



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Bases lingüísticas e informáticas
para la elaboración de tesauros

Dr. Jesús Valdez Ramos

Dra. Georgina Araceli Torres Vargas
Asesor



Ciencia Nueva
por los torres UNAM



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mis padres

Índice general	página
Sinopsis	II
Introducción	III
Capítulo 1. Bibliotecología y tesauros	1
1.1. El estudio del tesoro en la Bibliotecología	1
1.2. Definición del término tesoro desde la lingüística, la informática y la terminología	5
1.2.1. Desde la Lingüística	7
1.2.2. Desde la Informática	11
1.2.3. Desde la Terminología	13
1.2.4. Análisis de las definiciones del término tesoro	14
1.3. El tesoro en Bibliotecología y Estudios de la información	17
1.3.1. Rasgos comunes de las definiciones de tesoro	21
1.3.2. El tesoro en la indización	23
1.3.2.1. Revisión de los documentos	28
1.3.2.2. Selección de conceptos	28
1.3.2.3. Traducción de conceptos en descriptores	28
1.3.2.4. Establecimiento de enlaces entre descriptores conservados	30
1.3.2.4.1. Yuxtaposición	31
1.3.2.4.2. Ponderación	32
1.3.2.4.3. Especificación de un punto de vista	32
1.3.2.4.4. Especificación de un vínculo	32
1.3.2.4.5. Especificación de un rol o papel	33
1.3.2.4.6. Integración dentro de un enunciado estructurado	33

1.3.2.5. Adecuación de los descriptores al contexto	34
1.4. El tesoro en el almacenamiento y la recuperación de información	35
Capítulo 2. Bases lingüísticas y terminológicas para la elaboración de los tesauros	43
2.1. Bibliotecología, lingüística y terminología: un espacio de colaboración interdisciplinaria alrededor del tesoro	43
2.2. Procesos lingüísticos para la elaboración de tesauros	52
2.2.1. Procesos morfológicos	53
2.2.1.1. Derivación y flexión	55
2.2.1.2. Siglas	59
2.3. Procesos sintácticos	60
2.4. Procesos léxico semánticos	63
2.4.1. Relaciones de jerarquía	64
2.4.2. Relaciones de equivalencia	66
2.4.3. Relaciones de asociación	67
2.5. Procesos léxicos. Asociaciones, configuraciones y relaciones léxicas en un tesoro	68
2.5.1. Noción de relación en la lingüística y los tesauros	70
2.6. Configuraciones léxicas	71
2.7. Configuraciones jerárquicas	73
2.8. La Terminología: tres enfoques teóricos y puntos de convergencia alrededor del tesoro	78
2.8.1. Definición de Terminología	78
2.8.2. Teoría General de la Terminología, aspectos básicos	81
2.8.3. La Teoría Comunicativa de la Terminología, aspectos básicos	82

2.8.4. La Teoría Socio cognitiva de la Terminología	85
2.9. Terminología y tareas bibliotecológicas	88
Capítulo 3 El tesoro y la Informática	92
3.1. El tesoro y la Informática: definición del término y antecedentes	93
3.1.1. Definición del término	93
3.1.2. Antecedentes sobre el tesoro y la recuperación de la información	95
3.2. La recuperación de la información a través de los tesoros	96
3.3. La recuperación de la información y la evolución de los tesoros	97
3.4. Pasos básicos para la elaboración de un tesoro	101
3.4.1 Definir el rango y profundidad de un tesoro	102
3.4.2 Identificar las fuentes para construir el vocabulario	103
3.4.3 Extracción del vocabulario o términos candidatos	104
3.4.4 Organizar el vocabulario en grupos de términos asociados	105
3.4.5 Acomodo de grupos de términos	105
3.4.6 Establecer un sistema de notación	106
3.4.7 Tratamiento de los términos pre coordinados o de más de una palabra	107
3.4.8 Refinamiento del vocabulario	107
3.4.9 Creación de referencias entre los términos	108

3.4.10 Mantenimiento del tesoro	109
3.5. Herramientas lingüísticas para la indización y la recuperación de información	113
3.5.1. Analizadores morfológicos	115
3.5.2. Analizadores sintácticos	117
3.5.3. Analizadores semánticos	119
3.6. Herramientas de la termiología para la indización	123
3.6.1. Vocabularios o glosarios terminológicos en línea	124
3.6.2. Bases de datos terminológicas de léxicos especializados	124
Capítulo 4. Lineamientos metodológicos para la elaboración de tesauros	128
4.1. Definición y alcance de los lineamientos	128
4.1.1. Procesos y reglas morfológicas	129
4.1.2. Procesos y reglas para la creación de siglas	130
4.1.3. Procesos y reglas sintácticas	130
4.1.4. Los procesos semánticos	131
4.1.5. Procesos léxicos	133
4.1.6. Aspectos terminológicos	134
4.1.7. Aspectos informáticos	134
4.1.8. Analizadores morfológicos, sintácticos y semántico	135
5. Conclusiones	138
6. Obras consultadas	142

Índice de figuras y cuadros		Página
Cuadro 1	Rasgos comunes en las definiciones del término tesoro.	22
Figura 1	Puntos de contacto y convergencia.	44
Figura 2	Colaboración interdisciplinaria en la elaboración de tesauros.	46
Figura 3	Aspectos morfológicos: afijación.	56
Figura 4	Ejemplo de proceso morfológico de derivación: afijación	57
Figura 5	Ejemplo de proceso morfológico de flexión: sufijación	58
Figura 6	Ejemplo de proceso sintáctico: <i>sustantivo + modificador</i>	62
Cuadro 2	Procesos léxicos en un tesoro.	69
Figura 7	Tipos de configuraciones jerárquicas	73
Figura 8	Configuraciones jerárquicas en un tesoro	74
Figura 9	Relación partitiva	75
Figura 10	Configuración meronímica	76
Figura 11	Rango y profundidad de un tesoro	103
Figura 12	Tipos de relaciones entre los términos de un tesoro	106
Figura 13	Creación de referencias entre los términos	109
Figura 14	Modelo informático de un analizador	114
Figura 15	Propuesta de lineamientos metodológicos para la elaboración de tesauros	137

Reconocimientos

Deseo agradecer de manera muy especial a la Doctora Georgina Araceli Torres Vargas, directora de esta tesis y al Doctor Gerardo Eugenio Sierra Martínez cotutor, sus observaciones y comentarios siempre certeros que orientaron el proceso de la investigación, su lectura cuidadosa y apoyo continuo hicieron posible realizar este trabajo. Al Doctor Luis Fernando Lara Ramos, por sus comentarios y observaciones que hizo al inicio del trabajo de la investigación.

Mi agradecimiento especial también a la Doctora Elsa Barberena Blázquez, lectora de esta tesis. Su orientación y acompañamiento, desde hace varios años, han marcado mi desarrollo como bibliotecólogo en la UNAM.

Para la Doctora Marisela Colín Rodea y la Doctora Brenda Cabral Vargas, lectoras, cuyos comentarios me aclararon dudas puntuales.

Sinopsis

El presente estudio trata de las bases teóricas a partir de las cuales se explica el proceso de elaboración de un tesoro, considera a éste como una herramienta que se utiliza en dos tareas bibliotecológicas fundamentales: por una parte, la indización y por otra, la recuperación de información. La investigación que se realizó es de índole interpretativa. En ella, se exploran, explican, analizan y valoran las aportaciones teóricas con las que la lingüística, la informática y la terminología han contribuido para el diseño y construcción de un tesoro. El estudio se focaliza en la indagación de lo que sucede en las zonas limítrofes de estas disciplinas, esta área de fusión, en la que se entrecruzan y enlazan conceptos teóricos de los que se pueden derivar lineamientos metodológicos que iluminan caminos y enriquecen a la disciplina Bibliotecológica. Una investigación de este tipo parte de la idea de que la colaboración interdisciplinaria es un proceso complejo de intercambio a través del cual se establecen puntos de contacto y convergencia que enriquecen a las disciplinas que participan en el proceso.

El problema que se investiga en esta tesis es la identificación, análisis y valoración de las aportaciones teóricas y metodológicas que han hecho la Lingüística (L), la Informática (I) y la Terminología (T) con aproximaciones de solución al proceso de creación de un tesoro. El método que se utilizó fue el método analítico-sintético. La técnica utilizada fue el análisis documental para fundamentar todas y cada una de las opiniones de los autores de las fuentes que se utilizaron durante la investigación y la elaboración de esta tesis. Las preguntas de investigación alrededor de las cuales se desarrolló el trabajo fueron: ¿Cuáles son las aportaciones teóricas de la Lingüística, a partir de las cuáles se pueden estudiar las relaciones entre los términos que conforman un tesoro? ¿Cuáles son las aportaciones teóricas que desde la Informática se han estudiado y llevado a la práctica para la elaboración de tesauros? ¿Cuáles son las aportaciones teóricas que desde la Terminología ha contribuido para el diseño de un tesoro? Del estudio realizado se deriva una propuesta metodológica que contiene lineamientos necesarios para el diseño y elaboración de los tesauros, tomando en cuenta las aportaciones de la Lingüística, la Informática y la Terminología.

Introducción

En la Bibliotecología los tesauros han sido un tema de interés que puede considerarse como amplio y complejo. Su estudio se ha realizado desde diferentes enfoques, algunos de los cuales han producido una cantidad considerable de informes técnicos, libros, monografías y artículos especializados sobre aspectos puntuales, como la construcción de los tesauros, las normas técnicas y de procedimiento para el desarrollo de proyectos institucionales en los que se diseña y elabora un tesoro, manuales generales de procedimientos e informes de experiencias específicas sobre la construcción de tesauros en áreas específicas del conocimiento, entre otros.

Esta tesis muestra un estudio en el que se presentan las bases teóricas a partir de las cuales se explica el proceso de la elaboración de un tesoro, considerando a éste como una herramienta que se utiliza en dos tareas bibliotecológicas fundamentales, por una parte la indización y por otra la recuperación de información.

El estudio realizado es interpretativo porque en él se exploran, analizan y valoran las aportaciones teóricas con las que la lingüística, la informática y la terminología han contribuido para el diseño y construcción de un tesoro. Asimismo, el estudio centra su atención en indagar lo que sucede en las zonas limítrofes de estas disciplinas, esta área de fusión, en la que se entrecruzan y enlazan conceptos teóricos de los que se pueden derivar lineamientos

metodológicos que iluminan caminos y enriquecen a la disciplina Bibliotecológica.

Una investigación de este tipo parte de la idea de que la colaboración interdisciplinaria es un proceso complejo de intercambio a través del cual se establecen puntos de contacto y convergencia que enriquecen a las disciplinas que participan en el proceso.

Los tesauros se diseñan y elaboran por un grupo interdisciplinario de profesionales integrado por bibliotecólogos, lingüistas, informáticos y terminólogos, entre otros. Cada uno de estos profesionales hace una serie de aportaciones desde su disciplina, que enriquece el producto y el proceso de creación de la herramienta.

El problema que se investiga en esta tesis es un problema complejo y con diferentes aristas, mismo que puede ser abordado desde diferentes enfoques teóricos y con diversos métodos de trabajo. Por lo anterior, se plantea como problema central la identificación, análisis y valoración de las aportaciones teóricas y metodológicas que han hecho la Lingüística (L), la Informática (I), la Terminología (T) y que han colaborado con aproximaciones de solución al proceso de diseño y a la metodología para la elaboración de un tesoro, como herramienta para organizar información desde la Bibliotecología.

Los objetivos generales que se persiguen son los siguientes:

- Identificar, analizar y valorar cuáles son las aportaciones teóricas y metodológicas que han hecho la Lingüística (L), Informática (I), Terminología (T) para el diseño y elaboración de tesauros en el marco de la Bibliotecología.
- Proponer y elaborar lineamientos metodológicos necesarios para el diseño y elaboración de los tesauros, tomando en cuenta las aportaciones de la Lingüística, la Informática y la Terminología.

Como objetivos particulares se plantea:

- Identificar, analizar y valorar cuáles son los componentes lingüísticos: morfológicos, sintácticos, semánticos y lexicales que intervienen en la formación de un lenguaje documental del tipo tesoro.
- Identificar, analizar y valorar cuáles son los componentes terminológicos que intervienen en la elaboración de tesauros.
- Proponer una serie de lineamientos metodológicos para la Bibliotecología que reúna las aportaciones de la L, I y T para la elaboración de tesauros.

Durante la investigación se utilizó el método analítico-sintético, al momento de señalar las bases teóricas lingüísticas que intervienen en la elaboración de un tesoro y al momento de señalar las aportaciones que hacen la informática, la terminología para el diseño y elaboración de un tesoro. Asimismo se utilizó la

técnica de análisis documental para fundamentar todas y cada una de las opiniones de los autores de las fuentes que se utilizaron durante la investigación y en la elaboración de esta tesis.

Las preguntas de investigación alrededor de las cuales se desarrolló esta tesis fueron:

- ¿Cuáles son las aportaciones teóricas de la lingüística, a partir de las cuáles se pueden estudiar las relaciones entre los términos que conforman un tesoro?
- ¿Cuáles son las aportaciones teóricas que desde la informática se han estudiado y llevado a la práctica para la elaboración de tesoros?

Durante la investigación se contestaron estas preguntas de forma amplia y se estudiaron las diferencias que existen entre términos y conceptos y los tipos de relaciones que se establecen entre ellos. Se analizó este tema desde la lingüística y la terminología y las implicaciones que tiene para el tratamiento informático de los términos en la construcción de un tesoro.

Son varias las bases lingüísticas en las que se sustenta la estructura de un tesoro. Una de ellas tiene que ver con la morfología de las palabras, específicamente con el fenómeno de la flexión y el uso de los acrónimos. Otra, de las bases lingüísticas en las que se sustenta la estructura de un tesoro es el orden de palabras, este tema se abordó tomando en cuenta los estudios sintácticos sobre el orden canónico de la frase en español.

Una más de las bases lingüísticas que se estudió durante el desarrollo de la investigación es la que se refiere a los diferentes tipos de relaciones que se establecen entre las unidades léxicas que conforman un tesoro, se centró la atención en el estudio de las relaciones hponímicas a través de estructuras jerárquicas.

Finalmente se estudió la base lingüística derivada del fenómeno de la sinonimia y las relaciones de jerarquía, asociación y equivalencia entre los términos.

El estudio de la informática en la creación de tesoros ha sido abundante y exhaustivo. Las aportaciones que ha hecho la informática han sido fundamentales para el desarrollo de los tesoros. Entre las aportaciones que ha hecho esta disciplina se encuentran diferentes programas de cómputo para la construcción de los tesoros. Sin embargo, durante la investigación realizada, el estudio se centró en cuatro puntos concretos.

El primero fue una reflexión general sobre la relación entre la informática y el tesoro. El segundo punto consistió en la descripción de forma precisa y a detalle el procedimiento que sugiere la informática para la elaboración de los tesoros. El tercer punto consistió en la descripción de las características de los analizadores morfológicos, sintácticos y semánticos, herramientas informáticas con las que la informática colabora en el proceso de la creación de un tesoro. El cuarto y último punto consistió en la descripción general de las características de bancos y vocabularios terminológicos de léxicos especializados.

Esta tesis está organizada en cuatro capítulos. El primero es una reflexión sobre la importancia del tesoro en la bibliotecología. En este capítulo se define el término tesoro, se señalan las características del tesoro considerando a éste como una obra de consulta para otras disciplinas y se explica cuál es el papel que tiene el tesoro en la indización y en la búsqueda y recuperación de la información.

En el segundo capítulo se presenta un panorama general de los puntos de contacto y convergencia que se han identificado entre la lingüística, la terminología y la bibliotecología y, por otra, se analizan algunas de las aportaciones que han hecho la lingüística y la terminología para la construcción de un tesoro. Se realizó un análisis puntual de los rasgos distintivos que conforman los procesos lingüísticos y terminológicos que intervienen en la elaboración de un tesoro. Desde la lingüística se analizaron los procesos morfológicos, sintácticos, léxicos y semánticos. Desde la terminología se analizaron los términos como unidades del lenguaje especializado y como unidades de comunicación.

En el tercer capítulo se identifican, describen y analizan algunas aportaciones teóricas y metodológicas con las que la informática contribuye al proceso de indización y la recuperación de la información en los que interviene el tesoro.

Primero y de forma detallada se describe el procedimiento básico para la elaboración de un tesoro de acuerdo con la informática. Dicho procedimiento

está compuesto por los siguientes pasos, la definición del rango y profundidad del tesoro, la identificación de las fuentes para construir el vocabulario, la extracción de los términos candidatos, la organización del vocabulario en grupos de términos asociados, la organización en grupos de los términos, el establecimiento de un sistema de notación, el tratamiento de los términos pre coordinados o de más de una palabra, el refinamiento del vocabulario, la creación de referencias entre los términos y el mantenimiento del tesoro.

Enseguida, se describen y valoran algunas de las herramientas informáticas que se utilizan en la indización y la recuperación de la información. Las herramientas que se estudiaron son los analizadores morfológicos, sintácticos y semánticos y su relación con la indización y la recuperación de la información. Finalmente, se describen y valoran algunas de las herramientas terminológicas que se utilizan en la indización y la recuperación de la información.

En este capítulo se ha organizado la información de la siguiente manera, en primer lugar se presenta una reflexión sobre la relación entre la Informática y el tesoro. Se exponen los antecedentes de esta relación que se encontraron en la investigación documental que se realizó para tal propósito. Segundo, se describen y analizan las características de los analizadores morfológicos, sintácticos y semánticos. Tercero, se describen y analizan, de manera general las características de bancos y vocabularios terminológicos de léxicos especializados.

En el cuarto capítulo se exponen y analizan los lineamientos metodológicos derivados del estudio realizado en esta tesis. Dichos lineamientos idealmente servirán como una guía a los bibliotecarios que tienen a cargo proyectos relacionados con la elaboración y uso de los tesauros. Los lineamientos son de dos tipos. Por una parte, los que se relacionan con la indización y por otra los que están relacionados con la recuperación de la información. Asimismo, dichos lineamientos son de naturaleza diferente. Por una parte, se encuentran los que se derivan de las bases o fundamentos lingüísticos y terminológicos y, por otra los lineamientos que se derivan de las bases o fundamentos de la informática.

Capítulo 1. Bibliotecología y tesauros

Esto no es mío, sino de quien lo lee
FERNANDO PESSOA

1. Bibliotecología y tesauros

En este capítulo se presenta una reflexión sobre la importancia del tesoro en la Bibliotecología. Los propósitos específicos que se persiguen son los siguientes: definir el término tesoro, señalar sus características como una obra de consulta para otras disciplinas y explicar cuál es el papel que tiene tanto en la indización como en la búsqueda y recuperación de la información.

1.1. El estudio del tesoro en la Bibliotecología

Para iniciar esta reflexión plantearé la siguiente pregunta. ¿Los tesauros son objeto de estudio de la Bibliotecología? Veamos, hasta al final de la primera mitad del siglo XX, la práctica bibliotecaria estaba orientada hacia la organización de documentos —libros principalmente— a partir de instrumentos diseñados y contruidos previamente, por ejemplo los sistemas de clasificación.

El papel del bibliotecario era principalmente el de custodio o guardián de libros, no obstante que se tuviera la conciencia de que su función era la de intermediario entre las necesidades de información del usuario y la disponibilidad real de los materiales. Hoy en día la intermediación entre las necesidades de información del usuario y el bibliotecario está presente de una manera indirecta y más tenue

dado que existen algunas herramientas que facilitan el acceso a las diferentes colecciones documentales de las bibliotecas.

Durante ese mismo periodo en la formación profesional del bibliotecario se pueden distinguir, de acuerdo con Doyle Dodebei, V. L. (2002),¹ dos vertientes: La primera es la formación orientada hacia la posesión de elementos de cultura general o alta cultura, historia y filosofía, por ejemplo, y el dominio de diferentes lenguas extranjeras, inglés y francés principalmente. La segunda vertiente, una formación orientada hacia el aprendizaje de técnicas para la organización de los acervos, tales como normas, códigos de catalogación y tablas de clasificación. Esta última orientación limitaba el trabajo del bibliotecario a la clasificación del contenido de los libros a la luz de un instrumento codificador.

De acuerdo con Slype (1991)², a partir de la década de los años sesentas, con la especialización del saber y el aumento en la producción de documentos, el trabajo bibliotecario y uno de los objetos tradicionales de estudio de la Bibliotecología, obliga a los bibliotecarios a buscar nuevos procedimientos que se utilizan en la organización de acervos documentales. Hasta ese momento el libro era el objeto alrededor del que se organizaban una serie de normas y procedimientos. Ahora el lugar del libro lo ocupan otros formatos diferentes: artículos de revistas e informes o reportes de investigación.

¹ Doyle Dodebei, V. L. (2002) *Tesouro: linguagem de representação da memória documentaria*. Niterioi / Río de Janeiro: Interciencia.

² Slype, G. V. (1991) *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Pirámide. p. 17.

Con este cambio también surge la necesidad de crear nuevas herramientas que permitieran no sólo la organización de un dominio especializado del conocimiento, sino también la recuperación de la información de los mismos. Así se crearon lenguajes documentales diferentes a los tradicionales. Por lenguajes documentales tradicionales entenderemos a los sistemas de clasificación de la Biblioteca del Congreso (*Library of Congress*) o el Sistema de Clasificación Decimal Universal, por ejemplo.

Los lenguajes documentales creados, son capaces de representar el contenido de los nuevos formatos diferentes al libro, se les nombró inicialmente vocabularios controlados y después fueron conocidos como lenguajes artificiales. A este tipo de lenguajes pertenecen los tesauros, que también son objeto de estudio e interés de la Bibliotecología. Sin embargo, los tesauros no sólo han despertado interés en los bibliotecólogos, sino también en profesionales de otras disciplinas preocupados por la representación y recuperación de información, como lingüistas, informáticos, y terminólogos, entre otras profesiones.

Así entonces se percibe que los tesauros son y han sido objeto de estudio de la Bibliotecología a partir de esa época y su importancia radica al menos en dos razones. Como un instrumento apropiado para la organización del conocimiento y como una herramienta potencialmente eficiente para la recuperación de la información.

Los sistemas de clasificación son una de las formas en las que se organiza el conocimiento general para que, a su vez, en las bibliotecas se ordenen las colecciones documentales, en tanto que con los tesauros esta tarea se realiza en unidades de información, centros de documentación o bibliotecas que poseen colecciones documentales sobre dominios específicos del conocimiento.

De igual manera los tesauros resultaron importantes para la Bibliotecología. Por una parte en lo que se refiere a la indización de documentos, actividad que de acuerdo con Moreiro González (1994)³ “es la identificación de los documentos por conjuntos, de acuerdo con significaciones coincidentes, sus resultados se reflejan en términos relevantes y pertinentes con los conceptos que contiene el documento” y según Barité (1997)⁴ la define como “la operación más significativa del análisis documental que refiere al proceso de descripción y representación del contenido de un documento, mediante un número limitado de conceptos extraídos del texto de los documentos (palabras clave) o de vocabularios controlados (descriptores, términos o encabezamientos de materia). Estos conceptos deben ser los que más apropiadamente caractericen al documento, de modo de facilitar su futura recuperación temática”.

Más adelante en este capítulo se explicará, de manera más amplia, cuál es el papel que tiene el tesoro en el proceso de indización.

³ Moreiro González, J. A. (1994) “Documentación y lingüística: conceptos de relación esenciales” en *Ciencias de la Información*. 25 (4) p. 206.

⁴ Barité, M. (1997). *Glosario sobre organización y representación del conocimiento: clasificación, indización y terminología*. Montevideo: CSIC. p. 73 -74.

Y por otra parte, la importancia de los tesauros para la Bibliotecología, se manifiesta en la recuperación de información, que de acuerdo con Broughton (2006)⁵ se define como “el proceso de localización de información sobre un tema específico, generalmente de una colección organizada o de una base de datos. El término recuperación de información actualmente se utiliza, en un sentido amplio, para referirse a la obtención de información, que incluye la búsqueda, las técnicas de búsqueda, las herramientas de búsqueda y *software*”.

Por lo antes expuesto es claro que los tesauros son importantes para la Bibliotecología, pues constituyen una herramienta que permite organizar el conocimiento en áreas específicas del saber, además que ayudan a establecer lineamientos metodológicos en dos áreas fundamentales en la práctica bibliotecaria, como son los procesos de indización y los procesos de recuperación de información.

1.2. Definición del término tesoro desde la Lingüística, la Informática y la Terminología

Para ubicar en el contexto histórico el uso del término tesoro resulta pertinente señalar al menos tres momentos en un *continuum*. El primero, el término tesoro tiene su origen etimológico en la palabra latina *thesaurus*, misma que proviene de la palabra griega *thesaurós*, en ambos casos significaba tesoro o repositorio de palabras.

⁵ Broughton, V. *op cit.* p. 214.

El segundo momento, y de acuerdo con Arano, Silvia (2005)⁶ “fue la Lexicografía, en la época moderna, el primer campo en el que se aplicó el término tesoro y en el que se elaboraron las primeras definiciones. En 1852 Peter Mark Roget publica su obra *ROGET’S THESAURUS of English Words and Phrases*, cuya finalidad es la de proporcionar ayuda y asistencia a la expresión de ideas y composición literaria”.

Y el tercer momento, siguiendo a Arano, Silvia (2005)⁷, señala que “en 1957 se usa por primera vez el término tesoro en el campo de la Biblioteconomía y Documentación, en un trabajo presentado a la *Dorking Conference on Clasification*, por Helen Brownson, miembro de la *American National Science Foudation*. En dicha contribución, el término tesoro es utilizado al analizar la problemática de traducir conceptos y sus relaciones, tal y como se expresan en los documentos, aun lenguaje con mayor precisión y sin ambigüedades, para facilitar la recuperación de información”.

En esta sección se analizan una serie de “entradas” del término *tesoro* y sus respectivas definiciones. Estas últimas fueron tomadas de obras de consulta, generales y especializadas, tales como: diccionarios de lengua, diccionarios especializados, enciclopedias y obras especializadas de la lingüística, la informática y la terminología.

⁶ Arano, Silvia (2005) Los tesauros y las ontologías en la Biblioteconomía y la Documentación [en línea]. *Hipertex.net*, número 3, p2. <<http://www.hipertex.net>> [consulta 12 de abril 2010]

⁷ *Ibid* p. 2.

Tal selección de obras se realizó con la intención de contar con una definición apropiada, amplia y consistente con los diversos campos del saber que le conciernen.

El análisis se inicia con una definición de un diccionario general de lengua, se continúa con un diccionario especializado en lexicografía, área de la lingüística. Le sigue la definición de una obra especializada en terminología. Para terminar con una serie de definiciones que aparecen en diccionarios especializados de informática.

Al final de la sección se incluye una reflexión sobre la diversidad de definiciones del término tesoro. Dicha reflexión dará pie a una serie de preguntas amplias sobre los diferentes enfoques y características con las que se define el término tesoro.

1.2.1 Desde la Lingüística

En el campo de la lingüística y la lexicografía se destacan definiciones en las que la complejidad en la descripción del término es evidente, se ofrece información sobre el origen y desarrollo histórico del mismo, así como una diferenciación precisa de las acepciones y usos que tiene. Muestra de ello, son las siguientes definiciones, mismas que se presentan bajo un criterio cronológico.

Martínez de Sousa (1995)⁸ define al tesoro con cinco acepciones diferentes: “1. Tesoro, diccionario, catálogo, colección. 2. Lista alfabética de palabras clave para la clasificación de documentos. 3. Diccionario que muestra la equivalencia entre términos del lenguaje natural y términos normalizados y preferentes del lenguaje documental, así como las relaciones semánticas que existen entre otros términos. 4. Diccionario que registra las palabras u otras unidades léxicas extraídas de los textos de personas representativas de la lengua de una comunidad en una época determinada. 5. Vocabulario limitado, generalmente de palabras especializadas, dotado de sus correspondencias semánticas y elegido para que represente las nociones que figuran en un texto dado para su empleo en informática y en el establecimiento de índices”.

En las cinco acepciones de la definición del término tesoro que señala Martínez de Sousa, podemos observar lo siguiente:

En la primera acepción se muestra la utilización de términos equivalentes para nombrar a un tesoro. En la segunda acepción se señalan las características de la organización y el funcionamiento de un tesoro y se menciona que la estructura y organización de un tesoro se fundamenta en las relaciones semánticas que ocurren entre los términos. También se explica que las unidades léxicas que contiene un tesoro, provienen de una serie de textos de autores relevantes de una lengua en un periodo de tiempo determinado.

⁸ Martínez de Sousa, J. (1995) *Diccionario de lexicografía práctica*. Barcelona: Bibliograf. p. 330.

Y finalmente, es una definición en la que se muestra que un tesoro, es una serie de unidades léxicas especializadas entre las que se establecen correspondencias semánticas, por medio de la sinonimia. Los términos que contiene un tesoro representan las nociones que se encuentran en textos que pueden ser susceptibles de tratamiento informático para la generación de diferentes tipos de índices. Algunas de las características señaladas que aparecen en la definición de Martínez de Sousa; nos permiten conocer más sobre la complejidad de la estructura de un tesoro, su funcionamiento, aplicación y diseño.

Asimismo, Martínez de Sousa (1995)⁹ señala que el nombre de tesoro en su forma latina *thesaurus*, apareció en la época renacentista para denominar a los diccionarios monolingües. El autor muestra y explica las particularidades de por lo menos ocho diferentes diccionarios de este tipo.

Un tesoro de acuerdo con María Moliner (1999)¹⁰ se define de la siguiente forma: “Del latín (*thesaurus*). 1 (ant). M. Tesoro. 2 (ant) Se aplica al nombre particular, en vez de <tesoro> a algunos libros, tales como antologías o diccionarios. 3 Lista ordenada de palabras clave que sirven para clasificar un documento según su contenido”.

⁹ *Ibid.* p 330.

¹⁰ Moliner, M.(1999) *Diccionario de uso del español*. Madrid: Gredos. Vol 2. p. 1220.

Por su parte Hartmann (2006)¹¹ define el término señalando que éste “ha tenido diferentes extensiones en el significado a lo largo de los años. Ahora puede decirse que el término tesoro, tiene al menos tres diferentes sentidos: i) lista especial de palabras o *lexicon* ii) diccionario semántico o *nomenclatur* y iii) base de datos terminológica o índice”.

