



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



LICENCIATURA EN LENGUA Y LITERATURA HISPÁNICAS

U. N. A. M.
FACULTAD DE FILOSOFÍA Y LETRAS
Jefatura de la División del
Sistema Universidad Abierta

HACIA UNA OBTENCIÓN COMPUTARIZADA DE TÉRMINOS. (APLICACIÓN CONCRETA AL LÉXICO DE LA FÍSICA EN EL NIVEL BACHILLERATO)

Antonio Reyes Pérez

Tesis para optar por el grado de Licenciado en Lengua y Literatura Hispánicas

Asesor: Dr. Gerardo Sierra Martínez



Ciudad Universitaria 2002

TESIS CON FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



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.

PAGINACIÓN

DISCONTINUA

*Para mis padres
que siempre confiaron en mí;
y para mi familia y mi novia que siempre me han apoyado.*

*Una mención especial para todos y todas los que están y los que han entrado y
salido de mi vida dejando algo.*

Un agradecimiento especial para el Doctor Gerardo Sierra por apoyarme en éste y otros trabajos; a todos los integrantes del GIL por enriquecer a través de los seminarios y el trabajo diario esta tesis; también van mis gracias para la Doctora Ana María Cardero, la Maestra Margarita Palacios, la Maestra Adriana Ávila y la Maestra Alejandra Viguera por haber aceptado ser parte de este proyecto y haberlo mejorado con sus observaciones.

ÍNDICE

INTRODUCCIÓN

•	Antecedentes	i
•	Motivación: El Grupo de Ingeniería Lingüística (GIL) de la UNAM	iii
•	Objetivo	iv
•	Organización	iv
1.	FUNDAMENTOS BÁSICOS DE LA TERMINOLOGÍA	
1.1	La terminología	3
1.1.1	Definición de término y palabra	6
1.1.2	Tipos de términos	11
1.1.3	El área terminológica en nuestros días	13
1.2	Terminología y lingüística	15
1.3	La terminología como recurso de la ingeniería lingüística	18
2.	ESTUDIOS SOBRE EXTRACCIÓN TERMINOLÓGICA	
2.1	Extracción de terminología	21
2.2	Métodos para obtener términos de forma automática	25
2.2.1	Métodos estadísticos	26
2.2.2	Métodos basados en reglas	29
2.2.3	Métodos híbridos	31

2.3	Resultados importantes relacionados con la extracción automática de términos	33
3.	HERRAMIENTAS PARA EXTRAER TÉRMINOS A TRAVÉS DE PROCESOS AUTOMÁTICOS	
3.1	Procesos de análisis	36
3.1.1	Lista de palabras	37
3.1.2	Listas de detención	40
3.1.3	Palabras clave	43
3.1.4	Concordancia	45
3.1.5	Información mutua	47
3.2	Programas existentes	49
3.2.1	WordSmith	52
4.	OBTENCIÓN DE LA TERMINOLOGÍA	
4.1	Delimitación del objeto de estudio	55
4.1.1	Cómo se obtuvo el material en formato electrónico	56
4.1.2	El hipertexto	56
4.2	Listas de palabras	58
4.2.1	Procesos para la obtención de listas de palabras	58
4.2.2	Activación de la lista de detención	59
4.2.3	Pulido de las listas de palabras	60
4.3	Palabras clave	62
4.3.1	Elaboración de listas de palabras alternas	62
4.3.2	Comparación entre las diferentes listas de palabras	63
4.4	Índice de posibles términos simples	64
4.5	Concordancia	69
4.6	Análisis de palabras clave a través de agrupamientos y colocaciones	70
4.6.1	Terminología poliléxica a partir de listas de palabras múltiples e información mutua	73

5. RESULTADOS	
5.1 Presentación de los resultados	79
5.2 Validación de los resultados	83
6. CONSIDERACIONES FINALES	
6.1 Conclusiones	86
6.2 Trabajos futuros	90
BIBLIOGRAFÍA	92

INTRODUCCIÓN

- **Antecedentes**

El trabajo que enseguida se presenta es el resultado de la investigación realizada durante la estancia como becario del Grupo de Ingeniería Lingüística (GIL), dentro del Instituto de Ingeniería de la UNAM. Entre las distintas actividades y proyectos que se trabajan al interior del GIL, está un proyecto destinado a elaborar un diccionario onomasiológico en el área de física como material de apoyo para los estudiantes que cursan el nivel bachillerato. De dicho proyecto se desprende este trabajo, cuyo primer objetivo es el de entregar al GIL la metodología necesaria para extraer términos de forma automática y comprobar su eficiencia al extraer los términos concernientes a la subárea de mecánica. El objetivo particular que se busca conseguir con este trabajo es el de presentar la investigación realizada para cumplir con el objetivo del proyecto del GIL, pero dentro de un proyecto de tesis para obtener el grado de Licenciado en *Lengua y literatura hispánicas*; carrera que a través de su vertiente lingüística dota al alumno de conocimientos que le permiten ampliar la visión y el estudio de problemáticas que lo atañen directa e indirectamente en su relación consigo mismo y su entorno.

La información que a continuación se describe representa el desarrollo de la investigación y exposición del trabajo realizado para conseguir este objetivo particular.

El trabajo se titula: *HACIA UNA OBTENCIÓN COMPUTARIZADA DE TÉRMINOS. (APLICACIÓN CONCRETA AL LÉXICO DE LA FÍSICA EN EL NIVEL BACHILLERATO)*¹; tiene este título porque, como a lo largo del trabajo se explicará, tanto los análisis como las extracciones son tareas que hasta el momento son efectuadas de forma semiautomática, y que se pretende en un futuro, sean de carácter cien por ciento

¹ De aquí en adelante se hablará de extracción automática y semiautomática en lugar de computarizada

automático. El material que compone este trabajo está pensado y dividido implícitamente en tres funciones. En la primera se establecen los fundamentos básicos del tema que constituye la materia prima del trabajo, la terminología, y por consiguiente, los términos. La segunda función describe, una vez establecido lo que se pretende extraer, cómo la terminología es usada en apoyo a la ingeniería lingüística por medio de la aplicación de herramientas aprovechadas en programas de cómputo para extraer términos. La última función detalla el procedimiento que se llevó a cabo para obtener la terminología y los resultados obtenidos.

Cabe señalar que este trabajo, como se detallará en el capítulo 1, se engloba dentro de la disciplina de la ingeniería lingüística, de la cual en su momento se hablará. Esta interacción de disciplinas implica que la investigación no está ceñida exclusivamente a la lingüística, sino que participa de materias interdisciplinarias como la estadística, la ingeniería en computación y otras, que si bien no se detallan de forma profunda por cuestiones obvias de tiempo, espacio y necesidades, sí se describen los procedimientos básicos para lograr entender el desarrollo de los capítulos.

Es necesario hacer notar que los trabajos e investigaciones que se desarrollan dentro de ingeniería lingüística no son nuevos, ya que desde la segunda mitad del siglo pasado se comenzó a trabajar en cuestiones computacionales aplicadas a la explotación de la lengua, tal es el caso de los estudios sobre *recuperación de información, traducción automática y asistida por computadora, enseñanza de lenguas extranjeras*, etc. Las investigaciones y proyectos que tienen relación estrecha o indirecta con la explotación de la extracción automática de terminología² han sido numerosos también, tanto más en Europa y Estados Unidos que en América Latina; sin embargo, las investigaciones continúan y el desarrollo de esta área, así como de otras que tienen una interrelación muy estrecha, siguen en constante aumento.

La investigación que aquí se desarrolla es una muestra de la preocupación por distintos grupos de investigación interdisciplinaria en el desarrollo de técnicas, recursos y

² En un apartado del capítulo 2 se hace mención de algunos proyectos relacionados con la extracción automática de términos.

sistemas para extraer términos, no porque ésta tarea sea la panacea, sino porque resulta, además de una disciplina de vasta explotación, una herramienta bastante útil para el desarrollo y apoyo de otras áreas de investigación que no necesariamente tienen que ver con la lingüística.

- **Motivación: El Grupo de Ingeniería Lingüística (GIL) de la UNAM**

La presentación de este trabajo como material de tesis obedece al gusto e interés por dar impulso y, en cierto sentido (aunque mínimo), desarrollar un área de gran utilidad y provecho para el GIL, la extracción automática de terminología.

Esta tarea terminológica cumple un papel muy importante al interior del GIL, ahí se desarrollan sistemas, metodologías, recursos y aplicaciones que explotan los recursos que representa el aprovechamiento de la lingüística y sus distintas ramas de forma interdisciplinaria y en beneficio de investigadores, profesionales, estudiantes y público en general.

De la estancia en el GIL y de la colaboración para cumplir con los diversos proyectos que ahí se desarrollan se desprende este material, el cual responde a las necesidades que en su momento se tuvieron para desarrollar la fase respectiva a la obtención de la terminología que compondría el diccionario onomasiológico de física.

Para realizar esta investigación se contó con la infraestructura presente en el GIL, entendiéndose por infraestructura, tanto las instalaciones como el material bibliográfico, el programa WordSmith para extraer la terminología, la interacción con los diversos grupos interdisciplinarios que de forma directa o indirecta han y tienen que ver con la ingeniería lingüística, y el apoyo necesario para desarrollar este trabajo.

- **Objetivo**

El objetivo particular que se persigue con este trabajo (presentar esta investigación como proyecto de tesis), es el de estudiar la terminología “como teoría especializada en el análisis de vocabularios técnicos y científicos”³ para aplicarla como herramienta de la ingeniería lingüística, y aprovechar los usos que como herramienta de apoyo representa en un trabajo de extracción automática.

De la aplicación de la terminología como recurso de la ingeniería lingüística se pretende obtener, a través de la utilización de un programa computacional, los términos relativos a la rama de la física encargada de la *mecánica*.

En relación con los términos que se pretenden extraer, es necesario dejar muy en claro que éstos se restringen únicamente al nivel bachillerato y al material proporcionado por el Centro de Instrumentos, llamado *hipertexto*; de forma que cualquier término que no aparezca en él o que pertenezca a un nivel superior al bachillerato queda completamente justificado en el entendido de que la extracción está delimitada sólo al hipertexto y al nivel bachillerato.

- **Organización**

El desarrollo que este material sigue, responde a las necesidades del objeto de estudio, esto es, que mediante la aplicación y explotación de una herramienta computacional, WordSmith, se pretende extraer los términos relativos a mecánica pertenecientes al área de la física. Para complementar la información y dar un seguimiento más preciso a los temas que se tocarán en la investigación, el trabajo se encuentra dividido en cinco capítulos, los cuales a su vez cumplen tres funciones distintas:

³ Véase Ana María Cardero. El procesamiento de una terminología. Referencia especial a la terminología de control de satélites en el área de las telecomunicaciones en México. Tesis de doctorado. México: 2001.

1. describir los fundamentos básicos de la materia que se encarga del estudio de las unidades que se pretenden extraer, es decir, delinear los principios de la terminología;
2. explicar el uso de la terminología como recurso de la ingeniería lingüística a través de las herramientas diseñadas para ejecutar cuestiones de extracción terminológica y la aplicación de éstas a sistemas de cómputo;
3. aplicar las herramientas de extracción terminológica presentes en WordSmith para obtener la terminología requerida.

Estas tres funciones son complementarias y se desarrollan en los cinco capítulos de la siguiente forma.

En el capítulo 1 se delinear los fundamentos básicos de la terminología y los vínculos que relacionan a ésta con la lingüística; se detallan además las propiedades de la materia de estudio de la terminología, los términos, a fin de dejar en claro que lo que en esta investigación se pretende no es extraer léxico o información o algún otro rasgo lingüístico, sino que se delimita exclusivamente a los términos.

Se pretende que al final de este capítulo queden despejados los siguientes puntos:

1. qué es la terminología,
2. qué diferencias hay entre términos y palabras,
3. cuáles son las características y tipos de términos,
4. establecer los vínculos entre terminología y lingüística, y
5. dejar en claro cómo la terminología es aplicada como recurso de la ingeniería lingüística en este trabajo.

En el capítulo 2 se habla sobre la aplicación de la terminología como recurso de la ingeniería lingüística, esto es, se habla ya no sobre terminología desde un enfoque lingüístico, sino que se habla sobre extracción automática de terminología.

Está contemplado en este apartado especificar qué es la extracción automática o semiautomática de términos, qué métodos existen para extraer terminología y cómo funcionan. Para finalizar el capítulo 2 se mencionan distintos proyectos concluidos o por concluir que han trabajado ya sea de forma directa o indirecta con cuestiones de extracción terminológica.

El capítulo 3 hila el capítulo anterior. Se retoman las metodologías para extraer términos, pero aplicadas a sistemas de cómputo encargados de hacer automática la tarea extractiva.

En especial, se habla de las herramientas que presenta el programa WordSmith, con el cual fueron efectuados los análisis para obtener la terminología. No se deja de mencionar en un apartado especial las características que este programa presenta; además, se hace mención de algunos programas existentes en el mercado (aunque no se entra en un análisis sobre ellos) que realizan funciones de extracción de términos y cuentan con herramientas como las que posee WordSmith.

De esta forma, una vez finalizado este capítulo se tendrá en claro lo siguiente:

1. qué es una lista de palabras,
2. qué son y para qué sirven las listas de detención,
3. qué son y cómo se obtienen las palabras clave,
4. qué es un análisis de concordancia y para qué sirve,
5. qué es la información mutua y qué representa un análisis por medio de esta herramienta.

En el capítulo 4 se ponen en práctica las herramientas descritas en el 3 por medio del programa WordSmith. Este capítulo es la parte práctica de la teoría descrita en los tres primeros, ya que a lo largo de este capítulo se describen todos los procesos y análisis efectuados para extraer la terminología propuesta.

Para conformar este capítulo fue necesario realizar, con WordSmith, los análisis que debían arrojar la terminología pretendida. Se detalla desde la obtención del material, cedido por el Centro de Instrumentos, el cual mediante los análisis del programa proporcionaría la terminología deseada. Se describen los procesos y análisis llevados a cabo, tal y como son la elaboración de listas de palabras y de detención, los procesos de búsqueda de palabras clave, los análisis de concordancia, además de los hechos por medio de información mutua que dieron como resultado obtener un total de términos propuestos para conformar el bloque dedicado a mecánica.

El capítulo 5 y último consta de la presentación de los términos obtenidos en el capítulo anterior a la gente del Centro de Instrumentos⁴ para que evaluaran y avalaran la terminología derivada del trabajo.

Se finaliza este capítulo y el material de que consta este trabajo, con las observaciones y evaluaciones que los del Centro de Instrumentos sugieren a los resultados logrados en este trabajo.

De esta forma, las tres funciones implícitas de este trabajo quedan enmarcadas de la siguiente forma:

- a) la función primera dedicada a dar los fundamentos básicos de la terminología corresponde al capítulo 1,
- b) los capítulos 2 y 3 afectan a la segunda función, diseñada para establecer cómo la terminología es usada en esta investigación como herramienta de apoyo a la ingeniería lingüística,
- c) la última función, dedicada a la obtención de los términos, equivale a los capítulos 4 y 5.

Finalmente, se proporcionan las conclusiones derivadas de este trabajo, así como los trabajos futuros que se desprenden de este material...

⁴ Por razones de necesidad (había que obtener una respuesta pronta de parte de gente especializada no tanto en terminología, sino en el área de física, de los resultados conseguidos para no detener el proyecto principal), y luego de tiempo (en el tiempo y espacio disponibles para realizar una tesis de licenciatura no es posible abarcar todo lo que uno quisiera), no existió la posibilidad de evaluar la terminología obtenida por medio de reglas lingüísticas.

1. FUNDAMENTOS BÁSICOS DE LA TERMINOLOGÍA

1.1 La terminología

La investigación que a continuación se desarrolla tiene por objeto presentar una terminología concerniente al léxico empleado en la enseñanza de la física a nivel bachillerato. Para lograr este fin, a lo largo de los siguientes cinco capítulos se trazan los principios básicos de lo que es la materia de estudio de este trabajo: terminología, y por tanto, términos; así también, se describen las herramientas y algunos programas existentes sobre la extracción de términos, además de los procesos llevados a cabo para obtener la terminología de la que más arriba se habla.

De igual forma, en el desarrollo de este trabajo, y puesto que esta es una investigación interdisciplinaria debido a que se tratará la terminología no desde un punto estrictamente lingüístico, sino como una herramienta de apoyo para la ingeniería lingüística, se abordarán temas que tienen que ver con la terminótica, esto es, el trabajo terminológico mediante herramientas informáticas, como lo son los métodos existentes para extraer términos vía una computadora, las herramientas que presentan los distintos programas creados para trabajar con cuestiones de extracción de términos, recuperación de información, etc., sin dejar de mencionar algunos proyectos que han trabajado o trabajan con extracción de terminología, además de mencionar algunos programas que existen sobre el tema. Por último, luego de describir los procesos efectuados para obtener la terminología pretendida, en el capítulo 5 se pondrá de manifiesto la evaluación que los expertos en física dan sobre los términos presentados.

Para iniciar con esta investigación, el primer paso es describir en qué consiste y cuál es el objeto de estudio de la terminología.

Ésta es una disciplina que se ocupa de un aspecto importante de la lengua, el léxico, mas en este caso, el léxico especializado, y es, según Cabré, una disciplina que se divide en tres vertientes diferentes¹:

1. una vertiente que considera la terminología como una disciplina “autónoma, de carácter interdisciplinario, al servicio de las disciplinas científico – técnicas.”;
2. otra vertiente plantea la terminología a través de un aspecto filosófico, el cual se interesa por la “categorización lógica de los sistemas de conceptos y la organización del conocimiento.”;
3. y una tercera vertiente que se enfoca hacia el aspecto lingüístico de la terminología, el cual es el que nos ocupa y que considera la terminología como un “subcomponente léxico de la lengua, y los lenguajes de especialidad como subsistemas de la lengua.”

Tales vertientes se detallaran más adelante en el apartado dedicado a la *terminología y lingüística*.

Por otra parte, es necesario destacar que existen, respecto a la terminología, distintos enfoques que tratan de limitar su campo de estudio y la vinculación que con las diferentes disciplinas pudiese tener. Hay quienes plantean la terminología como una práctica ligada a la simple resolución de necesidades sociales del hombre; algunos otros más la consideran como una disciplina científica, autónoma y deudora de otras especialidades. Los más adoptan una posición entre los dos extremos, más neutral para con la terminología.

Por ejemplo, hay quienes definen las terminologías como vocabularios especiales, los cuales son la llave para los especialistas en sus procesos comunicativos, donde el uso de estas terminologías hace que su comunicación sea más eficaz. En cambio, hay quienes juzgan las terminologías como una jerga por demás incomprensible para el público. Baste como ejemplo mencionar las pláticas entre médicos.

¹ María Teresa Cabré. La terminología. Teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993. p. 32.

No deja de causar inquietud en esta investigación la interrogante en torno a la utilidad de las terminologías. ¿Por qué han surgido? Una respuesta muy convincente la presenta Christian Galinsky, quien argumenta que “la necesidad de aportar mayor precisión al pensamiento humano y a la misma comunicación fue lo que llevó al hombre a la creación de términos como representantes de conceptos claramente definidos”²; esto es, quizás de forma burda, etiquetar el mundo, encasillar todo lo que en él hay dentro de distintos márgenes, “cada uno con su lenguaje correspondiente”³, como lo dice Sager; para permitir un mejor nivel de comunicación y entendimiento, sin prestarse a ambigüedades.

A fin de cuentas, estas ideas podrían resumirse como competencia lingüística, que no es otra cosa que los conocimientos que un individuo tiene sobre una lengua; y en un punto en que esta competencia lingüística es llevada a tal grado de especialización como lo son las terminologías, resulta de gran utilidad y provecho para aquéllas que tienen su campo de estudio y trabajo en el uso diario de términos que deben representar sólo un concepto, objetos de sus desempeños laborales e intelectuales diarios, tal y como lo señala Cardero cuando apunta que “una terminología en su concepción ideal debe tratar de incluir todos los términos que empleen los profesionales de la ciencia o de la técnica de que se trate en una lengua”⁴ para facilitar la labor de éstos y propiciar un nivel vasto de entendimiento.

También es necesario aclarar que la terminología se ocupa de la designación de los conceptos de las lenguas de especialidad, y en cuanto a su estudio, Cabré hace hincapié en que deben tomarse en cuenta cuatro puntos a considerar sobre los diferentes enfoques desde los cuales puede estudiarse la terminología⁵:

1. enfoque lingüístico; donde la “terminología es una parte del léxico especializado por criterios temáticos y pragmáticos.”

² Christian Galinsky. *Terminología y sociedad moderna: el papel de INFOTERM. Las industrias de la lengua*. Madrid: Fundación Germán Sánchez Ruipérez, 1991. p. 243.

³ Juan Carlos Sager. *La terminología, puente entre varios mundos*. Op. Cit. pp. 12 – 13.

⁴ Ana María Cardero. *La integración del corpus de la terminología de control de satélites en México. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo*. Tomo II. México, (1996) p. 106.

⁵ María Teresa Cabré. *Op. Cit.* p. 37

2. Enfoque de los especialistas; donde la “terminología es el reflejo formal de la organización conceptual de una especialidad”, y por consecuencia, un medio por el cual debe de fluir la comunicación en el ámbito profesional.
3. Enfoque de los usuarios; para quienes la “terminología es un conjunto de unidades de comunicación.”
4. Enfoque de los planificadores lingüísticos; quienes ven a la terminología como un “ámbito del lenguaje donde se debe intervenir para reafirmar la existencia, la utilidad y la pervivencia de la lengua.”

Tomando en cuenta estos enfoques, y de acuerdo con las conveniencias de esta investigación, se optó por hacer uso de la terminología vista desde el enfoque lingüístico, del cual se hablará más adelante.

Ahora bien, una vez sentadas las bases de la terminología, es preciso dedicar un apartado a la explicación del objeto de estudio de la misma, los términos; y ahondar en lo posible sobre ella para, en la medida de nuestras posibilidades y de las necesidades de este trabajo, dejar en claro la disciplina que se ocupa de la materia prima de la actual investigación.

1.1.1 Definición de término y palabra

Si bien tanto la terminología como la lexicología tienen un campo de estudio en común, las palabras, tienen también la característica de que en la segunda el único interés es la palabra como tal, en tanto que en la primera, lo es el término. Pero ¿cuál es la diferencia que existe entre las palabras y los términos, en dónde empieza uno y acaba otra, cuáles son sus características para diferenciarlos?

Desde una perspectiva lingüística, y según palabras de Alicia Fedor, “la palabra es un símbolo lingüístico que admite matices semánticos y depende del contexto; el término

acusa un grado de precisión mucho más elevado y pertenece a un sistema de conceptos determinado.”⁶

En un grado muy similar de apreciación, José Vega apunta que “de un punto de vista lingüístico, los términos no son diferentes a las palabras (simples o compuestas) de la lengua general (tienen una categoría léxica, una representación morfológica, etc.). Sin embargo, dichos términos están en relación con otros objetos conceptuales que no están descritos en los modelos lexicográficos, como por ejemplo: los dominios, las definiciones, los contextos...”⁷.

Por otro lado, Cabré apunta que la diferencia entre palabra y término estriba en que “una palabra es una unidad descrita por un conjunto de características lingüísticas sistemáticas y dotada de la propiedad de referirse a un elemento de la realidad”⁸; mientras que un término es, “una unidad de características lingüísticas parecidas, utilizada en un dominio de especialidad. [Por tanto] una palabra que forme parte de un ámbito especializado sería un término”⁹.

Por su parte, Kulebakin y Klimovitskii, miembros de la *Escuela Rusa*, sugieren que un término es una palabra o combinación de ellas cuyo símbolo lingüístico se relaciona con el concepto que representa, esto es, simboliza su significado y su significante; unívoco dentro de un área específica y delimitada¹⁰.

Una característica por demás importante sobre las peculiaridades de un término la vierte otra rusa, L. Kutina, para quien un término debe ser monosémico y carente de sinonimia¹¹.

⁶ Alicia Fedor de Diego. *Terminología teoría y práctica*. Venezuela: Universidad Simón Bolívar / Unión Latina, 1995. p. 54 –55.

⁷ José Vega. *Herramientas lingüísticas para la terminología*. *Actas del IV simposio iberoamericano de terminología RI Term. Terminología y desarrollo*. Tomo II, Argentina: (1994). p. 26.

⁸ María Teresa Cabré. *Op. Cit.* p. 87.

⁹ Ídem. p. 87.

¹⁰ Ver V. S. Kulebakin y A. Klimovitskii. *Los trabajos para la elaboración de la terminología científico – técnica en la URSS y la Escuela de Terminología Soviética*. *Textos de terminólogos de la escuela rusa*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 61 - 87.

¹¹ Ver V. S. Kulebakin y A. Klimovitskii. *Op. Cit.* pp. 61 - 87.

Ahora bien, según análisis de las diferentes definiciones que se tomaron de diversos terminólogos, tenemos que en esencia una palabra es aquella forma que tiene un aspecto fónico y gráfico, posee una estructura morfológica, gramatical y de significación. Un término, visto también lingüísticamente, es otra forma con las mismas características, sólo que restringido a un lenguaje de especialidad. De aquí se deduce que las palabras se centran al total del repertorio léxico, exceptuando al de especialidad.

