un Id, por el hombre o por la máquina, distorsiona los productos del análisis.

Otro aspecto negativo de los Id es su concepción actual. Las estructuras arbóreas que presentan, incluso los más flexibles, y la rigidez de sus relaciones no recogen o encorsetan la libertad del texto o de la demanda. Por ello he realizado una investigación que propone un nuevo método de construcción de Id que no perturbe los modos de análisis, y por tanto, concurrente con los objetivos descritos aquí en el nivel analítico¹³.

Con la justificación de evitar la “anarquía del sentido común” del análisis efectuado por documentalistas y de reducir la relación n.º de documentos / tiempo de indización, surgen en los años sesenta los primeros paquetes de indización automática que podríamos denominar “modelos de representación automática” basados en la lectura secuencial de los textos¹⁴.

La preponderancia de la operación “representación” sobre la operación “análisis” mediante ordenador nos obliga a considerar, a lo que hasta ahora conocíamos por modos de indización, como meros modos de representación ya que no proponen reglas de lectura sino que circunscriben el texto a unas etiquetas de procedencia práctico-empíricas. Esta es la razón por la que las tesis de doctorado que oriento, desde 1986, van enfocadas a la problemática de la

¹³ Propuesta en Estructura lingüística... o.c.

¹⁴ Micrex, Macrex, Newsdex, son ejemplos de paquetes de indización automática lineal.

captación del texto con fines documentales, al complejo paso de unas estructuras de productor a otras de usuario con la mediación de documentalistas y tecnologías y, desde luego, cómo incide la representación, bajo nuevos enfoques, en esa fenomenología. De modo indicativo, citaré algunos de los títulos próximos que reflejan esas tesis: “Construcción de un modelo de análisis para la imagen de informativos de televisión”, “Superación de los operadores lógico-matemáticos de la recuperación automática mediante operadores naturales en el nivel de metalenguaje, aplicado al discurso jurídico”, “Aplicación de modelos lógico-semánticos para la determinación de la fiabilidad de las fuentes documentales” y, en proyecto: “Propuesta de estructura relacional para la construcción de lenguajes en el discurso periodístico”, “El arte como objeto del análisis documental” y “Reconocimiento artificial de las categorías funcionales”.

1.4 Crisis de la investigación actual en AD

En la práctica indizadora, manual o automática, a pesar de las mejoras incorporadas por los distintos sistemas, el problema persiste:

- la lectura lineal, palabra a palabra, enunciado a enunciado impide la comprensión del texto y, por tanto su representación.
- la falta de una metodología humana de interrogación del texto impide no sólo *normalizar* los resultados y aumentar la eficacia del análisis sino también diseñar un programa inteligente capaz de emular al documentalista en sus resultados (los métodos no parecen simulables).
- Ausencia de modelos teórico-conceptuales capaces de explicar los procesos habidos en el AD: del texto al metatexto de recuperación. Después de varias décadas de explotación de sistemas, contemplamos una parálisis de las investigaciones por lo que debemos hablar de crisis del Análisis documental tradicional. En mi opinión, esta crisis debe ser superada mediante otros objetivos específicos:
- Búsqueda de métodos de lectura contextual sistematizada tanto para el documentalista como para el ordenador inteligente.
- Superación del sistema de representación basado en la coincidencia de la expresión del análisis con la expresión de la demanda.

- Búsqueda de un modelo de diálogo interactivo en el nivel de estructura profunda. En ese nivel la segmentación del texto, libre del documento, interacciona y mezcla con otras segmentaciones recuperándose la intertextualidad o representación virtual del discurso global de los documentos analizados. A ese mismo nivel acceden las segmentaciones de la demanda, reguladas por un instrumento linguodocumental que también actuó en la fase de entrada.
- Propuesta de gramáticas que articulen las representaciones superando los recursos lógicos válidos en discursos unívocos pero no en Ciencias sociales o humanas, aceptables en lenguajes de alta formalización pero insuficientes para operar en lengua natural.
- Búsqueda de un marco interdisciplinar integrado, formado de recortes epistemológicos de varias Ciencias, en el que poder acomodar las estrategias y enfoques de la investigación en AD. Existe ausencia de objeto teórico: El análisis documental no ha sido formulado como objeto teórico, no han sido recortadas convincentemente sus influencias, ni ha sido adscrito a un universo referencial más amplio.
- Indefinición paramétrica del objeto pragmático: las unidades sometidas a análisis documental: palabras, enunciados, textos, discursos, relatos, documentos, en definitiva, unidades de contenido o unidades documentales no han sido definidas. Poco se sabe sobre qué parámetros intensionales y extensionales se aplican en el orden cuantitativo ni sobre qué parámetros intencionales o persuasivos deben ignorarse, obviarse o explicitarse adecuadamente.

Sin parametrización es inviable, por ejemplo, realizar estudios sobre muestras de unidades discretas que permitan testar un método y verificar los resultados. La vía estrictamente práctica queda anulada en Documentación a pesar de ser fundamental para la validación, dados los objetivos y la sustancia pragmática de esta disciplina y de la propia teoría. Solamente con ayuda de la simulación, de la especulación y de la heurística es posible realizar avances sustanciales en AD, por el momento, pero siempre con la adscripción a condiciones o delimitaciones de corpus empíricos. Esta aparente paradoja es lógica en un campo que trata de reducir problemas de Conocimiento en registros manipulables. La indefinición de objetos de análisis no es sólo propia del AD. En campos como la Lingüística, o el Análisis de contenido, muchos eruditos no están de acuerdo sobre los criterios para constatar la existencia de unidad

mínima para el trabajo empírico (cfr. los campos semánticos, o las unidades de contenido en AC).

El amplio espectro documental, por otra parte, no ayuda a definir las unidades de trabajo, superada la presencia del soporte material, y mucho menos a universalizarlas: así, en Ciencias afines será factible un objeto similar, pero la estructura del discurso periodístico o político exigirá distintas reglas. Debe llegarse, en todo caso, a reglas generalizables (si no generales).

- Aislamiento de las fases de AD: esta actitud, como si de división del trabajo se tratara, impide la concepción de un objeto teórico (P/PD) compacto, y excluye las influencias y nuevas situaciones que genera esta articulación. En el peor de los casos, como se ha señalado, se confunde el AD con una de sus fases intelectuales o materiales (se confunde todo el proceso con la representación, en el caso extremo) siendo partes del mismo método y actividad humana o automática y estando en mutua dependencia.
- Desvinculación de operaciones e instrumentos: tomando instrumento como equivalente en la lengua a sistema y norma, y operación a habla, en Documentación, según se refleja en los planes de estudio, se escinden, a pesar de ir articulados como significado/significante, no para justificadamente observar comportamientos o defectos, sino para preparar especialistas de una u otra disciplina. Esta forzada división causa problemas cotidianos como el de la adscripción inadecuada de conceptos o procesos, bien a operaciones bien a instrumentos, o a la repetición de la misma temática en una y otra asignaturas. Se observa, también, que las denominaciones de los planes de estudio españoles responden a criterios empresariales, por ejemplo, al reparto de tareas en la Institución y no a una división lógica de la Documentación.
- Bloqueo en el análisis automático de documentos: los ingenieros continúan construyendo artefactos para el análisis documental y los documentalistas sólo se preocupan de aprender las instrucciones de utilización sabiendo que a corto plazo los resultados no serán satisfactorios. Los papeles deben invertirse: los ingenieros deben estudiar las propuestas y necesidades sugeridas por los documentalistas e intentar materializarlas en equipos tecnológicos. Es preciso, para ello, un acercamiento de ambos sectores, comenzando por un mayor conocimiento de los documentalistas de su responsabilidad social, el

rol de una computadora inteligente y el control de las actividades para humanizar el proceso y el uso.

- Desconocimiento de modelos y métodos desarrollados en otros campos: El AD se sitúa en un amplio campo interdisciplinar; esto justifica, en cierto modo, el desconocimiento de las aportaciones de otras disciplinas. Como no parece posible que el documentólogo sea especialista simultáneo en Semántica, Lógica, Teoría del conocimiento, Semiótica, Informática, Análisis de contenido o Teoría de la Comunicación, es preciso trabajar cooperativamente, sobre los problemas formulados como prioritarios, con personas especializadas en esas disciplinas con conocimiento exacto de los objetivos, instrumentos y conceptos de la Documentación. En definitiva, la falta de investigadores capacitados multidisciplinariamente es determinante en la crisis de la Teoría Documental.

1.5 Condiciones previas a la formulación de un método analítico

El análisis de contenido documental adquiere grandes espacios para la representación de sus resultados en los formatos y hojas de trabajo de las bases de datos, campos de resúmenes y descriptores, de aproximación, de ponderación, etc. Sin embargo, ese lugar reservado para la extracción de la esencia documental no se ve correspondido por un conjunto de procedimientos que normalice y oriente la lectura del documentalista al abordar el texto, como ocurre con otros campos de naturaleza más descriptiva (por ejemplo, los previstos por ISBD) y, desde luego, más fácilmente regulables.

Debo insistir en que las normas del AD se reducen a unas cuantas pautas de representación y esta situación se repite tanto en los manuales generales sobre AD como en los específicos de software o de gestión del contenido de una base de datos. La paradoja es patente: se omiten las reglas para efectuar el análisis cuando la fase que permite mayor circulación de información en un sistema es precisamente el análisis dependiendo sus resultados del grado de normalización. La ausencia de reglas generales en las que se apoya el documentalista para la interrogación del texto crea una falsa situación de libertad de lectura porque, en realidad, todo el proceso documental, y en especial la interpretación, está sometido a estructuras ideológicas y a intereses de instituciones e individuos. Por lo tanto debemos preguntarnos lo siguiente: ¿Es posible regular un

procedimiento general o particular de análisis de contenido en documentación? Mi hipótesis de partida es afirmativa pero tampoco podemos dejar de preguntarnos esto otro: ¿Son favorables los gestores de la información a que tal procedimiento exista y podamos, en consecuencia, programar máquinas para que lo lleven a cabo con la mayor objetividad posible? Los efectos de la difusión de los resultados a través de medios masivos, en este caso, pudieran ser nocivos para los intereses de los mismos gestores. Por ello pienso que la respuesta a esta última cuestión es negativa.

La falta de un método de lectura empírica en AD plantea problemas de indefensión al documentalista que está al servicio del flujo de información sin discriminaciones. Este importante actor del proceso está actualmente en vías de ser sustituido por máquinas convencionales que realizan indización automática de manera secuencial lo que obliga al documentólogo a partir ex-nihilo en este campo al encontrarse desprovisto de la capacidad de poder criticar modelos anteriores y buscar ayuda en áreas aledañas. En esta situación enfoco el problema y, por ello, es necesario formular las condiciones generales previas para la proposición de cualquier método de lectura documental cuyo objetivo sea la recuperación:

- * El método debe describir la lógica del razonamiento en función de los objetivos trazados y explicar como efectuar las fases en vista de tales objetivos. Para la ejecución automática, las fases deben ser parametrizables y codificables para la programación.
- * Es inviable practicar un método fuera del contexto de las nuevas tecnologías y de las nuevas posibilidades tecnológicas, concretamente de la concepción de sistemas expertos e IA. Las pautas establecidas han de ser transportables (pragmática documental) por lo que cualquier estudio general debe prever el marco tecnológico en el que van a desenvolverse sus posibles aplicaciones.
- * Las profundas diferencias en los ámbitos del discurso, la temática, la estructura, los canales o los usuarios de un campo a otro sólo permiten generalizar tras haber enunciado las condiciones de esa misma generalización. Por tanto, podemos establecer que un método de AD no puede ser general sino tan sólo tener algunos parámetros generalizables (compatibles en otros campos). Es importante establecer los niveles mínimos de esa compatibilidad para reforzar la coherencia de los distintos métodos.
- * Los parámetros de lectura deben articularse entre sí pero no establecer dependencias, es decir, adaptarse a cada texto particular sufriendo las

irregularidades que éste pudiera presentar en lugar de mostrar un inventario inflexible de interrogantes.

- * El análisis no debe realizarse en función de una representación determinada, como ocurre con los métodos actuales, en especial con aquellos que se basan en la preordinación, sino en función de un uso (es, también, necesario definir qué significa uso documental en relación al paradigma, cuyo tronco se establece a través de la representación).
- * Deben observarse los métodos de análisis propuestos en otras disciplinas que, como la Documentación, trabajan en el marco de los procesos cognitivos y culturales, teniendo en cuenta que ninguna de ellas tiene los mismos objetivos de la Documentación: recuperación, control y circulación de la información.
- * Deben preverse mecanismos que permitan resultados aplicables a la producción de lenguajes documentales, además de la función habitual del AD que es describir los sentidos y proposiciones del texto.
- * El AD debe permitir la libertad de interpretación de determinados textos y facilitar las detecciones de manipulación y omisiones mediante un sistema codificado que alerte de la presencia de las mismas.
- * Para la construcción de un método es precisa la investigación empírica sobre muestras. La muestra tendrá como limitaciones el tema y la extensión del texto pero deberá poder representar a cualquier texto que cumpla las condiciones establecidas. Por lo tanto, deben tenerse en cuenta otras posibles variables: tipo de discurso, tipo de documento, extensión, estilos y géneros, escuelas y perspectivas, aspectos léxicos y sintácticos, formación de usuarios del sistema, naturaleza de los productores de textos e incidencia de la ideología.
- * Los resultados del AD inteligente simulado deben ser representables en lengua natural y permitir combinaciones sintácticas también en lengua natural.
- * Las condiciones anteriores no contradicen la necesidad de que los resultados se ajusten a una estructura preconstruída siempre que esta misma permita la automodificación en función de los cambios que proporcionen los resultados del análisis. Tal lenguaje documental, obviamente, no puede ser del tipo tradicional de nomenclatura o jerarquización.

- * Todo método de AD debe dar resultados en Documentación, modificando las condiciones anteriores de lectura y acelerando los procesos. Los métodos que generan explicaciones de los textos, deteniéndose en comparaciones, por ejemplo, no son válidos en Documentación porque impedirían la agilidad del análisis. Recuérdese que la finalidad de un procedimiento de lectura no es sólo optimizar los resultado sino también resolver los problemas de saturación de millones de datos en una memoria. Esto no contradice que las razones de la captación de una proposición, por ejemplo, sean explicitadas si así lo requiere un usuario.

- * Las categorías de interrogación textual, como método de lectura, deben reflejar las potenciales demandas del usuario modelo. No sólo es importante para el lector detectar las condiciones de producción sino también las posibilidades de consumo de la información. Por lo tanto, debe construirse un inventario de potenciales demandas sobre una muestra de usuarios seleccionada con criterios similares a los reseñados antes respecto a la muestra de textos. Deben establecerse las condiciones para la construcción de la muestra de usuarios y de sus unidades discretas así como preverse el nivel metalingüístico para controlar simbólica y sintéticamente tipos, categorías, estructuras, extensiones y temas.

- * Los resultados de un proceso lector dentro de un corpus documental deben ser interactivos, es decir, participar y ser participados por los resultados de otros documentos analizados (como ocurre en una red hipertexto aunque sólo en el nivel significante). El conjunto de resultados de un proceso repetido de lectura sobre documentos debe representar una situación discursiva determinada e informar al usuario sobre esa situación.

- * Los resultados de un método de AD, propuesto teóricamente, deben ser validados no sólo en términos de eficacia de la recuperación, para lo que es creado, sino también en cuanto a la tasa de representatividad, adecuación y objetivación del original respecto a su síntesis desde distintos planos de interés. La verificación fundamental debe basarse en la superación de la prueba de contraste de resultados con el cuadro empírico de demandas de usuario y del testado del método en textos fuera de la muestra. El índice de generalización del método puede detectarse mediante el testado sobre textos que gradualmente no cumplen las condiciones previas establecidas sobre tipo de lenguaje, temática, científicidad, etc. Para ello es preciso establecer la gradación de la importancia de las condiciones antes de proceder a la prueba.

En concordancia con las tendencias investigadoras del Grupo Temma de la Universidad de Sao Paulo, con objetivos que sostienen los apuntados hasta ahora, y teniendo en cuenta las restricciones impuestas por su área de interés sobre el AD y en la aplicación de la Ciencias Humanas, asumo la ayuda que presta el estudio de los Sistemas Expertos para la formulación circular de nuestros problemas ¹⁵:

1. Las bibliotecas de datos tienden a sustituir a las bibliotecas de documentos.
2. Los bancos de datos tienden a ser locales (temáticos) y totales (exhaustivos).
3. Los bancos locales tienden a la representación del conocimiento (status quaestionis).
4. Las representaciones del conocimiento tienden a ser simuladas.
5. Las simulaciones del conocimiento pueden constituirse en sistemas expertos.

Finalmente, debo señalar que el AD necesita instrumentos provistos por otras disciplinas de probada experiencia, en las que reconocemos unas preocupaciones similares, hasta un determinado nivel, a las planteadas aquí: la captación y explicación de los usos discursivos, la formación de redes semánticas intra e intertextuales, la profundidad de la significación y las articulaciones entre las diferentes bandas, la modelización de las estructuras, etc. Estos elementos son esenciales para una correcta formulación de las condiciones, de los problemas y de los métodos del AD. En el discurso que nos concierne en este trabajo, encontramos en Teun Van Dijk algunas propuestas sobre estas disciplinas ¹⁶. Para el autor holandés, los discursos poseen temas y se organizan en torno a ellos, por tanto, necesitamos una Macrosemántica que considere los significados globales con el fin de poder describir los significados de las unidades menores: párrafos, apartados o capítulos enteros del discurso escrito (pero también de proposiciones y datos en el ámbito documental). También es necesaria una Macrosintaxis para caracterizar las formas globales del discurso. Según Van Dijk, los titulares de la noticia son ejemplo de esta Sintaxis y Semántica globales. Incluso precisamos de una Macromorfología que dé cuenta

¹⁵ Cunha, I. y Kobashi, N.: Análise documentária e Inteligência artificial. 1990, 25 p.+ anexos (inéd. mecanograf.) Las autoras ejecutan las propuestas de Gardin sobre un texto, ensayo que les sirve para constatar la convergencia de objetivos entre AD e IA.

¹⁶ Las obras de Teun Van Dijk, consultadas en este trabajo son: Texto y contexto: semántica y pragmática del discurso. Madrid: Cátedra, 1980, 357 p. y La noticia como discurso: comprensión, estructura y producción de la información. Barcelona: Paidós, 1990, 284 p.

de la disposición total del texto impreso como determinante de la significación global mediante recusos tipográficos y otras estrategias de montaje y resaltado.

1.6 Usuario modelo

Abordo, en este epígrafe, la perspectiva que, en mi opinión puesto que es una campo que está por construirse y afirmarse, identifica con más exclusividad el objeto de la Documentología, la información desde la perspectiva de su uso. Para esta consideración nos servimos, nuevamente, de universos teóricos ajenos que nos prestan el aparato conceptual que utilizan para constituirse y operar, en este caso, especialmente la Pragmática.

La Pragmática se desarrolla dentro de la Semiótica cuando esta comienza a preocuparse por el uso de los signos y los sentidos que genera esta perspectiva, las relaciones de los individuos con el discurso y la manipulación de éste. Para Morris, la Pragmática es la Ciencia de las relaciones de los signos con sus intérpretes y más adelante añade a esa relación la de los signos con la situación de uso a través de todas las circunstancias que participan, definiéndola como "parte de la Semiótica que trata del origen, usos y efectos de los signos dentro de la conducta en que se hacen presentes"¹⁷. Según Bobes, la Pragmática ha aclarado definitivamente que el objeto propio de la Semiótica no es el signo, sino el signo en situación, es decir, no el producto objetivado en una forma sino todo el proceso de producción que lo crea y en el que se integra para adquirir sentido¹⁸.

Con la perspectiva pragmática se nos hace posible afianzar el objeto general de investigación documentológica, la información en situación de uso con vistas a su recuperación, objeto que presenta aún múltiples polémicas y escapa a las expectativas de este trabajo. Sin embargo, la influencia del uso en el proceso documental es evidente por lo que las investigaciones experimentales y de origen básicamente inductivo, tendentes a despejar mínimamente un corpus teórico general en nuestra disciplina, pasan por la construcción de una Teoría de Uso de la información documental basada en un modelo de síntesis de los comportamientos en cualquier aplicación.

¹⁷ En Morris, : *Linguistics and the Theory of Signs*. "Word", II, 1, cit. por Bobes, Carmen: *La semiología*. Madrid: Síntesis, 1989, p. 99.

¹⁸ Idem, p.102

Los intentos de modelizar al usuario han sido abundantes hasta el momento. Desde las anticuadas hojas de consulta, pasando por los perfiles de interés, a las bases de conocimiento sobre interés y satisfacción del usuario que consulta información, se ha buscado ese modelo para el que se construye y organiza la documentación. Esa búsqueda se basa en un objetivo: el usuario no sólo puede estar intuído, debe, además, ser conocido. La compartimentación de la disciplina en función de las aplicaciones ha hecho inviable la tarea de construir el modelo de usuario. Incluso a éste se le atribuye un código básicamente institucional que lo incompatibiliza con otros comportamientos homólogos dentro del mismo área de conocimiento.

Ahora bien, si existe unanimidad en cuanto a organizar las operaciones y los instrumentos documentales en función de su utilización, por otro lado se abren grandes polémicas en torno a las metodologías y, sobre todo, a las validaciones que apoyen unos u otros resultados. Las dos corrientes generales se distribuyen entre las que prefieren un usuario especulado puesto que es muy complejo modelizar por la vía empírica y, por lo tanto, nunca responde a la realidad (los propios productores del sistema se automodelizan para modelizar el mercado de información) y los que optan por un modelo de usuario representado en una base experimental que se autorregula con la práctica. En cualquier caso, vehicular un modelo de usuario como procedimiento requiere marcos tecnológicos inferenciales o inteligentes con programas de automodificación y, es a partir de ahí, que podemos pensar en una solución al problema.

Siendo coherente con mi interpretación de la Documentología, disciplina que opera en la práctica y, por tanto, sólo admite la experimentalidad por aplicaciones para construir sus teorías (aunque hasta ahora ha sobrevivido sin ellas, en una visión miope de los problemas del conocimiento), me alinee con la segunda opción mencionada, aunque haciendo una salvedad: la construcción de una Teoría del uso no puede obtenerse a través de la vía empírica manejada para solucionar los problemas de una determinada aplicación en una Institución, sino dirigida a la producción de la propia teoría, es decir, los corpus de necesidades y satisfacciones de usuarios conseguidos mediante cuestionarios, deben ser construídos científicamente y no mediante el aprovechamiento de algunos centenares de perfiles asistemáticos creados para canalizar las rutinas de una Institución concreta.

Desde la perspectiva de la teoría del uso, tienen que abordarse previamente las siguientes empresas:

1. Construcción de inventarios exhaustivos y contrastados, validados estadísticamente.
2. Inventario de temáticas y aproximaciones de usuario a una determinada área de conocimiento.
3. Inventario de funciones y representación de demandas por parte del usuario, es decir, reconocer las modalidades de interrogación y las diferentes categorías léxicas y enunciativas utilizadas por el usuario.
4. Búsqueda de interfaces con el método de representación y respuesta del sistema.

1.7 Aplicaciones del AD: las restricciones discursivas

La Documentología encuentra su mejor terreno de cultivo en la experimentalidad realizada en áreas de aplicación diferenciada. Observamos que en las tentativas de construir modelos operativos o instrumentales no solamente es persistente la idea de la especialización temática sino también la de delimitación del área de estudio. En este sentido, el documentalista práctico suplanta al propio científico en su labor de trazar las fronteras interdisciplinares al construir su corpus de trabajo, justificado en la imposibilidad de operar con áreas abiertas. En consecuencia, la aplicación determina los resultados y, por aplicación, entendemos no sólo temáticas y aproximaciones sino también Institución, o lo que es lo mismo, código.

La investigación documentológica centrada en aplicaciones no busca, en principio, la extrapolación o la compatibilidad con otras áreas del conocimiento, todo lo más, adopta o hereda, sin someterlos a observación, esquemas de operación que han tenido éxito o renombre en las disciplinas más consolidadas. No obstante, dentro de las Humanidades ha comenzado un cierto movimiento en búsqueda de propios modelos documentológicos que permitan generar métodos de lectura y representación específicos del área. Esta idea de generalizar algunos métodos documentales en diversas disciplinas sociales y humanas fue defendida por Cunha¹⁹ y contraargumentada por Kobashi²⁰. En efecto, la

¹⁹ Cunha, I.: Do mito...o.c.. Aplica sus propuestas sobre un texto de la política colonial portuguesa, en los niveles de análisis y construcción de vocabulario.

²⁰ Nair Kobashi discute las propuestas de Cunha detectando inconsistencias en su trabajo *Análise documentária: considerações sobre um modelo lógico-semântico*. In: Cunha, Isabel (org.): *Análise documentária: considerações teóricas e experimentações*. Sao Paulo: Febab, 1989, p.45-57

propuesta, en ese caso, de un método de análisis y construcción de vocabulario sólo era válido en el propio corpus utilizado por la autora.

En el mismo sentido abierto, de buscar extrapolaciones desde el discurso humanístico, hemos partido para la concepción de este trabajo, hecho que se comprueba en los modelos de análisis semántico descritos en el cap 4. Pero como podrá verse, pronto esta idea fue abandonada ante la realidad del corpus tomado en consideración aquí: el discurso periodístico. La distancia entre ambos es tal, que pocos elementos comunes son utilizables en el actual estado de conocimientos, llegando incluso a pensar en la mayor facilidad de extrapolación de las ciencias exactas a las Humanidades, que de éstas al discurso periodístico, tan decisivas son las variables de estructura y lenguaje para la generalización.

Por lo tanto, el estudio de las condiciones de producción de un discurso y de la estructura del mismo es esencial para proponer métodos de trabajo, aunque nos encontramos ahí en la misma frontera de la Documentología, desbordando el objeto puesto que, en pureza, pertenecería al objeto de investigación el estudio de las condiciones de producción del discurso documental o generado por documentalistas y usuarios dentro del mismo sistema y no los fenómenos que inciden sobre la producción discursiva que son objeto de la Semiótica textual o del Análisis del Discurso.

Procede, así y una vez más, un estricto recorte en el área de aplicación que nos interesa acudiendo a las disciplinas que, no siempre, tienen resueltos problemas que afectan directamente a la investigación y la práctica documentales.

Podemos proponer, grosso modo, una lista tentativa compuesta por grandes áreas discursivas en las que se practicarían, internamente, los primeros intentos de extrapolación:

- Discursos lógico-formalizados: Ciencias experimentales, tecnología y técnica.
- Discursos lógico-naturales: Ciencias sociales y humanas.
- Discursos enciclopédicos: Información de actualidad, Cultura, Historia, legislación.
- Discursos estéticos: Artes.

CAPITULO 2

2. CUADRO REFERENCIAL TECNOLÓGICO

Entendemos por cuadro referencial tecnológico, el conjunto de equipos existentes y potencialidades previstas que constituyen las condiciones de un sistema en el que pretendemos hacer viable la metodología del AD mediante simulación. La presentación de este universo paramétrico es necesaria para sostener algunos principios formulados en torno a la lectura empírica humana en relación con el reconocimiento artificial del texto.

Superado el carácter rudimentario de la lectura lineal y de la coincidencia de significantes se presumen dispositivos capacitados, mediante programas de autoeducación o autoconducción y memorias con experiencia utilizable para la inferencia, para la lectura en contexto y la comunicación con el usuario en el nivel de estructura profunda de los textos y demandas, lugar de significación en el que debe producirse la coincidencia. Estas exigencias descartan los soportes tradicionales tanto físicos como lógicos. Por ello, y en favor de nuestro interés de poder proponer y desarrollar, junto a tecnólogos, nuevos modelos de análisis documental por ordenador inteligente, descubrimos dos entornos investigados en la actualidad: la Inteligencia Artificial y los sistemas expertos acercando sus aplicaciones posibles hacia la gestión de textos con fines documentales.

Evidentemente no pretendemos, por desconocimiento, criticar los logros y proponer soluciones tecnológicas. Para ello, tendríamos que estar y conocer la vanguardia tecnológica de este campo, muchas veces alto secreto, y disponer de costosísimos equipos con los que no cuenta la Documentología en la Universidad. Nos conformamos por el momento, a pesar de la necesidad urgente de experimentación, con el acceso vicario a las habilidades que nos cuentan de la IA y al uso de unos pocos sistemas expertos conocidos más de cerca para poder, desde la reflexión teórica, formular cuáles son los problemas

de la Documentación, como pueden ser abordados y mediante qué métodos, apuntándose los requisitos mínimos que debe cumplir la tecnología para satisfacer las necesidades de la Documentología.

Doblegarnos a los avances de la tecnología de hoy supone hipotecar una investigación teórica que ofrezca resultados dentro de pocos años. Por ello, considero esencial una cierta "desvinculación" de las limitaciones actuales para la realización de propuestas. A buen seguro tendremos que efectuar importantes recortes para hacer compatible nuestras propuestas con la reducción que emana de cualquier autómata que simula decisiones.

2.1 Inteligencias artificiales

Aún estamos lejos del día en que una computadora inteligente pueda sustituir al hombre en actividades en las que existe una fuerte presencia de tomas de decisión, actitud posible a partir de la inferencia lógica ya desarrollada en algunos sistemas expertos, especialmente si para ello se necesitan ciertas variables basadas en la improvisación más cercana a la intuición y al juicio razonado. Pero, exceptuando los sentimientos, no es difícil concebir teóricamente, basándonos en el estado actual de desarrollo tecnológico y sus tendencias, máquinas que ejecuten tareas atribuidas con exclusividad al ser humano hasta el momento. Y esto incluye una tarea tópicamente humana como es la lectura y, por tanto, el análisis de texto. Debo señalar que, la lectura en documentación es de tipo empírico, atiende a unos objetivos específicos y debe obtener resultados análogos por parte de distintos lectores, lo que la hace, en principio, programable.

Ya se han dado los primeros pasos en indización automática, aún rudimentaria, y traducción automática hacia la delegación de los métodos de lectura, interpretación y representación humanos en una máquina que capta linealmente signos y espacios que codifica y decodifica en señales eléctricas millones de veces para un simple texto. El avance es pequeño pero el horizonte abierto por esta tendencia de la investigación tecnológica ofrece amplias posibilidades. Es justamente en este punto en el que la Documentología debe adelantarse y realizar propuestas, principios y descripciones. A partir de éstas, los tecnólogos habrán de intentar la transportabilidad electrónica o debatir, en caso inviable, las modificaciones de forma y estructura de los modelos propuestos que no afecten a los objetivos documentales.

En 1990, con el mercado saturado de ordenadores de cuarta generación, puede contarse con microprocesadores que realizan diez millones de operaciones por segundo, llegándose a treinta millones en los procesadores más avanzados. Estos constan de un circuito que efectúa las operaciones secuencialmente produciéndose mayor lentitud pero existen, desde hace algún tiempo, ordenadores que han logrado funcionar con 65.000 microprocesadores interconectados realizando doscientos millones de operaciones por segundo¹.

La 5ª generación la constituyen unos sistemas en vías de desarrollo, aunque muchas de sus funciones ya han sido aplicadas por los superordenadores y sistemas expertos, denominados Inteligencia Artificial (IA).

La IA emula el modo de pensar humano, para lo que debe cumplir dos condiciones:

- posibilidad de ejecutar operaciones en paralelo, logrando que los resultados de una deducción influyan, sobre la marcha, en otras operaciones.
- naturalización de la comunicación con la máquina, es decir, uso de recursos de la lengua natural tanto en interrogación como en respuesta. Obsérvese que este es uno de los objetivos que hemos planteado para la Documentación en sus aspectos analíticos y lingüísticos.

El LISP (USA) y el PROLOG (Europa) ya apuntan, aunque tímidamente, estas funciones. Desde hace una década, Estados Unidos, Europa y Japón han dado un fuerte impulso a las investigaciones en IA: En 1982 se crea en USA la MicroComputer Corporation, Instituto de investigación que agrupa una veintena de Empresas tecnológicas civiles que unen sus esfuerzos a la investigación militar en IA emprendida por la Institución federal Defense Advanced Research Projects Agency (DARPA)². La CEE contribuye al mismo campo a través de su programa ESPRIT (European Estrategic Program Research on Information Technology).

El objetivo de la IA lo describe J. Pazos³: “una máquina capaz de entender la inteligencia natural humana, adquirir conocimientos y resolver problemas

¹ Datos aportados por Colle, Raymond: *Tecnologías de la información*. Santiago de Chile: Pontificia Universidad Católica, 1988, 183 p.