La primera acepción, una lista especial de palabras, que de acuerdo con Hartmann, ahora es obsoleta y se asocia con los proyectos lexicográficos relacionados con diccionarios literarios como el de Thomas Cooper, *Thesaurus Linguae Romance et Britannicae* (1665); o el *Thesaurus Linguae Latine* (1894). Y señala que la segunda acepción está relacionada con la idea de un diccionario organizado semánticamente. Un producto típico de este enfoque es el famoso *Thesaurus of English Words and Phrases*, elaborado a mediados del siglo XIX por el famoso médico británico Peter Mark Roget, que continúa actualizándose hasta ahora. En las librerías se venden tesauros que tienen este nombre.

La tercera acepción se refiere a una de las tareas más relevantes de la bibliotecología contemporánea, como ya se ha reconocido antes en este trabajo, es decir, la indización y recuperación de la información.

Esa última acepción es la más actual y se vincula de manera estrecha con la idea de elaborar un índice para un vocabulario de un campo concreto del

¹¹ Hartmann, R.R.K ed. (2006). “Thesauruses” en *Encyclopedia of Language and Linguistics*. 2nd ed. vol 12. Amsterdam: Elsevier. p 668.

conocimiento especializado, que se puede utilizar para el estudio y la codificación de terminología técnica y la recuperación de la información.

Un ejemplo de una obra que ha sido elaborada de acuerdo con esta acepción es el *ASIS Thesaurus of Information Science and Librarianship Second Edition (1998)*, tesauro publicado por la *American Society for Information Science*, que se editó con el propósito de cubrir diferentes campos nucleares y periféricos de estudio e investigación de la Bibliotecología y Ciencia de la Información y para que sirviera de apoyo en las tareas de indización y de recuperación de información. Una característica que distingue a esta edición es que incluye descriptores relacionados con diferentes campos del saber como la ciencia de la computación, lingüística y ciencias cognitivas y del comportamiento. Asimismo incluye descriptores sobre campos periféricos como educación, economía, estadística y sociología.

1.2.2. Desde la Informática

En la Informática el término tesauro se define desde diferentes enfoques. En el *Diccionario de Informática (1993)*¹² se define como “característica de los sistemas de tratamiento de textos por la cual las expresiones similares y sinónimos pueden visualizarse en pantalla e incorporarse en el texto. Los sistemas de recuperación de texto completo pueden poseer *búsqueda de*

¹² Ilingworth, V (ed.) (1993). *Diccionario de informática*. Madrid: Oxford University Press/Diaz de Santos. p 638.

tesauro como opción por la cual pueden establecerse también términos similares en significado a los buscados”.

Asimismo y de acuerdo con Virga y Mesters (1994)¹³, un tesauro es “un léxico jerarquizado de términos comprensibles por el programa para la búsqueda de información. Un tesauro es, por tanto, una herramienta indispensable para consultar de forma eficaz un banco de datos. Para cada utilización adecuada, es necesario crear un nuevo tesauro (no se consulta de igual forma un banco de datos jurídico que uno médico)”. Finalmente, en el *Dictionary of Computing* (1995)¹⁴ un tesauro es “un archivo que contiene sinónimos que se presentan como alternativas para solucionar el deletreo erróneo de una palabra (*misspelet Word*) durante la verificación del deletreo (*spell-check*)”.

Otra definición es la que presenta Arano, Silivia (2005)¹⁵ y en la que caracteriza al tesauro en entornos digitales. En esta caracterización señala cuatro elementos que representan una innovación para el diseño, elaboración, gestión y uso de los tesauros en ambientes digitales. El primer elemento se refiere “al enriquecimiento de la funcionalidad de la estructura de los tesauros a partir de la hipertextualidad. Esto se traduce en el establecimiento de hipervínculos, entre

¹³ Virga y Mesters (1994) *Diccionario de Microinformática: todas las nuevas tecnologías informáticas*. Madrid: Thompson / Paraninfo p. 404 -405.

¹⁴ *Dictionary of Computing*. (1994). 2ª edición. Cambridge: Peter Collins p. 280

¹⁵ Arano, Silvia (2005) Los tesauros y las ontologías en la Biblioteconomía y la Documentación [en línea]. *Hipertex.net*, número 3, p 3-4. <<http://www.hipertex.net>> [consulta 12 de abril 2010]

todos los elementos estructurales (descriptores, no descriptores, notas de alcance, etc.) y también entre las distintas partes del tesoro”

El segundo elemento “es la reducción de costos de actualización y mantenimiento” El tercer elemento es “la integración del usuario al proceso de creación, gestión y optimización de los tesauros a través de test de usabilidad, uso de técnicas de modelado de usuario, etc”. Y el cuarto elemento, es “la posibilidad de aplicar medidas de reutilización e interoperabilidad en el momento de planificar y construir tesauros”

De acuerdo con la autora citada, se puede decir entonces, que para la generación de los tesauros digitales, a partir de la reutilización y la interoperabilidad y con el uso de información conceptual y lingüística almacenada en otro tipo de recursos, por ejemplo una ontología, se promoverá el enriquecimiento y solidez a nivel estructural de los tesauros. Asimismo se promoverá el incremento de las condiciones “amigables” para su uso por parte de los usuarios finales no especializados.

1.2.3. Desde la Terminología

Para la Terminología y de acuerdo con Cabré (1999)¹⁶, un tesoro “es uno de los instrumentos que normalmente se utilizan para clasificar la información, a través de una estructura de descriptores basada en relaciones que

¹⁶ Cabré, T. (1999). *La Terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. p. 237 – 238.

presumiblemente se establecen entre los conceptos de un documento, conceptos expresados básicamente mediante unidades de carácter terminológico. Los tesauros se caracterizan por tres aspectos: por su contenido, por su estructura y por su función. Desde el punto de vista del contenido, un tesoro contiene una serie de términos relacionados entre sí semánticamente en tres dimensiones: la sinonimia, la hiponimia y la asociación. Desde el punto de vista de la estructura, son documentos muy formalizados que constan de un vocabulario, controlado y dinámico, y de una serie de relaciones conceptuales, expresadas formalmente a través de expresiones codificadas. Desde el punto de vista de la función, son instrumentos de control que regulan el uso del lenguaje natural con objeto de facilitar la recuperación de la información”.

1.2.4. Análisis de las definiciones del término tesoro

Después de analizar las definiciones presentadas se pueden hacer una serie de consideraciones.

1. De manera general, se puede afirmar que hay una coincidencia en la idea de que el tesoro es una herramienta apropiada para las tareas de indización y recuperación de información. Se pueden observar diferencias en la forma y enfoque que presentan la información, es decir, en algunas de las definiciones se enuncian de forma sencilla las características generales del concepto y se utilizan enfoques diferentes para definir el concepto.

2. Las obras de consulta especializadas en lingüística y lexicografía que fueron consultadas definen el término *tesauro* con mayor profundidad y desde una perspectiva histórica y semántica, pues presentan información sobre su origen y desarrollo histórico; señalan, de manera precisa, sus diferentes acepciones y usos, e indican sus características, esto es, ser un léxico especializado organizado jerárquicamente bajo criterios semánticos. También mencionan las distintas formas en que se utiliza dicho léxico, es decir, que sirve para el estudio y codificación de terminología técnica y la recuperación de información.
3. Las definiciones que fueron consultadas en diccionarios de informática presentan información más escueta que la que ofrecen las definiciones en la lingüística y lexicografía, retoman elementos de otras disciplinas y sobre todo se enfocan en el empleo y aplicación de los tesauros en los sistemas para la recuperación de información. Asimismo se presenta una caracterización del tesauro en entornos digitales y sus implicaciones para el diseño, elaboración, gestión y uso.
4. La definición de tesauro que se presenta desde la terminología es más completa e integradora, dado que incluye aspectos sobre el contenido, la estructura interna de la herramienta y su función e importancia para la recuperación de información y, por supuesto, para el trabajo terminológico.

De lo anterior se puede decir que para la lingüística un tesauro no es más que un léxico especializado que se organiza de una forma diferente a un diccionario o a un glosario, es decir, su organización interna está regulada primordialmente

bajo criterios semánticos y con una estructura jerárquica del tipo género-especie, en la que se establecen relaciones de asociación, equivalencia y jerarquía.

Para la informática un tesoro es una herramienta lógico conceptual a través de la cual se representa el contenido de los documentos. Se trata de un sistema complejo en el que se establecen relaciones entre los términos para crear una red semántica que permita al sistema realizar tareas de indización y recuperación de información.

Para la terminología es una herramienta terminológica a través de la cual se clasifica la información, por medio de descriptores que están organizados semánticamente y cuya función es la regulación del lenguaje natural con el propósito de indizar y recuperar información.

Enseguida se expondrán una serie de definiciones del término desde la Bibliotecología, la cuales, al igual que las anteriores, se tomaron de fuentes documentales diferentes, entre las que se incluyen artículos relevantes del tema, diccionarios y enciclopedias especializadas en Bibliotecología y obras fundamentales sobre el estudio de los lenguajes documentales, específicamente sobre los tesauros. Finalmente, se presentan algunas preguntas que surgen del análisis realizado.

1.3. El tesoro en Bibliotecología y Estudios de la Información

Si bien es cierto que el tesoro tiene relevancia en la Lingüística, la Informática y la Terminología, parece ser claro que hoy en día en Bibliotecología y Estudios de la Información, su carácter de herramienta se acentúa junto a otros dos tipos de instrumentos característicos del campo que se utilizan para la organización del conocimiento y para el control del vocabulario en sistemas de almacenamiento y recuperación de información. Dichos instrumentos son los esquemas de clasificación y listas de encabezamientos de materia.

En esta sección se analizarán los términos antes mencionados, desde la óptica de la Bibliotecología y los Estudios de la Información, con el propósito de constituir una imagen más completa de conceptos e ideas que subyacen en la elaboración de un tesoro, con lo que se contribuye al enriquecimiento de la disciplina bibliotecológica.

Los esquemas de clasificación son, de acuerdo con el Glosario de la American Library Association,¹⁷ “listas que constituyen el esquema que dispone el ordenamiento de todas las materias del saber, dividiéndolas en clases. Pueden ir acompañadas de una notación, o sistema de símbolos que representan las clases. En un sistema de clasificación jerárquica, el cuadro muestra los niveles

¹⁷ Young, H. (ed.). (1988). *Glosario ALA de Bibliotecología y Ciencias de la Información*. Madrid: Díaz Santos. p 67.

jerárquicos de las clases fundamentales, las divisiones, subdivisiones, secciones en las que se dividen las clases”.

Las listas de encabezamientos de materia, por su parte, son de acuerdo con el mismo Glosario de la American Library Association,¹⁸ una “lista estándar de los términos que deben utilizarse como encabezamientos de materias o para el campo entero del conocimiento o para un área temática limitada, incluyendo referencias hechas a cada término y desde éste notas explicando el ámbito o uso de ciertos encabezamientos y ocasionalmente, los números de clasificación correspondientes”.

En Bibliotecología y Estudios de la Información el tesoro tiene una gran relevancia que se refleja en una multiplicidad de definiciones del término.

El tesoro en Bibliotecología y Estudios de la Información es relevante tanto de manera general como para el almacenamiento y la recuperación de información específicamente. Se puede definir de acuerdo con Miller (1997)¹⁹ “como un modelo léxico-semántico de una realidad conceptual o sus constituyentes, la cual se expresa como un sistema de términos y sus relaciones. Ofrece acceso a través de múltiples vías y aspectos y se utiliza como una herramienta en el procesamiento y búsqueda en una unidad de recuperación de información”.

¹⁸ *Ibid.* p 335 – 336.

¹⁹ Miller, U. (1997) “Thesaurus Construction: problems and their roots” en *Information Processing & Management*. 33 (4). Londres: Pergamon. pp. 481 – 493

En esta misma perspectiva para Aitchinson & Gilchrist (1987)²⁰ el tesoro se convierte en un vocabulario controlado, que forma parte de un lenguaje de indización y está previamente organizado con el fin de establecer de manera explícita relaciones entre los conceptos para propósitos de indización y recuperación de información.

De la misma manera Slype (1991),²¹ coincide con la definición anterior, al considerar al tesoro como “una lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de documentos y de las consultas dentro de un sistema documental determinado, y ayudar al usuario en la indización de los documentos y las consultas; los conceptos son extraídos de una lista finita, establecida *a priori*, sólo los términos que figuran en esta lista pueden ser utilizados para indexar los documentos y las consultas; la ayuda al usuario la proporciona la estructura semántica del tesoro: fundamentalmente las relaciones de equivalencia, de jerarquía y asociación”.

Con respecto a lo que señala Slype la UNESCO (1970)²² considera, que un tesoro es “un vocabulario controlado y dinámico de términos semántica y genéricamente relacionados que cubren ampliamente un campo específico del conocimiento. Este vocabulario es una colección sistemática o alfabética de

²⁰ Aitchinson, J. y Gilchrist, A. (1987). *Thesaurus Construction: a practical manual*. (2ª ed.). Londres: Asilb.

²¹ Slype, G. V. *op cit.* p. 23 – 24.

²² UNESCO (1970). *Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri for Information Rerieval*. París.

descriptores o no descriptores (términos auxiliares), así como de indicadores de relaciones entre los términos”.

Ampliando la información sobre la definición del tesoro y que se considera relevante para el análisis propuesto es la que COSATI (1973)²³ nos presenta y en la que se señala que “un tesoro es una compilación de términos selectos en los que se observa una relación apropiada y se manifiesta y hace ostensible esta relación entre los términos, de tal manera que promueva consistencia en la descripción de los conceptos para la indización y la búsqueda”.

Una aportación relevante para la definición del tesoro es la que propone Arano Silvia (2005)²⁴ y señala que un tesoro “es un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo de conocimiento. El tesoro proporciona una organización semántica principalmente a través de la explicitación de las relaciones establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los términos que los representan. Los tesoros, para la Documentación y la Bibliotecología, son por lo tanto una herramienta de control terminológico, puesto que la estructuración del tesoro se basa en conceptos”.

²³ COSATI (1973). *Guidelines for the development of information retrieval thesauri*. Washington: Government Print Office. Mencionado por: Kim, Ch. “Theoretical Foundations of Thesaurus – Construction and Some Methodological Considerations for Thesaurus–Updating” en *Journal of the American Society Information Science*. 24 (2). March – April p.149.

²⁴ Arano, Silvia (2005) Los tesoros y las ontologías en la Biblioteconomía y la Documentación [en línea]. *Hipertex.net*, número 3, p 3. <<http://www.hipertex.net>> [consulta 12 de abril 2010]

Por su parte Broughton (2006)²⁵ apunta que “el término tesoro en ciencia de la información se utiliza desde 1950, cien años después que el Roget”, y lo define “como una herramienta que se utiliza para la indización temática de documentos, que consiste en una lista estructurada de términos (generalmente en un campo específico del conocimiento), que un indizador puede usar para describir un documento que un usuario final podría recuperar ítems relevantes cuando busca documentos sobre un tema en particular”.

Finalmente, Barité (1997)²⁶ define los tesauros como “el tipo de vocabulario controlado que se integra con términos analizados y normalizados que guardan entre sí relaciones semánticas y funcionales. El tesoro se organiza bajo fuerte control terminológico, con objeto de proporcionar un instrumento idóneo para el almacenamiento y la recuperación de la información en áreas especializadas”.

1.3.1. Rasgos comunes de las definiciones de tesoro

En las definiciones que se presentaron anteriormente se pueden observar diferentes puntos de vista o maneras de abordar la definición del término tesoro. Aquí resulta pertinente analizar cuáles son los rasgos que se comparten en las definiciones y que caracterizan de manera substancial lo que es un tesoro. Los rasgos son los siguientes:

²⁵ Broughton, V. *Op Cit.* p. 4 – 5.

²⁶ Barité, M. *Op Cit.* p. 145.

En cuanto su forma, los rasgos que caracterizan al tesoro son:

- Un vocabulario controlado y dinámico.
- Una lista o compilación de términos.
- Una lista estructurada de conceptos.
- Un sistema de términos y sus relaciones.

En cuanto a su estructura, los rasgos que caracterizan al tesoro son:

- Relaciones semánticas y genéricas entre los términos.
- Formalmente organizado *a priori*.
- Relaciones de equivalencia, de jerarquía y asociación.
- Control terminológico.

En cuanto a su función, un tesoro es:

- Una representación de manera unívoca del contenido de documentos.
- Una herramienta para el procesamiento a través de la indización y para la búsqueda y recuperación de la información.

Forma	Estructura	Función
Vocabulario controlado y dinámico.	Relaciones semánticas y genéricas de términos.	Herramienta para la indización.
Lista o compilación de términos.	Formalmente organizado a priori.	Herramienta para la búsqueda de información.
Lista estructurada de conceptos.	Relaciones de equivalencia, asociación y jerarquía.	Herramienta para la recuperación de información.
Sistema de términos y sus relaciones.	Representación de manera unívoca del contenido de los documentos.	Herramienta para la indización y la recuperación de información.

Cuadro 1. Rasgos comunes en las definiciones del término tesoro. Elaboración, Valdez, R. J. (2013).

1.3.2. El tesoro en la indización

Para explicar cuál es el papel del tesoro en el proceso de indización y en qué momento del mismo se utiliza, a continuación se presentarán dos enfoques: uno de carácter teórico procedente de la Lingüística del Texto, y de los estudios de Dijk, Teun A. Van sobre el Análisis del Discurso y la Ciencia del Texto. Y otro de carácter práctico que proviene del procedimiento de indización.

Para explicar cuál es el papel del tesoro desde un enfoque teórico se plantea lo siguiente: por una parte la Lingüística del Texto se define, de acuerdo con Lewandowski (1982),²⁷ como “una sub disciplina lingüística relativamente nueva, que estudia las reglas y formas de organización del lenguaje que van más allá del límite de la frase; lingüística orientada al texto se entiende como consecuencia necesaria de los conocimientos modernos sobre el lenguaje y la comunicación”. La amplitud e interdisciplinariedad característica de la Lingüística del Texto que señala Theodor Lewandowski es importante para explicar el papel que tiene un tesoro en la indización. O bien, como lo señala Piller (2005)²⁸ “a la Lingüística del Texto se le puede atribuir la función de abrir nuevos horizontes para la lingüística, por medio de la cooperación interdisciplinaria con estudios literarios, educativos y otros campos del saber”.

Por otra, en la obra La Ciencia del Texto de Dijk Teun A. Van (1978)²⁹ se afirma que “el análisis de las estructuras y funciones de los textos requiere un modo de

²⁷ Lewandowski, T. (1982). *Diccionario de Lingüística*. Madrid: Cátedra. p. 215.

²⁸ Piller, I. (2005). “Text Linguistics” en Philipp, S (ed). *Encyclopedia of Linguistics*. Vol. 2. New York: Fitzroy Dearborn p. 1095

²⁹ Dijk Teun A. Van (1978). *La ciencia del texto. Un enfoque interdisciplinario*. Barcelona : Paidós p. 10.

proceder interdisciplinario, esto tiene validez aún más amplia para la investigación de la *utilización de la lengua y la comunicación*” Asimismo señala que “la tarea de la *Ciencia del Texto* consiste en describir y explicar las relaciones internas y externas de los distintos aspectos de las formas de comunicación y uso de la lengua”³⁰. Así al considerar que la indización es un proceso cognitivo a través del cual se formulan expresiones textuales a partir del análisis documental de resúmenes, títulos de artículos o artículos completos, en algunos casos, que representan el contenido de los documentos.

El modelo que ofrece Dijk Teun A. Van desde los estudios e investigaciones en el campo del Análisis del Discurso, que se puede definir de acuerdo Zaldúa Garoz, Alexei (2006)³¹ con como “un suceso de comunicación en el que los sujetos participantes utilizan el lenguaje para comunicar ideas y creencias y lo hacen como sucesos sociales más complejos en los que se produce un discurso que se basa en modelos mentales que se tienen de un acontecimiento y por motivos como:

- El conocimiento del receptor
- Las especificidades de los canales de comunicación y de los recursos de información
- Y en el que sólo se expresa parte de la información y se deja implícita otra parte de ella en la enunciación”

³⁰ *Ibid* p. 10

³¹ Zaldúa Garoz, Alexei (2006). El análisis del discurso en la organización y representación de la información-conocimiento: elementos teóricos. *ACIMED* [online] 14 (3) pp. 1-16. Disponible en: <http://scielo.sld.cu/scielo.php?script=sci> ISSN 1024-9435.

Si tomamos en cuenta lo anterior, es decir que para realizar diferentes procesos de indización, se deben conocer las características del receptor/ usuario, los canales específicos de comunicación y los recursos de información con que cuenta el receptor, entonces podemos considerar que el modelo del Análisis del Discurso de Dijk Teun A. Van resulta adecuado para explicar los procedimientos de indización de documentos.

Para ubicar el lugar que tiene el tesauro en el procedimiento de indización se reconoce lo dicho por Hutchins (1978)³² en cuanto que el objeto de la indización es esencialmente la formulación de una expresión en un lenguaje documental; expresión que deberá capturar el contenido semántico de un documento.

La indización entendida como un proceso que se realiza para alcanzar un fin determinado, se inicia con la descripción de un texto en lenguaje natural, esto es, que está integrado por un número determinado de párrafos, oraciones y palabras ligadas en una secuencia lineal, y termina con un descriptor-resumen o la condensación del texto.

En el proceso de indización la intervención del tesauro comienza con una serie de preguntas como: ¿qué pasa cuando un documento es indizado?, ¿cómo es que los indizadores utilizan un tesauro para describir el contenido de un documento? La respuesta a dichas preguntas se encuentra en la descripción detallada del proceso de indización.

³² Hutchins, W. J. (1978). *Languages of indexing and classification*. Inglaterra: Peter Peregrinus. p. 107.

Se pueden distinguir al menos dos tipos de indización: la indización humana y la indización automática. De acuerdo con Slype (1991)³³ la indización humana “es la operación que consiste en enumerar los conceptos sobre los que trata un documento y representarlos por medio de un lenguaje combinatorio, que puede ser una lista de descriptores libres; una lista de autoridades o un tesoro de descriptores”.

La indización humana es una actividad en la que, por diversas razones, la apreciación humana juega un rol muy importante. Una de ellas es la intervención del indizador al momento de distinguir, lo que Slype llama *unidades significativas*, es decir, conceptos o representaciones mentales de objetos de conocimiento. Otra razón es la selectividad a través del análisis crítico de los documentos, en donde el indizador reconoce los conceptos por medio de un proceso selectivo que puede ser de dos tipos: explícito o implícito.

El primero, o de carácter explícito, se realiza a partir de conceptos que están presentes en los documentos y que podrían interesar a los usuarios. Mientras que en el segundo, o implícito, el ser humano interviene de manera más contundente en el momento en que, de acuerdo con Slype (1991),³⁴ “busca los conceptos implícitos sobre los que, aunque no estén designados nominalmente en el documento, se aporta información dentro de él”.

³³ Slype, G. V. *op cit.* p. 115.

³⁴ *Ibid.* p. 116.

De acuerdo con Slype (1991) en el proceso de indización humana se distinguen, además, cuatro etapas:

- Revisión del documento
- Selección de conceptos
- Traducción de conceptos en descriptores
- Establecimiento de enlaces sintácticos entre los descriptores

Además de las cuatro etapas que reconoce Slype (1991) y de acuerdo con los avances realizados sobre el tema de la indización humana, una etapa quinta que deberá considerarse en el proceso de indización es la:

- Adecuación de los descriptores al contexto

Que está relacionada con la consistencia en el proceso de la indización, misma que ha sido estudiada por diferentes autores desde la década de los sesenta hasta la primera década del siglo XXI. Algunos de estos autores son los que señala Gil Leiva Isidoro (2008)³⁵ y que a continuación se mencionan: Zunde & Dexter (1969) ; Tar & Borko (1974) ; Markey (1984) ; Chan (1989) ; Saracevic (1991) ; Bertrand – Gastaldy y Lanteigne (1995) ; Fujita (1999) ; Matríns Lopes Naves (2002) y Fujita, Nardo y Fagundes (2003).

El segundo enfoque para explicar cuál es el papel del tesoro en el proceso de indización es de carácter práctico y se refiere a la descripción del procedimiento de indización. En seguida se describen cada una de las etapas de dicho proceso.

³⁵ Gil Leiva, Isidoro (2008). *Manual de indización. Teoría y práctica*. Madrid: Trea p. 74.

1.3.2.1. Revisión del documento

En la revisión del documento el indizador hace el contacto inicial a través de una lectura rápida y superficial de las diferentes partes del documento: el título, el índice, el resumen, la introducción general, las introducciones y conclusiones de los capítulos o secciones del documento, los anexos, las tablas, las figuras y las conclusiones generales.

1.3.2.2. Selección de conceptos

En esta etapa se puede realizar la selección de conceptos explícitos y conceptos implícitos. El indizador elige los conceptos a partir de una lectura profunda, pausada y analítica del documento. En esta lectura el indizador identifica y retiene las nociones que están explícitamente presentes y que guardan una relación directa con las necesidades de información del grupo de usuarios que utilizará el fondo documental. El indizador puede también identificar las nociones que están implícitamente contenidas en el documento y que no han sido enunciadas por el autor, pero que han sido tratadas de manera detallada por el autor y considere de interés para los usuarios.

1.3.2.3. Traducción de conceptos en descriptores

La traducción de conceptos en descriptores, de acuerdo con Slype (1991)³⁶ “es la operación que se realiza en varias fases, la primera consiste en la traducción

³⁶ *Ibid.* p. 117.

de los conceptos a la lengua del sistema, la segunda es la indización por medio de una lista de descriptores libres, la tercera es la indización por medio de una lista de autoridades y la cuarta y última es la indización por medio de un tesoro de descriptores". Cada una de estas fases tiene sus características y procedimientos propios, como se mencionan a continuación.

Primera fase. La traducción de conceptos de una lengua a otra, es simplemente eso, se traducen los conceptos a la lengua o lenguas que se utilizan en el sistema documental.

Segunda fase. La indización a través de una lista de descriptores libres. En esta fase el documentalista toma por lo menos un par de decisiones a fin de establecer algunos parámetros para el control del vocabulario. El primer nivel de control puede ser mínimo y consiste en tomar las designaciones de los conceptos tal y como se encuentran en los documentos analizados, tal y como han sido traducidos, tal y como han sido enunciados por el documentalista.

En un segundo nivel el control del vocabulario consiste en la transformación de las formas verbales y adjetivas en frases nominales. Por ejemplo *medir* se transforma en *medida* y *largo* se transforma en *longitud*. O bien la transformación puede consistir en poner formas en femenino y plural en un solo concepto masculino y singular, por ejemplo *doctora(s)* se transforma en *doctor*. Se trata aquí de un nivel de control del vocabulario más preciso, en el que el documentalista hace un ejercicio de comparación entre las designaciones normalizadas de los conceptos y una lista de descriptores libres.

Tercera fase. La indización por medio de una lista de autoridades. Esta fase consiste en la asignación que hace el documentalista de un término a través del cual se expresan el o los conceptos que trata el documento. Esta tarea pudiera ser similar a la que hace el documentalista en la traducción de conceptos en descriptores, la diferencia estriba, sin embargo, en que los términos los toma de una lista de términos autorizados con anterioridad y que han pasado por un proceso de normalización léxica como la que se describe en la segunda fase del proceso de traducción de los conceptos en descriptores.

Cuarta fase. Indización por medio de un tesoro de descriptores. De acuerdo con Slype (1991)³⁷ este proceso se compone de tres casos: el primero “cuando la expresión del concepto corresponde a un descriptor que lo represente y éste se transcribe en el listado de indización”; el segundo “cuando la expresión del concepto corresponde a un no-descriptor que lo represente, éste se envía al descriptor equivalente que se ha de usar”; y el tercer caso “cuando no existe una entrada dentro del tesoro que corresponda exactamente a la designación del concepto”.

1.3.2.4. Establecimiento de enlaces sintácticos entre los descriptores conservados

Establecer enlaces sintácticos entre los descriptores que se conservan y que derivan del trabajo de indización, es una tarea que consiste en separar los descriptores en diferentes categorías mediante el empleo de diferentes signos

³⁷ Slype, G. V. (1991) *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid : Pirámide. p. 119.

de puntuación, por ejemplo: punto y coma, puntos, comas, espacios en blanco o signos como “ / ” para establecer o señalar diferencias de jerarquía entre los descriptores. De acuerdo con ³⁸Slype (1991) “en los sistemas documentales se pueden utilizar seis tipos de enlaces:

- Yuxtaposición
- Ponderación
- Especificación de un punto de vista
- Especificación de un vínculo
- Especificación de un papel
- Integración dentro de un enunciado estructurado”

En los siguientes apartados se explicará cada uno de estos tipos de enlaces, que menciona Slype (1991)³⁹

1.3.2.4.1. Yuxtaposición

Consiste en el registro de los descriptores unos después de otros. Los descriptores se separan unos de otros por un signo, como el punto y la coma. Por ejemplo, si se trata de un documento sobre práctica de la lectura en adolescentes en el salón de clase y su relación con la ocupación profesional de sus padres, entonces la cadena de indización por simple yuxtaposición será

³⁸ Slype, G. V. *Op Cit.* p. 120.

³⁹ Slype, G. V. *Op Cit.* p. 119 – 122.

como sigue: *lectura; adolescentes; salón de clase; ocupación profesional; padres.*

1.3.2.4.2. Ponderación

Es el procedimiento a través del que se distinguen los descriptores en dos grupos: los principales y los secundarios. Los primeros reflejan el contenido principal del documento, en tanto que los segundos reflejan los conceptos sobre los que el documento aporta menos información. Siguiendo con el ejemplo del punto anterior, al identificar los descriptores principales y secundarios, los resultados de la ponderación quedarían de la siguiente manera: *lectura; adolescentes; salón de clase*, como descriptores principales y como descriptores secundarios, *ocupación profesional; padres.*

1.3.2.4.3. Especificación de un punto de vista

Este procedimiento consiste en la expresión de un punto de vista específico que asigna el indizador sobre el contenido de un documento, a través de un descriptor. Así el indizador establece un enlace entre el descriptor principal y el punto de vista que se expresa en un descriptor auxiliar. El indizador generalmente utiliza un signo de puntuación, como los dos puntos (:) para evidenciar el enlace. Por ejemplo, *lectura : hábitos de lectura : diagnóstico.*

1.3.2.4.4. Especificación de un vínculo

Esta tarea consiste en la creación de un índice que enlaza los descriptores de un documento cuyo contenido aborda diferentes temas. El índice contiene una o

varias cadenas de descriptores que pueden expresarse de la siguiente forma: *lectura (1); adolescente (1); salón de clase (1); ocupación profesional (2); padres (2)*.