Mas ahora surge una interrogante muy fuerte, ¿qué hay con aquellas palabras que pueden entrar dentro de la categoría tanto de palabras como de términos, o cómo diferenciar a los “términos [que] participan en varios campos”¹² de conocimiento? Este problema es al que Cardero se refería cuando hablaba del “comportamiento semántico de los términos, las creaciones neológicas, los problemas de normalización, los problemas sociolingüísticos y pragmáticos”¹³, que nos enfrentan al problema de la polisemia.

Los términos suelen tener un significado, solamente uno, e inequívoco, como propone Kutina, y estos términos se describen en el interior de un campo de especialidad concreto.

Un ejemplo es la palabra *síntesis*, la cual presenta un aspecto polisémico, puesto que su significado depende de los contextos en los que aparezca: si la situamos dentro del campo de la filosofía, su significado nos remitirá a operaciones mentales; si en cambio, se recurre a ella en un ámbito psicológico, representará un significado de conjunción de elementos psíquicos; y si es usada en química referirá su significado hacia la formación de sustancias.

Este argumento muestra que una palabra polisémica puede ser, dependiendo del sentido en que es utilizada, un elemento específico y diferenciador. En este caso, *síntesis* entra dentro de campos de especialización que le adjudican su carácter de término; pero si

¹² Ana María Cardero. Op. Cit. p. 110.

¹³ Ídem. p. 106.

esta misma unidad es usada en un contexto diferente, en donde por el carácter semántico de su utilización no sea aplicado como elemento de especialización, pierde la categoría de término, quedando sólo como palabra.

Dicho de forma más lingüística, una palabra polisémica que no designe en un nivel pragmático sólo un concepto, coincide únicamente con el significante, pero los significados pueden ser distintos; en tanto las palabras polisémicas o polifuncionales, como las denomina Danilenko, que en su uso cumplen con una designación específica a un sólo concepto, convergen tanto en significado como en significante¹⁴.

Pero no es tan fácil aceptar estos postulados; por ejemplo, Alpizar señala en *El léxico de la terminología. Algunas precisiones*, que:

Hay entre los teóricos de la terminología una corriente, que afirma que siempre los términos son monosemánticos, por lo que las diversas acepciones de un término no se analizan como casos de polisemia, sino como manifestaciones de homonimia, a partir del criterio de que las acepciones distintas responden a significados pertenecientes a especialidades diferentes, como es el ejemplo de *virus* en informática y *virus* en biología (dos unidades homónimas, cada una de las cuales es monosémica), pero que, en cada especialidad por separado el término es monosémico.¹⁵

No obstante que esa es sólo una posición entre muchas, no deja de causar inquietud o asombro. Por ejemplo, las ideas de García Palacios, quien plantea que “dentro de lo posible está también que en el fondo no nos encontremos ni ante palabras ni ante

¹⁴ Véase V. P. Danilenko. *Sobre la posición de la terminología científica en el sistema léxico de una lengua. Textos de terminólogos de la escuela rusa*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 19 – 28.

¹⁵ Rodolfo Alpizar Castillo. *El léxico de la terminología. Algunas precisiones. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo*. Tomo II, Argentina: (1994). p. 79.

términos, sino ante una especie de metalenguaje utilizado para referirnos a los nudos clasificatorios.”¹⁶

Sin embargo, para diferenciar entre estas unidades, Cabré señala que existe otro aspecto que permite diferenciar entre términos y palabras: el sentido pragmático, el cual determina la diferencia entre término y palabra a partir de cuatro puntos¹⁷:

1. Usuarios
2. Situaciones de uso
3. Temática
4. Contextos (discursos)

En cada uno de estos puntos tanto término como palabra son usados para referirse a asuntos diferentes; por ejemplo, así como los términos son usados por gente especializada o dentro de un contexto de especialidad, las palabras son usadas a un nivel más general en donde los usuarios no necesitan comunicarse con unidades inequívocas. Respecto a las situaciones de uso, un término se ciñe más a situaciones científicas que a coloquiales, que son más propias de las palabras. En cuanto a la temática, los términos cumplen una función muy específica que no cumplen las palabras; por último, en cuanto a los contextos, tanto términos como palabras, son utilizados(as) con una categoría particular que responde a las necesidades del discurso. Todo esto no excluye el que términos y palabras interactúen en cada uno en un sentido específico, en los puntos que señala Cabré.

El uso tanto de términos como de palabras en cada uno de los sentidos pragmáticos de que habla Cabré no es aleatorio; cada cual cumple una función comunicativa específica a diferente nivel, y aunque sea difícil diferenciarlos, existen rasgos y características (mencionados en este y el siguiente apartado) que permiten de algún modo, y con cierta seguridad, distinguirlos.

¹⁶ Joaquín García Palacios. *La terminología en los manuales de enseñanza media: hacia la determinación de la terminología básica del español. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología*. México, (1996). p. 153.

¹⁷ María Teresa Cabré. Op. Cit. pp. 88 – 89.

1.1.2 Tipos de términos

Uno de los resultados que se esperan obtener con este trabajo es la extracción de una lista de términos en el área de física; cabe destacar que en esta área, como en la gran mayoría de ellas, los términos no sólo se ven reflejados como unidades simples, acaso ni siquiera como unidades, sino, como siglas, como unidades compuestas por dos o más unidades, e incluso, por frases. Por tal motivo es menester señalar que existen diferentes tipos de términos, y dejar en claro en qué consisten sus diferencias

Cabré presupone tres diferentes enfoques de clasificación de términos: *Formal, funcional y de significado*¹⁸.

1. En cuanto a la forma, los términos siguen diversos criterios de clasificación:

a) Por el número de morfemas que lo conforman; en donde pueden ser términos simples o complejos.

gel *gélido*
radio *radioscopia*

b) Por el tipo de morfemas de que están constituidos los términos complejos. Éstos pueden ser “derivados (formados por la agregación de afijos a bases léxicas) y compuestos (formados por combinaciones de bases léxicas. Actuales o históricas, con la posibilidad de agregar posteriormente afijos a las mismas)”¹⁹.

pisapapeles
limpiabotas
microfilm *microfilmación*
dermatología *dermatológico*

c) En relación con los términos complejos, éstos pueden seguir estructuras sintácticas definidas (también conocidos como locuciones en los trabajos de terminología):

hora punta

¹⁸ Todos los términos para ejemplificar las diferencias entre los tipos de término fueron tomados de la obra de Cabré: La terminología. Teoría, metodología y aplicaciones. 1993.

¹⁹ María Teresa Cabré. Op. Cit. p. 176.

impuesto sobre la renta

d) Términos simples conformados por procesos de truncación:

- siglas (*ONU, UNAM, OTAN*),
- acrónimos (*Agrimed* = agricultura mediterránea, *informática* = información automática)²⁰,
- abreviaturas (*Etc., Excmo.*),
- formas abreviadas (*Auto* = automóvil, *Tele* = televisión)

2. En cuanto a la función, los términos pueden clasificarse en: *nombres, adjetivos, verbos y adverbios*.
3. En cuanto al significado, en donde la clasificación de los términos depende del concepto que denomina.

Rosa Estopà trabaja su clasificación de términos, o unidades terminológicas (UT) como ella les denomina, en unidades monoléxicas o poliléxicas, en cuanto a estructura; en cuanto a la forma, clasifica a “las unidades monoléxicas... en simples (*memoria, axón, hernia, dolor, mano*), derivadas (*drenaje, tratamiento, inyección, operación, diagnorático*), compuestas (*apendicitis, anorexia, radiografía..*) o siglas (*PET, TAC...*)”²¹.

Están también aquellos que se conforman por “la composición no cohesionada gráficamente que forma un término, es decir, los términos integrados por dos o más palabras”²². A este respecto, Cardero señala, en las conclusiones de su trabajo doctoral, que estos términos son “unidades mayores que una palabra desde la composición como sintagmas terminológicos. Formas lexicalizadas en las que el significado se ha petrificado y

²⁰ Es de notar lo que Guadalupe Aguado de Cea señala respecto a estas categorizaciones; ella dice que “no existe un acuerdo unánime en torno al concepto de sigla, y los límites con los conceptos de abreviatura y acrónimo”. Cit. en Franceschi, María Eugenia. *La terminología INTERNET. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología*. México, (1996): 144 – 149. p. 147.

²¹ Rosa Estopà. *Elementos lingüísticos de las unidades terminológicas para su extracción automática. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001. p. 68.

²² Ana María Cardero. *El procesamiento de una terminología. Referencia especial a la terminología de control de satélites en el área de las telecomunicaciones en México*. Tesis de doctorado. México, 2001. p. 45.

hay una fijación sintáctica. Aparecen, acompañadas, en ocasiones, por marcas que complementan su significado o que lo equivalen. Estas unidades léxicas adquieren valor de término en un contexto o situación especializada y representan un nudo de conocimiento compacto en la estructuración conceptual de un ámbito especializado.”²³

Una vez que se han aclarado en qué consisten las características y diferencias entre término y palabra, y se han mencionado los distintos tipos de términos que existen según los varios enfoques que dan los estudiosos, es necesario que antes de continuar se establezca bien claro qué es un término para esta investigación. Para hacerlo, se recurre a las ideas arriba expuestas, y que de acuerdo a las necesidades del presente trabajo, son las más adecuadas.

Por tanto, en esta investigación se adopta la definición de término que da Cabré, en la que señala que un término²⁴ es una “unidad que designa los conceptos propios de cada disciplina especializada, [y que] posee una vertiente sistemática (formal, semántica y funcional), toda vez que son [los términos] unidades de un código establecido, y manifiestan asimismo otra vertiente pragmática, puesto que son unidades usadas en la comunicación especializada para designar los ‘objetos’ de una realidad preexistente”²⁵.

1.1.3 El área terminológica en nuestros días

Como ya se ha mencionado, la terminología cumple una función, si no vital, sí muy importante en nuestra vida. Con el uso correcto de ésta, y haciendo una paráfrasis de las ideas de Galinsky, tenemos que la terminología puede ayudar en un nivel formal y desde el punto de vista de la ingeniería lingüística a la:

²³ Ídem. p. 163.

²⁴ También es necesario manifestar que cuando se hable sobre términos compuestos de dos o más palabras, o sintagmas, se adoptó el término que utilizan María Teresa Cabré y Rosa Estopà: *terminología poliléxica*.

²⁵ María Teresa Cabré. Op. Cit. p. 169.

1. ordenación del conocimiento humano a través del pensamiento conceptual;
2. aplicación del conocimiento;
3. formación y difusión de información;
4. traducción e interpretación;
5. almacenamiento y recuperación de información.

Debido a lo cual, los estudios y los órganos que voltean hacia ella son cada día más numerosos y exhaustivos.

El siguiente postulado, tomado de la norma DIN²⁶, muestra el grado de profundidad que deben presentar las terminologías, y dice que “el lenguaje especializado es el área de la lengua que aspira a una comunicación unívoca y libre de contradicciones en un área especializada determinada y cuyo funcionamiento encuentra un soporte decisivo en la terminología establecida”²⁷.

Hace tiempo que comenzó la profundización y difusión de la terminología a través de órganos tanto nacionales como internacionales; en Hispanoamérica, los trabajos más importantes sobre terminología han venido de España, en donde en 1982 se conforma la HISPANOTERM; en Venezuela, se creó el Banco de Terminología de la Universidad Simón Bolívar (BTUSB). Gracias al Primer Simposio Latinoamericano de Terminología realizado en Caracas, se funda en 1988 la Red Iberoamericana de Terminología (RITerm), que es uno de los organismos más importantes en cuestión de terminología²⁸.

En México, dentro del Grupo de Ingeniería Lingüística (GIL) de la UNAM, se hacen esfuerzos para aprovechar los recursos que representa el uso y explotación de la terminología en la ingeniería lingüística; muestra de ello son los diferentes proyectos que

²⁶ Deutsches Institut für Normung (Instituto Alemán de Normalización) encargado de la normalización en campos de la técnica y creada en Alemania en 1917.

²⁷ DIN 2342. Ver Reiner Arntz y Heribert Picht. Introducción a la terminología. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide, 1995. p. 28.

²⁸ Cfr. Alicia Fedor de Diego. Algunas reflexiones sobre las "políticas terminológicas". Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología. México, (1996): 11a - 11d.

participan de esta área para encaminarse a la consecución de las metas propuestas; además de la formación de profesionales comprometidos con el área. También en la UNAM, en el campus de Acatlán se desarrollan desde el inicio de la década de los 80 vocabularios especializados.

1.2 Terminología y lingüística

Como parte de la lingüística, la terminología tiene su fuente de trabajo en las lenguas, en este caso las lenguas de especialidad. Esta disciplina terminológica participa interdisciplinariamente de materias como la estadística o la informática. Asimismo, Cabré señala que la terminología puede ser vista desde el enfoque no sólo lingüístico, sino desde la perspectiva que ofrecen otras disciplinas que a su vez participan y se nutren de la terminología, como en este caso, la ingeniería lingüística.

Ahora bien, desde un enfoque lingüístico, conviene precisar las siguientes cuestiones respecto de la terminología en relación con la lingüística²⁹:

1. la terminología se interesa por el término mismo, independientemente de si éste presenta flexión o de la sintaxis que le confiera un valor gramatical específico, en tanto para la lexicología las palabras son las que tienen mayor relevancia, atendiendo siempre a la utilidad de éstas en un contexto;
2. a diferencia de la lingüística, que tiene por objeto la descripción, a la terminología le interesa normalizar las formas;
3. el punto de partida de la terminología es el concepto, mientras que para la lexicología es la designación;
4. la terminología no persigue el fin de explicar los términos, sino de crear una norma que fije la búsqueda, y la ordenación de los términos;
5. la terminología no pretende explicar las razones, sino, fijar normas.

²⁹ Ver M. Teresa Cabré. Op. Cit.

Ahora bien, más arriba se habla de que Cabré manifiesta una orientación de la terminología hacia un aspecto lingüístico³⁰, y dentro de éste, se le considera como una unidad autónoma respecto de la lingüística, y más específico, de la lexicología.

Mas, ¿por qué se habla de una relación específica mayor de la lexicología respecto de la terminología y no de la lingüística a la terminología, o a la lexicografía? La razón estriba en que no es la lingüística la disciplina que está más cerca de la terminología. Si bien es cierto que la lingüística engloba a la lexicología y a la terminología como subáreas, del mismo modo que a la fonología o la morfología, cada una se ocupa de un aspecto específico del lenguaje pero a diferente nivel.

La lexicología tiene su campo de acción en toda la gama de palabras de una lengua, mientras que la terminología sólo se centra en aquéllas específicas de cierta disciplina, estas últimas denominadas términos —con características específicas y delimitadoras que los constituyen una clase aparte y que más adelante se explicarán—; por otro lado, para dejar más clara la relación y/o diferencia de la terminología respecto de la lexicología, tengamos presente el razonamiento de que una parte es un componente de un todo; ahora bien, siguiendo este razonamiento, tenemos entonces que todo término es parte de un léxico; esto es, si bien ambas tienen un campo de trabajo en común, las palabras, sean éstas coloquiales en una o especializadas en otra, resulta, al fin, un grado de inclusión de la terminología dentro de la lexicología. De ahí que el tercer enfoque que maneja Cabré diga que la lexicología, como rama de la lingüística, es la disciplina más próxima a la terminología.

Pero ésta es sólo una aproximación y una teoría que pueden compartir un cierto grupo de personas; y que, aunque pudiese parecer general a simple vista, para el estudioso no lo es; por ejemplo, Sager³¹ habla de que “respecto a su aplicación, la terminología está relacionada con la lexicografía” y añade que la terminología “comparte con la lexicografía métodos sobre la estructuración y la descripción de palabras”.

³⁰ Ver nota 1.

³¹ Juan Carlos Sager. *Curso práctico sobre el procesamiento de la terminología*. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide, 1993. p. 23.

En otro sentido, Rodolfo Alpizar maneja una concepción muy particular, y válida sobre la terminología y el plurisemantismo que encuentra en esta voz, en la que señala que dado el carácter polisémico de la palabra, ésta ofrece una vaguedad y confusión que hacen necesario un riguroso reajuste de preceptos. Por ejemplo habla de³²:

- “Tecnocconceptología; que es el estudio de los conceptos de una o varias ramas de actividad científica o técnica, de las relaciones entre ellos y de sus formas de denominación en un sistema lingüístico.
- Tecnolexicología, que se enfoca al estudio de los términos de una o varias ramas de actividad científica o técnica, de sus relaciones como miembros del subsistema léxico de una lengua, y de sus relaciones con los conceptos que denominan.”

E incluso, habla sobre la:

- Tecnojerga, que no es otra cosa, que el “conjunto de recursos lingüísticos de los tecnolectos que suelen emplearse por los especialistas en el coloquio y en otras situaciones comunicativas informales.”

Para sostener el argumento de que en tanto el concepto como el término sean ocupaciones de la terminología, ésta debe ser estudiada en distintas vertientes. Habla también acerca del:

- Tecnolecto: “Conjunto de todos los recursos que la lengua pone a disposición del discurso científico o técnico.”

Y justifica el uso de éste, arguyendo que la terminología es el estudio de una parte de los recursos de un tecnolecto, que no es otra cosa que un conjunto de recursos lingüísticos.

Mas, como el deseo en esta investigación no es aclarar éste y otros problemas que surgen respecto de la terminología, sino dar los elementos básicos sobre un tema que forma parte del cuerpo central de la investigación, se hace necesario sólo el justificar la noción a priori de este apartado: el que la terminología en tanto subcomponente de la lexicología, e

³² Rodolfo Alpizar Castillo. Op. Cit. p. 86.

incluso, de la lexicografía, comparte criterios que la vinculan directa y estrechamente con la lingüística.

La terminología es entonces, para los fines de esta investigación, vista desde un enfoque lingüístico, la disciplina que se encarga de normar las unidades fuente de su objeto de estudio, los términos, para organizar la gnosis a través del pensamiento conceptual.

1.3 La terminología como recurso de la ingeniería lingüística

Cuando hablamos de lingüística, o de las áreas que participan de ésta, como la fonética, la semántica, la lexicología, y demás, es casi imposible concebir que tendría igualmente una relación de trabajo con un área de conocimiento tal como la biología, la medicina o la arquitectura. Esta incredulidad no es gratis, se debe a las categorizaciones en que se han encuadrado las distintas áreas de conocimiento, orientadas a áreas científicas, humanísticas, sociales, etcétera; y que tienen sus objetos de estudio delimitados con fronteras precisas, manejándose de acuerdo con criterios propios encaminados a resolver cuestiones diversas y a diferente nivel.

Hasta hace poco era extravagante el pensar en relacionar por ejemplo la ingeniería y la lingüística, dos áreas desvinculadas desde siempre y que en esencia no tienen nada en común; no obstante sus directrices opuestas, la combinación de sus recursos puede ayudar o coadyuvar, con un nuevo enfoque, a la solución de problemas. Ya antes se hizo mención de la diversidad de recursos con que cuenta la lingüística; si estos recursos se aplican a otras áreas como un apoyo, los horizontes de investigación se ensanchan con perspectivas diferentes, por ejemplo, una investigación computacional encaminada a resolver problemas de identificación de recursos semánticos, en donde las reglas y algoritmos matemáticos serían coadyuvados con criterios y recursos lingüísticos, en este caso semánticos y gramaticales, para resolver problemas o identificar situaciones que desde un punto de vista únicamente computacional, difícilmente serían contempladas. O la interacción de patrones lingüísticos con reglas informáticas para crear de forma rápida y eficiente un corpus digitalizado para el análisis gramatical del *Quijote* en forma electrónica. Esta interacción

permitiría obtener provecho tanto para las distintas áreas de conocimiento con sus especialistas, como para el común de la gente, con un resultado de beneficio muy grande. Si no fuese por esta interacción, no tendría este procesador de textos que facilita y ahorra cantidad de tiempo en la elaboración de este escrito, sin mencionar las repercusiones económicas que este simple procesador de textos representa a quienes vieron la capacidad y riqueza de investigación que trajo la unión de áreas de trabajo tan impensables: la ingeniería y la lingüística.

La interrelación de estas dos áreas de conocimiento es lo que hoy en día se conoce como *ingeniería lingüística*. Pero, ¿qué significa la ingeniería lingüística? Trataré de dar una respuesta rápida y que abarque en esencia lo que significa hoy la ingeniería lingüística. Ésta es un área de reciente creación, que a partir de la segunda mitad del siglo pasado ha visto incrementar sus estudios y que en la década pasada tuvo un auge muy interesante, sobre todo con la aparición de Internet y la cada vez mayor demanda en el uso de la computadora tanto en el ámbito institucional como en el ámbito casero.

El campo de estudio lo constituyen las diferentes ingenierías, entre las principales, la ingeniería en sistemas y la ingeniería en computación, y que busca la resolución de problemas y la creación de recursos prácticos y comerciales a través de la aplicación y explotación de recursos lingüísticos³³ y de ingeniería a problemas y necesidades que tienen que ver con el lenguaje y la explotación de éste, como la traducción automática o asistida, la inteligencia artificial, el procesamiento de datos, la extracción de información, conceptual y de terminología, etcétera. La ingeniería lingüística es investigación que no se detiene en el aspecto teórico, es investigación que se origina de necesidades y que busca la aplicación de sus resultados para resolver estas necesidades de forma práctica.

Ahora bien, arriba se habló de que la ingeniería lingüística se vale de los distintos recursos que le representa la ingeniería y la lingüística, con todas las áreas y subáreas que

³³ Tal es el caso de la terminología o la lexicología; ramas lingüísticas orientadas a la extracción de elementos para la creación de léxicos, vocabularios, diccionarios, etc.; que como herramientas de apoyo aplicadas a la ingeniería lingüística buscan la resolución de sus objetivos de forma automática.

las conforman; pues bien, una vertiente de esta investigación se encamina hacia la explotación de una de estas áreas, la terminótica; la cual representa el soporte terminológico y el campo de investigación para hacer que este soporte se vea alimentado en forma eficiente y en lo posible, automático, por procesos de extracción de terminología.

De esta forma se establece que es dentro de este ámbito, el de la ingeniería lingüística, en donde se desarrollará esta investigación que se encamina hacia la extracción automática de términos.

2. ESTUDIOS SOBRE EXTRACCIÓN TERMINOLÓGICA

2.1 Extracción de terminología

En el capítulo 1 se establecieron los principios básicos de la terminología¹, destacando tanto su objeto de estudio, los términos, como la forma en que será tratada esta disciplina en la presente investigación, esto es, vista como una herramienta de apoyo a la ingeniería lingüística.

Ahora bien, en el presente capítulo se tratará sobre la extracción de terminología como recurso que sirve a varias áreas de conocimiento, visto desde la disciplina de la ingeniería lingüística.

Cuando se habla de extracción en este trabajo se hace referencia a que de un todo se busca obtener una parte muy definida, esto es, que de un léxico tal se pretende conseguir sólo un reducto de palabras que por su carácter de especialización sean consideradas como términos. Al hacer mención de automático o semiautomático, se alude a los procesos que tienen que desarrollarse al interior de un programa² de cómputo, y que siguen determinadas reglas para realizar ciertas funciones para las que es creado.

Lo anterior aplica, como fin práctico, que al hablar de extracción auto o semiautomática se entienda que por medio de programas computacionales se intente reducir

¹ Si tanto la disciplina como el objeto de estudio (terminología y término respectivamente) no son tratados profusamente, se debe a que este trabajo, que si bien es cierto está vinculado con la terminología, no está diseñado ni trabajado para seguir lineamientos únicamente lingüísticos, sino que la terminología como disciplina sirve como herramientas para resolver cuestiones desde un enfoque de la ingeniería lingüística. Por tanto, baste con dejar en claro los elementos básicos, que son la materia prima de este trabajo (términos), y la disciplina que los estudia (terminología), para entender que lo que aquí se pretende es la extracción de términos, no de léxico, de palabras, de información, etc.

² En esta investigación, se hablará de programa y/o software, para referirse a todo el sistema computacional que trabaja con un conjunto de instrucciones que permiten a una computadora realizar determinadas operaciones. Cfr. D.R.A.E 1995.

los tiempos que llevan al especialista, sea terminólogo, lexicólogo, lexicógrafo, etc., a efectuar su labor y volverla más sencilla.

Para realizar tales tareas de extracción que redunden en beneficio de un sinnúmero de personas, es preciso contar con la infraestructura computacional y de conocimiento lingüístico, para crear programas³ que puedan llevar a cabo estas tareas; para ello, se ha trabajado desde hace algunos años en la investigación y creación de herramientas que suplan el trabajo manual por el efectuado por una máquina.

Los resultados que tantos años de investigación han producido son hasta cierto punto satisfactorios; sin embargo, existen tareas que tales programas no resuelven de modo eficaz; y aunque, como apunta Cardero, “la detección [y en este caso, la extracción] automatizada de términos resulta del mayor interés para la terminología, sobre todo en etapas de inicio de acopio”⁴, existe un consenso de parte de los usuarios de este tipo de programas, quienes manifiestan que todavía hay mucho sobre qué investigar con el fin de mejorar estos programas, puesto que presentan deficiencias que hacen que el trabajo no sea del todo automático.