² *Idem*, p.56

³ Cit. por Colle, R. *ibid.*, p. 120

considerados intelectualmente difíciles". Comprender la inteligencia humana es, desde luego, entender las producciones humanas entre las que incluimos los textos. Sin embargo, la comprensión es un concepto complejo y bastante subjetivo: el diálogo interpersonal entre dos individuos de la misma comunidad lingüística no da siempre como resultado una satisfactoria comunicación de ideas. Lo mismo ocurre con la lectura de textos que ya nos hizo concebir el documento como elemento subjetivo de comunicación (la objetividad de su existencia sólo se la concede el soporte material y esto no es relevante para el análisis de contenido)⁴.

La comprensión, parafraseando a Schank, es una gama de grados y niveles relativos: En un extremo encontraremos lo que entendemos por compenetración o comprensión en su grado más alto; en otro, tan sólo percibiremos una localización de sentidos⁵. La total afinidad sólo se daría entre individuos o máquinas con estructuras intelectuales, ideológicas, culturales y memorísticas, entre otras, muy similares. Es el aprendizaje empírico uno de los elementos que coadyuvan al entendimiento del otro. De esta forma, y aún separados, serán capaces de reaccionar y tomar decisiones parecidas ante acontecimientos idénticos.

Parte de este aprendizaje humano es individual y parte social. La captación del entorno determina el bagaje pedagógico del niño. Este aprendizaje se realiza fundamentalmente a través del lenguaje. Por ello, Chomsky y otros han contribuido con teorías especulativas acerca del proceso de captación de la naturaleza en el niño⁶ en trabajos, de corte más filosófico que lingüístico ortodoxo, que contribuyen a la configuración de la IA puesto que serán cerebros vacíos que hay que llenar a través de instrucciones y datos. Según Schank las IA deberán conformarse con encontrar sentido a través del reconocimiento de palabras, enunciados y contextos comparados a su experiencia acumulada⁷. Entre dos personas, en cualquier caso, es el nivel de comprensión habitualmente alcanzado. La afinidad total sólo se conseguiría dotando a la IA de sentimientos lo que para el investigador de Yale resulta imposible y del todo inconveniente.

⁴ La información y su transmisión y no el soporte del documento, aún teniendo en cuenta su importancia, es lo que da entidad al objeto y a la teoría documentológica: la captación, interpretación, segmentación y uso de los mensajes deben ser los campos de investigación que consoliden un estatuto científico autónomo para la Documentología.

⁵ Schank, R.: *El ordenador inteligente*. Barcelona: A. Bosch, 1986, p. 55 ss.

⁶ El investigador del MIT explica la facilidad del niño en el proceso de aprendizaje de la lengua a través de la teoría del innatismo. Chomsky, Noam: *Reflexiones sobre el lenguaje*. Barcelona: Planeta/Agostini, 1985, 278 p., en especial el cap. 4.

⁷ Schank, R. *ibid.*

La comprensión es un concepto esencial para el AD. La podríamos definir en el plano operativo como el núcleo de participación receptiva obtenido empíricamente como resultado de la lectura de varios documentos por varios analistas y sujeto a condiciones relativas. Los nuevos métodos analíticos y su proceso automático nos ayudarán a reformular lo que ahora entendemos documentológicamente por comprensión. Schank espera que para el año 2000 existan IA situadas en un nivel intermedio de la cadena sentido-afinidad, a la que denomina comprensión cognoscitiva. Una máquina con ese nivel de inteligencia podría realizar las siguientes operaciones:

- Aprender o modificarse como resultado de sus experiencias.
- Relacionar experiencias presentes con las pasadas de manera inteligente, a través de un proceso de acribia y utilidad.
- Llegar a conclusiones a partir de su propia experiencia.
- Poder explicar el proceso de razonamiento utilizado para llegar a una conclusión⁸.

Es evidente que para ello, los expertos deben acordar los árboles de decisión que ellos mismos adoptan en la lectura y obtener, a partir de métodos comunes, resultados comunes. Si esta primera fase, que podríamos llamar de lectura homogénea, no se alcanza en el nivel humano será imposible acceder a la segunda: emulación automática de comportamientos humanos ante el texto. Esto no implica, sino todo lo contrario, que los parámetros de los métodos de lectura válidos para la aplicación humana sean válidos para la simulación o viceversa.

2.1.1 *El acceso al significado en IA*

Para Colle, es importante reconocer la función gramatical para reconocer el significado; sin embargo, esto no es fundamental: el niño aprende gramática cuando ya sabe manejar los significados, elige lo que quiere decir y después busca la forma sintáctica de expresarlo. Este problema es el que han estado

⁸ Idem, p.50-60

involucrados grandes filósofos del lenguaje como Buhler o Chomsky aún está, a nuestro entender, en un ámbito que sólo se presta a la especulación y, en cualquier caso, no es extrapolable al caso de una IA que paralelamente a su instrucción con datos o significados también debe ser alimentada con reglas. No entendemos aquí por significado, el concepto referido por una palabra sino en el sentido más amplio: cualquier idea divisible o no. Ahora bien, el problema de la transportabilidad de la categorías léxicas, sustantivos en verbos o adjetivos y viceversa y las múltiples formas de parafraseo constituyen los ejes básicos de la investigación sintáctica aplicada a la tecnología. Esto confirma nuestra tesis sobre el rechazo a los métodos basados en la coincidencia formal: la coincidencia debe darse en el nivel funcional y de estructura profunda, cualquiera que sea la representación en el nivel de superficie⁹ y hacia ese objetivo deben dirigirse las investigaciones de Documentología aplicada, prioritariamente.

Debemos presumir que el significado no se integra en la memoria como un mero repertorio léxico sino formando campos conceptuales interrelacionados entre sí a través de redes semánticas construídas de la manera más flexible que posibiliten el sentido de acuerdo a la situación del texto, incluyendo polisemias o eufemismos. Así, por ejemplo, habrá que explicarle al ordenador que /ilustración/ se refiere a periodo histórico o a dibujos, enseñarle múltiples contextos con un sentido u otro para que esta experiencia le ayude al autoaprendizaje. Es una tarea ímproba pero plausible de carácter conceptual, semántico y morfosintáctico por un lado e inevitablemente cultural e ideológico, por otro: cualquier concepto tiene sentidos connotativos y denotativos. En la connotación se producen las mayores derivaciones de la significación. La IA debe ser capaz de generar distintas formas de los mismos significados lo que posibilitará comprender las distintas maneras enunciativas de un interlocutor humano de un texto leído. Se llega así a la creación de "modelos de conocimiento" que en IA orientados a la comunicación deben contar con analizadores o desestructuradores de preguntas y generadores o montadores de oraciones gramaticalmente correctas. En medio de estas rutinas actúa el sistema de inferencia empírica. El esfuerzo previo para estructurar la información, desde la lengua natural, en las primeras etapas, es evidente. Hacia esa labor habrá de orientar su técnica el nuevo documentalista.

⁹ García Gutiérrez, A.: Estructura lingüística...o.c., p.81 s.

2.1.2 *Primitivas, guiones y escenarios*

Una IA debe reconocer significantes, normas y significados, leerlos en contextos variables y realizar deducciones e inducciones para ofrecer resultados. Estas operaciones efectuadas instantáneamente por el hombre (sin autoexplicarse sus fases) requiere un laborioso proceso de programación y esto en el caso de hardwares liberados de las limitaciones actuales. La IA tendrá que ser alimentada no sólo de unos significados sino también de un sistema relacional entre ellos que permita flexibilidad en la interpretación. Sin embargo, ese sistema lingüístico necesitará todos los contextos posibles en los que puede funcionar un significado o combinación de ellos y las situaciones o hechos que se desprenden en relación a todos los contextos. Así, la IA debería ser capaz de reconocer /propina/ y ubicar el concepto en sus posibles categorías: sobresueldo, hostelería, caridad, paternalismo, los aspectos económicos o sociales derivados, etc. y entender cualquier situación en la que aparece el concepto: dinero que se da (o recibe) a alguien por ejecutar un servicio además del precio estipulado: botones, camareros y otras profesiones.

La IA habrá sido instruída ideológicamente (según los principios de la época o sociedad en la que se insertan sus educadores o expertos) para valorar positiva o peyorativamente lo que representa /propina/. Solamente con toda esta información sobre un concepto o idea, una máquina estaría en condiciones de dialogar con el hombre. Las preguntas y respuestas del interlocutor debe registrarlas y efectuar una reconducción de sus razonamientos. Pero la dificultad que encuentran los científicos para desarrollar el autoaprendizaje inteligente y la vulnerabilidad que presentan las IA al comportarse como ingenuos cerebros que lo asimilan todo, hacen que los más pesimistas duden sobre sus resultados.

Ahora bien, una cuestión de fondo planteable como elemento esencial para el abordaje empírico de un modelo de análisis inteligente es la validez de las pruebas específicas o parciales en corpus mayores, es decir la universalidad de los modelos (en el sentido de generalización en discursos distintos pero inmediatos, puesto que no creemos en la posibilidad de hacerlo para todo discurso), o la bondad de los estudios generales y sus resultados en enunciados particulares, es decir la aplicabilidad del método. El primer aspecto lo veremos al describir algunos métodos de análisis testados sobre textos particulares más adelante. En cuanto al segundo aspecto es importante introducir aquí el concepto de categoría primitiva¹⁰. Se ha demostrado que bastan once primitivas

¹⁰ Schank, R. o.c., p. 118

para abarcar y representar todas las acciones posibles mediante combinaciones: una primitiva puede representar el hecho de obtener, recibir, adquirir, otra, aplicar fuerza a objetos, etc. Cualquier significado o situación estaría encadenado a una de estas grandes células del pensamiento. Las clasificaciones jerárquicas ya aprovecharon este recurso de englobamiento sobre temas generales. Las IA seguirían un movimiento pendular que pasaría por miles de estratos distintos desde la estructura de superficie hasta la primitiva y viceversa. Debe ser estudiada, antes de lanzarse por esta vía en documentación, cuál es la influencia de los microcontextos en la estructura apriorística de las primitivas ya que como conjetura pragmática podríamos pensar que los problemas que genera un dispositivo arborescente en clasificación pueden aparecer también en el cuadro intermedio de primitivas.

Cada acción, prevista en alguna primitiva, se desarrolla a partir de una situación distinta y la IA debe reconocer todas ellas. A estas concreciones se les denomina "guión". A los guiones complejos se les llama "escenarios": todo lo que puede ocurrir en un taxi, en una reunión, en el cine, todo lo que el ser humano considera rutina debe ser administrado a la máquina: lo que nosotros conocemos por experiencia la IA debe asumirlo teóricamente: se excluye, en consecuencia, la capacidad de improvisación que tiene el ser humano aunque se liberan las situaciones de la condición de "repetitivas" en la tecnología tradicional. Las IA pueden decidir sobre situaciones si las reconoce realizando "reflexiones".

Los programadores de una IA deben prever cómo va a ser interrogada por el hombre. Por ello, no sólo deben ser nutridas con perfiles psicológicos humanos y con cuestiones estandar sino también con facilidades para el reconocimiento de cómo van a ser abordadas determinadas cuestiones (un inventario de interrogantes y sus reglas). La imposibilidad de universalizar una memoria de IA en este sentido para resultando cada vez más evidente. Por tanto, parece más factible realizar desarrollos parciales con guiones y preguntas controladas dentro de campos condicionados. Según Schank, si no queremos que las IA se pierdan en cuestiones irrelevantes será preciso que conozcan los objetivos que tienen las personas y los planos que trazan para lograrlos ¹¹.

¹¹ Idem. p.133

2.2 Sistemas expertos

Una variante de la IA, a la vez que propulsora de las indagaciones en la misma, es el Sistema Experto (SE). Sin pretender profundizar aquí sobre un campo que conoce una prolífica bibliografía técnica ¹² y justificando nuestro interés en sus aplicaciones y modos de estructuración cognitiva, pasaremos a exponer su arquitectura, objetivos, posibilidades e incapacidades para poder contar con una marco tecnológico más cercano en el que poder intuir la simulación de nuestras propuestas. En consecuencia, mi interés por los SE, y por la tecnología en general, no es más que el de desarrollar modelos de análisis y lenguajes documentales para evitar algunas tareas rutinarias humanas a la vez que, por qué negarlo, adquiero nuevas perspectivas que permiten remodelar algunos métodos. En ese sentido, la tecnología influye positiva y directamente en las reconducciones de la investigación aplicada.

El SE es una expresión polisémica que designa, por una parte, los generadores de SE, softwares que funcionan en ordenadores convencionales capaz de simular los procesos de raciocinio humano y, por otra, al propio campo del conocimiento organizado por los especialistas en una máquina. La presencia del especialista es esencial para aportar al sistema sus conocimientos, experiencia y métodos de razonamiento en un área específica (la presencia de varios especialistas con criterios contradictorios anularía la capacidad decisoria del sistema y, por tanto, ese tipo de discurso queda, por ahora excluido, de los SE). Surge, con ello, un nuevo profesional del área de Ingeniería cognitiva: el cognicista (galicismo) cuya labor consiste en entrevistar, conducir y traducir los datos suministrados por el especialista en un lenguaje sintético entendible por el sistema. De esa manera, al interrogar al SE obtendremos respuestas y argumentaciones (explicitables si son solicitadas) ofrecidas por la máquina y que no son otras que respuestas de los expertos que la programaron. La educación del SE se realiza en LN, admitiéndose todo tipo de información: hechos, datos, deducciones, dudas, etc.

Los SE se componen de una base de conocimiento (BCON) y de un motor de inferencia (MINF). En la BCON se encuentra la base de hechos o datos (BD) en la forma que hemos descrito antes y construída a semejanza del experto.

¹² Véanse, por ejemplo, dirigidos a nuestras aplicaciones, los libros de Wormell, Irene: *Knowledge Engineering*. London: Graham Taylor, 1987 y en la comunicaciones del congreso sobre *Image et Intelligence Artificielle dans l'Information scientifique et technique*. 6 au 10 juin 1988. Bénodet (Francia). Paris: Inria, 1988, 189 p.

Pueden organizarse los datos en función de estructuras más amplias denominadas facetas, esquemas o frames. Por otro lado, la BCOM cuenta con una base de reglas (BREG), elaborada en función de la lógica natural de razonamiento empleada por el mismo experto en su dominio de conocimiento. Pueden crearse, de manera más compleja, conjuntos de meta-reglas para organizar la base de reglas determinando, por ejemplo, el orden de intervención de las reglas cuando hay varias implicadas en el mismo razonamiento. Las reglas se construyen con un lenguaje sintético y una sintaxis que simplifican en exceso comportando importantes pérdidas de matices y de información en el proceso. Según sea modificable la BCOM durante el transcurso y después de un proceso, podremos clasificar los SE en monótonos (no cambian ni los datos ni las reglas ya que no hay influencia, a corto plazo, de los cambios en el área de conocimiento) utilizados, por ejemplo, en diagnóstico médico, meteorología y otras áreas cognitivas estables, y no monótonos (cambian los hechos y/o las reglas con la experiencia, a decisión del gestor del sistema) para áreas que precisan reflejar rápidamente sus cambios en el sistema. Es obvio que el discurso periodístico iría encuadrado en este tipo de generador.

Veamos, como ejemplo gráfico, una base de reglas tomadas de un sistema experto sobre derecho penal del informe de Desicy, A.E.: DESS Intelligence artificielle. Univ. Paris VI, 1985-86 presentado en el curso sobre Inteligencia Artificial de Bénodet (Francia) del 6 al 10 de junio de 1988¹³. Téngase en cuenta, para evaluar el ejemplo, que el sistema no utiliza variables, que es monótono, que los atributos pueden ser mono o polivalentes y que las proposiciones son positivas o negativas.

EJEMPLO DE BASE DE REGLAS SOBRE EL TEMA DE LA JUSTICIA PENAL

Regla 1

SI homicidio
Y acto voluntario
ENTONCES tipo de homicidio ← voluntario

Regla 2

SI homicidio
Y NO acto voluntario
ENTONCES tipo de homicidio ← involuntario

¹³ Cit. por Joab, Michelle: Les systèmes experts. In: "Image et Intelligence..." o.c., p.130-131

Regla 3

SI crimen
 ENTONCES tipo homicidio ← voluntario

Regla 4

SI crimen
 Y víctima = ascendiente legítimo
 ENTONCES parricidio

Regla 5

SI tipo de homicidio = voluntario
 ENTONCES crimen

Regla 6

SI parricidio
 ENTONCES crimen
 Y víctima <= ascendiente legítimo

Regla 7

SI infanticidio
 ENTONCES crimen
 Y víctima <= niño

Regla 8

SI crimen
 Y acechanza
 ENTONCES asesinato

Regla 9

SI crimen
 Y acto premeditado
 ENTONCES asesinato

Regla 10

SI culpable = madre
 SI víctima = hijo
 ENTONCES caso madre-hijo

Regla 11

SI crimen
 Y caso madre-hijo
 ENTONCES veredicto: reclusión ← de 10 a 20 años

Regla 12

SI asesinato
 Y caso madre-hijo
 ENTONCES veredicto: reclusión ← de 10 a 20 años

Regla 13

SI envenenamiento
Y caso madre-hijo
ENTONCES veredicto: reclusión ← de 10 a 20 años

Regla 14

SI privación de patria-potestad
Y edad de la víctima < 15
Y culpable = ascendiente
ENTONCES malos tratos

Regla 15

SI culpable = ascendiente legítimo
ENTONCES culpable <= ascendiente

Regla 16

SI culpable = madre
ENTONCES culpable <= ascendiente legítimo

(En la traducción del francés pueden producirse desajustes respecto al código penal español).

A partir de esta base de reglas se puede construir un diccionario de proposiciones utilizadas en las reglas. Obsérvese la presencia de negaciones en partes de la premisa o en partes del consecuente. También podemos generar un diccionario de atributos:

DICCIONARIO DE PROPOSICIONES

ENUNCIADO	PREM.POSIT./ CONSEC. NEG.	PREMISA NEGAT.	CONSEC. POSITIVA
Homicidio	R1, R2		
Acto voluntario	R1	R2	
Crimen	R3,R4,R8,R9,R11		R5,R6,R7
Parricidio	R6		R4
Infanticidio	R7		
Acechaza	R8		
Asesinato	R12		R8,R9
Acto premed.	R9		
Caso madre-hijo	R11,R12,R13		R10
Envenenam.	R13		
Privación patria-pot.	R14		
Malos tratos			R14

DICCIONARIO DE ATRIBUTOS

ATRIBUTO	MONO	MULTI	PREMISA	CONSECU.
Tipo homicidio	SI	NO	R5	R1,R2,R3
Víctima	NO	SI	R4,R10	R6,R7
Culpable	NO	SI	R10,R14,R15,R16	R15,R16
Veredicto: reclus. crimin.	SI	NO		R11,R12,R13
Edad de la víctima	SI	NO	R14	

El MINF es el elemento de unión de datos y reglas y se activa al ser solicitado por el usuario, generalmente, mediante preguntas condicionales (si...entonces...) produciendo respuestas distintas a las ofrecidas por las bases de datos tradicionales: se obtienen decisiones y no hechos o referencias.

Otros sistemas incorporan parámetros ad hoc como, por ejemplo, el I3R (*Intelligent Intermediary Information Retrieval System*), sistema experto de recuperación de información científica que cuenta con una base de perfiles de usuario que modeliza, desde la utilización, el sistema inferencial del motor ¹⁴.

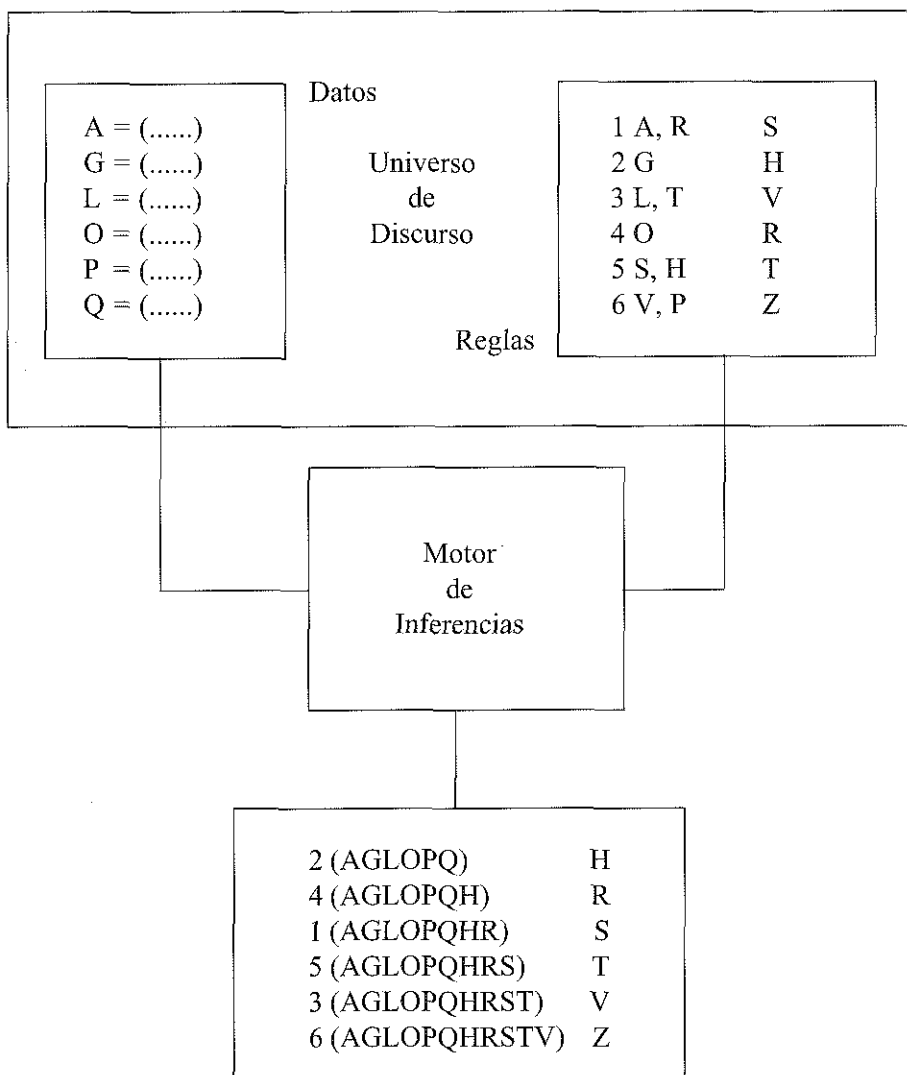
Atreviéndome a simplificar con la justificación de que los SE u otra tecnología, insisto, no son objetivo de este trabajo, paso a describir, por imperativos del corpus en discusión, cómo se representa el conocimiento en SE. Es fácilmente observable que las aportaciones de los SE a un sistema de documentación suponen nuevos productos (deducciones, ayuda a las decisiones, etc.) que complementan los productos clásicos (síntesis, datos, etc.) pero, desafortunadamente, no por una misma máquina, es decir, la estructuración a que es sometida la información en este caso determina un tipo de producto de interés documental pero incompatible con otros igualmente necesarios par el usuario que se obtienen de estructuraciones diferentes de la información.

La BD contiene objetos, hechos o fenómenos empíricos sobre los que actuará un proceso desencadenado de inferencias intermedias y finales. Los datos se digitan de manera simplificada, tanto en lenguaje y formato como en cuanto a su estructura proposicional: entidad-atributo, objeto-predicado, atributo-valor, etc. que son categorías que decide el experto para organizar su información. Tanto las categorías como los propios datos son reconocibles por el sistema en todo momento y se definen circularmente hasta el último nivel: X es presidente, Presidente es un cargo político en un país, un país es un espacio geográfico con fronteras (para otros, país puede tener un valor distinto, pero la polisemia causa problemas en el sistema).

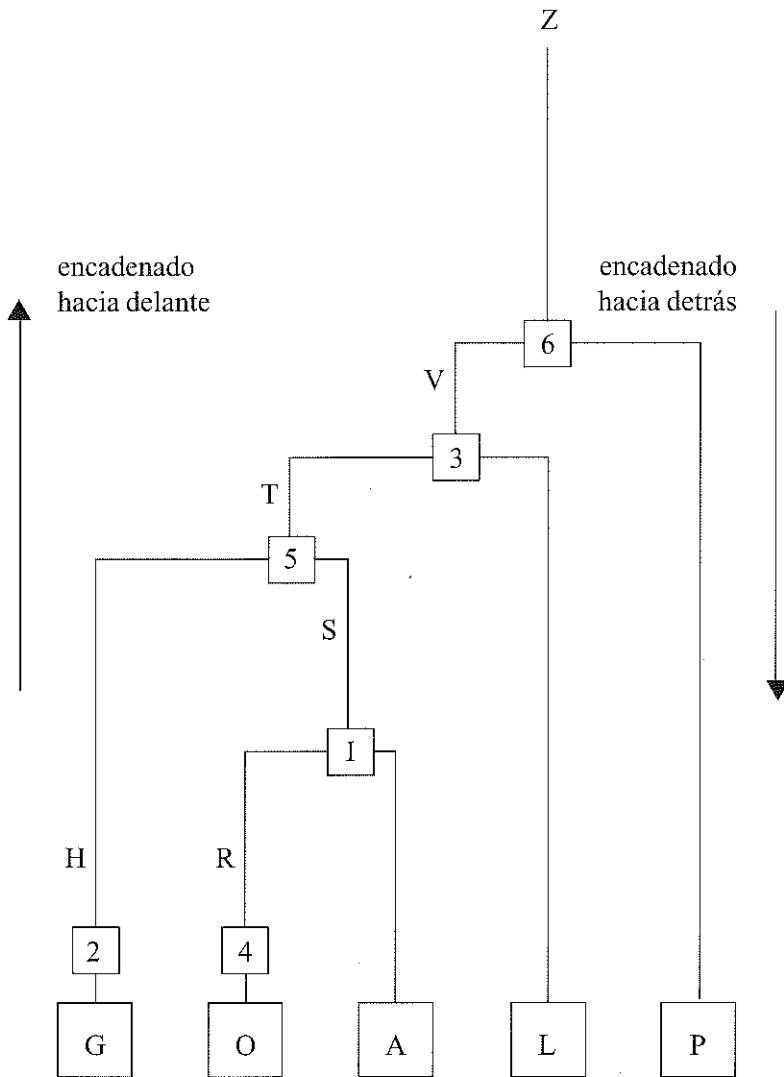
Las reglas también son predefinidas mostrando la siguiente estructura: Si $p \text{---} \text{---} > q$ (si entidad, relación, atributo). q es consecuencia de p en el sentido más abierto (deducción, implicación, inferencia, etc.) y se denomina consecuente o "parte derecha" de la proposición. p se denomina antecedente o "parte izquierda". En los SE pueden producirse inferencias en dos sentidos que se llaman encadenados a la derecha o hacia adelante, dando como resultado procesos deductivos, y encadenados a la izquierda o hacia atrás, que generan procesos inductivos. Veamos, en dos gráficos, el esquema del SE y la actuación de un motor de inferencia, usando símbolos tomados de Gardin:

¹⁴ Croft, W.B. and Thompson, R.H.: I3R: A New Approach to the Design of Document Retrieval Systems. "JASIS", 38 (6), 1987, p. 389-404.

BASE DE CONOCIMIENTO



Formación de un razonamiento a partir de una base de conocimientos que contiene los datos AGLOPQ y seis reglas de producción en desorden (las cifras 1-6 no son más que una manera de designación), bajo el control de un motor de inferencia. Mediante la aplicación sucesiva de las reglas 2, 4, 1, 5, 3, 6, el motor establece los datos nuevos H R S T V y, finalmente, Z, que marca la culminación del proceso¹⁵.



Representación del mismo razonamiento bajo forma arborescente, leído bien de abajo hacia arriba y de izquierda a derecha (encadenado adelante; reglas aplicadas en el orden 2, 4, 1, 5, 3, 6), bien de arriba hacia abajo y de derecha a izquierda (encadenado atrás, reglas aplicadas en el orden 6, 3, 5, 1, 4, 2)¹⁶.

¹⁵ Esquemas y ejemplos extraídos de Gardin, J.C. et al.: *Systèmes experts et Sciences humaines*. Paris: Eyrolles, 1987, p. 34

¹⁶ Idem, p.35

En general, se parte de hechos iniciales para llegar a conclusiones producto de la inferenciación. El camino inverso también es posible: se parte de unas hipótesis, por ejemplo, el conjunto de proposiciones designado por Z; el sistema busca una regla que contenga Z como consecuente para encontrar un antecedente posible. La operación se repetirá, por una especie de arrastre ascendente, hasta llegar a las deducciones GOALP de la hipótesis inicial, que son elementos de la base de datos. La hipótesis queda verificada por el sistema. El motor de inferencia consta de un compilador de hechos y reglas en un diccionario de términos con códigos de entidad, relación, valor, etc. verificador de sintaxis y un mecanismo de control que determina: reglas a desencadenar, justificación de la elección de una regla, ejecución y actualización de la base de datos (si el sistema es no monótono).

Para Shank, el propio nombre de experto para el sistema es confuso ya que, como se ha dicho, la respuesta es del instructor, lo que en un sistema de masas pueden incurrir en graves manipulaciones de la opinión. El SE no tiene capacidad para elaborar raciocinios que no fueran previstos aunque sí combinaciones entre ellos. Por lo tanto, se muestran altamente operativos en campos aplicados de alta objetividad como el seguimiento de síntomas, la geología, etc. En Ciencias humanas han dado algunos resultados en Arqueología, según demuestran diversos estudios del equipo de trabajo en torno al arqueólogo francés Gardin al testar el sistema SNARK (Symbolic Normalized Acquisition and Representation of Knowledge)¹⁷ y en la construcción automática de resúmenes de textos breves y preparados, según se desprende de los resultados conseguidos en la Universidad de Yale al evaluar el sistema FRUMP y CYRUS¹⁸, aunque con numerosos problemas y limitaciones para el tratamiento de textos integrales.

Tradicionalmente, los lenguajes de programación utilizados para la construcción de generadores de sistemas expertos son, como se ha dicho, LISP y Prolog. Ambos obligan al modo declarativo en las proposiciones, por lo tanto, se dice a la máquina qué hacer en lugar de cómo hacerlo. Según Cunha y Kobashi, actualmente se está trabajando con el lenguaje C que permite instruir al ordenador sobre cómo realizar las operaciones (modos de procedimiento)¹⁹ lo que hace surgir nuevas perspectivas para la aportación de métodos simulados.

¹⁷ Véanse algunas experimentaciones con SNARK en o. c. realizadas por colaboradores de Gardin.

¹⁸ Descritos por Raymon Colle en o. c., p. 143-151

¹⁹ Dato extraído de Schildt, M.: Inteligencia artificial utilizando linguagem C. Sao Paulo: Mc GrawHill, 1989 cit. por Cunha, I. y Kobashi, N.: o.c., p.5

Como se ve el panorama ofrece importantes soluciones pero también nuevos problemas. La tendencia general se dirige hacia los dominios aplicados lo que hace peligrar la simulación del discurso periodístico aunque este campo sería uno de los máximos beneficiarios de un sistema experto eficaz, no sólo por librar a los documentalistas de prensa de una gran cantidad de rutinas diarias sino porque como afirman Cunha y Kobashi, citando a Souza, "la exhaustividad de los bancos de datos compromete la especificidad. Por ello, la tendencia actual de la Documentación instauro como objeto no ya el documento o la información sino el conocimiento (con base en la información y en el documento), un conocimiento que se pretende específico y exhaustivo"²⁰.

Teniendo, finalmente, presentes las indudables ventajas de los SE, como nuevo marco referencial tecnológico, justificativa de su cita en este trabajo, señalemos también sus principales lagunas e incapacidades generales en lo que se refiere a la simulación del discurso periodístico:

1. Las proposiciones en modo declarativo son insuficientes cuando no nocivas.
2. Los buenos resultados se han logrado con representaciones de objetos (medicina, piedras, ánforas, etc) pero se intuyen diferentes y serios problemas para el control de conceptos abstractos.
3. La base de datos y hechos debe ser pequeña, crecer poco y altamente especializada, omitiendo opiniones no divergentes o de divergencia controlada. Estas condiciones invalidan el uso de los SE actuales sobre corpus de noticias a menos que sean reducidos y con objetivos experimentales.
4. Los productos de un SE son conclusiones, por lo tanto, debe esperarse un sistema que integre en salida: inferencias + referencias + datos factuales + comprobaciones de verdad y verosimilitud.