1.3.2.4.5. Especificación de un rol o papel

La especificación del rol o papel que juega un concepto en un documento, consiste en añadir un indicador del rol que se expresa a través de un código. Con este procedimiento se crean sub registros a partir de los descriptores y no a partir de los documentos como sucede en la especificación de un vínculo. Por ejemplo, un documento que trata sobre la exportación de calzado entre China y México. Si utilizamos la técnica de especificación del rol y asignamos un código el resultado de la indización será la siguiente. *Exportación; calzado; China (O); México (D)*, donde el código es O = Origen y D = Destino.

1.3.2.4.6. Integración dentro de un enunciado estructurado

Esta tarea consiste en integrar en el texto de un enunciado estructurado que elabora el documentalista, una serie de descriptores para expresar el contenido conceptual del documento. Generalmente se distinguen los descriptores con signos especiales, como las comillas.

Un ejemplo de la integración de descriptores en un texto de enunciados estructurados es el caso de un resumen. Sin embargo cabe señalar que no sólo en este tipo de textos se integran descriptores. Para ilustrar este proceso se presenta a continuación, un ejemplo de un resumen: “El progreso de la *sociedad*

reticular depende del desarrollo de las *redes telemáticas* ciudadanas – *community networks*- como servicios públicos de información para las *comunidades locales*. Surgen barreras formativas que requieren la oportuna alfabetización del ciudadano.” Donde los descriptores se encuentran en cursivas.⁴⁰

1.3.2.5. Adecuación de los descriptores al contexto

La adecuación de los descriptores al contexto es un parte importante del proceso de indización y se puede ubicar como una tarea que se concibe desde el plano pragmático del lenguaje, es decir el uso de la lengua en situaciones y contextos determinados, a diferencia del plano sintáctico de los procesos analizados hasta ahora. Considerando a la Pragmática de acuerdo con Serrano Albuja, A (2010)⁴¹ “como uno de los niveles de estudio lingüístico desde el cual puede analizarse el signo lingüístico. La pragmática se ocupa de las relaciones entre el signo lingüístico y los factores extralingüísticos que intervienen en el proceso de significación, entre ellos, por ejemplo, la clase social de los hablantes, la situación comunicativa, el tiempo y el lugar donde sucede un acto comunicativo”. A partir de esta definición podemos entonces decir que la adecuación de los descriptores al contexto es una etapa de la indización en la que se debe tomar en cuenta lo siguiente: el perfil y necesidades de los de usuarios, para identificar el las condiciones de comunicación entre el usuario y el bibliotecario. Y las políticas institucionales de indización, entre otras.

Ejemplo tomado del libro de María Pinto *et. al.* (2005) *Aprendiendo a resumir. Prontuario y resolución de casos*. Gijón, Asturias, España: Treda p.82.

⁴¹ Serrano Albuja, A (2010) *Cómo recuperar información en la biblioteca. Estrategia de indización con lenguajes combinados*. Caracas : Gráficas Iberia p. 33.

1.4. El tesoro en el almacenamiento y la recuperación de la información

Para explicar el rol que tiene el tesoro en el almacenamiento, búsqueda y recuperación de información, se iniciará esta sección con la definición de dichos términos y, enseguida, se explicará cuál es la relación que guarda el tesoro con estas actividades.

En el *Diccionario de Informática* (1993)⁴² se define el almacenamiento “como la actividad de registrar los datos para su uso posterior”. Asimismo se señala que “el almacenamiento es una de las operaciones fundamentales que pueden realizarse con la información, es la operación principal por la cual los datos se retienen durante un espacio de tiempo”.

La recuperación de la información se define desde dos enfoques diferentes. Por una parte es la actividad que consiste en recuperar los datos almacenados y constituye otra de las operaciones fundamentales que pueden llevarse a cabo con la información. Por otra, y desde el campo de la Documentación, el término recuperación de información, se emplea para referirse a técnicas (tales como palabras clave e índices) utilizadas para situar textos y/o referencias sobre un tema específico.

En la definición anterior podría parecer que el almacenamiento y la recuperación de información son dos actividades separadas, pero esto no es así, en realidad

³⁰ Illingworth, V. (ed.). *Op Cit.* p 301.

se trata de una actividad integradora. Vaswani (1970)⁴³ reconoce al almacenamiento y la recuperación de información como una actividad integradora de una serie de procesos que suceden en un sistema de información. Para Vaswani (1970) dicho sistema de información se puede definir como tal en términos de la respuesta que éste ofrece ante una solicitud de información, donde las solicitudes de información se expresan en lenguaje natural. Dicho lenguaje no siempre se presenta en la forma más idónea para que las solicitudes de información ingresen a un sistema automatizado. Sino que en cada caso será necesario expresar la solicitud de información en una forma estandarizada, por ejemplo reteniendo solamente las palabras clave de una solicitud.

En la definición anterior se atisba una relación estrecha e inicial entre el tesoro y la recuperación de la información, la cual se observa en la estandarización de un lenguaje natural en un lenguaje controlado en la expresión de una solicitud de información. Es decir, por una parte, en la creación de una serie de procedimientos normalizados que permitan al sistema de información la recuperación de datos precisos y consistentes con la necesidad de información de un usuario determinado —dichos procedimientos son los que se refieren a la tarea de la indización—; y por otra, en el establecimiento de un lenguaje creado de manera artificial cuya organización interna esté regulada por criterios semánticos primordialmente, aunque éstos no sean los únicos, pues también interviene una estructura jerárquica del tipo género-especie en la que se establecen relaciones de asociación, equivalencia y jerarquía entre los términos y, con ello, no se está hablando de otra cosa sino de un tesoro.

⁴³ Vaswani, P. T. K. (1970) *Encyclopedia of Linguistics, Information and Control*. Oxford: Pergamon Press p 223.

Otra dimensión de la relación entre el tesoro y la recuperación de la información, es la que se relaciona con una doble dimensión del tesoro, misma que reconoce Aitchinson, Jean, *et. al* (2000)⁴⁴ y señala “que el tesoro se utiliza para la indización de información o para la búsqueda de la misma en una base de datos con cuatro posibilidades diferentes:

- El tesoro puede ser utilizado en la indización y en la búsqueda de información
- El tesoro puede ser utilizado en la indización, pero no en la búsqueda de información
- El tesoro puede ser utilizado en la búsqueda, pero no en la indización
- Y el tesoro puede no ser utilizado en ninguno de los dos casos señalados”

Asimismo Aitchinson, Jean, *et. al* (2000)⁴⁵ plantea que con los recientes desarrollos en el campo de la Informática y el uso de la Internet el tesoro ha adquirido nuevas formas de uso intensivo de acuerdo a los cuatro posibles usos señalados. En la primera de las posibilidades, es decir, cuando se utiliza en la indización y en la búsqueda de información, el tesoro “permite una estrecha correspondencia en los procesos de indización y búsqueda, particularmente cuando ambos procesos se llevan a cabo por las mismas personas, como

⁴⁴ Aitchinson, J. ; Gilchirst, A. ; Bawden, D. (2000) *Thesaurus construction and use: a practical manual*. Chicago, Fitzory Dearborn Publishers, p. 1.

⁴⁵ *Ibid* p.2

sucede, por ejemplo, en una base de datos local y a la que sólo tienen acceso los especialistas en información que dan mantenimiento a la base de datos”.

En la segunda de las cuatro posibilidades, cuando el tesoro se utiliza en la indización, pero no para la búsqueda. En este caso el tesoro proporciona al usuario final del sistema de información o de una base de datos, un conjunto amplio y rico de términos, entre los que se incluyen términos sinónimos y de amplitud tal que permite incrementar las posibilidades exitosas en la recuperación de la información.

En la tercera posibilidad de uso, es decir cuando el tesoro se utiliza sólo para la búsqueda y no para la indización, su papel se reduce a apoyar a la búsqueda en bases de datos de texto libre sugiriendo términos de búsqueda, como sinónimos, términos específicos y términos relacionados.

Asimismo y de acuerdo con Aitchinson, Jean, *et. al* (2000)⁴⁶ se puede decir que para la búsqueda y recuperación de información se cuenta con dispositivos auxiliares, que si bien no forman parte del tesoro en sí mismo, sí guardan una estrecha relación con él. Dichos dispositivos auxiliares se pueden mencionar los siguientes:

- La coordinación de términos: Pre coordinación / Post coordinación

⁴⁶ Aitchinson, J. ; Gilchirst, A. ; Bawden, D. (2000) *Thesaurus construction and use: a practical manual*. Chicago, Fitzory Dearborn Publishers, p. 85

- Dispositivos para la búsqueda post coordinada: Operadores lógicos. Búsqueda a través de la fragmentación de palabras y de la distancia entre las palabras.

A continuación se definen y ejemplifican cada uno de ellos:

La coordinación de términos, es sin duda alguna un dispositivo necesario para la indización y para la búsqueda y recuperación de información. Existen dos tipos de coordinaciones, la pre – coordinación y la post - coordinación. Un ejemplo de los sistemas pre – coordinados son los índices impresos de las bibliotecas o los catálogos bibliográficos de temas, la coordinación de los términos la hace el catalogador o indizador. En cambio en la post coordinación, las posibilidades de acceso y recuperación de información son múltiples y sobre todo multidimensionales, es decir se asigna el mismo peso a cada uno de los descriptores en la indización.

Los dispositivos para la búsqueda post coordinada, de acuerdo con Aitchinson, Jean, *et. al* (2000) son tres:

El primer dispositivo son los operadores lógicos **O**, **Y** y **NO**.

El operador lógico **O** en una búsqueda de información, recupera ítems o documentos que contenga una o todas las características de un enunciado como el siguiente: *Salas de audiología O nivel de sonido*. Entonces se recuperaran documentos que contengan uno o más de estos términos. El uso de este

operador amplía la búsqueda, reduce el nivel de coordinación y tiende a incrementar el nivel de recuperación.

El operador lógico **Y** en una búsqueda de información, recupera ítems o documentos que contengan dos o más términos o grupos de términos que coincidan en un mismo ítem o documento, por ejemplo en *Salas de audiología Y nivel de sonido*, se recuperarán sólo aquellos ítems o documentos que contengan ambos términos. El uso de este operador reduce y restringe la búsqueda, incrementa el nivel de coordinación y mejora la precisión en la búsqueda y recuperación.

El operador lógico **NO** en una búsqueda de información, evita la recuperación de ítems o documentos que fueron indizados utilizando un término específico. Por ejemplo en *Salas de audiología NO nivel de sonido*, se excluyen los documentos que contengan el término *nivel de sonido*.

El segundo dispositivo es el uso de la técnica de la fragmentación de palabras en la búsqueda de información. Esta es la técnica a través de la cual el sistema reconoce la raíz o base de un término que trunca o suprime afijos (prefijos o sufijos) al realizar una búsqueda y de esta forma recupera los documentos o ítems relacionados con la raíz o base del término. A continuación se muestran ejemplos del uso de este dispositivo:

Siguiendo el ejemplo del término *Audio* como una raíz o base, al utilizar la técnica de la fragmentación de palabras, en el caso de fragmentación a la derecha de la

raíz o base, el sistema de información recuperará documentos que contengan términos como: *Audiología, Audiofonía, Audioterapia*, etcétera. O bien cuando la técnica de la fragmentación de palabras se realiza con la fragmentación a la izquierda de la raíz o base del término, ocurre que el sistema recuperará documentos que contengan términos como: *speech audio, noisy audio, multilingual audio*, etcétera.

El tercer dispositivo es la *distancia entre palabras*. A través del uso de este dispositivo el sistema de información recupera documentos que contengan términos que se encuentran entre una distancia determinada entre dos o más palabras. En un rango de específico de proximidad se incluyen palabras adyacentes. Por ejemplo, en el caso de la frase *Audiology (w) Rooms*, se recuperarán documentos que contengan los términos adyacentes de acuerdo con el orden señalado en la frase, es decir aquellos que se encuentren en una distancia determinada entre una o más palabras, como se puede observar en la frase *Audiology (w) y Rooms*. De acuerdo con Aitchinson, Jean, *et. al* (2000)⁴⁷ “el uso de este dispositivo incrementa la precisión y reduce la aparente disminución” en la cantidad de documentos que se recuperan.

En el panorama antes expuesto Broughton (2006)⁴⁸ reconoce dos formas en las que el tesoro interviene de manera directa. La primera se refiere a la formulación de la pregunta al sistema de información y la segunda a la expansión de la misma.

⁴⁷ Aitchinson, J. ; Gilchirst, A. ; Bawden, D. (2000) *Thesaurus construction and use: a practical manual*. Chicago, Fitzory Dearborn Publishers, p. 88.

⁴⁸ Broughton, V.(2006) *Op Cit.* p. 32 – 33.

La formulación de la pregunta hoy en día se realiza con el apoyo de los avances en el diseño de programas y *software* que se han dado en los últimos tiempos.

Anteriormente, sólo los indizadores utilizaban el tesoro para la construcción del sistema de términos controlado y sólo ellos tenían acceso a él, ahora con los enlaces de hipertextos que se pueden establecer cuando uno consulta una base de datos, el usuario final puede consultar el tesoro y los términos relacionados de un tema determinado a través de las referencias cruzadas que existen en el mismo tesoro. Esta condición permite que la formulación de la pregunta sea precisa y consistente con la necesidad de información que el usuario ha planteado.

La expansión de la pregunta es la otra tarea en la que el tesoro interviene de manera directa. Esto sucede cuando el tesoro está integrado al *software* de búsqueda y los términos que el usuario plantea en la pesquisa encajan con los términos del vocabulario controlado. En consecuencia el *software* de búsqueda puede expandir la indagación mediante el uso de los términos amplios, abriendo el rango hacia temas más específicos, o bien, buscando sinónimos o términos relacionados.

Capítulo 2

Bases lingüísticas y terminológicas para la elaboración de los tesauros

*Continuar significa ir lejos. Ir lejos significa retornar
Tao Te Ching*

En este segundo capítulo se analizarán los rasgos distintivos que conforman los procesos lingüísticos y terminológicos que intervienen en la elaboración de un tesoro. Para ello, en la primera parte del mismo, se dará un panorama general de los puntos de contacto y convergencia que se han podido identificar entre la Lingüística, la Terminología y la Bibliotecología, en tanto que, en la segunda se examinarán algunas de las aportaciones que han hecho la Lingüística y la Terminología para la construcción de tesauros.

Es preciso acotar que desde la perspectiva de la Lingüística se estudiarán los procesos morfológicos, sintácticos, léxicos y semánticos, mientras que desde el enfoque terminológico se estudiarán los términos como unidades del lenguaje especializado y como unidades de comunicación.

2.1. Bibliotecología, lingüística y terminología: un espacio de colaboración interdisciplinaria alrededor del tesoro

Los puntos de contacto y convergencia que se dan entre la Bibliotecología, la Lingüística y la Informática son diversos y de naturaleza variada, lo que repercute en los productos, modelos y herramientas necesarias para el trabajo cotidiano del bibliotecario. El espacio formado por estas relaciones, así como los productos, modelos y herramientas referidos se ilustra en la Figura 1 que se presenta a continuación:

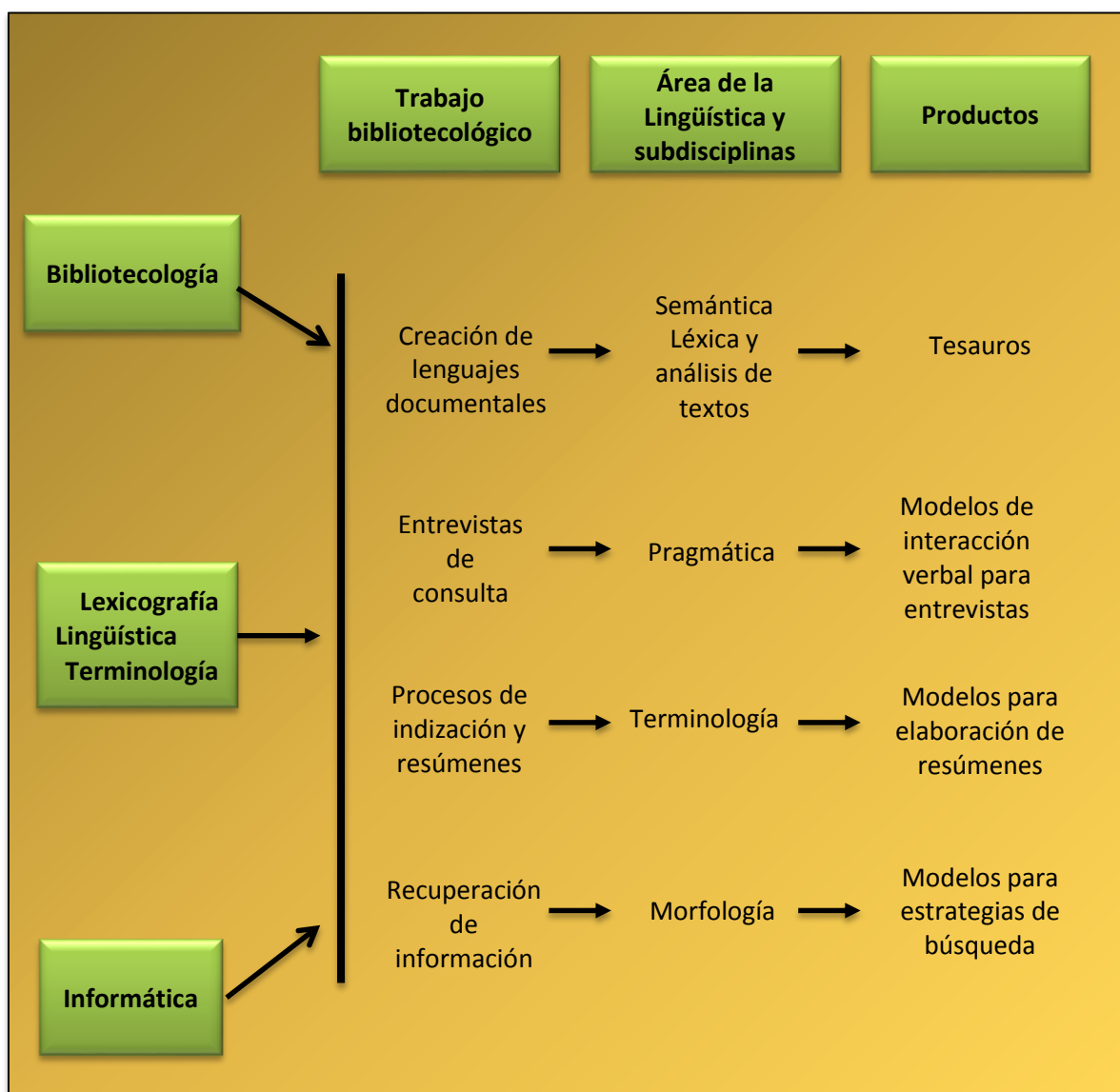


Figura 1. Puntos de contacto y convergencia. Elaboración, Valdez, R. J. (2013).

Se trata de un espacio constituido por cuatro bloques que en la Figura 1 se leen de izquierda a derecha. En el primero se ubican las tres disciplinas de las que parte este análisis. En el segundo se señalan las áreas del trabajo bibliotecológico en las que ocurren puntos de contacto y convergencia entre estas disciplinas, que son cuatro:

- Creación de lenguajes documentales a partir de la detección del perfil de necesidades de información de los usuarios
- Entrevistas de consulta,
- Procesos de indización y elaboración de resúmenes y
- Recuperación de información

En el tercer bloque se ubican las áreas de conocimiento de la Lingüística (Semántica Léxica y análisis de textos, Pragmática Terminología y Morfología,) que inciden en la creación de los productos relevantes para la creación de modelos metodológicos de trabajo que, eventualmente, pasarán a formar parte de las herramientas cotidianas del trabajo bibliotecario. Estos productos se mencionan en el cuarto bloque y son los siguientes: modelos para estrategias de búsqueda y recuperación de información, modelos para la elaboración de resúmenes, modelos de interacción verbal para el estudio de entrevistas de consulta y, de manera importante, los tesauros.

En el panorama que se describió en la figura 1 se ubica la elaboración de un tesauro y en las convergencias y puntos de contacto en diferentes áreas del conocimiento que marcan la complejidad de su naturaleza.

Ahora bien, el diseño y realización de un tesoro es un proceso de colaboración interdisciplinaria, hecho que supone la integración de un grupo de trabajo con integrantes de diferentes áreas disciplinares.

Las disciplinas que colaboran son diversas, aunque podemos señalar las siguientes al menos: la Lingüística, la Terminología, la Informática, la Teoría Comunicativa y la disciplina o área del conocimiento sobre la cual se está elaborando el tesoro.

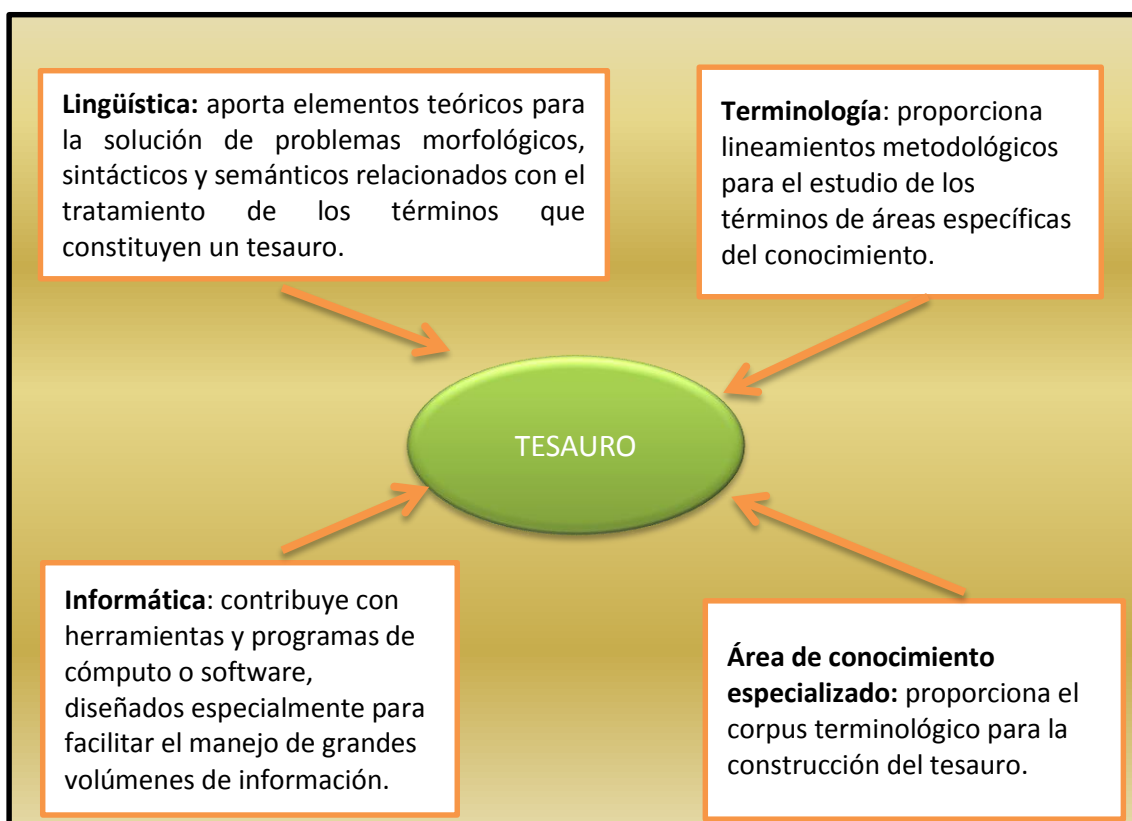


Figura 2. Colaboración interdisciplinaria en la elaboración de tesauros. Elaboración, Valdez, R. J. (2013).

En la Figura 2 se muestran las contribuciones que las disciplinas en cuestión hacen al tesoro. La Lingüística aporta elementos teóricos para la solución de

problemas morfológicos, sintácticos y semánticos relacionados con el tratamiento de los términos que constituyen un tesoro.

La Informática, por su parte, contribuye con herramientas y programas de cómputo o *software*, diseñados especialmente para facilitar el manejo de grandes volúmenes de información, además suministra metodologías de trabajo para la utilización de estructuras complejas en los diferentes tipos de relaciones que se establecen entre los términos de un tesoro.

La Terminología proporciona lineamientos metodológicos para el estudio de los términos de áreas específicas del conocimiento. De acuerdo con las necesidades específicas de un determinado grupo de usuarios para establecer canales y medios adecuados de comunicación entre especialistas. Asimismo la Teoría de la Comunicación, que en conjunto con la Teoría Comunicativa de la Terminología de Cabré, T (1999)⁴⁹ quien señala que “proporciona elementos para la descripción de diferentes tipos de situaciones en las que se da la comunicación en amplitud y diversidad”.

Y un área especializada de conocimiento, contribuye con un corpus terminológico específico para la construcción del tesoro.

⁴⁹ Cabré, Teresa (1999) Hacia una teoría comunicativa de la terminología: aspectos metodológicos. En: *La terminología representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos* Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabre p. 131 – 132.

La Lingüística, como disciplina, trata de dilucidar los principios genéricos del lenguaje humano a fin de construir una teoría que pueda explicarlo. Sin embargo, entendida de forma más amplia, la Lingüística coincide en sus objetivos con un número considerable de disciplinas, como la psicolingüística o la sociolingüística, que son más o menos próximas cuyos intereses son similares o superpuestos. De tal modo que cuando se piensa en la complejidad y versatilidad del lenguaje humano, y en cómo éste entronca con una doble vertiente: la de capacidad cognitiva y la del vehículo de comunicación, sucede entonces, como bien señala Payrató⁵⁰, “que la delimitación precisa de los objetivos teóricos de la lingüística se ve dificultada —particularmente cuando aumenta su grado de abstracción— por su integración inevitable con otros campos”.

Lo mismo acontece en otras áreas del saber, como la Lingüística aplicada que actualmente se define como un área de intersección con otras disciplinas, incluida la lingüística teórica y descriptiva, en la que se analiza y ofrecen soluciones a problemas en los que está presente directa o indirectamente el lenguaje humano. De esta manera, y siguiendo la consideración de Payrató,⁵¹ las demandas sociales y profesionales de la época, justifican la necesidad de profundizar e impulsar los estudios aplicados y, por consiguiente, los análisis en que converjan los aspectos teóricos y prácticos de las ciencias del lenguaje.

⁵⁰ Payrató, Lluís (1998) *De profesión, lingüista: panorama de la lingüística aplicada*. Barcelona: Ariel p. 46.

⁵¹ *Ibid.* p.16.

En este intercambio la Bibliotecología ha diseñado y elaborado instrumentos de apoyo para el trabajo diario en las bibliotecas, un ejemplo de la colaboración interdisciplinaria son los tesauros.

La interdisciplina ha sido estudiada de diferentes formas, Blázquez⁵² señala “que durante el siglo XX, fue definida, como metodología, como concepto, como proceso, como filosofía o como una ideología reflexiva”. Esta diversidad de aproximaciones y enfoques hace cada vez más difícil contar con una definición precisa del término. Sin embargo, para los propósitos de este trabajo, se considerará la interdisciplina como un proceso de colaboración y aporte de conocimientos de diferentes disciplinas para la resolución de problemas prácticos en el trabajo de un área específica del conocimiento bibliotecológico.

Según Follari,⁵³ cuando se habla de interdisciplina “se trata de un cambio de la interrelación orgánica de los conceptos de diferentes disciplinas hasta el punto de constituir una especie de nueva unidad que subsume en un nivel superior las aportaciones de cada una de las disciplinas particulares”.

Sobre este tema Morin, E (1997)⁵⁴ señala que la interdisciplinariedad es un término polisémico y etéreo que “puede también querer decir intercambio y cooperación, lo que hace que a interdisciplinariedad puede devenir en alguna

⁵² Blázquez Graf, N. et. al... (2000) “La interdisciplina en la historia de las ciencias” en *Historia a debate: Tomo III: problemas de historiografía, Actas del II Congreso Internacional “Historia a Debate”*. Vol. III. España: Santiago de Compostela. p.247

⁵³ Follari, Roberto (1982) *Interdisciplinariedad: los avatares de la ideología*. México: Universidad Autónoma Metropolitana–Azcapotzalco p. 27.

⁵⁴ Morin, Edgar (1997) “Sobre la interdisciplinariedad” en *Publicaciones ICESI (Instituto Colombiano de Estudios Superiores)* 62, enero – marzo p. 9 – 15.

cosa orgánica”. Asimismo Martínez García, S. (2009)⁵⁵ menciona que “la interdisciplinariedad se produce cuando se estudia un tema que se extiende sobre varias disciplinas sin formar un todo casi coherente”.

En las citas descritas anteriormente se pueden observar diferentes maneras para aproximarse a una noción de interdisciplinariedad, sin embargo en todas subyace la idea de cooperación, intercambio y colaboración.

El tema del proceso de colaboración entre la Bibliotecología, la Terminología y la Lingüística ha sido objeto de estudio en múltiples ocasiones, con enfoques diversos, con una larga trayectoria y en diferentes lenguas y países. Cabe mencionar que en el discurso de los documentos analizados para este apartado se utilizan diferentes términos para referirse a lo que no puede ser otra cosa que la Bibliotecología. Dichos términos son *documentación*, *Ciencia de la Información* y *Estudios de la Información*.

Considerando lo anterior, en la interrelación entre Lingüística y Bibliotecología, se pueden citar múltiples trabajos, por ejemplo el de Alicia Perales Ojeda titulado “La intersección lingüística en las ciencias de la información”,⁵⁶ otro más es la publicación ya clásica en el tema, de Karen Spark–Jones *Linguistics and Information Science*.⁵⁷ Asimismo Angelina Bosch de Roze publicó “Lingüística y

⁵⁵ Martínez García, S (2009) La representación y organización de la información a través de los tesauros. La interdisciplinariedad como nuevo paradigma: retos para la documentación y la bibliotecología, en *Contribuciones a las Ciencias Sociales*, octubre, p. 17. Consultado en mayo de 2013 en www.eumed.net/rev/cccss/06/smg.htm

⁵⁶ Perales Ojeda, A (1986). “La intersección lingüística en las ciencias de la información” en *Actas del II Congreso Internacional del Español de América..* México: UNAM, Facultad de Filosofía y Letras. p. 184–192.

