

TESIS DOCTORAL

Glosario militar de la Segunda Guerra Mundial.
Trabajo de extracción terminológica y de lingüística de corpus
mediante el análisis de textos de época.

LÍNEA DE INVESTIGACIÓN:

20360016 – Lingüística aplicada

Rodrigo Muñoz Cabrera

Directora: Dra. D^a. María
Gabriela Fernández Díaz

Dpto. de Filología Inglesa (Lengua Inglesa)



AGRADECIMIENTOS

Este trabajo no habría sido posible sin el alentador trabajo de la tutora de esta Tesis, María Gabriela Fernández Díaz. Deseo agradecerle no solo su atenta labor de dirección, sino también su constante ánimo desde que me otorgó la oportunidad de desarrollar este trabajo que en estos momentos es de dominio público. En sus manos recae el entusiasmo que me inspiró y que me hizo continuar, tras finalizar el máster de la que ella fue mi directora del Trabajo Fin de Máster, el desarrollo de mi capacidad investigativa.

Así mismo, deseo mostrar mi agradecimiento al revisor del glosario fruto de esta investigación: el brigada y maestro armero Lanny Hill, del Ejército canadiense. Todas sus ideas, sugerencias y propuestas han enriquecido sobremanera el trabajo, tanto desde el punto terminológico, como desde el aspecto lingüístico.

No debo olvidar a las personas que me ayudaron a la hora de buscar la documentación relevante para la compilación del corpus que sirvió como base de este trabajo. Ellos son Alberto Pérez Garrido y el difunto capitán de navío Francisco Bernal González-Villegas. Su apoyo, desde los campos de la informática y militar, hicieron posible la selección de los textos que sirvieron de base para la fundamentación de este trabajo de investigación.

Finalmente, y no por ello menos importante, agradezco a mi mujer, Cheli, el incondicional y constante apoyo que me ha brindado. Sin su ánimo no hubiera sido posible concluir esta Tesis. No hace falta detallar más mi gratitud pues, en este caso, las palabras no son capaces de describir lo que los hechos demuestran.

Por último, es mi deseo dedicar este trabajo a mis padres, Sasia y Rodrigo, que ya no están aquí, pero que siempre los llevaré conmigo.

ÍNDICE GENERAL

1. Índice de figuras.	9
2. Introducción.	14
3. Objetivos.	16
4. La lingüística de corpus.	18
4.1. Introducción.	18
4.2. Las diferentes generaciones de corpus.	22
4.2.1. Primera generación.	22
4.2.2. Segunda generación.	23
4.2.3. Tercera generación.	23
4.2.4. Panorama actual.	24
4.3. Definición de corpus.	24
4.4. Enfoques mayoritarios.	26
4.4.1. El análisis basado en el corpus (<i>corpus-based</i>).	26
4.4.1.1. Aislamiento.	27
4.4.1.2. Normalización.	27

4.4.1.3. Ejemplificación.	28
4.4.2. El análisis derivado del corpus (<i>corpus-driven</i>).	29
4.4.2.1. Unidad léxica y entorno.	31
4.4.2.2. Prosodia semántica.	32
4.4.2.3. <i>Deslexicalización</i>	32
4.4.2.4. Ideología.	33
4.5. Lingüística de corpus: ¿metodología o disciplina?	33
4.6. La anotación en el corpus.	35
5. La terminología.	38
5.1. Historia y evolución.	38
5.2. Definición y requisitos teóricos.	42
5.3. Una nueva percepción de la terminología.	44
5.4. Terminología: ¿ciencia, arte o práctica?	48
5.4.1. Las nuevas bases teóricas.	49
6. Uso del corpus en el proceso terminológico.	51
6.1. Lenguaje común y lenguaje especializado (LGP y LSP).	51
6.1.1. Diferentes usuarios y niveles de comunicación en los LSP.	52

6.1.2. Creación de un corpus del LSP.	54
6.1.3. Compilación de un corpus del LSP.	58
6.1.4. Proceso de creación de glosarios.	58
6.2. El trabajo terminológico.	60
6.2.1. El método tradicional.	60
6.2.2. El proceso de extracción terminológico actual.	62
6.3. Extracción terminológica monolingüe.	63
6.3.1. Términos monoléxicos.	63
6.3.2. Términos multiléxicos.	64
6.3.2.1. Enfoque lingüístico.	64
6.3.2.2. Enfoque estadístico.	65
6.4. Extracción terminológica bilingüe.	66
7. Compilación de glosario terminológico.	69
7.1. Bases del trabajo terminológico.	70
7.1.1. Competencias.	70
7.1.2. Supuestos básicos de la actividad terminológica.	70
7.1.3. Tipos de trabajo.	71

7.1.4. Tipos de aplicación.	72
7.1.5. Fases.	72
7.1.5.1. Presentación del tema y adquisición del conocimiento.	73
7.1.5.2. Delimitación del trabajo terminológico.	73
7.1.5.3. Selección y ampliación de la información.	78
7.1.5.4. Estructuración del conocimiento.	81
7.1.5.5. Búsqueda y formato de textos.	82
7.1.5.6. Vaciado.	89
7.1.5.7. Elaboración del fichero de vaciado.	90
7.2. Metodología y herramientas empleadas.	92
7.2.1. Extracción terminológica estadística.	93
7.2.1.1. Fase I: Configuración del programa.	93
7.2.1.2. Fase II: Primer análisis.	95
7.2.1.3. Fase III: Resolución de problemas.	103
7.2.1.4. Fase IV: Adición indirecta de candidatos a término.	111
7.2.2. Extracción terminológica lingüística.	113
7.2.2.1. Fase I: Identificación y análisis de candidatos a término.	114

7.2.2.2. Fase II: Identificación y análisis de unidades terminológicas complejas.	118
7.2.2.3. Fase III: Identificación y análisis de unidades terminológicas simples mediante la anotación gramatical.	121
7.2.2.4. Fase IV: Identificación y análisis de unidades terminológicas complejas mediante la anotación gramatical.	124
8. Consideraciones sobre el trabajo de investigación.	132
8.1. Observaciones generales.	132
8.1.1. Evolución de la terminología: apreciaciones genéricas.	136
8.1.2. Evolución de la terminología a nivel morfológico.	138
8.1.3. Evolución de la terminología a nivel semántico.	141
8.1.4. <i>Britanización</i> del lenguaje especializado en el inglés americano.	141
8.2. El trabajo de extracción terminológico: valoración.	143
9. Conclusiones.	156
10. Bibliografía.	164
11. Apéndice – Sistema de etiquetado gramatical UCREL CLAWS 7.	176

1. ÍNDICE DE FIGURAS

- Figura 1.** Primera página de la letra “b”. United States, War Department. *Dictionary of United States Army Terms*. Washington: GPO, 1944. Print. 76
- Figura 2.** Primera página de la letra “b”. United States, Department of Defense. *Department of Defense Dictionary of Military Terms and Associated Terms*. Washington: Joint Publication 1-02, 2010. Print. 77
- Figura 3.** Primera página del DIARIO OFICIAL DEL MINISTERIO DEL EJÉRCITO, de fecha 31, de enero de 1941 en su formato original. 84
- Figura 4.** Encabezamiento y primer párrafo de la primera página del DIARIO OFICIAL DEL MINISTERIO DEL EJÉRCITO, de fecha 31, de enero de 1941, tras su conversión a formato pdf. 85
- Figura 5.** Portada del *The Field Artillery Journal* de diciembre de 1944. 86
- Figura 6.** Vista parcial del primer artículo del *The Field Artillery Journal* de diciembre de 1944 tras su conversión a formato plano. 87
- Figura 7.** Encabezamiento del primer artículo de la publicación *Intelligence Bulletin* de septiembre de 1942, tal como aparece en el sitio *web* LoneSentry. 87
- Figura 8.** Portada del primer número de la publicación *Intelligence Bulletin*. 88
- Figura 9.** Ejemplo de ficha terminológica. 91
- Figura 10.** Ejemplo de ficha terminológica de SDL MultiTerm 2009. 93
- Figura 11.** Configuración del proyecto y activación del archivo de palabras de exclusión y la base de datos terminológica. 94

Figura 12. Selección de parámetros antes de la ejecución del proceso de extracción terminológica.	95
Figura 13. Fichero de vaciado terminológico.	97
Figura 14. Ejemplo de ficha terminológica.	97
Figura 15. Ejemplo de <i>balloon barrage</i>	99
Figura 16. Vista parcial del archivo de exclusión de términos en formato html.	100
Figura 17. Ejemplo de ficha terminológica con todos los campos establecidos.	102
Figura 18. Ejemplo de desambiguación.	108
Figura 19. Vista del fichero de vaciado terminológico tras la inclusión de todos los textos.	109
Figura 20. Vaciado terminológico con el programa SDL MultiTerm 2009 Extract. ...	110
Figura 21. Colocaciones compuestas por <i>follow, following</i> extraídas por el programa SDL MultiTerm 2009 Extract.	111
Figura 22. Vista parcial del archivo <code>excluded_terms.htm</code> y el verbo <i>follow</i> resaltado en color azul.	112
Figura 23. Búsqueda de frases en contexto para el término <i>dugout</i>	113
Figura 24. Visualización de textos anotados gramaticalmente.	115
Figura 25. Listado de palabras en AntConc 3.3.5.w (Windows).	115
Figura 26. Vista de las colocaciones de la unidad terminológica <i>armored</i>	116

Figura 27. Resultado de la búsqueda de <i>motpulk+</i> en AntConc 3.3.5.w (Windows). . .	117
Figura 28. Selección del contexto que será empleado para la definición de la unidad terminológica <i>Motpulk</i>	117
Figura 29. Colocaciones del término <i>gun</i>	119
Figura 30. Búsqueda de concordancias de <i>defense</i>	120
Figura 31. Búsqueda de colocaciones de <i>defense</i> por su margen izquierdo.	120
Figura 32. Resultado de una búsqueda mostrando las anotaciones gramaticales.	121
Figura 33. Búsqueda de sustantivos en AntConc 3.3.5.w (Windows).	122
Figura 34. Resultado de la indagación *_NP*.	123
Figura 35. Búsqueda por medio de la anotación gramatical de relaciones entre adjetivo y sustantivo.	124
Figura 36. Ejemplo de una unidad terminológica compuesta por un adjetivo y un nombre.	125
Figura 37. Diversos resultados de la búsqueda realizada mediante la anotación *_JJ *_NN* *_NN*.	126
Figura 38. Búsqueda de frases nominales conformadas por dos nombres comunes.	126
Figura 39. Vista de algunos resultados de la consulta formulada con la anotación *_NN* *_NN* *_NN*.	127
Figura 40. Resultado de la búsqueda de relaciones entre nombres propios y nombres comunes.	128

Figura 41. Resultados relevantes de la inspección nombre común+ adjetivo + nombre común.	129
Figura 42. Resultados de la búsqueda <i>nombre común, preposición “de”, nombre común.</i>	130
Figura 43. Muestra de los resultados de la búsqueda *_NN* *_IF* *_NN*	130
Figura 44. Inclusión en el proceso de extracción terminológica de unidades léxicas compuestas por la lexía <i>three</i>	144
Figura 45. Lematización incorrecta del verbo <i>punish</i> al incluir su forma nominal <i>punishment</i>	145
Figura 46. Interpretación errónea de un punto al extraer una frase en contexto.	146
Figura 47. Comparación de resultados entre SDL MultiTerm 2009 Extract y AntConc 3.3.5.w (Windows).	147
Figura 48. Interpretación equívoca en el uso de mayúsculas y minúsculas.	148
Figura 49. Vista del porcentaje (<i>Score</i>) que otorga el programa SDL MultiTerm 2009 Extract a los candidatos a término.	150
Figura 50. Resultado del recuento de palabras con el programa AntConc 3.3.5.w (Windows).	153
Figura 51. Recuento del archivo Test.txt con AntConc 3.3.5.w (Windows).	154
Figura 52. Resultado de la contabilización de voces del texto Test.txt con SDL MultiTerm 2009 Extract.	154
Figura 53. Relación genérica de los tipos de ametralladora según su peso.	160

Figura 54. Relación genérica de los tipos de granada según su medio de lanzamiento. .161

Figura 55. Estructuración conceptual de las granadas de mano atendiendo al tipo de carga explosiva. 161

Figura 56. Subdivisión de los tipos de granada de mano atendiendo a su forma.162

2. INTRODUCCIÓN

La Segunda Guerra Mundial tuvo como resultado una sorprendente evolución en lo que se refiere a tácticas, armamento y adoctrinamiento sin precedentes en el campo bélico. En consecuencia, la aparición de nuevas unidades conceptuales conllevó la búsqueda de unidades terminológicas para su representación. Ese número cada vez mayor de tecnicismos forzó la necesidad de acometer un trabajo terminológico inédito hasta la fecha con el propósito de difundir dicha información y de propiciar un correcto uso de los nuevos vocablos de especialidad.

Por todo ello, no resulta extraño que el Ejército norteamericano publicara un glosario en 1943 con el objetivo de dar a conocer (entre otras cosas) los sinónimos que se empleaban en el Ejército británico para homogeneizar la terminología empleada en el seno de la disciplina. Así mismo, se trataba también (a nuestro entender) de una medida de divulgación de la terminología castrense norteamericana, con la intención de preservar la univocidad de ésta y evitar el empleo de los sinónimos procedentes del Reino Unido.

Además de los capítulos dedicados a la introducción y a la exposición de nuestros objetivos, hemos delimitado el trabajo en seis grandes capítulos: la lingüística de corpus, la terminología, uso de corpus en el proceso terminológico, compilación de glosario terminológico, consideraciones sobre el trabajo de investigación y conclusiones.

En el capítulo número cuatro, dedicado a la lingüística de corpus, procederemos a su descripción, su historia y los enfoques mayoritarios que dominan dicha metodología, esto último unido a la diatriba existente entre los que la consideran una disciplina y los que la conciben como una útil herramienta.

El quinto capítulo, consagrado a la terminología, detalla someramente su historia y evolución, al igual que la discusión actual entre los que la valoran como un arte frente a los que la catalogan como una práctica.

El sexto episodio detalla el papel que desempeña el corpus dentro del ámbito terminológico y su versatilidad a la hora de identificar unidades terminológicas y para la compilación de glosarios. De igual modo, enumeramos los diferentes procesos de extracción terminológica al igual que sus diferentes enfoques.

El séptimo capítulo se encuentra dividido en dos apartados:

a) En el primero llevamos a cabo un estudio más detallado de nuestra investigación, que es la compilación de un glosario terminológico. En este punto destacamos los variados tipos de trabajo que se pueden acometer con el propósito de extraer terminología, sus aplicaciones, fases y destinatarios. De igual modo, describimos las diferentes formas de estructurar el conocimiento en cualquier disciplina, la manera de seleccionar los textos que conformarán nuestro corpus, al igual que la elaboración de un fichero de vaciado, y las fichas terminológicas.

b) El segundo apartado se centra en enumerar las consideraciones a tener presente durante la valoración de un trabajo terminológico. En dicha sección se pormenoriza sobre la evolución de la terminología, tanto a nivel general, como morfológico y semántico. Seguidamente, realizamos un juicio de valor de las herramientas empleadas para la extracción terminológica (SDL Multiterm 2009 Extract y AntConc 3.3.5.w -Windows-), al igual que una breve reflexión sobre la estructuración del conocimiento llevado a cabo por los redactores de los glosarios tomados como referencia, sus deficiencias y sus posibles mejoras. Posteriormente procederemos a detallar los hallazgos obtenidos, tanto a nivel de extracción estadística como lingüística, al igual que los fenómenos más característicos que hemos detectado.

A continuación, tanto en el capítulo dedicado a las consideraciones sobre el trabajo de investigación como en el de las conclusiones, evaluaremos nuestro análisis y la evolución que ha sufrido la terminología en esta disciplina durante dicho espacio de tiempo. Finalmente, expondremos las implementaciones que, a nuestro juicio, son necesarias para mejorar las herramientas a disposición de los terminólogos a la hora de desarrollar su empeño, las pautas que estimamos pueden facilitar dicha labor y el deseo de incrementar nuestro corpus con el objetivo de ampliar sobremanera el estudio terminológico militar sobre la Segunda Guerra Mundial.

3. OBJETIVOS

El propósito de esta tesis es la confección de un glosario terminológico en lengua inglesa que incluya los términos militares, históricos y afines empleados durante la Segunda Guerra Mundial. La finalidad de nuestra labor es la de cubrir un vacío descriptivo del uso real de la terminología militar en la franja temporal que nos ocupa. En nuestra fase de recopilación de información, solo hemos detectado dos glosarios dignos de relevancia, datados en 1943 y 1944, y ese es nuestro principal propósito: la compilación de un diccionario técnico que, no solo incluya la correspondiente definición de cada unidad terminológica, sino que aporte ejemplos en contexto, información léxica y adicional que ayude a una mayor comprensión de la unidad conceptual que describe cada unidad terminológica.

El corpus compilado para tal efecto ha quedado compuesto por 285.214 palabras en lengua inglesa, del que hemos podido evidenciar 3.090 términos, en donde quedan claramente reflejadas sus dos variedades más representativas: el inglés británico y el norteamericano. Dicho corpus está conformado únicamente por textos escritos. El motivo principal de tal decisión radica en la limitada fiabilidad que poseen las grabaciones orales encontradas y el laborioso trabajo que implica su correspondiente transcripción. Nuestra investigación quedó así mismo delimitada al estudio del teatro de operaciones europeo, del Mediterráneo y del norte de África y a las operaciones terrestres. No obstante, el empleo de diferentes cuerpos y ejércitos en ciertos tipos de ofensivas (como pueden ser las operaciones anfibia y las aerotransportadas) nos ha forzado a incluir un cierto número de términos procedentes de la Marina y del Ejército del aire al servir de apoyo a las unidades de tierra.

De igual modo, hemos incluido palabras y términos de índole político e histórico empleados con asiduidad en el campo militar, incorporando los nombres o denominaciones de armas utilizadas profusamente en la contienda, al igual que voces procedentes de otras lenguas de uso común en el seno de las tropas angloamericanas. No obstante, dichas unidades terminológicas debían albergarse en nuestro corpus quedando,

de esa manera, plasmado su uso específico en el lenguaje de especialidad objeto de estudio.

Nuestra labor se ha enfocado primordialmente en la variante norteamericana de la lengua anglosajona (la mayor y casi única fuente de información en la *web*) incluyendo, en la medida de lo posible, las equivalencias empleadas en el Reino Unido.

El trabajo de investigación se ha centrado en dos grandes objetivos:

a) Por un lado, y con el propósito de saber si se tratan de sinónimos parciales o si han sufrido una variación desde el punto de vista conceptual, nuestro interés ha sido el de corroborar el uso real de las unidades terminológicas, el número de ocasiones que han sido empleadas, su análisis dentro de un discurso de especialidad, su adecuación o su modificación y la búsqueda y estudio de sinónimos.

b) Por otro lado, dicha tarea quedaría incompleta si no añadiéramos términos que no han sido registrados en los glosarios empleados como referencia. En consecuencia, el trabajo terminológico debe centrarse en la aparición, uso y evolución de las unidades terminológicas y ello implica también una revisión, *a posteriori*, del espacio temporal en el que hemos centrado nuestro trabajo. La aparición de nuevas unidades conceptuológicas a las que atribuirle nuevas unidades terminológicas (u otras ya existentes) es un factor de vital importancia que estimamos que debía ser tomado en consideración como uno de las metas principales de nuestro estudio.

Por último, cabe añadir que este trabajo queda englobado en el ámbito de la lingüística aplicada, al emplear en su configuración dos de sus campos metodológicos: la terminología y la lingüística de corpus. De igual modo, esta tesis ha sido dividida en dos partes: una parte teórica (que es en la que nos hallamos) y una parte práctica (el glosario terminológico) en donde quedan manifestadas las hipótesis sobre las que hemos fundamentado nuestra investigación.

4. LA LINGÜÍSTICA DE CORPUS

En este capítulo describimos una de las dos metodologías empleadas en nuestro trabajo. Comenzaremos resumiendo brevemente la historia y desarrollo de la lingüística de corpus. Al mismo tiempo, destacaremos el papel que ha tenido (y que posee en la actualidad) la informática en su auge. Seguidamente, procederemos a detallar su definición y el motivo por el que la lingüística de corpus se considera disciplina o método, dependiendo de la importancia que se le atribuya a los resultados emanados del corpus. Finalmente, realizaremos una somera explicación de la anotación, la cual será descrita con mayor extensión en el capítulo dedicado a la compilación del glosario terminológico.

4.1. INTRODUCCIÓN

Aunque la lingüística de corpus como tal es relativamente reciente (tanto si es considerada como metodología o como disciplina), su práctica es tan antigua como la investigación lingüística. La lingüística histórica siempre se ha fundamentado en el corpus. Según Johansson (22), la prueba más relevante del cambio y la evolución del lenguaje se aprecia en las colecciones textuales realizadas en diferentes lugares y períodos de tiempo. Hace un siglo, y tomando como ejemplo el trabajo de Bréal, el estudio del lenguaje se limitaba a la observación de los datos; las leyes que gobernaban el desarrollo histórico del significado sólo se podían descubrir mediante el análisis de los fenómenos (McEnery y Wilson 1).

No obstante, y en un breve espacio de tiempo, el método de estudio se enfocó hacia la intuición e introspección (Biber y Finegan 205). Tras el histórico periodo evolucionista, la lingüística moderna cambia su concepción gracias a la investigación de Ferdinand de Saussure. A partir de entonces, el objeto de la lingüística se define mediante el sistema, abstracto por excelencia, y no por medio de símbolos; las reglas ya no estaban directamente relacionadas con fenómenos específicos.

Se denomina etapa inicial de la lingüística de corpus al periodo transcurrido antes de la divulgación de las ideas de Noam Chomsky (McEnery y Hardie 25). Aunque en esa época el enfoque metodológico con mayor vigor era la observación del uso del lenguaje, la lingüística de corpus era también empleada por numerosos lingüistas tales como Boas (1940). La mayoría de los que se inclinaban por el enfoque estructuralista aplicaban una metodología que se puede definir como basada en el corpus. No obstante, durante este espacio temporal el término *lingüística de corpus* no aparece ni en los textos ni en las investigaciones realizadas por aquel entonces.

La descripción basada en el corpus se empleó en los siglos XIX y XX, en particular para analizar el proceso de la adquisición del lenguaje. Según McEnery y Wilson (3), los estudios realizados sobre el lenguaje empleado por los niños durante el periodo de adquisición innata del lenguaje entre 1876 y 1926 estaban conformados por locuciones de infantes registradas por sus padres. Estos primitivos corpus (tales como los que compilaron Prever en 1889 y Stern en 1924) todavía se usan como fuente de datos normativos en estudios de adquisición del lenguaje. Por otro lado, el corpus también se ha empleado en el análisis de la ortografía y la redacción. En 1897, Käding recopiló un corpus en lengua germana de once millones de palabras para analizar y cotejar las distribuciones de frecuencia y las secuencias en cartas redactadas en alemán. En lo que respecta al aspecto pedagógico del lenguaje, Fries (1940) y Bongers (1947) son algunos de los investigadores que emplearon corpus para profundizar en la enseñanza de lenguas extranjeras (McEnery y Wilson 4).

En el campo de la lingüística comparada, cabe destacar el sofisticado trabajo de Eaton acometido en 1940, en el que contrastaba la frecuencia de significado de las palabras en neerlandés, francés, alemán e italiano. Sus investigaciones sirvieron como base para otros estudios realizados por lingüistas como Lorge (1949), cuyo objetivo fue el de realizar trabajos de descripción monolingüe. Así mismo, Fries fue uno de los pioneros a la hora de acometer un análisis descriptivo del inglés basado en el corpus en 1952 (McEnery y Wilson 4).

Por todo lo descrito hasta ahora, queda constatado que la metodología basada en el corpus fue práctica común durante muchos años en el campo de la lingüística. No obstante, los trabajos de Noam Chomsky provocaron que esta práctica cayera en desuso a

partir de finales de los años cincuenta. Chomsky promulgaba un distanciamiento del empirismo a favor del racionalismo; en vez de explicar el lenguaje mediante su observación, éste se debía justificar mediante la introspección. Hubo otros investigadores que abogaban por otra vía de acción, como Hockett (*Models* 234), que promulgaba validar los datos lingüísticos procedentes de los corpus con otros que habían sido analizados por medio de otros métodos con el propósito de demostrar su poder de predicción. No obstante, el rechazo de Hockett al método de la introspección le llevó a afirmar que si los datos procedentes del uso de dicho método coincidían con los resultantes de un corpus eran superfluos, y si eran divergentes, resultaban ofensivos (*State* 196).

El reconocimiento de que el número de oraciones que se pueden formular en cualquier lenguaje natural es infinito, gracias a su capacidad recursiva, invalidaba a la lingüística de corpus como la única fuente para explicar el lenguaje. Además, el corpus no puede describir la competencia de los hablantes, por lo que los datos que proporcionen mostrarán una imagen distorsionada de la realidad. Por otro lado, “el hecho de que las oraciones que componen un corpus sean construcciones repetitivas (por su frecuencia o simplemente fruto del azar) hace que la información que se extraiga de éste no tenga más valor que la de un simple listado” (Chomsky, *Independence* 159). Otro de los argumentos en contra era el porqué buscar en un corpus de billones de palabras hechos que podían fácilmente descubrirse mediante la introspección; “solo hace falta sentarse y pensar unos minutos para obtener ejemplos del todo relevantes” (Chomsky, *Modular Approaches* 44).

Existían otros investigadores contrarios al uso del corpus puntualizando otro tipo de inconvenientes. Abercrombie (115) acuñó el concepto de *seudométodo* en 1965 para describir la mayoría de los estudios basados en el corpus de aquella época. Abercrombie afirmaba que el estudio de un elevado volumen de datos por el hombre era costoso, lento e ineficaz. En consecuencia, en la actualidad resulta lógico considerar injustificado el desorbitado número de analistas empleados para procesar decenas de millones de palabras en los trabajos de West y de Käding realizados en 1953 y 1987 respectivamente (McEnery y Wilson 12). No obstante, debemos argumentar que tal reticencia se puede interpretar como una crítica a la lingüística de corpus previa a su digitalización, en donde su razonamiento es del todo lícito debido a la metodología empleada y a los ineficientes recursos a disposición del lingüista.

Afortunadamente, tales críticas no frenaron el estudio y la investigación basada en el corpus. Incluso Chomsky tuvo que reconocer que su rechazo a los datos resultantes del habla o actividad lingüística como fuente de evidencias era erróneo en lo que concierne a los estudios de adquisición del lenguaje (McEnery y Wilson 13). Debido a que ciertos temas no podían estudiarse mediante la introspección, el corpus siguió empleándose como metodología, aunque minoritaria, durante los años sesenta y setenta. Una de las ventajas que poseen los resultados procedentes del corpus es que son, intrínsecamente, más verificables que los juicios basados en la introspección. Otro aspecto a su favor reside en que el corpus es una fuente fiable de datos basados en la frecuencia y proporciona, de ese modo, un enfoque mucho más sistemático al análisis lingüístico, estando sujeto a una verificación objetiva de sus resultados (Leech, *Corpora* 105). En definitiva, algunas de las críticas de Chomsky eran válidas y tuvieron como respuesta redimensionar el enfoque actual de la lingüística de corpus, que ya ha dejado de ser un *seudométodo* gracias a la informática, convirtiéndose en una metodología del todo viable (McEnery y Wilson 17).

Al contrario que Chomsky, Leech (*Corpora* 107) abogaba por el uso del corpus resaltando sus características más destacadas: su enfoque en la actividad lingüística y no en la competencia; su interés en la descripción lingüística, más que en los principios universales lingüísticos; su orientación hacia los modelos cualitativos y cuantitativos del lenguaje; y su búsqueda de una visión más empírica, más que racional, de la investigación científica. La lingüística de corpus (o lingüística computerizada de corpus) comienza a tener auge a finales de la década de los cincuenta y principios de los sesenta. Gracias a dicho resurgimiento, esta metodología ha pasado de ser un enfoque marginado del (en su mayor medida) estudio de las humanidades bajo soporte informático, la *mecanolingüística*, y la lingüística inglesa y de los seguidores de J.R. Firth, a aplicarse a otras lenguas naturales y a las lenguas de especialidad.

Firth (*Papers* 32) sostenía que el contexto y propósito sociales de la comunicación son factores primordiales a tener en cuenta a la hora de estudiar el lenguaje. Su respaldo a la lingüística de corpus se basa en su afirmación de que el lenguaje autenticado, debidamente registrado, debe ser el foco de atención del lingüista. Por otro lado, fue él quien acuñó el término *colocación* (que describiremos con posterioridad) aunque fuera Sinclair el que propagara su uso en el ámbito de la lingüística de corpus. Su metodología

será desarrollada por otros, como por el mismo Sinclair, al igual que por Halliday y Hoey (McEnery y Wilson 12).

4.2. LAS DIFERENTES GENERACIONES DE CORPUS

4.2.1. Primera generación

La primera generación de corpus informatizado comienza en 1961, momento en el que Nelson Francis y Henry Kucera encabezaron un grupo de investigadores en la Universidad de Brown, y definieron los principios para la compilación de un corpus informatizado, el cual fue bautizado inicialmente como “una muestra de referencia del inglés americano a nivel escrito empleado en la actualidad para su uso en formato digital” (Francis y Kucera 1964). Su resultado será la creación del actual Brown Corpus en 1964. Dicho repositorio textual albergaba aproximadamente un millón de palabras extraídas de quinientos textos (de unas dos mil palabras cada uno) publicados en EE.UU. durante 1961.

Poco tiempo después, John Sinclair creó el primer corpus informático textual (inglés británico) en la Universidad de Edimburgo (Sinclair et al. 2004). En 1970 se inicia un trabajo similar al realizado por Francis y Kucera en la Universidad de Lancaster pero, en este caso, tomando como objeto de estudio el inglés británico. Para la compilación del que será conocido como el Lancaster-Olds-Bergen Corpus (cuya fase final se llevará a cabo en Noruega y que se prolongará hasta 1978), se empleó una metodología muy similar a la usada en el Brown Corpus para que pudieran realizarse estudios comparativos entre éstos.

Por otro lado, el primer corpus electrónico oral en lengua inglesa quedará concluido entre 1974 y 1975 gracias a la labor de un grupo de investigadores encabezados por Jan Svartvik en la Universidad de Lund. Su nombre será el de London-Lund Corpus y estará compuesto por quinientas mil palabras grabadas en cintas de audio y transcritas para su correspondiente anotación gramatical.

Esta primera generación de corpus tiene como peculiaridad su pequeño tamaño en lo que a datos se refiere, y que parte de su trabajo tuvo que realizarse a mano.

La informática se convertirá en el pilar de la lingüística de corpus inglesa a partir de la década de los setenta, hasta tal punto que Svartvik digitalizará el Survey of English Usage (SEU). El trabajo de Svartvik servirá, al mismo tiempo, como base de formación de investigadores especializados en el uso del corpus para llevar a cabo análisis gramaticales en dicho idioma. Esta generación también se caracteriza por el uso de ordenadores centrales (*mainframe*) y su restricción a la lengua inglesa (McEnery y Hardie 72).

4.2.2. Segunda generación

La segunda generación de corpus se inicia a partir de 1975 con la aparición del Cobuild Corpus. Dicho corpus albergaba alrededor de veinte millones de palabras y su trabajo de compilación se ejecutó en la Universidad de Birmingham. Una vez concluido, se empleó para la confección de diccionarios, siendo el más destacado el Collins Cobuild (1987).

Es en esta fase donde ya es posible emplear ordenadores personales y, por tanto, el investigador tiene la capacidad de realizar estudios en corpus de pequeñas dimensiones (Anthony 152). Por otro lado, otro efecto que produjo la proliferación del uso de la lingüística de corpus fue la aparición de centros de estudio de tal metodología por toda Europa (Bergen, Oslo, Berlín, Nimega y Potsdam, entre otros).

4.2.3. Tercera generación

La tercera generación se inicia en la década de los noventa y destaca por la amplia capacidad de almacenamiento, la recopilación, no solo a nivel escrito, sino a nivel oral y transcrito, y por su accesibilidad en *Internet*. La particularidad de esta fase reside en que ya es posible analizar otras lenguas (además de la inglesa), la aparición de herramientas que permiten realizar múltiples funciones, y el empleo de interfaces de fácil uso, incluso para los lingüistas con limitadas nociones de informática. Los corpus más conocidos son el British National Corpus (cien millones de palabras), el Bank of English (más de mil quinientos millones de palabras) y el OUP Corpus (mil millones de palabras). En lo que se refiere al español destacamos los corpus constituidos bajo el auspicio de la Real Academia

Española: CORDE, CREA (2008) y CORPES XXI (2013). De igual modo, es digno de mención el Corpus del español, creado en 2002 por Mark Davies (Brigham Young University).

4.2.4. Panorama actual

La cuarta generación se encuentra en pleno proceso de desarrollo. La dificultad de compilar y emplear corpus de más de cien millones de palabras conlleva, por un lado, la imposibilidad de realizar indagaciones en un ordenador personal y, por otro lado, la ineficacia de las herramientas empleadas en la tercera generación para el procesamiento de una información tan vasta. Ello ha traído como respuesta la creación de una nueva tecnología alojada en la *web*, que permite el almacenamiento de nuestro corpus objeto de estudio en *Internet*, o elegir una serie de datos almacenados en uno o varios corpus, seleccionando los criterios de búsqueda que más se adecúen a nuestro propósito (Anthony 153). Corpus.byu.edu (Davies 2013), CQPweb (Hardie 2013) y SketchEngine (Kilgariff 2013) son ejemplos de la nueva generación de herramientas que permiten el estudio de corpus de grandes dimensiones.

4.3. DEFINICIÓN DE CORPUS

El auge que ha experimentado la lingüística de corpus ha conllevado la reformulación de los criterios fundamentales por los que se guían los compiladores de corpus a la hora de realizar su trabajo. Ello se debe a que se cuestiona con más frecuencia qué se considera un corpus o cuáles son los requisitos indispensables para su creación (Bowker y Pearson 45).

Podemos definir *corpus* como una recopilación de textos (escritos, orales o transcritos) cuya cantidad y contenido se consideran *representativos* de una determinada lengua con el objeto de realizar diversos análisis lingüísticos (Tognini-Bonelli 53). Otra definición podría ser la del estudio del lenguaje tomando como base ejemplos de su uso en la *vida real* (McEnery y Wilson 1). Leech añade a esta definición que los datos lingüísticos que empleemos deben ser *auténticos* (*naturally-occurring*) para servir como base para una investigación lingüística. Además, Leech argumenta que, desde la década de los sesenta,

el término corpus se emplea para definir un cuerpo de material lingüístico en formato electrónico, capaz de ser procesado por un ordenador para un determinado propósito (*Corpora* 106). En lo que se refiere a los diferentes parámetros y criterios a la hora de elegir los textos, ellos dependerán de muchos factores, entre los que destacamos la tipología y el estudio que se desee acometer. No obstante, podemos afirmar que, ante todo, un corpus debe estar compuesto por *lenguaje natural y auténtico* (McEnery y Hardie 2).

Analicemos de manera sucinta las tres bases fundamentales de dicha definición: *autenticidad, representatividad y técnicas de selección*. Resulta evidente que el punto de partida de cualquier investigación lingüística realizada dentro del ámbito de la lingüística de corpus es el lenguaje y que se supone que todo el material que conforma un corpus ha sido extraído de comunicaciones reales entre personas que trataban de asuntos de su competencia (Aarts, *Intuition-based* 45; Johansson 19). Si damos como cierto que el corpus que hemos creado es una muestra representativa del lenguaje objeto de estudio, ello implica que lo consideremos como *auténtico* (Tognini-Bonelli 2). Por otro lado, y según Leech (*Corpora* 105), se juzga que un corpus es *representativo* cuando los resultados que se obtienen pueden generalizarse o contrastarse ante un corpus de mayores dimensiones. De igual modo, la mayoría de los usos lingüísticos que proporcionan los corpus se basan en la fiabilidad de éstos como elementos representativos del lenguaje (Biber 377). En suma, los resultados procedentes de un corpus son aceptados como una representación de las propiedades (emisión y recepción) de la percepción del lenguaje humano.

Por el contrario, las *técnicas de selección* tienen más relación con un enfoque teórico de la naturaleza del lenguaje que con el almacenamiento electrónico de datos. Por ejemplo, Stubbs (*British Traditions* 11) señala que, tradicionalmente, en el análisis de textos en lengua inglesa la unidad de estudio ha sido la de textos completos. Un corpus que contiene textos completos proporciona la posibilidad de elaborar una amplia gama de estudios lingüísticos comparado con otro que contenga muestras más pequeñas. Si tomamos como ejemplo el Brown Corpus, cuyos textos están limitados a dos mil palabras, cualquier estudio que busque patrones que solo aparecen en textos más extensos resulta del todo inadecuado (Sinclair, *Collocation* 23).

4.4. ENFOQUES MAYORITARIOS

4.4.1. El análisis basado en el corpus (*corpus-based*)

Partiendo de la premisa de que un corpus puede emplearse de diversas maneras con el objetivo de validar, ejemplificar o fundamentar una teoría lingüística, el término *basado en el corpus* se emplea para aludir a una metodología que hace uso del corpus para exponer, probar o ejemplificar teorías que ya habían sido formuladas antes de que los grandes corpus estuvieran a disposición de los investigadores (Tognini-Bonelli 65). Los datos emanados del corpus se consideran como un valor añadido, más que un factor determinante, ya que el análisis se ejecuta mediante unos parámetros ya preestablecidos. No obstante, los resultados obtenidos pueden implicar que se deba reformular una teoría ya existente o que aparezca una nueva teoría (Römer 8).

En el ámbito de la enseñanza de lenguas, Berry habla del problema de la *dirección única*, que considera como uno de los siete pecados capitales de la gramática pedagógica. Berry afirma que la dirección en sentido único es “un problema en la relación entre la descripción del lenguaje y la pedagogía del lenguaje”, que “aparece cuando los estudios descriptivos tienen como única motivación la teoría del lenguaje, el deseo de validar un aspecto de una teoría en particular” (39). El problema aparece cuando los datos no concuerdan con la teoría. Esta postura de dar prioridad a las postulaciones teóricas ya existentes está íntimamente relacionada con la metodología deductiva empleada en el aprendizaje gramatical y en la lingüística computacional, donde el modelo proporciona las etiquetas para el análisis de las diferentes categorías sintácticas o gramaticales.