La intervención ineludible de los expertos en la construcción de la base de reglas produce ciertos rechazos sobre la objetividad en el proceso de inferencia aunque las disputas de expertos pueden ser reflejadas en determinados casos o eliminadas por acuerdo. Entiendo que es inevitable la ideologización del proceso de lectura inteligente pero ¿acaso no lo es cuando los lectores son documentalistas? No podemos pretender que una máquina creada por el hombre

²⁰ *Ibíd.*, p. 24. La cita es del libro de Souza, B.: *Um discurso sobre as Ciências*. Lisboa: Afrontamento, 1988

supere los problemas que el consenso de grupo no ha podido superar. Y para llegar ahí, la coincidencia de resultados en lectores humanos de documentos o respecto a un sólo lector en diacronía debe ser aumentada como paso previo al diseño de un modelo de análisis automático (para poder modelizar la base de reglas). Si lo primero no es posible (a igual sistema de lectura, idénticos resultados) la creación de un sistema automático será, siempre, insatisfactoria.

Finalmente es necesario que el motor de inferencia o dispositivo de razonamiento tenga la posibilidad de explicitarse en aplicaciones y enunciados concretos, si así es requerido, facilitando el proceso seguido al usuario. Los resultados, cualesquiera que sean, deben ser aprovechados en primer lugar por la propia máquina (en sistemas no monótonos muy avanzados) para las aplicaciones de discursos sociales. Esta necesidad de la gestión inteligente, la automodificación, choca con las arquitecturas de los equipos actuales, siendo los circuitos neuronales, basados en soportes orgánicos que superan a los conductores minerales, los que presentan mayores esperanzas en este campo.

Por ahora los SE realizan procesos de razonamiento más basados en la mecánica que en la lógica: para determinadas condiciones atmosféricas la previsión es lluvia, para ciertos síntomas la respuesta es un tratamiento o medicamento. Sin embargo, estos programas expertos están encauzados y predefinen sus propios límites: si les preguntamos acerca de los beneficios de la lluvia cuando la finalidad del SE es la previsión, o los síntomas no son habituales de una enfermedad, no se obtiene respuesta.

En el AD, el objeto ni tan siquiera puede modelizarse para favorecer un dispositivo de deducción. Solamente una IA, que existe en hipótesis y es necesaria como marco tecnológico, será vehículo de la comprensión cognoscitiva a través del reconocimiento de las estructuras del texto mediante la lectura reglada de tipo empírico.

CAPITULO 3

3. EL ANALISIS DOCUMENTAL DEL DISCURSO PERIODISTICO: CONSIDERACIONES PREVIAS

Los objetivos de este estudio son los siguientes en el plano experimental: proponer las condiciones generales para la construcción de un método específico de análisis en AD en información de actualidad, tomado este campo como el banco de pruebas más complejo y enciclopédico. Para ello debemos partir de algunos parámetros teórico-pragmáticos de trabajo que nos permitan avanzar mediante conclusiones parciales en corpus elementales en los que funcionen las condiciones debatidas para, en otros estudios posteriores, ir ampliando el espectro de aplicación con las correcciones necesarias. De acuerdo con las argumentaciones anteriores, el método perseguido debe ser capaz de generar idénticos resultados para lecturas distintas sobre los mismos textos, sea el agente lector cualquier analista, hombre o máquina. Excluimos del corpus de ejemplos utilizados los géneros periodísticos literarios y la publicidad además de cualquier documento informativo no textual por requerir diferente tratamiento. Otras condiciones específicas sobre el corpus serán señaladas más adelante. A pesar de la concreción del estudio será necesario reflexionar sobre los procedimientos evaluados que sean generalizables en otros corpus discursivos.

El objetivo del análisis de actualidad en un sistema documental es desmontar la estructura de un texto mediante criterios objetivados en un método de tal manera que su representación responda al texto que refiere y admita el mayor número posible de demandas potenciales con las que mantenga cierta afinidad. La afinidad se establece en términos de relevancia de lo representado y de la satisfacción de lo obtenido, por parte del usuario. Es preciso partir, en consecuencia, identificando cómo el productor estructura su discurso (periodístico) y cómo formulan las demandas los consumidores del sistema.

El documentalista de prensa indiza documentos entendiendo por indizar el hecho de adscribir descriptores a conceptos o temas relevantes del texto. La operación indizadora, que debe consistir en la lectura empírica del texto, se identifica y determina a priori por arquetipos de representación. Nada se sabe o se dice sobre el camino lógico que sigue el documentalista para detectar y seleccionar los segmentos explícitos o tácitos del texto o por qué aglutina unos datos dispersos confiriéndoles el rango de descriptores o ecuación de descriptores. Sin poder explicarse con un argumento sistemático este proceso cognitivo que ejecuta diariamente sobre cientos de textos, el documentalista de prensa lee sin método y tampoco se plantea la responsabilidad de su lectura en el contexto del discurso ideológico diferido a cuya construcción está contribuyendo. Los documentalistas que se ayudan de procesadores automáticos de índices no cumplen los objetivos mejor que los anteriores ya que el programa selecciona en superficie. En tanto no se proponga un método objetivado de lectura en el plano manual no será posible generar árboles lógicos que reconozca una máquina inteligente para sustituir al documentalista en su trabajo rutinario.

Para establecer la primera base crítica del método analítico partiremos del acercamiento empírico que realiza el documentalista sobre el texto periodístico: la lectura por sentido común. Existen, sin embargo, demasiadas variables que impiden considerar válido el sentido común como único instrumento con que cuenta el analista. Más próximo de la intuición que de la realidad, al menos de la construida en un texto, y dosificado de distinta forma en cada profesional, las razones que validan tradicionalmente el método intuitivo son:

- la presencia material de un texto recortado físicamente en la publicación.
- la presencia psíquica de un usuario modelo que determina lo que debe ser seleccionado y cómo hacerlo (en base a la utilidad inmediata).
- la práctica rutinaria creada por normas externas que devienen forma de hacer o modo superficial de acceso al texto.

Las variables que invalidan el mismo método, son obvias:

- distintas capacidades intelectuales de los documentalistas que producen desigualdades en los resultados de los fragmentos sobre el mismo relato en función de sus niveles de comprensión del texto y del discurso en que participan.

- distintos recursos culturales, ideológicos, lingüísticos que determinan la prevalencia de una u otra óptica sobre el texto.
- la ausencia de un cuadro de categorías y reglas contrastadas y explícitas de abordaje del texto que enmarque o justifique el uso del sentido común o de la lógica en diferentes situaciones.
- la lectura en función de un código y de una Institución.

Se ha demostrado empíricamente¹ que las variaciones de coordenadas espacio-temporales y la situación de los analistas influyen en el sentido común, relativizando y, por tanto, invalidando este procedimiento a causa de la dispersión de resultados.

Por otra parte, y sin perjuicio de delimitar lo que consideramos, a efectos documentales, como información de actualidad, no podemos englobarla en un tipo discursivo homogéneo: la noticia, empleado aquí como sinónimo general, se enmascara en distintas tipologías discursivas puras o mixtas, pero en ningún caso neutras, que es preciso identificar ya que cada una de ellas influye en el método de captación.

En este trabajo no se expondrá una amplia tipología de géneros y estilos periodísticos, por no ser objeto de la Documentación pero, sobre todo, por no ser esencial ese nivel de especificidad. Nos basta con reconocer las características generales que afectan a las clases de información. Podemos, además, englobar los géneros periodísticos en estructuras discursivas determinadas por las condiciones de producción del texto de prensa. Inspirándonos en Angenot²:

- * Discurso del saber: propio de la Ciencia en la que se objetiva el conocimiento mediante el método científico de aproximación. Este es el único discurso que puede considerarse estructuralmente neutral (no incluimos como neutral la causa o los efectos de la Ciencia) al estar contrastado desde las conjeturas iniciales a las conclusiones. El método de análisis que tenga por objeto el discurso científico debería desentrañar esa estructura repetitiva que caracteriza a la investigación científica.

¹ Experimento llevado a cabo en el Departamento de Documentación de la UCM y recogido en García Gutiérrez, A. y Lucas Fernández, R.: Documentación automatizada en los medios informativos. Madrid: Paraninfo, 1987, p.60 s.

² Angenot, M.: La parole pamphlétaire: contribution à la typologie des discours modernes. Paris: Payot, 1982 analizado por Kobashi para sintetizar una tipología discursiva que afecta al AD en Análise documentária: tipologias discursivas. In: Cunha, I.: (org.) Análise documentária: considerações teóricas e experimentações. Sao Paulo: Febab, 1989. p. 31-44

- * Discurso doxológico o persuasivo: en prensa es habitual aunque, a veces, no lo construye personalmente o conscientemente el periodista, sino el contexto o el medio. No se explicitan los objetivos últimos del texto, ni la lógica argumental y se dan por asimilados ciertos principios o pautas ideológicas, políticas o sociales con las que comulgan los lectores y, como es obvio, el documentalista.

La apariencia neutral de ciertas noticias puede inducir a invalidar la posición que mantenemos sobre la inexistencia de estructura del saber en la información de actualidad. Podemos preguntarnos qué tipo de persuasión pretende una inocua información meteorológica o deportiva: pues bien, al margen de considerar la influencia del contexto en que se puedan producir, las limitaciones físicas del medio (páginas en prensa o minutos en radio y TV) hacen que la ausencia de ciertos temas produzca la no neutralidad de los presentes. Muchas veces los temas ausentes no son temas ignorados, sino simplemente no publicados ni publicables. En todo caso, el documentalista no puede ir más allá de sus objetivos profesionales que se fundan en hacer la información de actualidad recuperable en el menor tiempo posible. De los errores cometidos habrían de dar cuenta los historiadores o los tribunales.

En las noticias podemos considerar que la falsa neutralidad se vehicula en el denominado estilo informativo, método de discurso que aparenta una exposición aséptica de los hechos (*recordemos la campaña de lanzamiento de "EL Sol" el 22 de mayo de 1990: separamos la información de la opinión*). Diferenciaremos este tipo de información de la estructurada mediante el estilo de opinión basado en los fundamentos del discurso doxológico que explicita algunas bases del razonamiento aparentando ciertas formas del discurso del saber, sin llegar a serlo por carecer de método objetivo.

En este marco discursivo es necesario contrastar diversos métodos de análisis evaluando los resultados conforme a una misma parrilla de validación. Acotaremos el estudio a la primera fase del P/PD: la desestructuración de textos periodísticos, teniendo en cuenta en todo momento las influencias y condiciones de las tres fases restantes, pero persiguiendo como objetivo prioritario el establecimiento de las condiciones de un método de análisis para la captación del sentido de las noticias.

3.1 El objeto pragmático de observación

Hemos intuído y perfilado, durante meses, el objeto pragmático de esta aplicación del AD. A través de la identificación de una serie de características, efectuada en un trabajo de 1987³, pude observar la porción de universo materializada y el modo de plasmación en el discurso de tipo enciclopédico, centrífugo y efímero que acepto denominar, en este estudio y con alguna reticencia, información de actualidad. Aquellas características habían sido filtradas por el rol observador que hube de adoptar, el de documentalista para la captación de información en función de unos objetivos, conmutado en documentólogo en cuánto a la síntesis y descripción del problema observado.

Al documentólogo no le corresponde definir la información de actualidad, tomada aquí como expresión equivalente a información periodística o noticia, sino analizar y explicar su comportamiento en documentación. Sin embargo, es preciso formular los límites de la noticia en lo que atañe al documentalista para estar en condiciones de percibir su acción y responsabilidad. Por ello, exponemos a continuación algunos de los principios generales de la Documentación validables en este caso particular:

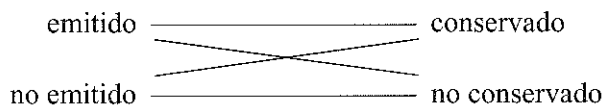
- * El objeto debe ser susceptible de recuperación: debe poder ser recuperado (capacidad del analista) y querer ser recuperado (voluntad de usuarios).
- * El soporte impone los límites de la investigación retrospectiva del documentalista, es decir, la verificación de datos, la detección de omisiones, errores o manipulaciones, las concatenaciones de unidades formando relatos, etc. se restringe a la información documentalizada (soportada y recuperable). De otro modo, el documentalista tendría que conjeturar hechos y datos suplantando el rol del periodista o del historiador, a pesar de la proximidad de las tareas.
- * Establecimiento de verdad⁴, y no exclusivamente de verosimilitud, y coherencia en la noticia fragmentada reconstruyendo el discurso de la realidad a través de otros fragmentos y de distintas fuentes.
- * Alerta permanente del documentalista como mediador y consolidador de la información convertida en conocimiento o cultura reciente y acumulada.

³ García Gutiérrez, A. y Lucas, R.: o. c., especialmente el primer capítulo.

⁴ Gabriel Galdón se extiende sobre esta cuestión en Principios operativos de la documentación periodística. Madrid: Dossat, 1989, 86 p.

- * Se restringe el corpus a la información escrita conservable. Esto no coincide siempre con lo publicado o emitido:

ejes de contrarios y análogos en documentación de actualidad



Pasemos a un análisis suficiente de la expresión información de actualidad y el concepto y perspectiva que nos interesan en Documentación para poder parametrizar la unidad mínima operativa en la materia.

3.1.1 *Reflexiones en torno a la expresión y al concepto de información de actualidad*

Emplearemos la expresión “discurso periodístico”, conscientes de las diferencias semánticas, de la apertura y complejidad de ambas locuciones, como sinónimo de “información de actualidad”. El discurso periodístico es adjetivado por su nivel de aplicación, bien entendido que tomamos aquí el periodismo escrito de carácter informativo y, por lo tanto, excluimos el periodismo literario, científico, etc. En una labor de delimitación más detallada extraemos, también, de nuestro corpus de referencia, las tiras de humor (a pesar de ser información de actualidad, su estructura y lenguaje hacen necesaria otra aproximación), la información gráfica, la información factual (horarios, divisas, bolsa, etc.) y, desde luego, los anuncios publicitarios o por palabras.

Nos queda, todavía, bastante material: el discurso periodístico escrito sobre hechos narrados en forma de textos. Esta última precisión implica publicación o, al menos, explicitación (despachos no publicados, por ejemplo, son considerados) porque existen, como dice Van Dijk al enfrentarse a una delimitación similar aunque con otros objetivos, “noticias que pueden no aparecer en los media”⁵

⁵ Van Dijk, Teun: La noticia como discurso: comprensión, estructura y producción de la información. Madrid: Paidós, 1990, p. 17.

El término "discurso" es empleado como sinónimo de texto o relato indistintamente aunque, en realidad, concebimos el discurso como una entidad superior y más abstracta que caracteriza, con reglas y estilo, el tipo de relato y texto. En las noticias da la impresión, tras un análisis superficial, que su dispersión temática impedirá cualquier tentativa de modelización y sistematización. Nuestro trabajo se basa en la presunción contraria: superficialmente es fácil identificar una noticia por el lector más profano y diferenciarla de otro tipo de texto, es decir, aún desconociendo la temática, cualquiera es capaz de reconocer y señalar una noticia. Hay rasgos empíricos peculiares que dan una naturaleza especial a este tipo de textos. Esto nos ofrece una pista para poder lanzar la hipótesis de que es posible construir un método de análisis ya que existen elementos homogéneos en la estructura de la noticia, desconsiderando como variable la importantísima pero no imprescindible temática. Bien es claro que no me refiero a una única estructura (aunque esta existe probablemente en un nivel superior, no es objetivo de la teoría del AD): es necesario detectar todas las variables estructurales desde una tipología microdiscursiva para que un AD simulado pueda reconocerlas y desencadenar y ordenar determinadas reglas de lectura según se trate de un tipo o de otro.

Existe un gran acuerdo, entre los autores revisados, en cuanto al significado de información de actualidad: noticia acerca de hechos sucedidos en el presente o que tienen relación con el presente. La mayoría de ellos relacionan el concepto con dos elementos inmutables: la novedad en cuanto al contenido y la pertenencia al tiempo presente. Sin embargo, ambos rasgos desaparecen al tratar de definir la documentación de actualidad (reutilización indefinida de las noticias) e, *incluso, cuando ponemos en entredicho su validez en relación a la propia noticia, según argumento de Haacke, cuyo pensamiento he usado para caracterizar el objeto ideológicamente.*

Para la Psicología social, la actualidad es un grado mensurable del interés que distintos sectores de población prestan a los diversos fenómenos del presente, pero a pesar de concebirse como algo estático, (para poder ser medido, por ejemplo, en una encuesta de opinión) existen "grados de actualidad"⁶. Esta posibilidad me lleva a identificar, en muchos casos, actualidad con motivación de la espectacularidad en el sentido de hacer de interés público determinadas cosas a la vez que se silencian otras. El documentalista, en consecuencia, no va a trabajar con textos generados en función de unos hechos espontáneos, objetiva o justificadamente actuales: los diversos niveles y condiciones de producción

⁶ Haacke, Wilmont: Escritos recientes sobre el concepto de actualidad. "Opinión pública". Madrid: Escuela oficial de Periodismo, s.a., p. 169.

de las fuentes de noticias y de las noticias mismas determinan las condiciones de análisis y selección documental, todos ellos ajenos a los conceptos que hemos visto como constitutivos elementales de la noticia. Tampoco "lo nuevo" parece atender a criterios objetivos porque muchas realidades son construidas o se identifica lo nuevo con lo actual, dando un sentido temporal a lo acrónico.

La actitud de los media, en manos de intereses económicos o ideológicos, es divulgar una serie de datos que despiertan la curiosidad del ser humano envuelto en un determinado paradigma de comportamiento que domina en las sociedades industriales del siglo XX y pretende servir de modelo a las no desarrolladas, fundamentalmente a través del vehículo de la información de actualidad transnacional. Las claves del mismo paradigma son tecnicismo y participación en la contemporaneidad, elementos a los que no escapa ningún uso discursivo, incluidos la Ciencia y el Arte. Para consolidar ese estado de cosas, el sistema socio-político permite y fomenta la movilización y el conocimiento masivo de hechos que se alimentan gracias a esa participación en el seguimiento. Sólo que, como ya se ha puesto de relieve, la participación es una falacia: el usuario o receptor de la información permanece estático ante los debates en torno a un discurso que le implica pero que no puede hacer cambiar (tómese el caso de la guerra, como el más obvio), a la sombra del discurrir contemporáneo, sin capacidad de reaccionar o emitir juicios que puedan imponer una actualidad sobre otra. La concentración de partidos y grupos de presión en los países democráticos son productores de contrainformación sobre puntos particulares del debate social pero alejados de poner en discusión el propio debate, por lo tanto contribuyen a la saturación.

El documentalista de prensa favorece la perpetuación de los relatos al registrarlos en memorias automáticas o al facilitar su reutilización difuminada en cientos de relatos diferentes y posteriores. Al intentar realizar su labor objetivamente, el documentalista es una de las principales víctimas de la realidad construida por el discurso de los media. En tanto que servidor de una institución con intereses concretos, su trabajo consiste en organizar los relatos diarios para constituir la prueba o el antecedente de los relatos futuros.

A la esencia de la actualidad se articula, y aquí introducimos el sustantivo información, la rapidez de la difusión y su rápida obsolescencia. La divulgación debe alcanzar inmediatamente al mayor número de personas. Este objetivo confirma dos hechos: el carácter puramente lucrativo, en términos económicos e ideológicos, de la actualidad y la superficialidad de su estructura discursiva, ya que el productor de noticias sabe perfectamente que los cientos de miles o millones de receptores a quienes se dirige no tienen ni el tiempo ni la capacidad

suficientes para desmontar el relato y contrastar los hechos. En los otros medios sólo encontrará, la mayoría de las veces, una información tal vez algo divergente, pero que en cualquier caso redonda la oportunidad del discurso.

El lector integrado busca la propia confirmación de sus ideas en su medio habitual incluso, como advierte pesimistamente Burgelin, el individuo comprará el periódico no sólo para estar informado sino para obtener un orden construido en el aparente desorden, es decir, delega la priorización y la postura ante los problemas sociales en los media los cuales, al decir de Martín Barbero, deciden qué acontecimientos serán noticia⁷.

La organización formal de la noticia de prensa en sumarios y titulares, recuadros, etc convierte a los medios en un inventario diario de eslóganes o reclamos basados en técnicas de captación de consumidores netamente publicitarias. Lo que ocurre es que el producto a consumir, en este caso, es ideología, pautas de comportamiento, hábitos y opiniones. El documentalista debe digerir esa masa de criterios impersonales fragmentados que crean la historia social reciente de una comunidad. Solamente el historiador, si utiliza un método científico historiográfico, será capaz, y ya a destiempo, de desentrañar alguna verdad en un cúmulo de formas creado y alimentado día a día. El periodista, carente de tiempo, antecedentes y método irá a nutrirse en una memoria en la que debe confiar que predetermina sus investigaciones y conclusiones sin someterlas a conjeturas o verificación.

Basten estas reflexiones sobre el concepto y la forma de la expresión información de actualidad para dar idea de la complejidad de la materia prima del discurso que debe canalizar el documentalista de prensa. Me remito al excelente trabajo que, sobre el concepto, ha realizado el alemán Wilmunt Haacke en cualquiera de sus obras sobre periodismo político y en particular la citada, desde una concepción etimológica hasta la interpretación filosófica, psicológica y sociológica. Sentencia Haacke: "el concepto de información de actualidad, en su acepción moderna, emana del mismo periódico... más tarde la categoría actualidad revierte a su vez sobre el periódico y emerge como su principal característica"⁸. En muchos casos, esta identidad facilita lo que llama Hagemann "actualidad artificial"⁹ que consiste en la reducción o exageración

⁷ La cita de Burgelin, O.: La comunicación de masas, es también de Martín Barbero, Jesús: Procesos de comunicación y matrices de cultura. Itinerario para salir de la razón dualista. México: Gustavo Gili/Felafacs, [1988], p. 49-61.

⁸ Haacke, W.: o.c., p.179.

⁹ Hagemann, Walter: Grundzüge der Publizistik. Münster, 1947, p.15-17, cit. por Haacke, W. o. c. p.186.

de lo actual por accidente o arbitrariedad o, más claramente, como afirma Auerback¹⁰, “no hay ninguna actualidad de por sí, sino sólo actualidad dentro de cierta agrupación y con respecto a cierto objetivo”.

Con esas precauciones debo abordar el corpus de información de actualidad susceptible de ser utilizado como campo para la investigación empírica. En definitiva, condicionándolo a su producción por algún medio de comunicación social sea o no publicado y siempre que tenga una supeditación temporal a lo que hemos llamado, en tono relativo, actualidad.

3.1.2 *La unidad discreta de trabajo: elementos para su formulación en información de actualidad*

No es posible proponer un método, ni las condiciones teóricas para el desarrollo o aplicación de éste, sin el conocimiento de los elementos que conforman el corpus que va a ser abordado por el método, esencialmente en lo que concierne a su validación empírica. Es claro que la materia prima con la que ha de enfrentarse el AD en el caso que nos ocupa, es un campo difuso que tomaremos como ejemplo generalizable por su complejidad a pesar de las condiciones de partida establecidas. Decidimos llamar a este epígrafe “Unidades discretas de trabajo” lo que, por una parte, no obliga a aceptar en principio ninguna otra denominación que pudiera confundirnos y, por otra, adelanta la hipótesis de la necesidad de trabajar con elementos separables en la teoría y en la práctica documentológicas. El trabajo y la validación empíricos en Documentación son ya tesis generales que he asumido a lo largo de este trabajo, y que confirman la hipótesis menor siguiente: es necesario construir o captar unidades discretas para el desarrollo teórico y la aplicación de métodos de análisis documental en el discurso periodístico. Ciertamente, para construir estructuralmente un corpus en Ciencias, debemos contar con elementos y estos han de ser separables y clasificables.

Partimos, también de un gran número de restricciones: Si la primera restricción que apuntamos es la de noticia escrita debemos indicar que nos referimos al soporte papel en concreto ya que videotex, microfilm o DON son tecnologías que ofrecen también textos entre sus aplicaciones. Sin embargo, el

¹⁰ Auerbach, Walter: *Presse und Gruppenbewusstsein. Vorarbeit zur Geschichte der deutschen Arbeiterbewegung*. Tesis. Köln, 1929. Leipzig, 1931, p.38-45, cit. por Haacke, W. o.c., p.190.

soporte no es lo que determina el tipo o las dimensiones del corpus que debe ser tomado en la propuesta de un método documental aunque supedita, incluso, la propia comprensión de los contenidos. Las unidades físicas en que se vehicula la noticia escrita, además de las reseñadas tecnologías, periódicos, revistas, panfletos, folletos, carteles, anuarios, índices y teletipos, entre otros medios, imponen unos formatos y una estructura diferentes sobre los relatos que contienen. La cuestión que planteo, a continuación, es si son necesarias unidades de trabajo, si deben y pueden ser discretas tales unidades, si se captan rápidamente o mediante procesos artificiales, como realizar su aislamiento y qué utilidad tienen en nuestro campo. Así será diferente el testado de un método sobre un anuario retrospectivo que sobre un diario. Tradicionalmente se identifican estas unidades de trabajo con unidades de contenido (temáticas) sobre las cuales realiza sus operaciones el documentalista condicionado por la forma de presentación que les da el productor (periodista, cronista, Agencia, etc).

Todos los soportes, en mayor o menor medida, realizan dos tipos de amputación sobre la realidad, además de adecuarla a los usos estilísticos y a los intereses del medio:

- * Segmentación: es la partición consciente de la continuidad de un relato en diferentes soportes físicos que configuran lo que se denomina serie o serialización (al estilo de una telenovela).
- * Fragmentación: es la atomización espontánea de la realidad que interesa destacar en su proyección hacia el futuro inmediato. En este caso denominaré secuencia a la plasmación material y a la expectativa de la fragmentación en diferentes soportes físicos (por ejemplo, distintos números del diario).

También podemos observar la fragmentación de la realidad en el mismo soporte, es decir, los datos que corresponden o emanan de un mismo hecho global son pulverizados (en este caso se produce segmentación en diacronía) en diferentes informaciones que los presentan como pertenecientes a distintos relatos o niveles del mismo relato. Solamente a partir de la construcción de una red de categorías conceptuales amplias en las que pudiéramos ajustar los distintos fragmentos y segmentos inter- e intradiscursivos sería posible captar la continuidad, la manipulación y las contradicciones de los relatos periodísticos. El análisis de contenido ya lo ha hecho con diferentes objetivos (Berelson, Krippendorf, Bardin y otros) pero es necesario establecer un cuadro de unidades discretas con el que podamos operar en Documentación. Lógicamente, parto influido por las unidades de contenido ya que tienen mayor trascendencia

práctica entre los documentalistas y los usuarios aunque esta influencia, tanto en el nivel de concepto como de denominación, sólo la acepto como punto de arranque, exponiendo objetivamente sus ventajas e inconvenientes, mediante un razonamiento que debe llevarme a la adopción en ambos niveles, de uno de ellos o al rechazo de ambos con nuevas propuestas.

Otra característica de la noticia, a la que debe supeditarse el AD, es la imposibilidad de homogeneizar, ni en extensión física ni en intensidad cada uno de los contenidos. Recordemos que en algunos métodos debemos partir de la hipótesis de verificación que nos advierte sobre los condicionamientos de la extensión sobre los resultados lo que hace que simplemente la cantidad de palabras valide o invalide un método. Esta situación ya tiene precedente para los analistas de contenido: la materia prima que utilizan no está, en muchos casos, estructurada, ni es homogénea y, además, se mezclan símbolos, objetos y actitudes. Así lo constata el norteamericano Krippendorf en uno de los mejores manuales sobre Análisis de Contenido que pude consultar y quien, no obstante, realiza una rigurosa aportación sobre el método empírico dentro de los objetivos de su interés¹¹.

En efecto, la diversidad de soportes y formatos, de estilos y géneros, plantea problemas tanto para la constitución de una muestra tomada como punto de partida por la investigación empírica como para el testado final de un modelo teórico de interrogación textual. Para el primer caso, proponemos la exhaustividad temática que se comprueba cuando a partir de un número de unidades de la muestra las interrogaciones efectuadas obtienen porcentajes gradualmente decrecientes de respuestas distintas, hasta alcanzar un umbral de equilibrio que corresponde a la idoneidad de la muestra. Las verificaciones realizadas por esa vía las puse en práctica sobre los resultados de un análisis infométrico de la Comunicación efectuado en 1988 tomando una muestra de mil quinientas referencias¹².

Otra opción es la "técnica de la mitad" descrita por Krippendorf¹³: para comprobar si el tamaño del corpus es adecuado se divide en dos mitades. Si de cada parte se extraen las mismas conclusiones e igual confianza, la muestra es adecuada. Si no es así, aumentamos gradualmente la muestra hasta cumplir la condición.

¹¹ Krippendorf, Klaus: Metodología de análisis de contenido: teoría y práctica. Barcelona: Paidós, 1990, 279 p. (Content Analysis: An Introduction to its Methodology. London: Sage, 1984)

¹² García Gutiérrez, A.: La productivité de la communication en Espagne: un approach infométrique. "DCI", 1989

¹³ Krippendorf, K. o.c.,

3.1.2.1 Unidades de contenido (UC): descripción y evaluación

Para construir una muestra es necesario, en cualquier caso, contar con los elementos menores que la constituyen. Estos elementos han de mantener interdependencia y ser independientes (unidades discretas e integrables) al mismo tiempo, es decir, cada unidad debe ser exclusiva y suficientemente representativa de un hecho o parte de un hecho (de tal manera que pueda ser analizable aisladamente) y no puede existir una excesiva dispersión de unidades que dificultaría su aglutinamiento en estructuras más amplias, es decir, deben situarse dentro de algunas categorías. En Documentación periodística, la categoría práctica que impera es la adscripción temática.

Todo lo anterior nos introduce en el concepto de unidad de contenido (UC) que pasamos a evaluar dentro del corpus discursivo de la información de actualidad. La UC es la representación textual unitaria de un hecho que permite la comprensión cabal, si bien parcial, de la totalidad, de una de sus partes o enfoques en un relato determinado. Podríamos decir que es una unidad autosuficiente de aportación de datos que excede las fronteras físicas y temporales del medio o bien el mismo medio puede contener muchas de ellas, por lo que podemos decir que las coordenadas de espacio físico y tiempo no son condicionantes de su existencia. Una misma noticia puede, a su vez, contener partes de diferentes unidades de contenido si reflejan en su desarrollo elementos integrantes de varios relatos de la actualidad. En definitiva la división de la información por el medio, debido a imperativos comerciales o para mantener la coherencia interna, no corresponde ni debe influir sobre el sistema de captación de la UC en la observación empírica. Veamos que condiciones y problemas se detectan para identificar las unidades de contenido en prensa según los objetivos del AD:

- * Continuidad temática: una UC debe presentar linealidad en la relación de datos que claramente contribuyan a la construcción de un tema. Empleamos el término tema, con las debidas precauciones, en su acepción corriente: "proposición o texto que se toma por asunto de un discurso"¹⁴ y que hacemos equivalente a la representación más general que podemos realizar a partir de varias unidades físicas. Se obtiene, por tanto, por vía estrictamente empírica. A su representación en una etiqueta lapidaria voy a llamarla macrodescriptor. El número de macrodescriptores reduce, en unos pocos centenares, todos los temas considerables de actualidad informativa cualquiera que sea la cultura,

¹⁴ Casares, Julio: Diccionario ideológico de la lengua española. Barcelona: GG, 1978, p. 804.

el medio o la lengua. Si esto es cierto, y creemos que lo es, necesitaríamos elevarnos a un nivel adecuado para poder proponer un paraguas de macros que cubra cualquier nueva información. Otro punto de discusión sería, por consiguiente, cuál debe ser el método de construcción del cuadro de macrodescriptores, qué utilidad tiene como universo de categorías necesario para el establecimiento de las UC y si es posible mantener una proximidad y una homogeneidad en la profundidad sémica de cada uno de los elementos constituyentes del inventario.

