

# El ejemplo en terminología

Caracterización y extracción automática

Jorge Adrián Lázaro Hernández

---

TESI DOCTORAL UPF / 2015

DIRECTORES DE LA TESI

Dra. Teresa Cabré Castellví, (Institut Universitari de Lingüística  
Aplicada, Universitat Pompeu Fabra)

Dr. Gerardo Sierra Martínez, (Instituto de Ingeniería, Universidad  
Nacional Autónoma de México)



***A Gerardo Sierra***

*Porque a pesar de todo, o quizá por todo ello, nunca dejaste de  
creer en mí.*

*Nunca acabaré de agradecerte todo lo que has hecho por mi  
persona, por ser mi padre académico.*

***A Esther Douriet***

*Porque, irónicamente, empecé a redactar esto cuando te conocí y le  
doy punto final ahora que está tan cercana tu llegada.*

*Gracias por enseñarme la belleza de la imposibilidad.*

*“Me sembraste una especie  
de puesta de sol  
por dentro”*



## Agradecimientos

Este trabajo no hubiera sido posible sin el gran apoyo académico, moral y personal de mi directora de tesis, la Dra. Teresa Cabré. Gracias por siempre estar presente y dispuesta a ayudar en todos los aspectos posibles. Gracias por enseñarme que la grandeza académica se lleva de la mano de un corazón enorme. Aunada a ella, la Dra. Mercé Lorente ha sido no sólo lectora y guía, sino también sostén anímico e inspiración teórica. Gracias por hacer del IULA mi segunda casa.

Al Dr. Juan-Manuel Torres-Moreno, el padre técnico de GENEX. Sin su ayuda y grandes conocimientos seguramente este proyecto no habría llegado a buen puerto. Gracias por recibirme en el grupo de trabajo TALNE del Laboratoire Informatique d'Avignon de la Université d'Avignon et des Pays de Vaucluse. Al Dr. Everardo Mendoza, por sus sabios consejos y por recibirme en la Unidad de Investigación Lingüística y Literaria de la Universidad Autónoma de Sinaloa. Al Dr. Ignacio Rodríguez, por su apoyo académico y por recibirme en la Universidad Autónoma de Querétaro; a la Mtra. Roxana Fitch, por hacer de aquella estancia una gran experiencia de conocimiento. Al Dr. Joaquín García Palacios y a la Dra. Teresa Fuentes Morán, por su invitación y atenciones durante mi estancia en la Facultad de Traducción y Documentación de la Universidad de Salamanca. Al Dr. Adam Kilgarriff de la Lexicography Master Class por habernos cedido el gran corpus esTenTen de Sketch Engine para esta investigación. Descanse en paz.

Agradezco también a toda la gente del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra y a mis compañeros y amigos del Grupo de Ingeniería Lingüística de la Universidad Nacional Autónoma de México. No los nombro a todos para no pecar de omisión. Sé que sabrán dispensar la memoria inmediata, pero saben que tenemos grandes recuerdos y los llevo en el corazón.

En este largo camino de ya cinco años quiero agradecer especialmente a Jesús De la Peña, hermano aunque no de sangre y psicólogo de cabecera. Gracias por acompañarme en la distancia, por los viajes y los consejos, por secar mis lágrimas. A Uana

Cornea por su gran amistad y por mostrarme que se puede querer demasiado la locura. A Octavio Sánchez por ser aeropuerto y amigo entrañable, además de excelente compañero de piso. A Andrés Torres, quien ciegamente cree en lo que hago, que discute y refuta mis ideas, que tiene hambre de conocimiento y disposición de corazón; que me acompaña en lo que era soledad. A los Milsuertes, Blanca y Paco, por hacer de estos últimos años una gran aventura y sacarme de una monotonía rabiosa. A Gemma Bernardó, la mejor catalana del mundo y amiga de verdad; gracias por darme un espacio en tu vida y tu casa, por darme la oportunidad de sentirme en un hogar.

A mi familia, por impulsarme, apoyarme y hacer de mí el hombre que soy, aunque me resista a ser bueno. A mi abuelita Guadalupe, por el carácter; a mi abuelito Benjamín por la sonrisa; a mi madre Adriana por la entereza; a mi tía Jessica por la templanza; a mi hermano Juan Carlos por la nobleza; a otra má, Tere Rodríguez, por no dejarme caer, por la inteligencia emocional; A Jorge De la Peña, por el consejo certero y fulminante; a Don Jesús, madrina Alejandra Rodríguez, Dany y Luis Jorge.

**Esta investigación ha sido llevada a cabo gracias a los siguientes proyectos y apoyos:**

Beca de estudios de doctorado otorgada por el Consejo Nacional de Ciencia y Tecnología (CONACyT) de México, durante el periodo 1 de octubre de 2011 a 31 de julio de 2015. CVU: 287775, registro: 214207.

“Detección y medición automática de similitud textual” a cargo del Dr. Gerardo Sierra Martínez, financiado por el Consejo Nacional de Ciencia y Tecnología (CONACyT), referencia: 178248.

"Análisis estilométrico para la detección de similitud textual". a cargo del Dr. Gerardo Sierra Martínez, financiado por el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la Universidad Nacional Autónoma de México. Registro: IN400312.

“Neologismos generales y neologismos especializados: análisis y contraste desde el punto de vista de su producción, recepción y circulación social” a cargo de la Dra. M. Teresa Cabré Castellví, financiado por el Ministerio de Economía y Competitividad (MINECO) de España, referencia: FF12012-37260





## Resumen

En esta tesis se presenta una propuesta para la extracción automática de ejemplos desde corpus lingüísticos a partir de criterios lingüísticos y estadísticos. Con base en una fórmula denominada *densidad lexicométrica* se creó una herramienta capaz de obtener candidatos a ejemplos en español: GENEX (Générateur d'Exemples). La *densidad lexicométrica* surge a partir de la propuesta teórica denominada *saturación semántica*, una noción que intenta explicar el proceso por el cual, idealmente, un concepto podría conocerse en todas sus realizaciones. Se hace una propuesta, también, de caracterización del ejemplo en terminología con el fin de demostrar que es una categoría de datos imprescindible en un diccionario ya que auxilia el proceso de aprehensión de nuevos conceptos y funciona como complemento conceptual de una definición terminográfica.

## Abstract

In this thesis, a proposal for automatic extraction of examples using a linguistic corpus, applying statistical and linguistic criteria, is presented. Using the *lexicometrical density* formula as basis, a program able to obtain candidates for examples in spanish was created: GENEX (Générateur d'Exemples). *Lexicometrical density* was born from a theoretical proposal called *semantic saturation*, a notion that tries to explain the process from which, ideally, a certain concept could be known in all its realisations. Also, a theoretical proposal for the characterisation of the example in terminology is made, with the purpose of proving that it's a vital data category in a dictionary, as it aids the process of apprehension of new concepts and works as a conceptual complement of the terminographic definition.



# Índex

Resumen .....	ix
Abstract .....	ix
Lista de figuras.....	xvi
<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
1.1. Objeto de estudio .....	1
1.2. Objetivos.....	1
1.3. Hipótesis .....	2
1.4. Metodología: los corpus masivos y la evidencia lingüística .....	3
1.5. Estructura de la tesis .....	9
<b>2. LA RELACIÓN ENTRE TÉRMINOS, DEFINICIONES Y CONCEPTOS .....</b>	<b>13</b>
<b>2.1. La relación concepto-término.....</b>	<b>14</b>
2.1.1. El concepto y el significado.....	18
2.1.2. El concepto y el mecanismo de activación de un término .....	20
2.1.3. El concepto en el discurso especializado.....	24
<b>2.2. La relación término-definición .....</b>	<b>28</b>
2.2.1. Los Contextos Definitorios en terminología .....	30
2.2.2. La definición terminológica.....	34
<b>2.3. La relación concepto-término-definición.....</b>	<b>36</b>
<b>2.4. Discusión: Las limitaciones de la definición terminográfica.....</b>	<b>40</b>
<b>3. LA SATURACIÓN SEMÁNTICA, MOTOR DEL EJEMPLO .....</b>	<b>47</b>
<b>3.1. La saciedad semántica, ¿una noción previa a la saturación? .....</b>	<b>47</b>
3.1.1. Saciedad vs saturación.....	49
<b>3.2. La noción multifacética de un término: el <i>Principio de poliedricidad</i>.....</b>	<b>52</b>
<b>3.3. El Principio de adecuación.....</b>	<b>53</b>
<b>3.4. La <i>saturación semántica</i> como mecanismo de complemento a la definición .....</b>	<b>54</b>
<b>3.5. La <i>saturación semántica</i> como mecanismo de operación del ejemplo.....</b>	<b>58</b>

<b>3.6. Discusión: la saturación semántica, prolegómenos para su tratamiento en terminología .....</b>	<b>60</b>
<b>4. EL EJEMPLO, HOY .....</b>	<b>67</b>
<b>4.1. La ejemplificación en diccionarios actuales, un estudio</b>	<b>67</b>
<b>4.2. Breve esbozo del ejemplo en lexicografía .....</b>	<b>71</b>
4.2.1. Zgusta .....	73
4.2.2. Meyer .....	74
4.2.3. Atkins & Rundell .....	75
4.2.4. Rey .....	76
4.2.5. Wiegand .....	77
4.2.6. Lara .....	77
4.2.7. Fitch .....	79
<b>4.3. El tratamiento actual del ejemplo en terminología .....</b>	<b>81</b>
4.3.1. Ahmad .....	81
4.3.2. Fuentes Morán y García Palacios .....	82
<b>4.4. Caracterización teórica de un ejemplo .....</b>	<b>83</b>
4.4.1. Funciones .....	83
4.4.2. Forma .....	84
<b>4.5. Los ejemplos definitorios oblicuos .....</b>	<b>87</b>
<b>4.6. Discusión: Un ejemplo siempre ha sido un recurso, pero no un elemento obligatorio.....</b>	<b>88</b>
<b>5. EL EJEMPLO EN TERMINOLOGÍA, UNA PROPUESTA</b>	<b>91</b>
<b>5.1. Experimento SKE .....</b>	<b>91</b>
<b>5.2. Determinando la estructura de un ejemplo en terminología .....</b>	<b>101</b>
5.2.1. El ejemplo no está ahí <i>a priori</i> .....	101
5.2.2. El ejemplo no es una concordancia .....	104
5.2.3. El ejemplo no es una colocación .....	105
5.2.4. El ejemplo no es una muestra de uso .....	106
5.2.5. El ejemplo no es una invención .....	107
5.2.6. El ejemplo se diseña .....	108
<b>5.3. Caracterización del ejemplo bajo criterios funcionales</b>	<b>111</b>
5.3.1. El ejemplo guarda una relación imprescindible con la definición .....	111
5.3.2. El ejemplo es una reformulación definitoria .....	112
5.3.3. El ejemplo como núcleo conceptual .....	113
5.3.4. El ejemplo es un elemento pragmático .....	116
5.3.5. El ejemplo como puente .....	116

<b>5.4. Caracterización del ejemplo bajo criterios sintáctico-semánticos.....</b>	<b>117</b>
5.4.1. El foco conceptual .....	117
5.4.3. La ampliación de los rasgos.....	118
5.4.4. La estabilización de la estructura.....	120
5.4.5. La delimitación de contextos.....	122
5.4.6. Los elementos prescindibles.....	123
<b>5.5. El ejemplo en terminología, su mecanismo de acción y una medida de asociación.....</b>	<b>124</b>
5.5.1. El mecanismo de acción del ejemplo en terminología: la <i>saturación semántica</i> .....	124
5.5.1. La medición de la <i>saturación semántica</i> en un ejemplo: la <i>densidad lexicométrica</i> .....	125
<b>5.6. Recapitulación.....</b>	<b>127</b>
<b>6. UN PROGRAMA PARA LA EXTRACCIÓN AUTOMÁTICA DE EJEMPLOS EN TERMINOLOGÍA: GENEX.....</b>	<b>129</b>
<b>6.1. Un antecedente no tan cercano: GDEX.....</b>	<b>129</b>
<b>6.2. Corpus de análisis.....</b>	<b>133</b>
6.2.1. Corpus <i>Jornada</i> .....	133
6.2.2. Corpus <i>Wiki</i> .....	134
6.2.3. Corpus <i>EsTenTen</i> .....	134
<b>6.3. Preprocesamiento del material de trabajo .....</b>	<b>134</b>
6.3.1. Recuperación .....	135
6.3.2. Extracción.....	138
6.3.3. Limpieza y recodificación .....	139
6.3.4. Segmentación.....	140
6.3.5. Etiquetado.....	142
<b>6.4. La formalización del ejemplo .....</b>	<b>143</b>
6.4.1. La desarticulación de la definición.....	143
6.4.2. El contexto nominal.....	145
6.4.3. El contexto verbal.....	146
<b>6.5. Reglas para la extracción de candidatos a ejemplos ....</b>	<b>148</b>
6.5.1. La utilidad de Cortex .....	148
6.5.2. La función coseno.....	148
6.5.3. El corte por longitud.....	150
6.5.4. La aplicación de la <i>densidad lexicométrica</i> .....	150
<b>6.6. El algoritmo general de extracción .....</b>	<b>152</b>
<b>6.7. El experimento .....</b>	<b>155</b>
<b>6.8. La evaluación .....</b>	<b>165</b>

6.9. Recapitulación.....	169
<b>7. CONCLUSIONES Y APORTACIONES.....</b>	<b>173</b>
7.1. Conclusiones por capítulo.....	173
7.2. Aportaciones al trabajo terminográfico y terminológico .....	179
7.2.1. La noción de <i>saturación semántica</i> .....	179
7.2.2. La noción de densidad lexicométrica.....	180
7.2.3. La redefinición del término <i>ejemplo</i> .....	181
7.3. Campo de aplicación.....	182
7.3.1. Integración a una herramienta de extracción conceptual (Describe <sup>®</sup> ).....	182
7.3.2. Integración a diccionarios.....	183
7.3.3. Integración a herramientas de organización de términos (Terminus).....	184
7.3.4. Aplicación comercial.....	185
7.4. Trabajo futuro.....	186
7.4.1. Afinación de la fórmula para medir la <i>densidad</i> <i>lexicométrica</i> .....	186
7.4.2. GENEX+ ( <i>plus</i> ).....	187
7.4.3. ¿Es posible crear automáticamente un ejemplo?.....	187
<b>Bibliografía.....</b>	<b>189</b>
<b>ANEXO 1. Resultados del experimento con GENEX.....</b>	<b>209</b>



## Lista de figuras

	Pàg.
Figura 2.1. Esquema de la comunicación de Jakobson...	26
Figura 2.2. Esquema general de un Contexto Definitorio	31
Figura 2.3. Esquema detallado de un Contexto Definitorio.....	31
Figura 2.4. Lista de verbos definitorios.....	32
Figura 4.1. Lista de términos.....	68
Figura 4.2. Recursos consultados.....	68
Figura 4.3. Captura de pantalla DRAE, 6ª Ed. 1822.....	85
Figura 5.1. Concordancias del término <i>virus</i> .....	93
Figura 5.2. Estructuras más comunes asociadas al término <i>virus</i> .....	94
Figura 5.3. Estructuras más comunes asociadas al término <i>virus</i> con el verbo <i>contraer</i> .....	95
Figura 5.4. Estructuras más comunes asociadas al término <i>virus</i> con el verbo <i>propagar</i> .....	95
Figura 5.5. Búsqueda compleja con el término <i>virus</i> como núcleo y el verbo <i>propagar</i> como contexto.....	97
Figura 5.6. Formas más comunes para las colocaciones de <i>virus</i> como núcleo y el verbo <i>propagar</i> como contexto.....	99
Figura 5.7. Colocaciones de <i>virus</i> como núcleo y el verbo <i>propagar</i> como contexto donde se focaliza la forma más productiva del verbo.....	100
Figura 5.8. Heurística de GDEX.....	103
Figura 5.9. Una definición ostensiva de acuerdo con Pozzi, ejemplo como núcleo conceptual, de acuerdo con nuestra propuesta.....	115
Figura 5.10. Lista simplificada de verbos definitorios.....	119
Figura 6.1. Algunos resultados de candidatos a ejemplo para <i>baby</i> obtenidos con GDEX.....	130
Figura 6.2. Lista de colocaciones más frecuentes para <i>virus</i> en esTenTen.....	132
Figura 6.3. Resultados más comunes del término <i>virus</i> y su colocación más frecuente ( <i>influenza</i> ).....	132
Figura 6.4. Fragmento de inicio del código	

<i>WikiExtractor.py</i> .....	135
Figura 6.5. <i>ProcesaWiki.sh</i> : Extractor y conversor a texto plano de artículos de la Wikipedia.....	136
Figura 6.6. Inicio del script <i>split_nonum.pl</i> , programa de segmentación cadena-por-línea.....	137
Figura 6.7. Script completo de <i>extractor02.sh</i> .....	139
Figura 6.8. Programa <i>elimina_NOutf8.pl</i> .....	140
Figura 6.9. Visualización del corpus Jornada separado por frases con Cortex.....	141
Figura 6.10. Fragmento de corpus etiquetado con TreeTagger.....	142
Figura 6.11. Lista simplificada de verbos definitorios....	147
Figura 6.12. Arquitectura del GENEX.....	152
Figura 6.13. Resultados de la evaluación del GENEX de acuerdo con el <i>Coficiente de correlación de Pearson</i> ...	166



# 1. INTRODUCCIÓN

## 1.1. Objeto de estudio

El presente trabajo tiene como objeto de estudio el ejemplo, en primer lugar, y la ejemplificación en diccionarios desde un enfoque terminológico, en segundo lugar. El trabajo se encuadra, además, en las potenciales aplicaciones informáticas derivadas de los resultados, esto es, el quehacer de la terminótica (Cabré, 1992).

El ejemplo como tema de estudio ha sido abordado por los más diversos autores, sin embargo, su caracterización ha sido siempre un tema polémico debido a lo complejo de su naturaleza. No hay un conjunto de criterios homogéneos para poder determinar qué es un ejemplo en una lengua dada, en un tipo de diccionario específico o en un área de estudio particular. Esto es, hay discrepancia en la caracterización de lo que es un ejemplo, primero, y no existe una propuesta de cómo obtenerlos, a la vez que en terminótica no existen métodos para extraerlos o generarlos automáticamente. En el apartado 6.1. podremos ver que existen algunas metodologías de trabajo eficientes para lexicografía, pero ninguna con un desempeño óptimo en general. El sostén de este argumento lo encontramos con la aplicación de la metodología utilizada por GDEX (Kilgarriff, Husák, Rundell & Rychly, 2008).

Por esa misma razón, este trabajo gira en torno al descubrimiento y establecimiento de un conjunto de criterios mínimos para un tipo de diccionario (terminológico), en un experimento para una lengua específica (el español), en tres distintas áreas de especialidad: finanzas, informática y sexualidad.

## 1.2. Objetivos

Esta tesis tiene dos objetivos principales. Por un lado se intentará demostrar que hay un conjunto de criterios mínimos para que un fragmento textual sea considerado *ejemplo*. Esto es, se desea

construir una lista de rasgos que puedan llegar a caracterizar, a homogeneizar, las diversas posturas que hay en torno al estudio de los ejemplos en diversos tipos de diccionarios. Como objetivo principal, se intentará describir y caracterizar el ejemplo en diccionarios de especialidad.

Por otro lado, una vez obtenidos los criterios de caracterización, el siguiente objetivo principal será formalizar la lista de rasgos en una heurística capaz de describir el funcionamiento del proceso de ejemplificación en los humanos para, posteriormente, diseñar un algoritmo computacional con el fin de crear una herramienta que extraiga automáticamente ejemplos desde corpus textuales.

### **1.3. Hipótesis**

Partiendo de la premisa de que toda lengua puede ser aprendida (en situaciones normales o estándar) por imitación, creemos que el ejemplo es el mecanismo por antonomasia por el cual una lengua cualquiera se transmite de una mejor manera. Sin embargo, el proceso por el cual operaría un ejemplo no es un elemento con una forma definida o con unas reglas de formación predecibles.

Con esta idea en mente, y teniendo en cuenta que aún así casi cualquier persona es capaz de dar un ejemplo acerca de una palabra, pensamos que dicho mecanismo no tiene que ver exactamente con un proceso cien por ciento lingüístico, sino que es el reflejo de un proceso cognitivo. Si esto es cierto, antes que buscar una estructura específica que funcione para ejemplificar un cierto número de palabras, invertimos el proceso y creemos que hay un conjunto más o menos amplio de procesos que operan bajo principios básicos cuya motivación se encuentra en algún proceso mental más amplio. Si dicho proceso puede ser emulado, entonces podría arrojar luz sobre cómo hacer que una computadora, imitando el comportamiento humano, extraiga ejemplos de forma automática. En una línea: si no hay pruebas de la existencia de un elemento (ejemplo), entonces se debería de imitar el proceso (la ejemplificación), para poder analizar los resultados arrojados por ese proceso (los candidatos a ejemplos) y así decidir si existe o no

una estructura definida de los elementos (estructura ejemplificativa) y finalmente, proponer esos resultados filtrados como ejemplo.

La hipótesis, así vista, se divide entonces en dos premisas básicas:

- a) Que el ejemplo se puede caracterizar, es decir, encuadrar en un conjunto mínimo de reglas de formación que atiendan a criterios más cognitivos que lingüísticos y,
- b) que si el ejemplo se puede caracterizar, entonces se puede formalizar y, por tanto, dar pie a una herramienta computacional que emule dicho proceso cognitivo o mental.

#### **1.4. Metodología: los corpus masivos y la evidencia lingüística**

Tomando en cuenta las hipótesis, la primera parte del trabajo corresponde a la aportación teórica que se pretende hacer, es decir, coadyuvar en el conocimiento de una categoría de datos muy importante en la confección de diccionarios que durante muchos años ha sido relegada a elemento secundario. La caracterización implica llegar a poder especificar qué es un ejemplo y cuáles son las características que debe tener para ser considerado como tal. El objeto de estudio es, por tanto, el ejemplo en terminología.

La segunda parte del trabajo se centra en el descubrimiento y aplicación de una técnica de extracción que derive en una herramienta de apoyo al trabajo terminográfico, es decir, el diseño de diccionarios de especialidad. El objeto de estudio aquí es, por tanto, la extracción automática del ejemplo postulado teóricamente en la primera parte de la tesis. Lo anterior da como resultado que la ejemplificación sea, en este contexto, un proceso computacional que intenta imitar el proceso cognitivo de un humano.

Los datos masivos (o Big Data) son desde hace algunos años una realidad indiscutible. Es cada vez mayor la tendencia a utilizar una cantidad inconmensurable de todo tipo de datos para poder

satisfacer necesidades tan distintas como el marketing o la salud, la física o la minería de opinión, la geografía o la lingüística, entre muchas otras áreas del conocimiento humano (Mayer-Schönberger & Cukier, 2013). Esta masividad de información (y los resultados de su procesamiento) nos llega día a día, por ejemplo, en las ofertas de productos y servicios o en la amplia cartelera de actividades que viaja directa hasta nuestros teléfonos móviles y en publicidad adaptada que utiliza nuestros archivos temporales para ofrecernos, en ciertos casos, un descuento considerable en productos sin gluten si acaso nuestro perfil en Facebook está lleno de recomendaciones para celíacos. Estos ejemplos, sin embargo, son sólo una muestra de lo importante que se pueden tornar millones de unidades de información para poder crear patrones que sirvan de impulso a programas que automaticen ciertas tareas que, hasta ahora, se vertían en el poder de persuasión de una persona o un conjunto de éstas.

Sin embargo, debido a que el acceso a todos estos datos era hasta hace poco tiempo libre o poco regulado, también se puede dar el caso de una mala utilización de la información almacenada para llevar a cabo ilícitos que desembocan algunas veces en delitos informáticos. Sin ir más lejos, el robo de información personal para acrecentar una cartera de clientes es frecuente e incluso se ha utilizado para incurrir en suplantación de identidad o *pishing*<sup>1</sup>. En otros casos, por ejemplo, el envío de correos entre personas durante mucho tiempo y bajo ciertas circunstancias puede ocultar también delitos de distinta naturaleza (Sierra et al., 2011).

¿Y por qué mencionamos todo esto? Resulta que en cada caso que analicemos dónde se usan datos masivos podremos encontrar por lo menos tres puntos básicos que convergen entre ellos:

- 1) Los datos masivos dan precisión a los estudios de todo tipo. Cuantos más datos disponibles, menor margen de error.

---

<sup>1</sup> Un caso interesante sobre suplantación de identidad para cometer un delito de fraude bancario es el que puede leerse en la sentencia del Tribunal Supremo de lo Penal de Madrid, en 2007, que dio pie a la formulación de varias leyes que ahora forman parte del Código Penal de España. La sentencia completa puede leerse en: [http://www.belt.es/jurisprudencia/anterior/seg\\_inf\\_y\\_prot\\_datos/com\\_electronico/120607-Sentencia-%20Phising.pdf](http://www.belt.es/jurisprudencia/anterior/seg_inf_y_prot_datos/com_electronico/120607-Sentencia-%20Phising.pdf)

- 2) Los datos masivos permiten predecir un comportamiento. Una vez que se puede cuantificar un fenómeno es mucho más fácil reproducirlo para después interpretar los resultados agregando variantes, y
- 3) Los datos masivos permiten determinar patrones. Hay fenómenos que a pequeña escala no presentan continuidad o recursividad aún cuando la intuición de quien los estudia le dicta que dichos patrones están ahí porque en los sistemas macro, como el lenguaje, los puede ver repetidos<sup>2</sup>.

Sobre este último punto es donde nos interesa detenernos. En este sentido, gracias a un experimento previo aplicado a dieciocho diccionarios generales, especializados y bases de datos terminológicas (que detallamos en 5.1.) se mostró que el ejemplo está asociado a una definición (Lázaro, 2011). Este primer acercamiento para la determinación de criterios que permitiesen definir la estructura de un ejemplo fue infructífero debido al tamaño mínimo del corpus utilizado (aproximadamente 1 millón de palabras) y las razones que encontramos nos llevaron a sopesar otras posibilidades de análisis lingüístico y el uso de corpus más grandes.

Un corpus lingüístico podía considerarse grande si contenía 10 millones palabras hace apenas unos años (Torruella & Llisteri, 1999), sin embargo, y como lo podemos experimentar cada día, el avance de la tecnología y sobre todo de los medios de almacenamiento de datos en equipos de cómputo ha permitido que apenas en poco menos de dos décadas hayamos pasado a decir que un corpus es mínimo si contiene 1 millón de palabras, o pequeño si contiene 10 millones, y a considerar que otro de 100 millones representa un tamaño adecuado o estándar.

---

<sup>2</sup> Pensemos aquí en las *colocaciones*, estas estructuras recurrentes a medio camino de lexicalización que no son posibles de ver en corpus pequeños. Un estudio interesante sobre la identificación de colocaciones en lexicografía, aplicado a 23 diccionarios a la vez, lo podemos encontrar en el estudio de (Fuertes-Olivera, Bergenholtz, Nielsen & Amo, 2012). Es interesante mencionar aquí que dicho estudio, aunque está basado en un tipo de diccionario “especializado” en particular, los *Account Dictionaries*, sigue siendo motivo de estudio de la lexicografía o la lexicografía especializada y por tanto la noción de *colocación* no se inserta en el campo de la terminología. Sobre este punto, el de las colocaciones en diccionarios de especialidad, ya volveremos en el apartado 5.1.3. de esta tesis.

Ahora bien, como los mismos autores hacen notar, el tamaño de un corpus siempre es un tema de discusión debido al carácter subjetivo de las afirmaciones que se pueden hacer con respecto al número de palabras o tokens que contienen, pero también es cierto que hay una gran diferencia entre los 100 millones que podemos procesar y almacenar en casi cualquier equipo doméstico y, por ejemplo, los 380 millones que representan el contenido de un diario a lo largo de su historia. No hay casi que mencionar, entonces, la abismal diferencia con aquel que contabiliza más de 2 mil millones de palabras provenientes de todas las variantes de una lengua. Es justo esta diferencia, la de un corpus como el *esTenTen* con respecto a *Wiki*, y de *Wiki* a las muestras presentadas por otros corpus como el *Corpus Tècnic de l'IULA*: 22 millones de palabras a 2004 (Cabrè & Bach, 2004), lo que nos lleva a integrarlos en el campo de estudio de los datos masivos.

También hay que considerar que los tamaños de los corpus de los que hablamos provocan que, si están almacenados en un archivo o en un conjunto de archivos, estos no pueden ser visualizados por casi ningún procesador de texto, tienen que ser vistos por fragmentos. Hablamos, pues, de *corpus masivos* cuando su tratamiento automático se complica o requiere herramientas y recursos computacionales más avanzados cada día<sup>3</sup>. Aún cuando una máquina puede llevar a cabo las tareas que se le encomiendan, el ojo humano no es capaz de ir monitoreando los pasos de un determinado proceso, esto nos lleva a diseñar algoritmos más precisos que tienen que ser probados varias veces en corpus de tamaños cada vez mayores para poder determinar si su funcionamiento es correcto o cercano a los resultados de las hipótesis planteadas. Esto nos lleva a la segunda parte de la discusión: la evidencia lingüística.

Un patrón en lingüística es, a veces, un problema igualmente complejo y subjetivo. Se trata de una noción que, por un lado,

---

<sup>3</sup> Para ilustrar esta complejidad baste mencionar que para almacenar un corpus como *Wiki* con una conexión estándar de Internet (descarga de hasta 20MB por segundo), se requiere un tiempo aproximado de 8 horas. Etiquetar 5 GB de texto con una computadora de última generación nos tomó alrededor de 36 horas. Descargar todo el corpus *esTenTen* nos requirió un día.

puede parecer intuitiva y sistematizable si nos ponemos a pensar que el proceso de aprendizaje de lenguas se sirve de ellos: ejemplos de patrones en lingüística son el número y el orden de fonemas en una palabra, las declinaciones verbales, el orden de los elementos en una oración (SVO, SOV...), la concordancia, los constituyentes sintácticos maximales, etcétera. Sin embargo, hay otros patrones que se alejan un poco más de la sintaxis, se encuentran en el orden de lo sintáctico-semántico y resultan más difíciles de aprehender o sistematizar: la *consequitio temporum*, la ergatividad o el proceso de palatalización de una yod -si incluimos fenómenos fonéticos-; etcétera. Dentro de estos patrones complejos encontramos también la estructura de las oraciones subordinadas o la relación y estructura de las oraciones introducidas por marcadores discursivos (da Cunha et al., 2012).

Como podemos ver, conforme el nivel de lengua que se analiza se inclina hacia lo pragmático, el descubrimiento o análisis de sus patrones se complica. Es por esto que los datos necesarios para describir y analizar los fenómenos correspondientes a cada nivel aumentan conforme se alejan de la norma y se insertan en el uso. Mientras más común es un elemento lingüístico, y por tanto más ejemplos de éste existen en la lengua del día a día, menos datos serían necesarios para poder transmitir su mecanismo de funcionamiento. Pensemos aquí en el funcionamiento de las frases que acompañan a los memes tan de moda ahora en Internet. No es lo mismo explicar el significado de YOLO, WTF o LOL que el rompimiento de la consecución temporal (*consequitio temporum*) en el español de algunos países americanos<sup>4</sup>. Aunque ambos fenómenos lingüísticos son identificables a simple vista (o también por el oído, en el segundo caso) porque están generalizados en la comunidad de hablantes hispanos, el segundo caso necesitaría una búsqueda y selección más minuciosa de ejemplos para poder explicarlo, ya que no es un fenómeno viral (no hay norma social que lo rija) y su aparición se da a nivel de habla y no tanto en lo escrito. Aparte, es un tema especializado del que no se habla o sobre el que no se repara en la lengua diaria. El primer fenómeno, en

---

<sup>4</sup> Por ejemplo: “Le dije que viniera a la casa para que saque al perro a pasear” por “Le dije que viniera a casa para que sacara al perro a pasear”. Sobre este tema se ha encontrado poca información y es un fenómeno extendido sobre todo en Ecuador y alrededores.

cambio, puede ser encontrado prácticamente con una simple búsqueda en Internet.

Pues bien, es justo esa cantidad de datos que se necesitan para documentar un fenómeno lo que hace compleja una tarea de, primero, caracterización, y luego de extracción. En nuestro caso particular podremos ver que después de una revisión bibliográfica sobre el concepto de ejemplo, no se encontraron pruebas de que éste tuviese una estructura fija o un conjunto de patrones que lo caracterizaran. Lo más cercano que se halló fue la descripción de las características *ideales* que debería tener y ciertas funciones que debía cumplir; sin embargo, en dos reproducciones de los experimentos documentados, no se hallaron pruebas fehacientes de que los argumentos mencionados fueran producto directo de la observación de datos, sino más bien interpretaciones acerca de los datos obtenidos. Dichos estudios están descritos en 4.2., para mostrar la ausencia/presencia de ejemplos en diccionarios generales y especializados; y en 5.2., para mostrar que la aplicación o requerimiento de las características “ideales” no aseguran que un fragmento textual sea un buen ejemplo como parte de ciertas estructuras como concordancias o colocaciones.

El objetivo central de esta discusión es mostrar que ciertos fenómenos lingüísticos, sobre todo semánticos, tienen su base en patrones que no pueden ser aprehendidos ni tratados como muchos otros donde la estructura de una lengua, su sintaxis o sus reglas fonológicas, los hacen palpables sino a simple vista, sí como un conjunto de características sistematizables.

El ejemplo en terminología, como veremos a lo largo de esta investigación, es un elemento semántico de base cognitiva que es difícil de determinar a través de patrones sintácticos. A diferencia de otros estudios que durante los últimos años han tratado de extraerlo a partir de colocaciones, concordancias y medidas de legibilidad –readability– (Kilgarriff et al., 2008), para nosotros se trata de una estructura mental que encuentra sus bases en las características conceptuales de un término y que se proyecta en las más diversas estructuras fraseológicas u oracionales. Es esta diversidad de proyecciones lo que hace de él un elemento difícilmente aprehensible y para lo cual se necesitan una cantidad de datos enormes: millones de frases, corpus de cientos de millones de

palabras, reestructuraciones sintácticas, reformulaciones y hasta sustituciones léxicas. El patrón que sigue un ejemplo en terminología, por tanto, es lo que hemos llamado tentativamente un *patrón conceptual ejemplificativo*. Dicho “patrón” sólo ha podido ser encontrado a través del análisis de cantidades enormes de datos en corpus masivos como veremos en el capítulo 6.

## 1.5. Estructura de la tesis

La metodología de trabajo para esta tesis se basó en tres etapas que cubrieron:

1. La preparación del material y los conceptos previos que fueron tomados en cuenta (Capítulos 1 y 2);
2. El ajuste y la adaptación de criterios para la identificación de ejemplos, así como la propuesta de un principio teórico para el quehacer terminológico (capítulos 3, 4 y 5) y;
3. La aplicación de ese principio a través de procedimientos computacionales para la generación de una herramienta extractora automática y la comprobación de resultados de dicha herramienta a través de una evaluación hecha por humanos (Capítulo 6).

En el capítulo 2, 3 y 4 veremos los conceptos previos tomados en cuenta como punto de partida para este trabajo. En 2 revisaremos la relación existente entre término, definición y concepto, esto es, intentamos descubrir cuáles serían las bases teóricas de las que se ha partido para poder establecer de dónde toma sus rasgos semánticos el ejemplo y cómo éstos son proyectados. Una vez establecidas las características de la definición, de su relación con el concepto y de la operación del mecanismo que activa un término, se podrán inferir los alcances de dicha definición terminológica y las relaciones que guarda con respecto a los elementos que la rodean en una entrada: las categorías de datos.

Este último análisis, empero, no se puede saber si desvelará el origen o la forma específica de un ejemplo en un diccionario

terminológico; así que una vez determinadas las características esenciales de un ejemplo y sus relaciones, nos daremos a la tarea de comparar y analizar cuáles de éstas coinciden o divergen en contraste con las características de la definición determinadas en el capítulo 2. Así, en el apartado 2.4., nuestra metodología intentará ponderar si, como lo mencionan varios autores, debido a su propia naturaleza, esto es, su estructura cuasi-estática, la definición terminológica conlleva ciertas limitaciones. Entre algunas de esas limitaciones que documentamos se encuentra una esencial: la incapacidad para reflejar de una sola vez, o en un conjunto de definiciones y acepciones, todos los rasgos semánticos de un concepto específico. Para poder superar este escollo veremos que se hace necesaria la postulación de un principio que rijan la conformación del ejemplo como elemento complementario a la definición. Como resultado podremos observar la postulación teórica de la noción de *saturación semántica*, la cual detallaremos en el capítulo 3. El desarrollo de esta noción teórica estará apoyado sobre dos principios extraídos de la Teoría Comunicativa de la Terminología (TCT en adelante): el *Principio de Adecuación* (Lorente, 2013) y el *Principio de Poliedricidad* (Cabré, 2008). Una vez que en la teoría la propuesta de *saturación semántica* sea comprobada y concatenada con los dos principios anteriores, nos dispondremos a intentar redefinir la noción de ejemplo para adaptarlo a la teoría terminológica y describiremos sus características mínimas, su funcionamiento, su origen y su pertinencia en diccionarios de especialidad.

En el capítulo 4 analizaremos distintas posturas acerca de cuál es la forma y cuáles son las funciones que el ejemplo puede cumplir en una entrada de diccionario. Por un lado se describirán los criterios funcionales que llevan a un especialista a elegir uno u otro fragmento textual dependiendo de sus necesidades. Por otro lado, se distinguirán las diferentes estructuras que se han utilizado para delimitar el ejemplo y las justificaciones para asociarlo a alguna de ellas; desde aquellas que resumen todo a que una estructura particular funciona como un ejemplo pleno, hasta aquellas que combinan dos o más estructuras bajo el argumento de que no siempre una de ellas es la óptima de acuerdo con la información dada por una palabra. Finalmente se podrá observar que ambos análisis dan como resultado la caracterización del

ejemplo en lexicografía y la relación que guarda con las otras categorías de datos de la entrada a la que pertenece.

En el capítulo 5 podremos ver la caracterización y la definición de nuestra propuesta de ejemplo en terminología. Describiremos minuciosamente los criterios funcionales, sintácticos y semánticos que tomamos en cuenta. Nos daremos a la tarea de mostrar sus posibles formas y las implicaciones que cada una de ellas proyecta a propósito de su papel complementario a la definición terminográfica. Mostraremos la heurística para la resolución de una estructura que funcione y finalmente el algoritmo: la propuesta teórica acerca del *diseño* del ejemplo con el fin de intentar establecer las reglas de formalización para el desarrollo de la herramienta computacional.

Finalmente, en el capítulo 6 veremos la preparación de los corpus que fueron utilizados para la parte experimental de la investigación. Se podrá notar que son de tres naturalezas completamente distintas y dicha variedad le dio riqueza a los resultados. Los procesos de obtención han sido automatizados en dos de ellos a través de crawlers (*Wikipedia* y *Jornada*) y el tercero se trata de una donación de los propios autores: *esTenTen* de Sketch Engine (Kilgarriff & Renau, 2013). Uno de ellos, *Jornada*, como veremos, tuvo que pasar además por un proceso de extracción, ya que sólo fue posible obtenerlo en formato HTML. Se explicará cómo los tres corpus fueron recodificados a formato UTF-8 y limpiados para que todo el contenido estuviese en texto plano. En el caso de *esTenTen*, además, se desetiquetó para que pudiese estar disponible de forma no anotada. Finalmente todos fueron segmentados con la herramienta *Cortex* (Torres-Moreno, Velázquez-Morales & Meunier, 2002) para poder tenerlos divididos en fragmentos de texto.

Para la parte aplicada, el diseño de la herramienta para la extracción automática de ejemplos, describiremos el enfoque que hemos tomado en cuenta: Recuperación de Información, basado sobre todo en los estudios contenidos en el libro *Modern Information Retrieval* (Baeza-Yates & Ribeiro-Neto, 1999), describiremos además la noción de *Información Mutua* aplicada a diccionarios (Ward Church & Hanks, 1990); y argumentaremos las razones por las que elegimos la *función coseno* (Salton, 1989)

aplicada a búsqueda de información (Spärck-Jones, 1972). Finalmente, para evaluar, explicaremos cómo funciona el *Coefficiente de correlación lineal de Pearson* (StatSoft, 2013), una medida de la relación lineal entre dos variables cuantitativas y cómo esta medida permite calcular el grado de relación y la pertinencia de los ejemplos generados de acuerdo con los criterios establecidos y la preferencia de los hablantes sobre la información aportada por los ejemplos con respecto a una definición dada<sup>5</sup>.

Describiremos en este capítulo, además, cómo aplicamos los análisis anteriores sobre la definición y el ejemplo vistos desde la lexicografía, el principio de *saturación semántica*, los criterios del ejemplo en terminología determinados por este estudio y la heurística aplicada a un algoritmo de extracción. Se mostrará que cada uno de los puntos anteriores tiene una representación formal que atañe a los procesos que realiza el *Generador de Ejemplos* (GENEX por sus siglas en francés: *Générateur d'Exemples*). Se muestra cómo la herramienta es capaz de ir seleccionando fragmentos textuales que cumplen con las exigencias teóricas. Además, se detallará una medida novedosa para la medición de la cercanía semántica entre una definición y un ejemplo: la *densidad lexicométrica*, fruto directo de la *saturación semántica*, corazón del GENEX. Con este capítulo la metodología de trabajo intentará cerrar como un círculo al enlazar cada una de las propuestas teóricas en una aplicación de apoyo para el quehacer terminográfico.

---

<sup>5</sup> Sobre la relación de estas dos medidas, el *Coefficiente de correlación lineal de Pearson* y la *Función de coseno*, mírese el interesante artículo de Egghe y Leydesdorff : “The relation between Pearson's correlation coefficient  $r$  and Salton's cosine measure” (Egghe & Leydesdorff, 2009).

## **2. LA RELACIÓN ENTRE TÉRMINOS, DEFINICIONES Y CONCEPTOS**

Para diseñar e implementar una herramienta de Procesamiento de Lenguaje Natural no sólo basta poder visualizar o determinar un patrón lingüístico, sino que también es imprescindible entender su funcionamiento y tratar de emular lo más de cerca posible el comportamiento humano en el desenvolvimiento de una lengua determinada. Hay que entender el fenómeno que se lleva a cabo detrás de esas estructuras y establecer cuáles son sus componentes básicos. En una línea: una vez determinado el patrón, se debe desintegrar en los elementos mínimos que lo componen para poder saber cómo se arma con el fin de aprender a imitarlo computacionalmente o formalizarlo en el lenguaje técnico.

El caso de la ejemplificación es un caso particular. Como hemos visto en el capítulo anterior, aun cuando se han podido perfilar algunos requerimientos o características que debe tener un fragmento textual para poder ser considerado ejemplo, lo cierto es que en la práctica dichos preceptos funcionan a medias. Además, los fragmentos elegidos se asocian con una palabra en un esquema libre. Es decir, se ha mencionado constantemente que la palabra debe aparecer en el fragmento que se elegirá como ejemplo, pero los argumentos que se dan para establecer esta regla siguen siendo interpretaciones de lo que refleja dicho fragmento y no explicaciones de los mecanismos que conlleva para poder hacer una generalización, aun cuando los sustentos de dichas afirmaciones se basan en procedimientos medibles, como en el caso de las colocaciones. Esto es, se dice que un fragmento funciona como ejemplo porque cumple ciertos preceptos teóricos, pero no se ha llegado a demostrar que, en la práctica, se pueda seguir de manera sistemática el camino que llevó a ese fragmento a ser un ejemplo y, menos aún, el mecanismo por el cual el emisor de dicho fragmento estructuró su lexicón mental para lanzar dicho acto de habla.

En el presente capítulo hablaremos de la relación que existe entre un término, la definición asociada a él y cómo es que tanto término como definición son reflejo de ciertas características de un

concepto (Cabré, 1992). La descripción de las relaciones entre estos tres elementos pretende mostrar que es justo en los lazos que se tienden entre ellos donde se encuentra la información necesaria para explicar la conformación de un posible patrón asociado a la ejemplificación en un discurso especializado. Es decir, se intenta mostrar teóricamente que el ejemplo de un término puede ser un producto de los rasgos que su definición proyecta, ya que es indudable que cuando queremos ejemplificar una palabra pensamos inmediatamente en su significado y, cuando la palabra es polisémica, en los distintos significados que tiene o que se conozcan. En este trabajo en particular, empero, no trabajaremos con palabras del léxico general, sino con términos, lo que desembocará en que el ejemplo se asociará a una definición y esta definición a un concepto (Cabré, Feliu & Tebe, 2001, p. 6).

De igual manera veremos si los rasgos del concepto de un término son suficientes para generalizar el patrón que se pretende encontrar y cómo es que funcionaría este mecanismo para, en conjunto, dar información adicional, complementaria, a la definición terminológica.

## **2.1. La relación concepto-término**

Los conceptos son mentalmente independientes de los términos y existen antes de que éstos sean nombrados, opuestos al significado que, de acuerdo con Saussure, es inseparable de la imagen que crea el sonido (Cabré, 1999). Si se entiende que un segmento de la realidad, como un proceso de cognición, se convierte en un concepto por un medio de abstracción (Cabré, 1992, p. 87), entonces la relación establecida entre el mundo real y las similitudes entre estos objetos reflejan la relación y las similitudes establecidas entre los conceptos. La capacidad de condensación y abstracción de la mente de un ser humano le permite reducir a una cantidad manejable el mundo de rasgos y propiedades que los objetos y sus representaciones mentales tienen (Fedor de Diego, 1995, p. 132). De aquí rescatamos que la cantidad de rasgos, aunque contable, es una abstracción humana que nos permite clasificar y ordenar conceptos, pero no es siempre igual para cualquier dominio. Un

concepto, por tanto, tiene tantas realizaciones como mentes humanas puedan pensarlo. Veamos cómo han concebido el concepto otras teorías y cómo lo aborda la TCT, que es finalmente la teoría sobre la que trabajaremos en esta investigación.

En la Teoría General de la Terminología [TGT], un concepto es un elemento de pensamiento, un constructo mental que representa una clase de objetos. Los conceptos consisten en una serie de características que comparten por las clases de objetos individuales; estas características que también son conceptos nos permiten a su vez estructurar el pensamiento, así como comunicarnos. Para comunicar los conceptos y las proposiciones que los respaldan, los hablantes usan signos lingüísticos orales y escritos para crear un término, un grupo de términos u otro tipo de símbolos (Wüster, 1931). La característica más importante de la TGT es que propone como prioridad de la disciplina los conceptos, por ende, los trabajos terminológicos están orientados hacia la normalización de términos y nociones. La base de la propuesta de Wüster está en las relaciones conceptuales como unidades de análisis para llegar a las denominaciones de dichos conceptos establecidos. En la Teoría General de la Terminología “el concepto es una unidad de pensamiento” (Wüster, 1996), existe a priori, y el término es una unidad designativa de ese concepto, cuya función es normalizar.

En la Teoría Sociocognitiva de la Terminología [TST] (Temmerman, 2000), los conceptos no existen como entidades aisladas e independientes, sino que existen gracias a los textos donde los distintos autores dan testimonio de su forma de entender unas categorías dentro de una determinada estructura. A la hora de hablar de terminología, la TST rechaza la idea de concepto y la nombra como la *unidad de comprensión*, ya que considera que la comprensión del mundo es posible a través de marcos cognitivos en los que relacionan las distintas unidades de comprensión estructuradas prototípicamente, y no mediante la estructuración lógica u ontológica de estas unidades en un sistema conceptual. Estas unidades cuentan con dicha estructura prototípica y sólo muy pocas pueden definirse estrictamente como conceptos, por lo que *categoría* resulta una noción más idónea para describir a la unidad de comprensión.

El enfoque socio-cognitivo considera ideal combinar tres perspectivas en la descripción terminológica:

- a) La nominalista, donde la unidad de comprensión es el sentido del mundo.
- b) La mentalista, en la que la unidad de comprensión es una idea que existe en la mente.
- c) La realista, en la que la unidad de comprensión es una entidad extrema que existe en el universo.

Finalmente, para la Teoría Comunicativa de la Terminología [TCT], los términos son “unidades léxicas activadas singularmente por sus condiciones pragmáticas de adecuación a un tipo de comunicación” (Cabré, 1999). Esto es, un término es un elemento lingüístico que depende en gran parte del contexto donde se encuentra. Es el contexto, lo que lo rodea, lo que le da su estatus de término. Al mismo tiempo hay que pensar que, por tanto, el significado que adquiere esa palabra no es el que se usa habitualmente en lengua general o el que está más adecuado a la multiplicidad de contextos de la vida diaria; es el que pocas veces surge y que está restringido a ciertas situaciones comunicativas. Este significado particular, el concepto, está representado por el término que, a su vez, es la designación: el nombre que toma el concepto en la realidad lingüística. Si hacemos un ejercicio de inferencia, el contexto es la condensación semántica del concepto que se ve reflejada en un término y cuya utilidad restringe al campo de especialidad donde ese contexto actúa para poder ejercer efectivamente un acto comunicativo.

Cabré menciona acerca del concepto en terminología que el conocimiento, como una manifestación de la realidad, no se inclina hacia lo general ni hacia lo especializado de manera sistemática, sino que la organización de los conceptos es lo que permite hablar de uno u otro según sea la situación comunicativa. Dentro de este mismo tenor apunta que la realidad puede conceptualizarse de distintas maneras y cada una de ellas representa una realización concreta. Así, dos o más conceptos se pueden referir a una realidad u objeto y su percepción varía de caso a caso. Con esta visión, se evita decir que el vocabulario general o especializado se mueve naturalmente entre conceptos y términos. Es decir, no hay un conjunto de reglas definidas para que una palabra se convierta en término, sino que la elección de una pieza léxica dentro de un

discurso especializado atiende a criterios meramente comunicativos y puede haber siempre, como en lengua general, un fenómeno de variación.

Para entender mejor esta postura y los preceptos que tomaremos en cuenta para tratar el concepto y su relación con el término en esta investigación hay que hacer notar algunos puntos clave:

1. Los conceptos pueden tener distintas denominaciones, es decir, puede haber variación terminológica, lo que prueba que el concepto tiene varias realizaciones (términos), y tanto conceptos como términos conforman una red que los relaciona (Montero Martínez, 2003).
2. El objeto de estudio de la terminología, visto desde esta perspectiva, es el término y sus realizaciones lingüísticas (Cabré, 1999, p. 25). El término, a su vez, es una unidad denominativo-conceptual que se encuentra siempre en relación con otros conceptos. No son unidades autónomas, ni términos ni los conceptos mismos (Cabré, 1999, p. 18).
3. Los términos son unidades que tienen a la vez una forma o denominación y un significado o contenido. El contenido del término es el concepto y éste es simultáneo a la forma (Cabré, 1999, p. 20).
4. Las mencionadas relaciones entre conceptos conforman una red de significados que da forma a una estructura conceptual. Dicha estructura conceptual se refiere a un mismo ámbito especializado, de modo que el valor de un término da fe del lugar que ocupa en dicha estructuración conceptual (Cabré, 1992, p. 91).

El énfasis en las relaciones existentes entre conceptos, términos y contexto es inherente a la TCT y es por ello que partimos del presupuesto de que es ahí donde los rasgos semánticos se proyectan con mayor fuerza y pueden dar indicios de otras estructuras equivalentes o complementarias. La definición no proyecta todos los rasgos del concepto, así como tampoco es esa definición la que da fe de toda la información contenida en un contexto especializado, sino que se trata de una condensación que, a efectos prácticos, intentaremos desarticular para poder describir su funcionamiento. Para lo anterior será necesario poder discernir entre significado y concepto, saber cómo es que el contexto activa el carácter especializado de una palabra convirtiéndola así en término

y la relación que guarda dicho concepto con el término en el discurso donde es emitido. A grandes rasgos, será el contexto especializado, como ya dijimos, un fragmento que activará el mencionado término.

### 2.1.1. El concepto y el significado

Como hemos visto en los apartados anteriores, el concepto es una noción distinta a la de significado ya que no designa una realidad o un objeto, no se trata de la descripción de los rasgos de una pieza léxica, sino que describe los rasgos de una idea, o un conjunto de ellas, una representación mental y/o las relaciones que se establecen entre ellas.

El significado, aunado a lo dicho por Saussure y sucesores, es básicamente una posible **designación** de una palabra, “un concepto y una imagen acústica” (Saussure, 1945, p. 91) de un elemento propio de la lengua establecido por convención. Se trata de la descripción de los rasgos de una unidad lingüística en tanto es el elemento necesario para transmitir convenciones sobre la realidad. El proceso de significación “dice algo” acerca de una palabra y la enmarca en una serie de conjuntos, llamados paradigmas, para hacerla funcional. Se trata de un proceso llevado a cabo dentro de un sistema lingüístico (Saussure, 1945).

El concepto, por otra parte, es una unidad de comprensión del mundo, un elemento cognitivo que nos permite categorizar e identificar ideas como parte de un subconjunto dentro de los objetos pensables (Cabré, 1992, p. 90). El concepto es, si se quiere ver así, un conjunto de rasgos mínimos agrupados que pueden dar paso a la significación, a la definición, a la descripción o a la designación de una idea en un plano lingüístico -cuando lo asociamos con una palabra-<sup>6</sup>. El concepto, por tanto, es la unidad básica de trabajo sobre la cual el terminólogo tendrá que pensar y analizar en tanto unidad a **denominar** y categorizar dentro de un dominio

---

<sup>6</sup> Hablando aquí estrictamente del *concepto* en lingüística y, específicamente, en terminología.

representado en un discurso de especialidad. Se trata de un proceso llevado a cabo dentro de un sistema cognitivo.

La terminografía, concebida como la aplicación de la terminología, atribuye denominaciones a los conceptos; va del concepto (A) al término (B).

Proceso onomasiológico

$A \rightarrow B$

Esto se diferencia de la lexicografía, concebida como la aplicación de la lexicología, que parte de la designación (entrada del diccionario) y establece sus características semánticas y funcionales; va de la palabra (B) al significado (A).

Proceso semasiológico

$A \leftarrow B$

Por tanto, hay que notar aquí que el trabajo terminológico está basado en procedimientos onomasiológicos, fundamentalmente, y parte de las características del concepto para nombrarlo.

Lo anterior desemboca en que los elementos que utilizará el terminólogo para denominar los conceptos serán aquellos que rodean al término en su contexto específico: las palabras a uno y otro lado, en las relaciones que mantiene con ellas y en el tipo y naturaleza del texto o del discurso hablado donde se usa. Por tanto, la terminología desde esta perspectiva, denomina conceptos basándose en un sistema lingüístico para describir situaciones comunicativas particulares donde una palabra adquiere un estatus, un uso distinto al que podría tener en la lengua general.

Ahora bien, el alcance y las repercusiones de las palabras que rodean al término pueden ser desde unas cuantas hasta un párrafo completo o un documento entero: el *texto especializado* (Ciapuscio, 2003). El papel que juega el contexto en la caracterización de un término o en el establecimiento de su significado especializado ha sido un tema central en la TCT. Una perspectiva interesante es aquella que sienta las bases para determinar cuáles de estas palabras activan dicho significado especializado aunado a un análisis de los posibles patrones que

intervienen en la reestructuración semántica de la palabra para convertirla en término. Veamos dicha propuesta.

### 2.1.2. El concepto y el mecanismo de activación de un término

Una palabra, puesta en un contexto específico, activará ciertas características léxicas, semánticas y contextuales que la convertirán en término (Kuguel, 2007). De acuerdo con la autora, esta activación puede verse a través del análisis de los elementos que rodean una palabra en un discurso especializado y propone que las estructuras recurrentes sugieren patrones de activación que después pueden ser recuperados y relevados. Para demostrar su hipótesis, analizó un corpus conformado por 6 textos del subdominio de la limnología, en el área de ecología. Sus resultados sugieren que los patrones encontrados para la activación de un término en un contexto específico “indican una de las relaciones conceptuales básicas entre entidades: la meronimia, que refiere a las relaciones parte-todo”, así como “una relación conceptual básica, la de función”. Específicamente habla de la relación existente entre *contenedor* y *contenido* y su relación con el significado de un *indicador*, cuyos patrones pudieron ser relevados en el corpus descrito y que mostraron que “este valor terminológico surge de una interpretación intencional introducida tanto por formas que denotan indagación como por unidades que suponen la evaluación por parte de un agente humano.” (Kuguel, 2007, p. 350).

Basicamente lo que Kuguel describe es que las palabras a uno y otro costado de una pieza léxica aportan información adicional que modifica su estructura semántica y que el resultado de esta adición es la inserción de dicha pieza en un campo semántico restringido delimitado por un discurso especializado. Los costados, los contextos, de esta nueva pieza léxica, ahora convertida en término, mantienen una regularidad que opera en el plano de lo conceptual, pero que tiene cierta estabilidad en el plano sintáctico. Al caer en la cuenta de este mecanismo, se propone en dicho estudio describir los patrones sintácticos para así dar fe del cambio semántico, conceptual, del término que se ha activado. Así, se tiene

que entender que la activación del significado especializado es un mecanismo esencialmente conceptual. Esto es importante porque, como veremos adelante, si lo especializado de un término está en ese plano (conceptual), dado por un contexto, entonces los elementos derivados de dicho término, definición y ejemplo, tendrán que atender primeramente al plano conceptual y después al sintáctico.

De lo anterior hemos resumido 3 aspectos interesantes que han sido tomados en cuenta para esta investigación y que conforman los primeros hallazgos para describir el comportamiento de un concepto y su relación con el término que lo denomina:

- a) La activación de un significado especializado está dado por los elementos que rodean a una palabra en un contexto de especialidad como lo ha determinado Cabré y lo demuestra Kuguel. Por tanto, existe información vital entre los elementos en el contorno de dicha palabra, que la convierten en término, y que modifican su estructura semántica para dar paso a nuevas interpretaciones. Esto ha sido demostrado a través del modelo Léxico-Generativo de Pustejovsky (1995) utilizado por la misma Kuguel.
- b) Dicha activación del significado especializado puede ser vista a través de estructuras sintácticas donde las categorías gramaticales focalizadas (contextos verbales, adjetivales, adverbiales y preposicionales) proyectan patrones recuperables que, a su vez, pueden ser sustituidos, lo que comprueba que la estructura de activación mantiene cierta estabilidad.
- c) Las relaciones reflejadas en los patrones atienden a relaciones semánticas del tipo meronimia, hiponimia, hiperonimia, etc., y dichos patrones conforman contextos que introducen aspectos funcionales sin alterar los tipos conceptuales del término; reflejan el núcleo semántico del dominio al que pertenecen, el dominio de especialidad.

Ahora, si bien es cierto que hay ciertos contextos que activan un término, también cierto es que dichos patrones, los contextos y las

relaciones que proyectan parecen formar parte de un conjunto de características *definitorias* que sirven para determinar el sentido y el significado de esas unidades. Sin embargo, como mencionaremos, los mecanismos que promueven esta activación no operan en los contextos que esperaríamos que lo hicieran: los contextos definitorios. Mirándolo por partes:

- a) Si las palabras que rodean a un término modifican su estructura semántica, entonces determinan también la manera en que el concepto se refleja en el sistema lingüístico. Esto es, cuando el término activado tiene que ser descrito, lo que se modifica en primer lugar es su definición. El punto aquí es saber si dicho contexto de activación aporta información adicional o si sólo *releva* – para utilizar las palabras de la autora- la estructura definitiva de dicho término. Es decir ¿la activación se da en un *contexto de uso* dentro del discurso especializado o se da en un *contexto definitorio* dentro de ese mismo discurso especializado? Al parecer, por los ejemplos presentados en dicho estudio, la activación es una alteración de algunos rasgos que afectan directamente a la estructura definitiva en cuanto a la presentación del concepto en el sistema lingüístico, pero que actúan directamente en un contexto de uso<sup>7</sup>.
  
- b) La activación del término se encuentra en función de patrones que focalizan ciertas categorías gramaticales y que afectan a dicho término en un segundo nivel. Es decir, que parece ser que estas estructuras activadoras influyen

---

<sup>7</sup> “El **Fe** es el *metal* pesado más *abundante* en estas aguas” (contexto adjetivo donde ‘abundante’ modifica a ‘metal’); “La *materia orgánica medida* por el método de permanganato es de una a cuatro veces superior a la DBO” (contexto verbal donde ‘medir’ modifica a ‘materia orgánica’), etcétera. Son contextos de uso que, sin embargo, en ese preciso momento, el de la activación de lo especializado, afectan a lo que se dice sobre la palabra, sobre el término: su definición. Es decir, el uso determina el sentido y la nueva definición partirá de este nuevo contexto. El contexto de uso es anterior al contexto definitorio en la caracterización de un nuevo término. El contexto de activación es una mezcla de ambos. Esto comprueba que no es sino hasta que se necesita comunicar un rasgo conceptual específico que se hace necesario el diseño de una nueva definición que echa mano de lo dicho por el contexto de activación y que lingüísticamente se refleja en un contexto de uso.

primero en el sentido de las palabras que la conforman y luego, en consecuencia y por acumulación, al término que rodean. De hecho, si miramos los contextos nominales de este estudio, veremos que es otra palabra la que se coloca como núcleo de la estructura y no el mismo término. Esto nos lleva a pensar que, ya que dichos patrones se dan en contextos de uso, como mencionamos arriba, entonces se podría encontrar información interesante al sustituir los núcleos en los patrones encontrados: el verbo más cercano al término en el contexto verbal, el adjetivo más común que modifica al término en el contexto adjetival, etc.

- c) Finalmente, si las relaciones que se presentan son semánticas (meronimia, hiponimia, hiperonimia, etc.), entonces la activación, en primer lugar, proyecta el movimiento de los rasgos semánticos de las palabras hacia otros campos semánticos y, finalmente, elige de ellos los que estén más cercanos al dominio de especialidad. Esto, a todas luces claro, deja también abierta la posibilidad de que esa elección pueda ser modificada en función del uso del término; o sea, que la activación del significado especializado de una palabra no considere estricta y exclusivamente el criterio del especialista, sino que sea una adaptación a una situación comunicativa donde el especialista sea el emisor, el que elige activar el término y el no-especialista el receptor, el que debe recibir el concepto como una mezcla entre rasgos comunes –que conozca, que use en el lenguaje general- y unos rasgos especializados –aquellos que puede inferir o que son inherentes al término-. Este último punto apoyaría el supuesto del inciso a). A saber: que para activar en un receptor un significado especializado, el concepto debe pasar por un contexto no especializado. El producto de dicha filtración dará como resultado un contexto de activación que al mismo tiempo es un contexto de uso, pero que está modificando una palabra x en su estructura semántica para adaptarla a un discurso de especialidad.