Por otro lado, otro problema que dificulta la exposición de los datos en categorías claramente diferenciadas radica en la inherente naturaleza cambiante del lenguaje producido de manera espontánea. El enfoque basado en el corpus intenta aislar dicha variabilidad, reducirla y, al mismo tiempo normalizarla. Por tanto, podemos puntualizar que cuando los datos no concuerdan con una teoría en particular se debe a dicha naturaleza cambiante, a la ambigüedad aparente y a la inestabilidad superficial del lenguaje en general, y a las evidencias procedentes del corpus en particular (Aarts, *Comments* 181).

Existen tres métodos para afrontar la falta de concordancia entre datos y postulaciones teóricas: el aislamiento, la normalización y la ejemplificación.

4.4.1.1. Aislamiento

Los datos quedan relegados a un segundo plano con respecto al postulado teórico. Ello ocurre cuando se plantea una dicotomía (como puede ser la de competencia y habla) y se decide que la comprensión derivada del corpus solo se relaciona con una de ellas. Aarts (*Intuition-based* 47) contrasta las gramáticas basadas en la intuición con las gramáticas basadas en la observación, dándoles prioridad a las gramáticas intuitivas. Seguidamente, se emplea el corpus a modo de banco de pruebas para la gramática formal y el modelo resultante de dicha confrontación es que la gramática se convierte en una gramática basada en la observación y, en consecuencia, explica los resultados del uso del lenguaje. Por tanto, ambos tipos de gramática son necesarias para obtener resultados positivos: la gramática basada en la intuición refleja la competencia y las modificaciones aparecidas a resultas de confrontarla ante el corpus, que refleja el habla, la convierten en una gramática basada en la observación.

Se ha aducido que mientras que en la lingüística teórica los datos perceptibles han sido apartados de manera radical de la investigación científica, la lingüística computacional ha proporcionado el uso de un modelo teórico sin excluir necesariamente la aportación de la estadística (Tognini-Bonelli 70). Sin embargo, se prioriza el modelo teórico inicial mediante el marcado y la anotación, cuyas etiquetas se extraen del modelo original (Bod 32). Todo ello se puede entender como una manera diferente de aislar la teoría de los resultados obtenidos de los corpus.

4.4.1.2. Normalización

Se trata de un intento de reconciliar las postulaciones de una gramática intuitiva con las evidencias del corpus. El propósito es reducir la variabilidad inherente de los datos mediante la anotación del corpus. Para Leech (*Corpora* 111), lo primordial es concebir una teoría y después contrastarla; de esa manera, el lingüista emplea su intuición, su

experiencia para experimentar y, finalmente, llega a un supuesto teórico al que le aporta una información cuantitativa extraída del corpus.

Para ello, Leech distingue entre el corpus de trabajo o *training corpus*, que ha sido anotado cuidadosamente a mano siguiendo las categorías gramaticales de la lengua objeto de estudio, y un corpus de prueba o *test corpus*, que será empleado en una segunda fase. La valoración del uso de la lengua sobre el corpus de prueba se realiza con referencia a un modelo gramatical ajeno a la gramática empírica y sujeto al proceso de anotación. Cada vez que exista un desacuerdo, éste se dilucidará por medio de las evidencias extraídas del corpus. El resultado de este proceso es un corpus anotado, *enriquecido* (Aarts, *Comments* 180) con la información adicional que le ha sido aportada. Ya no se trata de simples datos, sino de una nueva versión en la que se proporciona una abundante información lingüística (Leech, *State of the Art* 19).

4.4.1.3. Ejemplificación

Para Halliday (65), el supuesto básico que se esconde tras el trabajo con corpus radica en la equiparación entre la frecuencia en la que aparecen los ejemplos y la cuestión de la probabilidad en la gramática paradigmática, hecho que nos obliga a centrarnos en el grado de representatividad que se le otorga a un corpus. La frecuencia relativa de aparición de ejemplos que se observa en un corpus puede equipararse a la ejemplificación de la probabilidad en la gramática (Halliday y James, *Quantitative Study* 66). Por medio de la búsqueda en el corpus, los términos del sistema son validados y cuantificados a modo de elección en un momento determinado, desde un punto de vista paradigmático, pero también sintagmático. Ambos factores están relacionados con el léxico y la gramática que, para Halliday, se trata del mismo fenómeno observado desde distintos ángulos. La diferencia primordial estriba en que cuando se emplea el enfoque gramatical, los datos extraídos por un lexicólogo son fáciles de discernir, y los extraídos por un gramático son demasiado abstractos para su empleo dentro de un corpus. Los patrones gramaticales son más abstractos que los léxicos y, por tanto, su recuperación o extracción de un corpus resulta complicada a no ser que éste haya sido debidamente anotado. Según Firth, cuando una palabra se encuentra en una *colocación típica* “se muestra tal como es” (Palmer 182) y los patrones léxicos se perciben claramente con tan solo visualizar un listado de

concordancias. No obstante, Firth afirma que ambos aspectos están estrechamente relacionados y que la *coligación* solo es apreciable estudiando las pautas de colocación, consideración sobre la que Halliday discrepa.

4.4.2. El análisis derivado del corpus (*corpus-driven*)

Desde este planteamiento el compromiso del lingüista es preservar la identidad de los datos y sus descripciones tienden a estar en consonancia con las evidencias emanadas del corpus. El corpus se contempla como un repositorio de ejemplos que corroboran teorías ya existentes o como una dimensión probabilística de un sistema bien definido (McEnery y Hardie 151). Los ejemplos se toman sin el propósito de encajar con una de las categorías predefinidas por el lingüista. Se espera que los modelos recurrentes y las distribuciones de frecuencia conformen la evidencia básica para las categorías lingüísticas, considerándose que la ausencia de un modelo sea potencialmente de interés. La evidencias resultantes pueden ser rechazadas (detallando los motivos de tal decisión) o respetadas, pero nunca ignoradas (Tognini-Bonelli 84).

Se trata de “aceptar las evidencias y apreciar qué postulaciones teóricas y descriptivas reflejan la evidencia” (Sinclair, *Collocation* 4). Por otro lado, no se trata de una metodología mecánica, sino arbitrada constantemente por el lingüista, que aporta su conocimiento y su experiencia durante todo el proceso. Treinta años atrás no existía conciencia sobre la necesidad de idear un enfoque derivado del corpus; fue un proceso gradual en el que los investigadores se dieron cuenta de cómo las descripciones tradicionales del lenguaje anteriores al empleo generalizado del corpus estaban siendo cuestionadas de manera implícita a resultas de las evidencias procedentes de los corpus de grandes dimensiones. Podemos considerar que el proyecto COBUILD fue el primer estudio de lexicografía realizado mediante este planteamiento en donde, entre otras cosas, se refleja la postura de Sinclair sobre la teoría del lenguaje y la metodología descriptiva (Tognini-Bonelli 85).

El concepto de *representatividad* de un corpus adquiere una gran significación, debido a que es posible cometer una insalvable equivocación si el corpus no cumple con tal requisito. Las evidencias de un corpus se identifican con la descripción y la teoría de la

misma forma que cualquier otra evidencia científica que se rijan mediante procedimientos de muestreo. Es por ello que se estima que *la teoría contextual del significado* de Firth razona adecuadamente los supuestos en los que se basa el análisis derivado del corpus. Dicho lingüista concebía el lenguaje como el vector de “la continuidad de las repeticiones en el proceso social” (*Papers* 183), supuesto que encaja adecuadamente en una metodología en la que la frecuencia de los casos se hace tangible y cuantificable mediante el uso de la informática. Se trata de diferenciar lo impersonal de lo personal, que es lo que ocurre cuando se aíslan los patrones recurrentes que aparecen en el eje vertical de concordancias, extrayéndose de ellos conclusiones sobre su significado y comprensión del lenguaje. En el marco *firthiano*, lo típico no puede separarse del uso frecuente, y los *hechos recurrentes* son la principal prueba de lo que hacen las personas, de cómo funciona el lenguaje y de lo que trata (Tognini-Bonelli 84).

Con referencia a los hechos recursivos más proclives a aparecer en una concordancia, el primero de ellos es la *colocación*, que Clear define como “la *coocurrencia* recurrente de las palabras” (Firth, *Papers* 277). Las colocaciones se caracterizan por términos de frecuencia y por su posición, así como por su variación compositiva o idiomática. Otro tipo de pauta es el de la *coligación*, en donde el nodo de la palabra se integra dentro de su clase gramatical. Dicho de otro modo: una palabra no solo se coloca con unidades léxicas con clara significación, sino también con marcadores gramaticales, al igual que con diferentes categorías gramaticales (McEnery y Hardie 130). El concepto de coligación también abarca la preferencia sintáctica, en la que una palabra, vista como unidad léxica única, se funde con su entorno. A diferencia de la colocación, fácilmente perceptible en el eje vertical de una concordancia, la coligación resulta más abstracta y sólo se muestra con claridad si el corpus se encuentra anotado con su correspondiente información gramatical.

La colocación y la coligación forman la base para formalizar los hechos recurrentes y solo una exposición descriptiva que muestre sus interconexiones mostrará su dimensión semántica. El término *patrón* se convierte en un punto de unión entre léxico y gramática, de especial importancia en la lingüística derivada del corpus. Hunston y Francis (37) definen los patrones de una palabra como “todas las palabras y estructuras asociadas con ésta y que contribuyen a su significación”.

En lo que se refiere a la metodología empleada por los lingüistas derivados del corpus, Mindt (194) alega que es necesario tener en consideración las evidencias a nivel sintáctico para realizar las diferenciaciones semánticas y afirma que los métodos tradicionales han quedado obsoletos a la hora de analizar los datos lingüísticos cuya cifra significativa de datos se debe a la capacidad que poseen los corpus de acaparar una ingente cantidad de textos. A consecuencia de ello, este investigador constata que las diferenciaciones semánticas (al igual que sus descripciones) se realizan de manera intuitiva.

El lingüista llega a formular generalizaciones (por medio de los datos que se encuentran a su disposición) gracias a la inducción y no mediante la búsqueda de ejemplos que confirmen las hipótesis basándose en un reducido número de datos. Así mismo, el tipo de conclusiones procedentes de tales datos hace casi siempre necesaria la redefinición de las categorías lingüísticas, y conlleva reagrupar o reclasificar los datos. Las postulaciones de Mindt van más allá de lo que Leech definió como “el paradigma empírico de la lingüística de corpus” (*State of the Art* 29), puesto que el primero solo extrae conclusiones de fenómenos que hayan sido comprobados mediante evidencias procedentes de un corpus. Por otro lado, Mindt pone en tela de juicio la fiabilidad de la intuición e introspección: la intuición debe considerarse como un factor esencial a la hora de seleccionar el fenómeno que desee investigar el lingüista y en el momento de evaluar las evidencias extraídas de un corpus.

4.4.2.1. Unidad léxica y entorno

Uno de los aspectos más controvertidos dentro del trabajo derivado del corpus es la manera por la que las palabras interactúan entre ellas al situarse próximas las unas de las otras. Si aceptamos la fusión entre forma y significado debemos abandonar la noción de que cada palabra es una elección independiente y aceptar que las pautas de selección de palabras en un texto pueden conformar unidades complejas de significado nuevas y más complejas. A este nuevo fenómeno se le ha denominado *coselección* (Tognini-Bonelli 101), que describe que la tendencia que tienen las palabras de *volver a aparecer* conjuntamente es tan fuerte que éstas pueden adquirir significados diferentes. El ejemplo más evidente de coselección lo encontramos en los modismos y en las frases idiomáticas, en donde dos o más palabras, cada una con una carga semántica diferente, se unen para

crear un nuevo significado. Según Clear, se trata de un hecho en el que las palabras pierden la capacidad de combinarse libremente (Firth, *Papers* 272), mientras que Hunston y Francis (8) lo denominan *frases léxicas*, para describir las unidades en donde, de manera regular, el léxico y la gramática se *coseleccionan* automáticamente.

Por otro lado, Tognini-Bonelli (104) coincide con las postulaciones de Sinclair, el cual sostiene que el lenguaje se produce por medio de una relación lógica entre la *tendencia terminológica* y la *tendencia fraseológica*. El problema aparece a la hora de encontrar el límite entre una unidad léxica relativamente independiente y otra en donde el entorno forma parte de la frase: una unidad léxica formada por múltiples palabras.

4.4.2.2. Prosodia semántica

Una de las más obvias consecuencias de la ampliación de la unidad de significado es, probablemente, que las palabras *coseleccionadas* dejan de poseer independencia propia. Si una voz se emplea regularmente en contextos positivos o negativos, ésta lleva dicha carga semántica consigo. Mientras que la preferencia semántica identifica el campo semántico en el que se ubica una unidad léxica, Sinclair afirma que la prosodia semántica es “actitudinal y ubicada en el espectro semántico-pragmático”, y la define como “la elección funcional que conecta el significado con el propósito” (*Units of Meaning* 87). Representa el límite exterior de la unidad de significado donde el *cotexto* se funde con el contexto, momento en el que una unidad léxica alcanza un propósito particular en un entorno determinado.

4.4.2.3. Deslexicalización

Nos hallamos ante una de las ideas de Firth por la que concibe este planteamiento, que es el proceso que se origina al analizar una unidad de significado amplia (*extended unit of meaning*) en donde una unidad léxica pierde, al menos, parte de su valor léxico original, adquiriendo significados y funciones diferentes dentro de una unidad multiléxica. Se trata de una relación de colocación en la que una palabra pierde la mayor parte de su contenido semántico, al tiempo que la unidad con la que se *coloca* suele mantener casi todo su

significado original (Tognini-Bonelli 116). Cuando existe la necesidad, la palabra *deslexicalizada* desambigua el significado de la otra: selecciona un significado y acomete una especie de compensación al adoptar un papel más gramatical. Sinclair (*Millennium* 323) argumenta que se trata de un fenómeno muy común entre las palabras con una alta frecuencia que tienden a perder su significado propio.

4.4.2.4. Ideología

A menudo las palabras portan consigo una carga ideológica que, a veces, podemos imaginar, pero difícilmente probar. Firth (*Papers* 27) comentaba que “cada persona lleva su cultura... consigo adonde vaya”. Partiendo de una premisa similar, Stubbs (*Computer-assisted Studies* 194) admite que “al buscar las colocaciones más frecuentes podemos entrever las expresiones recurrentes que circulan en el mundo social, y vislumbrar cómo las categorías lingüísticas se convierten en categorías sociales”. Según Firth, el lenguaje refleja la cultura, haciéndose patente en la interacción de los diversos parámetros contextuales. Éste aboga por investigar la detallada distribución contextual de las palabras significativas desde el punto de vista sociológico, que podrían denominarse *palabras centrales* o *cruciales*. Este investigador situó su estudio en el ámbito de la semántica histórica, puesto que con anterioridad los corpus existían únicamente a modo de recopilaciones de textos para fines historiográficos. Lo que Firth denominó “la continuidad de las repeticiones en el proceso social” (*Papers* 183), hace que este tipo de análisis sea el idóneo para implementarse con programas informáticos, al ser posible la identificación de frecuencias de ocurrencia en el seno de un gran corpus casi al instante.

4.5. LINGÜÍSTICA DE CORPUS: ¿METODOLOGÍA O DISCIPLINA?

En la actualidad, la concepción de la lingüística de corpus se debate entre la condición de teoría o metodología. Se argumenta que no se trata de un campo de investigación específico, sino de una base metodológica con la que se estudia el lenguaje. Basándose en dicha premisa, Leech (*Introduction* 8) considera que un corpus (conjuntamente con su correspondiente *software*) es una herramienta con la que se puede procesar y analizar un enorme y complejo número de datos lingüísticos. McEnery y Hardie (2) afirman que la

lingüística de corpus no es una rama de la lingüística como lo son la sintaxis, la semántica, etc.; se trata de una metodología más que de un aspecto del lenguaje que requiere de una descripción o explicación. No obstante, la lingüística de corpus ha experimentado un cambio en su naturaleza, pasando de ser un elemento secundario de la lingüística a un componente y método indispensable de ésta. A partir de los años noventa, la lingüística de corpus ha comenzado a singularizarse claramente del resto de los campos de la lingüística y un cierto número de seguidores de las doctrinas de Firth desean considerarla como un dominio diferenciado de la lingüística.

Muchos lingüistas consideran que esta *disciplina* posee una serie de aspectos que traspasan el papel meramente metodológico. Este ámbito engloba todas las actividades relacionadas con la recopilación de datos, hecho que está provocando un cambio exponencial en nuestra comprensión del lenguaje (Halliday y James 24). Otros expertos justifican que ello también se debe al empleo de la informática y de diversos métodos estadísticos y algorítmicos, herramientas que han conllevado un cambio cualitativo en lo que se refiere a la extracción y examen de los datos objeto de estudio. En el siglo XXI, podemos afirmar que la lingüística de corpus se ha convertido en un nuevo campo de investigación y en un nuevo enfoque filosófico de los estudios lingüísticos. En la actualidad, la lingüística de corpus se emplea como herramienta de estudio en diferentes disciplinas tales como la historia, la geografía, la política, la educación, el derecho, el periodismo, la sociología, el aprendizaje de lenguas, etc.

De acuerdo con Cabré (*Terminología* 25), la famosa cita de Saussure, “es el punto de vista el que crea el objeto” puede ayudar a la hora de explicar este cambio de enfoque metodológico. Si las dimensiones del punto de vista cambian, el objeto que se crea es substancialmente diferente del anterior. Por tanto, es la metodología la que define el dominio. En consecuencia, se puede argumentar que la lingüística de corpus es mucho más que una base metodológica para el estudio que queda intrínsecamente unido a dicho enfoque. De esta manera, la lingüística de corpus se convierte en una disciplina por derecho propio, acaparando un dominio claramente especificado (Tognini-Bonelli 49). A pesar de pertenecer a la esfera de la lingüística aplicada, en la actualidad, la lingüística de corpus posee la capacidad de ser definida mediante una serie de reglas y elementos de conocimiento propios. En consecuencia, ésta posee un estatus teórico específico y, debido a ello, se encuentra en disposición de contribuir en otras disciplinas de manera específica.

Ambos planteamientos divergentes se basan en cómo considerar el corpus: algunos aducen que éste posee una base teórica y otros una base metodológica. Los partidarios que admiten que el corpus en sí es, al mismo tiempo, teoría y fuente de información, afirman que el corpus debe poseer un estatus teórico (*corpus-as-theory*). En otras palabras: la lingüística de corpus puede considerarse como “un enfoque teórico con el que estudiar el lenguaje” (Teubert 2). Por otro lado, otros autores estiman que el corpus no es más que un método (*corpus-as-method*). Estos últimos estiman desacertado mantener que los datos, por sí solos, son teoría. Para éstos, los datos son datos y la teoría, teoría. La observación y el estudio de los datos y ejemplos son importantes para contrastar los hechos y resultados, pero no constituyen de por sí una exposición o definición (McEnery y Hardie 147-78).

Independientemente de qué categoría se le deba otorgar a la lingüística de corpus, y cuál sea la metodología a emplear, se trata de una rama de la lingüística que traspasa las barreras de la misma para analizar uno de los aspectos más importantes que caracterizan al ser humano: el lenguaje.

4.6. LA ANOTACIÓN EN EL CORPUS

El estudio del lenguaje mediante un corpus empleando programas informáticos resulta ser un enfoque extremadamente útil. No obstante, su valor no se puede definir únicamente por sus dimensiones. Del mismo modo, la diversidad de los corpus (y de los criterios de su formulación) es un factor muy importante a la hora de analizar un campo determinado. Asimismo, otro factor determinante es su modo de compilación y codificación a la que ha sido sometido para enriquecerlo. Ese *valor añadido* (Leech, *Introduction* 9) es lo que se denomina anotación.

Se puede definir la anotación de corpus como la práctica de añadir información lingüística de relevancia a un corpus electrónico de datos lingüísticos a nivel oral y/o escrito (Leech, *Introduction* 2). De igual modo, la anotación también se refiere a los símbolos lingüísticos que se asocian o se adjuntan a la representación electrónica del material lingüístico. La anotación más extendida es el etiquetado gramatical, en donde se asocia una etiqueta a una palabra para indicar su categoría gramatical. La anotación es el producto de la comprensión que el ser humano tiene del texto, en donde no existe un

mecanismo que clasifique todos los fenómenos lingüísticos que puedan aparecer; pongamos como ejemplo la desambiguación, en donde se procede a la eliminación de la ambigüedad en el lenguaje (Baker, Hardie y McEnery 58). Además, la anotación posee un carácter metalingüístico al proporcionar información acerca del lenguaje que alberga el texto (McEnery y Hardie 29).

La anotación no tiene relación alguna con la representación ortográfica ni tipográfica de una obra escrita. Sin embargo, en un discurso oral la información figurativa y la interpretativa pueden dar lugar a confusión. En este caso, el transcriptor se ve obligado a *interpretar* el discurso a la hora de redactarlo. A causa de ello, el etiquetado prosódico es, a la vez, la representación de parte de los datos del discurso transcrito y una interpretación que el oyente hace de dicho discurso (Garside, Leech y McEnery 4). En lo que se refiere a un corpus escrito, el texto en sí constituye los datos y las anotaciones quedan superpuestas en él. En un corpus oral, la grabación se puede conservar a modo de transcripción textual a la que no se le añade información adicional alguna.

La anotación de corpus implica una serie de ventajas, de las que destacamos la posibilidad y facilidad en la extracción de información, su reutilización por terceros y la versatilidad en sus aplicaciones (Bowker y Pearson 88). Leech denomina corpus básico (*raw corpus*) al que en su forma ortográfica no contiene información alguna (McEnery y Hardie 137). Este modelo de corpus nos proporcionará la información necesaria para resolver la ambigüedad (mencionada anteriormente) de cuando una palabra funciona, por ejemplo, como sustantivo o como verbo.

Por otro lado, una vez que un corpus ha sido anotado éste recibe una información adicional muy útil para que, entre otras cosas, pueda ser empleado por otros usuarios. El trabajo de anotación es muy laborioso y costoso, y por ello es uno de los principales motivos para su reutilización. Una de las principales características de la anotación es su uso para múltiples aplicaciones, que van desde el campo de la lexicografía, la terminología, el análisis del discurso, hasta la traducción asistida por ordenador o la recuperación de información. Se trata de un *valor añadido*, en donde la anotación gramatical es el primer paso para acometer otros tipos de anotación, tales como el sintáctico o el semántico.

En lo que respecta a nuestro trabajo, y al no ser necesaria ningún otro tipo de anotación, detallaremos en el apartado 7.2.2., titulado *Extracción terminológica lingüística*, el sistema de anotación gramatical empleado (UCREL CLAWS 7) y los análisis que hemos llevado a cabo a la hora de identificar tanto unidades terminológicas simples como complejas.

El siguiente capítulo está dedicado a la otra metodología sobre la que se fundamenta esta investigación: la terminología. Procederemos a delinear su historia, evolución, premisas teóricas y la situación en la que encuentra, encrucijada a la hora de catalogarla como metodología o disciplina.

5. LA TERMINOLOGÍA

Iniciamos en este capítulo la descripción de la segunda metodología empleada en nuestro trabajo de investigación. Detallaremos, de manera sucinta, la historia y evolución de la terminología, procediendo seguidamente a su definición. De igual modo, y como ya hemos apreciado en el capítulo dedicado a la lingüística de corpus, han aparecido voces discordantes que ponen en tela de juicio las premisas teóricas sobre las que se fundamenta la percepción tradicional de la terminología. Es por ello que analizaremos las posturas que sostienen los defensores de la Teoría General de la Terminología y el razonamiento de los que promueven un cambio con el propósito de otorgar a la terminología el estatus de disciplina a todos los efectos.

5.1. HISTORIA Y EVOLUCIÓN

La terminología, a pesar de ser una labor centenaria, empezó a perfilarse como disciplina hace relativamente pocos años. Desde tiempos remotos, dicha labor se había realizado en el marco de las actividades de la lexicografía y la dialectología, y es en los siglos XVII y XVIII cuando se inician los trabajos de prácticas conscientes de terminología. En Europa Central, las primeras obras terminológicas se elaboran con el propósito de ordenar el vocabulario específico de un dominio en particular y en una sola lengua. Entre dichos trabajos, caben destacar los de Alberto Durero en el campo de las matemáticas (siglo XVI), los de Andrés Vesalio en anatomía (siglo XVI), los de Antoine Lavoisier (siglo XVIII) y Claude Louis Berthollet en química y los de Carlos Linneo en botánica y zoología (siglo XVIII). Algunas de las nomenclaturas compiladas por estos eruditos siguen empleándose en la actualidad, muestra de la validez y complejidad de los estudios que acometieron (Schmitz 3).

Es un hecho evidente que el inglés sigue apoderándose del papel desempeñado hasta hace más de dos siglos por el latín en la divulgación internacional del conocimiento, que ha alcanzado hoy en día una difusión social sin precedentes. No obstante, no es menos cierto que hubiera voces en el pasado que individuaran sobre la terminología, los neologismos, su uso y su impacto en las lenguas de destino. Así, por ejemplo, Macchiavello razonó sobre el problema que subyace en la aparición de préstamos tomados

de la lengua en la que han aparecido las nuevas unidades conceptuales. Dicho pensador reflexionaba sobre la necesidad de la asimilación de los neologismos al sistema lingüístico al que se incorporan, para así evitar el riesgo de que las lenguas parezcan parcheadas y pierdan su naturaleza (Blasucci 217-23).

Denis Diderot, figura clave de la Ilustración, publicó a partir de 1755 la que se puede denominar la primera gran enciclopedia terminológica de la Edad Moderna. Aunque se trataba, en teoría, de la traducción al francés de la obra escrita por Ephraim Chambers, Diderot incluirá nuevas aportaciones (tanto propias como de otros autores) en los diecisiete volúmenes de los que consta este lexicón. Aparte de ser un estudio pormenorizado de todos los campos del saber y de su vocabulario especializado, Diderot también habla de lo que hoy se conoce como *de-terminologization*: a saber, la modificación de la lengua, entendida como un lenguaje común o LGP (*Language for General Purposes*) producida mediante la asimilación de expresiones propias de las ciencias, en la medida en que el oído de los hablantes se va acostumbrando a las palabras nuevas o neologismos y las emplean en el lenguaje cotidiano (Adamo 86-7).

La primera gran obra terminológica multilingüe realizada en nuestros días fue obra de Alfred Schlomann. Este ingeniero alemán elaboró de manera sistemática una serie de diccionarios técnicos ilustrados. Cada glosario, que muestra los conceptos de un dominio específico, recoge cada término traducido en seis lenguas (alemán, inglés, francés, ruso, italiano y español). Entre 1906 y 1928, Schlomann publicó 16 volúmenes de su *Illustrierte Technische Wörterbücher in 6 Sprachen* (diccionario técnico ilustrado en 6 lenguas), constanding cada uno de entre cuatrocientas a dos mil páginas (Schmitz 5).

No obstante, los comienzos de la terminología como una rama de la lingüística se deben al trabajo realizado durante el pasado siglo por un investigador llamado Eugen Wüster. Como técnico, Wüster consideraba de vital importancia la normalización de las unidades conceptuales en unidades léxicas (terminológicas, en este caso) en el campo específico de la ingeniería. En 1930, durante la lectura de su tesis doctoral, Wüster detalló una serie de argumentos enfocados a sistematizar el trabajo metodológico de la terminología, estableciendo una serie de principios a aplicar en dicha tarea, además de describir las reglas básicas de una metodología para procesar los datos terminológicos. Dicha investigación obtuvo tanta notoriedad que, al año siguiente, la Asociación de

Ingenieros Alemanes publicó su libro *Internationale Sprachnormung in der Technik* (normalización internacional del lenguaje técnico) en el que abordaron los temas tratados en su trabajo de doctorado.

- Este primer periodo de la terminología moderna (1930-60) se caracteriza básicamente por la puesta a punto de métodos de trabajo terminológico que tienen en cuenta el carácter sistemático de los términos.

La influencia que tuvo Wüster fue tal que sus ideas sentaron las bases para la creación de una ciencia terminológica. Las tres escuelas clásicas de terminología (la de Viena, de la que fue promotor, la de Praga y la de Moscú o soviética) se fundaron tras conocerse el trabajo de Wüster, el cual alcanzó su cenit tras la aparición de su Teoría General de la Terminología (TGT), que será desarrollada y ampliada por sus seguidores. Wüster afirmaba que la terminología era un campo de estudio interdisciplinario, vinculando la lingüística, la lógica, la ontología y las ciencias de la información con los diferentes campos de especialidad. El elemento común entre estas disciplinas es que todas están interesadas en la organización formal de las complejas relaciones existentes entre conceptos y términos (Sager 2). Se trata de una disciplina autónoma y autosuficiente que no trata con el lenguaje propiamente dicho, sino que incluye el lenguaje.

En 1936, y gracias a los trabajos de Wüster sobre la comunicación técnica internacional, se instaura el *Technical Committee 37, International Federation of Standardizing Associations* ISA/TC 37, el predecesor del ISO/TC 37 (*Technical Committee 37, International Organization for Standardization*). Este investigador fue el que fundamentó los principios internacionales de normalización terminológica y contribuyó de manera significativa a forjar las bases de la sociedad de la información moderna. Así mismo, Wüster colaboró en la compilación y publicación de la primera edición del *International Electrotechnical Vocabulary*, editado en Londres en 1938. Dicho glosario constaba de dos mil conceptos divididos en catorce *subdominios*; los términos estaban definidos en inglés y francés, con sus traducciones en italiano, español, alemán y esperanto.

En lo que se refiere a la semántica, durante la primera mitad del siglo XX éste no era un tema de interés para los lingüistas, los cuales aducían que se trataba de una interacción

de la lingüística con otros campos cuyas bases eran demasiado débiles para fundamentar un marco teórico. Sin embargo, Wüster planteaba que la lingüística solo podía estudiar la estructura del significado por medio de la lógica, que estudia los conceptos y muestra cómo éstos se relacionan para formar dicha estructura. Este ingeniero austriaco continuó sus estudios terminológicos, tanto a nivel general como en su campo de especialidad: la ingeniería. Con el advenimiento y aplicación de las nuevas tecnologías al proceso terminológico, Wüster definió la terminología como una materia de encrucijada entre la lingüística, la ciencia cognitiva, las ciencias de la información, la comunicación y la informática (Cabré, *La terminología* 232). Su última gran obra, *The Machine Tool. An Interlingual Dictionary of Basic Concepts*, fue publicada en Londres en 1968 bajo el patrocinio de la OCDE, y su propósito era que sirviera como modelo para futuros glosarios técnicos. En 1971, será uno de los promotores de Infoterm (*International Information Centre for Terminology*), organismo auspiciado por la UNESCO, en el que participará activamente hasta su muerte en 1977.

- La segunda fase de la terminología moderna se desarrolla a mediados del siglo XX. Durante esos años, la tecnología y la ciencia sufrieron un desarrollo espectacular, siendo necesaria la invención de nuevos términos para nuevos conceptos en las nuevas y cada vez más especializadas ramas de la actividad humana. Tal como observó Alain Rey, solo en el siglo XX la terminología adquirió una orientación científica, siendo reconocida, al mismo tiempo, como una actividad social de relevancia (Cabré, *La terminología* 70). Gracias también a dicha evolución, la informática ha proporcionado herramientas diseñadas específicamente para estudiar, investigar y tratar correctamente las nuevas unidades terminológicas. El fruto de todo ello propició la aparición de los primeros bancos de datos y los primeros intentos para normalizar la terminología en cada una de las lenguas.

- La tercera fase que ha experimentado esta disciplina (la que comprende entre los años 1975 a 1985) está marcada por la proliferación de la planificación del lenguaje y por los proyectos terminológicos, aunque algunos países, como la extinta URSS e Israel, habían comenzado sus políticas lingüísticas con anterioridad. Es en este periodo en donde queda patente la importancia del papel de la terminología en la modernización del lenguaje, cuyo estudio deja de estar a manos de unos pocos investigadores gracias, entre otras cosas, a la proliferación de los ordenadores personales (Sager 129).

5.2. DEFINICIÓN Y REQUISITOS TEÓRICOS

Antiguamente, a la terminología se la relacionaba con el concepto medieval de trivio, o conjunto de las tres artes liberales de la elocuencia (gramática, retórica y dialéctica), y es por ello que en la actualidad ésta se considera una metodología puramente interdisciplinar (Adamo 88). Ya hemos reseñado que Wüster argumentaba que la terminología era una actividad interdisciplinar ligada a la lingüística, la lógica, la ontología y las ciencias de la información. El elemento común existente entre tales disciplinas es su interés por la organización formal de las complejas relaciones existentes entre unidades conceptuales y unidades terminológicas. Por todo ello, podemos afirmar que la terminología es el estudio y el trabajo de recopilar, describir, procesar y representar los términos pertenecientes a las distintas áreas de especialidad de una o más lenguas. Se diferencia de la lexicografía en la naturaleza de los datos objeto de estudio (unidades terminológicas, y no unidades lexicológicas), en la metodología a emplear y en la especialización que requiere dicho trabajo (Sager 3).

La terminología se relaciona hoy en día con la búsqueda y creación de información y con la compilación de diccionarios, glosarios y bancos de términos. Gracias a la ayuda de las nuevas tecnologías (más concretamente, a la informática), el proceso de recopilación terminológico se realiza actualmente de manera semiautomática. Por otro lado, el número de personas involucradas en este proceso ha aumentado considerablemente; ya no se trata de un selecto grupo de expertos de un campo especializado que trabajan de manera aislada, sino de un numeroso conjunto de profesionales cuyo propósito fundamental es compartir información y llegar a una armonización en lo que a procedimientos y técnicas se refiere (Sager 130-37). No obstante, dicha normalización se ha visto obstaculizada por una serie de condicionamientos culturales y lingüísticos propios de cada lengua, en donde las luchas de poder o el riesgo de su desaparición toman una importancia relevante. Sea como fuere, no cabe duda que una lengua sin terminología propia no puede considerarse en el mundo actual una lengua de cultura (Cabré, *terminología* 34).

La terminología se basa en el significado y, por ello, puede estudiarse desde tres puntos de vista diferentes. Por tanto, podemos afirmar que una teoría de la terminología tendrá tres dimensiones (Sager 13):

1. Una dimensión cognitiva, que relaciona las formas lingüísticas con su contenido conceptual.
2. Una dimensión lingüística, que examina las formas existentes y potenciales de la representación de la terminología.
3. Una dimensión comunicativa, que estudia el uso de la terminología y justifica la actividad humana en la tarea de compilación y procesamiento terminológicos.

Si la terminología estudia los conceptos con el objetivo principal de representarlos en el seno del lexicon de un lenguaje de especialidad o LSP (*Language for Special Purposes*), una teoría de la terminología debe estar enfocada primordialmente hacia la creación de un sistema referencial que someta a relación las estructuras del conocimiento con las léxicas y que defina los elementos constituyentes de cada una de las estructuras. A la hora de formular una teoría de referencia es imprescindible configurar una serie de principios para la clasificación de las unidades léxicas de acuerdo con las propiedades de los conceptos a los que aluden. Wüster afirmaba que “era imposible normalizar los términos de manera satisfactoria sin organizar previamente o al mismo tiempo todos los conceptos relevantes” (*Bibliography* 7). Por tanto, es necesario proponer en primera instancia un modelo de conocimiento que permita describir, relacionar y clasificar los conceptos.

La unidad terminológica, objeto primordial de la terminología, se concibe como el símbolo que representa una unidad conceptual. Por tanto, los conceptos deben crearse o existir antes de que se formen los términos para su representación. En un primer estadio definimos *concepto* como “aplicación de un proceso de cognición humana que motiva la clasificación de un objeto mediante una abstracción sistemática o arbitraria” (Sager 23). Pero, debido a la gran diversidad de definiciones formuladas con el mismo propósito, y a la considerable divergencia de opiniones, resulta más adecuado considerar, por el momento, que un concepto es un *axioma primario sin definir* (*Ibid.*).

5.3. UNA NUEVA PERCEPCIÓN DE LA TERMINOLOGÍA

No cabe duda que toda evolución implica una revisión rigurosa (si no crítica) de los pilares dogmáticos sobre los que se fundamenta un campo de estudio. El objetivo principal de la terminología tradicional era el de asignar una nueva unidad terminológica a una nueva unidad conceptual que aparecía en un campo de especialidad basándose en un enfoque onomasiológico. La terminología debía ser eficaz y carente de ambigüedad, teniendo como objetivo la búsqueda de la monosemia y la univocidad. No obstante, en algunos campos de especialidad existe la *ambigüedad intencionada* (Sageder 125), como aparece en numerosas ocasiones en el ámbito jurídico, la cual juega un papel importante, sobre todo a la hora de servir para su propósito, que es el de una herramienta eficaz por la que un letrado dispone de todas las opciones que le brinda la interpretación de una ley (Chovanec 9).

Después de muchos años en los que se ha debatido poco o nada de los aspectos teóricos de la terminología, han aparecido innumerables voces críticas sobre los principios establecidos, instando al estudio de nuevas alternativas a la teoría clásica. Desde la década de los noventa, y gracias a las publicaciones especializadas y a diversos seminarios y conferencias, se comenzó a abogar por sentar las bases desde las que debatir acerca de la conveniencia o no de reconfigurar el concepto de terminología. De todos ellos, cabe destacar el seminario sobre la teoría de la terminología celebrado en Barcelona en enero de 1999. En dicho encuentro participaron lingüistas, psicólogos, científicos y filósofos que se mostraban críticos con la denominada Teoría Tradicional de la Terminología (TGT) cuyo máximo representante era Eugen Wüster. Por otro lado, el encuentro que tuvo lugar en Vasa en 2001 bajo el título *Terminology Science at the Crossroads?*, el cual formaba parte del simposio europeo de lenguajes de especialidad o LSP, reunía a partidarios de la teoría tradicional teniendo como objetivo el de analizar el sentido y la importancia de las reprobaciones emitidas hacia ésta. Los expertos en terminología allí reunidos, que (supuestamente) no aceptaban del todo las máximas emanadas del trabajo de Wüster, propusieron complementar las deficiencias existentes en la TGT.