Para el establecimiento de los elementos y categorías anteriores, precisamos manejar temas (ya descritos) y conceptos, término que usamos pragmáticamente a pesar de las polémicas filosóficas sobre éste. El concepto, o idea abstracta o general captada o concebida por el entendimiento humano, es un ente sometido a permanente debate desde Platón, quien lo vincula al espíritu y no a la experiencia: los conceptos previstos por la experiencia son relativos (Juan es más alto que Pedro), hasta el pensamiento más reciente que se plantea la realidad del concepto: los empiristas y nominalistas, por ejemplo, para quienes el concepto no es más que una palabra (Locke, Hume) y los racionalistas que le confieren una realidad en el espíritu (Platón, Kant)¹⁵. Entrar en tales debates desborda mis conocimientos a la vez que supondría salir del corpus documentológico que propongo. Sin embargo, la necesidad del concepto en Documentación nos clarifica algo: por una parte, las implicaciones filosóficas de la Ciencia documental como constituyente de las llamadas Ciencias del Conocimiento y, por otra, la necesidad de la pragmática para poder avanzar en la investigación documentológica manteniendo la consciencia de las limitaciones y reducciones que impone ese objetivo.

- * Una UC no corresponde a una unidad física (recorte de noticia, reportaje, etc.). *El género periodístico puede abordar varias UC simultáneamente o apenas un fragmento o segmento de una de ellas.*
- * Las UC no se reconocen en el nivel de superficie discursiva, sino en su reconstrucción en el nivel de estructura profunda. Se analizan, por tanto, en estado virtual al que se accede tras un proceso de selección, aglutinación, análisis y síntesis: 1. Selección: se establecen unos criterios mínimos y objetivables de selección de noticias (excluyendo futuribles, rechazando superposiciones o superaciones de datos, etc.). 2. Aglutinación: se agrupan todas las noticias con afinidades evidentes. Esta agrupación debe hacerse en el marco de un cuadro teórico de categorías de macrodescripción que sirve

¹⁵ Dictionnaire de la Philosophie. París: Larousse, 1964, p.49

de referencia y garantiza la homogeneidad de la construcción de unidades de contenido. De la construcción de las categorías depende la mayor o menor parcialidad de los resultados del análisis. 3. Análisis: el conjunto de noticias se descompone de acuerdo a un método de desestructuración del texto. El resultado es una fragmentación de las noticias que no corresponde a la fragmentación de sus formas originales. 4. Síntesis: la detección de solapamientos en los resultados del análisis y la reducción de estos junto al rechazo de datos triviales y superfluos nos lleva a la creación de una síntesis que representa un relato o seguimiento temático al que se ha atribuido la característica de UC con una restricción temática unitaria. Este tipo de restricción no concuerda, en muchas ocasiones, con los criterios de los documentalistas o de los usuarios del sistema, por lo que utilizamos este punto como primer contraargumento para la determinación de unidades desde la perspectiva (exclusivamente temática).

- * La UC es participada por el principio de intertextualidad: en estructura profunda, las UC se interrelacionan a nivel de segmentos o partes de segmentos creando nuevos contenidos sólo atisbados cuando se adopta la actitud interrogadora y receptora. En cualquier caso las UC deben prever interrogaciones totales, parciales o mixtas para poder hacer frente a las demandas, siempre que éstas estén recogidas en la estructura de los textos y hayan sido asumidas al diseñar el sistema. Esto obliga a considerar un posible diseño de las UC desde su uso, desde la perspectiva de un usuario intuído; deben responder a cualquier interrogante recogida en un cuadro de metainterrogantes predefinidas lo que en principio supone mucha complejidad.
- * El método debe ser factible, es decir, atender a las posibilidades del documentalista y a sus restricciones de medios. Por ello, debe poder aplicarse a las unidades físicas no rechazadas por el documentalista generándose la construcción de unidades afines en el nivel de estructura profunda que se detecta al formular una demanda a una red semántica interrelacionada.
- * La existencia de UC se establece en régimen condicional o de dependencia. Por tanto, los resultados obtenidos de la aplicación de un método sobre un corpus de UC deben ser contrastados finalmente, para poder ser validados, en un corpus abierto de noticias. Las noticias presentan heterogeneidad a todos los niveles lo que obliga al investigador a elaborar, a priori, comprobaciones condicionales y designar las fases de tales comprobaciones en el plan de construcción de unidades previsto.

- * Las UC sólo interesan para la construcción del método y, por ello, se fabrican en laboratorio, mediante simulación o en condiciones ideales. Sin embargo, deben ser disueltas una vez construido aquél ya que será el usuario del sistema quien quizá las aborde desde una u otra perspectiva o tan sólo solicite un dato contenido en ellas. Este es un nuevo contraargumento ya que la dificultad de la construcción tiene un aprovechamiento relativo en el test (adscripción tópica) y ninguno en la aplicación del sistema.

- * A tenor de los principios anteriores, debe construirse un cuadro de macrodescripción que sirva de contexto extensional del corpus. Este cuadro condiciona los temas y la adscripción de fragmentos y segmentos a temas. No obstante, su modo de construcción debe ser también sometido a conjeturas. Puede provenir éste de la vía empírica para lo que se habrá de crear una muestra que cumpla los criterios de verificación aleatoria (todos los elementos incluidos y excluidos habrán tenido las mismas posibilidades de pertenecer a la muestra). La vía especulativa o apriorística también puede ser usada en exclusiva o combinada con la anterior. La elección de uno u otro método de construcción de las macrocategorías referenciales presenta ventajas e inconvenientes que he analizado en otro trabajo sobre elaboración de lenguajes combinatorios¹⁵. Será preciso justificar las causas de la elección de un método en función de los objetivos y establecer previamente los criterios para validar tanto la vía como los resultados y evaluar los equilibrios y la participación de todas las categorías que participen en el discurso en cuestión. El cuadro macrodescriptivo es de base temática, imprescindible para la construcción de las UC y para su organización en muestras, pero podría quedar invalidado en el caso de demostrarse innecesarias las UC. No obstante, hay una hipótesis, de origen distinto, para defender la existencia del inventario aludido: en los métodos de análisis inteligente mediante simulación es necesaria la presencia de esquemas (frames, para algunos autores como Van Dijk¹⁶) o macroadscripciones que en el discurso periodístico se inclinan abiertamente, entre otras perspectivas, por la temática.

- * Suficiencia estructural: las UC deben estar dotadas de suficiencia estructural, elemento que se percibe si contiene un conjunto mínimo de respuestas a posibles interrogantes (demandema) que incluyen la necesidad de captar ciertas omisiones de tipo contextual y encadenado (vid. más adelante, tipos de omisiones). Para una máquina sólo será posible, y en fase de inteligencia

¹⁵ Dictionnaire de la Philosophie. Paris: Larousse, 1964, p.49

¹⁶ Véase el concepto de frame en Van Dijk, Teun: Texto y contexto... o.c. Madrid: Cátedra, 1980, 357 p.

avanzada, captar las omisiones neutrales. La UC es al AD lo mismo que el concepto de texto al análisis semiótico o la frase al análisis sintáctico. Sin embargo, su constitución es divisible en unidades menores por motivos experimentales u otros que, en todo caso, la descontextualizan. Esta es la medida de su condición mínima: la manipulación de partes menores no representa la misma realidad, o la misma cantidad de actualidad que la UC completa. La suficiencia también depende del cuadro categorial de interrogaciones posibles. Por lo tanto, no podemos universalizar criterios para establecer una metodología práctica de estas unidades. En cada situación discursiva y en cada uso, sean estos establecidos mediante parámetros teóricos, empíricos o simulados, las unidades deben adaptarse a las condiciones de los indicadores macrodescriptivos, los cuales reflejan los temas generales del discurso y de los indicadores de interrogación que abarcan las demandas de usuarios modelos. En este sentido, las UC presentan un reduccionismo importante respecto a las categorías que deben asumir internamente.

- * Tematicidad: entendemos por tematicidad la claridad de la exposición o de la ubicación del tema. Es una categoría nocional que nos permite medir el grado de realización o culminación del tema. Por ejemplo, es preciso eliminar las previsiones y rumores y detectar las superposiciones. Por tanto, no es posible construir o testar un método sobre relatos en situación pendiente o inacabada. Este punto hace necesaria la presencia de una estrategia analítica en el trabajo cotidiano ya que no todas las informaciones cumplen el requisito de las UC. Los criterios de selección influyen y dependen de la calidad de los resultados obtenidos por la aplicación de los métodos. La tematicidad en campos enciclopédicos supone un argumento contrario al uso de UC ya que ninguna muestra de ellas podría representar todo contenido real o potencial. En este sentido, los temas son categorías poco operativas para la constitución de muestras lo que nos hace, sin rechazarlos totalmente, pensar en otras categorías estructurales.

- * Construcción de UC: la unidad de contenido se compone de una superficie textual a través de la cual se llega a la significación y, por inferencia, a las posibles connotaciones y omisiones. Las omisiones que, detectadas, sitúan a la unidad de contenido respecto a un contexto y la encadenan a otras unidades de contenido a través de múltiples perspectivas de la propia estratificación de los datos son de especial relevancia. En consecuencia, una UC queda conectada a un nivel discursivo superior e inmediato y a otras unidades de distinta adscripción temática dominante. Pero las conexiones o independencias no presentan fronteras evidentes. Los haces microtemáticos

de las UC se atraviesan mutuamente creando una estructura cognitiva/discursiva compleja. La delimitación sólo obedece a causas pragmáticas que compensan la pérdida de globalidad, de perspectiva y de información. Una UC es suficiente cuando comporta un tema y aporta información original a la construcción del discurso en cuestión, respondiendo a las expectativas de los usuarios. Estos receptores imponen sus interrogaciones al sistema pero, obligatoriamente, habrán de ser instruidos sobre cómo hacerlo y cuáles son las limitaciones y recursos.

- * Géneros, UC y estructura profunda: existen ciertas contribuciones periodísticas que alcanzan el principio de UC antes que multitud de fragmentos físicamente agrupados. Un editorial, por ejemplo, expone las condiciones de un asunto, desarrolla una argumentación y propone unas conclusiones. Todo ello se basa en la reflexión sobre conjeturas y hechos y no en un método científico, aunque intente emular formalmente el desarrollo de la exposición lógica en la investigación, lo que nos impide fiarnos de su objetividad. Sin embargo, el relato editorial es un buen ejemplo unitario y formal de infonema, en muchas ocasiones. Esta constatación nos lleva a la aceptación de perspectivas o tratamientos de una unidad temática para la formación de unidades analíticas diferenciadas, todas ellas bajo la misma macrodescripción. Debemos señalar en el discurso periodístico la necesidad de una homogeneidad de los fragmentos sobre un tema para poder constituir UC. En la aplicación práctica, se realizan análisis separados que mantienen en común conectadores¹⁷. En la observación teórica, afirmamos que la UC sólo existe a nivel de estructura y de significación, que a veces coincide con estructura de expresión, siendo ahí donde es necesario realizar el análisis, las adscripciones y las interconexiones. En superficie de representación, una vez aplicado el AD, las unidades quedarán separadas físicamente aunque conectadas mediante unos rasgos comunes de intertextualidad.

La UC pertenece, resumiendo, al dominio de la abstracción y puede ser observada y manipulada teóricamente. Se construye y fluctúa en el nivel de significación aunque la representación modela su extensión y número. En definitiva, la representación formal puede determinar la captación de UC por parte de usuarios pero, de ninguna manera, su observación teórica. Solamente a partir de la abstracción de unidades discretas en un discurso periodístico es posible proponer un método de análisis válido para ese mismo discurso. Pero esa retención está sometida al dictado de las macroestructuras temáticas que

¹⁷ Los conectadores son descriptores de contextualización del relato, concepto desarrollado en García Gutiérrez, A.: Estructura lingüística...o.c., p.102 s.

predefinen el marco discursivo de ejecución y validación si nos referimos a las UC. La dificultad de elaborarlas, dada la complejidad temática y aspectual de la noticia, nos induce a estudiar otras alternativas en cuanto a denominación y contenido de las unidades de trabajo para la investigación documental sobre la información de actualidad.

3.1.2.2 Rechazo de algunas conjeturas

- * Los elementos que analiza o produce el AD no tienen, generalmente, unidad temática en el discurso de actualidad. La realidad y las características de la prensa causan esta situación. Esto obliga a realizar construcciones artificiales de UC basadas en la tematicidad que pone en peligro el principio de generalidad aleatoria que debe cumplir una muestra.
- * No es operativo construir UC pragmáticas o para su utilización real : habría que ubicarlas en tantos lugares como temas las cruzan dependiendo la decisión de un usuario imaginado y del nivel de generalidad de la clasificación (cuánto más general menos riesgos pero también menor precisión), como queda demostrado en un archivo de recortes o acceso directo: cada fragmento es fotocopiado varias veces engrosando distintas carpetas temáticas. La UC existe pero permanece en estado virtual en una memoria automática.
- * Es inviable construir un inventario de macrodescripción representativo y homogéneo que sirva de matriz para la obtención de unidades de contenido.

Representación gráfica del procedimiento y de los problemas de obtención de UC temáticas:

N: Noticia. Unidad informativa limitada físicamente (recorte, imagen, etc.)

Letras: segmento o fragmento del relato.

Números: Adscripción temática del fragmento o segmento.

D: día (coordenada temporal) de aparición con n.º de correlatividad.

UC: Unidad de contenido con número de identidad.

D1	NA1	NB1	NC1
D2	NA2	NB2 fin	NC2
D3	NA3		NC3
D4			NC4 fin
—			
—			
—			
Dn	NA _n fin		

$$UC1 = NA(1-n)$$

$$UC2 = NB(1-2)$$

$$UC3 = NC(1-4)$$

Podemos observar la ubicuidad de muchas noticias. Si realmente son relevantes para diferentes UC en constitución no perjudica en absoluto la múltiple presencia. Por ejemplo, supongamos una noticia que trate de las actitudes de los consumidores ante una nueva crisis energética. En el caso de disponer de dos UC (consumo y energía) podemos adscribir la noticia a cualquiera de las dos o a ambas, puesto que en un sistema real la recuperación obtendrá sus frutos sin perjuicio de la localización de la noticia. Dejemos constancia de que la adscripción se realiza por una mera necesidad organizativa para la concepción teórica que en absoluto repercute en la aplicación posterior pero que crea el conflicto de la representatividad enciclopédica.

El tratamiento y la concepción de UC en Ciencias, o específicamente en Ciencias sociales y humanas, no es completamente extrapolable en otro tipo de discurso como el periodístico. No obstante, ciertos principios comunes pueden ser identificados:

La aparente facilidad de la construcción de UC nos ofrece un camino para confeccionar otro tipo de unidades. La presencia de condiciones y construcciones adhoc dan como resultado un modelo parcializado que debe ser enriquecido, ajustado o corregido de acuerdo con los resultados de una evaluación sobre informaciones no preparadas. En cualquier caso, el modelo debe ser abierto a modificaciones puesto que una vez superada la validación última, comprobaremos cómo la dinámica del propio discurso hace obsoletos o imprecisos algunos recursos que habían superado la verificación.

Las unidades deben ser abiertas e interrelacionadas, esto es lo que podríamos reconocer como campos referenciales al hacer una búsqueda tradicional: todo lo obtenido pertinente se adscribe a una unidad construida en la demanda. El carácter abierto permite enganchar y arrastrar las respuestas dispersas a la manera de los encadenados inferenciales de los sistemas expertos.

El planteamiento del desarrollo investigador en fases, hipótesis, método, descripciones y desarrollo, contrastaciones, conclusiones puede tener equivalencias en las fases del género dramático neoclásico: presentación, nudo y desenlace, estructura que se detecta de alguna forma en determinados tipos de información de actualidad.

Existen categorías conceptuales que deben estar presentes obligatoriamente en toda unidad de contenido aunque se omitan, por obvias o desconocidas, en los fragmentos originales. Estas cuestiones deben ser tratadas en un apartado específico.

Las extensiones de las unidades de contenido son variables en otros campos, sin embargo suelen tener cierta homogeneidad en cuanto a este aspecto. La naturaleza de la información de actualidad puede otorgar el mismo valor a una palabra (o a un silencio) como a miles de páginas. En cualquier caso, para la simulación de modelos de análisis no son aceptables unidades menores a la proposición es decir a la relación que puede establecer el emisor del mensaje entre una entidad y otra (equivalentes a enunciados simples). Este principio se basa en la mínima estructura lógico-sintáctica reconocible por una máquina aunque no debemos aquí implantar las limitaciones actuales que estriban en la restricción a enunciados declarativos. Este punto hace inviable la simulación en los sistemas conocidos ya que el modo declarativo sólo es uno más, y no el más importante, del discurso periodístico.

3.1.2.3 Alternativas y tipología de unidades: hacia la unidad de información.

El ángulo desde el que debemos vislumbrar y construir las unidades de trabajo con fines experimentales ha sido uno de los puntos más polémicos, y consultados en la práctica, de esta investigación, precisamente, por el hecho de que el cambio de perspectiva invalida unidades que se daban como seguras. Esto ocurrió al proceder a la creación de UC, influyéndome en exceso la metodología

sociológica (Bardin, Krippendorf) y documental tradicionales: contenido se identifica con temática, por lo tanto se reconoce una unidad en cuanto mantiene adscripción a un eje isotópico. A simple vista, esto era demasiado trivial como para ser operativo. En efecto, cambiando de eje desaparecía la UC lo que presentaba un incontrolable campo de temas. Para conseguir un mínimo nivel de coherencia interaxial era preciso mantenerse dentro de un mismo corpus, para lo que elaboré, por el método global, un inventario de macrotemas exhaustivo, si bien debo señalar la existencia de algunos recortes y limitaciones de poca importancia para la prueba, puesto que no buscábamos la exhaustividad de la representación de una muestra sino las dificultades de la construcción de unidades discretas por la vía temática.

El resultado de las experiencias fue el siguiente: al comparar noticias y categorías se iban formando, efectivamente, unidades de contenido que, al discurrir diacrónicamente, devenían unidades de relato. La diferencia de ambas expresiones no me parecía evidente ya que los relatos también eran unidades de contenido y, en este caso, la extensión física no es determinante. Un relato periodístico puede formarse a partir de pocas noticias, de una sola o a través de la intertextualidad de noticias, en principio, diferentes. Otros problemas surgidos y detectados fueron:

- * Dificultad objetiva para designar el nivel de profundidad o, al menos, el equilibrio de las categorías lo que incidía en la determinación de la UC.
- * Dificultad de objetivación para designar las distancias o aproximaciones sémicas entre las mismas y para hacerlas homogéneamente excluyentes mediante una regla.
- * Adscripción posible de las noticias a diferentes unidades, como se ha dicho, simultáneamente.
- * Necesidad de universalizar el inventario dado el carácter enciclopédico de las noticias, tarea compleja aunque factible, pero poco rentable en sentido metodológico.
- * Imposibilidad de cerrar la muestra, en cuanto a la inserción de UC temáticas, ya que en la adscripción temática de tipo enciclopédico los temas no se agotan (y, por tanto, el inventario no es finito) y queda siempre la posibilidad de que en el testado final una sola noticia imprevista no se adapta, lo que invalida el procedimiento.

* Finalmente, encontré el contraargumento más importante, al reflexionar sobre el origen de la producción de las UC: evidentemente, en este caso, mi estatuto de observador y mis claros objetivos me colocaban en el rol de receptor del texto, yo mismo me situaba como lector modelo (todo lector o cualquier lector) de temas modelo (todo tema o cualquier tema, al ser enciclopédica la aproximación). En ese punto encontré la respuesta para rechazar las UC como elementos constituyentes, en exclusiva o en gran parte, de las muestras destinadas a validación de métodos, en contra de algunas hipótesis defendidas al arrancar este estudio. Estas son las nuevas consideraciones:

Las UC son unidades temáticas pero puede haber otras unidades posibles para la medición. Incluso es más viable la construcción de un método utilizando la ejemplarización (necesaria debido al pragmatismo) junto a múltiples validaciones prácticas en la información de actualidad para provocar modificaciones en las reglas.

Las UC pueden ser construídas o percibidas según distintos enfoques y agentes. El propio productor del texto puede estar construyendo UC, a través de relatos, intencionalmente o, al menos, contribuyendo a ellas. Esas UC no coincidían con las que íbamos creando desde el puesto de observador. Del propio texto emanan también las componentes de claras unidades temáticas, prescindiendo de agente perceptor, lo que es difícil de probar ya que siempre es necesario un observador, incluso si la operación la realiza un autómat. El usuario o lector del texto define sus propias UC de acuerdo a sus necesidades: mi posición, por el contrario era aséptica en cuanto que no necesitaba temas sino unidades de cualquier categoría para operar. Sin embargo, en la práctica documental es el usuario el que modela los contenidos según la demanda. Es decir: los temas solicitados y los modos de solicitarlos organizan de una manera o de otra los, aparentemente, relatos establecidos: es el usuario quien establece qué es un relato en Documentación aplicada y, por tanto, la consecutividad diacrónica de noticias sobre un mismo "tema" desaparece como un espejismo cuando al usuario le interesa desde una perspectiva o modo distintos.

Estas consideraciones me hicieron proponer nuevas denominaciones para tales unidades con el fin de no confundirlas con las UC de adscripción temática en otros campos, a la par de constatar la imprecisión de estas últimas. Veamos algunas de las posibles alternativas:

— unidades documentales: denominación que comporta confusión sobre la unidad física (el documento o la unidad de registro),

máxime en los soportes anárquicos de la información de actualidad.

- unidades de búsqueda o recuperación: expresión que somete en exceso a una temática dictada por los usuarios y es difícilmente inventariable (coincidiría con cualquier posible demandema).
- unidades de información: denominación más aceptable al abarcar tanto a los posibles enunciados de usuario como a la entidad del propio texto y a los deseos de los productores e, incluso, de los observadores intencionados con objetivos experimentales, caso en el que nos situamos.

En consecuencia, propongo la expresión “unidad de información” como elemento abarcador de una amplia tipología y a la vez unidad mínima discreta de una muestra construida para la validación en AD aunque debo, todavía, clarificar su naturaleza: la unidad de información no se adscribe a un tema. El requisito es que contenga una mínima estructura, en el sentido de nuestra propuesta de “suficiencia estructural”¹⁸.

La unidad de información es la cristalización del infonema (nivel virtual de la esencia de un texto) pero, pese a existir esta mínima diferencia (explicitación o virtualidad), emplearemos ambas denominaciones como sinónimas a efectos generales¹⁹. La unidad de información tiene, como principal rasgo, organizarse en torno a una estructura. Son, pues, unidades que deben responder, por ejemplo, a una gramática mínima independientemente de la temática. Si esta gramática, en concreto, es la de casos, la extensión de la muestra podría conocerse tomando el número de casos y operando estadísticamente con todas las combinaciones posibles (2 a 2, 3 a 3, etc.).

La unidad de información debe expresar, al menos la extensión de una proposición, condición expuesta anteriormente, es decir, una relación entre entidades/entidades, entidades/atributos, etc. Este requisito dota al infonema de un cierto contexto, pero también debe encerrar la pragmática documental: ha de responder a alguna demanda. Por tanto, un dato (fecha, nombre, etc.) no puede ser infonema aunque sea elemento de respuesta porque para ello debió ser encadenado a un contexto mínimo (presencia del discursema en el infonema de

¹⁸ Concepto desarrollado en Idem: Suficiencia estructural... y tipología de la omisión en AD. “DCI”, vol. 13, 1990, p.81-84.

¹⁹ Idem: Estructura lingüística... o.c., p.94

manera explícita o tácita mediante relación estructural), pero sí puede serlo un enunciado simple: "La guerra del Golfo comienza el 17 de Enero de 1991", "Nunca nieva en Sevilla" (estos ejemplos señalan otro punto esencial de debate: como representar los enunciados). Ahora bien, es discutible el interés que puedan tener los enunciados simples para formar muestras ya que necesitaríamos tantos que la muestra sería poco manejable. Se sugiere trabajar con unidades mayores construídas a partir de enunciados validados. Hemos de tener en cuenta, en cualquier caso, que la alimentación de un sistema experto se realiza mediante proposiciones breves de tipo declarativo por lo que no deberían olvidarse esas estructura mínimas al simular los procedimientos en los sistemas.

Las unidades máximas, obviamente, no son definibles ya que el usuario puede solicitar grandes reportajes o, incluso, lo que aporta una síntesis de ellos al estado de la cuestión. En información de actualidad, sin embargo, los formatos y extensiones son conocidos según el medio y el género. Para el testado, los aspectos relacionados con la extensión deben ser definidos de acuerdo al tipo de método, de corpus y de objetivos fijados. Así, para métodos lógico-semánticos en amplios campos de significación deben tenerse en cuenta unidades de texto integral de varias decenas de párrafos e, incluso, de varias páginas. Para métodos que abordan el enunciado, deben construirse textos de la extensión de resúmenes descriptivos o analíticos con oraciones simples y compuestas que respondan a los interrogantes. En el caso de textos amplios, puede ponerse como condición la exigencia de responder a un número imprescindible de categorías: agentes, objetos y locativos, por ejemplo en la gramática de casos, para la aplicación del modelo de Lasswell debe exigirse una acción en relación predicativa con cualquier otro elemento, etc. (véase el cap. 4).

En los métodos de análisis formal la extensión no es obstáculo para unos resultados de *por sí confusos al basarse en presencia/frecuencia* y no en relevancia/representatividad de una proposición. No obstante, determinados sistemas estadísticos precisarían de un número abundante de ocurrencias para poder dar resultados. Los textos deben ser muy redundantes en el plano de expresión y, por tanto, son muy extensos.

Resumiendo, la propuesta de unidad de información o infonema es necesaria para la validación de métodos y, superando pero asumiendo la adscripción temática, debe cumplir con la suficiencia de estructura, características definibles pragmáticamente en función de los objetivos. Los infonemas deben poder ser adscritos a unidades temáticas (temas), estructurales (categorías de construcción textual) de búsqueda (categorías conceptuales y modales de la interrogación) y de soporte (dictadas por el medio físico de donde fueron

extraídos). El infonema no tiene presencia real, ni siquiera en la memoria automática sino que se compone a partir de partículas requeridas por uno u otro usuario a través de un proceso interrogativo que recorre la intertextualidad. La maleabilidad del infonema es tal que sus fronteras se establecen en función de otros infonemas, dentro de un contexto, y nunca en términos absolutos. No obstante, insistimos, su captación es imprescindible para la validación empírica de los métodos pragmáticos.

La muestra compacta de infonemas tiene por objeto verificar los resultados de los métodos de análisis y de lenguajes, como prueba de laboratorio, antes de obtener fracasos sin posibilidad de reaprovechar el test para las modificaciones necesarias. Pero estas argumentaciones sólo van destinadas a la documentación del discurso periodístico. Esto significa un nuevo abismo para la extrapolación general de propuestas desde las Ciencias a nuestro campo de interés y viceversa, por lo que es esencial buscar los mínimos de compatibilidad en esferas superiores de epistemología práctica.

Tipos de unidades de información

A efectos documentológicos y basándonos en la perspectiva estructuralista, en puridad, un infonema ocurre donde existe estructura o relación, que implican significación. Existe significación cuando dos elementos del lenguaje interactúan en el plano semántico en los modos paradigmático y sintagmático. Este principio es esencial para la construcción de lenguajes documentales, pero es insuficiente en los procedimientos del AD. Todo lo que produce significación no es de interés para el AD; tan sólo aquello que es susceptible de satisfacer a un usuario, conocido empíricamente o especulado, debe ser aislado y desmontado en el análisis. En consecuencia, el principio de relevancia determina pragmáticamente la significación en Documentación y aquella se establece a raíz de unos parámetros que emanan desde la selección al comportamiento del usuario, ajustada al dictado de un contexto.

En consecuencia, podemos afirmar que en una estructura intransitiva de dos significados /niño/+/come/ existe utilidad semántica, pero no documental. Si conocemos quién es el niño, por ejemplo, el primer niño que sufre un trasplante de cerebro o una persona determinada que por cualquier razón conocida (contexto) no podía comer etc., existe relevancia y, por tanto, significación para el AD. En cualquier caso, un enunciado de ese tipo no existe generalmente como documento sino como fragmento de un texto más amplio.

El documentalista (o el sistema experto) debe contrastar la relevancia contextual de /niño come/ con la descripción de esa situación y juzgar si el contenido aporta más información relevante, si satisface a un usuario interesado. En caso afirmativo, la estructura /niño come/ es objeto del AD. Vemos como cualquier enunciado atraviesa una rigurosa acribia antes de ser sometido al primer tramo del análisis.

No podemos, al hilo de lo anterior, considerar un sólo tipo de unidad de información en Documentación periodística, además de que existen otras denominaciones cuyo concepto ya ha sido incorporado, lo que nos lleva a concluir lo siguiente:

- * La unidad de información toma su naturaleza en virtud de ciertas variables que debe predefinir el documentólogo de acuerdo a unas aplicaciones prácticas. Así, el discurso científico determina unas unidades y el discurso periodístico, otras. Dentro de este último, los géneros y estilos impondrán criterios distintos para la captación. Todas las variables posibles tendrán que ser incorporadas en el sistema experto de manera codificada y situadas como metarreglas para el desarrollo de uno u otro árbol de decisión. La complejidad del discurso periodístico exige un sistema de navegación entre los diferentes despliegues de reglas que permita la intratextualidad de la búsqueda.

Si la hipótesis subrayada es cierta podemos afirmar que es posible crear una tipología de unidades de contenido documental:

- Texto: es una superficie discursiva en Documentación equivalente, a veces, a las noticias y en determinados casos a documentos (unidad física). El texto es la unidad material de trabajo analítico-práctico, y más concretamente el texto íntegro (original o no reducido). El conjunto de textos forma el corpus libre que debe superar cualquier método documentológico en la última fase de validación.
- Contexto: es la adscripción inmediata de un texto a un entorno superior y encadenado mediante coordenadas propias del texto que hacen inferir y asociar datos a relatos. En el discurso periodístico consideramos el contexto como equivalente al relato general en el que se inscribe una noticia. Unidad de contexto es igual a unidad de relato.
- Unidad de registro o de análisis: es igual a unidad textual y a texto, pero en el plano físico. El documentólogo realiza su trabajo sobre ellas

tras una previa selección. Equivale a documento aunque en éste pueden existir a veces distintas unidades de registro.

- Unidad de relato: es lo mismo que unidad de contexto para una noticia. Se detecta un relato cuando se observan macrodescriptores constantes o, dicho de otro modo, un relato se organiza en torno a un evidente eje isotópico.
- Discurso documental: es el conjunto de UC que representan el desarrollo del conocimiento en un ámbito determinado.
- Documento: unidad física diferenciada que contiene información. Podemos considerar documento tanto a un recorte o a una sección del periódico, como a todo el periódico, o una colección. La imprecisión de este concepto nos lleva a preferir unidad de registro como expresión más vinculada a los contenidos sobre todo en información de actualidad en la que las noticias se diferencian física y tipográficamente.
- Infonemas: en el máximo nivel de abstracción analítica es la síntesis virtual de un texto, cualquiera que sea su tipología. Es la esencia de un fragmento, segmento, documento o relato. Cuando el conjunto de infonemas representa un discurso nos referiremos a discursemas (síntesis discursiva), elementos que se hallan presentes en cada texto y que permiten su adscripción a un tipo de discurso o a otro. Cuando el infonema es la síntesis de las interrogaciones de un usuario modelo hablaremos de demandema. En el proceso de análisis y recuperación estos elementos participan en los objetos manipulados.
- Eje encadenante: es el hilo significativo de carácter isotópico que provoca la intertextualidad y obliga a la abstracción para obtener las unidades de información. Así, determinadas informaciones sobre la crisis económica generan relaciones con la contaminación. Los ejes encadenantes son los que hacen flexible la estructura de superficie del discurso y, a la vez, los que le otorgan unidad en el estado virtual.
- Dato: cada uno de los elementos significativos de la unidad de registro. No son analizables separadamente, ya que se descontextualizan, pero tienen utilidad, sobre todo, para atender interrogaciones de tipo factual.