Ahora bien, si tales programas sufren de limitaciones, es preciso señalar el porque de éstas. Una de estas insuficiencias que presentan los extractores automáticos es la que se refiere a la polisemia de las palabras, ya que un hablante no limita el conocimiento que goza de las palabras sólo a sus aspectos lingüísticos, sino, que junto a ellos, manipula datos exteriores, datos extralingüísticos, que determinan un uso real y particular que el hablante hace de cada palabra⁵, sin que le importe mucho o sea conciente del aspecto lingüístico; así, es difícil crear reglas lingüísticas que sean aplicables para todo el repertorio léxico de una lengua y adaptarlas a programas computacionales.

³ Sobre estos programas se hablará en un apartado del capítulo 3.

⁴ Ana María Cardero. El procesamiento de una terminología. Referencia especial a la terminología de control de satélites en el área de las telecomunicaciones en México. Tesis de doctorado. México: 2001. p. 163.

⁵ Véase Juan Carlos Sager. Curso práctico sobre el procesamiento de la terminología. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide, 1993.

En el capítulo anterior se habló sobre el problema de la palabra y/o término *síntesis*; dependiendo de los contextos en que es usada(o), presenta características que la delimitan en un rubro particular, pero, sin dejar de pertenecer a ambos. ¿Cómo entonces establecer y resolver esta problemática por medio de un sistema, que se valga de una o varias áreas de conocimiento para que permita que una máquina haga el trabajo de modo automático?

Sin embargo, la polisemia no es la única causa de ruido en este tipo de análisis, por ejemplo, Jorge Vivaldi⁶ establece como limitaciones el que este tipo de programas se proponen gran cantidad de palabras que el usuario debe desechar, y señala además que muchos candidatos reales no son propuestos como términos, o que no es frecuente encontrar términos monolexemáticos entre los candidatos.

En este mismo tenor, Rosa Estopà plantea en su estudio sobre extracción automática, su preocupación manifiesta al advertir que “las UT [unidades terminológicas] monoléxicas simples son difíciles de tratar automáticamente porque su carácter especializado es totalmente idiosincrático. Son, pues, un tipo de unidades que no poseen características, ni morfológicas, ni sintácticas explícitas, que permitan detectarlas automáticamente”⁷.

A problemas de este tipo han tenido que enfrentarse todos los que de alguna u otra forma hayan trabajado en la extracción automática, y no exclusivamente de términos, como por ejemplo, Luis Fernando Lara y el equipo que junto con él se ha visto involucrado en el desarrollo y elaboración del *Diccionario del Español de México*⁸ (DEM), quienes en un

⁶ Jorge Vivaldi. *Elaboración de una aplicación automática de reconocimiento y extracción de información terminológica en textos de dominios restringidos. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001. p. 230.

⁷ Rosa Estopà. *Elementos lingüísticos de las unidades terminológicas para su extracción automática La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001. p. 68.

⁸ Véase Luis Fernando Lara, Roberto Ham Chande y María Isabel García Hidalgo. *Investigaciones lingüísticas en lexicografía*. México: COLMEX, 1979.

principio del diccionario se vieron ante la problemática de elegir qué tipo y cantidad de documentos trabajarían para la elaboración del diccionario, la pertinencia de éstos, o la necesidad de delimitar sincrónicamente sus fuentes para contrarrestar el problema de la aparición de nuevos datos y documentos. Problemas que si bien no son estrictamente del orden de la extracción automática de terminología, si se relacionan con los preliminares que anteceden el proceso de extracción automática, en este caso de vocabulario.

De aquí que no sólo sea útil, sino necesario profundizar y tratar de resolver estas carencias o fallas que representan los extractores automáticos, ya que los campos de acción de las disciplinas cada día interactúan de una manera más activa y en beneficio de todos, aunque, como indica Carlos Sager, los resultados de estos trabajos no tengan como fin inmediato el hombre, sino las diferentes disciplinas que puedan satisfacer sus necesidades a través de los resultados terminológicos, como es el caso de los estudios sobre ingeniería lingüística, que se socorren de la terminología para resolver las exigencias que se le presentan en:

- traducción automática y asistida por computadora,
- correctores de ortografía,
- sistemas de recuperación de información,
- sistemas expertos y basados en el conocimiento,
- inteligencia artificial, etc.⁹

No son mínimas las cuestiones y problemáticas que implica el estudio y trabajo con cuestiones de extracción de términos, ni tampoco son fáciles las posibles soluciones; mas con todo, resulta una labor bastante interesante esta de extraer terminología, pese a todas las dificultades y sufrimientos que conlleva su estudio.

Mientras estos problemas se resuelven, es preciso describir, en este material, los distintos métodos con los que se trabaja para obtener terminologías de modo automático, o semiautomático, propiamente dicho.

2.2 Métodos para obtener términos de forma automática

Existen tres métodos para realizar la tarea de extracción de terminología automática; éstos basan sus análisis en distintos tipos de conocimientos:

- estadísticos,
- basados en reglas lingüísticas, e
- híbridos (estadísticos y lingüísticos).

Dichos sistemas de extracción, como señalan en su artículo sobre los avances de la terminología computacional, Cabré, Estopà y Vivaldi¹⁰, trabajan a partir del siguiente principio:

All systems analyse a corpus of specialized texts in electronic form and extract list of word chunks (i.e. candidate terms) that are to be confirmed by the terminologist. To make the terminologist's task easier the candidate terms is provided with its context and, when available, with any other further information (frequency, relationship between terms, etc.)

[Todos los sistemas analizan un corpus de textos especializados en formato electrónico y extraen listas de palabras (o sea, listas de posibles términos o candidatos a ellos), las cuales deben ser corroboradas por los especialistas. Para hacer más sencilla la labor del terminólogo, los términos candidatos son incluidos con su contexto, cuando está disponible, además de información útil, (como frecuencia, relación entre palabras, etc.)]

⁹ Véase Juan Carlos Sager. Op. Cit.

¹⁰ M. Teresa Cabré, Rosa Estopà, Jordi Vivaldi. Recent Advances in Computational Terminology. Amsterdam / Philadelphia: John Benjamins B. V., 2001. p. 54.

2.2.1 Métodos estadísticos

Estos métodos han sido ampliamente usados por la bibliotecología y ciencias de la información para encontrar las palabras clave o descriptores en textos especializados. De igual forma, tales métodos han sido utilizados para encontrar los probables términos de un área determinada. Esta metodología opera con una base que, sin entrar en detalles matemáticos o estadísticos, detecta las unidades que mayor incidencia tienen dentro del universo de donde se extraerán los términos, el cual se conforma exclusivamente por los textos que conforman el corpus¹¹, y los presenta como posibles términos.

La lógica para identificar términos en un texto especializado consiste en comparar las palabras que ocurren en el texto sobre el tema que se busca, con referencia a las palabras que ocurren en un texto sobre otra área cualquiera. De esta manera, las palabras que son diferentes serán las consideradas como específicas de esa área en particular.

Una vez que el programa aplica reglas estadísticas para realizar los procesos de búsqueda, los resultados se presentan en diferentes tipos de listas: listas alfabéticas, de frecuencia o estadísticas; cada una con una función a desempeñar, según las necesidades del terminólogo o la investigación, dentro del análisis. Lara les atribuye los siguientes beneficios¹²:

- “La ordenación por frecuencias permitirá identificar el vocabulario más usual — estadísticamente hablando.
- La ordenación alfabética permitirá conocer las características estadísticas de cualquier vocablo”.

El fin de este método estadístico consiste en presentar candidatos a términos, sustentados en el criterio de mayor aparición y reincidencia. Debido a este criterio, existen investigadores que no lo consideran una opción efectiva, ya que la labor de este método

¹¹ Véase Luis Fernando Lara, Roberto Ham Chande y María Isabel García Hidalgo. Op. Cit.

¹² Ídem. p. 34.

termina al desplegar las palabras que estadísticamente tienen más o menos posibilidades de ser tenidas como términos.

Un problema muy común con el empleo de este tipo de métodos es con respecto a los términos que aparecen con muy baja frecuencia¹³:

“The problem with this kind of approach is that there are low-frequency terms difficult to be managed by extraction systems. Here it is important to note that these systems use basically numerical information and thus are prone to be language independent.”

[El problema con este tipo de acercamiento lo constituye el difícil manejo que representan para estos sistemas los términos de baja-frecuencia. Aquí es importante señalar que estos sistemas usan información básicamente numérica y, por consiguiente, tiende a ser un idioma independiente.]

Términos estos que no por tener una baja frecuencia dejan de ser términos, independiente de si representan mayor aporte para el correcto entendimiento de los textos o no; recuérdese que esto no es extracción de información documental, en donde se pretende la extracción de información representativa del texto; sino extracción de terminología, y el fin de ésta es presentar los términos que aparezcan en un corpus, aunque sólo sea nombrado una vez.

Otro punto en contra que tiene el método estadístico, reside en el hecho de que es muy probable que las unidades que conforman los términos compuestos aparezcan muy por debajo en los índices de frecuencia, y por tanto, sean discriminadas con base en un criterio

¹³ M. Teresa Cabré, Rosa Estopá, Jordi Vivaldi. Op. Cit. p. 55.

numérico—estadístico. Por ejemplo, en el estudio *Combining linguistics with statistics for multiword term extraction: A fruitful association?*¹⁴, los autores señalan esta deficiencia:

“On the other hand, as they only allow the acquisition of binary associations, these systems must apply enticement techniques to acquire multiword terms with more than two words. Unfortunately, such techniques have shown their limitations.”

[Por otro lado, cuando estos programas permiten la adquisición de asociaciones binarias, estos sistemas deben aplicar diferentes técnicas para adquirir términos de más de dos palabras. Desgraciadamente, tales técnicas han mostrado sus limitaciones.]

Además, mencionan que:

“Statistical methodologies extract multiword lexical units that can not be considered terms.”

[Las metodologías estadísticas extraen unidades léxicas de dos o más palabras que no pueden ser consideradas términos.]

Sin embargo, este método no es del todo desdeñable, por ejemplo, el equipo del *DEM* optó por este método, ya que ellos señalan como atributo el que éste es “el único capaz de dar los registros necesarios y la cantidad de datos suficientes (...) de un modo objetivo e imparcial”¹⁵.

Mas con todo, la aplicación de la estadística resulta de provecho, sobre todo en la elaboración de nuevos programas que empiezan a tomar en cuenta estas deficiencias; o

¹⁴ Gaël Dias, Sylvie Guilloré, Jean-Claude Bassano, José Gabriel Pereira Lopes. Combining Linguistics with Statistics for Multiword Term Extraction: A Fruitful Association? [En línea.] <<http://citeseer.nj.nec.com/context/1674205/410737>>

¹⁵ Luis Fernando Lara, Roberto Ham Chande y María Isabel García Hidalgo. Op. Cit. p. 12.

cuando se trabaja vinculando la estadística con la lingüística, como en el caso de los métodos híbridos que más adelante se explicarán.

2.2.2 Métodos basados en reglas lingüísticas

En contraposición al método anterior, que basa sus análisis en la probabilidad de ocurrencia de términos en un mismo texto, y que por tanto dependen de la cantidad de información disponible, los métodos basados en reglas lingüísticas presuponen criterios cualitativos y no requieren una gran cantidad de textos para extraer información.

Este tipo de análisis propone la extracción de términos a partir del uso exclusivo de reglas lingüísticas: sintaxis, morfología, etcétera., esto es:

Linguistic systems propose to extract relevant terms by using techniques that analyse specific syntactical structures in the texts¹⁶.

[Los sistemas completamente lingüísticos proponen extraer términos usando técnicas que analizan estructuras sintácticas específicas en los textos.]

Estos sistemas basan su funcionamiento en:

“Use noise and silence measure of its efficiency. Noise attempts to assess the rate between discarded candidates and accepted ones; silence attempts to assess those terms contained in an analysed text that are not detected by the system. Noise is the common problem of those systems using this approach. Errors in the assignation of morphological category are also shared by these systems.”¹⁷

¹⁶ Véase Gaël Dias, et al. <http://citeseer.nj.nec.com/contexto/1674205/410737>

¹⁷ M. Teresa Cabré, Rosa Estopá, Jordi Vivaldi. Op. Cit. p.56.

[Usar medidas de ruido y de silencio para su eficiencia. El ruido intenta evaluar el promedio entre los candidatos desechados y los aceptados; el silencio intenta evaluar aquellos términos contenidos en un texto analizado que no son detectados por el sistema. El ruido es el problema común de los sistemas que usan este acercamiento. Los errores en la asignación de categoría morfológica también son deficiencias presentadas por estos sistemas.]

Este tipo de análisis, según Cabré, puede presentar dos tipos de información¹⁸:

- a. "Term specific: it consists in the detection of the recurrent patterns from complex terminological units such as noun-adjective and noun-preposition-noun. This calls for the use of regular expressions and techniques of finite state automata.
 - b. Language generic: it consists in the use of more complex systems of NLP that start with the detection of more basic linguistic structures: noun phrase (NP), prepositional phrase (PP), etc."
-
- a. [Término específico: consiste en la detección de modelos recurrentes de las unidades terminológicas complejas como nombre-adjetivo y nombre-preposición-nombre. Esto requiere el uso de expresiones regulares y técnicas de autómatas de estados finitos.
 - b. Genérico del lenguaje: consiste en el uso de sistemas más complejos de PLN que empiecen con la detección de estructuras lingüísticas más básicas: nombre - frase (FS), frase preposicional (FP), etc.]

¹⁸ Ídem. p. 55.

“In both approaches each word is associated to a morphological category.”

[En ambas aproximaciones cada palabra se asocia a una categoría morfológica.]

Esta metodología basada en el uso de reglas lingüísticas también tiene sus inconvenientes, ya antes se mencionaron las dificultades de categorización morfológica; otro problema que no puede superar totalmente este método, reside en el hecho del idioma para el que esté elaborado tal sistema, en la mayoría de los casos estos programas son monolingües, ya que sólo presentan resultados importantes en la lengua para la que fueron creados, y menos importantes cuando se aplican a otras lenguas. Baste mencionar para los que han trabajado en la aplicación de estos u otros programas diseñados casi en su mayoría para el inglés, que los resultados aplicados para el español distan mucho de los obtenidos en la lengua madre del programa. O bien, hay que manejar el corpus amoldándolo a las características de una lengua en la que no están escritos, por ejemplo, la flexión de los verbos en español, o el género o la acentuación; rasgos que en algunos otros idiomas no están presentes y por lo tanto no están contemplados para resolverse en otra lengua diferente a la que tienen por base.

2.2.3 Métodos híbridos

Una metodología que aúna los beneficios y reduce, aunque no elimina totalmente, las carencias de los dos métodos mencionados más arriba, es la que trabaja a partir de criterios híbridos; es decir, interactúa la metodología estadística con la metodología lingüística.

Los métodos híbridos de algún modo solventan las carencias que por separado presenta tanto el método estadístico como el lingüístico. Trabaja a partir de la interacción de criterios estadísticos y lingüísticos para los análisis de corpora.

Algunos programas que trabajan con este método funcionan de la siguiente forma¹⁹:

“In order to overcome the problems evidenced by most of the statistical approaches, hybrid linguistic-statistical methods define co-occurrences of interest in terms of syntactical patterns and statistical regularities. Some approaches reduce the searching space to groups of words that correspond to *a priori* defined syntactical patterns (Noun+Adj, Noun+Prep+Noun etc...) and then apply statistical measures to classify the pertinent sequences. Other approaches first identify statistical word regularities and then apply *a priori* defined syntactical filters to extract multiword term candidates.

[Para superar los problemas evidenciados por la mayoría de los acercamientos estadísticos, los métodos lingüístico-estadísticos híbridos definen una co-ocurrencia de interés en términos de modelos sintácticos y controles estadísticos. Algunos acercamientos reducen el espacio de búsqueda a grupos de palabras que corresponden a los modelos sintácticos *a priori* definidos (Sust+Adj, Sust+Prep+Sust, etc...) y entonces aplican medidas estadísticas para clasificar las secuencias pertinentes. Otros acercamientos, primero identifican controles estadísticos de la palabra y después aplican filtros sintácticos definidos *a priori* para extraer los candidatos a término de dos o más palabras.]

¹⁹ Véase Gaël Dias, et al. <http://citeseer.nj.nec.com/context/1674205/410737>

No obstante, al aunar estos dos métodos, los resultados de extracción automática de términos no complacen al cien por ciento a los terminólogos. Hay muchas lagunas que los métodos antes señalados, aun y cuando trabajen de forma conjunta, no llenan satisfactoriamente, y más todavía cuando se trata de cuestiones de automatización; sin embargo, y a pesar de todas estas dificultades, se sigue trabajando y analizando métodos, teorías, aplicaciones, recursos y demás para mejorar los programas de extracción con que ahora se cuenta; mas, es necesario esperar todavía tiempo para ver los resultados. Por ahora y para este trabajo, el uso de estas tres metodologías es bienvenida; aunque por razones obvias no será posible adentrarse en investigaciones más exhaustivas respecto a los problemas y soluciones que estas metodologías presentan.

2.3 Resultados importantes relacionados con la extracción automática de términos

En la actualidad muchos son los proyectos e investigaciones que se han visto tanto vinculados como beneficiados con la aplicación de sistemas de extracción de términos. Mucho de este beneficio se debe en gran medida a la creación de bancos de datos terminológicos, que desde que se hace un uso mayor de éstos, las labores de numerosos profesionales, sin excluir al terminólogo, se han visto facilitadas y optimizadas hasta el grado de que en la actualidad su uso, ya sea en investigación como en industria, se hace cada vez más imprescindible.

Mas no sólo es la explotación de los citados bancos de datos terminológicos lo que hace de la extracción de términos una herramienta útil y de provecho: está también la ayuda en automatización de procesos, la investigación en tareas metodológicas lingüísticas y estadísticas, la planeación interdisciplinaria e interactiva de distintas áreas de conocimiento, el ahorro de tiempo y recursos económicos, la facilitación para la elaboración de diccionarios, léxicos, terminologías, etcétera.

Por lo dicho anteriormente, es fácil deducir que la tarea de extracción de términos, vía procesos automáticos, representa un gran aporte en el proceso de investigación y realización de las tareas que simplifcan y conllevan al cumplimiento del fin de la empresa.

Para concluir este capítulo, se citan algunos proyectos²⁰ en donde la extracción automática de términos ha sido participe con algunas de sus herramientas, y en los cuales se ven los alcances de ésta.

- Proyecto CRATER²¹: creación de herramientas para el manejo de corpus y la extracción de terminología.
- Proyecto GLOSMETAL²²: Repertorio terminológico trilingüe (español, inglés, griego) sobre el sector industrial del Metal.
- Proyecto RITerm BD²³: constitución de una red de bancos de datos terminológicos.
- Proyecto HAL²⁴: Herramientas de ayuda al lexicógrafo.
- Proyecto BN ARGENTINA²⁵: Catalogación y edición electrónica del Fondo Foulché_Delbosc de la Biblioteca Nacional de la República Argentina
- Proyecto PROTEUS²⁶: Sistema de extracción de información y traducción automática.
- Proyecto ACORDEON²⁷: Aplicaciones cooperativas de recuperación de información.

²⁰ Por razones de tiempo y espacio sólo se hará mención de algunos proyectos. Para más información acerca de éstos, se ponen a disposición ligas de Internet a pie de página, en donde pueden ser consultados tales proyectos.

²¹ Proyecto CRATER. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/crater.html>>

²² Proyecto GLOSMETAL. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~flora/projects/glosmetal.html>>

²³ Proyecto RITerm BD. [En línea]. <<http://www.riterm.net/es/index.htm>>

²⁴ Proyecto HAL. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <http://www.lllf.uam.es/docs_es/proyectos/APC.html>

²⁵ Proyecto BN ARGENTINA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~fmarcos/informes/BNArgentina/BN.htm>>

²⁶ Proyecto PROTEUS. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/proteus.html>>

- Proyecto C-ORAL-ROM²⁸: Corpus oral de las lenguas romances en formato multimedia.
- Proyecto ADMYTE²⁹: Archivo digital de manuscritos y textos españoles.
- Proyecto ATILA³⁰: Aplicaciones telemáticas de ingeniería lingüística.
- Proyecto CREA³¹: Corpus de referencia del español actual.
- Proyecto CON-TEXT³²:: Corrector gramatical.
- Proyecto EUROTRA³³: Traducción automática.
- Proyecto SIIT VIRTUAL³⁴: Servicio de información en línea.

²⁷ Proyecto ACORDEON. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~ares/acordeon.html>>

²⁸ Proyecto C-ORAL-ROM. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~sandoval/coralrom.html>>

²⁹ Proyecto ADMYTE. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~fmarcos/informes/admyte/admyteix.html>>

³⁰ Proyecto ATILA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/atila.html>>

³¹ Proyecto CREA. [En línea]. <<http://www.rae.es/NIVEL1/CREA.HTM>>

³² Proyecto CON-TEXT. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/context.html>>

³³ Proyecto EUROTRA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <http://www.lllf.uam.es/docs_es/proyectos/APC.html>

³⁴ Proyecto SIIT VIRTUAL. [En línea]. <<http://www.ritem.net/es/index.htm>>

3. HERRAMIENTAS PARA EXTRAER TÉRMINOS A TRAVÉS DE PROCESOS AUTOMÁTICOS

3.1 Procesos de análisis

En un apartado del capítulo anterior se habló acerca de las distintas metodologías que existen para realizar la extracción de términos (métodos estadísticos, lingüísticos e híbridos); ahora, en este capítulo, toca el turno de hablar sobre la aplicación de estas metodologías en programas computacionales, y en específico, en el programa con el cual se efectuaron los procesos de extracción: WordSmith¹.

La aplicación de estos sistemas se ve traducida en herramientas con las cuales WordSmith y algunos programas desarrollan los procedimientos que llevan hacia la extracción de términos; es decir, representan el punto práctico de la teoría.

Con el uso de estas herramientas surgen los primeros resultados: unidades candidatas a términos; además de que es viable trabajar direccionando el análisis hacia puntos específicos; por ejemplo, es posible cercar los elementos en el corpus para estrechar los límites de lo que interesa y de aquello que no; o comprobar de una forma, aunque no segura cien por ciento, si fiable, de saber que se trabaja con unidades que representan, como en recuperación de información, la esencia del corpus; etcétera.

¹ Se hablará, para la aplicación de los métodos expuestos en el capítulo 2, básicamente sobre WordSmith, porque es el programa que sirvió a esta investigación; sin embargo, no se dejarán de lado la extensión de beneficios a otros programas existentes, sin que esto signifique un dominio absoluto sobre el funcionamiento de estos otros programas, ya que como en la introducción se señala, el interés de este trabajo no reside en la evaluación ni de los métodos ni de los programas que trabajan en la extracción de términos, ni tampoco hacer una ponderación de WordSmith, sino simplemente, señalar los diferentes usos, aplicaciones y beneficios que presenta la extracción automática de términos, independiente del programa o el método utilizado.

Las herramientas² que a continuación se enumeran son, si no todas las que existen, sí las más representativas y ocurrentes en los programas; y fueron aquéllas que además guiaron los procesos de esta investigación por estar presentes y ser la base del programa WordSmith.

3.1.1 Lista de palabras

Los estudios lingüísticos por medio de análisis estadísticos se caracterizan por un conteo aritmético de las palabras en el texto, incluso, algunos de estos análisis, como el de los procesadores de texto, llegan sólo a este tipo de análisis: contar el número de palabras y el número de caracteres en un texto dado; lo cual no deja de ser útil, pero en un proceso como el de extracción de términos no sólo es necesario identificar la cantidad de palabras en el texto, cuanto el número de ocurrencias de cada una.

Las listas de palabras no son más que el resultado de un conteo estadístico de unidades en un corpus, y la aplicación de éstas da como resultado cualquiera de dos alternativas: una lista en orden alfabético de las palabras o de las unidades a considerar (constituyentes o partes de la oración), junto con su correspondiente frecuencia o número de ocurrencias (ilustración 1)³; o una lista de las mismas unidades, pero ordenada por la frecuencia, de manera que primero aparezcan las unidades que tienen mayor frecuencia, hasta llegar a las de mínima frecuencia.

Los rendimientos y provechos de estas listas ya los señalaba Lara⁴ en *Investigaciones lingüísticas en lexicografía*, donde el uso de estas listas da fruto en la realización del *DEM*; mas como el enfoque que en el *DEM* se daba al uso de estas listas era dentro de un ámbito más bien lexicológico, porque el interés era el léxico representativo del

² Cabe señalar que en la actualidad la mayoría de los programas que hacen uso de estas herramientas, emplean una metodología híbrida, apoyando las reglas lingüísticas con elementos estadísticos o viceversa, con la intención de arrojar un mejor y más sustentado resultado; aunque hay investigadores y, por tanto, programas que prefieren hacer uso de un sólo método, sin que por ello dejen de ofrecer resultados.

³ Todas las ilustraciones que aparecen en este capítulo corresponden a las ventanas con que se trabaja en WordSmith.

⁴ Véase nota 12 en capítulo 2.