Los investigadores que se mostraban partidarios de un revisionismo comenzaron a atraer más adeptos interesados en conocer las alternativas a los antiguos dogmas. De entre ellos, cabe mencionar a Gaudin (1993), Cabré (1999), Temmermann (2000) y Cabré y

Feliu (2001). 2003 fue un año crucial para este grupo de renovadores, los cuales tuvieron la oportunidad de expresar sus ideas en diversas conferencias celebradas en Praga, Surrey, París y Lisboa. Desde entonces, la terminología se debate entre la defensa de los principios de la Teoría General de la Terminología y la búsqueda de nuevos caminos que puedan dar cuenta de la complejidad de los términos. Las causas más importantes de este revisionismo son la concepción global de la unidad terminológica, la reducción a su condición denominativa, el olvido de los aspectos sintácticos, la ignorancia de los aspectos comunicativos y la negación de su variación formal y conceptual. Además, su supuesta homogeneidad no queda reflejada en los datos empíricos relativos a los distintos ámbitos temáticos, las finalidades aplicadas o la defensa de una única metodología. De igual manera, la aplicación de las nuevas tecnologías al trabajo terminológico ha comportado un replanteamiento de la metodología y de la estrategia onomasiológica de la TGT, a la que se le achaca su carácter reduccionista, que le hace incapaz de dar cuenta de la complejidad del fenómeno terminológico en el marco de la comunicación especializada.

Las voces críticas a la terminología tradicional, en lo que se refiere a la psicología cognitiva, argumentan que los interlocutores juegan un papel relevante en la creación del conocimiento a través del discurso y que la cultura está omnipresente durante la percepción de la realidad. Tanto desde la terminología como desde la filosofía se enfatiza la dificultad de separar con claridad el conocimiento general del especializado, puesto que el primero contribuye a la adquisición del segundo. La teoría tradicional se fundamenta principalmente en el sistema conceptual, siendo éste la base del lenguaje especializado; no se le presta atención a la evolución del lenguaje, al emplear como referencia el enfoque sincrónico. No obstante, los conceptos evolucionan con el tiempo, al igual que sus designaciones (Temmermann 15). Desde la vertiente social, se enfatiza el carácter comunicativo de la terminología, la importancia de la implantación social y la diversidad nocional del concepto *normalización*. En lo que se refiere al campo de la lingüística, se estima erróneo defender que la terminología es lenguaje natural, al igual que considerar que los términos son unidades sin interés sintáctico e irrelevantes para el discurso, y defender la univocidad y la monosemia de los términos, cuando los datos extraídos de la realidad revelan lo contrario. Por tanto, Cabré afirma que la teoría clásica no ha sido capaz de crear un marco teórico en el que asentar sus propios métodos y principios, quedando frenada la investigación en detrimento de la normalización (*Theories of Terminology* 182).

La necesidad y el deseo de renovación de las bases en la que se sustenta la terminología quedan resumidos en palabras de Picht y Schmitz:

...los progresos futuros de la terminología se encuentran sitiados por numerosos problemas, para los cuales todavía no se han propuesto soluciones adecuadas. Una cuestión que se debe plantear es la formulación de una amplia (sic) teoría general de la terminología, capaz de adaptar los nuevos conceptos que han surgido durante los últimos catorce años, de manera que se incorporen a una estructura teórica completa (278).

Según Cabré (*Théorie des portes* 3-7), la terminología no se había beneficiado hasta ahora del desarrollo natural que han experimentado otras disciplinas. Esta anómala evolución se debe a que se trata de una disciplina joven, que no ha sido hasta hace muy poco cuando se han comenzado a debatir con seriedad sus conceptos básicos (Cabré, *Theories of Terminology* 169). Otro motivo a destacar es que en las pasadas dos décadas las discusiones teóricas se habían limitado a simples conjeturas sin tener en consideración que una teoría sólo es válida en la medida en que permita la descripción del objeto de estudio y, en consecuencia, la de los datos por los que dicho objeto se manifiesta. Así mismo, la inexistencia de opiniones confrontadas hasta ahora ha conllevado un estricto control por parte de los seguidores de la teoría tradicionalista, desvirtuando cualquier atisbo de disconformidad con los parámetros establecidos. A todo ello se le debe añadir la enorme influencia de la Escuela de Viena, por la que se daba a entender que todos los investigadores aceptaban sus dictados sobre la descripción del significado (Temmerman 22).

La reacción entre los correligionarios de la doctrina dominante no se hizo esperar. Según Myking (63), las diferentes posiciones epistemológicas y paradigmáticas podían resultar un peligro potencial que impidiera la comunicación entre las diversas comunidades terminológicas, y para evitar contratiempos sería necesario, al menos, leer las propuestas de los disidentes. Como ya hemos aducido, tal actitud poco conciliadora argumentaba que las críticas se debían al desconocimiento de las doctrinas de Wüster y de los trabajos de sus seguidores. Sin embargo, esta postura dio paso a tres formas de afrontar la disidencia: la línea negativa, que abogaba por ignorar y rechazar la propia existencia del oponente; la línea constructiva, dispuesta a revisar la teoría y a valorar las nuevas ideas, o

a reinterpretar el trabajo de Wüster para acallar las críticas; y la línea probabilista, que propugna la cooperación para la constitución de una teoría que dé cuenta de toda la complejidad que entraña la terminología (Cabré, *Theories of Terminology* 173). Esta última posición, la más conciliadora, es la que se va imponiendo, teniendo como seguidores a Budin (2001) y a Myking (2001), entre otros.

En lo que se refiere a la actitud de los que promueven el cambio, Myking divide a éstos en tres grupos: moderados y leales, radicales y subversivos y radicales y leales (Antia 103). El primero propone integrar la terminología con la lingüística, sin abandonar sus bases teóricas y metodológicas. Se trata de una postura ecléctica que implica un entendimiento liberal de la lingüística, alejada del estructuralismo y del generativismo. Sus seguidores más representativos son Laurén, Myking y Pitch (1998). El segundo grupo destaca por su rechazo abierto hacia la terminología tradicional. La tercera postura de la disidencia, representada por, entre otros, Toft (128), pone énfasis en la necesidad de adaptar y complementar el trabajo de Wüster con las nuevas vías abiertas en la lingüística funcional y cognitiva.

Debido a las diferentes percepciones existentes sobre la materia, Laurén y Pitch (1993) examinaron las concepciones que cada grupo sostenía sobre el concepto, el término, la relación entre éstos, su posición con respecto a la planificación y normalización, el lugar en el que ubicaban a la terminología en el mundo de las ciencias,... Su conclusión fue que la mayoría de los grupos y escuelas terminológicas comparten un buen número de supuestos teóricos, y que sus diferencias estriban en una serie de intereses, prioridades e investigaciones diferentes. Por consiguiente, el objetivo sería el de hablar de una sola teoría de la terminología que compartan todos los investigadores y a la que contribuyen de diversas formas y en diferentes lenguas. De igual manera, Budin (20) habla de la creación de una única teoría de la terminología, colectiva y multidimensional, que se fundamente en los descubrimientos realizados por los pioneros de la terminología, y en el trabajo de una nueva generación de investigadores que pone de manifiesto la necesidad de discutir sobre ciertos aspectos conceptuales.

5.4. TERMINOLOGÍA: ¿CIENCIA, ARTE O PRÁCTICA?

Además de la existencia de nuevas propuestas para la formulación de un nuevo marco teórico, existe otro punto de divergencia con referencia a la terminología. En lo que concierne a su estatus, la percepción de la terminología en la actualidad se encuentra polarizada entre dos posiciones antagónicas:

a) Para algunos, entre los que cabe destacar a Sager (1), la terminología es una práctica que atiende a una serie de necesidades sociales relacionadas con fines políticos o comerciales.

b) Para otros, tal como afirman Cabré (*terminología* 22) y Kageura (*Dynamic of Terminology* 2) entre otros, se trata de una disciplina científica a todos los efectos que debe mucho a otros campos del saber. Este último enfoque sostiene que la Terminología ha sufrido una revisión y que se han sintetizado sus fundamentos originales para que pueda constituir su propio campo disciplinario (de ahí que la escribamos por tal motivo con mayúscula).

c) Por otro lado, existe un tercer grupo que sostiene una posición intermedia. De entre ellos, cabe destacar a Temmerman (37-8), quien argumenta que la terminología solo se concibe como una tarea que se encuentra dentro del marco de una disciplina ya consolidada. En consecuencia, podemos afirmar que la terminología se concibe dependiendo de la manera en la que se emplea.

El segundo grupo de renovadores considera a la terminología como una disciplina científica independiente y desea crear un marco teórico para la Terminología con el que describir sus dinámicas. Sus usuarios serían los lingüistas, los sociolingüistas y los expertos de las ciencias cognitivas, siendo su objetivo el de crear una sólida Teoría de la Terminología (Sageder 128). La primera y tercera corrientes perciben la terminología como una práctica o como un arte. Su objetivo es su uso en la comunicación en los campos de especialidad, por medio de intermediarios, y para la compilación de glosarios y bancos terminológicos en cada una de las disciplinas.

5.4.1. Las nuevas bases teóricas

Las tesis que han tenido una mayor divulgación han sido, a nuestra opinión, las formuladas por Cabré y Kageura. No es nuestra intención limitar las nuevas concepciones de la terminología a las ideas de estos dos investigadores, sino plasmar y ejemplificar la actual concepción que se posee de la terminología mediante estas dos figuras destacadas.

a) Cabré (*Théorie des portes* 6-9) considera que su idea está concebida como una propuesta para la discusión. Su objetivo es proponer una teoría de las unidades terminológicas. El motivo principal por el que no se propone una *teoría de la terminología*, sino una *teoría de los términos*, es que al hablar de una teoría de la terminología se concibe principalmente su carácter disciplinario. Por el contrario, Cabré considera que son los objetos los que constituyen el elemento sujeto a teorizar. Siguiendo este enfoque, se sostiene que la terminología no pertenece al campo de la lingüística, sino que puede ser tratada a partir de las teorías lingüísticas, así como a partir de una teoría del conocimiento o de una teoría de la comunicación.

Los términos (unidades del campo de conocimiento terminológico) se pueden analizar desde perspectivas diferentes y, en tanto que objetos poliédricos, pueden participar de su campo de estudio y convertirse en parte central del objeto de análisis y de estudio teórico. Por otro lado, la comunicación especializada no se puede estudiar de forma autónoma, al margen de otros signos dotados de capacidad referencial y denominativa, ni alejada de las teorías que proponen explicar la comunicación y la cognición; pero aún así, debemos describir el carácter multidisciplinario de sus unidades. Su propuesta de construcción teórica se denomina *la teoría de las puertas* (*Ibid.*). Se trata de una concepción de la terminología como un campo de conocimiento intrínsecamente multidisciplinar, que se ocupa de las unidades terminológicas y que integra sus aspectos cognitivos, lingüísticos, semióticos y comunicativos.

b) Por otro lado, cabe destacar el trabajo realizado por Kageura (2002) y su propuesta de marco teórico para la terminología. Éste intenta clarificar una serie de conceptos relacionados con las dinámicas de la terminología (formación de términos o desarrollo terminológico) que sirvan para la conceptualización de esa nueva base sobre la que fundamentar la terminología. Cuando Kageura se refiere a término y terminología como

objetos empíricos los escribe con minúscula, y con mayúscula cuando denota el estudio que toma la forma teórica que puede, *de facto*, considerarse un estudio de la terminología *per se*, distinto a un estudio de terminología pertinente a términos o a terminología fundamentado en datos terminológicos. Este investigador razona el estado teórico del estudio de la terminología por medio de dos nuevos conceptos: *quid iuris* y *quid facti* (*Dynamic of Terminology* 1-8). *Quid iuris* estudia el aspecto teórico de la Terminología y las condiciones por las que ésta aspira a obtener un estatus independiente como estudio teórico de la Terminología, y *quid facti* describe a las unidades terminológicas como objetos observados de manera empírica.

Llegados a este punto de explicitación de la tesis objeto de análisis, podemos constatar que los dos pilares sobre los que se basa nuestra investigación están sujetos a un revisionismo de sus premisas teóricas. Por un lado, en la lingüística de corpus existen dos modos de pensar antagónicos (tanto en lo que se refiere a la importancia que se proporciona a los datos extraídos del corpus como a su estatus teórico) y no se ha planteado un acercamiento entre ambas escuelas. A pesar de ello, no estimamos vinculante el consenso entre ambas corrientes para el empleo de dicha metodología o disciplina. Por otro lado, existe una evidente falta de acuerdo entre los valedores de la TGT y los que demandan una reformulación de los pilares sobre los que ésta se fundamenta. No obstante, entendemos que, a largo plazo, sus dos vertientes lograrán consensuar una nueva Teoría de la Terminología para, no solo su propio beneficio, sino también para todos los lenguajes de especialidad.

Una vez descritas las dos metodologías aplicadas en esta labor investigadora, iniciamos la descripción de los pasos que hemos seguido a la hora de llevar a cabo el resultado práctico de este trabajo: el glosario terminológico. En el siguiente capítulo procederemos a detallar las premisas y los métodos de análisis usados para identificar las unidades terminológicas, el factor clave que caracteriza el lenguaje de especialidad. En este caso, los términos que hemos extraído proceden de discursos especializados que fueron almacenados en un corpus. Detallaremos, a continuación, las pautas que deben tomarse en consideración a la hora de realizar dicho proceso.

6. USO DEL CORPUS EN EL PROCESO TERMINOLÓGICO

En el capítulo cuarto hemos pormenorizado la definición, historia y enfoques existentes en el seno de la lingüística de corpus. La lingüística de corpus se emplea primordialmente en el estudio del lenguaje común o LGP. No obstante, esta disciplina o metodología resulta de vital importancia a la hora de individuar, analizar y definir las representaciones léxicas de las unidades conceptuales de cada uno de los lenguajes de especialidad.

A continuación, detallaremos los diferentes niveles de comunicación en los lenguajes de especialidad, para continuar enumerando los pasos a seguir en la creación y compilación de un corpus que cumpla con las expectativas a cumplir. Posteriormente, comentaremos los diferentes métodos empleados en el trabajo terminológico, delineando las diferencias entre la identificación de las unidades terminológicas monoléxicas y las multiléxicas. Finalmente, describiremos los dos tipos de herramientas a disposición de los terminólogos, que han sido diseñadas mediante dos enfoques diferentes: el lingüístico y el estadístico.

6.1. LENGUAJE COMÚN Y LENGUAJE ESPECIALIZADO (LGP y LSP)

Resulta más didáctico describir el lenguaje especializado al compararlo con el de uso cotidiano. El lenguaje común (LGP) es el empleado a la hora de hablar de hechos consuetudinarios en diferentes situaciones y contextos. Por el contrario, el lenguaje especializado, de especialidad (LSP), es el utilizado para tratar de los distintos campos del conocimiento (Bowker y Pearson 25). En todos los idiomas existen estos dos tipos de lenguaje. Un hablante nativo emplea el LGP en su vida diaria gracias a sus conocimientos de gramática y de vocabulario, que le permiten formular frases comprensibles para los oyentes y entender, al mismo tiempo, a sus interlocutores. Sin embargo, para hablar de un tema específico en mayor profundidad, es necesario emplear su correspondiente LSP.

Es más que inevitable un cierto grado de solapamiento entre el LGP y los diferentes LSP. Un LSP contiene una cantidad de términos especializados que, empleados con palabras del lenguaje común, conforman el discurso de especialidad (Cabré, *terminología* 25). De la misma manera que en el lenguaje especializado se emplean un buen número de palabras del lenguaje general, ya hemos mencionado que algunos términos se convierten en palabras del LGP gracias al proceso de la *de-terminologization* (Meyer y Mackinstosh 111), el cual ocurre cuando los términos pertenecientes a un campo de especialidad se abren paso hasta ser empleados por la mayoría de las personas en su vida diaria.

El vocabulario especializado es uno de los factores que caracterizan los LSP. Existe una disciplina o metodología denominada terminología, cuyo propósito es el de recopilar o describir el vocabulario de los distintos campos de especialidad, siendo uno de los trabajos primordiales de los terminólogos el de compilar glosarios para cada ámbito del saber. Por todo ello, y en lo que se refiere al contexto profesional de la terminología, hoy en día es impensable realizar actividades terminológicas sin el concurso de herramientas informáticas (Vargas Sierra 77).

6.1.1. Diferentes usuarios y niveles de comunicación en los LSP

Partiendo de la premisa de que un LSP es el lenguaje usado para tratar de un campo específico del conocimiento, su propósito debe ser pues el de facilitar la comunicación entre las personas que desean hablar o discutir de un tema especializado. Debido a que el nivel de sapiencia dentro de un determinado ámbito es diferente entre las personas que emplean un LSP determinado, debemos distinguir entre ellos a tres grandes tipos (Bowker y Pearson 27-8):

1. - Expertos: son los que, por medio de su formación académica o profesional, poseen un conocimiento profundo del campo en cuestión. En este grupo cabe también incluir a los aficionados (a la caza, pesca, numismática, etc.) que, aun no siendo profesionales, emplean también un discurso y vocabularios técnicos.

2. - Semiexpertos: estudiantes, personal en proceso de formación o expertos de campos afines que desean familiarizarse con la terminología y las unidades conceptuales de un campo especializado.

3. - No expertos: son las personas que, por un motivo u otro, deben emplear un LSP desconocido para ellos.

De igual manera, cuando se habla de un determinado campo de especialidad, no todas las personas poseen el mismo grado de entendimiento. Los expertos en una materia se expresan a través de un lenguaje muy especializado puesto que, al compartir un conocimiento similar en la materia y en el lenguaje empleados, conocen el significado que entrañan las frases y los términos, sin necesidad de aclararlos para una mayor comprensión.

Cuando la comunicación se realiza entre expertos y no expertos, los primeros emplearán menos términos y más palabras del lenguaje general para dar descripciones simplificadas de los conceptos de una materia dada. El experto no espera que los receptores posean sus mismos conocimientos y el propósito de la comunicación es que el mensaje enviado sea comprensible para el oyente. Este tipo de comunicación se observa primordialmente en artículos periodísticos o divulgativos.

En otro orden de cosas, para convertirse en experto de un LSP es necesario adquirir dos tipos diferentes de conocimiento: el lingüístico y el conceptual (Bowker y Pearson 30). El conocimiento lingüístico consta de los términos especializados, las *colocaciones*, las estructuras gramaticales y unas características estilísticas específicas. El conocimiento conceptual consiste en la información de los conceptos especializados descritos mediante el empleo de los LSP, lo que conlleva la comprensión de las unidades conceptuológicas que entrañan las unidades terminológicas.

Las herramientas de análisis de corpus son muy eficaces a la hora de estudiar un lenguaje de especialidad, en particular, para identificar las palabras y los términos que más se emplean. La manera más sencilla de llevar a cabo dicha tarea es solicitar un listado de palabras, que es la enumeración de todas las unidades léxicas y terminológicas que aparecen en un texto o corpus. Dicho listado suele mostrarnos la frecuencia de repetición

de éstas ordenadas alfabéticamente o por su número de frecuencia (Baker, Hardie y McEnery 76). En este tipo de listados aparecen siempre en los primeros lugares artículos, preposiciones y pronombres, entre otros: palabras pertenecientes al LGP. No obstante, también aparecerán términos específicos de los LSP objeto de estudio. Una vez identificados dichos términos potenciales, se procede a observarlos en contexto mediante el empleo de la concordancia, proceso que nos permitirá conocer su uso y significado.

Otro aspecto a investigar es el de las *colocaciones*. Dentro de un lenguaje de especialidad, las palabras se unen unas con otras para crear unidades multiléxicas que poseen una significación específica dentro de un contexto determinado. En lo que se refiere al estilo, resulta fácil asumir que los LSP siguen las mismas reglas que el LGP. Sin embargo, en muchos de ellos se emplean construcciones sintácticas de uso poco frecuente en la vida cotidiana y es imprescindible conocer su empleo para expresarse y entender a otros expertos de una materia determinada.

No cabe duda que es de especial importancia conocer las unidades conceptuales que se esconden detrás de las unidades terminológicas. En algunas ocasiones, los términos pueden resultar familiares porque forman parte de nuestro léxico general, pero es necesario hacer especial hincapié en que los vocablos pueden poseer significados diversos en un LSP al comportarse como unidades terminológicas y no como unidades léxicas. Para ello, un corpus resulta útil a la hora de suministrar información conceptual.

6.1.2. Creación de un corpus del LSP

Un corpus no consiste en una recopilación de textos realizada al azar, sino que éste debe seguir unos criterios específicos determinados por los objetivos que se desean obtener en cada estudio. Los patrones de diseño para un LGP no resultan prácticos a la hora de investigar un LSP, que deben estar sujetos a las necesidades específicas que debemos contemplar en cada investigación (Bowker y Pearson 45).

Uno de los factores a considerar es el tamaño del corpus, el cual se describe como una gran recopilación de textos. No obstante, el adjetivo *grande* resulta ambiguo, puesto que no existen reglas básicas que determinen el tamaño ideal de un corpus. Ésta deberá ser una

decisión personal fundamentada en las necesidades a satisfacer en cada estudio, la disponibilidad de datos y el tiempo estipulado para acometer la investigación. Resulta digno de mención que la máxima *cuanto más grande, mejor* (McEnery, *Introduction* 2) no se adecúa a las necesidades de configuración de un corpus. Los datos que se extraen de un corpus de pequeñas dimensiones, pero bien diseñado, serán más y mejores que los procedentes de uno mayor que no ha sido creado conforme a nuestras necesidades (Bowker y Pearson 48). Ello no quiere decir que no sea aconsejable compilar un corpus de gran tamaño; todo depende de la *calidad* de los textos. Cuanto más textos satisfagan las exigencias que configuran nuestra investigación, mejor. No obstante, siempre es imprescindible otorgarle mayor valor a la *calidad* que a la *cantidad*.

La búsqueda de información relevante de un LSP es más laboriosa que la de un LGP, al ser el LSP un pequeño dominio dentro de una lengua natural. Así mismo, se asume que la información textual es el indicador más eficaz para la búsqueda, revisión y análisis de las unidades terminológicas (Sager 131). Un problema añadido a la cuestión del tamaño es que si sus dimensiones son demasiado pequeñas puede que no contenga todos los términos, conceptos y patrones lingüísticos relevantes de un dominio de especialidad. Si ello fuere así no sería posible formular generalizaciones válidas (Bowker y Pearson 48). De igual modo, es necesario que una palabra aparezca dos o tres veces para formular cualquier hipótesis basada en la frecuencia estadística. Otro aspecto a tener en consideración es el tipo de corpus a compilar. Un *corpus cerrado* o *estático* es aquél que es finito, de dimensiones fijas, que contiene un ejemplo de uso y del estado del lenguaje en un periodo de tiempo determinado. Por el contrario, un *corpus abierto, dinámico* o *monitorizado* es aquél al que se le añaden y extraen textos durante el paso del tiempo para reflejar los cambios que se aprecian en el lenguaje y su uso (Baker, Hardie y McEnery 64). El lenguaje especializado destaca por su dinamismo; éste se encuentra en continua evolución, al igual que los campos de especialidad. En consecuencia, y debido a su naturaleza, el estudio de los LSP se suele realizar mediante corpus abiertos que se actualizan regularmente para registrar y dejar constancia de los cambios que aparecen.

Otro factor a tener en cuenta es la decisión acerca de la inclusión en un corpus de textos parciales o completos. Los primeros corpus de LGP que se crearon en lengua inglesa (Brown, LOB, y los que se han compilado con posterioridad siguiendo las mismas pautas para realizar análisis contrastivos entre ellos) contenían extractos de unas 2.000

palabras, aunque actualmente existen corpus de varios cientos de millones de voces (como, por ejemplo, el BNC). Cuando se analiza un texto de un LSP, los conceptos, términos, patrones lingüísticos y contextos de interés pueden mostrarse en cualquier parte del texto. No obstante, ciertas unidades conceptuales suelen aparecer definidas en la introducción de la obra, empleándose su acrónimo o sigla en los sucesivos cuerpos que conforman el artículo o informe. Por ello, si se decide incluir solo extractos de textos, podemos caer en el error de omitir partes de éste que contenga información valiosa para nuestro estudio. En consecuencia, se considera conveniente incluir obras completas en un corpus destinado a descubrir las unidades terminológicas que se emplean en un LSP determinado (Bowker y Pearson 49).

Además del número de obras a incluir en el corpus, también es importante conocer la autoría de ellas. Si su mayor parte están escritas por un mismo investigador solo tendremos su particular punto de vista y únicamente aparecerán los términos que suele emplear en sus estudios, pero no todos. Con la idea de obtener una visión lo más global posible de un LSP, es mucho más coherente recopilar textos que hayan sido manuscritos por diferentes autores; cuanto más, mejor, siempre que sean expertos en la materia. De igual forma, cuando se busca información en *Internet* hay que tener especial cuidado a la hora de seleccionar la información. Un artículo editado en una página *web* o en un *blog* de un particular no suele cumplir con los requisitos de representatividad, todo lo contrario si se da el caso de que éstos se encuentran en el portal de una reconocida organización profesional o editorial (Atkins, Clear y Ostler 4-5).

En lo referente al *medio* (si se trata de un escrito o de la transcripción de una disertación oral), existen una serie de usuarios (principalmente traductores y terminólogos) que suelen trabajar por medio de la escritura, mientras que otros (como los expertos en la materia) se comunican tanto a nivel oral como escrito. La decisión de emplear material textual, oral o ambos es una decisión personal del compilador. Un dato cierto es que los corpus orales confieren una mayor dificultad en su creación y que es necesario que los interlocutores no sepan que se está grabando su conversación de manera que ésta resulte natural y no forzada (O'Keeffe y Walsh 167). No obstante, el hecho de grabar una comunicación oral sin el consentimiento de sus hablantes puede acarrear problemas éticos y legales (McEnery y Hardie 62). Debido a ello, la compilación de un corpus textual evita los problemas que acabamos de destacar.

Aunque resulte paradójico e irónico, debemos reiterar que los textos que se incluyan en un corpus deben tratar del campo de especialidad objeto de nuestro estudio. El motivo de esta aseveración radica en que la mayoría de las especialidades se caracterizan por ser multidisciplinarias, por lo que no sería injustificable incluir textos de otras disciplinas afines que traten de temas relacionados con nuestro campo de estudio. Este aspecto nos lleva a otro que es la elección del tipo de texto a escoger. Las publicaciones que muestran los resultados de una investigación suelen estar escritas por expertos y para ellos, con un estilo y vocabulario muy especializados. Por el contrario, en los artículos de divulgación y en los libros de texto, cuyos destinatarios no son especialistas en la materia, los términos suelen aparecer intercalados con sinónimos procedentes del LGP y por definiciones o explicaciones de éstos. El tipo de textos a recopilar en un corpus depende de la naturaleza del estudio, pero si lo que se desea es cubrir un espectro lingüístico y conceptual lo más amplio posible se deben incluir estudios de investigación, libros de textos, divulgativos, de opinión, folletos promocionales y artículos periodísticos, entre otros (Bowker y Pearson 51).

Dependiendo de si el estudio será de carácter monolingüe, bilingüe o plurilingüe, el corpus deberá incluir las lenguas que van a ser objeto de investigación. Un dato importante es que las obras deben haber sido escritas por autores nativos (Bowker y Pearson 71), ya que si éstos son hablantes no nativos o si se trata de traducciones pueden contener expresiones no idiomáticas que tienden a desvirtuar la verdadera realidad del LSP (Adamo 86). A menos que el autor que redacta en una lengua que no es su lengua materna haya recibido la ayuda de un nativo para corregir el texto o cuya experiencia en el conocimiento de esa lengua extranjera esté del todo demostrada, es preferible compilar solo las obras realizadas por hablantes nativos.

Por último, destacar la importancia de la fecha de publicación de las obras. Tanto si nuestro propósito es apreciar el estado actual en un campo especializado, el de hace cincuenta años atrás o su evolución para conocer la variación conceptual de los términos, la aparición de los nuevos o la inclusión de éstos en el LGP, es imprescindible conocer cuándo fueron editados los textos. El momento en el que fueron escritos los textos a incluir en nuestro corpus posee una gran trascendencia (Sager 132), independientemente de si nuestro estudio es sincrónico o diacrónico.

6.1.3. Compilación de un corpus del LSP

Una vez diseñados y definidos los parámetros para la compilación de un corpus, se debe proceder a la búsqueda de información relevante en formato electrónico. Según el campo que procedamos a estudiar, resultará más sencillo o más complicado encontrar textos en revistas científicas, artículos, libros, tesis doctorales, *Internet* etc. Dependiendo de la fortuna que tengamos, se debe tener en consideración la posibilidad de realizar ciertos ajustes en nuestro diseño inicial.

Es por todo ello importante ajustarse a la realidad de los hechos y sopesar el tiempo y el esfuerzo que implicará la compilación de un corpus con los resultados que éste nos proporcionará. En bastantes ocasiones nos veremos forzados a emplear textos almacenados en CD-ROM o, en el peor de los casos, tener que escanear libros, que posteriormente serán digitalizados mediante un programa de reconocimiento óptico de caracteres. Este proceso no es solo laborioso, sino que dicho *software* es también proclive a malinterpretar las voces y a espaciar en exceso las letras y las palabras (al igual que incrementar los saltos de línea), haciendo necesaria la corrección manual, hecho que ralentizará sobremanera nuestro trabajo.

Como ya hemos reseñado con anterioridad, hay también que considerar si los textos (impresos o en formato electrónico) que deseamos utilizar están sujetos a derechos de autor o no. Si es así, ante todo es preciso aclarar este punto y obtener el permiso del escritor para el uso de sus trabajos (Atkins, Clear y Ostler 4).

A la hora de compilar un corpus con textos procedentes de la red hay que tener precaución con los gráficos, fotografías o vídeos añadidos (e incluso con los hipervínculos), los cuales pueden retardar el proceso de descarga al ordenador, debiendo ser previamente eliminados.

6.1.4. Proceso de creación de glosarios

Según el diccionario de la Real Academia Española, un glosario es un catálogo de palabras de una misma disciplina, de un mismo campo de estudio, de una misma obra,

etc., definidas o comentadas, en una o varias lenguas. La cantidad de información que éstos pueden albergar depende de los objetivos del autor. Por ejemplo, un glosario básico contiene un listado de términos y sus equivalencias en una o varias lenguas, quedando destinado a personas que tienen un amplio conocimiento del dominio que trata dicho lexicón, y limitándose a mostrar las traducciones terminológicas. No obstante, una entrada en un glosario corresponde a un concepto en particular y, lógicamente, su definición debe estar redactada en la lengua en la que se compila el glosario (Sager 139). En el otro extremo del espectro se encuentran los que contienen una profusa información que, además de detallar la definición, incluyen ejemplos de uso, sinónimos, acrónimos, subdominios, notas de uso, abreviaturas, etc. (Bowker y Pearson 138). Tales obras son las preferidas por estudiantes y traductores para conocer y emplear correctamente los términos al igual que sus equivalencias en otros idiomas. Lamentablemente, estos detallados glosarios no son comunes al ser costosos de elaborar tanto en lo que se refiere a tiempo como a dinero.

Como norma general, las editoriales que encargan la elaboración de glosarios carecen de los recursos necesarios para compilar unas obras muy detalladas y las que se comercializan suelen contener la definición del término, su equivalencia en otro idioma y, a lo sumo, información sobre su uso. Así mismo, existe una amplia variedad en lo referente a su contenido y a la presentación de los datos que albergan, por lo que ha sido necesaria la previa revisión de algunos de ellos antes de iniciar la composición de nuestro proyecto.

Los diccionarios y los glosarios (al igual que los bancos de términos) se pueden dividir en diferentes categorías: diccionarios de lenguaje general monolingüe, glosarios de lenguaje de especialidad monolingüe, diccionarios de lenguaje general bilingües y glosarios de lenguaje de especialidad bilingües. Los diccionarios de lenguaje general monolingües nos proporcionan información gramatical, definiciones, sinónimos, antónimos, ejemplos de uso, etc. Aunque centrados en las palabras que conforman un LGP, a menudo contienen un cierto número de términos genéricos. No obstante, si lo que se desea es encontrar información sobre el significado y el uso de términos más especializados, estas obras no son las más indicadas. Los glosarios de lenguaje de especialidad monolingües (diccionarios que estudian un dominio específico del conocimiento) suelen enfocarse en proporcionar información sobre el significado, más que

en el uso de los términos. En consecuencia, estos volúmenes no suelen incluir información gramatical o sobre sus pautas de uso.

Los diccionarios de lenguaje general bilingües o plurilingües tienden a enfocarse en el uso de las palabras y no incluyen definiciones. Ello puede representar un problema, en particular cuando una voz es polisémica. El contenido de los glosarios de lenguaje especializado bilingües o plurilingües varía de una entrada por término con equivalencias en varias lenguas a los que cubren menos lenguas e incluyen información más detallada. Como ocurre en los diccionarios bilingües o plurilingües, apenas se encuentran definiciones y el énfasis recae en mostrar equivalencias y ejemplos de uso. En definitiva, no existe un solo diccionario o glosario que nos facilite toda la información, general o técnica, que necesitemos (Bowker y Pearson 138-40).

6.2. EL TRABAJO TERMINOLÓGICO

En esta sección resumiremos los diferentes métodos empleados en el trabajo de búsqueda e identificación de unidades terminológicas. Iniciaremos dicho apartado detallando el proceso antaño empleado hasta la aparición de los programas informáticos de extracción terminológica. Seguidamente, describiremos las diferencias existentes entre la búsqueda de unidades terminológicas sencillas y de unidades terminológicas complejas y los distintos enfoques por los que se acomete dicho trabajo. Finalmente, comentaremos de manera concisa las diferencias existentes entre la extracción terminológica monolingüe y la bilingüe. Se trata de una breve aproximación a la labor previa a la compilación del glosario terminológico, que trataremos en el siguiente capítulo.

6.2.1. El método tradicional

La labor terminológica tradicional comienza por decidir con qué dominio de especialidad y con qué lengua o lenguas se va a trabajar. Seguidamente, es necesario contactar con un experto en dicho campo para que nos aconseje a la hora de seleccionar los términos, definirlos, el ámbito que abarca en comparación con otros, etc. Si el campo de especialidad seleccionado es muy amplio, el terminólogo debe elegir uno de sus

subdominios como punto de partida. Tras ello, el especialista comienza a recopilar documentación de diversa índole relacionada con el subdominio en cuestión. Su posterior análisis nos permitirá seleccionar los términos y apreciar tanto su uso típico como el inusual. El terminólogo lee concienzudamente toda la documentación, subrayando los candidatos a término (que pueden constar de una palabra o de varias, denominadas unidades multiléxicas o unidades terminológicas complejas) y las colocaciones características y poco comunes. Al mismo tiempo, éste comienza a indagar cómo los candidatos a término se relacionan entre ellos para comenzar a delinear la estructura del subdominio, por medio de un sistema conceptual, una representación jerárquica o un diagrama arbóreo (Sager 30-1).

Una vez identificados los posibles términos, el terminólogo se centra en el sistema conceptual. Es en esta fase donde se aprecia si ciertos candidatos a término no lo son o si no pertenecen a la subárea objeto de estudio. Puede que en esta fase el sistema conceptual no se encuentre plenamente delineado, aunque dicho factor no reviste mucha importancia siempre y cuando el especialista tenga una noción aproximada del número de conceptos existentes en el subdominio. Ello nos proporcionará una idea del tiempo que deberemos emplear para completar el proyecto. El siguiente paso consiste en consultar al experto en la materia para aclarar las pertinentes dudas y para dejar concluido el listado de términos. El experto revisará el sistema de conceptos que hemos elaborado, siendo él el que valide o rechace el sistema propuesto (Bowker y Pearson 141-44).

Seguidamente, el terminólogo realizará las mejoras oportunas a su listado de términos y al sistema conceptual y pasará a la siguiente fase, que es la de buscar información sobre los términos seleccionados. Dicha información quedará almacenada en una ficha terminológica, la cual poseerá un cierto número de entradas (definición, dominio, información gramatical, información fraseológica, notas de uso, sinónimos, pseudo-sinónimos, antónimos, etc.) que deberán completarse, aunque no todas al mismo tiempo (Sager 145-53). Es más: habrá ciertas entradas para algunos términos que no podrán rellenarse debido a la ausencia de información relevante. Las fichas terminológicas suelen ser monolingües, aunque se emplean también para glosarios bilingües y multilingües. El motivo es que el terminólogo tiende a iniciar su labor en una lengua y, posteriormente, procede a buscar las equivalencias en otras lenguas.

La definición de las entradas es un tema de relevancia ya que, en la mayoría de los casos, los terminólogos están sujetos a seguir unas pautas que especifican la manera de detallar las definiciones (Elkin 11). Ello hace poco viable el uso de una definición formulada previamente. Por ello, se suele proceder a comparar varias definiciones existentes o a configurar una nueva que cumpla con los requisitos estipulados. En términos generales, la estructura jerárquica de un dominio debe quedar reflejada en su definición haciendo, de dicha forma, más fácil la comprensión del término. Al estructurar las definiciones de esta manera, un término que se encuentre en la base del sistema conceptual adquiere automáticamente todas las características del término con el que está relacionado en la cima de la estructura jerárquica.

6.2.2. El proceso de extracción terminológico actual

Hoy en día no es necesario llevar a cabo esta tarea de forma manual, ya que existen una serie de programas informáticos diseñados para tal efecto: las denominadas *herramientas de extracción terminológica, de reconocimiento de términos o de identificación de términos*. Estos programas resultan de gran utilidad durante la primera fase de la compilación, puesto que buscan todos los términos potenciales en los textos almacenados en un corpus, creando un listado con todos ellos. Un dato a clarificar es que, aunque se denominen herramientas de extracción automática, dicha aseveración resulta cuestionable. Es cierto que con ellas se ejecuta una primera búsqueda de posibles términos (o vaciado de términos) que el terminólogo deberá estudiar. No obstante, es más que probable que el *software* no identifique todos los candidatos a término y sea necesario un examen visual para una correcta verificación. Lamentablemente, estos sistemas aún no son capaces de identificar las unidades terminológicas complejas y confunden las unidades léxicas con las terminológicas al no estar provistos de información que elimine la desambiguación (Cabré, Estopà y Vivaldi 54). En definitiva, este proceso debería denominarse asistido por ordenador o semiautomático para evitar interpretaciones erróneas.