⁵⁷ Spark – Jones, K. (1973). *Linguistics and Information Science*. New York: Academic Press, p. 3- 4.

Ciencias de la Información: un acercamiento integrador”.⁵⁸ En Brasil, el tema fue abordado por Ulf Gregor Baranow en su artículo “Perspectivas na contribuição da Lingüística e de áreas afins à Ciência da Informação”⁵⁹ y por Ercilia Severina Mendonça en “A lingüística e a ciência de informação: estudos de uma interseção”.⁶⁰ Recientemente, en 2004, Silvia Arano escribió “La ontología: una zona de interacción entre la Lingüística y la Documentación”.⁶¹

Igualmente notoria es la relación entre la Bibliotecología y la Terminología. Se pueden citar trabajos como los de Emilia Currás, quien en 1991, en su obra *Thesaurus Lenguajes Terminológicos*,⁶² escribe dos capítulos dedicados a la Terminología, su evolución histórica y su relación con los tesauros. Por su parte Teresa Cabré y Lluís Codina, en 2000, coordinaron el libro *Terminologia i Documentació*.⁶³ De igual forma Reiner Arntz y Heribert Picht publicaron en 1995, “Terminología y Documentación”.⁶⁴ Por otra parte, De Irazazábal escribió en 1996 “Terminología y documentación”.⁶⁵ Juan Sager, en 1993, escribió *La terminología y las ciencias de la información*.⁶⁶ Teresa Cabré en 1999 escribió

⁵⁸ Bosch de Roze, A. (1986) “Lingüística y Ciencias de la Información: un acercamiento integrador” en *Revista Española de Documentación Científica*. 9:2. p. 167- 171.

⁵⁹ Baranow, Ulf Gregor (1983) “Perspectivas na contribuição da Lingüística e de áreas afins à Ciência da Informação” en *Ciência da Informação*. Brasilia. 12:1. p. 23 – 35.

⁶⁰ Mendonça, Ercilia Severina (2000). “A lingüística e a ciência de informação: estudos de uma interseção” en *Ciência da Informação*. Brasilia. 29:3 p. 50–70.

⁶¹ Arano, Silvia (2004). “La ontología: una zona de interacción entre la Lingüística y la Documentación” [En línea] [Acceso en 12 de abril 2010] Disponible en <http://www.hipertext.net>

⁶² Currás, Emilia (1991). *Thesaurus Lenguajes Terminológicos*. Madrid: Paraninfo. p.284.

⁶³ Cabré Castellví, Teresa & Lluís Codina (2000). *Terminologia i Documentació. I Jornada de Terminologia i Documentació*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabre.

⁶⁴ Arntz, Reiner & Heribert Pitch (1995). “Terminología y Documentación” en *Introducción a la Terminología*. Madrid: Fundación Germán Sánchez Ruipérez, Pirámide. p. 299 –310.

⁶⁵ De Irazazábal, A. (1996). “Terminología y documentación” en *Jornada Panllatina de Terminologia. Perspectives i camps d'aplicació*. Barcelona: Institut Universitari de Lingüística Aplicada. p. 47-55.

⁶⁶ Sager, Juan (1993). *Curso práctico sobre el procesamiento de la terminología*. Madrid: Fundación Germán Sánchez Ruipérez, Pirámide. p.25–27.

un capítulo sobre el tema⁶⁷ y, en 2005, publicó “Terminología y Documentación”.⁶⁸ Asimismo en el año 2000⁶⁹, Rita Teemerman escribió una obra importante que marca el inicio del estudio de la Terminología con un enfoque diferente y novedoso, el enfoque sociocognitivo de la Terminología. Finalmente, Montero Martínez y Pamela Faber Benítez en 2008⁷⁰ reconocen a los documentalistas y profesionales de la información como uno de los usuarios de la Terminología.

2.2. Procesos lingüísticos para la elaboración de tesauros

De acuerdo con la definición que se presenta en el Diccionario del Español Usual en México⁷¹, un proceso “es el conjunto de cambios y modificaciones que se producen en las características o en la naturaleza de algún objeto o de algún fenómeno durante su evolución o desarrollo”. Si atendemos a esta definición se puede establecer una relación entre el objeto tesoro, su creación y desarrollo y los cambios que se producen durante el proceso de diseño del mismo. Los procesos lingüísticos que se describen y analizan, a continuación son de tres tipos:

⁶⁷ Cabré Castellví, Teresa (1999). “Terminología y Documentación” en *La Terminología: representación y comunicación elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra Institut Universitari de Lingüística Aplicada. p. 231–247.

⁶⁸ Cabré Castellví, Teresa (2005). “Terminología y Documentación” en *Grup IulaTerm. Introducció a la Terminología*. Barcelona: IULA. Universidad Pompeu Fabra [En línea] [Acceso el 16 de marzo de 2005] Disponible en <<http://www.iulaonline.org>>

⁶⁹ Teemerman, Rita (2000). *Towards new ways of terminology description: the sociocognitive approach*. Amsterdam: John Benjamins.

⁷⁰ Montero Martínez, S. & Pamela Faber Benítez (2008). *Terminología para traductores e intérpretes*. Granada: Tragacanto. p. 26.

⁷¹ Lara Ramos, Luis Fernando (dir.) (2002). *Diccionario del Español Usual de México*. México: El Colegio de México. p.729.

- Procesos morfológicos
- Procesos sintácticos
- Procesos léxico – semánticos

2.2.1. Procesos morfológicos

Los procesos morfológicos que ocurren en la creación de un tesoro se refieren a la normalización de los términos en función de principios morfológicos, la explicación de dichos principios resulta ostensiblemente necesaria para mostrar de forma clara y precisa cómo es que los fenómenos morfológicos, como la derivación y la formación de siglas intervienen en la elaboración de un tesoro.

La Morfología de acuerdo con la definición que presenta Pérez Saldanya (1998),⁷² es la disciplina lingüística que estudia la estructura interna de las palabras. En el estudio de la Morfología, normalmente se distinguen dos campos de trabajo, por un parte la Morfología flexiva y por otra la Morfología derivativa o léxica. La derivativa o léxica es la parte de Morfología que estudia la formación de las palabras, es decir el proceso de composición de nuevas palabras a partir de las existentes. En tanto la flexiva es la parte de la Morfología que estudia el proceso de la formación de palabras, que consiste en la unión de afijos a un tema o radical con el propósito de analizar y explicar las variaciones que pueden experimentar diferentes categorías gramaticales como el caso, el género, el tiempo, la persona, etc.

⁷² Pérez, Saldanya *et al.* (1998). *Diccionario de Lingüística*. España: Valencia, Colomar. p. 218-219.

Asimismo y de acuerdo con Cabral, Leonor (1973)⁷³ “la Morfología se divide en dos ramas la morfología flexional y la morfología lexical”, esta división en su opinión es apenas una cuestión de grado. Siguiendo a esta autora, la morfología flexional (o gramatical) estudia las relaciones entre las diferentes formas de una misma palabra, esto es, su paradigma flexional. Y la morfología lexical trata de la estructura de las palabras y de sus procesos de formación, de las relaciones entre formas de paradigmas diferentes, esto es de formas como paradigmas separados.

Los procesos morfológicos que intervienen en la elaboración de un tesoro son la derivación, la formación de siglas y la flexión, al menos. Enseguida se presentan las definiciones de estos tres términos para después ejemplificar como intervienen dichos procesos en la elaboración de un tesoro.

La derivación de acuerdo con Bauer ⁷⁴ “es el proceso de agregar afijos para cuatro propósitos al menos: a) para crear nuevos lexemas; b) para poder cambiar una parte de la palabra, cuya base cambió; c) para cambiar su significado regular y d) para poder ser plenamente productiva y al mismo tiempo no podrá ser generalizable”.

⁷³ Cabral, Leonor (1973). *Introdução à lingüística*. Porto Alegre: Globo, p 129.

⁷⁴ Bauer, L. (1998). *Introducing Linguistic Morphology*. Edinburg: U.K. Edinburg University Press. p 241.

Una sigla es de acuerdo con Pérez Saldanya ⁷⁵ “una abreviatura formada por la letra o letras iniciales o finales de palabras que conforman un enunciado o sintagma y que constituyen una nueva palabra”.

La flexión en el Diccionario básico de lingüística (2005)⁷⁶ se define como “el proceso mediante el cual se unen al lexema uno o varios sufijos denominados desinencias o gramemas cuyo significado puede ser de número, tiempo, modo o persona”

Los procesos morfológicos señalados intervienen en la elaboración de un tesoro y se pueden encontrar evidencias de la aplicación de dichos procesos en un tesoro impreso. A continuación se presentan ejemplos en los que se ilustra el uso de los procesos morfológicos de la derivación, la formación de siglas y la flexión en un tesoro. La fuente documental de donde proceden es el *Tesoro Latinoamericano en Ciencia Bibliotecológica y de la Información*.⁷⁷

2.2.1.1. Derivación y flexión

La derivación, según la define Manuel Seco (1999)⁷⁸ es “el procedimiento de formación de palabras que consiste en las modificaciones de un término

⁷⁵ Pérez, Saldanya. *Op. Cit* p. 296

⁷⁶ *Diccionario básico de lingüística* (2005) México: UNAM, Instituto de Investigaciones Filológicas. p.546

⁷⁷ Naumis Peña, C. *et. al.* (1999). *Tesoro latinoamericano en ciencia bibliotecológica y de la información*. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas. p. 309

⁷⁸ Seco, Manuel (1999). *Diccionario del Español Actual*. Madrid: Aguilar. p. 1459.

mediante la adición de afijos”. Los afijos, de acuerdo con el autor señalado, ⁷⁹ “son el elemento que añadido a una raíz modifica su sentido o su valor gramatical formando un lexema”.

Los afijos a su vez se dividen en prefijos y sufijos. Un prefijo es el elemento que se antepone a la raíz y un sufijo el elemento que se pospone a la raíz. En la Figura 3 se muestran los elementos que componen el fenómeno morfológico de la afijación

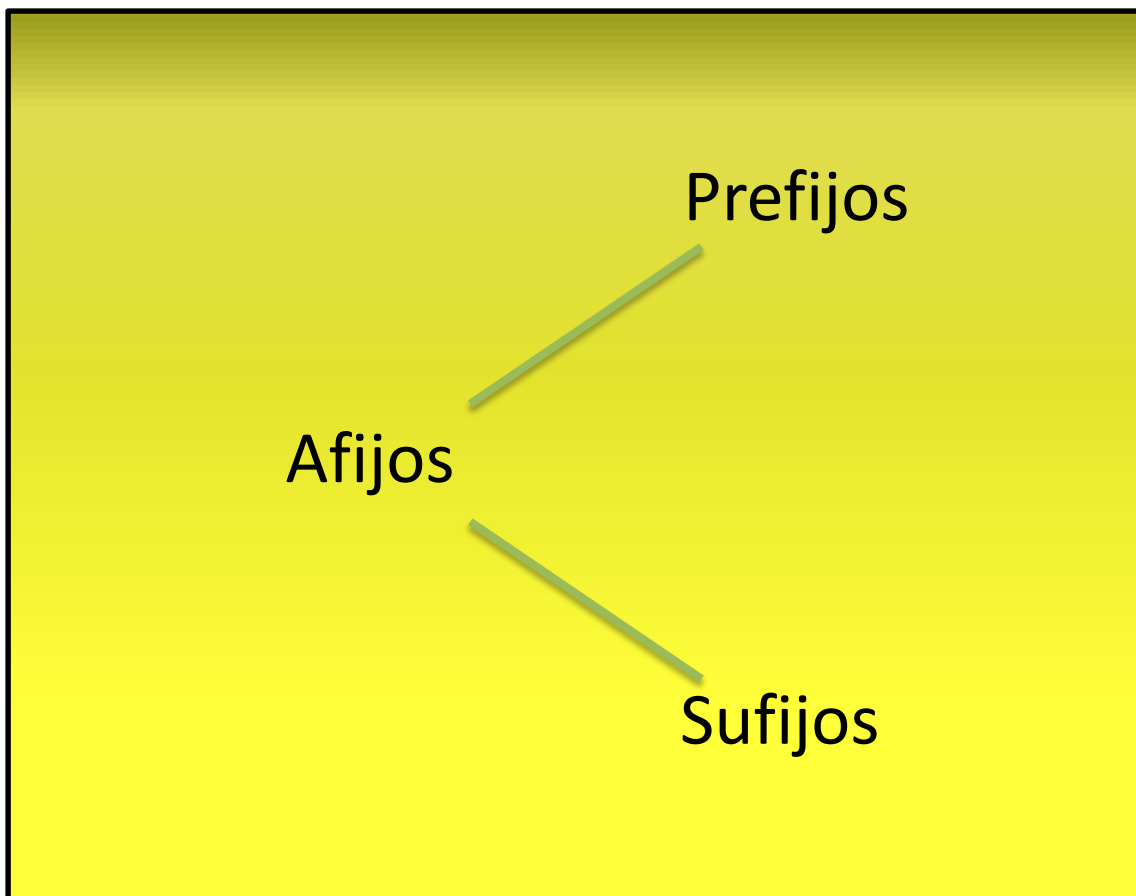


Figura 3. Aspectos morfológicos: afijación. Elaboración, Valdez R. J. (2013).

⁷⁹ Seco, Manuel. *Op. Cit.* p 125

En el tesoro analizado se identificó el fenómeno de la derivación, a través del cual se forman nuevos términos, por medio de la afijación. A continuación se muestran dos ejemplos tomados de las páginas 26 y 37 del tesoro.



Figura 4. Ejemplo de proceso morfológico de derivación: afijación. Elaboración, Valdez R. J. (2013).

En este ejemplo la derivación sucede en la formación de nuevas palabras o lexemas a partir de la base BIBLIO a la que se añade sufijos como BUS, FAGOS, FILIA, FILOS, o GRAFÍA. De dicho proceso derivativo se forman entonces BIBLIOBUS, BIBLIOFAGOS, BIBLIOFILIA, BIBLIOFILOS o BIBLIOGRAFÍA. En el caso de añadir prefijos a la raíz, como ocurre en el ejemplo se forma el lexema BIO/BIBLIOGRAFÍAS.

La flexión, como se señaló anteriormente es el proceso mediante el cual se forman palabras a través de la suma a la base o radical de uno o varios sufijos llamados desinencias y que pueden variar el significado atendiendo al número, tiempo, modo o persona.

En el tesoro analizado se identificó el fenómeno de la flexión, a través del cual se forman nuevos términos, por medio de la suma de uno o varios sufijos a la base o radical y que en el tesoro que se analizó, el ejemplo que se presenta se refiere a un proceso flexivo de número en un sustantivo. A continuación se muestran un ejemplo tomado de la página 147 del tesoro.

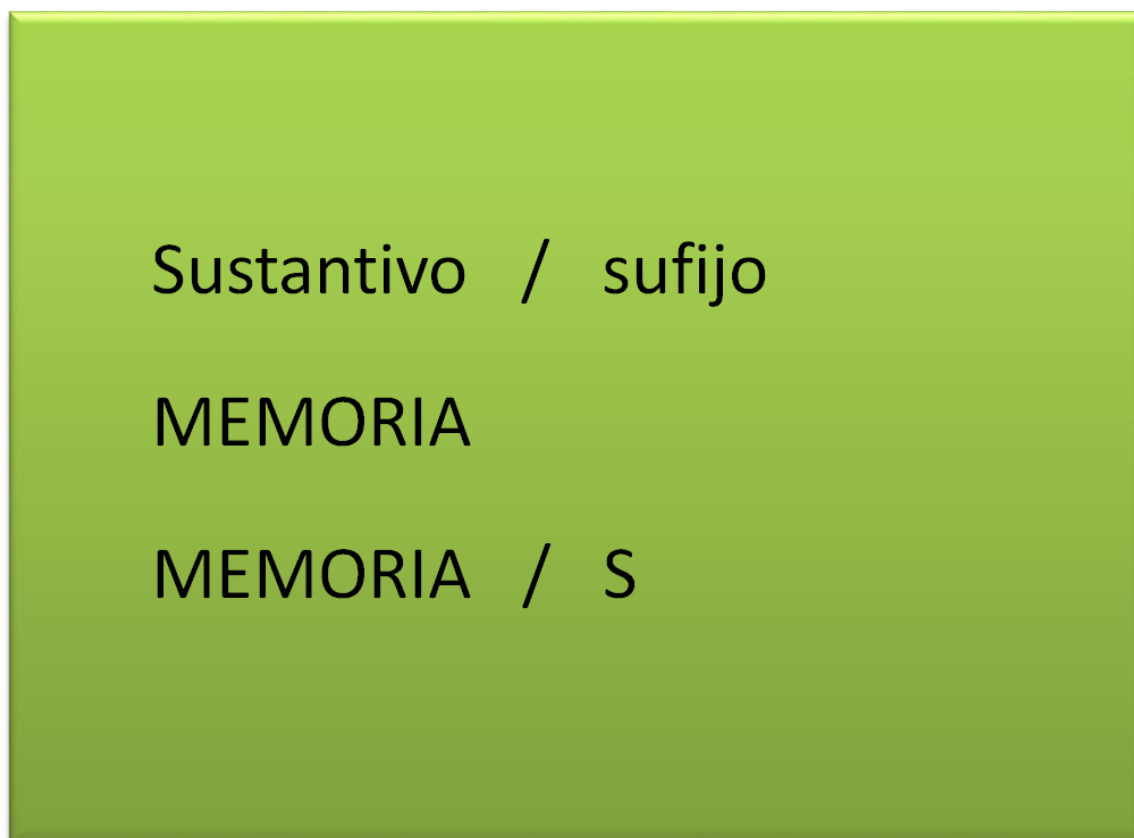


Figura 5. Ejemplo de proceso morfológico de flexión: sufijación. Elaboración, Valdez, R. J. (2013).

2.2.1.2. Siglas

Una sigla es de acuerdo con Moreno de Alba (1995),⁸⁰ “la que se forma con las letras iniciales de las voces fundamentales de alguna denominación”. En un tesaurus es común que las siglas formen parte de los términos que lo integran. Generalmente se escriben con letras mayúsculas, se separan con: (dos puntos), y en seguida se escribe el nombre completo del organismo, entidad gubernamental, asociación o compañía comercial que significa la sigla.

A continuación se presenta una muestra de las siglas que se encontraron en el tesaurus analizado.

Sigla 1

ASCII/ASCII

UP: American Standard Code for Information Interchange

TR: Computadores

Intercambio de información

Sigla 2

ISBN/ISBN

UP: International Standard Book Number

Número Internacional Normalizado para Libros.

TR: Control Bibliográfico Universal

ISBD (M)

Libros.

⁸⁰ Moreno de Alba, J. G (1995). *Minucias del lenguaje*. México: FCE. p. 451.

2.3. Procesos sintácticos

La Sintaxis de acuerdo con Dubois J⁸¹ “es la parte de la gramática que describe las reglas por las que las unidades significativas se combinan en oraciones”. En tanto que Mito, Carlos (2013)⁸² define a la sintaxis “como la disciplina de la lingüística que estudia como combinamos palabras para formar sintagmas y como combinamos sintagmas para formar oraciones. Esta concepción de la sintaxis se apoya en lo que se llama hipótesis lexicalista, es decir, una sintaxis que comienza a actuar donde termina la actuación de la morfología”

Un tesoro está compuesto por términos o unidades significativas, según Dubois, que combinados entre sí construyen oraciones. Esas oraciones están construidas con una serie de reglas sintácticas que dan solidez y consistencia a la herramienta, entre otras condiciones.

En la creación de lenguajes documentales Hutchins⁸³ identificó los siguientes cuatro aspectos sintácticos:

- la concordancia entre lexemas,
- los indicadores de las relaciones de casos,
- el uso de palabras funcionales, y
- el uso del orden de las palabras.

⁸¹ Dubois, Jean *et al.*(1998). *Diccionario de Lingüística*. Versión española de Inés Ortega y Antonio Domínguez. Madrid: Alianza. p. 575.

⁸² Mito, Carlos. (2013) *Novo manual de sintaxe*. Sao Pablo: Contexto, p. 32.

⁸³ Hutchins, W.J (1978). *Languages of indexing and classification: a linguistic study of structures and functions*. Londres: Peter Peregrinus. p. 25–26.

De los cuatro aspectos señalados, sólo el último está presente en la elaboración de los tesauros. Enseguida se explicará en qué consiste dicho *uso del orden de las palabras*, se mostrarán algunos ejemplos de su uso derivados del análisis realizado en el *Tesaurus Latinoamericano en Ciencia Bibliotecológica y de la Información*⁸⁴ y se argumentará sobre las ventajas que significa la aplicación de las reglas sintácticas en la elaboración de un tesaurus.

Desde la perspectiva del aspecto sintáctico *el uso del orden de las palabras*, en español el orden canónico de las palabras en un sintagma nominal es:

(S) *sujeto + modificador*

Este orden aparece en la lengua natural; sin embargo, en lenguajes controlados, como es el caso de los tesauros, el orden de las palabras adopta el siguiente esquema:

Sustantivo + Modificador

En un nivel de frase nominal el Sustantivo es el tema de la frase y el Modificador explica o modifica al sustantivo.

La aplicación de este aspecto adquiere mayor relevancia en la elaboración y desarrollo de un tesaurus, porque si las frases nominales que conforman sus entradas —entiéndase términos genéricos, términos específicos o términos

⁸⁴ Naumis Peña, C. *et. al. Op Cit.*, p. 309

relacionados—, atienden las reglas sintácticas de la lengua en la que está escrito, tendrá mayor consistencia interna y será una herramienta útil para la indización y para la recuperación de información por parte del usuario.

Ahora bien, a fin de identificar el uso del aspecto sintáctico señalado, enseguida se muestran algunos ejemplos tomados del *Tesaurus Latinoamericano en Ciencia Bibliotecológica y de la Información*.⁸⁵

En los Términos Específicos (TE) del tesaurus analizado se cumple el patrón sintáctico de:

(S)Sustantivo + (M)Modificador.

Ejemplos:

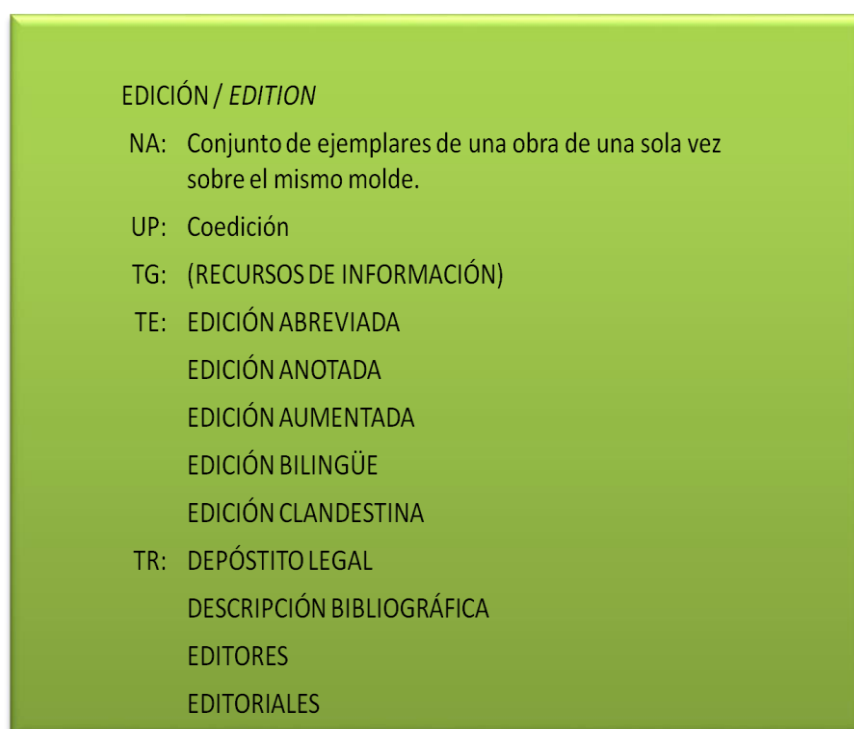


Figura 6. Ejemplo de proceso sintáctico *Sustantivo + Modificador*. Elaboración, Valdez, R. J. (2013).

⁸⁵ *Ibid*, p.89.

En este ejemplo el patrón sintáctico *S+ M* ocurre entre los términos específicos (TE) y de uno de los términos relacionados (TR) del tesoro analizado.

2.4. Procesos léxico - semánticos

El aspecto léxico - semántico es el componente lingüístico en el que es más evidente la relación de colaboración entre la Bibliotecología y la Lingüística, especialmente en la elaboración de un tesoro. El contacto entre estas dos disciplinas se presenta en los diferentes tipos de relaciones que se establecen entre los términos que conforman la herramienta; en este apartado se analizarán cuáles son las relaciones semánticas que ocurren entre los términos que conforman un tesoro y cómo es su dinámica. Asimismo se se analizarán y ejemplificarán las asociaciones, relaciones y configuraciones de tipo léxico.

La semántica de acuerdo con Werner Abraham⁸⁶ “es la ciencia que estudia las relaciones de las formas lingüísticas con los contenidos y significados y, además, también con el pensamiento, el mundo exterior y la conducta”. En otra acepción del término dicho autor señala que “es la ciencia que describe las relaciones entre el signo lingüístico con la imagen mental”.

⁸⁶ Werner Abraham, *et al.* (1981). *Diccionario de Terminología Lingüística Actual*. Madrid: Gredos. p. 404.

Las relaciones semánticas son, de acuerdo con Broughton (2006),⁸⁷ “otra forma de describir las relaciones jerárquicas o relaciones paradigmáticas” que se presentan en un tesoro. Ejemplos de este tipo de relaciones son las siguientes:

Tipos de relaciones	Ejemplos
todo - parte	cuerpo - brazo
proximidad	metal - oro
causa – efecto	agua - humedad
género – especie	mamífero/animal - caballo
sinonimia	hondo – profundo.

En los apartados que siguen se explicará en qué consisten las relaciones semánticas y qué tipos de relaciones se pueden encontrar en un tesoro, además se darán algunos ejemplos de uso de este proceso. Los ejemplos que se muestran se derivan del análisis realizado en el *Tesoro Latinoamericano en Ciencia Bibliotecológica y de la Información*⁸⁸. También se argumentará sobre las ventajas que tiene la aplicación del proceso semántico en la elaboración de un tesoro.

2.4.1. Relaciones de jerarquía

El fenómeno de la sinonimia, en el sentido de Theodor Lewandowski,⁸⁹ es decir, entendiéndolo como “una relación semántica de identidad de significado (en sentido estricto o amplio) o de similitud de significado / parentesco de sentido

⁸⁷ Broughton, Vanda. *Op. Cit* p. 222.

⁸⁸ Naumis Peña, C. *et. al. Op Cit*. P 309.

⁸⁹ Lewandoswki, Theodor (1982). *Diccionario de lingüística*. Madrid: Cátedra. p.321.

de formas distintas, es decir un sentido con varios nombres”, resulta ser una referencia indispensable para el desarrollo de un tesoro. Con base en lo anterior, se podría decir que la sinonimia es quizá la forma más evidente de relación semántica y que si se establecen relaciones entre los significados de los términos.

Una relación jerárquica según refiere Broughton (2006),⁹⁰ “es aquella que ocurre entre conceptos cercanos en una jerarquía, indicando si hay una relación o *status* de subordinación, super ordinación o coordinación. En un tesoro las relaciones de jerarquía se expresan por medio de los términos genéricos (TG) y términos específicos (TE)”. A continuación se presentan algunos ejemplos de este tipo de relaciones:

TG: ÍNDICES

TE: ÍNDICES ACUMULATIVOS

ÍNDICES ALFABÉTICOS

ÍNDICES ANALÍTICOS

ÍNDICES AUTOMATIZADOS⁹¹

En el ejemplo anterior se presenta una relación jerárquica entre los términos, es decir una relación que va de un término genérico hacia uno o varios términos específicos. La identificación, el conocimiento y uso de este tipo de relaciones, por parte de los indizadores, tiene diferentes ventajas, una de ellas ocurre

⁹⁰ Broughton, Vanda. *Op Cit.* p. 213

⁹¹ Naumis Peña, C. *et. al. Op Cit.* P. 120-121.

cuando se establecen enlaces sintácticos al indizar documentos y se señalan las diferencias jerárquicas entre los descriptores.

2.4.2. Relaciones de equivalencia

Broughton (2006)⁹² define las relaciones de equivalencia como “la relación que se establece entre los términos de un tesoro en la que los términos tienen más o menos el mismo significado y solamente uno se utiliza en la indización. La equivalencia existe entre los sinónimos y los cuasi sinónimos”. Véanse los siguientes ejemplos de este tipo de relaciones:

GRUPOS MINORITARIOS

USE: MINORÍAS ÉTNICAS
MINORÍAS RACIALES⁹³

CARACTER

UF: TEMPERAMENTO

CONFLICTO

UF: DISPUTA⁹⁴

AACR

USE: REGLAS DE CATALOGACIÓN ANGLOAMERICANAS

⁹² *Ibid* p.212

⁹³ Colby, Anita et al. (1992) *Thesaurus of Linguistic Indexing Terms*. San Diego, California. Sociological Abstracts, p. 53.

⁹⁴ Viet Jean (1976) *EUDISED Thesaurus Multilingüe para el tratamiento de la información relativa a la educación*. Paris, Consejo de Europa. Centro de Documentación para la Educación en Europa, p 131 y

ABSTRACTS

USE: RESÚMENES⁹⁵

En los ejemplos anteriores se hace explícita una relación de equivalencia entre, i) frases que puede utilizarse como equivalentes o cuasi sinónimos ii) una sigla en inglés y su equivalente traducido al español y iii) un término en inglés y su equivalente traducido al español. Asimismo se puede observar que durante el proceso de elaboración del tesoro el diseñador de la herramienta toma una decisión sobre el cuál o cuáles son los términos que se recomiendan para la indización o para la búsqueda de información.