Resumiendo, el contexto indudablemente funciona como un activador del significado especializado de una palabra y la convierte

en término por medio de una modificación funcional sin alterar el tipo conceptual. Esto es, restablece su estructura semántica en relación con las piezas léxicas que la rodean y la inserta definitivamente en un discurso especializado. Este mecanismo permite que el término activado refleje el núcleo semántico al que pertenece, que no es ya cien por ciento el que tenía fuera de ese discurso, sino que sólo conserva algunos rasgos. Sin embargo, este análisis parece indicar que, al modificar dicha estructura semántica, se afecta directamente al mecanismo de descripción del concepto de ese término: la definición terminológica. Ahora bien, la definición terminológica es producto de un discurso organizado de una manera particular donde el conocimiento de una temática es profundo y, por tanto, las estructuras en que se presentan los rasgos de los conceptos de su terminología tienen también una estructura estable. Esto es, para poder determinar el presupuesto de que es necesaria la utilización de un contexto de uso que proyecte rasgos de activación especializada, habría que analizar no sólo el término y su contexto inmediato, sino también el discurso completo -en medida de lo posible-, el texto donde se inserta o el dominio donde se activan los términos de un área. Para esto habrá que hacer algunas anotaciones teóricas sobre el discurso especializado.

### 2.1.3. El concepto en el discurso especializado

El estudio del discurso de especialidad ha tenido un desarrollo particular dentro de los estudios terminológicos. Se han tenido en cuenta factores ajenos a la lingüística en un principio, pero con el paso del tiempo se han vuelto imprescindibles uno del otro. Hasta hace unos años, acerca del lenguaje de especialidad, se versaba:

“Los lenguajes especiales son sistemas semióticos complejos semiautónomos basados y derivados del lenguaje general; su empleo presupone educación especial y está restringido a la comunicación entre especialistas en el mismo campo o en uno estrechamente relacionado” (Sager, 1980, p. 69)

Esta posición, derivada directamente de la TGT, nos muestra que el enfoque principal era transmitir conceptos entre un grupo mínimo de personas y la preocupación principal era que los

especialistas de un área dominaran alguna terminología. Hasta ese momento y todavía algunos años más adelante, no se pensaba en el usuario final del diccionario de especialidad como un lector lego al que se tuviese que instruir a la vez de transmitir dicho conocimiento especializado. A diferencia de la TCT, esas escuelas han contextualizado el estudio de la terminología y sus productos (los diccionarios especializados) sólo a especialistas de un área. Esto es, el discurso de emisión de los términos y el análisis sobre ellos recaía básicamente en el especialista del área a tratar y el especialista en lingüística. Con el enfoque comunicativo de la terminología se ha podido acceder a un tipo de trabajo multidisciplinario donde además se toman en cuenta las necesidades de los especialistas (Fathi, 2014), la multiplicidad de realizaciones de un término (Cabré, 2008), la variación de estos términos dentro de su propio discurso (Freixa, 2002), entre otras variables y, por supuesto, en las necesidades intrínsecas para la generación de un producto terminográfico (presupuesto, distribución, plataforma, recursos, etc.), como las necesidades “extrínsecas” de ese mismo producto: los potenciales lectores. Hacemos hincapié en la noción “extrínseca” ya que, hasta este momento, el diseño y la creación de diccionarios especializados dentro de la TCT cubre en buena medida este tema al tomar en cuenta el *Principio de Adecuación* (Cabré, 1999; 2000), pero no se ha ahondado de manera tal que se tenga una clasificación sobre los lectores lego que potencialmente consultarían un diccionario especializado. Es decir, se habla de un *público* receptor y se toman en cuenta sus generalidades (básicamente que son especialistas en ciernes o técnicos), pero no hay disponible un estudio profundo y detallado sobre la naturaleza de estos individuos y sus necesidades particulares<sup>8</sup>.

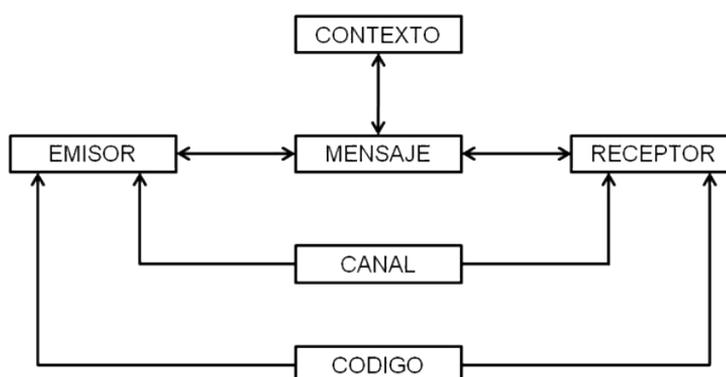
Con todo, la TCT ha funcionado de manera fructífera para el desarrollo de herramientas y productos terminológicos, ya que su núcleo se basa en las unidades especializadas dentro de un discurso específico. Lo que no sabemos a ciencia cierta es si un término, dentro de dicha teoría, podría ser activado o utilizado en un discurso no especializado. Creemos que sí y que las unidades léxicas de un área de especialidad conservan los rasgos que las hacen distintas de otras unidades no especializadas (e incluso de sí mismas -en la red

---

<sup>8</sup> Como sí lo hay para los usuarios expertos. A propósito de esta parcelación y análisis véase el artículo citado arriba, (Fathi, 2014)

de conceptos-), aun cuando son emitidas en un contexto no especializado.

Pensemos, por ejemplo, en el discurso de un médico en una consulta. Incluso cuando su posición de especialista y su registro léxico le indican que debe de referirse a los conceptos que desea explicar al paciente con unidades especializadas, existe también ese otro impulso que lo caer en la cuenta de que dicho conjunto de términos no son totalmente adecuados para el tipo de discurso comunicativo que debe emitir en ese momento. Si lo explicáramos en el contexto del esquema comunicativo básico:



**Figura 2.1. Esquema de la comunicación de Jakobson (1981, p. 360) adaptado al discurso especializado**

Nos daríamos cuenta de que esta comunicación que va del *emisor/especialista* al *receptor/no especialista* tiene más variables que el simple contexto. Hay un mensaje que se quiere transmitir y el especialista puede hacer que el canal se establezca con éxito, pero debe adaptar el código de acuerdo con el contexto para que la comunicación sea efectiva. A propósito de esto Cabré menciona que existen tres situaciones de discurso de este tipo<sup>9</sup>:

- a) Aquel que está destinado a pares especializados. Es decir, de especialista a especialista.
- b) Aquel que está destinado a la difusión de lo especializado hacia un público semiespecializado. Es decir, situaciones

---

<sup>9</sup> *ut supra*

como el discurso de la enseñanza o las presentaciones en congresos.

- c) Aquel que está destinado a la difusión de lo especializado hacia un público no inmerso en la disciplina o disciplinas sobre las que versa dicho discurso: artículos de difusión, presentaciones en medios de comunicación masiva, etc.

Por tanto, el discurso domina el modo en que el canal debe ser entendido y modifica el código con el cual el especialista se ha de referir al campo de especialidad. No puede omitir el efecto que el contexto de comunicación está creando y debe de modificar las dos variables anteriores para que el mensaje pueda ser transmitido.

Hay que tomar en cuenta, también, que la mayoría del trabajo que se lleva a cabo para diseñar un diccionario especializado en estos días recae con mucha fuerza en el corpus de análisis. Este corpus está formado, en su mayoría, por textos especializados que han sido minuciosamente elegidos por los especialistas que trabajarán en él. Por un lado, en la cuestión formal y organizativa, el corpus de trabajo es preparado por el lingüista atendiendo a una metodología determinada, que en este caso es el trabajo de la terminografía; y por otro lado, en la parte del contenido, el corpus es tratado como un recipiente de información textual validada por especialistas del área.

Esta organización, sin embargo, no carece de fallas y no es exclusiva en cuanto al trabajo de sus agentes. Es decir, tanto el terminógrafo influye en la elección del material que conformará el corpus, como el especialista del área en la forma que el corpus debe tener en cuanto a distribución, tamaño, organización, etc. Esto en ocasiones encuentra un punto de conflicto que ha sido tratado como la ideología en el diccionario especializado. A propósito de este tema, una referencia imprescindible en los últimos años es el libro preparado por Félix San Vicente, Cecilio Garriga y Hugo Lombardini (San Vicente, Garriga & Lombardini, 2012). Esto es importante debido a que el modo y la manera en que el discurso especializado sea emitido ejercerá una gran influencia en la conformación y selección de unidades (términos y conceptos), relaciones (conceptuales, temáticas, semánticas) y estructuras

(contextos definitorios, de uso, colocaciones, ejemplos, etc.). El corpus es un reflejo del discurso especializado.

Como vemos, el proceso por el cual un contexto activa esas características de unidad léxica especializada y el proceso por el cual los términos son transmitidos, son complementarios e indispensables uno para el otro. Así, al momento de establecer comunicación y mencionar los rasgos pertinentes del término en función del tipo de receptor, ha de tenerse en cuenta dicho término no sólo dentro de su campo definitorio y lo que esto implica (contexto definitorio y contexto de activación), sino también su contexto de uso en la lengua general. Es decir, hay que llevar a cabo un proceso por el cual se relacione, en medida de lo posible, los rasgos semánticos de una unidad léxica especializada con una estructura o frase que, sin ser altamente especializada, es decir, que contenga otros términos, conserve la mayoría de rasgos del concepto que se desea transmitir. Esto debería coadyuvar a que un receptor, que carece de los recursos discursivos que le permitan entender al especialista dentro del campo de especialidad, pueda tener una aprehensión de los conceptos que se están tratando transmitir y que le permitan enlazar el discurso especializado con el propio discurso lego para tener una comunicación efectiva.

Para poder ver más de cerca esta problemática pensemos en el primer elemento que crea una incomunicabilidad especializada entre emisor-especialista y receptor-no especialista: la definición terminológica.

## **2.2. La relación término-definición**

El proceso de denominación de un término tiene su raíz en la relación que establece el concepto con el término y que dicha relación se ve afectada por el contexto de activación de tal término, en un nivel micro, como menciona Kuguel (2007) y las características inherentes al discurso especializado en que ese término es utilizado, en un nivel macro, en el texto en que se inserta (Ciapuscio, 2003).

Pues bien, la relación término-definición se puede notar claramente en el contexto en que se define, esto es, en el contexto donde el concepto se hace presente a través de fórmulas definitorias que lo caracterizan y que lo posicionan en la red de conceptos a la que pertenece (como ya lo menciona Cabré). A veces este contexto coincide con el contexto de activación que propone Kuguel y otras veces se trata de entramados más complejos donde no necesariamente se puede ver la proyección de esta definición del concepto que finalmente se va a fijar en el término. Veamos un ejemplo<sup>10</sup>:

*El As es un elemento altamente **nocivo** para la salud. (Texto 3)*

Si atendemos al criterio de que el verbo copulativo por antonomasia en el español (ser) introduce elementos que delimitan o definen una palabra, caeríamos en la cuenta de que la frase anterior podría ser una potencial definición de *As* (arsénico). Sin embargo, lo cierto es que dicha frase no aporta información acerca del concepto al que se refiere, la relación con algún otro concepto o los hipónimos o hiperónimos que lo rodean (por nombrar algunas características definitorias). Aun así, se trata de un contexto donde se activan las características terminológicas de la unidad focalizada (*As*) en un contexto adjetival ya que “Un nombre denota ya sea una entidad indicadora o un evento indicador cuando es modificado por un adjetivo calificativo ligado con la cuantificación” (Kuguel, 2007).

Lo anterior muestra que aún cuando el contexto de activación da fe del uso de un término en un discurso especializado, dicho contexto de activación aporta información adicional sobre ese término sin mostrar necesariamente características definitorias, aun cuando la estructura sintáctica parece promover lo contrario. Se acerca más bien a un contexto de uso.

En este punto es imprescindible detenerse para delimitar lo que hemos considerado un contexto de uso basado en corpus con orientación terminológica. Un contexto de uso constituye una muestra real de lengua. Los fragmentos que se analizan aquí como contextos de uso son parte de un corpus que ha recogido muestras

---

<sup>10</sup> Muestra tomada de Kuguel (2007)

de habla y escritura de una determinada lengua. Son trazas de un discurso, de una conversación transcrita o epistolar, de un tratado académico, de una información comercial, de un estudio científico, etc. En nuestro caso se tratará de contextos de uso especializados, si parten de un corpus de especialidad (el de sexualidad, por ejemplo), y contextos de uso (a secas) cuando se han recuperado de un corpus general (*Jornada, Wiki y esTenTen*). El punto más fuerte de ellos es que ya han sido resguardados en un formato que permite recuperarlos en cualquier momento y además analizarlos desde diferentes puntos de vista (de acuerdo con el estudio a realizar), sin necesidad de contar con el emisor en todo momento. Sobre este mismo tema De Bessé menciona que el contexto de uso (el que hemos denominado “especializado”) es el punto de inicio de cualquier trabajo terminográfico y tiene dos funciones: aclarar el significado de un término e ilustrar su funcionamiento (De Bessé, 1991).

Para poder mirar a detalle la diferencia que un contexto definatorio guarda con respecto a un contexto de uso, es necesario delimitar la noción de contexto definatorio y sus características.

### 2.2.1. Los Contextos Definitorios en terminología

Investigaciones recientes han mostrado, con base en análisis de extracción de información y corpus lingüísticos, que efectivamente se puede extraer información conceptual vía contextos definatorios (Sierra, Medina, Alarcón & Aguilar, 2003) y que las definiciones de una palabra pueden ser divididas de acuerdo con su naturaleza – atendiendo a la división aristotélica- en sinonímicas como las que menciona Seco, extensionales (que define por las partes que constituyen al término), funcionales (para qué sirve el término) y analíticas (que proveen información característica sobre un término y presentan el género próximo al cual pertenecen). Además, se ha encontrado que dentro de los contextos definatorios las definiciones mantienen patrones distintivos que se asocian a distintos verbos (Aguilar, 2009). La aportación en cuanto al análisis de dichas definiciones es que son vistas a través de la mirada de los contextos definatorios y no sólo de la definición en sí.

Un Contexto Definitorio es aquel fragmento textual en un documento especializado donde se define un término (Sierra, Aguilar, Alarcón & Bach, 2008). Las partes constitutivas de este contexto son: Término (T), Definición (D) y Patrón Definitorio (PD), que es el elemento encargado de mostrar la relación entre un término y su definición. De manera satelital se encuentran también los Patrones Pragmáticos (PP), que son elementos que delimitan las condiciones de uso de un término. Un esquema general del contexto definitorio es el siguiente:

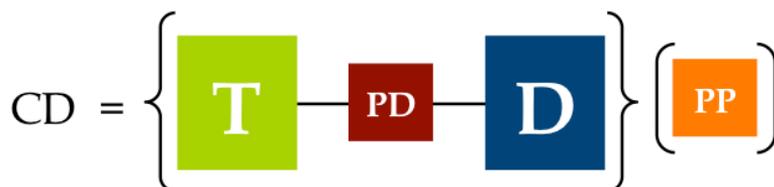


Figura 2.2. Esquema general de un Contexto Definitorio [Tomado de (Alarcón, 2006)]

Para poder establecer la relación entre término y definición, los contextos definitorios funcionan, bajo esta perspectiva, a través de recurrencias sintácticas determinadas por los distintos patrones que pueden proyectar una definición y estos, a su vez, dependen en gran parte de los verbos que dichos patrones tengan como núcleo. En la siguiente imagen podemos observar un condensado de los distintos patrones que se pueden encontrar en un contexto definitorio:

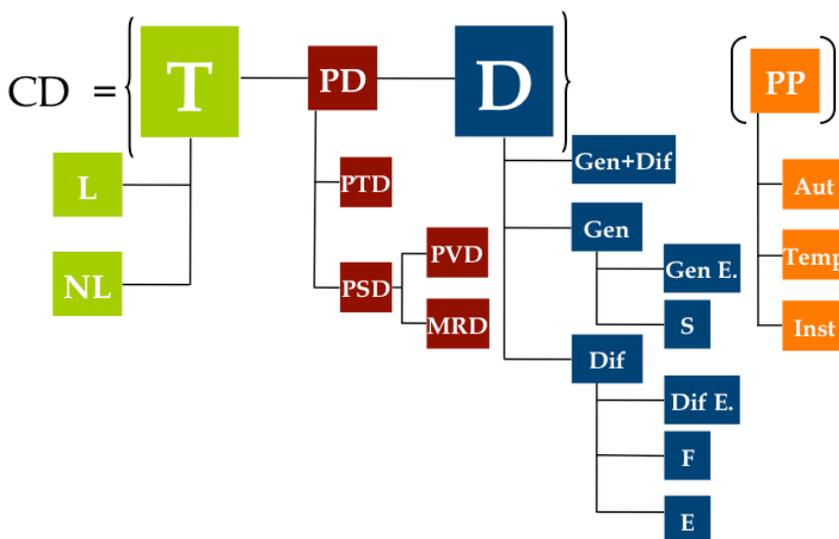


Figura 2.3. Esquema detallado de un Contexto Definitorio [Tomado de (Alarcón, 2006)]

Donde:

- El Término (T) puede ser léxico (L) o no léxico (NL). Los términos léxicos son aquellos constituidos por una palabra con un uso especializado que Alarcón (2009) ha dado en tomar como equivalente de una USE (Unidad de Significación Especializada), siguiendo a Estopà (2001) y que considera las frases nominales, adjetivales, verbales, etcétera, o sus núcleos: nombres, adjetivos, verbos, respectivamente. Los términos no léxicos (NL) son aquellos que constituyen fórmulas, símbolos, operadores, etcétera.

- El Patrón Definitorio (PD), puede ser del tipo tipográfico definitorio (PTD) cuando el término al que se refiere tiene una marcación implícita en el documento (estar en versalitas o en negritas, por ejemplo); o un sintáctico definitorio (PSD), cuando término y definición son unidos por estructuras gramaticales. Dentro de este segundo tipo de patrones encontramos los Patrones Verbales Definitorios (PVD) cuyo núcleo es un verbo definitorio, y los que son introducidos por Marcadores Reformulatorios Definitorios (MRD) que, de acuerdo con los autores (Sierra et al., 2008), se trata de estructuras sintácticas asociadas a un proceso metalingüístico que sirven para explicar lo dicho sobre el término dentro de los propios contextos definitorios. A propósito de los patrones verbales definitorios, se ha propuesto una lista de los verbos que funcionan como núcleos de dichas estructuras y que se han descrito detallada en una investigación que parte de esta propuesta teórica. A continuación reproducimos dicha lista (Aguilar, 2009):

<b>Género + Diferencia:</b>	<b>Funcional:</b>	<b>Sinonimia:</b>
Predicación secundaria <i>Caracterizar</i> <i>Comprender</i> <i>Concebir</i> <i>Conocer</i> <i>Considerar</i> <i>Definir</i> <i>Denominar</i> <i>Describir</i> <i>Entender</i> <i>Identificar</i> <i>Visualizar</i>	Predicación simple <i>Emplear (se) en</i> <i>Permitir</i> <i>Servir (en/para)</i> <i>Usar (en/para)</i> <i>Utilizar (en/para)</i>	Predicación simple <i>Equivaler</i> <i>Llamar (también)</i> <i>Nombrar</i> <i>Ser_igual</i> <i>Ser_similar</i>
Predicación simple <i>Emplear (se) en</i> <i>Permitir</i> <i>Servir (en/para)</i> <i>Usar (en/para)</i> <i>Utilizar (en/para)</i>	<b>Meronomimica/por Extensión:</b>  Predicación simple <i>Comprender</i> <i>Consistir (de)</i> <i>Constar (de/con)</i> <i>Constituir</i> <i>Contar (de/con)</i>	<b>Genus exclusivo:</b>  Predicación simple <i>Ser</i>

**Figura 2.4. Lista de verbos definitorios**

- La Definición (D), que, de acuerdo con el modelo aristotélico, puede ser del tipo *genus + diferencia* (Gen+Dif), sólo *genus* (Genus), o sólo *diferencia* (Dif)<sup>11</sup>.

- El Patrón Pragmático (PP), es un fragmento textual limítrofe al contexto definitorio y su función es principalmente instruccional (Inst), de autoría (aut) o de tiempo (Temp), entre otros. Cumplen el papel de indicadores de alguna cualidad delimitatoria en cuanto a fechas, autores, matices o condiciones de uso tanto del término como del mismo patrón definitorio. Son relevantes en algunos casos debido a su carácter actualizador, entendiendo por esto la naturaleza específica que le dan al término en cuanto a posicionamientos dentro de esferas más allá de la conceptual. Es decir, especifican aún más su uso y sus relaciones con la red de conceptos a la que dicho concepto del término se refiere.

Como podemos ver, los contextos definitorios funcionan a un nivel gramatical y están basados en estructuras sintácticas semifijas que aportan información imprescindible para el entendimiento del concepto de un término. Al mismo tiempo, proveen la información necesaria para el diseño y creación de una definición adecuada y demarcan los límites que ese concepto tiene con respecto a otros pares del mismo ámbito. De igual manera, hacen palpable la red de sentidos que comporta el concepto y lo ubica de manera idónea dentro del discurso especializado.

Una vez delimitado el contexto que define un término, la tarea de creación de definiciones se vuelve un trabajo más fluido, aunque la definición en sí es un tema aparte que ha de abordarse de acuerdo con el campo de especialidad y la finalidad del material donde será inserta. Para efectos de este trabajo, como hemos mencionado, tomaremos en cuenta la noción de definición terminológica dentro de la TCT.

---

<sup>11</sup> Para Meyer (2001), este último tipo de contextos son llamados *KRC* (Knowledge Rich Contexts) *explicativos*.

### 2.2.2. La definición terminológica

La definición terminológica, en la TCT, parte de una distinción entre tres tipos de definiciones. A saber: la definición lingüística, que es aquella que generalmente se usa en la labor lexicográfica y se refiere al significado del signo lingüístico, como hemos visto anteriormente; la definición ontológica que se refiere a la realidad inmediata y designa significados compartidos por una comunidad de hablantes, incluso cuando dichos significados contengan rasgos totalitarios, que se refieran a todo el conjunto de las características de un concepto; y la definición terminológica, que se refiere al concepto y a las relaciones que estos forman para referirse exclusivamente a un campo de especialidad, es decir, focalizan sólo algunos de sus rasgos (Cabré, 1992, p. 174).

Para nosotros, la definición terminológica debe tener las características formales que hemos descrito a propósito de los contextos definatorios, esto es, un patrón asociado a dicha definición que esté en relación con la lista de verbos definatorios antes mostrada; y que comporte los preceptos teóricos que marca la TCT para la correcta designación de un concepto a través de un término, es decir, que sea la designación de ese término y que se enmarque en un contexto de especialidad.

Del mismo modo, hay que entender la definición terminológica como la condensación del concepto dentro de una estructura sintáctica. Aunque, como vimos, hay varios actos lingüísticos que pueden activar el significado especializado de un término, es el contexto definatorio el que organiza dicha activación y el que la provee de sistematicidad. Esta característica, si lo miramos detalladamente, no es arbitraria. El discurso especializado tiene una organización definida que afecta directamente, a su vez, la organización de los conceptos que abarca (la red de conceptos de la que habla Cabré) (Ciapuscio, 2003) y la organización de las unidades que los designan (su terminología). Así, no es raro pensar que aunque la activación, como ha mostrado Kuguel, se da en contextos de uso dentro de discursos especializados, al momento de intentar definir estos conceptos, ambos especialistas, el de la materia y el terminólogo, traten de echar mano de la estructura

conceptual de la materia que abordan. Así, el camino más obvio será el de una estructura sintáctica sistemática y casi fija que desembocará en uno de esos contextos, que a su vez contiene los rasgos definitorios del concepto: el contexto definitorio. Finalmente, dicho contexto sienta las bases para obtener esa estructura deseada: la definición terminológica.

Se ha de hacer hincapié, aquí, que esta definición terminológica puede o no estar en un diccionario especializado siempre y cuando cumpla con estos requisitos. Por lo anterior, hemos decidido llevar a cabo el análisis y el procesamiento de las piezas léxicas elegidas y sus definiciones atendiendo al contenido del Diccionario del Español de México (Lara, 2010). En este diccionario, los argumentos antes mencionados se cumplen de manera satisfactoria, esto es, aun cuando los términos se encuentran en un diccionario de habla general, los términos y sus definiciones han sido tratados de manera independiente de acuerdo con las metodologías actuales para el tratamiento del lenguaje de especialidad y se ha consultado a terceros para la validación de las definiciones. En palabra de sus autores:

“Hemos puesto especial atención a las terminologías técnicas usuales en México. Éstas, como toda convención con que se unifican los textos de un mismo campo para hacer eficaz y precisa su comunicación, están sujetas tanto a aportes de diferentes tendencias científicas y pedagógicas como a cambios producidos por la influencia del tiempo y de las situaciones sociales en que se crean. En el caso de las terminologías de las ciencias naturales, se han tomado en cuenta los consejos de los especialistas que asesoran al cuerpo de redacción del *Diccionario* y que representan a las instituciones más autorizadas de la república. Para la definición de cada uno de esos términos se han consultado, también, varias obras contemporáneas en otras lenguas, con el objeto de ofrecer las informaciones más seguras y modernas posibles.” (Lara, 2010)

Esta elección tiene que ver, también, con el alcance posible que tiene una definición terminológica dentro del diccionario donde se inserta. Esto lo podemos ver si nos detenemos a analizar la relación de dicha definición con el concepto al que se refiere y el término elegido para designarlo.

### 2.3. La relación concepto-término-definición

Normalmente, el conjunto de términos que conforman un lecionario o una lista de entradas es extraído del mismo corpus donde se extraen o identifican las definiciones. Corresponde a criterios selectivos de corte *exclusivo*. Esto es, aunque la definición no sea extraída literalmente del corpus, sí atiende a los criterios que el mismo especialista que ha estado inmiscuido en el diseño del corpus sea uno de los consultores para la validación de las definiciones y, por tanto, de la red conceptual de la terminología (del conjunto de términos). El carácter *exclusivo*, de esta manera, recae en que terminólogo, especialista y corpus influyen directamente en la forma y el contenido de la definición.

Por otro lado, cuando la definición terminológica está inserta en un diccionario general, el concepto al que designa por medio del término se encuentra en relación con una red más amplia de conceptos que permiten una mayor riqueza léxica y semántica. El terminólogo y el especialista validan su forma y su contenido, por decirlo así, de una manera parcial, pues no tienen acceso al entramado completo que el concepto en cuestión tiene con los demás conceptos del corpus; sólo validan lo que está bajo su control dentro del área de especialidad que tratan en ese momento. Así, el mismo procedimiento para diseñar una definición terminológica tendrá más o menos alcance de acuerdo con el tipo de corpus desde donde sea extraída.

Lo anterior nos da fe de la relación establecida entre concepto, término y definición. A lo largo de este capítulo hemos visto que este trinomio tiene a su vez características de sus elementos por separado, y características de esos mismos elementos cuando se relacionan unos con otros.

Se puede ver, por consiguiente, que en la relación existente entre concepto y término, el proceso de denominación, meramente terminológico, es un fenómeno muy distinto a su par en lexicografía, el proceso de designación. Atendiendo a esta perspectiva, la denominación es el mecanismo por el cual nosotros

como seres humanos y particularmente los especialistas de un área, en tanto tales, asignan una representación lingüística a un concepto, a un significado delimitado por sus condiciones pragmáticas de uso en un discurso especializado. Esta denominación del concepto está a su vez comprometida con el discurso de especialidad donde es emitida. Es decir, la denominación del término depende también de los otros términos en el discurso que a su vez denominarán conceptos de la red que delimita al concepto que se está denominando. El mecanismo por el cual la denominación se da con mayor efectividad, es el mecanismo de activación del término, que no es otra cosa que el descubrimiento y análisis de aquellos contextos de uso donde los rasgos del concepto referido emergen y dirigen la vista hacia la unidad que mejor los denomina, el término. Esta primera relación nos permite, por tanto, validar al término como el elemento de trabajo nuclear en la labor terminográfica.

En la segunda relación mencionada, la que corresponde al binomio término-definición, hemos visto que hay que tomar en cuenta cuatro aspectos importantes: El contexto de activación, el contexto definitorio, la definición y el contexto lingüístico. Definir la definición terminológica es ya un trabajo complejo que teorías de distinta naturaleza han abordado, de las cuales hemos mencionado aquí sólo tres y de ellas nos hemos quedado con la que corresponde a la Teoría Comunicativa de la Terminología. Así, para efectos de este trabajo y la definición terminológica es una estructura lingüística que refleja las características de un concepto a través de la utilización de un término, esto es, una unidad léxica derivada de un campo de especialidad.

Nos dimos a la tarea de rastrear qué punto o ubicación de ese discurso especializado es de donde se desprenden término y definición. Para la ubicación de los términos hay ya numerosas investigaciones, pero lo que nos ocupaba aquí era poder distinguir la relación término definición, por lo que necesitábamos ubicar el instante inmediato anterior a la creación de la definición. Este momento es, pues, el contexto definitorio<sup>12</sup>. Una estructura que

---

<sup>12</sup> Otras investigaciones, empero, han incluido los contextos definitorios en un grupo denominado “Contextos Ricos en Conocimiento” (KRC por sus siglas en inglés), ya que no se limitan sólo a aquellos que contienen información definitoria sino que además aportan información pragmática o metalingüística, como los

focaliza los rasgos más importantes del concepto y que a su vez lo hace a través de estructura semiestables llamados patrones definitorios. Si bien una relación conceptual es difícil de analizar, un patrón representa una regularidad que coadyuva a explicar un fenómeno y a lanzar generalizaciones comprobables, es decir, es más sencillo de analizar. El resultado de dicho análisis es una identificación de las relaciones que establece un término con cada uno de los elementos de la estructura que lo define, es decir, con la definición terminológica.

A manera de conclusión podríamos decir que la relación concepto-término-definición es una codependencia semántica lineal que, dependiendo de la utilización o el análisis de alguno de sus tres elementos, el terminógrafo podrá apoyarse para diseñar las diversas categorías de datos de una entrada en un diccionario de especialidad. Si se focaliza el concepto, se podrá mirar el conjunto de rasgos asociados a él y las relaciones que dicho concepto tiene con otros conceptos de la misma área de especialidad; esto dará como resultado la distinción más minuciosa del término o los términos que se asocien a dicho concepto. Si se focaliza el término, se podrá observar la representación lingüística de un concepto y se podrán definir las características morfológicas y semánticas que lo hacen distinto de otras piezas léxicas utilizadas por los hablantes de una lengua. Aquí se podrá ver, además, la diferencia que puede guardar con alguna palabra similar utilizada en el lenguaje general si se coloca en algún contexto de activación a través de la relación que guarde con otros términos o palabras en dicho contexto ya que debemos recordar que el contexto de activación guarda una distancia muy cercana a un contexto de uso. Si se focaliza la definición, se podrá observar la estructura sintáctica más adecuada para representar las características del término, esto es, el conjunto de rasgos conceptuales que se han elegido como imprescindibles para poder definir el alcance semántico de dicho término.

Finalmente, vemos que la relación concepto-término da fe de las elecciones tomadas por los especialistas para representar por medio de palabra o término los rasgos inherentes de un concepto

---

patrones que menciona Alarcón, y relacional o atributiva, como también contempla Meyer (2001)<sup>12</sup>, pero a este tema volveremos en el apartado 2.4.

con los matices que el propio discurso de especialidad, y la relación término-definición da fe de la representación lingüística del término a través de un acto comunicativo (definitorio) que condensa el conjunto de los rasgos elegidos en una estructura lingüística organizada: la definición terminográfica<sup>13</sup>.

El conjunto de características esenciales y complementarias presentes en la definición terminográfica es un punto que crea controversia constante, ya que por un lado el especialista verá como obligatorias aquellas características reflejadas en otros términos dentro de la corriente teórica en la que se formó, o la estructuración previa de su formación, mientras que el terminólogo tendrá en cuenta, además, relaciones semánticas (hiperónimos, hipónimos, cohipónimos, etc.), relaciones morfosintácticas y léxicas (colocaciones, concordancias, contextos) y relaciones conceptuales (árbol de campo, mapa conceptual, campos semánticos). Estas variables harán que la definición se incline hacia uno u otro campo semántico y, en consecuencia, declare como obligatorios ciertos rasgos del concepto para conformar los descriptores genéricos, dejando de lado las demás características como opcionales. En cualquier caso, los argumentos imperantes siempre parecen ser los de estos especialistas y no estos mismos argumentos en relación con el potencial usuario.

---

<sup>13</sup> De aquí en adelante nótese que nos referiremos a la definición en el diccionario especializado como *definición terminográfica*, siempre atendiendo al criterio de De Bessé (1997, p. 65), quien hace una distinción entre definiciones terminológicas y definiciones terminográficas. Para él, las primeras crean conceptos mientras que las segundas sólo los describen en referencia a un sistema conceptual preexistente. Sobre esta noción, que compartimos, Pérez apunta: “Esta distinción entre definiciones terminológicas y definiciones terminográficas está en consonancia con la diferente orientación que normalmente poseen ambas actividades: normativa en el primer caso y descriptiva en el segundo. En consecuencia, los encargados de formular las definiciones terminológicas son los propios expertos en la materia y los organismos normalizadores correspondientes, mientras que las definiciones terminográficas se encuentran en los trabajos de orientación descriptiva realizados por los terminógrafos.” (Pérez, 2002).

## 2.4. Discusión: Las limitaciones de la definición terminográfica

Algunos estudios han demostrado que la información encontrada alrededor del término en una definición puede aportar información valiosa. Seco se refiere a esta información como “contorno” cuando habla sobre definiciones de verbos. Para él, la definición tiene dos partes esenciales: la palabra como signo (categoría gramatical, contextos de uso, etc.) y la explicación de su contenido (la definición propiamente dicha) (Seco, 1979). Por tanto, si la palabra como signo es de alguna manera “mutilada” al eliminar sus elementos constitutivos, sólo nos quedará la parte del contenido, que si bien es importante y la mejor desarrollada por la lexicografía de acuerdo con Seco por medio de la sinonimia<sup>14</sup>, no es completamente abarcadora y por ello la definición se encuentra incompleta.

Maldonado menciona, después de hacer un análisis sobre las posibilidades del principio de sustitución que postula Seco a propósito de la sinonimia en definiciones, que muchas palabras no pueden cumplirlo y que estamos ante un escenario donde caemos en la cuenta de que algunas palabras “no se definen (no se describe su significado) sino que se explican (se describe qué son, cómo se usan y para qué se emplean), son palabras que se explican por su contexto de uso en vez de por su *contenido* (significado)”. (Maldonado, 1997).

En García de Quesada es cuando ya nos sumergimos en terminología y encontramos un interesante análisis acerca de la sinergia específica entre término y concepto: la definición terminográfica. En dicha investigación, los conceptos “no son más que una formalización posible de una parte del conocimiento. Estos conceptos serán punto de referencia para el término”, y describe al término como la “formalización de uno o varios conceptos, para su uso en la comunicación experta” (García de Quesada, 2001). De tal manera que podemos entender que, el concepto como una parte del

---

<sup>14</sup> Es decir, que la parte del contenido, la definición, puede sustituir al término en una oración pero eso no significa que la definición sea la correcta o mejor. Ver Seco (1979, p. 186).

conocimiento, y el término como la formalización de ese concepto, dan como resultado la simplificación y reducción de un significado en un dominio específico. Pues bien, la definición asociada a un término ha de dar cuenta de este proceso de simplificación y lo ha de desglosar en un ejercicio de concentración semántica que Lara ha definido como ecuación sémica (Lara, 1997).

Un análisis entre ambas posiciones nos puede explicar que este es un proceso mental, reflejado en estructuras lingüísticas que a su vez da paso a las definiciones terminográficas. Sin embargo, en este mismo sentido, hay que mencionar que de acuerdo con otros autores (Galinski & Picht, 1997), y como la misma Quesada muestra, la definición terminográfica no cubre todas las posibles realizaciones de un concepto debido a la problemática que sugiere definir el concepto *concepto*.

Estas limitaciones tienen diversos efectos y explicaciones. En primer lugar hay que tener en cuenta que el redactor de la definición terminográfica no es sólo uno. Se trata de un trabajo conjunto que parte del especialista y el terminólogo pule echando mano de sus conocimientos lingüísticos, o que es una creación del terminólogo que se valida a través de la revisión del contenido por parte del especialista. Por tanto, el diálogo unívoco entre expertos impide delimitar el concepto de tal manera que pueda ser aprehendido por el grueso de la población. Así, la representación que se haga de dicho concepto, proyectado en uno u otro término, atenderá a criterios estrictamente especializados y no en la adecuación de su definición a otros registros léxicos. Basado en lo anterior, si el grado de formación de los usuarios es mínimo, la decodificación de la definición será mínima también y, por ende, la comprensión del término a la que está asociada no se da eficazmente.

En segundo lugar, la parcelación del discurso especializado en el que se basa el diccionario hará que las definiciones contengan piezas léxicas o relaciones semánticas específicas. Es decir, que la redacción de las definiciones depende de cómo el terminólogo divida el campo semántico que trabaja. Dicha división, generalmente, está dada previamente por un conjunto de criterios con base en comentarios hechos por los especialistas, por lo cual los términos y sus respectivas definiciones ya tienden a una estructuración conceptual *ad hoc* que en muchas ocasiones puede

chocar con las estructuras que la sociedad tiene sobre ese mismo campo de especialidad, lo que desemboca, nuevamente, en una aprehensión fragmentada del concepto en cuestión.

En este sentido, podríamos hablar de cuatro restricciones esenciales de la definición terminográfica:

1. No cumple con el *Principio de adecuación*, esto es, debido a que es una condensación de algunos rasgos del concepto, los que el experto dicta, no siempre es entendida por otros lectores no expertos. Esto debería ser tomado en cuenta de manera sistemática si pensamos que es más probable que un especialista en ciernes consulte este tipo de material para conocer las relaciones conceptuales y los términos de su campo de especialidad, que aquel especialista formado que probablemente lo consulte para reafirmar conocimiento ya adquirido o para refinar definiciones que ya conoce y tiene en mente. No hay adecuación sino adaptación de formato. Nos dice Lorente:

“Esta adaptación no parece que esté realmente orientada hacia la adecuación de los productos terminográficos a los perfiles y las necesidades reales y contrastadas de sus usuarios, ya que es una práctica habitual que algunos autores manifiesten, a través de sus textos introductorios, que sus vocabularios están dirigidos a diversos usuarios a la vez [...] Sin embargo, desde el punto de vista de la metodología de aplicaciones, estos cambios no han supuesto realmente una alteración de los principales postulados teóricos que sostienen el modelo terminográfico clásico. Sólo han supuesto tímidos intentos de adaptación de la metodología imperante a las circunstancias nuevas, más cercana a una ampliación de contenidos que a una verdadera innovación.” (Lorente, 2001).

2. El discurso especializado impacta muy profundamente en la constitución de la definición terminográfica. La manera en que el discurso haya sido emitido propiciará que se enaltezcan o apaguen algunos rasgos conceptuales en el conjunto de términos que formarán el diccionario de esa área. Es decir, tanto corpus como experto pueden influir en

la parcelación del campo de especialidad dependiendo de sus intereses, ideologías o escuelas<sup>15</sup>.

3. La definición terminográfica no siempre da fe del proceso de activación del término al que se refiere. Como vimos anteriormente, el contexto de activación de un término es mucho más cercano a un contexto de uso que el propio contexto definitorio. El primero muestra la manera en que los hablantes han interiorizado un aspecto de la realidad y cómo han colocado un concepto dentro del conjunto de sus apreciaciones de la realidad y lo han enlazado por medio de relaciones conceptuales que finalmente representan lingüísticamente en un término; el segundo es un contexto que pertenece a la realidad lingüística de esa comunidad de hablantes y está formado por esas unidades léxicas que se han convenido en utilizar como representantes de los conceptos. Por ello, la suma de las unidades de la definición no es sino la suma de las condensaciones de los conceptos que son representados con las palabras que la forman. A esto hay que aunar el hecho de que la estructura misma de la definición está limitada por el conjunto de constituyentes que la pueden formar. De ahí que una definición terminográfica no pueda describir, lo que conlleva, entonces, a que esos rasgos del concepto que se explican, se describen o se ejemplifican, no puedan ser aprehendidos por una definición. Lo que nos lleva al siguiente punto:
4. Dado que el concepto es una entidad mental difícilmente formalizable, es casi imposible poder mostrar todos sus rasgos en un acto lingüístico mínimo, en una sola predicación o como una sentencia simple. Un concepto es una entidad de varias caras, de diferentes realizaciones que no siempre tienen interpretación en la realidad, una

---

<sup>15</sup> En casos extremos encontramos casos como el siguiente acerca del término *sexualidad*: “Denominación propia de la reproducción humana; hace referencia a que los fenómenos biológicos que generan un nuevo hombre son insuficientes para explicar su espiritualidad (inteligencia y libertad), que proceden de un acto directo de creación por Dios. Esto es sostenido por la tradición judeocristiana, por la filosofía pagana clásica y por las religiones de muchos pueblos primitivos, y es negado por los científicos en su optimismo. La explicación meramente biológica de la procreación deja fuera aspectos psicológicos y espirituales más profundos de la sexualidad humana.” (Diccionario Espasa de Medicina, 1999)

interpretación. El concepto es tanto conjunto de rasgos semánticos como conjunto de relaciones con otros conceptos. A este problema Cabré lo ha llamado *Principio de poliedricidad* (Cabré, 2008) y básicamente dicta, como ya lo podemos perfilar, que el concepto de un término no se puede representar en su totalidad en un acto comunicativo. La definición terminológica, como un acto restringido y condensatorio, no escapa, pues, de esta limitación: es imposible que pueda representar todos los rasgos de un concepto de una sola vez.

Y sin embargo se ha intentado. Un estudio realizado en 2011 (Lázaro, 2011) ha mostrado que hay una fuerte tendencia a tratar de cubrir todos los aspectos de un concepto a través de la definición terminológica. El modo de operación no es otro que la adición de definiciones que vayan cubriendo parcialmente los aspectos más importantes que el especialista/terminólogo considera a propósito del concepto al que se desea referir. Sin embargo, como ya lo apuntamos, si la limitación de la definición terminológica es en cuanto a aspectos semánticos (los verbos que la componen y la forma que esta adopta al agotar los espacios funcionales de dichos verbos), no tanto a aspectos estructurales, la adición no será una mecanismo que pueda llegar a cubrir, en ningún caso, los aspectos que son más efectivos en ocasiones para la transmisión del conocimiento; esto es, si la definición no puede describir ni ejemplificar, sería inocente pensar que con cinco o diez definiciones especializadas estaremos cumpliendo con el papel que cumplen otras categorías de datos en una sola línea.

Este último punto es justo el que se desea enfocar en esta discusión: si la definición terminográfica es incapaz de llevar a cabo mecanismos para facilitar la transmisión del conocimiento especializado por la misma naturaleza de su forma y contenido, entonces ¿cómo transmitir esos rasgos imprescindibles del poliedro que supone un concepto para poder entenderlo con mayor facilidad y en una dimensión aprehensiva que no sea exclusiva del especialista? A propósito de esto Lorente menciona:

“La poliedricidad es una noción que supera la adición de facetas conceptuales, ya que con la simple recursividad secuencial, propia del lenguaje, se perdería la estructura. Por

otro lado, cada hablante desde su perspectiva sólo puede observar o tratar algunas de las caras del poliedro (aproximación parcial).

En este sentido, para superar las limitaciones del texto, para describir un concepto, se deberían construir definiciones múltiples, parciales y complementarias, atendiendo a la parcialidad de perspectivas. Para superar la falta de estructura poliédrica que tendría un conjunto de definiciones, convendría algún mecanismo que las interrelacionase.” (Lorente, 2001, p. 105)

Dicho mecanismo, sin embargo, aún no ha sido descrito o formalizado. De ahí que, para poder saltar este escollo, se haga necesario un replanteamiento teórico sobre el tratamiento del concepto a través de las categorías de datos de los diccionarios terminológicos. Con esta idea en mente, se propone analizar primero el problema de, para decirlo en palabras de las autoras arriba citadas, cubrir todas las caras del poliedro que es un concepto, todas las realizaciones posibles en una estructura complementaria a la definición; poder saturar el concepto de un término o por lo menos acercarse a ese ideal para que éste pueda ser aprehendido ya no sólo por el usuario experto a quien va dirigido principalmente, sino también a todo aquel usuario con una instrucción menor que ya se está formando, ya se está documentando para resolver un problema conceptual o semántico particular. De aquí el nacimiento de la noción de *saturación semántica* que abordaremos a continuación.



### **3. LA SATURACIÓN SEMÁNTICA, MOTOR DEL EJEMPLO**

En este apartado se hablará de la propuesta teórica bajo la cual se intentan resolver las problemáticas derivadas de las relaciones establecidas entre concepto, término y definición. Se parte del supuesto de que la definición no satura el concepto y a partir de esta premisa se sugieren una vía, primero, para determinar el concepto de saturación, y más tarde la búsqueda de un mecanismo que la lleve a cabo si es teóricamente posible. Para llegar a conclusiones confiables, se parte de la descripción y enlazamiento con dos principios de la TCT: el *Principio de adecuación* y el *Principio de poliedricidad*. Una vez determinados los alcances e implicaciones de ambos principios, se intenta complementar ambos principios con el fin de llegar al ideal de la saturación. El análisis hecho desemboca en que podría proponerse una noción llevada a cabo por un mecanismo cuyo ideal fuese saturar un concepto con cualesquiera categorías de datos disponible en un diccionario.

#### **3.1. La saciedad semántica, ¿una noción previa a la saturación?**

Hacia 1782 Boswell<sup>16</sup> mostraba lo endeble de la asociación entre una palabra de uso común y su significado con un simple mecanismo de repetición, pero no es hasta más de un siglo después que Severance y Washburn abordan este tema desde el punto de vista científico (Severance & Washburn, 1907). En dicha investigación, las autoras exploran las consecuencias de la lectura o fijación visual prolongada (3 minutos) en una palabra escrita y documentan los resultados. El inicial desmembramiento de la palabra en sílabas, el intento por crear otras palabras a partir de dicha disección, la disociación fonética y el final desconocimiento

---

<sup>16</sup> "Words, the representations, or rather signs of ideas and notions in the human race [...] who in a fit of musing, have tried to trace the connection between a word of ordinary use and its meaning, repeating the word over and over again, and still starting in a kind of foolish amazement, as if listening for information from some secret power in the mind itself." (Boswell, 1782; versión revisada: 1928)

del conjunto de letras como una palabra son los principales resultados de esta investigación. Sus resultados sugieren que los sujetos de prueba muestran una tendencia, después de unos segundos y hasta un par de minutos, a no reconocer como familiares ciertas palabras que son de uso diario y que, además, la primera relación que se ve afectada es aquella que se da entre grafemas y fonemas. Dicha investigación representa el punto de partida para el fenómeno que poco más de cincuenta años después Jakobovits y Lambert (1960, 1961) afianzarían con el nombre de “saciedad semántica” (*semantic satiation*).

Durante las siguientes dos décadas el fenómeno fue ampliamente investigado y presentado desde diferentes perspectivas (Severance & Washburn 1907; Basset & Warne, 1919; Jakobovits & Lambert, 1961; Amster, 1964; Esposito & Pelton, 1971; Collins & Lofthus, 1975; Neely, 1977; Cohene, Smith & Klein, 1978; Smith, 1984; Smith & Klein, 1990; Kounios, Kotz & Holcomb, 2000; Black, 2001). Los estudios sobre la caracterización de dicho fenómeno son aquellos que corresponden a la determinación de medidas de pérdida de significado (Jakobovits & Lambert, 1962), es decir, cuánto tarda una persona en sentir que la palabra que repite ya no tiene significado –que para estos autores gira en torno a los 30 segundos- y que además es un fenómeno que se da menos en personas bilingües que en monolingües (Jakobovits & Lambert, 1961). Hubo otros estudios que se centraron en la interpretación de datos (Hull, 1943; Osgood, 1953) y determinaron que la saciedad semántica es un tipo cognitivo de una reacción inhibitoria, es decir, que la estimulación repetida de una palabra automáticamente la coloca como la menos adecuada para representar un concepto. Además, hubo quien se dedicó a analizar la asimilación del fenómeno en distintas teorías interpretativas (Amster, 1964) y quien midió los efectos de la saciedad semántica por medio de un ejercicio de decisión léxica (Neely, 1977), descubriendo que si colocamos una palabra relacionada semánticamente (un óptimo) cerca de una palabra que deseamos identificar como parte o no de una lengua (objetivo), el óptimo acelera el proceso de decisión incluso cuando dicho óptimo haya sido saturado anteriormente. Algunos más sugieren que se trata de una adaptación o fatiga mental (Smith & Klein, 1990) y otros que los efectos de la saciedad semántica impiden la resolución de ambigüedades léxicas cuando se presentan con palabras relacionadas conceptualmente.

Al final, en esta revisión histórica del término *semantic satiation* encontramos que en ningún caso se le compara o asimila al término *semantic saturation* que, dicho sea de paso, ni siquiera existe en este contexto, lo que sugiere que *semantic saturation*, en todo caso, se trataría de una traducción errónea ya que sólo se encuentra en un par de páginas de “curiosidades lingüísticas”, en Wikipedia como una nota tangente y como nombre de una banda de rock progresivo. En todo caso, el término *saturation*, que traduciríamos como ‘saturación’, no existe como parte de este campo de estudio ni como un término lingüístico. Lo que sí sugiere, en cambio, es que no se trata para nada de una relación con ‘saciedad’, traducción de *satiation*, y justo es aquí donde nos detenemos para re-pensar el fenómeno de la saciedad semántica, sus efectos en la apropiación de significados (o desconocimiento de ellos) y si, en un ejercicio de traducción -pero también teórico lingüístico-, tendría algo que ver con la palabra “saturación”.

### 3.1.1. Saciedad vs saturación

El fenómeno que hemos revisado es, entonces, el de una saciedad de grafemas o fonemas, pues todos los autores se basan en la repetición oral de una palabra o su exposición prolongada en lectura. El proceso del que se habla en la mayoría artículos implica un desconocimiento temporal de la relación existente entre un grafema y su correspondiente sonoro, entre un grafema y la clase gramatical o paradigma flexivo al que pertenece, y la relación entre una palabra y los elementos con los que se relaciona conceptualmente en un campo semántico determinado. Hay que mencionar, empero, que la idea, el significado de la pieza léxica, queda intacto en la mayoría de los casos, es recuperable después de un tiempo breve. Estos experimentos muestran, además, que dicho desconocimiento temporal no es nunca permanente aunque el condicionamiento se prolongue (ni por número de repeticiones ni por tiempo de exposición continua). La saciedad semántica se refiere, como cuando comemos, a un límite al que llegamos física y mentalmente donde ya no aceptamos más de lo mismo, pero que pasado un tiempo podemos recordarlo o desearlo.

En los casos de exposición prolongada, esto es, de una lectura insistente y continua de una palabra en papel durante algún tiempo, lo que el ojo humano desvanece es la relación entre cada grafo, no entre los rasgos semánticos que representa el conjunto de dichos grafos. El cerebro comienza a ver sólo garabatos y no una palabra, una pieza de un sistema que supuestamente domina. Hacer una afirmación de que se pierde significado o de que se pierde la capacidad lectora o interpretativa equivaldría a decir que la saciedad semántica, así, sería un mecanismo que “desaparece” el significado de una palabra a fuerza de concentrar la mirada, durante un cierto lapso, en la convención gráfica que se ha acordado en cada lengua: la palabra escrita.

De igual manera, en los casos de repetición oral continua, a diferentes velocidades, volúmenes o soportes (hojas, diapositivas, lectura en voz alta, etc.), sería equivalente a decir que la pronunciación o repetición sostenida de una palabra cualquiera “desvanece” su significado, incluso hacia dos lenguas distintas (*cross-satiation*) y además en una gradación medible (*semantic differential*), como lo perfilan Jakobovits y Lambert (1961).

En ambos casos, de ser cierto que existe una saciedad semántica absoluta que resultaría en la pérdida o reducción del significado de la palabra, dicha pérdida asociativa entre grafía o sonido y significado daría paso a otro fenómeno: la resemantización; en cuyo caso, el mecanismo imperante en toda lengua por rellenar espacios cognitivo-semánticos con un referente en el sistema lingüístico daría paso a la variación —el uso de otro grafema o palabra para designar dicho significado— o a la neología — surgimiento de un nuevo grafema o palabra— para designar una realidad no-nombrada (o desnombrada, si se quiere ver así).

Así, debemos entender en todos los casos que la saciedad semántica siempre es parcial y dicho término, por tanto, atiende a la definición de ‘saciedad’ que versa: “Hartura producida por satisfacer con exceso el deseo de una cosa.” (DRAE), ya que el concepto de la palabra, y sus relaciones con otros conceptos, no desaparecen, sino que “hartan” el cerebro. La saciedad semántica es el ‘hartazgo’ de significado que, pasado un tiempo, se disuelve y deja en condiciones normales al hablante de una lengua sin consecuencias aparentes. Es decir, puede recuperar los significados

desvanecidos o las relaciones entre grafías y conceptos, sin problemas y sin afectar su capacidad lingüística.

Ahora bien, la saturación, en sentido estricto, es “llenar una medida” (DRAE); en teoría del color “la diferencia entre los valores de los canales [...] una saturación máxima tiene una luminosidad de 100% en el canal rojo y 0% en los otros canales” (Wikipedia); en química es “el punto en que una solución de una sustancia no puede disolver más de dicha sustancia”; y también, en fisicoquímica, “denota el grado en el que un sitio de fijación está completamente ocupado” (Wikipedia). Como vemos, en todos los casos se trata de una inadmisibilidad irrevocable, limita completamente o llena de manera plena algo. La saturación es la no-aceptación de algo en un lugar destinado a ello, ya que se encuentra totalmente ocupado. Visto así, mientras la saciedad es un hartazgo, la saturación es un límite. El hartazgo se puede corromper, se puede ir más allá de él; se puede ir más allá de la saciedad. De hecho, la saciedad como la hemos visto hasta ahora, desprovee de significado, aunque sea de manera temporal, a una palabra. Harta el cerebro y lo vuelve incapaz de adherir más información sobre esa pieza léxica y la lleva al límite de lo irreconocible lingüísticamente. No sucede así con la saturación, pues ésta plantaría un alto definitivo e incorruptible de adición de significado. Ergo, la saciedad y la saturación, en semántica, operan en direcciones contrarias. Mientras la saciedad semántica corrompe, desconoce, difumina las relaciones entre un concepto y la unidad que lo representa en un sistema lingüístico; la saturación semántica intentaría agregar significado, contextualidad, variación –hasta los límites posibles–, a una pieza léxica para que el entendimiento del concepto al que se asocia fuera total.

Volviendo a la semántica, podríamos entender que la saturación semántica sería aquel estado en que el recipiente de un concepto estaría completamente lleno y sería inadmisibile “agregar más significado”, es decir, sería el conocimiento total de dicho concepto asociado a, por ejemplo, un término. Como nos interesa aquí perfilar esta propuesta designativa en terminología, la saturación equivaldría a decir que una persona conoce un término en todas sus realizaciones, sus definiciones, sus contextos, y en todas las variantes conceptuales que pueda tener dentro de un discurso de especialidad. Si englobamos lo anterior dentro de la TCT, esto sería lo mismo a decir que una persona ha logrado saber

lo suficiente como para cubrir todas las caras del poliedro que supone el concepto de un término, esto es, cumpliría idealmente el *Principio de poliedricidad*.

### **3.2. La noción multifacética de un término: el *Principio de poliedricidad***

El *Principio de poliedricidad* (Cabré, 2008), como noción imprescindible en la TCT, postula que un concepto está organizado en redes y es muy difícil poder aprehenderlo en su totalidad (Cabré 1999, p.137; 2003). De acuerdo con este principio, debemos entender que solamente conoceremos una parte, algunos rasgos de los conceptos tratados en un diccionario de especialidad, cada vez que hagamos un acercamiento a ellos ya sea vía discursiva (creación del corpus, elección de datos, consulta a expertos), explicativa (diseño de ejemplos, estudio sobre adecuación, diseño del árbol de campo) o definatoria (diseño y elección de definiciones y acepciones). La poliedricidad, por tanto, se refiere a que cada concepto, en cualquier área de especialidad, tiene muchas caras, es decir, varias realizaciones, dependiendo de algunos factores cognitivos y de categorización que tienen que ver con aspectos psicológicos, antropológicos y sociológicos que se encuentran en el contexto de la realidad y, también, en relación con la adquisición del conocimiento especializado.

Ahora bien, el término, como una representación condensada del concepto, refleja en su morfología, en su sintaxis, en su definición y en su contexto de activación, cuáles rasgos conceptuales son inherentes a él y las relaciones que mantiene con ellos (la relación concepto-término de la que hablamos antes). Cuando el término, por alguna razón pragmática, tiene que modificar sus características morfológicas o intercambiarse por una pieza léxica que sea su equivalente para denominar un mismo concepto, hablaremos de variación denominativa. Dicho fenómeno, visto desde la lexicología, y por tanto desde la lexicografía, es llamado sinonimia. Este fenómeno dará fe de la focalización de un porcentaje de rasgos conceptuales compartidos, pero al mismo tiempo indicará que hay una variación en cuanto a la forma de referirse a dicho concepto en un sistema lingüístico.

Por otro lado, cuando las modificaciones se dan al nivel de las relaciones que mantiene el término con el concepto, y del concepto mismo con otros conceptos de la misma área de especialidad, estaremos hablando de una variación conceptual. Cuando la variación conceptual entra en juego, los elementos que se modificarán son aquellos que se refieran a las cualidades semánticas del término, esto es, contexto de activación, contexto definitorio y, por supuesto, definición terminográfica.

### **3.3. El Principio de adecuación**

El ideal de la terminografía es ofrecer un diccionario capaz de proveer información relevante de un cierto campo del conocimiento humano en pro del entendimiento de su red de conceptos, a la vez de cubrir las distintas perspectivas desde donde es visto. Es, además, una herramienta que busca la mejor comprensión del léxico utilizado en un discurso especializado y una de sus funciones es la mejor comprensión de dicho discurso en especialistas noveles o en formación. Este prototipo ideal de diccionario especializado es modificado por diferentes variables que actúan dependiendo de la amplitud del corpus y el alcance o sustento teórico que le otorgue el terminólogo como coordinador de las personas involucradas: tanto lingüistas como especialistas consultados. Este ideal, tal cual lo hemos explicado, es planteado y desarrollado siguiendo otro principio dentro de la misma TCT: el *Principio de adecuación*.

El *Principio de adecuación* (Cabré, 1999; 2000) es un conjunto de pautas de trabajo que permiten al terminólogo diseñar un diccionario especializado que cubra varias necesidades del usuario final a la vez. Entre ellas se debe contemplar el tipo de público al que va dirigido, los recursos existentes (financieros), las fuentes de apoyo (corpus, especialistas), la macroestructura y, de manera particular, la microestructura desde el punto de vista de la elección y del diseño de las categorías de datos.

Además, este principio debería tener también en cuenta el tipo de definición terminográfica más adecuada atendiendo a la multiplicidad de rasgos que pueden ser reflejadas en ella, dependiendo de cuáles de ellos son reflejados por el término que atiende, a su vez, el *Principio de poliedricidad*. Es decir, la

adecuación debe tomar en cuenta, al mismo tiempo, aspectos semánticos provenientes del análisis del terminólogo y especialista en el área, y aspectos pragmáticos, que es la manera de representar dicho conocimiento especializado en una definición y la manera en que será transmitido.

Si la definición terminográfica, como vimos antes, tiene limitaciones en este sentido, entonces no se podrá cumplir el *Principio de adecuación* y, en consecuencia, tampoco se podrá observar el *Principio de poliedricidad*. Esto finalmente viene a comprobar que dicha categoría de datos es insuficiente y, por consiguiente, hace falta o bien otro mecanismo de operación para una definición terminográfica o bien otra estructura que complemente a la definición y que tenga en cuenta los dos principios anteriores a la par del mecanismo que hemos mencionado. Como efecto imprescindible, también será necesario que la haga funcionar de una manera tal que se puedan reflejar a la vez la variación conceptual, la variación denominativa, el contexto de activación y el contexto definitorio. Esto es, necesitamos un mecanismo que intente saturar un concepto a través de la condensación de sus posibles realizaciones y que se adecue a registros menos especializados con el fin de ayudar a una mejor comprensión y transmisión del conocimiento especializado. La propuesta, aquí, es empezar por describir el mecanismo de funcionamiento y después la categoría de datos idónea para hacerlo funcionar.

### **3.4. La saturación semántica como mecanismo de complemento a la definición**

Los dos principios que se han tratado anteriormente son elementos fundamentales para el desarrollo del trabajo terminográfico dentro de la TCT, pero a veces no son tomados en cuenta de manera sistemática y se hace una reinterpretación *ad hoc* de ellos dependiendo del material y del producto final que se desea diseñar. Para ilustrar lo anterior, tenemos que algunos diccionarios a pesar de pretender ser de iniciación utilizan una metodología para la redacción de sus entradas llena de otros términos (tanto en las remisiones como en las definiciones) que opacan en gran medida el

entendimiento del concepto al que se refieren. Es decir, desean cumplir el *Principio de Poliedricidad* (al focalizar las relaciones de un concepto con otros concepto a partir de la aparición de los términos que los representan), pero trasgreden completamente el *Principio de adecuación*. Miremos un ejemplo<sup>17</sup>:

**“INVENTIO”** (o “eúresis”, o invención y persuasión \*, prueba, lugar, “topoi”, “locus”, lugar común, lugar propio, “quaestio”, premisa, silogismo, entimema, epiquerema, sorites, dilema).