Los programas de extracción terminológica pueden realizar dicha tarea a nivel monolingüe y bilingüe. La extracción en una sola lengua se ejecuta en todos los archivos almacenados dentro de un corpus de un lenguaje de especialidad y escritos en un idioma

determinado (SDL). Las herramientas de extracción terminológica bilingüe analizan corpus alineados bilingües para identificar términos potenciales y sus equivalencias en un segundo idioma. Algunos de estos programas permiten la extracción de parte del contexto en donde se ubica el posible término, proporcionando información adicional muy útil para incluir en la ficha terminológica, datos que serán comprobados en subsecuentes exploraciones.

Ya hemos mencionado previamente que un término puede constar de una o varias unidades léxicas. Dependiendo del programa, esta labor se ejecutará de diferentes maneras.

6.3. EXTRACCIÓN TERMINOLÓGICA MONOLINGÜE

6.3.1. Términos monoléxicos

Como ya hemos aludido previamente, la manera más sencilla de buscar términos monoléxicos es mediante su frecuencia de aparición, a saber: por medio de listado de frecuencia de palabras. Se trata de contar cuántas veces aparecen las palabras en un corpus (McEnery y Hardie 243). No obstante, ciertos programas operan mediante la coincidencia exacta, desestimando las veces que un término aparece en mayúsculas o en plural, contabilizándolos como palabras diferentes (Bowker y Pearson 166). Así mismo, es necesario establecer un valor mínimo de frecuencia, que son las veces que una unidad léxica debe repetirse en un corpus para ser considerada candidata a término. Si acordamos cinco como valor mínimo, el programa nos mostrará un listado en donde se enumerarán todas las palabras que aparezcan, al menos, en cinco ocasiones en el corpus. En teoría, partimos de la base de que una unidad terminológica relevante en un campo del saber se empleará en más de cinco ocasiones. No obstante, ello no quiere decir que si un término aparece en pocas ocasiones en un corpus no sea importante, ya que puede ser muy específico o que su uso entrañe dificultades porque no aparezca en los glosarios o en publicaciones previas por diversos motivos. Este factor debe considerarse como un incentivo para incluir términos poco recurrentes, pero manifiestamente relevantes en nuestros glosarios.

En consonancia con lo que acabamos de describir, existirá un cierto número de vocablos especializados que se emplean con asiduidad que no aparecerán al inicio de nuestra investigación. Ello puede ser debido al uso de sinónimos, abreviaturas, o a poseer diversas formas de escritura (aparte de las formas de singular y plural) y, a veces, las herramientas de extracción no son capaces de agrupar los términos en torno a un lema (Bowker y Pearson 167). Otra forma de identificar términos monoléxicos es comparar un corpus de un LSP con otro corpus de referencia de lenguaje general. El proceso es el de buscar las palabras clave. Un corpus de referencia es un corpus que no se considera representativo de ninguna variedad del lenguaje ni de ningún dominio en particular, sino que se trata de un intento de representación de una lengua (considerada como un LGP) por medio de la creación de un corpus de grandes dimensiones (Baker, Hardie y McEnery 137). Si el término en cuestión aparece con menos frecuencia en el corpus de referencia, ello puede ser un hecho indicativo de que se trata de un término del LSP sujeto a estudio.

6.3.2. Términos multiléxicos

Existe la posibilidad de que un término lo compongan dos o más unidades léxicas. La mayoría de estos programas emplean técnicas lingüísticas o estadísticas (o una combinación de ellas) para identificar las posibles unidades léxicas complejas. Pasemos a explicar ambos métodos.

6.3.2.1. Enfoque lingüístico

Los programas que siguen esta metodología intentan buscar combinaciones de palabras que correspondan a un patrón específico de partes del discurso. Para ello, es necesario que los textos que compongan el corpus estén etiquetados con su correspondiente anotación gramatical (Bowker y Pearson 168). Algunos programas están diseñados para buscar (al menos, en lengua inglesa) diversas combinaciones gramaticales. Por ejemplo, si se solicita una búsqueda de combinaciones adjetivo + nombre común, el *software* creará un listado con todas las combinaciones que coincidan con estos patrones de búsqueda. Lamentablemente, un alto porcentaje de los posibles candidatos a término no lo son, provocando lo que se denomina *ruido* (palabras extraídas erróneamente que no son

términos) y *silencio* (patrones que constituyen términos que no son identificados por el programa) (Cabré, Estopá y Vivaldi 56).

Las unidades multiléxicas que no son términos y que han sido identificadas por error son eliminadas manualmente. Otro problema potencial es que algunos términos pueden conformarse de acuerdo con patrones que no han sido programados previamente y, por tanto, no son reconocidos por las herramientas de extracción terminológica. Para ello será necesario reconfigurar el programa para recuperar tales candidatos y evitar el silencio. Otro fenómeno que puede darse es el de la *elipsis*, que ocurre cuando un elemento lingüístico se suprime sin contradecir las reglas gramaticales. Este hecho imposibilita al *software* identificar una unidad multiléxica como posible candidato.

Un defecto adicional asociado con este enfoque es que depende en exceso del lenguaje. Las pautas de formación de términos varían de una lengua a otra y estos programas están diseñados, en términos generales, para funcionar en una sola lengua o lenguas de una misma familia (Bowker y Pearson 169). De igual manera, algunos patrones suelen ser característicos en ciertos campos de especialidad, pero no en otros, siendo ésta otra debilidad del sistema.

6.3.2.2. Enfoque estadístico

Desde este enfoque, el programa informático busca las unidades léxicas más recurrentes. Dicho análisis se denomina valor mínimo de frecuencia, ya descrito con anterioridad y que, en ocasiones, puede especificarse por el usuario (Cabré, Estopá y Vivaldi 55). Si el valor mínimo de frecuencia se establece en dos, éste será el número de veces que una voz debe aparecer en un texto para que se la considere candidata a término. No obstante, ya hemos comentado que esta simple estrategia acarrea problemas ya que el lenguaje en sí se encuentra pleno de repeticiones y no todas las series recurrentes de unidades léxicas son consideradas unidades terminológicas. Ello implica la aparición de ruido, que deberá ser eliminado manualmente. El empleo de un listado de palabras vacías, *stop list*, o listado que contenga las palabras que se deseen excluir del análisis (SDL), puede reducir el número de candidatos a término. De igual modo, una lista de exclusión de términos puede indicar al programa que ignore las series léxicas que contengan una

palabra funcional al principio o al fin de la unidad de significación (artículos, preposiciones, conjunciones,...), ya que éstas raramente constituyen términos, aunque siempre existen excepciones.

Un aspecto en detrimento del empleo del enfoque estadístico es el hecho de que no todos los términos se repiten dentro de un escrito, factor que puede provocar silencio si, por ejemplo, aparecen solo en una ocasión y hemos situado el valor mínimo de frecuencia en dos. Un problema similar ocurre si se emplean abreviaturas en un texto; estas herramientas de extracción terminológica no son inteligentes y se basan únicamente en la coincidencia de patrones, por lo que habrá que buscar el término y su abreviatura y/o su acrónimo de forma separada y vincularlos manualmente. A pesar de todo ello el enfoque estadístico posee un factor determinante: al trabajar sobre patrones recurrentes no depende del lenguaje en sí, por lo que, en teoría, se puede emplear para el procesamiento de textos en diferentes idiomas.

6.4. EXTRACCIÓN TERMINOLÓGICA BILINGÜE

Previamente hemos esbozado que el proceso de extracción terminológica a nivel de dos lenguas se lleva a cabo por medio de corpus alineados (textos escritos en una lengua origen con su correspondiente traducción en una lengua meta). Con el objetivo de obtener una identificación lo más certera posible, estos corpus deben haber sido previamente alineados a nivel de oración (Bowker y Pearson 171). Las herramientas de extracción terminológica intentan, mediante el uso de la estadística, identificar las equivalencias entre las parejas de oraciones alineadas. Una vez que se encuentra un término en el texto origen, el programa busca en todos los segmentos del texto origen todas las ocasiones en las que se repite dicha unidad de significación especializada para, a continuación, analizar todos los segmentos alineados en el texto meta que albergan la traducción del término identificado en el texto origen.

Una parte del análisis se enfoca en buscar qué elemento o elementos en los segmentos del texto meta se repiten con una frecuencia similar al término que aparece en el texto origen. Otros de los métodos que se emplean es el de formateo (*formatting*), proceso que interpreta que si el término en el texto origen está escrito en letra cursiva o en negrita, es

probable que su equivalencia aparezca escrita de la misma forma en el texto meta (Bowker y Pearson 171-72). Al igual que los métodos anteriores, el programa muestra una lista de posibles candidatos en ambas lenguas a modo de vaciado terminológico. A partir de ahora, el factor humano sigue siendo determinante, al aceptar o desestimar qué términos serán los que compondrán el futuro glosario. Además, no resultaría extraño encontrarse ante la posibilidad de que un término en la lengua origen carezca de su correspondiente equivalencia en la lengua meta. En este caso, el terminólogo deberá paliar esta deficiencia buscando fuera de su corpus documentación que le ayude a solventar dicho problema.

Otro fenómeno que puede aparecer en la extracción terminológica bilingüe es el de la homonimia. Cuando una palabra posee más de un significado, es probable que cada una de tales significaciones quede reflejada en la otra lengua mediante diferentes palabras o términos. Si ese fuere el caso, y deseáramos incluirlos en el glosario, se debe incluir una entrada distinta para cada unidad conceptual.

Por último, puntualizar que, en general, los términos no solo aparecen en su forma canónica (singular para los nombres, e infinitivo para los verbos). Por tanto, es necesario modificar los patrones de búsqueda para encontrar todas las variantes de los términos. En este caso, la anotación del corpus resultaría de gran ayuda, a pesar del esfuerzo que ello requiere.

Hemos expuesto en este capítulo las diferencias entre LSP y LGP y los aspectos a tener en consideración a la hora de compilar corpus y glosarios de una lengua de especialidad. Seguidamente, procedimos a describir los métodos que se han empleado para realizar la extracción terminológica: la tradicional y la actual. Gracias a la informática, dicho proceso se efectúa hoy en día mediante programas informáticos que acometen dicha tarea mediante un enfoque lingüístico o estadístico. Este último enfoque resulta el más eficaz, en particular, a la hora de identificar unidades terminológicas complejas. Finalmente, hemos explicado las diferencias existentes entre la búsqueda de terminología en una o en varias lenguas y cómo se identifican los candidatos a término en una lengua origen y una lengua meta.

Una vez realizada una aproximación a las bases teóricas que conforman nuestro trabajo, procederemos al desarrollo del capítulo en donde iniciamos nuestra investigación

per se, detallando los pasos que hemos llevado a cabo, de manera pormenorizada, al igual que describiendo todos los resultados y los puntos de interés que consideramos dignos de mención.

7. COMPILACIÓN DE GLOSARIO TERMINOLÓGICO

En este capítulo enumeraremos los pasos que hemos efectuado durante el proceso de extracción terminológica. Comenzaremos con la descripción de las premisas teóricas que hemos seleccionado a la hora de delimitar nuestro trabajo, su aplicación y objetivos. Posteriormente, comentaremos el proceso de selección de la información que hemos utilizado, enumerando los motivos que nos han motivado a ello para obtener un grado de especificidad adecuado a nuestras necesidades. Asimismo, especificaremos cómo se ha estructurado el conocimiento de la disciplina militar por medio de los subdominios que hemos acordado para la elaboración del fichero de vaciado terminológico. Con posterioridad, describiremos los dos programas empleados para tal fin, SDL Multiterm 2009 Extract y AntConc, versión 3.5.5.w (Windows).

El primero de ellos, concebido para analizar el lenguaje desde un enfoque estadístico, requiere de una configuración previa al proceso de análisis, hecho que comentaremos y valoraremos, al igual que las decisiones tomadas para tal fin. A continuación, detallaremos el proceso de detección y vaciado terminológicos, así como el método de elaboración de las fichas terminológicas. Del mismo modo, describiremos los distintos procesos por los que hemos identificado los términos (tanto los que aparecían en los glosarios de época consultados como los no normalizados, con los que fue necesario emplear otros métodos para su identificación) y el modo de configurar su correspondiente definición.

En lo que se refiere a la extracción terminológica lingüística, descubriremos el uso que se ha llevado a cabo del *software* AntConc, versión 3.5.5.w (Windows). En primer lugar, enumeraremos el método empleado para la identificación y el examen de los candidatos a término (tanto los compuestos por unidades léxicas simples como complejas) con los textos carentes de etiquetado para, posteriormente, describir el proceso ejecutado con los textos anotados gramaticalmente, describiendo los aciertos y los errores que se han cometido en cada uno de los pasos que hemos efectuado.

7.1. BASES DEL TRABAJO TERMINOLÓGICO

7.1.1. Competencias

Como ya hemos detallado en el apartado dedicado a la terminología, la práctica terminológica presupone el dominio de las tres grandes competencias que configuran la disciplina (Sager 13):

- La competencia cognitiva: el conocimiento del ámbito especializado que va a ser objeto de nuestro trabajo.
- La competencia lingüística: el conocimiento sobre la lengua en la que se lleva a cabo la investigación.
- La competencia sociofuncional, o comunicativa: el conocimiento necesario para acometer un trabajo terminológico con eficacia y adecuado a las necesidades de los destinatarios finales.

No obstante, debemos incluir una cuarta competencia cuando el objetivo final es la compilación de glosarios terminológicos (Cabré y Estopà 3):

- La competencia metodológica: costumbre adquirida y necesaria para llevar a cabo un proceso ordenado y sistemático con el propósito de presentar los datos de manera adecuada y eficiente.

7.1.2. Supuestos básicos de la actividad terminológica

Un término es la asociación de una forma y un contenido. Resulta fundamental preservar el carácter semiótico del objeto terminológico. Forma y significado constituyen dos caras indisociables de la unidad terminológica, al ser unidades sónicas (Cabré y Estopà 4).

Para todo dato terminológico debe existir una fuente real. La terminología planificada es la muestra de que si la definición de un término no es satisfactoria puede acuñarse en forma de propuesta, aunque ello solo será legítimo y eficiente en condiciones

excepcionales y nunca como acto de creatividad individual. Los terminólogos, en muy pocas ocasiones encuentran descripciones formuladas de acuerdo con las normas establecidas, siendo necesaria su formulación empleando exposiciones existentes en glosarios o en documentación especializada (Bowker y Pearson 154). Para todo ello, entre otros factores, es imprescindible citar correctamente las fuentes de toda la información.

7.1.3. Tipos de trabajo

Los métodos de trabajo en terminología se adecúan dependiendo del estudio a realizar, teniendo en cuenta los condicionantes siguientes:

- El trabajo puede ser monolingüe o plurilingüe, aunque es necesario recordar que el resultado de este último es el de tantos procesos de trabajo monolingüe como lenguas incluya (Sager 223).
- De acuerdo con las características intrínsecas del trabajo, éste puede ser sistemático y/o puntual en el tiempo.

Dependiendo de la especificidad metodológica que requiere cada tipo de área de conocimiento, Kocourek (9) clasifica las especialidades de la siguiente forma:

- Ciencias teóricas.
- Ciencias experimentales.
- Ciencias humanas y sociales.
- Técnicas.
- Materias desde el punto de vista de la producción.
- Materias desde el punto de vista del consumo.

En lo referente a nuestra investigación, se trata de un trabajo monolingüe, de carácter puntual, centrado en un área de estudio técnica.

7.1.4. Tipos de aplicación

La terminología, que incluye teoría y aplicación, tiene dos posibles finalidades que dan salida al trabajo terminológico (Cabré, *terminología* 239):

- Descriptiva: La recopilación de términos reales en el uso y su representación en glosarios.
- Prescriptiva: La recopilación de términos reales para su normalización.

No obstante, Schmitz (31-2) añade otras tres finalidades:

- Ad hoc*: Búsqueda terminológica que depende de las necesidades existentes en un momento determinado.
- Textual y puntual: Trabajo preliminar de investigación terminológica. Búsqueda e identificación de unidades terminológicas previa al trabajo de traducción.
- Específico: Propio de una disciplina. Búsqueda de todas las unidades conceptuales de un área de especialidad, al igual que sus correspondientes unidades terminológicas.

En este estudio, partimos del método descriptivo. Sin embargo, podría considerarse el empleo del enfoque prescriptivo ya que incluimos términos que no se encuentran en los glosarios, al igual que sinónimos o variantes empleadas con mayor profusión que los que aparecen en ellos. No obstante, nuestro propósito no es la normalización, al tratarse de un estudio sincrónico, sino el de destacar el uso de la terminología militar durante un periodo de tiempo determinado y la valoración de la competencia y eficacia metodológicas de los especialistas encargados de la compilación de los glosarios tomados como referencia. De igual manera, nuestra labor se enfocará hacia una metodología específica, al tratarse del estudio de una parcela del saber claramente delimitada con el objetivo de identificar las unidades terminológicas existentes en el seno de una disciplina.

7.1.5. Fases

Centraremos nuestro trabajo de investigación en las fases significativas que, según Cabré, constituyen el trabajo terminológico sistemático (*terminología* 201).

7.1.5.1. Presentación del tema y adquisición del conocimiento

El estudio de la terminología militar empleada durante la Segunda Guerra Mundial es un tema altamente especializado en el que, mediante la lectura progresiva de los textos, se amplía la información, se seleccionan los datos relevantes, se reordena el conocimiento y se busca una definición precisa para cada término.

No cabe duda que el conocimiento especializado se adquiere y aumenta a medida que se tiene acceso a la información pertinente, que en este caso ha sido una labor que se ha dilatado en el tiempo a causa de su grado de especificación, y que dicha competencia se estructura a través de los textos especializados. Como es lógico, en dichos escritos existe información significativa (se trata del objetivo principal de nuestro estudio: la búsqueda de unidades terminológicas, su definiciones y ejemplos de uso en contexto) e información no relevante. El léxico especializado sobre una materia concreta aumenta a medida que se obtienen más datos que se procesan sobre ella, factor indispensable si nuestro propósito es estructurar su conocimiento (Cabré y Estopà 13).

7.1.5.2. Delimitación del trabajo terminológico

La delimitación de todo trabajo terminológico conlleva reflexionar sobre una serie de cuestiones (Cabré y Estopà 11):

- El tema en concreto, en qué punto específico o *microtema* del tema más amplio o *macrotema* va a centrarse el trabajo.
- Los destinatarios, a qué colectivo va dirigido el glosario.
- Qué finalidad o actividad lingüística se pretende facilitar o cubrir con el glosario.
- El número de lenguas de trabajo.
- La dimensión, cuántas unidades terminológicas debe comprender aproximadamente el glosario.

Todo ello quiere indicar que, para una elaboración rigurosa de un trabajo terminológico sistemático, es necesario plantearse una serie de condicionantes previos que influirán a lo largo de las diferentes fases del proceso. En el caso que nos ocupa, la

compilación de un glosario se encuentra íntimamente influenciada por el microtema (Segunda Guerra Mundial, operaciones terrestres en el teatro de operaciones europeo, del Mediterráneo y del norte de África) de la disciplina a investigar o macrotema (terminología militar), los destinatarios, las funciones, las lenguas y las dimensiones de nuestro estudio. Es por ello necesario que el trabajo esté claramente delimitado.

Los destinatarios deben ser homogéneos y estar bien definidos, ya que influirán en el tipo de función que debe cubrir el glosario, así como su finalidad (como la de cualquier trabajo terminológico sistemático), que es siempre lingüística o sociolingüística. De igual manera, las dimensiones del trabajo dependerán tanto de la delimitación de éste como de las características y de los recursos disponibles durante el periodo de su compilación.

En lo que respecta a las lenguas de trabajo, la falta de documentación nos ha obligado a llevar a cabo la parte práctica de nuestro estudio únicamente en lengua inglesa. Por lo cual, se trata de una investigación monolingüe, tomando como base la variante norteamericana. No obstante, han quedado reflejadas todas las unidades terminológicas empleadas por el contingente británico, al igual que los neologismos procedentes de otras lenguas y usados por los ejércitos aliados de habla anglosajona.

El microtema que vamos a tratar resulta lo suficientemente vasto para sopesar la idea de compilar un glosario, independientemente de que en la guerra en el Pacífico se emplearan muchos de los tecnicismos que se han incluido en nuestra obra. No obstante, una de las formas más típicas de combate en ese teatro de operaciones eran las operaciones aeronavales (que no anfibas) y la guerra en la jungla, cuyos métodos han sido excluidos de nuestra investigación. Ya hemos puntualizado que nuestra investigación se ha centrado, primordialmente, en la terminología empleada en el seno del ejército de Tierra y en las operaciones llevadas a cabo en los continentes europeo y africano. Por tanto, la terminología específica relacionada con la Marina y el ejército del Aire ha quedado excluida de nuestra investigación, a excepción de las voces especializadas de ámbito general y las relacionadas con armas y técnicas de apoyo a las operaciones terrestres.

En lo que concierne a nuestra labor, el propósito es que la información sea accesible a los destinatarios circunscritos en los tres modelos de comunicación para cada lengua de

especialidad, en la que los ejemplos en contexto facilitarán la comprensión de las unidades terminológicas y sus acepciones. No obstante, nuestro propósito ha sido el de centrarnos en la comunicación entre experto y no experto, con el objetivo de hacer comprensible las unidades terminológicas de la disciplina militar al mayor número de destinatarios posible. Consideramos que cualquier interesado con un nivel cultural medio y con unas nociones no muy profundas del ámbito de nuestra investigación, será capaz de comprender las entradas y definiciones de nuestro glosario. Ello no ocurriría si tratáramos la terminología militar actual. En la actualidad, el grado de especialización y su carácter interdisciplinar, no solo en ámbitos más generales (ingeniería, química, física, geopolítica, electrónica, diplomacia, estrategia,...), sino en campos más específicos (fusión y fisión nucleares, ingeniería espacial, robótica,...) obliga a que el destinatario sea, como mínimo, un semiexperto en la materia.

Como ya hemos comentado en el capítulo dedicado a los objetivos, la finalidad de nuestra labor ha sido la de cubrir un vacío descriptivo del uso real de la terminología militar durante el espacio de tiempo que nos compete, donde solo tenemos constancia de la existencia de un glosario fechado en 1943, y de un segundo editado en 1944. El propósito de ambas obras, que hemos tomado como referencia terminológica y base de partida de nuestra investigación, se limitaba a proporcionar la definición de cada término, dejando a un lado las unidades terminológicas que, por aquel entonces, eran consideradas coloquiales y otras que, por su carácter histórico, o por su fecha de aparición, se han perpetuado a lo largo del tiempo *a posteriori*. Por otro lado, y dado el carácter lingüístico que posee un glosario, es ineludible añadir información de la subárea temática de cada término, su procedencia, sus posibles abreviaturas o acrónimos, ejemplos en contexto, la fuente donde aparece la definición y los ejemplos aportados y cualquier otra información relevante que nos sirva para una mayor comprensión e identificación de cada entrada.

Es por todo ello que consideramos que los glosarios que hemos encontrado ofrecen una información muy limitada y estimamos que su estructuración puede ser ampliada y mejorada. Uno de nuestros objetivos principales ha sido el de mostrar el uso de cada unidad terminológica; si no se aporta un solo ejemplo en nuestro corpus textual implicará que ha caído en desuso o que su intento de normalización ha sido infructuoso.

bachelor officers' quarters. *Abbrev:* BOQ

back azimuth, angular expression of the opposite direction from a given azimuth along a given line. When the azimuth is known, the back azimuth is calculated by adding or subtracting 180 degrees or 3200 mils to or from the azimuth. For example, if the azimuth from the observer to the target is 85 degrees, the back azimuth, which is equal to the azimuth from the target to the observer, is 265 degrees.

back-azimuth method, method of locating an observer's position on a map or chart with a very accurate angle-measuring instrument. It is done by measuring the azimuth of three visible points that are located on the map or chart, and drawing through each point a line or ray with its calculated back azimuth. The point where the lines or rays intersect is the location of the observer's position. The back-azimuth method is one kind of resection.

back pack parachute, parachute attached to a harness so that it can be worn across the wearer's upper back, thus permitting greater freedom of movement.

back plate, plate at the rear of the breech mechanism of certain automatic guns.

backsight, sighting in a backward direction or toward a former station.

back step, single step 15 inches straight to the rear of a given standing position. The back step, like the side step, is used for short distances only and is never executed at double time.

badge, medal, ribbon, or other device worn on the uniform to show that the wearer has been awarded a decoration or other honor, or to show his organization, rank, or, in some cases, the type of duty he is performing.

baffle, wall or screen that controls the flow of air, water, or gas, by hindering its movement or changing its course. Baffles are used extensively in motor mufflers, gun muzzle brakes, and air-cooled motors.

baffle grease trap, container that is divided into an inlet and outlet chamber by a baffle or screen by which grease and debris are separated from liquid waste matter.

bag, baggage.

baggage. *Abbrev:* bag

baggage train, group of motor trucks or other conveyances carrying the equipment of military units.

bakery. *Abbrev:* bkry

balance, steadiness; steady condition or position. The balance of a rifle is the point at which the weight is evenly divided lengthwise, or its center of gravity. The balance of an aircraft is a condition leading to steady, straight, horizontal flight.

balk, timber put lengthwise in a trestle or floating bridge.

ball ammunition, cartridges containing solid bullets. Ball ammunition is general purpose small-arms ammunition.

ballast, 1. heavy material carried in a ship to steady it.
2. weight carried in a balloon or dirigible to control its buoyancy.
3. crushed gravel or rock used in making the bed for a railroad track or a road.

ball cartridge, projectile that consists of a cartridge case, a primer, powder, and a solid bullet. Ball cartridges are general purpose small-arms ammunition for standard service.

ballistic, having to do with the motion or throwing of projectiles; as, ballistic curve, ballistic wind.

ballistic area, space lying between the centers of impact of two groups of shots, one consisting entirely of shots over the target, the other entirely of shots short of the target.

ballistic coefficient, number that represents the power of a projectile to overcome air resistance and keep up its speed during flight. The coefficient is calculated from a formula that makes allowances for all factors that affect the flight of the projectile, such as its shape, weight, diameter, density of air, direction of the wind, and temperature.

ballistic correction, adjustment in firing data that is based on conditions affecting the flight of a projectile. It allows for such factors as wind, temperature, etc. It does not include adjustment based on observation of fire.

ballistic curve, actual path or trajectory of a bullet or shell as influenced by wind, etc.

ballistic density, assumed constant air density that would have the same total effect on a projectile during its flight as the varying densities actually encountered.

Figura 1. Primera página de la letra "b". United States, War Department. *Dictionary of United States Army Terms*. Washington: GPO, 1944. Print.

As Amended Through 15 December 2013

B

backfill — Reserve Component units and individuals recalled to replace deploying active units and/or individuals in the continental United States and outside the continental United States. See also **Reserve Component**. (JP 4-05)

bale cubic capacity — The space available for cargo measured in cubic feet to the inside of the cargo battens, on the frames, and to the underside of the beams. (JP 4-01.2)

ballistic missile — Any missile which does not rely upon aerodynamic surfaces to produce lift and consequently follows a ballistic trajectory when thrust is terminated. See also **guided missile**. (JP 3-01)

bare base — A base having minimum essential facilities to house, sustain, and support operations to include, if required, a stabilized runway, taxiways, and aircraft parking areas. A bare base must have a source of water that can be made potable. Other requirements to operate under bare base conditions form a necessary part of the force package deployed to the bare base. See also **base**. (JP 3-05.1)

barrage — 1. A prearranged barrier of fires, except that delivered by small arms, designed to protect friendly troops and installations by impeding enemy movements across defensive lines or areas. 2. A type of electronic attack intended for simultaneous jamming over a wide area of frequency spectrum. See also **electronic warfare; fires**.

barrier — A coordinated series of natural or man-made obstacles designed or employed to channel, direct, restrict, delay, or stop the movement of an opposing force and to impose additional losses in personnel, time, and equipment on the opposing force. (JP 3-15)

barrier combat air patrol — One or more divisions or elements of fighter aircraft employed between a force and an objective area as a barrier across the probable direction of enemy attack. See also **combat air patrol**. (JP 3-01)

barrier, obstacle, and mine warfare plan — A comprehensive, coordinated plan that includes responsibilities; general location of unspecified and specific barriers, obstacles, and minefields; special instructions; limitations; coordination; and completion times; and may designate locations of obstacle zones or belts. (JP 3-15)

base — 1. A locality from which operations are projected or supported. 2. An area or locality containing installations which provide logistic or other support. 3. Home airfield or home carrier. See also **facility**. (JP 4-0)

base boundary — A line that delineates the surface area of a base for the purpose of

Figura 2. Primera página de la letra “b”. United States, Department of Defense. *Department of Defense Dictionary of Military Terms and Associated Terms*. Washington: Joint Publication 1-02, 2010. Print.

En la figura 2 apreciamos el grado de especificidad del glosario editado en el año 2010. Los términos que se citan en dicha página se centran en las subáreas de la estrategia,

la balística y la logística. Por el contrario, a destacar que los que aparecen en la figura 1 (publicado en 1944) son de carácter más genérico, incluyendo entradas pertenecientes al equipo de combate (y a partes de éste), al arma de artillería (íntimamente relacionada con la física) y a la intendencia. Las únicas unidades terminológicas que aparecen en la página mostrada del glosario editado en el año 1944, y que se repiten en el publicado en 2010, son *barrage*, *barrier* y *base*.

Como hemos mencionado con anterioridad, nuestro trabajo práctico es monolingüe, en lengua inglesa, enfocándonos en la variante norteamericana, aunque incorporando todos los términos empleados por el Ejército británico y las fuerzas de la *Commonwealth* que aparezcan en nuestro corpus. No obstante, en nuestro estudio hemos incluido las unidades terminológicas procedentes de otras lenguas y que fueron empleadas por los ejércitos aliados de habla inglesa. Nuestro objetivo ha sido el de compilar un glosario de una cifra no inferior a tres mil términos (incluyendo sinónimos, unidades conceptuales descritas por unidades terminológicas diferenciadas entre británicos y norteamericanos, extranjerismos, al igual que coloquialismos). Inicialmente, han quedado excluidas las abreviaturas y acrónimos, que fueron incluidos en cada ficha terminológica, cuando procedía, pero no creando una entrada específica para cada uno de ellos, a no ser que el programa de extracción terminológica o su frecuencia de uso nos hubiera indicado su particular relevancia y fuera adecuada su significación.

Por otro lado, y como deseábamos que nuestra labor no se limitara a buscar términos especificados en los glosarios de época consultados, nuestro mayor interés se centró en la búsqueda de nuevas unidades terminológicas, de sinónimos o de otras variantes cuya definición o su empleo quedarán reflejados en nuestro corpus o en obras publicadas tras la Segunda Guerra Mundial.

7.1.5.3. Selección y ampliación de la información

Las obras empleadas en el corpus proceden, exclusivamente, de informes desclasificados por el Ejército norteamericano, fechados y redactados en el espacio temporal que nos ocupa. Como ya hemos analizado anteriormente, además de su carácter representativo, originalidad, actualidad y nivel de especialización (Cabré, *terminología*

233-35), hemos deseado reflejar los tres modelos de comunicación existentes en el lenguaje especializado, aunque poniendo un mayor énfasis en el tercer grupo:

-Comunicación experto-no experto (manuales para reclutas, folletos informativos y periódicos o revistas destinados a la tropa en general, bibliografía básica destinada a su consulta por los cadetes aspirantes a suboficial u oficial,...). Estos escritos tenían la finalidad de describir, de forma clara y sencilla, los procedimientos militares empleados por los aliados y por sus enemigos con el propósito de mantener informados a los soldados destinados en el frente. Dichos textos tienen la particularidad de estar siempre redactados por medio de una sintaxis clara y sencilla (características del lenguaje general), profusa en explicaciones y definiciones para que todo el personal comprenda el mensaje que se desea difundir. Estas publicaciones son las más idóneas, tanto para encontrar definiciones de los términos, como ejemplos dentro de un contexto especializado.

De igual modo, hemos tomado la decisión de emplear textos completos. No obstante, en diferentes ocasiones hemos optado por eliminar artículos o capítulos de los boletines empleados que no trataban de nuestro tema objeto de estudio (batallas navales, la guerra en el Pacífico, campañas aéreas, empleo de submarinos en el Atlántico,...), determinando su eliminación siguiendo los parámetros que detallamos a continuación (Cabré, Estopà y Vivaldi 56):

- Eliminación de ruido en el proceso de extracción terminológica.
- Delimitación del tamaño de las obras y empleo únicamente de la información relevante.

Otro de nuestros propósitos ha sido el de utilizar información redactada por un variado grupo de autores con el objeto de identificar, dentro de un abanico lo más amplio posible, el empleo de los términos de especialidad. Ya habíamos reseñado que un autor circunscribe el empleo de la terminología a gustos personales y a situaciones específicas, hecho que hace que sea imprescindible compilar un corpus con el mayor número de autores que tengamos a nuestra disposición (Bowker y Pearson 49).

No obstante, ello no ha quedado reflejado de manera explícita en algunas ocasiones por motivos de seguridad. Por ejemplo, en el caso de los boletines publicados por el US Military Intelligence Service (que son los textos que hemos empleado para nuestro estudio), los nombres de los autores de los artículos no aparecen en los textos (ni tampoco los redactores ni los editores) por ser material clasificado. Cuando se trata de publicaciones destinadas al personal que se encontraba en el frente, y en donde se detallan estrategias empleadas por el enemigo y se proporciona información obtenida por incautación o por interrogatorios a prisioneros de guerra, o temas relacionados con el armamento o con las tácticas empleadas por los aliados o por las Fuerzas del Eje, es lógico omitir a sus autores y el lugar o modo de obtención, en caso de que la publicación cayera en manos enemigas. Por ello, quedaba terminantemente prohibido llevar dichos manuales a primera línea debido al peligro potencial que ello acarrearía.

Dicho todo ello, nuestro corpus textual se ha conformado por artículos extraídos de la revista *Intelligence Bulletin* y escritos por personal militar de habla inglesa cuya autoría se mantuvo en secreto por razones lógicas.

En suma, se trata de un estudio sincrónico en donde nuestro objetivo ha sido el de recopilar textos publicados durante la Segunda Guerra Mundial, hasta poco más de un año después de su conclusión. Sin embargo, también podemos catalogar nuestro trabajo como diacrónico al analizar la evolución de las unidades terminológicas, y cómo algunas han pasado de ser meros sinónimos a obtener la condición de candidato a término *per se*, con una nueva significación durante ese estrecho periodo de tiempo. En lo que respecta al estatus de nuestro corpus, una vez concluida nuestra labor de investigación, ha sido catalogado como cerrado, aunque nuestro propósito es el de continuar con nuestra labor e incluir nuevos textos, por lo que, en un futuro próximo será catalogado como corpus abierto y así estudiar, de ese modo, la evolución que ha sufrido la terminología de la ciencia militar hasta nuestros días.

Como ya hemos reseñado, hemos utilizado artículos completos siempre que ello fuere posible (Bowker y Pearson 49). No obstante, no se han tenido en cuenta algunas secciones de las publicaciones que trataban de temas que no estaban vinculados directamente con nuestro estudio. Independientemente de ello, siempre hemos empleado capítulos íntegros.

7.1.5.4. Estructuración del conocimiento

En todo trabajo terminológico es necesaria la estructuración conceptual del campo objeto de investigación previamente definido. La elaboración de la estructura conceptual de un glosario se puede sintetizar en los siguientes puntos (Cabré y Estopà 17):

- Toda área de especialidad transmite una determinada visión de la realidad.
- Un sistema conceptual está integrado por un conjunto estructurado de conceptos organizados en clases conceptuales.
- La estructuración de un vocabulario terminológico no debe confundirse con la estructuración conceptual de la materia.
- Todo trabajo terminológico representa una parcela de la estructura conceptual de una materia especializada.
- Una misma materia puede estructurarse desde puntos de vista diferentes.
- La estructuración conceptual de un glosario suele representarse mediante diagramas arbóreos.
- La estructuración conceptual debe ser equilibrada, detallada y completa en relación al espacio conceptual que pretende cubrir el glosario.
- La estructuración conceptual debe servir como mecanismo de control para:
 - La delimitación del tema y su alcance.
 - El control del vaciado.
 - La selección de términos.
 - La atribución de subárea temática a los términos.
 - La elaboración de definiciones.
 - El control de los descriptores clave.
 - La incorporación de los términos a una base de datos.
 - La recuperación automática posterior de la información.
 - El establecimiento de nuevos campos conceptuales afines.

Todo ello debe servir como evidencia para afirmar que cualquier glosario de un área de especialidad concreta será distinto a otro, dependiendo del enfoque sobre el que basemos nuestra compilación y de nuestro propio juicio y percepción de dicha realidad. Por otro lado, nuestro deseo ha sido el de mejorar la estructuración conceptual de los glosarios empleados como referencia. *Grosso modo*, su estructuración conceptual, una vez

revisada por medio de diagramas arbóreos, comienza por el término genérico, pero no deja conclusa las relaciones subordinadas, quedando incompletos algunos estratos inferiores de especificidad. Nuestro propósito ha sido el de paliar dicho vacío y completar todas las relaciones subordinadas que consideramos que no se encuentran clara o correctamente definidas.

7.1.5.5. Búsqueda y formato de textos

Nuestra búsqueda de textos se centró exclusivamente en la *web*. En este vasto repositorio de información, la variedad de textos y de configuraciones ha dificultado nuestro progreso, ya que no todos ellos son compatibles con su uso por medio de herramientas de lingüística de corpus y de extracción terminológica. Por tanto, fue necesaria una eliminación sistemática de todos los escritos de los que no ha sido posible su transformación a formato simple o sencillo.

Como ya hemos puntualizado, los escritos finalmente seleccionados pertenecen exclusivamente a la publicación denominada *Intelligence Bulletin*, editada por el Servicio de Inteligencia del Ejército norteamericano. Se trata de treinta y nueve números publicados entre septiembre de 1942 y junio de 1946. Se caracterizan por su sintaxis clara y descriptiva en la que se detallan las últimas tácticas e ingenios bélicos, empleados durante la franja temporal que nos ocupa, tanto por aliados como por enemigos. Otro de los motivos de dicha selección se debe a su grado de representatividad y de especialización, factores muy a tener en consideración, a tenor de los resultados poco esperanzadores obtenidos durante nuestra búsqueda de obras especializadas. A causa de la falta de material a disposición en *Internet* que poseyera una relevancia tal para ser empleado en nuestra investigación, tuvimos que otorgar especial importancia a la calidad de los textos en detrimento de la cantidad de éstos (Bowker y Pearson 48).