Word	Frequency	Weight
AB	1	
ABAJÓ	4	0,03
ABANDONÓ	1	
ABSOLUTA	1	
AC	4	0,03
ACABABA	1	
ACABÓ	1	
ACCESIBLE	1	
ACCION	14	0,11
ACCIONES	1	
ACELERA	1	
ACELERACION	25	0,19
ACELERACIONES	1	
ACELERADO	11	0,08
ACELERAN	1	
ACELERAR	1	
ACELERARSE	1	
ACENSO	1	
ACEPTABA	1	
ACEPTAR	2	0,02
ACERCAN	1	
ACERCAR	1	
ACERO	4	0,03
ACLARACION	2	0,02
ACLARAR	1	
ACORTA	1	
ACTIVIDAD	1	
ACTIVOS	1	
ACTO	1	
ACTUA	10	0,08
ACTUALICE	1	
ACTUAN	4	0,03
ACTUANDO	1	
ACTUAR	1	
ACTUE	1	
ACTUEN	1	
ACUERDO	8	0,06
ADQUIRIDA	1	

Ilustración 1: Lista de palabras por ordenación alfabética

español de México, y no terminológico, como es el caso de este trabajo, se ponía un peso equivalente a ambas listas; sin embargo, para este trabajo son las listas de frecuencia las que presentan mayor riqueza de trabajo y son la base para los primeros estudios lingüísticos. Luhn señala la importancia de este tipo de listas cuando dice:

“The lists may alternatively be given in the order of frequency of occurrence of the words so as to facilitate the recognition of word ranking and the selection of useful keywords in accordance with such ranking.”⁵

⁵ H. P. Luhn. Potentialities of auto – encoding of Scientific Literature. Yorktown Height: CBM, 1959. p. 5.

[Las listas pueden darse alternativamente en el orden de frecuencia de ocurrencia de las palabras tanto para facilitar el reconocimiento de la jerarquía de la palabra y la selección de palabras clave útiles de acuerdo con tal clasificación jerárquica.]

Estas listas de palabras son la base por medio de la cual WordSmith y algunos programas realizan los procesos de extracción e identificación de términos; dado que estos programas trabajan a partir de listas de palabras; esto es, una vez que todas las palabras del corpus están organizadas, ya sea en orden alfabético o de frecuencia, el programa puede lematizar⁶, identificar las palabras clave, analizar una palabra en su contexto, calcular la probabilidad de que dos palabras puedan aparecer juntas, etcétera, y coadyuvar a la extracción de términos.

No obstante la utilidad de las listas, la ordenación por frecuencia, tal y como se puede apreciar en la ilustración 2, muestra como unidades de mayor aparición palabras que no aportan mayor relevancia al análisis; artículos, preposiciones, conjunciones e incluso verbos, que si bien podrían ser parte de algún término, como en los términos poliléxicos; la aparición de estas palabras no esclarece en nada esta duda y, en cambio, vuelve mayor una lista de palabras, de por sí grande, que hacen más lento el trabajo del terminólogo.

Para solucionar este problema, existe la herramienta *palabras de detención*⁷ (stop list), que cumplen la función de eliminar del conteo estadístico palabras que de antemano se saben que no son útiles o no interesa que aparezcan.

⁶ La lematización es una herramienta que ofrece la posibilidad de simplificar las palabras resultantes mediante el conteo por derivación léxica, esto es, permite mediante procesos lingüísticos unir dos o más unidades a una forma canónica o a alguna otra forma según las necesidades; se pone por caso las palabras *medimos*, *media*, *mide*, y *medirá*, que pueden ser consideradas variaciones de la forma *medir*, y analizadas todas esas formas como una única forma, en este caso, como *medir*.

⁷ La adaptación del término inglés *stop list* es propio.

LA	806	6,17
QUE	473	3,62
EL	424	3,24
EN	332	2,54
Y	272	2,08
SE	261	2,00
A	221	1,69
ES	215	1,64
POR	214	1,64
UN	180	1,38
DEL	176	1,35
UNA	151	1,16
LO	145	1,11
LAS	139	1,06
LOS	111	0,85
PARA	109	0,83
MOVIMIENTO	107	0,82
SI	96	0,75
CONDO	94	0,72
FUERZA	93	0,71
VELOCIDAD	81	0,62
ENERGIA	79	0,60
AL	77	0,59
CON	76	0,58
FIGURA	75	0,57
SOERE	66	0,50
NO	62	0,47
MASA	57	0,44
TIEMPO	54	0,41
EDUCACION	53	0,41
CUERPO	48	0,37
SU	47	0,36
NEWTON	45	0,34
DOS	43	0,33
ESTE	42	0,32
ESTE	42	0,32
SISTEMA	42	0,32
DISTANCIA	41	0,31

Ilustración 2: Lista de palabras por ordenación de frecuencias

3.1.2 Listas de detención

Las listas de detención son archivos que contiene todas aquellas palabras que no se desea que aparezcan en los resultados del conteo estadístico y que, por tanto, mientras el programa efectúa el análisis, se impide que sean tomadas en cuenta las unidades que se encuentren incluidas en la lista de detención; más no así de todos los procesos que el programa realiza, a menos que sean habilitadas tales listas en otros módulos (ilustración 3).

Una lista de detención está compuesta, tal y como Lancaster⁸ lo apunta, por:

⁸ F. W. Lancaster. Vocabulary Control for Information Retrieval. Washington D.C.: Information Resources Press, 1972. p. 154.

“Common words, pronouns, articles, conjunctions, conjunctive adverbs, copula and auxiliary verbs, quantitative adjectives are eliminated from further processing by means of a “stop list.”

[Palabras comunes, pronombres, artículos, conjunciones, adverbios, verbos auxiliares y copulativos, adjetivos cuantitativos son eliminados del proceso por medio de una "lista de detención".]

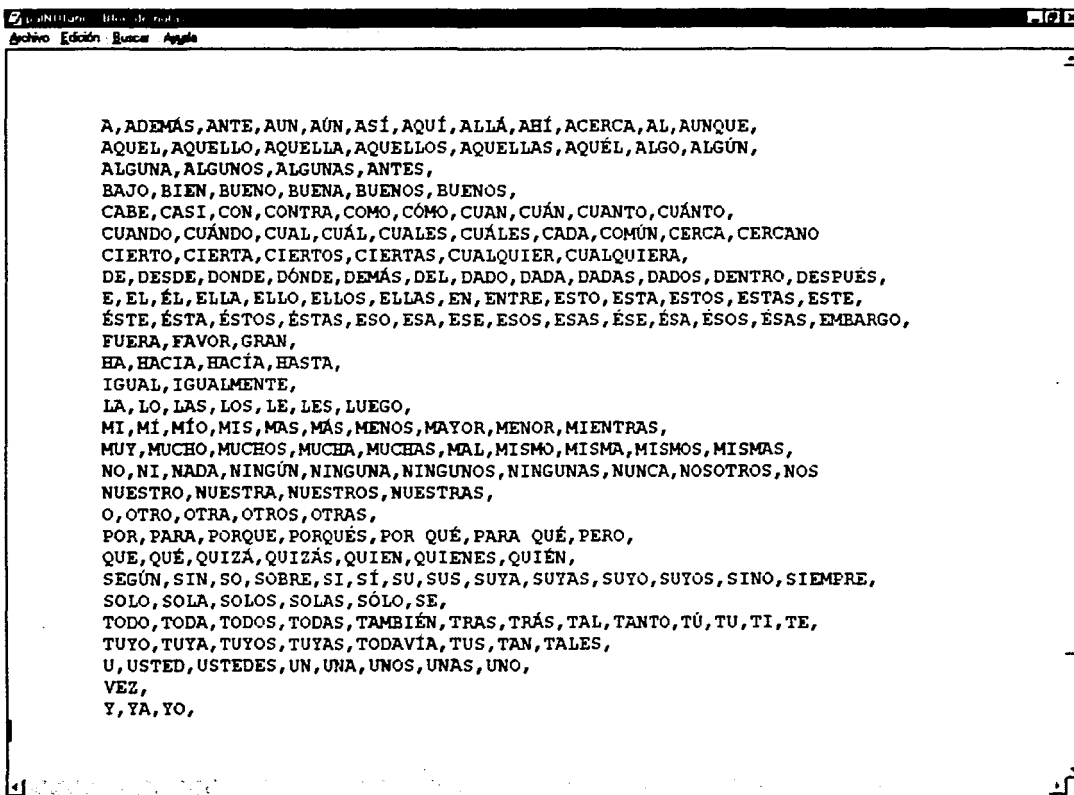


Ilustración 3: Ejemplo de las palabras que constituyen una lista de detención

La aplicación de estas listas logra reducir hasta un 50 por ciento del corpus, según palabras del mismo Lancaster; y el beneficio es obvio e inmediato. Al ser eliminado un porcentaje alto de palabras como artículos, pronombres, etcétera, que no presentan mayor relevancia al análisis, se puede al mismo tiempo y en un porcentaje alto estar seguro de que se trabaja con palabras que representan mayor riqueza de análisis (ver ilustración 4 en donde se aprecia como la lista se enriquece al desaparecer del conteo palabras que no significativas, conservando ahora, en los primeros sitios, palabras que aportan mayor significación al análisis); además, en caso de que una palabra se haya dado por sentado de que no es significativa y en verdad lo sea, el programa puede tomarla en cuenta para otros procesos, como en *palabras clave* o *concordancia*, si la lista de detención no es activada en tales módulos.

	Term	Count	Value
1	MOVIMIENTO	107	0.82
2	FUERZA	93	0.71
3	VELOCIDAD	81	0.62
4	ENERGÍA	78	0.60
5	FIGURA	75	0.57
6	MASA	57	0.44
7	TIEMPO	54	0.41
8	EQUILACIÓN	53	0.41
9	CUERPO	48	0.37
10	NEWTON	45	0.34
11	DOS	43	0.33
12	SISTEMA	42	0.32
13	DISTANCIA	41	0.31
14	X	41	0.31
15	T	40	0.31
16	TIENE	38	0.29
17	FUERZAS	37	0.28
18	FRICCIÓN	36	0.28
19	MOMENTO	36	0.28
20	PODEMOS	36	0.28
21	V	36	0.28
22	FORMA	36	0.27
23	ANGULAR	34	0.26
24	CENTRO	34	0.26
25	LEY	33	0.25
26	CONSERVACIÓN	32	0.24
27	PUEDE	32	0.24
28	TRABAJO	32	0.24
29	OBJETO	30	0.23
30	TENEMOS	30	0.23
31	ANTERIOR	29	0.22
32	CASO	29	0.22
33	PLANO	29	0.22
34	PRINCIPIO	28	0.21
35	DIRECCIÓN	26	0.20
36	SOL	26	0.20
37	ACELERACIÓN	25	0.19
38	TIERRA	25	0.19

Ilustración 4: Resultado de la aplicación de una lista de detención

3.1.3 Palabras clave

Entre la enorme cantidad de palabras que pueden aparecer en una lista de frecuencias, aunque se haga uso de una lista de detención, no siempre dentro del resultado se verá reflejada una terminología total y única de un área específica. Por ejemplo, en la ilustración 4 se observa que *dos*, *tiene* y *podemos*, dentro de un contexto como el de física, aparecen en la lista de frecuencias con un porcentaje de apariciones considerado, pero el resultado no excluye que estas palabras aparezcan en otro contexto temático con un porcentaje que lo manifieste como palabra significativa, y por lo tanto, como candidata a término. Una herramienta bastante útil que permite en gran medida eliminar, o por lo menos poner a consideración si tal palabra, polisémica o no polisémica, es dentro de un área tal, una palabra significativa o no, son las listas de palabras clave.

Estas listas no son otra cosa que una representación de los resultados que se obtienen cuando se comparan dos listas de frecuencia de diferentes áreas de conocimiento; esto es, presentar la relación de palabras significativas que resultan de una comparación entre dos listas de palabras a partir de un análisis de frecuencias del total de palabras que componen la lista (ilustración 5), por ejemplo, si *movimiento* aparece en una lista de frecuencias con un porcentaje de apariciones alto, y en la lista con la que va a compararse, *movimiento* aparece también, pero con muy poco porcentaje o no aparece, inmediatamente *movimiento* será tomada como palabra significativa en una lista de palabras clave.

Dicho de otra forma:

If a certain word occurs significantly more often in a given text than it occurs, on average, in a large selection of “normal” texts (the reference corpus), this word is identified as a keyword.⁹

⁹ Ralf Steinberger. *Cross - lingual Keyword Assignment*. Procesamiento del lenguaje natural. España: SEPLN / Universidad de Jaén, 2001. p. 274.

Palabra	Frecuencia	Promedio	Desviación	Significancia	Valor	Probabilidad
MOVIMIENTO	107	0,82	0		163,3	0,000000
FUERZA	93	0,71	0		141,9	0,000000
VELOCIDAD	81	0,62	2	0,01	107,2	0,000000
FIGURA	75	0,57	1		105,0	0,000000
ECUACIÓN	53	0,41	0		80,8	0,000000
MASA	57	0,44	2	0,01	71,9	0,000000
NEWTON	45	0,34	0		68,6	0,000000
X	41	0,31	0		62,5	0,000000
DISTANCIA	41	0,31	0		62,5	0,000000
T	40	0,31	0		60,9	0,000000
FUERZAS	37	0,28	0		56,4	0,000000
FRICCIÓN	36	0,28	0		54,8	0,000000
Y	36	0,28	0		54,8	0,000000
CENTRO	34	0,26	0		51,8	0,000000
LEY	33	0,25	0		50,3	0,000000
CONSERVACIÓN	32	0,24	0		48,7	0,000000
TIEMPO	29	0,22	7	0,05	47,6	0,000000
PLANO	29	0,22	0		44,2	0,000000
ANGULAR	34	0,26	1		44,0	0,000000
DIRECCIÓN	26	0,20	0		39,6	0,000000
OBJETO	30	0,23	1		38,1	0,000000
TENEMOS	30	0,23	1		38,1	0,000000
TIERRA	25	0,19	0		38,1	0,000000
ACELERACIÓN	25	0,19	0		38,1	0,000000
MOMENTO	26	0,28	3	0,02	37,5	0,000000
GLUCOSA	0		29	0,19	36,5	0,000000
ÁCIDOS	0		29	0,19	36,5	0,000000
ORGANISMO	0		31	0,21	39,1	0,000000
AMINOÁCIDOS	0		36	0,24	45,4	0,000000
NECESIDADES	0		39	0,26	49,2	0,000000
VITAMINAS	0		39	0,26	49,2	0,000000
EMBARAZO	0		45	0,30	55,7	0,000000
DIETA	0		52	0,35	65,6	0,000000
ALIMENTOS	0		53	0,36	66,8	0,000000
NUTRIENTES	0		61	0,41	76,9	0,000000
HTM	0		77	0,52	97,1	0,000000
PROTEÍNAS	0		63	0,56	104,7	0,000000

Ilustración 5: Lista de palabras clave

[Si cierta palabra ocurre significativamente más a menudo en un texto dado de lo que ocurre, en promedio, en una selección grande de textos "normales" (el corpus de referencia), esta palabra se identifica como una palabra clave.]

Para obtener una lista de palabras clave es necesario contar con dos listas de palabras de diferentes áreas de conocimiento, esto, con el fin de eliminar en un porcentaje alto las posibles coincidencias que una o varias palabras pudiesen tener en ambas listas, como el ejemplo de *movimiento*, el cual pudiese aparecer tanto en física como en poesía; y ya que estos inconvenientes no pueden eliminarse radicalmente, es recomendable elaborar

más listas de palabras de diferentes áreas con que comparar la lista principal, para así reducir al mínimo el margen de coincidencias que pudiesen presentarse.

Una vez tenidas las listas de palabras, con las herramientas que en este caso WordSmith ofrece, se realiza una comparación de listas de palabras que mostrará en los resultados un grupo de palabras significativas de cada lista; estas palabras representan la columna vertebral del texto analizado, puesto que muestran en esta serie de palabras clave, a grandes rasgos, el contenido del texto; tal y como sucede en *Recuperación de Información*, donde esta herramienta es utilizada en gran medida.

3.1.4 Concordancia

A concordance, in its simplest form, is an alphabetical listing of the words in a text, given together with the contexts in which they appear. The most common form of concordance today is the Keyword-in-Context (KWIC) index, in which each word is centered in a fixed-length field.¹⁰

[Una concordancia, en su forma más simple, es una inscripción alfabética de las palabras en un texto, dado junto con los contextos en los que aparecen. La forma más común de concordancia hoy en día es el índice de Palabra clave-en-contexto, en el que cada palabra se centra en un campo posición-longitud.]

Lara señala en la investigación para la creación del *DEM* que “no sólo [ies] interesa obtener del análisis de[I] corpus la lista alfabética de los vocablos incluidos en él, sino que también [interesa] conocer los contextos en que aparece utilizada cada palabra”¹¹. Esto es, ver en los resultados cómo se comporta una palabra dada, a través del análisis del

¹⁰ Para mayores referencias sobre el tema y sobre KWIC, Véase:

<http://www.georgetown.edu/cball/corpora/tutorial3.html>

¹¹ Véase Luis Fernando Lara, Roberto Ham Chande y María Isabel García Hidalgo. Op.cit. p. 30.

contexto en el que aparece, es decir, aplicar los procesos que la herramienta de concordancia ofrece.

servación del impetu lineal	ecuación de movimiento	ecuación del movimiento pl	254	d:\msdoc-1\vesis\docume-1\hpart-1\ejemplo.txt	96
cción del momento angular	dirección del movimiento	distancia Espacio o tiempo	256	d:\msdoc-1\vesis\docume-1\hpart-1\ejemplo.txt	84
se necesitaba una fuerza para mantener un movimiento a un objeto que se mueve por u			13	d:\msdoc-1\vesis\docume-1\hpart-1\mvec3.txt	2
podemos explicar varias de las causas del movimiento al tener un principio causal expli			767	d:\msdoc-1\vesis\docume-1\hpart-1\mvec5.txt	81
e sobre de el y lo modifique. La cantidad de movimiento angular se relaciona con el me			258	d:\msdoc-1\vesis\docume-1\hpart-1\mvec3.txt	35
			0	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	0
			436	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	59
			84	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	12
			418	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	56
			167	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	24
			37	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	7
			7	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	1
			18	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	3
			0	d:\msdoc-1\vesis\docume-1\hpart-1\mvec4.txt	0
			130	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	19
			647	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	50
			1365	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	98
			71	d:\msdoc-1\vesis\docume-1\hpart-1\mvec3.txt	11
			145	d:\msdoc-1\vesis\docume-1\hpart-1\mvec2.txt	25
			33	d:\msdoc-1\vesis\docume-1\hpart-1\mvec3.txt	6
			102	d:\msdoc-1\vesis\docume-1\hpart-1\ejemplo.txt	34
			77	d:\msdoc-1\vesis\docume-1\hpart-1\mvec12.txt	5
			660	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	51
			313	d:\msdoc-1\vesis\docume-1\hpart-1\mvec2.txt	38
			78	d:\msdoc-1\vesis\docume-1\hpart-1\mvec5.txt	8
			423	d:\msdoc-1\vesis\docume-1\hpart-1\mvec5.txt	46
			1098	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	82
			803	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	61
			778	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	59
			732	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	56
			736	d:\msdoc-1\vesis\docume-1\hpart-1\mvec5.txt	77
			717	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	55
			652	d:\msdoc-1\vesis\docume-1\hpart-1\mvec8.txt	50
			312	d:\msdoc-1\vesis\docume-1\hpart-1\mvec3.txt	42
			119	d:\msdoc-1\vesis\docume-1\hpart-1\mvec10.txt	17
			138	d:\msdoc-1\vesis\docume-1\hpart-1\mvec4.txt	24
			378	d:\msdoc-1\vesis\docume-1\hpart-1\mvec5.txt	41
			787	d:\msdoc-1\vesis\docume-1\hpart-1\mvec11.txt	39

Ilustración 6: Análisis de concordancia de la palabra *movimiento*

Los procesos que en concordancia pueden realizarse involucran el análisis de las palabras en su contexto, esto es, una vez seleccionada una palabra *X*, ésta es tomada desde su contexto para ser analizada individualmente y ver el comportamiento que sufre, ver alrededor de qué palabras se encuentra y cuáles son las que más a menudo aparecen junto a ella (ilustración 6).

3.1.5 Información mutua

Una herramienta más que se tiene a disposición tanto en extracción automática de términos como en otras disciplinas como recuperación y extracción de información, es la fórmula tomada de ciencias de la computación, denominada información mutua (Mutual Information), la cual permite identificar qué tan conectadas se encuentran dos palabras que ocurren contiguamente en un texto dado. Esta fórmula evalúa, para un par de palabras dadas, la probabilidad de que ambas ocurran juntas como un evento conjunto, en comparación con la probabilidad de que éstas ocurran individualmente, de tal manera que su contigüidad haya sido dada por pura casualidad¹².

Word	Description	potencia	luz	distancia	rio	ley
ABAJA	1 AB se aproxima a una recta. Co					
ABANDONÓ	4 abajo la distancia D y regresa c					
ABSOLUTA	1 abandonó esta idea y trató de aj					
AC	1 absoluta de un objeto ya que to					
ACABABA	4 ac. La aceleración es: co					
ACABÓ	1 acababa de entender cómo podí					
ACCESIBLE	1 acabó por generar el mito de qu					
ACCION	1 accesible y conveniente. Por un					
ACCIONES	14 acción y reacción aceleraci	potencia 4,79	luz 4,58	distancia 4,13	rio 3,99	ley 3,86
ACELERA	1 acciones instantáneas a distanc					
ACELERACION	1 acelera al movimiento. El peso					
ACELERACIONES	25 aceleración. Aumento de la velo	centrípetas 5,07	diferente 4,48	q 4,26	ahora 4,26	acuerdo 4,07
ACELERAR	1 aceleraciones en el sistema vale					
ACELERARSE	11 acelerado por lo que tenemos ti	tormentoso 7,25	rectilíneo 5,66	curva 5,66	aje 4,93	uniforme 4,83
ACELERAR	1 aceleran por la acción de la fuer					
ACELERARSE	1 acelerar. En un instante de tiem					
ACENSO	1 acelerarse ningún objeto materia					
ACEPTABA	1 acenso sea cero. Por lo que hue					
ACEPTAR	1 aceptaba su existencia sin ning					
ACEPTAR	2 aceptar la especulación filosófi					
ACEPTAN	1 acercan al cuerpo el momento d					
ACERAR	1 acercar las cosas al cuerpo la c					
ACERAR	4 acero sobre un rol de canal					
ACERAR	2 aclaración. Sin embargo, la acta					
ACERAR	1 aclarar, ya que, cuando salió a					
ACORTA	1 acorta y el observador externo v					
ACTIVIDAD	1 actividad era difícil de aceptar, s					
ACTIVOS	1 activos" o "atractivos". Esta idea					
ACTO	1 acto que produce la acción sobr					
ACTUA	10 actúa sobre el móvil es $F = ma$,	ma 6,07	magnitud 5,80	aplicada 5,07	fuerza 4,17	momento 3,22
ACTUALICE	1 actualice. 3 El acto que produ					
ACTUAN	4 actúan sobre las partículas es ig					
ACTUANDO	1 actuando sobre el sistema, el c					
ACTUAR	1 actuar sobre la potencia para lo					
ACTUJE	1 actué sobre el sistema, entonc					
ACTUEN	1 actúen sobre el sistema, entonc					
ACUERDO	R acuerdo con la firma (11) se an	analizamos 6,39	aceleración 5,07	anterio 3,86	temos 3,80	fuerza 3,80

Ilustración 7: Ejemplo del análisis de información mutua

¹² Sobre este tema y sobre ingeniería lingüística cfr. el manuscrito La informática en el trabajo terminológico. Sierra, 2002.

La herramienta de información mutua trabaja, como se describe en el análisis que Cabré, Estopà y Vivaldi hacen a los diversos programas de extracción de términos, a partir de reglas estadísticas y describen la información mutua como¹³:

Statistically based systems function by detecting two or more lexical units whose occurrence is higher than a given level. This is not a random situation, but it is related to a particular usage of these lexical units. This principle, called Mutual Information.

[Los sistemas basados en estadística funcionan detectando dos o más unidades léxicas, las cuales, ocurren constantemente más arriba de un nivel establecido. Ésta no es una situación aleatoria, sino que se relaciona a un uso particular de esas unidades léxicas (ilustración 7). El principio que rige estas ocurrencias se conoce como Información Mutua.]

La utilidad de esta herramienta en este trabajo, consiste en que:

- permite extraer términos de más de una palabra;
- ayuda a identificar diferentes sentidos de la palabra, esto con referencia al aspecto polisémico de las palabras.¹⁴

Dentro de la investigación, esta herramienta, aunque no ofreció grandes aportes de los ya ofrecidos por las anteriores herramientas, puede mostrar algunos resultados bastante interesantes además de que abre otras líneas de investigación para quien esté interesado en el tema.

¹³ María Teresa Cabré, Rosa Estopà, Jordi Vivaldi. Recent advances in computational terminology. Amsterdam / Philadelphia: John Benjamins B. V., 2001. p. 55.