En la tradición grecolatina, primera de las partes de la *retórica \**, que corresponde a la primera fase preparatoria del *discurso \** oratorio: la concepción de su *contenido \**, que abarca la selección de los *argumentos \** y las ideas sobre las que después habrá de implantarse un orden considerado por otra de las partes de la retórica: la “*dispositio*” (disposición). Los argumentos y las ideas funcionan como instrumentos intelectuales (que convencen) o como instrumentos afectivos (que conmueven) para lograr la *persuasión \** mediante un alto grado de credibilidad. La “*inventio*” no pertenece pues a la creación sino a la preparación del proceso discursivo, pues consiste en localizar en los compartimientos de la *memoria \** (“*loci*”) los *temas \**, asuntos, pensamientos, nociones generales allí clasificados y almacenados mediante constantes ejercicios. Establecida por CÓRAX, la “*inventio*” fue reglamentada por

Tenemos también el caso inverso, donde podemos encontrar un conjunto numeroso de definiciones o acepciones sobre un término, pero que apenas aportan información nueva; es decir, pretenden adecuarse reformulando un concepto con los mismos elementos de una definición, o varias de ellas, en lugar de buscar otros caminos como las explicaciones o los ejemplos. En este sentido, se acercan más al *Principio de adecuación*, pues intentan cumplir con la premisa de llegar al mayor público posible a través de diferentes estructuras con combinaciones léxicas distintas, pero no muestran muchas más caras de un concepto, sino apenas algunas

---

<sup>17</sup> Tomado del *Diccionario de retórica y poética* (Beristáin, 1995), esta entrada excede las cuatro páginas en la misma dinámica de la explicación enciclopédica, exhaustiva, llena de referencias a otras entradas del mismo diccionario aún cuando se refiere a esta obra, expresamente como que “ha sido planeado como material de apoyo para la docencia en cursos y seminarios [...] que se imparten en el bachillerato o en las escuelas de enseñanza superior”. Por lo que, desde el punto de vista de la adecuación, esto significaría que dichos estudiantes (entre 15 y 18, y 19 y 22 años) tendrían que comprender estas entradas. En la realidad ¿esto es posible? Creemos que no y que por esta simple razón, se corrompe el *Principio de adecuación*.

de ellas reformuladas con distintas palabras o un conjunto de informaciones definitorias que, sobre todo, opacan el concepto antes que aclararlo. Un ejemplo de este caso lo podemos ver a continuación (Mosby, 2003):

*erotismo. n. m.*

1. *Impulso o deseo sexual*
2. *Despertar o intento de despertar el instinto sexual por métodos sugestivos o simbólicos*
3. *Expresión del instinto o deseo sexual*
4. *Impulso sexual anormalmente persistente*

Ante esta asistematicidad y la problemática que nos plantea mostrar un conjunto amplio de realizaciones o contextos donde un término se puede desarrollar, la noción de *saturación semántica* puede jugar un papel importante. Si volvemos a los planteamientos arriba mencionados, la *saturación semántica* sería un caso hipotético e ideal donde un hablante conoce un concepto en todas sus realizaciones, es decir, los términos, sus definiciones, sus contextos y todas las variantes conceptuales y denominativas que pueda tener dentro de un discurso de especialidad. Este caso hipotético cubriría, al mismo tiempo, los principios de poliedricidad y adecuación pues, en el primer caso, se podrían ver todas las caras de dicho concepto y, al conocer todas esas realizaciones, ya no admitiría ni una más, el poliedro estaría completo; en el segundo caso, al tener disponible toda la información semántica y léxica de un término y sus relaciones conceptuales, se podrían formar oraciones o cláusulas dependiendo del registro del consultante para mostrar los rasgos más relevantes de un término y que puedan ser apprehendidos por éste.

También podríamos mencionar que esta propuesta puede operar en las distintas categorías de datos de un diccionario. No olvidemos que el concepto de un término está repartido en varios contextos y estos pueden ser a veces inaprehensibles por una lengua o difícilmente vertidos en una oración. Recordemos que el concepto es, ante todo, una representación mental, no lingüística, de objetos e ideas que no necesariamente existen en la realidad palpable. Habría que echar mano de otros elementos como imágenes, sonidos o fórmulas, lo que podría llevarnos hacia un trabajo muy cercano a la

enciclopedia que, a su vez, nos desviaría de la meta inicial, que es el diseño de un diccionario de especialidad.

Ante este escenario, la saturación semántica intenta condensar los conceptos en estructuras lingüísticas no definatorias y, por ello, prepondera el uso de la lengua como vehículo de transmisión de los rasgos del concepto. Esto desemboca en un estudio profundo sobre la distribución y características de dichos rasgos en las diferentes categorías de datos. Así, esta noción tiene en cuenta dos puntos imprescindibles:

- a) Que la definición terminográfica no logra saturar el concepto de un término y que para la efectiva transmisión del conocimiento especializado hacen falta complementos conceptuales, y
- b) Que los elementos lingüísticos disponibles en un diccionario de especialidad –esto es, otras categorías de datos– deberían complementar dicha definición a través de la puesta en escena de los rasgos del concepto que no han sido mostrados por la definición.

Como vemos, las limitaciones de la definición terminográfica, en tanto vehículo del conocimiento especializado, se pueden superar al encontrar, en esas otras categorías de datos donde están repartidos los rasgos del concepto, alguna que la complemente de tal manera que cumpla varias funciones a la vez, pero que no se aleje demasiado de ella. Es decir, en tanto complemento, no debe ser una reformulación de la misma definición, pero tampoco un acto independiente. La saturación siempre parte de una insuficiencia, no opera como mecanismo aislado, es decir, la saturación semántica es una complementación conceptual al proceso definitorio.

Mencionábamos en los primeros apartados de esta tesis que la lengua y el conjunto de los significados asociados a los signos que utilizamos como palabras son transmitidos, esencialmente, sobre una base de imitación. Aprendemos la lengua por imitación de los padres o la familia y uno de los mecanismos más efectivos para construir una realidad del mundo, es decir, una red conceptual, es la ejemplificación. Esta ejemplificación, como proceso, desemboca en varios tipos de ejemplos dependiendo de la

naturaleza del significado que desean destacar en un momento dado. Es decir, siempre están asociados a un significado ya sea léxico – como en los diccionarios de lengua general– o conceptual –en los diccionarios especializados–, por nombrar los dos productos de los que nos ocuparemos desde ahora. Por tanto, un ejemplo siempre está asociado a una definición y los constituyentes de dicho ejemplo serán perfilados por lo dicho en la definición sin ser éste una reformulación de aquella, sino una complementación semántica que, en el caso del diccionario especializado será conceptual.

En este sentido, el elemento que más aprovecha lo postulado en la noción de saturación semántica es la ejemplificación. Esta saturación es, pues, el mecanismo por el cual un ejemplo complementa de manera armónica y sistemática cualquier definición terminográfica desde sus diferentes realizaciones. Dichas realizaciones se pueden observar desde el punto de vista del contenido en explicaciones, reformulaciones, extensiones, símiles, etcétera; y desde el punto de vista formal en concordancias, colocaciones, contextos, etcétera. Las razones por las cuales el ejemplo es la categoría de datos que mejor se adapta a los requerimientos de la saturación semántica para complementar conceptualmente una definición terminográfica se detallan a continuación.

### **3.5. La *saturación semántica* como mecanismo de operación del ejemplo**

Ya hemos dicho que la saturación semántica es un proceso complementario al de la definición y que su finalidad ideal, en teoría, es la de mostrar todas las posibles realizaciones de un concepto a través de estructuras lingüísticas. También hemos dicho que esta saturación semántica está en concordancia con los principios de poliedricidad y de adecuación debido a que, el primero, al intentar saturar un concepto, realiza la tarea de intentar cubrir todas las caras de ese poliedro imaginario que supone el concepto; y el segundo, estas realizaciones son complementarias a la definición ya que no parten de un mecanismo definitorio, sino que intentan completarlo con aquellas estructuras lingüísticas más generales, aprehensibles a un lector lego.

Si se quiere ver de esta manera, la saturación semántica completa el mapa mental de un concepto a través de la puesta en escena de esas estructuras menos especializadas que la definición, pero que conservan rasgos de aquella al tener siempre como núcleo un mismo término. Ergo, la saturación semántica explica por qué frases u oraciones que contienen un término conservan rasgos del concepto que representa y que no son tomados en cuenta en el proceso definitorio de dicho término. La razón, como lo miramos en el capítulo 2, es que la definición terminográfica es la condensación de un concepto. El proceso definitorio hace eso justamente, delimita un significado dejando dentro de ella, de la definición, los rasgos más pertinentes de acuerdo con los criterios de los especialistas involucrados en su diseño. Estos rasgos son mucho más específicos en tanto descriptores del concepto, sí, pero recordemos que un concepto es difícilmente aprehensible totalmente en la realidad. Su comprensión y, por tanto, su decodificación, sólo es posible cuando se tiene una cantidad suficiente de información sobre él. Esta comprensión, hasta ahora, se ha propuesto como un logro posible sólo a través de la especialización.

Si quisiéramos transmitir ese mismo concepto a personas con un nivel de especialización menor, o mínimo, es obvio que el mecanismo definitorio nos supondría un obstáculo al ser una estructura demasiado densa y especializada. Ergo, no es la definición el elemento que idealmente se debería utilizar al inicio de dicha transmisión de conocimiento, sino estructuras lingüísticas alternas con un nivel de especialización menor. Ahora bien, estas estructuras tampoco se pueden separar completamente de la definición, ya que ella tiene la mayor parte de la información que queremos transmitir. Así, se debe hacer un análisis profundo para saber cuáles de las estructuras alternativas mantienen un mayor nivel de cercanía con la definición, pero mayor lejanía con respecto a la especialización del léxico utilizada en ella. Si volvemos atrás otra vez, veríamos que el ejemplo es quien cumple con estas características.

Distintas aportaciones teóricas han mostrado que el ejemplo siempre trata de ayudar a la mejor comprensión del significado de una palabra. Son muchas y muy variadas las posiciones y los estudios con respecto a su forma y sus funciones, pero todas ellas concluyen, de una u otra manera, en que se trata de un apoyo a lo

que existe en el diccionario; y en el diccionario existen esencialmente definiciones.

Si la saturación semántica opera desde y para el concepto, entonces el resultado de su operación coadyuvará a comprenderlo mejor. Si el ejemplo es un elemento que ha sido siempre utilizado como apoyo en la comprensión de un significado, y particularmente en terminología en la comprensión de un concepto, entonces es lógico pensar que el mecanismo por el que opera el ejemplo es el de la saturación semántica. El ejemplo intenta completar, pero sólo algunos de ellos lograrán acercarse más a la saturación. Por tanto, ¿en qué medida el ejemplo puede saturar o intentar saturar un concepto y cómo mediríamos esto?

Para poder comprobar el supuesto anterior, veremos en el siguiente capítulo cómo se ha tratado el ejemplo en los diccionarios, partiendo brevemente del de lengua general, hasta enfocarnos en el de especialidad, en tanto en el capítulo 5 analizaremos la propuesta de caracterización del ejemplo en terminología para responder las preguntas que hemos lanzado.

### **3.6. Discusión: la *saturación semántica*, prolegómenos para su tratamiento en terminología**

Como podemos ver, la saturación semántica operaría en varias direcciones a la vez y enlazaría los principios básicos del diseño de diccionarios especializados con el reanálisis de los roles de las categorías de datos, específicamente el del ejemplo y, al mismo tiempo, tendría en cuenta a los usuarios y la metodología de trabajo para cada material especializado sin corromper los dos principios ya establecidos en la TCT. Es decir, sería un complemento teórico (terminológico) para una práctica específica (terminográfica).

Hay que tener en cuenta, sin embargo, que saturar un concepto es una tarea poco menos que imposible debido a la naturaleza exhaustiva del trabajo que supondría abarcar todas sus definiciones, acepciones y descripciones. Pero, si tenemos en cuenta que gran parte de ese camino ya es cubierto por la definición terminográfica, entonces la tarea a seguir sería la de elegir fragmentos textuales lo suficientemente densos para que la

complementen y así intenten llegar a un porcentaje más cercano a la saturación. Esta tarea, empero, debe tomar como bandera la adecuación y no acercarse demasiado a la definición en el sentido de la elección de piezas y estructuras léxicas especializadas. Es decir, debe mostrar ciertas características que focalicen al término, que digan algo más de él, pero con un léxico más sencillo. Si vamos un poco más atrás, veríamos que de los escenarios posibles para analizar las relaciones y los mecanismos operantes en el trinomio concepto-término-definición, el de los contextos es el que mejor ha funcionado. Hemos hablado de contextos definitorios, de contextos de activación, de contextos de uso, etcétera. De estos, aquel que cumple este papel de mostrar los rasgos de un concepto, representado por un término, en una estructura sencilla distinta a la definitoria, es el contexto de activación (Kuguel, 2007).

La existencia de este contexto nos permite inferir dos preceptos pilares para esta tesis:

1. Que la activación de un término, la puesta en escena de rasgos conceptuales, no atañe a unas cuantas estructuras, sino que ocurre en diferentes realizaciones léxicas que tienen patrones específicos, y
2. Que el carácter especializado de una palabra se da en muchas ocasiones en contextos menos densos que los contextos definitorios. Recordemos que los contextos de activación de los que habla Kuguel tienen mucho más que ver con un contexto de uso que con un contexto definitorio.

Por consiguiente, si hay estructuras densas conceptualmente, que contienen mucha información semántica, pero que por su léxico forman parte de estructuras más sencillas como lo son los contextos de uso, entonces podríamos pensar que la cualidad de especializado de un término se mantiene (o se activa) a pesar de encontrarse en un contexto no especializado. A efectos prácticos: se comprueba que se puede aportar información conceptual valiosa desde estructuras más sencillas; ergo, se cumplen los dos postulados básicos de la saturación semántica.

Llegados a este punto, parece que apuntamos a que un contexto de activación puede ser equivalente a un ejemplo o que ambas estructuras cumplen un mismo papel. Sin embargo, las

funciones llevadas a cabo por ambos elementos son distintas y su utilidad en la labor terminográfica dista en buena medida. El contexto de activación funciona básicamente desde la sintaxis de la oración que contiene el término y se basa en los patrones que siguen determinadas palabras dentro de ella. Según este modelo, dichos patrones forman configuraciones semánticas determinadas que, vistas bajo la teoría del modelo léxico generativo de Pustejovski (1995), activan o desactivan ciertos sentidos de algunas palabras a partir de la alteración de su estructura argumental; el resultado es la activación de rasgos específicos de esa palabra que la convierten en término.

Por otro lado, el ejemplo actúa desde un nivel conceptual. No tiene un patrón determinado y la aparición de los elementos que contiene en su estructura sintáctica están determinados en gran medida por el significado que se le da al término a través de la definición terminográfica. Las bases sobre las que descansan estos argumentos están detalladas en el capítulo 5 de este trabajo. Baste saber que en este caso no existe un patrón sintáctico, sino un patrón de comportamiento que atiende más a criterios cognitivos; se trata, proponemos aquí, de un patrón conceptual.

Este patrón conceptual viene a complementar otro patrón, también conceptual, que ya conocemos bien: el de la definición terminográfica. La definición terminográfica, en tanto patrón, focaliza las relaciones existentes entre el concepto (que define a través del término) y otros conceptos dentro de su mismo campo de especialidad. Dicho patrón, sin embargo, tiene limitaciones que responden a su estructura sintáctica, pues, como hemos mencionado y mostrado, la estructura de una definición es bastante estática. El cambio más probable en este tipo de estructura sólo se dará si se cambia su referente inmediato, esto es, el término, lo que dará paso a una variación denominativa.

El ejemplo, al no tener estas restricciones, por no tener una estructura sintáctica fija, puede mostrar más libremente los rasgos de un término, sus características semánticas y los significados asociados al concepto que éste representa. Es decir, el campo semántico de un ejemplo es mucho más amplio que el de una definición, pero menos preciso. La imprecisión puede dar paso a, o permitir el ingreso de, la variación conceptual. El resultado de esta

imprecisión será, empero, benéfico si pensamos que nunca deja de estar en relación con el concepto, sino que simplemente muestra cómo se comporta el término que se le ha asociado en un lenguaje más general, no definido, no definitorio, e incluso cómo este concepto puede proyectar diferentes estructuras y representaciones léxicas que se correspondan con los rasgos focalizados (que serán aquellos que ha dejado fuera la definición).

Resumiendo: para efectos de este trabajo hemos decidido, pues, que la noción de saturación semántica es imprescindible, ya que nos otorga una explicación satisfactoria sobre el proceso de complementación de un concepto con el fin de hacerlo más aprehensible por un público que no se había considerado en su totalidad. Explica, además, las limitaciones que son inherentes a la categoría que debe complementar (la definición) y al mismo tiempo explica el funcionamiento de la categoría complementadora (el ejemplo).

Así pues, debemos entender la saturación semántica como una de las posibles explicaciones del funcionamiento del ejemplo, sobre todo en terminografía, y que funciona bajo los siguientes principios:

- 1) La saturación semántica es un proceso cognitivo por el cual un hablante intenta conocer la totalidad de un concepto a través de estructuras lingüísticas que lo acerquen al proceso de abstracción de la realidad que realiza la mente en el momento de formar dichos conceptos.
- 2) La saturación semántica, como indica su nombre, intenta llenar el significado de una pieza léxica a través del descubrimiento y organización de los rasgos semánticos de dicha pieza en diferentes estructuras léxicas.
- 3) En terminografía, la saturación semántica se refiere a la saturación de un concepto, es decir, al ideal que supone el *Principio de poliedricidad*, de la TCT.

- 4) La saturación semántica es un esfuerzo en la labor terminográfica por acercar el diccionario de especialidad a un público más amplio. En este sentido, intenta llegar al ideal del *Principio de adecuación* de la TCT.
- 5) La saturación semántica explica el mecanismo por el cual el ejemplo funciona como tal y lo distingue de otras estructuras como contextos, concordancias y colocaciones.
- 6) La saturación semántica es un complemento teórico-terminológico para una práctica específica: la terminografía.
- 7) La saturación semántica explica los alcances y limitaciones de otras estructuras sintácticas de base conceptual como las definiciones terminográficas.
- 8) La saturación semántica es un mecanismo de complemento conceptual que se lleva a cabo sobre todo en el proceso de la ejemplificación.
- 9) La saturación semántica es un mecanismo de funcionamiento semántico de ciertas categorías de datos que buscan un ideal: dar fe del uso y variación de una palabra o término en todas sus realizaciones contextuales. Por tanto, como ideal, es poco probable que se cumpla, sin embargo, en tanto mecanismo, ayuda a justificar la aparición de elementos complementarios a la definición y permite especificar un conjunto de pautas que hace de ciertos fragmentos los mejores candidatos para complementar dicha definición a la par de hacerlo de una manera más sencilla, con un registro más cercano al léxico general.

Con todo lo anterior debemos entender, entonces, que la saturación semántica es una noción teórica que permite explicar el proceso ejemplificativo. Esto es importante porque, como veremos a continuación, aún cuando los estudios sobre el ejemplo han profundizado bastante sobre su identificación, sus funciones, sus formas y sobre todo los criterios que se pueden tomar en cuenta para determinar si se trata de un “buen ejemplo” (de la tradición inglesa *good examples*), su mecanismo de acción y el proceso por el que se forman no han sido descritos. En el capítulo 5 veremos que un ejemplo, como lo hemos sugerido aquí, no tiene una forma determinada ni un patrón o un conjunto de patrones sintácticos asociados a él. Tomando en cuenta esto, y bajo el argumento de que todo hablante puede emitir un ejemplo sobre una palabra con relativa facilidad, entonces inferimos de que quizá sí existe un patrón para formar esta categoría de datos pero no se encuentra en la lengua misma. Ahora bien, todo patrón se forma bajo algún mecanismo de formación. Si el mecanismo de formación no ha sido descrito, pero los efectos son reconocibles, entonces es posible encontrarlo a través del análisis del conjunto de tales efectos. Entiéndase efectos, aquí, como funciones y características. Todas estas funciones, descritas y analizadas por los estudiosos que hemos citado y que citaremos en los siguientes apartados, han permitido conformar la propuesta que describe el mecanismo de funcionamiento del ejemplo y, a través del análisis de ese mecanismo, una metodología plausible para la identificación automática de dichos ejemplos. La base del análisis es, por consiguiente, la relación existente entre los elementos que lo conforman. Por eso la necesidad de explicar de qué manera se relacionan término y concepto, término y definición, hasta dónde llega la definición y finalmente cuál es la función principal del ejemplo que la acompaña. Como vemos, aún sin contar con uno de los pilares de estudio de la lingüística aplicada, como lo es la sintaxis, se puede llegar a conclusiones aceptables a través de los efectos de las relaciones conceptuales, de eso que no vemos directamente en las estructuras del lenguaje. Si volvemos atrás una vez más, esto sólo viene a comprobar que si el concepto es una abstracción de un segmento de la realidad, un proceso cognitivo, no se podría esperar que uno de sus descriptores, el ejemplo, fuese un proceso alejado de mismo proceso cognitivo. La propuesta es la saturación semántica.

Para poder mirar a detalle el camino que seguimos para conformar esta propuesta, será necesario ver qué se ha dicho antes sobre el ejemplo (capítulo 4), qué de eso se ha aprovechado en terminología y finalmente cuáles serían las condiciones ideales para formar un ejemplo en terminografía (capítulo 5). Si todo esto es posible, si la saturación semántica funciona, entonces es razonable pensar que se puede automatizar dicho proceso (capítulo 6).

## 4. EL EJEMPLO, HOY

En este capítulo abordaremos los estudios referentes a las características del ejemplo en diccionarios. En primer lugar mostraremos cómo un estudio previo nos permitió ver que el ejemplo no es actualmente una categoría de datos obligatoria y que, de hecho, se utiliza menos en ciertos diccionarios de especialidad que en diccionarios generales. Después haremos un recorrido sobre lo que han mencionado algunos autores a propósito de las características de un ejemplo en lexicografía. Continuaremos con un par de estudios referidos al ejemplo en terminología, uno independiente de lengua y uno dirigido al castellano, y a partir de ellos, y de la información aportada en los apartados anteriores, se hará una caracterización teórica del ejemplo. Finalmente se abordará una propuesta teórica acerca de un elemento entre una definición y un ejemplo, que Lara (2006) ha llamado *ejemplo definitorio oblicuo*, y veremos si la existencia de dicho elemento nos puede ayudar a discernir con mayor precisión las funciones de un ejemplo.

### **4.1. La ejemplificación en diccionarios actuales, un estudio**

A través de la revisión de dieciocho diccionarios divididos en: cinco generales en formato papel (grupo P1), cinco especializados en formato papel (grupo P2), cinco generales en formato digital (grupo D1) y tres especializados en formato digital (grupo D2); y un monitor, obtenido del *Corpus de las Sexualidades en México* (CSMX) (Sierra, Medina & Lázaro, 2009), se diseñó un estudio para determinar la aparición o ausencia de ejemplos y contexto de uso referidos a quince términos en el área de sexualidad.

TÉRMINO	EQUIVALENTES
bisexualidad	preferencia bisexual
coito	
diafragma	
erotismo	
género	
heterosexualidad	
homosexualidad	
identidad de género	identidad sexual
intersexualidad	
preferencia sexual	orientación sexual
sexo	
transexualidad	
travestismo	
sexualidad	
papel genérico estereotipado	rol sexual, papel genérico

**Figura 4.1. Lista de términos**

Clave	Diccionario	Grupo
DSMX	Diccionario de la Sexualidad en México. Prototipo. 2011.	Monitor
DRAE	Diccionario de la Real Academia Española. 22a ed.	P1
DUEMM	Diccionario de Uso de María Moliner. Gredos. 2008	
DUEAyE	Diccionario de uso del español de América y España. VOX. 2002.	
DIME	Diccionario Inicial del Español de México. Trillas. 2004	
DTCM	Diccionario Terminológico de Ciencias Médicas. Masson. 2003.	
ESPASA	Diccionario Espasa Medicina. Espasa. 1999.	P2
MOSBY	Diccionario Mosby. Medicina, enfermería y ciencias de la salud. Mosby. 2003.	
TABER	Diccionario enciclopédico TABER de ciencias de la salud. DAE. 2008.	
MARBÁN	El gran Harper Collins Ilustrado. Diccionario Médico. MARBÁN. 2006.	
Wref	Wordreference. <a href="http://www.wordreference.com/es/">http://www.wordreference.com/es/</a>	D1
MUNDO	Diccionario de la Lengua Española de "El mundo". <a href="http://diccionarios.elmundo.es/diccionarios">http://diccionarios.elmundo.es/diccionarios</a>	
LAR-VOX	Diccionario Larousse-VOX en línea. <a href="http://www.diccionarios.com/">http://www.diccionarios.com/</a>	
PAIS-SANTLL	Diccionario castellano de "El País". Powered by Santillana. <a href="http://www.elpais.com/diccionarios/castellano/">http://www.elpais.com/diccionarios/castellano/</a>	
CLAVE	Diccionario CLAVE en línea. <a href="http://clave.librosvivos.net/">http://clave.librosvivos.net/</a>	
IATE	InterActive Terminology for Europe. <a href="http://iate.europa.eu/iatediff/SearchByQueryEdit.do">http://iate.europa.eu/iatediff/SearchByQueryEdit.do</a>	D2
DeCS	Descriptores en Ciencias de la Salud. Biblioteca virtual en salud. <a href="http://decs.bvs.br/cgi-bin/wxis1660.exe/decsserver/">http://decs.bvs.br/cgi-bin/wxis1660.exe/decsserver/</a>	
TERMIUM	TERMIUM Plus. <a href="http://www.btb.termiumplus.gc.ca/">http://www.btb.termiumplus.gc.ca/</a>	

**Figura 4.2. Recursos consultados**

La metodología constó de tres etapas. En la primera etapa creamos una base de datos que llamamos “monitor”. La nombramos así por ser aquella con la cual comparamos los datos de los diccionarios que elegimos analizar. El monitor, así, sugiere una base de datos “ideal” para nosotros, lo que significa que cada término contiene al menos una definición y un contexto de uso.

En la segunda etapa creamos tablas con las definiciones, ejemplos y contextos de uso, en caso de que existiesen, por cada uno de los diccionarios que utilizamos. Estas tablas contienen información acerca del tipo de diccionario del cual procede la información (especializado o de lengua general), el tipo de definición asociada a su término y las definiciones, ejemplos y contextos que fueron encontrados para cada uno de los términos.

Finalmente en la tercera etapa creamos gráficas con los datos obtenidos de las tablas para un posterior análisis contrastivo. Para esta etapa tomamos en cuenta el número de definiciones, contextos de uso y ejemplos presentes en cada diccionario.

Los resultados de la búsqueda para los términos en diccionarios generales proporcionan una imagen de estabilidad en cuanto a distribución, aunque su aparición sea ligeramente inferior con respecto a los diccionarios de especialidad. Esto quiere decir que aunque la aparición de los términos elegidos fuera menor en diccionarios generales, entre ellos hay una distribución uniforme – casi en todos se encuentra la misma cantidad– mientras que para los especializados hay polos muy claros: algunos los contienen casi todos los términos y algunos carecen casi totalmente de ellos.

Cuando los términos cuentan con más de una definición, los ejemplos asociados a ellos en los diccionarios generales disminuyen a un ritmo de 50% más o menos (49%>22%>10%) en los diccionarios digitales y a un ritmo de 65% (33%>13%>4%) en los diccionarios formato papel. Este es un posible acercamiento a cómo se comportan las definiciones de los términos referidos a sexualidad cuando comparamos diccionarios generales en formato papel y digital. Lo anterior quiere decir que mientras más definiciones proporciona un diccionario, menor es la posibilidad de encontrar un ejemplo que la refuerce; de acuerdo con nuestro análisis para un orden del tipo primera definición, segunda definición, tercera definición, en los diccionario digitales la posibilidad de encontrar

ejemplos es: 1 de cada 2, 1 de cada 4 y 1 de cada 10, respectivamente; y para los diccionarios digitales sería: 1 de cada 3, 1 de cada 10 y 4 de cada 100, respectivamente.

Por otro lado, suponemos que los encargados de los materiales consultados han actuado bajo el supuesto de que “a mayor especialidad menor explicación”, como lo presupone Zgusta al hablar de la aparición de ejemplos asociados a términos en diccionarios generales (Zgusta, 1971, p. 263). Bajo nuestra observación, los diccionarios especializados cuentan con un menor número de ejemplos y contextos de uso que los diccionarios de lengua general.

El análisis que hemos realizado nos ha permitido observar dos cosas principalmente:

1) Que el porcentaje de aparición de términos de especialidad en el área de la sexualidad humana entre diccionarios generales como especializados se mantiene (más en los primeros que en los segundos), pero entre aquellos en formato papel contra los de formato digital tiende a variar bastante. Dichos términos son más proclives a aparecer en diccionarios generales, para nuestra sorpresa, que en diccionarios de especialidad; lo que se traduce en que los diccionarios generales, por el momento, le dan un mayor peso a esta área que aquellos dedicados a la salud humana propiamente dicho. Pensamos que esto se debe a que los segundos lo toman desde un ángulo completamente reproductivo y de salud y los primeros, además, como un asunto social y cultural.

2) Que la aparición de los ejemplos asociados a las definiciones dadas a esos términos son cada vez menores. Esto quiere decir que hemos encontrado que los ejemplos disminuyen en la mayoría de los diccionarios cuando debiera de ser al revés, pues nuestras muestras no son de diccionarios antiguos, sino de una tradición de 40 años atrás, más o menos pero con publicaciones y ediciones de los últimos 15 años. Además, esta tendencia se nota más en diccionarios especializados que generales y más aquellos en formato papel que en digitales.

## 4.2. Breve esbozo del ejemplo en lexicografía

Desde el punto de vista teórico, el ejemplo ha sido visto por lexicógrafos y lingüistas de maneras distintas de acuerdo con las necesidades que surgen en un momento dado de su investigación. Hacia 1726 el *Diccionario de Autoridades* era un modelo a seguir en los estudios de lengua española y su pilar más fuerte era, justamente, la inclusión de “autoridades” o citas, como lo llamamos hoy. Desde 1780, la supresión de la mayoría de estas citas parece haber marcado un camino bastante común en la planeación de diccionarios tanto generales como especializados. Si bien antes el uso de categorías de datos más allá de la definición constituía un ejercicio constante para poder comunicar conceptos y significados, hoy en día la supresión de elementos que no tengan que ver con el contexto definitorio (definiciones y acepciones) es una práctica común.

Se pensaría que los diccionarios, al tratar términos, recurrirían a otro tipo de información, debido a lo complejas que pueden llegar a ser sus definiciones, pero no es así. Lo que sí hay, sin embargo, es una evolución en el tratamiento de este tipo de información y la manera en que se hace presente en una entrada de diccionario. Desde las citas –textuales u orales– copiadas a mano, que aún encontramos, por ejemplo, en el diseño del *Diccionario de Mexicanismos* de la Academia Mexicana de la Lengua (Company, 2010), hasta técnicas avanzadas que conjuntan semántica y extracción de información para diseñar ejemplos en otras áreas distintas a la lingüística (Kim et al., 2013).

Las propuestas actuales parten de la identificación de estructuras lingüísticas que sirvan como apoyo informativo a la definición dentro de un artículo lexicográfico. Esta búsqueda e identificación se ha llevado a veces como una tarea rigurosa, por ejemplo en diccionarios como el COBUILD dirigido por Sinclair (1987) y el Diccionario Planeta (1982). Aunque la mayoría de los estudios que destacan la importancia del ejemplo tienen que ver más con diccionarios orientados a traducción y adquisición de segunda lengua (L2) (Alvar 1982: 191-201, Cowie, 1989; Drysdale, 1987; Fox, 1987; Humble, 1998; Laufer, 1992; Minaeva, 1992; Nesi, 1996; Paquot, 2008; Sinclair, 1987), no desarrollaremos a fondo

aquí su contenido debido a que este trabajo está orientado al ejemplo en diccionarios monolingües.

Actualmente, la mayoría de los especialistas que se dedican a la confección y desarrollo de diccionarios coinciden en que la elección de un fragmento textual que será utilizado como ejemplo puede partir de diversas fuentes y es tratado a través de distintas técnicas atendiendo básicamente a una especie de adecuación (a quién va dirigido el diccionario). Así, en algunos casos se echa mano de las colocaciones y se modifican mínimamente. Es decir, se conservan fragmentos textuales amplios para aportar suficiente contexto de uso, dando lugar a un diccionario que da importancia vital al contexto (Mott & Mateo, 2009). En otros casos hay una mezcla entre ejemplos que parten de métodos computacionales, como la generación de concordancias y colocaciones (Kilgarriff et al., 2008), ejemplos confeccionados *ad hoc* -inventados por el lexicógrafo- (Fuentes Morán & García Palacios, 2002, p. 79) y fragmentos tomados del habla popular (Lara, 2008; 2010).

Casos menos prototípicos encontramos cuando la adecuación se encuentra a medio camino entre lo especializado, lo general y la enseñanza (Estopà, 2013). De este último caso cabe destacar que la técnica utilizada para crear las entradas del diccionario incluye la fusión entre diferentes tipos de definiciones para crear una sola y la inclusión de fragmentos discursivos que ejemplifican (que son ejemplos en sí), pero que de igual manera se fusionan con la definición. Esto da paso a un híbrido capaz de transmitir conceptos o significados, es decir, está entre una definición y un ejemplo. Hay casos que vale la pena mencionar por lo acertado en la intuición de la transmisión de un concepto atendiendo principalmente al uso: el *Diccionario de Mexicanismos* (Company, 2010). En este diccionario en particular, una de las características que salta a la vista es el hecho de que una buena parte del corpus de trabajo está compuesto de muestras de habla espontánea atendiendo a “la competencia lingüística, memoria, agudos oídos y creatividad del equipo de colaboradores”, lo que es una ventaja si se quiere aplicar a otro tipo de análisis como el diseño de ejemplos, porque la confección multifuente permite, por un lado, echar mano de expresiones reales captadas sin ningún control –documentadas anónimamente– y por otro enriquecer la estructura del ejemplo a través de la confirmación de datos por

fuentes escritas o documentadas. Se trata, pues, de un material rico en contextos *actuales* que muestran rasgos de los conceptos propios de una pieza léxica concreta y que sirven para identificarla y ejemplificarla, aunque pobre en una metodología para el diseño de sus definiciones (Lara, 2011).

Un trabajo interesante que recoge dos factores importantes para esta investigación es el *Breve Diccionario Ejemplificado de Mexicanismos* (Steel, 1999), cuya confección está dedicada tanto a hablantes de lengua inglesa, como a hablantes de la variante peninsular del español. La importancia de esta obra, publicada originalmente como e-book, radica en que intenta, por un lado, mostrar palabras de uso exclusivo en México para las cuales superpone los equivalentes y ejemplos, extraídos en forma de citas desde fuentes cultas y coloquiales; y por otro lado suprime las fórmulas definitorias –son mínimas y no prototípicas–, lo que desemboca en que el lector debe “descubrir” el significado de las unidades que selecciona para las entradas a través de un ejercicio de inferencia semántica. Se trata del otro extremo del caso de la transmisión de un concepto, la desaparición casi total de las definiciones, que da paso sólo a los ejemplos.

Hablando específicamente del ejemplo como una categoría de datos independiente, han sido muchos los trabajos y autores que lo han tratado. Veamos algunos de los estudios más representativos orientados a la descripción del ejemplo en diccionarios monolingües.

#### 4.2.1. Zgusta

Zgusta (1971, p. 265) describe dos tipos de ejemplos dependiendo de su origen. Para él, los denominados *quoted examples* son aquellos fragmentos provenientes de un corpus textual que, bajo ciertos criterios establecidos por el lexicógrafo, tienen una longitud determinada, como primera característica inherente. Dicha longitud atenderá al tamaño del diccionario<sup>18</sup>. La segunda característica

---

<sup>18</sup> Aunque apenas hay diez años de distancia entre el trabajo de Zgusta y el de Sinclair, podemos notar aquí, todavía, la visión clásica del trabajo lexicográfico. Es decir, históricamente el ejemplo tuvo también una restricción de longitud o

inherente a este tipo de ejemplos es lo que denomina la “citación”. Este rasgo se refiere a indicar, de la manera más precisa posible, el lugar de donde ha sido extraído el fragmento textual (autor, obra, fecha, página, etcétera). El segundo tipo de ejemplos que menciona son los *constructed examples*, que son aquellas frases que el lexicógrafo construye para mostrar ciertas combinaciones contextuales de una palabra y que idealmente debe contrastar con informantes que le otorguen validez.

#### 4.2.2. Meyer

Meyer (2001), dentro de su concepción de *Contextos Ricos en Conocimiento* (KRC’s), incluye los ejemplos y las definiciones como las representaciones implícitas comunes de una red conceptual. Explícitamente dice:

As a result of analyzing conceptual relations, a high-quality terminology project ultimately aims to illustrate the network of concepts underlying the terms of the domain. In traditional term banks and specialized dictionaries, the conceptual network is usually represented implicitly, through the definitions and examples provided to elucidate the meaning of a term. (Meyer, 2001, p. 280).

Es decir, coincide con lo que hemos apuntado anteriormente sobre los términos, que se refieren a conceptos y las relaciones entre ellos, pero también incluye aquí a los ejemplos, lo que desemboca en que existe una función compartida entre definiciones y ejemplos; por ello no es raro pensar que el ejemplo es también, como se muestra en el capítulo 3, un elemento conceptual que está relacionado íntimamente con la definición terminográfica.

---

aparición debido a que la extensión de las entradas estaba limitada por el presupuesto y el tamaño máximo del diccionario. Incluso en este mismo texto se menciona que es “normal” que los diccionarios de *technical terminology* no presenten ejemplos debido a la naturaleza específica de sus unidades” (Zgusta, 1971, p. 263), como si la pura definición fuese elemento suficiente para aprehender significado y contexto de uso . Sobre este tema volveremos más adelante.

### 4.2.3. Atkins & Rundell

Para Atkins y Rundell (Atkins & Rundell, 2008) un *buen ejemplo* “has to get the right balance between too much context and too little.”, es decir, se basan en la naturaleza contextual de un fragmento para proponerlo como ejemplo. Dentro de su estudio, más allá que una forma definida, el ejemplo debe cumplir con ciertas funciones, a saber: la naturalidad/tipicalidad, la informatividad y la inteligibilidad. *Naturalidad* es aquello que se basa en la intuición del hablante y que no tiene una base objetiva sólida, se trata de la preferencia que cierta palabra o estructura tiene hacia otras palabras o estructuras con las que se une y donde la recurrencia es altamente importante, ya que en la medida en que estas combinaciones se den, podríamos afirmar que es más natural que un hablante use cierta estructura o prefiera ciertas palabras para relacionar significados. Lo natural, entonces, es un asunto de preferencia léxica antes que un asunto premeditado o medible. Cuando se habla de lo natural en un ejemplo se hace referencia a lo más usual, lo más común en la lengua, pero que sea suficientemente informativo. Ahora bien, un ejemplo *informativo* es aquel que complementa la definición y ayuda a comprenderla mejor. Además, este ejemplo tiene que estar en concordancia con la definición con la que se relaciona, no puede describir una situación o un conjunto de características que no coincidan con el conjunto de características de la definición. El ejemplo, para ser informativo, debe tener una función clara dentro de la entrada: debe reflejar el uso y agregar un valor. Finalmente, la *inteligibilidad* se refiere al uso adecuado de un registro estándar para la identificación de un buen ejemplo. Se sugiere que el contexto elegido no esté lleno de otros términos, cuando se refiere a una palabra especializada, pero también se recomienda fervientemente que no se utilice un registro demasiado coloquial. Ambos polos, para los autores, representan una contrademanda a la naturaleza misma de un buen ejemplo; por tanto, podríamos decir que la inteligibilidad es el uso indicado del registro léxico para cada situación comunicativa.

#### 4.2.4. Rey

El ejemplo en la tradición francesa, resumida certeramente por Rey, es un elemento derivado del análisis semántico, en relación con la definición, que tiene una fuerte carga de ideología cultural (Rey, 1995). Además, tiene mucho más que ver con las realizaciones de habla (*parole*). Con base en los planteamientos de Rey-Debove<sup>19</sup>, agrega, el ejemplo es un elemento metalingüístico que además de funcionar dentro del discurso lexicográfico, también plantea que es en sí un autónimo, es decir, que significa (o connota), además de su propio contenido, características propias del discurso que representa:

D'après J. Rey-Debove — dont je m'inspire ici — , le propre du discours lexicographique est de mettre en rapport, à des fins métalinguistiques, des éléments autonimes, tels l'entrée, et, à des fins didactiques, des éléments de langage non autonimes. L'exemple est lui-même autonome, en ce qu'il signifie (ou connote), outre son contenu propre, une caractéristique du discours qu'il représente. Cependant, ces éléments (mots, termes, locutions...) qui composent l'exemple sont eux-mêmes en usage, ne sont pas autonimes. Sur ce point, les thèses classiques des logiciens, Carnap, Tarski, Strawson, selon lesquels tout énoncé autonome est formé d'éléments autonimes, sont clairement contre battues. (Rey, 1995, p. 103)

Hay que agregar, que, de acuerdo con esta concepción, el ejemplo puede ser entonces un objeto de doble lectura: hacia el significado mismo de la unidad que describe o complementa y hacia el discurso en que está inserta dicha unidad. Sin embargo, cuando el tratamiento es menos cuidadoso, agrega Rey, puede leerse también como un no-autónimo y se tratará de un ejemplo de la constitución de un tipo particular de uso en un discurso.

---

<sup>19</sup> Cabe mencionar que en 1971, dentro de la tradición francesa, Rey-Debove describe minuciosamente las funciones inherentes a la naturaleza del ejemplo lexicográfico y establece una taxonomía para organizarlos (Rey-Debove, 1971). De ahí que Rey se remita a ella como punto de partida para la descripción del ejemplo en la teoría lexicográfica francesa contemporánea. De dicho estudio, sin embargo, no rescatamos todos los presupuestos y nos quedamos con lo que aquí citamos a propósito de lo que Rey dice de Rey-Debove.

#### 4.2.5. Wiegand

En la tradición alemana el ejemplo se retoma poco tiempo después (en 1977 para ser precisos) de la mano de Wiegand, quien lo aborda desde la concepción francesa a propósito de los debates surgidos a principios de esa década (Robles & Sabater, 2011). De acuerdo con este estudio, en etapas posteriores<sup>20</sup>, se deduce que el ejemplo debe ser tomado en cuenta como una parte esencial del diccionario y atiende a criterios de adecuación ya que es un “objeto de estudio de la lexicografía teórica y [se tiene] la obligación de adecuar esta parte esencial de una entrada a las exigencias de los usuarios del diccionario” (Robles & Sabater, 2011, p. 249). De acuerdo con esta misma escuela, el rol principal del ejemplo “es su función, es decir, la finalidad que el lexicógrafo persigue mediante la inclusión del ejemplo en una entrada del diccionario”, además de señalar cuáles son sus condiciones ideales de uso.

#### 4.2.6. Lara

Volviendo a la tradición hispánica, en un estudio ya más reciente, Lara (2006) menciona que el ejemplo en el artículo lexicográfico tiene dos características centrales:

a) consiste de un enunciado que contiene el vocablo *mencionado* en la entrada y

b) introduce ese enunciado para devolver el vocablo de la entrada al *uso*; es decir, recupera el vocablo tal como aparece en la práctica social de la lengua.

---

<sup>20</sup> En los estudios de Hermanns (Das lexikographische Beispiel. Ein Beitrag zu seiner Theorie», en: HARRAS, G. (ed.), Das Wörterbuch: Artikel und Verweisstrukturen. Düsseldorf: Schwann 1988, 161-195) y Harras («Zu einer Theorie des lexikographischen Beispiels», en: HAUSMANN, F. J. et al. (eds.), Wörterbücher. Ein internationales Handbuch zur Lexikographie. Berlin/New York: Walter de Gruyter 1989, vol. 1, 607-614.) para ser precisos. La referencia directa fue tomada del estudio de Robles i Sabater y nos atenemos a su interpretación. Mírese dicho artículo en la bibliografía para mayor información: (Robles i Sabater, 2011)

Para él, el ejemplo en una obra lexicográfica y en una obra enciclopédica cumple funciones distintas. Mientras en el diccionario general su función es la de mostrar al lector algunas maneras en que los vocablos significan en el uso, en la enciclopedia su función es la de explicar hechos del mundo y por ello sirve de herramienta tanto para introducir el vocablo al que se refiere en un contexto explicativo, como para citar alguna explicación a propósito de todo lo que el vocablo mismo refiere, no sólo por lo que significa. En contraste, sin embargo, el lingüista mexicano niega que la función del ejemplo sea complementar la definición o completar la falta de información que da la definición, ya que “por corresponder al conocimiento de las cosas y no del signo, no pudo formar parte de la definición”.

Como podemos ver, son muchas las posiciones acerca de las funciones del ejemplo y sus fuentes, pero casi todas tienen puntos de encuentro o cruces teóricos que nos permiten recuperar ciertos criterios específicos vitales para nuestra investigación. Mencionamos tres de los aspectos más importantes:

1. Los ejemplos pueden partir de fragmentos donde la palabra muestra su aparición común, las concordancias;
2. Algunos fragmentos donde las palabras se insertan, revelan su preferencia hacia ciertas combinaciones con otras piezas léxicas, como en el caso de las colocaciones; y
3. Aun teniendo en cuenta el material que nos arrojan los corpus tales como concordancias y colocaciones, es necesaria la depuración manual atendiendo a los criterios que, como especialista o hablante nativo de una lengua, tiene el lexicógrafo. Este conocimiento, muy variado, y hasta cierto punto subjetivo, comprende conocimiento histórico, contextual, social, habla espontánea no controlada o modas lingüísticas, entre otros.

#### 4.2.7. Fitch

Fitch (2013) ha realizado un estudio a propósito de la ejemplificación en dos diccionarios de lengua general para el español de México atendiendo principalmente a los preceptos de la lingüística de corpus. En este estudio se pone en juego principalmente la pertinencia y el valor de los ejemplos de uso para la mejor comprensión de la definición a la que se asocian.

Para ella, es indispensable establecer parámetros que indiquen la pertinencia y la calidad de los ejemplos. Por ello decide tomar en cuenta cinco rasgos esenciales. A saber:

- 1) La *colocación*, cuyas características toma principalmente de los estudios de Firth (1957) y dos de los continuadores de su línea de pensamiento: Sinclair y Hoey. Es decir, entiende colocación en el sentido de “la palabra o serie de palabras que tienden a aparecer antes o después del lema en una oración” (Fitch, 2013, p. 42).
- 2) La *coligación*, que es una noción tomada de Hoey (2005) donde se entiende que es un “tipo específico de combinatoria léxica en la cual se percibe una marcada tendencia de aparición de vocablos que tienen relación o preferencia gramatical con la palabra nodo. Es decir, es un tipo de colocación en la cual se evidencia la compañía gramatical que mantiene (o evita) una palabra o secuencia de palabras ya sea dentro de su propio grupo o a un nivel más alto; la función gramatical preferida o evitada por el grupo en que participa la palabra o secuencia de palabras; el sitio en una secuencia que una palabra o secuencia de palabras prefiere o evita” (Hoey, 2005, p. 43).
- 3) La *asociación semántica*, noción también tomada de Hoey (2005), cuya característica principal es que “se refiere a las pistas que transmiten ciertas concurrencias habituales que acompañan a una palabra” (Fitch, 2013, p. 44).

- 4) La *pista extralingüística* es una propuesta de Fitch para referirse a las asociaciones semánticas que resultan de otras asociaciones culturales y pragmáticas de un vocablo en alguna variedad dialectal particular. Así, utilizando su ejemplo, “En México, resulta natural asociar a un clavadista con la ciudad costera de Acapulco, y en particular, con la Quebrada, el nombre del acantilado desde donde se arrojan clavadistas”. Para ella, esto “demuestra el carácter enciclopédico o cultural que los ejemplos pueden evocar” (Fitch, 2013, p. 45).
  
- 5) Finalmente, la variable que ha denominado *correspondencia gramatical* se refiere al acuerdo que idealmente debería darse en la categoría de la palabra a ejemplificar cuando está dentro de la definición y cuando está dentro del ejemplo, “como el de poner que *fuereño* es adj. y dar un ejemplo usándolo como sustantivo.” (Fitch, 2013, p. 45).

Estas cinco variables fueron aplicadas en el análisis de 36 entradas para ambos diccionarios, el *Diccionario de Mexicanismos* (Company, 2010) y el *Diccionario del Español de México* (Lara, 2010). Sus resultados sugieren que el mejor ejemplo es una “combinación de la introspección y experiencia de hablante del lexicógrafo y un análisis cuidadoso basado en datos de corpus para verificar que contenga por lo menos uno de los elementos o variables que ayudan a hacer del ejemplo un factor ilustrador confiable y útil”. Es decir, se inclina como hemos visto en otras perspectivas, por la combinación entre fragmentos encontrados en corpus, la aplicación de criterios homogéneos a todo un producto lexicográfico y la adaptación de dichos fragmentos para reflejar cierta información que se desee focalizar.

Valga este pequeño esbozo para mostrar que, aunque sin intención aparente, varios estudios llegan a ciertas conclusiones compartidas y dichas conclusiones son las que más adelante retoman algunos teóricos para describir el comportamiento de esta misma categoría de datos, pero en diccionarios más específicos, que al final es el punto clave de nuestra investigación.

### 4.3. El tratamiento actual del ejemplo en terminología

Los estudios sobre el ejemplo en diccionarios de especialidad han sido muy pocos y su orientación, como mencionamos al principio de este capítulo, es casi siempre hacia las áreas de adquisición de lenguas o traducción. En este apartado abordaremos dos estudios que tomamos como representativos del tratamiento y análisis del ejemplo en diccionarios monotemáticos, terminológicos y monolingües.

Para el caso del tratamiento general del ejemplo en terminografía hemos tomado el estudio de Ahmad, y para el caso del ejemplo en diccionarios terminológicos en español, el estudio de Fuentes Morán y García Palacios.

#### 4.3.1. Ahmad

El trabajo de Ahmad et al. (1992) se centra específicamente en la construcción de ejemplos referidos a términos y uno de los puntos clave es que un término, o una palabra relacionada de manera cercana a él, proyecta un contexto donde se pueden encontrar diferentes clases de información. En la mayoría de los casos el término es utilizado como parte de una cita, pero la información que lo rodea es rica en datos contextuales, lo que la hace un elemento importante para aportar rasgos acerca del concepto. A este tipo de fragmento textual se ha llamado *contextual examples* y se hace la nota aclaratoria de que en otros estudios a esta misma estructura se le denomina sólo *citation*<sup>21</sup>. Para mirar mejor esto, en un estudio posterior de Bergenholtz y Tarp (1995, p. 137-143) se hace la distinción entre *citations* (fragmentos sin adaptación), *citations examples* (manipulados parcialmente) y *competence examples*, que son los que encontraremos como “ejemplos inventados”, es decir, que son diseñados por el lexicógrafo. De aquí sólo rescataremos el hecho de que el término a ejemplificar aparece comúnmente dentro del fragmento elegido como una parte imprescindible de éste para conformar un criterio de selección.

---

<sup>21</sup> Por ejemplo en (Zgusta, 1971)

Hay que notar, sin embargo, que la anterior es una perspectiva tomada de la tradición lexicográfica y por tanto una adaptación de ella. La propuesta de ejemplificación parte de la noción de *lexicografía especializada* y no de la terminología como tal. Se trata, si lo miramos cronológicamente, de los estudios sobre ejemplificación hasta mediados de la década de los noventa.

#### 4.3.2. Fuentes Morán y García Palacios

Fuentes Morán y García Palacios (2002) hacen hincapié en la necesidad de los ejemplos en diccionarios de especialidad y en el tratamiento profundo de la adecuación para elegir los más apropiados como parte de la tipología de información que este tipo de diccionarios deben ofrecer (que aquí llamamos categorías de datos: registro, equivalentes, sinónimos, pronunciación, etc.). Del análisis que hacen los autores reproducimos un conjunto de características que el ejemplo debería cumplir y que están basadas, además, en lo dicho anteriormente por Gutiérrez Cuadrado (1999, p. 81) y Drysdale (1987, p. 213). A saber, que el ejemplo:

- Complementa información de la definición
- Muestra la palabra en su contexto
- Distingue diferentes acepciones
- Muestra colocaciones típicas
- Indica registros y niveles estilísticos
- Muestra diferentes cuestiones gramaticales
- Incluye ciertas orientaciones ideológicas
- Incluye información enciclopédica
- Da información sobre costumbres y realidades específicas

Hasta aquí, podemos ver que las características de un ejemplo van acompañadas por una reflexión profunda sobre su utilidad pero, como los mismos autores de los dos estudios mencionan, el interés teórico en cuanto a la tipologización de estas unidades aún no se ha dado de manera sistematizada. Dicha tipología, creemos, es necesaria para poder establecer las características que se acaban de enumerar a la par de distinguir en la

práctica cuándo dichas características son activadas y se formalizan en una estructura que podemos llamar ejemplo. La enumeración de funciones, fuentes de origen, forma y posición que ellos proponen, sin embargo, arroja sólo una luz media sobre el papel que los ejemplos juegan y la relación que guardan con la definición. Por lo anterior, creemos que es necesario un análisis desde una perspectiva distinta que privilegie la aplicación de estos criterios a la par de reestructurar y armonizar una posible tipologización de los ejemplos asociados a definiciones terminológicas.

#### **4.4. Caracterización teórica de un ejemplo**

Hemos visto que los ejemplos cumplen funciones específicas que han sido tratadas desde diversas perspectivas y que básicamente ayudan al lector a situarse en un contexto más amplio desde donde pueden observar el campo semántico al que pertenece una palabra y sus diversos usos y acepciones. En terminografía esta posición no cambia y, de hecho, es más productiva, pues el tratamiento independiente de la relación entre concepto y término, y entre términos y definiciones, permite aprehender un concepto desde un camino más natural como lo es la ejemplificación. Todos aprendemos por similitudes y ejemplos antes que por definiciones. La leche, antes de ser “un producto lácteo extraído de la vaca”, es ese alimento que se le ofrece al niño acompañado de chocolate en polvo, por ejemplo<sup>22</sup>.

##### **4.4.1. Funciones**

*Adición de información.* Ante las restricciones que hemos mencionado acerca de la definición, el ejemplo cumple una función complementadora al agregar información sobre el significado de una palabra al evitar la estructura estática de aquella. En el ejemplo podemos encontrar información de tipo contextual, sobre el significado en tanto elemento lingüístico, al colocar una palabra en un uso específico, como menciona Lara. Desde otras perspectivas, incluye también información enciclopédica que explica el uso de

---

<sup>22</sup> Ejemplo tomado parcialmente de Lara (2001)

esa palabra en el léxico general y lo enlaza con la realidad donde la pieza léxica es utilizada en conjunto con los elementos que la rodean. Muestra además particularidades gramaticales al mostrar la forma más usual del paradigma propio de la categoría a la que pertenece la palabra ejemplificada y recupera rasgos semánticos que no son abarcados por la definición debido a criterios del creador o los creadores de la definición.

*Distinción de acepciones.* El ejemplo puede mostrar variaciones conceptuales al estar ligado a definiciones concretas. Así, dependiendo de la estructura definitoria de la que se desprende, el ejemplo puede mostrar esas particularidades teóricas o de uso que los hablantes otorgan en el momento de su uso bajo distintas situaciones comunicativas.

*Muestra viva.* Desde esta perspectiva, basada en corpus, el ejemplo representa una muestra viva del uso de palabras contenidas en el diccionario, pero también de las relaciones que mantiene con otras palabras. Si tenemos en cuenta que el uso es el que motiva el cambio lingüístico en cualquier lengua, entonces el ejemplo es la herramienta principal de análisis donde se puede constatar la evolución que va teniendo una palabra y que finalmente, de concretarse el cambio, será el primero que cambie. El ejemplo, antes que la definición o cualquier otra categoría de datos, será el testigo de las variaciones contextuales, de pronunciación, de sintaxis o discursivas.

*Orientación ideológica.* Como hemos mostrado antes, el ejemplo no depende sólo del uso que tiene una palabra en el léxico general, sino que también a través del proceso de selección puede mostrar la inclinación teórica o prescriptiva de los creadores del diccionario. Por una parte da fe de las preferencias que el hablante de una lengua tiene con respecto a la acepción de dicha palabra y por otro lado da fe de la constatación o refutación de dichas preferencias al momento de ser documentada y vaciada en un diccionario.

#### 4.4.2. Forma

*La cita.* Es quizá la forma más antigua de ejemplificación en los estudios lexicográficos con base en corpus. La cita representa al

mismo tiempo una prueba de uso de una palabra así como un elemento prescriptivo. Con la citación, como vimos arriba, se documenta de manera sistemática y precisa la aparición de una palabra en contextos específicos, ya que sus fuentes son casi siempre obras publicadas. Estas fuentes son de diferentes registros, sobre todo cultos: obras literarias, prensa, tratados científicos, etcétera. Este último rasgo les da su carácter prescriptivo. Una desventaja de la cita es, sin embargo, que puede no mostrar el uso actual de la palabra a la que se refiere, debido a que el periodo temporal entre la emisión de la obra y su documentación para la confección del ejemplo puede distar bastante. Baste ver algunos ejemplos de los diccionarios de la RAE:

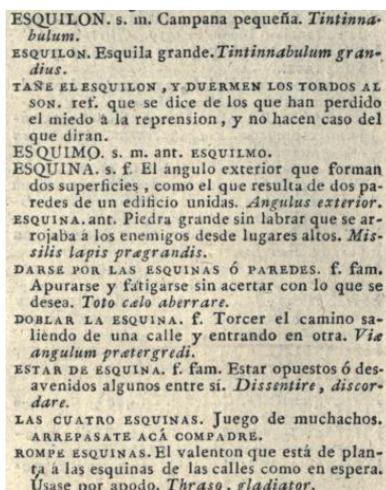


Figura 4.3. Captura de pantalla DRAE, 6ª Ed. 1822

*El contexto.* Utilizar un contexto como ejemplo fue hasta bien entrado el s.XX una práctica común, pero debido a la noción imprecisa de lo que es un contexto en lingüística, se pueden hallar dificultades metodológicas. Es decir, un contexto puede ser una concordancia, pero también puede ser una colocación, e incluso puede ser una cita. Estudios actuales han hecho reflexiones profundas sobre la naturaleza y la definición misma de contexto para diferenciarlo de otras estructuras lingüísticas (Van Dijk, 2012), aun así es arriesgado proclamar que “contexto” es igual a “ejemplo”, sobre todo cuando las relaciones entre éste y la palabra que ejemplifica no son del todo claras. El contexto, como

desventaja, es que no crea relaciones demasiado estrechas con la definición de la palabra que, valga la redundancia, contextualiza.

*La colocación.* Desde que Firth las describiera en 1957 como “actual words in habitual company” (Palmer, 1968), hasta ahora las colocaciones han sido una herramienta indiscutible en la conformación de ejemplos debido a sus características lingüísticas, ya que muestra combinaciones lingüísticas frecuentes, fijadas en el uso (y en la norma) que sin embargo no llegan a ser frases totalmente lexicalizadas. Esta doble naturaleza, si nos fijamos detenidamente, coincide con la concepción de ejemplo que se tiene en los estudios arriba referidos. Actualmente las colocaciones de una palabra pueden ser fácilmente extraídas por medios computacionales a través de medidas de asociación como la información mutua (Ward Church & Hanks, 1990). Sin embargo, hay que tener en cuenta, nuevamente, que estas estructuras guardan una relación estrecha con el uso de una palabra en el léxico común, pero su relación con un significado específico no se puede deducir nítidamente<sup>23</sup>. Las colocaciones son candidatas fuertes para la conformación de ejemplos, pero tienen que ser analizadas desde el significado de cada una de las piezas que la conforman para comprobar si se trata efectivamente del uso de la palabra con la acepción que se desea ilustrar.

*El ejemplo “inventado”.* A este tipo de ejemplos se les ha llamado con diferentes nombres dependiendo de la corriente lexicográfica que los trate. Así, podemos encontrar denominaciones como *ejemplos cita*, *ejemplos de uso*, *ejemplos contruidos ad hoc* (Fuentes Morán & García Palacios, 2002) y, como mencionamos arriba, también *constructed examples* (Zgusta, 1971). El punto que aquí nos interesa mostrar es que se trata de construcciones basadas en el conocimiento del lexicógrafo acerca de la unidad que desea ejemplificar a través de la comprobación de sus características semánticas en los textos o corpus que tiene a su alcance. Algunas veces son fragmentos totalmente nuevos y otras tantas son adaptaciones de fragmentos encontrados en las fuentes documentales.

---

<sup>23</sup> Sobre la adopción de un significado específico de una palabra a través de colocaciones mírese el interesante artículo de Jaqueline León “Meaning by collocation: The Firthian filiation of corpus linguistics” (León, 2007)

*Los equivalentes y sinónimos.* Son la forma más sencilla y directa de ejemplificar el uso de una palabra, pero los más imprecisos y menos ricos en información. Muchas veces son remisiones a palabras de un registro más culto o estándar, cuando se trata de sinónimos. El caso de los equivalentes es más usado en diccionarios bilingües debido a la precisión que supone otras palabras del mismo campo semántico en la lengua original, cuando la segunda lengua no cuenta con una pieza léxica asociada al significado o concepto de una manera suficientemente informativa.

#### 4.5. Los ejemplos definatorios oblicuos

Los *ejemplos definatorios oblicuos* son una noción introducida por Lara (2006, p. 145) y hace referencia a aquellas estructuras en que “el vocablo cuyo uso se ejemplifica forma parte de definiciones de otros objetos”. Si miramos los siguientes casos tomados de ese mismo artículo:

s.v. **hojuela**: La madera contrachapada está constituida por capas de *hojuelas*. [DEM, 1ª. Red.]

s.v. **espía**: Matahari fue una *espía* famosa. [DEM, 1ª. Red.]

s.v. **perenne**: Los árboles y los arbustos son plantas *perennes*. [DEM, 1ª. Red.]

s.v. **fascismo**: Los militantes del *fascismo* se denominaban ‘camisas negras’. [DUEAE]

veríamos que lo que se coloca después de los vocablos como definición en realidad son ejemplos y forman parte de las definiciones de la madera contrachapada, de Matahari, de árboles y arbustos y de los camisas negras, respectivamente.