Ya habíamos destacado que el hecho de que ciertos números fueran editados una vez finalizada la contienda no implicaba dudar sobre su idoneidad. Los temas tratados en tales escritos siguen ubicándose en la misma franja espacial que nos ocupa. Además, la información que aparece en las últimas entregas era más profusa y detallada, motivada por el descubrimiento de archivos y de información clasificada acerca de nuevas invenciones

desconocidas hasta el momento de su descripción. De igual modo, con el cese de las acciones bélicas, el propósito de los autores se centró en ampliar información de temas previamente comentados o de hechos que precisaban un estudio más pormenorizado.

Por otro lado, estimamos conveniente resaltar el hecho de que, finalmente, conseguimos localizar la mayoría de los títulos de *Intelligence Bulletin* en su formato original en el sitio *web* titulado *Internet Archive*. Ya habíamos comentado que íbamos a eliminar los artículos directamente relacionados con las campañas en el Pacífico y las operaciones navales. No obstante, hemos constatado que no todas las secciones de los manuales han sido digitalizadas en la *web LoneSentry* (se trata de un sitio *web* especializado en donde es posible encontrar documentos impresos y gráficos de especial relevancia sobre la Segunda Guerra Mundial) perdiendo, con ello, una cierta dosis de información del todo relevante para nuestro estudio.

Ello no quiere decir que hayamos realizado el proceso de extracción terminológica de manera tradicional; nuestro interés ha tenido como resultado el estudio, por medio de la lectura, de ciertos capítulos de las publicaciones que no han sido incluidos en la página de *Internet* de donde los hemos extraído quedando, de ese modo, fuera de nuestro estudio. Debido a que nuestro corpus posee unas reducidas dimensiones (285.214 palabras), hemos considerado oportuno constatar que su grado de representatividad sea el adecuado para asegurarnos de que el número de unidades terminológicas que logremos extraer sea una cifra aceptable.

Los ejemplos que seguidamente mostramos en las figuras 3 y 4 proceden de la Biblioteca Virtual de Defensa. Con ellos queremos constatar que no todos los escritos que se pueden encontrar en *Internet* son adecuados para su empleo mediante herramientas de lingüística de corpus y de extracción terminológica.

No obstante, un cierto número de documentos que se encuentran en la red de redes en formato pdf no poseen tales deficiencias. Existen otros que son fáciles de convertir a texto plano o sencillo; la única dificultad consiste en eliminar los pies de página, las fotografías y los dibujos (la mayoría de ellos suprimidos durante la conversión) para que no queden insertados en la literatura.

DIARIO OFICIAL



DEL MINISTERIO DEL EJÉRCITO

ORDENES

Subsecretaría

Hoja de servicios

Las reales ordenes de 11 de enero de 1909 (C. L. núm. 17) y 30 de mayo de 1910 (C. L. núm. 78), en que se disponia que las hojas de servicios y de hechos de los jefes y oficiales destinados en comisión en Centros, Dependencias y en cualquier Cuerpo o Unidad del Ejército, radiquen en éstos, se considerarán ampliadas haciéndoles extensivas a todos aquellos que presten servicio como agregados.

Madrid, 24 de enero de 1941.

VARELA

CUERPO GENERAL DE LA ARMADA

Destinos

A propuesta del Ministro de Marina se designa para cubrir vacante de su empleo en el Consejo Supremo de Justicia Militar al capitán de corbeta (escala complementaria) D. Rafael Aguilar Ojeda:

Madrid, 27 de enero de 1941.

VARELA

Dirección General de Reclutamiento y Personal INFANTERIA

Destinos

Para cubrir las vacantes de provisión normal en Unidades de Infantería, anastadas por orden de fecha 14 del ac-

tual (D. O. núm. 11), pasan a servir los destinos que se indican, con arreglo a la orden de fecha 3 de julio último (D. O. núm. 148), los jefes y oficiales de dicha Arma que a continuación se relacionan, debiendo los que se encuentren destinados en las guarniciones de Baleares, Canarias y territorio de África, continuar en las mismas Unidades hasta la incorporación de los que han de sustituirles.

Destinados voluntarios

Teniente coronel D. Julió Rodríguez Gómez, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 85.

Otro, D. Valeriano Laclaustra Valdés, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 81.

Otro, D. José Miralles González, de disponible forzoso en la primera región militar, a la Caja de Recluta núm. 54.

Otro, D. Juan Selva Mergelina, de disponible forzoso en la tercera región militar, al regimiento de Infantería número 16.

Otro, D. Alejandro Álvarez López-Baños, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 8.

Otro, D. Carmelo Porqueras Bañeres, de disponible forzoso en la cuarta región militar, al regimiento de Infantería núm. 36.

Otro, D. Ramón Osende Fernández, de disponible forzoso en la octava región militar, al regimiento de Infantería núm. 57.

Otro, D. Víctor Flores Horraach, de la Caja de Reclutas núm. 71, al regimiento Infantería núm. 60.

Otro, D. José Imaz Echavarrri, del batallón de Infantería núm. 33, al regimiento de Infantería núm. 54.

Otro, D. Fernando Iturralde León, del Gobierno Militar de Burgos, al regimiento de Infantería núm. 22.

Teniente coronel habilitado (comandante efectivo) D. Antonio Pavón Ro-

dríguez, del regimiento de Infantería núm. 64, al regimiento de Infantería número 55.

Destinados forzosos

Teniente coronel D. José Villagrán Gancinotto, de disponible forzoso en la segunda región militar, al regimiento de Infantería núm. 47.

Otro, D. Fernando Alvarado Maldonado, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 19.

Otro, D. César Moneo Ranz, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 72.

Destinados voluntarios

Comandante D. Miguel Jareño Hernández-Vaquero, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 81.

Otro, D. Alejandro Rodríguez del Rivera, de disponible forzoso en la primera región militar, a la Plana Mayor de la Infantería de la 92 División.

Otro, D. Julio Esteban-Infantes Martín, de disponible forzoso en la primera región militar, al batallón Ciclista número 1.

Otro, D. Luis Gutiérrez Fernández, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 1.

Otro, D. Manuel Martínez Vivanco, de disponible forzoso en la primera región militar, a la Plana Mayor de la Infantería de la División núm. 102.

Otro, D. Moisés González-Besada Jiráldez, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 42.

Otro, D. Baltasar Castelló Agüera, de disponible forzoso en la segunda región militar, al regimiento de Infantería número 58.

Otro, D. Miguel Fernández Mucientes, de disponible forzoso en la primera región militar, al regimiento de Infantería núm. 43.

Otro, D. Emeterio Marcos Abarca,

Figura 3. Primera página del DIARIO OFICIAL DEL MINISTERIO DEL EJÉRCITO, de fecha 31 de enero de 1941 en su formato original.

Año LII.-Nilln. ~5 Viernes .
 .11 .de cuero .dc I').tI
 t
 Tonlo L-púg. 45i'

DIARIO OFICIAL
 DEL
 MINISTERIO DEL EJÉRCITO

,i' '
 ORDENES'
 Subsecretaría

tual '(6. O; núm. 1I). pasan a servir los
 destinos que se indican, con ar rculo va
 la orden de fecha .1' de julio último
 , (D, O, núm. ;-IS), los j,fes y oficiales,
 de dicha Arma que a continuación se relacionan.
 debiendo los que se cncuen-
 drienez del t'cg'ithiento de
 LIÚ~, ' (;4, al rC'g-imiento "lt'
 número ..S5.

l tcst intuios [orzosos

ln fal\te rí a
 Iníantcr in

Hoja de serv.icios

tren destinados en las guarniciones' (te
 ~ Baleares. Canarias v territorio de Afri-
 Teniente coronel D. José' Vülagrún
 Gancinotto de disponible forzoso en la
 L~" reales ordenes <le JI de enero de l')<>() fe. L.
 núm. ii) y 30 do mayo {le J?IO (e. L. núm. i;;;1,
 en (J\le se dispo-
 ca; contiuar. en lá,;, mismas Unidades hasta la
 incorporación dLO lo; que han
 de sustituirles,
 segunda región militar, al regimiento de
 Infautcria núm. 4i,
 Otro, D. Fernando Alvarado Xlaldo-
 lila (ill~ las hojas de servicios v de, he-
 'rhos de los j cíes')' oficia les destinados
 en comisión en Centros. Dependencias \'
 <':1 c\lalqt~i<,Cr uerpo o l!ilidad del Ejé;:_
 CItO. radiquen en ésto-, se considcrur áu
 an:,pliadas haciéllldol~;i' pxiensi\as a te-
 ,1,,; aquello, que presten .scrvcio CO!110
 a~rega;JG3. .' , , ,'

Figura 4. Encabezamiento y primer párrafo de la primera página del DIARIO OFICIAL DEL MINISTERIO DEL EJÉRCITO, de fecha 31, de enero de 1941, tras su conversión a formato pdf.

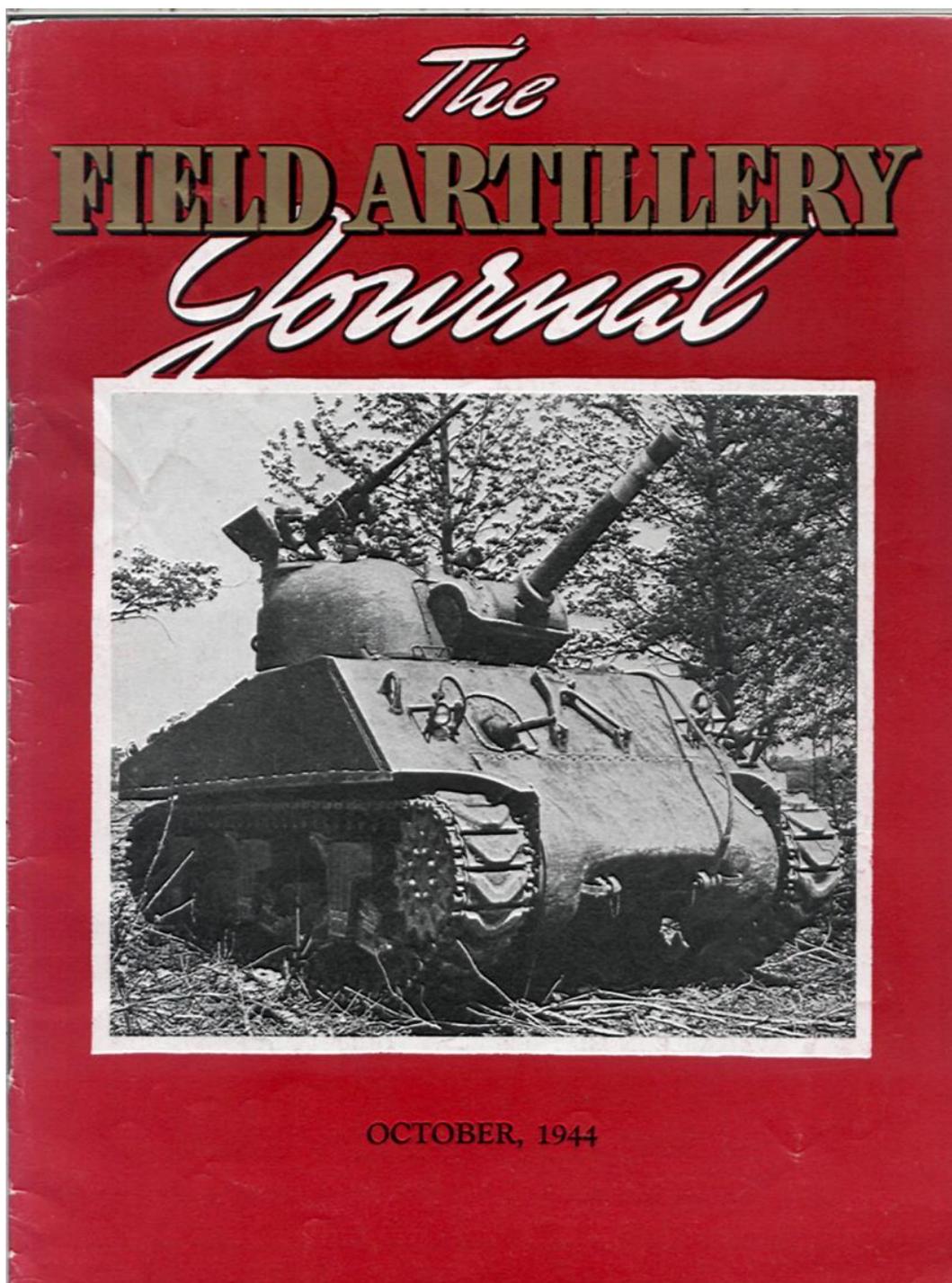


Figura 5. Portada original de *The Field Artillery Journal*, de diciembre de 1944.

NEW TIME FIRE TECHNIQUES

By Lt. C. A. Chapin, USMC, and Lt. H. W. Oliver, USMCR

EDITOR'S NOTE

As a result of research by Lieutenants Chapin and Oliver, the Marine Corps has adopted the methods here described. Due to the close association between Army and Marine artillery, field artillerymen should be familiar with this system. It is well to bear in mind that "complementary site" is misnamed: it is really "complementary elevation," realization of which makes this treatment of fuze setting appear most logical.

If time fire is to be most effective, it must be used in massed fire of battalions or larger units. With methods now in use corrections are determined and applied so that such massed fires can be executed with devastating effect. Much of our artillery, however, both Army and Marine Corps, is confronted with mountainous terrain in the several theaters of operation; and it must be noted that these massed time fires are accurately adjusted for height of burst only where the point of registration was at about the same altitude as the target. When considerable changes in angle of site are introduced an adjustment has to be made for each target unless a

Figura 6. Vista parcial del primer artículo del *The Field Artillery Journal* de diciembre de 1944 tras su conversión a formato plano.

ARMORED FORCES (German)

1. THE 88-MM. GUN

Field Marshal Erwin Rommel's use of the 88-mm. antiaircraft gun as an offensive antitank weapon in Libya has caused so much discussion among American soldiers everywhere that this seems a good time to describe it. In plain English, there is nothing strange or unusual about the 88. To the question "Is it vulnerable?" the answer is "Yes!"

Back in the 10-year period before Hitler came into power, the German 88-mm. antiaircraft gun was designed and built in secret. In those days the German Army was rigidly limited as to men and materiel. It is known that the designers of the gun were chiefly interested in constructing a double-purpose antiaircraft and antitank weapon. The news of the gun's antitank capabilities was not allowed to leak out, however, and not until the Nazis invaded Poland did the world discover what the German designers had perfected.

Figura 7. Encabezamiento del primer artículo de la publicación *Intelligence Bulletin* de septiembre de 1942, tal como aparece en el sitio *web* LoneSentry.

VOLUME 1 * NUMBER 1

**INTELLIGENCE
BULLETIN**

September 1942

**FILE COPY
C & GSS LIBRARY
PLEASE DO NOT REMOVE**

**MILITARY INTELLIGENCE SERVICE
WAR DEPARTMENT**

Figura 8. Portada del primer número de la publicación *Intelligence Bulletin*.

Los ejemplos presentados en las figuras 5 y 6 muestran cómo, en esta ocasión, un texto convertido a formato pdf editable es fácilmente transformado en tipo plano o sencillo. No obstante, esta publicación no ha sido empleada en nuestro trabajo al poseer un excesivo grado de especificidad del arma de artillería de campaña y al no disponer de todos sus números.

La figura 8 (véanse también las figuras 3 y 4) muestra un tipo de texto inapropiado para su conversión a pdf de ciertas obras (al contener, entre otras cosas, tachaduras, correcciones, manchas, etc.), haciendo necesario el uso únicamente de los tratados disponibles en la *web* en el formato que pudiera ser convertido con fiabilidad a configuración sencilla.

7.1.5.6. Vaciado

La fundamentación de todo trabajo terminológico consiste en el reconocimiento y delimitación de las unidades terminológicas a partir de un corpus textual. Si se emplea el método tradicional, se procederá a leer los textos que constituyen el corpus de vaciado y a marcar todos los segmentos que, debido a su competencia léxica y cognitiva especializada, se consideren candidatos a término del campo de especialidad en el que se trabaja. Si aplicamos el proceso de extracción terminológica automática, el programa de extracción se encargará de llevar a cabo dicha tarea de manera mecánica (Sager 155).

Durante esta fase se procedió a diferenciar entre unidad léxica y fraseológica, entre unidad léxica y terminológica, a determinar las estructuras morfosintácticas de los términos, el reconocimiento de la importancia cognitiva especializada para la identificación de las unidades terminológicas, la pertinencia de los términos en función de la estructuración conceptual elaborada y el hecho de que no todos los términos que aparecen en los textos de vaciado son pertinentes para su inclusión en el glosario que se desea compilar.

7.1.5.7. Elaboración del fichero de vaciado

Una vez seleccionadas las unidades terminológicas que deben formar parte del glosario, éstas deben registrarse en una ficha de vaciado manual o electrónica (en nuestro caso, esta última). Dichas fichas se elaboran siguiendo una serie de convenciones aunque, en términos generales, suelen comprender los siguientes campos (Cabré y Estopà 19):

- La unidad terminológica o entrada.
- El subárea temática.
- La categoría gramatical.
- La definición.
- La fuente donde aparece la definición (glosario o documento de origen).
- La abreviatura (si procede).
- El acrónimo (si procede).
- El contexto.
- La fuente (archivo) de donde procede el ejemplo en contexto.
- El número de ejemplos en contexto.
- Información adicional (si procede).
- La fecha de elaboración de la ficha.

No obstante, cuando una unidad terminológica detectaba más de una equivalencia, procedimos a dar su definición en la entrada más empleada o en la que ha sido normalizada. Por ello, solo el término principal posee una definición, apareciendo en las fichas de cada uno de sus sinónimos la leyenda *véase*, seguida de la denominación de la unidad terminológica principal. Con ello tratamos de no repetir innecesariamente las explicaciones y de otorgarle la importancia que merece a la unidad terminológica principal. No obstante, y para que un sinónimo fuera incluido en el glosario, debía encontrarse alojado en nuestro corpus en frases en contexto, de lo contrario no ostentaría una entrada independiente en el glosario.

En lo que concierne a la estructuración del conocimiento, concebimos éste como un espacio multidimensional entrecruzado por ejes que representan diferentes características sobre las que se agrupan las unidades conceptuales, las cuales se identifican y describen con referencia a las coordenadas que ocupan en cada uno de los ejes. Una ordenación

taxonómica puede fundamentarse en los valores que posee un concepto con respecto a cada uno de los ejes, o de sus propias características, lo cual equivaldría a definir su posición dentro del espacio conceptual. No obstante, también es plausible reflexionar por medio de clases o subáreas, en donde las unidades conceptuales se agrupan dentro en un mismo subespacio (Sager 15). Partiendo de dicha premisa, hemos considerado oportuno agrupar las unidades terminológicas en torno a las siguientes subáreas temáticas:

- Organización.
- Instrucción y Adoctrinamiento.
- Tácticas y Estrategia.
- Logística y Construcción.
- Armamento.
- Vehículos.
- Batallas, Campañas y Personajes.
- Interdisciplinario.

TERM: SUBDOMAIN: PART OF SPEECH: WORDFORMS: DEFINITION: DEFINITION SOURCE: RELATED TERMS: ABBREVIATION: ACRONYM: SYNONYMS: CONTEXT SENTENCE: FILENAME: FILENAME SOURCE: NOTE: NUMBER OF HITS: DATE ENTERED:

Figura 9. Ejemplo de ficha terminológica.

Debido a que la terminología se caracteriza por su carácter multidisciplinar, añadimos el subdominio *Interdisciplinario* para incluir unidades terminológicas empleadas también en otras disciplinas y que no constituyen una característica única y unívoca de este campo

del saber. Por otro lado, y motivado por el carácter histórico de la investigación, incluimos el subárea *Batallas, Campañas & Personajes* para constatar la importancia que ciertos hitos han tenido y tienen en el seno del campo militar.

Cabe recalcar que el fichero de vaciado supuso el punto de partida para la elaboración de las fichas terminológicas que conforman nuestro glosario, y que su composición varía dependiendo del trabajo a realizar y de los objetivos a cumplir. Hemos empleado el modelo de ficha terminológica descrito con anterioridad, del que hemos eliminado los campos que no se pudieron cumplimentar, ya fuere por falta de información o por su falta de pertinencia.

7.2. METODOLOGÍA Y HERRAMIENTAS EMPLEADAS

En esta sección detallaremos nuestro *modus operandi* mediante el empleo de diferentes técnicas de extracción terminológica.

En primer lugar, destacar que, en ciertas ocasiones, hemos empleado el método tradicional de búsqueda de terminología. Es innegable que a veces resulta necesario leer un párrafo con detenimiento para obtener una perspectiva más clara acerca de la idoneidad de incluir o no un candidato a término en el archivo de vaciado terminológico. No obstante, no se trata de una labor sistemática, sino auxiliar para ampliar la información concerniente a las unidades terminológicas objeto de búsqueda.

En segundo lugar, cabe precisar que hemos utilizado dos tipos de herramientas informáticas para el proceso semiautomático de extracción terminológica: una fundamentada en la estadística y otra en la lingüística. En este apartado analizaremos su funcionamiento, la metodología empleada y los resultados obtenidos en nuestra investigación.

7.2.1. Extracción terminológica estadística

Una vez recopilados la mayor parte de los textos, se llevó a cabo el proceso de conversión de éstos a texto plano, que es el formato empleado a la hora de realizar trabajos de extracción de terminología y análisis de lingüística de corpus. El programa informático que utilizamos para la búsqueda de términos fue SDL MultiTerm 2009 Extract. Se trata de un programa de extracción terminológica que forma parte del paquete de traducción asistida por ordenador SDL Trados Studio 2009¹, la herramienta comercial más utilizada por los traductores y mejor valorada por las empresas de traducción.

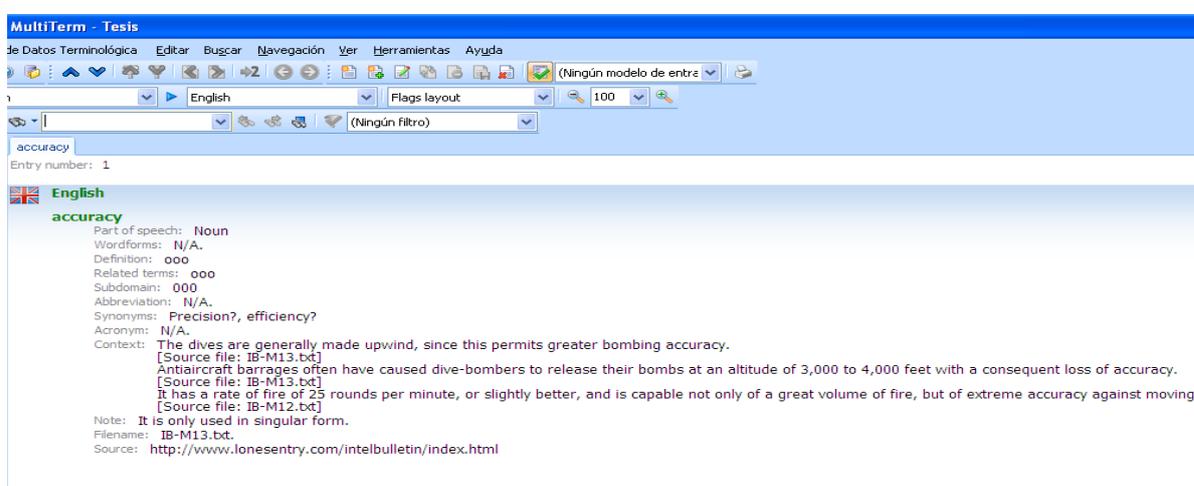


Figura 10. Ejemplo de ficha terminológica de SDL MultiTerm 2009.

7.2.1.1. Fase I: Configuración del programa

SDL MultiTerm 2009 Extract está concebido para realizar la extracción mediante fórmulas estadísticas, que no lingüísticas. Para la búsqueda sobre base lingüística hemos empleado AntConc 3.5.5.w (Windows), que será detallado en el apartado 7.2.2., dedicado a este tipo de método de extracción. SDL MultiTerm 2009 Extract permite acometer indagaciones monolingües y bilingües y posibilita la exportación de los datos obtenidos a SDL MultiTerm 2009. Se trata esta última de una base de datos terminológica informatizada que permite archivar glosarios con la posibilidad de modificar los diferentes campos que componen las fichas terminológicas atendiendo a las necesidades particulares

¹ Para más información, véase <http://www.translationzone.com>

de cada traductor. Así mismo, es posible su uso durante el proceso de traducción para un acceso rápido a las unidades terminológicas.

Procedimos a configurar el proyecto monolingüe en inglés norteamericano aprobando que un término pudiera estar compuesto de hasta por diez palabras, no delimitando el número máximo de ejemplos de unidades terminológicas que fuera posible extraer de los textos; situando la relación entre silencio y ruido al cincuenta por ciento y acordando que el número de términos extraídos no tuviera límite (100 por defecto). Una vez definidos los parámetros, procedimos a aceptar el empleo de lo que SDL Multiterm 2009 Extract denomina “archivo de exclusión de términos”. Se trata de un archivo en formato de texto en donde aparecen las palabras y verbos de uso más común del lenguaje general del inglés americano, cuyo propósito es la eliminación, en la medida de lo posible, de ruido durante la extracción. No obstante, cabe precisar que fue necesaria su revisión, ya que existen unidades léxicas que en el ámbito militar poseen la particularidad de emplearse también como unidades terminológicas debido a su adecuación a la economía lingüística (Schmitz 39). Posteriormente, mostraremos dos ejemplos para denotar la importancia de dicha verificación.

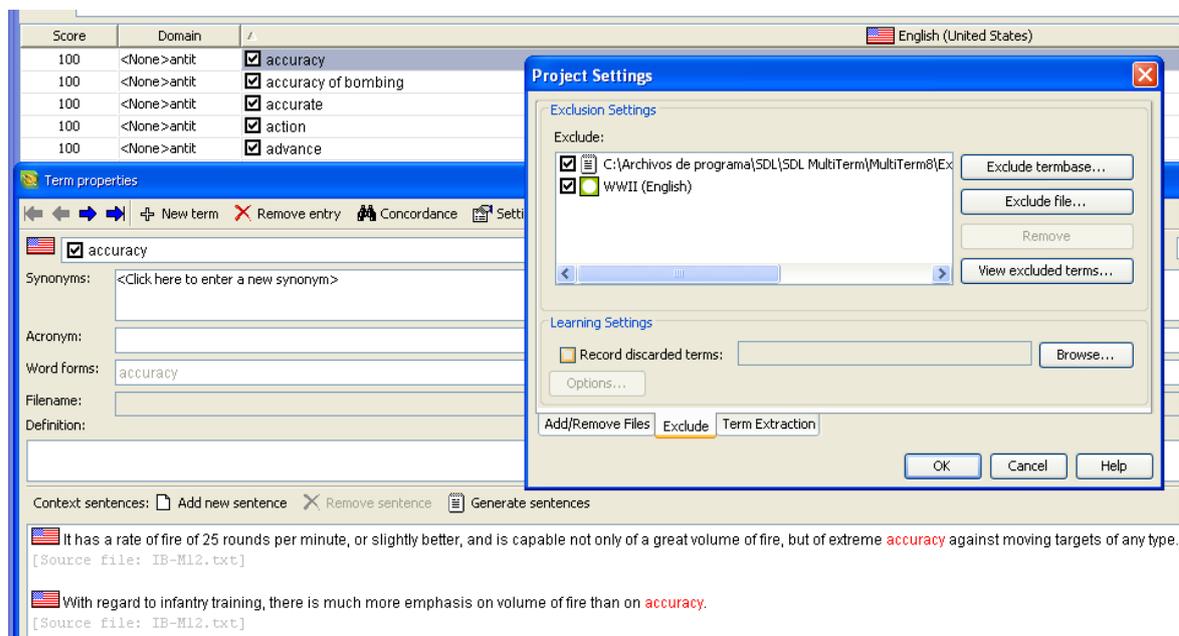


Figura 11. Configuración del proyecto y activación del archivo de palabras de exclusión y la base de datos terminológica.

De igual modo, activamos el uso del denominado “listado de palabras vacías” (*Stopword Lists*) en inglés americano. Dicho listado está conformado por palabras funcionales tales como artículos, pronombres y adverbios que no suelen convertirse o emplearse en la formulación de unidades terminológicas complejas, aunque siempre existen excepciones.

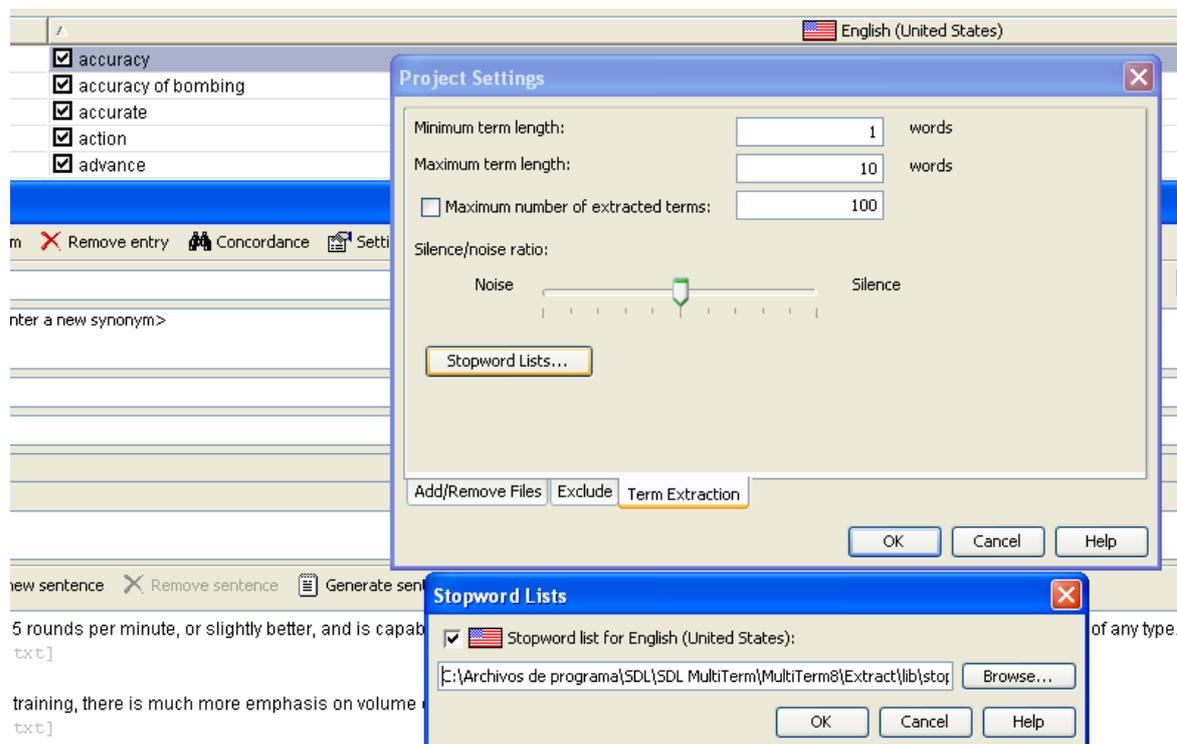


Figura 12. Selección de parámetros antes de la ejecución del proceso de extracción terminológica.

7.2.1.2. Fase II: Primer análisis

Hemos procedido a acometer el proceso de extracción terminológica en varias etapas para, de esa manera, familiarizarnos con el programa SDL Multiterm 2009 Extract. La inclusión de textos en el trabajo de extracción se ha llevado a cabo en cuatro fases en las que hemos apreciado una serie de insuficiencias que padece el citado *software* y que no habríamos detectado si hubiéramos añadido todos los textos en una sola ocasión. Bien es cierto que la inclusión al unísono de todos los escritos que conforman nuestro corpus textual nos hubiera ahorrado tiempo, aunque es bastante probable que nos hubiera sido

difícil identificar el motivo (entre otras cosas) de la existencia de un alto porcentaje de ruido y de silencio.

Una vez configurados todos los parámetros, procedimos a examinar el primer texto, compuesto por 7.599 palabras. Tras su análisis, el programa informático nos mostró una serie de unidades léxicas y unidades fraseológicas consideradas candidatas a término. Estos candidatos conformaron el archivo de vaciado terminológico. Partiendo de este repositorio, iniciamos el análisis detallado de dichos candidatos, siendo posteriormente catalogados como candidatos a término, unidades terminológicas o unidades lexicológicas (Bowker y Pearson 149-51).

No obstante, el conocimiento que un investigador posee del campo de especialidad en el que lleva a cabo su investigación no resulta del todo concluyente para validar o rechazar un candidato a término, por lo que, en nuestro caso, fue necesario analizarlo en su contexto y, no menos importante, corroborar su categoría como unidad terminológica por medio de glosarios editados en la franja temporal que nos ocupaba o consultando a un experto. Sin embargo, y debido a la continua evolución que sufren las diferentes disciplinas, todo ello no ha sido suficiente, ya que hemos podido apreciar la existencia de unidades terminológicas desconocidas para nosotros hasta el momento y que no aparecían en ninguno de los glosarios consultados. En no pocas ocasiones hemos localizado nuevos términos dentro de las obras objeto de nuestro análisis e, incluso, hasta su propia definición, hecho que nos confirmó el descubrimiento de un nuevo término (esta tarea se detallará más a fondo en el proceso de análisis lingüístico), cuya definición ha sido, en ocasiones, registrada en obras editadas con posterioridad.

Mediante un estudio sistemático y pormenorizado de cada entrada, procedimos a completar los diferentes campos que nos proporciona el programa: término, sinónimos, acrónimo, formas nominales, frases en contexto, nombre del archivo origen, definición y notas. En otro orden de cosas, y para llevar a cabo un recuento riguroso de los candidatos a término, todos ellos fueron adscritos al dominio *Military* hasta que se les fue reconocido su estatus de unidad terminológica. Tras ello, pasaron a formar parte del dominio *<None>antit*, hasta su inclusión en el glosario, hecho que se verá reflejado al rellenar el recuadro que aparece a la izquierda del término.

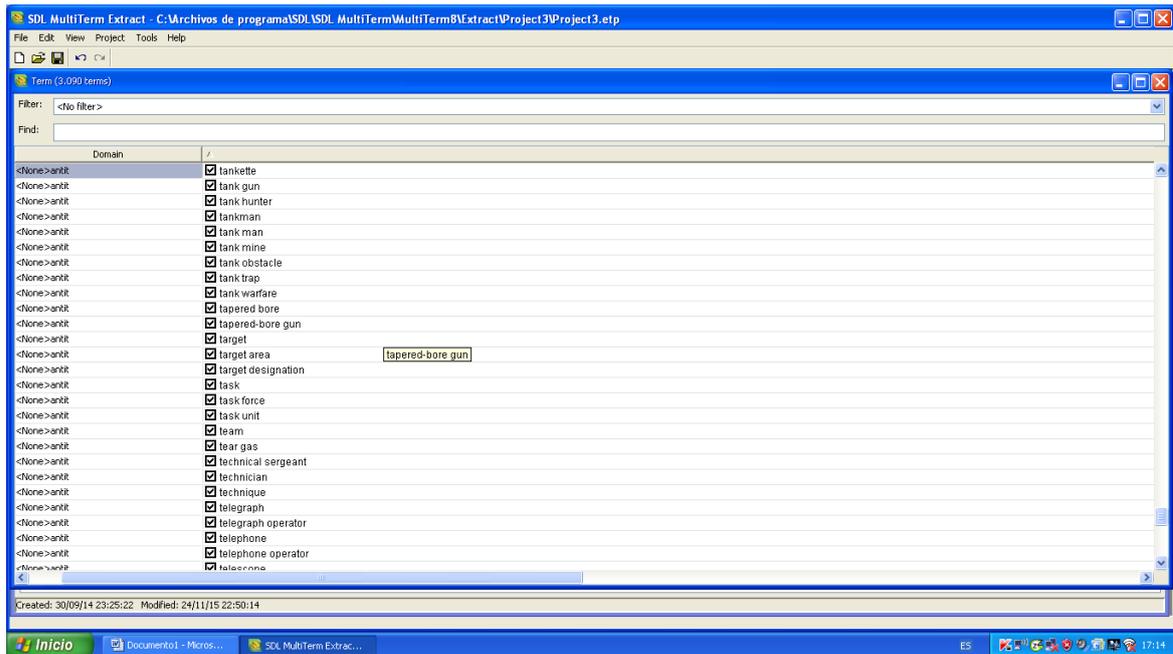


Figura 13. Fichero de vaciado terminológico.

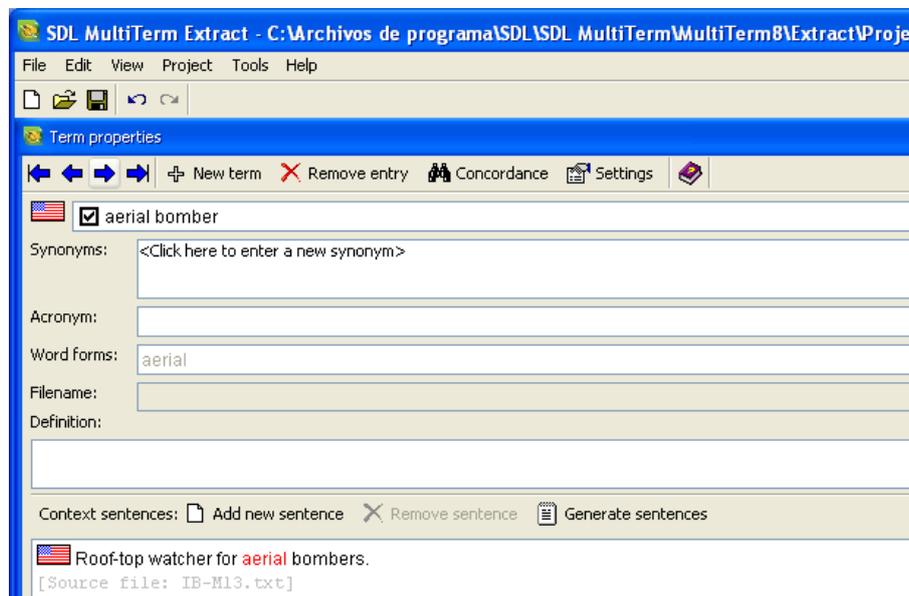


Figura 14. Ejemplo de ficha terminológica.