¹⁴ Para profundizar en este tema y ver algunos resultados interesantes relacionados con información mutua y otras medidas de extracción, consultar el siguiente enlace:
<http://www.ling.lancs.ac.uk/monkey/ihe/linguistics/corpus3/3fra1.htm>

3.2 Programas existentes

Esta necesidad por extraer términos de manera automática surge a finales de los años ochenta; la aparición del programa TERMINO en 1990, marca un avance significativo en el área de extracción terminológica vía procesos automáticos y muestra resultados alentadores¹⁵. Muchos son ahora los grupos¹⁶ que tienen interés sobre el tema y que se encuentran en el desarrollo de recursos para aprovechar los estudios, programas y la infraestructura existentes, para dar a conocer nuevas técnicas que redunden en el beneficio de los terminólogos, de los profesionales que necesitan una terminología inequívoca en sus labores, de los estudiantes e incluso de la misma gente que carece de elementos con que poder entender estas jergas.

Sin embargo, existen muchos problemas que aún no encuentran solución, entre estos problemas, los principales se refieren a:

1. "identification of complex terms, that is, determining where a terminological phrase begins and ends;
 2. recognition of complex terms, that is, deciding whether a discursive unit constitutes a terminological phrase or free unit;
 3. identification of the terminological nature of a lexical unit, that is, knowing whether in a specialised text a lexical unit has a terminological nature or belongs to general language and
 4. appropriateness of a terminological unit to a given vocabulary (this has scarcely been addressed from the point of view of automatization)"¹⁷.
-
1. [la identificación de términos complejos, es decir, determinar dónde empieza y acaba una frase terminológica;

¹⁵ Véase María Teresa Cabré, Rosa Estopà, Jordi Vivaldi. Op. Cit. p. 54.

¹⁶ Véase capítulo 1. *El área terminológica en nuestros días.*

¹⁷ María Teresa Cabré, Rosa Estopà, Jordi Vivaldi. Op. Cit. p. 54.

2. el reconocimiento de términos complejos, es decir, decidir si una unidad discursiva constituye una frase terminológica o una unidad libre;
3. la identificación de la naturaleza terminológica de una unidad léxica, es decir, saber si en un texto especializado una unidad léxica tiene una naturaleza terminológica o pertenece al idioma general y
4. la adecuación de una unidad terminológica a un vocabulario dado (esto se ha dirigido escasamente desde el punto de vista de automatización).]

Otro gran problema de los extractores automáticos de términos es, como señala Estopà, que “la mayoría de extractores de terminología se han centrado en la recuperación de las unidades terminológicas poliléxicas a través de estrategias formales basadas en sus patrones morfosintácticos”¹⁸. Y como apunta Vivaldi, “ningún sistema comprueba si el término propuesto pertenece o no al ámbito de especialidad que se está analizando”¹⁹; de ahí que salga a colación la necesidad del uso correcto de las herramientas que dichos programas ofrecen, puesto que la labor del terminólogo se incrementa en la medida en que el programa presente estas carencias o en que estas carencias se hagan más grandes si no se hace una explotación apropiada de los recursos existentes.

Mas con todo, el trabajo ahí está y son constantes las investigaciones y el interés por aprovechar las ventajas que la extracción automática de términos presenta.

Ahora bien, el interés en este apartado es mencionar algunos de los programas que tienen que ver directa o indirectamente con la extracción de términos. Y he dicho

¹⁸ Rosa Estopà. *Elementos lingüísticos de las unidades terminológicas para su extracción automática. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001. p. 68.

¹⁹ Jorge Vivaldi. *Elaboración de una aplicación automática de reconocimiento y extracción de información terminológica en textos de dominios restringidos. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001. p. 230.

mencionar, porque por cuestiones de tiempo, espacio y por no salir del tema que ahora nos ocupa y entrar en otro, se hace imposible una evaluación de tales programas; mas para quien esté interesado se indican enlaces en donde se puede encontrar información detallada y evaluaciones de los programas aludidos.

Cabe señalar que los programas referidos no siguen orden alguno para su inclusión en este trabajo, ni son todos los que existen en el mercado; es sólo una muestra representativa de los recursos disponibles, algunos de ellos incluso son gratuitos en Internet.

- MonoConc Pro:
<http://www.ruf.rice.edu/~barlow/mono.html#monopro>
- TACT Text Analysis Computing Tools:
<http://etext.lib.virginia.edu/helpsheets/tact.html>
- CONC:
<http://www.sil.org/computing/conc/>
- Oxford Concordance Program (OCP):
<http://www1.oup.co.uk/E-P/Humanities/Micro-OCP/>
<http://info.ox.ac.uk/ctitext/resguide/resources/o125.html>
- ParaConc:
<http://www.ruf.rice.edu/~barlow/parac.html>
- Hyperbase:
<http://info.ox.ac.uk/ctitext/resguide/resources/h123.html>
<http://lolita.unice.fr/~brunet/hyperbase.html>
- Lexa 6.0:
<http://info.ox.ac.uk/ctitext/resguide/resources/l120.html>
- ANA²⁰
- CLARIT

²⁰ A partir de este programa y hasta el último señalado, se puede encontrar una evaluación detallada por terminólogos y referencias hacia los documentos originales en: Cabré, María Teresa, Rosa Estopà, Jordi Vivaldi. Recent advances in computational terminology. Op. Cit. pp. 57-84.

- Daille-94
- FASTR
- LEXTER
- NAULLEAU
- NEURAL
- NODALIDA-95
- TERMIGHT
- TERMINO
- TERMS

3.2.1 WordSmith

Como arriba queda sentado, el programa con el cual fueron efectuados todos los análisis de esta investigación es WordSmith. Dicho programa es un software desarrollado por Mike Scott en la Oxford University Press en la década de los 90,

La elección de este programa no obedece a ningún análisis previo de los beneficios y potencialidades que ofrece; tampoco fue por interés comercial; o porque sea el mejor y más completo programa. Simplemente, fue el que se tenía disponible en el proyecto del cual se desprende esta investigación, y como arriba se señala, no es éste el lugar ni el trabajo para evaluar los programas y elegir el mejor. Es un recurso disponible que para los intereses de este trabajo y del proyecto base ha funcionado, pero no se ha estudiado si es mejor, igual o peor que otros. Sin embargo, para el desarrollo de esta investigación ha representado un medio eficiente, aceptable y aplicable para llevar a cabo los fines planeados en el área de física y en otras áreas de conocimiento, y por lo tanto, válido para ser usado en esta investigación.

WordSmith es un programa que se encuentra dividido en tres módulos básicos de donde se desprenden las distintas operaciones que permite:

1. lista de palabras
2. palabras clave
3. concordancia

En estos módulos el usuario puede hacer uso de las herramientas con que el programa cuenta para reducir, dirigir y, en cierto sentido, comprobar los resultados que el programa arroja mediante los propios recursos del programa. Entre los procesos con que cuenta WordSmith para extraer términos y que fueron utilizados dentro de esta investigación se encuentran:

- las listas de palabras de una o más palabras,
- elaboración de listas de detención aplicables a los tres módulos o alguno de ellos solamente,
- lematización automática y manualmente,
- etiquetado de palabras,
- comparación ente listas de palabras para obtener palabras clave,
- análisis por medio de agrupaciones,
- análisis gráfico de la posición que guarda una palabra dentro de un archivo,
- observación del comportamiento de una palabra a través de colocaciones,
- análisis de una palabra en su contexto,
- creación de índices de palabras,
- aplicación de información mutua,
- etcétera.

Algunos de estos procesos ya han sido descritos, algunos otros se detallarán más adelante, como las colocaciones y el análisis por medio de agrupaciones, los cuales sirvieron de pauta junto con la elaboración de listas de palabras de dos y tres unidades para la obtención de los difíciles términos poliléxicos, de los que tanto dicen, es arduo conseguir de forma automática.

Fue la aplicación de WordSmith en esta actividad, a resumidas cuentas, productiva, valiosa e interesante, puesto que, como más adelante se explicará, se obtuvieron una buena cantidad de términos simples y poliléxicos, simplificó, aunque sin llegar a ser automático, el trabajo de extracción y abrió líneas de investigación prometedoras en el área de ingeniería lingüística.

4. OBTENCIÓN DE LA TERMINOLOGÍA

4.1 Delimitación del objeto de estudio

Como en la introducción queda asentado, el fin último del presente trabajo consiste en presentar una terminología básica, obtenida con herramientas informáticas y avalada por expertos, relativa a la física a nivel bachillerato; esta investigación se desprende de un proyecto que el *GIL* realiza para implementar un diccionario de búsquedas onomasiológicas en el área de física. Una de las metas del proyecto consiste en obtener la terminología relativa al área de mecánica. Para lo cual, se realizó esta investigación y se desprendió el presente trabajo, el cual se enfoca hacia la extracción de los términos de mecánica en física. Las siguientes metas en este rubro terminológico, y que no son tratadas en esta tesis, pretenden probar la metodología empleada en este trabajo, y según los resultados obtenidos, implementar los trabajos de extracción de términos a completar todas las disciplinas que conforman la física y extenderlos hacia otras áreas de conocimiento como la lingüística, la química, etc.

Debido a las necesidades y premisas del proyecto del *GIL*, y como en la introducción se menciona, el objeto de estudio y los resultados de esta investigación se ceñirán únicamente a la terminología de mecánica del área de física.

De esta forma, se reafirma el objeto de estudio de esta investigación, el cual se limitará a los términos concernientes a la mecánica; dicho lo anterior, es conveniente recapitular sobre lo que hasta ahora se ha desarrollado en este estudio.

En los capítulos precedentes se han presentado los fundamentos teóricos de la materia prima de esta investigación, terminología y términos, direccionando el estudio que se les dará a éstos, con un enfoque visto desde la disciplina de la ingeniería lingüística. Dentro de esta área, se explotará el recurso de la extracción de términos con la metodología

descrita en el capítulo 2 y, mediante el uso de las herramientas explicadas en el capítulo 3, contenidas en el programa WordSmith, se describirán en este capítulo los procesos llevados a cabo para efectuar los análisis que arrojarán la terminología que se pretende obtener mediante el uso de los recursos antes mencionados.

4.1.1 Cómo se obtuvo el material en formato electrónico

Para desempeñar el propósito de este estudio y cumplir con los requisitos que deben contener las terminologías, esto es, ser palabras que dentro de cierto dominio estén adscritas a un criterio especializado, los expertos en física del Centro de Instrumentos de la UNAM proporcionaron un grupo de quince archivos en formato electrónico y con la extensión txt (texto), con información relativa y especializada a mecánica.

El material proporcionado fue elaborado con los fines de representar los recursos terminológicos y en algunos casos conceptuales del proyecto del *GIL*.

Tal material es la raíz de la información con que se contó en esta investigación, y a partir de él, fueron realizados todos los análisis que aquí tienen lugar. EL grupo de quince archivos que proporcionó el Centro de Instrumentos fue denominado *hipertexto*.

4.1.2 El hipertexto

El hipertexto¹ es la conformación de los quince archivos proporcionados por el Centro de Instrumentos, compuesto de 76,991 palabras (ilustración 1). Fue conformado por especialistas de física con la consigna de que el material que facilitasen debería ser exclusivo para la enseñanza y el estudio de la física en el nivel bachillerato, sin que esto excluya que deban ser textos especializados en el área, en este caso, de mecánica.

¹ Es preciso señalar que toda la información que contenía el hipertexto fue proporcionada bajo la premisa de que el diccionario estaría dirigido a los alumnos del nivel bachillerato. Por lo tanto, no se espere encontrar en la terminología resultante términos que pertenezcan a un dominio más profundo de la física y de sus subáreas.

Una vez obtenido el material del que partirían los análisis se iniciaron los procesos que conforman la fase de extracción.

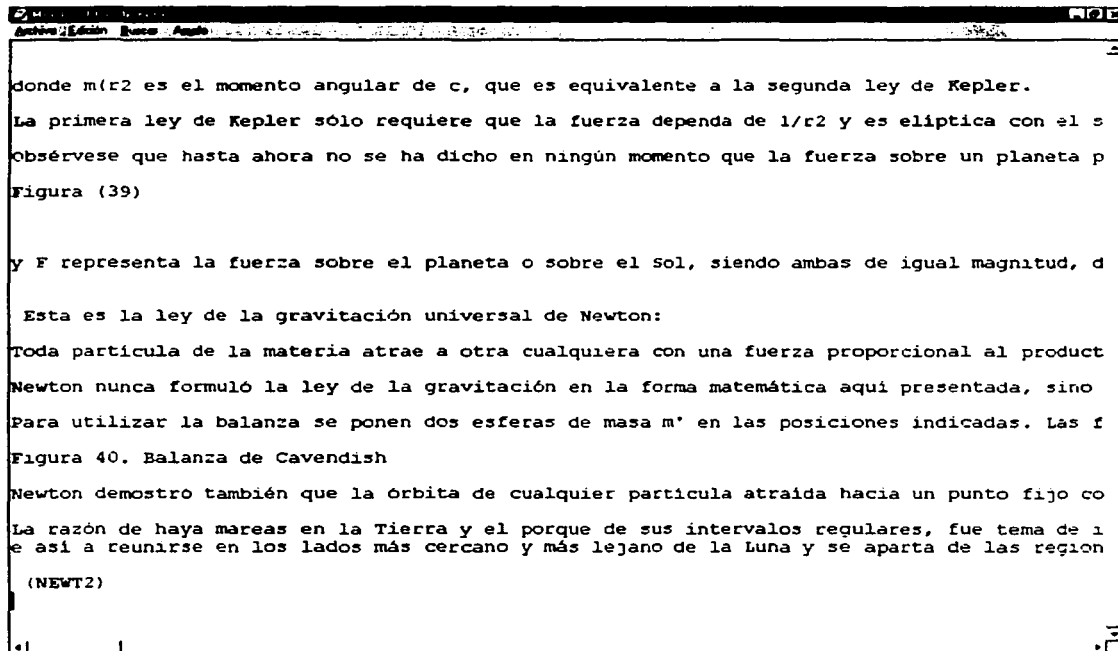


Ilustración 1: Fragmento de un archivo del hipertexto

Cabe hacer mención de que el análisis del hipertexto nunca fue manual, en ningún momento se dio una lectura de la información que contenía, como tampoco se hizo en los materiales conseguidos para hacer comparaciones con éste. En este sentido, se realizó un proceso automático.

A continuación se detallarán los procesos realizados² que condujeron a obtener la terminología que se presentará en el capítulo 5.

4.2 Listas de palabras

4.2.1 Procesos para la obtención de listas de palabras

Con el hipertexto empezaron los primeros análisis que se efectuaron con WordSmith. Se procedió primero a elaborar una lista de palabras que mostrara el contenido del hipertexto. Esta lista era grande en exceso, pero mostraba palabras interesantes como *movimiento, fuerza, velocidad, energía* en posiciones de frecuencias no muy bajas; mas como en esta primera lista no se había activado una lista de detención, la cantidad de palabras, significativas y funcionales, que componían al hipertexto, resultaba considerable.

Con la idea de reducir esta lista a sus elementos más significativos, se decidió eliminar todas aquellas palabras que mediante criterios personales resultaban inútiles, esto se logró al incluir en una lista de detención todas las palabras que fueron consideradas poco relevantes en el conteo estadístico del hipertexto.

Es preciso resaltar que en el análisis de esta lista de palabras sólo se buscó, mediante las herramientas de WordSmith, obtener una lista de términos simples, por lo que la lista presenta el conteo sólo por palabra; por lo que respecta a los términos poliléxicos, más adelante se explicará que se efectuaron listas de palabras de dos o más palabras que junto con los análisis por medio de agrupamientos, permitieron obtener la terminología poliléxica.

² Hubo toda una serie de procesos y análisis de prueba realizados con WordSmith para ver el funcionamiento de éste y los posibles resultados que arrojaba, tanto en mecánica como en otras áreas, sin embargo, tales análisis no serán detallados en este lugar.

4.2.2 Activación de la lista de detención

El siguiente paso consistió en activar una lista de detención compuesta por 250 palabras, entre las que se encuentran pronombres, artículos, preposiciones, conjunciones y los verbos ser y estar, los cuales fueron incluidos respondiendo a los análisis de prueba³ que para este fin se hicieron en WordSmith.

A, ADEMÁS, ANTE, AUN, AÚN, ASÍ, AQUÍ, ALLÁ, AHÍ, ACERCA, AL, AUNQUE, AQUEL, AQUELLO, AQUELLA, AQUELLOS, AQUELLAS, AQUEL, ALGO, ALGÚN, ALGUNA, ALGUNOS, ALGUNAS, ANTES, BAJO, BIEN, BUENO, BUENA, BUENOS, BUENAS, CABE, CASI, CON, CONTRA, COMO, CÓMO, CUAN, CUÁN, CUANTO, CUÁNTO, CUANDO, CUÁNDO, CUAL, CUÁL, CUALES, CUÁLES, CADA, COMÚN, CERCA, CERCAÑO, CIERTO, CIERTA, CIERTOS, CIERTAS, CUALQUIER, CUALQUIERA, DE, DESDE, DONDE, DÓNDE, DEMÁS, DEL, DADO, DADA, DADOS, DADOS DENTRO, DESPUÉS, E, EL, ÉL, ELLA, ELLO, ELLOS, ELLAS, EN, ENTRE, ESTO, ESTA, ESTOS, ESTAS, ESTE, ÉSTE, ÉSTA, ÉSTOS, ÉSTAS, ESO, ESA, ESE, ESOS, ESAS, ÉSE, ÉSA, ÉSOS, ÉSAS, EMBARGO, FUERA, FAVOR, GRAN, HA, HACIA, HACÍA, HASTA, IGUAL, IGUALMENTE, LA, LO, LAS, LOS, LE, LES, LUEGO, MI, MÍ, MÍO, MIS, MAS, MÁS, MENOS, MAYOR, MENOR, MIENTRAS, MUY, MUCHO, MUCHOS, MUCHA, MUCHAS, MAL, MISMO, MISMA, MISMOS, MISMAS, NO, NI, NADA, NINGÚN, NINGUNA, NINGUNOS, NINGUNAS, NUNCA, NOSOTROS, NOS NUESTRO, NUESTRA, NUESTROS, NUESTRAS, O, OTRO, OTRA, OTROS, OTRAS, POR, PARA, PORQUE, PORQUÉS, POR QUÉ, PARA QUÉ, PERO, QUE, QUÉ, QUIZÁ, QUIZÁS, QUIEN, QUIENES, QUIÉN, SEGÚN, SIN, SO, SOBRE, SI, SÍ, SU, SUS, SUYA, SUYAS, SUYO, SUYOS, SINO, SIEMPRE, SOLO, SOLA, SOLOS, SOLAS, SÓLO, SE, TODO, TODA, TODOS, TODAS, TAMBIÉN, TRAS, TRÁS, TAL, TANTO, TÚ TU, TI, TE, TUYO, TUYA, TUYOS, TUYAS, TODAVÍA, TUS, TAN, TALES, U, USTED, USTEDES, UN, UNA, UNOS, UNAS, UNO, VEZ, Y, YA, YO,

Ilustración 2: Palabras funcionales de la lista de detención aplicada al hipertexto

³ Ver nota 2.

El fin pretendido al activar la lista de detención fue el de reducir a un número más significativo las palabras del hipertexto, y sólo fue activada en el módulo de listas de palabras, lo que significa que para los análisis de los siguientes módulos (palabras clave y concordancia) todas las palabras incluidas en la lista de detención quedaron habilitadas para ser tomadas en cuenta.

La ilustración 2 muestra las palabras incluidas en la lista de detención aplicada al análisis del hipertexto.

4.2.3 Pulido de las listas de palabras

El resultado después de la aplicación de esta lista de detención significó una lista más compacta de palabras, aunque esto no quiere decir que hayan quedado sólo palabras significativas; el total de palabras de que se compuso esta segunda lista, llamada *mecánica*, es de 1674 palabras.

Las palabras que se obtuvieron en *mecánica* constataban la información que la gente del Centro de Instrumentos había dado, que el hipertexto contenía información especializada sobre mecánica; un análisis más detallado de los resultados de *mecánica*, revelaba que gran parte de las palabras que la componían y que se ubicaban entre las frecuencias más bajas, esto es, que contaban con menos de 10 en aparición, se trataban de verbos y sustantivos que acompañan al discurso, lo que descartaba gran parte de la lista, sin embargo, se tuvo la precaución de revisar a fondo las 1674 palabras que componían *mecánica* para que ninguna de ellas pasase desapercibida. De esta forma, casi un tercio de las palabras obtenidas parecía ser en verdad significativo.

Para corroborar esta idea, mejorar los supuestos que se tenían y aprovechar los recursos disponibles, se hizo uso del módulo dedicado a extraer palabras clave.

Número	Palabra	Frecuencia	Número	Palabra	Frecuencia
1	MOVIMIENTO	107	39	CONSTANTE	24
2	FUERZA	93	40	M	24
3	VELOCIDAD	81	41	R	23
4	ENERGÍA	78	42	F	21
5	FIGURA	75	43	PUNTO	21
6	MASA	57	44	CINÉTICA	20
7	TIEMPO	54	45	CUERPOS	20
8	ECUACIÓN	53	46	ECUACIONES	20
9	CUERPO	48	47	INCLINADO	20
10	NEWTON	45	48	ALTURA	19
11	DOS	43	49	CANTIDAD	19
12	SISTEMA	42	50	MUESTRA	19
13	DISTANCIA	41	51	PARTÍCULAS	19
14	X	41	52	DECIR	18
15	T	40	53	VALOR	18
16	TIENE	38	54	D	17
17	FUERZAS	37	55	ÍMPETU	17
18	FRICCIÓN	36	56	LEYES	17
19	MOMENTO	36	57	LINEAL	17
20	PODEMOS	36	58	MUEVE	17
21	V	36	59	POTENCIAL	17
22	FORMA	35	60	CERO	16
23	ANGULAR	34	61	H	16
24	CENTRO	34	62	MECÁNICA	16
25	LEY	33	63	PROBLEMA	16
26	CONSERVACIÓN	32	64	TOTAL	16
27	PUEDE	32	65	UNIFORME	16
28	TRABAJO	32	66	VE	16
29	OBJETO	30	67	RÍO	15
30	TENEMOS	30	68	ACCIÓN	14
31	ANTERIOR	29	69	ÁNGULO	14
32	CASO	29	70	DEBE	14
33	PLANO	29	71	DICE	14
34	PRINCIPIO	28	72	ESCRIBIR	14
35	DIRECCIÓN	26	73	GALILEO	14
36	SOL	26	74	HEMOS	14
37	ACELERACIÓN	25	75	LADO	14

Ilustración 3: Lista de las primeras 75 palabras después de la aplicación de la lista de detención

El resultado que se obtuvo con el uso de esta herramienta se detalla en el siguiente apartado.

4.3 Palabras clave

4.3.1 Elaboración de listas de palabras alternas

En el análisis para la obtención de las palabras clave de *mecánica* se realizó un proceso de comparación entre listas de palabras; para conseguir este fin, se partió del criterio de alejamiento entre áreas de conocimiento. Con este criterio como premisa para realizar las listas de palabras con que se compararía *mecánica*, se obtuvieron de Internet dos archivos, de los cuales, el primero estaba dedicado a *Octavio Paz*⁴ (ilustración 4); y el segundo, a un artículo sobre *salud alimenticia*⁵; compuestos por 27,844 y 14,966 palabras respectivamente, los cuales fueron sometidos a los mismo procesos que sufrió el hipertexto.

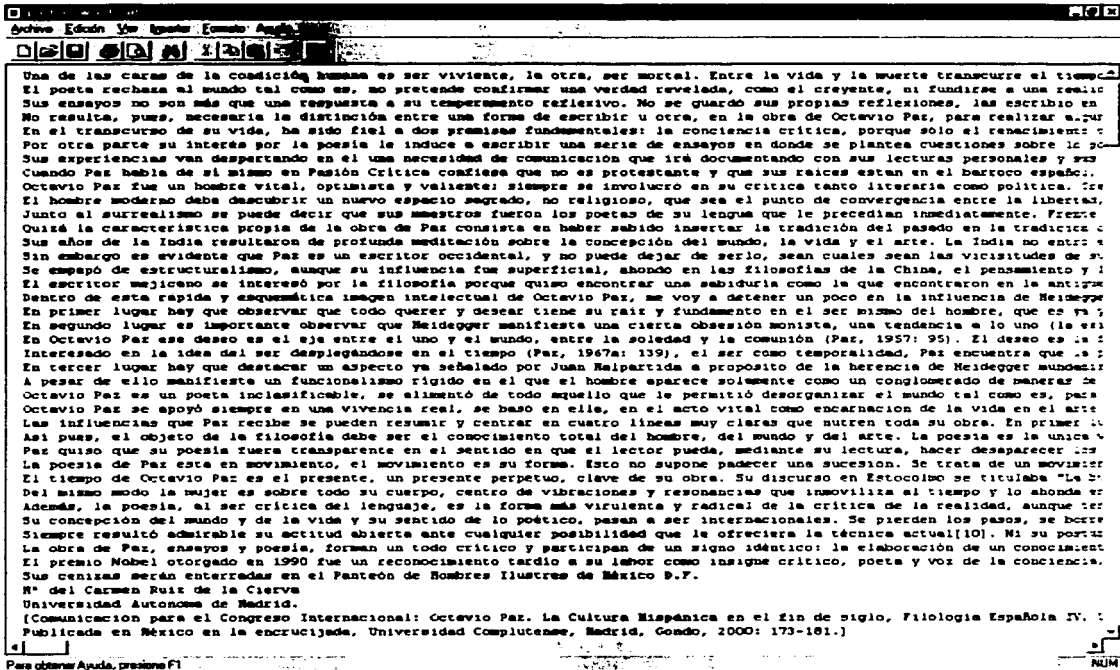


Ilustración 4: Fragmento de uno de los archivos obtenidos de Internet

⁴ María del Carmen Ruiz de la Cierva. *Imagen intelectual de Octavio Paz*. [En línea]. En: Proyecto Ensayo Hispanoamericano (2002). <<http://ensayo.rom.uga.edu/filosofos/mexico/paz/ruiz/>>

⁵ Guía de alimentación y salud [En línea]. En: Sociedad española de nutrición básica y aplicada (2002). <<http://www.uned.es/pea-nutricion-y-dietetica-l/guia/index.htm>>

4.3.2 Comparación entre las diferentes listas de palabras

Una vez obtenidas las listas de palabras de los archivos⁶ bajados de la red (de aquí en adelante se hará referencia a estas listas como *Paz y alimentación*), se efectuó un proceso de comparación entre listas.