- Eje isotópico: significación redundante en diferentes fragmentos informativos que otorgan el rango de relato. Puede estar representado por uno o varios macrodescriptores.
- Macrodescriptor: unidad de representación contextual que adscribe un fragmento a un relato mediante una isotopía o eje isotópico.
- Situación o escenario: todos los entornos en que un tema, persona, lugar o tiempo se desarrollan. Por ejemplo, todas las actividades privadas o públicas, obvias o desconocidas de un presidente. Las situaciones se detectan por la isotopía que genera, en este ejemplo, el presidente o sus distintas formas de denominación o reconocimiento: González, Felipe, el Jefe del Gobierno, el inquilino de la Moncloa, etc.
- proposición: contenido de un enunciado. Si es básica debe contar, al menos, con una entidad relacionada con otra entidad o atributo.
- unidades de búsqueda o recuperación: unidades de información que responden a ciertas categorías de interrogaciones.
- unidades de contenido: unidades construidas mediante la adscripción temática conseguida a tenor de un cuadro de macrodescripción universal (en un discurso determinado).
- unidades de soporte: equivalentes a documentos y, también, a unidades de registro, aunque su polivalencia las hace concepto vago y poco operativo.
- unidades estructurales: unidades elaboradas a partir de los elementos presentes y posibles en la estructura de un discurso específico, ej. la disposición y relación de elementos en Ciencias, el paso de las ideas a las conclusiones, de los hechos a las teorías, etc. Las condiciones de producción de los reportajes, la cadencia de los datos y su organización, etc. Se construyen en torno a combinaciones funcionales (objeto, agente, producto, modo, etc.) realizadas exhaustivamente.
- unidades de información: elemento unitario y separable para la validación de métodos documentales. Pueden basarse en una perspectiva, tema, soporte, estructura, demanda, etc o combinaciones binarias o múltiples de los elementos anteriores, siempre que mantengan homogeneidad entre ellas y respondan al principio de aleatoriedad y representatividad en las muestras.

3.1.3 Estructura de la información de actualidad

La noticia se organiza de distinta forma, según sus objetivos. Hemos descrito anteriormente los dos tipos de discurso que nos interesan en documentación periodística: el neutral (que, al menos, no explicita intencionalidad) y el doxológico (que explicita la opinión y la argumentación).

En la información neutral (generalmente impersonal, breve, sin contrastar, incompleta: flash, suelto, teletipo) los datos se organizan en torno a una pirámide invertida que distribuye la información más novedosa o estrepitosa en primer lugar, ya avanzada en titulares, descendiendo en detalles, antecedentes o con datos ya conocidos (reiteración). El estilo es narrativo y descriptivo pretendiendo mostrar los hechos como son, como ocurren y adjudicándose una falsa objetividad.

En este tipo de noticias es donde encontramos mayor fragmentación pero no segmentación, pues esta última es un recurso que utiliza el diario con fines comerciales que no puede permitirse en una información que envejece rápidamente o sobre la que puede perder la primicia. Los datos se estructuran en torno al acontecimiento, persona, lugar o tiempo. Debemos señalar la fuerte presencia del componente locativo en la actualidad. En varios periódicos analizados aleatoriamente pudimos comprobar que el ochenta por ciento de la información tenía una adscripción geográfica concreta. De hecho, la propia división del periódico (internacional, nacional y local) se presta a ello. En las restantes secciones las noticias ofrecen un locativo en las primeras líneas. Por locativo entendemos no sólo el espacio físico sino también la mención de temporalidad. Solamente en las secciones culturales o literarias no existe la adscripción locativa como imperativo. De hecho es el tiempo el único distintivo en el desarrollo de los relatos: la variación de situaciones adscritas a lugares se debe al transcurso del tiempo.

Las adscripciones onomásticas son menos frecuentes que las anteriores, al menos tienen menos presencia en lo que se refiere a intensidad o dedicación de la noticia, pero constantemente se observan nombres citados directa o coyunturalmente. En realidad sólo podríamos hablar de la presencia dominante de un macrodescriptor onomástico en el caso de perfiles biográficos, necrológicas, etc. Este último caso es un distintivo de UC: todo lo relativo a una persona gravita sobre el campo de interés de esa persona.

Sin embargo, a pesar de la posible multiplicidad de locativos y personas en una misma noticia, la ubicuidad es mucho más reducida que en el caso de los

temas. Por ello, debe partirse de la organización de estos, en un cuadro referencial, con el fin de hacerlos excluyentes y reflejar todas las ubicaciones posibles de las noticias.

Por otro lado, existen géneros informativos personalizados, como los reportajes, que no plantean la misma urgencia de los sueltos y presentan un contexto amplio de los hechos. Se aproximan más a lo que entendemos por UC. Muchos reportajes se nutren (y anulan al mismo tiempo) de una acumulación de noticias que se han hilado en secuencias narrativas interdependientes realizando ya la depuración de previsiones y reiteraciones.

Otros géneros, como las crónicas, adoptan una personalización pero, a pesar de ser consideradas como híbridas, entre la información y la opinión, tienen mayor adscripción a la pretendida objetividad manejada por la visión específica de una persona identificada. El estilo es diferente pero los objetivos concuerdan más con la fragmentación de las noticias breves que con la amplitud de los géneros de opinión.

El teletipo o despacho de agencia es el más lapidario de los mensajes periodísticos. Las agencias vigilan férreamente el principio de economía lingüística y han llegado, salvo en contadas excepciones, a una cierta homogeneidad en estructura y longitud de los despachos. Esto se debe a la ingente cantidad de palabras que genera una agencia, que incluso somete a estudio métrico, lo que las convierte en verdaderas industrias léxicas. No es extraño que la Real Academia e ilustres personajes de las Letras españolas se interesen en el control terminológico de Efe como una de las mayores empresas exportadoras de castellano.

Dada la austeridad del despacho, por una parte, y por ser portador de la fragmentación, de forma ordenada dentro de lo imprevisible de los acontecimientos, por otra, además de carecer de recursos tipográficos que desvirtúan la lectura al atraer la atención físicamente, y por estar escrito en estilo conciso informativo, el teletipo de agencia se ofrece como una unidad de registro con condiciones ideales para testar y construir en laboratorio modelos y métodos de análisis y lenguajes documentales. Claro está que tras la validación sobre muestra queda la prueba más difícil de la verificación sobre otro tipo de noticias, pero ésta última será útil para realizar correcciones sobre un prototipo más o menos consolidado. Partir de una muestra anárquica de géneros requiere mayor trabajo para establecer unidades de información de la máxima homogeneidad, de las que surgirá el desarrollo del método, algo que nos facilitaría enormemente el uso de información de agencias como primera muestra experimental.

En la información doxológica, la estructura se organiza en torno a unos objetivos, opiniones o recomendaciones que se pretenden demostrar o exponer. La argumentación está presente (*editoriales, críticas, artículos firmados*) y no se pretende ocultar. El estilo evidencia el deseo de convencer mediante la exposición razonada. En este tipo de información de actualidad, el tiempo no es vital aunque los contenidos se vinculan a hechos actuales o del pasado con repercusión actual. Personas, lugares y temas desarrollados, comparados, demostrados verdaderos o falsos confluyen en el espacio narrativo de la opinión. El análisis debe extraer la información que funda el motivo de reflexión, supuestamente conocida aún siendo del mismo día, separándola de las técnicas argumentativas utilizadas y, de las conclusiones o recomendaciones propuestas. Estos últimos datos también constituyen información, pero información asumida por un intermediario que no tiene inconveniente en descubrirse. En cualquier caso, el comentario o la crítica son tan informativos como la noticia o el reportaje (o éstos tan persuasivos como aquéllos) y, por ello, deben ser motivo de análisis para el documentalista siempre que los resultados obtenidos de la opinión, o información explícitamente marcada, sean identificables por el usuario.

Somos conscientes de la dificultad de proceder a la disección anterior. La información periodística sitúa al usuario en una posición fácil para la manipulación ya que el acceso se establece vicariamente. El documentalista, adscrito a una segunda construcción vicaria, sume al usuario ante un espejismo de la realidad. Pero tampoco el documentalista, como el usuario, puede traspasar el umbral de sus funciones, comprobando directamente los hechos: tal vez los hechos ya se difuminaron, o nunca fueron observables directamente. La información construye una realidad fragmentada pero verosímil en la que los lectores deben confiar. El documentalista compone una realidad firme a partir de retales de datos y opiniones, como si de la declaración de haber visto extraterrestres se tratara, ofreciendo una *para-realidad* en los documentos que almacena. No obstante, el documentalista no es el único en hacerlo: los individuos de una determinada sociedad alinean su comportamiento a la misma *para-realidad* que, con ayuda de los media, han ido legitimando en el subconsciente.

3.1.4 Aproximaciones a la noticia desde el análisis del discurso

El Análisis del Discurso, campo interdisciplinar en el que intervienen desde la Psicología cognitiva hasta la comunicación social, como instrumentos

teóricos capaces de dar cuenta de su naturaleza, ha sido objeto de estudio por muchos autores (con objetivos y enfoques diferentes). Por su labor sintética y la aplicación que propone, nos centramos aquí en los trabajos de Teun Van Dijk en torno a una disciplina que sitúa en un amplio marco epistemológico en el que intervienen “los procesos mentales de la interpretación y la formulación, la recuperación y el uso del conocimiento y de otras estrategias de la dimensión cognitiva del discurso (...) además de los contextos comunicativos y socioculturales (pragmática discursiva)”²⁰. Van Dijk realiza un análisis sumamente útil y aprovechable en la Documentación general y en sus aplicaciones a la prensa ya que dedica sus propuestas de modelos a la noticia, considerada como discurso.

El Análisis del Discurso es heredero de la Retórica, disciplina que ya ocupó el interés de Aristóteles, en el sentido de buscar y describir el contenido persuasivo de los textos. Nótese la importancia de este aspecto en un campo como el discurso periodístico. Actualmente, el Análisis del Discurso incorpora entre sus recursos la Retórica, disciplina que debiera usarse instrumentalmente en Teorías del Análisis documental (como análisis retórico) como paso previo a la construcción de síntesis una vez identificados los recursos que los autores utilizan en sus composiciones, a la Psicología cognitiva, al ocuparse de los modelos de producción y comprensión del discurso, del rol de las creencias, etc., la Lógica para la formalización y, en concreto, la proposicional, variantes de la Semántica (cognitiva, lógica, estructural, textual, etc) y de la Sintaxis (lógica, micro, macro, etc.) y entre otros por su interés para nuestro trabajo, la Inteligencia Artificial como Ciencia de la simulación de los raciocinios humanos (a la vez complejo cruce interdisciplinar) especialmente en sus aplicaciones a la lengua natural (Schank, Abelson etc).

Hecha esta presentación indicativa de la disciplina que vamos a utilizar en los siguientes epígrafes, pasemos a identificar las propuestas de Van Dijk que consideramos de interés para nuestros objetivos. La organización temática desempeña un rol crucial en la naturaleza del propio discurso periodístico. Este principio no entra en contradicción con nuestra hipótesis sobre los inconvenientes de organizar muestras de unidades de trabajo en torno a la temática en información de actualidad. Mi convicción es que el carácter enciclopédico y disperso de la temática periodística impide construir muestras universales por lo que es necesario acudir a recursos estructurales y categoriales, más fácilmente generalizables. Este aspecto no tiene interés para Van Dijk ya que propone su método para el análisis de textos individuales (siempre consideran-

²⁰ Van Dijk, Teun: La noticia como discurso. o.c., p.52 s.

do la importancia del contexto) aunque en Documentación aplicada las decisiones metodológicas se toman teniendo presente colectivos de autores, textos y usuarios en interacción dinámica y constante. No obstante, debo admitir la influencia decisiva de la temática en la construcción de la noticia o de los relatos, por lo que uno de los objetivos centrales del método de análisis es la captación y priorización de temas, pero en la modelización deben emplearse categorías en las que cualquier tema es representable.

Ahora bien, el propio concepto de tema no es obvio y la definición pragmática de Van Dijk no ayuda en exceso a enunciarlo: "asunto de que trata un discurso"²¹. Tema y asunto son sinónimos y a este nivel de generalidad no es posible validar una definición sin afectar los niveles específicos. Por ello, es preciso reivindicar la necesidad de un análisis del concepto de tema dentro de una Teoría Semántica aplicada a los objetivos y prácticas de la Documentación. Intuimos, anticipadamente, que los temas deben ser definidos en Documentación mediante valores que conjuguen los intereses de los autores, expresados en el texto, y la pragmática de uso, expresada en las demandas.

El concepto de tema en el discurso que nos atañe aquí es aún más complejo de aislar y se presta poco, todavía menos en el nivel de ejemplos, para edificar o justificar conjeturas y métodos de análisis. Abordamos la cuestión, por consiguiente, asumiendo esta laguna de nuestra disciplina pero con la consciencia de que su ausencia no perjudica la sistematización de elementos y procesos, especialmente si los sostenemos sobre recursos estructurales y en el nivel de abstracción. Van Dijk, además, nos da algunas claves para la comprensión de la temática discursiva (insisto, distinta a la documentológica a pesar de que coincidan en ocasiones), al introducirse en las diversas estructuras del discurso: los temas pertenecen al macronivel global de la descripción discursiva. El concepto teórico que utiliza el autor para los temas es el de macroestructura semántica, ofreciéndonos, a continuación, la siguiente definición: tema es una macroproposición subjetiva estratégicamente deducida²². Para llegar a este punto es preciso explicar conceptos intermedios que nos sean de utilidad en nuestro campo de actuación.

Los textos se articulan en proposiciones (sobre las que volveremos para su descripción en el nivel microdiscursivo). Las proposiciones son los constructos de significado más pequeños e independientes del lenguaje y el pensamiento. Pero son, también, en el nivel referencial, las unidades semánticas más

²¹ Idem, p.54

²² Idem, p.59

pequeñas que pueden ser verdaderas o falsas o que deben aportar suficiencia para poder deducir esos valores. Ahora bien, esto es sólo válido para secuencias asertivas pero el texto funciona con otros tipos de enunciados: órdenes, futuribles, amenazas, preguntas, etc que también son proposiciones que aparecen tanto en niveles enunciativos como globales de la noticia ²³.

Una macroestructura es un conjunto organizado de proposiciones. Si el conjunto es unitario podemos denominarlo macroproposición. Si son varios conjuntos, diremos que la macroestructura del texto se compone de varias macroproposiciones: cada tema del texto es representado por una macroproposición. Notemos, también aquí, que utilizamos con cierta vaguedad y en su sentido palmario el término "representación" (dar cuenta de una entidad a través de otra entidad) y señalemos la necesidad de un estudio profundo de la cuestión en sus aplicaciones documentológicas (a partir, por ejemplo, de las Teorías de la imagen, del Psicoanálisis y de otros campos en los que ha sido considerada).

En los discursos extensos suele haber varias macroproposiciones, lo que confirma la relación directa entre extensión y variable de método, que hemos planteado como hipótesis. Claro es que cualquier extensión puede ser considerada como un conjunto de segmentos, es decir, puede ser reducida a extensiones de diversa magnitud. La cuestión es poder demostrar cuál es el grado de diferencia del análisis documental sumatorio respecto al análisis documental integral de la noticia.

La macroestructura se organiza, según Van Dijk, jerárquicamente en una ordenación que subsume las proposiciones en otras más genéricas: las relaciones jerárquicas, en este sentido, se denominan macrorreglas. Las macrorreglas son transformaciones que relacionan proposiciones de nivel inferior con las de nivel superior ²⁴. A través de las macrorreglas podemos reducir (resumir) el texto. Estos elementos de la gramática del significado son elementales y necesitan ser sistematizados empíricamente para poder operar con un sistema experto que pretenda la transformación de un texto integral en metatextos reducidos. Veamos algunos conceptos que nos interesan de la metodología discursiva:

²³ Véase la delimitación que hace Raymond Colle para el discurso fáctico: Técnica de evaluación fáctica. Extracto del libro en preparación "Fundamentos y métodos de análisis de contenido de mensajes verbales e icónicos". Santiago de Chile: Pontificia Universidad Católica, Escuela de Periodismo [1991], 8 p.

²⁴ Van Dijk, T. o.c. p. 56.

- reglas de supresión: consisten en eliminar datos irrelevantes, localistas, etc., algo muy común en el nivel enunciativo. Al carecer de estas reglas, los programas de indización automática generan gran cantidad de erróneos resultados.
- reglas de generalización: son adscripciones clasemáticas a la manera de las cuasisinonimias de los Thesaurus para englobar elementos parciales.
- reglas de construcción: reducción de los detalles constituyentes de los hechos habituales que pertenecen al dominio del mundo (o que el productor presupone que pertenecen al mundo del lector). Por ejemplo, todas las situaciones lógicas que atravesamos para hacer algo son resumidas en el hecho realizado y, por lo tanto, es una reducción por obviedad.

Estas reglas son recursivas, es decir, pueden ser aplicadas sobre los resultados obtenidos en aplicaciones anteriores sobre niveles inferiores del mismo texto.

La macroestructura asignada o derivada de un texto nos introduce en otro concepto esencial del Análisis del Discurso: la coherencia global. Los temas garantizan la unidad semántica del texto, por lo tanto, una proposición que no puede ser subsumida en otra anterior o más amplia produce incoherencia²⁵. El principio de coherencia global nos hace pensar en el lenguaje documental como discurso (o metadiscurso) en el que, teóricamente, no es necesaria la adscripción temática (excepto en los lenguajes jerárquicos). ¿Quiere esto decir que los lenguajes basados en relaciones asociativas son incoherentes?. Entendemos aquí que Van Dijk se refiere a la estructura sintáctica (o macrosintáctica) y el nivel proposicional (o macroproposicional) para proponer los conceptos y no a la estructura paradigmático-sintagmática propia de un lenguaje documental. No obstante, esta cuestión deberá ser reflexionada con más detalle al abordar la fenomenología de los lenguajes documentales.

Así, el AD debe partir del reconocimiento de proposiciones empíricamente relevantes, tras un expurgo sistemático de irrelevancias, y proceder a englobamientos jerarquizados de los temas. Si el método resulta válido en Análisis del Discurso, también nos parece vago en Análisis documental, sobre todo con vistas a su modelización: la relevancia de los temas para los analistas

²⁵ Idem, p.57

del discurso es distinta que para el documentalista que lee para otros; el procedimiento se encuentra aún en niveles lejanos a una posible automatización, en términos de Pêchex ²⁶, como paso previo a la simulación. Un problema por resolver es, por ejemplo, como prever las necesidades del usuario cuando éste se interesa más por un dato o relación que por una macroproposición o por una compleja síntesis intertextual.

3.1.5 Gramática y semántica de la noticia

En una teoría sobre el análisis documental aplicado al discurso periodístico debe explicarse la naturaleza de la estructura en cuestión y, aprovechando otras aplicaciones afines, en qué y por qué la noticia es diferente para poder precisar los métodos particulares de abordaje. Tengamos en cuenta que el objetivo es realizar aportaciones a una Teoría general de la Documentación.

Intuitivamente y de una manera empírica informal sabemos que la información de actualidad en prensa escrita (nuestro objeto) se explicita y estructura de modo diferente no solamente en relación a otros discursos sino también a otros canales y medios informativos. Esta presunción inicial nos hace suponer una organización sistemática de la noticia en torno a unos criterios más o menos estables, es decir, podemos considerar una Gramática del texto periodístico (no sólo en el nivel de enunciado sino también en el de estructura semántica) que debe cumplir los requisitos propios del medio y una Semántica del mismo cuyo nivel descriptivo y categorías básicas sean más fácilmente generalizables o extrapolables en otros medios y situaciones. Contenidos y esquemas formales constituyen, por tanto, los dos ejes esenciales para la teorización sobre las estructuras del discurso periodístico con objetivos de modelizar algunos sistemas de interrogación textual para el documentalista.

3.1.5.1 Contenidos de la noticia

La suma de los temas insertos en una noticia o relato constituye su contenido o macroestructura. Notemos la diferencia sobre este punto de partida entre analistas del discurso y documentales en los intereses de ambas disciplinas

²⁶ Véase el trabajo de Michel Pêchex: *Hacia el análisis automático del discurso*. Madrid: Gredos, 1978, 374 p.

expuestos, en cuanto a la Documentación, en epígrafes anteriores y especialmente al referirme al problema de determinar la unidad de contenido documental. El periodista determina qué contenidos son de actualidad de acuerdo a la situación que posee en un marco institucional determinado (empleo institución en el más amplio sentido: lengua, cultura, ideología, etc y además en el sentido orgánico) y qué temas y subtemas (concepto que nos ayuda a jerarquizar pero que goza de la misma indefinición que tema) utiliza para informar o argumentar su exposición. No obstante, esta exposición no se ofrece globalmente ya que está sometida a la expectativa que producen los acontecimientos en el nivel de discurso periodístico. En el nivel discursivo documental esta expectativa no existe puesto que se trabaja sobre relatos ya contruidos, aunque puedan ser modificados desde el nivel primario.

En este estudio no vamos a cuestionar el importantísimo aspecto de la selección de temas y argumentos en la fase de producción y cómo afectan, los elegidos o los omitidos, al trabajo del documentalista aunque dedicamos un epígrafe a la tipología de omisiones que debemos tener en cuenta. El periodista realiza una estrategia discursiva en la producción de la noticia insertándola en un tipo de discurso que es (o cree que es) reconocido por sus lectores. La presuposición, que en términos de Van Dijk, es "el conjunto de proposiciones que el hablante supone que el oyente debe conocer"²⁷ desempeña un rol fundamental en la noticia diaria acusando numerosas situaciones de omisiones no intencionales. Todos los individuos involucrados en el proceso han adquirido las técnicas de producción o de entendimiento en procesos culturales y cognitivos previos dentro de una sociedad determinada, si bien los recursos discursivos del periodismo y de la propaganda gozan de un alto grado de utilización universal e intercultural.

Un texto periodístico ortodoxamente construido, según las Escuelas y Facultades de Periodismo o las propias empresas informativas, debe constar, al menos, de dos elementos: titulares + hechos novedosos. Lo primero se incluye también en lo segundo provocando redundancia. Un posible contexto del mismo texto se encuentra en otras noticias del mismo día en el propio medio, en otros medios, en días pasados o, tal vez, pertenecen al conocimiento de los lectores o al imaginario colectivo. Podemos diferenciar, por lo tanto, dos niveles de información en la noticia: un resumen previo realizado por el productor y reflejado en titulares (piénsese en el título como un reclamo cuya morfología gráfica es equivalente al grito del vendedor de periódicos americano de los cuarenta) y un corpus textual en el que, además de repetirse (en realidad, la

²⁷ Van Dijk, T. o.c., p.97

repetición es el título) el resumen previo utilizando la retórica, el estilo y la profusión de datos disponibles, se incluyen argumentaciones y recursos auxiliares en torno a la esencia de los hechos principales.

3.1.5.2 Títulos y macroestructuras

Incluiremos, en el apartado de títulos, todos los elementos destacados delante del corpus textual, es decir, antetítulos, subtítulos y sumario, esencialmente. En este nivel el periodista coloca lo que considera más relevante de la información realizando una estratégica labor de síntesis y priorización determinada por imperativos de la actualidad y comerciales. La omisión del corpus, por parte del lector, no va a provocar una lectura parcial del texto sino tan sólo la pérdida de argumentaciones o datos que corroboran los hechos principales leídos en los títulos (excepto en el periodismo sensacionalista en el que abultados titulares no responden a sus desarrollos posteriores por falta de información). En términos de Van Dijk, podemos decir que los títulos expresan una macroproposición en la que el periodista presupone, para su entendimiento, que el lector se sitúa en un nivel de conocimiento suficiente de ese mundo concreto.

Supuestamente, los títulos recogen la esencia de todo el corpus, aunque sabemos por experiencia que determinados párrafos son añadidos con el fin, por ejemplo, de no situarlos en lugares aislados, entre noticias de mayor envergadura, sin estar representados en la macroestructura sugerida. Tales "collage" se articulan mediante las conectivas o conectores: "en otro orden de cosas", "por otro lado", etc. en párrafos diferenciados, exigiendo la atención del documentalista, puesto que pueden contener información relevante, en todos los niveles de la noticia. Sin embargo, la tendencia general del texto ortodoxo es subsumir cualquier proposición particular de la noticia en la macroestructura provista en titulares. Esto, como es evidente, no es una exclusiva de la prensa: también en ciencias se tiende a describir sintéticamente en títulos generales, de capítulos etc., la esencia de los respectivos textos aunque es fundamentalmente en la formación de periodista donde las técnicas de titulación adquieren mayor grado de importancia. Estas técnicas basculan entre dos polos: la síntesis real del texto que asume y los fines comerciales del medio. En otro lugar, proponíamos, no sin cierta ironía, la introducción de los modos documentales en las prácticas de redacción de noticias y en la propia

organización de los textos, lo que facilitaría la disposición de los datos para el lector del diario²⁸.

Para producir títulos (y los títulos ya tienen valor en sí mismos como producto comercial del medio: véanse los índices de títulos de los periódicos de mayor audiencia) el periodista utiliza técnicas de reducción de lo irrelevante pero, cuanto mayor es la reducción mayor conocimiento precisa, por parte del lector, para situarse en el mismo eje isotópico. Esta constatación es de especial importancia para la simulación del reconocimiento de títulos por máquinas. La alimentación de la memoria, con todos los mundos y situaciones posibles, para el entendimiento de una información excesivamente sintética, sería un trabajo muy laborioso y, acaso, inviable. En consecuencia, si los titulares recogen lo más importante del corpus textual, la representación parcial de aquéllos y la descontextualización que ofrecen, en cuanto al propio relato y al mundo en general, hacen necesaria la lectura más o menos cabal de las noticias para obtener resultados fiables en el proceso de análisis.

3.1.5.3 El corpus de la noticia

Según concluye Van Dijk, tras demostrarlo empíricamente, los textos periodísticos contienen esquemas generalizados que utilizan los lectores y los periodistas para la comprensión y la producción de las noticias²⁹. Es decir, las noticias se adaptan a unos esquemas formales de presentación de los datos, llamados superestructuras, definidos por categorías y reglas que se asocian a una macroproposición. Las macroproposiciones son inscritas por el periodista en determinadas partes del texto mediante una macromorfología que se explicita en señalizaciones, títulos, ladillos y otros recursos. Sin embargo, los temas que las constituyen no son listados en un orden cronológico (excepto en necrológicas y biografías, pero esto no es información sino documentación de actualidad): la ordenación estructural de los temas es esencialmente jerárquica y, aunque Van Dijk no especifica los tipos, creemos que la pirámide invertida (general-particular) y la composición en paralelo (jerarquía de ideas colocada en párrafos independientes que posteriormente entrarán en interdependencia) conforman la mayoría de las noticias cotidianas de un diario.

²⁸ Propuesta hecha en García Gutiérrez, A. y Lucas, R.: Documentación automatizada...o.c., p.31.

²⁹ Van Dijk, T.: o.c., p.89. La demostración empírica es realizada con sus colaboradores en Structures of International News. A Case Study of the World's Press. Informe para la UNESCO preparado en la Universidad de Amsterdam en 1984.

La ordenación de temas y subtemas por el periodista no determina, en cualquier caso, el orden semántico de éstos con respecto a la macroestructura principal. Por ejemplo, es usual que los antecedentes aparezcan justamente al final del texto, que se hable de los personajes ya avanzada la noticia, etc. Queremos decir que la organización semántica de las categorías temáticas no coincide con la organización que introduce el periodista ni con la de la representación documental de la noticia. Para Van Dijk, no obstante, las macroestructuras y otras estructuras semánticas pueden organizarse atendiendo a un número de categorías fijas que incluyen las causas, antecedentes o consecuencias de orden temático³⁰. Es esencial, para el AD, descubrir esas categorías y modelizarlas en función de las posibles variables que hemos tenido en cuenta en torno a la tipología de noticias. Este recurso, las categorías temáticas generales del texto periodístico, es lo que permite teóricamente superar el reduccionismo de la coincidencia formal y formular métodos lógico-semánticos de AD. En consecuencia, las veremos nuevamente dentro de algunos de los métodos evaluados posteriormente. Van Dijk nos da una pista sobre ello: causas, finalidad, acciones, agentes, etc son categorías que aparecen con frecuencia en las noticias. Ello nos induce a testar la gramática de casos y el paradigma de Lasswell, derivada de aquélla, entre otros modelos posibles para el reconocimiento de la estructura del texto periodístico.

Para Van Dijk, los temas son temas de un texto cuando se realizan de manera efectiva a través de oraciones expresadas en las respectivas proposiciones de los textos, es decir en episodios. Dentro de los episodios se encuentra “troceado” el discurso periodístico en función de los objetivos generales del periodismo: lo principal (para el medio) aparece destacado o en primer lugar³¹. Esto obliga a que el análisis deba ser hecho en la práctica en el nivel de párrafos y de enunciados dentro de párrafos, lo que impone criterios y metáforas señalizados por el productor en el proceso de lectura. La presencia de la forma y de la estructura es, así, determinante en el método y en el resultado de la lectura documental de la noticia.

Ahora bien, si como dice Van Dijk, “la estructuración de la noticia es el resultado de una estrategia de producción”³², cómo afecta esto a la posibilidad de homogeneizar los resultados del AD es algo que debe ser profundizado mediante instrumentos semióticos y discursivos auxiliares de la Documentación que traten de simplificar las diferentes estrategias para hacerlas reconocibles

³⁰ Idem, La noticia como discurso...o.c., p.68

³¹ Idem, p.70-71.

³² Idem, p.71

por la máquina que procederá, también, con una u otra estrategia de desestructuración. El proceso documental se desarrolla en el marco de instituciones, de las que emanan las estrategias, aunque en la Teoría general de la Documentación el marco institucional debe existir sólo como abstracción y variable respecto a los métodos, pero no debe formar parte de las condiciones de sus leyes a pesar de que pertenezcan inevitablemente al nivel de la producción de los textos que esa misma teoría debe desentrañar y modelizar en sus conexiones con el Análisis del Discurso.

3.1.5.4 Esquema formal del discurso periodístico

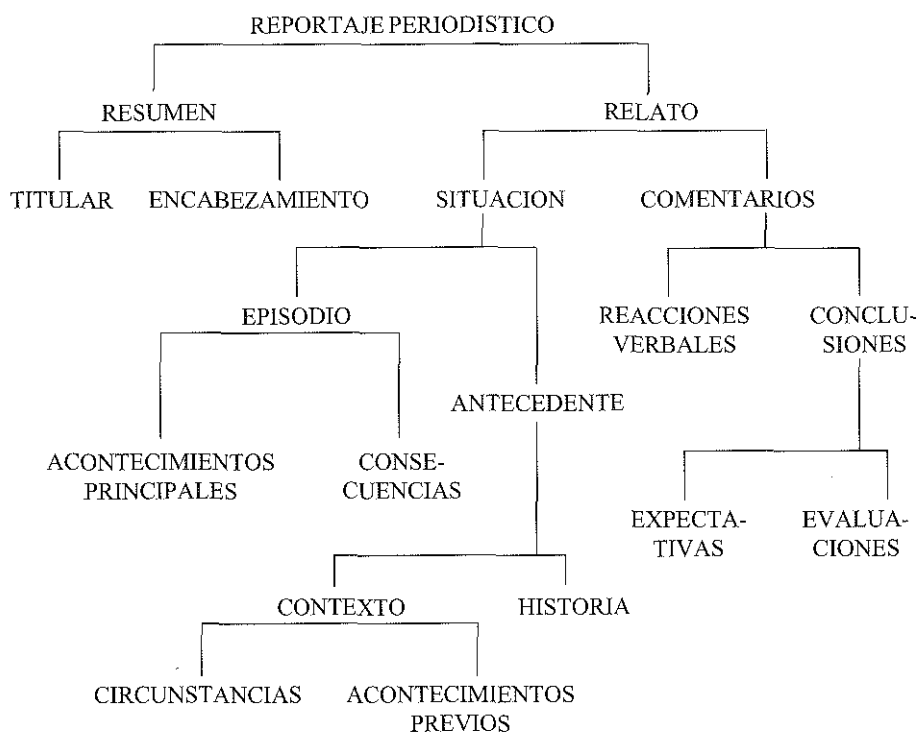
Van Dijk concede gran importancia al aspecto macrosintáctico de la información de actualidad: "las macroestructuras o significado total poseen algo más que principios organizativos propios: necesitan algún tipo de sintaxis total que defina las formas posibles en que los asuntos pueden insertarse y graduarse en el texto real"³³. Esta forma interna del texto es definida por el autor como esquema formal basado en reglas: "el esquema se basa en categorías jerárquicamente ordenadas, específicas para los distintos tipos de discurso y convencionales en las distintas sociedades y culturas"³⁴. Precisamos señalar, sobre esta decisiva aportación de Van Dijk a nuestros objetivos, varios puntos:

- Es un esquema teórico y, en consecuencia, se mencionan todas las categorías posibles aunque muchos textos sólo contengan unas pocas.
- El esquema se adapta, con la cautela anterior, a todo tipo de noticias incluidas en nuestro universo empírico de referencia.
- Las categorías no son temáticas (no pertenecen a la macroestructura) sino formales (pertenecen a la superestructura) y organizativas.
- Las categorías son recursivas (al igual que las categorías temáticas).