2.4.3. Relaciones de asociación

Una relación asociativa es aquella que se expresa, de acuerdo con Broughton (2006)⁹⁶ “cuando para un término dado, existe otro que está relacionado con el anterior de manera no jerárquica”. Obsérvense los siguientes ejemplos:

METADATOS / *METADATA*

TR: DATOS

MERCADO LABORAL / *LABOR MARKET*

TR: FORMACIÓN PROFESIONAL

PERFIL PROFESIONAL⁹⁷

⁹⁵ Broughton, Vanda. *Op Cit.* p. 1.

⁹⁶ *Ibid.* p. 221.

⁹⁷ *Ibid.* p. 148.

En el ejemplo anterior se hace explícita una relación de asociación entre términos a los que se asocia o asocian de forma no jerárquica otro (s) términos y en este caso se presenta la traducción al español del término asociado. El asociar términos de esta forma no jerárquica, facilita al usuario la búsqueda y recuperación de documentos al ampliar y diversificar el espectro terminológico disponible.

2.5. Procesos léxicos: Asociaciones, configuraciones y relaciones léxicas en un tesoro

De acuerdo con el *Diccionario básico de lingüística*⁹⁸ la Lexicología es “la disciplina que se ocupa del estudio, análisis y descripción de las palabras y de las unidades significativas que las componen, y de las relaciones que existen entre ellas”. Es precisamente en estas últimas en las que se sustenta la estructura, solidez y consistencia de un tesoro, por lo tanto es importante explicar y analizar las relaciones lexicales que ocurren cuando se elabora uno, ya que un tesoro está integrado por términos o palabras entre los cuales se establecen relaciones lexicales de diversos tipos.

Como se señaló anteriormente la Lexicología es la disciplina que se ocupa del estudio de las relaciones entre las palabras y, de acuerdo con Luis F. Lara

⁹⁸ Luna Trail, *et al.*(2005). *Diccionario básico de lingüística*. México: Universidad Nacional Autónoma de México. p 834.

(2006),⁹⁹ para el estudio de estas relaciones existe el análisis semántico cualitativo del léxico del cual se ocupa la semántica léxica.

Para continuar con el estudio de los y procesos que intervienen en la elaboración de un tesoro, a continuación se analizarán y ejemplificarán las asociaciones, relaciones y configuraciones de tipo léxico. En el siguiente cuadro se observan los diferentes tipos de asociaciones, configuraciones y relaciones que ocurren en un tesoro.

Asociaciones	Configuraciones	Relaciones
Heteronímicas	Es parte de (a kind of)	Meronímicas Partitivas
	Es un tipo de (a type of)	Género - especie

Cuadro 2. Procesos léxicos en un tesoro. Elaboración, Valdez, R. J. (2013).

Resulta pertinente ahora definir las asociaciones léxicas, como punto de partida para el análisis de cada una de ellas, esto es, saber en qué consisten, qué elementos las caracterizan y distinguen, para enseguida presentar ejemplos. Una asociación léxico-semántica es la relación que se da entre términos por su significado, el ejemplo típico de este tipo de asociaciones son los sinónimos o cuasi sinónimos, que fueron estudiados en el apartado 2.4. de esta tesis.

⁹⁹ Lara, Luis Fernando (2006). *Curso de Lexicología*. México: El Colegio de México. p 181.

La heteronimia, de acuerdo con Moreno Martínez (2005),¹⁰⁰ es el “fenómeno por el cual dos palabras que corresponden a dos términos gramaticales en oposición proceden de raíces diferentes: *nuera/yerno; toro/vaca; caballo/yegua*”. Esta misma autora¹⁰¹ define la hiponimia, “como la relación de inclusión semántica entre unos elementos léxicos de significación reducida (los hipónimos) y otro de amplia significación (hiperónimos): *gorrión, águila, codorniz*, etc., quedan incluidos bajo el hiperónimo *ave*”.

2.5.1. Noción de relación en la lingüística y los tesauros

La noción de relación en la lingüística moderna ha sido objeto de estudio profundo y diverso, por ejemplo, se ha investigado sobre las relaciones entre los signos y los objetos reales, la relación entre el significado y el significante del signo lingüístico o las relaciones entre los hechos gramaticales y los hechos fonológicos.

Para el caso de los tesauros y su vínculo con la Lingüística, conviene destacar dos tipos de relaciones: las sintagmáticas y las paradigmáticas. Estas relaciones han sido estudiadas ampliamente desde la corriente estructuralista. Dubois *et al* (1979)¹⁰² en su *Diccionario de Lingüística* señala que “según Saussure, (habla) en el discurso, las palabras contraen entre sí relaciones basadas en el carácter lineal de la cadena hablada; el sintagma es una combinación de unidades entre las que existen estas relaciones”. De manera más simple se puede decir que las

¹⁰⁰ Moreno Martínez, M. (2005). *Diccionario lingüístico-literario*. Madrid: Castalia. p. 175.

¹⁰¹ *Ibid.* p. 176

¹⁰² Dubois, Jean *et al.* (1979). *Diccionario de Lingüística*. Madrid: Alianza. p. 531.

relaciones sintagmáticas tienen lugar en un eje horizontal imaginario y que en un tesoro este tipo de relaciones sólo están presentes en los descriptores de dos o más palabras como:

TR: VOCABULARIOS CONTROLADOS

TG: ESTACIONES DE TRABAJO

TG: ESTADÍSTICAS BIBLIOTECARIAS

En tanto que el propio Dubois (1979)¹⁰³ menciona que “una relación paradigmática opone términos cuyo valor depende de esta oposición y de los cuales sólo uno se realizará en el enunciado producido; por ejemplo, *enseñanza* está en relación paradigmática en la lengua, con *educación*, *aprendizaje*, por una parte, y con *enseñar*, *enseñamos*, etc., por otra”. Siguiendo la imagen de los ejes imaginarios, la relación paradigmática se daría en un eje vertical. En un tesoro se presentan como relaciones género, especie y partitivas, por ejemplo las siguientes:

TG: BIBLIOTECAS

TE: BIBLIOTECAS ACADÉMICAS

TE: BIBLIOTECAS CENTRALES

TE: BIBLIOTECAS PÚBLICAS

TE: BIBLIOTECAS UNIVERSITARIAS

2.6. Configuraciones léxicas

Pasemos ahora al tema de las configuraciones léxicas. El estudio de las configuraciones léxicas es un campo de interés de la Semántica Léxica, ésta se

¹⁰³ *Ibid.* p. 531-532.

puede definir de acuerdo con Pustejovsky (2006)¹⁰⁴ como “el estudio de lo que significan las palabras y de cómo se estructuran estos significados”. A partir de esta definición se puede establecer un nexo claro e inmediato entre el tesoro y la Semántica Léxica, porque un tesoro, de acuerdo con Uri Miller (1997),¹⁰⁵ es “un modelo léxico–semántico de una realidad conceptual o sus constituyentes, la que se expresa como un sistema de términos y sus relaciones. Ofrece acceso a través de múltiples vías y aspectos y se utiliza como una herramienta de búsqueda en una unidad de recuperación de información es decir el significado y su estructura en un nivel léxico”. A partir de esta definición se plantean las siguientes preguntas:

- ¿Cuáles son las configuraciones léxico–semánticas que están presentes en el modelo que señala Miller, o sea, en un tesoro?
- ¿Qué características tienen estas configuraciones?
- ¿Cómo se describen y se presentan en un tesoro?

En el modelo que plantea Miller se identifican diferentes tipos de configuraciones léxico-semánticas. En un primer grupo están las jerárquicas, en el segundo las partitivas y en el tercero las meronímicas. Enseguida se explicará en qué consiste cada una de ellas.

¹⁰⁴ Pustejovsky, J (2006). “Lexical semantics: Overview” en Keith Brown (ed.) *Encyclopedia of Language & Linguistics*. 2a ed. Amsterdam: Elsevier. p. 98.

¹⁰⁵ Miller, Uri (1997). “Thesaurus Construction: problems and their roots” en *Information Processing & Management*. 33:4. p. 481 – 493.

2.7. Configuraciones jerárquicas

De acuerdo con Cruse (1989)¹⁰⁶ “una jerarquía es un conjunto de elementos relacionados entre sí de un manera característica”. Cabría preguntarse, ¿cuál es esta forma característica? Cruse reconoce dos: una ramificada o arbórea (I) y otra no ramificada (II), que gráficamente puede expresarse de la siguiente forma:

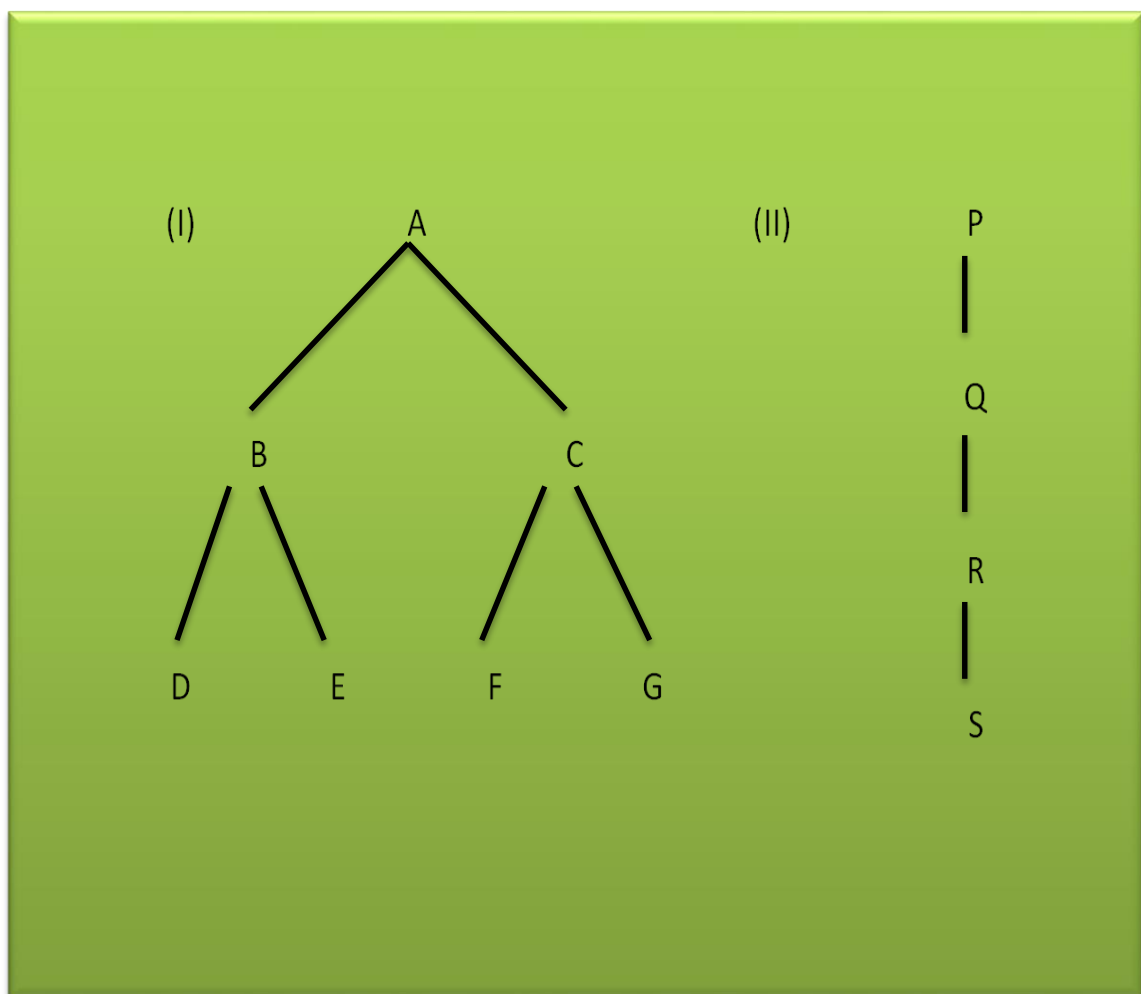


Figura 7. Tipos de configuraciones jerárquicas. Elaboración, Valdez, R. J. (2013).

¹⁰⁶ Cruse, D. A (1979). *Lexical Semantics*. Cambridge: CUP. p. 112.

En una configuración jerárquica siempre está presente una *relación de dominancia*, esto es, una relación vertical como en (II) y (I), es decir, A se conecta con B y C, B con D y E, y C con F y G; y en (II) P con Q, Q con R y R con S. En una jerarquía bien formada la relación de dominancia es una constante a través de la estructura. Por otro lado, en una estructura arbórea se requiere además de una *relación de diferencia*, esto es, la relación horizontal que ocurre por ejemplo entre B y C, D y E y en F y G en (I). La relación de diferencia también debe ser constante a través de una jerarquía bien construida.

Ahora veamos cómo se presenta en un tesoro este tipo de configuraciones jerárquicas.

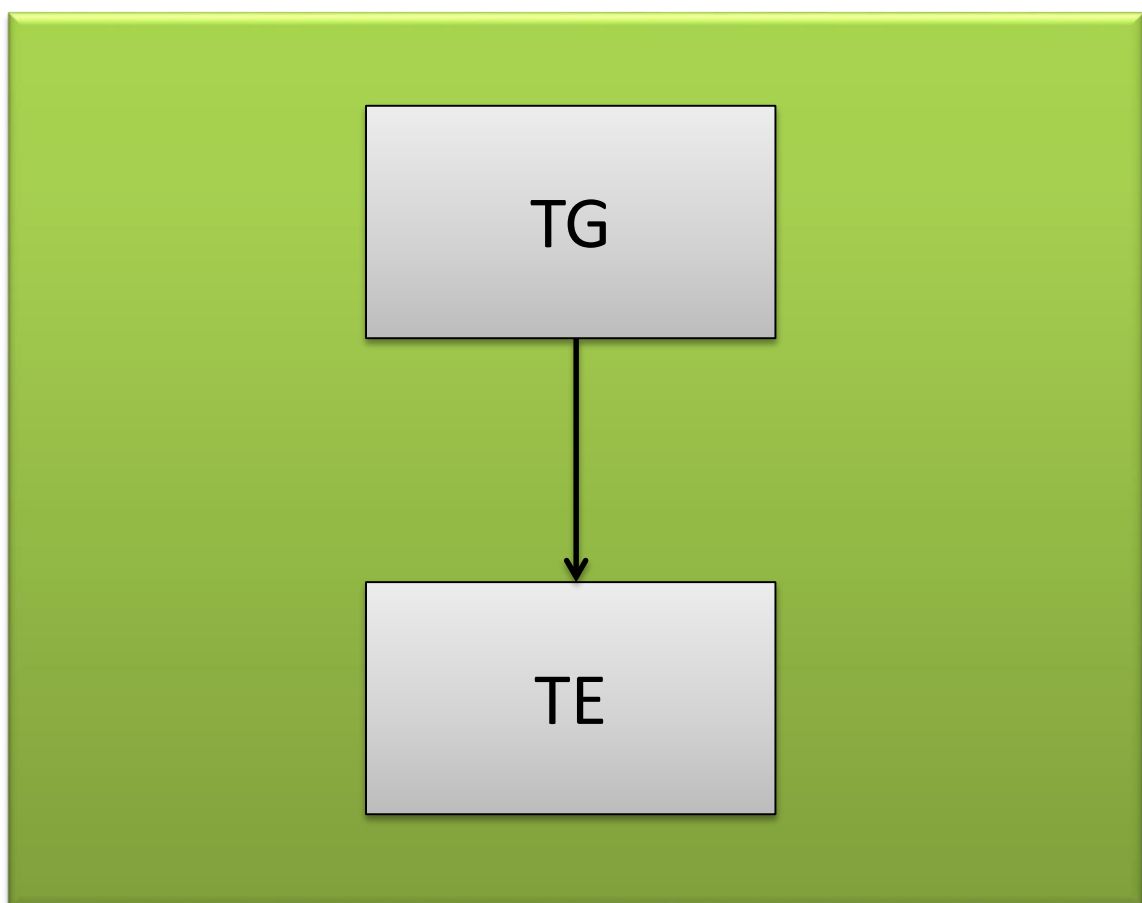


Figura 8. Configuraciones jerárquicas en un tesoro. Elaboración, Valdez, R. J. (2013).

Las configuraciones jerárquicas se pueden describir a través de una fórmula algebraica como la que sigue:

$$A > B$$

En los tesauros se identifican dos tipos de configuraciones jerárquicas: las relaciones todo–parte y las relaciones género-especie.

Una relación partitiva es de acuerdo con Lamarca (2008)¹⁰⁷ aquella en la que “el nombre de un parte implica y evoca la del todo poseedor. Se aplica a objetos, lugares, disciplinas y sus partes de estudio, estructuras sociales, entre otras”. Este tipo de relación jerárquica se puede expresar de la siguiente manera:

X es parte de Y

Y gráficamente se representa:



Figura 9. Relación partitiva. Elaboración, Valdez, R. J. (2013).

¹⁰⁷ Lamarca Lapuente, María de Jesús (2008). *Hipertexto: el nuevo concepto de documento en la cultura de la imagen*. [Consultado el 25 de mayo 2008]. Disponible en: <http://www.hipertexto.info/documentos/tesauros.htm>

Una configuración de este tipo se llama meronímica y se usa para la clasificación de cosas vivientes, el ejemplo clásico que se utiliza es sobre el cuerpo humano y puede expresarse en el siguiente esquema:

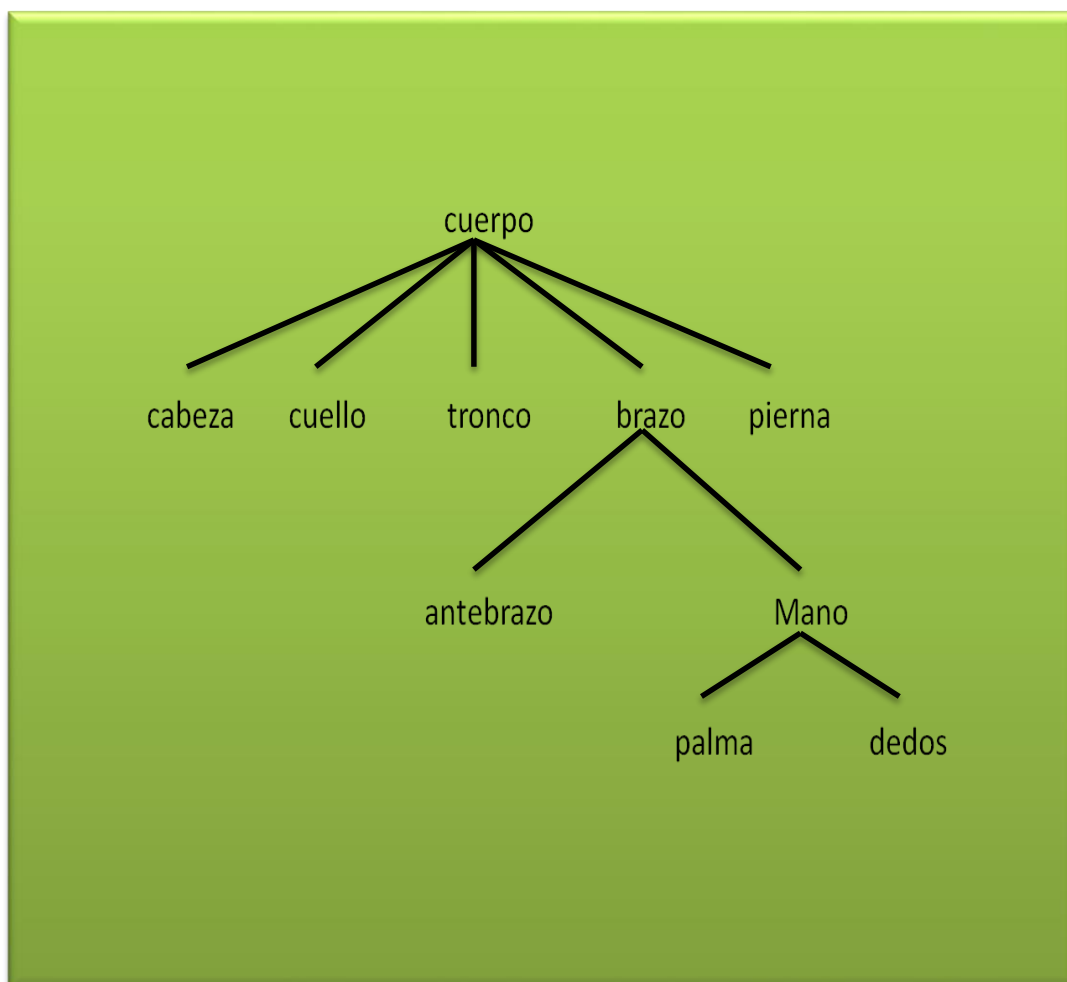


Figura 10. Configuración meronímica. Elaboración, Valdez, R. J. (2013).

Otro ejemplo de este tipo de relaciones es el de las disciplinas y sus partes de estudio, como se muestra a continuación:

TG: EDUCACIÓN

TE: EDUCACIÓN A DISTANCIA

EDUCACIÓN CONTINUA

EDUCACIÓN DE ADULTOS

EDUCACIÓN DE USUARIOS

Según Lamarca (2008)¹⁰⁸ una relación género–especie es aquella que “se aplica a acciones, propiedades y agentes”. Un ejemplo de este tipo de relaciones, que corresponden a acciones, es el siguiente:

TG: CONTROL

TE: CONTROL BIBLIOGRÁFICO

CONTROL DE ADQUISICIONES

CONTROL DE AUTORIDAD

CONTROL DE EXISTENCIAS

CONTROL DE PRÉSTAMO

Otro tipo de asociación léxica es la hiperonimia, al respecto, Luis Fernando Lara (2006)¹⁰⁹ apunta: “un hiperónimo es todo vocablo cuyo significado, por ser más general, puede sustituir parte del significado de otros vocablos, ayudando a su comprensión, aunque perdiendo precisión, por ejemplo vocablos como *cosa* y *objeto*, que puede servir como hiperónimos, sin que sean sinónimos”. Este tipo de asociación léxica se puede expresar con una fórmula como la siguiente:

X es de un tipo de Y

¹⁰⁸ *Ibid.*

¹⁰⁹ Lara, Luis Fernando (2006). *Curso de Lexicología*. México: El Colegio de México. p.202.

En un tesoro este tipo de asociación se presenta de la siguiente forma:

TG: DIAGRAMAS

TE: DIAGRAMAS DE FLUJO

TE: DIAGRAMAS DE OPERACIONES

TE: DIAGRAMAS DE VENN

2.8. La Terminología: tres enfoques teóricos y puntos de convergencia alrededor del tesoro

En este apartado se presentarán y analizarán tres enfoques teóricos de la Terminología: la teoría general, la teoría comunicativa y la teoría sociocognitiva. Las tres teorías mencionadas han marcado, de manera general, avances importantes en el estudio de la Terminología y su relación con la Bibliotecología y, de forma específica, en su relación con la elaboración y uso de los tesoros. Para iniciar esta sección se ofrece una definición de Terminología, posteriormente, se analizarán las tres teorías señaladas y su relación con los tesoros, para concluir se presentará una reflexión breve sobre la indización y la recuperación de información y cómo esto se articula con la Terminología.

2.8.1. Definición de Terminología

¿Qué es la Terminología? De acuerdo con Juan Sager (1993)¹¹⁰ la Terminología es “el estudio y el campo de actividad relacionado con la recopilación, la

¹¹⁰ Sager, Juan. *Op Cit.* p 21.

descripción y la presentación de términos, es decir, los elementos léxicos que pertenecen a áreas especializadas de uso en una o más lenguas”. Si se toma en cuenta esta definición, se puede identificar un primer punto de encuentro o acercamiento entre la Bibliotecología y la Terminología, es decir, en el estudio de los términos de una especialidad. A partir del estudio profundo de los términos de una especialidad, es que se lleva a cabo el diseño y elaboración de cierto tipo de herramientas lingüísticas como los glosarios, vocabularios controlados, tesauros y ontologías, dichas obras son herramientas útiles para el trabajo bibliotecario en el análisis bibliográfico, en la clasificación, en la indización y en la recuperación de la información.

Por su parte Alsina, V ; Estopa, R (1997)¹¹¹ define a la Terminología “como una materia interdisciplinaria que, en su vertiente teórica, toma conceptos de otras teorías para elaborar la suya propia, y, en su vertiente aplicada, es integrada en otras disciplinas a la vez que se sirve de ellas” Asimismo señala que en la Terminología “no todas las disciplinas relacionadas (con la Terminología) utilizan las unidades terminológicas del mismo modo y con los mismos fines y que por ello se conciben tres funciones distintas de la Terminología: La función de representación, la función de transferencia y la función de investigación”

De las tres funciones que mencionan las autoras la función de representación es en la que se puede observar una relación estrecha entre la Bibliotecología y la Terminología. Dicha relación se hace evidente, de manera general y de acuerdo

¹¹¹ Alsina, V ; Estopa, R (1997) “Las profesiones y los usuarios de la terminología” en *Terminómetro : la Terminología en España*. Número especial N° 2, p. 85.

con Arntz, R.; Pitch, H. (1995)¹¹² “cuando se utilizan las unidades terminológicas para elaborar, ordenar y almacenar el conocimiento y ponerlo al alcance de los usuarios” y de forma particular “en la aplicación de principios terminológicos en la construcción de sistemas de clasificación y tesauros”.

Para profundizar más sobre la relación que se da entre la Bibliotecología y la Terminología, es pertinente analizar al menos tres de las posturas teóricas que ha tenido la Terminología. Asimismo y derivado de dicho análisis, se señalarán algunos puntos de contacto entre las posturas teóricas y la elaboración de los tesauros.

Después de la investigación bibliográfica realizada, se puede afirmar que existen, al menos, tres posturas y concepciones diferentes en el estudio de la Terminología, aunque éstas, desde mi punto de vista, son complementarias y tienen cierta continuidad ya que ambas estudian las diferentes formas que existen para el tratamiento de los términos. Dichas formas enriquecen, retroalimentan y promueven el avance y crecimiento respecto a lineamientos y procedimientos metodológicos de la disciplina.

Enseguida se presentan algunos rasgos básicos de las tres corrientes teóricas más importantes que existen en terminología, es decir, la teoría general de la terminología (TGT), la teoría comunicativa de la terminología (TCT) y la teoría sociocognitiva de la terminología (TST)

¹¹² Arntz, R ; Pitch, H. (1995) *Introducción a la terminología*. Madrid, Fundación Germán Sánchez Ruipérez, Pirámide p. 22-26.

2.8.2. Teoría General de la Terminología, aspectos básicos

La Teoría General de la Terminología (TGT) de Eugene Wüster tiene, según María Pozzi (2005),¹¹³ los siguientes rasgos que la caracterizan:

- El objeto de estudio de la terminología son los conceptos de un campo del conocimiento y su denominación mediante términos.
- Los términos se analizan a partir del concepto que representan y por lo tanto se entiende que el concepto precede al término.
- Los conceptos de un campo del conocimiento están relacionados, y el conjunto de conceptos interrelacionados constituye el sistema de conceptos de ese campo del conocimiento.
- Los sistemas de conceptos se construyen de manera jerárquica, sin tomar en cuenta otro tipo de relaciones entre los conceptos.
- Cada concepto, y por lo tanto cada término que lo denomina, tiene un lugar específico en el sistema de conceptos. Todos los campos del conocimiento se tratan de la misma manera sin considerar las diferencias propias de cada materia ni las diferencias sociales, geográficas, culturales, económicas, etc. Los términos son “etiquetas” de los conceptos que denominan, por lo tanto se analizan fuera de contexto y su sintaxis no es relevante para la terminología.
- La evolución de los conceptos no interesa a la terminología ya que se consideran estáticos.
- No se considera la variación semántica.

¹¹³ Pozzi, María (2005). “La teoría comunicativa de la terminología” en *II Coloquio inter-universitario de terminología*. [en prensa]. México: Veracruz.

De acuerdo con Pozzi y Cabré, la TGT tiene limitaciones y deficiencias, a continuación se refieren tres de ellas que guardan relación con los tesauros:

La primera se puede enunciar de la siguiente forma: cada término pertenece a un campo del conocimiento y cada campo del conocimiento tiene sus propios términos que no comparte con otros campos del conocimiento. Esta condición permite que un concepto no pueda ser visto desde diferentes disciplinas.

La segunda se refiere a que, según la TGT, los sistemas de conceptos se construyen de manera jerárquica, sin tomar en cuenta otro tipo de relaciones entre los conceptos, esta característica restringe la relación entre términos a una sola, cuando en un tesoro están presentes relaciones de asociación y equivalencia, además de la relación jerárquica.

La tercera limitación se relaciona con el hecho de que la TGT defiende la univocidad y la monosemia como característica intrínseca de los términos, es decir, que para cada concepto existe un término preciso y deja de lado fenómenos semánticos como la polisemia, homonimia y sinonimia. Cabe mencionar que dichos fenómenos forman parte de las relaciones que están presentes entre los términos en un tesoro.

2.8.3. La Teoría Comunicativa de la Terminología, aspectos básicos

La Teoría Comunicativa de la Terminología (TCT) propuesta por Teresa Cabré del Instituto Universitario de Lingüística Aplicada de la Universidad “Pompeu

Fabra” de Barcelona, tiene de acuerdo con Pozzi (2005),¹¹⁴ los siguientes rasgos que la caracterizan:

- A la terminología se le concibe como una materia interdisciplinaria y se explica dentro de una teoría del lenguaje, que a su vez, se inserta en una teoría de la comunicación y del conocimiento.
- La terminología es multifuncional. La TCT debe dar cuenta de las unidades terminológicas en relación con la diversidad de visiones que de ellas tienen los especialistas. Cada unidad terminológica puede ser analizada desde la perspectiva lingüística, comunicativa y cognitiva.
- El objeto de estudio de la terminología es el término, considerado éste como: unidad de conocimiento, unidad de comunicación social y como unidad de lenguaje.

La condición interdisciplinaria de la TCT es un rasgo que permite el enriquecimiento de la teoría debido al intercambio recíproco que se establece entre las diferentes disciplinas. En el caso de la elaboración de los tesauros, la TCT aporta una serie de lineamientos metodológicos que permiten el tratamiento sistemático de los términos de un lenguaje de especialidad.

La condición multifuncional de la TCT y el tratamiento de los términos desde diferentes puntos de vista por parte de los especialistas permite que un tesoro sea un producto más acabado y enriquecido.