El autor menciona que este tipo de ejemplos hacen referencia a una entrada en uso, pero cuyo valor es esencialmente enciclopédico, quizá como una intención del lexicógrafo de mejorar el conocimiento del mundo por parte de sus lectores. Su propuesta es no incluirlos como definiciones sino asociarlos a los vocablos a los que se refieren ya como una subentrada, ya como un ejemplo tal cual o una acepción.

En nuestro caso volveremos sobre este tema en el próximo capítulo para mostrar que los ejemplos definitorios oblicuos forman parte de una esfera, muy pequeña por cierto, de vocablos o términos que basan su significado en el conocimiento común, cultural y social del concepto al que se refieren. Es decir, son palabras que se definen por su uso antes que por su significado. La existencia de este tipo de ejemplos, sobre todo en terminología, comprueba que las realizaciones del concepto no siempre pueden ser aprehendidas por la definición terminográfica y por consiguiente echan mano de estructuras alternas que consideran como inmediatas. Si tenemos en cuenta que, de acuerdo con lo que hemos dicho, hay una relación lineal de codependencia entre concepto-término-definición-ejemplo, no será complicado entrever que cuando uno de los elementos de esa cadena es insuficiente, los elementos adyacentes intenten conservar la cohesión a través de la asunción del rol del elemento que sufre dicha insuficiencia. En 5.3.3. ahondaremos sobre este caso.

#### **4.6. Discusión: Un ejemplo siempre ha sido un recurso, pero no un elemento obligatorio**

Hasta aquí hemos mostrado que el ejemplo siempre ha acompañado a la tradición lexicográfica y que ha permeado también en el trabajo terminológico. Se trata de un elemento imprescindible en casi todos los estudios generales sobre la confección de diccionarios y su importancia no se pone en tela de juicio. Hay que notar, sin embargo, que la realidad dista mucho de la teoría pues, también históricamente, el ejemplo se ha ido relegando a un estado de categoría menor, en tanto la definición y las acepciones han obtenido el mayor interés de los especialistas.

Para mostrar lo anterior, hemos descrito un estudio hecho a propósito de la información encontrada en diccionarios terminológicos y generales en referencia a un conjunto de términos y los resultados han sido contundentes. La tendencia a no usar ejemplos asociados a los vocablos que conforman las entradas de un diccionario es evidente. De hecho, si nos detenemos a pensar por simple intuición o con un conocimiento mínimo sobre diccionarios, caeríamos en la cuenta de que esta afirmación no sólo es verdadera,

sino generalizable. Diccionarios como el DRAE, uno de los de mayor referencia mundial con respecto al español, no tiene una metodología sistemática en cuanto al tratamiento del ejemplo en ninguna de sus ediciones. Los diccionarios terminológicos apenas muestran interés en esta categoría de datos.

De este mundo de diccionarios hay que hacer notar algunos que desde su gestación han tenido en cuenta todo lo que decimos anteriormente y la aparición de ejemplos es una constante en sus ediciones. De entre ellos podemos contar el Diccionario de Uso de María Moliner, el Diccionario del Español de México coordinado por Lara y el propio COBUILD de Sinclair como prueba histórica.

Por todo esto, es posible afirmar que aunque en la teoría el ejemplo ha tenido un peso importante, lo cierto es que en la práctica no se hace palpable de ninguna manera. Quizá se deba a cuestiones externas a los propios materiales lexicográficos y terminológicos, como el espacio en el papel o los recursos económicos, pero, como ya lo apuntamos antes, esas limitaciones se van difuminando cada día más y ya no son obstáculo para que el ejemplo retome la importancia que antaño, como en el diccionario de autoridades, tenía.

No se quiere decir aquí que la técnica lexicográfica o terminográfica deba volver a esos inicios, sino que si hasta ahora hemos visto que los estudios muestran todas las ventajas que tiene el ejemplo para la confección de materiales con mayor cantidad, quizá sea hora de declarar como obligatoria una categoría de datos que incluso tome la importancia de la definición en casos en que el significado o el concepto escapen a la aprehensión que de ellos hacemos con el lenguaje.

Como un primer acercamiento a esta premisa de obligatoriedad como complemento, y como un primer esfuerzo de caracterizar el ejemplo desde el discurso especializado, hemos retomado todos los argumentos descritos en los capítulos anteriores, y en este mismo, para diseñar una propuesta tanto de forma como de contenido del ejemplo que se podría utilizar sistemáticamente en terminografía como complemento a la definición.



## 5. EL EJEMPLO EN TERMINOLOGÍA, UNA PROPUESTA

En este capítulo desarrollamos una propuesta sobre la caracterización del ejemplo en terminología. Nos basamos en el marco teórico descrito en el capítulo anterior acerca del ejemplo en diccionarios generales y algunos acercamientos al ejemplo en diccionarios especializados. Aquí perfilamos una propuesta basada, en primer lugar, en un experimento hecho con la suite de herramientas de *Sketch Engine* para determinar las condiciones formales y sintácticas de un candidato a ejemplo; en segundo lugar analizamos los criterios a cumplir acerca de ciertos requerimientos semánticos a propósito de su relación con la definición de un término; y finalmente aplicamos noción teórica de *saturación semántica* y la medida de *densidad lexicométrica* con el fin de evaluar la cercanía semántica a la definición y la pertinencia a propósito de la adecuación de los posibles candidatos a ejemplo que se encontrarían siguiendo esta metodología.

### 5.1. Experimento SKE

Un experimento llevado a cabo en 2014 con la suite de herramientas de *Sketch Engine* (SKE en adelante)<sup>24</sup> nos proporcionó los datos necesarios para reafirmar los preceptos teóricos de los capítulos anteriores y para llegar a la heurística que pudiese funcionar en el diseño de un algoritmo inicial de extracción de ejemplos.

La metodología consistió en varias técnicas de extracción de fragmentos textuales de manera automática (detallamos adelante) y las aplicamos recursivamente a fragmentos que, en principio, solo tenían dos criterios:

1) que el término que se deseaba ejemplificar estuviese presente en dicho fragmento, y

---

<sup>24</sup> Presentado como ponencia en el XIV Simposio Iberoamericano de terminología, RITerm 2014 (Lázaro, J; Sierra, G; Kilgarriff, A; Cabré, T., 2014).

2) que el sentido del fragmento correspondiera con el uso dentro de un área de especialidad (a través de una revisión manual, con una definición en mente).

Estos fragmentos, a los cuales llamamos de manera genérica “frases”, nos permitieron trabajar con una gran cantidad de datos que después fuimos restringiendo. Como hemos mencionado en el capítulo 1, trabajar desde adentro hacia fuera resulta improductivo, esto es, tratar de extraer colocaciones de un corpus especializado; por ello decidimos partir de un corpus general, el *esTenTen* (Kilgarriff & Renau, 2013), con el fin de ir aplicando restricciones a los fragmentos que obtuvimos desde concordancias, después colocaciones a partir de esas concordancias, después colocaciones complejas, etcétera, con las mismas opciones que proporciona SKE. Los criterios de elección de dichos fragmentos, basados en lo que hemos mencionado anteriormente, fueron estos<sup>25</sup>:

- Que complementara información de la definición.
- Que mostrara un término en un contexto de uso.
- Que mostrara colocaciones típicas.
- Que indicara otros registros y niveles estilísticos.
- Que mostrara diferentes accidentes gramaticales.
- Que, idealmente, fuera ideológicamente neutro.
- Que incluyera información enciclopédica (más descriptivo que definitorio).

Con los quince términos de tres áreas temáticas distintas que elegimos (ver Tabla 4.1.), se aplicó el siguiente flujo de trabajo. Veamos el caso de *virus* del área de informática. Dada la definición: “*Programa elaborado para alterar y dañar el funcionamiento normal de una computadora; virus cibernético*” [Diccionario del Español de México, DEM] (Lara, 2010).

---

<sup>25</sup> Esta lista es una adaptación de los criterios que se pueden encontrar en (Fuentes Morán & García Palacios, 2002) y (Atkins & Rundell, 2008)

## I. Extraer todas las concordancias donde el término aparecía

Query <b>virus</b> 389,944 (34.9 per million)	
Page 1	of 19,498 <a href="#">Go</a> <a href="#">Next</a>   <a href="#">Last</a>
#77706	de cava. En los años 90 d.N. (Después del Nintendo), un <b>virus</b> llamado Nintenditis aguda se propaga por el universo hasta
#257303	otro organismo. El vehículo utilizado en la transferencia es un <b>virus</b> incapaz de realizar ciclos infectivos, pero también capacitado
#342201	hipótesis de que el scrapie de las ovejas estaba causado por un <b>virus</b> lento. Esta enfermedad típica del ganado ovino y caprino se
#342344	reforzándose la hipótesis de que el kuru estaba causado por un <b>virus</b> lento que colonizaba el cerebro de las víctimas. Muchos años
#342837	Este concepto echaba por tierra las viejas hipótesis de los <b>virus</b> lentos de Hadlow y Sigurdsson, los virus no convencionales
#342845	viejas hipótesis de los virus lentos de Hadlow y Sigurdsson, los <b>virus</b> no convencionales de Gajdusek, los virinos, los viroides y
#342916	enfermedades priónicas. Prusiner demostró que los priones no eran <b>virus</b> , tanto por sus características infectivas como por su estructura
#372245	es una enfermedad frecuente causado por la reactivación del <b>virus</b> de la varicela zoster, un virus de la familia de los herpesvirus
#372252	causado por la reactivación del virus de la varicela zoster, un <b>virus</b> de la familia de los herpesvirus que es morfológicamente idéntico
#372264	familia de los herpesvirus que es morfológicamente idéntico al <b>virus</b> de la varicela pero antigénicamente distinto. De igual forma
#372282	distinto. De igual forma, el cuadro clínico producido por el <b>virus</b> de la varicela (cuyo acrónimo es VVZ) es diferente del producido
#372298	varicela (cuyo acrónimo es VVZ) es diferente del producido por el <b>virus</b> del herpes zoster (VHZ). Después de un cuadro de varicela
#372314	herpes zoster (VHZ). Después de un cuadro de varicela, el <b>virus</b> permanece latente en los ganglios radicales sensitivos a
#372539	lesión ocular puede estar causada por una invasión directa de los <b>virus</b> o por difusión del virus desde el ganglio trigémino ( * ) a
#372544	causada por una invasión directa de los virus o por difusión del <b>virus</b> desde el ganglio trigémino ( * ) a través del nervio nasociliar
#373598	zoster, la vacuna ZOSTAVAX contra el Herpes Zoster a base de <b>virus</b> vivos atenuados, es desde hace poco, el único tratamiento eficaz
#504934	extiende a un amplio grupo de microbios, que engloba bacterias, <b>virus</b> , hongos y parásitos, al bloquear la interacción entre las células
#690081	Conserva el marco original. Secuenciadas más de 200 cepas del <b>virus</b> de la gripe El Proyecto de Secuenciación del Genoma de
#690123	determina la secuencia genética completa de 209 cepas distintas de <b>virus</b> de la gripe que afectan al ser humano, información que ayudará
#690145	ayudará a los científicos a comprender mejor cómo evolucionan los <b>virus</b> de la gripe, cómo se propagan y como causan enfermedad.
Page 1	of 19,498 <a href="#">Go</a> <a href="#">Next</a>   <a href="#">Last</a>

Figura 5.1. Concordancias del término *virus*

Como podemos ver, las concordancias más comunes en términos cuantitativos son las que tienen que ver con el término *virus* en el área de biología, por tal motivo, las concordancias simples fueron descartadas como posibles candidatos a ejemplos. Si bien muestran el término en un contexto de uso, no pueden reflejar certeramente que se use en el área de informática. Lo anterior nos llevó a reflexionar en cómo delimitar el área de especialidad para que la extracción de concordancias fuese productiva en un discurso específico.

Aquí es donde entra en juego la propuesta de Kuguel (2007): Si un término es activado por sus condiciones sintácticas específicas y las clases de palabras alrededor de él determinan su cualidad de pieza léxica especializada, entonces aquellas frases que recurrentemente tuviesen una clase particular de palabra que proyectase una estructura específica, podrían ser mejores candidatos; serían las estructuras activadoras del término. Para poder determinar qué clase de palabra activaba esa función, decidimos hacer una extracción más detallada.

**II. A partir de las concordancias anteriores, hacer un análisis con Wordsketch**, esto es, determinar las relaciones sintácticas del término con las palabras que lo rodean con el fin de conocer cuáles son las estructuras más comunes asociadas a él:

virus (noun)														
esTenTen11 (Eu + Am, Freeling, Lempos) freq = 389,244 (34.9 per million)														
object of	50,034	-0.7	subject of	37,929	-1.3	n_modifier	100,690	-0.7	modifies	37,601	-0.3	y_o	16,432	-0.6
contraer	2,701	9.12	mutar	422	8.23	influenza	12,321	11.17	subtipo	316	7.77	bacteria	3,371	9.69
propagar	667	8.25	infectar	374	7.74	inmunodeficiencia	7,693	10.99	cepa	768	7.75	troyano	557	9.36
contagiar	391	7.49	arn	74	5.98	hepatitis	7,226	10.68	propagación	532	7.44	gusano	586	8.53
aislar	490	7.41	atacar	396	5.93	gripe	8,823	10.41	detección	925	7.0	parásito	412	8.1
detectar	2,196	7.28	circular	342	5.9	isa	3,459	10.01	brote	399	6.5	hongo	467	7.55
transmitir	1,704	7.22	penetrar	177	5.89	sincicial	1,112	8.48	replicación	131	6.43	spam	268	7.33
inocular	234	7.13	hanta	60	5.69	informático	4,747	8.48	contagio	241	6.3	viroides	67	7.05
portar	671	6.77	adn	46	5.19	papiloma	1,171	8.4	escaneo	118	6.28	hackers	118	6.89
diseminar	190	6.73	contagiar	57	4.98	respiratorio	2,288	8.11	creador	452	6.25	protozoo	58	6.69
eliminar	1,817	6.7	isa	35	4.88	pandémico	869	8.07	serotipo	95	6.08	spywares	47	6.53
combatir	889	6.62	sobrevivir	112	4.8	gripal	838	7.98	ataque	1,148	5.87	malwares	47	6.52
erradicar	326	6.29	evolucionar	78	4.54	causante	856	7.82	vector	149	5.86	microbio	59	6.3
inactivar	125	6.28	permanecer	265	4.52	rabia	912	7.59	programador	121	5.86	antivirus	79	6.06
neutralizar	162	5.98	persistir	83	4.5	hanta	549	7.47	vacuna	371	5.84	phishing	42	5.95
esparcir	136	5.96	pertenecer	149	4.38	hiv	530	7.41	mutación	162	5.83	microorganismo	130	5.9
contener	1,945	5.9	reaparecer	38	4.36	fiebre	959	7.13	eliminación	437	5.77	priones	32	5.9
hanta	88	5.84	poder	4,254	4.25	rábico	420	7.09	genoma	112	5.69	germen	94	5.79
excretar	89	5.8	afectar	399	4.18	troyano	424	7.06	diseminación	89	5.68	patógeno	50	5.7
transmitir	140	5.77	ingresar	229	4.18	sincitial	398	7.01	busca	348	5.55	protozoarios	25	5.54
inyectar	133	5.68	invadir	85	4.13	circulante	406	6.84	infección	519	5.47	intruso	37	5.41
circular	306	5.65	entrar	378	3.99	viruela	372	6.84	aislamiento	169	5.27	adenovirus	23	5.41
mutar	82	5.53	koofrace	18	3.95	letal	481	6.84	desinfección	70	5.21	nematodo	23	5.25
atacar	261	5.24	soler	141	3.94	varicela	356	6.78	epidemia	137	5.17	espia	40	5.08
incubar	62	5.14	causar	342	3.85	mortal	659	6.78	proliferación	133	5.13	toxina	55	5.08
temer	155	5.11	vacunar+le	17	3.84	vivo	1,416	6.68	transmisión	512	5.11	retrovirus	18	5.03

**Figura 5.2. Estructuras más comunes asociadas al término *virus***

En esta figura podemos ver que la herramienta nos muestra los verbos más comunes asociados a *virus* en la columna “object of” cuando el término es el objeto directo o indirecto (por ejemplo, *contraer un virus*); los verbos más comunes cuando el término es el sujeto de dichos verbos en “subject of” (*virus muta*); los elementos que modifican al término en una relación adjetiva en “n\_modifier” (por ejemplo: *virus de la influenza*); también en función adjetiva pero en posición anterior en la columna “modifies” (*subtipo de virus*) y, finalmente, aquellos elementos potencialmente intercambiables en una relación de coordinación en la columna “y\_o” (por ejemplo: *virus y bacterias / virus o bacterias*).

De esas columnas decidimos utilizar aquella que coloca al término como objeto del verbo (“object of”), ya que así podemos asegurarnos de que guardan una relación estrecha de predicación. Para el caso de *virus*, los verbos con mayor relación dentro de este corpus fueron *contraer* y *propagar*. Si miramos las concordancias de *virus* aunado a ambos verbos, veremos que sucede algo interesante:

Word sketch item 2,701 (0.2 per million)

Page 1 of 136 [Go](#) [Next](#) | [Last](#)

#1821091	entre los cinco y los diez años después de haber <b>contraído</b> el <b>virus</b> , y en ese tiempo pueden, sin saberlo, haberlo transmitido a
#3362623	un buen hogar. </p><p> ¿Cuáles son las ventajas de <b>contraer</b> el <b>virus</b> del SIDA?. </p><p> La compresión y cariño de toda la sociedad
#3376016	quedarse en casa inhalando tizas (forma de la cual se <b>contrae</b> el <b>virus</b> ). Nótese que el virus es tan astuto (nótese, un virus astuto
#8132495	</p><p> Hasta 5,7 millones de estadounidenses han <b>contraído</b> el <b>virus</b> A/H1N1 entre abril y finales de julio, según un estudio realizado
#21573551	</p><p> Cada día aumenta el número de personas que <b>contraen</b> el <b>virus</b> en todo el mundo. Los datos sobre el contagio del VIH son alarmantes
#26154280	lectores les sorprenderá saber que el riesgo de <b>contraer</b> un <b>virus</b> como el que causa la hepatitis C (VHC) en el medio sanitario
#27838500	pero tuvo que dejar el trabajo después de haber <b>contraído</b> el <b>virus</b> . A partir de entonces apenas se mantuvo con una pensión equivalente
#40712781	35.000 están aún sin diagnosticar. Como desde que se <b>contrae</b> el <b>virus</b> hasta que aparecen los primeros síntomas de sida puede pasar
#41579029	cacheo de toxicomanos (sobre todo por no <b>contraer</b> el temible <b>Virus</b> del Sida), ni chalecos anti-balas anti-pícazos, ni una insignificante
#63693826	bastante simples por lo general, que ayudan a no <b>contraer</b> el <b>virus</b> , y por otro en la mayor difusi/> e la informaci/> isponible
#63983330	minimizando con ello los riesgos de infecci/> , caso de <b>contraer</b> el <b>virus</b> , adelantando el plazo de curaci/Acaso ha salido la Ministra
#64112849	VIH de la OMS explican además que los riesgos de <b>contraer</b> el <b>virus</b> a una edad avanzada son mayores que si se adquiere en la juventud
#70254471	<p> Cualquier individuo que corra un alto riesgo de <b>contraer</b> el <b>virus</b> de la hepatitis B debe recibir la vacuna correspondiente. En
#73180720	menores de 5 aro los ni/> m/> mayores tambi/> pueden <b>contraer</b> el <b>virus</b> y transmitirlo/> . </p><p> MSF ha decidido apoyar al sistema
#73186808	comienzan el tratamiento con antirretrovirales, cinco <b>contraen</b> el <b>virus</b> . </p><p> Por su parte, el director ejecutivo del Fondo Mundial
#73240620	er que al menos 167 personas han muerto ya tras <b>contraer</b> el <b>virus</b> . Se han registrado adem/> otros 400 casos. </p><p> La directora
#76408581	en cuenta", protesta Valentina. Su teoría es que <b>contrajo</b> el <b>virus</b> "más por falta de información que por otra cosa. No creo que
#83674845	infección. Los pequeños (mayores de cinco años) <b>contraen</b> el <b>virus</b> , lo contagian en el eje familiar (progenitores entre 30 y 39
#87224829	pacientes con problemas respiratorios que hayan <b>contraído</b> el nuevo <b>virus</b> , tal y como explica el Dr. Molero "a todo paciente que llegue
#89776963	las personas no vacunadas o mal vacunadas, pueden <b>contraer</b> el <b>virus</b> durante un viaje a un país en el que la higiene y la salubridad

Page 1 of 136 [Go](#) [Next](#) | [Last](#)

Figura 5.3. Estructuras más comunes asociadas al término *virus* con el verbo *contraer*

Word sketch item 667 (0.1 per million)

Page 1 of 34 [Go](#) [Next](#) | [Last](#)

#2524602	aplique todo el peso de la Ley contra los que crean y <b>propagan virus</b> por Internet, afectando tanto a Empresas como a usuarios particulares
#51370643	psicópatas únicamente interesados en continuar <b>propagando</b> el <b>virus</b> sin tener en cuenta su propia integridad física. No serán controlables
#75668916	información con nuestro portátil infectado, ya que este <b>propagaría</b> el <b>virus</b> por todos los equipos informáticos de la empresa pudiendo llegar
#78890112	/H1N1, Fukuda destacó la rapidez con que se <b>propagó</b> el nuevo <b>virus</b> , a una velocidad "sin precedentes", según Fukuda, y la mayor
#80522667	un servidor central del hospital sin el riesgo de <b>propagar virus</b> o verse inundados de spyware. </p><p> Dos meses después de la
#98659974	nacional es acusado por las autoridades sanitarias de <b>propagar</b> un <b>virus</b> letal que ha atacado a la sociedad española. </p><p> El estreno
#100570622	muerte de su hija en Albania. Descubre que FarCorp <b>propagó</b> el <b>virus</b> en el país a través de una persona con el nombre en clave de
#105161776	autobuses: uno al norte y otro al sur, y en ellos <b>propagó</b> el <b>virus</b> . Por cierto, después de ver el episodio uno se pregunta, ¿cómo
#106019738	no haya quien, de manera consciente, se dedique a <b>propagar</b> el <b>virus</b> como si de un arma bacteriológica se tratara. Sea por esto
#128637223	demasiado pronto para determinar con qué rapidez se <b>propaga</b> el <b>virus</b> de la gripe porcina. </p><p> La muerte de este paciente, sin
#155904823	titular de Sanidad, "es la mejor garantía de que no se <b>propague</b> el <b>virus</b> ". </p><p> "Posponer" los viajes a México </p><p> La ministra reiteró
#165246138	pero eficaz, que desborda pasión por los libros y <b>propaga</b> el <b>virus</b> de la lectura entre los más pequeños. </p><p> Ramón </p><p> Jesús
#171184034	King relata con precisión esos 19 días en los que se <b>propaga</b> el <b>virus</b> de la que él llama la supergripe o el Capitán Trotamundos.
#172064367	aún no ha desarrollado los síntomas. </p><p> ¿Cómo se <b>propaga</b> el <b>virus</b> VIH? </p><p> El VIH, virus del SIDA, se transmite de una persona
#202025523	Prefieren la suya propia. Pues bien: nosotros ayudamos a <b>propagar</b> el <b>virus</b> de este saludable inconformismo entre los usuarios de nuestras
#254123035	el mundo físico del mismo modo que una computadora <b>propaga</b> un <b>virus</b> por las redes cibernéticas como afirmara uno de los fundadores
#273297041	abrirlo, y así, repetir su ciclo de vida. </p><p> ¿Como <b>propagar</b> un <b>virus</b> ? </p><p> Primero que nada, se debe ser un Hacker profesional,
#287706609	forma de una organizaci?n bioterrorista que trata de <b>propagar</b> el <b>Virus</b> Gemini. La misi?n del usuario ser? viajar a cualquier lugar
#298276179	esencial para descubrir la función que desempeñan para <b>propagar virus</b> en la población humana", afirma Shawn Babliuk, director de la
#301957902	parte del principio en la que describe como se va <b>propagando</b> el <b>virus</b> por el mundo. Creo que de llegar a ocurrir una gran epidemia

Page 1 of 34 [Go](#) [Next](#) | [Last](#)

Figura 5.4. Estructuras más comunes asociadas al término *virus* con el verbo *propagar*

Cuando el término *virus* se relaciona con el verbo *contraer*, como vemos en la figura 5.3., los contextos muestran el uso de ese término casi exclusivamente en el área de biología; sin embargo, cuando ese mismo término se relaciona con el verbo *propagar*, la mayoría de los contextos muestran su uso en el área de informática.

Esto nos lleva a pensar que las relaciones establecidas entre un término y los verbos asociados mantienen una relación temática. Mírese detenidamente que los contextos donde *virus* se combina con *contraer*. Veremos que aparecen otros términos como AH1N1, SIDA, Hepatitis C, infección, OMS, etcétera; mientras que en los contextos donde *virus* se combina con *propagar* aparecen términos como internet, [ordenador] portátil, spyware, etcétera. Es decir, un verbo asociado a un término proyecta el campo semántico del término y muestra su uso. Depende de ese verbo que los contextos donde se combinan ambos (verbo y término) funcionen a su vez como proyecciones de las relaciones que el concepto mantiene con otros conceptos y términos de su misma área de especialidad.

Hay que notar, sin embargo, que si bien la primera combinación recupera indiscutiblemente fragmentos que tienen que ver con biología, la segunda combinación, que es la que nos importa, no es tan definitiva, y la lista de fragmentos navega entre ambos campos de conocimiento. Ante esta dificultad, los criterios de recuperación de los fragmentos tuvieron que ser revisados. El resultado fue que por ser una herramienta automática, el *Wordsketch* puede ponderar la cercanía de los verbos con respecto al término, pero en el momento de la recuperación, ambos elementos (término y verbo elegido) no tienen una preferencia sintáctica específica. Es decir, simplemente se puede solicitar su aparición. Esto, si bien tiene también implicaciones semánticas, no es lo suficientemente preciso como para poder establecer la temática de un fragmento. Para superar este escollo se estableció, con base en las observaciones y las combinaciones posibles de los dos elementos implicados, que habría que indicar cuál de ellos sería el más importante al momento de elegir el fragmento (el núcleo) y cuál el satelital.

### ***III. Hacer una búsqueda de concordancia compleja donde el query fuese el término elegido y el contexto el verbo con la asociación más estrecha.***

Sketch Engine permite hacer búsquedas complejas, es decir, extraer fragmentos con dos coincidencias. Una de ellas es el query o la palabra que se colocará como núcleo, totalmente inalterable, y la otra palabra será el “contexto”, esto es, lo que puede aparecer tanto

a la izquierda como a la derecha del query sin importar su distancia y sin importar su conjugación, flexión o derivación<sup>26</sup>.

Una vez determinado que *propagar* era el verbo más cercano al término, y más productivo para el área de informática, decidimos que el query sería el término y, el contexto, dicho verbo. Los resultados fueron los siguientes:

Query **virus** 389,944 > Positive filter (excluding KWIC) **propagar** 3,252 (0.3 per million)

Page 1 of 163 [Go](#) [Next](#) | [Last](#)

#77706 de cava. </p><p> En los años 90 d.N. (Después del Nintendo), un **virus** llamado Nintenditis aguda se **propaga** por el universo hasta

#690145 ayudará a los científicos a comprender mejor cómo evolucionan los **virus** de la gripe, cómo se **propagan** y como causan enfermedad. </p>

#735674 causado la muerte de 39 personas. </p><p> Los expertos temen que el **virus** mute de forma que se **propague** entre los seres humanos y derive

#2524602 aplique todo el peso de la Ley contra los que crean y **propagan virus** por Internet, afectando tanto a Empresas como a usuarios particulares

#10569123 por cualquiera de los medios empleados normalmente por otros **virus** (correo electrónico, redes de ordenadores, transferencias de

#11601391 BAT sin reiniciar el sistema. </p><p> Aunque, al parecer, este **virus** aún no se está **propagando** masivamente, conviene tener presentes

#17159327 radio y televisión para anunciar la aparición de un extraño **virus** en el país. Un virus que se **propagaba** sin control, llegado

#17159333 para anunciar la aparición de un extraño virus en el país. Un **virus** que se **propagaba** sin control, llegado del exterior y del que

#20589021 idea que tuvo de pequeño: vengarse de Solid Snake **propagando** su virus de computadora llamado "Virus T" que a su vez se subdividía

#20589026 de Solid Snake **propagando** su virus de computadora llamado "Virus T" que a su vez se subdividía en "Virus G", "Virus A lpha"

#20589384 violaciones al ejército alemán. Albert tuvo éxito al **propagar** sus **virus** por internet, especialmente en Ciudad Mapache la cual es considerada

#23007727 <p> Proteja su piel de una infección dañina. </p><p> Evita que el **virus** se **propague** a otras áreas de su cuerpo. </p><p> Los extractos

#27583686 Luisiana, quien asegura que "esto es sólo el comienzo". </p><p> El **virus** del Nilo se **propaga** a los humanos por la picadura de un mosquito

#27755285 embarca en una misión para encontrar la verdad. En poco tiempo, el **virus** se **propaga** como un reguero de pólvora y se convierte en una

#33260026 medidas que les ayudarían a impedir que se **propague** entre ellos el **virus** . </p><p> Tad Carper , portavoz de los Cavaliers, indicó que "

#43615087 de enero con tres miembros de una familia de Tailandia, este **virus** no se **propaga** entre los humanos de forma eficiente. Por el

#50214651 inflación se **propaga** furibunda por la economía global, cual **virus** del ébola. </p><p> Greenspan es un hombre muy listo. Es absurdo

#51370643 psicópatas únicamente interesados en continuar **propagando** el **virus** sin tener en cuenta su propia integridad física. No serán controlables

#52564908 tan leve como se dedujo con posterioridad. Ya se conoce que el **virus** se **propaga** a una gran velocidad y que el mal afecta con especial

#54317999 que rotos en fragmentos no sirven para **propagar** . </p><p> Los **virus** que son nocivos son los que no son atacados por las restricciones

Page 1 of 163 [Go](#) [Next](#) | [Last](#)

**Figura 5.5. Búsqueda compleja con el término *virus* como núcleo y el verbo *propagar* como contexto**

Se puede notar que hay una recurrencia mayor de contextos que se refieren a informática, sin embargo, todavía aparecen algunos contextos referidos a biología. Para agregar un poco más de precisión a dichos fragmentos, apelamos a un criterio importante para la extracción de estructuras recurrentes: las colocaciones.

<sup>26</sup> Ambos argumentos, distancia con respecto al query y accidentes gramaticales, son configurables en la herramienta. Se puede determinar el límite de tokens a la derecha o a la izquierda donde puede estar posicionada la palabra-contexto, y se puede elegir si se desea que sea la forma introducida o si el sistema puede buscar todos sus derivados a partir de la determinación de su lema.

#### ***IV. Hacer una extracción de colocaciones a partir de los fragmentos que cumplieren las restricciones anteriores con el fin de determinar qué forma o conjugación del verbo asociado al término es más productiva***

Toda vez que los fragmentos ya cumplieran con los criterios anteriores y se acercaban visualmente a un contexto en el área de informática, lo que nos faltaba era poder determinar si existía alguna constante que nos permitiera saber cuál de las relaciones establecidas entre el término y las formas del verbo elegido era la más productiva. Las colocaciones, en tanto que son combinaciones frecuentes entre palabras y pueden tener una asociación con flexibilidad sintáctica, fueron las más idóneas para tratar de identificar estructuras más comunes y cercanas al término. Miremos por ejemplo los casos mencionados por Školníková (2010, p. 28) a propósito de diferentes categorías gramaticales:

##### **a) Modificación adjetival**

*Ganó un premio importante* en el mundo de la literatura [*ganar un premio*]

##### **b) Pronominalización**

*La multa* ha prescrito y ya no tengo que *pagarla* [*pagar la multa*]

##### **c) Nominalización**

*La derogación de las leyes* de la dictadura es normal en una democracia [*derogar las leyes*]

##### **d) Transformación en pasiva**

*El cadáver fue inhumado* en el cementerio del pueblo [*inhumar el cadáver*]

##### **e) Uso atributivo y predicativo de algunos adjetivos**

*invierno crudo* > *el invierno es crudo*, *un cuchillo afilado* > *el cuchillo está afilado*.

##### **f) Inclusión de cuantificadores**

*El cuchillo es muy agudo* [*cuchillo agudo*]

### g) Relativización

*seguidor asiduo* > *el que sigue asiduamente, comportamiento*  
*ejemplar* > *el que se comporta ejemplarmente.*

Como vemos, las combinaciones posibles de una colocación no atienden siempre a un patrón específico y sus elementos se pueden encontrar distribuidos a lo largo de una frase u oración. Con esta idea en mente, extrajimos la lista de las colocaciones asociadas a la combinación de *virus* + *propagar* con el fin de saber cuál de las formas del verbo (palabra-contexto) y cuáles otros términos eran más comunes. El resultado fue el siguiente:



	Freq	T-score	MI	logDice
P   N propagar	62	7.874	19.353	7.856
P   N propaguen	3	1.732	17.589	6.016
P   N propagan	10	3.162	17.542	6.034
P   N propaga	8	2.828	16.383	4.887
P   N maliciosos	3	1.732	14.820	3.325
P   N informáticos	10	3.162	13.928	2.444
P   N gusanos	3	1.732	13.915	2.426
P   N hackers	3	1.732	13.816	2.327
P   N spam	6	2.449	13.446	1.962
P   N Potter	3	1.732	13.197	1.711
P   N antivirus	3	1.732	13.156	1.670
P   N infectados	3	1.732	12.952	1.466
P   N piratas	3	1.732	12.602	1.117
P   N Harry	3	1.732	12.095	0.612
P   N enlaces	3	1.732	11.833	0.350
P   N computadoras	6	2.449	11.785	0.303
P   N curiosidad	3	1.731	11.612	0.129
P   N códigos	3	1.731	11.470	-0.013
P   N direcciones	3	1.731	11.135	-0.347
P   N archivos	5	2.235	10.885	-0.597
P   N mensajes	6	2.448	10.835	-0.647
P   N electrónico	6	2.448	10.659	-0.823
P   N correo	8	2.827	10.655	-0.827
P   N amenazas	3	1.731	10.417	-1.065
P   N utilizados	3	1.731	10.411	-1.071
P   N Windows	3	1.731	10.320	-1.163
P   N Web	3	1.730	10.034	-1.448

Figura 5.6. Formas más comunes para las colocaciones de *virus* como núcleo y el verbo *propagar* como contexto

Si tomamos el primer resultado referente al verbo (*propagar* –infinitivo-) y recuperamos los fragmentos, obtendremos una lista como la que se muestra ahora:

Query **virus** 315,923 > Positive filter (excluding KWIC) **propagar** 93 > Positive filter (excluding KWIC) **propagar** 62 (0.0 per million)

Page 1 of 4 Go Next Last

#36222135	utilizadas y prevé que la tendencia de <b>propagar virus</b> a través de estos servicios continuará
#50757487	configuración del equipo , dañarlo o <b>propagar virus</b> informáticos . El usuario no puede instalar
#343811255	cuatro años , que han pasado a <b>propagar virus</b> con una finalidad personal o ideológica
#490927675	que se utilizará para poder <b>propagar virus</b> , será La facturación electrónica . Mensajes
#723295579	por otras entidades , enviar o <b>propagar virus</b> informáticos o realizar cualquier otra
#928815240	adjuntos comúnmente usados para <b>propagar virus</b> , tales como .vbs , . bat , . exe , . pif
#997446837	<p> Crean un falso Google para <b>propagar virus</b> 3 . </p><p> PowerPoint es afectado por
#1053260816	tipos en el mundo que tratan de <b>propagar virus</b> y spam por la red . </p><p> - Las mejoras
#1306278106	Métodos para cultivar células y <b>propagar virus</b> </p><p> La presente invención se refiere
#1306278125	en particular a métodos para <b>propagar virus</b> , y más particularmente aún , a métodos pana
#1306278136	particularmente aún , a métodos pana <b>propagar virus</b> recombinantes para terapia de genes . </p>
#1408669815	Es también una forma común de <b>propagar virus</b> , gusanos y Troyanos . Por ejemplo , algunas
#1431161434	de solo lectura sirven para <b>propagar virus</b> . </p><p> Esta actualización es para Windows
#1552445539	creación de una nueva técnica para <b>propagar virus</b> en la red , que incluso , ya tiene nombre
#1607195093	- Crean un falso Google para <b>propagar virus</b> ==== http://www.laflecha.net/canales/seguridad/noticias/20060724/
#1729857027	textos pueden ser utilizados para <b>propagar virus</b> , son extremadamente vulnerables a estos
#2050664604	algunos ciberdelincuentes para <b>propagar virus</b> cuando los enlaces vienen en un correo
#2108246758	acusado el primero de ellos por <b>propagar virus</b> informáticos en redes académicas y militares
#2207060643	similar al acostumbrado para <b>propagar virus</b> . </p><p> Este problema afecta a los clientes
#2684247301	Esta también es una forma de <b>propagar virus</b> , generalmente a través de los archivos

Page 1 of 4 Go Next Last

**Figura 5.7. Colocaciones de *virus* como núcleo y el verbo *propagar* como contexto donde se focaliza la forma más productiva del verbo**

No es difícil observar que de los resultados obtenidos, apenas tres (#1306278106, #1306278136, y #1306278125 que más bien es indeterminado por falta de contexto) no corresponden al área de informática, lo que nos hizo pensar que esta combinación de extracción recurrente es un método bastante fiable para la obtención de fragmentos pertenecientes a una temática específica.

Sirva este caso como prototipo de la metodología seguida para el descubrimiento de coincidencias y recurrencias de la búsqueda de fragmentos como potenciales ejemplos asociados a un término. En total, se repitió este experimento quince veces, es decir, tantas como los términos elegidos; se obtuvieron tres mil frases asociadas a las quince definiciones respectivas, doscientas por definición, extraídas del DEM.

Sin embargo, estos criterios basados en estadística, que fue lo que básicamente hizo la suite de herramientas de Sketch Engine, han de ser complementados con bases teóricas y criterios lingüísticos derivados de esta investigación. Veámoslo a detalle.

## 5.2. Determinando la estructura de un ejemplo en terminología

En este apartado plantearemos las reflexiones hechas a partir de la observación de los datos en los experimentos realizados y del análisis de algunas posiciones que colocan al ejemplo como una categoría de datos proveniente de un conjunto de características particulares que no siempre son suficientes ni exhaustivas. Habrá que leerse, empero, que ninguna de esas posiciones se contravienen y menos aún son descalificadas por esta investigación, sino que se trata de un reanálisis combinatorio donde se toman las características más importantes de cada técnica para complementar otras, o se refuerzan los argumentos de ciertas posiciones con otros tantos que, antes que contradecir, complementan.

### 5.2.1. El ejemplo no está ahí *a priori*

Es posible que las dificultades mayores al momento de elegir fragmentos de corpus textuales que serán utilizados como ejemplos en diccionarios sean la pertinencia de su registro, la adecuación de su léxico y la cantidad de información aportada, entre otras. Si lo miramos de cerca, el corpus no es más que un recipiente de un conjunto de discursos provenientes de las más diversas fuentes, aun cuando éste sea un corpus de especialidad. Esta diversidad discursiva tenderá siempre a un determinado punto de acuerdo con ciertas escuelas de pensamiento, de acuerdo con ciertas temáticas que se deseen colocar como relevantes e incluso de acuerdo con el momento histórico en que se emitan o se diseñe dicho corpus. Por tanto, hay que pensar que las emisiones lingüísticas escritas son siempre resultado de variables que el lexicógrafo o terminólogo no podrán controlar completamente, así, no podemos pensar que el ejemplo puede ser un fragmento recuperado íntegramente de un discurso para referirse a otro discurso, siempre habrá interferencia. Uno podría pensar que la solución sería emparejar el discurso especialista con los criterios de selección de ejemplos, o sea, que el especialista sea quien elija dichos ejemplos, pero, ¿tendrá dicho especialista los criterios lingüísticos suficientes para elegir el mejor fragmento? La respuesta es clara: no. Es tarea del terminólogo determinar si un contexto cumple, o no, con lo dicho por la

definición para poder ejemplificarla. Recordemos que un ejemplo no es sólo mostrar el término de esa definición en un contexto de uso, sino que además es un mecanismo para mostrar los rasgos que la definición no abarca y, por tanto, para ampliar la información asociada a dicho término. Así, hay que hacer una pausa al momento de mirar fragmentos que contienen el término que deseamos ejemplificar y preguntarnos: ¿es este fragmento textual adecuado para ampliar la información del término? ¿es esta información fruto del discurso donde se inserta dicho término o es información asociada al concepto del término que desea ejemplificar? ¿Los rasgos de la definición y los rasgos del ejemplo se corresponden?

El discurso de especialidad no se enfoca en explicar o ejemplificar conceptos, sino determinarlos; el discurso general no se enfoca en determinar significados específicos de palabras, sino en explicarlas en cláusulas universales. Por ello, como podemos ver, sería inconveniente pensar que 1) el ejemplo se encuentra en el corpus de especialidad donde fueron extraídas las definiciones; 2) el ejemplo se encuentra en un corpus general, pero dicho discurso no se corresponde con el discurso especializado de quien puede emitir la definición de un término; 3) y el ejemplo puede ser un fragmento cualquiera que contenga el término sin atender a las razones discursivas contextuales que lo emiten.

Por otra parte, hemos visto que una simple concordancia o una colocación no son suficientes para determinar la pertinencia de un fragmento como ejemplo de un término, no contienen información suficiente o discriminatoria. Si se ha trabajado así es porque la supervisión humana ha estado presente, pero además esa supervisión humana está plagada de influencias discursivas o ideológicas (San Vicente, Garriga & Lombardini, 2012), tanto, que hay incluso un trabajo dedicado exclusivamente a ponderar la elección de ejemplos por estas técnicas en algunos diccionarios para el español de México (Fitch, 2013).

En los trabajos actuales para la elección de ejemplos en lexicografía vemos que, aunque se trate de alejar lo más posible la influencia del humano, es todavía necesaria su supervisión para cuestiones sobre todo lingüísticas. Véase la heurística del extractor automático de ejemplos GDEX:

## GDEX heuristics

- Sentence length (10-26 words)
- Mostly common words is good
- Rare words are bad
- Sentences
  - Start with capital, end with one of .!?
- No [, ], <, >, http, \
- Not much other punctuation, numbers
- Not too many capitals
- Typicality: *third collocates* is a plus

Galtür 2012

Kilgarriff: GDEX

10

**Figura 5.8. Heurística de GDEX (*Good Dictionary Examples in Corpus*) (Kilgarriff, Husák, Rundell, & Rychly, 2008)**

A propósito de esto mismo, el estudio de Humble (1998) hace énfasis en que el ejemplo no es uno solo para cada tipo de usuario, y que las palabras que rodean el vocablo que se desea ejemplificar dependen del nivel de formación (en este caso aprendices de L2) que tenga el usuario. Los aprendices iniciales necesitan que el ejemplo esté conformado por palabras muy frecuentes con significados muy frecuentes; en cambio, los aprendices avanzados necesitan que el ejemplo esté compuesto por palabras poco frecuentes en sentidos poco frecuentes; que los ejemplos inventados (*made-up*) ilustren usos nuevos y antiguos, y que los ejemplos recuperados de corpus (o *authentics*, auténticos) ilustren el “buen uso”.

Aunado a esto, debemos considerar que, como mencionamos en el capítulo anterior, no existe un consenso sobre la fuente exacta de los ejemplos, aunque todos ellos apuntan a corpus generales, y esto es importante porque incluso las unidades terminológicas o especializadas son tratadas igual cuando se habla del ejemplo asociadas a ellas (Zgusta, 1971).

Finalmente, los estudios referidos al ejemplo en terminología no mencionan a menudo la fuente de donde estos fragmentos deberían ser extraídos aun cuando se mencionen todas las características que deben tener. Sin embargo, con una mirada rápida podemos darnos cuenta de que, por la información o función que deben de cumplir dichos fragmentos (Fuentes Morán & García

Palacios, 2002, p. 80-87), su fuente más probable es un corpus de lengua general.

Como podemos ver, en la bibliografía destinada al estudio del ejemplo, que describimos en el capítulo 4, no se encuentra un punto en que todas las perspectivas converjan y se pueda indicar un punto exacto donde idealmente puede ser localizado un ejemplo. El ejemplo no está ahí *a priori*, los fragmentos elegidos siempre tienen que ver con un conjunto de funciones que deben cumplir con respecto al discurso y a su referente (vocablo o término), sus características formales sacrificando su contenido semántico (en las concordancias, por ejemplo) o su orientación semántica (en las colocaciones) que no siempre coincide con el sentido de la palabra que se quiere ejemplificar.

Ahora bien, aún sin estar ahí *a priori*, el ejemplo puede ser identificado gracias a las necesidades que cubre y a las características formales y semánticas, además de las funciones que desempeña con respecto a la aclaración del significado de un vocablo. Así, repararemos en que el trabajo actual sobre identificación y extracción de ejemplos desde corpus echa mano de los criterios que hemos mencionado, es decir, mira qué tan funcionales y potencialmente adaptables son los fragmentos documentados (lo que les dará su categoría de testigos) para poder ser asociados a un vocablo y no las asociaciones o relaciones entre un vocablo, su definición y la información faltante para poder elegir dicho fragmento. En esta propuesta miraremos que los ejemplos actuales sean una adaptación de algo existente –o documentado–, una necesidad pragmática.

Veamos a continuación como funcionan ciertas técnicas para la extracción de candidatos a ejemplos, sus características y algunos autores que las han descrito.

### 5.2.2. El ejemplo no es una concordancia

La concordancia es un fragmento textual donde se focaliza un elemento de cualquier categoría gramatical para poder mirar su contexto de aparición independientemente de otros criterios. Para el caso de la ejemplificación en terminología, las concordancias son

importantes debido a que muestran un término en un contexto donde se usa y puede ser que contengan rasgos que el término en sí no proyecta por sí solo, como las relaciones con otras palabras fuera del campo de especialidad.

La desventaja que tienen las concordancias es que justamente no son siempre contextos donde un término polisémico se desarrolla con el sentido que se busca, pues no tiene restricciones de tipo semántico o sintáctico. Se trata simplemente de una prueba de aparición. A efectos prácticos, las concordancias pueden ser un primer paso para acercarnos a un potencial ejemplo, de una manera sencilla, pero deben ser revisados de manera exhaustiva, lo que conlleva una supervisión humana obligatoria.

### 5.2.3. El ejemplo no es una colocación

La colocación muestra un conjunto de rasgos asociados a una pieza léxica como la cercanía de las palabras que la rodean y la utilización típica de esa pieza en contextos específicos. La ventaja de basarse en las colocaciones para la identificación de ejemplos en terminología es que dan una idea concreta de la fijación discursiva que tiene un término. Así, cuando extraemos las colocaciones de un término podemos ver que hay otras piezas asociadas que influyen en su carácter especializado y que, dependiendo del contexto, pueden o no activar sus capacidades como término. En el punto IV del experimento realizado en 5.1. vimos además que las colocaciones son flexibles en cuanto a su organización sintáctica y esto permite inferir que las relaciones establecidas entre las palabras que la conforman y el término al que se refieren mantienen una estabilidad que es posible identificar para el descubrimiento de patrones que nos aporten información semántica.

Sin embargo, la colocación en sí no es una prueba determinante de que el término se encuentra en una estructura ejemplificativa o explicativa, simplemente muestra que dicho término se relaciona más o menos frecuentemente con las unidades que conforman los conceptos de su área de especialidad. Incluso, como las concordancias, las colocaciones pueden tener defectos temáticos cuando el término es polisémico, pues estas recurrencias de aparición no distinguen si los campos semánticos de las unidades

satelitales se corresponden con el uso específico del término en áreas determinadas o si, incluso, el término que se está buscando es un derivado conceptual de otro término ya existente (como en el caso de *virus*). Es decir, la colocación no supera los escollos que presuponen la variación conceptual, como hemos mencionado anteriormente, y no distinguen la variación denominativa, pues términos polisémicos pueden compartir algunas estructuras sintácticas.

#### 5.2.4. El ejemplo no es una muestra de uso

Aunque en un principio se puede pensar que el ejemplo, como afirman varias de las investigaciones que hemos citado, es una muestra de uso, lo cierto es que hay que agregar aquí que *no sólo* es una muestra de uso. En primer lugar porque dicha muestra puede ser producto de una extracción no estudiada a fondo cuando parte de un corpus –y en este sentido son equivalentes a las concordancias– e incluso puede ser producto de una “invención”, cuando es el diseñador del diccionario quien determina que ese uso es el más común, asesorado por un experto o no. De aquí se desprende la siguiente inferencia que venimos repitiendo a lo largo de esta tesis: el ejemplo en terminología muestra el uso de un término en un contexto específico, pero ese contexto puede no ser especializado, lo que implica que puede no contener rasgos necesarios para que un especialista (formado o en ciernes) pueda distinguir las relaciones conceptuales del término y, por tanto, el campo semántico específico al que pertenece; o puede ser demasiado especializado, lo que implicaría que el uso se restringe sólo a los contextos del discurso donde se inserta y en consecuencia que las funciones, explicativas sobre todo, no se cumplen. Una muestra de uso que no está en relación con criterios bien definidos de adecuación muy probablemente será un elemento no pertinente en un diccionario de especialidad.

El ejemplo, por tanto, para ser considerado una muestra de uso, debe estar relacionado con conocimiento previo que permita distinguir dicho uso de otros posibles. Este conocimiento previo debe ser ponderado de acuerdo con las relaciones que el ejemplo ejerza y con respecto al nivel de especialización del lector en un

caso ideal, pero más estrechamente con el conocimiento previo que el mismo diccionario ponga en juego para darle la calidad de muestra de uso a dicho ejemplo, es decir, la definición. Desde el inicio de este trabajo hemos insistido en este punto, ya que es una consideración imprescindible. Siempre que el proceso de ejemplificación se pone en marcha, lo primero que se trae a la mente son los rasgos conceptuales del término a ejemplificar que se conocen. Si no hay un antecedente específico, esto es, si no hay una acepción del término como punto de partida, el ejemplo que muestre un uso puede diferir bastante de lo que se conoce previamente acerca de este concepto. En una línea, el ejemplo en terminología es una muestra de uso en tanto pueda mostrar los rasgos conceptuales del término. Es decir, el ejemplo terminológico es también un producto derivado del proceso de denominación del concepto proyectado en un término. Es un proceso complementario de comprensión y, por ende, atiende a un proceso onomasiológico; debe verse como un recurso para llegar al significado especializado de esa palabra y no sólo como una muestra de uso de esa palabra en tanto elemento lingüístico (que sería el caso de un ejemplo en lexicografía, por su dependencia de un proceso semasiológico)<sup>27</sup>.

### 5.2.5. El ejemplo no es una invención

Los ejemplos inventados, *made-up* en la tradición inglesa (Humble, 1998) o *ad-hoc* de acuerdo con lo dicho por algunos autores en la tradición hispánica (Fuentes Morán & García Palacios, 2002, p. 79), por mencionar algunos, son admisibles para efectos pedagógicos y tienen una gran aceptación entre la comunidad lexicográfica, lo cierto es que algunas veces carecen de esa calidad “natural” que se podría esperar para mostrar el uso de un vocablo; son una reconstrucción del conocimiento previo del que hablábamos y por ello la decodificación de su significado se puede ver seriamente afectada por motivos que ya hemos mencionado, como la ideología del diseñador del ejemplo o la escuela a la que pertenece<sup>28</sup>. Esta

---

<sup>27</sup> Véase Lara, “El ejemplo en el artículo lexicográfico”, 2006.

<sup>28</sup> Habría que mencionar aquí, como bien lo apunta Calderón Campos (1993, p. 112), que incluso la habilidad del lexicógrafo influye en la creación de ejemplos funcionales, pues se puede caer en casos poco informativos como “la casa es bonita” para representar un ejemplo de *casa* o de *bonito*.

aceptación del ejemplo inventado tiene su justificación en la calidad y la cantidad de información que se quiere mostrar en el ejemplo. Quien escribe el ejemplo intenta reflejar de la manera más cercana posible los rasgos semánticos del vocablo y los rasgos del concepto al que alude el término; la invención es un recurso válido, en los casos en que no se puede encontrar un fragmento que tenga estas características en un corpus. Una prueba inmediata de ello es el caso que nos ataño a propósito del primer esfuerzo por extraer candidatos a ejemplos con un corpus pequeño. Ahora, si bien es más probable encontrar candidatos en corpus grandes como el *esTenTen* o el *Wiki* y así evitar la invención, lo cierto es que también la revisión de millones de frases para encontrar la más adecuada es una tarea bastante compleja que conlleva demasiado tiempo. Para evitar, pues, que dicha tarea sea un obstáculo para la obtención de candidatos relevantes a partir de fragmentos ya existentes en corpus, habrá que analizar no el fragmento en sí, sino el mecanismo por el cual es generado, como se ha mencionado en el capítulo sobre la *saturación semántica*. Así, podríamos ver que ese mecanismo recae en una premisa que no se ha tratado hasta ahora y que es clave en la elección de ejemplos asociados a términos: el ejemplo se diseña. Pero no se diseña en términos sintácticos o a base de patrones lingüísticos, sino que se trata de un diseño de base cognitiva cuyo resultado puede ser observado en una realización lingüística. Veámoslo a continuación.

#### 5.2.6. El ejemplo se diseña

Cuando se pregunta a alguien sobre un ejemplo para determinada palabra, es muy probable que la persona responda algo muy cercano a “¿en qué sentido?”. Esto sucede debido a que lo primero que se necesita para poder emitir una respuesta a un estímulo lingüístico es un contexto. En palabras de uso general (Humble, 1998, p. 598) no hace falta más que buscar en el lexicón mental otras tantas palabras comunes que normalmente rodean a la palabra que se desea ejemplificar, pero recordemos que este es un proceso esencialmente lexicológico que en lexicografía se explica por el proceso semasiológico mismo con el que se definen las piezas léxicas de una lengua, esto es, yendo de la palabra al significado. Sin embargo, cuando se pregunta por un ejemplo en terminología, lo más

probable es que se necesite un contexto más amplio y especializado. Dicho contexto parte del significado de la palabra que se quiere ejemplificar, el término, y, de acuerdo con las relaciones conceptuales que dicho término guarda con otros términos del mismo dominio, se puede inferir un sentido específico. Ahora, ese sentido específico tiene una densidad terminológica que vuelve compleja la emisión, porque parte del discurso especializado. Hay que hacer un reanálisis de los contextos posibles donde ese término puede ser emitido y trasladarlo al que sea más aprehensible, más fácil de comprender, debido a que el ejemplo funciona como una reformulación de la definición de dicho término. Esto comporta la tendencia a seleccionar los contextos más cercanos al léxico general, pero se procura encarecidamente que dicho contexto mantenga, en la medida de lo posible, visibles los rasgos del concepto al que se alude. Esto sólo se logra si se da una combinación armónica entre palabras de uso común y otros términos asociados a la pieza que se quiere ejemplificar. Dicho proceso no es otra cosa que el camino mental que sigue un especialista cuando se dirige a un público lego y se demuestra lingüísticamente en los textos de divulgación científica.

En una línea: el ejemplo en terminología es un contexto general donde un término mantiene relaciones con otros términos o palabras que activan su significado especializado. Este contexto es, idealmente, un fragmento textual con un registro léxico distinto al especializado, más cercano al coloquial.

Todo esto puede desembocar en la siguiente premisa: si lo que se busca en un ejemplo es que esté relacionado con el término, pero que no se encuentre en el discurso especializado donde dicho término es emitido (un corpus especializado) y tampoco se busca que el potencial ejemplo sea una realización “genérica” –un fragmento del discurso general donde aparece el término–, entonces hay que buscar un fragmento lejano del discurso especializado que conserve los rasgos especializados del término. Esto sólo se logra si, como en el procesamiento humano, se atiende a la definición del término antes que a su misma aparición. Por consiguiente, el ejemplo en terminología es una representación lingüística de los rasgos conceptuales de un término que complementan y explican una definición y que está estrechamente relacionado con ella. Esa representación lingüística es, también,

resultado del proceso de ir eliminando contextos de ambos polos: aquellos demasiado especializados, densos en cuanto a la aparición de otros términos, y aquellos demasiado genéricos, densos en cuanto a la utilización de un léxico coloquial.

Atendiendo al experimento descrito al principio de este capítulo y con base en los preceptos teóricos anteriores, sobre todo en el capítulo 4, se pudo determinar un conjunto de condiciones, que aquí resumimos, para la identificación fragmentos relevantes que potencialmente podrían usarse como ejemplos. A saber:

- El ejemplo debe contener el término al que se refiere.
- El ejemplo siempre está asociado a una definición.
- El ejemplo debe ser un contexto general de un término.
- El ejemplo es la consecuencia de las restricciones sintácticas y semánticas aplicadas a un contexto donde aparece el término elegido.
- El ejemplo no tiene una estructura sintáctica fija.

Con todo lo anterior vemos entonces que el diseño de un ejemplo terminológico se da, inicialmente, en el plano conceptual, tal cual se hace en el proceso de denominación de un concepto a través de un término, y los resultados pueden variar de caso a caso. Estos casos tendrán implicaciones distintas si se dan en el habla, pues se tendría que tener en cuenta los conocimientos de quien formula el ejemplo, que si se dan en el texto, que es donde nos enfocamos en esta investigación. Ahora bien, si ya tenemos el punto de partida y conocemos el mecanismo por el cual se diseña un ejemplo, lo que hay que hacer es caracterizar ese ejemplo y describir las condiciones funcionales y estructurales (sintáctico-semánticas) que lo hacen parte de un artículo terminográfico.

### **5.3. Caracterización del ejemplo bajo criterios funcionales**

En este apartado abordaremos la propuesta de caracterización de un ejemplo en terminología atendiendo las funciones que puede llevar a cabo y describiremos brevemente cómo las hemos analizado.

#### **5.3.1. El ejemplo guarda una relación imprescindible con la definición**

En apartados anteriores hemos visto que el ejemplo funciona desde el concepto y no siempre desde la definición. Su papel principal es mostrar un conjunto de rasgos del concepto al que se refiere que sea más amplio que el conjunto que muestra la definición. Sin embargo, para que su funcionamiento pueda ser productivo, siempre se debe de tener en cuenta los rasgos conceptuales que ya ha cubierto aquella a través de la ecuación definitoria. Es decir, el punto de partida del ejemplo, desde esta perspectiva, son aquellos rasgos importantes que pueden aclarar el sentido de un concepto y que la definición, por las limitaciones que ya hemos visto anteriormente, no alcanza a cubrir, siempre entendiendo por “cubrir” como el reflejo de las distintas realizaciones de dicho concepto en el plano discursivo. Así, la función del ejemplo será mostrar los usos y las variaciones conceptuales que un término puede adoptar en el discurso especializado. En este sentido, lo dicho por el ejemplo complementará en el plano semántico a una definición y la aclarará, después la explicará y finalmente la trasladará a un plano lingüístico más sencillo donde las realizaciones conceptuales no se pierden, pero pueden ser aprehendidas más fácilmente por el lector. Así, dicho lector estará recuperando las caras no cubiertas del poliedro que supone un concepto a través de estructuras no definitorias pero siempre en ligazón con ellas. El mecanismo por el cual se puede llevar a cabo esta aprehensión es una reformulación definitoria, pero no en el campo sintáctico, sino en el semántico.

### 5.3.2. El ejemplo es una reformulación definitoria

Atendiendo a lo anterior, podríamos decir que el ejemplo es, de alguna manera, una reformulación conceptual de la definición en el sentido que intenta transmitir el mismo conocimiento especializado que supone el concepto, pero lo hace desde otro mecanismo. Mientras la definición se vale de un conjunto de estructuras estáticas que le dan estabilidad, el ejemplo es más flexible y puede adoptar formas explicativas, descriptivas, reformulativas e inclusoseudodefinitorias.

Cuando se trata de formas explicativas, descriptivas o reformulativas, el ejemplo parece adoptar estructuras que se inician con marcadores cuya función consiste en cohesionar enunciados cercanos semánticamente pero alejados en el plano léxico: *a saber, esto es, por ejemplo, en otras palabras, es decir*, etcétera. (Fernández Bernárdez, 1994-1995). En dichas cohesiones, uno de los enunciados tiene una carga conceptual densa, es decir, se acerca más a la definición y el segundo elemento reformula lo dicho por el primer enunciado en palabras más cercanas al léxico general. Cuando los marcadores textuales están ausentes, entran en juego marcadores fónico gráfico para indicar que lo que se dice a continuación constituye una reformulación implícita (dos puntos, paréntesis o guiones).

Sin embargo, como ya hemos visto, una reformulación textual no es siempre una estructura que agregue información adicional ligada al concepto, se trata más bien de un mecanismo que dice con otras palabras lo mismo que ya se ha dicho en la definición. En este sentido, los ejemplos que se pudiesen encontrar bajo esta perspectiva no cumplen satisfactoriamente con la noción de saturación semántica, más bien parecen acercarse a una fusión entre la enciclopedia y el diccionario y se dan siempre en el plano discursivo. No encontramos nunca un ejemplo en un diccionario que inicie con fórmulas como *por ejemplo, de otra manera, es decir*, etcétera. Y esto tiene una explicación sencilla. Como la misma Fernández Bernárdez indica, los marcadores textuales de ejemplificación conllevan una naturaleza de *doble adherencia*, esto es, por una parte se refieren a lo que el hablante acaba de decir y por otra parte a lo que dirá después del marcador (1994-1995, p. 116). Es decir, este tipo de ejemplos están casi siempre ligados a sus

marcadores (léxicos y fónicos) y ante su ausencia “se pueden producir, sin más, *enunciados incoherentes, incompletos*” (Fernández Bernárdez, 1994-1995, p. 114). El experimento que detallamos en 5.1, nos permitió comprobar lo anterior, ya que cuando aislamos todas las concordancias complejas del término *virus* y después a ese conjunto de concordancias le aplicamos una segunda búsqueda con los marcadores *por ejemplo, es decir y en otras palabras*; los resultados fueron casi nulos. Cuando un fragmento se “especializa”, esto es, cuando su núcleo es un término, y el contexto un verbo que lo identifica como parte de un área determinada, es muy poco probable que aparezcan estructuras reformulatorias para explicar ese mismo término ya que un texto especializado determina, prescribe el uso de un término dentro de su propio discurso.