El procedimiento típico fue el de identificar dicha información de manera paulatina, ya que resultaba muy improbable que en una primera búsqueda se pudieran detectar todos los datos requeridos (Sager 140). Por otro lado, estimamos que la subdivisión que nos proporcionaba el programa informático no era todo lo amplia que deseábamos. En

consecuencia, y una vez concluido el proceso de extracción terminológica, procedimos a realizar un nuevo fichero terminológico en Microsoft Word, *software* que nos permitió incluir fichas con los campos detallados anteriormente en la sección 7.1.5.7. Además, dicho programa fue el seleccionado para editar nuestro glosario, debido a su facilidad de manejo y adaptabilidad.

Tras agregar el segundo y tercer textos (19.319 palabras en total), y acometer su consecuente análisis, alcanzamos la cifra de 661 candidatos a término. Ello nos indicó que (como ya hemos apuntado con anterioridad) el uso de textos representativos es de vital importancia para el trabajo de extracción terminológica. No obstante, hemos de hacer dos matizaciones: un diez por ciento de dichos posibles términos fueron excluidos al no obtener una adecuada definición y/o ejemplos en contexto, y fue necesaria la adición y eliminación de ciertos vocablos que pertenecían, simultáneamente, al lenguaje general, a otros lenguajes de especialidad y al ámbito castrense. Así mismo, la aparición de nuevos candidatos a término fue decreciendo y, a medida que se fueron añadiendo nuevos textos, su identificación se convirtió en una tarea más compleja.

El trabajo continuó con la adición de tres nuevos escritos para su consiguiente análisis (43.975 palabras en total).

Seguidamente, estimamos oportuno alimentar el programa de extracción terminológica con los treinta y tres textos restantes que teníamos preparados en nuestra investigación que, sumados a los seis anteriores, conformaron un total de treinta y nueve.

En lo que se refiere al proceso de extracción terminológica, SDL MultiTerm 2009 Extract seleccionó un total de 5.314 términos y candidatos a término, entre los anteriormente validados y los incorporados de los nuevos textos. No obstante, el programa volvió a eliminar del fichero de vaciado terminológico todas las voces que se encontraban almacenadas en su fichero de exclusión de términos, a pesar de que ya estaban catalogadas como unidades terminológicas dentro del dominio (a saber: subdominio) <None>*antit* y no en el de *Military*, al que pertenecían los candidatos a término antes de su correspondiente estudio. En consecuencia, y con el propósito de no duplicar el trabajo, éstos no fueron recuperados hasta que dimos por finalizado el estudio.

Concerniente a las nuevas entradas, más del cincuenta por ciento fueron eliminadas con rapidez, al ser notoria su pertenencia al lenguaje general, mientras que aproximadamente un veinticinco por ciento fueron incluidas como unidades terminológicas, quedando el veinticinco por ciento restante almacenadas para un estudio más detallado con el objeto de averiguar su pertenencia al LGP, al LSP objeto de análisis, o a un LSP afín.

Como muestra de nuestro análisis, obsérvese a continuación que en la figura 15 apreciamos que, en el caso del candidato a término *balloon barrage*, pudimos aportar una definición, describir su fuente e incluir un sinónimo, al igual que dos muestras de su uso.

Ya hemos aludido la necesidad de revisar el archivo de exclusión de términos. En particular, en el caso de algunas unidades terminológicas complejas, ciertas (o todas) de sus piezas léxicas que la conforman se encontraban incluidas en el mencionado fichero. Tomemos como ejemplos los sustantivos *action* y *company*. Al consultar el mencionado archivo hallamos que ambos vocablos se encontraban ahí y, por tanto, no aparecieron en el listado inicial de candidatos a término.

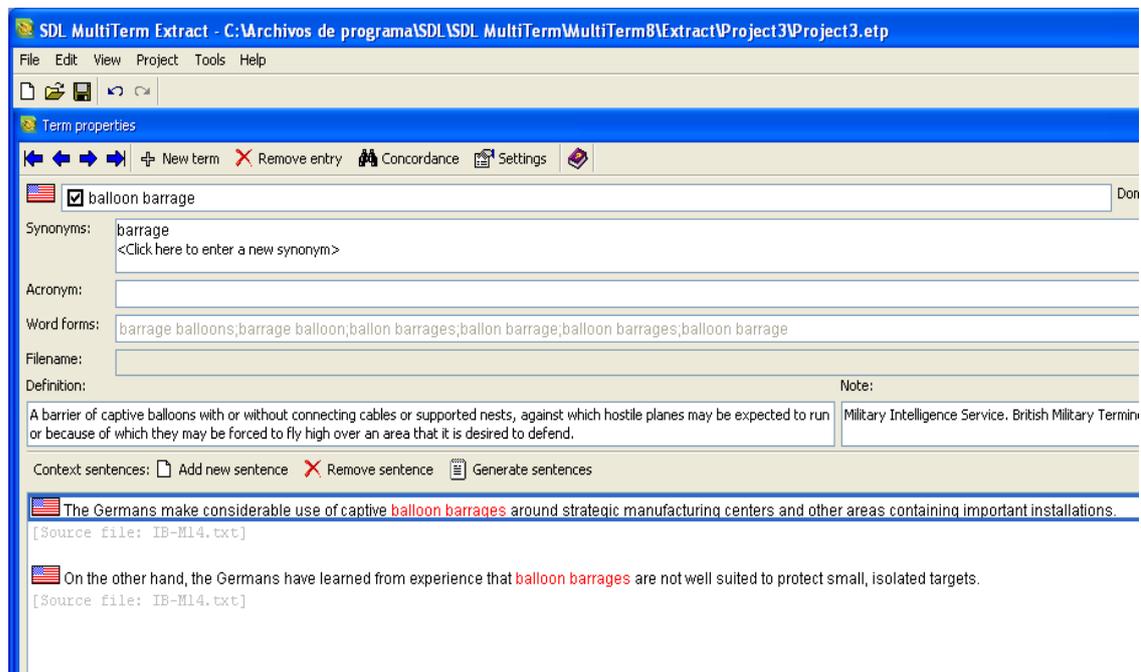


Figura 15. Ejemplo de *balloon barrage*.

Obsérvese la aparición en la figura 16 de la unidad lexicológica *action* (resaltada en azul) en el listado de exclusión de términos. Todo ello hizo necesario una continua revisión del mencionado fichero con el propósito de no pasar por alto unidades conceptuales referentes a nuestro campo de especialidad cuyas representaciones léxicas también se emplean en el lenguaje general.

```
Excluded Terms

Created On: sábado 27 de septiembre de 2014 17:55:28

Exclusion Sources:

  • Exclusion List: C:\Archivos de programa\SDL\SDL MultiTerm\MultiTerm8\Extractlib\BasicVo

Exclusion List: C:\Archivos de programa\SDL\SDL MultiTerm\MultiTerm8\Extractl

1.224 terms excluded

abandon
ability
absolute
absorb
accompany
accomplish
accord
account
accuracy
accurate
achieve
acknowledge
across
act
action
...
```

Figura 16. Vista parcial del archivo de términos de exclusión en formato html.

En consecuencia, añadimos dichas voces a nuestro fichero de vaciado terminológico y realizamos una exploración de palabras en contexto. Este proceso fue el primer paso para considerar si una voz poseía el rango de término. Seguidamente, si las frases en las que aparecía una unidad léxica dada nos indicaban la posibilidad de que nos encontráramos ante un posible vocablo especializado, proseguíamos con su búsqueda en glosarios editados en dicha época y registrábamos su entrada y definición (si existían), validándolos e incorporándolos a nuestro fichero de vaciado. Como ya hemos señalado en más de una ocasión, hallamos unidades terminológicas que, aun no siendo descritas en los glosarios editados en el periodo de tiempo que nos ha ocupado, sí habían sido incluidas en otros publicados a finales del siglo XX y principios del siglo XXI.

A continuación, mostramos a modo de ejemplo, las unidades terminológicas *action* y *company* citadas con su definición y una frase en contexto:

TERM: **action**

SUBDOMAIN: Tactics & Strategy / Ordnance

PART OF SPEECH: noun

WORDFORMS: action, actions

DEFINITION: 1. An engagement or battle, usually one on a small scale. 2. The mechanism of a firearm involved in presenting the cartridge for firing and in ejecting the spent case and introducing a fresh cartridge.

DEFINITION SOURCE: (1) United States, War Department. *British Military Terminology*. Washington: GPO, 1943. Print. (2) McNab, Chris. *The Great Book of Guns*. London: Salamander, 2004. Print.

CONTEXT SENTENCE: (1) *When an attack is in preparation, orders as a rule will be issued first to the heavy weapons, so that the attack will not be delayed while the heavy weapons are getting ready to come into action.* (2) *There are very few moving parts in the rifle. The bolt operates straight to the rear, and has no turning action. The locking action takes place when the locking lugs on the rear of the bolt are forced down into a locking recess by the bolt housing.*

FILENAME: IB-M14.txt, IB-M48.txt

NUMBER OF HITS: N/A

DATE ENTERED: 14/7/2014

TERM: **company**

SUBDOMAIN: Organization

PART OF SPEECH: noun

WORDFORMS: company, companies

DEFINITION: Basic administrative and tactical unit of most branches of the military service, larger than a platoon, smaller than a battalion. A company is equivalent to a battery of artillery, a troop of cavalry, or an aviation squadron, and is usually commanded by a captain.

DEFINITION SOURCE: United States, War Department. *Dictionary of United States Army Terms*. Washington: GPO, 1944. Print.

ABBREVIATION: co (AE), coy (BE)

CONTEXT SENTENCE: *For the battle of Berlin, the Russians organized combined assault detachments, consisting of one tank battalion, a rifle battalion, a company or*

platoon of engineers, a battalion of artillery (not less than 122-millimeter), and a platoon of flame throwers.

FILENAME: IB-M49.txt

NUMBER OF HITS: N/A

DATE ENTERED: 8/8/2014

<p>TERM: arm of service SUBDOMAIN: Organization PART OF SPEECH: noun WORDFORMS: arm of service DEFINITION: See branch (2) DEFINITION SOURCE: ABBREVIATION: ACRONYM: SYNONYMS: (2) arm of the service, (2) branch CONTEXT: <i>They select especially promising officers and men for this arm of service, give them the most thorough training, and provide them with the best possible weapons.</i> FILENAME: IB-M13.txt. NOTE: NUMBER OF HITS: 2 DATE ENTERED: 14/11/2014</p>

Figura 17. Ejemplo de ficha terminológica con todos los campos establecidos.

Sirva ello como evidencia para constatar la necesidad de revisar cuidadosamente el archivo de exclusión de términos y así evitar la pérdida de información relevante.

Con referencia a la posibilidad de usar una base de datos terminológica diseñada previamente o a la creación de una nueva con SDL MultiTerm 2009 Extract, hemos de constatar finalmente que los formatos de los archivos de éste y de SDL MultiTerm 2009 son diferentes (exd y sdltb respectivamente) y que son incompatibles, imposibilitando la exportación de los datos. Éste ha sido otro de los motivos por los que tomamos la decisión de transferir las entradas a Microsoft Word con el objeto no de perder información y de facilitar el proceso de la elaboración de las fichas terminológicas.

De igual modo, y tal y como habíamos destacado, el glosario también ha sido elaborado mediante este programa informático al permitirnos incorporar todos los campos que consideramos oportunos para nuestro trabajo.

7.2.1.3. Fase III: Resolución de problemas

En este apartado detallaremos el tratamiento otorgado a las unidades terminológicas que carecían de definición o de ejemplos de su uso, su consiguiente análisis y las medidas que se adoptaron en cada caso.

- **Unidad terminológica *a posteriori*.** Ocurre cuando identificábamos una unidad terminológica compleja perteneciente a nuestro dominio objeto de estudio de la que no encontrábamos definición en ninguno de nuestros glosarios de consulta. Para resolver la situación (reseñada con anterioridad), buscábamos su descripción en algunos de los glosarios editados tras la segunda conflagración mundial. En esta ocasión, ponemos como ejemplo *dual-purpose weapon*:

TERM: **dual-purpose weapon**

SUBDOMAIN: Ordnance

PART OF SPEECH: noun

WORDFORMS: dual-purpose weapon, dual-purpose weapons

DEFINITION: A weapon designed for delivering effective fire against air or surface targets.

DEFINITION SOURCE: United States, Department of Defense. Dictionary of Military and Associated Terms. Washington: Department of Defense, 2005. Print.

SYNONYMS: double-purpose weapon

CONTEXT SENTENCE: *The high velocity of anti-aircraft guns makes them suitable for anti-tank missions, and, since 1940, German designers have paid special attention to the possibility that any German anti-aircraft gun may be used as a dual-purpose weapon. The smaller guns --the 20-mm 2-cm Flak 30 and 38, and the 37-mm 3.7-cm Flak 18 and 36-- are now of little value in an anti-tank role because of their lack of power.*

FILENAME: IB-M37.txt

NUMBER OF HITS: 1

DATE ENTERED: 13/7/2014

- **Candidatos a término no incluidos en glosarios.** Otro caso diferente ocurría cuando aparecía un candidato a término del que no disponíamos de su definición y no se encontraba en ninguno de los glosarios consultados; ni siquiera en los editados después de la guerra. Ante tal hecho, iniciamos su búsqueda en algunos de los diccionarios monolingües y enciclopédicos a disposición libre en *Internet*. Por un lado, estos recursos electrónicos nos sirvieron a modo de corpus monitorizado o de prueba. Si en alguno de esos repertorios aparecía su definición, ello podría indicarnos lo siguiente:

- Un candidato a término es, en realidad, una unidad léxica.
- Una unidad terminológica ha dejado de emplearse en el LSP objeto de estudio y se emplea en el lenguaje cotidiano (*de-terminologization*) (Bowker y Pearson 230).
- Un candidato a término ha obtenido su condición con posterioridad, aunque no ha sido normalizado.
- Un candidato a término ha obtenido finalmente dicha condición, siendo incluido en subsecuentes glosarios.

Los candidatos a término fueron automáticamente descartados si se catalogaban como unidades lexicológicas. No obstante, en los otros tres casos, fue necesario conocer si durante los años cuarenta el candidato a término había sido reconocido como unidad terminológica. Si ello hubiera sido así o hubiera obtenido la categoría de término *a posteriori*, quedó finalmente representado en el glosario. A continuación mostramos un ejemplo:

TERM: **G.I.**

SUBDOMAIN: Organization

PART OF SPEECH: acronym

WORDFORMS: G.I.

DEFINITION: A private soldier in the US Army.

DEFINITION SOURCE: "Oxford Dictionaries - Dictionary, Thesaurus, & Grammar."
Oxford University Press, n.d. Web. 23 Oct. 2014.

SYNONYMS: galvanized iron, general issue, Government Issue

CONTEXT SENTENCE: *Though many an American G.I. will grunt derisively when told that other people walk more than he does, it is true that the Soviet infantryman must depend upon his feet for much of his transportation.*

FILENAME: IB-M49.txt

NOTE: The context sentences confirm that the term is also used to denote a private soldier in any army.

NUMBER OF HITS: 5

DATE ENTERED: 23/10/2014

En este caso, se trata de un acrónimo que no se encuentra incluido en los glosarios empleados como referencia, al ser considerado un coloquialismo. No obstante, y a pesar de estar descrito en un diccionario del lenguaje general de la lengua inglesa, su empleo abundante durante la contienda (incluso para denotar a soldados de otros ejércitos) nos indicaba su idoneidad para incorporarlo a nuestro trabajo. La aparición de un vocablo sujeto a nuestro estudio en un diccionario de LGP no indicaba, necesariamente, su consideración de unidad terminológica, sino que sugería la necesidad de una valoración pormenorizada para considerar qué estatus se le debía otorgar durante el periodo temporal sujeto a análisis.

- **Candidatos a término con definición por implicación.** Otro hecho a resaltar es la identificación de candidatos a término cuya definición (independientemente del método empleado para su formulación) fue posible hallar en los textos empleados para la compilación del corpus:

TERM: **additional armor**

SUBDOMAIN: Vehicles

PART OF SPEECH: noun

WORDFORMS: additional armor

DEFINITION: Additional armor plates provided to armored vehicles to reduce the penetration of hollow-charge shells.

DEFINITION SOURCE: IB-M28.txt.

SYNONYMS: additional armour (BE)

CONTEXT SENTENCE: *By the summer of 1941, the weight of the Pz.Kw. 3 had already been increased by additional armor, and it must have been clear to the Germans that*

future developments in armor and armament would necessarily involve still further increases in the weight of this tank.

FILENAME: IB-M24.txt

NUMBER OF HITS: 4

DATE ENTERED: 10/8/2014

En este caso, se trata de una definición por implicación (Sager 43). Sea como fuere, la mayoría de los candidatos extraídos automáticamente, y de los que fue posible facilitar su definición gracias a la exploración del contexto, fueron incluidos en nuestro glosario.

- **Candidatos a término sin definición.** Es necesario mencionar aquellas unidades almacenadas en nuestro fichero de vaciado terminológico a las que no fue posible encontrar una definición. Para las que, después de consultar a los expertos en la materia, se consideró acertada su inclusión en el glosario, se le procuró una definición a modo de propuesta, y siempre tras su aceptación por parte de los especialistas (Bowker y Pearson 155). En tales definiciones se procedió generalmente a proporcionar las relaciones genéricas propias de los términos que se encontraban en las ramas superiores de la estructuración del conocimiento al que pertenecían, incluyendo los suyos propios situados a un estadio inferior de su estructuración arbórea. En consecuencia, el grado de especificidad aumentaba, al igual que su intensidad (Sager 31).

TERM: **field howitzer**

SUBDOMAIN: Ordnance

PART OF SPEECH: noun

WORDFORMS: field howitzer, field howitzers

DEFINITION: Field artillery piece; howitzer mounted on a carriage for use in the field.

DEFINITION SOURCE: Author's proposal.

CONTEXT SENTENCE: *Hollow-charge shells have been furnished for standard infantry guns and artillery of German divisions of all types—in particular, the 75-mm and 150-mm infantry howitzers, the 105-mm and 150-mm field howitzers, and the 105-mm guns.*

FILENAME: IB-M37.txt

NOTE: The definition has been based on that of **field gun**.

NUMBER OF HITS: 3

DATE ENTERED: 24/11/2014

En el ejemplo previamente mostrado se le ha atribuido la condición de pieza de artillería de campaña, adoptando la condición de los términos que se encontraban en el plano superior de la estructura conceptual de *piezas de artillería* y de *artillería de campaña*. En consecuencia, y siempre dentro de la perspectiva onomasiológica, el método empleado para su descripción se puede denominar de *abajo a arriba*.

Todo lo detallado hasta el momento ha dado como resultado la corroboración (entre otras cosas) de que los programas informáticos no son infalibles (Cabré, Estopà y Vivaldi 82), o que deberían denominarse de *extracción terminológica semiautomática o asistida por ordenador* (Bowker y Pearson 165).

Debido a lo expuesto hasta ahora, en ocasiones sopesamos la posibilidad de emplear la antigua técnica de revisión tradicional de los documentos para constatar que se había obtenido el máximo partido de cada uno de ellos. No obstante, y gracias a las herramientas de extracción terminológica y de lingüística de corpus, ello no resultó necesario. Bastó con realizar una rápida lectura por medio del programa AntConc 3.3.5.w (práctica que detallaremos con posterioridad), con el propósito de identificar los posibles candidatos a término, para proceder, a continuación, a corroborar su estatus con SDL Multiterm 2009 Extract mediante el estudio del candidato en su contexto.

• **Ejemplos procedentes del LGP y del LSP.** Finalmente, el impedimento a solventar resultó cuando para una unidad terminológica las frases que encontrábamos en contexto procedían del LGP y no tenían relación alguna con la definición otorgada para su uso en el seno de esta disciplina. En ese caso, mantuvimos la entrada y su definición a la espera de que aparecieran frases adecuadas en futuros textos a incluir. Ello implicó que a la hora de detallar el número de coincidencias de una unidad terminológica, podríamos indicar una cifra propensa a equívocos, puesto que ésta incluía tanto los casos procedentes del LGP como del LSP objeto de nuestra investigación. Por tanto, fue necesario distinguir entre los ejemplos en los que aparecía la unidad terminológica con su correspondiente significación, de los que procedían del lenguaje común. Cuando ello no fue posible, procedimos a reflejarlo al usar el acrónimo N/A.

El ejemplo mostrado en la figura 18 indica la aportación de frases en contexto procedentes del LGP para el término *firmes* en inglés. Fue necesario tratar estos casos con

cautela y verificar todos los ejemplos en contexto antes de su incorporación a una ficha terminológica.

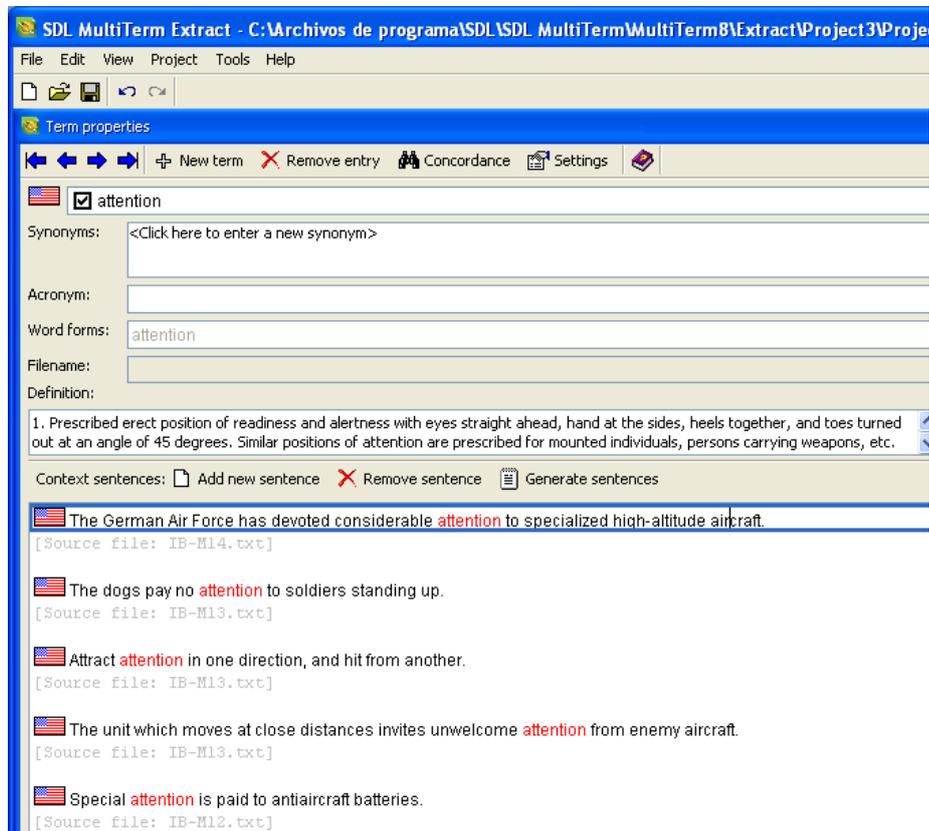


Figura 18. Ejemplo de desambiguación.

Una vez finalizado el proceso de extracción, se añadieron automáticamente las nuevas entradas a nuestro archivo de vaciado terminológico. En éste se encontraban los sancionados con anterioridad, conjuntamente con los susceptibles a estudiar, los cuales poseían la casilla de validación en blanco. Algunos de los nuevos vocablos fueron eliminados con celeridad al provenir del lenguaje común; otros fueron sometidos a estudio para su posible inclusión, mientras que un tercer grupo (alrededor de un veinte por ciento) fue añadido al fichero tras proceder a corroborar su condición de término por medio de frases en contexto, y a buscar su consiguiente definición.

En el ejemplo mostrado previamente aparecen las nuevas inclusiones con cuadros de validación en blanco. Algunas unidades léxicas, como *alternate* o *ankle*, fueron eliminadas de inmediato; otras (*ammunition handlers*), fueron valoradas para su posible

inclusión, mientras que las restantes (como *all-around defense* y *alternate position*) formaron parte de nuestra base de datos terminológica.

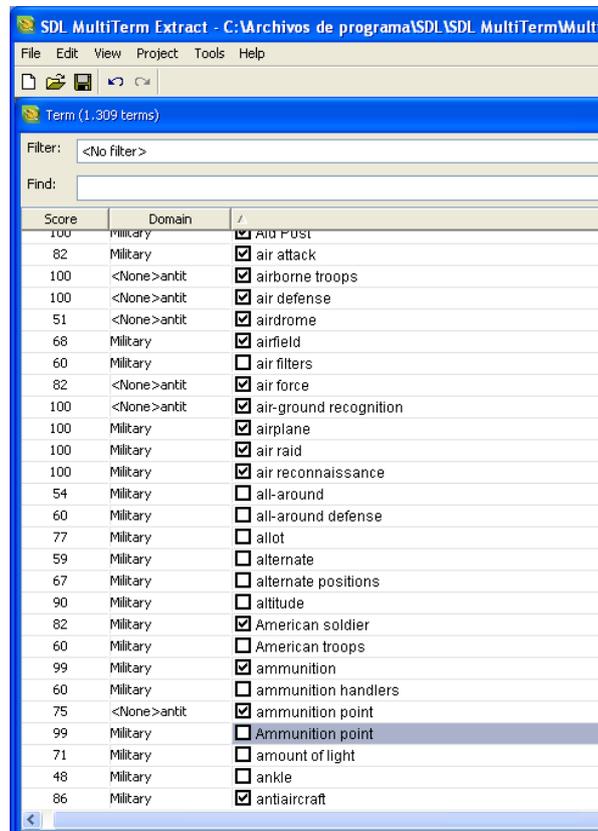


Figura 19. Vista del fichero de vaciado terminológico tras la inclusión de todos los textos.

- **Tratamiento de sinónimos.** Otro aspecto que tuvimos en consideración fue el tratamiento a otorgar a los sinónimos. Al tratarse de un tema delicado, estimamos oportuno incluir todos los que aparecían en nuestro corpus, tanto los que estaban normalizados como los que no. Su uso quedaba claramente reflejado en su aparición en el corpus y podía ser debido a la preferencia que tenían los diferentes autores de emplear unos en detrimento de otros. Por otro lado, es posible que su uso variara dependiendo de la clase social o laboral (en este caso, dependiendo del rango) a la que se pertenecía y a sus conocimientos sobre la disciplina (Picht y Schmitz 228).

- **Identificación de colocaciones equívocas.** Lamentablemente, la extracción terminológica fundamentada en la estadística provoca, independientemente de los ajustes que se realicen, un alto porcentaje de ruido. Según sus propios parámetros, el programa

considera que el número de ocasiones en las que aparece una palabra o verbo (o una flexión de éste) puede ser un hecho indicativo de que se produce lo que Firth denominó *colocación* (*Sinopsis* 196). El hecho de que ciertas palabras se combinen con otras con mayor asiduidad en ciertos contextos puede, en consecuencia, extrapolarse al ámbito terminológico. Como se trata de un hecho basado en la frecuencia, es de suponer que SDL MultiTerm 2009 Extract asume que las relaciones que más se repitan puedan ser consideradas unidades terminológicas complejas. Tomemos como ejemplo *follow* y *following*, unidades léxicas que el programa consideraba como palabras al estar (al menos, su forma canónica) incluida en su archivo de palabras de exclusión. No obstante, cuando éstas aparecían formando una unidad multiléxica con un porcentaje superior al 48 (el mínimo que hemos observado para su inclusión en el archivo de vaciado terminológico), el programa informático las catalogaba como candidatos a término. A destacar que el valor porcentual de estas entradas oscilaba entre 58 y 83, y puede ser ello el motivo de su inclusión.

The screenshot shows the SDL MultiTerm Extract application window. The title bar reads "SDL MultiTerm Extract - C:\Archivos de programa\SDL\SDL Mu". The menu bar includes "File", "Edit", "View", "Project", "Tools", and "Help". Below the menu bar is a toolbar with icons for file operations. The main area is titled "Term (5,314 terms)" and contains a "Filter:" field with the text "<No filter>" and a "Find:" field. Below these fields is a table with three columns: "Score", "Domain", and a column for the term name with a checkbox. The table lists various terms, most with a score between 49 and 100 and a domain of "Military". Some terms are checked, such as "trench", "trip wire", "trip-wire", and "troop-carrier".

Score	Domain	Term
54	Military	<input type="checkbox"/> transmitter
73	Military	<input type="checkbox"/> transport plane
58	Military	<input type="checkbox"/> transport vehicle
92	Military	<input type="checkbox"/> traverse
58	Military	<input type="checkbox"/> traversing mechanism
57	Military	<input type="checkbox"/> tread
49	Military	<input type="checkbox"/> treatise
99	Military	<input checked="" type="checkbox"/> trench
57	Military	<input type="checkbox"/> trestle
49	Military	<input type="checkbox"/> triangle
49	Military	<input type="checkbox"/> triangular
54	Military	<input type="checkbox"/> tribes
49	Military	<input type="checkbox"/> trickery
96	Military	<input type="checkbox"/> tried
49	Military	<input type="checkbox"/> trifles
58	Military	<input type="checkbox"/> trigger lever
57	Military	<input type="checkbox"/> triple
58	Military	<input type="checkbox"/> triple concertina
61	Military	<input type="checkbox"/> tripod
100	Military	<input checked="" type="checkbox"/> trip wire
54	Military	<input checked="" type="checkbox"/> trip-wire
58	Military	<input type="checkbox"/> trip-wire alarm
49	Military	<input type="checkbox"/> trivial
100	<None>antit	<input checked="" type="checkbox"/> troop-carrier
63	Military	<input type="checkbox"/> tronn concentratinn

Figura 20. Vaciado terminológico con el programa SDL MultiTerm 2009 Extract.

No obstante, un alto porcentaje de ocurrencias entre dos palabras no implica que nos encontremos ante una unidad terminológica compleja; la relación exclusiva de dos voces con un bajo porcentaje de frecuencia también existe (Cabré, Estopà y Vivaldi 62) y por ello se han desarrollado otras técnicas en el seno de la lingüística de corpus (*Mutual Information, T-Score, Log-Likelihood,...*) en las que se emplea la estadística bajo unos criterios que buscan analizar la fortaleza de las colocaciones a investigar.

The screenshot shows a window titled "Term (3,896 terms)". It has a "Filter:" field with "<No filter>" and a "Find:" field. Below is a table with three columns: "Score", "Domain", and "/". The table lists various compound terms, each with a checkbox in the third column.

Score	Domain	/
70	Military	<input type="checkbox"/> follow close
71	Military	<input type="checkbox"/> follow close behind
58	Military	<input type="checkbox"/> follow immediately
58	Military	<input type="checkbox"/> following article
58	Military	<input type="checkbox"/> following comment
70	Military	<input type="checkbox"/> following comments on German
63	Military	<input type="checkbox"/> following day
58	Military	<input type="checkbox"/> following dimensions
58	Military	<input type="checkbox"/> following enemy
77	Military	<input type="checkbox"/> following extract
75	Military	<input type="checkbox"/> following extract from a German Army
58	Military	<input type="checkbox"/> following German
73	Military	<input type="checkbox"/> following information
58	Military	<input type="checkbox"/> following methods
83	Military	<input type="checkbox"/> following note
58	Military	<input type="checkbox"/> following observations
63	Military	<input type="checkbox"/> following order
63	Military	<input type="checkbox"/> following paragraph
75	Military	<input type="checkbox"/> following points
67	Military	<input type="checkbox"/> following precautions
58	Military	<input type="checkbox"/> following questions
67	Military	<input type="checkbox"/> following statement
70	Military	<input type="checkbox"/> following table
58	Military	<input type="checkbox"/> following types
66	Military	<input checked="" type="checkbox"/> follow it closely
71	Military	<input type="checkbox"/> follow the tank

Figura 21. Colocaciones compuestas por *follow, following* extraídas por el programa SDL MultiTerm 2009 Extract.

7.2.1.4. Fase IV: Adición indirecta de candidatos a término

Otra manera eficaz de realizar la búsqueda de unidades terminológicas ocurre durante el estudio de frases en contexto de cualquier término o candidato a término. Una de las peculiaridades que poseen los documentos técnicos de cualquier área de especialidad es el profuso empleo de terminología. Las unidades conceptuales se expresan mediante unidades terminológicas y son estas últimas su expresión escrita. Si se trata de un texto

representativo, resulta de lo más frecuente que los conceptos se vayan relacionando unos con otros para explicar un ensayo, método, resultado o fin de una acción, por lo que no hace falta nada más que prestar un poco de atención y registrar los candidatos a término (o susceptibles de serlo) que observemos para comprobar si el programa los ha detectado. Un dato favorable radica en que estas unidades léxicas o terminológicas ya se encuentran almacenadas en nuestro corpus, y no hace falta más que incorporarlas a nuestro fichero de vaciado para comprobar, mediante la búsqueda de frases en contexto, si pertenecen al ámbito militar o al lenguaje general.

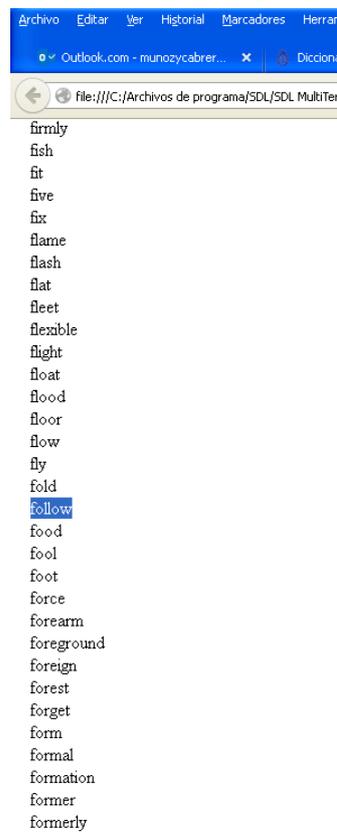


Figura 22. Vista parcial del archivo *excluded terms.htm* y el verbo *follow* resaltado en color azul.

Durante la búsqueda de frases en contexto que podemos observar en la figura 23, constatamos la aparición de unidades terminológicas tales como *crawl trench, gun, crew, quarters, trenches, pillboxes, personnel, supplies, forward positions, communication trenches, machine gunners, defense, shelters, field telephone, pyrotechnic equipment, Very pistol* y *grenade launcher*. Simplemente fue necesario anotarlas para, seguidamente,

corroborar si se encontraban en nuestro fichero de vaciado de términos. En caso contrario, las incorporamos a éste y procedimos a su estudio.

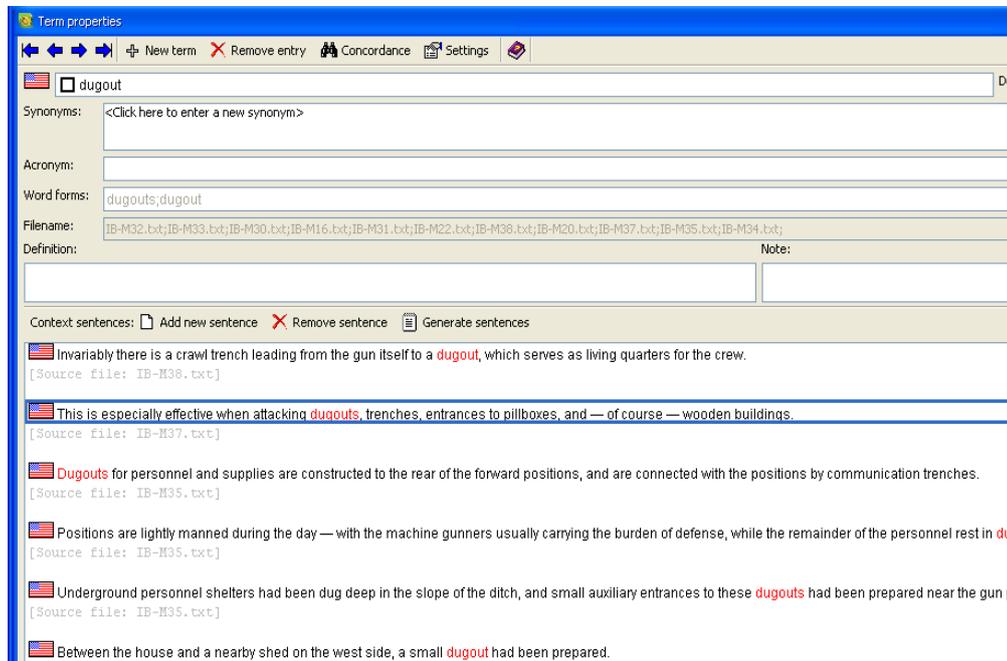


Figura 23. Búsqueda de frases en contexto para el término *dugout*.

Este fue, de forma resumida, el proceso llevado a cabo para la configuración del proyecto, la metodología empleada durante la extracción y búsqueda terminológica estadística, las dificultades relevantes dignas de mención que han aparecido durante la ejecución de dicho procedimiento y el resultado de dicho trabajo detallado, en mayor medida, en la fase dedicada a la resolución de problemas. En la siguiente sección, comentaremos el desarrollo y evolución de la extracción terminológica ejecutada sobre una base lingüística.

7.2.2. Extracción terminológica lingüística

El proceso de extracción terminológica mediante una base lingüística se llevó a cabo empleando el programa AntConc, versión 3.3.5.w (Windows). Anthony, su creador, trabaja en la actualidad en la Facultad de Ciencias e Ingeniería de la Universidad de

Waseda (Japón), y pone a libre disposición dicho *software*, al igual que sus correspondientes actualizaciones. Nuestro propósito ha sido el de emplear la lingüística de corpus para identificar términos mediante su búsqueda en contexto y por medio de la anotación gramatical con el objetivo de reconocer, mediante los patrones con lo que se combinan habitualmente, las unidades terminológicas complejas de la disciplina militar en lengua inglesa.