Veamos el esquema propuesto, que asumimos como elemento básico para elaboración de metodologías destinadas a la comprensión del texto periodístico con vistas a su reducción y recuperación:

³³ Idem, p.77

³⁴ Idem, p.78



3.1.5.5 Proposiciones

Si la relevancia de la significación documental (dictada pragmáticamente) se encuentra en el nivel de macroestructura, solamente podemos acceder a los temas globales a partir de las microestructuras: es decir, el acceso a las estructuras totales se realiza a través de enunciados y conjuntos enunciativos en los niveles menores del texto. Tomamos, para explicar el paso del micro al macronivel, el concepto de proposición del mismo autor, siempre con la precaución de nuestros distintos objetivos. La proposición se construye a partir de un predicado y varios argumentos ligados a aquél y corresponde al significado conjunto de un enunciado. Los argumentos se encadenan mediante operadores del tipo de los conectores (si bien, cuando, a pesar de, etc.) y contienen diferentes roles o casos: agentes, beneficiarios de una acción, objetos, etc. Nuevamente encontramos la presencia de la gramática de casos en estos roles esenciales de las proposiciones del discurso periodístico lo que nos

decide definitivamente a utilizarla como modelo probable. La importancia de los conectores en la articulación de las proposiciones, y de los nexos en la modulación de sus partes, nos lleva también a confirmar la necesidad de su estudio del que deben surgir aportaciones parciales a los métodos de lectura documental obtenidos mediante otros procedimientos.

Según Van Dijk, existen tres tipos de proposiciones:

- 1) atómicas, que contienen un predicado y un atributo o acción simple:
A es B, A hace B.
- 2) coordinadas: varias proposiciones unidas por conjunciones, disyunciones y yuxtaposiciones. Juan tiene hambre. Juan come.
- 3) subordinadas: las proposiciones se supeditan mediante nexos o conectores: Juan tiene hambre, en consecuencia, come.

Véase como la tipología de proposiciones sólo afecta a la forma enunciativa, es decir, puede expresarse la misma proposición mediante enunciados coordinados, subordinados, etc. Al margen de las deducciones lógicas que podamos hacer (por ejemplo, en el último caso inferenciamos una cosa de la otra) y de los matices diferenciales entre las diferentes alternativas, lo que es importante destacar aquí es la complejidad que supone para el sistema inteligente extraer la misma conclusión de diferentes modos de exposición: para el documentalista este tipo de transfert enunciativo no representa dificultades ya que acude al contexto inmediato. Para la máquina, sería preciso disponer de todas las realizaciones posibles que derivan en significados idénticos y divergentes.

Las proposiciones menores o intermedias (ya adquiridas por reducción) deben mantener coherencia respecto a las macroproposiciones dentro del texto para que puedan ser entendidas a pesar de que, acertadamente, Van Dijk señala que el periodista presupone un conocimiento en el lector que éste no posee, por lo que “el discurso no tiene coherencia sino que ésta es asignada por los usuarios del lenguaje”³⁵. El control del nivel de coherencia intratextual (micro-macro) es esencial para conseguir coherencia en y entre los resultados del análisis.

³⁵ Idem, p.96

3.1.6 *Palabra, enunciado y texto: relevancia documental*

Hemos sostenido, en varios lugares, la tesis de que la significación en Documentación se encuentra a nivel textual desde la perspectiva del sentido que concede el usuario³⁶. En efecto, las palabras y las frases pueden ser significativas pero su significación global no se confirma más que a través de la lectura del texto y en función de un uso. Así, podemos comprobar en cualquier información como muchas palabras significativas no presentan significación global o significación que interesa en Documentación (significado pragmático). A este último concepto se le conoce por relevancia.

Los enunciados, frase simple o compuesta, son altamente significativos para el lingüista (para muchos de ellos es, incluso, su objeto de trabajo), para analistas de contenido y semióticos pero no necesariamente para el documentalista: deben tener relevancia enunciativa o globalización de significado dentro del texto.

Evidentemente, la extensión de un texto condiciona el hecho de que un número mayor o menor de palabras y enunciados sean relevantes. En el flash, por ejemplo, eliminando términos coyunturales, nexos y reflexiones, casi todas las palabras son significativas. En un reportaje de varias páginas, los enunciados diluyen su significado como si fueran tan sólo palabras aisladas. En términos de Van Dijk, en los textos muy reducidos, el nivel de macroestructura coincide con el de microestructura.

Por ello, la fijación del documentalista para captar lo relevante no puede situarse, en primer lugar, sobre palabras o frases ya que se dejará influenciar por ellas. El plan de interrogación textual debe soslayar estos elementos facilitando el reconocimiento del mensaje más allá de la forma. Debemos pensar en el texto, por tanto, como una información que un productor representa con unas palabras y ecuaciones sintácticas, como estrategia expresiva, y otro productor de manera distinta. Esa es toda la fiabilidad que puede ofrecer un texto ante un plan interrogador: los resultados del análisis deben ser los mismos para la misma información redactada por diferentes individuos. Esta aseveración condiciona el acercamiento del lector al texto ya que, generalmente nos dejamos influir por la estructura de superficie.

³⁶ García Gutiérrez, A.: Suficiencia estructural... o.c. En palabras de Umberto Eco, contribución interpretativa del lector en la producción del texto: Lector in Fabula. Barcelona: Lumen, 1987, 330 p.

Cuando hemos obtenido la esencia del texto, y no fijamos límites mínimos ni máximos para éste, podremos observar sin peligro la información que nos ofrecen las unidades menores: los párrafos, los enunciados, los títulos y ladillos, los ejemplos, las comparaciones concretas, las palabras, los números. Si la aportación de estas unidades no contradice la síntesis virtual obtenida previamente, esta información pormenorizada puede ser también incorporada ya que, seguramente, será de utilidad para algunos usuarios. Este punto, prioritario en el análisis documental, plantea problemas de sumisión a las macroestructuras ya que su procedencia o ubicación presenta desvinculación respecto a la isotopía central.

En consecuencia, accedemos a la significación a través de la forma, pero en AD debe realizarse una abstracción que supere temporalmente la influencia de la misma. Una vez obtenidos resultados válidos, de representación global del texto analizado, es posible que determinados datos del nivel de superficie, ligados a los ejes esenciales de la significación sean de interés para el lector. Estamos, en consecuencia, antes dos planos de lectura:

- 1.º Lectura de captación de las proposiciones y ejes de significación del texto cuyo resultado es una síntesis globalizada que abarca las propuestas del productor y/o hechos referidos estableciendo el contexto superior y los contextos paralelos.
- 2.º Lectura de captación de datos cuyo resultado es un conjunto de conceptos pulverizados que no tienen capacidad de representación textual pero sirven para satisfacer demandas puntuales.

Según la extensión, estilo y estructura del texto estos planos de lectura pueden separarse o confundirse, situación que será visible en los resultados obtenidos: un lugar puede ser todo el objeto de un texto, una fecha puede ser el distintivo de una reflexión, etc. Las omisiones del texto periodístico son, también, elementos esenciales de significación. Veamos, dentro de la tipología general desde la perspectiva documentológica ³⁷, cuáles afectan al discurso periodístico.

³⁷ García Gutiérrez, A.: o.c., p.78-81

3.1.7 Tipología de omisiones en el discurso periodístico

El texto despliega, en la estructura de superficie, un conjunto de ideas connotadas propuestas por el autor o deducidas y, tal vez, creadas por la estructura ideológica del lector que se aplica y explicita en la lectura. Por tanto, de alguna manera la connotación es legible y puede ser adscrita a una macroestructura determinada. Pero el documentalista está también afectado por lo no dicho en el documento, lo que nos lleva a identificar una tipología de omisiones con el fin de poder observar de qué modo lo ausente influye en la integridad y en la fiabilidad del análisis realizado exclusivamente sobre lo denotado y lo connotado. Sin la categorización del cuadro de posibles omisiones en la Documentación no podrían proponerse modelos científicos de AD transportables a sistemas expertos sin correr el riesgo de que la asepsia se convierta en manipulación desmedida.

La tipología de la omisión es una contribución más al concepto de *condiciones del enunciado* o de *la producción*, cuando el productor del texto voluntaria o involuntariamente manipula, informa o desinforma y a las condiciones de consumo de información. Nuevamente debemos plantear si el alcance extensional e intensional/intencional del autor debe ser captado por el lector con funciones de documentalista en su lectura o debe, simplemente, restringirse al plano de la denotación o de las connotaciones evidentes. La inhibición, en este caso, está al servicio de la fuente. Asumir, en Documentación, la importancia de las omisiones y de la actitud del documentalista en la reconstrucción de la realidad puede suponer la descalificación de muchas propuestas técnicas e informáticas sobre AD que simplifican la lectura no sólo en lo explícito sino también en la representación mediante significantes.

Propongo, a continuación, un inventario pragmático de las omisiones que afectan a las noticias y relatos periodísticos (un inventario exhaustivo desborda el interés de la documentación) sin poder vincularlas, por falta de sistematización, a determinados usos discursivos analizados por la documentóloga Kobashi³⁸, principalmente a partir de los trabajos de Angenot, Charaudeau, Bronckart, Eco y Greimas. Las categorías de omisión descritas no son excluyentes entre sí. Según la perspectiva del autor de la omisión (productor o lector), podemos reconocer dos tipos generales:

³⁸ Kobashi, Nair: *Análise documentária: tipologias discursivas...o.c.*

- * omisión onomasiológica: producida en el transcurso de la emisión. El autor del texto las sitúa en lugares estratégicos de la cadena enunciativa.
- * omisión semasiológica: resultado del efecto interpretador producido por el receptor al presuponer, rechazar o situarse en el mismo universo conocido por el autor.

El autor del texto se encuentra, desde la óptica onomasiológica, en tres situaciones posibles de omisión:

- * omisión doxológica: deliberadamente oculta información porque desea desinformar o lo considera obvio. Podemos clasificar estas omisiones en manipulativas (lo ausente como refuerzo adicional de lo expresado) y neutrales (se da por conocido lo omitido, presenta lo omitido bajo otra forma o estructura anterior, etc.). La connotación no puede ser plenamente identificada con omisión ya que existe intencionalidad de captación por parte del destinador aunque, en muchos casos se confunden. La omisión implica una ocultación en la dicotomía signica pero también la connotación puede contribuir a la desinformación.
- * omisión agnósica: omite lo que desconoce, por tanto esta omisión es involuntaria al no tener intencionalidad aunque produce un efecto sobre el lector.
- * omisión impuesta: son omisiones ejercidas por causas ajenas al deseo del autor: falta de espacio o tiempo en un medio informativo.

En cuanto a las omisiones que afectan directamente al documento, encontramos:

- * omisiones enunciativas: corresponden dentro de un mismo texto a la estructura de superficie y, a pesar de ser las más fácilmente detectables, invalidan las técnicas de análisis automático. Estas omisiones son de tipo formal, es decir, presuponen la existencia de connotaciones o mensajes expresados bajo otros significantes no identificables por el ordenador. Por ejemplo, las omisiones por sustitución, como las causadas por sinonimia (la polisemia supondría una presencia de multisentidos igualmente indescifrable para el ordenador descontextualizado). Son también omisiones por sustitución las producidas por los canales de emisión informativa que podríamos denominar sinónimos multimedia (sustituciones de iconos por fonemas o viceversa). Mucho más compleja es la omisión estilística (metafórica,

eufemística, etc.) que no es exclusiva del discurso literario. La Ciencia está invadida de metáforas y, en consecuencia, la Documentación debe interesarse por ellas como formas del lenguaje que producen potenciales pérdidas de información en el proceso de representación sintética.

- * omisiones cognitivas: se producen en el proceso de captación de la naturaleza o de la formación de un juicio en el plano de la cognición por parte de autores y lectores. Estos no tienen los medios lingüísticos o ideológicos suficientes para materializar o percibir informaciones relevantes. No obstante, el principio de autocompactación de la superficie discursiva apenas deja visibles estas lagunas.
- * omisiones redundantes: el autor reitera determinadas palabras, frases o estructuras para encubrir la evidencia o difuminar el sentido. Transformar en obvio lo peculiar, en consumo la cultura. En muchos softwares una alta frecuencia de repetición anula la posibilidad de representación textual.
- * semiomisiones: el autor da pistas difusas (como en el cuento policíaco) para confundir al lector. La información construye una realidad metatextual que absorbe temporalmente al lector, o definitivamente si se trata del discurso de poder o de noticias fragmentadas sobre un caso polémico.
- * pseudo-omisiones: el autor hace creer que omite un tema cuando, en realidad, se refiere a ello.
- * omisiones locativas: se omite, por obviedad o intencionalidad, la ubicación territorial, temporal o situacional. Las anáforas y deixis contribuyen a este tipo de omisión.
- * omisiones clasemáticas: se producen cuando el autor no hace alusión al contexto inmediato, perdiéndose la perspectiva general (clasema ascendente) o cuando, manteniéndose en un discurso abierto, no menciona intencionalmente casos o elementos particulares que ilustrarían o aclararían el texto (clasema descendente partitivo o de especie). Estas omisiones son un recurso habitual de la información periodística debido a la presuposición (?) de que el lector continúa en la isotopía fragmentada.
- * omisiones asociativas o encadenadas: son ocultaciones de elementos afines o colaterales que pertenecen a la misma banda de profundidad del hilo discursivo, que aclararían el texto por contraposición o comparación. En ocasiones se combina la omisión por redundancia de asociaciones para

provocar, del mismo modo, una omisión de la realidad. Por ejemplo, la alusión comparativa y reiterada al modelo europeo de desarrollo como creadora de una falacia futura o para desviar la atención de determinados problemas.

Las omisiones semasiológicas ofrecen un cuadro categorial más complejo de sistematizar aunque determinadas omisiones atribuidas al autor son compartidas también por el lector. El documentalista, como se ha dicho, aparece en el proceso de contribución al texto de todo lector modelo, como coautor. El documentalista, muy a pesar del autor, no es sólo el otro polo de la comunicación que da forma personalizada al texto. Es, también, un autor secundario que transforma textos ajenos. Por tanto, atendiendo a los objetivos de la Documentación el lector/coautor no presenta el mismo comportamiento semasiológico que el lector final (a su vez lector del sistema o usuario modelo para el documentalista) que individualiza (aunque no lo cierra) el ciclo comunicativo del texto puesto que aquél realiza una lectura derivada, convirtiéndose en productor vicario que entra en competencia con la fuente y trata de persuadir a otro lector, erigiéndose como intermediario. El documentalista actúa como conmutador en un circuito de aproximaciones aparentes al discurso que finalmente resulta una realidad distinta a la pretendida originariamente: el uso de resúmenes analíticos como referencias bibliográficas en el proceso de construcción del texto científico, la cita de obras citadas, las fuentes que resumen comunicados de otras fuentes son prueba de esa afirmación.

El documentalista conocedor de avanzadas técnicas de análisis podría intervenir doblemente en la canalización de la transmisión del texto y de sus posibles réplicas diferidas. A pesar de la división (conmutadores y finales) podemos observar un conjunto de causas omitivas generales:

- * interpretación errónea: omisión de sentidos por causas específicas imputables a un sólo lector. No debe confundirse con la interpretación múltiple que es una de las posibles riquezas del texto. Esta se debe a insuficiencias cultural o lingüística.
- * lectura rápida o de zonas aparentemente neurálgicas: el lector omite una gran cantidad de enunciados al evitar la aproximación lineal. El AD debe resolver como proceder en esta lectura. En los diarios, la posición y el destaque de la noticia captará o no la atención del lector. El periódico Pravda publicaba, en 1978, una noticia de tres líneas, con aspecto de anuncio por palabras: En Roma se ha elegido un nuevo Papa de la ciudad de Cracovia. Tanto la

ubicación física de la información como el localismo geográfico impedirían que muchos ciudadanos soviéticos supieran que se trataba de un polaco.

- * omisión de lo que se cree o intuye conocer. A partir de un reconocimiento del texto, el lector llega a un compromiso o complicidad que le permite presuponer ciertos temas o tratamientos por afinidad o antagonismo ideológico con el autor.
- * sustitución de proposiciones por omisiones clasemáticas de recepción, consideradas obvias, demasiado generales o excesivamente particulares.
- * omisión, por rechazo, de argumentos que no le convencen.
- * omisión selectiva: es la más importante en Documentación al corresponder al principio de la síntesis textual. La síntesis omite para economizar la representación de modo que el documentalista silencia unas proposiciones para que otras puedan ser expresadas. A pesar de normativas estrictas sobre selección documental, la lectura aparente y las argucias del conmutador ponen en manos subjetivas el concepto de relevancia y el proceso de formación de síntesis debe basarse en generalizaciones y no en ocultaciones.

Existen omisiones no imputables a productores o documentalistas pero que afectan la comprensión del texto debido a aspectos como la producción insuficiente por falta de medios técnicos, económicos, de libertad política o de expresión, factores que generan una documentación parcial de la realidad.

Finalmente, la escasez de desarrollo de los modelos de análisis, síntesis, representación y recuperación documentales produce importantes omisiones que podríamos denominar técnicas: por ejemplo, la falta de investigación documental provoca que la simplificación excesiva de los operadores booleanos en la búsqueda o la ausencia de elementos de función en la representación generen situaciones de extrema incomunicación con la memoria automática.

3.2 Los agentes del AD en el discurso periodístico

Entendemos por agente, todo individuo que participa en la circulación de la documentación periodística. Los agentes pueden adoptar dos posturas: activa o productiva y pasiva o consumidora. Evidentemente, un agente conmuta estos

roles constantemente nutriendo o sirviéndose del sistema. Siguiendo un esquema cronológico de la circulación de documentación periodística observamos que el documentalista de prensa es el primer agente activo en aparecer: el documentalista produce (transforma) un material con fines distintos a los perseguidos por el autor del mismo contenido. El periodista pretendía informar (o desinformar), culturizar o, simplemente, entretener y, dentro del flujo documental, se convierte en consumidor de la propia estructura discursiva engendrada corporativamente. Con el uso de la documentación, el periodista genera nueva información que alimenta el sistema dentro de una especie de semiosis indefinida.

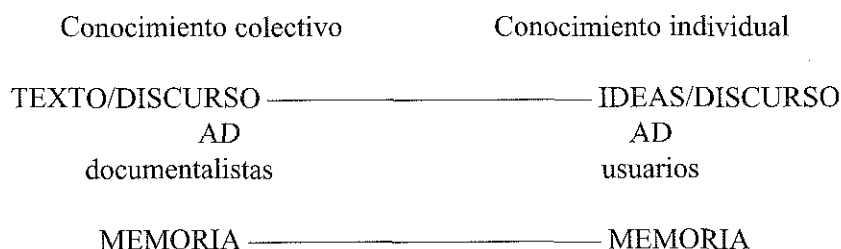
El periodista, y su postura ante los hechos, escapa del ámbito de la Documentación periodística. Los textos producidos por éste, y de interés para el sistema, suponen el universo informativo que maneja y limita el trabajo del documentalista. La participación masiva de estos agentes en la circulación de datos, el valor de cambio que tiene la información en nuestros días y la, acertadamente, denominada Industria de la Información justifican, sobradamente, las denominaciones que uso para los agentes que intervienen en los procesos documentales: productores y consumidores en el sentido más economicista de los términos. En la información de actualidad, tanto unos como otros son especialistas de modos y técnicas de interpretar y relatar pero no de temáticas concretas. Por ello, el campo de la documentación periodística no puede absorber, entre otras razones, modelos practicados en las disciplinas científicas: al documentalista de información de actualidad es al único que se le pide estar especializado en el medio más que en el mensaje. La complejidad de la estructura articulada del discurso periodístico y la incapacidad de los documentalistas de abordarlo globalmente son las causas principales del estancamiento de esta aplicación documentológica.

No debemos olvidar a los consumidores indirectos y finales de la documentación periodística: al lector o espectador que ratifica la denominación anterior y a los usuarios esporádicos que utilizan las fuentes con otros fines: búsqueda personal, investigación histórica, etc. El periodista, sin embargo, se consolida como el usuario inmediato y directo del sistema y, en consecuencia, la planificación de la memoria y las potencialidades de captación y respuesta deben ser pragmáticas y supeditarse a su perfil: no debe ocurrir, en un moderno sistema documental, que los usuarios sean creados y educados (después de provocarles adicción) en función de unos productos del sistema sino, por el contrario, que el sistema se construye en función de las necesidades y actitudes de los usuarios. El uso de una memoria es el verificador real de su diseño. Este pragmatismo impide, evidentemente, un método universal. Por lo tanto, en este

trabajo planteado en el plano teórico y apuntando al mayor grado posible de generalización, en el que no es posible especificar o individualizar, necesitamos, al menos, la figura de un usuario modelo que condicione y delimite algunos aspectos de la argumentación.

El usuario modelo nos impone el establecimiento de condiciones modélicas que aporten un cierto cariz de simulación en los resultados. Este hecho no afecta la finalidad del estudio que persigue proponer alternativas y llenar lagunas en primera instancia, para poder tener una base crítica y elaborar, desde una posición más cómoda a partir de las condiciones modélicas en un discurso determinado, métodos concretos para usuarios reales.

El segmento de los actores que participan en el ciclo documental de la actualidad queda así acotado:



Vemos que el documentalista está afectado por la presencia del macrodiscurso en su labor a pesar de individualizarla en el acto analítico. El usuario, por su parte, sólo está preocupado por su universo personal reflejado en una estructura mental que determina el contenido y la forma de lo buscado, pero debe doblegarse a la colectividad que representa y transforma el sistema interrogado.

He identificado agentes con seres humanos, pero estos roles bien pudieran ser desempeñados por máquinas inteligentes. Las reglas de captación de las máquinas sobre la información deben ejecutarse en base a cientos de miles de perfiles empíricos o especulados que configuran, superpuestos, el usuario modelo o prototipo. Pero ¿Qué interés hace que el documentalista o el sistema tengan en cuenta al usuario? Parece evidente que, junto a cientos de fisonomías psicológicas y hábitos o gustos de los usuarios, de importancia secundaria, los productores desean que sus productos sean localizados y utilizados en medio de millones de informaciones inútiles para lo que necesita conectar con sus

consumidores y conocer qué información solicitan y cómo van a solicitarla. En los perfiles de interés tradicionales, en consecuencia, encuentro insuficiente la especificación de temas, que sólo nos lleva al inventario temático, sin los modos de interrogación, que nos permite construir un inventario de categorías funcionales de expresión. El ajuste del sistema a unos fines y usuarios determina su éxito o fracaso. La propia construcción de la noticia hace del documentalista de prensa un caso peculiar:

- * realiza funciones de bibliotecario, documentalista y archivero indistintamente.
- * se le supone especialista en formas narrativas y tipologías discursivas de la actualidad pero no en contenidos (como el periodista, es un paradójico generalista especializado).
- * mantiene atención perenne, no sólo a las nuevas fuentes, sino también a las ya procesadas sobre las que el expurgo de unas referencias es la medida de la supervivencia de otras.
- * recopilador creyente de hechos y opiniones, contrastados o no y, en caso de contrastación, utiliza fuentes impersonales o métodos dudosos.
- * realiza un proceso que no depende de un discurso sino de los ritmos de periodicidad comercial del medio y del conocimiento de hechos o de la simple estimulación del medio para despertar el interés sobre los mismos.
- * practica sobre memorias de rápido crecimiento que no le permiten realizar modificaciones sobre sus métodos excesivamente concretos.
- * utiliza un lenguaje difícilmente modelizable tanto en la representación de textos como en la expresión de las demandas.

3.3 Objetivos específicos del AD en información de actualidad

Los objetivos del AD en este campo no difieren excesivamente de los objetivos generales. Solamente la peculiaridad de los textos, de los usuarios y el destino último del proceso productivo de la información obliga a diseñar adaptaciones o rechazar métodos que pueden ser validados en las áreas científicas.

Saturación y dispersión de datos y usuarios y demandas, improvisados, conforman el marco de trabajo del documentalista de prensa. Este debe proporcionar información sistematizada a individuos de comportamiento asistemático desde el punto de vista documental. Por ello, en información de

actualidad, a diferencia de otros ámbitos, se tienen que interrogar los textos en representación de todos los usuarios y demandas potenciales, menos restringidos que en Ciencias. La demanda de un periodista, por ejemplo, puede referirse a un conjunto de originales, a un análisis comparativo, a una cifra, a la confirmación de una verdad o falsedad, etc. El documentalista debe tener en su poder un marco universal de preguntas que abarque exhaustivamente cada texto en particular y que facilite la conexión de ellos en la estructura cognitiva de la memoria, o intertextualización, de tal manera que el sistema automático responda, dentro de ciertos límites predefinidos, como si fuera el propio analista, reflejando, incluso, los pasos seguidos al cuestionar y desmontar el documento.

Por ello, partimos de la hipótesis de que ningún método analítico conocido satisface suficientemente la exigencia planteada debiendo llegarse a un modelo integrado en el que cada interrogación sea excluyente y el conjunto de ellas sea capaz de abarcar cualquier texto seleccionado para la memoria.

El diseño del método de análisis debe ir acompañado por una previsión sobre los problemas de la representación de los resultados y si ésta cumple los requisitos de captación de las representaciones de la demanda.

Cada texto analizado debe explicitar, en los resultados, lo siguiente:

- contexto inmediatamente superior: unidad de contenido
- textos relacionados o encadenados: intertextualidad
- descripción ajustada al contenido: quién participa, qué sucede, repercusiones, lugar, etc.
- detalles destacados en el texto
- aportación del texto al discurso general abarcado en la memoria.

Cada uno de estos objetivos se obtiene formulando una interrogación concreta, por tanto, es necesario elaborar un cuadro de interrogaciones basado especulativamente en un usuario modelo para que el documentalista o la máquina inteligente puedan proceder a la lectura objetivada del texto.

3.3.1 *Nuevos productos para nuevos métodos*

El modelo de AD no persigue ofrecer un original o un recorte concreto como respuesta a una demanda. Esto, obviamente, ya está conseguido en las

facilidades automáticas actuales y la indización mediante coincidencia formal. El nuevo producto es la respuesta espontánea de acuerdo a una interrogación específica que el usuario deberá formular y situar en un contexto que irá definiendo con ayuda de la computadora.

Así, el método de análisis simulado por un sistema inteligente al servicio del discurso periodístico debe estar capacitado para ofrecer nuevos tipos integrados de respuesta, algunos de los cuales los podemos ya encontrar aisladamente:

- 1.º texto original, si así es integrado en memoria, o resumen de original mediante coincidencia formal de significantes o reconocimiento de sinónimos y reducción lógica de eufemismos. Enunciados simples y compuestos ofrecidos por el propio original o reconstruidos por la máquina tras un proceso inferencial.
- 2.º datos factuales, para preguntas planteadas a partir de pronombres interrogativos acompañados o no de preposiciones: Quién hace, con quién firma, para quién es, cómo ocurre, cuándo muere.
- 3.º Afirmación o negación de acuerdo a enunciados que el usuario desea comprobar: ¿Nace X en 1945? ¿El segundo apellido de Y es Z? etc. La mayoría de estas cuestiones pueden ser expresadas en las formas previstas en el punto anterior.
- 4.º Respuestas condicionadas: ¿Si A hace B, qué hace C?
- 5.º Confirmación o rechazo de verdad o falsedad con referencia al universo considerado.

En definitiva, se espera del sistema cualquier respuesta posible en la comunicación interpersonal para lo que debe cumplir varias condiciones:

- que la información sea memorizada con estructura flexible y múltiple.
- que exista información para proveer la respuesta y mecanismos de transformación de la información en unas u otras formas enunciativas.
- que el programa sea instruido con todas las preguntas posibles mediante modelos experimentales o simulados de respuesta que le sirvan de referencia, además de un sistema de reconducción de

respuestas en función del grado de satisfacción expresado por los usuarios.

- que el sistema tenga posibilidad de duda y que pueda expresarla porcentualmente.

3.3.2 *Objetivos del análisis: tipos de segmentos relevantes*

Exponemos los objetivos concretos del análisis como primera fase metodológica del proceso interactivo del AD. Entendemos por análisis, en este contexto, toda lectura pragmática de información. Las siguientes fases del AD, siendo coherentes con nuestro paradigma, se incardinan a ésta y cumplen, a su vez, objetivos específicos. El análisis es, pues, un procedimiento de lectura documental que tiene por finalidad extraer unos resultados en el nivel virtual que serán procesados en fases posteriores. Para establecer los objetivos, por tanto, debemos atenernos a dos premisas previas:

- Deben ser compatibles con los objetivos generales del AD.
- Deben presentar plausibilidad de proceso en las fases complementarias del paradigma.
- Debe existir la presencia de un contexto mínimo.

Paso a describir las condiciones de los objetivos del análisis:

- 1.º Detección de segmentos significativos del texto. Entiéndase por significación, la relevancia o representatividad de los datos en un texto. Necesitamos un enunciado para decidir si un dato o una información es relevante y si está suficientemente representada. Necesitamos el relato para conocer su aportación.
- 2.º Los segmentos significativos pueden accederse bien en la estructura denotativa bien en la connotativa. Deben tenerse en cuenta, además, algunos tipos de omisiones propias de la información de actualidad.
- 3.º Es necesaria una tipología de segmentos relevantes:

- * Un dato factual de orden onomástico, territorial, temático o cronológico siempre que presente una mínima vinculación contextual. Ej. una fecha referida al nacimiento de una persona.

- * Un enunciado: microtexto de una ley, norma, de un discurso político, etc.
- * Tomas de posición de personas o entes ante acontecimientos u objetos. Decisiones.
- * Ejecutores de acciones (animados o inanimados) entendiendo por ejecutores seres individuales, colectivos o cosas sean motivadores o pacientes de la acción.
- * Acciones ejecutadas.
- * Instrumentos usados para ejecutar una acción.
- * Lugar, tiempo y condiciones situacionales de seres, objetos o acontecimientos.
- * Cualidades o características de seres, objetos, acciones o situaciones, introducidas por los verbos de estado. Actitudes de seres individuales y entidades.
- * Aportación del texto al contexto inmediato superior; vinculación a antecedentes registrados.
- * Detección de omisiones de tipo neutral y doxológicas si el análisis es efectuado por documentalistas.
- * Finalidades y causas de las acciones.
- * Implicaciones de seres, objetos, lugares, tiempos y causas y finalidades de las implicaciones. Independencia o interdependencia de actores, roles y cosas. Efectos producidos de hecho y potenciales.

Para obtener alguno de estos segmentos, el texto debe leerse lineal y contextualmente. Como empíricamente se demuestra que la mera lectura lineal es insuficiente para detectar los segmentos repertoriados, sobre todo si es efectuada por una computadora, proponemos que, para proceder al análisis, sea preciso un modelo metodológico amplio que contemple un cuadro estructurado de interrogaciones en base a un inventario de respuestas posibles o pertinentes como el mostrado anteriormente. En consecuencia, la validación de los métodos analíticos tendrá en cuenta la mayor o menor validación de éstos a la obtención de respuestas y a la formulación de preguntas. Un proceso de síntesis posterior

hará viable la transportación de interrogantes y la presentación e interrelación de respuestas así como la demostración razonada del proceso lector efectuado por un sistema inteligente.

CAPITULO 4

4. EVALUACION DE METODOS DE ANALISIS: CONDICIONES GENERALES

El análisis es un momento del AD que consiste en una lectura pragmática del texto en el contexto de los procedimientos y objetivos del paradigma global propuesto anteriormente. Usamos, en consecuencia, el concepto de lectura como sinónimo de captación de información aprovechable para los fines documentales e insertamos el producto de esta operación en el conjunto de todas las fases del AD: el resultado de la abstracción empírico/selectiva que es el análisis debe ser, en primer lugar, objeto de descripción documentológica y, a continuación, sintetizable, representable formalmente y, desde luego, de uso social. De esta forma implicamos los principios señalados para el AD, obviamente, en su primera fase: nos servimos de los elementos del paradigma para la primera evaluación general de cualquier método propuesto.