Este proceso de búsqueda de palabras clave permite que dos listas sean equiparadas a través de procesos estadísticos y matemáticos que actúan estableciendo que si una palabra aparece en un texto *X* pero no aparece en un texto *Y* o aparece muy poco en relación con el texto *X*, entonces esta palabra será clave en el texto *X*, dado que el contraste de aparición en los textos *X* y *Y* supone que tal palabra debe ser significativa; pero si esta misma palabra aparece en ambos textos, el proceso las anulará de forma automática, si es significativa en ambos, bajo la regla de que como su aparición es regular en uno y otro texto, la posibilidad de que sea significativa se reduce sobremanera.

En capítulos anteriores se hace mención al problema polisémico de las palabras *movimiento* y *síntesis*, las cuales pueden ser significativas en distintas áreas dependiendo del sentido pragmático de su uso; para remediar en parte este problema, se sugirió la aplicación de más de una lista de comparación y que tales listas sean de distintas áreas para reducir al mínimo el problema de la polisemia. El resultado que en este trabajo se dio tras la comparación de *Paz y alimentación* contra *mecánica* muestra una serie de datos que a continuación se anotan.

⁶ Conviene hacer mención que los dos archivos obtenidos de la red con que se comparó *mecánica* son más grandes en la extensión de su corpus; esto permite que el número de comparaciones de las palabras que conforman *mecánica* sea mayor y las palabras clave resultantes más específicas y confiables.

4.4 Índice de posibles términos simples

El proceso de comparación abarcó cuatro cotejos con distintas variantes de búsqueda de las listas *Paz* y *alimentación*. En total, a *mecánica* se le realizaron ocho comparaciones. El número de palabras clave resultantes varía de acuerdo con los rangos de búsqueda aplicados en cada comparación.

La ilustración 5 muestra los rangos de búsqueda que se aplicaron en las comparaciones de *mecánica* en contra de *alimentación* y *paz*:

Porcentaje máximo de búsqueda	Lista 1	Lista 2	Palabras clave
0.1	mecánica	alimentación	529
0.001	mecánica	alimentación	113
0.00001	mecánica	alimentación	85
0.0000000001	mecánica	alimentación	25
0.005	mecánica	Paz	499
0.0001	mecánica	Paz	128
0.000001	mecánica	Paz	60
0.000000000000001	mecánica	Paz	5

Ilustración 5: Valores con que se efectuaron las comparaciones

La variación en los criterios de búsqueda refleja una oscilación de resultados, que va desde 5 y hasta 529 palabras clave. Las palabras clave que en cada comparación resultan no varían, sólo disminuyen o aumentan según los valores de búsqueda, como se observa en las ilustraciones 6 y 7.

De las 529 palabras clave, algunas se repetían, en mayor o menor grado dependiendo de los valores de búsqueda en cada comparación, de forma que la suma de palabras clave de las ocho comparaciones arrojó un total de 1444 palabras. Con este cúmulo de palabras se realizó una lista de palabras para observar el comportamiento y cantidad total de palabras clave.

Los resultados de esta lista de 1444 palabras mostraban que el número total de palabras diferentes era de sólo 606; esto quiere decir que algunas palabras estaban presentes en todas las comparaciones y tenían una frecuencia de ocho —debido a las ocho

comparaciones— y había otras que sólo aparecían con frecuencia de uno. Este contraste en el número de frecuencias se debía a los valores de búsqueda de las palabras clave, en donde, en algunas comparaciones se tenían más de 500 palabras y en otras apenas 5.

Valor de búsqueda	0.1	0.001	0.00001	0.000000001
Palabra clave	MOVIMIENTO	MOVIMIENTO	MOVIMIENTO	MOVIMIENTO
	FUERZA	FUERZA	FUERZA	FUERZA
	VELOCIDAD	VELOCIDAD	VELOCIDAD	VELOCIDAD
	FIGURA	FIGURA	FIGURA	FIGURA
	ECUACIÓN	ECUACIÓN	ECUACIÓN	ECUACIÓN
	MASA	MASA	MASA	MASA
	NEWTON	NEWTON	NEWTON	NEWTON
	DISTANCIA	X	X	X
	X	DISTANCIA	DISTANCIA	DISTANCIA
	T	T	T	T
	FUERZAS	FUERZAS	FUERZAS	FUERZAS
	V	FRICCIÓN	V	FRICCIÓN
	FRICCIÓN	V	FRICCIÓN	V
	CENTRO	CENTRO	CENTRO	CENTRO
	LEY	LEY	LEY	LEY
	CONSERVACIÓN	CONSERVACIÓN	CONSERVACIÓN	CONSERVACIÓN
	TIEMPO	TIEMPO	TIEMPO	TIEMPO
	PLANO	PLANO	PLANO	PLANO
	ANGULAR	ANGULAR	ANGULAR	ANGULAR
	DIRECCIÓN	DIRECCIÓN	DIRECCIÓN	DIRECCIÓN

Ilustración 6: Primeras 20 palabras clave obtenidas de las cuatro comparaciones entre *mecánica* y *alimentación*

La tarea que a continuación se presentaba, era que de esas 606 palabras se debían descartar todas aquellas que no se relacionasen con mecánica o con física. Para realizar la tarea se procedió a hacer una depuración de las 606 palabras. La depuración consistió, en primer lugar, en eliminar todos los verbos, flexionados o en su forma canónica, que aparecían como palabras clave, además de todas las palabras que sólo cumplían una función discursiva⁷.

⁷ Para constatar que tales palabras cumplían sólo una función discursiva, fueron sometidas a un proceso de análisis de concordancia para observar el comportamiento que presentaban en relación con los contextos en que aparecían dentro de *mecánica*.

Valor de búsqueda	0.005	0.0001	0.000001	0.0000000000000001
Palabra clave	VELOCIDAD	VELOCIDAD	VELOCIDAD	VELOCIDAD
	MOVIMIENTO	MOVIMIENTO	MOVIMIENTO	MOVIMIENTO
	ENERGÍA	ENERGÍA	ENERGÍA	ENERGÍA
	FUERZA	FUERZA	FUERZA	FUERZA
	MASA	MASA	MASA	MASA
	FIGURA	FIGURA	FIGURA	
	ECUACIÓN	ECUACIÓN	ECUACIÓN	
	NEWTON	NEWTON	NEWTON	
	T	T	T	
	X	X	DISTANCIA	
	DISTANCIA	DISTANCIA	X	
	FRICCIÓN	FRICCIÓN	FRICCIÓN	
	ANGULAR	ANGULAR	ANGULAR	
	CONSERVACIÓN	CONSERVACIÓN	CONSERVACIÓN	
	PLANO	PLANO	PLANO	
	SISTEMA	SISTEMA	SISTEMA	
	V	PODEMOS	PODEMOS	
	PODEMOS	V	V	
	ACELERACIÓN	ACELERACIÓN	ACELERACIÓN	
	LEY	LEY	LEY	
	TENEMOS	TENEMOS	TENEMOS	
	TRABAJO	TRABAJO	TRABAJO	
	R	R	R	
	CENTRO	CENTRO	CENTRO	
	PRINCIPIO	PRINCIPIO	PRINCIPIO	
	CUERPO	CUERPO	CUERPO	
	ECUACIONES	ECUACIONES	CINÉTICA	
	CUERPOS	CUERPOS	ECUACIONES	
	CINÉTICA	CINÉTICA	CUERPOS	
	FUERZAS	FUERZAS	FUERZAS	

Ilustración 7: Primeras 30 palabras clave obtenidas de las cuatro comparaciones entre *mecánica* y *paz*

El resultado de esta depuración dio como resultado una lista compuesta por 218 palabras clave, todas ellas simples (ilustración 8); sin contar con las siglas las cuales fueron eliminadas después de aplicarles un análisis de concordancia que revelaba que éstas no eran usadas de manera general; cada una aparecía en archivos diferentes del hipertexto y no se repetía en algún otro; por tanto, no eran parte de una norma que fijase que tales siglas deberían ser tenidas como términos.

El procedimiento siguiente consistía en constatar que las 218 palabras clave obtenidas eran términos dentro del área de mecánica; para comprobar lo anterior, se realizaron varios análisis a través del módulo de concordancia, el cual abarca análisis de agrupamientos y colocaciones; para de esta forma, verificar que tales palabras clave podrían ser términos, además de que con tales análisis se permitiría avanzar con la identificación de la terminología poliléxica.

ACCIÓN	COTES	GRAVITACIONAL	OBJETO	RELACIÓN
ACELERACIÓN	CUALIDAD	HIELO	OBSERVACIÓN	RELATIVIDAD
ACELERADO	CUERPO	HIPOTENUSA	ONDA	REPOSO
ACERO	CURVA	HIPÓTESIS	OPUESTO	RESONANCIA
AFELIO	CHOQUE	HOOKE	ÓRBITA	ROTACIÓN
AGUA	DESPLAZAMIENTO	HORIZONTAL	OSCILACIÓN	ROTACIONAL
AIRE	DESTELLO	ÍMPETU	OSCILADOR	SATÉLITE
ALTURA	DÍA	INCIDENTE	PAR	SEGUNDA
AMORTIGUADO	DIÁMETRO	INCLINACIONES	PARABÓLICO	SENOIDAL
AMPLITUD	DIFRACCIÓN	INCLINADO	PARALELA	SENTIDO
ANGULAR	DILATACIÓN	INERCIA	PARTÍCULA	SIMPLE
ÁNGULO	DIMENSIÓN	INFLUENCIA	PENDIENTE	SIMULTA- NEIDAD
APLICADA	DIRECCIÓN	INICIAL	PÉNDULO	SISTEMA
ARCO	DISIPATIVAS	INTERACCIÓN	PERIHELIO	SOL
ÁREA	DISTANCIA	INTERFERENCIA	PERÍODO	SOLAR
ARISTÓTELES	ECUACIÓN	INTERVALO	PERPENDICULAR	SUPERFICIE
ARMÓNICO	EJE	KEPLER	PESAS	TANGENCIAL
ÁSPERA	ELÉCTRICA	LADO	PESO	TEORÍA
ATRACCIÓN	ELIPSE	LÁMPARA	PLANETA	TÉRMICA
BALANZA	ELÍPTICAS	LEY	PLANETARIO	TÉRMINO
BALÍN	ELONGACIÓN	LIBRE	PLANO	TIEMPO
CAÍDA	ENERGÍA	LÍMITE	POSICIÓN	TIERRA
CALOR	EQUILIBRIO	LÍNEA	POSTULADO	TIPO
CAMBIO	ESCALA	LINEAL	POTENCIA	TIRO
CANTIDAD	ESFERA	LONGITUD	POTENCIAL	TORSIÓN
CARTESIANOS	ESPACIO	LORENTZ	PRECISIÓN	TRABAJO
CAUSALIDAD	ESPECIAL	LUMINOSO	PREFACIO	TRANSFORMA- CIÓN
CAUSA	ESTADO	LUNA	PRINCIPIO	TRASLACIÓN
CAVENDISH	ESTÁTICA	LUZ	PROPIEDAD	TRAYECTORIA
CELESTE	ÉTER	MAGNITUD	PROPORCIONAL	TRIÁNGULO
CENTRÍPETA	EVENTOS	MARCO	PROYECCIÓN	TRILOGÍA
CENTRO	EXPERIMENTO	MAREA	PROYECTIL	UNIDAD
CINEMÁTICA	FASE	MASA	PTOLOMEO	UNIFORME
CINÉTICA	FENÓMENOS	MATERIA	PULSOS	UNIFORME- MENTE
CIRCULAR	FÍSICA	MÁXIMA	PUNTO	UNIVERSAL
CÍRCULO	FOCO	MAXWELL	RADIO	VACÍO
CIRCUNFERENCIA	FORMA	MECÁNICA	RAPIDEZ	VALIDEZ
COEFICIENTE	FRECUENCIA	MECANISMOS	RAYO	VALOR
CONSERVACIÓN	FRICCIÓN	MEDIO	RAZÓN	VARILLA
CONSERVATIVA	FUERZA	METAL	REACCIÓN	VECTOR
CONSTANTE	FUNCIÓN	MOMENTO	REALIZADO	VELOCIDAD
CONTRACCIÓN	GALILEO	MÓVIL	RECTA	VIBRACIÓN
COORDENADAS	GRAVEDAD	MOVIMIENTO	RECTÁNGULO	
COPÉRNICO	GRAVITACIÓN	NEWTON	RECTILÍNEO	

Ilustración 8: Lista final de palabras clave

4.5 Concordancia

Cada una de las 218 palabras clave finales fue sometida a un análisis de concordancia con el fin de verificar que tuviesen una relación con el área pretendida. Los análisis fueron hechos desde *mecánica* para advertir cómo era el comportamiento de las 218 palabras en cada uno de los archivos del hipertexto.

El estudio de estas palabras a través del módulo de concordancia permitió saber que las palabras clave extraídas de *mecánica* formaban parte de un área en particular, sin embargo, aún y cuando éstas fueron identificadas como pertenecientes a *mecánica*, todavía distaba de poder manifestarlas como términos.

<p> cia en y es un movimiento uniforme en el eje x y un Sin embargo el movimiento a ecuación del movimiento en el caso del movimiento ia al caso del movimiento del tipo de un movimiento niforme y el que llamamos a ecuación del movimiento e el sistema, el centro de ección y es un movimiento </p>	<p> uniformemente ACELERADO uniformemente ACELERADO uniformemente ACELERADO uniformemente ACELERADO uniformemente ACELERADO uniformemente ACELERADO Uniformemente ACELERADO. uniformemente ACELERADO. ACELERADO. En nuestro caso c es el orige ACELERADO. Por lo tanto las ecuaciones d </p>	<p> cuya ecuación es: en el eje y, generamos lo que implica la presencia de fuerza obtenemos resolviendo para el por lo que tenemos una curva d el área bajo la curva será la ya que es el que estudiamos. ¿Podríamos describir un movim ACELERADO. </p>	<p> de la que fuerza para el d la la que estudiamos. movim ACELERADO. </p>
---	---	--	---

Ilustración 9: Análisis de concordancia de *acelerado*

El módulo de concordancia, además de comprobar la pertenencia de las palabras clave, principalmente las de mayor frecuencia como *movimiento*, *fuerza*, *masa* o *momento* entre otras, a un área como *mecánica*, o por lo menos relacionada con la física, reveló. por el contexto, que algunas palabras eran componentes de un término poliléxico. Por ejemplo, en la ilustración 9 se ve el comportamiento que sufre *acelerado* en relación con *movimiento*, mientras que en la ilustración 10 se aprecia el comportamiento de *velocidad* y *angular*.

Estos datos hicieron necesario que, a cada palabra que mostraba un comportamiento como el antes mencionado, se le realizara una exploración por medio de agrupamientos para estudiar cómo era la relación que guardaba con respecto a determinadas palabras antes y después de ellas.

cial que tenía a la altura h se ve como VELOCIDAD a la altura cero. Por lo tanto Sol está también en movimiento, nuestra VELOCIDAD a través de éter será todavía Podríamos demostrar que el valor de la VELOCIDAD al caer la altura h por la caída hora tenemos que un cuerpo gira con una VELOCIDAD angular ω cada una de sus partículas angular (ω en el tiempo $t = 0$ y la VELOCIDAD angular ω , cuando a transcurrido masa m en otra de radio r con la misma VELOCIDAD angular ω . Como la fuerza centzozos extendidos y se le pone a girar con VELOCIDAD angular (ω nos podemos dar cuenta pecto del tiempo. Si un cuerpo tiene VELOCIDAD angular (ω en el tiempo $t = 0$ ior nos da el tiempo su inverso será la VELOCIDAD angular en ese momento y si: r ejemplo entre tierra y sol: La VELOCIDAD angular en términos del periodo locidad angular: por lo tanto la VELOCIDAD angular será: aplicando : y por lo tanto: (ω se denomina VELOCIDAD angular y es un vector cuya dirección producto del momento de inercia por la VELOCIDAD angular, así: $L = I\omega$ El ese momento y si: sustituyendo la VELOCIDAD angular: por lo tanto L se tiene que: como se ve la VELOCIDAD aumenta, conforme aumenta la altura h y t_B será: Si conocemos la VELOCIDAD común V de los barcos medimos ovimiento, este continuará moviéndose a VELOCIDAD constante en línea recta. Un objeto se movería sobre la superficie a VELOCIDAD constante sin tener que aplicar os marcos de referencia que se mueven a VELOCIDAD constante unos de otros. 2.

Ilustración 10: Fragmento del análisis de concordancia de velocidad

4.6 Análisis de palabras clave a través de agrupamientos y colocaciones

El análisis por medio de agrupamientos implica realizar un estudio de la ocurrencia de dos o más palabras juntas; por ejemplo, en la lista *mecánica* aparece *centro* con una frecuencia de 34 y *masa* con 57 apariciones; sin embargo, y a pesar de las diferencias de frecuencia, ambas resultaron ser, luego de las comparaciones efectuadas, palabras clave; mas cuando fueron sometidas al análisis de concordancia mostraron resultados que hacían pensar que, no obstante las frecuencias de cada una, podrían ser más que términos simples.

Con la idea de despejar esta duda que las concordancias habían arrojado, se procedió a ejecutar en cada una de las 218 palabras clave un análisis por medio de agrupamientos para examinar la conducta de cada una de ellas en relación con el contexto en que se encontraban dentro de los archivos del hipertexto⁸.

⁸ Estos análisis fueron hechos sobre la lista *mecánica*, la cual contenía los quince archivos del hipertexto.

CENTRO de <i>masa</i>	21	<i>centro</i> de MASA	19
el CENTRO de	10	el <i>centro</i> de	8
del CENTRO de	6	a la MASA	5
en el CENTRO	5	la MASA del	5
el CENTRO del	3	de MASA m	4
<i>equilibrio</i> CENTRO de	3	de MASA m#	4
su CENTRO de	3	de una MASA	4
<i>causalidad</i> CENTRO de	2	la <i>velocidad</i> del	3
CENTRO de <i>equilibrio</i>	2	proporcional a la	3

Ilustración 11: Análisis de agrupamientos de *centro* y *masa*

Este análisis de concordancia, en una forma muy restringida como lo es el agrupamientos, ofreció mucha riqueza de análisis; por ejemplo, en la ilustración 11 se puede ver que el estudio de una palabra clave en su contexto, en este caso *centro* y *masa*, (en mayúsculas), no aparecen solas, sino que muy cerca de ellas se suceden y anteceden otras palabras clave (en cursivas), y que lejos de pensarlas como posibles términos simples como *resonancia* o *marea* —que en el análisis de agrupamientos no mostraban sino palabras funcionales alrededor de ellas—, podrían ser términos poliléxicos.

Ahora bien, de los posibles 218 términos simples con que se contaba después de las comparaciones, resultaba interesante que los resultados de agrupamientos mostraran que un porcentaje alto de la supuesta terminología simple podría ser sólo componente de la terminología poliléxica.

Para corroborar esta hipótesis se realizó un análisis con la herramienta de colocaciones. Esta herramienta permite observar qué palabras preceden y anteceden a una palabra *X*; un análisis muestra, dependiendo de los criterios de búsqueda, qué palabras y con qué frecuencia aparecen, tanto a la izquierda como a la derecha de una palabra analizada.

PALABRA	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
MASA	60	1	2	0	0	1	0	0	57	1	0	1	0	0
LA	42	28	14	5	5	1	0	17	0	1	1	4	5	3
CENTRO	21	20	1	1	0	0	19	0	0	0	1	0	0	0
EL	21	16	5	4	3	9	0	0	0	0	1	1	0	3
M	19	1	18	0	0	1	0	0	0	9	2	2	2	3
DEL	17	6	11	0	1	5	0	0	0	7	2	0	0	2
UNA	14	10	4	4	2	0	0	4	0	0	0	1	1	2
QUE	13	2	11	0	1	0	1	0	0	1	4	1	2	3
EN	12	3	9	1	1	0	1	0	0	0	5	0	3	1
SE	11	4	7	2	1	1	0	0	0	0	0	3	1	3
Y	11	3	8	1	1	0	1	0	0	3	3	0	1	1
A	10	6	4	1	0	0	5	0	0	0	2	0	1	1
ES	10	5	5	2	2	0	1	0	0	1	0	2	2	0
POR	8	3	5	0	1	0	2	0	0	1	1	1	1	1
VELOCIDAD	6	3	3	1	2	0	0	0	0	0	0	0	1	2
COMO	5	2	3	0	1	1	0	0	0	0	1	1	0	1
SI	5	1	4	0	1	0	0	0	0	0	3	0	1	0

 Ilustración 12: Colocaciones de *masa*

El número en los valores de búsqueda de las colocaciones así como de agrupamientos, se determina según las necesidades y los resultados que se vayan presentando.

En las ilustraciones 12 y 13 se observa cuales son, en un rango de cinco palabras a la derecha y cinco a la izquierda, las palabras que en los textos rodean a las palabras analizadas.

Esta representación reafirmaba la hipótesis de que la gran mayoría de palabras clave candidatas a términos eran componentes de términos poliléxicos; estos resultados hacían más imperiosa la necesidad de realizar análisis que demostraran lo que se pensaba; para lograrlo, se realizaron listas de palabras de dos y hasta ocho palabras, además de analizar todo el hipertexto con la herramienta de información mutua.

Los resultados obtenidos se muestran a continuación.

PALABRA	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
DE	43	10	33	2	2	1	5	0	0	25	0	2	4	2
CENTRO	38	2	2	0	0	2	0	0	34	0	0	2	0	0
MASA	21	1	20	0	0	0	0	1	0	0	19	0	0	1
EL	19	16	3	0	1	1	0	14	0	0	0	1	0	2
DEL	16	10	6	1	1	0	1	7	0	3	1	2	0	0
LA	16	11	5	2	4	4	1	0	0	0	2	1	0	2
EN	9	6	3	0	1	0	5	0	0	1	0	1	1	0
POR	7	6	1	2	2	0	2	0	0	0	0	1	0	0
ES	6	3	3	0	2	1	0	0	0	0	0	2	1	0
Y	6	1	5	0	1	0	0	0	0	0	0	4	0	1
POSICIÓN	5	2	3	0	0	1	1	0	0	0	2	0	0	1
SE	5	3	2	1	1	1	0	0	0	1	0	1	0	0
VELOCIDAD	5	4	1	1	0	1	2	0	0	0	0	0	0	1

Ilustración 13: Colocaciones de centro

4.6.1 Terminología poliléxica a partir de listas de palabras múltiples e información mutua

Tanto los resultados de los análisis por medio de agrupamientos como a través de colocaciones encaminaban las investigaciones hacia un punto específico: comprobar que las palabras clave obtenidas eran parte de la terminología que se proponía, y que las más de ellas eran constituyentes de términos poliléxicos.

Mas no bastaba el que esta hipótesis surgiera y, en cierto punto, se comprobara por medio de los análisis de concordancia, entendiéndose este análisis como módulo integral que involucra los estudios de concordancia, agrupamientos y colocaciones; el punto ahora consistía en corroborar este supuesto a través de las herramientas que a disposición se tenían para aumentar a un porcentaje mayor y más sustentado los resultados que se venían dando.

El procedimiento que en primer lugar se llevó a cabo fue la elaboración de ocho listas más a partir del hipertexto, en las cuales se contemplaba extraer listas de palabras que se compusieran de dos y hasta ocho palabras.

Todos los términos que aparecen en las ilustraciones 14 a 18 fueron seleccionados de las listas de 2 hasta 8 palabras, a partir de los siguientes criterios:

1. mayor frecuencia. (PLANO INCLINADO: frecuencia = 20);
2. recuperación de serie de palabras coherentes, aunque su frecuencia no fuese tan alta. (CONSERVACIÓN DEL MOMENTO ANGULAR: frecuencia = 3);
3. eliminación de series de palabras que no tenían significado en sí mismas. (DE FRICCIÓN ESTÁTICA. El término como tal es: FUERZA DE FRICCIÓN ESTÁTICA).