De esta manera, ante lo infructífero de esta metodología en el discurso especializado, se tuvo que buscar otra manera de traer a flote esos fragmentos reformulatorios de un concepto a través de estructuras léxicas. El método por el cual comprobamos que el ejemplo funciona desde el concepto, fue tomar unidades de contenido de la propia definición y utilizarlas como marcadores de identificación de fragmentos potencialmente ejemplificativos. Es decir, si las palabras que definen un término tienen una carga conceptual densa y reflejan tanto los rasgos del concepto como las características del discurso especializado y algunas características del contexto de activación de dicho término, entonces no es raro pensar que las concordancias y colocaciones de esas palabras están semánticamente y estrechamente ligadas con término y concepto; forman parte del mismo campo semántico. Este proceso, esencialmente sintáctico basado en recuperación de información, lo explicamos en 5.4.

### 5.3.3. El ejemplo como núcleo conceptual

Retomando a Maldonado (1997), sabemos que hay palabras que “no se definen (no se describe su significado) sino que se explican (se describe qué son, cómo se usan y para qué se emplean), son palabras que se explican por su contexto de uso en vez de por su *contenido* (significado)”. Una palabra como *desmenuzar*, se suele

definir como una técnica empleada para dividir ciertas piezas de carne en virutas, pero para hacerla comprender, basta con un ejemplo del tipo: “Mi madre está desmenuzando el pollo para la ensaladilla rusa”<sup>29</sup>

De acuerdo con Pozzi (2004, p. 634) una imagen puede ser una definición ostensiva cuando se trata de un complemento de otro tipo de descripción conceptual. Para nosotros, recordemos, cuando ya existe una definición asociada a un término, el complemento en la entrada amplía el conocimiento y la aprehensión de dicho concepto. Así, la imagen de una célula que padece una enfermedad coronaria, para utilizar el caso que la autora describe, es conocimiento adicional que puede funcionar como núcleo conceptual en un diccionario de medicina si tomamos en cuenta que esa imagen distingue de todas las demás células, sanas y enfermas, aquella que tiene un padecimiento específico. Este mecanismo es el utilizado por los diccionarios gráficos donde sólo se coloca una palabra y una imagen que ejemplifica, muestra una de las realizaciones del concepto que representa. No es siempre la imagen prototípica, sino la que más se acerca a los rasgos que se quieren transmitir. En esta propuesta el ejemplo, como imagen, toma el papel definitorio cuando se trata de un complemento a un discurso especializado o a otra definición.

---

<sup>29</sup> Incluso hay que mencionar que, en el plano del habla, el ejemplo asume el papel de elemento definitorio, pero ni siquiera en una realización lingüística, sino en un acto definitorio ostensivo, se hace la imitación con manos y dedos imitando la acción misma.

### enfermedad coronaria

La enfermedad coronaria es un tipo de enfermedad del corazón. Es causada por la obstrucción gradual de las arterias coronarias. El corazón obtiene oxígeno y nutrientes de la sangre que fluye a través de estas arterias. En la enfermedad coronaria, (EC), se forman depósitos gruesos de tejido graso en el interior de las paredes de las arterias coronarias. Estos depósitos se llaman placas. Al engrosarse la placa, disminuye el flujo de sangre. Esto hace que el corazón reciba menos oxígeno. La placa, o un coágulo de sangre que se forma sobre esa la placa, pueden bloquear totalmente la arteria. Este adelgazamiento de las arterias por la placa, y a veces la obstrucción completa, aumentan el riesgo de sufrir un ataque cardíaco y de morir súbitamente.

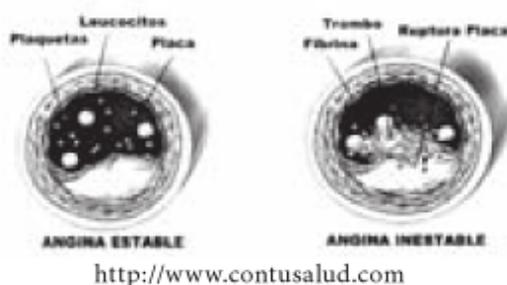


Figura 1. Definición ostensiva de la enfermedad coronaria

**Figura 5.9. Una definición ostensiva de acuerdo con Pozzi (2004), ejemplo como núcleo conceptual, de acuerdo con nuestra propuesta**

Para Meyer (2001), este tipo de estructuras son llamados *Knowledge Rich Contexts* (KRC) explicativos y se trata de aquellos contextos que, aunque cercanos a los *Contextos Definitorios*, carecen de género próximo y sólo presentan diferencias distintivas (lo que en el modelo aristotélico sería equivalente a la diferencia específica). En este sentido representan una aportación semántica al sentido del término sin necesariamente ser definitiva. En palabras de Alarcón, el tipo de información que aportan “sólo permite crear una clasificación de éste [el término] a partir de la relación conceptual que establece con otros términos de su misma clase [...] encontramos relaciones conceptuales como meronimia o sinonimia, por citar algunas.” (Alarcón, 2009)

Esto da fe de que una estructura que presente sólo diferencias específicas acerca de un término no puede ser tomada como definición pero, como aporta riqueza conceptual, puede ser tomada

como complemento de aquella, es decir, puede haber fragmentos a medio camino entre una definición y una explicación que pueden ser utilizados como ejemplos. Mírese el siguiente caso<sup>30</sup>:

El junco, que mantiene los hilos de urdimbre separados, ayuda a determinar el ancho de tela.

Aquí se provee información sobre los atributos del término *junco*, los cuales pueden ser vistos como una función: *que mantiene los hilos de urdimbre separados*, y como una utilidad: *ayuda a determinar el ancho de tela*.

#### 5.3.4. El ejemplo es un elemento pragmático

En nuestra propuesta el ejemplo tiene relación con una determinada definición pero, como vimos anteriormente, ya que se trata de una emisión que no depende de algún marcador y funciona desde el concepto de un término, puede ser un fragmento textual autónomo y no pierde fuerza comunicativa. Muy al contrario, si miramos el apartado anterior, veremos que incluso puede llegar a sustituir a una definición cuando el concepto es difícilmente transmisible por medio de una ecuación definitoria. En este sentido, entiéndase que, como complemento a la definición, tendrá una relación estrecha conceptualmente, pero puede que esté alejada gramaticalmente. Lo que conserva el ejemplo de la definición son rasgos comunes expresados de distinta manera, es decir, es otra forma de expresar significados asociados al término y pueden ser encontrados en el discurso de manera aislada. Es un elemento pragmático, de uso.

#### 5.3.5. El ejemplo como puente

Decíamos en el capítulo 2 que la relación concepto-término-definición es una codependencia semántica lineal y que el análisis de alguno de sus tres elementos dará al terminógrafo el apoyo necesario para diseñar las diversas categorías de datos de una entrada en un diccionario de especialidad. Pues bien, si recalamos

---

<sup>30</sup> Ejemplo tomado de Alarcón (2009).

la importancia de focalizar las relaciones entre estos elementos antes que en los elementos mismos, veremos que el ejemplo puede funcionar en todas ellas. Si el ejemplo está orientado a la relación entre concepto y término, puede mostrar el contexto de activación de dicho término, pues los elementos constitutivos del ejemplo darán fe de los otros términos que se usan para describir el comportamiento del concepto y las redes que establece con otros conceptos de su misma área de especialidad. Si por otro lado el ejemplo focaliza la relación entre término y definición, puede dar cuenta de los contextos que no están en el discurso especializado y en los que el término conserva sus rasgos de representante de un concepto. En este sentido, aunque el ejemplo muestra un término en un contexto donde el registro léxico no es el prototípico, el culto, que es lo que hace la definición, entonces sirve como apoyo para que el lector tenga un camino alternativo para comprender la definición y así acercarse más al concepto. Finalmente, si el ejemplo focaliza la relación entre definición y concepto, veremos que puede ser una reformulación conceptual que aporte información valiosa que no ha sido aprehendida por dicha definición, ya que recordemos que el término al que se refiere es sólo una representación lingüística de un concepto determinado. El ejemplo cubre las caras del poliedro conceptual que la definición, por sus limitaciones semánticas y sintácticas, no cubre; la complementa.

#### **5.4. Caracterización del ejemplo bajo criterios sintáctico-semánticos**

En este apartado abordaremos la propuesta de caracterización de un ejemplo en terminología atendiendo a la estructura que puede adoptar con el fin de sentar las bases para su reconocimiento automático y su extracción posterior.

##### **5.4.1. El foco conceptual**

El primer punto que no ha de pasarse por alto para la elección de un ejemplo es tener un núcleo en torno al que giren los elementos de la estructura. Mientras que en la definición, como hemos visto, pueden

ser el género próximo y la diferencia específica, las partes, el sinónimo o los hiperónimos e hipónimos, en el ejemplo el elemento imprescindible es el propio término. Esto no es raro si atendemos a que como hemos dicho, el ejemplo es una reformulación definitoria conceptual. A través de la puesta en escena del término en la estructura que deseamos buscar, llamamos al mismo tiempo al concepto que deseamos saturar y al elemento que representa los rasgos ya cubiertos de dicho concepto: el término. También ha de verse, pues, que con un término como núcleo, apelamos a su definición y delimitamos los contextos que pueden ser utilizados para ejemplificarlo. Así, pues, el foco conceptual pasa de ser un conjunto de rasgos definitorios a un conjunto de relaciones conceptuales.

#### 5.4.3. La ampliación de los rasgos

Una de las premisas que hemos marcado en esta investigación es que el ejemplo funciona básicamente ampliando los rasgos conceptuales y reduciendo el número de contextos. Pues bien, la ampliación de rasgos, con base en las observaciones, se lleva a cabo de una manera sencilla si, como lo hemos hecho hasta ahora, nos basamos en las relaciones que se establecen entre las palabras tanto de la definición como de los potenciales fragmentos a considerarse ejemplos.

El mecanismo determinado aquí tiene que ver con la asunción de que la estabilidad y la densidad de la definición es el punto de partida para diseñar el ejemplo. Pensemos en cómo relacionar un ejemplo con una definición.

Hemos visto que para efectos de la práctica lexicográfica y de extracción automática no podemos basarnos en patrones sintácticos y, por tanto, en los marcadores que introducen reformulaciones textuales. Entonces, si analizamos de cerca la definición y los *Contextos Definitorios*, hemos de caer en la cuenta que las palabras que lo forman, tanto otros términos como palabras de uso común, establecen una simbiosis que delimita el campo de acción del término. Esta simbiosis la entendimos en la investigación de Kuguel como el *Contexto de Activación* y la entendemos en Meyer como el *Contexto Rico en Conocimiento-explicativo*. De esto inferimos que

todas las piezas son importantes y aportan una porción de lo especializado al término.

Sin embargo, también hay que tener en cuenta que el uso de todas las palabras de una definición o un contexto de activación acarrearán ruido, ya que la mayoría de las frases recuperadas bajo estas condiciones serán variaciones sintácticas de lo dicho por aquellos elementos (definición o contexto de activación) o paráfrasis de cada uno de ellos, por lo que se incumpliría nuestro propósito de no reformular textualmente una definición. Por esta razón dejamos de lado la lista de verbos definitorios mostrados en 2.2., que aquí resumimos:

Caracterizar	Describir
Comprender	Entender
Concebir	Emplear (se) en
Conocer	Permitir
Considerar	Servir (en/para)
Definir	Usar (en/para)
Denominar	Utilizar (en/para)

**Figura 5.10. Lista simplificada de verbos definitorios**

Esta eliminación de los verbos romperá la estructura definitoria y dejará sólo los elementos que aportan información adicional y pertinente acerca del término. Nos quedaremos pues con los elementos activadores del *contexto de activación* y muy probablemente con todos los *contextos explicativos* (de base KRC) ya que éstos no tienen género próximo y, por tanto, es muy probable que tampoco tengan verbo definitorio. Ahora bien, si a nivel micro sabemos que los elementos remanentes de la definición dicen algo del término y lo delimitan semánticamente por las relaciones que establecen con él; no es raro pensar que a nivel macro, es decir, en las realizaciones (los contextos) de cada una de las piezas de ese remanente, se pueda encontrar información relativa al término. Esta reflexión viene acompañada y sustentada por el hecho de que, al referirse a un término, estas realizaciones se están refiriendo también a un concepto, y un concepto es a su vez significado especializado y significado determinado por redes. Así, si recuperamos estas realizaciones de las palabras del remanente como concordancias, entonces muchas de esas concordancias guardarán

cierta relación con el término. Las más cercanas representarán el campo semántico que sugiere el área de especialidad donde se inserta el término, y las más alejadas representarán otros campos semánticos donde esas otras palabras o términos funcionan<sup>31</sup>. A este campo semántico de especialidad, conformado por todas las realizaciones de las piezas del remanente en un corpus, lo hemos llamado *contexto nominal*, ya que básicamente está formado por nombres, pero también contiene otras categorías de palabras como adjetivos, cuya carga semántica es destacable.

Finalmente, por el hecho de que las palabras del remanente se insertan en un discurso especializado y por tanto se relacionarán a través de sus conceptos, entonces cada una de las combinaciones o fragmentos donde aparezcan y que compartan más o menos piezas léxicas, pero no las mismas, atraerán el conjunto de rasgos particulares que esa realización hace emerger. Así, el contexto nominal, entendido como la suma de concordancias del conjunto de las palabras remanentes de una definición, ampliará los rasgos del término en un ejercicio de desdoblamiento semántico conceptual. Descondensará el significado de una definición.

#### 5.4.4. La estabilización de la estructura

Ahora bien, ¿será suficiente ese ejercicio de descondensación para asegurar que los fragmentos recuperados sean ejemplos? Evidentemente no, pues son fragmentos genéricos y hasta ahora, aunque lo hemos mencionado, no podemos medir la cercanía de dichas concordancias con respecto al término y, menos aún, la cantidad informativa que le aporta a la definición como complemento.

Para hacer una depuración efectiva hay que tener en cuenta una vez más las relaciones establecidas entre concepto, término y definición. Si bien es cierto que por su propia naturaleza la definición sólo da fe de algunos verbos para referirse al término,

---

<sup>31</sup> Hay que tomar, empero, esta afirmación con pinzas, ya que los límites entre campos semánticos no son nítidos. Para poder medir la cercanía de un fragmento con respecto al término hemos desarrollado una técnica que explicamos al final de este capítulo.

también es cierto que esos verbos tienen poco que ver con el concepto en sí. Son herramientas de una estructura definitoria, no son caracterizadores de las posibles realizaciones de un concepto en un discurso. Por ello, una vez que nos hemos liberado de la estructura definitoria, hay que determinar cuáles son las estructuras que el concepto prefiere para mostrar sus rasgos. Aquí hay que tener en cuenta dos argumentos a nuestro favor: 1) Es altamente probable que todas las concordancias recuperadas a propósito de los elementos remanentes tengan al menos un verbo, pues en tanto elementos nominales, casi siempre forman parte de frases, oraciones o cláusulas. 2) Existen diversos métodos para determinar la cercanía de unas palabras con otras. De ellas destacamos aquí la *Información Mutua* (Ward Church & Hanks, 1990).

Estos dos argumentos marcan la pauta para el primer filtro de restricción que utilizaremos para la elección de candidatos a ejemplos. A saber: una vez que hemos podido descondensar la definición y al mismo tiempo eliminar las estructuras potencialmente definitorias de nuestro conjunto de concordancias que forman el *contexto nominal*, entonces podemos conservar sólo aquellas que se acerquen más a la definición. Si sabemos que, además de los verbos definitorios, puede haber otros verbos en la definición terminográfica, entonces podemos tomar esos verbos para limitar el número de elementos que comprenden nuestro *contexto nominal*. A este conjunto de verbos asociados al término que también podemos encontrar en la definición terminográfica lo hemos llamado *contexto verbal*. Los elementos de este *contexto verbal*, como hemos visto en el experimento realizado con *Sketch Engine*, pueden funcionar como argumentos de exclusión de fragmentos textuales semánticamente alejados del término y, así, tendríamos una primera aproximación a una medida de cercanía del fragmento que se refiere a un término. Esto desemboca en una estructura más estable que ya podríamos llamar “candidato a ejemplo”. Desgraciadamente no siempre se encuentran verbos en las definiciones más allá de los que la propia definición utiliza.

#### 5.4.5. La delimitación de contextos

Delimitar los contextos con la técnica que hemos descrito antes no parece una operación compleja y tiene la ventaja de que en todo momento se está trabajando con los mismos elementos de la definición, por lo que se puede estar seguro de que se trabaja dentro de un mismo campo semántico. Siempre se conserva por lo menos una relación, ya sea término-elemento remanente, término-verbo no definitorio o término-elemento remanente-verbo no definitorio. Este último es el caso ideal y el que finalmente usamos como filtro para decidir sobre la pertinencia de un fragmento a propósito de su inclusión en la lista de nuestros ejemplos potenciales.

Sin embargo, no todas las definiciones tienen verbos adicionales a los definitorios, o los tienen pero son irrelevantes para obtener información adicional o conceptual (pensemos, por ejemplo, en el caso de *dícese de...* o *dicho de un...*). Cuando esto sucede, es necesario echar mano de técnicas que recuperen los verbos más cercanos al concepto a través de la cercanía con el elemento que lo representa, esto es, con el término. Si ya hemos visto que de un corpus masivo podemos extraer sólo aquellos fragmentos que contengan las palabras remanentes de una definición, lo que conformará un subcorpus, y este subcorpus sabemos que está inserto en un campo semántico específico relativo a un área de especialidad, entonces ese subcorpus puede dar fe de las relaciones entre los elementos que lo componen. Como se ha dicho, muestra las relaciones de unos términos con otros y así el entramado conceptual que le confiere su calidad de campo semántico especializado. Con todo esto en mente, podemos entender que habrá una cierta relación de cada elemento con todos los demás elemento de este nuevo corpus. El resultado de esas relaciones –que pueden ser sintácticas, de aparición inmediata (lado a lado), o de aparición en los mismos contextos (en combinación con terceros elementos), etcétera– puede ser medido por la *Información Mutua*.

En lingüística de corpus, y particularmente en lexicografía, esta medida de *Información Mutua* se hizo especialmente famosa debido al enfoque estadístico, y por ende certero y palpable, que le daba a las investigaciones sobre léxico. En una investigación sobre asociación de palabras, Ward Church & Hanks (1990) mostraron que era posible medir la fuerza con que dos palabras se unen e

incluso se pueden predecir estructuras sintácticas sencillas como las frases (nominales) y mostraron las potenciales aplicaciones que podía tener en la confección de diccionarios. Desde entonces el uso de la *Información Mutua* ha sido una constante en los estudios lingüísticos, por ejemplo en traducción (Jang, Myaeng & Park, 1999), extracción de colocaciones (Bouma, 2009), parsing (Magerman & Marcus, 1990), entre muchos otros. Baste decir, aquí, que dicha medida la utilizamos para determinar qué verbos sería más conveniente utilizar para poder llevar a cabo la delimitación de contextos de la que hemos hablado. Así, superamos la falta de verbos y podemos instaurar un filtro más para la determinación del candidato a ejemplo.

#### 5.4.6. Los elementos prescindibles

Hemos hablado aquí del contexto nominal (que incluye sustantivos y adjetivos, esencialmente) y el contexto verbal (que pueden ser verbos no definitorios que están en la definición o verbos que se recuperan a partir de la cercanía semántica con el término), pero hay otra clase de palabras que si bien le dan estructura tanto a la definición como al ejemplo, son elementos prescindibles al momento de llevar a cabo los procedimientos de recuperación y aplicación de filtros: las palabras funcionales.

En semántica, uno de los teóricos que analizó a fondo más sistemáticamente las palabras funcionales fue Ullmann (1962). Para él, este tipo de palabras tienen una función estrictamente gramatical y no significan por sí mismas, sino en relación con las palabras con las que interaccionan en el discurso. Las palabras funcionales constituyen categorías finitas como las preposiciones, las conjunciones y los determinantes. En oposición, palabras de contenido, que son aquellas piezas léxicas con significado autónomo, constituyen categorías abiertas (siempre están en constante cambio, crecimiento y desuso) como los sustantivos, los verbos, los adjetivos, etcétera, y no dependen de otras palabras en el discurso para significar.

A estas palabras funcionales les hemos llamado aquí elementos prescindibles, ya que no son tomadas en cuenta para la

elección de los candidatos a ejemplos. De hacerlo, podrían suceder dos cosas:

a) que los fragmentos recuperados crecieran exponencialmente hasta infinito (pues casi siempre se usan en emisiones lingüísticas)<sup>32</sup>, lo que echaría abajo los filtros de restricción, o

b) que los candidatos a ejemplos perdieran la flexibilidad sintáctica de la que hablábamos antes y se redujeran drásticamente, pues estarían condicionados por la aparición de este tipo de palabras en posiciones determinadas. A efectos prácticos, como no aportan información conceptual y su función estructural no es una limitante para extraer candidatos a ejemplos, simplemente no se tomaron en cuenta para la extracción de los candidatos ni para el diseño de los filtros.

## **5.5. El ejemplo en terminología, su mecanismo de acción y una medida de asociación**

En este apartado veremos la síntesis de todos los preceptos teóricos que hemos venido explicando y se explicará cómo el conjunto de criterios, filtros y metodologías pueden dar como resultado una medida confiable para la extracción de candidatos a ejemplos en terminología. Al final de este apartado, se sentarán las bases para el diseño de la herramienta extractora.

### **5.5.1. El mecanismo de acción del ejemplo en terminología: la *saturación semántica***

Todos los criterios mencionados anteriormente para la caracterización del ejemplo en terminología han sido encaminados, como hemos apuntado, a enfocar las relaciones existentes entre concepto, término y definición. Esto dio como resultado que el ejemplo, visto bajo esta lupa, siempre tenga una relación estrecha con la definición asociada al término que tiene como núcleo. Por

---

<sup>32</sup> Baste saber que en todas las listas de frecuencias del español la preposición *de* siempre ocupa el primer lugar

otro lado, se trata de una emisión libre, independiente del acto definitorio y puede ser separado sin restarle importancia semántica; empero, si aparece, amplía dichos rasgos y aporta información nueva.

Esta aportación de información y esta independencia sintáctica permiten al ejemplo ser un elemento permutable y complementario. No necesita de una definición pero sí está relacionada con ella y solventa sus limitaciones sintácticas y actúa siempre desde el concepto. La aportación de información, continuamos, es fragmentada y depende en gran parte de los elementos que rodean al término en la definición, pero también en un contexto de uso general. Cuando este contexto de uso es además un contexto que refleja rasgos de activación especializada, entonces podemos decir que está cubriendo aspectos conceptuales. A mayor cobertura de rasgos conceptuales, mejor comprensión. A mejor comprensión, mayor aprehensión de conocimiento sobre dicho concepto. Este mecanismo de adición paulatina y fragmentada de rasgos es lo que hemos llamado *Saturación semántica*. Es decir, cubrir el mayor número de rasgos conceptuales posibles con el menor número de estructuras necesarias. Ahora bien, para poder determinar qué tantos rasgos estamos tomando en cuenta de la definición y qué tantos rasgos conceptuales se están adicionando, es necesario ponderar la cantidad de información que contiene el ejemplo.

### 5.5.1. La medición de la *saturación semántica* en un ejemplo: la *densidad lexicométrica*

Llegados a este punto es justo decir que tenemos las características y los filtros de un ejemplo en terminología y que podemos, por ello, explicar cómo se diseña, identificarlo e incluso ponderarlo frente a otros fragmentos. Retomando: un ejemplo en terminología, bajo esta propuesta, es un fragmento textual independiente (digamos de punto a punto), inserto en un discurso general (en corpus lingüístico general), que guarda una relación estrecha con la definición que se asocia a un término. Esta relación está explicada porque el término siempre aparece como núcleo en el fragmento que se considera candidato a ejemplo, pero además dicho fragmento contiene una o

más palabras que forman la definición sin importar su flexión o derivación; mientras más palabras comparten, más cerca están semánticamente; forman un *contexto nominal*. Además, como se trata de emisiones lingüísticas, deben de tener un núcleo sintáctico. Si, como apuntamos antes, el ejemplo es una predicación sobre el término, entonces dichos fragmentos contienen verbos que les dan estabilidad sintáctica; dichos verbos conforman el *contexto verbal*. Estos verbos idealmente son también los que vienen en la definición mientras no sean definitorios. Cuando una definición es descondensada y no quedan verbos que recuperar para formar el contexto verbal, entonces se utiliza el subcorpus que supone el contexto nominal, esto es, todas las frases de un corpus más grande que contienen el contexto y por lo menos una palabra de la definición. Este subcorpus, a través de la ponderación de la *Información Mutua* del término con todos los verbos, podrá aportar un *contexto verbal* confiable.

En una línea, el ejemplo de un término es un fragmento que contiene el mismo término, recuperado de un conjunto de concordancias, todas ellas con el término como núcleo, donde la más adecuada será aquella que contenga más palabras de la definición de ese término y el verbo más cercano semánticamente. Mientras más de estas condiciones cumpla el ejemplo, más denso será semánticamente. Ahora, si la densidad está en relación con un procedimiento léxico que será utilizado para estructuras de un diccionario, entonces estaremos hablando de *densidad lexicométrica*. La densidad que mide la importancia del léxico para extraer una unidad lexicográfica: el ejemplo.

Como podemos ver, la *densidad lexicométrica* es una noción teórica comprobable. Baste con elegirse un término, una definición de ese término y aplicar la metodología descrita. La metodología, para fines de esta investigación, puede ser formalizada a través de procesos automatizados y vaciada en una fórmula general para la ejemplificación. De ahí la denominación “lexicométrica”, más general, y no “terminométrica” por mencionar una opción más cercana. El mecanismo de la *densidad lexicométrica* actuará y obtendrá resultados basados en la definición que se le asocie, sea esta general o terminográfica.

Aplicado a terminología, se puede decir que esta medida expresa qué tan denso es un contexto general en relación con una definición terminográfica, qué tan cercano está de ella. Cuan más denso, y por ello cercano, sea un contexto general, más adecuado para ejemplificar un término (y por tanto para cumplir el *Principio de adecuación*). Cuanto más adecuado, mayor la posibilidad de que cumpla con el *Principio de poliedricidad*, ya que los fragmentos más densos, en este sentido, serán los más informativos y se necesitarán menos de ellos para que el lector obtenga una visión más global y clara del concepto de un término. Cuanto más cercano se esté de obtener esa visión global del concepto, más cercano se estará de saturarlo semánticamente.

## 5.6. Recapitulación

En este capítulo hemos visto que un experimento nos ayudó a sentar las bases para la caracterización de un ejemplo en terminología. Aunque ya la teoría nos había proporcionado datos suficientes para suponer que un ejemplo en terminología es distinto a un ejemplo en lexicografía –tanto en forma como en mecanismo de acción y diseño–, hubo que echar mano de criterios sintácticos, semánticos y pragmáticos para describir su funcionamiento e indagar la fuente donde se puede encontrar.

Quizá el asunto más importante aquí sea que se puso en juego la noción de *saturación semántica* para comprobar si efectivamente sus presupuestos teóricos funcionaban en la práctica, en el experimento, y finalmente si se podría derivar de ellos una explicación productiva y razonable del mecanismo por el cual funciona un ejemplo en terminología.

El siguiente punto en orden de importancia, hay que hacerlo notar, es el de la *densidad lexicométrica*. Esta medida, producto directo de la *saturación semántica*, permite ponderar de una manera sencilla y empírica la cantidad de información que puede tener un ejemplo a propósito de su relación con la definición.

Hay que enfatizar, sin embargo, que estas dos cuestiones no hubiesen sido posible sin las observaciones hechas a propósito de un experimento basado en corpus masivos. Así, se ha de justificar

que los resultados aquí obtenidos son producto de las aplicaciones de nociones teóricas a una gran cantidad de datos que, como dijimos al principio, son necesarios para la determinación de un patrón ejemplificativo, ya que éste no tiene una estabilidad sintáctica comprobable.

Ahora bien, la caracterización del ejemplo en terminología, como propuesta, sienta al mismo tiempo bases teóricas y prácticas ya que, como veremos a continuación, los resultados obtenidos de la observación de datos y el análisis del funcionamiento del ejemplo bajo esta perspectiva, abren la puerta a la formalización de esta categoría de datos, lo que permite su potencial extracción automática. Así, hemos de destacar aquí que el ejemplo que nosotros mostramos conjuga a su vez nociones que tienen que ver con patrones y reglas de restricción como lo veremos con GDEX, la metodología más seguida actualmente para la identificación de ejemplos: nociones cognitivas asociadas a procesos semánticos que explican un fenómeno lingüístico particular.

Finalmente, hemos de decir que la ejemplificación es un proceso cognitivo que tiene implicaciones lingüísticas. Ejemplificar es al mismo tiempo un proceso lingüístico y un proceso semántico que tiene como finalidad reformular y explicar significados en contextos específicos. En terminología, la ejemplificación es una reformulación de un concepto que relaciona al mismo tiempo concepto, término y definición terminográfica con el fin de hacer más aprehensible un discurso especializado. Con todo lo anterior, podemos inferir entonces que el ejemplo, en terminología, es la capacidad de traer al presente inmediato, con la misma lengua y en una sola oración o frase, un conjunto de rasgos asociados a un término con el fin de ampliar la información del concepto que representa en una situación contextual particular que ayudará a saturarlo, a comprenderlo mejor.

## **6. UN PROGRAMA PARA LA EXTRACCIÓN AUTOMÁTICA DE EJEMPLOS EN TERMINOLOGÍA: GENEX**

En este capítulo explicaremos el diseño y el funcionamiento de la herramienta GENEX (*Générateur d'Exemples*), un prototipo para la extracción automática de ejemplos con fines terminográficos. A partir de lo expuesto por una metodología productiva para la identificación de ejemplos (GDEX) que ha sido aprovechada para algunas lenguas, y con base en las propuestas teóricas sobre *saturación semántica* y *densidad lexicométrica*, diseñamos un algoritmo capaz de trabajar con corpus masivos de una manera aceptable. Presentamos un experimento hecho en tres áreas de especialidad a partir de 15 términos y evaluamos la herramienta derivada de dicho algoritmo. Los resultados sugieren que es posible extraer ejemplos asociados a definiciones terminográficas desde corpus de lengua general.

### **6.1. Un antecedente no tan cercano: GDEX**

GDEX o *Good Example Extraction* es un sistema que intenta cambiar el antiguo modelo de extracción semiautomática de ejemplos a partir de corpus. En el antiguo modelo, los ejemplos seguían, sobre todo, los criterios que hemos expuesto antes y que están expuestos en el trabajo de Atkins & Rundell (2008). La metodología para extraerlos ha consistido en seguir esos criterios y obtener concordancias para colocaciones que son guardadas y analizadas una por una hasta encontrar un buen ejemplo. El método de GDEX, en cambio, obtiene concordancias clasificadas y las presenta en orden de importancia. Para poder clasificarlas les asigna pesos que obtiene a través de una lista previa que varios estudiantes calificaron (con base en su conocimiento lingüístico) y que después fueron evaluadas de tal manera que el sistema pudiese elegir las mismas opciones que los estudiantes, con lo que se crearon reglas para elegir la mejor colocación<sup>33</sup>.

---

<sup>33</sup> “Once the features have been identified, the question arises: how should they be weighted? Which features are most important, and by how much? With this in

Los resultados de GDEX permiten ver que hay patrones en las frases que son declaradas buenos ejemplos, pero lo cierto es que en su lectura parece hacer falta una estructura funcional. Es decir, muchas veces se trata de algo cercano a una frase lexicalizada y otras tantas de palabras que se unen comúnmente por alguna cuestión temática aunque en el léxico diario no sean tan frecuentes. La siguiente lista presenta algunos resultados de GDEX para el inglés cuando es utilizado con el *British National Corpus*<sup>34</sup>:

<i>and/or</i>	toddler	○	○	○	○	○	<u>649</u> babies and toddlers [48.7 %]
<i>modifier</i>	unborn	○	○	○	○	○	<u>1710</u> unborn baby [63.2 %]
<i>modifier</i>	newborn	○	○	○	○	○	<u>1542</u> newborn baby [53.2 %]
<i>modifier</i>	premature	○	○	○	○	○	<u>904</u> premature babies [67.9 %]
<i>modifier</i>	aborted	○	○	○	○	○	<u>255</u> aborted babies [78.4 %]
<i>modifies</i>	boomer	○	○	○	○	○	<u>4111</u> baby boomers [84.3 %]
<i>modifies</i>	sitter	○	○	○	○	○	<u>400</u> baby sitter [70.8 %]
<i>modifies</i>	boom	○	○	○	○	○	<u>535</u> baby boom [96.4 %]
<i>object_of</i>	bear	○	○	○	○	○	<u>5159</u> babies born [24.7 %]
<i>object_of</i>	breastfeed	○	○	○	○	○	<u>766</u> breastfed babies [22.2 %]
<i>object_of</i>	breast-feed	○	○	○	○	○	<u>478</u> if you are breast-feeding a baby .<p><p> [39.7 %] [3, 2]
<i>object_of</i>	abort	○	○	○	○	○	<u>413</u> abort the baby [7.3 %]

**Figura 6.1. Algunos resultados de candidatos a ejemplo para *baby* obtenidos con GDEX**

Los resultados sugieren un intento interesante, pero lo cierto es que la contextualización de la palabra a ejemplificar es todavía un problema difícil de superar. A propósito de esto, sus autores

---

mind, we asked two students to select good examples for 1000 collocations. We then used those “known good” examples to set the weights, by automatically finding the combination of weights that would give the “known good” examples the highest average rank. The first two features, sentence length and word frequencies, were given greatest weight” (Kilgarriff et al., 2008, p. 3)

<sup>34</sup> <http://www.natcorp.ox.ac.uk/>

mencionan que hay que tener ciertos criterios en mente a la hora de hacer un acercamiento a la extracción y que es el método que siguen. A saber (Kilgarriff et al., 2008):

- Whole sentences – identified as beginning with a capital letter and ending with a full stop, exclamation mark, or question mark, were preferred.
- Sentences with ‘third collocates’, that is, words that occurred with high salience in sentences containing the node and primary collocate, were preferred.
- We note that good examples often first introduce a context, and then contain the collocation which, to speak figuratively, fits into the space that the context has created for it: this is helpful as a user who is unsure of the meaning of the collocation will be able to make inferences about what it must be from the context in which it appears. In sentences having this structure, the collocation is likely to be towards the end of the sentence. Sentences with the target collocation towards the end were given credit.

Algunas investigaciones para otras lenguas han obtenido resultados parecidos, satisfactorios, pero no siempre adecuados. De los casos que se pueden nombrar están el esloveno (Kosem, Husák, McCarthy, 2011) y el sueco (Volodina, et al. 2012). Para el caso del alemán (Didakowski, Geyken & Lemnitzer, 2012), los autores recurrieron al trabajo previo hecho para GDEX y acoplaron las nociones de frecuencia, longitud y aparición de la palabra “matriz”, que resumieron en dos criterios: legibilidad (readability) y complejidad (complexity)<sup>35</sup>. Hay que mencionar, empero, que este sistema funciona de manera aceptable cuando se utiliza para fines lexicográficos, pero cuando se trata de términos los resultados no son tan satisfactorios. Miremos un ejemplo usando su metodología.

En la primera figura veremos que a un nivel de profundidad, esto es, *virus* como palabra de búsqueda para concordancias y después la extracción de sus colocaciones, resultaría ser que *influenza* es la más común:

---

<sup>35</sup> “[...] examples are ordered in respect to their goodness with the help of some soft criteria which are listed in the order of their importance: a) including words should be among the 17000 most frequent words of our balanced corpus; b) including words should be no longer than 15 characters; c) finally, the keyword should be within the matrix clause.”

alez corpus: esTenTen11 (Eu + Am, Freeling, Lempos)

Collocation candidates

Page 1 of 686 Go Next >

	Freq	T-score	MI	logDice
P   N influenza	13,710	117.067	12.336	9.913
P   N H	14,008	118.284	10.704	9.508
P   N bacterias	11,150	105.556	11.474	9.507
P   N gripe	10,852	104.130	11.239	9.421
P   N hepatitis	8,158	90.301	12.126	9.240
P   N papiloma	6,907	83.105	14.468	9.149
P   N VIH	8,563	92.473	10.504	8.972
P   N infección	9,047	95.042	10.325	8.968
P   N inmunodeficiencia	5,197	72.085	13.874	8.734
P   N virus	8,680	93.029	9.403	8.553
P   N N	14,103	118.487	8.785	8.467
P   N AH	4,050	63.626	12.231	8.324
P   N vacuna	5,049	70.988	10.016	8.287
P   N ISA	3,696	60.784	12.446	8.210
P   N infectados	3,847	62.000	11.317	8.185
P   N Papiloma	3,390	58.221	14.685	8.141
P   N propagación	3,714	60.911	10.903	8.094
P   N herpes	3,307	57.499	12.841	8.071
P   N hongos	3,593	59.900	10.484	7.994
P   N inmunodeficiencia	2,861	53.485	14.058	7.892
P   N sida	3,227	56.771	10.647	7.884
P   N informáticos	3,528	59.337	9.951	7.876

Figura 6.2. Lista de colocaciones más frecuentes para *virus* en *esTenTen*

Y sólo hasta la posición veintidós encontraríamos algo que tuviese que ver con informática. Por tanto, no podríamos encontrar un ejemplo adecuado a nuestros fines:

Query *virus* 389,944 > Positive filter (excluding KWIC) *influenza* 13,710 (1.2 per million)

Page 1 of 686 Go Next Last

#4326463	características diferentes (coronavirus en el caso del SRAS, <i>virus</i> de la <i>influenza</i> en el caso de la gripe), el periodo de incubación
#10150469	realidades </p><p> Zanamivir : inhibidor de la neuraminidasa del <i>virus</i> de la <i>influenza</i> . Se utiliza para el tratamiento de la gripe
#10565466	asafoetida han mostrando una actividad antivirica frente al <i>virus</i> de la <i>influenza</i> A1 superior a la de amantadina. </p><p> La medicina
#43614710	inmunización específica. </p><p> Los virus de la gripe </p><p> El <i>virus</i> de la gripe o <i>influenza</i> es a menudo subestimado. Sin embargo
#43614847	controlable, advierten los expertos </p><p> De los tres tipos de <i>virus influenza</i> que se conocen, A, B y C, sólo el primero puede desatar
#43615149	para mutar </p><p> Una de las características más increíbles del <i>virus influenza</i> de tipo A es su facilidad para mutar. Cuando estos
#43615476	. </p><p> Para defenderse de estas peligrosas alteraciones del <i>virus influenza</i> , la OMS estableció en 1947 un programa internacional
#52613277	la ciudad, si bien el impacto de una pandemia causada por el <i>virus de influenza</i> porcina es dif /> l de predecir, pues depende de
#52613323	ruzada por anticuerpos adquiridos a trav /> de la infec /> on <i>virus de influenza</i> estacional y de factores propios del portador.
#66178692	Sun, exploraron esta posibilidad mediante la disección de los <i>virus de influenza</i> H1N1 y H9N2 y la mezcla sistemática de genes de
#74376881	de lanimivir para el tratamiento de pacientes adultos con <i>virus influenza</i> . Métodos: Estudio randomizado y doble ciego que
#108441265	declaró recientemente en la mayor parte del planeta por motivo del <i>virus</i> de la <i>influenza</i> H1N1, hace menester una profunda reflexión
#116570746	sobre el manejo diagnóstico y terapéutico de infección por el <i>virus de la influenza</i> tipo A (H1N1) y la organización de la asistencia
#134118922	comunicación oral sobre inmunología de circovirus, calicivirus, PRRSV y <i>virus de influenza</i> . </p><p> La Sociedad Española de Inmunología (SEI
#139254995	dedicó su tesis al estudio de los mecanismos moleculares del <i>virus de la influenza</i> , bajo la supervisión del Dr. Juan Ortín en
#148498429	de acción </p><p> inhibe selectivamente las neuraminidasas del <i>virus influenza</i> , importantes para la entrada de virus en células
#161708561	Laboratorios NBS3 de CRESA realizó dos comunicaciones, sobre el <i>virus de influenza</i> aviar y el virus del Nilo Occidental. </p><p> La
#161708567	realizó dos comunicaciones, sobre el virus de <i>influenza</i> aviar y el <i>virus</i> del Nilo Occidental. </p><p> La sexta reunión de los grupos de
#168543823	infecciosas argentinos, cuyo paravies por una fase /> ida del <i>virus influenza</i> AH1N1, determinan que la vacuna contra el neumococo
#175136493	<i>virus</i> . Aunque la amantadina inhibe la replicación de todos los <i>virus</i> de la <i>influenza</i> A, no es un sustituto de la vacuna. </p><p> Mecanismo

Page 1 of 686 Go Next Last

Figura 6.3. Resultados más comunes del término *virus* y su colocación más frecuente (*influenza*)

Como vemos, para la palabra *virus*, en su acepción más común o primaria, esta metodología funciona bien. Sin embargo para la desambiguación de contextos el proceso se debe de complejizar y utilizar reglas para determinar verbos y patrones asociados a un tipo específico de *virus*, ya que esta metodología

sólo consideró, hasta ese momento, las palabras más comunes asociadas para formar las reglas de ejemplificación.

De este esfuerzo rescatamos el hecho de que se demostró que se puede encontrar un cierto tipo de patrón (ellos lo hacen con clases funcionales, como podemos ver en la columna de la izquierda de la figura 6.1.) y que la combinación de *concordancias* + *colocaciones* realiza una serie de recortes que pueden ser utilizados para determinar aparición de cierto elementos léxicos. Se trata de concordancias complejas que lanzan contextos específicos y sobre las cuales hablaremos más adelante.

## 6.2. Corpus de análisis

Como vimos al principio de este trabajo, el primer acercamiento al estudio sobre ejemplificación en terminología fue llevado a cabo con el *Corpus de las Sexualidades en México* (Sierra, Medina, Lázaro, 2009), sin embargo, debido a su pequeño tamaño –cerca de un millón de palabras– fue imposible determinar si existía algún patrón recurrente en la aparición de un término en ciertos contextos que pudiesen ser aprovechados como ejemplos. Por esto, fue necesario buscar corpus más grandes para explorar las posibilidades contextuales de un término y poder perfilar, a través de una revisión manual, cuáles de ellos eran potenciales candidatos a ejemplos. Al final se optó por trabajar con lo que hemos llamado *corpus masivos*, esto es, corpus de tamaño mayor a los cien millones de palabras.

### 6.2.1. Corpus *Jornada*

*La Jornada* es un diario mexicano de circulación nacional, fundado en 1996. El corpus que se creó está formado por todo el contenido del sitio que contiene la versión impresa (<http://www.jornada.unam.mx>), desde el 21 febrero de 1996 hasta el 30 de mayo de 2014. Incluye comentarios de usuarios, ya que fue extraído automáticamente de su página de Internet. Básicamente es un corpus periodístico, sincrónico, de registro culto, con inclinación hacia la política, economía y cultura. Variedad dialectal del español: mexicana. Este corpus contiene más de 382 millones de palabras

(382,119,353) distribuidas en aproximadamente 15 millones de frases.

### 6.2.2. Corpus *Wiki*

Este corpus está formado por los artículos de la enciclopedia Wikipedia en español (*dump* de la versión septiembre 2014)<sup>36</sup>. Se trata de un corpus de lengua general, enciclopédico, de registros varios, sincrónico. Variedad dialectal del español: todas. Está formado por casi 393 millones de palabras (392,530,981) distribuidas en aproximadamente 16 millones de frases.

### 6.2.3. Corpus *EsTenTen*

Es una versión reducida del original *esTenTen* de Sketch Engine (Kilgarriff & Renau, 2013) donde sólo se utilizaron los primeros 5 gigabytes de texto. Es un corpus de lengua general que agrupa todas las variantes del español. Esta muestra tiene un tamaño de más de 2 mil millones de palabras (2,443,447,212) distribuidas en aproximadamente 78 millones de frases.

## 6.3. Preprocesamiento del material de trabajo

En esta fase del trabajo, los tres corpus mencionados tuvieron que ser preprocesados para uniformizar aspectos como la codificación, el etiquetado y la segmentación. A continuación explicamos el procedimiento utilizado para poder tenerlos todos en un mismo formato con el fin de que la herramienta informática GENEX pudiese llevar a cabo su tarea de manera eficiente.

---

<sup>36</sup> <https://es.wikipedia.org>

### 6.3.1. Recuperación

Para el caso de los corpus *Wiki* y *EsTenTen*, el procedimiento que utilizamos lo llamamos *recuperación*. Esto es, bajo ciertos permisos, uno explícito y otro libre, nos dimos a la tarea de guardar en un servidor todos los datos que contenían sus fuentes. Explicamos a continuación el proceso.

Para el corpus *Wiki* utilizamos Kiwix<sup>37</sup>, que es el gestor estándar de descargas de la Wikipedia. Un script en Python (*WikiExtractor.py*) facilita la tarea y es proporcionado gratuitamente por los autores, Attardi y Fuschetto<sup>38</sup>. Aquí el inicio de dicho script:

```
17 # -----
18 # Copyright (c) 2009. Giuseppe Attardi (attardi@di.unipi.it).
19 # -----
20 # This file is part of TanL.
21 #
22 # TanL is free software; you can redistribute it and/or modify it
23 # under the terms of the GNU General Public License, version 3,
24 # as published by the Free Software Foundation.
25 #
26 # TanL is distributed in the hope that it will be useful,
27 # but WITHOUT ANY WARRANTY; without even the implied warranty of
28 # MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
29 # GNU General Public License for more details.
30 #
31 # You should have received a copy of the GNU General Public License
32 # along with this program. If not, see <http://www.gnu.org/licenses/>.
33 # -----
34 # Example of Use
35 #
36 # The following commands illustrate how to apply the script to a Wikipedia dump:
37 #
38 # wget http://download.wikimedia.org/itwiki/latest/itwiki-latest-pages-articles.xml.bz2
39 # bzcat itwiki-latest-pages-articles.xml.bz2 | WikiExtractor.py -cb 250K -o extracted
40 #
41 """Wikipedia Extractor:
42 Extracts and cleans text from Wikipedia database dump and stores output in a
43 number of files of similar size in a given directory.
44 Each file contains several documents in TanL document format:
45 <doc id="" url="" title="">
46 ...
47 </doc>
48 """
49 Usage:
50 WikiExtractor.py [options]
51
52 Options:
53 -c, --compress : compress output files using bzip
54 -b, --bytes= n[KM] : put specified bytes per output file (default 500K)
55 -B, --base= URL : base URL for the Wikipedia pages
56 -l, --link : preserve links
57 -n NS, --ns NS : accepted namespaces (separated by commas)
58 -o, --output= dir : place output files in specified directory (default
```

Figura 6.4. Fragmento de inicio del código *WikiExtractor.py*

Al ser un corpus masivo<sup>39</sup>, esto es, con un tamaño mayor a los 12 Gigabytes, la propia página lo proporciona en ZIM, un

<sup>37</sup> <http://www.kiwix.org>

<sup>38</sup> La versión libre y completa así como la documentación de este script se pueden encontrar en:

<http://medialab.di.unipi.it/Project/SemaWiki/Tools/WikiExtractor.py>

<sup>39</sup> Sobre la noción de *corpus masivo* mírese el apartado 1.4. de esta tesis

formato de alta compresión<sup>40</sup>. La licencia para el uso y la distribución de datos de la Wikipedia en español es libre y está protegido por la *Creative Commons Attribution/Share-Alike License 3.0*<sup>41</sup>, así como por la *GNU Free Documentation License*<sup>42</sup>.

Para la descompresión y la organización en archivos de texto plano editables se creó un programa sencillo en Bash<sup>43</sup> que básicamente extrae cada bloque de la Wikipedia, lo limpia y lo concatena en un solo archivo. La primera parte del proceso se puede ver a continuación:

```
1 #!/bin/sh
2 # ---Segmentador de corpus grande WIKIPEDIA---
3 #Juan-Manuel Torres & Adrian Lazaro 16 sept 2014
4
5 # Procesa cada bloque: segmenta en frases, agrega punto al final, ordena alfabeticamente y elimina frases duplicadas
6 # Todo ello simplificado usando pipes |
7 #clear
8
9 #cd ../Programas
10
11 #----- Ya esta hecho
12 #for i in ../CORPUS/CorpusWiki/*/*
13 #do
14 #   echo "Procesando fichero... $i"
15 #   perl split_nonum.pl < $i | perl punto.pl | sort | uniq > $i.seg
16 #done
17 #-----
18
19 #Concatenar todos los .seg para crear un solo CORPUS.txt
20 cat ../CORPUS/CorpusWiki/Descomprimido/*/*.*.seg > ../CORPUS/CorpusWiki/CORPUS_Wiki.utf8.txt
21
22 #Eliminar caracteres no utf8 para tree-tagger-spanish-utf8
23 perl elimina_utf8.pl < ../CORPUS/CorpusWiki/CORPUS_Wiki.utf8.txt > ../CORPUS/CORPUS_Wiki.txt
24
25 done
```

Figura 6.5. *ProcesaWiki.sh*: Extractor y conversor a texto plano de artículos de la Wikipedia

Este segmentador, llamado *ProcesaWiki.sh* echa mano, a su vez, de un programa en Perl para la segmentación de cadenas pegadas, el *split\_nonum.pl*, que las coloca una por una en cada línea; a este proceso lo hemos llamado *división*. Como se menciona dentro en los comentarios, este programa está basado en el algoritmo de Manning & Schutze (2000). Una parte del código la podemos ver a continuación:

<sup>40</sup> [http://www.openzim.org/wiki/Main\\_Page](http://www.openzim.org/wiki/Main_Page)

<sup>41</sup> Código legal: <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

<sup>42</sup> <https://www.gnu.org/licenses/fdl-1.3.en.html>

<sup>43</sup> <https://www.gnu.org/software/bash/>

```

1 #!/usr/bin/perl -w
2 use strict;
3 use encoding 'utf8';
4 # Version 0.1 Juan Manuel Torres
5 # usage:
6 # perl split.perl < TEXTE.txt
7
8 my @string = <>;
9 my $string = join " ",@string; # Coller les morceaux dans un string
10
11 ----- INEX 2013 -----
12 $string =~ s/\{([\^]+\}\)/ /g; # enlever les {}
13 ----- INEX 2013 -----
14
15 $string =~ s/[0-9]+\./ /g; # JORNADA 2014.11.sep Para segmentar líneas pegadas en noticias (sin punto)
16 $string =~ s/^\s*/ /g; # JORNADA 2014.11.sep Para segmentar líneas pegadas en noticias (xxx * xxx)
17
18 $string =~ s/\n/ /g; # preservar les sauts de ligne
19 $string =~ s/ +/ /g;
20 $string =~ s/>+ / /g; # eliminer les >
21 $string =~ s/-+ / - /g; # eliminer les -
22
23 $string =~ s/ \. e\./ ;?e? /g; # Exception i.e.
24 $string =~ s/ \. S\./ U?S? /g; # Exception U.S.
25 $string =~ s/ (Jan|Feb|Mar|Apr|Jul|Aug|Sep|Oct|Nov|Dec)\. / $!? /g; # Exception Months
26
27
28 # The sentence boundary algorithm used here is based on one described
29 # by C. Manning and H. Schütze, 2000. Foundations of Statistical Natural
30 # Language Processing. MIT Press: 134-135.
31
32 # abbreviations that (almost) never occur at the end of a sentence
33 my @known_abbr = qw/Gen al prof Prof ph d Ph C D H K dr Dr M Mre mr Mr mrs Ms vs vol cit pp ed cap p P R S W cf L N E v chap ch I No St T J F 1o
34 . 2o/;
35
36 # abbreviations that can occur at the end of sentence
37 my @sometimes_abbr = qw/etc jr Jr sr Sr/;
38
39 my $pbm = '<pbound/>'; # tentative boundary marker
40
41 # JMT Let ... as non boundaries
42 $string =~ s/\.\.\./ /g;
43
44 # JMT put a tentative sent. boundary marker after \n\n. Important all sentences must to have a space at the end
45 $string =~ s/\n\n*/ $pbm/g;

```

Figura 6.6. Inicio del script *split\_nonum.pl*, programa de segmentación cadena-por-línea

Finalmente, para pegar todos los archivos que contenían las cadenas de texto encontradas, una vez que se eliminaron las repetidas, las direcciones web, los metadatos de fechas, número de ejemplar, volumen, etc.; se utilizó el comando **cat**<sup>44</sup>, que ejecuta el programa de concatenación que viene por defecto en los equipos con base UNIX. Con esto se obtuvo un solo archivo en texto plano (.txt) con todos los documentos de Wikipedia, separados por cadenas, sin cadenas repetidas, y ordenados alfabéticamente. A este proceso lo hemos llamado *concatenación*.

Para el corpus *EsTenTen* fue necesario pedir permisos explícitos a los autores, quienes lo han donado amablemente para esta investigación bajo palabra de no distribución y bajo las licencias de la *Lexical Computing Ltd*. La descarga fue directa de servidor a servidor. El método de compresión que utiliza este corpus para ser transferido es el XZ de Tukaani<sup>45</sup>. Los archivos que venían dentro ya estaban en formato de texto plano (.txt), pero dicho corpus estaba en formato vertical, esto es, una palabra por línea (WPL, por sus siglas en inglés). Además, ya venía anotado

<sup>44</sup> <http://unixhelp.ed.ac.uk/CGI/man-cgi?cat>

<sup>45</sup> <http://tukaani.org/xz/>

con el tagset del PennTreeBank con el método del *Stanford Natural Language Processing Group*, cuyo POS-Tagger está documentado, sobre todo, en el artículo *Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network* (Toutanova, Klein, Manning & Singer, 2003).

### 6.3.2. Extracción

A diferencia de los dos corpus anteriores, el que hemos nombrado *La Jornada*, cuyo nombre original como conjunto de archivos es *La Jornada en Internet*, tuvo que ser obtenido de manera distinta. En palabras de sus propietarios legales, DEMOS, Desarrollo de Medios S.A. de C.V, “se trata de la edición digital del diario *La Jornada*” y por ello todos los textos y el contenido multimedia obedecen a una organización web definida (HTML). Los archivos de este corpus fueron descargados directamente del sitio atendiendo al Aviso Legal disponible en su plataforma<sup>46</sup>, lo que nos permitió no tener que gestionar una serie de permisos para la utilización de este material periodístico de distribución nacional en México. Sin embargo, como mostraremos más adelante, necesitábamos que todos los documentos obtenidos que conforman los tres corpus estuviesen disponibles en texto plano simplificado y sin etiquetado de ningún tipo. Esto desembocó en el diseño de un crawler programado en Bash (*extractor02.sh*) que fuese capaz de extraer todo el texto pertinente, esto es, las noticias de cada una de las carpetas en que se organiza el diario: una carpeta por sección (8 secciones), de cada día del mes (28-31 días), de cada mes del año (12), de cada uno de los años (18). Además de buscar las carpetas –y luego los archivos– que contuvieran sólo texto (no imágenes ni multimedia), el programa eliminó todas las etiquetas HTML y convirtió cada archivo en uno equivalente en texto plano simple (.txt) a través del comando `–dump` del navegador *Lynx*<sup>47</sup>. Esta es la manera en que trabaja:

---

<sup>46</sup> <http://www.jornada.unam.mx/aviso.php>

<sup>47</sup> <http://www.skrenta.com/rt/man/lynx.1.html>

```

1 #!/bin/sh
2 # ---Extractor de HTML a partir de CORPUS Jornada---
3 #Juan-Manuel Torres & Adrian Lazaro 12 sept 2014
4 # Usa el directorio /tmp para almacenar temporalmente
5 #-----
6 clear      # Limpiar la pantalla
7 echo "Extractor HTML recursivo version 0.002"
8 echo "Juan-Manuel Torres LIA/Avignon & Adrian Lazaro UPF "
9
10 rm /tmp/*html # Eliminar *html de /tmp
11 rm CORPUS_bruto.txt # Eliminar CORPUS.txt para recrearlo
12 |
13 #----- RECORRE DIRECTORIOS RECURSIVAMENTE BUSCANDO html
14 DIR=./CORPUS/CORPUS_comprimido/JORNADA # Directorio raiz donde esta el corpus
15 find ${DIR} -type f | while read A ; do
16     if test "${A##*}" = 'html'
17     then
18         b=`basename "$A"`      # Obtiene nombre
19         cp "$A" /tmp/"$b"      # Lo copia al temporal /tmp
20
21         a=`echo /tmp/"$b" | tr "[:blank:]" "_"` # Elimina blancos del nombre y lo pone en $a
22         mv /tmp/"$b" "$a"      # renombre temporal con el nombre sin espacios
23         mv "$a" /tmp/x.html     # renombra archivo sin espacios en x.html para no crear millones de ficheros
24
25         lynx -dump /tmp/x.html >> ../CORPUS/CORPUS_bruto.txt # Llama a Lynx con cada x.html y guarda en CORPUS.txt el resultado
26     fi
27 done

```

Figura 6.7. Script completo de *extractor02.sh*

Como se puede ver, además de una recuperación, en este caso también tuvimos que hacer una extracción sencilla para obtener texto utilizable. A los archivos obtenidos en texto plano aplicamos también el programa *split\_nonum.pl* para separar las cadenas. Para pegar todos los archivos .txt obtenidos volvimos a utilizar el comando *cat*.

### 6.3.3. Limpieza y recodificación

En los tres casos las codificaciones del texto plano eran distintas entre sí (UTF-8, UTF-16, etcétera). Para que un programa cualquiera de procesamiento de texto funcione correctamente es necesario que el intérprete pueda leer el conjunto de caracteres que conforman los documentos de dichos textos en una sola codificación. Si un documento o una parte de ellos difiere de la codificación estándar que se haya determinado para trabajar y que se utiliza en el intérprete, no sólo arrojará resultados anómalos poco legibles para el ojo humano, sino que también mermará el desempeño global del programa al hacer pasar unos caracteres o símbolos por otros que nada tienen que ver ya con la lengua de trabajo, ya con la información adicional. El discernimiento de otras lenguas insertadas en el mismo texto como parte de citas literales, traducciones, equivalentes, etc. también se ve afectado por la



*limpiado*) no fueron suficientes para obtener los corpus en el formato deseado. Debido a que nuestra intención ha sido utilizar en la menor medida posible recursos externos (desambiguadores, etiquetadores, etc.), necesitábamos que los textos estuviesen ya divididos en frases antes de poder analizarlos con GENEX (ya que este programa funciona analizando segmentos previamente divididos por criterios establecidos por otras herramientas o de punto a punto). Sin embargo, estudios actuales han comprobado que no existe un método cien por ciento eficaz para dividir cadenas de texto en frases cortas y que mantengan una independencia semántico-sintáctica. Este tema ha sido profundamente estudiado sobre todo para el desarrollo de resumidores automáticos y compresores de frases, pero nosotros hemos tomado sólo un método para este estudio y utilizamos el programa que deriva de esa investigación: el algoritmo Cortex (Torres-Moreno, Velázquez-Morales & Meunier, 2002). De este algoritmo, utilizamos sólo una parte, la que tiene que ver con el preprocesamiento. Estos módulos, básicamente, hacen la labor de cortar las cadenas que obtuvimos en los corpus descritos y las divide en frases comprensibles a partir de medidas estadísticas. Los resultados de la segmentación mostraron que la herramienta hace su trabajo de manera adecuada con las cadenas (a través de una revisión manual) y presenta una notable ventaja al momento de discernir entre fragmentos de texto. Por ejemplo, sirvió para desechar frases no pertinentes, como fechas, direcciones electrónicas o publicidad. Una muestra de cómo quedan divididas las cadenas en frases se puede ver a continuación:

```

taIne@caracole:~/works/Lazaro/CORPUS/Corpus_JORNADA$ sed -n 15p JORNADA_bruto.txt.seg3
Se van porque no hay nada ahí , excepto su todo , que no ofrece lo suficiente para alimentar a sus niños .
taIne@caracole:~/works/Lazaro/CORPUS/Corpus_JORNADA$ sed -n '1,15p' JORNADA_bruto.txt.seg3
dónde hallar nuestro lugar .
Dónde hallar nuestro lugar .
Nunca antes ha sido tan extensa como hoy la devastación ocasionada por la búsqueda de la ganancia , según la define el capitalismo .
Casi todo mundo lo sabe .
Cómo entonces es posible no hacerle caso a Marx , quien profetizó y analizó tal devastación .
Sin mapa alguno , no saben a dónde se dirigen .
Todos los días , la gente sigue señales que apuntan a algún sitio que no es su hogar , sino a un destino elegido .
Señales carreteras , señales de embarque en algún aeropuerto , avisos en las terminales .
Donde se encuentran tiene la latitud , la longitud , el tiempo local y la moneda correctos , y no obstante no tiene la gravedad específica del destino que escogieron .
Se hallan junto al lugar al que escogieron llegar .
Puede ser únicamente la anchura de un vía pública , puede estar a un mundo de distancia .
Ha perdido su territorio de experiencia .
Por órdenes de los gobiernos británico y francés el albergue fue cerrado recientemente .
Varios cientos de personas se albergaban ahí , muchos de ellos con la esperanza de llegar a Gran Bretaña .
Se van porque no hay nada ahí , excepto su todo , que no ofrece lo suficiente para alimentar a sus niños .
taIne@caracole:~/works/Lazaro/CORPUS/Corpus_JORNADA$

```

**Figura 6.9. Visualización del corpus *Jornada* separado por frases con Cortex**

Como podemos ver, la mayoría de lo que nos presenta como frases son cadenas cortadas de punto a punto y, en varios casos, los cortes se hacen también por algunos marcadores discursivos (*nunca, casi, sin, etc.*). Que se inicien todos con mayúscula obedece a la presentación de las frases como segmentos independientes y los puntos al final con un espacio obedece a la segmentación del *stemmer* incluido en el propio Cortex. Esta última anomalía se corrige después al pasar por segunda vez el programa *eliminaNOutf8.pl*.