Como ya se ha mencionado, tradicionalmente el AD se identifica con la acción de un sistema de representación. Este error ha venido condicionando los acercamientos a los modos de comprender y desmontar un texto ya que se descarga tal responsabilidad (contribuyendo al mito de la objetividad) sobre la propia fuerza estilística del autor y sus propuestas de formas, de enunciados y significantes o bien se confía en los dictámenes de un lenguaje preconstruido. Ya hemos demostrado que la fiabilidad de estos últimos es refutable en virtud de la inconsistencia del método coincidente y, en cuanto a la presencia del productor, sin caer en utopías o divagaciones, es constante por lo que la labor de filtro del documentalista es más que necesaria. Para ello: —Debe asumirse la presencia de un intermediario. —Debe aceptarse la intersubjetividad del análisis. —Debe explicitarse el modo de lectura seguido por el agente lector, hombre o autómeta.

En un sistema cualquiera no es posible la asistematicidad. La lectura es un sistema de captación de un sistema de signos (discurso) para producir otro (un metasistema o síntesis). La Documentación, como objeto, es también un sistema discursivo, formado por datos y relatos. La construcción de una memoria humana se realiza merced a la acción lectora del propio individuo quien mantiene unos intereses y unos criterios que pueden variar en determinado momento. Así, nos enfrentamos al dilema de construir una memoria que se alimenta de los resultados de lecturas realizadas por varios individuos con diferentes criterios para ser consultada por individuos distintos con, también, diferentes criterios. Parece evidente que unos criterios de procedimiento facilitarían la homogeneidad del trabajo. El método de análisis debería dar como resultado, y esta sería una forma de verificación, respuestas semejantes para agentes lectores diferentes. En caso contrario debemos rechazar la propuesta o, en todo caso, aceptar la que más se aproxime a ese principio. He aquí un segundo criterio de evaluación general. La validez, en sentido teórico, se configura como el resultado más próximo a la descripción sistemática de la intersubjetividad de la lectura y, en sentido práctico, se calcula porcentualmente. A menor porcentaje de resultados distintos más ajuste debemos reconocer en el procedimiento.

Cuando la expresión no evidencia suficientemente el contenido (y esto es habitual en lenguajes de bajo nivel de formalización) como ocurre en el discurso periodístico, el lector/documentalista precisa, de acuerdo con Menon¹ conocimientos extraídos de la experiencia del mundo (experiencia que también debe asimilar el SE) y procedimientos lógico-semánticos evolucionados a fin de poder dar cuenta de los sentidos más complejos.

Los procesos de indización automática no se dirigen a la significación a nivel de texto sino a nivel de palabra o conjuntos grafémicos, por lo que se generan falsos resultados de lectura. Por ello, no podemos considerarlos simulaciones del proceso humano sino, tan sólo, mecanizaciones rudimentarias de los textos. En función de estas premisas Joab presenta una clasificación de los modos de indización que nos ayuda a agrupar los métodos que vamos a evaluar. Para esto, parte de la detección de los problemas que deben superar los procedimientos aplicados al texto²:

1. Variables morfológicas y semánticas: anáforas, sinonimias, polisemias, etc.

¹ Menon, Bruno: Indexation automatique et Intelligence Artificielle: quelques questions de stratégie. In: "Image et IA..." o.c., p.145-175

² Joab, Michelle: Les systèmes experts. o.c., p.155-158

2. Variables sintácticas: categorías diferentes y posibles para igual significado: activa y pasiva, disposición de elementos en el enunciado, etc.
3. Ambigüedades del lenguaje: elementos que pueden adoptar la naturaleza de sustantivos o adjetivos según el contexto, contracciones enunciativas, léxias complejas, etc.

Es necesario efectuar un inventario de estos problemas en un discurso e idioma determinado para alimentar la memoria léxica del sistema experto. En este trabajo tendremos en mente qué tipo de método supera o no tales problemas como criterio de evaluación en el discurso periodístico. La clasificación, referida al caso de la indización automática, es la siguiente:

- * medios mecánicos: permiten indización libre (por palabras clave).
- * medios algorítmicos: permiten indización léxica (con thesaurus, etc.).
- * medios empíricos: permiten ponderar los resultados de la lectura y ordenarlos por prioridad conceptual.
- * medios inteligentes: permiten una indización en contexto y situación.

Los tres primeros existen y son viables industrialmente, pero Joab se pregunta si el último es factible e interesa el esfuerzo de elaborarlo. Mi respuesta a esa interrogante se basa en la solución de varias incógnitas: hasta que no se agoten diversas aproximaciones sobre el AD aplicado no sabremos si ha merecido la pena un esfuerzo que, además, ha comenzado muy recientemente; los resultados positivos o negativos obtenidos en esa línea optimizarán los procedimientos conocidos como empíricos y de rango inferior; la profundización es precisa, al menos, sobre el discurso periodístico, desde el momento en que se producen masivos procesos rutinarios que atañen a la transmisión de ideología y que afectan diariamente a millones de receptores; la experimentación contribuye al refuerzo teórico del AD y a la construcción de la Documentología.

Quiero demostrar, en los epígrafes que siguen, cómo algunos procedimientos analíticos representativos (el universo de métodos no se agota en el estudiado) testados en áreas científicas, los modos tradicionales de análisis automático y otros métodos extraídos de campos aledaños no funcionan plenamente en el discurso periodístico cuando interponemos objetivos documentales y cómo, no obstante, obtienen resultados parciales de la descripción de noticias (conjetura practicada a la hora de seleccionar los métodos presentados). Esto nos induce a pensar en la posibilidad de fusionar estos métodos en

un modelo integrado en un trabajo posterior: con ese fin, comprobaremos el problema que resuelven particularmente en torno al texto periodístico y sus niveles de complementación e insuficiencia y si existen superposiciones e incompatibilidades entre ellos. Para ello, hemos admitido previamente que la noticia tiene una estructura o parte de ella que responde a estos modelos, de otra forma, no sería posible la evaluación.

El modelo integrado debe ser formalizado en una investigación posterior para permitir la parametrización en un sistema simulado con base de reglas. El procedimiento está concebido para funcionar con instrucciones empíricas modelizadas en reglas de acuerdo al tipo de estructura textual, su extensión, género y otras variables.

Basta comprobar cada método sobre un ejemplo aislado puesto que si no cumple las condiciones, o no proporciona los resultados adecuados, es considerado inmediatamente insuficiente. No debe confundirse esta forma de proceder vía ejemplarización con la propugnada en este trabajo, uso de muestras representativas de unidades discretas, ya que estas últimas deben ser utilizadas al construir y validar un método. Sin embargo, es importante señalar qué aportación realizan sobre las lagunas detectadas en los otros. Así, la evaluación que presento se basa en criterios de globalidad de la aportación de los distintos métodos.

4.1 Métodos de análisis basados en la representación formal

La denominación que precede corresponde a los métodos tradicionales basados en la coincidencia formal de significantes. Los sistemas de representación determinan los procedimientos lectores con la consiguiente pérdida de información de lo no representado y de las distorsiones emergentes debido al ajuste forzado de etiquetas. No persiguen la detección de macroproposiciones o del *raciocinio montado por el autor* y apenas sirven para la recuperación simple de referencias. Son métodos sumamente intuitivos y prácticos entre los documentalistas que hacen recaer todo el peso del conocimiento en el azar de las formas. Se describen, a continuación, los métodos más comunes en una revisión rápida puesto que el título del presente epígrafe los invalida para nuestros objetivos. Así, presentaré una síntesis crítica para, seguidamente, detenerme con detalle en los métodos, provenientes de otras disciplinas, que comportan innovación respecto a la práctica.

4.1.1 *Indización mediante unitérminos y descriptores en lenguaje libre*

Unitérminos y descriptores son los productos de la indización tradicional realizada manual o automáticamente. Los primeros se basan en el sistema ideado por Taube en los años cincuenta: segmentación del texto en sustantivos aislados (unitérminos) considerados relevantes o altamente significativos. Nuevamente, la representación impera sobre el método de análisis porque no se explica lo que se entiende por significación o relevancia ni como el documentalista accede al conocimiento de lo relevante. Debo insistir en que mi actitud crítica no rechaza las formas de representación mediante unitérminos o descriptores sino el hecho de identificar productos con operaciones y de no preocuparse por el camino seguido desde el texto a su representación documental. Evidentemente, el ordenador realiza la lectura linealmente, encontrándose una secuencia de signos y blancos desprovista de sentido. En cualquier caso, el resultado, la ecuación de unitérminos, no sólo no representa de manera fiable la esencia de un documento apenas rozado en la estructura de superficie sino que, dependiendo de la combinación realizada con los elementos, se obtienen unos sentidos u otros:

<Centros de Documentación sobre Política Económica>

/Centros/*/Documentación/*/Política/*/Economía/

Sentidos = real (Centros de Doc. sobre Política económica)
 falso (Centros de Doc. sobre Economía Política)
 falso (Centros de Doc. de Economía y Política)
 falso (Economía y Política de los Centros de Doc.)
 falso (Política de los Centros de Doc. económicos)
 falso (Economía de los Centros de Doc. sobre Política)

todos ellos sentidos lógicos que representan falsamente el texto matriz, u originan grandes dosis de ruido en la recuperación. Este, no obstante, no es un problema que afecta al análisis (captación del sentido) sino al usuario y su resolución debe provenir de la introducción de operadores sintácticos en los enunciados.

No obstante, la representación por unitérminos debe ser revisada en aras de explotar los beneficios generados por la poscoordinación. En el caso de los descriptores, términos o conjuntos de términos (sintagmas nominales) que representan conceptos, Mooers da un salto cualitativo, se pasa del significante al significado, lo que no es realmente aprovechado en la práctica. Sólo se asimila de la idea original el hecho de poder representar mediante una ecuación de

palabras con núcleo sustantivo, un concepto. Y, en efecto, los resultados son palpables: no sólo se elimina la posible ambigüedad del unitérmino, sino que se reducen considerablemente las falsas combinaciones en el ejemplo anterior:

/Centros de Documentación*/Política Económica/

Sentidos: real (Centros de Documentación sobre Política económica)

falso (Política Económica de los Centros de Documentación)

El problema es que se sustituye el peso de la función que debe recaer sobre operadores sintácticos por la precoordinación que es un recurso de orden morfosemántico (mediante uniones morfológicas y semánticas).

En cualquier caso, el modo de indización por descriptores sólo explica cómo representar conceptos en formas alternativas, simples o compuestas, pero no cómo detectar los verdaderamente relevantes, a veces sólo connotados, entre una multitud de ellos. No abundo más sobre esta cuestión que pertenece a un enfoque distinto del propuesto en este trabajo y sobre el que se ha generado bastante bibliografía.

Nos interesa destacar, en lo que concierne a los modos de representación tradicionales que :

- Desconsideran el paradigma propuesto como globalidad, potencian-do la fase de representación y la dependencia de la operación respecto a los instrumentos.
- Están desconectados los productos de sus métodos de obtención.
- Se detectan mediante la lectura lineal de ordenador.
- Restringen la libertad de análisis al encorsetar el resultado en secuen-cias prefabricadas.
- No ha sido explotada su concepción con las aportaciones de otros campos que han resuelto problemas relacionados con la lectura y la identificación de sentido.
- Responden exclusivamente a las necesidades de los bancos bibliográ-ficos tradicionales pero no a las demandas más complejas o a los nuevos productos descritos en el capítulo 3.

- Presentan excesivas diferencias en los resultados dependiendo del agente lector y de las variables discursivas y textuales.
- El resultado no refleja la estructura, la profundidad o las aportaciones del texto al relato en cuestión.
- En realidad no podemos hablar de método de análisis, sino de una fase que afecta a los resultados de un análisis intuitivo, si es humano, y suplanta totalmente el proceso lector, si es automático.
- Ni siquiera el nivel de representación consigue tener reglas generales de normalización: los descriptores sintagmáticos, como ejemplo de preocupación constante de los documentalistas, no atienden a criterios lógicos y generales de composición.

4.1.2 *Cálculo frecuencial automático*

Los sistemas de indización automática mediante cálculo frecuencial son los primeros en aparecer. Consisten en la captación simultánea de los términos de un texto y contabilizar sus repeticiones dentro del mismo texto y con respecto a otros textos ya integrados en memoria. Si no existe una normalización previa del texto, esta tarea es imposible, por ejemplo, en información de actualidad o Ciencias Sociales. Sólo en campos de alta univocidad y estabilidad terminológica es factible la utilización. Por lo tanto, la preparación del texto por el documentalista hace inservible el análisis automático aséptico.

En cualquier caso, el hecho de repertirse un término no es síntoma de relevancia, ni siquiera de importancia. Hay gran cantidad de términos coyunturales no útiles para la recuperación. Podemos encontrar, también, términos no coyunturales, nombres de personas o temas específicos que no presentan relevancia porque ésta se determina en el nivel de texto y tal vez esos términos no están suficientemente desarrollados como para representar grandes segmentos textuales.

No obstante, no debemos menospreciar radicalmente el método estadístico en el AD siempre que vaya acompañado por recursos semánticos y lógicos. Sobre la utilidad del cálculo de frecuencia volveremos, más adelante, en las propuestas de Guiraud.

4.1.3 *Indización mediante léxicos*

En una base de datos podemos contar con una memoria auxiliar en la que habremos grabado un listado de términos de mayor o menor complejidad. Esta característica nos ayuda a diferenciar varios sistemas que pueden incorporar o mezclar las técnicas que describimos.

4.1.3.1 Términos stop

La memoria se nutre con partículas del léxico: artículos, preposiciones, conjunciones e incluso adverbios etc. El ordenador omite estos elementos en su lectura lineal, al contrastar el enunciado con el fichero de términos stop. De esta manera evitamos la extracción de cientos de términos desprovistos de significación. Obviamente los términos restantes pueden tener significación, pero no relevancia y, por lo tanto, el problema persiste.

4.1.3.2 Antidiccionarios

Bastante más compleja resulta la elaboración de un glosario negativo o antidiccionario: en él se inscriben todas las palabras que con total seguridad no interesan para la indización de textos en una área determinada. Con criterios semánticos, lógicos, científicos o documentales se eliminan de algún diccionario y de varios léxicos los términos que pueden ofrecer algún interés. De esa operación resulta un listado de términos no relevantes que es grabado en la memoria. Cuando el ordenador procede a la lectura lineal compara (match) los términos del texto con los provistos por el antidiccionario. Solamente cuando existe una ausencia en este último, el ordenador procede a identificarla como palabra clave.

Vemos, en este caso, como el sistema se ha sofisticado algo más pero aún encontramos problemas que se añaden a los vistos en técnicas anteriores, de difícil solución mediante el uso de antidiccionarios: Además de la complejidad de fabricar tal herramienta, una ausencia no significa palabra clave relevante. Los sinónimos aparecerán como distintos conceptos, el usuario no tiene forma de conocer, debido a las ausencias, cuales son los términos que debe utilizar para interrogar a la máquina. Recordemos que todos estos sistemas se basan en la coincidencia de significantes.

4.1.3.3 Lenguajes combinatorios

Incorporando un glosario de términos llenos en la memoria o, incluso, *estableciendo redes semánticas entre ellos formando*, entonces, una estructura de thesaurus, el ordenador captará de la secuencia lineal solamente aquellos términos admitidos previamente por los diseñadores del sistema. Sin embargo, un concepto relevante para un thesaurus, que es una red conceptual global, no tiene por qué ser relevante en un texto, red conceptual específica. Por tanto, esta técnica de representación frustrará a muchos usuarios que encontrarán documentos en los que efectivamente se halla el concepto buscado pero sin desarrollo suficiente o en contextos no deseados.

A pesar de la complejidad del procedimiento de construcción de un thesaurus, y de su coste, su utilización no da resultados óptimos ya que los motivos que distorsionan el proceso son de otra naturaleza. Ni siquiera un thesaurus es instrumento suficiente, con las estructuras conceptuales que ofrece, para ayudar al ordenador a captar el sentido o la importancia significativa. En muchos casos tampoco lo es para el documentalista.

No debemos, sin embargo, depreciar las ventajas que aporta un lenguaje combinatorio de este tipo, sobre todo, en las nuevas propuestas basadas en la igualdad de la relación y la explicación al usuario y a la propia Inteligencia artificial, de las razones y mecanismos que constituyen las micro y las macroestructuras. Incluso en los casos tradicionales, de rigurosa elaboración, los problemas de sinonimia o el desconocimiento de campo por parte del usuario, son superados. Podemos pensar el thesaurus como un rudimentario (pero conflictivo) sistema experto puesto que recoge conceptualmente el conocimiento de un campo estructurado por especialistas. Sin embargo, la producción de tal estructura no obedece a reglas cognitivas ni explícitas sino a criterios intuitivos y formales de donde surgen los inconvenientes para su aplicación. En cualquier caso, es evidente que un thesaurus es un potente objeto ideológico por lo que propongo dos ejercicios prácticos: describir un estado de ideas o de cosas a través de estos instrumentos de representación tomados como pulso social, económico, cultural, etc, e intentar construir metodologías para averiguar como la intuición y el sentido común han suplantado el uso de reglas sistemáticas en la construcción de estos lenguajes a lo largo de décadas de utilización.

4.1.3.4 Ponderación conceptual.

Algunos software se suministran con una capacidad complementaria: la de reconocer la valoración del concepto extraído en escalas codificadas, aunque tal ponderación debe ser efectuada por el ser humano. Este sistema evita las decepciones al recuperar información ya que, si por ejemplo, la escala valorativa es 1-10, un concepto solicitado a partir del nivel 8 recupera documentos donde cuantitativa y cualitativamente está desarrollado. Pero dos problemas alumbran aún este avance:

- No se consigue que la máquina, por sí misma, detecte acertadamente la relevancia, ni siquiera en el marco de criterios prefijados (frecuencial, thesaurus, etc.). Además, no supera los problemas inherentes a las estructuras lingüístico-documentales clásicas, que son las que utiliza.
- Continúa sin sistematizarse por qué se seleccionan unos términos sobre otros y cuál es el criterio objetivo de adjudicación de los valores.
- Este método tiene gran interés futuro en sus aplicaciones a la información de actualidad ya que, a pesar de ser una innovación continuista de los anteriores, propone el punto de partida para una nueva concepción del análisis. No obstante, todo el proceso tendría que ser ejecutado por una máquina de manera simulada.

4.2 **Métodos de análisis basados en la significación**

Los métodos que siguen, la mayoría completamente ajenos a la Documentación en sus objetivos, hasta épocas recientes, ofrecen un común denominador: pretenden captar o explicar textos grandes o pequeños, literarios o técnicos, con diferentes variables formales, desde el plano de la significación. Son métodos que proponen una aproximación al texto para obtener resultados válidos en sus respectivas disciplinas.

Nuestro interés por ellos es obvio: de estos métodos extraemos algunos procedimientos de interrogación textual a partir de los cuales podemos obtener respuestas o resultados que pudieran ser representados como elementos de recuperación. Es decir, en ellos hemos encontrado claves esenciales para avanzar en la construcción metodológica del AD.

En la selección realizada aparecen métodos que abordan el texto como unidad abierta o cerrada, que toman como objeto el enunciado o que consideran el texto como la sumatoria simple de aquéllos. Pasamos desde el estudio individualizado de conceptos al de estructura y macroestructura. Todo ello nos ha de servir para acercarnos a la formulación de los principios que deben regir los procesos del AD en su primera fase: el reconocimiento del texto.

Incluimos aquellos métodos de lectura más representativos dirigidos al análisis del texto. Por lo tanto, la forma y la expresión son meros, aunque importantes, canales por los que se accede al sentido. Los enunciados de superficie, incluso, pueden ser perjudiciales para los resultados del análisis ya que influyen con algunos usos coyunturales, metáforas, etc., en la captación del documentalista. Es, por ello, que se desea un método que interrogue el texto en múltiples formas homologando los resultados de distintos lectores y posibilitando respuestas del sistema diferentes a las provistas por los métodos basados en la coincidencia formal.

Estos métodos parten del reconocimiento de la construcción del texto y tienen en cuenta tanto la estructura, elemento que también contiene significación, como la significación misma. Dos tareas deben resolver: detectar las propuestas del autor y lograr priorizarlas lo más posible para facilitar los mecanismos de selección.

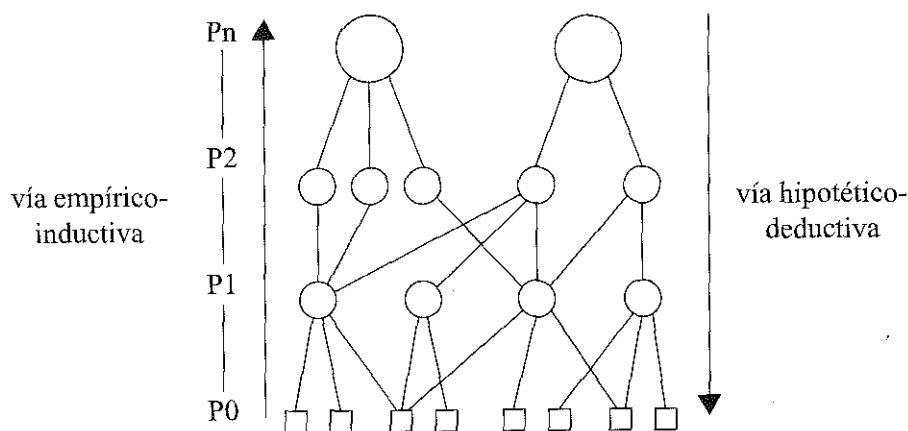
Los métodos conceptuales, en general, están en fase de experimentación, precisan de tecnología inteligente para ser realizados y, a diferencia de los logros habidos en los métodos tradicionales, presentan problemas de representación formal de los resultados del análisis para la práctica documental. Ahora bien, su sistematización puede despejar numerosas incógnitas sobre el paso del texto a su representación cuando el lector es el ser humano. Es decir, los documentalistas podrán encontrar de forma dispersa, en estos métodos, algunas reglas y criterios que les permitan explicarse y explicitar los resultados de sus lecturas.

4.2.1 *El análisis logicista*

El análisis logicista consiste en la detección de las construcciones del discurso en Ciencias humanas. Jean Claude Gardin usa este método, en sus experimentaciones actuales con sistemas expertos, aplicado al discurso de las

Ciencias que se expresan en lengua natural, que utilizan dificultosamente la formalización por imperativos de su propia naturaleza y que crean construcciones basadas en la argumentación lógica natural, es decir, no es preciso decodificar la estructura lógica del raciocinio como en las Ciencias Exactas o Experimentales. Con la denominación empleada, Gardin busca la clasificación de este tipo de análisis con respecto a otros realizados dentro del marco del análisis del discurso: análisis estructural (mitos, leyendas, etc.) muy utilizado en antropología cultural, análisis de contenido de tendencia sociológica, análisis semiológico, análisis conceptual, etc. Entre todos ellos, el análisis logicista pretende la comprensión de los fundamentos y de la arquitectura de los textos en Ciencias humanas. Es evidente, sin embargo, que mantiene algunas relaciones con los otros procedimientos aludidos. Gardin justifica la denominación "logicista" para diferenciar este método del análisis lógico, ya que logicista connota "a la manera lógica" en el sentido natural y abierto distinto del empleado en Lógica³.

El análisis logicista produce esquematizaciones de los razonamientos realizados en lengua natural (cfr. la distancia con los esquemas de Van Dijk). Gardin toma ese concepto del lógico Grize⁴ quien lo diferencia de formalización: aquél se basa en generar modelos en discursos en LN mientras que la formalización se utiliza en discursos que usan representaciones más complejas y propias de las ciencias experimentales. La esquematización es, por tanto, una formalización poco estricta y adaptable a las características de las Humanidades. Supone una reducción de las proposiciones del autor en un esquema articulado. Para ello, se parte del esquema general del razonamiento en las Ciencias humanas:



³ Gardin, J.C. et al.: Systèmes experts et... o.c., p.23

⁴ Idem, p.26

De la observación de los datos o hechos empíricos (Po) a la formación de la hipótesis teóricas (Proposiciones Pn) se genera la vía empírico-inductiva. De las hipótesis (Pn) a la observación de los hechos se produce la vía hipotético-deductiva. En ambos casos, el raciocinio se basa en proposiciones intermedias (Pi) que se articulan unas sobre las otras⁵.

El análisis logicista consiste en la observación de esas articulaciones y en la reconstrucción de las cadenas de inferencias sucesivas que recorren el texto entre los polos citados. Detectadas las cadenas inferenciales de una construcción, mediante proposiciones simples del autor, es posible simularlas en un sistema experto según han demostrado varios colaboradores de Gardin⁶ en el discurso arqueológico y Cunha y Kobashi en textos de contenido más abstracto⁷.

El análisis logicista es un procedimiento innovador para el Ad, es decir, se muestra consistente en cuanto a la realización del principal objetivo documental que es la recuperación. Sin embargo, su ámbito de acción probada son las Ciencias humanas y, concretamente en lo que se refiere a simulación, la Arqueología, campo en el que la incidencia de abstracciones conceptuales es escasa ya que proliferan los objetos de inventario y el nivel de formalización de las construcciones es más alto que en otras Ciencias.

En lo que respecta al discurso periodístico, el análisis logicista se muestra poco adecuado para esquemas de noticias netamente informativas (incluso si son persuasivas, contienen opinión, etc). ya que no es detectable el raciocinio de los productores: sólo se presentan hechos sin alegar argumentos ni estar articulados lógicamente. En el caso de editoriales y reportajes es posible la aplicación del método aunque sin garantía de reproducción puesto que los autores no atienden a criterios de argumentación repetibles de una estructura a otra, a pesar de que tales estructuras sí sean repetibles. No obstante, la detección de inferencias en las informaciones que las contengan y, sobre todo, entre las noticias adscritas a un relato, es un mecanismo que debe ser incorporado en cualquier método integrado para el análisis de la noticia a pesar de la superficialidad del raciocinio en ese tipo de discurso.

⁵ Idem, p.19

⁶ Véanse, en la misma obra, los trabajos de Guillaume, Herman, Hesnard y Zadora-Rio, en el área de Arqueología.

⁷ Cunha, I. y Kobashi, N.: Análise documentária e Inteligencia Artificial. o. c. Las autoras comprueban las propuestas de Gardin sobre un texto de gran nivel de abstracción conceptual (influencia del entorno en la construcción de la razón)

4.2.2 *El modelo de Lasswell*

Como proponíamos en un trabajo anterior⁸, las conocidas 5W del modelo de Lasswell en el mundo periodístico responden a la manera en que los periodistas se acercan a los acontecimientos y organizan sus textos y relatos en el estilo informativo del periodismo occidental contemporáneo. Por ello, parece evidente, que el lector que interroga el texto con las mismas preguntas debe obtener las informaciones relevantes suministradas por el autor. Si ampliamos el espectro de interrogantes: Qué (acción), Quién (agentes), Cómo (modo), Dónde (lugar), Cuándo (tiempo), Por qué (causa), Para qué (finalidad), A través de qué (medio), Con quién o qué (asociación), etc. obtenemos todavía una mayor comprensión de la significación del texto. De hecho, si desmontamos una noticia con estas categorías y escribimos un resumen de ella a partir de las respuestas, los resultados obtenidos reconstruyen, a nivel metatextual, el texto de origen de manera fiel. El lector puede, incluso, reconocer determinadas omisiones al no obtener respuestas.

Aunque este método ha sido testado sobre sumarios y despachos comprobándose su buen resultado, la variable extensional condiciona las respuestas. En textos de gran extensión, las respuestas se superponen (ej. de varios quién, qué, etc.) lo que hace compleja la priorización de las mismas. Además, no todos los estilos y géneros periodísticos se adaptan por igual a esa estructura por lo que el modelo es operativo en textos breves condicionados al desarrollo tradicional de datos en estilo informativo.

El agente analista debe ser, forzosamente, el ser humano mediante lectura contextual. Las máquinas sólo podrían ejecutar este método si poseen criterios de reconocimiento funcional y lógico-semántico para detectar la importancia de estos elementos, estén o no los mismos en el texto y esté o no explicitada su relevancia.

4.2.3 *El modelo de casos*

Los trabajos de Fillmore y Pottier⁹ sobre casos conceptuales de la lengua fueron tomados por Cunha como hipótesis metodológica para la construcción

⁸ Desarrollado en García Gutiérrez, A. y Lucas, R.: Documentación...o.c., p.47-49

⁹ Pottier, B.: *Lingüística general*. Madrid: Gredos, 1976, p. 135-146

de vocabulario¹⁰ en el área de política colonial. Los resultados fueron demostrados inoperantes en otras áreas, incluso de las ciencias sociales, aunque el modelo de casos debe ser aceptado como uno de los posibles modos estructurantes de los lenguajes documentales¹¹.

En el presente trabajo, sin embargo, nos interesa la aportación de Cunha en otro sentido: evaluar la adaptación que hace la autora de las tesis de Fillmore en su aplicación al discurso periodístico y, en concreto, a sus posibilidades de operación, es decir, si el análisis del texto, desmontado mediante categorías que funcionan como interrogantes, responde a las expectativas que el autor y el usuario han depositado en el texto. Por lo tanto, vamos a averiguar si pueden ser utilizados los casos en AD aplicado a la información de actualidad, con qué restricciones y si es posible la simulación de lectura basada en los mismos.

Para Fillmore, las nociones de casos incluyen un conjunto de conceptos universales que identifican cierto tipo de juicios que los seres humanos suelen hacer sobre temas tales como quién hizo, con quién ocurrió o qué cambió¹². De los once casos conceptuales propuestos para las lenguas naturales, Cunha realiza una selección y síntesis en función de los objetivos pragmáticos de la Documentación. Así, algunos quedan subsumidos y, otros, desdoblados o descartados:

- El Agente que practica o sufre la acción se distribuye en diversos casos: ergativo, nominativo, dativo, agentivo, acusativo, asociativo y beneficiativo. Para Cunha, esta pulverización de lo humano-agente dificulta la formación de categorías unitarias de conceptos, interesa la formación de una categoría en torno al atributo humano más que distinguir entre los roles actanciales en primera instancia. La pregunta práctica sería: ¿Quién ejecuta o sufre la acción?
- *Instrumental y causativo pueden confundirse en lo no-humano, por lo tanto, Cunha los fusiona en la pregunta ¿Qué medios permiten a la acción consumarse?*
- El caso locativo asume, para Pottier, tiempo, espacio y situación: ¿Dónde y cuándo ocurre la acción?. Como se ve, esta pregunta no recoge la noción de situación en sentido abstracto, diferente de lugar.

¹⁰ Cunha, I.: Do mito... o.c. p.111-137

¹¹ García Gutiérrez, A.: Estructura lingüística...o.c., p. 114-116

¹² Cunha, I.: o.c., p.115

- El caso acusativo puede considerarse, según Cunha, como modo de relación entre la acción y la no-acción y controlarse con la doble pregunta: ¿Qué objeto sufre o motiva la acción? y ¿Cómo se realiza la acción?. El acusativo se desglosa, en consecuencia, en dos categorías: objeto y modo.
- Los casos dativo, finalidad y beneficiativo, extraído anteriormente el rol humano exclusivo para el caso Agente, se transforman. Así, el dativo queda suprimido al comportar ser animado. El beneficiativo se obtiene con la pregunta ¿Cuál es el beneficiario, producto o resultado de la acción?. La pluralidad del interrogante lleva a Cunha a proponer la categoría Producto. La finalidad, equiparable a intencionalidad, rasgo que diferencia este caso del Producto (intención y no sólo ejecución o competencia) se obtiene mediante ¿Cuál es el fin o intención de la acción?.
- El asociativo es integrado por Cunha en la categoría Agente considerándolo inadecuado para formar bloques conceptuales en un vocabulario. No obstante, a efecto de análisis, la pregunta ¿Con quién o en compañía de quién? parece importante y aislable respecto al ¿Quién?.

Este es el cuadro de síntesis propuesto por Cunha:

Instrumento (no humano, concreto, material). Qué medios.

Agente (humano). Quién practica o sufre la acción.

Objeto (no humano). Qué sufre o motiva la acción.

Modo (no humano, competencia). Cómo se realiza.

Lugar (tiempo, espacio, situación). Dónde y cuándo.

Producto (no humano, material, competencia).Cuál es el resultado.

Finalidad (no humano, material, intención).Cuál es la intención o fin.

En el test realizado sobre diversos recortes de prensa y despachos de agencia puede observarse que la aplicación del modelo de casos en la interrogación de textos periodísticos genera un alto grado de respuestas en cuanto a lo que consideramos información relevante desde el punto de vista documental. Sin embargo, podemos hacer varias objeciones en cuanto a su aplicación incondicional:

1. El método no sirve para el análisis simulado, en tanto no se resuelvan problemas previos: reconocimiento artificial de las relaciones significante-significado y de las categorías funcionales puesto que las respuestas a los casos

se basan en una lectura sobre la estructura de significación no reproducible por máquina.