¹¹⁴ Pozzi, M. *Op. Cit.*

La condición diversa en el estudio del término que plantea la TCT, es decir, como unidad de conocimiento, como unidad de comunicación social y como unidad de lenguaje, promueve una visión amplia en el tratamiento de los términos y aporta pautas metodológicas con diversos puntos de vista para la elaboración de los tesauros.

En los párrafos siguientes se analizarán los términos como unidades del lenguaje y como unidades de comunicación social, a fin de identificar y señalar cuáles son los puntos de contacto entre la Terminología y la Bibliotecología.

Como unidades de lenguaje. Desde el punto de vista de la TCT los términos forman parte del lenguaje natural. Los términos que se seleccionan como candidatos para que formen parte de la estructura de un tesoro surgen del lenguaje natural contenido en las fuentes, aquí un primer punto de contacto entre la terminología y la elaboración de los tesauros.

Otra característica que tienen los términos según la TCT, en su momento histórico más avanzado es que pueden ser polisémicos, es decir, pueden tener diferentes significados según el contexto de la especialidad en la que se usan, esta característica es un aspecto muy importante que se toma en cuenta en el momento de establecer las relaciones entre los términos en un tesoro.

De acuerdo con la TCT, los términos tienen propiedades morfológicas, sintácticas y semánticas, estas características y su relación con la elaboración

de los tesauros han sido estudiadas con mayor profundidad en los apartados 2.1, 2.2 y 2.3 de esta tesis.

Como unidades de comunicación. Desde el punto de vista de la TCT los términos están determinados por las situaciones en que se usan. La comunicación especializada, esto es, aquella que se da entre especialistas de una disciplina, es una de esas situaciones. De acuerdo con Cabré (1999)¹¹⁵ la comunicación especializada se da a través de dos formas al menos, la comunicación directa y la comunicación indirecta. La primera se refiere a la comunicación entre expertos, al discurso didáctico o el de divulgación científica realizado por expertos; mientras que la comunicación indirecta se realiza por la vía de un proceso mediatizado como la traducción o el periodismo especializado. Considero importante agregar a esta lista el uso de un tesoro, ya que, desde mi punto de vista, constituye una situación de comunicación especializada en la medida que éste —el tesoro— sirve de puente entre las necesidades de información de una comunidad de usuarios y la colección documental representada por los términos descriptores y no descriptores del tesoro.

2.8.4. La Teoría Sociocognitiva de la Terminología (TST), aspectos básicos

Un antecedente importante en el tema de la TST y su relación con la Bibliotecología es el que señala Naumis, Peña, C¹¹⁶ quien dice que para las tareas de acceso temático a los contenidos documentales en las que interviene

¹¹⁵ Cabré Castellví, Teresa. (1999) "Elementos ... *Op. Cit.* p. 69-92

¹¹⁶ Naumis Peña, C. (2012) *Acceso temático a los contenidos de las colecciones de bibliotecas de la UNAM: Historias y perspectivas*. Documento no publicado. (Comunicación personal con la autora) p. 15.

la Bibliotecología, ésta disciplina “en el año de 1977 comienza a reconocer en el socio – cognitivismo como un nuevo método de trabajo” en la elaboración y diseño de herramientas para la indización y la recuperación de información, como los tesauros.

Estudios y experiencias sobre el inicio y desarrollo de este método de trabajo que ha permeado en el trabajo terminológico y bibliotecológico se encuentran en Hjørland, B. (2002)¹¹⁷ quien señala las ventajas que tiene el utilizar el enfoque socio – cognitivo en la recuperación de la información al reducir la distancia semántica entre los documentos y buscadores de información. A continuación se mencionan algunos rasgos característicos de la TST.

La Teoría Sociocognitiva de la Terminología (TST) surge a fines del siglo XX y al inicio del XXI, y su principal exponente es Rita Temmerman. De acuerdo con Durán Muñoz, Isabel (2012)¹¹⁸ dicha teoría “surge como una alternativa a los principios de la TGT, los cuales son considerados irreales por concentrarse solo en la estandarización y por obviar la influencia comunicativa y sociocognitiva en el trabajo terminológico”. Durán Muñoz, Isabel (2012)¹¹⁹ señala que Rita Temmerman “establece su teoría a partir de un estudio empírico sobre los procesos de categorización y lexicalización de un corpus de publicaciones y propone una metodología inspiradas en la herméutica y en la semántica cognitiva”.

¹¹⁷ Hjørland, B. (2002) Epistemology and the Socio – Cognitive Perspective in Information Science. En *Journal of the American Society for Information Science and Technology*, 53 (4), 257 – 270.

¹¹⁸ Durán Muñoz, Isabel. (2012) *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*. Frankfurt am Main : Peter Lang p. 53 -54.

¹¹⁹ Durán Muñoz, Isabel. (2012) *Op. Cit* p. 53.

Durán Muñoz, Isabel (2012)¹²⁰ señala que para Rita Temmernam los términos son multidimensionales y flexibles a partir de la influencia que ejerce la situación comunicativa en la que se utilizan y que la TST ha valorado.

Asimismo se reconoce que en la TST el lenguaje tiene un papel relevante en la comprensión del mundo y dicho mundo se encuentra en la mente humana. En esta línea de pensamiento Naumis Peña, C (2012)¹²¹ señala que en la TST “se habla de unidades de comprensión para describir los aspectos flexibles que hacen la variedad de definiciones de un término y que están relacionados con el nivel de especialización del emisor y del receptor del mensaje”.

Desde la perspectiva de la Teoría Socio – Cognitiva de la Terminología y de acuerdo con lo que señala Naumis Peña, C (2012)¹²² “se recomienda que se construyan sistemas de acceso a la información basados en la cultura propia y la traducción debe ser utilizada como guía, pero siempre adaptada a los valores sociales del grupo al que se dirige. Ello sólo se puede hacer mediante estudios lingüísticos mucho más profundos que categorizar el material en decisiones rápidas y personales tomadas en el momento de la catalogación”

¹²⁰ Durán Muñoz, Isabel. (2012) *Op. Cit* p. 55.

¹²¹ Naumis Peña, C. *Op. Cit* p.17

¹²² Naumis Peña, C. (2012) *Op. Cit* p.21.

2.9. Terminología y tareas bibliotecológicas

La relación que se establece entre la Terminología y la Bibliotecología es, de acuerdo con Teresa Cabré (2003),¹²³ de naturaleza bidireccional, es decir, una relación recíproca e imprescindible. De entre ese tipo de relaciones se destacan algunas actividades documentales que realiza el bibliotecario o documentalista. En este apartado se describirán algunas de las tareas y herramientas que se utilizan en el trabajo bibliotecario y en las que considero interviene la terminología.

La relación recíproca que se da entre la Terminología y la Bibliotecología sucede en dos momentos diferentes, el primero es la descripción y representación del contenido de los documentos o indización, la segunda tiene lugar durante el proceso de recuperación de la información. En un punto intermedio entre estos dos momentos, está presente el tesoro, que juega un doble papel: por una parte, controla y uniforma el uso de términos en la indización y, por otra, precisa y facilita al usuario final la tarea de la recuperación de información.

La indización es, de acuerdo con Hutchins (1975),¹²⁴ “la formulación de una expresión en un lenguaje documental que captura el contenido semántico de un documento”. Dicha formulación no es otra cosa más que la explicitación de los elementos más representativos de la información que se transmite en un

¹²³ Cabré Castellví, Teresa (2000). “Terminología y Documentación” en *Documentación Terminología y Traducción*. Madrid: Síntesis, p. 31–43.

¹²⁴ Hutchins, W. J. (1975) *Op Cit.* p 107.

documento. Según Teresa Cabré¹²⁵ la descripción del contenido contempla las siguientes dos actividades:

- *La indización o identificación explícita de las unidades y expresiones representativas del contenido de los documentos.* Según esta autora “la indización es en el fondo una operación terminológica, es una actividad de detección de la terminología más indicativa del contenido documental, por la que se atribuye un valor de contenido a determinadas unidades y expresiones del texto, que son las que servirán de anclaje para la recuperación de la información”. Asimismo señala que “la identificación explícita mediante la indización de las unidades más representativas del contenido del texto, se suele controlar a través de un tesoro, para reducir la dispersión de las expresiones lingüísticas espontáneas.
- La elaboración de resúmenes, es una tarea a través de la cual los términos sintetizan el contenido informativo de un documento. El resumen es el resultado de una operación de condensación, que consiste en seleccionar la información más relevante sobre el contenido y expresarla de manera sintética.

Se puede decir que las herramientas que se utilizan en la tarea de clasificación documental y, por lo tanto, en la recuperación de la información contenida en los documentos son al menos tres. Uno de estos instrumentos es la clasificación conceptual; el segundo es el análisis de textos y el otro, los tesauros.

¹²⁵Cabré Castellví, Teresa (2000). “Terminología y Documentación” en *Documentación Terminología y Traducción*. Madrid: Síntesis. p. 36.

Para concluir esta sección del segundo capítulo se presenta a continuación la comparación de las tres corrientes de la Terminología que se han presentado y su relación con los tesauros.

La Teoría General de la Terminología (TGT) tiene una relación estrecha con la elaboración de tesauros, en la medida que aporta elementos teóricos y metodológicos que permiten el estudio claro y preciso de los términos que constituyen un tesoro y de las relaciones entre ellos, sobre todo las relaciones de tipo jerárquico. Dicha teoría tiene, entre otras limitaciones, el de considerar a los términos como exclusivos para una disciplina o campo de conocimiento, condición que no promueve el estudio de los términos de forma amplia y diversa con la visión multidisciplinar.

En cambio la Teoría Comunicativa de la Terminología (TCT), estudia a los términos con un enfoque amplio y diverso, al considerarlos como unidades de conocimiento, unidades de lenguaje y unidades de comunicación. Asimismo dicha teoría muestra un avance respecto a la TGT, cuando considera a los términos como un elemento importante para promover la comunicación entre diferentes campos del conocimiento a partir de la elaboración de herramientas lexicográficas como los tesauros.

Por su parte la Teoría Sociocognitiva de la Terminología (TST) tiene una relación estrecha con el trabajo bibliotecológico, especialmente con tareas sobre acceso temático a los contenidos documentales, la indización y por lo tanto con la

elaboración de herramientas para la indización y la recuperación de información, como los tesauros.

La TST al igual que la TCT considera que los términos son multidimensionales y flexibles y en su estudio y uso deben considerarse los elementos que intervienen en un evento comunicativo, es decir características del emisor, receptor, medio y mensaje. El avance y aportación de la TST estriba en agregar información y datos sobre aspectos de tipo social y procesos cognitivos del grupo meta hacia quien están orientados los esfuerzos para el diseño de productos u objetos, como los tesauros. Esta suma de información y datos sociales y cognitivos se puede lograr a través de investigaciones y estudios de tipo lingüístico a profundidad que permita establecer categorías sobre los términos que formarán parte de un tesoro.

Capítulo 3

El tesoro y la Informática

*T.S. Eliot decía “¿Dónde está el conocimiento que perdemos con la información?”
El conocimiento no es conocimiento si no es organización.
Puesta en relación y en contexto con las informaciones.
Las informaciones constituyen parcelas del saber dispersas.
Edgar Morin.*

En el capítulo anterior se analizaron dos temas: las bases lingüísticas para la elaboración de un tesoro, en primer lugar, y en segundo, la relación que se establece entre la terminología y la elaboración de un tesoro. El tema del presente capítulo es la Informática y su relación con los tesauros.

Este tercer capítulo tiene cuatro diferentes propósitos. El primero es identificar, describir y analizar algunas aportaciones teóricas y metodológicas con las que la Informática contribuye al proceso de indización y recuperación de la información en los que interviene el tesoro.

El segundo busca describir el procedimiento básico para la elaboración de un tesoro de acuerdo con la informática. Dicho procedimiento está compuesto por los siguientes pasos: la definición del rango y profundidad del tesoro, la identificación de las fuentes para construir el vocabulario, la extracción de los términos candidatos, la organización del vocabulario en grupos de términos asociados, la organización en grupos a los términos, el establecimiento de un sistema de notación, el tratamiento de los términos precoordinaados o de más de una palabra, el refinamiento del vocabulario, la creación de referencias entre los términos y el mantenimiento del tesoro.

El tercero es describir y valorar algunas de las herramientas informáticas que se utilizan en la indización y la recuperación de la información. Las herramientas que se estudiarán son los analizadores morfológicos, sintácticos y semánticos.

Finalmente, el cuarto es describir y valorar algunas de las herramientas de la termiología que se utilizan en la indización y la recuperación de la información.

La información que se expone en este capítulo se ha organizado de la siguiente manera: se inicia con una reflexión sobre la relación entre la Informática y el tesoro, luego se exponen los antecedentes que se encontraron sobre esa relación en la investigación documental que se llevó a cabo para tal propósito. Posteriormente, se describen y analizan las características de los analizadores morfológicos, sintácticos y semánticos. Por último, se describen y analizan, de manera general, las características de bancos y vocabularios terminológicos de léxicos especializados.

3.1. El tesoro y la informática: definición del término y antecedentes

3.1.1. Definición del término

El término *tesoro* se encuentra en diversas obras de consulta especializadas en informática, las cuales lo definen desde diferentes enfoques, por ejemplo, en el *Diccionario de informática* de Valerie Illingworth (1993)¹²⁶ se define como “característica de los sistemas de tratamiento de textos por la cual las expresiones similares y sinónimos pueden visualizarse en pantalla e

¹²⁶ Illingworth, Valerie (ed.) (1993). *Diccionario de Informática*. Madrid. Oxford University Press/Diaz de Santos. p 638.

incorporarse en el texto. Los sistemas de recuperación de texto completo pueden poseer *búsqueda de tesauros* como opción por la cual pueden establecerse también términos similares en significado a los buscados”.

Asimismo, y de acuerdo con Virga y Mesters (1994),¹²⁷ un tesoro es “un léxico jerarquizado de términos comprensibles por el programa para la búsqueda de una información. Un tesoro es, por tanto, una herramienta indispensable para consultar de forma eficaz un banco de datos. Para su uso adecuado, es necesario crear un nuevo tesoro (no se consulta de igual forma un banco de datos jurídico que uno médico)”. En el *Dictionary of Computing* (1995)¹²⁸ un tesoro es “un archivo que contiene sinónimos que se presentan como alternativas para solucionar *misspelt word* durante un *spell-check*” (sic).

Las definiciones consultadas en diccionarios de informática presentan información escueta y, sobre todo, enfocada hacia el uso y aplicación de los tesauros en los sistemas para la recuperación de la información, es decir, el uso de la herramienta para propósitos acotados.

Sin embargo, para propósitos de esta tesis un tesoro es *una herramienta lógico-conceptual a través de la cual se representa el contenido de los documentos, se trata de sistema complejo en el que se establecen relaciones entre los términos*

¹²⁷ Virga y Mesters (1994). *Diccionario de microinformática: todas las nuevas tecnologías informáticas*. Madrid: Thompson / Paraninfo. p 404 – 405.

¹²⁸ *Dictionary of Computing* (1994). 2ª edición. Cambridge: Peter Collins. p. 280.

para crear una red semántica que permite al sistema realizar tareas de indización y recuperación de información.

3.1.2. Antecedentes sobre el tesoro y la recuperación de la información

El tema de la informática en la elaboración de los tesauros ha sido estudiado de manera abundante y exhaustiva por ingenieros en informática y cómputo, así como por científicos de la información. Entre los trabajos realizados se puede encontrar un acercamiento desde distintos puntos de vista, por ejemplo: el diseño interno de la herramienta, es decir, las características y condiciones que contribuyen a que el tesoro sea una herramienta sólida y consistente con los principios teóricos del diseño de instrumentos informáticos para la indización y la recuperación de la información. Otra vertiente de estudio ha sido vista desde la perspectiva del uso, aplicación y eficacia de la herramienta en la recuperación de la información. Una más ha sido desde la óptica de las ventajas que tiene el uso del tesoro en la indización, lo que —como se señaló en el primer capítulo—, es la contraparte del proceso del análisis documental en un sistema de almacenamiento y recuperación de la información.

Asimismo, se ha investigado la construcción automática de los tesauros y sus efectos en los niveles de eficiencia y calidad en la recuperación de información en un sistema. También se ha indagado sobre los tesauros electrónicos en los que se incluyen interfaces que utilizan herramientas de apoyo multimedia, entre otros aspectos.

Ante esta variedad de acercamientos o enfoques para el estudio del tesoro desde la Informática, resulta pertinente señalar y acotar el alcance de esta apartado, en el que sólo se mencionarán datos sobre algunas experiencias de investigaciones relacionadas con el tesoro y la recuperación de la información.

3.2. La recuperación de la información a través de los tesauros

Los tesauros son un constructo conceptual complejo en cuya creación y uso se han utilizado nociones fundamentales de la Informática. Por ejemplo, para la recuperación de la información, los tesauros han sido empleados frecuentemente para aclarar las dudas de los usuarios en el momento en que éstos plantean una pregunta al sistema y agregan palabras y frases.

Asimismo, han sido utilizados para incrementar el texto de los documentos a través de la indización. El incremento al texto se puede utilizar para realizar tareas relacionadas con el proceso de indización de los documentos, es decir, la agrupación y construcción de redes, a través de las que se establecen relaciones y nexos con otros documentos.

La recuperación de la información ha tenido cambios fundamentales desde sus etapas tempranas, esto es, desde los días en los que el procesamiento de conjuntos de datos se realizaba en equipos que almacenaban registros de apenas algunos cientos de documentos, hasta el advenimiento de las computadoras personales en las que se da una retroalimentación interactiva para la búsqueda de textos completos de entre *gigabytes* de datos en el Internet.

De acuerdo con Evens (2003),¹²⁹ “el proyecto Tipster de la DARPA *Defense Advanced Research Projects Agency* ha tenido un papel preponderante en estos cambios, orientando su atención hacia la comunidad de los investigadores, utilizando documentos académicos, bibliografías y artículos de publicaciones periódicas de diferentes extensiones y tamaños, y utilizando también *softwares* para el ordenamiento y rastreo de cientos de miles de documentos, en lugar de algunos cientos o miles que se procesaban al inicio de los trabajos de investigación sobre la recuperación de la información”.

Información más amplia sobre estos cambios se puede encontrar en los reportes publicados en la TREC *Text Retrieval Conference*, celebrada en 1993. Los cambios que se mencionan en el proyecto Tipster han tenido impacto en otros desarrollos para la recuperación de información en ambientes multilingües, por ejemplo, se han llevado a cabo experimentos en pequeña escala en lenguas tan distantes como el japonés, finlandés, árabe, chino e italiano. Estos cambios también han tenido gran impacto en el desarrollo de los tesauros en dos sentidos, uno sobre el uso de la herramienta y otro sobre su construcción.

3.3. La recuperación de la información y la evolución de los tesauros

Son cuatro momentos o etapas importantes las que marcan el desarrollo de investigaciones sobre la construcción y diseño de tesauros en los que interviene la Informática. Las etapas y trabajos que se señalarán, a continuación, se

¹²⁹ Evens, Martha (2002). “Thesaural Relations in Information Retrieval” en Rebecca, Green; Carol, A. Bean y Sung Hyon Myaeng, (eds.) *The Semantics of Relationships. An Interdisciplinary Perspective*. Netherlands. Kluwer Academic. p. 144.

orientan hacia y en el contexto de la recuperación de la información. En ellos se pueden identificar de manera general el uso de conceptos y herramientas teórico–metodológicas derivadas de la estadística, la semántica y la morfología.

Un primer momento se identifica con los experimentos que realizaron Gerard Salton y sus discípulos en los años cincuenta. Salton creó la prueba llamada *Smart system* con el propósito de desarrollar algunos experimentos sobre la recuperación de la información, en los que se incluía la construcción de un tesoro. Una característica del enfoque que utilizó este autor, en esa etapa de sus investigaciones, es la normalización consistente del lenguaje para la recuperación de la información por medio del establecimiento de clases de conceptos. Salton experimentó con algunos métodos para la construcción automática de los tesauros, partiendo del análisis estadístico de los términos y el estudio de la construcción de jerarquías entre ellos.

El segundo momento es el de la construcción de los tesauros a partir de los conocimientos derivados de la semántica, esta etapa la representan los trabajos realizados por Karen Spark–Jones en los años setentas. La investigadora estaba convencida de la utilidad del tesoro, siempre y cuando éste estuviera organizado con base en un sistema automático de clasificación, encontrando así una alternativa al enfoque estadístico de concurrencia de términos desarrollado por Salton. Spark–Jones desarrolló su propia teoría sobre la clasificación semántica para la recuperación de la información. El elemento básico en la teoría de la clasificación semántica es un *run*, es decir, un grupo de formas de palabras

que son intercambiables en algunos contextos. Sus experimentos sobre la recuperación de la información han sido calificados como efectivos.

El tercer momento tuvo lugar en los años ochenta y está representado por Edgard Fox con su enfoque de los inventarios de relaciones entre los conceptos. Fox decidió que era importante considerar el uso de un amplio rango de relaciones en el proceso de la recuperación de la información. Su postura teórica se deriva de los conocimientos y experimentos realizados a partir del Modelo Textual de Significados de Mel'cuk's en los años setentas. El modelo contempla un gran inventario de relaciones, dichas relaciones son de naturaleza diversa, morfológicas por ejemplo, que se dan en el eje paradigmático, es decir, relaciones entre las palabras que se expresan con una misma carga semántica o de significado. O bien otro grupo de relaciones que se dan en el eje sintagmático, es decir, aquellas que se expresan como un nexo entre las palabras que aparecen juntas en una oración. A partir de la implementación y uso del modelo, Fox encontró que al combinar ambos tipos de relaciones en las preguntas que se formulaban al sistema de recuperación de información y utilizando el tesoro, se enriquecían y mejoraban significativamente los resultados de la búsqueda en el sistema.

El cuarto momento es el de la construcción de tesauros con la ayuda de la computadora. Este sucedió a partir de la primera mitad de los años setentas, en el contexto de las investigaciones realizadas para la construcción de algunos diccionarios electrónicos. Sin embargo, fue hasta los años ochenta que Fox y Evens llevaron a cabo algunos experimentos sobre la construcción de un tesoro

para la recuperación de la información a partir de los datos almacenados en un diccionario electrónico, específicamente del *Webster's Seventh Collegiate Dictionary*, cuyos resultados no fueron satisfactorios en un inicio.

La ambigüedad en las definiciones contenidas en el diccionario electrónico fue un obstáculo importante al que se enfrentaron Fox y Evens. No obstante, un hallazgo que se puede derivar de dicho obstáculo fue la creación de una herramienta de la lingüística computacional que permitiera superar el fenómeno de la ambigüedad. Sólo hasta entonces les fue claro que para la construcción automática de un tesoro se debía incluir una herramienta que permitiera hacer transparente la ambigüedad del sentido en una definición, por ejemplo, que permitiera distinguir e identificar la diferencia entre *pine cones* y *ice-cream cones* (piñas y conos de nieve).

Otro avance en el tema fue el que realizó Calzonari en 1988¹³⁰ (Universidad de Pisa), quien señala que “el diccionario y el tesoro pueden y deben combinarse”. Ella construyó una base de datos de léxico en italiano de más de un millón de palabras que formaron 120 000 radicales o raíces. Esta base de datos ha sido utilizada para diferentes propósitos, como la traducción automática, la generación de textos, el análisis sintáctico y la recuperación de la información. Asimismo se han desarrollado proyectos para la recuperación de información a partir del estudio de fenómenos semánticos como la sinonimia, y su contraparte la antinomia, así como la derivación morfológica y las relaciones morfológicas explícitas que aparecen en un diccionario.

¹³⁰ Calzonari, N. (1988). “The dictionary and the thesaurus can be combined” en M. Evans (ed). *Relational Models of the Lexicon*. Reino Unido : CPU. p. 75 -96

3.4. Pasos básicos para la elaboración de un tesoro

Si como se señaló anteriormente un tesoro es un constructo conceptual complejo en sí mismo, su diseño y construcción es un proceso que demanda de un trabajo metodológico riguroso y sustentado en principios teóricos que proviene de diferentes disciplinas. Desde la perspectiva de la Informática, la construcción de un tesoro tiene como punto de partida el reconocimiento de grupos de conceptos que comparten características.

Cada grupo está organizado como una jerarquía de conceptos y representado por términos en lenguaje natural adecuados para el contexto en el que se utilizará el tesoro. Asimismo y para una comunicación eficiente entre el usuario y el sistema de búsqueda, se requiere de un vocabulario regulado con relaciones conceptuales definidas de manera clara y notas de alcance que favorezcan el uso consistente de la herramienta.

Un paso previo y necesario que se sugiere, antes de iniciar la construcción de un tesoro, es el de plantearse las siguientes preguntas:

- ¿Quiénes y cuántos serán los usuarios del tesoro?
- ¿Cuáles son las expectativas de los usuarios potenciales del tesoro?
- ¿Saben usar un tesoro para la búsqueda de la información o para la indización?
- ¿Podrán usar un mapa conceptual para encontrar información?

Después de contestar de forma amplia, precisa y exhaustiva las preguntas planteadas, se sugiere continuar con los pasos básicos que Batty, David (1989) en su ya clásico artículo *Thesaurus construction and maintenance: a survival kit* propone para la elaboración de un tesoro. Dichos pasos son los siguientes:

3.4.1. Definir el rango y profundidad de un tesoro

La profundidad de un tesoro se define por el número de términos y su especificidad. Un campo de información puede ser dividido en tres áreas: 1) *área núcleo*, que contiene temas y términos totalmente pertinentes al tema principal, 2) *área limítrofe*, que contiene temas y términos que apoyan al tema principal, pero que no están directamente relacionados, y 3) *área externa*, que contiene temas y términos que pueden apoyar al tema principal, pero que no aportan conocimiento experto al mismo.

De manera gráfica podría expresarse de la siguiente forma:

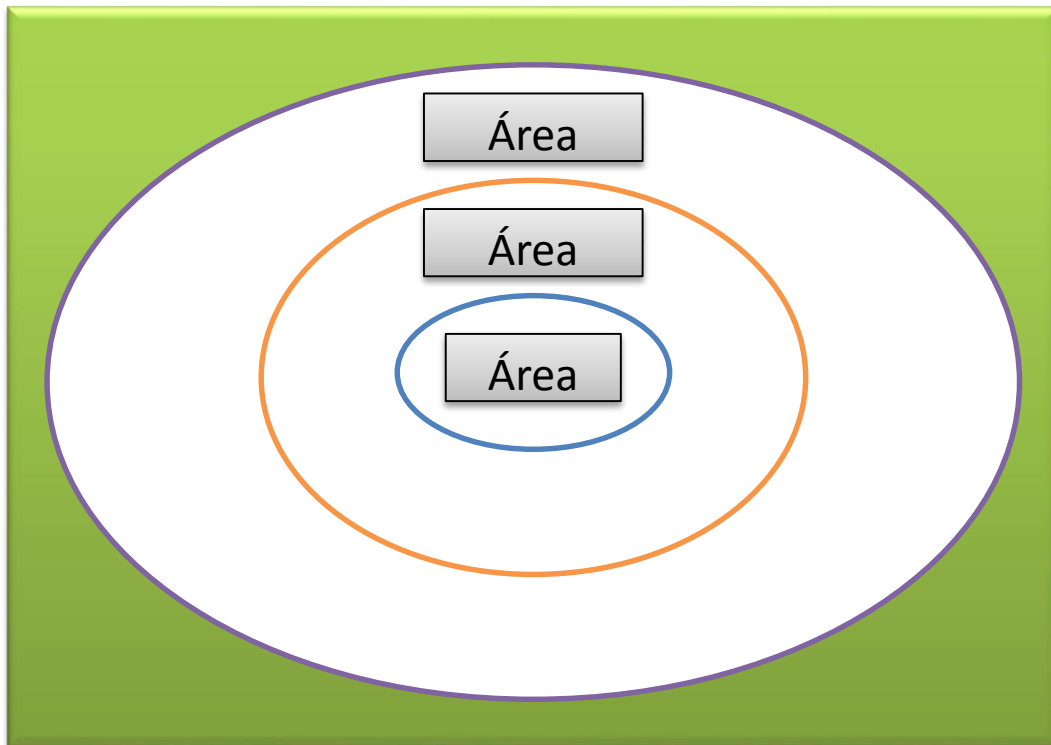


Figura 11. Rango y profundidad de un tesoro. Elaboración, Valdez, R. J. (2013).

3.4.2. Identificar las fuentes para construir el vocabulario

Para construir el vocabulario de un tesoro se recomienda utilizar las siguientes fuentes: una muestra de la literatura clásica sobre el tema del tesoro; una muestra de la literatura general del tema, como: memorias de congresos, publicaciones académicas reconocidas del tema central, publicaciones de organismos oficiales y reconocidos, resúmenes de las principales publicaciones periódicas del tema; una muestra de las preguntas de los usuarios y tesauros e índices existentes sobre el tema.

Con lo descrito hasta este momento, pareciera que se trata de la forma tradicional de elaborar un tesoro, todavía no interviene la Informática en la tarea, es decir, se describió una tarea manual. No obstante, a partir de esta etapa comienza a participar la Informática. A continuación presento algunos datos sobre la manera en la que se involucra dicha disciplina.

3.4.3. Extracción del vocabulario o términos candidatos

El criterio inicial para seleccionar los términos que formarán el tesoro es tomarlos directamente de las fuentes documentales, por ejemplo, si aparece *salario, jornal, cuota* que son palabras que significan *pago por un trabajo hecho*, se deben tomar la(s) que aparecieron en los documentos fuente.

Actualmente existen herramientas informáticas que pueden extraer términos de algunas fuentes documentales, como es el caso de sistemas informáticos para el tratamiento de la información textual, ejemplos de este tipo de programas son el software para la extracción automática de términos, por ejemplo *eXtraTerm* de Trados que menciona Pozzi, M (2003) o *software* más sofisticados como programas informáticos para la obtención de concordancias también mencionados por Pozzi, M (2003), por ejemplo el *software KWIC (keywords in context)*. En esta misma línea se pueden mencionar los *softwares* utilizados para la extracción automática de contextos definitorios en textos especializados que señala Sierra, Martínez, G (2003).