Para el primer proceso empleamos los mismos textos usados en SDL MultiTerm 2009 Extract, pero incluyendo su correspondiente anotación gramatical. Dicha tarea se acometió gracias a la inestimable ayuda proporcionada por UCREL (University Centre for Computer Corpus Research on Language) de la Universidad de Lancaster, que ofrece, de manera gratuita, varios sistemas de marcado. Para nuestro trabajo hemos optado por emplear el sistema de anotación CLAWS (the Constituent Likelihood Automatic Word-tagging System) en su séptima versión. Dicho sistema fue desarrollado a inicios de los años ochenta y es el que se emplea para las búsquedas gramaticales en el British National Corpus o BNC (en particular, una versión actualizada del CLAWS 6 tagset).

7.2.2.1. Fase I: Identificación y análisis de candidatos a término

El primer paso consistió en extraer un simple listado de palabras de cada una de las obras empleadas en nuestro estudio basado en su frecuencia de aparición, mostrándose al inicio de la lista los vocablos que más se repetían (preposiciones, artículos, pronombres,...). Sin embargo, conforme descendíamos en la tabla, comenzamos a descubrir términos y candidatos a término. Este sencillo paso (realizado también en el proceso de extracción estadístico) es muy útil para identificar voces especializadas y comprobar su recurrencia. No obstante, no se trata de un procedimiento concluyente, puesto que existen unidades terminológicas que se emplean con menor recurrencia y para su localización es necesaria una búsqueda mucho más exhaustiva.

Simplemente examinando una de las palabras (por ejemplo, en *armored*), nos encontramos con diferentes colocaciones que eran, o podían ser, susceptibles de ser consideradas unidades terminológicas. Ello implicó que tuvimos que proceder a examinar en contexto unidades multiléxicas como *armored forces*, *armored-car patrol*, *armored*

troop carrier, armored division, armored unit, armored warfare y armored formation entre otras.

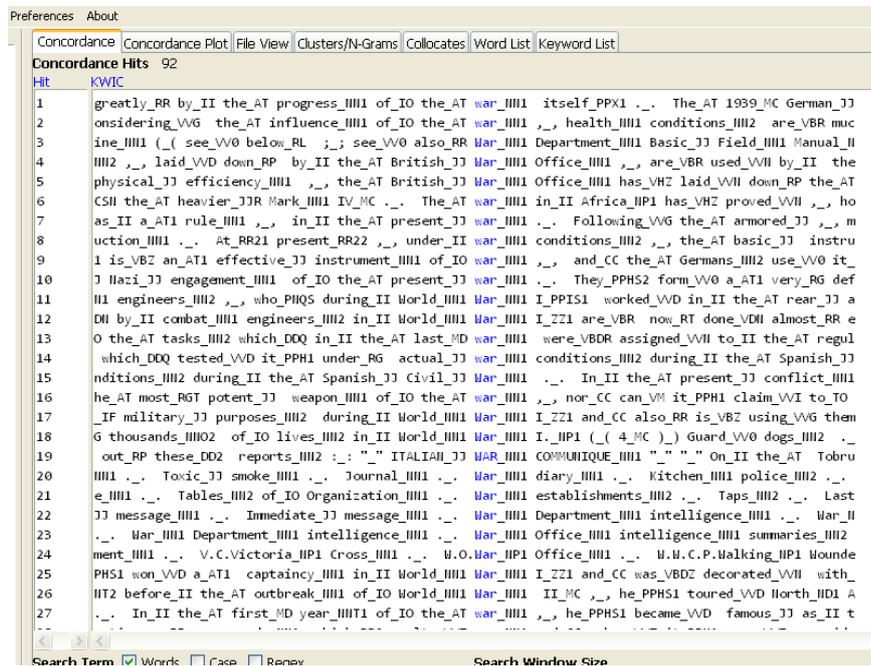


Figura 24. Visualización de textos anotados gramaticalmente.

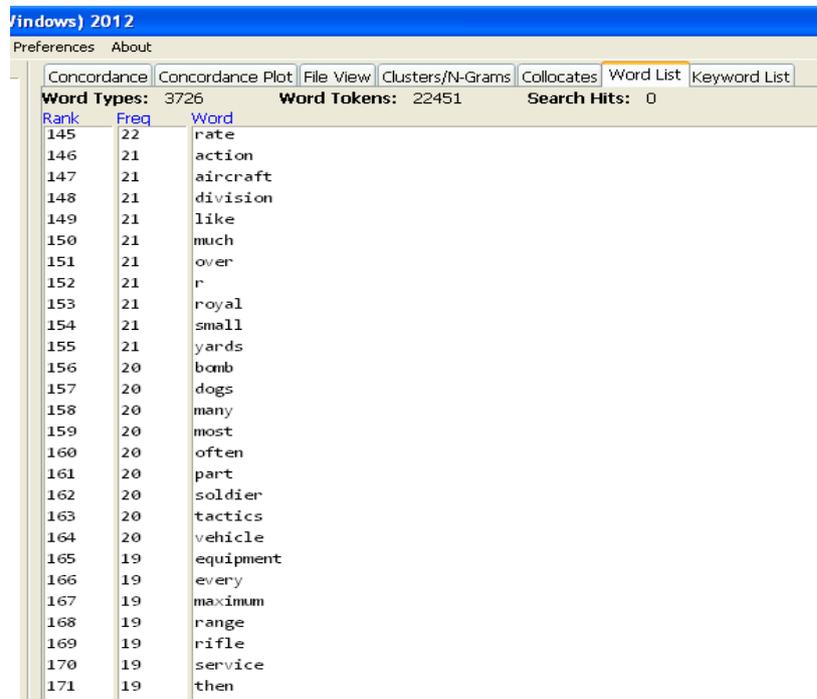


Figura 25. Listado de palabras en AntConc 3.3.5.w (Windows).

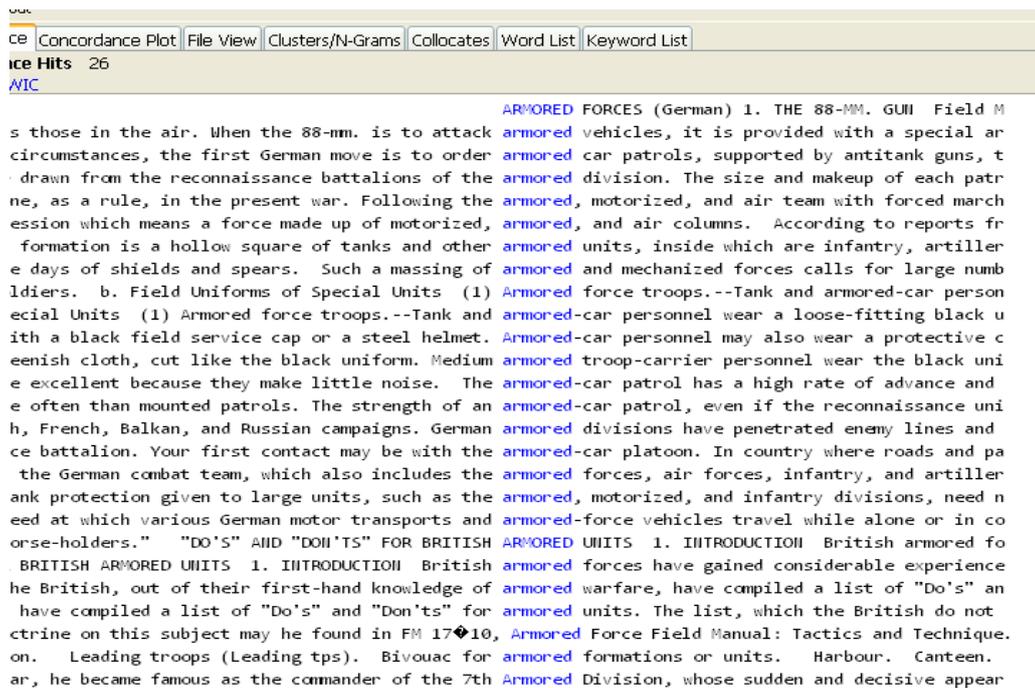


Figura 26. Vista de las colocaciones de la unidad terminológica *armored*.

En el caso de haber identificado un candidato a término en SDL MultiTerm 2009 Extract, y corroborado su condición como tal al examinar la palabra en contexto, pero del que carecíamos de su definición, procedimos a examinarlo con mayor detenimiento en AntConc 3.3.5.w (Windows).

Usando como caso *Motpulk*, realizamos una búsqueda de concordancia escribiendo el término en minúscula (sin seleccionar la casilla *Case*, que implicaría que la búsqueda se limitaría a obtener las veces que aparece únicamente en minúscula), y añadiéndole el comodín “+” (cero o una letra más) para incluir su forma plural.

Seguidamente analizamos las ocho ocasiones en las que apareció dicho candidato a término, con el cometido de averiguar el significado de éste. En esta ocasión, la literatura nos facilitó la comprensión de la unidad conceptual que entrañaba la unidad terminológica *Motpulk*, y nos permitió, además, emplear parte de ésta para configurar su definición.

Este simple método permitió agilizar el proceso de búsqueda terminológica, al obviar la necesidad de crear una definición o de solicitar su formulación a un experto en la materia. Así mismo, hemos enriquecido nuestro conocimiento de este campo de

especialidad al descubrir unidades terminológicas hasta ahora desconocidas. De igual manera, si considerábamos que la definición encontrada en uno de los glosarios consultados era insuficiente, procedíamos a incluir también la hallada en el contexto del término para hacerla más completa.

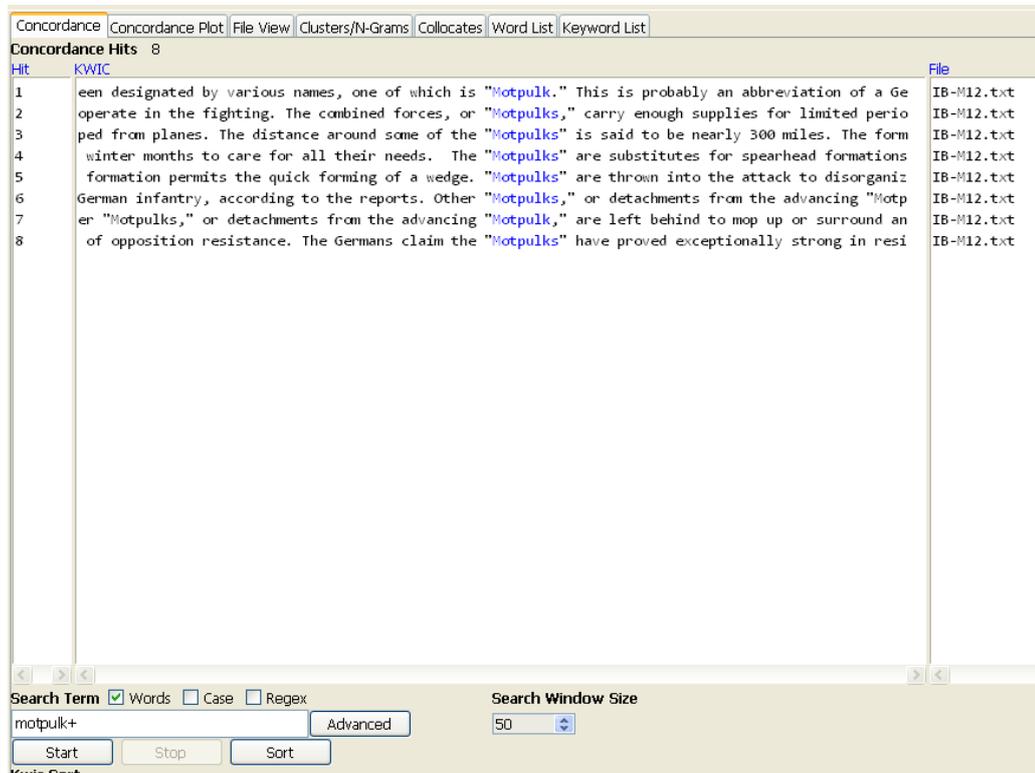


Figura 27. Resultado de la búsqueda de *motpulk+* en AntConc 3.3.5.w (Windows).

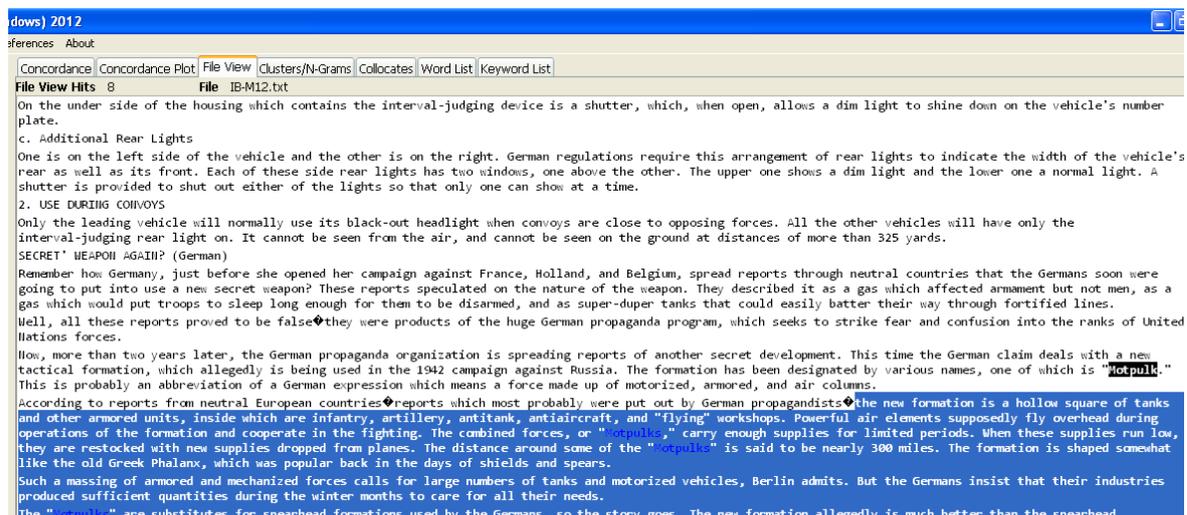


Figura 28. Selección del contexto que será empleado para la definición de la unidad terminológica *Motpulk*.

En este caso, es importante ratificar la importancia de no emplear la casilla *Case* durante el primer estadio de búsqueda. En muchas ocasiones desconocíamos si una unidad terminológica se empleaba en mayúsculas o en minúsculas (o en ambas formas). Tras la primera exploración, y dependiendo de los resultados obtenidos, optamos por realizar una siguiente búsqueda escribiendo el posible candidato a término en mayúscula o minúscula, realizando un estudio más minucioso.

7.2.2.2. Fase II: Identificación y análisis de unidades terminológicas complejas

El siguiente proceso de búsqueda de términos fue el de averiguar qué palabras aparecían con mayor probabilidad en compañía de otras en particular: las denominadas *colocaciones* (Baker, Hardie y McEnery 36). Empleando textos anotados (habiendo previamente elegido la opción de que no aparecieran las etiquetas gramaticales en pantalla para que no fueran consideradas como palabras por el *software*), y usando el término *gun* como punto de partida, realizamos una exploración de sus *collocates* fundamentándonos en su frecuencia de aparición estadística. Esta exploración nos mostró, por ejemplo, que existían 107 voces que aparecían, conjuntamente con *gun*, justo a su izquierda o a su derecha al menos en una ocasión. El consiguiente listado enumeró los resultados basándose en la metodología estadística y nos señaló, de igual modo, en cuántas ocasiones aparecían en total y cuántas a cada lado de la palabra de búsqueda.

Esta indagación también se podía haber llevado a cabo mediante su frecuencia de aparición en general, a su izquierda o a su derecha en particular, siendo clasificadas por palabras o por la terminación de éstas. No obstante, estimamos que la exploración estadística es la más idónea al mostrarnos los resultados estadísticos, de frecuencia en términos absolutos, de frecuencia a la izquierda (*submachine gun, machine gun, tommy gun, ...*) y a la derecha (*gun nests, gun howitzer, gun carts, gun crews, gun barrels, ...*). Todo ello nos proporcionó una información fundamental para explorar y corroborar qué colocaciones podían ser catalogadas como unidades terminológicas, añadiéndolas a nuestro archivo de vaciado terminológico, máxime si habían sido obviadas por el programa SDL MultiTerm 2009 Extract.

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
8	5	1	4	9.41033	breda
9	2	0	2	9.31079	howitzer
10	1	0	1	8.89576	weigh
11	1	0	1	8.89576	pointers
12	1	1	0	8.89576	neighboring
13	1	0	1	8.89576	fiat
14	3	3	0	8.89576	defenders
15	1	0	1	8.89576	carts
16	38	38	0	8.76864	machine
17	1	0	1	8.31079	whereas
18	1	1	0	8.31079	favorite
19	1	1	0	8.31079	experienced
20	1	0	1	8.31079	dismounted
21	1	1	0	8.31079	defending
22	5	0	5	8.31079	crews
23	1	1	0	7.89576	tonny
24	1	0	1	7.89576	automatically
25	1	0	1	7.57383	fast
26	1	0	1	7.57383	destroyed
27	1	1	0	7.57383	camouflaged
28	1	0	1	7.57383	barrels
29	1	0	1	7.31079	lies
30	2	0	2	7.08840	fires
31	1	1	0	7.08840	big
32	1	1	0	6.89576	propelled
33	16	16	0	6.84047	mm
34	2	0	2	6.80829	muzzle

Figura 29. Colocaciones del término *gun*.

Así mismo, AntConc 3.3.5.w (Windows) permitió la ordenación de los resultados y resaltar las palabras que aparecían en su entorno inmediato. Si buscábamos concordancias para (en este caso) *defense*, podíamos optar por seleccionar el casillero *Sort* y hacer que aparecieran la primera, segunda y tercera palabras a su derecha destacadas en diferentes colores. Ello se ejecuta al posicionar el *Kwic Sort* u ordenación de palabras en contexto en la primera palabra a la derecha (1R) en la primera casilla o *Level* y la segunda y la tercera palabras en las dos siguientes. Este fácil método nos permitió buscar unidades fraseológicas y unidades terminológicas complejas. En este caso, *defense area* apareció en más de cuarenta ocasiones, además de otras colocaciones.

Por el contrario, cuando seleccionamos la ordenación hacia la izquierda, también aparecieron colocaciones terminológicas. En este caso, identificamos el término *coast/coastal defense*, al igual que con *coast-defense gun*, que podría considerarse sinónimo parcial de *coast artillery* aunque, en este caso, fue necesaria una investigación más pormenorizada. El propósito de estos ejemplos es el de mostrar la utilidad y

versatilidad que poseen las herramientas puestas a disposición para realizar estudios basados en la lingüística de corpus durante el proceso terminológico, sin necesidad de emplear la anotación gramatical. La simplicidad y facilidad de uso de este programa facilita la compilación sistemática terminológica, que desde hace lustros se encuentra sólidamente fundamentada en el uso del corpus (Sager 130).

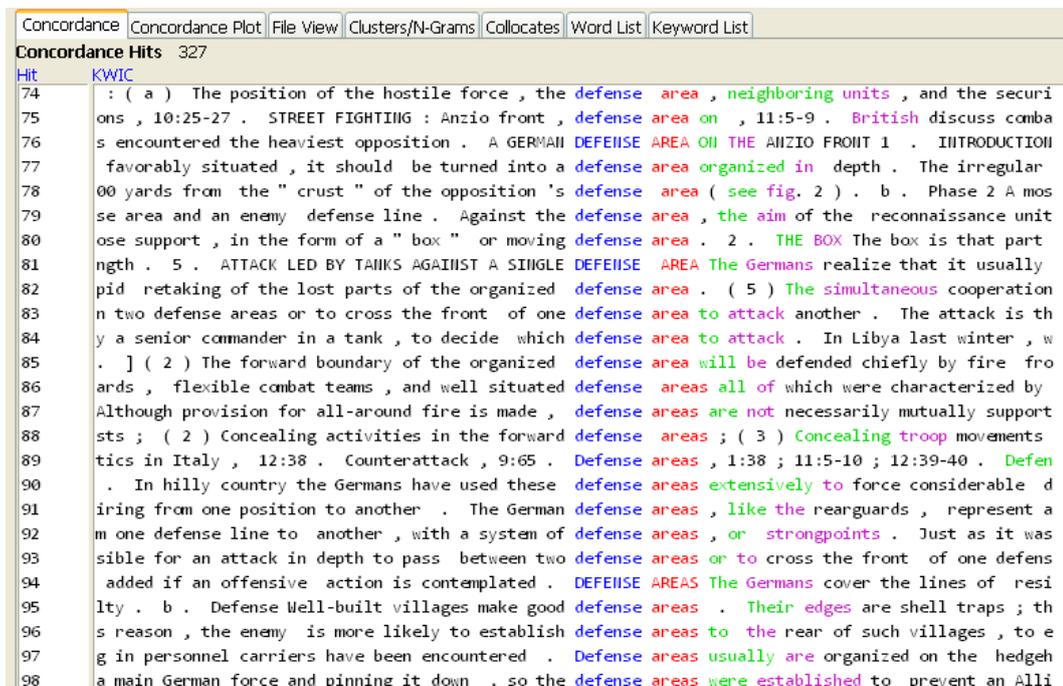


Figura 30. Búsqueda de concordancias de *defense*.

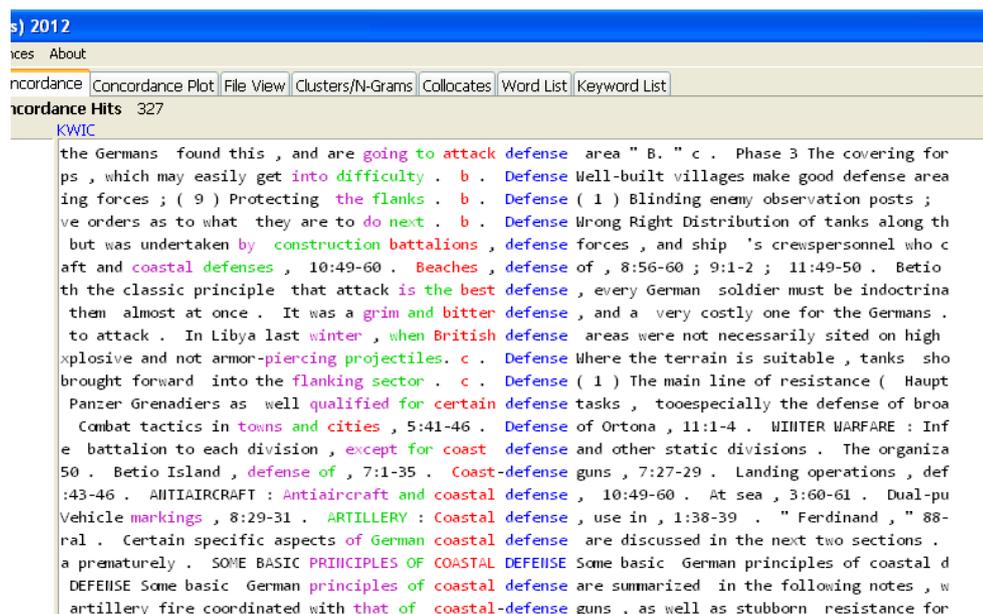


Figura 31. Búsqueda de colocaciones de *defense* por su margen izquierdo.

7.2.2.3. Fase III: Identificación de unidades terminológicas simples mediante la anotación gramatical

a) A la hora de implementar el estudio terminológico mediante la anotación gramatical, el primer paso fue una sencilla búsqueda de voces para iniciar la recopilación de nuestro fichero de vaciado terminológico e integrar en él los posibles candidatos a término que aparecieran. Como ya hemos puntualizado con anterioridad, dentro de los ajustes generales, seleccionamos la opción de no mostrar las etiquetas gramaticales puesto que, al menos en nuestro caso, tendían a confundir y a disturbar nuestra atención. De igual modo, se puede comprobar que cada anotación se une a la palabra que categoriza mediante un guión bajo, y que en los signos de puntuación se aprecia que éstos se duplican, y que, por el contrario, aparece el guión entre ellos (Garside, Leech y McEnery 2). Ese guión bajo funciona como nexos de unión entre palabra y etiqueta y permite investigar los escritos sin necesidad de que aparezcan las anotaciones gramaticales en la pantalla. Afortunadamente, no era necesaria la presencia física de dicho etiquetado, puesto que el *software* reconocía dicha codificación durante el proceso de búsqueda sin necesidad de que el usuario la tuviera que visualizar.



Figura 32. Resultado de una búsqueda mostrando las anotaciones gramaticales.

En consecuencia, y empleando el sistema de anotación gramatical de UCREL CLAWS 7 (ver Anexo I) comenzamos empleando la etiqueta NN, equivalente a nombre común. No obstante, y al ser el guión bajo lo que une a una palabra con la etiqueta, tuvimos que añadir delante de ella el comodín “*” para que el programa AntConc 3.3.5.w (Windows) nos mostrara todas las palabras que estaban unidas a dicho etiquetado. Al mismo tiempo, y como deseábamos incluir las formas plurales, unidades de medida, etc., empleamos también el comodín “*” a la derecha de la anotación. En consecuencia, el etiquetado de búsqueda fue el siguiente: *_NN*.



Figura 33. Búsqueda de sustantivos en AntConc 3.3.5.w (Windows).

El resultado de la exploración nos mostró unidades terminológicas tales como *tank*, *infantry*, *support*, *attack*, *fortress*, *reserves*, *formation*, *artillery*, *engineers*, *teams* y *advance*. Debemos puntualizar que ésta fue la primera fase en la identificación de términos y que fue, a partir de este estadio, cuando iniciamos el proceso de análisis de los vocablos de especialidad en contexto, con el propósito de encontrar su propia definición, y eliminando aquellas voces que procedían del lenguaje común.

b) Una segunda indagación se realizó entre los nombres propios (NP), incluyendo el comodín asterisco a la derecha para incluir los nombres propios en singular (NP1) y en plural (NP2). En consecuencia, la anotación resultante fue *_NP*.

Cabe destacar que entre los resultados obtenidos se encontraban nombres propios (al igual que nombres comunes) tales como *Degtyarev, bomber, officers, wing, lieutenant, offensive, aircraft,...*, que también se empleaban como unidades terminológicas. La aparición de nombres comunes se debió a un error en el proceso de etiquetado automático que no revistió mayor gravedad.

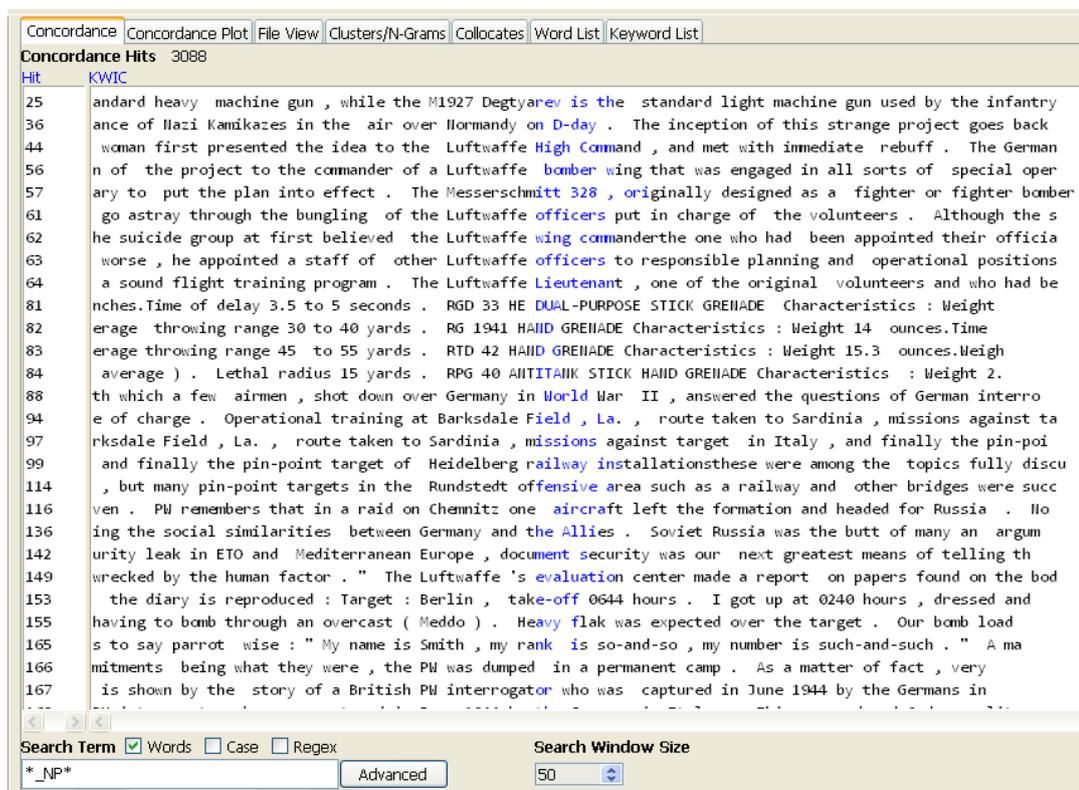


Figura 34. Resultado de la indagación *_NP*.

Estas dos primeras exploraciones sirvieron para completar la búsqueda ejecutada con SDL MultiTerm 2009 Extract. Se efectuó el mismo proceso para el resto de las categorías gramaticales (verbos, adjetivos, adverbios, conjunciones, preposiciones,...) pero, al estimar que su empleo ni poseía una excesiva dificultad ni requería de una formación especializada, consideramos más oportuno pasar a detallar la exploración de unidades terminológicas complejas.

7.2.2.4. Fase IV: Identificación y análisis de unidades terminológicas complejas mediante la anotación gramatical

➤ **Adjetivo + nombre común:** Constatamos que la exploración mediante el etiquetado simple no era del todo concluyente. Es por ello que el mayor reto implicaba el descubrimiento de las unidades terminológicas complejas. En un primer estadio, debíamos reflexionar sobre qué modelos sintácticos se repetían en este campo de especialidad. En inglés, como en cualquier otro idioma, aparecen una serie de patrones por los que se conforman las unidades fraseológicas y las unidades terminológicas complejas. Podríamos decir que los más comunes son *Noun + Noun* (Cabré, Estopà y Vivaldi 63) y *Adjective + Noun*, aunque existen otros más intrincados y/o menos comunes. No obstante, es necesario destacar que en cada lenguaje de especialidad se emplean unas relaciones sintácticas características que no tienen que ser las mismas en todos los campos del saber (Bowker y Pearson 173).

En primer lugar, acometimos una búsqueda adjetivo/nombre. Para ello, empleamos la etiqueta “JJ” (*general adjective*) sin comodín alguno a su derecha, pues resulta muy concreto, al que le siguió la etiqueta “NN” (*common noun*) con el asterisco a la derecha para que nuestra búsqueda fuera lo más amplia posible, quedando nuestra anotación de búsqueda tal como sigue: *_JJ*_NN*.

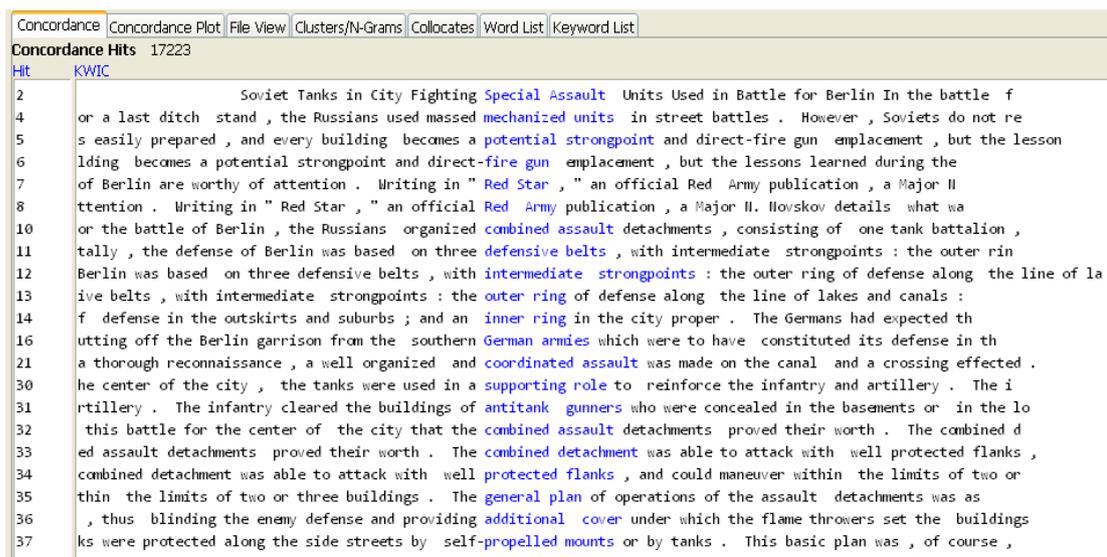


Figura 35. Búsqueda por medio de la anotación gramatical de relaciones entre adjetivo y sustantivo.

Dicha búsqueda de concordancias nos proporcionó 17.223 resultados, entre los que podemos destacar, *mechanized unit, Red Army, defensive belt, outer ring, inner ring, attacking force, defensive zone y main effort*, entre otros. Al igual que con el programa de extracción terminológica sobre base estadística, pudimos seleccionar cualquiera de las colocaciones para encontrar su correspondiente frase en contexto. De igual modo, fue posible ejecutar una exploración tan solo de una entrada en particular, para estudiar todas las ocasiones en las que ésta aparecía con la misión de corroborar su estatus o, mejor aún, buscar su definición, la empresa más difícil de acometer.

TERM: vertical envelopment SUBDOMAIN: Tactics & Strategy PART OF SPEECH: noun WORDFORMS: vertical envelopment, vertical envelopments DEFINITION: The capture of an area by air-borne troops. DEFINITION SOURCE: IB-M12.txt CONTEXT SENTENCE: <i>Airfields and railway and highway junctions are likely to be among the foremost objectives of vertical envelopments.</i> FILENAME: IB-M12.txt NUMBER OF HITS: 4 DATE ENTERED: 6/8/2014

Figura 36. Ejemplo de unidad terminológica compuesta por un adjetivo y un nombre.

➤ **Adjetivo + nombre común + nombre común:** Por otro lado, sabedores de la existencia de términos compuestos por tres unidades léxicas, realizamos una búsqueda más compleja, como fue la de *Adjective + Noun + Noun*. Partíamos de la premisa que cuanto más extensa fuera la pesquisa, mayor sería el ruido resultante y más cuidadosos debíamos ser a la hora de analizar los resultados.

Esta indagación tuvo como resultado la obtención de algunas unidades terminológicas complejas que pasaron a formar parte de nuestro fichero de vaciado terminológico, como fueron *antitank grenade launchers, Military Service Law, heavy machine gun, light machine gun, armored car patrols, United Nations forces, general staff officer, small arms fire, advanced dressing station y protective gas cape*.

➤ **Nombre común + nombre común:** A continuación, procedimos a ver qué secuencias de dos sustantivos eran susceptibles de considerarse candidatos a término. Para ello, utilizamos la etiqueta de nombre común (NN) en dos ocasiones: *_NN* *_NN*.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 2775						
Hit	KWIC					
2	lding becomes a potential strongpoint and direct-fire gun emplacement , but the lessons learned during the bat					
9	trong machine gun and automatic fire and also by antitank grenade launchers installed in houses at the barrica					
13	iable for military service . Under the Universal Military Service law of 1939 , all male citizens " regardless					
14	e Soviet Constitution , as well as the Universal Military Service Law , emphasizes the liability of every citi:					
24	ine gun . The water-cooled Maxim is the standard heavy machine gun , while the M1927 Degtyarev is the standar					
25	gun , while the M1927 Degtyarev is the standard light machine gun used by the infantry squad . Some automati					
52	h Command , and met with immediate rebuff . The German Air Force was not interested in an idea they considere					
53	d Marshall Milch , at that time the head of the German Air Ministry . Again no progress was made . After more					
60	mission of the volunteersto destroy the eventual Allied invasion fleet . Instead , they were continually foster					
90	Germans also make wide use of the Mark III , a light medium tank of 18 to 20 tons (fig. 3) . Formerly it v					
91	ormerly it was armed with one 37-mm. gun and two light machine guns , but in most cases the 37-mm. has been req					
98	ircumstances , the first German move is to order armored car patrols , supported by antitank guns , to do a th					
122	wer half contains the usual red tail-light and a yellow " stop " light , which goes on when the brakes are put					
125	to strike fear and confusion into the ranks of United Nations forces . Now , more than two years later , the					
129	sidelight , an American observer asked a German general staff officer immediately after the campaign in France					
141	itank guns . These boats are very vulnerable to small arms fire . The part played by the engineers must not l					
189	ents . Protected papers . Collecting station . Advanced dressing station . Combat aviation . Operational sq					
190	Conduct of fire.Fire control . Daily telegram . Daily ration strength state . Danger space . Danger area . I					
224	Royal Artillery . R.A.C.Royal Armoured Corps or Royal Automobile Club . R.A.M.C.Royal Army Medical Corps . F					
233	rule the world . 2 . HOW TO IDENTIFY HIM a . Standard Field Service Uniform (see fig. 1b) Privates and ne					
236	er-half with ropes , a canteen , a gas mask and protective gas cape , an entrenching tool , and side arms . C					
237	sonnel wear a loose-fitting black uniform with a black field service cap or a steel helmet . Armored-car perso					

Figura 37. Diversos resultados de la búsqueda realizada mediante la anotación *_JJ* *_NN* *_NN*.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 11113						
Hit	KWIC					
3417	on . The current method is to send two or three recovery vehicles forward with the fighting					
3418	two or three recovery vehicles forward with the fighting units . These vehicles advance in					
3419	line of attack and cruise across the width of the battle front . The Germans believe that he					
3420	e German tanks and therefore will not attack the recovery vehicles , even when they come ver					
3421	en when they come very close . If a member of a tank crew orders the driver of a recovery v					
3422	f a member of a tank crew orders the driver of a recovery vehicle to tow his tank to the rear					
3423	rtant and could have been fixed on the spot by a repair section) . It is always permissible					
3424	forward alone , and tows a disabled tank away by tow ropes . Towing is used in preference to					
3425	as 20 minutesand time is precious in front-line recovery . Prisoners state that trailers ar					
3426	maintained , and the trailers neither cut up the road surfaces nor weave as much as a towed					
3427	much as a towed tank . In roadless parts of the desert , trailers are resorted to where the					
3428	the ground affords reasonably good going . The recovery platoon is not given the whole res					
3429	tanks . In case of retirement , the Germans use combat tanks to tow disabled tanks . Instar					
3430	t , the Germans use combat tanks to tow disabled tanks . Instances have been reported in whi					
3431	in the towing . Recovered tanks are towed to an assembly point behind the combat area . Tr					
3432	tanks are towed to an assembly point behind the combat area . Trailers may be used to take					
3433	may be used to take the disabled tanks from the assembly point to a workshop company . Acc					
3434	the disabled tanks from the assembly point to a workshop company . According to prisoners					
3435	According to prisoners of war , the drivers of recovery vehicles have done front-line duty					
3436	t a time , and then worked at the rear , between assembly points and workshops . One prisor					
3437	intelligence and plenty of courage . c . Light Workshop Platoon A German document from Hor					
3438	ica gives detailed instructions for organizing a workshop platoon in a two-battalion tank re					
3439	organizing a workshop platoon in a two-battalion tank regiment (which normally would not hav					
3440	was obtained for the platoon by breaking up the battalion headquarters repair sections of th					
3441	o battalions . This platoon was smaller than the workshop platoon designated by the pre-war					
3442	oon designated by the pre-war organization for a tank regiment of three battalions , and was					
3443	battalions , and was to operate in place of the battalion headquarters repair sections , ur					

Figura 38. Búsqueda de frases nominales conformadas por dos nombres comunes.