2. En el caso de solución semántica y tecnológica a las lagunas mencionadas o en el caso de que el método fuera utilizado exclusivamente por documentalistas, encontramos los siguientes problemas:

- a) El interrogador “qué” (objeto) no se corresponde al what (qué) lasswelliano que busca la acción de la noticia. Por tanto, la propuesta de Cunha no detecta la acción o acciones que determinan las preguntas y se produce confusión del caso objeto en cuanto a la organización que los periodistas hacen de sus textos en torno al “qué”.
- b) Se aprecia cierta dificultad de categorización de conceptos en forma excluyente. Por ejemplo, la extensión del texto determina que un concepto desempeñe varios roles en el mismo. Además, el lector no ve claramente la diferencia entre objetos, instrumentos o modos en determinados textos de redacción compleja. Si se da respuesta errónea a un caso todos los demás sufren una rotación falsa.
- c) Algunas respuestas pueden tener interés para objetivos del análisis de textos, pero son irrelevantes para la recuperación.
- d) Algunas respuestas son difícilmente representables, al menos, en los sistemas de representación documental convencionales, ya que los resultados adoptan formas enunciativas complejas o explicativas.
- e) Espacio y tiempo son locativos esenciales en la actualidad, por lo tanto deberían ser separados y desglosados en distintos operadores (nación, accidente geográfico o geopolítico, años o periodos, etc.). La situación requiere una pregunta específica: ¿En qué situación ocurre la acción? El resultado sería, por ejemplo, el marco que envuelve un hecho en torno a un personaje.

En consecuencia, es aceptable la interrogación mediante la propuesta de Cunha, siempre que se haga una revisión en función del discurso periodístico y una fusión con el modelo de Lasswell, si bien la puesta en práctica artificial ha de pasar por la resolución de múltiples problemas intermedios como nuevas formas de estructurar la información o la búsqueda del reconocimiento artificial de las categorías en el enunciado.

4.2.4 *El modelo facetado: los campos conceptuales*

En terminología tradicional de lenguajes documentales, denominamos facetas a vórtices conceptuales en torno a los cuales se organiza el vocabulario creando un ensamblado de estructuras. El campo conceptual sería una estructura de significación organizada, a su vez, en función de una temática o perspectiva. Enfoque y asunto son, en consecuencia, los distintivos de las facetas y de los campos conceptuales respectivamente. Lo que nos interesa destacar aquí es cómo puede aprovecharse la faceta o el campo como método o marco del método de reconocimiento textual, es decir, cómo podemos convertirlos en operaciones extrayéndolos de su vinculación constante a la representación por su condición de elementos instrumentales. Un sistema de indización automática mediante thesaurus usa criterios selectivos en función de las representaciones.

Tanto un lenguaje documental clásico o los nuevos lenguajes, cuyo método de construcción proponemos en otro lugar¹³, conciben algún tipo de organización del vocabulario en unidades estructurales amplias. Tales unidades son generadas en una especialidad de acuerdo a los criterios de un especialista y siempre en torno a un código. Parece, por tanto, que podemos referirnos a estos instrumentos como sistemas expertos, es decir, son representaciones de cómo el experto entiende o piensa un universo de ideas y consigue explicitarlo en forma de mapas conceptuales. Si el experto es capaz de construir ese sistema usando conceptos, enunciados y relaciones diversas entre ellos, es lógico pensar que ese esquema, esa idea del mundo que plasma, por ejemplo en un thesaurus, es la que marca su entendimiento y sus tomas de decisión en la lectura. Por tanto, el texto, universo organizado por otro experto, se somete en la lectura a esas facetas que el documentalista hace o utiliza a partir de un lenguaje fuera de contexto. He ahí como el instrumento se convierte en acción, pero no queda claro cómo puede salir de su reducido papel verificador y dar el paso de convertirse en un sistema de reconocimiento de estructuras enunciativas en contexto a partir de estructuras conceptuales descontextualizadas.

Una máquina que hubiese sido instruída con miles de campos y facetas abiertos e interrelacionados con complejas reglas basadas en procesos inferenciales dispondría de una "mente estructurada" capaz de ayudar a reconocer estructuras textuales en los procesos lectores, pero no parece evidente que puedan suplantarlos. Es decir, su método de interrogación no se basará en

¹³ Las propuestas metodológicas para la construcción de nuevos lenguajes documentales, especialmente el método vectorial, se encuentran en García Gutiérrez, A.: *Estructura lingüística...* o.c. cap.5

preguntas sino en comparaciones múltiples entre lenguaje/sistema y texto de la misma manera que el sistema puede dar cuenta de las incorrecciones del habla.

En el caso del discurso periodístico, un lenguaje documental como sistema experto unido a un dispositivo interrogador parece una condición necesaria aunque esta hipótesis debe ser contrastada en corpus testados empíricamente. La justificación se basa en que la superficialidad del texto periodístico, su brevedad habitual y la simplicidad de su sintaxis pueden ser abarcadas sin grandes problemas por un método de reconocimiento, si bien la complejidad ideológica y la tendenciosidad aparecen como inconvenientes difícilmente superables para la lectura automática ya que los expertos no estarán de acuerdo, entre ellos o con la audiencia, sobre los escoramiento de la noticia. La connotación deberá ser abandonada en los primeros ensayos, a favor de lo denotado.

En este modelo precisaríamos del aparato conceptual de la psicología cognitiva en su interface con la Inteligencia artificial: así, primitivas, escenarios, guiones, situaciones y macrodescriptores son elementos que deben ser desarrollados experimentalmente e incorporados a nuevas estructuras de lenguaje de tal manera que permitan reconocer en sus articulaciones y léxico cualquier enunciado del autor del texto.

4.2.5 *El modelo del análisis del discurso*

El modelo adoptado para describir la estructura de la noticia en el capítulo 3 procede de investigaciones empíricas realizadas mediante técnicas del Análisis del Discurso por el holandés Teun Van Dijk y algunos colaboradores de la Universidad de Amsterdam. Aunque debemos observar que el esquema no se ajusta a todo tipo de género periodístico, especialmente a aquéllos de corte persuasivo o de estilo de sollicitación de opinión, responde adecuadamente a la estructura de la noticia en el periodismo escrito actual y occidental por lo que podemos considerar el modelo como ampliamente generalizable.

Teniendo en cuenta sus limitaciones, en consecuencia, veamos cómo la jerarquización de la información articulada mediante macrorreglas por Van Dijk puede funcionar como metodología de lectura de noticias, qué resultados se obtienen de su aplicación y cuáles son sus principales inconvenientes o insuficiencias.

En un trabajo de investigación, presentado como tesis en septiembre de 1991, la profesora Amaro¹⁴ de la Universidad de Sao Paulo realiza un test de la propuesta de Van Dijk sobre un corpus de cinco textos periodísticos para obtener la macroproposición global, síntesis del texto, que la investigadora equipara al concepto de información principal en AD. Pasamos a describir el experimento y sus resultados y analizar por qué existen algunas lagunas de tipo conceptual y operativo para poder ser usado en corpus abiertos de noticias. Las fases de análisis a las que el texto debe ser sometido, son las siguientes:

- 1) Eliminación de proposiciones a través de operaciones selectivas:
 - Supresión de predicados atributivos.
 - Supresión de informaciones constitutivas mediante:
 - a) Localización de acontecimientos y acciones.
 - b) Razones y consecuencias de los acontecimientos y acciones.
 - c) acciones preliminares y auxiliares.
- 2) Sustitución de proposiciones específicas por proposiciones supraordenadas a través de operaciones constitutivas:
 - a) Sustitución por generalización simple.
 - b) Sustitución por construcción propiamente dicha.
- 3) Concretización de la macroproposición global del texto.

Tras efectuar el experimento, Amaro llega a realizar una validación de los resultados obtenidos comparándolos con los obtenidos a su vez por un documentalista experto utilizando los criterios intuitivos y de sentido común y con los generados por el sistema de indización PRECIS¹⁵. Entre las conclusiones de la autora, destacamos las siguientes:

- El concepto de macroproposición de Van Dijk coincide con el de información principal en AD.
- Las macroproposiciones se encuentran en el sumario o lead de las noticias que usan estructura narrativa del tipo pirámide invertida.

¹⁴ Amaro, Regina: Contribuição da análise do discurso para à análise documentária: o caso da documentação jornalística. Sao Paulo: ECA / USP, 1991, 87 p. + anexos.

¹⁵ Fujita, M.S.L.: Precís na língua portuguesa: teoria e prática de indexação. Brasília UNB/ABDF, 1989.

- Existen coincidencias entre los resultados obtenidos por sentido común y por el método usado, por lo que este supone un modo válido de objetivar y explicitar procedimientos de análisis y síntesis.

Sobre estas conclusiones es necesario exponer algunas observaciones:

1. El sistema de validación usado es insuficiente, habida cuenta de que el corpus no es representativo, ni en número de unidades, ni en temática y la elección de un documentalista o de PRECIS para la comparación tampoco es definitivo porque no se justifica la utilización de un sólo lector o de un sistema con muchas carencias, suficientemente. Parece más adecuado, a falta de una metodología de validación en el área, comparar resultados con otros métodos a saber, modalidades, casos, lenguajes documentales, etc., para observar la precisión o la mejora que aporta el propuesto respecto a los demás.
2. El concepto de macroproposición se equipara, globalmente, al de información principal sin haberse encontrado una acepción convincente y general en Documentación. En cualquier caso, de acuerdo a los resultados, discrepo del sentido que se le da ya que es la Pragmática, por tanto el uso y no el documentólogo, lo que hace importante o no una información. Así, identificar información principal con síntesis del texto no parece adecuado en el discurso periodístico ya que para el usuario un dato aparentemente secundario, una comparación, etc., pueden constituir la información principal del texto.
3. La errónea equiparación se debe a que la autora considera como fines de su estudio la obtención de referencias convencionales, productos claramente insuficientes para satisfacer la demanda de documentación periodística. Por lo tanto, la generación de síntesis mediante este método comporta una excesiva pérdida de información relevante, aunque confirmamos la eficacia para obtener el sentido global del texto. En cualquier caso, la aportación de Amaro es esencial para comprender el paso del texto periodístico a su representación sintética por el documentalista y apuntar determinadas áreas previas y colaterales de urgente investigación.

4.2.6 *Método cuantitativo-semántico (Guiraud)*

En su trabajo "Las estructuras elementales de la significación", Guiraud¹⁶ propone un método de análisis textual para la detección de la significación, a través de un estudio cuantitativo del nivel de superficie. Una vez elaborada una lista de los sustantivos relevantes de un texto, se observan las repeticiones para reconfigurarla. Sobre ese corpus significativo que representa fragmentadamente el texto, Guiraud establece tres zonas que denomina arritmo-semánticas:

- Núcleo de conceptos-clave que representan globalmente el discurso, cuyo índice de repetición es notablemente mayor al de otras zonas constituyendo un eje isotemático.
- Zona de palabras-tema que constituyen la esencia del texto.
- Zona de palabras de baja frecuencia, poco relevantes o coyunturales.

El procedimiento es el siguiente: se señalan, linealmente, las palabras con alguna significación estudiándose la frecuencia de repetición. Se dividen los sustantivos en tres categorías: Claves, temas y de caracterización. El sustantivo clave es el de mayor frecuencia. Los sustantivos-tema tienen una frecuencia cercana a la anterior pero su papel consiste en reforzar la coherencia del propio texto. Los sustantivos de caracterización son los restantes.

Sobre esta metodología que pretende objetivar la significación, podemos objetar lo siguiente:

- No se explicita la condición del texto analizado ni su extensión. Se supone que esta última determina un resultado u otro, incluso, la aplicabilidad o no del método.
- Se basa en lectura lineal, a la manera de la indización automática convencional, despreocupándose por los roles enunciativos de los sustantivos.

¹⁶ Guiraud, P.: Les structures élémentaires de la signification. "Bulletin de la Société de Linguistique de Paris" 60. 1965, p. 97-114. Véase también: Sens et information. "Colloque sur la valeur des méthodes d'analyse quantitative en Linguistique. Strasbourg, 20-22 abril 1964

- Sólo se admiten sustantivos, lo que supone una pérdida de información transportada en otras categorías léxicas.
- Deben ignorarse eufemismos, metáforas y otras figuras al realizar el recuento.
- El texto periodístico es un terreno propicio para este método dada su extensión (editoriales, reportajes), aunque en sueltos y despachos las frecuencias son poco significativas. El análisis, de este tipo, sobre noticias, queda sometido al lenguaje del productor que propone un campo de referencia que no es siempre el más indicado para la fiabilidad de la indización.
- El estilo informativo fluye mediante palabras altamente significativas, fácilmente inventariables, pero muchas de ellas son neologismos o modismos imposibles de controlar.
- La presencia de determinados sustantivos en superficie no es síntoma de relevancia documental.

4.2.7 *Método de las modalidades (Greimas)*

Al proponer un procedimiento del análisis del discurso en Ciencias humanas que persiga la detección de las formas y tipologías discursivas y explique las condiciones de este, Greimas, en su teoría de las modalidades¹⁷, ofrece algunos elementos aprovechables por el AD. Para el autor francés, el predicado es una función que puede ser investida de unos mínimos semánticos, lo que permite establecer la distinción entre dos funciones predicativas, hacer y ser, y proponer dos formas posibles de enunciados elementales: enunciados de estado y enunciados de acción. El ser atribuye una propiedad y el hacer supone una transformación.

Cunha realiza una adaptación de la propuesta greimasiana, provocando un recorte de acuerdo a los fines del AD: “A los enunciados de estado corresponderían las macroproposiciones semánticas / trazos descriptivos con

¹⁷ Greimas, A.J.: Pour une théorie des modalités. “Langages” 43, 1976, p.90-107

mayor capacidad de traducir el sentido total del texto”¹⁸, hipótesis que llega a demostrar sobre un texto de política colonial portuguesa, dentro de un contexto histórico más amplio.

El método, sin embargo, fue testado por Kobashi, quien comprobó sus principales defectos que lo hacen inviable fuera del corpus en que fue validado¹⁹. Los argumentos empleados sirven, también para el discurso periodístico:

- Existen enunciados de estado que no contienen información relevante.
- La transportabilidad de formas enunciativas de estado y acción, es carnívoro = come carne, plantea problemas de operación.

El procedimiento a seguir es :

- Se segmenta el texto en enunciados que contienen los verbos ser o estar. Se aísla el resultado.
- Se segmenta el texto en enunciados que contienen verbos que implican acción y se aíslan los resultados.
- Se realiza la misma operación con los verbos reflexivos y recíprocos.

Aplicado sobre noticias de un diario, se observan otros inconvenientes:

- Los resultados de la segmentación no son válidos para la recuperación en múltiples ocasiones.
- Al realizar la lectura, debe imperar un criterio de sentido común e intuición, lo que hace compleja la simulación, al no tener instrumentos suficientes para explicitar el proceso.
- Los verbos ser y estar pueden comportarse como verbos de acción en enunciados como “estar haciendo”.
- El verbo haber comporta una problemática distinta, no recogida por los dos tipos descritos.
- El dinamismo de la información de actualidad opta por los enunciados de acción sobre los declarativos.
- El enunciado de estado se encuentra, sobre todo, en el género editorial y no siempre representa elementos relevantes de la argumentación. El hecho de pertenecer o estar vinculado a un enunciado de estado no convierte en relevantes a las categorías léxicas respecto al texto total.

¹⁸ Cunha, I.: Do mito...o.c., p.51

¹⁹ Kobashi, N.: Análise documentária: considerações sobre um modelo lógico-semântico. o.c., p. 45-57

4.3 Otras aportaciones metodológicas

Describimos, a continuación, algunos métodos, categorías y elementos propios del lenguaje que pueden servir parcialmente para el enriquecimiento de un modelo integrado de análisis, teniéndose en cuenta que sus contribuciones son mucho menores que las de los denominados métodos semánticos puesto que, en su mayoría, consideran parcelas muy pequeñas de la enunciación.

4.3.1 *Conectivas lógicas*

Los conectores o conectivas lógicas establecen relaciones lógicas entre enunciados determinando, generalmente, una consecuencia a tenor de un antecedente. Por ello, son de especial interés para la Lógica proposicional que, no obstante, se centra en cinco: negación, conjunción, disyunción, implicación y equivalencia. El resto de conectivas, ampliamente recogidas por Pêcheux en su propuesta de análisis automático del discurso²⁰ no han sido objeto de estudio de esta disciplina²¹.

En Documentología nos interesan las conectivas en dos diferentes planos:

- para estudiar el rol que representan al formar enunciados documentales compuestos y la articulación entre ellos.
- como ayuda para desmontar automáticamente los enunciados con vistas al análisis documental.

Al ser objeto de estudio de la Lógica, entre otras disciplinas, debemos acercarnos a esta Ciencia teniendo en cuenta, sin embargo, que el objeto de estudio compartido representa una intersección muy pequeña y poco investigada respecto al resto de sus respectivos corpus. Es necesario, por lo tanto, comenzar por un recorte y señalar el aparato teórico conceptual que puede ser aprovechado en Documentación y sus aplicaciones.

²⁰ Pêcheux, M.: o.c., p.191-198

²¹ Para este epígrafe ha sido consultado, por la ventaja del recorte ya realizado a efectos de la Lingüística, el libro de Allwood, Jens et al.: *Lógica para lingüistas*. Madrid: Paraninfo, 1981, 203 p., especialmente el cap. 4.

Un problema evidente de la aplicación de la Lógica proposicional a la Documentación del discurso periodístico es centrarse en enunciados y, más concretamente, en las relaciones entre enunciados, alejándose del sentido del texto/discurso. No obstante, debe reconocerse la aplicación de esta disciplina en la formación de las bases de reglas de los sistemas expertos que manejan lengua natural. La Lógica no se preocupa de la relación de un enunciado con la realidad (verdad real) sino sus relaciones internas (verdad lógica), es decir, no es de interés lo que es realmente verdadero sino lo que tiene que ser verdadero si las premisas son supuestamente verdaderas. De acuerdo a la misma limitación: ¿Debe desinteresarse el documentalista por el contexto?. Hemos argumentado la respuesta negativa en otro trabajo²².

La proposición como objeto del análisis lógico no puede ser objeto del AD ya que la segunda fase del proceso, la síntesis, sería inviable tras un AD de todos los enunciados de un texto. La macroproposición y la isotopía son, por el contrario, ejes de interés de la metodología y de la Teoría del AD. En ese sentido interesan más las aportaciones de la Lógica aplicada al texto/discurso aunque sus avances han sido menores. En cualquier caso, el AD precisa de la Lógica sus sistemas de notación para representar enunciados y trabajar más fácilmente en laboratorio con lenguaje formalizado: vocabulario (variables de enunciados, constantes lógicas o conectivas) y reglas (cercanas al generativismo)²³. La notación lógica permite representar la estructura de los enunciados mediante símbolos o diagramas arbóreos y cuadrados lo que es de gran utilidad para el enunciado documental.

Las conectivas tienen una presencia constante en los textos y recogen algunas claves de la cadena argumental y de las condiciones de producción puesto que se presentan en la articulación de frases simples a nivel sintáctico y es la piedra angular de la argumentación a nivel semántico. Un estudio contextual de las conectivas de un texto puede informarnos acerca de lo que pretende decir o demostrar el autor. Por tanto, son esenciales para desmontar determinadas argucias del discurso del productor con el fin de conocer más claramente sus propuestas y poder sintetizarlas. Sin embargo, en función de los objetivos del AD, una metodología basada exclusivamente en la detección de conectores presentaría los siguientes problemas:

²² Suficiencia estructural...o.c. al referirme a la responsabilidad del documentalista en la lectura y transmisión diferida de lo leído.

²³ Allwood, J. et al.: o.c., p.57 s.

- Es un procedimiento inviable para el análisis humano ya que requiere demasiado tiempo para obtener una información insuficiente. Automáticamente es viable combinado con los métodos semánticos, después de solucionar problemas teóricos que culminen en un inventario exhaustivo de sinónimos y encontrar procedimientos para el reconocimiento artificial.
- Existe un gran abuso de los funtores en la lengua natural, en especial en el discurso periodístico, usados como muletillas o de forma irrelevante.
- Presentan un fuerte localismo enunciativo cuando al AD le interesarían las articulaciones que producen funtores de tipo textual (macrofuntores).

4.3.2 *Análisis sintáctico*

La detección de los roles funcionales de sujetos y complementos y su comportamiento en la oración ha sido una preocupación de la Gramática clásica. Sin embargo, la contribución del análisis sintáctico al AD es de escasa incidencia puesto que la información obtenida de enunciados simples y compuestos no representa las macroproposiciones del texto. En cualquier caso, para la información de tipo factual el análisis sintáctico ofrece un modo de estructurar la información al detalle, de gran utilidad, siempre que se combine con un análisis gramatical a nivel de texto (los casos, por ejemplo).

Debemos destacar la necesidad de inventariar las categorías funcionales del enunciado con el fin de modelizarlas para poder dar solución a algunos problemas de la lectura artificial del texto, que al fin y al cabo, se efectúa linealmente a un nivel morfosintáctico y semántico, simultáneamente. El hecho de que un concepto aparezca transportado en diferentes lugares y funciones del enunciado no es problemático para la lectura humana, pero un ordenador tendría que asumir previamente todas las posibilidades categoriales puesto que es la única vía de acceder al significado. Actualmente, la detección de funciones (previamente declaradas al sistema) en enunciados simples es la única realizable por máquina con resultados eficaces (Lisp, Prolog, etc.).

4.3.3 *Otros métodos de análisis del relato*

En Propp encontramos los fundamentos del Análisis textual moderno a pesar de los errores de su modelo, detectados por los investigadores del área. El método propuesto por el investigador soviético perseguía la reducción de cualquier cuento de hadas ruso en treinta y una funciones y siete personajes. Para conseguirlo, aplicó el método inductivo, es decir, partió de un corpus de cien cuentos para elevar sus categorías generales. De ahí surgen unas funciones que explican cualquier cuento de hadas aunque revista, cada uno de ellos, un ropaje de anécdotas y detalles distintos pero que a la vez los hace reconocibles. No obstante, el modelo de Propp es poco generalizable ya que no funciona en otros textos, ni siquiera en relatos de distinta temática con narrativa similar, además de haber sido acusado de excesivamente formalista por Levi-Strauss²⁴.

Hendricks, más recientemente, propone normalizar el texto con el fin de llegar a su estructura profunda. Las operaciones que sugiere son:

- Reemplazar las anáforas por referentes.
- Eliminar los enunciados de descripción dejando sólo los de acción que constituyen materia narrativa.
- Se subdividen los enunciados de acción atribuyéndolos a los sujetos implicados.

Así se alcanza un metatexto carente de detalles y un esquema general del relato. Sin embargo, el mismo investigador autocritica su método: "la operación segunda no garantiza la objetividad ya que depende en alto grado de la intuición"²⁵.

Volvemos a encontrar en estos autores el interés por las categorías generales (en nuestro caso macrodescriptores y primitivas), por la normalización del texto (Guiraud, Greimas) y su clasificación (estructuralismo). Todo ello nos corrobora la necesidad de crear un macrosistema categorial y temático para el discurso periodístico como marco de referencia lectora, por la vía empírica, capaz de dar cuenta y absorber cualquier texto no considerado en la muestra.

²⁴ Metodologías para el análisis del relato recogidas por Bobes, M^a Carmen: *La Semiología*. o.c., p.88-90

²⁵ Idem, p.90-91

CAPITULO 5

CONCLUSIONES

- * Cualquier investigación en materia de análisis o lenguajes documentales, independientemente de la aplicación, debe ser realizada y contar con la presencia simultánea de todos los elementos del paradigma del proceso documental, a saber: Análisis (lectura o reconocimiento), síntesis (transformación), representación y recuperación.
- * Es posible explicitar el proceso de lectura empírica efectuado por el documentalista. La explicitación se obtiene mediante la interrogación y reconocimiento objetivable del texto basada en métodos de análisis semántico utilizados en otras disciplinas.
- * No es posible extrapolar los parámetros de explicitación de lectura humana en los procesos de lectura artificial (por tanto, la hipótesis de partida es negada). Los métodos humanos interrogan el contenido pero su simulación es inviable. Los métodos artificiales actuales se basan en la linealidad de la lectura, obteniéndose resultados pobres, erróneos o confusos.
- * La propuesta de un método de lectura artificial con fines documentales sólo puede ser realizada tras encontrar solución a las siguientes cuestiones: inventario del comportamiento funcional de los conceptos en el nivel enunciativo, método de construcción de un cuadro de macrodescripción (universo de situaciones), inventario de modalidades de interrogación, métodos de estructuración de la información en función de los productos.

- * Los resultados de un método de lectura artificial deben ser semejantes a los obtenidos por varios documentalistas expertos al leer los mismos documentos usando métodos de explicitación para la lectura humana. Esto supone un sistema de validación final.
- * Es necesario construir muestras compuestas por unidades discretas que representen el universo de la aplicación y recojan cualquier variable posible. En el discurso periodístico, la unidad de contenido, como unidad discreta, no es operativa (rechazo de hipótesis de partida).
- * Es necesaria una investigación específica sobre metodología de la validación en análisis y lenguajes documentales.
- * Cada corpus de aplicación precisa de una sistematización de todas las variables y condiciones para el trabajo teórico y para la práctica.
- * En el discurso periodístico las principales variables para el AD son: extensión, estilo, género, temática, estructura, ideología y productos que desea el usuario: síntesis, datos, inferencias, comparaciones y verificaciones.
- * Es necesario realizar recortes interdisciplinares en Psicología cognitiva y Pragmática como campos que pueden aportar aparato teórico y conceptual de interés para avanzar en las investigaciones documentológicas que afectan la representación paramétrica del conocimiento.

BIBLIOGRAFIA

Allwood, Jens; Lars, Gunnar y Dahl, Osten: *Lógica para lingüistas*. Madrid: Paraninfo, 1981, 203 p.

Amaro, Regina: *Contribuição da análise do discurso para à análise documentária: o caso da documentação jornalística*. Sao Paulo: ECA/USP, 1991, 87 p. + anexos.

Amat, Nuria: *De la información al saber*. Madrid: Fundesco, 1990, 208 p.

Angenot, M.: *La parole pamphlétaire: contribution à la typologie des discours modernes*. Paris: Payot, 1982

Bardin, Laurence: *Análisis de Contenido*. Madrid: Akal, 1986, 183 p.

Barthes, Roland: *S/Z*. Madrid: Siglo XXI, 1980, 221 p.

Idem: *El placer del texto*. México: Siglo XXI, 1989, 150 p.

Bobes Naves, M^a Carmen: *La semiología*. Madrid: Síntesis, 1989, 167 p. (col. *Lingüística*;12)

Cintra, A.M.: *Estratégias de leitura em Documentação*. In: Smit, J. (org.): *Análise documentária: a análise da síntese*. Brasília: Ibict, 1987, p. 27-35.

Colle, Raymond: *Tecnologías de la Información*. Santiago de Chile: Pontificia Universidad Católica, 1988, 183 p.

Cunha, Isabel (org.): *Análise documentária: considerações teóricas e experimentações*. Sao Paulo: Febab, 1989. 191 p.

Idem: *Do mito à análise documentária*. Sao Paulo: Edusp, 1990, 163 p.

Idem y Kobashi, N.: Análise documentária e Inteligência artificial. 1990, 25 p.
+ anexos (ined. mecanograf.)

Chaumier, J.: El análisis documental. Barcelona: Mitre, 1986, 172 p.

Chomsky, Noam: Reflexiones sobre el lenguaje. Barcelona: Planeta/Agostini,
1985, 278 p.

Eco, U.: Conceito de texto. Sao Paulo: Taq/Edusp, 1984

Idem: Lector in Fabula. Barcelona: Lumen, 1987, 330 p.

Idem: Signo. Barcelona: Labor, 1988, 217 p.

Ellis, Davis: New Horizons in Information Retrieval. London: Library
Association, 1990, 133 p.

Foucault, Michel: La arqueología del saber. México: Siglo XXI, 1990,
355 p.

Galdón, Gabriel: Principios operativos de la documentación periodística.
Madrid: Dossat, 1989, 86 p.

García Gutiérrez, Antonio: Lingüística documental: aplicación a la comunica-
ción social. Barcelona: Mitre, 1984, 279 p.

Idem: La productivité de la Communication en Espagne: une approximation
infométrique. "Documentación de las Ciencias de la Información", vol.12,
1989

Idem: Estructura lingüística de la documentación: teoría y método. Murcia:
Servicio de Publicaciones de la Universidad de Murcia, 1990, 166 p.

Idem: Suficiencia estructural y tipología de la omisión en Análisis documental.
"Documentación de las Ciencias de la Información", vol 13, 1990, p. 73-86.

Idem y Lucas, R.: Documentación automatizada en los medios informativos.
Madrid: Paraninfo, 1987, 264 p.

Gardin, J.C.: Les analyses de discours. Neuchâtel: Delachaux et Niestlé, 1974,
178 p.

Idem et al: Systèmes experts et Sciences humaines. Paris: Eyrolles, 1987, 269 p.

Idem et al: La Logique du plausible. Essais d'Epistémologie pratique en Sciences humaines. 2.^a ed. Paris: Maison des Sciences de l'Homme, 1987, 300 p.

Greimas, A.J.: Semántica estructural. Investigación metodológica. Madrid: Gredos, 1976, 398 p.

Guiraud, Paul: Les structures élémentaires de la signification. "BLS", 60, 1965, p. 97-114.

Haacke, Wilmont: Escritos recientes sobre el concepto de actualidad. "Opinión pública". Madrid: Escuela oficial de Periodismo, (s.a.), p.169-193.

Image et Intelligence artificielle dans l'information scientifique et technique. 6 au 10 juin 1988. Bénodet (Finistère). Paris: Inria, 1988, 189 p.

Izquierdo, José M^a: Esquemas de Lingüística documental. Barcelona: PPO, 1990. t.I: 242 p.

Joab, Michelle: Les systèmes experts. In: "Image et Intelligence artificielle dans l'information... o.c., p. 115-141.

Kobashi, Nair: Análise documentária: tipologias discursivas. In: Cunha, I. : Análise documentária: considerações...o.c., p.31-44.

Idem: Análise documentária: considerações sobre um modelo lógico-semântico. In: Cunha, I.: o.c. p.45-57.

Krippendorf, Klaus: Metodología de análisis de contenido: teoría y práctica. Barcelona: Paidós, 1990 (Paidós comunicación;39), 279 p. (Content analysis: An introduction to its methodology. London: Sage, 1984).

Lopez Yepes, José: Teoría de la Documentación. Pamplona: Eunsa, 1978, XXXI+337 p.

Idem (comp.): Fundamentos de Información y Documentación. 2.^a ed. Madrid: Eudema, 1990, 485 p.

Martín Barbero, Jesús: Procesos de comunicación y matrices de cultura: itinerario para salir de la razón dualista. México: Gustavo Gili/ Felafacs, 1988, 212 p.

Menon, Bruno: Indexation automatique et intelligence artificielle: quelques questions de stratégie. In: Image et ...o.c. p.143-175.

Pêcheux, Michel: Hacia el análisis automático del discurso. Madrid: Gredos, 1978, 374 p.

Pottier, Bernard: Lingüística general. Madrid: Gredos, 1976, 426 p.

Sagredo, Félix e Izquierdo, José M^a: Concepción lógico-lingüística de la Documentación. Madrid: Ibercom- Red Comnet, 1983, 440 p.

Saussure, Ferdinand de: Curso de Lingüística general. Barcelona: Planeta/ Agostini, 1985, 292 p.

Schank, R.: El ordenador inteligente. Barcelona: A. Bosch, 1986, 281 p.

Schiller, Herbert: El poder informático. Imperios tecnológicos y relaciones de dependencia. Barcelona: Gustavo Gili, 1983, 225 p.

Shannon, Claude y Weaver, Warren: Teoría matemática de la Comunicación. Madrid: Forja, 1981, 159 p.

Smit, Joanna (org.): Análise documentária: a análise da síntese. Brasília: Ibtict, 1987, 133 p.

Tálamo, Fátima: A definição semântica para a elaboração de glossários. In: Smit, J.: o.c. p.87-98.

Van Dijk, Teun: Texto y contexto: semántica y pragmática del discurso. Madrid: Cátedra, 1980, 357 p.

Idem: La noticia como discurso: comprensión, estructura y producción de la información. Barcelona: Paidós, 1990, 284 p.