3.4.4. Organizar el vocabulario en grupos de términos asociados

La organización de términos consiste en el agrupamiento de éstos, atendiendo a criterios semánticos de afinidad. Por ejemplo, *madera, plástico, metal*, se agruparían bajo el concepto de materiales. De forma general podemos decir que los términos se agrupan en tres categorías: asociación, equivalencia y jerarquía. Cuando se hace de forma manual, esta etapa del trabajo es la que consume la mayor parte del tiempo de construcción de un tesoro. Sin embargo, existen herramientas informáticas con las que esta etapa del trabajo es más rápida. En el mercado hay *softwares* disponibles que organizan los términos a partir de los hallazgos hechos por Karen Spark-Jones en los años setentas. Algunos programas de este tipo son los que menciona López Alonso, Miguel Ángel (1998), por ejemplo ASTUTE (*Automated System for Thesurus Updating, Testing and Editing*), diseñado por el Centro de Cálculo de la Comisión de las Comunidades Europeas en Luxemburgo o MINISIS (*Micro computer Integrated Set of Information System*) a través del módulo THES diseñado en la UNESCO.

3.4.5. Acomodo de grupos de términos

El acomodo de grupos de términos se realiza a través del establecimiento de relaciones entre ellos. Las relaciones que se establecen son jerárquicas del tipo padre–hijo o de tipo hermanos, mismas que corresponden a las entradas de TG, TE y TR. Que de manera gráfica podría expresarse de la siguiente forma:

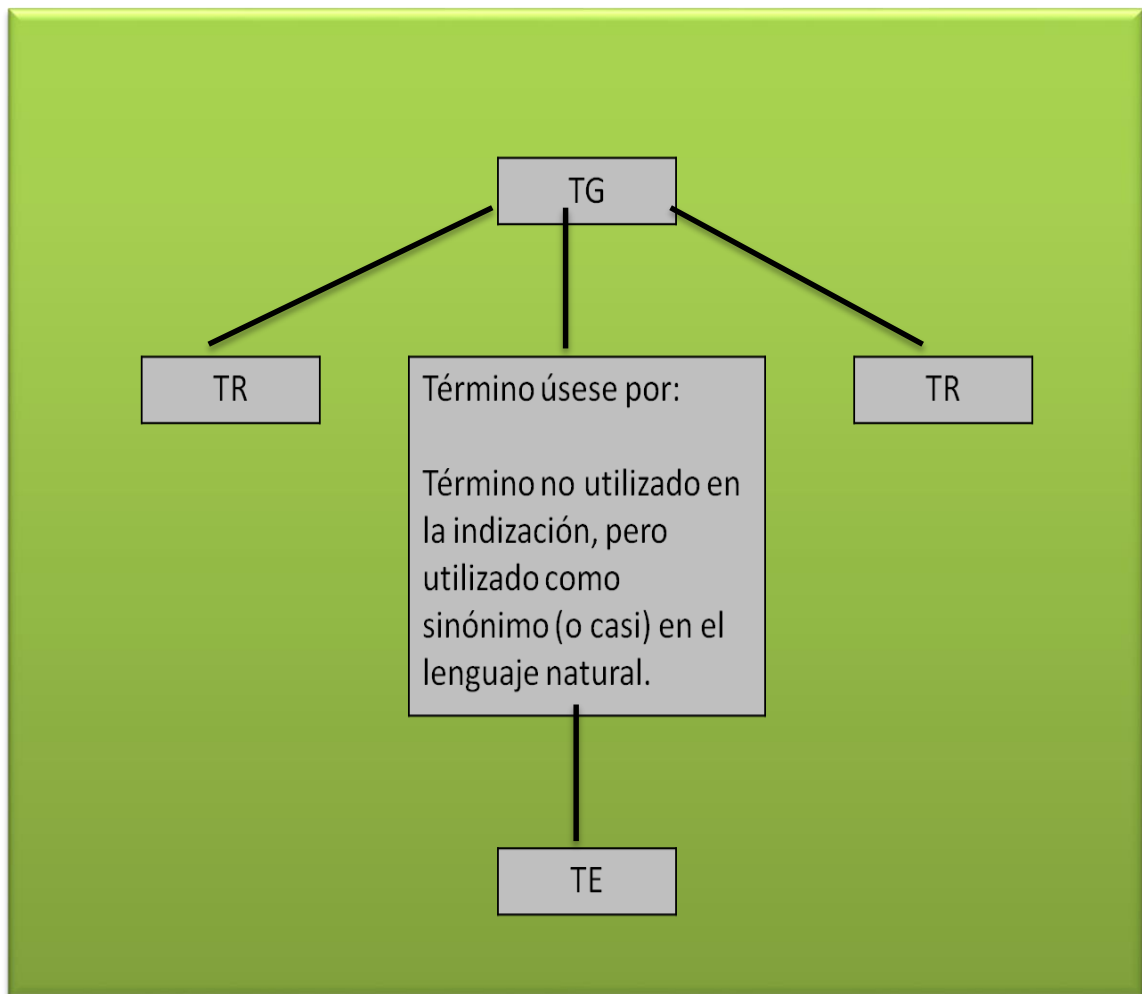


Figura 12. Tipos de relaciones entre los términos de un tesoro. Elaboración, Valdez, R. J. (2013).

3.4.6. Establecer un sistema de notación

Una vez que se han agrupado los términos del tesoro, es necesario definir una notación a través de un código simbólico en el que se enuncie la jerarquía de los términos. Dicha notación debe ser fácil de pronunciar, escribir y recordar. Al definir un sistema de notación se deben tomar en cuenta las perspectivas de crecimiento del tesoro.

3.4.7. El tratamiento de los términos precoordinados o de más de una palabra

Cuando se elabora un tesoro de forma automática, pueden presentarse situaciones o momentos en los que es necesario incluir términos formados por más de una palabra. Este hecho tiene algunas implicaciones importantes para el tratamiento automático de los términos. Entre estas implicaciones se pueden distinguir al menos tres situaciones: i) cuando el término de más de una palabra se utiliza en un campo de conocimiento determinado, ii) cuando la frase de más de una palabra aparece frecuentemente en los documentos que se han tomado como muestra para la indización y iii) cuando es necesario utilizar la frase de más de una palabra para hacer más claro el significado de la misma. Algunas herramientas informáticas que realizan esta tarea son los analizadores sintácticos.

3.4.8. Refinamiento del vocabulario

En esta fase de elaboración del tesoro, se deben realizar al menos las siguientes acciones: i) decidir sobre el nivel de detalle necesario para el área del conocimiento que se quiere representar, ii) transformar o normalizar el vocabulario extraído de los documentos a través de la indización en términos aceptados por la comunidad de usuarios meta del tesoro, iii) distinguir de entre la misma palabra los diferentes significados, usos y aplicaciones que tiene y iv) hacer más clara la forma de uso de un término añadiendo siempre una nota de alcance. Cabe mencionar aquí, que las notas de alcance no son definiciones más

amplias de los términos, aunque en algunas ocasiones las definiciones pueden ser buenas notas de alcance, por ejemplo, la definición psicológica de adolescencia puede no ser útil para un indizador o para los usuarios, sin embargo, si se utiliza una forma del tipo *Úsese para un(a) joven de entre 13 y 18 años* puede resultar una manera útil de nota de alcance.

3.4.9. Creación de referencias entre los términos

La creación de referencias en un tesoro alude al establecimiento de las relaciones entre los términos, si son del tipo género–especie (TG–TE) o de jerarquía básica y de semejanza o similitud como: Padre–Hijo o semejanza de familia. A partir del reconocimiento de estos tipos de relaciones es necesario hacer un análisis para determinar entre cuáles términos se pueden establecer más de una relación con un término general, esto sucede cuando varios tipos de jerarquías ocurren al mismo tiempo, como en el caso de Cosa–Tipo de cosa y Todo–Parte.

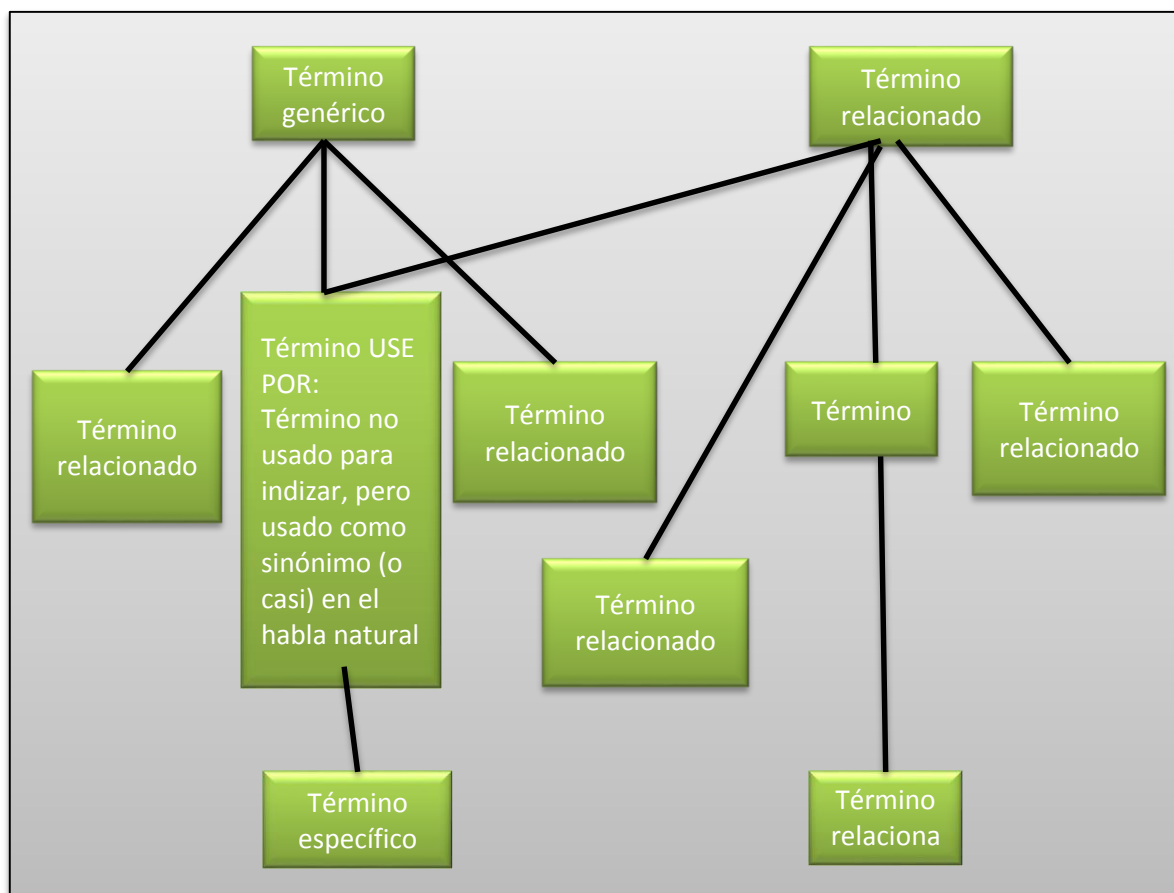


Figura 13. Creación de referencias entre los términos. Elaboración, Valdez, R. J. (2013).

La creación de este tipo de referencias entre los términos no siempre se pueden establecer por medios automáticos, entonces será necesario recurrir al trabajo humano de los documentalistas o lexicógrafos, según sea el caso del equipo de profesionales encargados de la elaboración del tesauro.

3.4.10. Mantenimiento del tesauro

Una de las características del tesauro es su condición dinámica y evolutiva, es decir, se trata de un organismo vivo que se modifica y crece. El mantenimiento

del tesoro es un proceso mediante el cual se agregan nuevos términos al mismo y se suprimen los términos que resultan obsoletos. Realizar dichos cambios es una tarea que demanda rigor sistemático y cuidado extremo para evitar que se fracture la estructura del tesoro.

Emilia Currás (1991)¹³¹ y Vanda Broughton (2006)¹³² proponen dos formas para la actualización y mantenimiento de un tesoro. La primera, la de Emilia Currás, parte de la noción de integración de los términos para el mantenimiento del tesoro y propone las siguientes tareas: añadir, sustituir, suprimir y modificar los términos para reestructurar la presentación, además señala que para el mantenimiento, deberán hacerse revisiones periódicas por personal especializado. En general, se puede decir que este enfoque para el mantenimiento está centrado en la descripción de un proceso manual.

La segunda forma, propuesta por Vanda Broughton, tiene un enfoque sistémico en el que se utiliza un *software* diseñado para la construcción de tesauros. La autora reconoce que el mantenimiento es un proceso fácil de entender, si se toma en cuenta que suprimir, sustituir, añadir o modificar un término puede ser controlado desde las normas del diseño mismo del *software*, es decir, a través de un proceso similar, pero en sentido contrario al que sucede cuando se organizan los términos durante la etapa de construcción de un tesoro.

¹³¹ Currás, Emilia (1991). *Thesaurus Lenguajes Terminológicos*. Madrid: Paraninfo. p. 229.

¹³² Broughton, Vanda (2006). *Essential thesaurus construction*. Londres: Facet publishing. p. 197.

Broughton describe el proceso de la siguiente forma: primero es necesario decidir e identificar a qué faceta o categoría pertenece el término, después establecer si existe una ordenación apropiada o si es necesario crear una nueva faceta o categoría. Por ejemplo, si queremos agregar el término *hurones* al tesoro, sería necesario entrar a la categoría de animales, después a la de mustélidos y seguir la ruta que se muestra enseguida, hasta llegar a la adición del término nuevo:

T	carnívoros
TC	cánidos
TCF	zorros
TCG	zorros azules
TCH	coyotes
TG	perros domésticos
TL	osos
TL/K	osos de zoológico
TM	musteliade ¿ver traducción especializada?
TMM	visón
TMN	tejones
TMO	nutrias
TMP	nutrias euroasiáticas
TMR	<i>hurones</i> -----Término agregado

Al agregar un término, el *software* convierte una taxonomía en un formato alfabético y el resultado final sería el que a continuación se presenta:

Hurones **TMR**

BT mustelides

RT visión

Tejones

Hurones

Hasta aquí se han detallado los pasos básicos para la construcción de un tesoro, se han mencionado algunas herramientas informáticas, como software especializados para extracción automática de términos de colecciones documentales y para el establecimiento de relaciones semánticas (equivalencia, asociación y jerarquía).

El procedimiento que se describió es el que se realiza de acuerdo con los principios que la informática propone. A continuación se presentarán información y datos en los que se describen y valoran algunas de las herramientas informáticas que se utilizan en la indización y la recuperación de la información, dichas herramientas son los analizadores morfológicos, sintácticos y semánticos. También se explicará cuál es la relación que estas herramientas guardan con la indización y recuperación de la información.

Además se mencionarán algunas de las herramientas terminológicas que se utilizan para la indización y recuperación de la información.

3.5. Herramientas lingüísticas para la indización y la recuperación de la información

En el proceso de indización y recuperación de la información intervienen factores tanto humanos como tecnológicos, en esta sección no se abordarán los aspectos humanos de manera puntual y directa, sólo se estudiarán tres herramientas lingüísticas que tienen una relación estrecha con la indización y la recuperación de la información, las cuales surgen y se desarrollan a partir de conceptos propios de la Informática.

Asimismo cabe señalar que, para propósitos de esta tesis, el estudio de las tres herramientas lingüísticas se hará por separado, aunque en artículos¹³³ relacionados con el tema en ocasiones se mezclen aspectos sintácticos, semánticos y morfológicos.

Las herramientas a las que se alude son los analizadores morfológicos, sintácticos y semánticos. Pero ¿qué es un analizador? En Informática un analizador forma parte de un proceso más amplio y complejo que se llama compilación, al respecto Karen Lemone (1998)¹³⁴ define la compilación como

¹³³ Por ejemplo en el artículo de Betancourt Rodríguez, Osvaldo y Patricia Arias García (1987). “Diseño de un analizador semántico–sintáctico de un sistema automatizado de información” en *Actualidades de la información científico técnica* 18 (2) abril. p 98 – 107.

¹³⁴ Lemone, Karen A. (1998). *Fundamentos de compiladores. Cómo traducir al lenguaje de computadora*. México: CECSA. p. 3- 4.

“una secuencia de fases seriadas que comienzan con el análisis lexicográfico y finalizan con la generación de un código, con el manejo de tablas, módulos de error y de E/S que interactúan con más de una fase”, de manera gráfica se puede representar así:

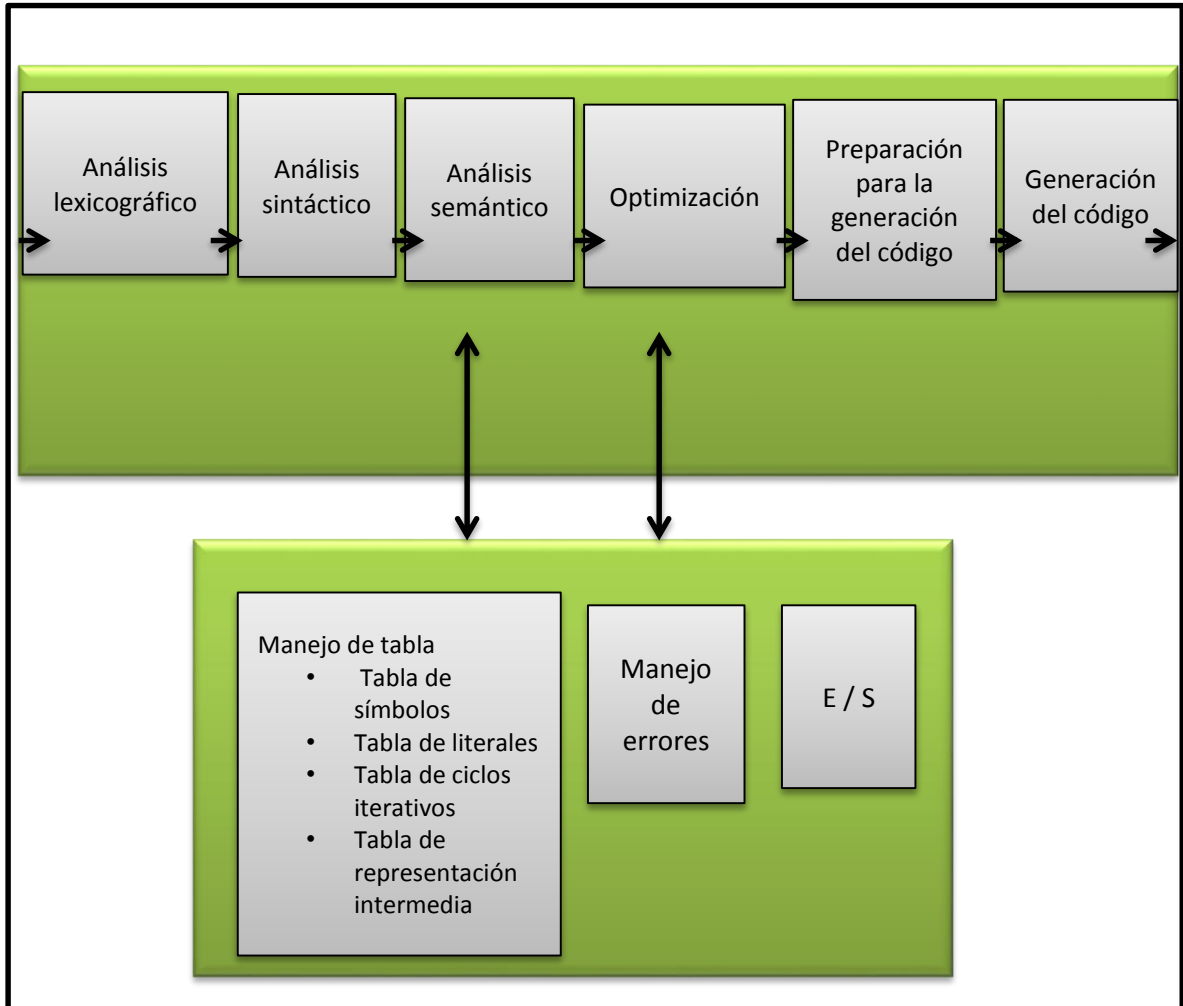


Figura 14. Modelo informático de un analizador. Lemone, K. A. (1998)

Este proceso tiene el propósito de traducir un lenguaje de alto nivel a un lenguaje ensamblador para que la computadora pueda ejecutar una serie de instrucciones. Que para el caso de los analizadores morfológicos, sintácticos y semánticos que se utilizan en la indización y la recuperación de la información se trata de tareas como la afijación (sufijos y prefijos), el orden de palabras y la

sinonimia. Un analizador se puede utilizar como una interfase entre un lenguaje natural y formalizado con un sistema de información.

En un sistema de información, como puede ser una base de datos bibliográfica, por ejemplo, Gil Leiva (2008)¹³⁵ distingue entre dos tipos de analizadores: los etiquetadores morfológicos y los etiquetadores sintácticos, además señala “que este tipo de herramientas de la lingüística computacional son las que se utilizan en los sistemas automatizados de la indización”.

3.5.1. Analizadores morfológicos

Un analizador morfológico es una herramienta informática que permite el reconocimiento de reglas morfológicas para la formación de una palabra. Para Gil Leiva (2008)¹³⁶ dicha tarea se realiza “a través de la descomposición de las palabras en sus raíces y de esta forma se agrupan en una raíz común” lo que permite el reconocimiento de prefijos y sufijos en las palabras. Este trabajo se refleja en la recuperación de información, entre otras tareas.

A continuación se plantean dos preguntas con el propósito de señalar y analizar la relación que existe entre los analizadores morfológicos, la indización y recuperación de información, y el tesauro.

La primera pregunta es la siguiente:

¹³⁵ Gil Leiva, Isidoro (2008). *Manual de indización. Teoría y práctica*. Madrid: Trea. p.327.

¹³⁶ Gil Leiva, Isidoro. *Op. Cit.* p.324.

- ¿Cuál es el papel que tiene un analizador morfológico en la indización y la recuperación de información?

Un analizador morfológico tiene un papel importante en la indización y la recuperación de información, ya que permite el reconocimiento de las categorías gramaticales de las palabras, tales como sustantivo, adjetivo, verbos y otras. A partir de ese primer nivel de reconocimiento, el analizador lleva a cabo la descomposición de las palabras. En un segundo nivel de reconocimiento, identifica algunas de las características morfológicas, por ejemplo, las marcas de género, número, persona, tiempo y modo, dichas marcas son útiles en algunas de las tareas de indización automática, por ejemplo. Otro de los usos básicos que tiene un analizador morfológico es el reconocimiento de la raíz o base de la palabra, con lo que se puede calcular la aparición de un mismo término en un texto determinado (en la indización automática).

En la recuperación de información el papel que tiene un analizador morfológico es también importante, pues hace que las búsquedas sean más rápidas y eficientes para el usuario, ya que reconoce la raíz o base del término que busca, además el sistema muestra las palabras que tienen una raíz común y, a partir de este hecho, el usuario selecciona la palabra adecuada para continuar con la tarea de búsqueda y recuperación de información. Asimismo permite que el usuario conozca los términos que tienen un origen común mediante el reconocimiento automático de los fenómenos de flexión y derivación.

La segunda pregunta es la siguiente:

- ¿Cómo se representa en un tesoro el reconocimiento de reglas morfológicas en una palabra?

En un tesoro el reconocimiento de las reglas morfológicas de una palabra se representan a través de sus derivaciones. Veamos unos ejemplos:

Sufijos

MAP/

/AS

/EO

/OTECAS

Prefijos

/BIBLIO/

BIO/ ----- /GRAFÍAS

3.5.2. Analizadores sintácticos

A continuación se plantean dos preguntas con el propósito de señalar y analizar la relación que existe entre los analizadores sintácticos, la indización, la recuperación de información y el tesoro.

La primera pregunta es la siguiente:

- ¿Cuál es el papel que tiene un analizador sintáctico en la indización y la recuperación de información?

El papel que tiene un analizador sintáctico es principalmente la creación de patrones sintácticos o sintagmas de al menos dos tipos:

- Sustantivo + Adjetivo
- Sustantivo + de + Sustantivo

La segunda pregunta es la siguiente:

- ¿Cómo se presentan en un tesoro los patrones sintácticos señalados?

Ejemplos de patrones sintácticos que aparecen en un tesoro son los siguientes:

Ejemplo 1: Sustantivo + Adjetivo

EDICIÓN / *EDITION*

TG: (RECURSOS DE INFORMACIÓN)

TE: EDICIÓN ABREVIADA

EDICIÓN ANOTADA

EDICIÓN AUMENTADA

EDICIÓN BILINGÜE

EDICIÓN CLANDESTINA

Ejemplo 2: Sustantivo + de + Sustantivo

REGISTROS DE AUTORIDAD

TG: REGISTROS

TR: CATÁLOGOS DE AUTORIDAD

CONTROL DE AUTORIDAD

Los patrones sintácticos anteriores son reconocidos por los programas informáticos y son útiles para las tareas de indización; sin embargo, como se ha señalado, en dicho proceso interviene el ser humano que es quien al final del proceso de indización selecciona las frases o siglas que describirán el contenido del documento indizado.

En la recuperación de información el usuario escribe las frases o siglas que han sido indizadas. Dichas frases o siglas se integran y forman parte de un tesoro, de este modo, cuando el usuario las utilice para formular su estrategia de búsqueda la recuperación de información estará garantizada.

3.5.3. Analizadores semánticos

Los analizadores semánticos son los dispositivos informáticos que permiten el reconocimiento de diferentes tipos de relaciones que se presentan en un corpus lingüístico de determinado dominio o área del conocimiento. Como ya se ha señalado en los capítulos anteriores, los tesauros se construyen a partir de la colaboración interdisciplinaria y del trabajo sistemático que realiza un grupo de

trabajo. Las relaciones semánticas que se encuentran en un tesoro son las siguientes:

- Jerarquía o generalización
- Equivalencia
- Todo parte o meronimia
- Entidad-localización
- Autoría-entidad
- Entidad-proceso

Ejemplos de este tipo de relaciones que aparecen en un tesoro¹³⁷ son las siguientes:

- Jerarquía o generalización

DICCIONARIOS/

TG: OBRAS DE CONSULTA

TE: DICCIONARIOS BILINGÜES

DICCIONARIOS ELECTRÓNICOS

DICCIONARIOS ENCICLOPÉDICOS

DICCIONARIOS ESPECIALIZADOS

- Equivalencia

¹³⁷ Los ejemplos fueron tomados de Naumis Peña, Catalina, *et al* (1999). *Tesoro Latinoamericano en Ciencia Bibliotecológica y de la Información*. México: UNAM. Centro Universitario de Investigaciones Bibliotecológicas. p. 307.

PALABRAS CLAVE

TG: TÉRMINOS DE INDIZACIÓN

TE: DESCRIPTORES

ENCABEZAMIENTOS DE MATERIA

- Todo parte o meronimia

TECNOLOGÍA

TG: TECNOLOGÍA DE LA INFORMACIÓN Y DE LAS
COMUNICACIONES

TE: TECNOLOGÍA ADAPTADA

TECNOLOGÍA BIBLIOTECARIA

TECNOLOGÍA COMPUTACIONAL

TECNOLOGÍA DE LA INFORMACIÓN

TECNOLOGÍA DIGITAL

TECNOLOGÍA EDUCATIVA

TR: INFORMACIÓN TECNOLÓGICA

LITERATURA TÉCNICA

RECURSOS TECNOLÓGICOS

- Entidad-localización

ARCHIVOS (INSTITUCIÓN)

TG: UNIDADES DE INFORMACIÓN

TE: ARCHIVOS CENTRALES

ARCHIVOS NACIONALES

- Entidad-proceso

UNIDADES DE INFORMACIÓN

TG: ADMINISTRACIÓN Y GESTIÓN DE LA INFORMACIÓN

TE: ARCHIVOS (INSTITUCIÓN)

CENTROS DE ANÁLISIS DE LA INFORMACIÓN

CENTROS DE DISTRIBUCIÓN E INTERCAMBIO

CENTROS DE DOCUMENTACIÓN

CENTROS DE RECURSOS PARA EL APRENDIZAJE

El proceso de reconocimiento de las relaciones es complejo y de acuerdo con Sánchez Cuadrado *et al* (2004)¹³⁸ tiene dos propósitos fundamentales, por una parte, la organización de la información, y por otra, la organización del conocimiento. Para tales propósitos se han creado diferentes herramientas como los tesauros, las taxonomías y las ontologías. En el proceso de reconocimiento de las relaciones “se utilizan las tecnologías aplicadas para la construcción de tesauros y ontologías de forma automática en la que se emplean (*sic*) modelos mixtos de Procesamiento del Lenguaje Natural (PLN)¹³⁹ y teorías matemáticas

¹³⁸ Sánchez –Cuadrado, S. *et. al.* (2004). “Extracción automática de relaciones semánticas” en *Revista Iberoamericana de sistemas, cibernética e informática*. 1(1) p. 37.

¹³⁹Una definición del Procesamiento del Lenguaje Natural (PLN), es la proporciona el grupo de investigación en Procesamiento del Lenguaje y Sistemas de Información de la

de la comunicación aplicados al análisis automático de contenido.” Es a partir del reconocimiento de este tipo de relaciones que se estructura un tesoro. Una de las consecuencias del uso de los analizadores semánticos en la construcción de tesauros, es la normalización de los términos a partir de las relaciones que se establecen entre ellos. Otra es la que incide en la calidad de la información que recupera el usuario al momento de hacer una consulta.

3.6. Herramientas de la terminótica para la indización

Actualmente en los procesos de indización intervienen algunas herramientas de la terminótica tales como los vocabularios o glosarios y las bases terminológicas.

Universidad de Alicante, España en el año 1998 y señala que el PNL “es una parte esencial de la Inteligencia Artificial que investiga y formula mecanismos computacionalmente efectivos que faciliten la interrelación hombre-máquina y permitan una comunicación mucho más fluida y menos rígida que los lenguajes formales y sistemas de menús utilizados tradicionalmente. Todo sistema de Procesamiento del Lenguaje Natural intenta simular un comportamiento lingüístico humano; para ello debe tomar conciencia tanto de las estructuras propias del lenguaje, como de un conocimiento general acerca del universo de discurso. De esta forma, una persona que participe en un dialogo sabe cómo pueden combinar las palabras para formar una oración, conoce los significados de las mismas, sabe cómo éstas afectan el significado global de la oración y poseen un conocimiento del mundo en general que permite participar de la conversación. Así el problema principal del Lenguaje Natural se puede resumir en una palabra: Ambigüedad. El PLN tiene una serie de fases o niveles de análisis:

- Análisis morfológico-léxico: Transforma la secuencia de caracteres de entrada en una secuencia de unidades significativas haciendo uso del diccionario y reglas morfológicas.
- Análisis sintáctico: Analiza la secuencia de unidades léxicas y produce una representación de su estructura (árbol, red, ...).
- Análisis semántico: A partir de la estructura generada por el proceso sintáctico genera otra estructura o forma lógica asociada que representa el significado o sentido de la sentencia.
- Análisis contextual o función pragmática: Utiliza la forma lógica o estructura semántica de la fase anterior para desarrollar la interpretación final de la oración, en función de las circunstancias de contexto”. Consultado en junio 2013 en <http://gplsi.dlsi.ua.es/gplsi/areasf.htm>

A continuación se presenta la definición de estos dos tipos de herramientas, sus características y la relación que tienen con la indización y la recuperación de información.