Palabra	Frecuencia		
ACELERACIÓN ANGULAR	3	FUERZAS CONSERVATIVAS	3
ATRACCIÓN		FUERZAS DISIPATIVAS	3
GRAVITACIONAL	6	ÍMPETU LINEAL	17
CAÍDA LIBRE	11	MOMENTO ANGULAR	16
ENERGÍA CINÉTICA	15	MOVIMIENTO ARMÓNICO	2
ENERGÍA INICIAL	3	MOVIMIENTO RECTILÍNEO	7
ENERGÍA POTENCIAL	13	ÓRBITA CIRCULAR	2
ENERGÍA TÉRMICA	3	ÓRBITAS ELÍPTICAS	2
FRICCIÓN CINÉTICA	4	ÓRBITAS PLANETARIAS	2
FRICCIÓN ESTÁTICA	6	PLANO INCLINADO	20
FUERZA CENTRÍPETA	2	PULSOS LUMINOSOS	6
FUERZA CONSERVATIVA	2	RELATIVIDAD ESPECIAL	2
FUERZA GRAVITACIONAL	3	VELOCIDAD ANGULAR	11
FUERZA PROPORCIONAL	3	VELOCIDAD CONSTANTE	7

Ilustración 14: Candidatos a términos obtenidos de la lista de dos palabras

Los resultados que estas ocho listas arrojaron muestran toda una serie de candidatos a términos poliléxicos muy interesantes; en las Ilustraciones 14 a 18 se observa cómo la mayoría de las palabras clave se mezclan entre ellas para conformar posibles términos poliléxicos que ya en los análisis de concordancia se insinuaban.

Palabra	Frecuencia
ACCIÓN A DISTANCIA	2
ARCO DE CIRCUNFERENCIA	3
BALANZA DE CAVENDISH	3
CAMBIO DE ENERGÍA	3
CENTRO DE MASA	19
COEFICIENTE DE FRICCIÓN	5
CONSERVACIÓN DEL ÍMPETU	3
CONSERVACIÓN DEL MOMENTO	3
DIRECCIÓN DE MOVIMIENTO	4
ECUACIÓN DE CONSERVACIÓN	7

ECUACIÓN DE MOVIMIENTO	4
ESTADO DE MOVIMIENTO	5
FUERZA DE ATRACCIÓN	2
FUERZA DE FRICCIÓN	9
LEY DE CONSERVACIÓN	3
LEY DE KEPLER	3
LEYES DE NEWTON	16
LEYES DEL MOVIMIENTO	3
MARCO DE REFERENCIA	7
MECÁNICA DE NEWTON	3
MOMENTO DE INERCIA	5
MOMENTO DE TORSIÓN	4
MOVIMIENTO CIRCULAR UNIFORME	5
MOVIMIENTO DE TRASLACIÓN	4
MOVIMIENTO RECTILÍNEO UNIFORME	3
MOVIMIENTO UNIFORMEMENTE ACELERADO	8
PRINCIPIO DE CONSERVACIÓN	8
PRINCIPIO DE GALILEO	7
TIPO DE MOVIMIENTO	3

Ilustración 15: Candidatos a términos obtenidos de la lista de tres palabras

Las listas de dos hasta ocho palabras ayudaron a esclarecer la duda que el comportamiento de las palabras clave en los análisis de concordancia suponía; y no obstante que luego de los análisis que se habían efectuado para obtener la terminología simple y poliléxica, los resultados en cierta medida resultaban satisfactorios, se optó por realizar una última prueba que constatará los términos simples y sustentara de forma más sólida los términos poliléxicos. Esta última prueba fue efectuada con la herramienta de información mutua, explicada en capítulos anteriores, con la cual se esperaba, si así lo demostraban los resultados, cerrar los análisis de esta investigación.

Palabra	Frecuencia
CONSERVACIÓN DE LA ENERGÍA	16
CONSERVACIÓN DEL ÍMPETU LINEAL	3
CONSERVACIÓN DEL MOMENTO ANGULAR	3
ECUACIONES DE LA ENERGÍA	2
FUERZA DE FRICCIÓN ESTÁTICA	2
MOVIMIENTO RECTILÍNEO Y UNIFORME	2
TEORÍA DE LAS CUALIDADES	2

Ilustración 16: Candidatos a términos obtenidos de la lista de cuatro palabras

Palabra	Frecuencia
CASO DE LA CAÍDA LIBRE	2
CASO DEL MOVIMIENTO UNIFORMEMENTE ACELERADO	2
CUALIDADES OCULTAS EN LA MATERIA	2
ECUACIÓN DEL MOVIMIENTO UNIFORMEMENTE ACELERADO	2
LEY DE LA GRAVITACIÓN UNIVERSAL	2
TEORÍA DE LAS CUALIDADES OCULTAS	2
TEORÍA ESPECIAL DE LA RELATIVIDAD	2

Ilustración 17: Candidatos a términos obtenidos de la lista de cinco palabras

Para realizar este análisis final se volvió a hacer uso de *mecánica*, ya que en todos los procesos practicados, como palabras clave, concordancia y demás, fue necesario contar con los textos de los cuales se desprendían todas las palabras con las que se venía trabajando; por consiguiente, en información mutua era necesario contar con toda la información que se encontraba en el hipertexto, para que los cálculos de probabilidad con que trabaja esta herramienta estuviesen respaldados por los contextos en los que aparece cada palabra clave.

Palabra	Frecuencia
ECUACIÓN DE CONSERVACIÓN DE LA ENERGÍA	3
PRINCIPIO DE CONSERVACIÓN DE LA ENERGÍA	3

Ilustración 18: Candidatos a términos obtenidos de la lista de seis palabras

El proceso final constó de la elaboración de un índice alfabético en el cual se realizó el análisis de información mutua⁹. Para lograr este índice se partió de *mecánica*, para la que se guardó la lista alfabética y la de frecuencia. Con estas listas hechas, WordSmith elaboró un índice en el que aparecía alfabéticamente cada palabra del hipertexto con sus respectivas frecuencias. Una vez que se tuvo listo el índice alfabético se procedió a realizar el análisis de información mutua; éste es un proceso tardado debido a que se efectúa la aplicación del logaritmo base 2 en cada palabra del índice, y esta palabra es confrontada con todas las palabras de todos los textos para calcular la probabilidad de

⁹ Información mutua trabaja a partir de la siguiente fórmula:
 $(\text{frecuencia de unión} * \text{número de unidades del corpus}) / (\text{frecuencia de cruce} * \text{frecuencia de colocación}) = \text{IM}$. Cfr. WordSmith versión 2.0 Mike Scott & Oxford University Press.

que dos o más palabras puedan aparecer juntas en un contexto determinado (ilustración 19, en donde los valores entre paréntesis son los porcentajes de información mutua).

PALABRA	RELACIÓN 1	RELACIÓN 2	RELACIÓN 3	RELACIÓN 4	RELACIÓN 5	RELACIÓN 6	RELACIÓN 7
ACELERACION	centrípeta (5,07)	diferente (4,48)	q (4,26)	ahora (4,26)	acuerdo (4,07)	cero (3,65)	cambio (3,61)
ACELERADO	uniformemente (7,25)	rectilíneo (5,66)	curva (5,66)	eje (4,93)	uniforme (4,83)	tipo (4,79)	movimiento (3,83)
AFELIO							
AIRE							
ANGULAR	pesas (4,88)	momento (4,77)	i (4,04)	satélite (3,81)	ahora (3,81)	gira (3,75)	
ARMÓNICO	simple (7,32)	tipo (5,66)	ecuaciones (5,39)	movimiento (3,38)			
ATRACCIÓN	gravitacional (6,47)	luna (4,66)	tierra (4,48)	cuerpos (4,39)	partículas (3,88)		
CENTRIPETA	sustituimos (6,12)	aceleración (5,8)	fuerza (3,17)				
CENTRO	círculo (4,3)	Masa (4,11)	planetas (4,04)	circunferencia (3,88)	posición (3,75)	línea (3,62)	
CIRCULAR	uniforme (6,19)	supongamos (6,07)	órbita (5,39)	llama (4,93)	móvil (4,8)	mueve (4,32)	movimiento (3,97)
CIRCULO	perpendicular (6,8)	Llama (5,93)	móvil (5,8)	dirección (4,69)	velocidad (3,05)		
CIRCUNFERENCIA	arco (6,58)	Iguales (5,58)	radio (5,22)	mueve (4,3)	centro (3,88)		
COEFICIENTE	fricción (5,22)						
CONSERVACION	física (4,71)	Principio (4,6)	problema (4,03)	siguiente (3,59)	energía (3,59)	dice (3,49)	ley (3,25)
CUALIDADES	materia (7,65)	Teoría (5,58)					
CUERPO	pesas (4,39)	Realizado (3,9)	peso (3,8)	estado (3,8)	define (3,8)	aplicar (3,54)	mueve (3,36)

Ilustración 19: Fragmento de los resultados del análisis de las palabras clave efectuado a través de información mutua

Los valores de probabilidad de que dos palabras lleguen a aparecer juntas se expresa mediante porcentajes con el título mutua; entre más alto sea este porcentaje, mayor es la posibilidad de que dos o más palabras sean términos poliléxicos e, incluso, locuciones.

Una vez que se revisaron los resultados de información mutua y luego de comprobar que el material obtenido en vez de distar de lo que los análisis previos habían arrojado, los confirmaba, se dieron por concluidos los análisis y procesos de extracción de términos.

La cuestión que ahora privaba consistía en ordenar la terminología obtenida para presentar los resultados a los expertos del Centro de Instrumentos y esperar la evaluación que ellos diesen.

Pero antes de presentar las evaluaciones de los expertos, se hace necesario presentar los resultados obtenidos de todos los análisis de esta investigación. Trabajo que se detalla en el capítulo siguiente.

5. RESULTADOS

5.1 Presentación de los resultados

Lancaster señala que “está justificada la inclusión de un término en un índice sólo si es de interés para el usuario”¹. Tomando en cuenta el razonamiento de Lancaster, es necesario recalcar que la terminología obtenida ha estado desde el principio de este trabajo estrictamente delimitada, debido a que los términos a obtener se ciñen a un nivel muy preciso, el bachillerato. Por tanto, cualquier carencia de términos que se encuentren en un nivel superior a éste está justificada bajo el razonamiento de que si esta terminología es un material de apoyo para el alumno de bachillerato, es de poco interés y beneficio considerar una terminología de un nivel superior, la cual se encontraría fuera de las necesidades de tales alumnos.

Por otro lado, es conveniente precisar que la terminología que se obtuvo pretende ser lo más completa en el nivel especificado; sin embargo, no es exhaustiva debido a que el campo de extracción de los términos se limitó al material proporcionado por la gente especializada del Centro de Instrumentos y que consta, como ya se ha anotado antes, de sólo quince archivos. También es necesario aclarar que no se efectuó un proceso de extracción en materiales más específicos como diccionarios o enciclopedias sobre el tema, puesto que ahí la extracción, y desde luego, la identificación, hubiesen sido procesos de antemano asegurados a un cien por ciento, aunque fuesen como en el caso presente, semiautomáticos; mas la cuestión es que se pretendía además de extraer los términos de mecánica, probar una herramienta como WordSmith que fuese aplicable no sólo a material absoluto e íntegro sobre un tema o área específica, sino aprovecharlo en otras vertientes que no necesariamente sean especializadas.

¹ Frederick W Lancaster. El control del vocabulario en la recuperación de información. València: Universitat de València, 1995. p. 43.

Ahora bien, los análisis que se llevaron a cabo para obtener la terminología pretendida son, como en capítulos anteriores se menciona, efectuados con el programa WordSmith. Con este programa se pretendía, a partir del uso y explotación de las distintas herramientas con que cuenta, extraer la terminología de mecánica presente en el hipertexto, por medio de los módulos de listas de palabras, palabras clave y concordancia.

Los procesos que se realizaron en estos módulos y que se detallan en el capítulo 4 dieron por resultado la terminología que a continuación se presenta.

ACCIÓN
 ACELERACIÓN
 AFELIO
 ALTURA
 AMORTIGUADO
 AMPLITUD
 ANALOGÍA
 ANGULAR
 ÁREA
 ARISTÓTELES
 BALÍN
 CALOR
 CANTIDAD
 CARTESIANO
 CAUSALIDAD
 CAUSALIDAD
 CAVENDISH
 CENTRÍPETA
 CHOQUE
 CINEMÁTICA
 CINÉTICA
 CIRCULAR
 CIRCUNFERENCIA
 CONSERVACIÓN
 CONSTANTE
 CONTRACCIÓN
 COPÉRNICO
 COTES
 CRONÓMETRO
 CUADRADO
 CUALIDAD
 CUERPO
 CURVA
 DESPLAZAMIENTO

DIFRACCIÓN
 DILATACIÓN
 DIMENSIÓN
 DINÁMICA
 DIRECCIÓN
 DISTANCIA
 ECUACIÓN
 EINSTEIN
 EJE
 ELONGACIÓN
 ENERGÍA
 EQUILIBRIO
 ESFERA
 ESPACIO
 ESTÁTICA
 ÉTER
 EXPERIMENTO
 FÍSICA
 FRECUENCIA
 FRICCIÓN
 FUERZA
 FUNCIÓN
 GALILEO
 GIRO
 GRAVEDAD
 GRAVITACIÓN
 HIPOTENUSA
 HOOKE
 ÍMPETU
 INCLINACIÓN
 INERCIA
 INTERFERENCIA
 KEPLER
 LEY

LÍNEA
 LONGITUD
 LORENTZ
 LUMINOSO
 LUNA
 LUZ
 MAGNITUD
 MAREA
 MASA
 MATERIA
 MECÁNICA
 MECANISMO
 MOMENTO
 MÓVIL
 MOVIMIENTO
 NEWTON
 NEWTON (UNIDAD)
 OBJETO
 OBSERVACIÓN
 OBSERVADOR
 ONDA
 ÓPTICA
 ORBITA
 OSCILACIÓN
 OSCILADOR
 PAR
 PARABÓLICO
 PARTÍCULA
 PÉNDULO
 PERIHELIO
 PERÍODO
 PERPENDICULAR
 PESA
 PESO

PLANETA	RECTA	TIEMPO
PLANO	RECTÁNGULO	TIERRA
POSTULADO	RECTILÍNEO	TORCA
POTENCIA	RELATIVIDAD	TORSIÓN
PRINCIPIO	REPOSO	TRABAJO
PROPIEDAD	RESONANCIA	TRANSFORMACIÓN
PROPORCIONAL	RESORTE	TRASLACIÓN
PROYECTIL	REVOLUCIÓN	TRAYECTORIA
PTOLOMEO	ROTACIÓN	TRIÁNGULO
PULSOS	SATÉLITE	UNIDAD
PUNTO	SEGUNDO	UNIFORME
RADIO	SIMULTANEIDAD	VECTOR
RAPIDEZ	SISTEMA	VELOCIDAD
RAYO	SOL	VIBRACIÓN
RAZÓN	SUPERFICIE	
REACCIÓN	TEORÍA	

Ilustración 1: Terminología simple

Las ilustraciones 1 y 2 muestran, respectivamente, la terminología simple y poliléxica obtenida en esta investigación. Consta de 307 términos, de los cuales 148 son simples, y 159 poliléxicos. Estos últimos están constituidos por dos o más términos simples, tal es el caso de los términos *aceleración* y *centrípeta*, cada uno término simple con un significado y un significante, pero la unión de ellos da un término más, *aceleración centrípeta*, que designa otro concepto distinto al que cada uno se refería por separado.

ACCIÓN A DISTANCIA	CONSERVACIÓN DEL ÍMPETU LINEAL
ACELERACIÓN ANGULAR	CONSERVACIÓN DEL MOMENTO
ACELERACIÓN CENTRÍPETA	CONSERVACIÓN DEL MOMENTO ANGULAR
AMPLITUD DE UNA OSCILACIÓN	CUALIDADES OCULTAS EN LA MATERIA
ARCO DE CIRCUNFERENCIA	CUERPO CELESTE
ATRACCIÓN GRAVITACIONAL	DÍA SOLAR
BALANZA DE CAVENDISH	DÍA SOLAR MEDIO
CAÍDA LIBRE	DILATACIÓN DEL ESPACIO
CAMBIO DE ENERGÍA	DILATACIÓN DEL TIEMPO
CENTRO DE EQUILIBRIO	DIRECCIÓN DE UN VECTOR
CENTRO DE MASA	DIRECCIÓN DEL MOVIMIENTO
CHOQUE ELÁSTICO	ECUACIÓN DE CONSERVACIÓN
CHOQUE INELÁSTICO	ECUACIÓN DE CONSERVACIÓN DE LA ENERGÍA
COEFICIENTE DE FRICCIÓN	ECUACIÓN DEL MOVIMIENTO
COEFICIENTE DE FRICCIÓN CINÉTICA	ECUACIÓN DEL MOVIMIENTO UNIFORMEMENTE ACELERADO
COEFICIENTE DE FRICCIÓN ESTÁTICA	ECUACIONES DE LA ENERGÍA
COMPONENTE DE UN VECTOR	ECUACIONES DE MOVIMIENTO
CONSERVACIÓN DE LA ENERGÍA	EJE DE GIRO
CONSERVACIÓN DEL ÍMPETU	EJE DE ROTACIÓN

ELÍPTICA
 ENERGÍA CINÉTICA
 ENERGÍA CINÉTICA DE ROTACIÓN
 ENERGÍA ELÉCTRICA
 ENERGÍA INICIAL
 ENERGÍA MECÁNICA
 ENERGÍA POTENCIAL
 ENERGÍA POTENCIAL GRAVITACIONAL
 ENERGÍA QUÍMICA
 ENERGÍA TÉRMICA
 ESTADO DE MOVIMIENTO
 ESTADO DE REPOSO
 EVENTOS SIMULTÁNEOS
 FRECUENCIA DE RESONANCIA
 FRECUENCIA NATURAL
 FRICCIÓN CINÉTICA
 FRICCIÓN ESTÁTICA
 FUERZA CENTRÍPETA
 FUERZA CONSERVATIVA
 FUERZA DE ATRACCIÓN
 FUERZA DE ATRACCIÓN GRAVITACIONAL
 FUERZA DE FRICCIÓN
 FUERZA DE FRICCIÓN CINÉTICA
 FUERZA DE FRICCIÓN ESTÁTICA
 FUERZA DE INERCIA
 FUERZA DE MAREA
 FUERZA DISIPATIVA
 FUERZA FICTICIA
 FUERZA GRAVITACIONAL
 FUERZA NORMAL
 FUERZA PROPORCIONAL
 FUERZA RESULTANTE
 FUERZAS CONSERVATIVAS
 FUERZAS DE FRICCIÓN
 FUERZAS DISIPATIVAS
 ÍMPETU LINEAL
 LEY DE CONSERVACIÓN
 LEY DE LA CONSERVACIÓN DE LA ENERGÍA
 LEY DE LA CONSERVACIÓN DEL ÍMPETU
 LEY DE LA CONSERVACIÓN DEL ÍMPETU LINEAL
 LEY DE LA CONSERVACIÓN DEL MOMENTO ANGULAR
 LEY DE LA GRAVITACIÓN UNIVERSAL
 LEY DE LA POTENCIA DE MOVIMIENTO
 LEY DE LAS ÁREAS
 LEY DE LOS PERÍODOS
 LEY DEL EFECTO
 LEY DE LA GRAVITACIÓN UNIVERSAL
 LEYES DE KEPLER
 LEYES DE NEWTON
 LEYES DE NEWTON
 LEYES DEL MOVIMIENTO
 LÍNEA DE ACCIÓN DE FUERZA
 MAGNITUD DE UN VECTOR
 MARCO DE REFERENCIA
 MARCO DE REFERENCIA INERCIAL
 MARCO DE REFERENCIA NO ACELERADO
 MASA PUNTUAL
 MECÁNICA DE NEWTON
 MEDIO ELÁSTICO
 MOMENTO ANGULAR
 MOMENTO DE INERCIA
 MOMENTO DE TORSIÓN
 MOVIMIENTO ARMÓNICO
 MOVIMIENTO ARMÓNICO SIMPLE
 MOVIMIENTO CIRCULAR
 MOVIMIENTO CIRCULAR UNIFORME
 MOVIMIENTO DE CAÍDA LIBRE
 MOVIMIENTO DE TRASLACIÓN
 MOVIMIENTO LINEAL
 MOVIMIENTO RECTILÍNEO
 MOVIMIENTO RECTILÍNEO Y UNIFORME
 MOVIMIENTO ROTACIONAL
 MOVIMIENTO UNIFORMEMENTE ACELERADO
 MOVIMIENTO UNIFORMEMENTE RETARDADO
 ONDA LUMINOSA
 ONDA SENOIDAL
 ORBITA CIRCULAR
 ORBITA ELÍPTICA
 ÓRBITAS PLANETARIAS
 OSCILACIÓN FORZADA
 OSCILACIÓN LIBRE
 OSCILADOR ARMÓNICO
 OSCILADOR ARMÓNICO SIMPLE
 PLANO INCLINADO
 PRIMERA LEY DE KEPLER
 PRIMERA LEY DE NEWTON
 PRINCIPIO DE CONSERVACIÓN
 PRINCIPIO DE CONSERVACIÓN DE LA ENERGÍA
 PRINCIPIO DE GALILEO
 PRINCIPIO DE INERCIA DE GALILEO
 PRINCIPIO DE LA CAUSALIDAD
 PRINCIPIO DE LA CONSERVACIÓN DE LA ENERGÍA
 PRINCIPIO DE TRABAJO Y ENERGÍA
 PULSOS LUMINOSOS
 PUNTO DE EQUILIBRIO
 RELATIVIDAD ESPECIAL

SEGUNDA LEY DE KEPLER
 SEGUNDA LEY DE NEWTON
 SENTIDO DE UN VECTOR
 SISTEMA AMORTIGUADO
 SISTEMA CONSERVATIVO
 SISTEMA DE PARTÍCULAS
 SISTEMA DE REFERENCIA INERCIAL
 SISTEMA DISIPATIVO
 SISTEMA FÍSICO
 SISTEMA MECÁNICO
 SISTEMA VIBRATORIO
 TEORÍA DE LA RELATIVIDAD
 TEORÍA DE LAS CUALIDADES
 TEORÍA DE LAS CUALIDADES OCULTAS

TEORÍA ESPECIAL DE LA RELATIVIDAD
 TERCERA LEY DE KEPLER
 TERCERA LEY DE NEWTON
 TIPO DE MOVIMIENTO
 TIRO PARABÓLICO
 TRANSFORMACIÓN DE LA MASA
 TRANSFORMACIÓN DE LORENTZ
 VELOCIDAD ANGULAR
 VELOCIDAD CONSTANTE
 VELOCIDAD DE LA LUZ
 VELOCIDAD INSTANTÁNEA

Ilustración 2. Terminología poliléxica

Esta terminología fue presentada en su totalidad a los expertos del Centro de Instrumentos de la UNAM para que evaluaran los resultados obtenidos y, a partir de su dictamen, comprobar si la investigación concluyó en resultados aceptables.

5.2 Validación de los resultados

Con el fin de obtener una evaluación sobre cualquier trabajo realizado es necesario recurrir a gente o material especializado para cotejar y sustentar los avances y resultados que se van obteniendo. En este trabajo, que tiene un carácter ciertamente terminológico, Lancaster² propone que, para comprobar los términos que componen una terminología (para que ésta pueda resultar objetiva y confiable), es conveniente y hasta necesario juzgar y evaluar los datos obtenidos con trabajos realizados sobre el tema, con diccionarios y enciclopedias, con material especializado y con expertos en el ramo donde la terminología está dada.

Ahora bien, para evaluar la terminología que se obtuvo mediante el desarrollo de este trabajo se recurrió a uno de los puntos que Lancaster señala: evaluar los datos a través de expertos en la materia sobre la que trata la terminología. Por lo tanto, sólo los 307

² Idem.

términos³ que componen, según los procesos realizados en este trabajo, la terminología de mecánica para el nivel bachillerato, fueron presentados a los expertos del Centro de Instrumentos para que, como gente especializada sobre el tema, diesen su evaluación. Para la evaluación de los expertos se siguió el método Delphi⁴, consistente en este caso de un solo ciclo, en donde cada uno de los expertos evaluó la terminología en privacidad, con el fin de no haber influencias entre los expertos.

Cabe señalar que estas personas no se basaron en criterios lingüísticos para emitir su evaluación, sino que ésta fue dada con base en sus propias apreciaciones, sustentadas por el conocimiento y experiencia en el área, y bajo la premisa que se ha manifestado a lo largo de este estudio: la terminología está pensada sólo para satisfacer las necesidades de aprendizaje y fungir como material de apoyo para los estudiantes del nivel *bachillerato*.

Dicho lo anterior, estas son algunas de las observaciones⁵ que los expertos⁶ señalan a la terminología obtenida en este trabajo:

- “si se incluye *área*, por qué no se incluyen términos como *círculo* o *triángulo*”
- “ser uniforme con el orden de la inscripción de nombres propios: o utilizan *Ptolomeo Claudio* o *Claudio Ptolomeo*”
- “deben ser consecuentes con el idioma en los nombres propios, esto es, si usan *Albert Einstein* por qué no siguen el mismo criterio en *Juan Kepler*”
- “*cotes* : de dónde sacaron este término”
- “el término *causalidad* aparece repetido”
- “*cronómetro* no resulta necesario”
- “considero que no son pertinentes ni *observación* ni *observador*”

³ Es preciso señalar que lo único que se entregó a los expertos del Centro de Instrumentos fue la lista que contenía los 307 términos resultantes. A excepción de esta información que era necesaria para obtener una evaluación conforme a los requerimientos del proyecto del GIL, nada del material que compone esta tesis fue puesto en manos de las personas del dicho centro.