### 6.3.5. Etiquetado

El etiquetado de todos los corpus se llevó a cabo con *TreeTagger*, un etiquetador basado en árboles de decisión desarrollado por Helmut Schmid (1994) en la Universidad de Stuttgart a mediados de los noventa, dentro del Institute for Computational Linguistics. La decisión de usar este etiquetador fue porque su precisión es alta y porque el conjunto de etiquetas (tagset) que utiliza es adecuado para el trabajo que llevamos a cabo<sup>48</sup>. Dentro de la comunidad lingüística, además, es de gran aceptación y relevancia al ser uno de los etiquetadores más trabajados y mejorados para el español.

En la siguiente figura podemos ver cómo luce un fragmento de corpus etiquetado con este método:

```

en      SP500  en-i
aqueí  D00M50 aquel-x
preciso A00M50 preciso-j
momento NCM5000 momento-n
<g/>
,      Fc      , -x
el     DA0M50  el-x
timbre NCM5000 timbre-n
de     SP500   de-i
la     DA0F50  el-x
puerta NCF5000 puerta-n
en     SP500   en-i
la     DA0F50  el-x
casa   NCF5000 casa-n
contigua AQ0F50 contiguo-j
y      CC      y-c
el     DA0M50  el-x
ladrido NCM5000 ladrido-n
del    SP500+DA0M50 de+el-i
perro  NCM5000 perro-n
<g/>
,      Fc      , -x
quien  PR0C5000  quien-p
anunciaba VMII350 anunciar-v
así    RG      así-r
su     DP3C50  su-x
alegría NCF5000 alegría-n

```

**Figura 6.10. Fragmento de corpus etiquetado con *TreeTagger***

<sup>48</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/spanish-tagset.txt>

## 6.4. La formalización del ejemplo

A través de un procedimiento de exclusión, hemos podido dar cuenta de que el mecanismo por el cual un hablante forma, o más bien, elige e identifica un ejemplo, no atiende estrictamente a criterios sintácticos, sino que es un procedimiento cognitivo que echa mano de las relaciones entre palabras y el cruce de los campos semánticos a los que estas palabras pertenecen. A lo largo de esta tesis hemos visto tres condiciones formales básicas para poder identificar un ejemplo y extraerlo automáticamente, a saber:

- 1) que el término siempre se encuentra en el fragmento que se utiliza como ejemplo,
- 2) que el fragmento guarda una relación estrecha con la definición; por tanto comparte con ella algunas palabras, y
- 3) que el fragmento es una unidad sintáctica independiente con un verbo que la estructura, ligado al término y que este fragmento va de punto a punto o está separado en ocasiones por marcadores discursivos.

Si hay fragmentos que cumplen estas reglas, entonces podemos decir que son los más adecuados para utilizarse como ejemplos según nuestro método. Veamos a continuación las bases de los criterios de selección utilizados para programar GENEX.

### 6.4.1. La desarticulación de la definición

El Contexto Definitorio es uno de los tantos tipos de contexto que se pueden referir a determinadas operaciones mentales y lingüísticas que llevamos a cabo en la comunicación. Sus bases han sido analizadas anteriormente y hemos visto que tienen cierta estructura que les permite funcionar como delimitadores del significado de un término. Hemos visto también que hay otro tipo de contexto, llamado *de activación especializada*, que permite observar cómo la interacción de algunas piezas léxicas en un orden particular dentro del discurso especializado modifica la estructura semántica de un término. Pues bien, con la idea de que en el momento de crear la definición terminográfica estos dos procesos han sido llevado a

cabo o han sucedido, inferimos que las palabras que están dentro de dicha definición ya tienen un papel asignado en la emisión del mensaje que quieren hacer llegar, esto es, describir el significado de un concepto. El orden de estas palabras sigue un patrón determinado, ya sea porque definen (Alarcón, 2009, p. 125) o porque activan (Kuguel, 2007, p. 2), pero siempre están en relación con el término y dicen algo importante sobre él. Esto que dicen y la relación que mantienen con el término es lo que nos interesa. Para poder aprovechar la carga semántica de dichas palabras, primero hay que sacarlas de su estructura, es decir, quedarnos con lo definitorio semánticamente, pero deshacernos de la estructura estática que supone una definición. A este proceso es a lo que llamamos antes una “reformulación conceptual”. Para que dicha reformulación se pueda llevar a cabo, es necesario trasladar el conjunto de significados de la definición a otra estructura que conserve las relaciones de esas palabras con el término, pero que no sea tal u otra definición. Aquí es donde llegamos al primer momento de la metodología del GENEX: la desarticulación de la definición.

Para desarticular una definición es necesario desvanecer la estructura estática que hemos visto que conserva en el capítulo 2. Es decir, hemos apuntado a que un Contexto Definitorio y en particular una definición tienen como núcleos verbos definitorios. También sabemos que un verbo es la parte central de una oración en un sentido ideal y canónico. Y finalmente sabemos que las palabras funcionales en el lenguaje permiten establecer relaciones entre palabras para poder emitir mensajes coherentes sintácticamente. Pues bien, cuando hablamos de la desarticulación de una definición queremos decir que tenemos tres argumentos en mente:

- a) Los fragmentos que potencialmente pueden ser considerados ejemplos deben conservar las palabras de contenido de la definición, pero sin los verbos definitorios que especificamos más adelante en la figura 6.11.
- b) Se debe encontrar un nuevo conjunto de verbos asociados al término que o bien estén en la propia definición o bien se calculen dentro de un campo semántico determinado.

- c) Se debe encontrar un fragmento textual que cumpla con los dos argumentos anteriores, pero que además no sea demasiado extenso para facilitar su lectura y comprensión.

El conjunto de palabras conservadas, según su naturaleza y función, forman parte de contextos que sugieren campos semánticos. Esto es, determinamos que todos los fragmentos de texto donde aparece por lo menos una de esas palabras además del término, representan un contexto de uso. El conjunto de los contextos de uso conforman campos semánticos. Veamos cómo funciona este proceso a continuación.

#### 6.4.2. El contexto nominal

Para GENEX, el contexto nominal son todas aquellas palabras encontradas en la definición que no son verbos ni palabras funcionales. Lo hemos llamado así debido a que se trata esencialmente de nombres y adjetivos, aunque ocasionalmente puede admitir otras clases de palabras que no sean verbos con el único fin de dar riqueza al vocabulario utilizado. Veamos un caso extraído del DEM:

*virus*

s m sing y pl

**II**

*(Inform)* **Programa elaborado** para alterar y dañar el **funcionamiento normal** de una **computadora**; virus **cibernético**

En este caso resaltamos las palabras que formarían parte del contexto nominal de *virus* y con las que se trabajaría inicialmente. Si nos damos cuenta, las palabras elegidas conforman representaciones del campo semántico al que se refiere la nota (*Infor*), informática. Por consiguiente, no es difícil creer que la mayoría de estas palabras en combinación con el término se refieran a este campo si las buscamos en algún contexto específico. Es decir, la combinación de los significados de las unidades que conforman la definición inclinaría a pensar que el contexto en el que se combinan tiene que ver con el campo semántico *Informática*.

Si realizáramos este ejercicio de manera recursiva, veríamos que mientras más de estas palabras aparecen en combinación con el término en otros contextos, más probable es que ese contexto pertenezca a un área determinada. Es así como funciona el contexto nominal en el GENEX.

### 6.4.3. El contexto verbal

Los contextos encontrados por coincidencia de ciertas palabras, sin embargo, no son siempre representantes del campo semántico de un término cuando este presenta polisemia. Hay una dificultad evidente en la recuperación de fragmentos temáticos que empaña la identificación de ejemplos, ya que, aunque cumplen con características léxicas, su semántica y el orden de aparición de esos elementos no se tiene controlado hasta cierto punto. Por esta razón, nos vimos en la necesidad de incluir un segundo filtro para extraer fragmentos y para dar estabilidad a las frases que se podrían usar como ejemplos. Para este filtro procedimos de manera similar que con el contexto nominal. A saber: tomada una definición, se eliminan todas aquellas palabras no pertenecientes a la categoría de verbo, guardando los verbos restantes para poder hacer búsquedas en concatenación con el término y las palabras del contexto nominal. Hasta ese punto no hay problema si consideramos el mismo caso de virus:

*virus*

s m sing y pl

**II**

(*Inform*) Programa elaborado para **alterar** y **dañar** el funcionamiento normal de una computadora; virus cibernético

donde los verbos *alterar* y *dañar* conformarían el contexto verbal, con lo que, si nos damos cuenta, sería muy probable encontrar alguna frase referida al campo semántico *informática*. Pero si encontramos definiciones que están conformadas de acuerdo con los criterios de los contextos definitorios, entonces los verbos asociados al término recuperarían frases muy similares a dichas definiciones. Esto acarrearía el problema de que los potenciales candidatos a ejemplos sean una paráfrasis de la definición, lo que

los invalidaría como aportadores de información adicional, pues dirían casi lo mismo acerca del concepto. Por esta razón, decidimos eliminar los verbos de la lista simplificada de verbos definatorios:

Caracterizar	Describir
Comprender	Entender
Concebir	Emplear (se) en
Conocer	Permitir
Considerar	Servir (en/para)
Definir	Usar (en/para)
Denominar	Utilizar (en/para)

**Figura 6.11. Lista simplificada de verbos definatorios**

Esto nos permitió “quitarle” lo definatorio a los fragmentos candidatos y alejó el ejemplo de la definición en un sentido sintáctico.

Finalmente, cuando la definición no presentaba verbos o cuando sólo contenía verbos definatorios, estos eran eliminados con este filtro.

Cuando la lista de verbos daba como resultado cero, es decir, que no había verbos asociados al término en la definición después de aplicar el filtro anterior, se procedió a la búsqueda de verbos asociados al término para poder buscar correctamente los candidatos. La determinación de estos nuevos verbos se dio por el cálculo de la *información mutua* entre el término y todos los verbos del subcorpus que se forma a partir del contexto nominal. Esto da dos ventajas importantes:

1) que se sabe que el verbo está íntimamente asociado con el término y por ello es más probable encontrarlos juntos en una emisión comunicativa especializada, es decir, el término transfiere sus rasgos de especialidad al conjunto completo de palabras que lo rodean, lo que supone que la activación de la que habla Kuguel (2007) es bilateral y,

2) que se sabe que la relación entre el término y dicho verbo se referirá, probablemente, a un campo especializado específico, debido a que el campo semántico de donde se extrae pertenece ya en gran medida al campo de especialidad por estar conformado por

líneas de texto que contienen al término y una o más palabras de la propia definición.

## **6.5. Reglas para la extracción de candidatos a ejemplos**

En esta sección veremos las tres fases principales que lleva a cabo el sistema de extracción de candidatos a ejemplo y describiremos cómo funciona.

### **6.5.1. La utilidad de Cortex**

Cortex es un sistema de resumen automático desarrollado en la Université d'Avignon et des Pays de Vaucluse (Torres-Moreno, Velázquez-Morales & Meunier, 2002). En nuestro caso fue utilizado para poder determinar los límites de las frases que serían analizadas con GENEX, a través de un filtrado que aprovechó los módulos de preprocesamiento y presentó frases segmentadas para que el procesamiento posterior fuese más sencillo. Cortex funciona determinando límites entre frases a través de signos de puntuación y marcadores discursivos, entre otras funciones, para crear resúmenes automáticos. Dentro del sistema GENEX, Cortex es de suma utilidad por ser la herramienta de filtrado que determina cuáles frases serán las que conformen los subcorpus que, a su vez, componen los contextos nominal y verbal. Por lo anterior, hablamos de corpus de frases, antes que de corpus textual a secas. Un corpus de frases sólo es posible en la medida que éstas son debidamente segmentadas y ordenadas. Con la utilización de dichos módulos de Cortex se puede decidir si ordenar las líneas en orden alfabético o por longitud.

### **6.5.2. La función coseno**

En los algoritmos de Recuperación de Información (Baeza-Yates & Ribeiro-Neto, 1999), es común utilizar el apareamiento entre un conjunto de documentos  $\{D\}$  y una consulta ( $q$ , por su denominación en inglés query) del usuario. Los documentos

retenidos por este proceso son llamados relevantes y guardan una relación de contenido con la consulta. La consulta es expresada como un conjunto de términos:

$$\mathbf{q} = (t_1, t_2, \dots, t_n)$$

La representación en modelo vectorial tanto del conjunto  $\{D\}$  como de  $q$  es adecuada para los propósitos de búsqueda de información. En este modelo (Salton, 1989), los  $P$  documentos son representados como una matriz  $M[P \times N]$ , de  $P$  líneas y  $N$  columnas (el vocabulario), y la consulta  $q$  como un vector igualmente de  $N$  dimensiones. Cada celda de  $M_{i,j}$  corresponde al peso del término  $j$  en el documento  $i$ . Este peso puede ser binario, ya sea Term-Frequency (TF) o bien una ponderación de tipo TFxIDF (Sparck Jones, 1972). A partir de la matriz  $M$ , se pueden usar varias métricas para obtener un apareamiento entre los documentos y la consulta.

Una medida clásica de la proximidad del documento  $i$  a la consulta  $q$  es el cálculo del coseno:

$$\cos(D_i, q) = D_i \cdot q / \|D_i\| \|q\|$$

El coseno entre dos vectores representa una distancia entre ellos: cuanto más pequeña es la distancia (el coseno del ángulo tiende a 1) más próximos se encuentran en el espacio de representación. Una hipótesis fuerte es que si dos puntos en el espacio  $N$ -dimensional están próximos entre sí, probablemente están próximos igualmente en el espacio semántico.

En el caso que nos interesa, los documentos corresponden a las frases del corpus recuperadas, y la consulta es el conjunto de palabras de la definición (sin las palabras funcionales) y que además incluyen el término a ejemplificar. De esta manera, las frases cuyo coseno respecto a la consulta es más próximo de 1, serán bonificadas. Esto permite obtener una clasificación de los candidatos a ejemplo.

### 6.5.3. El corte por longitud

De acuerdo con esta metodología, los fragmentos textuales que contengan el término, una cantidad  $n$  de palabras contenidas en la definición y un verbo asociado al término serán los candidatos más fuertes para utilizarse como ejemplos. Sin embargo, esto es hasta aquí una verdad a medias, ya que otro de los principios que determinan que un ejemplo sea bueno es justamente su longitud. Observaciones sobre la marcha en el diseño de esta herramienta nos hicieron caer en la cuenta de que la cantidad de información que aporta un ejemplo a veces es proporcional a la distribución de sus elementos y su longitud. Es decir, hubo fragmentos largos, de más de veinte o veinticinco tokens que cumplían con las condiciones de los filtros que hemos descrito, pero lo cierto es que no siempre un ejemplo muy extenso es el más adecuado. Hay que tener en cuenta, también, que las frases más extensas son más propicias a contener información de otros campos, lo que, lejos de ayudar al lector a comprender el concepto, lo pueden opacar más, y que en el mejor de los casos es información adicional que muchas veces es redundante o prescindible.

Por lo anterior, se decidió que habría que marcar un umbral tanto en el número de candidatos elegibles como en la extensión en número de tokens de cada candidato. Con esta idea en mente, una vez que el procesamiento del GENEX llegó al momento de la elección, se decidió que habría que combinar los métodos de evaluación para que, en conjunto con los métodos de extracción, presentaran los mejores resultados. A esta combinación y su medida asociada fue lo que llamamos *densidad lexicométrica*.

### 6.5.4. La aplicación de la *densidad lexicométrica*

La *densidad lexicométrica* es un producto directo de la noción de *saturación semántica*. Para medirla, como hemos visto hasta ahora, es necesario tener en cuenta que el ejemplo guarda una relación estrecha con la definición y que los elementos de la primera deberían aparecer idealmente en el segundo, aunque no en un estricto orden sintáctico ni bajo las mismas reglas lexicológicas, esto es, puede haber variación de forma.

La aparición de los elementos de la definición en el ejemplo atiende a dos cuestiones principales.

Por un lado, la inclusión del término en el ejemplo le da al fragmento un núcleo conceptual sobre el cual giran los elementos de la definición que al final serán los satélites y los límites de significación de ese término. Si se quiere ver de este modo, la combinación de las palabras de una definición en una nueva estructura no definitoria permite que el contexto de activación del término siga operando, y concede al fragmento una pertenencia a un campo semántico determinado, que en este caso será un área de especialidad.

Por otro lado, al quitarle a la definición su estructura a través de la eliminación de los verbos definitorios, conlleva que, en consecuencia, el nuevo fragmento carezca también de estructura. Para solventar esto se ha decidido colocar en esta nueva estructura el verbo semánticamente más cercano al término. Dicho verbo puede existir *a priori* en la definición o puede ser producto de una búsqueda hecha con la medida de *información mutua*. Todo esto, si se quiere ver así, sirve para dar estabilidad a la nueva estructura.

La *densidad lexicométrica* calcula qué tanta información tiene un fragmento textual y qué tanta cercanía semántica tiene con respecto a otro fragmento, de tal manera que lo puede complementar en el plano conceptual. Opera bajo un mecanismo que da fe de un núcleo conceptual, el término, y un núcleo sintáctico, el verbo, que se ven envueltos en una estructura cuyo léxico lo inclina a una temática determinada (el contexto nominal). La *densidad lexicométrica* se puede considerar como el producto del valor coseno de una frase con respecto a otra por el inverso del logaritmo de su longitud. Gráficamente la fórmula para determinarla se presenta a continuación y se describe en los siguientes apartados:

$$score_i = \cos(\text{sentence}_i, q) \times \frac{1}{\log|\text{sentence}_i|}$$

## 6.6. El algoritmo general de extracción

El sistema GENEX utiliza cuatro recursos básicos:

- 1) Un diccionario general de la lengua (DEM)
- 2) Un conjunto de corpus de tamaño adecuado (*Jornada, Wiki o esTenTen*)
- 3) Un etiquetador morfológico (TreeTagger)
- 4) Un programa para calcular la *información mutua* (minfo.pl, cuyo funcionamiento detallamos adelante)

La arquitectura general del GENEX se muestra en la siguiente figura:

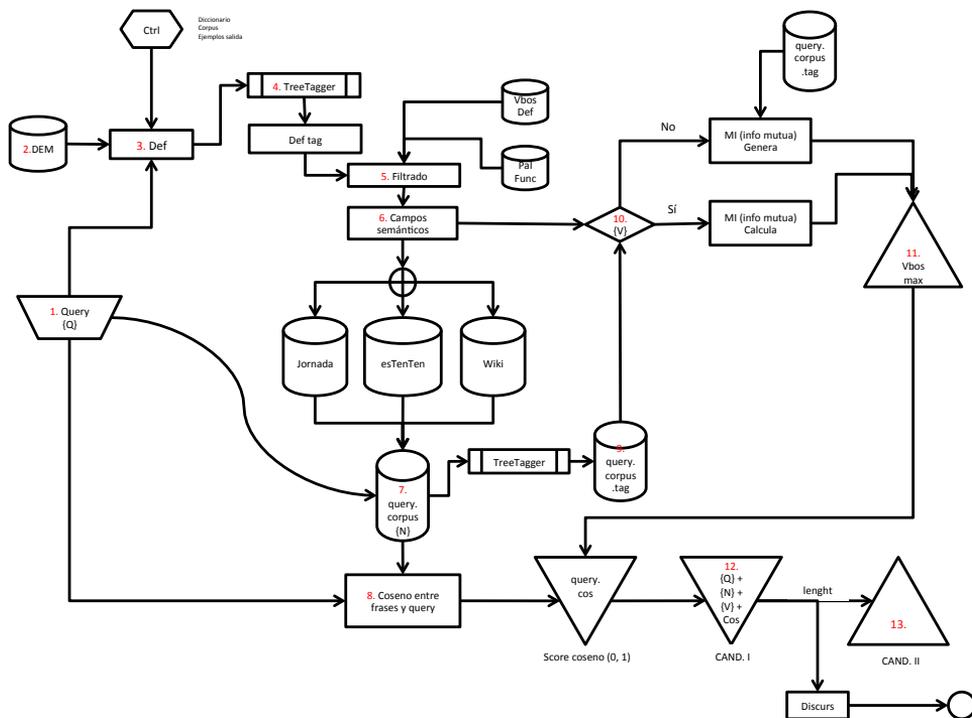


Figura 6.12. Arquitectura del GENEX

Bajo este esquema, el funcionamiento de GENEX es el siguiente:

1. INPUT > Se introduce una consulta q (en este caso, un término).
2. El sistema extrae del DEM las definiciones y acepciones asociadas a esa consulta. En el archivo de control se pueden cambiar el diccionario de entrada, la forma de normalización: lematización, stemming o ultrastemming (Torres-Moreno, 2012), el número de resultados y el umbral del coseno.
3. El sistema pide que se elija una de las acepciones que ha traído.
4. La acepción elegida es etiquetada con TreeTagger.
5. Una vez etiquetada, la acepción es filtrada:
  - a. Elimina todas las palabras funcionales (preposiciones, conjunciones, etc.) de una lista de paro predeterminada (PAL FUNC).
  - b. Elimina todos los verbos definitorios y sus derivados (VBOS DEF).
  - c. Guarda la lista de palabras restante junto con sus etiquetas.
6. Crea dos nuevas listas que determinarán los campos semánticos {N} y {V}, donde:
  - a. {N} o “Campo semántico nombre”, está compuesto por el conjunto de líneas encontradas en el corpus de entrada (*Jornada, Wiki o esTenTen*) que contienen una o más palabras de la lista de los elementos obtenidos en el paso anterior, exceptuando los verbos, es decir, aquellas unidades que tengan las etiquetas /NC, /NP o /ADJ y sus derivados.
  - b. {V} o “Campo semántico verbo”, está compuesto por el conjunto de los verbos asociados al término. Esto es, todos aquellos encontrados en la definición.
7. Una vez que ha obtenido {N} guarda ese subcorpus como query.corpus.
8. Calcula el coseno (cos) de las frases y la consulta q. Esto es, determina cuántas palabras dadas en {N} están presentes en cada

frase y que además contengan q. De acuerdo con este criterio identifica una frase ideal, después crea una matriz para, a partir de ella, asignar una calificación a cada frase extraída. Desecha las que no contengan el criterio  $\geq I\{N\}+q$  (la frase tiene por lo menos un elemento de la lista  $\{N\}$  y además la consulta). Guarda el resultado en *query.cos*.

9. Toma el subcorpus *query.corpus* y etiqueta con TreeTagger. Guarda como *query.corpus.tag* y se prepara para realizar el análisis morfosintáctico.

10. Toma una decisión:

SÍ) Si hay un conjunto de verbos extraídos de la definición, calcula la información mutua que guardan con q tomando como base *query.corpus.tag* y los almacena ordenándolos de mayor a menor valor.

NO) Si no hay una lista de verbos extraídos, entonces genera una lista nueva calculando la *información mutua* de q con todas las palabras en *query.corpus.tag*. Retiene los 3 primeros ítems con la etiqueta “verbo” (/V y sus derivados) que tengan la calificación más alta y los guarda.

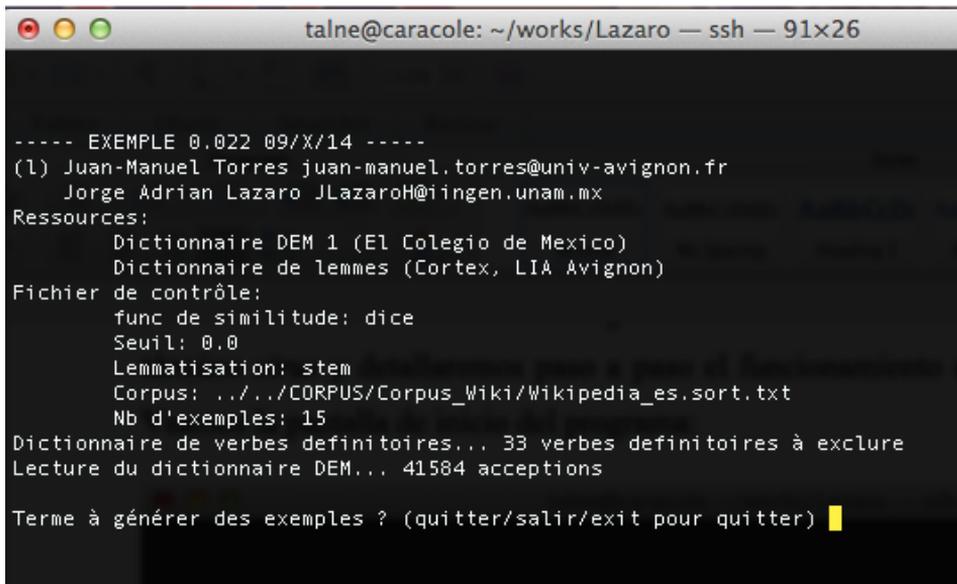
11. De los verbos que obtiene, toma el primero y busca todas las líneas en *query.cos* que lo contengan en cualquiera de sus conjugaciones. Guarda esa lista y presenta como CANDIDATOS I (*candidatos coseno*).

12. Calcula la cantidad de tokens que tiene cada línea y ordena la lista de CANDIDATOS I de tal manera que el primero sea el más corto con más alta calificación, es decir, ordena de forma ascendente por el número de tokens, pero descendente por su calificación de *cos* (*candidatos length*) que es igual al producto de *length-1(cos)* o lo que es lo mismo, su *densidad lexicométrica*.

13. OUTPUT > Toma la lista anterior y guarda sólo los 15 primeros elementos, los llama CANDIDATOS II y los presenta como “Ejemplos de \$query”.

## 6.7. El experimento

En este apartado describiremos cómo se ha realizado el experimento para cada uno de los 15 términos elegidos<sup>49</sup>. Tomaremos como modelo el término *virus* y detallaremos paso a paso el funcionamiento de GENEX<sup>50</sup>. Veamos la pantalla de inicio del programa:



```
----- EXEMPLE 0.022 09/X/14 -----
(1) Juan-Manuel Torres juan-manuel.torres@univ-avignon.fr
    Jorge Adrian Lazaro JLazaroH@ingen.unam.mx
Ressources:
    Dictionnaire DEM 1 (El Colegio de Mexico)
    Dictionnaire de lemmes (Cortex, LIA Avignon)
Fichier de contrôle:
    func de similitude: dice
    Seuil: 0.0
    Lemmatisation: stem
    Corpus: ../../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
    Nb d'exemples: 15
Dictionnaire de verbes definitives... 33 verbes definitives à exclure
Lecture du dictionnaire DEM... 41584 acceptations

Terme à générer des exemples ? (quitter/salir/exit pour quitter) █
```

Al momento de iniciar el programa podemos mirar las especificaciones bajo las cuales trabajará. En primer lugar vemos que los recursos que utilizará son el diccionario *Diccionario del Español de México* (DEM) del Colegio de México (Lara, 2010), un conjunto de módulos del programa Cortex (Torres-Moreno, Velázquez-Morales & Meunier, 2002). Para el caso del fichero de control contamos con la función de similitud *dice*; el umbral para la medida de coseno que en francés se denomina *seuil*, el lematizador utilizado que hemos mencionado, *ultrastemming* (Torres-Moreno,

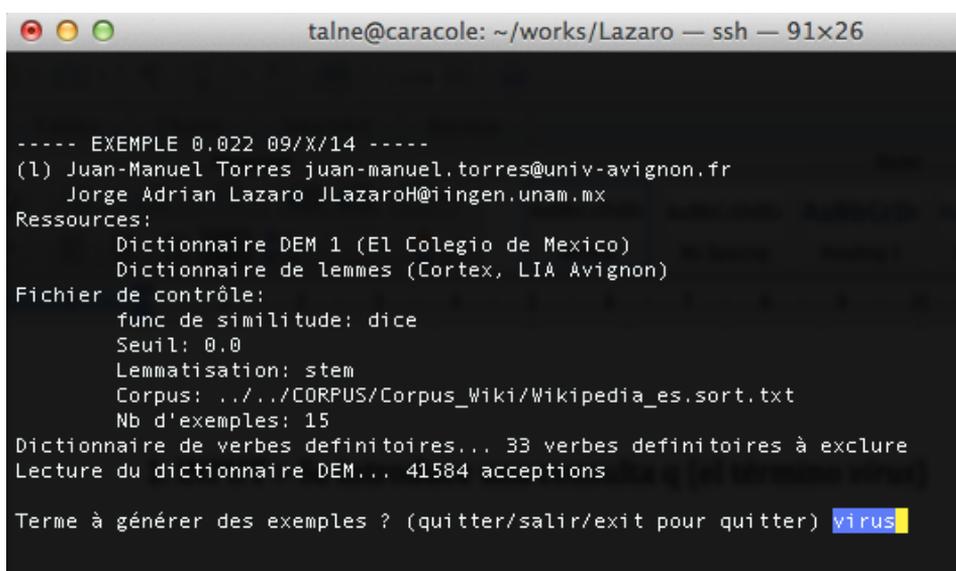
---

<sup>49</sup> Los resultados de los 15 términos elegidos se pueden consultar en el Anexo I

<sup>50</sup> Es indispensable mencionar que, aunque la mayoría del proceso se puede ver reflejado en las etapas que mostraremos, no todos los cálculos son visibles en pantalla, por lo que en los pasos correspondientes haremos la aclaración de los procesos que no se ven pero que se reflejan en los resultados que irá arrojando el programa.

2012); el corpus, que en este caso será *Wiki*, y el número de candidatos a ejemplos que queremos que conserve después de hacer la recuperación, en este caso los 15 primeros. Veremos también que toma en cuenta un diccionario de verbos definitorios a excluir, que son los que mostramos en la figura 2.4. de esta tesis y que están documentados en Aguilar (2009). Finalmente se puede notar que indica que el número de acepciones con las que se puede trabajar del DEM son 41 584.

1. INPUT > Se introduce una consulta q (el término *virus*).



```
talne@caracole: ~/works/Lazaro — ssh — 91x26
----- EXEMPLE 0.022 09/X/14 -----
(1) Juan-Manuel Torres juan-manuel.torres@univ-avignon.fr
    Jorge Adrian Lazaro JLazaroH@ingen.unam.mx
Ressources:
    Dictionnaire DEM 1 (El Colegio de Mexico)
    Dictionnaire de lemmes (Cortex, LIA Avignon)
Fichier de contrôle:
    func de similitude: dice
    Seuil: 0.0
    Lemmatisation: stem
    Corpus: ../../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
    Nb d'exemples: 15
Dictionnaire de verbes definatoires... 33 verbes definatoires à exclure
Lecture du dictionnaire DEM... 41584 acceptions

Terme à générer des exemples ? (quitter/salir/exit pour quitter) virus
```

2. El sistema extrae del DEM las definiciones y acepciones asociadas a esa consulta.

```
talne@caracole: ~/works/Lazaro — ssh — 89x26

----- EXEMPLE 0.022 09/X/14 -----
(1) Juan-Manuel Torres juan-manuel.torres@univ-avignon.fr
    Jorge Adrian Lazaro JLazaroH@ingen.unam.mx
Ressources:
    Dictionnaire DEM 1 (El Colegio de Mexico)
    Dictionnaire de lemmes (Cortex, LIA Avignon)
Fichier de contrôle:
    func de similitude: dice
    Seuil: 0.0
    Lemmatisation: stem
    Corpus: ../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
    Nb d'exemples: 15
Dictionnaire de verbes definitives... 33 verbes definitives à exclure
Lecture du dictionnaire DEM... 41584 acceptations

Terme à générer des exemples ? (quitter/salir/exit pour quitter) virus
virus possède les suivantes acceptations:
1.- Partícula o ser microscópico compuesto por proteínas que rodean a un ácido nucleico,
parásito de las células, donde puede reproducirse y mutar; es capaz de infectar a todo ser
viviente
2.- Programa elaborado para alterar y dañar el funcionamiento normal de una computadora;
virus cibernético
0.- AUTRE MOT
Acception?
```

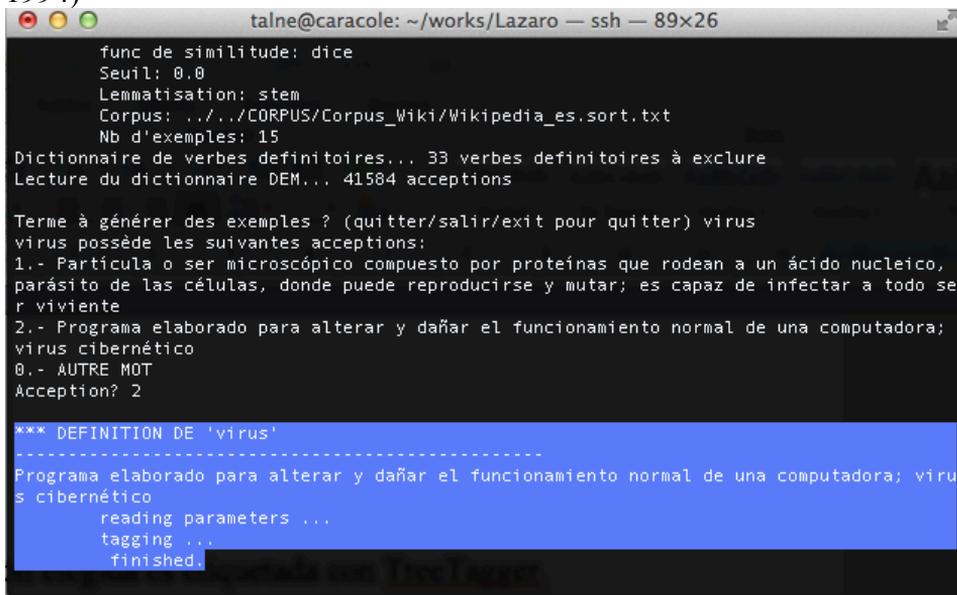
3. El sistema pide que se elija una de las acepciones que ha traído. En este caso elegimos la 2, que se refiere a *virus* en el área de informática.

```
talne@caracole: ~/works/Lazaro — ssh — 89x26

----- EXEMPLE 0.022 09/X/14 -----
(1) Juan-Manuel Torres juan-manuel.torres@univ-avignon.fr
    Jorge Adrian Lazaro JLazaroH@ingen.unam.mx
Ressources:
    Dictionnaire DEM 1 (El Colegio de Mexico)
    Dictionnaire de lemmes (Cortex, LIA Avignon)
Fichier de contrôle:
    func de similitude: dice
    Seuil: 0.0
    Lemmatisation: stem
    Corpus: ../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
    Nb d'exemples: 15
Dictionnaire de verbes definitives... 33 verbes definitives à exclure
Lecture du dictionnaire DEM... 41584 acceptations

Terme à générer des exemples ? (quitter/salir/exit pour quitter) virus
virus possède les suivantes acceptations:
1.- Partícula o ser microscópico compuesto por proteínas que rodean a un ácido nucleico,
parásito de las células, donde puede reproducirse y mutar; es capaz de infectar a todo ser
viviente
2.- Programa elaborado para alterar y dañar el funcionamiento normal de una computadora;
virus cibernético
0.- AUTRE MOT
Acception? 2
```

#### 4. La acepción elegida es etiquetada con TreeTagger (Schmid, 1994)



```
talne@caracole: ~/works/Lazaro — ssh — 89x26
func de similitude: dice
Seuil: 0,0
Lemmatisation: stem
Corpus: ../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
Nb d'exemples: 15
Dictionnaire de verbes definitoires... 33 verbes definitoires à exclure
Lecture du dictionnaire DEM... 41584 acceptions

Terme à générer des exemples ? (quitter/salir/exit pour quitter) virus
virus possède les suivantes acceptions:
1.- Particula o ser microscópico compuesto por proteínas que rodean a un ácido nucleico,
parásito de las células, donde puede reproducirse y mutar; es capaz de infectar a todo se
r viviente
2.- Programa elaborado para alterar y dañar el funcionamiento normal de una computadora;
virus cibernético
0.- AUTRE MOT
Acception? 2

*** DEFINITION DE 'virus'
-----
Programa elaborado para alterar y dañar el funcionamiento normal de una computadora; viru
s cibernético
reading parameters ...
tagging ...
finished.
```

#### 5. Una vez etiquetada, la acepción es filtrada:

- a. Elimina todas las palabras funcionales (preposiciones, conjunciones, etc.) de una lista de paro predeterminada (PAL FUNC).
- b. Elimina todos los verbos definitorios y sus derivados (VBOS DEF).
- c. Guarda la lista de palabras restante junto con sus etiquetas.

```

talne@caracole: ~/works/Lazaro — ssh — 89x26
Corpus: ../../CORPUS/Corpus_Wiki/Wikipedia_es.sort.txt
Nb d'exemples: 15
Dictionnaire de verbes definitoires... 33 verbes definitoires à exclure
Lecture du dictionnaire DEM... 41584 acceptions

Terme à générer des exemples ? (quitter/salir/exit pour quitter) virus
virus possède les suivantes acceptions:
1.- Particula o ser microscópico compuesto por proteínas que rodean a un ácido nucleico,
parásito de las células, donde puede reproducirse y mutar; es capaz de infectar a todo se
r viviente
2.- Programa elaborado para alterar y dañar el funcionamiento normal de una computadora;
virus cibernético
0.- AUTRE MOT
Acception? 2

*** DEFINITION DE 'virus'
-----
Programa elaborado para alterar y dañar el funcionamiento normal de una computadora; viru
s cibernético
  reading parameters ...
  tagging ...
  finished.

*** DEFINICION FILTRADA de 'virus'
programa\NC alterar\VLinf dañar\VLinf funcionamiento\NC normal\ADJ computadora\NC virus\N
C cibernético\ADJ

```

6. Crea dos nuevas listas que determinarán los campos semánticos {N} y {V}, donde:

- a. {N} o “Campo semántico nombre”, está compuesto por el conjunto de líneas encontradas en el corpus de entrada (*Wiki*) que contienen una o más palabras de la lista de los elementos obtenidos en el paso anterior, exceptuando los verbos, es decir, aquellas unidades que tengan las etiquetas /NC, /NP o /ADJ y sus derivados: **programa\NC**, **funcionamiento\NC**, **normal\ADJ**, **computadora\NC**, **virus\NC** y **cibernético\ADJ**
- b. {V} o “Campo semántico verbo”, está compuesto por el conjunto de los verbos asociados al término. Esto es, todos aquellos encontrados en la definición: **alterar** y **dañar**.

```

talne@caracole: ~/works/Lazaro — ssh — 89x26
r viviente
2.- Programa elaborado para alterar y dañar el funcionamiento normal de una computadora;
virus cibernético
0.- AUTRE MOT
Acception? 2

*** DEFINITION DE 'virus'
-----
Programa elaborado para alterar y dañar el funcionamiento normal de una computadora; viru
s cibernético
  reading parameters ...
  tagging ...
  finished.

*** DEFINICION FILTRADA de 'virus'
programa\NC alterar\VLinf dañar\VLinf funcionamiento\NC normal\ADJ computadora\NC virus\NC
C cibernético\ADJ

*** CAMPO SEMANTICO: VERBOS ASOCIADOS
V* = ( alterar, dañar, )

*** CAMPO SEMANTICO: NOMBRES ASOCIADOS
N* = ( cibernético, computadora, funcionamiento, normal, programa, virus, )
  reading parameters ...
  tagging ...
  finished.

```

7. Una vez que ha obtenido {N}, lematiza cada una de las formas, busca todas las concordancias con esas formas y guarda ese subcorpus como *virus.corpus*. Aquí {N} es igual a {**cibern| comput| funcio| normal| progra| }**

```

talne@caracole: ~/works/Lazaro — ssh — 95x25
Acception? 2

*** DEFINITION DE 'virus'
-----
Programa elaborado para alterar y dañar el funcionamiento normal de una computadora; virus cibe
rnético
  reading parameters ...
  tagging ...
  finished.

*** DEFINICION FILTRADA de 'virus'
programa\NC alterar\VLinf dañar\VLinf funcionamiento\NC normal\ADJ computadora\NC virus\NC cibe
rnético\ADJ

*** CAMPO SEMANTICO: VERBOS ASOCIADOS
V* = ( alterar, dañar, )

*** CAMPO SEMANTICO: NOMBRES ASOCIADOS
N* = ( cibernético, computadora, funcionamiento, normal, programa, virus, )
  reading parameters ...
  tagging ...
  finished.

*** Creando virus.corpus usando terminos: ( cibern| comput| funcio| normal| progra) + ' virus '
258 ../GENEX/virus.corpus

```

8. Calcula el coseno (cos) de las frases y la consulta *virus*. Esto es, determina cuántas palabras dadas en {N} están presentes en cada frase y que además contengan *virus*. De acuerdo con este criterio

identifica una frase ideal, después crea una matriz para, a partir de ella, asignar una calificación a cada frase extraída. Desecha las que no contengan el criterio  $\geq I\{N\}+q$  (la frase tiene por lo menos un elemento de la lista {**cibern**| **comput**| **funcio**| **normal**| **progra**} y además la consulta *virus*). Guarda el resultado en *virus.cos*. [Proceso oculto]

9. Toma el subcorpus *virus.corpus* y etiqueta con TreeTagger. Guarda como *query.corpus.tag* y prepara para realizar el análisis morfosintáctico.

```

talne@caracole: ~/works/Lazaro — ssh — 95x25
tagging ...
finished.

*** DEFINICION FILTRADA de 'virus'
programa\NC alterar\VLinf dañar\VLinf funcionamiento\NC normal\ADJ computadora\NC virus\NC cibe
rnético\ADJ

*** CAMPO SEMANTICO: VERBOS ASOCIADOS
V* = { alterar, dañar, }

*** CAMPO SEMANTICO: NOMBRES ASOCIADOS
N* = { cibernético, computadora, funcionamiento, normal, programa, virus, }
reading parameters ...
tagging ...
finished.

*** Creando virus.corpus usando terminos: ( cibern| comput| funcio| normal| progra) + ' virus '
258 ../GENEX/virus.corpus

*** Etiqueta subcorpus virus.corpus...
reading parameters ...
tagging ...
11000 finished.

*** Sub-corpus virus.tag etiquetée

```

10. Toma una decisión:

SÍ) Si hay un conjunto de verbos extraídos de la definición, calcula la *información mutua* que guardan con *virus* tomando como base *query.corpus.tag* y los almacena ordenándolos de mayor a menor valor.

NO) Si no hay una lista de verbos extraídos, entonces genera una lista nueva calculando la *información mutua* de *virus* con todas las palabras en *virus.tag*. Retiene los 3 primeros ítems con la etiqueta “verbo” (/V y sus derivados) que tengan la calificación más alta y los guarda.

```

talne@caracole: ~/works/Lazaro — ssh — 95x25
rnético\ADJ

*** CAMPO SEMANTICO: VERBOS ASOCIADOS
V* = { alterar, dañar, }

*** CAMPO SEMANTICO: NOMBRES ASOCIADOS
N* = { cibernético, computadora, funcionamiento, normal, programa, virus, }
      reading parameters ...
      tagging ...
      finished.

*** Creando virus.corpus usando terminos: ( cibern| comput| funcio| normal| progra) + ' virus '
258 ../GENEX/virus.corpus

*** Etiqueta subcorpus virus.corpus...
      reading parameters ...
      tagging ...
11000 finished.

*** Sub-corpus virus.tag etiqueté
Mutual information alter-virus...
Mutual information dañ-virus...
dañ      -230.853231334983
alter    -420.497823377825

```

11. De los verbos que obtiene, toma el primero y busca todas las líneas en *virus.cos* que lo contengan en cualquiera de sus conjugaciones. Guarda esa lista y presenta como CANDIDATOS I (*candidatos coseno*).

```

talne@caracole: ~/works/Lazaro — ssh — 95x25
*** Sub-corpus virus.tag etiqueté
Mutual information alter-virus...
Mutual information dañ-virus...
dañ      -230.853231334983
alter    -420.497823377825

*** CANDIDATOS score: COSENO
1.      En este rubro se incluyen : adware , dialers , herramientas de acceso remoto , programa
s de bromas ( " Virus joke " ) , programas para conferencias , programa de mensajería instantá
nea , spyware y cualesquiera otros archivos y programas no bienvenidos que no sean virus y que p
uedan llegar a dañar el funcionamiento de una computadora o de una red .      0.10958763535409
8
2.      Algunos de estos cracks pueden ser en realidad " software " malintencionado que puede i
ncorporar vulnerabilidades al programa original o contener virus o troyanos , lo cual puede pro
vocar serios daños a un computador , hasta que incluso también puede convertir un computador en
parte de una " botnet " .      0.103941057523732
3.      Algunos de los primeros programas infecciosos , incluido el Gusano Morris y algunos vir
us de MS - DOS , fueron elaborados como experimentos , como bromas o simplemente como algo mole
sto , no para causar graves daños en las computadoras .      0.0875754513580922
4.      Ella es una joven de 14 años , una estudiante de secundaria no muy buena en todo lo que
tenga que ver con las matemáticas y la informática , Yui soñaba con ser una súper heroína para
salvar al mundo de las fuerzas del mal , hasta que un día mientras hacia su tarea en su comput
adora , algo muy extraño le pasó , ya que ella había sido absorbida por el computador , estando
dentro una criatura semejante a un Mapache llamado I. R. se le acerca y le dice que ella es l
a encargada de salvar a la RedCom y combatir contra los corruptores que son un grupo de softwar

```

12. Calcula la cantidad de tokens que tiene cada línea y ordena la lista de CANDIDATOS I de tal manera que el primero sea el más corto con más alta calificación, es decir, ordena de forma ascendente por el número de tokens, pero descendente por su

calificación de  $\cos$  (*candidatos length*) que es igual al producto de  $\text{length} \cdot 1(\cos)$  o lo que es lo mismo, su *densidad lexicométrica*.

13. OUTPUT > Toma la lista anterior y guarda sólo los 15 primeros elementos, los llama CANDIDATOS II y los presenta como “Ejemplos de *virus*”.

```

talne@caracole: ~/works/Lazaro — ssh — 95x25
portal . 0.0773626427600916
*** CANDIDATOS score: LONGITUD
Tambi?n analiza los correos que recibimos en busca de virus y programas da?inos . 0.08472
97992509235 15
- A no introducir ni difundir virus inform?ticos o programas da?inos en o a trav?s del portal .
0.0773626427600916 19
( f ) Da?os ocasionados por virus inform?ticos , o programas similares , transmitidos a trav?s
del Portal . 0.0789233886036442 20
contenga virus u otro componente da?ino capaz de interrumpir , destruir o limitar la funcionali
dad de cualesquiera programas o equipos inform?ticos . 0.119936173616495 23
o Subir archivos que contengan virus , archivos corruptos , o cualquier otro software o program
as que puedan da?ar la operaci?n del computador de otro . 0.0938128845247817 27
Cargar o adjuntar archivos que contengan virus , archivos da?ados o cualquier otro software sim
ilar o programas que puedan da?ar el funcionamiento de otros ordenadores . 0.0926231433647
117 27
Descargar o cargar archivos que puedan da?ar la operaci?n de la computadora de otros , como vir
us de computadora , archivos corruptos o alg?n software similar . 0.0837807854847396 2
8
El usuario se compromete a no introducir programas , virus , controles active-x o cualquier otr
o dispositivo que causen da?o al funcionamiento inform?tico de esta p?gina . 0.0878700291230
402 28
Introducir virus inform?ticos , archivos defectuosos , o cualquier otro programa inform?tico qu
e pueda provocar da?os o alteraciones en los contenidos , programas o sistemas de AluminioMallo
rca . 0.0837807854847396 29

```

Los resultados finales son ordenados por su *densidad lexicométrica*, ésta se obtiene a través del producto de su coseno (valor en negritas) por su longitud (valor en negritas y subrayado). A efectos prácticos hemos dejado ambos valores para una mejor visualización de dichos resultados. Para el caso de *virus* son los siguientes:

1. También analiza los correos que recibimos en busca de virus y programas dañinos. **0.0847297992509235** 15
2. - A no introducir ni difundir virus informáticos o programas dañinos en o a través del portal. **0.0773626427600916** 19
3. ( f ) Daños ocasionados por virus informáticos , o programas similares , transmitidos a través del Portal. **0.0789233886036442** 20

4. contenga virus u otro componente dañino capaz de interrumpir, destruir o limitar la funcionalidad de cualesquiera programas o equipos informáticos. **0.119936173616495 23**
5. Subir archivos que contengan virus, archivos corruptos, o cualquier otro software o programas que puedan dañar la operación del computador de otro. **0.0938128845247817 27**
6. Cargar o adjuntar archivos que contengan virus, archivos dañados o cualquier otro software similar o programas que puedan dañar el funcionamiento de otros ordenadores. **0.0926231433647117 27**
7. Descargar o cargar archivos que puedan dañar la operación de la computadora de otros, como virus de computadora, archivos corruptos o algún software similar. **0.0837807854847396 28**
8. El usuario se compromete a no introducir programas, virus, controles active-x o cualquier otro dispositivo que causen daño al funcionamiento informático de esta página. **0.0878700291230402 28**
9. Introducir virus informáticos, archivos defectuosos, o cualquier otro programa informático que pueda provocar daños o alteraciones en los contenidos, programas o sistemas de Aluminio Mallorca. **0.0837807854847396 29**
10. Pero no porque, como los virus, dañe programas de la computadora, borre datos cruciales o sature casillas de correo propias y ajenas con mensajes infectados. **0.103760492651546 30**
11. Los programas culpables de infecciones de virus además también pueden dañar el sector de carga del disco duro de la computadora, y esto puede evitar que el sistema arranque. **0.0929754232206007 32**
12. contengan virus , caballos de Troya, gusanos, bombas de tiempo o cualquier otro material o programa que pueda dañar o alterar el funcionamiento de equipos o sistemas informáticos. **0.0774859072901217 32**
13. Los virus, en general son programas que se ejecutan en nuestro ordenador y generan instrucciones que ocasionan daños en los

programas y / o sistema operativo, e incluso en el propio aparato.  
**0.0808929217232955 36**

14. Programación: Uso Emacspeak para escribir programas al durante mi aprendizaje de Navegar la Internet: Orca y Firefox me permiten navegar la internet sin temor a contraer un virus electrónico que me dañaría mi computadora bajo Windows.  
**0.0894537400462454 41**

15. Tal y como explica web de la gran compañía Microsoft los virus informáticos son: "programas diseñados expresamente para interferir en el funcionamiento de una computadora, registrar, dañar o eliminar datos, o bien para propagarse a otras computadoras...  
**0.13998770789824 45**

## **6.8. La evaluación**

Aunque existen varios métodos automáticos para medir el rendimiento de un sistema informático, para este experimento nos basamos casi completamente en el criterio de hablantes nativos de español con el fin de hacer un primer acercamiento a la efectividad del sistema GENEX. A través de la comprobación de lo extraído por GENEX y la confirmación o refutación humana, hemos derivado algunas conclusiones que nos permitirán, en el futuro, escalar y mejorar el sistema.

Para determinar la efectividad del sistema se hizo una encuesta que constaba de 15 ejemplos por cada término (5 de cada corpus), es decir, una muestra de 225 elementos cuya función era determinar qué tanto el sistema se acercaba a lo que un hablante medio de español considera un buen ejemplo. Los sujetos de evaluación, 50 hablantes nativos de español, tenían entre 18 y 70 años, 31 mujeres y 19 varones; su rango de educación es entre bachillerato (6), universidad (31) y posgrado (13), es decir, diastáticamente pertenecientes al rango de hablantes cultos. Su función fue elegir entre 5 opciones (de los candidatos que generó el sistema), el que ellos consideraban como el mejor ejemplo para cada término. Tuvieron 3 secciones, una por cada corpus. Los ejemplos en dichas secciones se ordenaron de manera descendente de acuerdo con la puntuación determinada por su *densidad lexicométrica*  $[\cos(\text{length}-1)]$ , esto es, los más cercanos

semánticamente a la definición y más cortos de acuerdo con la longitud de palabras.

La medida que utilizamos para medir la efectividad del sistema fue el *Coficiente de correlación de Pearson* (Egghe & Leydesdorff, 2009), que es una medida de la relación lineal entre dos variables cuantitativas. Esta medida nos permitió mirar el grado de relación de dos variables: la pertinencia de los ejemplos que hace el sistema de acuerdo con los criterios mencionados, y la preferencia de los hablantes con respecto a la información que esos ejemplos aportan acerca de la definición. El *Coficiente de correlación de Pearson* se simboliza como  $\rho_{X,Y}$  y se puede determinar a través de la fórmula:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

donde  $\sigma_{XY}$  es la covarianza de (X,Y),  $\sigma_X$  es la desviación típica de la variable X, y  $\sigma_Y$  es la desviación típica de la variable Y.

Los resultados obtenidos se pueden observar en la siguiente tabla:

		COEFICIENTE DE CORRELACIÓN DE PEARSON					
		JORNADA		WIKIPEDIA		esTenTen	
		Término	Area	Término	Area	Término	Area
sexualidad	sexualidad	-0,2	-0,2	0,8	0,04	0,9	0
	sexo	-0,3		0,5		-0,7	
	aborto	0,7		-0,5		0	
	matrimonio	-0,8		-0,8		0,2	
	embarazo	-0,4		0,2		0,2	
finanzas	banco	-0,1	0,24	0,5	0,16	-0,1	0,1
	crédito	0,3		1		-0,8	
	nómina	0,5		-0,6		0,7	
	préstamo	0,3		0,2		0,2	
	pensión	0,2		-0,3		0,5	
informática	virus	0,6	-0,14	0,1	0,26	0	0,08
	computadora	-0,1		-0,3		0,3	
	programa	-0,6		0,9		-0,2	
	teclado	0,2		0,2		0,7	
	informática	-0,8		0,4		-0,4	
<b>PROMEDIO</b>		<b>-0,0333333</b>		<b>0,15333333</b>		<b>0,1</b>	

**Figura 6.13. Resultados de la evaluación del GENEX de acuerdo con el *Coficiente de correlación de Pearson***

En la tabla de resultados podemos ver, en a primera columna, las tres áreas que a las que pertenecen los términos que hemos utilizado para esta evaluación. A cada una de las áreas le

corresponden cinco términos. Estos términos están enlistados en la segunda columna. La fila superior indica los resultados obtenidos en cada uno de los tres corpus. Cada uno de los corpus muestra dos columnas. La primera indica el *Coefficiente de correlación de Pearson* para cada término, y la segunda indica la misma medida pero para cada área temática. Así, se habrá de leer que, por ejemplo, *sexo* tiene un coeficiente de -0.3 y el conjunto de términos referidos al área de sexualidad, esto es, *sexualidad, sexo, aborto, matrimonio* y *embarazo* tienen un coeficiente de -0,2. El promedio que se muestra en la última fila corresponde al coeficiente obtenido por cada corpus.

Los datos arrojados deben leerse como la correlación existente entre lo que dice el sistema y lo que prefieren los hablantes. Cuando el resultado se acerca a 1 quiere decir que el sistema ha acertado en la selección de un candidato a ejemplo, cuanto más alejado esté significa que el sistema no logró ordenar tan precisamente los candidatos.

Es de notar que para el corpus *JORNADA* el sistema tuvo serias dificultades para encontrar buenos candidatos a ejemplo. Esto se puede explicar por dos razones básicas:

1) es un corpus periodístico que tiene muy acotadas sus áreas y su lenguaje se acerca mucho a lo especializado. Los resultados, así, se ven reflejados en frases muy especializadas para un hablante lego

2) las temáticas centrales de este diario son las finanzas, la economía y la política: los resultados del área *finanzas* son positivos, mientras el coeficiente de las áreas de sexualidad e informática cae drásticamente.

Por otra parte, el corpus *Wikipedia* fue en el que el sistema pudo trabajar mejor y obtener mejores candidatos a ejemplo de acuerdo con la evaluación humana. Incluso es considerable el hecho de que no se dio una correlación lineal ( $r=0$ ) como sí ocurrió en *esTeTen*, o se tuvo una correlación negativa ( $-1 < r < 0$ ) como en *JORNADA*.

Finalmente, en el corpus *esTenTen*, GENEX tuvo resultados poco alentadores pero encontró información para generar algunos ejemplos, sobre todo en el área de informática. Suponemos que su bajo rendimiento se debe al hecho de que es una muestra pequeña (5 Gigabytes) de un corpus mucho más grande (16 Gigabytes) lo que pudo ocasionar que la muestra extraída sea poco representativa ya que no se controló su variedad ni su balance.

De manera general podemos anotar que, de acuerdo con el método con el que se programó la herramienta, un buen ejemplo se encuentra en el umbral arriba de 0.2 de acuerdo con el *Coefficiente de correlación de Pearson*, lo que indica que los mejores candidatos a ejemplos son aquellos cuya *densidad lexicométrica* se encuentra sobre el 0.24 y que tienen una extensión media de 15 tokens. Aún así, una revisión manual nos mostró que los ejemplos cuya extensión rozaba los 30 tokens de longitud tenían una precisión de acuerdo con la misma *densidad lexicométrica* de 0.5 (recordemos que 1 es ideal y 0.0 es nula cercanía con la definición a la que se asocian).

El sistema GENEX ha demostrado que es posible obtener ejemplos automáticamente con apenas unas cuantas reglas lingüísticas, sobre todo de delimitación de contextos, lo que comprueba el hecho de que un ejemplo no tiene una estructura sintáctica fija, sino que es el resultado de acotaciones que tienen que ver con criterios semánticos y de asociación de palabras. Aunque su funcionamiento aún no es del todo eficiente, los resultados indican que es un buen camino para explorar la extracción automática.

El funcionamiento básico del sistema ha tratado de imitar el proceso mental que llevamos a cabo como humanos para seleccionar de un conjunto ilimitado de posibles contextos donde aparece un término, aquel que guarde una relación semántica estrecha con la definición de ese mismo término, teniendo como criterios de selección el conjunto de palabras que conforman una definición determinada; en el humano está dada por su conocimiento previo del concepto del término que desea ejemplificar, y en el sistema lo encontramos en un repositorio organizado de conceptos que se realizan formalmente en la definición terminológica. Por tanto, un contexto ejemplificativo existe, contiene un candidato a ejemplo, y es aquel contexto que,

teniendo como núcleo un término, y a través de un conjunto de restricciones, proyecta rasgos distintos a los de una definición y la complementa. En esta investigación lo observamos directamente en los resultados arrojados por el corpus *JORNADA*. Mientras los otros corpus eran compatibles con el sistema y se pudieron generar candidatos a ejemplos, el contenido de *JORNADA* impedía que el sistema obtuviera resultados positivos ( $r > 0$ ), lo que comprueba que un corpus especializado, que a su vez tendrá sobre todo contextos especializados, pocas o nulas veces puede funcionar para diseñar un ejemplo. El ejemplo es el resultado de llevar un término a un contexto general manteniendo las características que lo hacen perteneciente a un dominio especializado. El ejemplo sirve para diseminar la opacidad comprensiva que puede crear una definición terminológica. Hace más sencillo aprehender un concepto y su generación no es una tarea imposible.

Como nota final podemos mencionar que si bien el sistema es efectivo para la extracción de candidatos a ejemplos, una buena parte de su rendimiento tiene que ver con la extensión del corpus (mientras más datos, mejor), con la naturaleza del corpus (mientras más general, mejor), con la fuente primaria de análisis (el diccionario de entrada), y con la calidad de la definición.

## 6.9. Recapitulación

En este capítulo hemos hecho un recorrido que va desde el antecedente inmediato de la herramienta que diseñamos hasta un método de evaluación para el sistema GENEX. Vimos que el GDEX, aunque funcional y preciso, consta de varias reglas que conllevan un procesamiento que compromete al mismo sistema (tarda varios segundos en procesar las peticiones) y su precisión temática está determinada por la aptitud de la gente que establece lo que es un buen ejemplo, de manera subjetiva, es decir, sin un conjunto de criterios específicos más allá de lo que comprenden a propósito de los criterios de Atkins y Rundell (2008). Además, está basado en las palabras más comunes y por tanto no se toma en cuenta la situación contextual especializada de una palabra. De este método tomamos la precisión del proceso *concordancia + colocación* y la ventaja de trabajar con corpus grandes.

Ahora bien, como mencionamos al principio de esta investigación, los corpus que debimos utilizar han sido de tamaño bastante considerable debido a que el proceso utilizado requiere mirar la lengua desde una perspectiva macro, es decir, en la mayor amplitud posible. Y esto es necesario debido a la naturaleza del análisis que se realiza. Es cierto que un corpus general más o menos grande, digamos dos o tres millones de palabras, puede considerarse como una muestra representativa de una lengua, ya que, si está bien diseñado, contendrá una buena cantidad de léxico y muy probablemente la mayoría de las palabras más comunes de esa lengua; también hay que tomar en cuenta que las combinaciones posibles entre las palabras pueden llegar a ser infinitas. El GENEX, en tanto herramienta basada en contextos, toma más en cuenta las posibles combinaciones de palabras antes que la frecuencia o aparición de éstas. Por tanto, aunque muchas piezas léxicas se repitan en los corpus que hemos utilizado, rescatamos de ellos la ventaja que nos dan en cuanto a la información sintáctica de sus constituyentes. Así, podremos decir con cierta seguridad que para hacer una extracción exitosa, bajo este panorama, siempre *más es mejor*. Vimos también que la tarea del preprocesamiento es un trabajo que conlleva bastante tiempo y que sirve para adaptar el formato de un corpus para los fines que convenga. En nuestro caso la segmentación correcta de frases fue crucial.

Por otro lado, vimos que pudimos aprovechar una parte de la metodología utilizada por Kilgarriff et al. (2008) a propósito de las concordancias complejas; y también las nociones de *legibilidad*, cuando hablamos de la longitud de los ejemplos, y *complejidad*, cuando incluimos siempre al término en la frase, propuestas hechas por Didakowski, Geyken y Lemnitzer (2012).

Con este preprocesamiento y estas variables a tomar en cuenta a propósito de trabajos anteriores, pudimos formalizar la propuesta que describimos en el capítulo anterior para poder diseñar el extractor. Así, la desarticulación del contexto definitorio tiene que ver con un alejamiento de lo especializado para hacer búsquedas donde las frases se acerquen más al lenguaje general y así reduzcan su complejidad. El contexto nominal es la recuperación de contextos donde se puede encontrar potencialmente el candidato a ejemplo pero no está formado por las palabras más comunes de una lengua de acuerdo con las frecuencias de un corpus, sino que

está formado por la totalidad de las frases que cumplen con los criterios de combinación de los elementos nominales que aparecen en una definición. El contexto verbal tiene como finalidad darle al ejemplo una estructura específica y acercar este candidato a un área especializada determinada. Este último punto, aun así, fue un obstáculo para discernir un área de especialidad de otra cuando el término era polisémico, basta mirar el mismo caso de *virus* que abordamos anteriormente pero en un corpus distinto, como *Jornada*<sup>51</sup>. Por esta razón, en las reglas para la elección de candidatos fue necesario utilizar dos medidas de distancia semántica que en conjunto con lo propuesto antes, nos permitieron lanzar la propuesta de la noción de *saturación semántica*. Con esta propuesta, vaciada en una fórmula de medición, pudimos programar GENEX para comprobar después si su funcionamiento atendía a los criterios que necesitábamos responder, esto es, que el ejemplo se refiriera al término, y que el contexto de dicho ejemplo perteneciera al campo de especialidad determinado por la definición a la que se asociaba dicho término. Con esto nos dimos cuenta que la reducción de contextos para la identificación de ejemplos es un camino fiable, pues la comprobación arrojó resultados positivos aunque no irrevocables. Esto, en una sola línea, significa que la identificación y la extracción son posibles.