3.6.1. Vocabularios o glosarios terminológicos en línea

Los vocabularios o glosarios terminológicos en línea son aquellas listas de términos que se encuentran al final de una obra especializada en algún área específica del conocimiento y sirven para explicar algunos términos técnicos que son difíciles de entender y que pertenecen a un oficio, profesión o jerga. Un ejemplo de este tipo de obras es el DigiDocIndex, diseñado por el Grupo de Investigación en Documentación Digital y Comunicación Interactiva del Instituto Universitario de Lingüística Aplicada de la Universidad Pompeu Fabra de Barcelona, cual se utiliza como términos normalizados en la indización y como palabras clave que facilitan la recuperación de información sobre un tema especializado.

3.6.2. Bases de datos terminológicas de léxicos especializados

Las bases de datos terminológicas son, de acuerdo con Montero Martínez y Faber Benítez (2008),¹⁴⁰ “terminologías especializadas en varias lenguas que están almacenadas en bases de datos de organismos internacionales y en algunas empresas multinacionales”. Dichas autoras señalan que existen tres generaciones de estas bases de datos: la primera es aquella en la que estaban

¹⁴⁰ Montero Martínez, S. y P. Faber Benítez (2008). *Terminología para traductores e intérpretes*. Granada: Ediciones TragaCanto. p. 153.

orientadas al término, eran productos término gráficos que se utilizaban principalmente en tareas de traducción, automática o manual.

La segunda generación de bases de datos terminológicas incorpora algunas ideas de estructuras a fin de conformar jerarquías en redes semánticas entre los términos. Un rasgo que distingue estas bases de datos es su orientación hacia el concepto y no hacia el término, como sucede en las de la primera generación.

Una definición de este tipo de bancos terminológicos es la que presentan Meyer *et al.* (1992)¹⁴¹ cuando señalan que “son repositorios terminológicos en los que los datos lingüísticos están enriquecidos con componentes de conocimiento formalizado que facilitan tareas compartidas y multifuncionales, además de aplicaciones diversas en la documentación, manejo de información, entrenamiento, escritura automatizada, sistemas expertos y traducción automática”. De la definición anterior, llama la atención que ese tipo de bancos terminológicos tienen aplicaciones en algunas tareas de documentación y manejo de información, dichas tareas pudieran estar relacionadas con la indexación y la recuperación de la información de forma directa y con la construcción de un tesoro de forma indirecta.

Los bancos terminológicos de la tercera generación son, según Montero Martínez y Faber Benítez (2008),¹⁴² “aquellos que presentan una estructura

¹⁴¹ Meyer, *et al* (1992). “Towards a New Generation of Terminological Resources: An Experiment in Building a terminological Knowledge Base” en *COLING-92: Proceedings of the 15^o International Conference on Computational Linguistics*. p. 129–138.

¹⁴² Montero Martínez, S y P. Faber Benítez. *Op Cit.* p. 157.

mucho más compleja, ya que la representación no incluyen sólo datos, sino posibles reglas que un sistema más desarrollado, por ejemplo, un sistema experto, o agente de inteligencia artificial podría ejecutar. Es decir se trata de solucionar problemas en un dominio determinado a través de las inferencias hechas a partir de una base de conocimiento”. Este tipo de bancos terminológicos se utilizan en tareas de traducción automática.

En este capítulo se identificaron, describieron y analizaron algunas aportaciones teóricas y metodológicas con las que la Informática contribuye a los procesos de indización y recuperación de la información en los que interviene el tesoro. Asimismo se describieron los pasos necesarios que conforma el procedimiento básico para la elaboración de un tesoro de acuerdo con la Informática.

De igual forma se describieron y valoraron los analizadores morfológicos, sintácticos y semánticos, herramientas informáticas que se utilizan en la indización y la recuperación de la información. Finalmente se describieron y valoraron los bancos y vocabularios terminológicos de léxicos especializados como herramientas de la terminótica que se utilizan en la indización y la recuperación de la información.

El estudio que se realizó promovió una reflexión sistemática por parte del autor de esta tesis, en la que se muestran las aportaciones que hace la Informática. A partir de dicha reflexión se puede decir que la colaboración que ha tenido la Informática para la elaboración de los tesoros, representa una paradoja.

Porque, por una parte, el uso de las herramientas, como los analizadores morfológicos, sintácticos y semánticos, o el uso de los bancos y vocabularios terminológicos ha permitido que la cantidad y rapidez en el tratamiento y análisis de datos sea cada vez mayor.

Y por otra, el trabajo del bibliotecólogo en el tema de la elaboración de los tesauros se oriente hacia la colaboración con otros campos del conocimiento, como la Terminología y la misma Informática. Finalmente, los avances de la Informática y otras disciplinas, ha promovido como se señala en 3.3.5 la creación de otras herramientas para recuperación de información y la organización del conocimiento como son las ontologías, que serán en el futuro el objeto de estudio de los bibliotecólogos

Capítulo 4

Lineamientos metodológicos para la elaboración de tesauros

*Nosotros estamos hechos de la misma materia que los sueños
Próspero. La tempestad. W. Shakespeare*

El propósito de este capítulo es exponer y analizar los lineamientos metodológicos derivados del estudio realizado en esta tesis. Dichos lineamientos idealmente servirán como una guía para los bibliotecarios que tengan a su cargo proyectos relacionados con la elaboración y uso de los tesauros. Los lineamientos son de dos tipos: los que se relacionan con la indización y los que se vinculan con la recuperación de la información. Además son de naturaleza diferente, ya que por una parte, se encuentran los que se derivan de las bases o fundamentos lingüísticos y terminológicos que se estudiaron en el segundo capítulo de esta tesis, y por otra, los que emanan de las bases o fundamentos de la informática que se estudiaron en el tercer capítulo de esta tesis.

4.1. Definición y alcance de los lineamientos

Un lineamiento se puede definir, de acuerdo con el *Diccionario del Español Usual de México* (2002),¹⁴³ como “cada uno de los pasos, reglas, objetivos, etcétera, que se establecen con el fin de organizar, regular o lograr el desarrollo de algo o alguien”. Para propósitos del desarrollo de este capítulo, se tomarán en cuenta algunos elementos de esta definición. Los lineamientos que se presentan a continuación son sólo algunos de los pasos y reglas que se pueden derivar del

¹⁴³ Lara Ramos, Luis Fernando (dir.) (2002). *Diccionario del Español Usual de México*. México: El Colegio de México. p.560.

estudio realizado sobre los procesos lingüísticos que están presentes en la elaboración de un tesoro. Así como de las herramientas terminológicas e informáticas que se utilizan para la indización y la recuperación de la información.

Los lineamientos que se presentan son de tipo metodológico, no tienen carácter exhaustivo y sólo pretenden orientar y regular el desarrollo de un tesoro, considerándolo como una herramienta que se utiliza para dos de las tareas bibliotecarias fundamentales, esto es, la indización y la recuperación de información. Con aplicación de estos lineamientos se hará evidente la inserción del enfoque lingüístico, terminológico e informático en la elaboración y uso de un tesoro.

Para la elaboración de los lineamientos se identificaron algunos problemas que impiden (o que no promueven) el uso y aplicación de nociones teóricas derivadas de la Lingüística, la Terminología y la Informática en el uso de un tesoro. A continuación se presentan ocho lineamientos metodológicos.

4.1.1. Procesos y reglas morfológicas

El proceso morfológico que está presente en la elaboración de un tesoro es la derivación, que se definió, analizó y ejemplificó en el apartado 2.2.1.1 de esta tesis. Del estudio realizado sobre dicho proceso se puede formular lo siguiente:

En la integración de nuevos términos que formen parte de un tesoro, deberán atenderse las normas de la derivación, como la

colocación de prefijos y sufijos en los términos, descriptores o no descriptores de un tesoro, con el propósito de mantener consistencia y solidez en la estructura del mismo.

4.1.2. Procesos y reglas para la creación de siglas

La integración de siglas es uno de los procesos que suceden cuando se crea un tesoro. De la información que se presentó en el apartado 2.2.1.2 de esta tesis se puede formular lo siguiente:

En la escritura de las siglas que forman parte de un tesoro se procurará, al menos, que se conserve el orden y la puntuación siguiente:

SIGLA: ORGANISMO O ENTIDAD.

4.1.3. Procesos y reglas sintácticas

El proceso y dispositivo sintáctico que está presente en la elaboración de un tesoro es el orden de las palabras. Mismo que se definió, analizó e ejemplificó en el apartado 2.3. de esta tesis.

Del estudio realizado se puede formular lo siguiente:

El orden canónico de las palabras en español es: (S) *sujeto*, (V) *verbo*, (O) *objeto*. En los tesauros, el orden de las palabras en una frase nominal tiene el siguiente esquema: *Sustantivo + Modificador*.

El conocimiento y aplicación de este dispositivo debe ser atendido durante la fase de construcción y mantenimiento de un tesoro, poniendo atención en la realización correcta de las frases nominales, mismas que deberán estar escritas de acuerdo con las reglas sintácticas de la lengua en la que esté escrito el tesoro. De esta forma, el tesoro tendrá consistencia interna y será una herramienta útil para el bibliotecario en la indización y para el usuario en la recuperación de información.

4.1.4. Los procesos semánticos

Los procesos semánticos están constituidos por las relaciones semánticas que se establecen entre los términos que conforman un tesoro, y que son fundamentalmente relaciones a partir de la sinonimia o semejanza de familia entre los términos. Se pueden distinguir relaciones de los siguientes tipos: todo–parte, proximidad, causa–efecto, género–especie, y sinonimia.

Los procesos semánticos son la dinámica que se establece entre las categorías de los términos, es decir, la red de relaciones que se establecen entre los términos genéricos (TG), términos específicos (TE) y términos relacionados (TR) que conforman un tesoro. Dicha red de relaciones dará consistencia y solidez al tesoro.

Las relaciones semánticas se describieron, analizaron y ejemplificaron en el apartado 2.4. de esta tesis, en el apartado 3.4.4 se muestra el procedimiento a través del cual se organiza el vocabulario en grupos de términos asociados con el apoyo de algunos programas de cómputo *ad hoc* y en 3.5.3 se presenta información sobre los analizadores semánticos, que son los dispositivos informáticos que permiten el reconocimiento de los diferentes tipos de relaciones (semánticas) que se suceden en un corpus lingüístico de determinado dominio o área de conocimiento.

Del estudio realizado se puede formular lo siguiente:

Las relaciones que se establecen entre los términos que conforman un tesoro deben constituirse a partir de la sinonimia o semejanza de familia. En dichas relaciones son, fundamentalmente, de tres tipos: asociación, equivalencia y jerarquía. Esta tipología de relaciones, más las notas de alcance de los términos y notas de tipo *útese o use por*, constituyen la forma normalizada que presentan las relaciones semánticas en un tesoro.

Una de las funciones que tiene las notas del tipo *útese* o *use por* en las relaciones de equivalencia, es el control de la sinonimia.

4.1.5. Procesos léxicos

Los procesos léxicos que se estudiaron en el apartado 2.6. de esta tesis, se analizaron y ejemplificaron las asociaciones, configuraciones y relaciones de tipo léxico. Las asociaciones lexicales son de tres tipos: heteronímicas, hiperonímicas y hiponímicas, además corresponden a dos tipos de configuraciones. La primera configuración tiene la forma “es parte de”, mientras que la segunda tiene la forma “es un tipo de”. A su vez, estas configuraciones corresponden a dos tipos de relaciones que se dan entre los términos, por una parte las relaciones partitivas o también llamadas meronímicas, y por otra, las relaciones género–especie.

Del estudio realizado se puede formular lo siguiente:

Las asociaciones, configuraciones y relaciones que conforman la estructura fundamental de un tesoro, deben establecerse tomando en cuenta las asociaciones semánticas que se presentan entre los términos y las diferentes formas de jerarquía entre los términos. Estas dos condiciones permitirán establecer equilibrio topológico y, por lo tanto, solidez en la estructura del tesoro.

4.1.6. Aspectos terminológicos

Este lineamiento se deriva del análisis realizado en el apartado 2.8 de esta tesis. Dicho análisis consistió en la presentación de tres teorías relevantes para el estudio de la terminología y su relación con dos tareas bibliotecarias, es decir, la indización y la recuperación de la información, tareas en las que está presente el tesoro.

Del análisis realizado se puede formular lo siguiente:

El tesoro guarda una relación estrecha con la terminología en la medida que se le considera una estructura de descriptores basada en relaciones que se establecen entre los conceptos de un documento. Dichos descriptores son identificados por el bibliotecario mediante la indización y son utilizados por el usuario cuando recupera la información del sistema. Los descriptores se expresan básicamente mediante unidades terminológicas. Por lo tanto en la elaboración de un tesoro deberá considerarse un tratamiento de los descriptores como unidades de lenguaje y como unidades de comunicación social.

4.1.7. Aspectos informáticos

Este lineamiento proviene del análisis realizado en esta tesis. Entre otras cuestiones, se estudió la relación que existe entre la informática y los tesoros, y cómo dicha relación se presenta, a través de las aportaciones teóricas y

metodológicas, en la recuperación de información y en la indización. Asimismo, se describió de manera general el procedimiento básico para la construcción de un tesoro a partir de la informática.

Del análisis realizado se puede formular lo siguiente:

En la elaboración de un tesoro, un principio procedente de la informática sostiene que deberán reconocerse grupos de conceptos que se expresen por medio de términos derivados del lenguaje natural y que compartan ciertas características. Asimismo, dicho principio, señala que las relaciones entre los conceptos deberán organizarse de manera jerárquica y que los términos deberán obtenerse del léxico especializado de la disciplina que trata el tesoro. Un procedimiento básico para la elaboración de un tesoro deberá observar el principio señalado y la observación de dicho principio tendrá como resultado un tesoro sólido en su estructura de relaciones entre los términos que lo integran.

4.1.8. Analizadores morfológicos, sintácticos y semánticos

Anteriormente se presentaron las características de los analizadores morfológicos, sintácticos y semánticos que se utilizan en la indización y en la recuperación de información, así como la relación que tienen dichas herramientas con los tesoros.

Del análisis realizado se puede formular lo siguiente:

Las herramientas informáticas como los analizadores, morfológicos, sintácticos y semánticos promueven el reconocimiento automático de estructuras morfológicas, de patrones sintácticos y relaciones semánticas en grandes corpus lingüísticos. Estos procesos de reconocimiento automático resultan relevantes en el uso de los tesauros, por el ahorro de tiempo que significa para el usuario en la recuperación de la información y para el bibliotecario durante el proceso de indización de los documentos.

Finalmente y para su mejor comprensión, los lineamientos descritos y enunciados en esta sección, se muestran de forma gráfica en la figura número 15 que a continuación se presenta:



Figura 15. Propuesta de lineamientos metodológicos para la elaboración de tesauros. Elaboración, Valdez, R. J. (2013).

5. Conclusiones

A partir de lo enunciado en esta tesis, se concluye lo siguiente:

Los objetivos de la tesis se alcanzaron. Por una parte se identificaron, analizaron y valoraron las aportaciones teóricas y metodológicas que han hecho la Lingüística (L), Informática (I), Terminología (T) para el diseño y elaboración de tesauros. Y por otra se propusieron y elaboraron lineamientos metodológicos necesarios para el diseño y elaboración de los tesauros, tomando en cuenta las aportaciones de la Lingüística, la Informática y la Terminología.

Se ha revisado y analizado la definición de tesoro desde tres visiones diferentes, primero desde la Lingüística, después desde la Informática y por último desde la Terminología, con el propósito de conocer y señalar la manera en que las disciplinas señaladas definen el tesoro. Asimismo, y a partir del análisis que se realizó del término tesoro en el campo de la Bibliotecología, se puede señalar que un tesoro tiene rasgos distintivos de tres tipos: formales, estructurales y funcionales.

Los tesauros son y han sido objeto de estudio de la Bibliotecología, al menos, por dos razones: porque se utilizan como instrumento para la organización del conocimiento y porque se utilizan como herramienta para la indización y la recuperación de la información.

Se puede decir que las asociaciones, relaciones y configuraciones léxicas que se han analizado en esta sección y que forman parte de la estructura de un tesoro son una evidencia de la colaboración interdisciplinaria que se da entre la Bibliotecología y la Lingüística en la tarea de la construcción de herramientas, como son los tesauros. La solidez en la construcción de las relaciones léxicas permitirá que el uso de los tesauros en tareas de indización y recuperación de información sea más eficiente.

El tesoro guarda una relación estrecha con la Terminología en la medida que es una estructura de descriptores basada en relaciones que presumiblemente se establece entre los conceptos de un documento, conceptos expresados básicamente mediante unidades terminológicas.

La relación que existe entre la Informática y los tesauros, puede evidenciarse, primero a través de las aportaciones teóricas y metodológicas con las que la Informática contribuye al proceso de indización y la recuperación de la información en la que interviene el tesoro. Segundo, mediante la descripción, de manera general, del procedimiento básico para la elaboración de un tesoro a partir de la informática.

Tercero, se describieron y valoraron algunas de las herramientas informáticas, como son los analizadores morfológicos, sintácticos y semánticos que se utilizan en la indización y la recuperación de la información. Y cuarto, se describieron y valoraron algunas de las herramientas terminológicas que se utilizan en la indización y la recuperación de la información.

Como producto final de esta tesis, se proponen ocho lineamientos metodológicos:

1. Procesos y reglas morfológicas
2. Procesos y reglas para la creación de siglas
3. Procesos y reglas sintácticas
4. Los procesos semánticos
5. Los procesos léxicos
6. Aspectos terminológicos
7. Aspectos informáticos
8. Analizadores morfológicos, sintácticos y semánticos

Los lineamientos expuestos consideran elementos de la Lingüística, la Terminología y la Informática que, desde nuestro punto de vista, ofrecen un método más completo para la construcción de tesauros.

Hasta el momento el bibliotecólogo se ha preocupado más por la revisión de las normas, y no por el enriquecimiento de la metodología para la creación de tesauros. A partir del rescate de las propuestas hechas desde otras disciplinas, es posible enriquecer y replantear la forma en que se diseñan los tesauros.

Para concluir esta tesis, resulta pertinente señalar una de las limitaciones de esta tesis es que el estudio se circunscribió al análisis detallado de las aportaciones y colaboración que hacen la lingüística, la informática y la terminología para la elaboración de tesauros y que a partir del análisis se propusieron una serie de lineamientos metodológicos para optimizar dicha tarea. A futuro y como una línea

de investigación que pudiera derivarse de esta investigación y como se señaló en la página 124, habrá que estar atento al desarrollo de herramientas para la organización del conocimiento, como son las ontologías. Entonces una de las tareas del bibliotecólogo, como es la elaboración de tesauros, sería mucho más precisa y orientada a la colaboración y trasvase de conocimientos que enriquezcan a la propia disciplina bibliotecológica y al usuario de herramientas informáticas para la indización y la recuperación de información.

6. Obras consultadas

- Aitchinson, J. ; Gilchirst, A. ; Bawden, D. (2000) *Thesaurus construction and use: a practical manual*. Chicago, Fitzory Dearborn Publishers.
- Alsina, V ; Estopa, R (1997) “Las profesiones y los usuarios de la terminología” en *Terminómetro : la Terminología en España*. Número especial N° 2, p. 85.
- Arano, Silvia (2003) La ontología: una zona de interacción entre la Lingüística y la Documentación [en línea]. *Hipertex.net*, número 2 <<http://www.hipertex.net>> [consulta 12 de abril 2010]
- Arano, Silivia (2005) Los tesauros y las ontologías en la Biblioteconomía y la Documentación [en línea]. *Hipertex.net*, número 3 <<http://www.hipertex.net>> [consulta 12 de abril 2010]
- Arntz, R ; Pitch, H. (1995) *Introducción a la terminología*. Madrid, Fundación Germán Sánchez Ruipérez, Pirámide.
- Batty, David (1989) “Thesaurus construction and manitenance: a survival kit” en *Database* 12(1) pp. 13 -20.
- Bauer, L. (1988). *Introducing Linguistic Morphology*. Edimburgo: Edinburg University Press.
- Betancourt Rodríguez, Osvaldo y Patricia Arias García (1987). “Diseño de un analizador semántico–sintáctico de un sistema automatizado de información” en *Actualidades de la información científico técnica*. (páginas 98–107). 18 (2) abril.
- Broughthon, Vanda (2006). *Essential thesaurus construction*. Londres: Facet Publishing.
- Cabral, Leonor (1973). *Introdução à lingüística*. Porto Alegre: Globo.

- Cabré Castellví, Teresa (2000). "Terminología y documentación" en *Documentación Terminología y Traducción*. (página 36). Madrid: Síntesis.
- Cabré Castellví, Teresa (2006). "Morfología y terminología" en *La Morfología a debate*. (página 131) Jaén: Universidad de Jaén.
- Cruse, D.A (1979). *Lexical Semantics*. Cambridge: CUP.
- Currás, Emilia (1991). *Thesaurus Lenguajes Terminológicos*. Madrid: Paraninfo.
- Diccionario básico de lingüística* (2005) México: UNAM, Instituto de Investigaciones Filológicas.
- Dictionary of Computing* (1994). 2ª ed. Cambridge: Peter Collins.
- Dijk Teun A. Van (1978). *La ciencia del texto. Un enfoque interdisciplinario*. Barcelona : Paidós.
- Dubois, Jean *et al.* (1998). *Diccionario de Lingüística*. Versión española de Inés Ortega y Antonio Domínguez. Madrid: Alianza.
- Durán Muñoz, Isabel. (2012) *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*. Frankfurt am Main: Peter Lang.
- Evens, Martha (2002). "Thesaural Relations in Information Retrieval" en Rebecca Green, Carol A. Bean y Sung Hyon Myaeng (eds). *The Semantics of Relationships. An Interdisciplinary Perspective*. Países Bajos: Kluwer Academic.
- Fernández Smith, Gerard (2007) *Modelos teóricos de la lingüística del texto*. Cádiz, España: Servicio de Publicaciones de la Universidad de Cádiz.
- Foster Allen ; Rafferty, Pauline (2011) *Innovations in Information Retrieval: perspectives for theory and practice*. London: Facet Publishing.
- Gil Leiva, Isidoro (2008). *Manual de indización. Teoría y práctica*. Madrid: Trea.

- Hutchins, W.J. (1978). *Language of indexing and classification. A linguistic study of structures and functions*. Reino Unido: Peter Peregrinus.
- Illingworth, Valerie (ed.). (1993). *Diccionario de Informática*. Madrid: Oxford University Press/Díaz de Santos.
- Lancaster, F.W (2002) *El control del vocabulario en la recuperación de información*. 2ª edición. Valencia, España: Universitat de Valencia.
- Lamarca Lapuente, María de Jesús. (sf). *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Consultado en mayo de 2008 en <http://www.hipertexto.info/documentos/tesauros.htm>
- Lara, Luis Fernando (2006). *Curso de Lexicología*. México: El Colegio de México.
- Lefevre, Philippe (2000). *La recherche d'informations, du texte integral au thesaurus*. Paris: Hermes Sciences.
- Lemone, Karen A. (1998). *Fundamentos de compiladores. Cómo traducir al lenguaje de computadora*. México: CECSA.
- Lewandowski, Theodor (1982) *Diccionario de lingüística*. Madrid: Cátedra.
- López Alonso, Miguel Ángel (1998). "La gestión automatizada de tesauros: estado de la cuestión" en: *Revista Interamericana de Bibliotecología* 21 (1) pp. 49 – 68.
- Martí, M.A. ; Llisterri, J (2002). *Tratamiento del lenguaje natural: Tecnología de la lengua oral y escrita*. Barcelona: Ediciones de la Univesitat de Barcelona
- Martínez García, S (2009) La representación y organización de la información a través de los tesauros. La interdisciplinariedad como nuevo paradigma: retos para la documentación y la bibliotecología, en *Contribuciones a las Ciencias Sociales*, octubre 2009. Consultado en mayo de 2013 en www.eumed.net/rev/cccss/06/smg.htm

- Meyer, Ingrid *et al.* (1992). "Towards a New Generation of Terminological Resources: An Experiment in Building a terminological Knowledge Base" en *COLING-92: Proceedings of the 15° International Conference on Computational Linguistics*. (páginas 129–138).
- Miller, Uri (1997) "Thesaurus Construction: problems and their roots" en *Information Processing & Management*. (páginas 481 – 493). 33:4.
- Mioto, Carlos. (2013) *Novo manual de sintaxe*. Sao Pablo: Contexto.
- Montero Martínez, S y Pamela Faber Benítez (2008). *Terminología para traductores e intérpretes*. Granada: Ediciones Tragacanto.
- Moreiro González, J.A. ; Méndez Rodríguez, E.M. (1999) "Lenguaje natural e indización automatizada" en *Ciencias de la Información*. 30 (3) pp. 11 – 24.
- Moreiro González, J. A. ; Marzal García – Quismondo, M.A. (2001) "Modelos teóricos y elementos funcionales para el análisis de contenido documental: definición y tendencias" en *Investigación Bibliotecológica*. 15 (31) pp. 125 – 162.
- Moreno de Alba, J. G. (1995). *Minucias del lenguaje*. México: FCE.
- Múnera Torres, M. T. (2001). "La terminología: un campo del saber transdisciplinario" en *Revista Interamericana de Bibliotecología* 24 (1) pp. 91 -115.
- Naumis Peña, Catalina, *et al.* (1999). *Tesaurus Latinoamericano en Ciencia Bibliotecológica y de la Información*. México: UNAM. Centro Universitario de Investigaciones Bibliotecológicas.
- Naumis Peña, Catalina (coordinadora) (2009). *Memoria del I Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología*. México: UNAM. Centro de Investigaciones Bibliotecológicas.

- Pinto María, *et. al.* (2005) *Aprendiendo a resumir. Prontuario y resolución de casos*. Gijón, Asturias, España: Treda.
- Pozzi, María. (2005) “La teoría comunicativa de la terminología” en *II Coloquio inter-universitario de terminología*. México: Veracruz (en prensa).
- Pozzi, M. (2003) “Hacia la armonización de la terminología usada en las normas del comité ISO/TC37: Identificación semi – automática de términos problemáticos en un corpus” en: *Estudios de Lingüística Aplicada*, 38 pp.109 -127.
- Pustejovsky, J. (2006) “Lexical semantics: Overview” en . Brown Keith (ed.). *Encyclopedia of Language & Linguistics*. 2a ed. Amsterdam: Elsevier. Vol 7 p. 98.
- Ribeiro Pinheiro, L. V. y J. M. Matheus Loureiro (1995). “Traçados e limites da Ciencia da Informação” en *Ciencia da Informação*. (páginas 42 -53) Brasilia: 24(1).
- Ruge, Greda. (1999) “Combining corpus linguistics and human memory models for automatic term association” en Tomek Strazalkowski (ed.) *Natural Language Information Retrieval*. (páginas 75 – 98). Países Bajos: Kluwer Academic.
- Ruiz Pérez, Rafael. (1992) *El análisis documental. Bases terminológicas, conceptualización y estructura operativa*. Granada: Universidad de Granada; Grupo de trabajo de información y documentación de la Comisión Nacional de España de la UNESCO.
- Salvador Olivan, José A. (2008) *Recuperación de la Información*. Buenos Aires: Alfagrama.

- Sager, Juan (1993). *Curso práctico sobre el procesamiento de la terminología*. Madrid: Fundación Germán Sánchez Ruipérez, Pirámide.
- Sánchez–Cuadrado, S. *et al.* (2004). “Extracción automática de relaciones semánticas” en *Revista Iberoamericana de sistemas, cibernética e informática*. (página 37). 1(1).
- Saracevick, T (1995). “Interdisciplinary nature of information science” en *Ciencia da Informaçao*. (páginas 36–41)Brasilia: 24(1).
- Seco, Manuel (1999). *Diccionario del Español Actual*. Madrid: Aguilar.
- Serrano Albuja, Alba (2010). *Cómo recuperar información en la biblioteca. Estrategia de indización con lenguajes combinatorios*. Venezuela: Gráficas Iberia.
- Sierra Martínez, G. (2003) Extracción automática de contextos definitorios en textos especializados en: Inchurralde, C. y Ibarretxe, I. (eds). *Memorias del XXI Congreso de la Sociedad Española para el procesamiento del lenguaje natural*. Universidad de Zaragoza. Zaragoza, España p. 351 – 352.
- Sparck Jones, Karen (1999). “What is the role of NLP in text retrieval?” en Tomek Strazalkowski (ed.) *Natural Language Information Retrieval*. (páginas 1–24) Países bajos: Kluwer Academic.
- Tálamo, M.F.G.M. ; Lara, M.L.G (2006). “O campo da Lingüística Documentária” en *TransInformaçõ* (páginas 203 -211) Campinas, 18(3)
- Thelen, M ; Frieda Steurs, (editores) (2001) *Terminology in everyday life*. Amsterdam: John Benjamins.
- Temmerman, Rita (2000). *Towards new ways of terminology description: the sociocognitive approach*. Amsterdam: John Benjamins.

- Trask, Robert Lawrence (1999) *Key concepts in Language and Linguistics*. London: Routledge.
- Virga y Mesters (1994). *Diccionario de Microinformática: Todas las nuevas tecnologías informáticas*. Madrid: Thompson / Paraninfo.
- Vizcaya Alonso, Dolores (1997) *Lenguajes documentarios*. Rosario, Argentina: Nuevo Paradigma.
- Werner, Abraham *et al.* (1981). *Diccionario de Terminología Lingüística Actual*. Madrid: Gredos.
- Zaldua Garoz, Alexei (2006). El análisis del discurso en la organización y representación de la información-conocimiento: elementos teóricos. *ACIMED* [online] 14 (3) p. 1-16.