⁴ Referencia al método Delphi. Cfr. Makridakis, Spyros y Wheelwright. Forecasting Methods for management. Wiley, 1989.

⁵ Las observaciones que los expertos hicieron luego de las evaluaciones no siguen un patrón de importancia para insertarlas en este trabajo.

⁶ En total fueron cuatro expertos quienes evaluaron los términos presentados.

- “*hipotenusa, frecuencia, cuadrado*, entonces también podría estar *matemáticas*, sean más precisos”
- “no se ve aporte alguno al incluir: *Segundo*”
- “ni al caso que incluyan *pulsos luminosos*”
- “han presentado algunos términos matemáticos que bien podrían ser eliminados”
- “podría ser el caso que resultaran demasiados términos para un nivel bachillerato”

Estas son las observaciones que los del Centro de Instrumentos devolvieron en las copias que se les entregaron sobre la terminología presentada. Hubo, de entre los cuatro expertos, uno que no hizo ninguna observación; sólo al final de la terminología, que fue dada en forma alfabética, ponía un signo de paloma para los términos que consideraba adecuados, y un tache para los términos que consideraba descartar.

En cuanto a la evaluación general, de los 307 términos propuestos, opinan que se deberían eliminar ciertos términos como *hipotenusa, observación, cuadrado* o *elongación*. O por el contrario, agregar términos que justifiquen la aparición de los anteriores y de algunos otros. Alguno sugirió que sería buena idea disminuir los 307 términos y sintetizar esta lista en una más simple en cuanto a la dificultad de términos. En suma, la evaluación que dieron resultó ser positiva y, en cierta medida, sobrepasó las expectativas que se tenían respecto a que se pensaba que serían desechados gran cantidad de términos por parte de los expertos, cosa que no sucedió.

En números, de los 307 términos, se descartaron, por reconvencción de los expertos, 12 términos⁷, esto es, que casi fue aprobado un 98% del total de términos propuestos para integrar el diccionario onomasiológico.

Con esta evaluación de los resultados se dio por concluido este trabajo.

⁷ Para este trabajo no se excluyeron los términos que proponían los expertos y que para el diccionario se descartaron. Aquí se presentan los 307 términos finales que arrojó la investigación.

- “*hipotenusa, frecuencia, cuadrado*, entonces también podría estar *matemáticas*, sean más precisos”
- “no se ve aporte alguno al incluir: *Segundo*”
- “ni al caso que incluyan *pulsos luminosos*”
- “han presentado algunos términos matemáticos que bien podrían ser eliminados”
- “podría ser el caso que resultaran demasiados términos para un nivel bachillerato”

Estas son las observaciones que los del Centro de Instrumentos devolvieron en las copias que se les entregaron sobre la terminología presentada. Hubo, de entre los cuatro expertos, uno que no hizo ninguna observación; sólo al final de la terminología, que fue dada en forma alfabética, ponía un signo de paloma para los términos que consideraba adecuados, y un tache para los términos que consideraba descartar.

En cuanto a la evaluación general, de los 307 términos propuestos, opinan que se deberían eliminar ciertos términos como *hipotenusa, observación, cuadrado o elongación*. O por el contrario, agregar términos que justifiquen la aparición de los anteriores y de algunos otros. Alguno sugirió que sería buena idea disminuir los 307 términos y sintetizar esta lista en una más simple en cuanto a la dificultad de términos. En suma, la evaluación que dieron resultó ser positiva y, en cierta medida, sobrepasó las expectativas que se tenían respecto a que se pensaba que serían desechados gran cantidad de términos por parte de los expertos, cosa que no sucedió.

En números, de los 307 términos, se descartaron, por reconvención de los expertos, 12 términos⁷, esto es, que casi fue aprobado un 98% del total de términos propuestos para integrar el diccionario onomasiológico.

Con esta evaluación de los resultados se dio por concluido este trabajo.

⁷ Para este trabajo no se excluyeron los términos que proponían los expertos y que para el diccionario se descartaron. Aquí se presentan los 307 términos finales que arrojó la investigación.

6. CONSIDERACIONES FINALES

6.1 Conclusiones

A pesar de los contratiempos y vicisitudes que supone la elaboración y conclusión de un proyecto de tesis, la etapa correspondiente a este trabajo se da por concluida, no sin antes puntualizar lo que se llevó a cabo a lo largo del desarrollo de cada capítulo, y de dar, a modo de conclusión, una serie de puntos que en la investigación resultaron de gran interés.

Las premisas contempladas para ser resueltas mediante el capítulo 1 quedaron despejadas. Se establecieron los fundamentos básicos de la terminología, resumiendo que ésta es “una disciplina que permite identificar el vocabulario de una especialidad en forma sistemática, analizarlo y si es necesario crearlo entre el especialista y el terminólogo, además de normalizarlo en una situación concreta de funcionamiento con la finalidad de responder a las necesidades de expresión de sus usuarios.”¹

Quedó claro qué es un término, tanto para los estudiosos como para esta investigación (*término es una palabra que dentro de un dominio específico, sea de cualquier índole [técnico, científico, etc.], tiene la característica de que sólo puede referirse a un concepto, no acepta sinonimia; y esta referencia debe ser inequívoca*²), y las diferencias y características que lo distinguen de una palabra.

Además, se despejaron las dudas de por qué este trabajo está enmarcado en un interés lingüístico, ya que como en el capítulo 1 se señaló, la terminología, como disciplina encargada del estudio de las palabras de especialidad, comparte características que la relacionan con la lexicología, lexicografía y, en un nivel más alto, con la lingüística, que a fin de cuentas, se encarga del estudio del lenguaje. Esto quiere decir, según Cabré, que en tanto los lenguajes de especialidad sean entendidos como

¹ Cardero, Ana María. El procesamiento de una terminología. Referencia especial a la terminología de control de satélites en el área de las telecomunicaciones en México. Tesis de doctorado. México: 2001. p. 5.

² Ver capítulo 1.

subsistemas de la lengua general, la terminología puede ser considerada como parte de la lexicología, y por tanto, estudiados desde un plano lingüístico³.

Por último, en este capítulo quedó establecida la forma en que sería tratada la terminología, esto es, estudiar esta disciplina para aplicarla como herramienta de apoyo a la ingeniería lingüística.

Con el desarrollo del capítulo 2, en el que se habla de forma más concreta de lo que es la extracción automática de términos, se dio pie a la aplicación de la terminología como recurso de la ingeniería lingüística; para esto, se explicó qué es la extracción auto o semiautomática de términos, señalando las ventajas y desventajas que representa esta tarea, dando un vistazo rápido a lo que es y ha sido la extracción automática de terminología.

Por otra parte, para complementar la información de este capítulo, se habló de los métodos que existen para ser aplicados a programas computacionales (*métodos estadísticos, lingüísticos e híbridos*) y en qué consiste cada método, señalando los beneficios y carencias que presenta cada uno. Asimismo, se señalaron algunos proyectos que han trabajado y se han beneficiado de esta tarea de extracción automática de términos, como por ejemplo el *Proyecto CRATER* o el *Proyecto GLOSMETAL*, encargados de creación de herramientas y repertorios terminológicos.

El capítulo 3 complementó la información del capítulo 2. Luego de haber sido explicadas las metodologías existentes para extraer términos de forma automática, las cuales son aplicadas en programas de cómputo para realizar los procesos de extracción.

En el 3, se señalaron las distintas herramientas que presentan estos programas de cómputo, citando esta descripción al programa WordSmith, que como ya se ha señalado en repetidas ocasiones, no es el mejor ni el peor, pero sí el que se tuvo a la mano para esta investigación.

³ Véase María Teresa Cabré. La terminología: teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993.

Se explicó en qué consiste y cuál es la función de las herramientas:

1. lista de palabras,
2. listas de detención,
3. palabras clave,
4. concordancia,
5. información mutua.

Con el capítulo 4 se puso en práctica la teoría descrita en los anteriores capítulos. En este capítulo se desarrollaron los procesos y análisis de extracción, efectuados según la aplicación de las herramientas que presentaba WordSmith.

Un punto muy importante a resolver respecto de la terminología que fue extraída, consistía en, como arguye Martinet, que “una característica esencial de todo léxico es su carácter ‘abierto’ al constante aumento de vocablos dentro de una lengua, con lo que la identificación del conjunto se hace imposible”⁴. Esta consideración, tomando en cuenta lo que se dijo en el capítulo 1 sobre que la terminología de un área de conocimiento es un componente del léxico, representaba una cuestión prioritaria. Sin embargo, una vez que fueron sucediéndose los análisis, los resultados confirmaban una idea que anulaba las consideraciones de Martinet; ésta es, que los términos, como parte de un léxico, son en cierta medida dentro de un área científica, como lo es la física, y en este caso, mecánica, un círculo cerrado, en el cual no es frecuente la inclusión de nuevos términos de forma periódica.

Con este argumento que era apoyado por los resultados y la delimitación muy específica del objeto de estudio, se realizaron todos los procesos y análisis que dieron como resultado la terminología que se presentó en el capítulo 5.

La terminología final que se obtuvo, luego de todos los procesos y análisis efectuados, consistió de un total de 307 términos, entre simples y poliléxicos. Tal terminología se incluyó en su totalidad en el capítulo 5; además, fue enviada a gente especializada en física del Centro de Instrumentos, quienes evaluaron de forma positiva los resultados presentados.

⁴ Citado en Luis Fernando Lara, Roberto Ham Chande y María Isabel García Hidalgo. Investigaciones lingüísticas en lexicografía. México: COLMEX, 1979. p. 15.

Con la presentación, evaluación y observaciones que los del Centro de Instrumentos hicieron a los 307 términos se dio por concluido este trabajo.

Así que, una vez concluida esta investigación, es necesario, por consiguiente, destacar algunos puntos importantes, a juicio personal, de lo realizado.

- Sin que el aspecto terminológico haya sido tratado extensa y profundamente, existe, según lo investigado en la documentación bibliográfica, una serie de dudas respecto a los límites precisos entre término y palabra; principalmente, cuando éstos entran dentro del campo de la polisemia.
- Otro aspecto importante es que algunos terminólogos plantean una duda muy fuerte en relación con los límites precisos de lo que es un término; hay quienes incluso aducen que no se trata siquiera de términos o palabras, sino, de unidades aún no catalogadas.
- Con respecto a las metodologías que existen para extraer términos de forma auto o semiautomática, es preciso recalcar que aún con la interacción de ellas, es muy difícil que presenten resultados cien por ciento satisfactorios.
- Los métodos estadísticos tienen la ventaja de que proponen una cantidad importante de candidatos a términos, sin que esto quiera decir que en verdad lo sean.
- Puesto que la metodología estadística se basa en la premisa de que a mayor frecuencia de aparición mayor posibilidad de ser término, muchos términos que tienen una frecuencia que no entra dentro de los rangos proporcionados por el usuario, son descartados automáticamente.
- Los métodos lingüísticos deben estar apoyados sobre gramáticas que en sus reglas quede muy claro, a la hora de ser aplicadas por un programa computacional, los límites precisos y las características que debe presentar cada elemento de la oración.
- De lo anterior se deduce que hay una necesidad por adaptar, sino crear, una gramática que se ajuste a las necesidades de los extractores de términos para que la tarea, a fin de cuentas, sea lo más automática posible, además de confiable.
- Resulta también necesario la interacción de más metodologías o herramientas que redunden en recursos útiles y prácticos para extraer términos: por ejemplo, las herramientas que ahora se estudian en el GIL, en otro proyecto de tesis que

tiene como fin identificar contextos definitorios, los cuales pueden coadyuvar a la identificación y extracción de terminología.

- Los programas de cómputo que trabajan en este rubro de extracción de términos, si bien es cierto que no ofrecen un resultado cien por ciento satisfactorio, deben presentar interfaces amigables, prácticas y sencillas para explotar al máximo los recursos y herramientas que presentan.
- Por último, es necesario hacer notar que debe haber una mayor adecuación, investigación y compromiso sobre este tema, lo cual contribuiría en beneficio de terminólogos, profesionales, expertos, investigadores y público en general.

6.2 Trabajos futuros

Si bien con este trabajo se lograron los objetivos que se tenían pensados al inicio de la investigación, e incluso del proyecto del diccionario onomasiológico, como son:

1. estudiar la terminología para aplicarla como recurso de la ingeniería lingüística;
2. proponer un método eficiente y rápido que proporcione la terminología de cualquier área de conocimiento;
3. comprobar tal método mediante la extracción de la terminología de la rama de la física dedicada al estudio de la mecánica, la cual sería incluida en el diccionario onomasiológico de física;
4. obtener una evaluación positiva y satisfactoria de los términos obtenidos dada por gente especializada en el área a la que pertenece tal terminología.

Aún y cuando se lograron estos objetivos, quedan muchas líneas de investigaciones abiertas, muchas dudas por despejar, muchos temas por explorar, además de investigaciones, estudios y análisis para quien esté interesado en el tema.

Los apartados siguientes son un ejemplo de la riqueza de explotación del tema en un futuro:

- evaluar los términos mediante cánones lingüísticos;
- aprovechar las herramientas que se tienen a disposición para crear o mejorar los recursos para extraer términos;

- ahondar en la investigación y delimitación de las fronteras entre términos y palabras;
- planear gramáticas que puedan ofrecer mejores resultados en la tarea extractiva;
- adecuar los métodos estadísticos y lingüísticos a los requerimientos de los terminólogos;
- sistematizar los procesos y análisis de extracción para optimizar la calidad y el tiempo de la presentación de resultados.

BIBLIGRAFÍA

- Anzaldi, Clementina. *Un método para contribuir a la puesta al día de la terminología técnica. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo.* Tomo I, Argentina: (1994): 37 – 42.
- Alpizar Castillo, Rodolfo. *El léxico de la terminología. Algunas precisiones. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo.* Tomo II, Argentina: (1994): 79 – 87.
- Arntz, Reiner y Heribert Picht. Introducción a la terminología. trad. Amelia de Irazazábal, et al. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide, 1995.
- Arroyo Hidalgo, Susana. Aplicación de la lingüística computacional al *Primero sueño* de Sor Juana. México: ITESM-CEM, 1993.
- Ball, Catherine. Concordances and Corpora. [En línea] En: Department of linguistics Georgetown University (1997).
<<http://www.georgetown.edu/cball/corpora/tutorial3.html>>
- Barriga Villanueva, Rebeca y Claudia Parodi. La lingüística en México 1980 – 1996. México: COLMEX / UCLA, 1998.
- Cabré, María Teresa. La terminología: teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993.

- Cabré, María Teresa y Judit Feliu, eds. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001.
- Cabré, María Teresa, Rosa Estopà, Jordi Vivaldi. Recent Advances in Computational Terminology. Didier Bourigault, Christian Jacquemin, Marie—Claude L’Homme eds. Amsterdam / Philadelphia: John Benjamins B. V., 2001.
- Cammelli, Antonio. *Hacia una normalización terminológica para una mejor redacción normativa.* Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo. Tomo II, Argentina: (1994): 37 – 42.
- Cardero, Ana María. El procesamiento de una terminología. Referencia especial a la terminología de control de satélites en el área de las telecomunicaciones en México. Tesis de doctorado. México: 2001.
- Cardero, Ana María. *La integración del corpus de la terminología de control de satélites en México.* Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología. México, (1996):106 - 110.
- Danilenko, V. P. *Sobre la posición de la terminología científica en el sistema léxico de una lengua.* Textos de terminólogos de la escuela rusa. Cabré, María Teresa, et al. Eds. trads. Manuel Barreiro y Delia Vázquez. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 19 – 28.
- De los Ríos Porras, E., et al. *Aplicación de las bases de datos orientadas a objetos a los trabajos terminológicos.* Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo. Tomo I, Argentina: (1994): 33 – 36.

- Dias Gaël, Sylvie Guilloré, Jean-Claude Bassano, José Gabriel Pereira Lopes. Combining Linguistics with Statistics for Multiword Term Extraction: A Fruitful Association? en línea. <http://citeseer.nj.nec.com/context/1674205/410737>
- Diéguez, María Isabel, e Ileana Cabrera Ponce. *Estado de avance proyecto FONDECYT n° 1960441: "Traducción automática versus traducción humana: variables que inciden en la elección de uno u otro método de traducción con miras a optimizar el tiempo, costo y calidad de la traducción. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología.* México, (1996):134 – 143.
- Doyle, Lauren. *Automatic Indexing. Information Retrieval and Processing.* Los Angeles, California: Melville Publishing Company, 1975. 291 – 300.
- Doyle, Lauren. *Word Frequency Methods in Automatic Indexing and Abstracting. Information Retrieval and Processing.* Los Angeles, California: Melville Publishing Company, 1975. 301 – 316.
- Estopà, Rosa. *Elementos lingüísticos de las unidades terminológicas para su extracción automática. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica.* María teresa Cabré y Judit Feliu, eds. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 67-80.
- Fedor de Diego, Alicia. *Algunas reflexiones sobre las "políticas terminológicas". Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología.* México, (1996): 11a - 11d.
- Fedor de Diego, Alicia. *Terminología teoría y práctica.* Venezuela: Universidad Simón Bolívar / Unión Latina, 1995.

- Franceschi, María Eugenia. *La terminología INTERNET. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología.* México, (1996): 144 – 149.
- Galinsky, Christian. *Terminología y sociedad moderna: el papel de INFOTERM. Las industrias de la lengua.* trad. E. Lavín. Madrid: Fundación Germán Sánchez Ruipérez, 1991.
- García Palacios, Joaquín. *La terminología en los manuales de enseñanza media: hacia la determinación de la terminología básica del español. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología.* México, (1996): 150 – 156.
- Guía de alimentación y salud [En línea]. En: Sociedad española de nutrición básica y aplicada (2002). <<http://www.uned.es/pea-nutricion-y-dietetica-I/guia/index.htm>>
- Heaps, H. S. y K. V. Leung. *A System for Automatic Classification of Scientific Literature. Ordering Systems for Global Information Networks.* India: FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. 224 – 234.
- Heaps, H. S. *Information Retrieval. Computational and Theoretical Aspects.* New York: Academic Press, 1978. 263 – 290.
- Helbich, Jan. *Experimental Comparison of Nine Statistical Procedures for Measuring Selective Power of Single Words. Ordering systems for global information networks.* India: FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. 246 – 255.
- Kofnovec, Ladislav y Dusan Simandl. *Quantitative Structure of World Scientific and Technical Literature and its Use for Optimizing General Classifications. Ordering*

Systems for Global Information Networks. India: FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. 363 – 368.

Kulebakin, V. S., y A. Klimovitskii. *Los trabajos para la elaboración de la terminología científico – técnica en la URSS y la Escuela de Terminología Soviética*. Textos de terminólogos de la escuela rusa. Cabré, María Teresa, et al. Eds. trads. Manuel Barreiro y Delia Vázquez. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 61 - 87.

Lancaster, Frederick W. El control del vocabulario en la recuperación de información. València: Universitat de València, 1995.

Lancaster, F. W. Vocabulary Control for Information Retrieval. Washington D.C.: Information Resources Press, 1972.

Lara, Luis Fernando, Roberto Ham Chande y María Isabel García Hidalgo. Investigaciones lingüísticas en lexicografía. México: COLMEX, 1979.

Las industrias de la lengua. Dir. José Vidal Beneyto. trad. Manuel Alvar Ezquerro, et al. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

Luhn, H. P. Potentialities of Auto – Encoding of Scientific Literature. Yorktown Height: CBM, 1959.

Makridakis, Spyros y Wheelwright. Forecasting Methods for management Wiley. 1989.

McEnery, Tony y Andrew Wilson. Corpus Linguistics [En línea]. En: Department of linguistics Lancaster University
<<http://www.ling.lancs.ac.uk/monkey/ihe/linguistics/corpus3/3fra1.htm>>

Nedobity, Wolfgang. *Perspectivas para la elaboración de una metodología común en la descripción del lenguaje. Las industrias de la lengua.* trad. E. Lavín. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

Nedobity, Wolfgang. *Terminología de las ramas e ingeniería del conocimiento. Las industrias de la lengua.* trad. E. Lavín. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

Pérez, J. L., Miranda A., Garcés A., y Lara N. *Hipertexto de física.* En Ruiz G. (Ed.). *SOMI XIII*, México: Sociedad Mexicana de Instrumentación, 1998, 325-329.

Prado, Daniel. *RITerm, proyecto de nodos nacionales y regionales. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo.* Tomo II, Argentina: (1994): 43 – 45.

Proyecto ACORDEON. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~ares/acordeon.html>>

Proyecto ADMYTE. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/~fmarcos/informes/admyte/admyteix.html>>

Proyecto ATILA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/atila.html>>

Proyecto BN ARGENTINA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y

Teoría de la literatura y Literatura
<<http://www.ullf.uam.es/~fmarcos/informes/BNArgentina/BN.htm>>

Proyecto CON-TEXT. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.ullf.uam.es/proyectos/context.html>>

Proyecto C-ORAL-ROM. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.ullf.uam.es/~sandoval/coralrom.html>>

Proyecto CRATER. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.ullf.uam.es/proyectos/crater.html>>

Proyecto CREA. [En línea]. <<http://www.rae.es/NIVEL1/CREA.HTM>>

Proyecto EUROTRA. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <http://www.ullf.uam.es/docs_es/proyectos/APC.html>

Proyecto GLOSMETAL. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.ullf.uam.es/~flora/projects/glosmetal.html>>

Proyecto HAL. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <http://www.ullf.uam.es/docs_es/proyectos/APC.html>

Proyecto PROTEUS. [En línea]. En: Universidad Autónoma de Madrid. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia y Teoría de la literatura y Literatura <<http://www.lllf.uam.es/proyectos/proteus.html>>

Proyecto RITerm BD. [En línea]. <<http://www.riterm.net/es/index.htm>>

Proyecto SIIT VIRTUAL. [En línea]. <<http://www.riterm.net/es/index.htm>>

Ruiz de la Cierva, María del Carmen. *Imagen intelectual de Octavio Paz*. [En línea]. En: Proyecto Ensayo Hispanoamericano (2002). <<http://ensayo.rom.uga.edu/filosofos/mexico/paz/ruiz/>>

Sager, J. C. A Practical Course in Terminology Processing. Amsterdam: John Benjamins Publishing Company, 1990.

Sager, Juan Carlos. Curso práctico sobre el procesamiento de la terminología. trad. Laura Chumillas Muya. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide, 1993.

Sager, Juan Carlos. *La terminología, puente entre varios mundos*, prólogo de la versión española del libro de María Teresa Cabré. La terminología. Teoría, metodología y aplicaciones. Barcelona: Antártida/Empúries, 1993; pp. 11 – 17.

Scott, Mike. WordSmith [En línea]. <<http://www.liv.ac.uk/~ms2928/wsmhomep.htm>>

Sierra Martínez, Gerardo. *Avances en el desarrollo del diccionario electrónico de búsqueda onomasiológica*. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología. México, (1996): 184. 191.

- Sierra Martínez, Gerardo. *Bases del diccionario onomasiológico computarizado en el área de desastres*. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo. Tomo II, Argentina: (1994): 55-59.
- Sierra Martínez, Gerardo. *Estructura semántica del léxico en un diccionario onomasiológico práctico*. Estudios de lingüística aplicada. 23, 24 (1996): 417-425.
- Sierra Martínez, Gerardo. La informática en el trabajo terminológico. Manuscrito. 2002.
- Sparck Jones, K. y R. M. Needham. *Automatic Term Classifications and Retrieval*. Information Storage and Retrieval. 4 (1968): 91 – 100.
- Steinberger, Ralf. *Cross - Lingual Keyword Assignment*. Procesamiento del lenguaje natural. España: SEPLN/Universidad de Jaén, 2001.
- Textos de terminólogos de la escuela rusa. Cabré, María Teresa, et al. Eds. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001.
- Thesen, Gottfried y José Soler. *Iniciativas de la Comunidad Europea en el campo de la terminología*. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo. Tomo II, Argentina: (1994). 33.
- Torres Vargas, Georgina Araceli. *Hacia una metodología terminológica en bibliotecología*. Actas del V simposio iberoamericano de terminología. Terminología, ciencia y tecnología. México, (1996): 192 – 196.
- Vega, José. *Herramientas lingüísticas para la terminología*. Actas del IV simposio iberoamericano de terminología RITerm. Terminología y desarrollo. Tomo II, Argentina: (1994): 25 – 32.

Vivaldi, Jorge. *Elaboración de una aplicación automática de reconocimiento y extracción de información terminológica en textos de dominios restringidos. La terminología científico – técnica: reconocimiento, análisis y extracción de información formal y semántica.* María teresa Cabré y Judit Feliu, eds. Barcelona: Institut Universitari de Lingüística Aplicada Pompeu Fabra, 2001, 299-238.

WordSmith. Lexical Analysis Software for Data-Driven Learning and Research [En línea]. <<http://www.liv.ac.uk/~ms2928/wordsmith>>

WordSmith [En línea]. <<http://www.ndirect.co.uk/~lexical/wordsmi.htm>>

WordSmith [En línea]. <<http://www.ndirect.co.uk/~lexical/wsmhomep.htm>>

WordSmith [En línea]. En: Oxford University Press (2002). <<http://www.oup.co.uk/elt/catalogu/multimed/4589846/4589846.html>>