---

<sup>51</sup> Ver Anexo 1



## **7. CONCLUSIONES Y APORTACIONES**

En este apartado haremos una recapitulación general del trabajo presentado y nos enfocaremos en los puntos más destacables de cada capítulo con el fin de hacer notar las aportaciones teóricas y aplicadas que hemos intentado proveer al trabajo terminográfico y terminológico actual.

### **7.1. Conclusiones por capítulo**

En el **capítulo 1** vimos que hay una diferencia entre el proceso que supone la ejemplificación y el producto de ese proceso, que es el ejemplo. A lo largo de la tesis fuimos desentrañando por diversos argumentos y técnicas que el ejemplo, en general, y el ejemplo en terminología, de manera particular, son, por tanto, producto de dos procesos distintos. Ahora bien, en concordancia con otras investigaciones que tratan sobre el ejemplo en general, que englobamos en la labor lexicográfica, vimos que dichos procesos tienen que echar mano de recursos lingüísticos muy grandes, esto es, de corpus lingüísticos con una gran cantidad de palabras. Primeramente procedimos a comprobar esto con un experimento sencillo para el área de sexualidad y los resultados casi nulos nos llevaron a la conclusión de que en un corpus pequeño, y en un corpus de especialidad, es difícil encontrar estructuras que se asemejen a un ejemplo debido a la precisión de su lenguaje y a su tamaño relativamente pequeño (1 millón de palabras). Así, concluimos con este primer acercamiento que la utilización de corpus de lengua general y que además fuera masivo (varios millones de palabras) era un asunto imprescindible.

En el **capítulo 2**, una vez establecido que el primer paso para acercarnos a la caracterización y la posible extracción de ejemplos era comprender el proceso o el mecanismo de la ejemplificación, nos dimos a la tarea de analizar las relaciones existentes entre los elementos que conforman dicho proceso.

En primer lugar analizamos la relación entre un concepto y un término. Con base en lo expuesto por varios autores, y sobre todo en Freixa (2002, p. 54-56), dijimos que el concepto está representado por un término, pero no de una forma completa. Es decir, el término sólo puede representar un conjunto de rasgos de dichos conceptos porque dicho concepto puede ser percibido de distintas maneras y, en consecuencia, conlleva distintas realizaciones (Cabré, Feliu & Tebe, 2001). Se expone que cuando hay una variación en el acuerdo sobre qué rasgos representa, y por tanto qué significa, estaremos hablando de una variación conceptual; por otro lado, cuando ese conjunto de rasgos es expresado por más de un término y sus variantes implican significados muy parecidos, estaremos hablando de variación denominativa, lo que da paso a la sinonimia (Freixa, 2002, p. 56). De esta primera relación concluimos que las variaciones del concepto o de la forma en que se aborda el concepto afecta las realizaciones de éste en el sistema lingüístico, como proponen Cabré et al. (2001, p. 6). Su contexto de activación (Kuguel, 2007), es decir sus relaciones con otros conceptos, y el discurso especializado donde se puede realizar (Ciapuscio, 2003) también juegan un papel imprescindible.

En segundo lugar analizamos la relación del término con la definición. A partir de lo propuesto por Sierra et al. (2008) a propósito de los contextos definitorios, apuntamos que las relaciones que mantiene el término con los elementos definitorios que lo enmarcan dentro de una u otra definición, hacen presentes las relaciones del concepto con otros conceptos del mismo campo de especialidad. Este tipo de relaciones (como la hiponimia, la hiperonimia, la meronimia, etcétera), de acuerdo con esta perspectiva, da paso a diferentes tipos de definiciones como las analíticas, las extensionales, las sinonímicas y las funcionales. Ahora bien, aunque dicho estudio está basado en textos especializados, lo cierto es que los contextos definitorios también se pueden referir a lexías que no sean términos, por esta razón se ha aclarado que los contextos definitorios a los que nos referimos hacen referencia exclusivamente a definiciones de tipo terminológico. Para poder determinar qué entendimos por definición terminológica, volvimos a lo propuesto por la TCT y tomamos lo dicho por Cabré, a saber, que la definición terminológica “se refiere al concepto y a las relaciones que estos

forman para referirse exclusivamente a un campo de especialidad” (Cabré, 1999). Vimos, además, que otros estudiosos se han dado en distinguir de este tipo de definición dos variantes: la *definición terminológica*, que se refiere al concepto o puede llegar a crear conceptos en la misma emisión, y la *definición terminográfica* que se refiere a un sistema conceptual ya existente (De Besse, 1997, p. 65). Esto fue importante ya que partimos de la premisa de que un ejemplo siempre está asociado a una definición y es mucho más sencillo que esa definición esté asociada a un sistema preexistente que a uno que esté en formación, es decir, a una definición terminográfica.

Finalmente, en tercer lugar, nos referimos a las limitaciones que esta definición terminográfica tiene al momento de referirse al concepto. De acuerdo con las nociones de contexto definitorio y de las características formales que tiene una definición terminográfica, dimos cuenta de que los rasgos de un concepto sólo podrán ser reflejados en parte. Es decir, esta definición no da cuenta de todos los rasgos del concepto ni de todas las relaciones que establece con otros conceptos de su misma área de especialidad. Así, nuestra conclusión ha sido que la definición terminográfica no satura el concepto al que se refiere. Para sostener esta conclusión acudimos a la noción que supone el *Principio de poliedricidad* (Cabré, 2008), que básicamente reza que el concepto de un término no se puede representar en su totalidad en un acto comunicativo, ya que supone una entidad de varias caras, de diferentes realizaciones que no siempre tienen interpretación en la realidad. El concepto es tanto conjunto de rasgos semánticos como conjunto de relaciones con otros conceptos. En palabras de Lorente:

“La poliedricidad es una noción que supera la adición de facetas conceptuales, ya que con la simple recursividad secuencial, propia del lenguaje, se perdería la estructura. Por otro lado, cada hablante desde su perspectiva sólo puede observar o tratar algunas de las caras del poliedro (aproximación parcial)” (Lorente, 2001, p. 105)

Ante este escenario, el capítulo cierra con la puesta en escena de una necesidad tanto teórica como práctica que pudiese dar fe de la posibilidad de saturar un concepto con otros elementos que no fuesen definiciones. Se propone, así, que el primer paso es encontrar el mecanismo que diese pie a esa saturación y después

encontrar un elemento que funcionase bajo ese mecanismo. La conclusión ha sido que quizá lo que se necesita es una propuesta teórica, que desarrollamos en el capítulo 3: la noción de *saturación semántica*.

En el **capítulo 3** abordamos la propuesta teórica que hemos denominado *saturación semántica*. En este capítulo mostramos que aunque hay investigaciones profundas sobre la manera en que debe funcionar un ejemplo en un diccionario, y aunque incluso hay métodos para identificarlos o extraerlos, no se detectó en la bibliografía revisada algún método para explicar la manera en la que diseñamos un ejemplo naturalmente en el lenguaje. Ante este escenario y debido a que nuestra intención fue desde un principio sentar las bases para la creación de un extractor de ejemplos, nos dimos a la tarea de modelar y explicar el mecanismo por el cual podría ser que los humanos elegimos ciertos fragmentos textuales como ejemplos de palabras. El resultado fue la *saturación semántica*, una propuesta que argumentamos como la puesta en escena de un conjunto de elecciones entre palabras y sus combinaciones que nos permiten complementar la información previa que tenemos acerca de un concepto. Esa información previa sería la definición, y específicamente la definición terminográfica. La manera de complementar esa definición, entonces, vimos que sería tratar de cubrir todas las caras que supone el poliedro del concepto de un término, esto con base en el *Principio de poliedricidad*. Ahora bien, para saber cuáles de esas caras sería más conveniente cubrir o abarcar, recurrimos a las pautas que se siguen en la TCT a propósito de la planeación, la redacción y la compilación del material necesario para diseñar un diccionario terminológico y que está desarrolladas por el *Principio de adecuación*. Este análisis nos hizo concluir dos cuestiones principales: 1) Cuando se trata de complementar una definición, se intenta decir algo acerca del concepto que la definición no logra decir. La información complementaria tiene como ideal decirlo todo de dicho concepto, de forma que cuando hablamos de “saturar” queremos decir que el complemento tendría idealmente que aportar toda la información posible para que ese concepto pudiese ser comprendido totalmente y por todos. Aunque sabemos que esto es una tarea poco menos que imposible, lo cierto es que esta primera conclusión nos permitió saber qué características semánticas debería tener ese complemento. 2) Debido a lo anterior,

sabemos entonces que hay que elegir una sola opción, la más informativa, para que sea el mejor complemento. En conclusión, hay que poder discernir la informatividad de un complemento y en consecuencia diseñar una medida que lleve a cabo esta tarea.

Ahora bien, el complemento del que hablamos lo vimos realizado más adecuadamente en el ejemplo. Por ello hay que pensar que el ejemplo como complemento a la definición debe tener información que esta no tenga, pero que esté enlazada con ella de alguna manera, y que además hay que medir qué tan informativo y adecuado puede ser.

Todo lo anterior resulta que, si la *saturación semántica* es el mecanismo con el que se intenta llegar a saber la mayor cantidad de rasgos sobre un concepto, entonces la misma *saturación semántica* es el mecanismo por el cual podría operar el ejemplo. De esto concluimos que sería necesario medir dicha saturación con el fin de aplicarla a distintos fragmentos textuales para saber si pueden ser o no buenos candidatos a ejemplos.

En el **capítulo 4** hicimos un esbozo sobre el tratamiento del ejemplo en lexicografía a través de la elección y el análisis de algunos estudios que nos aportaron los criterios para la elección y caracterización de un ejemplo en general. Mostramos que muchos de esos estudios están orientados a la traducción y enseñanza de lenguas y que otra gran parte están dedicados al estudio del ejemplo en general. Como nuestra intención es mostrar un estudio sobre el ejemplo en terminología, nos basamos en dos estudios para discernir las características de éste con respecto del ejemplo en general. En conclusión vimos que el comportamiento y las funciones que cumple dicho ejemplo en terminología varían en buena medida del ejemplo lexicográfico. Hemos cogido esas características para formar nuestra propia propuesta de caracterización del ejemplo cuando éste se utiliza para el diseño de diccionarios de especialidad.

En el **capítulo 5** presentamos nuestra propuesta de caracterización del ejemplo en terminología desde el punto de vista teórico. A partir de los preceptos del capítulo anterior y con una clara diferenciación con respecto a la terminología, demostramos a partir de un experimento hecho con *Sketch Engine* que las relaciones de un término con los elementos que lo acompañan dan

paso a la conformación de las diferentes categorías de datos de una manera particular. En este experimento se pudo ver, sobre todo, que los contextos que rodean a los términos sirven como catalizadores para la localización de información adicional y complementaria a la definición del término en cuestión. Sin embargo, no todos los contextos son productivos, sino que son el resultado de restricciones sintácticas, semánticas y combinatorias. El experimento manual pudo hacernos llegar a las conclusiones sobre las que basamos la descripción de las características del ejemplo en terminología de acuerdo con nuestra propuesta, desde su naturaleza, pasando por el establecimiento de sus características funcionales, hasta llegar a la descripción de sus características sintáctico-semánticas. Las conclusiones de este capítulo nos han permitido formular la propuesta de que el ejemplo en terminología podría operar bajo los preceptos de la noción de *saturación semántica* y que la metodología para elegirlos desde corpus textuales puede echar mano de una medida confiable como lo es la *densidad lexicométrica*. Para poder comprobar la propuesta teórica de este capítulo se diseñó una herramienta que pusiera en marcha todos sus preceptos y que formalizara, sobre todo, la heurística seguida en el experimento manual aunado a la medida de la *densidad lexicométrica*.

En el **capítulo 6** mostramos el procesamiento llevado a cabo para poder utilizar corpus de gran tamaño para una tarea terminográfica específica. Vimos que, sobre todo, en nuestro caso, mientras más datos hubiese, mejor, debido a que no buscábamos patrones sintácticos específicos, sino que seguimos una metodología que intenta imitar un proceso cognitivo. Este proceso, empero, es propenso a la sistematización, como lo perfilamos en *la formalización del ejemplo*. Vimos que a través de ciertas pautas de selección y corte de fragmento textuales a partir de criterios previos, los rasgos de la definición, es posible ir identificando un ejemplo sin necesidad de contar con patrones sintácticos o discursivos. Para poder aplicar dichos cortes, sin embargo, ya que no se cuenta en todo momento con texto etiquetado, hubo que echar mano de medidas de asociación y similitud que dieron como resultado el diseño de la ecuación para medir la *densidad lexicométrica* de un fragmento textual, que al final del capítulo es una de las aportaciones que más nos sirvió para la identificación de candidatos a ejemplos.

Una vez formalizada esta medida, se procedió a diseñar un experimento de extracción automática a través de la creación de un algoritmo que aprovechara todos los recursos anteriores. Los resultados fueron evaluados por humanos y se determinó la eficiencia del programa a través de una medida de correlación. A través de dicha evaluación se pudo concluir que GENEX trabaja de manera eficiente, aunque todavía con algunas fallas. Los resultados sugieren que la mejor manera de obtener resultados satisfactorios es utilizar corpus grandes, de gran variedad de registros, de lengua general con una considerable variedad de temáticas y, sobre todo, que el diccionario que se utilice como referencia contenga definiciones de calidad. El hecho de que el término que se desea ejemplificar no sea polisémico ayuda de manera significativa a la precisión de GENEX.

## **7.2. Aportaciones al trabajo terminográfico y terminológico**

En general podríamos decir que se ha logrado obtener una herramienta totalmente independiente que trabaja de manera satisfactoria y que ahorra una cantidad considerable de tiempo para poder identificar ejemplos asociados a términos. Esto tiene dos ventajas trascendentales y que son las principales preocupaciones de este estudio: que se pueda aportar una herramienta que dote de sistematicidad la integración de ejemplos en diccionarios de especialidad, y que a través de esta sistematicidad se ayude al lector lego o en ciernes de especialización a la mejor comprensión de la red de conceptos que sugiere un diccionario de especialidad.

### *7.2.1. La noción de saturación semántica*

La *saturación semántica* ha sido a lo largo de este estudio un pilar para el estudio de la ejemplificación en diccionarios de especialidad y para la construcción de la herramienta GENEX. Esta propuesta teórica tiene la ventaja de tomar como punto de partida nociones como la *adecuación* y la *poliedricidad* que ya han sido trabajadas desde la Teoría Comunicativa de la Terminología, pero

las actualiza y complementa. Las actualiza en el sentido de que aunque dichos principios de trabajo se toman en cuenta para el diseño y conformación de algunos diccionarios de especialidad, lo cierto es que no siempre son cumplidos. De tal manera que el aprovechamiento de la *saturación semántica* en la confección de este tipo de diccionarios implica que aquellos dos principios serán utilizados y aprovechados al mismo tiempo. Si nos damos cuenta, la *saturación semántica* no sólo intenta resolver el problema de la saturación de un concepto, sino que también intenta que en ese proceso no se pierda información o se vuelva demasiado opaca, especializada, y que al final no pueda ser decodificada por un lector cualquiera.

Por otro lado, las ventajas de la *saturación semántica* son varias. Da fe al mismo tiempo del proceso que puede llevarse a cabo para discernir las diferentes realizaciones de un concepto en la lengua, pues no sólo opera como mecanismo de conformación e identificación de ejemplos, sino que también puede ser utilizada para discernir la cantidad de información que tiene la definición a propósito de un concepto vía el término que lo representa. Asimismo, y he aquí una de sus ventajas más importantes, es que sienta las bases para poder comprender cómo funciona una categoría de datos. Insistimos a lo largo de esta investigación que el ejemplo siempre fue visto desde las perspectivas de su “buena elección” o sus funciones, pero casi nadie se ha detenido a describir cómo es que opera y cuáles son los alcances de dicho mecanismo. Por tanto, vemos en esta investigación que la *saturación semántica* hace un acercamiento a este proceso y da fe de él como un mecanismo de complemento conceptual. No sólo opera desde la definición, sino que llega hasta el concepto mismo.

### 7.2.2. La noción de densidad lexicométrica

La noción de *densidad lexicométrica* es imprescindible en este estudio debido a que representa la formalización de la *saturación semántica*. Se trata de una medida que concentra cercanía semántica, cercanía conceptual y respeta los principios de *poliedricidad* y *adecuación*. Representa una medida confiable que toma en cuenta las teorías anteriores sobre ejemplificación y pone

en juego un elemento firme, numérico, que puede ser de gran ayuda tanto a lexicógrafos como a terminólogos. Es indispensable ver que esta propuesta es un esfuerzo que ha sido diseñado dentro del campo de la terminótica y la lexicografía computacional para la correcta planeación de un diccionario cualquiera, ya que al no estar condicionada por un solo tipo de definición o discurso (corpus), puede ser aprovechada en otros campos fuera del diseño de diccionarios de especialidad.

### 7.2.3. La redefinición del término *ejemplo*

Para efectos de este trabajo, como hemos visto, fue necesario no sólo tomar en cuenta muchos de los estudios sobre ejemplificación en varias lenguas para finalmente llegar a nuestra propuesta en español, sino que también fue necesario diseñar una noción teórica que nos permitiese reformular lo dicho por otros expertos y adaptarlo a una metodología de trabajo orientada a la terminótica. El resultado atendió a las necesidades que conllevaba el diseño de la herramienta computacional, pero también nos hizo darnos cuenta, en el camino, que el ejemplo como complemento de una definición tiene un mecanismo propio de operación. Dicho mecanismo no sólo atiende a las características semánticas, formales y funcionales que las investigaciones precedentes ya han anotado, sino que, adicionalmente, muestra que el ejemplo puede tomar muchas y distintas formas. En este sentido podríamos decir que el patrón de un ejemplo, bajo nuestra perspectiva, no tiene que ver exactamente con un conjunto de estructuras sintácticas, sino que más bien se trata de una especie de patrón conceptual. Al sumergirnos en el estudio del funcionamiento de dicho patrón, hemos notado características adicionales que nos permiten aportar un par de características más a la definición del término ejemplo: que se trata de una elección razonada entre los contextos de aparición de un término, que está ligado estrechamente a la definición por medio de redes conceptuales, que complementa la definición y actúa desde el concepto, que es un elemento formalizable y extraíble, que se diseña y no sólo se elige o inventa, entre otros rasgos que hemos detallado a lo largo de esta investigación.

Quizá esta redefinición, que nos ha servido para llevar a cabo una tarea particular, pueda ser un referente optimizable para realizar otras tareas que tienen que ver, sobre todo, con el comportamiento de los términos y los conceptos dentro de un campo de especialidad. Nuestra intención fue, aquí, sobre todo, que era posible formalizar la teoría sobre ejemplificación existente para poder desarrollar una herramienta de ayuda a la terminografía.

### 7.3. Campo de aplicación

Pensado como herramienta de apoyo al quehacer terminográfico, el GENEX se concibe como una potencial aplicación de varias herramientas ya consolidadas y en desarrollo en el ámbito de la terminología y la lexicografía. Al ser un elemento independiente de lengua con una gran flexibilidad para sustituir los elementos con los que trabaja, permitirá al usuario intermedio, esto es, a quien desee integrarlo a su aplicación o diccionario, personalizarlo para mejorar los resultados de acuerdo con su propia aplicación. Así, si alguien deseara extraer ejemplos de ecología, no haría falta más que cambiar el diccionario de entrada por su propio diccionario de ecología y permitir al GENEX realizar la búsqueda en los corpus generales que se le indique. Para usuarios más expertos en cuestiones computacionales, se podría modificar también la medida de similitud (*dice* o *información mutua*) para obtener resultados basados puramente en estadística, por decir algo. Lo mismo sucederá si se quiere cambiar el corpus o los corpus de búsqueda e incluso el umbral de la *densidad lexicométrica* que permitiría encontrar ejemplos más informativos, aunque más extensos en cuanto a cantidad de tokens, si se quisiera integrar, digamos, a sistemas de resumen automático. Veamos algunas de las aplicaciones en las que se ha pensado.

#### 7.3.1. Integración a una herramienta de extracción conceptual (Describe<sup>®</sup>)

*Describe* es un buscador de definiciones en internet desarrollado por el Grupo de Ingeniería Lingüística de la Universidad Nacional

Autónoma de México (Sierra, Alarcón, Molina, & Aldana, 2009). Su característica principal es que a partir de texto libre en la web, y con el aprovechamiento de los patrones descritos a propósito de la investigación en contextos definitorios que ya hemos abordado, puede extraer candidatos a definiciones en tiempo real y sin consultar diccionarios. Las ventajas de contar con un sistema así es que se pueden diseñar y crear, por ejemplo, glosarios sencillos de una materia en un tiempo mínimo o puede ser utilizado como auxiliar en el diseño de diccionarios generales y de especialidad. Asimismo, su aportación para la búsqueda y el hallazgo de nuevos sentidos en las palabras ya conocidas es una ventaja destacable, pues busca definiciones antes de que los sentidos que estas indican sean fijadas en un diccionario.

GENEX funcionaría en este caso como un complemento que permitiría, a partir de los candidatos a definición extraídos por Describe, confirmar el sentido y el uso de un vocablo con la acepción correspondiente. Dentro de este tenor, GENEX estaría trabajando con internet como corpus de análisis y Describe como diccionario de entrada. Se podría pensar, entonces, que una búsqueda con ambas herramientas integradas, ayudaría en gran medida a la conformación confiable y actualizada de entradas de diccionario.

### 7.3.2. Integración a diccionarios

Uno de los problemas principales que observamos en esta investigación es que los ejemplos han sido relegados a elementos satelitales en la planeación de diccionarios, sin embargo, como hemos mostrado, estos elementos son imprescindibles para la transmisión del conocimiento especializado. Por esta razón, GENEX fue programado como un módulo independiente y como una herramienta que puede ser adaptada a cualquier diccionario y a cualquier corpus. Así, tanto diccionarios ya terminados o en proceso, como aquellos de nueva planta, pueden hacer uso de esta herramienta para extraer los ejemplos de sus propias definiciones sin comprometer demasiado el tiempo de desarrollo y con la ventaja de ofrecer mayor información y de calidad en las entradas que conforman o conformarán sus trabajos.

Se ha pensado particularmente en la integración de la herramienta GENEX al *Diccionario de las Sexualidades en México* (DSMX) (Sierra, Medina & Lázaro, 2009), con el fin de ayudar a escolares a la mejor comprensión de los términos en el área de sexualidad y reproducción humana. Los resultados preliminares de esta tesis han demostrado que es posible extraer algunos buenos ejemplos a partir de las definiciones en el DEM. El siguiente paso será utilizar las definiciones del propio DSMX para complementar todas sus entradas con ejemplos.

### 7.3.3. Integración a herramientas de organización de términos (Terminus)

Otro caso donde son necesarios los ejemplos es en herramientas para la gestión de terminologías. Existen bases de datos terminológicas como IATE, CRITER, FMI, etcétera, que reservan un espacio en sus entradas para ejemplos de uso de los términos que describen. Esto es importante porque, en algunos casos son pruebas que permiten la mejor traducción de los términos ya que sirven como desambiguadores contextuales, y en otros casos como reafirmaciones del uso con un cierto sentido.

Nosotros hemos pensado específicamente en una herramienta que sirve para la gestión de terminología: *Terminus*, que en palabras de sus autores es “una estación de trabajo para la terminología. Integra la gestión de corpus y de terminología. Permite crear y gestionar grupos de trabajo y modelar las categorías de datos.” (Cabré, Montané & Nazar, 2011). Justo en este último punto es donde nos detenemos, ya que la modelación y la gestión de las categorías de datos es una práctica manual. Es decir, aun cuando *Terminus* concede un espacio para la inclusión de ejemplos, éstos tienen que ser ingresados a mano.

En este caso la integración de GENEX a una herramienta de este tipo generaría un valor agregado al agilizar el diseño y la gestión de los glosarios o diccionarios que se diseñen con este recurso. Su inclusión permitiría al usuario reducir tiempo de trabajo y serviría, al mismo tiempo, para sistematizar el uso de esta

categoría de datos en los productos terminográficos que se deriven de su investigación.

#### 7.3.4. Aplicación comercial

Como nos podemos dar cuenta en estos últimos apartados, GENEX tiene una potencial aplicación académica, ya como complemento para agregar valor a aplicaciones existentes, ya como herramienta independiente para agilizar trabajos lexicográficos y terminológicos en curso o en fase de planeación. Pero lo cierto es que siempre será mejor poder diseñar un diccionario más completo sin comprometer horas de trabajo y sin aumentar demasiado los costos de producción, lo que significa que GENEX también tiene una aplicación comercial.

Dentro de esta línea, hay que mencionar que GENEX no sólo representa una aplicación dirigida a diccionarios. Si pensamos en que muchas veces mientras navegamos en internet encontramos palabras que no entendemos, y sobre todo cuando estamos haciendo investigación esas palabras se refieren a términos, no es raro inferir que la utilización de GENEX sería altamente provechosa. Una de las ideas en las que estamos trabajando es poder crear un módulo flotante o aplicación que se instale en la computadora, ya con un corpus precargado, y poder activar GENEX desde el menú de un click secundario sobre la palabra que no entendemos. Entonces GENEX traería ejemplos de uso de esa palabra en registros más sencillos y ayudaría al usuario a seguir su lectura con una mejor comprensión del contenido que está consultando. Una variante de esta posible aplicación se puede integrar en textos didácticos o enciclopedias, e incluso en traductores automáticos en línea, tema del que hablaremos más adelante como trabajo futuro.

Como vemos, la potencial aplicación comercial de esta herramienta tiene un camino atractivo y es sólo cuestión de que haya una inversión que permita continuar con la investigación encaminada a mejorarla.

## 7.4. Trabajo futuro

Hasta la versión final de este trabajo han quedado claras las ventajas y las aportaciones que se han querido ofrecer al estudio de la ejemplificación y el ejemplo en terminología. Aun así, el trabajo ha dejado algunas líneas que si bien no representaron un obstáculo para el desarrollo de este trabajo, sí son asuntos importantes que hay que tomar en cuenta tanto para aplicación y mejoramiento de la propia herramienta, como para la distinción con otros sistemas incipientes o futuros.

### 7.4.1. Afinación de la fórmula para medir la *densidad lexicométrica*

Actualmente se está desarrollando un experimento para mejorar los resultados generales del GENEX y tratar de reducir su margen de error, sobre todo en términos polisémicos. El experimento consta básicamente de la aplicación recursiva de la medidas de *información mutua* en fragmentos que son candidatos a ejemplos de acuerdo con GENEX. Tomemos por caso el mismo término *virus*. De acuerdo con los resultados obtenidos, el campo semántico {V}, verbal, estaría conformado por los elementos *alterar* y *dañar*. Por otra parte, el campo semántico {N}, nominal, estaría compuesto por los elementos *programa*, *funcionamiento*, *normal*, *computadora*, *virus* y *cibernético*. Hemos visto que GENEX calcula la *información mutua* de los verbos con respecto al término para determinar cuál de ellos es el más cercano semánticamente y después hacer una selección que incluya *término + verbo elegido + \*uno o más elementos de {N}\**. También hemos visto que los candidatos que obtienen una mejor calificación son aquellos que respetan esta condición y que tienen más elementos de {N}. Pues bien, dichos elementos no están ordenados y no se toma en cuenta su posición dentro del candidato a ejemplo.

La afinación de la fórmula consiste en agregar un nuevo cálculo a partir de tomar la lista {N}, calcular la información mutua de cada uno de sus elementos con respecto al término y asignarles un peso. De esta manera, si el fragmento a considerarse ejemplo tiene elementos de {N} con mayor peso, entonces será bonificado.

Un proceso similar se piensa hacer con los verbos de la lista {V}. Nuestra hipótesis sugiere que con este procedimiento adicional GENEX podría distinguir mejor entre los contextos pertenecientes al sentido del término en un área de especialidad, aun cuando se utilice en otra área de especialidad.

#### 7.4.2. GENEX+ (*plus*)

Otra investigación en curso es la aplicación de GENEX a investigaciones de diccionarios bilingües, multilingües o herramientas de traducción automática. En esta segunda etapa implementaremos los cambios necesarios para que la entrada sea bilingüe. Esto es, a partir de dos palabras o términos en lenguas distintas, con sus respectivas definiciones, obtendríamos ejemplos en ambas lenguas. La hipótesis sugiere que los mejores ejemplos en un diccionario bilingüe estarían asociados a la definición de la lengua meta y no a los de la lengua materna (Laufer & Levitzky-Aviar, 1992), pues el sistema lingüístico es distinto. Sin embargo, los ejemplos en ambas lenguas se complementarían para dar fe de la forma en que la realidad es aprehendida por la lengua que se quiere conocer, como ya lo ha mostrado Humble (1998). Esto ayudaría al aprendizaje de una segunda lengua a poder discernir los contextos donde un equivalente puede ser utilizado y cuándo esto no es posible, ya que las restricciones contextuales de uso habrían sido aplicadas previamente de forma automática por el propio GENEX.

#### 7.4.3. ¿Es posible crear automáticamente un ejemplo?

Toda nuestra investigación ha girado en torno a la extracción automática de ejemplo. Hemos visto cómo la metodología utilizada analiza el texto de un corpus y, a través de un conjunto de reglas, va eligiendo los mejores candidatos. También hemos insistido en que estos candidatos no tienen una estructura prototípica o que, al menos, no la encontramos a lo largo de la investigación. Por tanto, hay que hacer notar que, en sentido estricto, en ningún momento se generó un ejemplo. Lo anterior contravendría a nuestra noción de ejemplo como “muestra real de uso documentada extraída de un corpus”, que fue básicamente la premisa que hemos seguido hasta

ahora. Sin embargo, como vimos, hay buenos argumentos para pensar que un ejemplo diseñado, generado o creado puede ser un buen recurso dentro de ciertos diccionarios (sobre todo los de aprendizaje de segundas lenguas). Por esto mismo es que, también como trabajo futuro, pensamos ponderar la posibilidad de combinar el método de extracción que lleva a cabo el GENEX con los incipientes métodos de generación automática de frases que se utilizan, por ejemplo, en sistemas inteligentes de pregunta-respuesta (QA, por sus siglas en inglés).

Una idea inicial es que, si ya podemos obtener un conjunto bastante grande de candidatos a ejemplos con la ayuda de GENEX, podríamos aplicar un análisis de creación automática de patrones léxicos. Esta idea está sostenida sobre todo por los trabajos de Hanks (2004) y Renau & Nazar (2011). En dichos estudios se propone que a partir de la información que aportan los elementos léxicos que rodean a un vocablo, es posible inferir su significado y que el conjunto de contextos que le otorgan ese significado o sentido tienen un patrón que es posible determinar. Si esto es comprobable con los candidatos a ejemplo que arroja GENEX, entonces podríamos generar una lista de patrones que, en adelante, nos permitirían colocar un término a buscar, extraer sus ejemplos, después dichos patrones y generar un ejemplo prototípico. ¿Entonces, es posible la generación? Nosotros creemos que sí, pero esa sería una hipótesis que habría que comprobar y que correría a cargo de otra investigación sobre el proceso de ejemplificación.

## Bibliografía

Academia Mexicana de la Lengua. (2000). *Índice de mexicanismos*. México D.F.: Fondo de Cultura Económica/Consejo Nacional para la Cultura y las Artes/Academia Mexicana de la Lengua.

Acosta, O. (2009). Automatic Extraction of Lexical Relations from Analytical Definitions Using a Constraint Grammar. *Canadian AI '09 Proceedings of the 22nd Canadian Conference on Artificial Intelligence* (pp. 262-265). Berlin: Springer-Verlag.

Adelstein, A. (2004). *Unidad léxica y valor especializado: estado de la cuestión y observaciones sobre su representación*. Barcelona: Institut Universitari de Lingüística Aplicada. Serie Tesis.

Adelstein, A. (2005). Los modelos de texto en la teoría terminológica. *VI Jornadas de Traducción y Terminología del CETRATER*. Tucumán: Departamento de Idiomas Modernos, Facultad de Filosofía y Letras, Universidad Nacional de Tucumán. 1-11.

Aguilar, C. A. (2009). *Análisis lingüístico de definiciones en contextos definatorios* (Tesis doctoral). México: UNAM.

Ahmad, K., Fulford, H., y Rogers, M. (1992). The elaboration of special language terms: the role of contextual examples, representative samples and normative requirements. En H. Tommola, K. Varantola, T. Salmi-Tolonen, J. Schopp (eds.), *5th EU- RALEX International Congress*, University of Tampere. 139-149.

Ahumada, I. (1989). *Aspectos de lexicografía teórica: aplicaciones al diccionario de la Real Academia Española*. Granada: Universidad de Granada.

Alarcón, R. (2003). *Análisis lingüístico de contextos definatorios en textos de especialidad*. México: Universidad Nacional Autónoma de Letras, Facultad de Filosofía y Letras.

Alarcón, R. (2009). *Descripción y evaluación de un sistema basado en reglas para la extracción automática de contextos definatorios* (Tesis doctoral). Barcelona: Universitat Pompeu Fabra.

- Alatorre, A. (2001). Sobre americanismos en general y mexicanismos en especial. *Nueva Revista de Filología Hispánica*, XLIX (1), 1-51.
- Alejo, J. (21 noviembre de 2010). ¡Chida noticia: salió el Diccionario de Mexicanismos! *Milenio*.
- Alvar Ezquerro, M. (1982). Diccionario y gramática. *Lingüística española actual*, 4 (2), 151–212.
- Álvarez, A. (2006). *La variación lingüística y el léxico: conceptos fundamentales y problemas metodológicos*. Hermosillo: Universidad de Sonora.
- Amster, H. (1964). Semantic satiation and generation: Learning? Adaptation?. *Psychological Bulletin*, 62 (4). 273-286.
- Atkins, B. T. y Rundell, M. (2008). *The Oxford guide to Practical Lexicography*. Oxford: Oxford University Press.
- Atkins, B. T., y Rundell, M. (2008). Exemplification in learner writing: a cross-linguistic perspective. En F. Meunier, S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins Publishing. 101–119.
- Ausubel, D. (1983). *Psicología Educativa. Un punto de vista cognoscitivo*. México: Trillas.
- Bauer, L. (1992). *Introducing Linguistic Morphology*. Edinburgh University Press.
- Bach, C. (2001). *Els connectors reformulatius catalans: Anàlisi i proposta d'aplicació lexicogràfica* (Tesis doctoral). Barcelona: IULA, UPF.
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York: ACM Press Addison-Wesley.
- Basset, M. F. & Warne, C. J. (1919). On the lapse of verbal meaning with repetition. *The American Journal of Psychology*, 30 (4). 415-418.

Bergenholtz, H., y Tarp, S. (1995). *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins Publishing.

Beristáin, H. (1995). *Diccionario de Retórica y Poética*. México: Editorial Porrúa.

Black, S. R. (2001). Semantic satiation and lexical ambiguity resolution. *The American Journal of Psychology*, 114 (4). 493-510.

Bosque, I. (2011). Deducing collocations. En *Fifth International Conference on Meaning Text Theory*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. 6–23

Bouma, G. (2009). Normalized (Pointwise) Mutual Information in Collocation Extraction. En Biennial GSCL Conference. Stuttgart: Gunter Narr Verlag, 31-40.

Cabré, M. T. (1992). *La terminologia. La teoria, els mètodes, les aplicacions*. Barcelona: Les Naus d'Empúries.

Cabré, M. T. (1998). El discurs especialitzat o la variació funcional determinada per la temàtica: noves perspectives. *Caplletra: revista internacional de filologia*, 25, 173-194.

Cabré, M.T. (1999). *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Cabré, M. T. (2002). Textos especializados y unidades de conocimiento: metodología y tipologización. En J. García Palacios, y M.T. Fuentes (eds.), *Texto, terminología y traducción*. Salamanca: Ediciones Almar. 15-36.

Cabré, M. T. (2003). Theories of terminology: their description, prescription and explanation. *Terminology*, 9 (2), 163-199.

Cabré, M. T. (2004). La terminología en la traducción especializada. En C. Gonzalo y V. García, *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco libros. 89-122.

Cabré, M. T. (2008). El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en terminología (I). *Ibérica*, 9-36.

Cabré, M. T., & Feliu, J. (. (2001). *La terminología científico-técnica*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.

Cabré, T., & Estopá, R. (2002). El conocimiento especializado y sus unidades de representación: diversidad cognitiva. *Sendebarr*, 141-153.

Cabré, M. T., & Bach, C. (2004). El corpus tècnic del IULA: corpus textual especializado plurilingüe. *Panace@*, V (16), 173-176.

Cabré, M. T., Montané, A., Nazar, R., & Reus, R. (2010). Estació Terminus: a Web Application for Terminology and Corpus Management. *Proceedings of TKE 2010 Conference (Terminology and Knowledge Engineering)*. Dublin: TKE.

Calderón Campos, M. (1993). Sobre la elaboración de diccionarios monolingües de producción: las definiciones, los ejemplos y las colocaciones léxicas. *Foro Hispánico. Revista hispánica de los Países Bajos* (6), 105-120.

Caravedo, R. (1999). *Lingüística del corpus. Cuestiones teórico-metodológicas aplicadas al español*. Salamanca: Ediciones Universidad de Salamanca.

Church, K. W., y Hanks, P. (1990). Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, 16 (1), 22–29.

Ciapuscio, G. (2003). *Textos especializados y terminología*. Barcelona: Institut Universitari de Lingüística aplicada.

Cohene, L. S.; Smith, M. C. & Klein, D. (1978). Semantic satiation revisited with a lexical decision task. *Memory & Cognition*, 6 (2). 131-140.

Collins, A. M. & Lofthus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82 (6). 407-428.

Company, C. (coord.) (2010). *Diccionario de Mexicanismos*. México D.F.: Siglo XXI editores/Academia Mexicana de la Lengua.

Company, C., y Cuétara, J. (2007). *Manual de gramática histórica*. México: Universidad Nacional Autónoma de México. Facultad de Filosofía y Letras.

Comisión Nacional para la Defensa de los Usuarios de las Instituciones Financieras (CONDUSEF) (2009). *ABC de Educación Financiera*. México: Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros.

Cowie, A. (1989). The Language of Examples in English Learners' Dictionaries. En G. James (ed.), *Lexicographers and their Work*, Exeter Linguistic Studies, 14. Exeter: University of Exeter. 55–65.

Egghe, L., y Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient  $r$  and Salton's cosine measure. *Journal of the American Society for Information Science and Technology*, 60 (5), 1027-1036.

Eres Fernández, I., y Almeida Seemann, P. (2009). Un estudio sobre los cambios lingüísticos del español escrito en las charlas informales por internet. *Trabalhos em Linguística Aplicada*, 48 (1).

Espasa. (1999). *Diccionario Espasa de Medicina*. Espasa.

Estopà, R. (2001). Elementos lingüísticos de las unidades terminológicas para su extracción automática. En M. T. Cabré, y J. Feliu (eds.), *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut Universitari de Lingüística Aplicada. 67-80.

Estopà, R. (dir.) (2013). *Mi primer diccionario de ciencia*. Barcelona: Publicacions de l'Abadia de Montserrat.

da Cunha, I. (2008). *Hacia un modelo lingüístico de resumen automático de artículos médicos en español* (Tesis doctoral). Barcelona: Institut Universitari de Lingüística Aplicada.

da Cunha, I., y Sierra, G. (2010). Hacia la búsqueda de rasgos discursivos específicos de dominios especializados. En M. T. Cabré

y B. Rodríguez (eds.), *Actas del XII Simposio de RITerm 2010: La terminología: puente ineludible de una sólida mediación cultural (RITERM 2010)*. Red Iberoamericana de Terminología/ Colegio de Traductores Públicos de la Universidad de Buenos Aires.

da Cunha, I., San Juan, E., Torres-Moreno, J., Cabré, M. T., y Sierra, G. (2012). A Symbolic Approach for Automatic Detection of Nuclearity and Rhetorical Relations among Intra-sentence Discourse Segments in Spanish. *Lecture Notes in Computer Science (LNCS)*, 7181, 462-474.

De Bessé, B. (1991). Le contexte terminographique. *Meta*, 36 (1), 111-120.

Díaz, H. M. (1996). Las estructuras paradigmáticas secundarias por desarrollo predicativo y la teoría de los esquemas sintáctico semánticos. *Contextos*, 14 (27-28), 65-105.

Drysdale, P. (1987). The role of examples in a learner's dictionary. En A. Cowie (ed.), *The Dictionary and the Language Learner*. Papers from the EURALEX Seminar at the University of Leeds, 1985 (Lexicographica. Series Maior, 17. Tübingen: Max Niemeyer. 213-223

Esposito, N. J. & Pelton, L. H. (1971). Review of the measurement of semantic satiation. *Psychological Bulletin*, 75 (5). 330-346.

Fuentes Morán, M. T., y García Palacios, J. (2002). Los ejemplos en el diccionario de especialidad. En J. García Palacios, y M. T. Fuentes Morán. *Texto, terminología y traducción*. Salamanca: Almar. 75-98.

Fuertes-Olivera, P., Bergenholtz, H., Nielsen, S., y Amo, M. N. (2012). Classification in Lexicography: The Concept of Collocation in the Accounting Dictionaries. *Lexicographica*, 28 (1), 293-308.

Faber, E.; Montero Martínez, M.; Castro Prieto, M.R.; Senso Ruiz, J.; Prieto Velasco, J.A.; León Araúz, P.; Márquez Linares, C.; Vega Expósito, M. (2006). Process-oriented terminology management in the domain of Coastal Engineering. *Terminology*, 12 (2), 189-213.

- Fathi, B. (2014). Experts and Specialised Lexicography: perspectives and needs. *Terminàlia*, 9, 12-21.
- Fedor de Diego, A. (1995). *Terminología. Teoría y práctica*. Maracay, Venezuela: Unión Latina/ Ediciones de la Universidad Simón Bolívar.
- Fellbaum, C. (ed.) (1998). *WORDNET. An electronic lexical database*. Massachusetts: The MIT press.
- Fernández Bernárdez, C. (1994-1995). Marcadores textuales de <<ejemplificación>> textual. *Estudios de Lingüística de la Universidad de Alicante*, 10, 103-144.
- Fernández Silva, S. (2011). Variación terminológica y cognición: factores cognitivos en la denominación del concepto especializado (Tesis doctoral). Barcelona: Institut de Lingüística Aplicada, Universitat Pompeu Fabra.
- Fitch, R. (2013). *Un análisis de los ejemplos de uso en la lexicografía mexicana contemporánea*. (Tesis de Maestría). Querétaro: Universidad Autónoma de Querétaro.
- Fox, G. (1987). The case for examples. En J. Sinclair (ed.), *Looking up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins ELT. 137-149.
- Freixa, J. (2002). *La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient* (Tesis doctoral). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Galinski, C., y Picht, H. (1997). Graphic and Other Semiotic Forms of Knowledge Representation in Terminology Management. En S. E. Wright, *Handbook of Terminology Management*. Amsterdam: John Benjamins. 42-62.
- García, J. (1905). *Vocabulario de mexicanismos*. (Pimentel, L.G. ed.) México: La Europea.

García, J. (2002). El artículo lexicográfico en el diccionario de especialidad. *Diccionarios y lenguas de especialidad V Seminario de Lexicografía Hispánica*. Jaén: Universidad de Jaén. 21-47

García de Quesada, M. (2001). Estructura definicional terminográfica en el subdominio de la oncología clínica. *Estudios de Lingüística del Español*, 14.

Garduño, G., Sierra, G., y Medina, A. (2004). Herramientas de análisis para el Corpus Lingüístico en Ingeniería. *Encuentro Internacional de Ciencias de la Computación*, México: Universidad de Colima.

Garrido Almiñana, J. M. (1996). *Modelling spanish intonation or text-to-speech applications*. (Tesis doctoral). Barcelona: Universitat Autònoma de Barcelona, Departament de Filologia Espanyola.

Gaudin, F. (1993). Socioterminologie: propos et propositions épistémologiques. *Intercommunications. Le langage et l'homme*, 28 (4). France: De Boeck.

Gómez, G. (2001). *Diccionario breve de mexicanismos*. México D.F.: Fondo de Cultura Económica/Academia Mexicana de la Lengua.

Gutiérrez Cuadrado, J. (1999). Notas a propósito de la ejemplificación y la sinonimia en los diccionarios para extranjeros. En J. L. Orduña López (coord.), *Así son los diccionarios*. Lleida: Universitat de Lleida. 77-98.

Hernández, A. (2009). *Análisis lingüístico de definiciones analíticas para la búsqueda de reglas que permitan su delimitación automática*. (Tesis de licenciatura). México D.F.: UNAM.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. New York: Routledge.

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century.

Humble, Ph. (1998). The use of Authentic, Made-up, and Controlled Examples in Foreign Language Dictionaries. En T.

Fontenelle et al. (eds.), EURALEX '98 Proceedings. Liège: University of Liège. 593-599.

International Standard Organization (1999). *Computer applications in terminology. Data categories: ISO 12620*. Ginebra: ISO.

Jacobovits, L. A. (1962). *Effects of Repeated Presentation on cognitive Aspects of Behavior: Some Experiments on the Phenomenon of Semantic Satiation*. Doctoral Dissertation, Department of Psychology, McGill University.

Jacobovits, L. A. & Lambert, W. E. (1961). Semantic satiation among bilinguals. *Journal of Experimental Psychology*, 62 (6). 576-582.

Jang, M. G., Myung, S., Park, S. (1999). Using Mutual Information to Resolve Query Translation Ambiguities and Query Term Weighting. En Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Maryland, USA: ACL. 223-229.

Jacinto García, E. J. (2013). *El principio de autoridad en los diccionarios generales del español (siglos XVIII-XX)* (Tesis doctoral). Madrid: Universidad Complutense de Madrid.

Kilgarriff, A., y Renau, I. (2013). esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia, a Journal of the Social and Behavioral Sciences*, 95. 12-19.

Kilgarriff, A., Husak, M., McAdam, K., Rundell, M., Rychly, P. (2008). GDEX: Automatically finding good dictionary examples in a corpus. En E. Bernal, J. DeCesaris (eds.), 13th EURALEX International Congress, Euralex. Barcelona: Universitat Pompeu Fabra. 425-433.

Kim, J., Lee, S., y Hwang, S. W. (2013). Enriching Documents with Examples: A Corpus Mining Approach. *ACM Transactions on Information Systems*, 31 (1), 1-26.

Kounios, J.; Kotz, S. A. & Holcomb, P. J. (2000). On the locus of the semantic satiation effect: evidence from event-related brain potentials. *Memory & Cognition*, 28 (8). 1366-1377.

Kuguel, I. (2007). La activación del significado especializado. En M. Lorente, R. Estopà, J. Freixa, J. Martí, y C. Tebé (eds.), *Estudis de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Barcelona: Institut Universitari de Lingüística Aplicada. 337-354.

Kuguel, I. (2010). *La semántica del léxico especializado: los términos en textos de ecología*. (Tesis doctoral). Barcelona: Institut Universitari de Lingüística Aplicada.

Lakoff, G. (1987). *Women, fire and dangerous things. What categories reveal about the mind*. Chicago: The University of Chicago Press.

Lara Ramos, L. F. (2001). *Ensayos de teoría semántica. Lengua natural y lenguajes científicos*. México: El Colegio de México.

Lara Ramos, L. F. (1979). *Investigaciones lingüísticas en lexicografía*. D.F., México: El Colegio de México.

Lara Ramos, L. F. (1997). *Teoría del diccionario monolingüe*. México: El Colegio de México.

Lara Ramos, L. F. (2008). *Diccionario del español usual en México* (2ª ed.). México: El Colegio de México.

Lara Ramos, L. F. (2010). *Diccionario del Español de México (DEM)*. México: El Colegio de México.

Lara Ramos, L. F. (2011). Diccionario de mexicanismos, *Letras Libres*, 146 (ed. México), 68-72.

Lara Ramos, L. F. (2006). El ejemplo en el artículo lexicográfico. En E. Bernal, y J. DeCesaris (eds.), *Palabra por palabra. Estudios ofrecidos a Paz Battaner*. Barcelona: Institut Universitari de Lingüística Aplicada. 139-146.

Lara Ramos, L. F., y García Hidalgo, M. I. (1974). El uso de la computadora electrónica en la elaboración del Diccionario del español de México. En Instituto Nacional de Antropología e Historia/Centro de Investigación en Matemáticas Aplicadas y Sistemas (ed.), *Reunión de trabajo sobre la aplicación de las computadoras en el área de las ciencias sociales*. México D.F. : Cuadernos de trabajo del INAH. 1-12.

Laufer, B. (1992). Corpus-Based versus lexicographer Examples in Comprehension and Production of New Words. En H. Tommola et al (eds.), *EURALEX '92 Proceedings*. Tampere: University of Tampere. 71-76.

Lázaro, J. (2011). *Análisis de relevancia cuantitativa y cualitativa de ejemplos y contextos de uso en definiciones de términos referidos a sexualidad*. (Trabajo de fin de máster). Barcelona: Institut Universitari de Lingüística Aplicada.

Lázaro, J. (2010). *Extracción de la terminología básica de las sexualidades en México a partir de un corpus lingüístico*. (Tesis de licenciatura). México: Universidad Autónoma de México.

Lázaro, J. (2012). *Mecanismos de apoyo conceptual en la definición terminográfica: el caso de los ejemplos y contextos de uso*. (Proyecto de Tesis doctoral). Barcelona: Universitat Pompeu Fabra.

León, J. (2007). Meaning by collocation: The Firthian filiation of corpus linguistics. En D. A. Kibbee (ed.), *History of Linguistics 2005: Selected papers from the Tenth International Conference on the History of the Language Sciences (ICHOLS X)*. Illinois: John Benjamins Publishing. 404-415.

López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco libros.

Lorente, M. (2001). Teoría e innovación en terminografía: la definición terminográfica. En M. Cabré, y J. Feliu, *La terminología científico-técnica*. Barcelona: Institut Universitari de Lingüística Aplicada. 81-112.

Lorente, M. (2013). Terminología in Vivo y variación funcional. *Ugarteburu Terminologia Jardunaldiak*, V. 2-18.

Magerman, D., y Marcus, M. (1990). Parsing a Natural Language Using Mutual Information Statistics. *Association for the Advancement of Artificial Intelligence, AAAI-90*. 984–989.

Maldonado, C. (1997). La descripción del contexto de uso en la definición lexicográfica. En M. Martínez H., D. García P., D. Corbella D., C. Corrales Z., y C. R. Francisco, *Cien años de*

*investigación semántica: de Michel Bréal a la actualidad*. Madrid: Ediciones Clásicas. 647-658.

Manning, C., y Schutze, H. (2000). *Foundations of Statistical Natural Language Processing*. Massachusetts: MIT Press.

Marcos, M. (2006). *Evaluación de la usabilidad en sistemas de información terminológicos online*. Consulta 2 de junio de 2015: <http://www.hipertext.net/web/pag268.htm>

Marsá, F. (1982). *Diccionario Planeta de la lengua española usual*. Indiana: Planeta Publishing Corporation.

Martín, M. (1999). Los marcadores del discurso. En I. Bosque, y V. Demonte, *Gramática descriptiva de la lengua española*. Madrid: Espasa. 4051-4214.

Mateos-Vega, M. (14 de enero de 2011). No estamos obligados a utilizar el léxico de otra sociedad: lingüista. *La Jornada*.

Mayer-Schönberger, V., y Cukier, K. (2013). *Big data: la revolución de los datos masivos*. Madrid: Turner.

McEnery, T., y Wilson, A. (1996). *Corpus Linguistics*. Edinburgo: Edinburgh University Press.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. En D. Bourigault, C. Jacquemin, y M.-C. L'Homme, *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins Publishing Company. 279-302.

Medina, A., y Méndez, C. (2006). Arquitectura del Corpus Histórico del Español de México. En A. Hernández y J. Zechinelli (Eds.), *Avances en la Ciencia de la Computación*. México: Sociedad Mexicana de Ciencia de la Computación. 248-253.

Medina, A., y Sierra, G. (2004). Criteria for the Construction of a Corpus for a Mexican Spanish Dictionary of Sexuality. *Proceedings 11th Euralex International Congress*. 2. Lorent: Université de Bretagne-Sud.

Mejía P., J. (1984). *Así habla el mexicano*. México D.F.: Panorama.

- Minaeva, L. (1992). Dictionary examples: friends or foes? En H. Tommola et al (eds.), 5th EURALEX '92 Proceedings. Tampere: University of Tampere. 77-80.
- Molina, A. (2008). Deja de buscar, mejor DESCRIBE. *Actas del 4to Seminario de Ingeniería Lingüística*. México D.F.: Instituto de Ingeniería, UNAM.
- Molina, A. (2009). *Agrupamiento semántico de contextos definitorios*. (Tesis de Maestría). México D.F.: UNAM.
- Montané March, A. (2009). TERMINUS, gestión de corpus y terminología en línea. (R. Red Iberoamericana de Terminología, ed.). *Debate Terminológico*, 6. 113-115.
- Mosby. (2003). *Diccionario Mosby. Medicina, enfermería y ciencias de la salud*. España: Elsevier/Mosby.
- Mott, B.; Mateo, M. (2009). *Diccionario- guía de traducción español-inglés, inglés-español* (Diccionarios UB). Barcelona: Ediciones de la Universitat de Barcelona.
- Neely, J. H. (1977). The effects of visual and verbal satiation on a lexical decision task. *The American Journal of Psychology*, 90 (3). 447-459.
- Nesi, H. (1996). The Role of Illustrative Examples in Productive Dictionary Use, Dictionaries: *Journal of the Dictionary Society of North America*, 17 (1), 198–206.
- Osgood, C. E. (1953). The nature of measurement of meaning. *Psychology Bulletin*, 49. 197-237.
- Palmer, F. (1968). *Selected papers of J.R. Firth (1952-59)*. London: Longmans.
- Paquot, M. (2008). “Exemplification in learner writing: a cross-linguistic perspective”. En F. Meunier y S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins Publishing. 101-119.

Pecina, P. (2009). *Lexical Association Measures. Collocation Extraction*. Prague: Studies in Computational and Theoretical Linguistics, Institute of Formal and Applied Linguistics.

Pérez, C. (2002). Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del Español (ELiEs)*, 18.

Pietersma, A. (2007). Context is King in Septuagint Lexicography, Or is It? En *Biblical Lexicography of the Society of Biblical Literature, Annual Meeting*. San Diego, USA. 1–10.

Pineda, L.A. (2009). The Corpus DIMEx100: Transcription and evaluation. En *Language Resources and evaluation*, 44. Netherlands: Springer.

Pozzi, M. (2004). Estudio de los diferentes tipos de descripción de conceptos especializados de acuerdo con las diferentes necesidades comunicativas del destinatario final: un estudio de caso en el campo de la cardiología. En T. Cabré, R. Estopà, y C. Tebé, *La terminología en el siglo XXI: contribución a la cultura de la paz, la diversidad y la sostenibilidad*. Barcelona: Institut Universitari de Lingüística Aplicada. 529-642.

Procházková, P. (2006). *Fundamentos de Lingüística de corpus*. Consulta 2 de junio del 2015: [http://www.prochazkova.de/fundamentos\\_de\\_la\\_linguistica\\_de\\_corpus.pdf](http://www.prochazkova.de/fundamentos_de_la_linguistica_de_corpus.pdf)

Pustejovski, J. (1995). *The generative lexicon*. Cambridge: The MIT Press.

Ramos, M. (2009). Hacia un nuevo recurso léxico: ¿Fusión entre corpus y diccionario? *Panorama de investigaciones basadas en corpus*. Murcia: AELINCO. 1191-1207.

Real Academia Española. *Diccionario de la Real Academia Española de la lengua* (22a Ed.).

Rey, A. (1995). Du discours au discours par l'usage: pour une problématique de l'exemple. *Langue française*, 106. 95-120.

Reyes, T., Medina, A., y Sierra, G. (2011). Un corpus para la investigación en la extracción de términos y contextos definitorios: hacia un diccionario de las sexualidades desde México. *Debate Terminológico* (7). 24-35.

Rey-Debove, J. (1971). *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris: Mouton.

Rigau, G. (1999). La estructura del sintagma nominal: Los modificadores del nombre. En I. Bosque, y V. Demonte (eds.), *Gramática descriptiva de la lengua española* (Vol. 1). Madrid: Espasa. 311-362.

Robles i Sabater, F. (2011). ¿Qué es un buen ejemplo? La ejemplificación en la teoría lexicográfica alemana. *Revista de Filología Alemana*, 19, 247-261.

Rychly, P. (2008). A lexicographer-friendly association score. En Sojka, P; Horák, A. (Eds.) *Recent Advances in Slavonic Natural Language Processing*, RASLAN 2008. 6-9.

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Boston: Wesley Longman Publishing Co., Inc.

Sager, J. C. (1996). *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Sager, J. C.; Dungworth, D.; McDonald, P. (1980). *English Special Languages: principles and practice in science and technology*. Wiesbaden: Oscar Brandstetter.

San Vicente, F., Garriga, C., y Lombardini, H. (2012). *Ideolex: Estudios de Lexicografía e Ideología*. Monza, Italia: Polimetrica.

Sánchez, A. (2008). *John Sinclair (1933-2007). En memoriam*. Consulta 2 de junio del 2015: <http://www.aedean.org/NEXUS-Archive/Nexus2008.1/Nexus%202008.1-94.pdf>

Severance, E. & Washburn, M. F. (1907). The loss of associative power in words after long fixation. *The American Journal of Psychology*, 18 (2). 182-186.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*. Manchester.

Schumann, A. (2012). Towards the Automated Enrichment of Multilingual Terminology Databases with Knowledge-Rich Contexts. Experiments with Russian EuroTermBank Data. *CHAT 2012 Workshop Proceedings*. Linköping: Linköping Universitet. LiU Electronic Press. 27-34.

Seco, M. (1979). El "contorno" en la definición lexicográfica. En VOX, *Homenaje a Samuel Gili Gaya (In memoriam)*. Barcelona: Bibliograf S.A. 183-192.

Sierra, G. (2006). Diseño de corpus textuales para fines lingüísticos. *Memorias del IX Encuentro internacional de Lingüística del Noroeste, 2*. Sonora: Unison. 445-462.

Sierra, G., y Rosas, A. (2008). Clasificación de corpus textuales. *X Encuentro Intenacional de Lingüística del Noroeste*. Sonora: UNISON.

Sierra, G., Aguilar, C., Alarcón, R., y Bach, C. (2008). Definitional verbal patterns for semantic relation extraction. *Terminology*, 14 (1), 74-98.

Sierra, G., Alarcón, R., Aguilar, C., Bach, C. (2010). Definitional verbal patterns for semantic relation extraction. En Auger, A. Barriere, C. (eds.), *Probing Semantic Relations: Exploration and Identification in Specialized Texts*. Amsterdam/Philadelphia: John Benjamins Publishing. 73-96.

Sierra, G., Alarcón, R., y Aguilar, C. (2006). Extracción automática de contextos definitorios en textos especializados. *Procesamiento de lenguaje natural*, 37. 341-352.

Sierra, G., Alarcón, R., Molina, A., y Aldana, E. (2009). Web exploitation for Definition Extraction. En E. F. Edgar Chávez (ed.), *2009 Latin American Web Congress*. Los Alamitos, California: IEEE.

Sierra, G., Medina, A., y Lázaro, J. (2009). Determinación de la terminología básica en sexualidad a partir de la Web como corpus.

*Panorama de investigaciones basadas en Corpus*. Murcia: AELINCO. 374-385.

Sierra, G., Medina, A., y Lázaro, J. (2010). Terminótica y sexualidad: un proyecto integral. *XII Simposio RiTerm 2010: La terminología: puente ineludible de una sólida mediación cultural*. Buenos Aires: UMSA.

Sierra, G., Medina, A., Alarcón, R., y Aguilar, C. (2003). Towards the Extraction of Conceptual Information from Corpora. En D. Archer, P. Rayson, y A. y. Wilson, *Corpus Linguistics*. Lancaster: UCREL.

Sierra, G., Torres-Moreno, J.-M., López, F., Molina, A., Velázquez, P., Rosas, A. (2011). *Análisis de correos electrónicos para la detección de filtrado de información, seguimiento de un caso de filtración*. Instituto de Ingeniería, Universidad Nacional Autónoma de México, Grupo de Ingeniería Lingüística. México D.F.: IINGEN-UNAM.

Sinclair, J. (1987). *Collins COBUILD English Language Dictionary*. Birmingham: Collins.

Školníková, P. (2010). *Las colocaciones léxicas en el español actual*. (Tesina de fin de máster). Brno: Masarykova univerzita.

Smith, L. C. (1984). Semantic satiation affects category membership decision time but not lexical priming. *Memory & Cognition*, 12 (5). 483-488.

Smith, L. & Klein, R. (1990). Evidence for semantic satiation: repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16 (5). 852-861.

Spärck-Jones, K. (1972). A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation*, 28 (1), 11-21.

StatSoft, I. (2013). *Electronic Statistics Textbook (Electronic Version)*. Tulsa: StatSoft.

Steel, B. (ed.). (1999). Breve Diccionario Ejemplificado de Mexicanismos. Consultado el 2 de junio del 2015: <<http://www.briansteel.net/articsylibros/bdemsamp1ab.htm>>

Teubert, W. (2001). *Text corpora and multilingual lexicography*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Temmerman, R. (2000). *Towards new ways of terminology description: the sociocognitive-approach*. Amsterdam: John Benjamins Publishing.

Toutanova, K., Klein, D., Manning, C., y Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *Proceedings of HLT-NAACL 2003*. Edmonton: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. 252-259.

Torruella, J., y Llisterri, J. (1999). Diseño de corpus textuales y orales. En G. C. J. M. Blecua, *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Editorial Milenio. 45-77.

Torres-Moreno, J.-M., Velázquez-Morales, P., y Meunier, J.-G. (2002). Condensés de textes par des méthodes numériques. *Journées internationales d'Analyse statistique des Données Textuelles (JADT)*. Saint-Malo: Lexicométrie. Revue électronique. 1-12.

Torres-Moreno, J.-M. (2012). Beyond Stemming and Lemmatization: Ultra-stemming to Improve Automatic Text Summarization. *Computing Research Repository abs/1209.3126*.

Ullman, S. (1962). *Semántica. Introducción a la ciencia del significado*. Madrid: Aguilar.

Van Dijk, T. (2012). *Discurso y contexto*. Barcelona: Gedisa.

Vargas, C. (2009). *Bases de datos terminológicas en Internet*. Consulta el 2 de junio del 2015: [http://rua.ua.es/dspace/bitstream/10045/3823/1/BDT\\_enInternet.pdf](http://rua.ua.es/dspace/bitstream/10045/3823/1/BDT_enInternet.pdf)

Vieyra, J. (2011). *Adaptación, optimización y expansión de ECODE, un sistema extractor de contextos definitorios*. (Tesis de

licenciatura). México D.F.: Universidad Nacional Autónoma de México.

Wüster, E. (1931). *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik [La normalización internacional de la terminología técnica, particularment en electrotécnica]* (Tesis doctoral). Stuttgart: Bonn: Bouvier y Co. (reedición, 1970).

Wüster, E. (1996). «Die vier Dimensionen der Terminologearbeit» [Les quatre dimensions del treball terminològic]. En M. T. Cabré, *Terminologia. Selecció de textos d'E. Wüster*. Barcelona: Servei de Llengua Catalana de la Universitat de Barcelona. 69-106.

Wüster, E. (1979). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. (Vols. Teil 1-2. ). Springer-Verlag.

Ward Church, K., y Hanks, P. (1990). Word association, mutual information and lexicography. *Computational Linguistics* , 16 (1), 22-29.

Ward Church, K., y Hanks, P. (1990). Associat Sager, J. C. (1996). *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.

Zaid, G. (2011). Lejía. *Letras Libres*, 145. 64-65.

Zgusta, L. (1971). *Manual of Lexicography*. Paris: Walter de Gruyter.



# **ANEXO 1. Resultados del experimento con GENEX**

## **SEXUALIDAD**

### **Sexualidad**

1. Son los relacionados con la sexualidad .8
2. La incómoda relación de Howerd con su sexualidad , 10
3. Curiosidades Casualmente su nombre está relacionado ala sexualidad . 10
4. Este concepto tiene una estrecha relación con el concepto homosexualidad , 12
5. Los desórdenes mentales relacionados con la sexualidad están incluidos en el .13
6. La relación entre religión y sexualidad es variada en las distintas religiones . 14
7. La mayoría de sus libros están relacionados con la sexualidad de los chilenos 15
8. La masculinización es socialmente relacionada como un atributo clave de la homosexualidad femenina , 15
9. El afeminamiento es socialmente relacionado como un atributo clave de la homosexualidad masculina , 15
10. En la cultura secreta se desarrolló una estrecha relación entre la homosexualidad y el transgénero 16
11. Sus trabajos generalmente tratan hechos relacionados con el género , sexualidad y el cuerpo humano . 17
12. Debido a ello , los investigadores cuestionan una relación como respuesta evolutiva o de sexualidad , 17
13. La relación con las mariliendras permite a los participantes separar la intimidad de la sexualidad . 17
14. El erotismo denota todo lo relacionado con la sexualidad y no simplemente con el acto sexual físico 18
15. Su trabajo se relaciona con la espiritualidad , el judaísmo , la sexualidad y el Derecho . 18
16. Cabría preguntarse si carencias de esta hormona podrían tener alguna relación con la bisexualidad en humanos . 18
17. Era la diosa que eliminaba del mundo el pecado y la diosa más relacionada con la sexualidad . 19
18. Impacto social , seguridad , convivencia , relaciones interpersonales y sexualidad embebida alrededor de los eventos sonideros . 19

19. No es posible rebajar al erotismo todo lo relacionado con la sexualidad y con el acto sexual físico o 20
20. Dentro de la cultura sexual relacionada con la homosexualidad masculina , suele identificarse un fetichismo hacia los suspensorios . 20
21. Tiene una relación evidente con la sensualidad , la sexualidad y las capacidades de atracción entre seres humanos . 20
22. Se centra en las cuestiones relacionadas con el género , la sexualidad , el VIH / SIDA y el Islam . 22
23. En su pensamiento relacionaba la lucha feminista con la de la clase obrera y hacía aportes sobre la sexualidad femenina . 22
24. Los patrones de comportamiento sexual animal relacionados con la homosexualidad y la bisexualidad se manifiestan en una menor recurrencia poblacional que la heterosexualidad . 25
25. Las relaciones sexuales voluntarias entre hombres adultos dejaron de ser delito ; por primera vez en casi 100 años la homosexualidad era legal , 25
26. Los términos " zoosexualidad " y " zoosexual " se utilizan aproximadamente desde 1980 , en correlación con las orientaciones sexuales homosexualidad y heterosexualidad . 26
27. La relación entre religión y sexualidad implica una moral sexual , entendida no tanto como parte de la moral general o común a todos , 26
28. El judaísmo tradicional prohíbe las relaciones sexuales fuera del matrimonio , y se contempla como graves pecados el adulterio , el incesto y la homosexualidad . 27
29. Incluía diferentes secciones relacionadas con la literatura , el cine , la sexualidad ... Comenzó a emitirse en 2002 en el segundo canal de Telemadrid . 27
30. Sus temas de investigación se relacionan con los estilos de aprendizaje en educación física y el amor y la sexualidad en las relaciones de pareja . 27
31. Joyce tenía relaciones frecuentes con prostitutas y había adquirido cierta lujuria y agresiva sexualidad , cosa que no caracterizó el inicio del romance con Nora . 27
32. Con su madre desarrolla una relación de mujer a mujer , en la que hablan del rol de mujer en esos años y la sexualidad ; 27
33. Se reflejan distintas prohibiciones y mandatos de abstinencia relacionados con la dieta y la sexualidad , centrándose especialmente en la prohibición de comer carne de cerdo .

34. En el 2003 publicó el diario de género de cultura , se centró en la relación entre el género y la sexualidad en la cultura musical .  
28
35. Es un especializado en la relación entre religión y sexualidad , religión y derechos humanos , y en la relación entre religión griega y cristianismo primitivo . 28
36. A parte de los hombres , en varios episodios , Lois se ha revelado que ha mantenido relaciones sexuales con mujeres haciendo hincapié en su posible bisexualidad . 29
37. Para Freud lo relevante era establecer la causalidad psíquica de los fenómenos histéricos y presentar por vez primera una relación causal entre la sexualidad y la neurosis . 29
38. Esto puede ser relacionado con la similitud , el amor , romance , la intimidad , sexualidad , actividad sexual , espiritualidad , o compatibilidad y confianza . 29
39. Sin embargo , las relaciones entre " nadies " y varones o mujeres no merecen ningún tipo de categorización , ni como homosexualidad , ni como heterosexualidad , ni 30
40. La Pornopedia es una enciclopedia on - line orientada hacia temas relacionados con la pornografía , erotismo , sexualidad y todo lo relacionado con el género del sexo , 30
41. A pesar de una nueva disciplina , un número creciente de universidades han comenzado a ofrecer programas académicos relacionados con el sexo , la sexualidad y orientación sexual .  
30

## Sexo

1. Se originan en una pareja de distinto sexo que se reproduce y genera una descendencia más o menos numerosa . 21
2. Las " características sexuales secundarias " o " caracteres sexuales secundarios " son aquellos signos físicos y fisiológicos de madurez sexual que distinguen entre los dos sexos de una especie , 32
3. Es difícil distinguir los machos de las hembras , dado que los únicos dimorfismos existentes entre ambos sexos es el tamaño y que los machos pueden presentar sus gargantas de pigmentación más oscura . 35
4. La apariencia de los osiconos permite distinguir el sexo o la edad de una jirafa : los osiconos de las hembras y los jóvenes son delgados y tienen un pequeño penacho de pelo en la parte superior , 39
5. A veces se usa el término " prostitución " en el sentido mucho más amplio , de mantener relaciones sexuales con un fin distinto

de la reproducción o el placer de una de las partes , incluyendo formas de prostitución religiosa , en las que se practica sexo en cumplimiento de preceptos religiosos específicos . 56

6. Con numerosos ejemplos del reino animal y de culturas distintas de la occidental , muestra que la naturaleza y las diferentes sociedades ofrecen soluciones sorprendentes a la sexualidad : peces con varios tipos diferentes de machos o cuyos componentes cambian de sexo en caso de necesidad ; mamíferos que tienen a la vez órganos reproductores masculinos y femeninos , etc.

61

7. De igual modo se ha considerado que los baños de barro tienen un importante papel en las relaciones sociales de la especie e incluso se ha descrito un papel en la selección sexual , de modo que si mientras en el verano usan los baños de barro todos los jabalíes , sin distinciones de sexo ni edad , durante la época de celo parecen reservadas casi exclusivamente a los machos adultos , de modo que se ha considerado ( Pedro Fernández - Llarío , 2004 ) , que estos baños pueden estar ligados a la persistencia de los olores corporales sobre un sustrato estable como el que proporciona una capa de barro adherida al pelo ( incluso para perder rastros con los perros de caza ) , sin olvidar las funciones de marca territorialista , y sobre todo sanitarias que tienen para la especie los baños en barro . 150

## **Aborto**

1. Durante un tiempo Ruth se fue a vivir a la ciudad donde conoció al joven rico Brett Macklin de quien quedó embarazada , cuando este le dijo que abortara Ruth se negó y huyó , poco después empezó a salir con Frank Jonathan Morgan a quien engañó diciéndole que él era el padre del bebé , cuando se iban a casar Ruth no pudo más con su mentira y le cuenta la verdad a Frank .
2. Durante un tiempo Ruth se fue a vivir a la ciudad donde conoció al joven rico Brett Macklin de quien quedó embarazada , cuando este le dijo que abortara Ruth se negó y huyó , poco después empezó a salir con Frank Jonathan Morgan a quien engañó diciéndole que él era el padre del bebé , cuando se iban a casar Ruth no pudo más con su mentira y le cuenta la verdad a Frank .
3. Durante un tiempo Ruth se fue a vivir a la ciudad donde conoció al joven rico Brett Macklin de quien quedó embarazada , cuando este le dijo que abortara Ruth se negó y huyó , poco después empezó a salir con Frank Jonathan Morgan a quien engañó diciéndole que él era el padre del bebé , cuando se iban a casar

Ruth no pudo más con su mentira y le cuenta la verdad a Frank .  
77

## **Matrimonio**

1. Solo en 1966 pudo contraer matrimonio con Matilde 9
2. La pareja contrajo matrimonio el 26 de abril de 1967. 11
3. Las mujeres eran literalmente compradas para contraer matrimonio con ellas . 12
4. El apellido de una mujer cambia tradicionalmente tras contraer matrimonio en algunas culturas , 15
5. Tanto su padre como su abuelo contrajeron matrimonio con mujeres de familias adineradas . 15
6. Al muchacho joven que contraía matrimonio con el guerrero se le consideraba una mujer .16
7. Después contrajo matrimonio con Kurt Beringhoff , un hombre de negocios ; murió en 1987. 16
8. Anatol ha contraído matrimonio con una mujer polaca que ha abandonado su país natal , 16
9. Larreta perteneció a una familia adinerada y contrajo matrimonio con una mujer de familia aristocrática . 17
10. Su última mujer fue la actriz inglesa Lynne Frederick , con la que contrajo matrimonio en 1977. 18
11. A pesar de esta oposición la pareja contrajo matrimonio en Berlín el 28 de agosto de 1928 , 19
12. En 1919 contrae matrimonio con Marie Nys , una mujer belga refugiada en Inglaterra durante la Primera Guerra Mundial . 21
13. En la Academia Nacional de Compositores conoció a Blythe Newlon , una mujer quince años mayor , con quien contrajo matrimonio , 23
14. Así conoció a su primera mujer , la mezzo - soprano estadounidense Cathy Berberian con quien contraería matrimonio en 1950 , poco 23
15. Mientras , se había divorciado de su primera mujer y en 1935 contrajo matrimonio con Virginia Ruth Rogers , que duró poco . 24
16. Las principales leyes versaban principalmente en el terreno personal : no podían contraer matrimonio con alguna mujer perteneciente a la jurisdicción que administraba , 25
17. En julio de 1943 , Giulia presenta a Federico a sus padres , y la pareja contrajo matrimonio en octubre de ese mismo año . 26

18. J. Jonah Jameson contrajo matrimonio por primera vez , con una mujer de nombre Joan , con la cual tuvo un hijo , John , que  
27
19. Erasmus Darwin , abuelo paterno de Charles Robert Darwin , contrajo matrimonio con dos mujeres : Mary Howard en 1757 y Elizabeth Chandos - Pole en 1781. 28
20. Finalmente , pudo matricularse en la facultad de Económicas de Praga y en 1964 contrajo matrimonio con Olga Šplíchalová , una mujer de origen obrero , a quien años 30
21. El único punto oscuro de su vida , que fue objeto de crítica por parte de sus contemporáneos , fue el hecho de que Epaminondas nunca contrajera matrimonio , 30
22. En 1502 había contraído matrimonio en Santo Domingo con una mujer indígena que servía de mesonera en Santo Domingo , que con su bautizo vino a llamarse Leonor . 30
23. En 1871 contrajo matrimonio por compromiso con una mujer llamada Urilla Sutherland y era pariente de las esposas de sus hermanos , cuyo matrimonio de dos años sería desventurado .  
31
24. También tiene otra hermana , Salma , nacida en 2012 , fruto de la relación de su padre con su actual pareja Carmen Moreno , con la que contrajo matrimonio en 2014. 33
25. El 23 de diciembre de ese año , bajo un total hermetismo , la pareja se dirigió al Registro Civil , previo aviso a un par de conocidos suyos , y contrajeron matrimonio . 35
26. En contraste , la gobernante arregló el matrimonio de la hija de Ronglu , con el príncipe Chun , prometiendo que si la pareja tenía un hijo varón , éste sería designado heredero al trono . 37
27. Según ellos Blair , durante su gobierno , había adoptado medidas de carácter legislativo que van en contra de las creencias católicas tales como el apoyo al aborto y el matrimonio civil para las parejas homosexuales . 38
28. Segundo hijo de Pedro IV de Aragón y de su tercera mujer Leonor de Sicilia , Martín contrajo matrimonio en 1372 con María de Luna , hija de Don Lope , el primer conde de Luna , en 1374.  
40

## **Embarazo**

1. Las mujeres diabéticas deben cuidar su embarazo .  
0.138823395035964 9
2. Las mujeres que deciden interrumpir su embarazo ?  
0.138823395035964 9

3. Cada embarazo es diferente y cada mujer también .  
0.13138171562999 10
4. Cada mujer es única y cada embarazo también .  
0.13138171562999 10
5. La mujer está obligada a comunicar su embarazo ?  
0.13138171562999 10
6. Es un estado bastante habitual en la mujer durante el embarazo .  
0.189707134805748 13
7. Muchas mujeres experimentan inestabilidad en su estado de ánimo durante el embarazo . 0.142360850364503 14
8. Ser padre también es saber acompañar a la mujer durante el embarazo . 0.159164269374853 14
9. El embarazo representa un estado fisiológico normal para las mujeres y no una condición patológica .  
0.131699436249287 17
10. El estado nutricional de la mujer embarazada influye en el resultado final de su embarazo . 0.147244446025903  
17
11. Objetivo específico 1.3 Establecer nuevas acciones que beneficien a las mujeres en estado de embarazo . 0.131699436249287  
17
12. El embarazo es una época en el que un nuevo ser está creciendo en tu interior .0.14409373417654 18
13. es / s / bebés / calendario-embarazo El embarazo es uno de los estados más bonitos de toda mujer . 0.136276627276688  
21
14. Alrededor de 17 % de las mujeres británicas y 20 % de las estadounidenses declaran fumar durante el embarazo , señalaron los autores . 0.130202214499345 25
15. Más de 1 de cada 10 mujeres que tuvo hijos en Estados Unidos en el 2004 dijo que había fumado durante el embarazo .  
0.128458582597126 25

## **FINANZAS**

### **Banco**

1. La Sociedad Mexicana de la Dalia es muy reciente , se formó el 12 de octubre de 1995 , con el fin de investigar los antecedentes históricos , diagnóstico , ubicación taxonómica , importancia económica , propagación , cultivo y difusión de las dalias silvestres y cultivadas , así como la formación de un banco de germoplasma , el cual constituirá la base de los trabajos sobre el

género e incluirá las más de 30 especies hasta ahora reportadas , asimismo establecer un banco de información sobre dalias disponible para todos los socios y amigos aficionados a ellas , así como a los cultivadores y comerciantes .

2. Como gobernante Villa mandó imprimir diversos billetes , como los " dos caritas " y las " sábanas " ; embargó tiendas y sustituyó a los comerciantes venales por administradores honorables , llenó el mercado de reses de las haciendas en las que intervenía ; abarató los precios del maíz , frijol y carne ; asumió funciones federales en materia de telégrafos y ferrocarriles ; expulsó del estado a muchos españoles , acusados de que habían ayudado a Victoriano Huerta ; reabrió el Instituto Científico y Literario y decretó el establecimiento del banco del estado .
3. A través de sus bancos asociados , la Reserva Federal emite dinero , actúa como banco de liquidación financiera y establece las reservas de efectivo que los bancos deben mantener .
4. Juró apoyar la liberalización continua de la economía , aumentar la inversión en capital humano e infraestructuras , establecer un banco central independiente , e incrementar los ingresos por la imposición de recaudaciones fiscales más estrictas , en vez de aumentar impuestos .
5. No pueden celebrar tratados parciales de carácter político ; ni expedir leyes sobre comercio , o navegación interior o exterior ; ni establecer aduanas provinciales ; ni acuñar moneda ; ni establecer bancos con facultad de emitir billetes , sin autorización del Congreso Federal ; ni dictar los códigos Civil , Comercial , Penal y de Minería ,
6. Recientemente , ha establecido restricciones sobre las comisiones que los bancos pueden cobrar a sus clientes
7. Para formalizar la asociación , Enron establecía los términos del contrato y creaba una SPE que obtenía el dinero de los bancos para los productores de gas , a cambio de las futuras entregas de este recurso .
8. El nombre datáfono , erróneamente llamado TPV , se da al equipo que se instala en los establecimientos comerciales , empresas de servicios , bancos , hoteles , líneas aéreas ...
9. En 2011 el municipio contaba con 158 establecimientos comerciales , 25 restaurantes , 71 bares , 5 bancos o cajas de ahorro , 3 farmacias , 3 gasolineras y 10 hoteles u hostales .
10. En la política de descuento el banco central establece cuales son las condiciones por las que está dispuesto a conceder créditos a los banco comerciales a corto plazo .
11. El establecimiento de un banco nacional que promoviese una moneda única en todo el país , facilitando el comercio y emitiese

12. Recientemente , ha establecido restricciones sobre las comisiones que los bancos pueden cobrar a sus clientes 17
13. El establecimiento de un banco nacional que promoviese una moneda única en todo el país , facilitando el comercio y emitiese 22
14. En la política de descuento el banco central establece cuales son las condiciones por las que está dispuesto a conceder créditos a los banco comerciales a corto plazo . 30
15. El nombre datáfono , erróneamente llamado TPV , se da al equipo que se instala en los establecimientos comerciales , empresas de servicios , bancos , hoteles , líneas aéreas ... 32
16. A través de sus bancos asociados , la Reserva Federal emite dinero , actúa como banco de liquidación financiera y establece las reservas de efectivo que los bancos deben mantener . 32
17. En 2011 el municipio contaba con 158 establecimientos comerciales , 25 restaurantes , 71 bares , 5 bancos o cajas de ahorro , 3 farmacias , 3 gasolineras y 10 hoteles u hostales . 35
18. Para formalizar la asociación , Enron establecía los términos del contrato y creaba una SPE que obtenía el dinero de los bancos para los productores de gas , a cambio de las futuras entregas de este recurso . 39
19. Juró apoyar la liberalización continua de la economía , aumentar la inversión en capital humano e infraestructuras , establecer un banco central independiente , e incrementar los ingresos por la imposición de recaudaciones fiscales más estrictas , en vez de aumentar impuestos . 44
20. No pueden celebrar tratados parciales de carácter político ; ni expedir leyes sobre comercio , o navegación interior o exterior ; ni establecer aduanas provinciales ; ni acuñar moneda ; ni establecer bancos con facultad de emitir billetes , sin autorización del Congreso Federal ; ni dictar los códigos Civil , Comercial , Penal y de Minería , 59
21. Como gobernante Villa mandó imprimir diversos billetes , como los " dos caritas " y las " sábanas " ; embargó tiendas y sustituyó a los comerciantes venales por administradores honorables , llenó el mercado de reses de las haciendas en las que intervenía ; abarató los precios del maíz , frijol y carne ; asumió funciones federales en materia de telégrafos y ferrocarriles ; expulsó del estado a muchos españoles , acusados de que habían ayudado a Victoriano

Huerta ; reabrió el Instituto Científico y Literario y decretó el establecimiento del banco del estado . 97

22. La Sociedad Mexicana de la Dalia es muy reciente , se formó el 12 de octubre de 1995 , con el fin de investigar los antecedentes históricos , diagnóstico , ubicación taxonómica , importancia económica , propagación , cultivo y difusión de las dalias silvestres y cultivadas , así como la formación de un banco de germoplasma , el cual constituirá la base de los trabajos sobre el género e incluirá las más de 30 especies hasta ahora reportadas , asimismo establecer un banco de información sobre dalias disponible para todos los socios y amigos aficionados a ellas , así como a los cultivadores y comerciantes . 108

## **Crédito**

1. La entidad financiera pone a disposición cierta cantidad de dinero , de la cual se podrá disponer hasta ese límite , de forma que sólo se pagarán los intereses por las cantidades efectivamente dispuestas y no por la totalidad del crédito concedido . 44
2. Caitlin Cooke - Se une al grupo de amigos al comienzo de la serie , trabaja vendiendo limonada en " El Limón " , una sede con forma de limón para pagarle a su padre todo el dinero que gastó en su tarjeta de crédito .47
3. La liquidación está constituida por todas las operaciones posteriores a la disolución , que son necesarias y precisas para dar fin a los negocios pendientes , pagar el pasivo , cobrar los créditos y reducir a dinero todos los bienes de la sociedad , para repartirlo entre los socios . 51
4. Esa manipulación se lleva a cabo de forma irresponsable y sin escrúpulos , obteniendo dinero de la confianza y esperanza de los seres humanos , estableciendo el crédito del que cada uno puede gozar y el precio que debe pagar por él llegando incluso a imponer el crédito de los estados , que , según Agamben , han abdicado dócilmente de su soberanía . 65
5. A nivel más general , la principal es que el dinero moderno es , en última instancia , un pagaré estatal , y , como cualquier otro pagaré , su aceptación depende de la confianza que se tenga que , por un lado , el emisor lo aceptará como remisión de deuda por ese monto , y , por el otro , la capacidad de ese emisor en adquirir créditos con otros , 74
6. Incluyen el poder para : establecer y cobrar impuestos , pagar deudas , proveer para la defensa común y el bienestar general de los Estados Unidos , tomar prestado dinero en base al crédito de

los Estados Unidos , regular el comercio con otras naciones , establecer una regulación uniforme para la naturalización , acuñar moneda y regular su valor , proveer para el castigo de las falsificaciones , establecer oficinas postales y carreteras , promover el progreso y la ciencia , crear tribunales inferiores a la Corte Suprema , definir y castigar actos de piratería y felonía , declarar la guerra , hacer levas y mantener ejércitos , proveer y mantener una flota , dictar reglas para la regulación de las fuerzas terrestres y navales , proveer la milicia , armar y disciplinar la milicia , ejercitar la legislación exclusiva en Washington , y dictar las leyes necesarias para ejecutar las facultades del Congreso .

157

## **Nómina**

1. Por ejemplo , un archivo de nóminas puede contener datos sobre todos los empleados de una empresa y los detalles de su nómina ; cada registro del archivo de nóminas se refiere únicamente a un empleado , y todos los registros tienen la característica común de estar relacionados con las nóminas -
2. González fue premiado en 2010 con el Premio Action Award por la Asian American Legal Defense and Education Fund y un segundo premio George Polk Award por una serie de columnas en el " Daily News " en las cuales revelaba diversos sucesos relacionados con el proyecto CityTime , un sistema computerizado de control de nóminas , del alcalde Michael Bloomberg , que terminó en la acusación por fraude de cuatro consultorías .
3. En 1949 comienza su relación con la Galería Maeght entrando en la nómina de esta prestigiosa galería una de las más importantes del mundo ,
4. En La relación de Diego Palomino y en la relación anónima de la tierra de Jaén , se da cuenta de algunos aspectos de la vida social , política , económica de los pobladores del bajo Utcubamba en el siglo XVI .
5. El concepto de " hechicera " se relaciona " más con la manipulación de ensalmos , hierbas , nóminas , bendiciones , filtros , polvos , pelo , ropa , incienso , tierra de cementerio , agua bendita , conjuros , ligar y desligar , etc. ,
6. Pike se casó con Clarissa Harlow Brown en 1801 y siguió una carrera militar nada reseñable desempeñando tareas relacionadas con la logística y nóminas y sueldos en una serie de puestos fronterizos .

7. En 1949 comienza su relación con la Galería Maeght entrando en la nómina de esta prestigiosa galería una de las más importantes del mundo , 26
8. Pike se casó con Clarissa Harlow Brown en 1801 y siguió una carrera militar nada reseñable desempeñando tareas relacionadas con la logística y nóminas y sueldos en una serie de puestos fronterizos . 34
9. En La relación de Diego Palomino y en la relación anónima de la tierra de Jaén , se da cuenta de algunos aspectos de la vida social , política , económica de los pobladores del bajo Utcubamba en el siglo XVI . 43
10. El concepto de " hechicera " se relaciona " más con la manipulación de ensalmos , hierbas , nóminas , bendiciones , filtros , polvos , pelo , ropa , incienso , tierra de cementerio , agua bendita , conjuros , ligar y desligar , etc. , 48
11. Por ejemplo , un archivo de nóminas puede contener datos sobre todos los empleados de una empresa y los detalles de su nómina ; cada registro del archivo de nóminas se refiere únicamente a un empleado , y todos los registros tienen la característica común de estar relacionados con las nóminas - 53
12. González fue premiado en 2010 con el Premio Action Award por la Asian American Legal Defense and Education Fund y un segundo premio George Polk Award por una serie de columnas en el " Daily News " en las cuales revelaba diversos sucesos relacionados con el proyecto CityTime , un sistema computerizado de control de nóminas , del alcalde Michael Bloomberg , que terminó en la acusación por fraude de cuatro consultorías . 74

## **Préstamo**

1. De acuerdo a Douglas y sus seguidores , ese sistema tiene el resultado de transferir , incrementalmente , recursos al sector financiero -o quien sea que tenga el poder para crear ese dinero- en desmedro tanto de los productores como de los consumidores : en la medida que ambos sectores se ven obligados a recurrir a la deuda para solventar la diferencia económica entre lo que se produce y lo necesario para comprarlo , están , por lo tanto , obligados a pagar una porción de sus ingresos como interés por el uso de esos préstamos .
2. Mientras tanto , el fascismo lanzó la campaña para sostener el crecimiento demográfico : los varones solteros deben pagar un impuesto especial , para los matrimonios el Estado regala un

premio en dinero y se establecen préstamos , agilización de trámites económicos y exenciones de impuestos para las familias numerosas .

3. Su objetivo era conseguir el dinero necesario para pagar el préstamo y
4. M - PESA buscaba la creación de un servicio que permitiese a los beneficiarios de microcréditos recibir el dinero y pagar los préstamos con comodidad utilizando la red de revendedores de recargas de móviles de Safaricom .
5. Pagar más en cualquier mensualidad disminuye la cantidad de intereses que tendrá que pagar durante el resto de la vida del préstamo y
6. Si un estado puede pagar sus préstamos sin problemas tendrá menos intereses que uno que pueda tener problemas de dinero .
7. En este sentido , la tasa de interés es el precio del dinero , el cual se debe pagar/ cobrar por tomarlo prestado/ cederlo en préstamo en una situación determinada .
8. Desafortunadamente para él , no tiene bastante dinero para pagar el préstamo que
9. Para poder pagar esta cantidad los lugareños solicitan diversos préstamos , a veces a los vicarios de las parroquias vecinas .
10. Necesitaba el dinero para pagar algunos préstamos de colegio y deudas , entonces
11. Griffin asegura : " Si todos pagaran todo lo tomado en préstamo , no habría dinero en existencia " .
12. En el último episodio nos enteramos de que a causa de la mala economía , La Pizzería Scavo tiene bajas ventas , y Tom se vio obligado a vender su negocio para reunir el dinero para pagar sus préstamos y pagar a Bree su dinero .
  
13. Su objetivo era conseguir el dinero necesario para pagar el préstamo y 13
14. Necesitaba el dinero para pagar algunos préstamos de colegio y deudas , entonces 14
15. Desafortunadamente para él , no tiene bastante dinero para pagar el préstamo que 14
16. Griffin asegura : " Si todos pagaran todo lo tomado en préstamo , no habría dinero en existencia " . 21
17. Para poder pagar esta cantidad los lugareños solicitan diversos préstamos , a veces a los vicarios de las parroquias vecinas .

18. Si un estado puede pagar sus préstamos sin problemas tendrá menos intereses que uno que pueda tener problemas de dinero .  
22
19. Pagar más en cualquier mensualidad disminuye la cantidad de intereses que tendrá que pagar durante el resto de la vida del préstamo y 24
20. En este sentido , la tasa de interés es el precio del dinero , el cual se debe pagar/ cobrar por tomarlo prestado/ cederlo en préstamo en una situación determinada .32
21. M - PESA buscaba la creación de un servicio que permitiese a los beneficiarios de microcréditos recibir el dinero y pagar los préstamos con comodidad utilizando la red de revendedores de recargas de móviles de Safaricom . 38
22. En el último episodio nos enteramos de que a causa de la mala economía , La Pizzería Scavo tiene bajas ventas , y Tom se vio obligado a vender su negocio para reunir el dinero para pagar sus préstamos y pagar a Bree su dinero . 47
23. Mientras tanto , el fascismo lanzó la campaña para sostener el crecimiento demográfico : los varones solteros deben pagar un impuesto especial , para los matrimonios el Estado regala un premio en dinero y se establecen préstamos , agilización de trámites económicos y exenciones de impuestos para las familias numerosas . 52
24. De acuerdo a Douglas y sus seguidores , ese sistema tiene el resultado de transferir , incrementalmente , recursos al sector financiero -o quien sea que tenga el poder para crear ese dinero- en desmedro tanto de los productores como de los consumidores : en la medida que ambos sectores se ven obligados a recurrir a la deuda para solventar la diferencia económica entre lo que se produce y lo necesario para comprarlo , están , por lo tanto , obligados a pagar una porción de sus ingresos como interés por el uso de esos préstamos . 98

## **Pensión**

1. El 17 de noviembre fue citado a declarar por primera vez ante el Juzgado Militar número 6 de León , para responder de la misma , acusado de haber violado el punto 4 del artículo 329 del Código de Justicia Militar ( según dicho artículo del código de 1890 , vigente entonces , se consideraba falta grave , penada con arresto militar o suspensión de empleo , que los militares asistieran a manifestaciones políticas o acudieran a la prensa en relación con asuntos del servicio ) .

2. Otros personajes como el fundador de Macondo , José Arcadio Buendía ( quien muere solo , atado a un árbol ) , Úrsula ( quien vive la soledad en la ceguera de su vejez ) , José Arcadio ( hijo del fundador ) y Rebeca ( quienes se marchan a habitar solos en otra casa por haber " deshonrado " a la familia ) , Amaranta ( quien permanece y muere soltera y virgen ) , Gerineldo Márquez ( quien espera una pensión que nunca llega y el amor de Amaranta ) , Pietro Crespi ( quien se suicida ante el rechazo de Amaranta ) , José Arcadio Segundo ( quien desde que vio un fusilamiento nunca tuvo relación con nadie y pasó sus últimos años encerrado en el cuarto de Melquíades ) , Fernanda del Carpio ( quien fue criada para ser reina y la primera vez que sale de su casa es a los 12 años de edad ) , Remedios 'Meme ' Buendía ( la cual fue enviada a un convento , en contra de su voluntad , pero completamente resignada luego de la desgracia que sufrió Mauricio Babilonia y se condena al eterno silencio ) , y Aureliano Babilonia ( quien pasa encerrado en el cuarto de Melquíades ; e incluso hubo un momento en que habitó completamente solo en la casa de los Buendía , después del asesinato del último José Arcadio y antes de la llegada de Amaranta Úrsula ) entre otros , sufren las consecuencias de su soledad y abandono .
3. En caso de igualdades en este parámetro se utilizan los siguientes en orden sucesivo en caso de haber más igualdades estadísticas : mayor diferencia de goles , mayor número de goles marcados y menor puntaje de juego limpio ( 1 punto por tarjeta amarilla o suspensión de un partido ; 2 puntos por tarjeta roja ) .
4. Los vehículos que hayan sufrido un daño importante a consecuencia de un accidente que pueda haber afectado a algún elemento de seguridad ( dirección , transmisión , frenado , suspensión , bastidor o alguna estructura de anclaje ) pueden ser obligados a realizar una revisión según el dictamen del atestado , retirándoles en este caso el permiso de circulación a sus conductores y enviándolo a Tráfico hasta que la avería este solventada .
5. no encontrarse en suspensión de dicho derecho ( por interdicción en caso de demencia ; por hallarse procesado por delito que merezca pena aflictiva o por delitos calificados por la ley como conducta terrorista , y por haber sido sancionado por el Tribunal Constitucional de acuerdo con el inciso séptimo del número 15. Así pues , Luis pasó toda su infancia y adolescencia en Zaragoza , donde cursó la educación primaria y secundaria , primero en Corazonistas ( con mayoría de franceses ) y en 1908 , durante siete años , en el colegio jesuita de El Salvador , al comienzo del paseo de la Constitución , donde hoy se encuentra la sede

principal de Ibercaja , cerca de la plaza de Aragón ; como alumno a media pensión , no vestía el uniforme completo de los internos , sino solamente la gorra con un galón .

6. El 17 de noviembre fue citado a declarar por primera vez ante el Juzgado Militar número 6 de León , para responder de la misma , acusado de haber violado el punto 4 del artículo 329 del Código de Justicia Militar ( según dicho artículo del código de 1890 , vigente entonces , se consideraba falta grave , penada con arresto militar o suspensión de empleo , que los militares asistieran a manifestaciones políticas o acudieran a la prensa en relación con asuntos del servicio ) .
  
7. El 17 de noviembre fue citado a declarar por primera vez ante el Juzgado Militar número 6 de León , para responder de la misma , acusado de haber violado el punto 4 del artículo 329 del Código de Justicia Militar ( según dicho artículo del código de 1890 , vigente entonces , se consideraba falta grave , penada con arresto militar o suspensión de empleo , que los militares asistieran a manifestaciones políticas o acudieran a la prensa en relación con asuntos del servicio ) . 88

## **INFORMÁTICA**

### **Virus**

23. Existen diversos tipos de virus , varían según su función o la manera en que este se ejecuta en nuestra computadora alterando la actividad de la misma , entre los más comunes están : Otros tipos por distintas características son los que se relacionan a continuación : La característica principal de estos virus es que se ocultan en la memoria RAM de forma permanente o residente .
24. Los malwares que tienen por objeto alterar el normal funcionamiento del ordenador en forma de virus informático son en su mayoría programas homebrew también .
  
25. Existen diversos tipos de virus , varían según su función o la manera en que este se ejecuta en nuestra computadora alterando la actividad de la misma , entre los más comunes están : Otros tipos por distintas características son los que se relacionan a continuación : La característica principal de estos virus es que se ocultan en la memoria RAM de forma permanente o residente .

26. Los malwares que tienen por objeto alterar el normal funcionamiento del ordenador en forma de virus informático son en su mayoría programas homebrew también .
27. Los malwares que tienen por objeto alterar el normal funcionamiento del ordenador en forma de virus informático son en su mayoría programas homebrew también . 26
28. Existen diversos tipos de virus , varían según su función o la manera en que este se ejecuta en nuestra computadora alterando la actividad de la misma , entre los más comunes están : Otros tipos por distintas características son los que se relacionan a continuación : La característica principal de estos virus es que se ocultan en la memoria RAM de forma permanente o residente .

68

## Computadora

1. La limitación más evidente del diodo de Schottky es la dificultad de conseguir resistencias inversas relativamente elevadas cuando se trabaja con altos voltajes inversos pero el diodo Schottky encuentra una gran variedad de aplicaciones en circuitos de alta velocidad para computadoras donde se necesiten grandes velocidades de conmutación y mediante su poca caída de voltaje en directo permite poco gasto de energía , otra utilización del diodo Schottky es en variadores de frecuencia ( inversers ) para que la corriente que vuelve desde el motor al variador no pase por los transistores IGBT del chopper , lo cual conduciría a su rápido deterioro .
2. Gracias a esta maniobra de la defensa Mitnick fue sentenciado a solo un año de prisión , y al salir de allí debía seguir un programa de seis meses para tratar su " " adicción a las computadoras " " .
3. En sus mensajes en línea , Titor proclamaba ser un soldado estadounidense del año 2036 , comisionado en Tampa en Hillsborough County , Florida , quien fue asignado a un proyecto gubernamental de viajes por el Tiempo , en donde supuestamente se le asignó la misión de regresar al año 1975 y conseguir una computadora IBM 5100 para usarlo , según dijo , en " editar " varios programas de computadora antiguos en el año 2036 ; con problemas del código de error " timeout " UNIX 2038. Un e - comic no es más que un archivo de computadora , una versión electrónica o digital de un documento gráfico .
4. Deep Junior , pese a conseguir el campeonato mundial , es un programa más orientado a combatir contra jugadores humanos , y

en buscar y descubrir nuevos recursos en el juego ciencia , su condición de buscar un juego dinámico pese al sacrificio de material es una de sus características principales , actualmente se espera al mercado el Deep Junior 11 , para mantener dentro del ciclo competitivo de las computadoras y con nuevas herramientas de aprendizaje como hizo Chessbase con Fritz 11. En la lista de la SSDF Junior 6 es el primer programa en romper la barrera de los 2700 , apareciendo con un ELO de 2706 , en una computadora de 128MB K6 - 2 450 MHz [ <http://web.archive.org/20010427125834/www.geocities.com/CapeCanaveral/Launchpad/2640/ssdf/2000/ssdf0001.htm> El 26 de enero del 2003 , Kaspárov se enfrentó contra Deep Junior , al final el resultado quedó empatado con marcador de 3 - 3 , Kaspárov ganó la primera partida y perdió la tercera .

5. Incluso después de agregar todos esos solenoides e interruptores , conseguir que la Selectric pueda conectarse con una computadora fue un proyecto grande .
6. Junto con Gerald Jay Sussman , Abelson desarrollado introductoria del MIT tema ciencias de la computación , la estructura e interpretación de los programas de ordenador , un tema organizado en torno a la idea de que un lenguaje de programación es principalmente un medio formal para expresar sus ideas acerca de la metodología , y no sólo una manera de conseguir una computadora para realizar operaciones .
7. El código fuente de un programa informático ( o software ) es un conjunto de líneas de texto que son las instrucciones que debe seguir la computadora para ejecutar dicho programa .
  
8. La limitación más evidente del diodo de Schottky es la dificultad de conseguir resistencias inversas relativamente elevadas cuando se trabaja con altos voltajes inversos pero el diodo Schottky encuentra una gran variedad de aplicaciones en circuitos de alta velocidad para computadoras donde se necesiten grandes velocidades de conmutación y mediante su poca caída de voltaje en directo permite poco gasto de energía , otra utilización del diodo Schottky es en variadores de frecuencia ( inverters ) para que la corriente que vuelve desde el motor al variador no pase por los transistores IGBT del chopper , lo cual conduciría a su rápido deterioro .
9. Junto con Gerald Jay Sussman , Abelson desarrollado introductoria del MIT tema ciencias de la computación , la estructura e interpretación de los programas de ordenador , un tema organizado en torno a la idea de que un lenguaje de

programación es principalmente un medio formal para expresar sus ideas acerca de la metodología , y no sólo una manera de conseguir una computadora

10. El código fuente de un programa informático es un conjunto de líneas de texto que son las instrucciones que debe seguir la computadora para ejecutar dicho programa .
11. El código fuente de un programa informático es un conjunto de líneas de texto que son las instrucciones que debe seguir la computadora para ejecutar dicho programa .29
12. Junto con Gerald Jay Sussman , Abelson desarrollado introductoria del MIT tema ciencias de la computación , la estructura e interpretación de los programas de ordenador , un tema organizado en torno a la idea de que un lenguaje de programación es principalmente un medio formal para expresar sus ideas acerca de la metodología , y no sólo una manera de conseguir una computadora 65
13. La limitación más evidente del diodo de Schottky es la dificultad de conseguir resistencias inversas relativamente elevadas cuando se trabaja con altos voltajes inversos pero el diodo Schottky encuentra una gran variedad de aplicaciones en circuitos de alta velocidad para computadoras donde se necesiten grandes velocidades de conmutación y mediante su poca caída de voltaje en directo permite poco gasto de energía , otra utilización del diodo Schottky es en variadores de frecuencia ( inverters ) para que la corriente que vuelve desde el motor al variador no pase por los transistores IGBT del chopper , lo cual conduciría a su rápido deterioro . 105

## **Programa**

1. Erlang es otro lenguaje funcional de programación concurrente . 10
2. En 1941 Konrad Zuse presentó el Z3 , primera computadora completamente funcional y programable del mundo 17
3. Intentó modernizar , impulsar y racionalizar el funcionamiento de la administración con un programa de reformas 17
4. La generación de pruebas para cada funcionalidad hace que el programador confíe en el código escrito . 18
5. Se han desarrollado muchas otras bibliotecas para proporcionar una funcionalidad equivalente a la de otros lenguajes de programación . 20

6. Las instrucciones de lenguaje ensamblador LDB y DPB sobreviven como funciones en el lenguaje de programación Common Lisp .20
7. El Z3 se convirtió así en la primera computadora funcional , de todo propósito , controlada por programa . 20
8. Los interruptores generalmente son botones , que permiten efectuar funciones determinadas por el programa que se esté ejecutando . 20
9. Posee cientos de funciones matemáticas y la posibilidad de integrar programas en los lenguajes más usados . Kenneth Iverson , 21
10. En algunos simples diseños de computador con programas almacenados , un mal funcionamiento del programa puede dañar el computador .21
11. El " Computer Language Shootout Benchmarks " compara el funcionamiento de implementaciones de problemas de programación típicos , en diversos lenguajes . 23
12. Los " scripts " Tcl son a menudo más compactos y legibles que los programas funcionalmente equivalentes en otros lenguajes de programación . 24
13. Las principales funciones que los sistemas de tipificación ofrecen son : Un programa normalmente asocia cada valor con un tipo de dato determinado . 25
14. Dicha computadora asume menos funciones y , por tanto , puede atender a un número mayor de programas cliente con los mismos recursos . 25
15. Del mismo modo , la idea de los datos inmutables de la programación funcional se incluye a menudo en lenguajes de programación imperativa , 25
16. Estos micro controladores tienen un programa instalado para su funcionamiento , estos mismos programas son ejecutados y realizan la exploración matricial de las teclas 25
17. El teorema del aumento de velocidad de Blum dice que por cualquier medida de complejidad hay funciones computables que no tienen un programa mínimo . 26
18. Entre las ventajas que suelen citarse de usar un paradigma funcional en la programación de computadoras , están las siguientes : Simulación de estados . 26
19. Los programas escritos en un lenguaje funcional están constituidos únicamente por definiciones de funciones , entendiendo éstas no como subprogramas clásicos de un lenguaje imperativo , 27

20. Es un depurador portable que se puede utilizar en varias plataformas Unix y funciona para varios lenguajes de programación como C , C++ y Fortran .27
21. La combinación de las mismas pueden resultar en algunas funciones casi triviales cuya versión en lenguajes imperativos pueden llegar a resultar extremadamente tediosas de programar .  
27
22. Los símbolos en la tabla de símbolos pueden referirse a constantes , a funciones o a tipos de datos en el código fuente de un programa . 28
23. Curry es un lenguaje experimental de programación en lógica funcional , basado en el lenguaje Haskell , que combina elementos de programación funcional y programación lógica . 28
24. Los factores neurotróficos son secretados por determinados tejidos , siendo una de sus funciones impedir a las neuronas diana que inicien la muerte celular programada , permititiendo  
28
25. En arquitectura de computadoras , una unidad funcional o unidad de ejecución es una parte de la CPU que realiza las operaciones y cálculos llamados por los programas . 30
26. En la actualidad , la funcionalidad de la instrucción sigue existiendo en todos los lenguajes de programación , bajo nombres específicos y con un campo de uso delimitado . 30
27. El funcionamiento de los programas de ajedrez consiste , esencialmente , en explorar un número muy elevado de posibles futuros movimientos y aplicarles una función de evaluación al resultado , 31
28. DDD es usado principalmente en sistemas UNIX , y su funcionalidad se puede complementar con Plug - in desarrollados por programadores de la comunidad del software de código abierto . 31
29. Fue el primer lenguaje puramente funcional en ser destinado al uso comercial y no al académico La solución a muchos ejemplos de programas resulta más simple en Miranda que en 31
30. Buena parte de la funcionalidad de los archivos de scripts ha sido ahora asumida por la inserción del lenguaje de programación Python en las rutinas de sintaxis del SPSS . 31
31. El software de KDE está escrito principalmente en C++ , un lenguaje derivado del lenguaje de programación C con muchas funcionalidades añadidas , entre ellas programación orientada a objetos . 31
32. En ciencias de la computación , la programación funcional es un paradigma de programación declarativa basado en la utilización

- de funciones aritméticas que no maneja datos mutables o de estado .32
33. Se refiere al sistema informático que tiene el hospital y que soporta su funcionamiento en redes de computadoras y programas diseñados especialmente para el correcto funcionamiento de todas las áreas . 32
  34. Object Pascal expande las funcionalidades del Pascal estándar : La potencia del lenguaje de programación Delphi , se basa principalmente en la gestión y administración de Base de Datos , concepto 32
  35. La programación funcional moderna y la programación orientada a objetos se preocupan mucho menos de “empujar tantas palabras hacia un lado y otro” que los anteriores lenguajes como era Fortran . 32
  36. El gran inconveniente es que a menudo no funcionan especialmente bien en conjunto , normalmente son los propios programadores familiarizados con las diferentes bibliotecas quienes consiguen sacarles el máximo partido , 32
  37. En informática , una biblioteca es un conjunto de implementaciones funcionales , codificadas en un lenguaje de programación , que ofrece una interfaz bien definida para la funcionalidad que se invoca . 33
  38. Una interfaz de funciones foráneas es un mecanismo por el cual un programa escrito en un lenguaje de programación puede realizar llamadas a funciones o usar los servicios escritos en otro . 33
  39. Frecuentemente se clasifica a ML como un lenguaje funcional impuro dado que permite programar imperativamente con efecto colateral , a diferencia de otros lenguajes de programación funcional , tales como Haskell . 33
  40. Recientes declaraciones de funcionarios del gobierno polaco indican que existen planes para acelerar el programa de construcción , que pretenden reducir la finalización de la autopista al año 2012 ante Eurocopa 2012. 33
  41. Casi toda la funcionalidad del editor , desde las operaciones básicas de edición hasta la configuración de la interfaz de usuario , es controlada por un dialecto del lenguaje de programación Lisp . 34
  42. Los analizadores sintácticos fueron extensivamente estudiados durante los años 70 del siglo XX , detectándose numerosos patrones de funcionamiento en ellos , cosa que permitió la creación de programas generadores de analizadores sintácticos

43. Como resultado , las funciones de Lisp pueden ser manipuladas , alteradas o aún creadas dentro de un programa Lisp sin un extensivo análisis sintáctico o manipulación de código de máquina binario . 34
44. El grado de agregación de las cuentas públicas se corresponde con los niveles de gestión : el funcionario que administra determinado programa gubernamental precisará trabajar con una mayor desagregación que el Ministro . 34
45. Entre ellas se encuentran las siguientes instrucciones : Las instrucciones de entrada/ salida son : Además están las cientos de funciones matemáticas y gráficas integradas en la calculadora están disponibles para la programación . 35
46. Los sistemas operativos proveen un conjunto de funciones necesarias y usadas por diversos programas de aplicaciones de una computadora , y los vínculos necesarios para controlar y sincronizar el hardware de la misma . 35
47. Logo es un lenguaje de programación de alto nivel , en parte funcional , en parte estructurado ; de muy fácil aprendizaje , razón por la cual suele ser el lenguaje de programación preferido 35
48. Generalmente en los lenguajes de programación funcional y en otros lenguajes de programación , una tupla es un objeto que bien puede tener datos o diversos objetos , de forma similar a una tupla definida matemáticamente . 38
49. Este amplio colectivo ha dedicado su esfuerzo en forma de trabajo individual o en proyectos comunes que van acumulando en internet funcionalidades y librerías de código abierto que son utilizadas y ampliadas por nuevos programadores autodidactas . 38
50. El lenguaje en sí era poderoso , contando con una gran cantidad de sentencias , funciones , administración de memoria y variables que permitían al programador desarrollos muy flexibles y eficientes ; en forma razonablemente rápida . 38
51. Xlib es una biblioteca que reúne un conjunto de funciones y macros realizadas en el lenguaje de programación C y utilizadas por un cliente para hacer de interfaz con el servidor gráfico de X Window System . 38
52. Los usuarios pueden crear nuevas funcionalidades escribiendo un " dialplan " en el lenguaje de script de Asterisk o añadiendo módulos escritos en lenguaje C o en cualquier otro lenguaje de programación soportado en GNU/ Linux . 38
53. A cambio de tantas responsabilidades la recompensa es el buen funcionamiento de la red como un medio que vincula personas y

- de los computadores y programas como herramientas para agilizar algunas labores que dan tiempo y dar tiempo 39
54. Además de las funcionalidades básicas que son habituales en un editor de texto , como copiar , cortar y pegar texto , imprimir , etc. , gedit incorpora coloreado de sintaxis para diversos lenguajes de programación y marcado . 40
  55. Las funciones `include` , `include_once` , `require` , `require_once` son utilizadas para incluir en una misma página otras páginas a la vez , por la necesidad de utilizar un código fuente o por otras diversas razones de programación . 40
  56. Una característica bastante inusual del i860 era que las pipelines en las unidades funcionales eran accesibles por los programas , requiriendo a los compiladores ordenar cuidadosamente las instrucciones en el código de objeto para mantener las pipelines llenas . 40
  57. El sistema operativo tiene otras funciones que ofrecer a otros programas , como los códigos que sirven a los programadores , escribir programas para una máquina sin necesidad de conocer los detalles internos de todos los dispositivos electrónicos conectados . 41
  58. El entorno o ambiente Smalltalk es primordialmente gráfico y funciona como un sistema en tiempo de ejecución que integra varias herramientas de programación , utilidades multimedia , interfaces para ejecutar código no nativo a Smalltalk y servicios del sistema operativo . 42
  59. En el estudio y uso de los semáforos se puede ver que las llamadas a las funciones necesarias para utilizarlos quedan repartidas en el código del programa , haciendo difícil corregir errores y asegurar el buen funcionamiento de los algoritmos . 42
  60. Pero en la práctica , desde su fundación en 1969 , Televisión Boliviana siempre ha funcionado como un instrumento de propaganda del gobierno de turno en Bolivia , y su programación ha dependido de los vaivenes del partido en función de gobierno 43
  61. Uno de los más modernos y completos compiladores para lenguaje C es , que es un ambiente de desarrollo con editor de texto , bibliotecas con múltiples funciones para todos los módulos y herramientas incorporadas para facilitar enormemente el proceso de programación . 44
  62. Esta importante sede comienza a funcionar el 23 de enero de 1995 y dispone de aulas con diseño aclimatado a la zona ; biblioteca , igualmente adaptada al medio ambiente , modernos

- laboratorios de computación con sofisticados equipos y programas en distintas áreas . 45
63. UML ofrece un estándar para describir un " plano " del sistema , incluyendo aspectos conceptuales tales como procesos de negocio , funciones del sistema , y aspectos concretos como expresiones de lenguajes de programación , esquemas de bases de datos y compuestos reciclados . 46
  64. Una JAL cumple funciones concernientes con los planes y programas distritales de desarrollo económico y social de obras públicas , vigilancia y control a la prestación de los servicios públicos en su localidad y las inversiones que se realicen con los recursos del Distrito Capital , 47
  65. Una JAL cumple funciones concernientes con los planes y programas municipales de desarrollo económico y social de obras públicas , vigilancia y control a la prestación de los servicios municipales en su comuna o corregimiento y las inversiones que se realicen con los recursos públicos , 47
  66. Los visitantes empresa e instituciones en grupos de funcionarios , empleados o trabajadores podrán hacer uso de las instalaciones y programas de outdoors a cargo de instructores part time dependientes del proyecto y guías preferente formados entre jóvenes de la zona y capacitados para este propósito . 48
  67. Bibliotecas C. Una biblioteca de C es una colección de funciones utilizadas en el lenguaje de programación C. Las bibliotecas más comunes son la biblioteca estándar de C y la biblioteca del estándar ANSI C , la cual provee las especificaciones de los estándares que son ampliamente compartidas entre bibliotecas .  
52
  68. El entonces Consejo Nacional de la Universidad Peruana , transcribe un oficio del Ministerio de Pesquería por el que se pide que la Universidad Nacional de Piura opine respecto al funcionamiento de un programa de Pesquería en la Ciudad de Tumbes ; esta petición determinó el nombramiento de una comisión multisectorial 52
  69. Para programas que realizan cálculos numéricos intensivos , los lenguajes funcionales tales como OCaml y Clean son algo más lentos que C. Para programas que manejan grandes matrices y bases de datos multidimensionales , los vectores de los lenguajes funcionales , como J y K , fueron diseñados optimizando su velocidad . 53
  70. Al mismo tiempo , Bjarne Stroustrup empezó a trabajar con algunos compañeros de Bell Labs para añadir funcionalidades de programación orientada a objetos a C. El lenguaje que crearon , llamado C++ , es hoy en día el lenguaje de programación de

aplicaciones más común en el sistema operativo Microsoft Windows ; 53

71. Es una herramienta escrita en lenguajes C y Python que permite crear y reproducir guiones o " scripts " que repliquen el comportamiento de una aplicación informática en aquellas que utilicen las bibliotecas de accesibilidad propias del entorno Linux Permite generar casos de prueba repetibles que permiten revisar la consistencia funcional del programa . 55
72. Las primeras computadoras Apple II estuvieron a la venta el 5 de junio de 1977 , basadas en un microprocesador 6502 de MOS Technology funcionando a 1 MHz , con 4 KiB de RAM , 12 KiB de ROM con el lenguaje de programación Integer BASIC , y una interfaz para grabadores de casetes de audio . 58
73. Está disponible para las plataformas Unix , Windows , Mac OS X y GNU/ Linux Entre sus prestaciones básicas se hallan : la manipulación de matrices , la representación de datos y funciones , la implementación de algoritmos , la creación de interfaces de usuario y la comunicación con programas en otros lenguajes y con otros dispositivos hardware . 60
74. La transparencia referencial es un término utilizado en la programación funcional que se refiere a la propiedad por la cual " en un programa , una expresión " E " del lenguaje puede ser sustituida por otra de igual valor " V " , resultando en un programa cuya semántica no va a diferir de la del original " . 61
75. El material , inicialmente acordado por el Consejo Universitario , era la publicación de las disposiciones legales de la universidad y la instrucción pública , las memorias aceptadas que son publicadas " in extenso " , un resumen de las memorias rechazadas , cuenta del Consejo ante el Gobierno , nota sobre funcionarios fallecidos y los programas de la Universidad . 62
76. Ésta técnica usada de forma maliciosa permitiría a un atacante modificar el funcionamiento normal de nuestro programa y nuestro sistema , y es al menos una técnica útil si no lo evitamos en lenguajes muy populares como el ejemplo C++ . Muchos lenguajes de programación ofrecen clases o funciones integradas de biblioteca en tiempo de ejecución que rodean las API de Windows subyacentes , permitiendo 66

## Teclado

1. el teclado es francamente pequeño , se escribe bien pero las teclas están demasiado juntas y pequeñas . 0.0770213406956418  
19
2. Me gusta el híbrido que hace entre el teclado moderno y el de las primeras máquinas de escribir . 0.0741118738679954 20
3. Especialmente ingenioso el teclado hecho a la forma de las clásicas máquinas de escribir , algo recargado pero bonito .  
0.0741118738679954 21
4. El teclado es un poco confuso y cuesta escribir , pero al menos hay varias máquinas instaladas por toda la terminal .  
0.0716755730002658 23
5. El teclado numérico se dice que es ultrafino , pero cuando se deslizan las teclas Qwerty es más fácil para escribir .  
0.0716755730002658 23
6. Tecla Espacio Tecla Sym : al escribir texto , pulse la tecla Sym para introducir caracteres que no aparecen en el teclado .  
0.087784290436129 24
7. Para escribir este símbolo con el teclado , tenemos que pulsar la tecla " Alt Gr " y la tecla con el número dos .  
0.07521764407644 26
8. Recuerdo mi anterior teclado Packard Bell , cuyas teclas eran casi tan pesadas como una máquina de escribir , sobre todo después de unos años de uso . 0.0820298719800126 29
9. [ 10 ] La distribución de teclado QWERTY se ha convertido en el estándar de facto para máquinas de escribir y teclados de computadoras ingleses y españoles . 0.0717380104183071  
29
10. As? , se encuentra el teclado normal para las teclas que se muestran inicialmente , el teclado caps con las teclas que se escriben al pulsar sobre la tecla Bloq . 0.0869451443135413  
32
11. Las instrucciones del lenguaje máquina realizan tareas muy sencillas , como , por ejemplo , sumar dos números , detectar qué tecla se ha pulsado en el teclado o escribir algo en la pantalla del ordenador . 0.108097713608463 38
12. El aprendizaje consta de 3 partes : PRIMERA PARTE : Empieza con el teclado del medio , que son las teclas guías : para empezar a escribir a máquina siempre los dedos van en esas teclas .  
0.0866709875606738 38
13. Lleva un teclado de 48 teclas , una tecla especial para mensajes y dos programables , el estilo de teclado es QWERTY , es decir que sigue el orden de un teclado de ordenador o máquina de escribir .  
0.0707534355958477 40

14. El teclado consta de teclas mecánicas con muelle que hace que se produzca el efecto que los más tradicionales escritores , los que lo han hecho durante años con una máquina de escribir mecánica , aprecian allá donde quiera que van . 0.0774528659090905  
43
15. Los textos se escribían mediante un teclado de 26 teclas dispuestas en 3 hileras , en forma semejante a los teclados de las máquinas de escribir alemanas , y a cada tecla correspondía una bombilla eléctrica tipo linterna de 3,8 V. , y 0,20 A. , que al accionarse la tecla , se encendía y señalaba el nuevo valor de la letra cifrada . 0.0735540881082685 65

## **Informática**

1. Desarrollo histórico de la informática . 0.101896593248487 7
2. La informática es muy divertida , desarrollar software es especial .  
0.112134353050738 12
3. Conocimientos ?tiles para el desarrollo del trabajo : informática , idiomas . 0.100296014413765 13
4. Las decisiones son complicadas dado el rápido desarrollo de la tecnología informática . 0.100664322668224 14
5. Un negocio pensado y desarrollado para los amantes de la informática y las nuevas tecnologías . 0.107532139407652  
17
6. La informática al servicio de las comunicaciones , entendida como computadoras y programas que desarrollan tareas de comunicaciones . 0.119107580327359 20
7. También , el ocio es otro ámbito al que se está dirigiendo el actual desarrollo de la tecnología informática .  
0.117648952471021 21
8. Lafont Tecnología y Desarrollo SL es una empresa formada por profesionales de amplia experiencia en diferentes ámbitos de la informática y nuevas tecnologías .0.097984362219099325
9. Este conjunto de información se ha grabado en una base de datos informática desarrollada al efecto , para asegurar su almacenamiento , recuperación y actualización .  
0.0979843622190993 27
10. Me gusta la informática , me inquieta , sobre todo , el desarrollo de las tecnologías de la información y la comunicación y su impacto social . 0.108325788267676 28
11. Este cambio ha sido posible gracias al desarrollo tecnológico que se ha producido desde mediados del siglo pasado en dos ámbitos : la tecnologías de la comunicación y la informática .  
0.107358771248866 32

12. Tanto si quieres ser un científico , o dedicarte a desarrollar las últimas aplicaciones informáticas , o simplemente entender qué significa cuando alguien dice que " el computador ha fallado " , estudiar informática te proporcionar? los conocimientos necesarios . 0.101123420831995 41
13. Capacidad para poder aplicar en las sesiones los conocimientos y habilidades adquiridos en informática , nuevas tecnologías de la información , gestión de recursos ... Adquisición de conocimientos y técnicas informáticas para poder aplicarlos en el desarrollo de los trabajos . 0.102869111493963 42
14. urjc es Introducción En los tiempos actuales hay varios campos y tecnologías que están sufriendo un desarrollo vertiginoso , entre ellos la informática y las comunicaciones ( de aquí en adelante TIC , Tecnologías de la Información y la Comunicación ) . 0.10185601869751 43
15. Capacidad para poder desarrollar en las sesiones de trabajo los conocimientos y habilidades adquiridos en informática , nuevas tecnologías de la información , gestión de recursos ... Conocimientos sobre técnicas de informática que favorezcan la aplicaci?n de las mismas a los ?mbitos de la pedagog?a social . 0.110762911734059 48