Como suponíamos, esta exploración proporcionó una inestimable información que nos sirvió como punto de partida para la compilación de nuestro glosario.

➤ **Nombre común + nombre común + nombre común:** Empleando el mismo procedimiento ejecutado entre adjetivos y nombres, iniciamos la búsqueda de relaciones entre tres nombres, concedores de que existen términos que siguen dicho patrón.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 1538						
Hit	KWIC					
27	f body 2.2 inches.Time delay 3 to 4.4 seconds.F1 FRAGMENTATION HAHD GRENADE Characteristics : Weight (fuzed) 1 1/4 pounds					
28	ge) . Lethal radius 15 yards . RPG 40 AHTITANK STICK HAHD GRENADE Characteristics : Weight 2.7 pounds.Length over-					
29	rds.Average throwing range 25 to 30 yards.RPG-43 HEAT HAHD GRENADE Characteristics : Weight About 2.65 pounds.Avera-					
30	pe of Fuze Instantaneous impact . M1917 CHEMICAL STICK HAHD GRENADE Characteristics : Weight 1.9 pounds.Weight of fi					
31	head 2.6 inches.Time delay 3.5 to 4 seconds.RGDI SNOKE HAHD GRENADE Characteristics : Weight 18 to 18.3 ounces.Filli					
32	gth of smoke screen 27 to 33 yards . " 5 " SNOKE HAHD GRENADE Characteristics : Weight 1.25 pounds.Length 6.3 inches.Diameter					
45	some important locality in our area , such as a tank repair center . There may be more than one of these thrusts					
64	ll the arms and services operating together in a task force combat team . The " secret " weapon was cooperation .					
185	R.A.M.C.Royal Army Medical Corps . R.A.O.C.Royal Army Ordnance Corps . R.A.P.Regimental Aid Post . R.A.P.C.Royal Arm					
186	rps . R.A.P.Regimental Aid Post . R.A.P.C.Royal Army Pay Corps . R.A.S.C.Royal Army Service Corps . R.A.V.C.Ro					
187	. R.A.P.C.Royal Army Pay Corps . R.A.S.C.Royal Army Service Corps . R.A.V.C.Royal Army Veterinary Corps . R.E.Cor					
188	R.A.S.C.Royal Army Service Corps . R.A.V.C.Royal Army Veterinary Corps . R.E.Corps of Royal Engineers . R.E.M.E.Royal					
197	e world . 2 . HOW TO IDENTIFY HIM a . Standard Field Service Uniform (see fig. 1b) Privates and noncommissioned off					
198	ainted with a gray , rust-resistant paint , or a field service cap made of greenish-gray cloth . This cap may be w					
199	lt is of soft black leather and has a dull white metal buckle . Cartridge pouches are attached to the belt . The gray clot					
200	protective gas cape , an entrenching tool , and side arms . Officers ' field uniforms are similar to those of line so					
201	wear a loose-fitting black uniform with a black field service cap or a steel helmet . Armored-car personnel may al					
202	niform , cut like the black uniform , and a gray field service cap or steel helmet . (2) Parachute troops(see					
203	ose-fitting gray-green coveralls with very short legs , gauntlet gloves , and ankle-length boots laced at the sides . L					
204	These troops wear a special cap (similar to the field service cap , but with the addition of a cloth visor) , th					
222	e of the fully motorized units . Also , portable radio communication sets may be allotted to the patrols . It must be rem					
225	ght and heavy antiaircraft artillery units , and barrage balloon units . 2 . USE OF SEARCHLIGHTS In addition to seekin					
230	ge is coordinated with light- and medium-caliber antiaircraft gun defense , which protects the larger gaps in the belt .					
254	ision headquarters for tactical interrogation by division intelligence officers . The officer or noncom in charge of the front-					
255	es with him all captured documents , and informs division intelligence officers as to the local tactical situation so that they					
257	econd in command . He fires the turret gun , the turret machine gun , or the machine carbine , as ordered by the tan					
263	l voice communication . The same microphones and telephone receiver headsets are used in both cases . When the control switc					

Figura 39. Vista de algunos resultados de la consulta formulada con la anotación *
NN* * NN* * NN*.

Al contrario de lo que esperábamos, los resultados fueron infructuosos. Ello tiende a confirmar que los patrones por los que se crean las unidades terminológicas complejas varían de una disciplina a otra (Schmitz 33-5). Por tanto, es necesario estudiar las asociaciones de las unidades terminológicas complejas para establecer cuáles son las más recurrentes en la disciplina en la que se lleva a cabo una investigación.

➤ **Nombre propio + nombre común:** Otro factor a analizar fue la localización de los nombres propios que anteceden a nombres comunes. En múltiples campos de especialidad, los términos incluyen el descubridor de una especie, o creador de un artilugio o procedimiento. Por ello, procedimos a llevar a cabo la búsqueda de concordancias entre nombres propios seguidos de nombres comunes (*_NP *_NN*):

AntConc 3.3.5.w (Windows) nos presentó una serie de relaciones en las que se observaba (en la mayoría de las ocasiones) el nombre del fabricante o del creador seguido de su producto o metodología. Cabe precisar que con la exploración de las relaciones entre nombres propios y comunes aparecieron un cierto número de relaciones entre parejas de nombres comunes, no apreciables con anterioridad. Tanto Garside (1997), como Hardie (2013) o Anthony (2013) afirman que la anotación gramatical no es del todo concluyente y que, en particular, la versión CLAWS 7 posee entre el 96 y el 97 por ciento de precisión (UCREL), por lo que ello provocó la aparición de pequeños errores. No obstante, ya hemos destacado que se trata de fallos de poca importancia que no impidieron el desarrollo de nuestra investigación.



Figura 40. Resultado de la búsqueda de relaciones entre nombres propios y nombres comunes.

➤ **Nombre común + adjetivo + nombre común:** Una nueva exploración consistió en averiguar la relación en la secuencia nombre/adjetivo/nombre. Para ello, configuramos las etiquetas gramaticales de la siguiente manera: *_NN* *_JJ* *_NN*.

En esta ocasión quedó contrastada la ineficacia de nuestra búsqueda. No obstante, de ella pudimos destacar *effective bursting radius, maximum lethal radius, district*

administrative authority y *Army Medical Department*. A pesar del alto número de coincidencias (993), fue necesario un análisis minucioso de todos los resultados y el número de candidatos a término fue claramente inapreciable.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 993						
Hit	KWIC					
1	East and had concentrated their defenses in that area . Soviet tank units , however , attacked from the south					
15	on jacket 2.2 inches.Time delay 3.2 to 4 seconds.Effective bursting radius (with fragmentation jacket) 15 yar					
16	ng radius (with fragmentation jacket) 15 yards.Maximum lethal radius (without fragmentation jacket) 10 yard					
33	hute machine-gun battalion (three companies) . Parachute antitank battalion (three companies) . Parachute a					
34	te engineer battalion . Parachute signal unit . Parachute medical unit . Parachute supply unit . Captured loa					
37	ere is little emphasis on its use in training . NIGHT DRIVING EQUIPMENT 1 . GENERAL German vehicles have speci					
114	, and even then the matter is in the hands of a district administrative authority . When the families of offic					
344	ation acquired from local health authorities and Army medical officers , should serve to indicate which of the					
350	treated under the supervision of British or U.S. Army Medical Department personnel is safe and should be used ,					
351	day may suffice , but on such a restricted water ration , physical efficiency is reduced after a short interva					
352	he sun is necessary for an appreciable length of time . Outer garments should be loose-fitting so that sweat w					
353	e day , especially indoors , but from dusk until morning long trousers and long sleeves are imperative . The c					
354	. The often-repeated suggestions to wear stomach bands , spinal pads , and similar devices to prevent diarrhea					
355	; as a result , meat from animals infected with tuberculosis , undulant fever , anthrax , trichinosis , and sc					
356	ion (by food handlers--cooks and waiters--or by flies , other insects , or small animals) and thus may become					
357	kinned fruits , and so forth , be carried . The Army canned field ration is a practical , safe food for the at					
358	following : dysentery , typhoid and paratyphoid fevers , common diarrheas , diphtheria , tuberculosis , undula					
359	typhoid fevers , common diarrheas , diphtheria , tuberculosis , undulant fever , septic sore throat , and other					
360	dical Department standards) , canned evaporated milk , condensed milk , powdered milk prepared with boiled wat					
361	revalence of certain eye diseases--for example , trachoma , gonorrhoeal ophthalmitis , and pink eye--it is neces					
362	f the body noted above is also a good preventive measure . Fresh waters , such as lakes , rivers , streams , s					
363	e scorpions , spiders , and other insects prefer dark , warm places for rest , and may crawl into such places					
364	of the health problems found in the Middle East area . PHYSICAL TRAINING NOTES FOR BRITISH AA UNITS 1 . INTF					
365	et 6 inches in full equipment without sustaining injury . Progressive instruction should be given in how to la					
366	ghts . It will be found that in the majority of cases elementary instruction will be necessary in plimsols (s					

Figura 41. Resultados relevantes de la inspección nombre común + adjetivo + nombre común.

➤ **Nombre común + preposición + nombre común:** Así mismo, un esmerado estudio nos sirvió para corroborar la formación de unidades terminológicas empleando la pauta sustantivo/preposición/sustantivo. Empleamos inicialmente la preposición *of* cuya etiqueta es “IO” sin ningún tipo de comodín, conjuntamente con la de nombre común. La búsqueda quedó constituida de la siguiente forma: *_NN* *_IO* *_NN*.

Como habíamos supuesto, de las 2.671 concordancias resultantes pudimos incluir un cincuenta por ciento de candidatos a término que seguían dicho patrón (entre los que destacamos *center of resistance*, *prisoner of war*, *volume of fire*, *theater of operations*,...). Otra preposición empleada en este campo de especialidad es *for*, por lo que usamos su etiqueta (IF) en medio de las de nombre común (*_NN* *_IF* *_NN*):

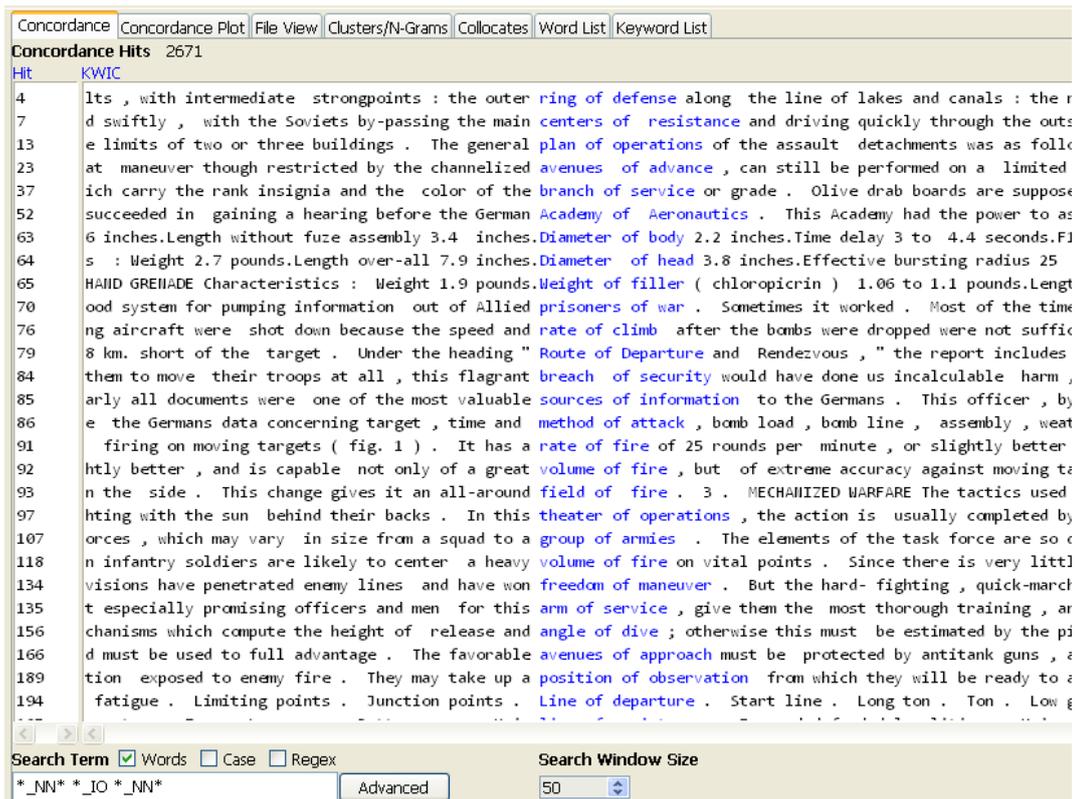


Figura 42. Resultados de la búsqueda *nombre común, preposición "de", nombre común.*

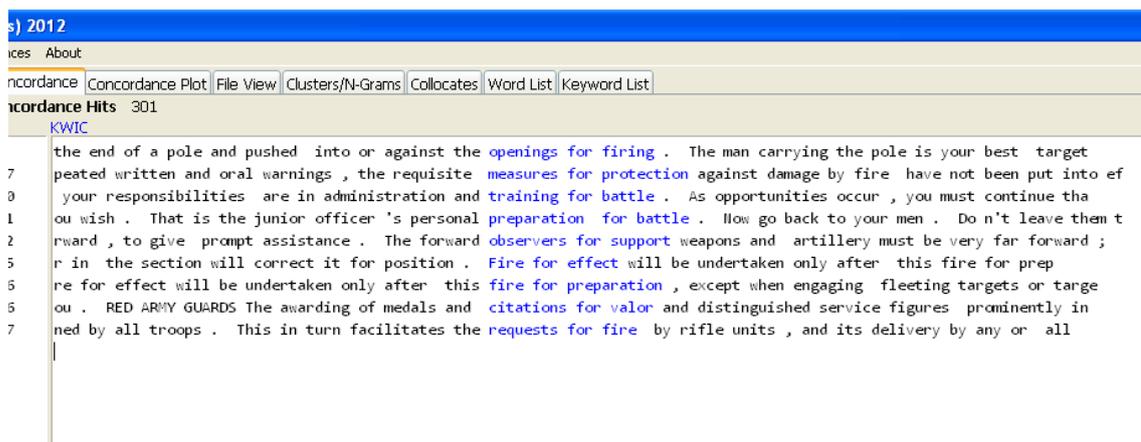


Figura 43. Muestra de los resultados de la búsqueda ** NN* * IF * NN*.*

Desafortunadamente, resulta más que evidente que de las 301 concordancias obtuvimos muy pocos ejemplos (a destacar *opening for firing, fire for effect* y *fire for*

preparation) que acabaron finalmente por engrosar nuestro fichero de vaciado terminológico.

A lo largo de este capítulo hemos pormenorizado el proceso de extracción terminológica estadística mediante el uso del programa SDL MultiTerm 2009 Extract. Gracias a él se configuró el archivo de vaciado de términos, punto de partida de nuestra investigación. En una segunda fase, comenzamos a realizar un estudio conceptual de los candidatos a término, analizamos el uso del vocabulario especializado en contexto para seleccionar ejemplos de uso, agregamos posibles candidatos que no fueron reconocidos por el mencionado *software*, y procedimos a la formulación de sus posibles definiciones.

Por otro lado, hemos empleado AntConc 3.3.5.w (Windows) a modo de programa de extracción terminológica lingüística ya que contribuye de manera exponencial a identificar unidades terminológicas (simples y complejas), al posibilitar búsquedas que no eran posibles de acometer con el programa de base estadística. El enfoque lingüístico permite el estudio de patrones recurrentes a nivel sintáctico imposibles de realizar desde una base estadística. Por ello, es primordial el etiquetado (al menos, gramatical) de los textos que conforman el corpus para realizar dicha tarea con éxito. No obstante, no sería necesaria dicha anotación si, por medio de métodos más sencillos, deseamos analizar un corpus textual para identificar unidades lexicológicas o terminológicas. En definitiva, las herramientas de lingüística de corpus poseen la capacidad de contribuir eficazmente al estudio de las unidades terminológicas de cualquier disciplina, y no solo de las unidades léxicas de los LGP (Bowker y Pearson 137).

8. CONSIDERACIONES SOBRE EL TRABAJO DE INVESTIGACIÓN

8.1. OBSERVACIONES GENERALES

Una vez concluido el proceso de compilación del glosario, procedemos a detallar los datos dignos de mención, al igual que los hechos o teorías que hemos tenido la oportunidad de apreciar durante nuestro trabajo.

En términos generales, hemos constatado que los glosarios, una vez publicados, dejan al poco tiempo de estar actualizados (Bowker y Pearson 141). Un glosario en formato tradicional no solo requiere de un largo proceso de compilación, sino que desde que queda finalizado hasta que aparece en las librerías especializadas transcurre un periodo temporal demasiado amplio en el que dejan de emplearse ciertas unidades terminológicas y, al mismo tiempo, se aprecia la acuñación de otras nuevas. Ello queda más patente en cualquier ámbito de especialidad de las ciencias (tales como la medicina o la ingeniería), en donde los nuevos hallazgos provocan la aparición de nuevas unidades conceptuales y, por ende, de nuevas unidades terminológicas que las denoten.

No obstante, la informática y la digitalización han agilizado en cierta medida la actualización terminológica. En lo que concierne a la terminología militar, el Ministerio de Defensa de los Estados Unidos pone (desde el ocho de noviembre de dos mil diez) a libre disposición en *Internet* la siguiente obra: United States, Department of Defense. *Department of Defense Dictionary of Military Terms and Associated Terms*. Washington: Joint Publication 1-02, 2010. Print. La finalidad de dicha obra es descriptiva, ya que recopila los términos de este campo de especialidad en un glosario. En dicho caso, éste se actualiza periódicamente, y su última revisión se llevó a cabo el quince de diciembre de dos mil trece. Por otro lado, la finalidad de esta publicación también es prescriptiva (Cabrè y Estopà 8), al tener como objetivo la normalización de los términos, no solo a nivel nacional, sino también en el seno de la OTAN. Además, la red informática, en este caso, propicia su libre divulgación, poniendo a manos de cualquier interesado en dicha disciplina una información actualizada y rigurosa.

Otro dato digno de mención resultado de este trabajo ha sido el de constatar que la manera más idónea de estudiar y almacenar la terminología es por medio de bancos terminológicos informatizados. El inmenso coste que conlleva la compilación y actualización terminológicas tiene como resultado la necesidad de poner al alcance de los potenciales usuarios bancos de datos terminológicos (Sager 172). Al igual que ocurre con los grandes corpus de última generación, estos bancos de datos se encuentran disponibles en la *web* propiciando, mediante diferentes herramientas, la realización de consultas terminológicas que se adecúen a cada una de nuestras necesidades. En cualquier caso, el fruto de nuestra investigación no requiere necesariamente estructurarse en este tipo de formato, al tratarse de un estudio puntual dentro de un periodo determinado en el tiempo.

Aunque este trabajo ha tenido como objetivo el análisis sincrónico de la terminología militar empleada por las fuerzas terrestres en los teatro de operaciones europeo, del Mediterráneo y del norte de África durante la Segunda Guerra Mundial, no podemos obviar la evolución que sufren las unidades terminológicas durante ese espacio temporal. Como ya hemos aludido previamente, los glosarios que hemos usado como manuales de referencia fueron editados en 1943 y 1944. Es digno de mención que, en tan solo un año de diferencia, ciertas unidades terminológicas sufrieron una profunda transformación, en particular a nivel morfológico y semántico. Cabe destacar que algunos de los sinónimos que poseían ciertas unidades terminológicas, procedentes del inglés británico y/o del americano, se convirtieron posteriormente en unidades terminológicas *per se* con una significación propia.

El uso profuso de sinónimos (tanto los normalizados como los que han sido descubiertos durante el análisis de nuestro corpus) da clara fe de la multiplicidad de autores que escribieron artículos en el *Intelligence Bulletin*. No solo se trata de escritores de procedencia norteamericana, sino británica y de todos los rincones de su antiguo imperio. Una de las peculiaridades de la publicación que hemos empleado para la creación de nuestro corpus es el análisis de las nuevas tácticas empleadas por parte de los aliados y por el enemigo, además del estudio de las nuevas armas y de su uso, factor que propicia la observación de nuevas unidades terminológicas a nivel escrito pudiendo, de ese modo, obtener incluso sus definiciones en contexto.

A nuestro juicio, en la década de los cuarenta no existía un adecuado consenso con respecto al empleo de la terminología en el seno de la ciencia militar, y su normalización no comenzará a apreciarse hasta la segunda mitad del siglo XX. Esta falta de acuerdo y el continuado uso de sinónimos que no aparecen en los glosarios puede, de igual modo, poner en duda la limitada adecuación de ciertas unidades terminológicas. Todo ello es comprensible puesto que, en un conflicto que duró casi seis años, se crearon un alto porcentaje de ingenios y tácticas bélicas, unidades conceptuales que tuvieron que asociarse a nuevas representaciones léxicas.

El número de candidatos a término que conformaban el archivo de vaciado terminológico al cargar todos los textos en SDL MultiTerm 2009 Extract era de 5.382. Tras su pertinente análisis, obtuvimos un total de 3.090 unidades terminológicas (abreviaturas y acrónimos incluidos). En lo que atañe al método empleado para su obtención, alrededor de un cuarenta por ciento de las unidades terminológicas fueron detectadas por el programa SDL MultiTerm 2009 Extract, siendo el sesenta por ciento restantes identificados por el autor durante el estudio pormenorizado de las frases en contexto, o mediante la revisión de los textos por medio de AntConc 3.3.5.w (Windows).

En lo que se refiere a la descripción de las unidades terminológicas, nuestras tres fuentes principales han sido los glosarios elaborados por el Ejército norteamericano en 1943 y 1944, y nuestro corpus. No obstante, también hemos empleado diccionarios disponibles en *Internet* (tanto generales como enciclopédicos), y glosarios del *US Army* editados en 1950, 1983, 2005 y 2010 respectivamente. Hemos utilizado, en la medida de lo posible, una metodología cronológica a la hora de seleccionar cada definición. *Grosso modo*, las descripciones editadas en 1943 suelen ser más breves y más genéricas comparadas con las que aparecen en 1944 aunque, en ciertas ocasiones, éstas prácticamente no varían (*emergency barrage, encode,...*). Por el contrario, las aparecidas en 1944 suelen considerarse como más didácticas, con una cierta tendencia a la paráfrasis y a la síntesis y a mostrar las diferencias existentes entre las unidades terminológicas pertenecientes a un subespacio común. Como norma general, hemos empleado la definición más antigua, a no ser que ésta haya caído en desuso o que la publicada en fechas posteriores fuera más específica. No obstante, si su descripción no quedó correctamente especificada, procedimos a incluir varias definiciones de diferente procedencia para una mejor comprensión.

Con respecto a las definiciones obtenidas de nuestro corpus, nuestra labor se ha visto limitada a la hora de seleccionar una frase en contexto para mostrar el uso de una unidad terminológica en particular. A veces no hemos tenido otra opción que emplear una oración que también constituye parte de la propia definición aunque, en la medida de lo posible, hemos mostrado una diferente en la que se aprecie únicamente su uso y que proporcione información adicional sobre dicho término. En ciertas ocasiones, un término solo ha aparecido en nuestro corpus en una ocasión y, a modo de adjetivo o modificador de otro término (*emergency barrage / emergency barrage area*). Cuando ello ha ocurrido, se ha procedido a su pertinente comentario mediante una nota aclaratoria en la correspondiente entrada.

En lo que concierne a la manera de detallar las definiciones que hemos redactado a modo de propuesta, éstas se han realizado (en la medida de lo posible) usando definiciones de otras unidades conceptuales ubicadas en un estrato superior dentro de la estructura jerárquica sujeta a estudio, o de otra unidad conceptual perteneciente, al menos, al subárea temática objeto de observación. En términos generales, no ha sido siempre posible describirlas mediante el clásico modelo de *genus et differentia* o de análisis (Sager 42), sino mayoritariamente empleando la paráfrasis, la síntesis y la implicación. Tras ello, el experto en la disciplina valoró la propuesta, aceptándola, realizando las correcciones, oportunas o formulando una nueva.

Como pauta general, el número de frases en contexto que hemos acordado en SDL MultiTerm 2009 Extract con el propósito de seleccionar un enunciado en donde se aprecie el uso de cada unidad terminológica ha sido de diez. No obstante, en los casos en los que las unidades léxicas poseían además la categoría de unidad terminológica, dicha indagación fue de hasta cincuenta oraciones, puesto que era necesario encontrar las locuciones en las que se empleaba dicha palabra o recurso idiomático como término y no existía otro método que el de incrementar el número de frases para obtener un ejemplo significativo en el ámbito de la ciencia militar.

Por otro lado, resultó inevitable añadir unidades terminológicas de carácter histórico al quedar detalladas y definidas en los artículos que constituyen nuestro corpus. Términos tales como *Volksstrum*, *Hitler Youth*, *Molotov cocktail*, *Kamikaze* y *Panzer*, entre otros, definen unidades conceptuales que, en algunas ocasiones, han traspasado ese espacio

temporal y se emplean en la actualidad, tanto a modo de unidad terminológica perteneciente al lenguaje de especialidad del ámbito militar, como de unidad léxica del lenguaje común anglosajón. A nuestro juicio, la inclusión de unidades terminológicas que no se encontraban incluidas en los glosarios consultados (una cifra inferior al ocho por ciento del total de las unidades terminológicas registradas) son el aspecto de originalidad y de mayor peso investigativo en nuestro proyecto (*tank warfare, stool pigeon, plastic explosive, Motpulk,...*); sin ello, el resultado hubiera sido una mera búsqueda de las unidades terminológicas en contexto, a las que se le añade una definición previamente formulada. El propósito de todo este estudio ha sido el de aportar información nueva, ya fuera tanto mediante ejemplos en contexto como por medio de la inclusión de nuevas unidades terminológicas o de carácter histórico íntimamente relacionadas con el conflicto en cuestión.

A continuación, procedemos a detallar las variaciones más representativas que hemos apreciado en nuestro trabajo:

8.1.1. Evolución de la terminología: apreciaciones genéricas

a) El primer objeto de estudio ha sido el caso de la sinonimia entre las unidades terminológicas. Hemos constatado el hecho de que a veces se empleaba con mayor asiduidad el sinónimo en detrimento del término principal o normalizado (*secondary attack / holding attack*). Otro suceso se produce cuando la unidad terminológica normalizada se utilizaba al igual que otra no normalizada (*combat patrol / fighting patrol, combat troops / fighting troops, combat unit / fighting unit*), y que esta última cayó en desuso en años posteriores. Por otro lado, en más de una ocasión se empleaba una unidad terminológica no normalizada en detrimento de la que aparecía en los glosarios consultados (*signal lamp / signal light, service cap / field service cap, rope ferry / rope railway*). Ello tiene dos interpretaciones evidentes: el reducido tiempo que transcurre para que un glosario quede anticuado, y la rápida evolución de la terminología militar en tiempos de conflagración.

b) Una segunda evolución se produce a favor de la precisión o adecuación de la unidad terminológica. En algunos casos, su definición se amplía o se desambigua (a veces

mediante notas aclaratorias) para confirmar que una acción puede emplearse para más de un tipo de vehículo (*radius of action*) o, por el contrario, se altera la denominación de la unidad terminológica con el propósito de delimitar su especificidad y, por ende, su carácter unívoco (*heavier-than-air / heavier-than-air aircraft*). Dentro de este grupo podemos añadir a los términos que eliminan el adjetivo *military* al ser empleados en dicha área de especialidad (*military information / information*), aunque dicha norma no siempre se cumpla (*military intelligence* e *intelligence* son considerados sinónimos en 1944 compartiendo una misma definición).

c) El tercer caso implica la eliminación de ciertas unidades terminológicas de los glosarios consultados (*arrive, cooperation, hold, seize, situation, strongpoint, stronghold*), a pesar de que algunas de ellas mantengan una clara significación bélica. En la mayoría de las ocasiones, estos vocablos pasarán a formar parte del lenguaje general inglés, mediante el proceso denominado *de-terminologization*, ya comentado con anterioridad.

d) Un cuarto grupo lo constituyen las unidades terminológicas que perdían su definición en 1944 y a las que solo se les aportaba su abreviatura o acrónimo (*entrucking point, sanitary, ...*). Como norma general, este método se empleaba con unidades léxicas que, al emplearse profusamente en el campo militar, no procedía incluir su definición (*miscellaneous - misc, low altitude - LA, machinist - machst, forward - fwd, sick - sk, ...*). No obstante, sí hemos aportado su correspondiente descripción a cada abreviatura y acrónimo que aparecía en nuestro corpus y que estaba incluido en los glosarios empleados como guías de uso. El propósito principal de tal decisión estriba en la necesidad de detallar una explicación a cada entrada que hemos incluido en nuestro trabajo (a excepción de los sinónimos, cuya definición ha quedado explicitada en el término principal o normalizado) sin dejar vacía ninguna casilla significativa de las que está compuesta nuestra ficha terminológica y nuestro glosario (una mención especial se aplicará a los sinónimos, abreviaturas y notas aclaratorias, campos que fueron eliminados si no eran necesarios).

e) El quinto proceso evolutivo radica cuando una unidad terminológica es reemplazada por otra. Por ejemplo, en 1943 *time of attack* era el término normalizado y "*H*" *hour* (AE) y *zero hour* (BE) sus sinónimos. Sin embargo, a partir de 1944 *H-hour* y *zero hour* son considerados términos normalizados por el *US Army*. No obstante, en nuestro corpus

hemos encontrado la variante “*H*” *hour*, pero no la de *H-hour* (ni la de *zero hour*) descrita en el glosario. Ello indica que el uso generalizado de las modificaciones en las representaciones conceptuales se llevaba a cabo de manera paulatina y que los especialistas empleaban, con bastante asiduidad, las unidades terminológicas que habían quedado anticuadas.

f) Así mismo, debemos hacer mención de las nuevas unidades terminológicas, en particular, aquéllas que fueron consideradas como tales mucho después de que concluyera la Segunda Guerra Mundial (*pin-point target, muzzle brake, order of battle,...*). Ante este sexto caso no es nuestro cometido debatir si hubiera sido procedente incluir tales términos *de facto* en los glosarios de la época; nuestra misión es la de dar a conocer un hecho relevante que indica, como ocurre en cualquier lenguaje de especialidad, que un candidato a término suele estar en uso varios años antes de obtener el estatus de unidad terminológica.

g) Finalmente, el séptimo y último grupo lo conforman las unidades terminológicas que carecen de significación. Una vez ratificada su consideración como término tras consultar a un experto en la disciplina, formulamos su definición. En un alto número de ocasiones, ésta ha sido extraída del contexto en donde aparecía la unidad terminológica objeto de estudio. En tal caso, detallamos el archivo fuente de nuestra definición. No obstante, y como ya hemos puntualizado, si la información procedente del corpus resultaba insuficiente para una adecuada explicación, procedimos a formular una definición a modo de propuesta. En este caso, la labor del experto fue indispensable a la hora de ratificar la adecuación o no de la definición formulada. Tras su aprobación, su condición de propuesta quedó reflejada en el glosario (*author’s proposal*) con el propósito de permitir a otros terminólogos o expertos una reformulación o su aceptación como válida (Bowker y Pearson 154).

8.1.2. Evolución de la terminología a nivel morfológico

a) El primer cambio apreciado en las unidades terminológicas entre los años 1943 y 1944 es el paso de su descripción en plural a singular (*identifications / identification, portable obstacles / portable obstacle*). Debido a que el método convencional de

presentación de unidades terminológicas es de singular para nombre, infinitivo para verbos, masculino singular para adjetivos y minúscula para las iniciales (Sager 145), hemos estimado más oportuno continuar esta metodología y mostrar los términos siempre en singular, independientemente del glosario de donde extraigamos su definición.

b) Un segundo caso a destacar es la paulatina eliminación del guión en ciertas unidades terminológicas complejas, a excepción de cuando se emplean como modificador o adjetivo (*second-in-command* / *second in command*). Como ya hemos destacado con anterioridad, nuestro propósito ha sido el de mostrar (siempre que ha sido posible) la unidad terminológica en contexto en su condición como tal aunque, en ciertas ocasiones, la mostramos como modificador, al carecer de otros ejemplos más adecuados. Cuando ello ha ocurrido se procedió a incluir una nota aclaratoria explicando dicho motivo.

c) El tercer caso a destacar es que ciertas unidades terminológicas que aparecen en los glosarios con mayúscula inicial se emplean también en minúscula (quedan descartadas las unidades terminológicas que poseen una diferenciada significación cuando se emplean con mayúscula o minúscula, como es el caso de *Army* / *army* y *Ordnance* / *ordnance*). En algunas ocasiones, su representación conceptual sigue siendo la misma (*Tables of Organization* / *tables of organization*, *General Staff* / *general staff*), aunque otros ejemplos muestran un grado de especificidad más restrictivo que el término genérico (*mark* / *Mark*), mientras que en otros casos se trata de préstamos que se escriben en mayúsculas en la lengua origen y acaban empleándose en minúscula en lengua inglesa (*Flak* / *flak*).

d) En cuarto lugar se observa la alteración de una unidad terminológica compuesta por dos palabras que acaban uniéndose y conformando un término monoléxico. No obstante, existen casos en las que ambas variantes se emplean de manera profusa, siendo incapaces de determinar cuál es la forma correcta, predominante y, por ende, incluyendo ambas en el glosario (*mine field* / *minefield*, *strong point* / *strongpoint*). Otra pauta similar se aprecia cuando una unidad terminológica multiléxica contiene un guión, eliminándose éste con posterioridad (*counter-attack* / *counterattack*, *anti-personnel* / *antipersonnel*), y convirtiéndose en un término monoléxico. Por otro lado, también ocurre la evolución contraria, por la que unidades terminológicas complejas que se empleaban sin guiones comienzan a utilizarse con éste. Se trata más bien de un hecho aislado y (como ya hemos

reseñado arriba) relacionado con el empleo de términos compuestos empleados como modificadores (*close order drill / close-order drill, high explosive shell / high-explosive shell*), aunque no existe consenso alguno sobre su empleo (*lung irritant gas / lung-irritant type*)

e) Una quinta evolución compete a la eliminación de signos de acentuación, debido a que se trata de un préstamo procedente de otra lengua y tiende a obviarse para su adecuación a las normas gramaticales de la lengua receptora, en este caso, la anglosajona (*matériel / materiel*) (Blasucci 217). No obstante, el empleo de dichos préstamos se mantendrá hasta bien entrada la Guerra Fría.

f) Una sexta variación se observa en ciertas unidades terminológicas compuestas por dos sustantivos. Ésta consiste en la permuta del primer sustantivo por un adjetivo con el propósito de poner un mayor énfasis en el uso o funcionamiento de un dispositivo o táctica (*elevation quadrant / elevating quadrant, standard operating procedure / standing operating procedure*). Se trata de una modificación típica en el seno del idioma inglés, en donde la relación adjetivo/nombre describe una acción o procedimiento. No obstante, en ciertas ocasiones ambos sinónimos conviven hasta que uno de ellos, en ocasiones el conformado por dos sustantivos, es el seleccionado definitivamente (*defense area / defensive area*) como unidad terminológica normalizada.

g) El séptimo caso de alteración, similar al anterior, se aprecia cuando ciertas unidades terminológicas compuestas por dos sustantivos cambian el primero de ellos por otro con el que comparten el mismo lema. Esta lematización puede ser definitiva (*Transportation Corps / Transport Corps*) o temporal, en donde ambas versiones subsisten como sinónimos durante la conflagración.

h) La octava variación consiste en un tipo de hipóstasis, que es la sustantivación. Durante el periodo sujeto a estudio se aprecia en los glosarios una clara preferencia por representar las unidades conceptuales por medio de sustantivos en detrimento del empleo de verbos (*defend / defense, repatriate / repatriation, secure / security*), incluso cuando éstos comparten la misma clase gramatical (*command*). Ello no significa que en nuestro corpus no aparezcan ejemplos de tales unidades terminológicas en forma verbal, sino que se trata de una tendencia que ha ido manteniéndose hasta nuestros días.

i) Para concluir este apartado, consideramos digno de destacar la única unidad terminológica en la que se emplea el genitivo sajón: *no man's land* y/o *no-man's-land*. Ésta fue incluida en el glosario editado en 1944. No obstante, y a causa de su profuso uso, dejará de ser considerada como unidad terminológica a finales del pasado siglo, quedando definitivamente integrada en el lenguaje general anglosajón.

8.1.3. Evolución de la terminología a nivel semántico

Debido a que las unidades terminológicas se caracterizan por su condición unívoca, tan solo hemos apreciado dos tipos de cambio en lo que al carácter conceptual se refiere.

a) El primer ejemplo de variación semántica se aprecia cuando la definición de una unidad terminológica se amplía o se detalla con mayor especificidad para delimitar su significación, implementado así su univocidad (Cabré, *Théorie des portes* 8), tal como apreciamos en *Tables of Organization* y *minimum range*.

b) El segundo proceso identificado aparece cuando el sinónimo de una unidad terminológica es empleado para definir una unidad conceptual diferente, convirtiéndose en un nuevo término (*zone defense* mantiene su definición, mientras que *defense in depth* adquiere en 1944 una significación distinta). En otras ocasiones, mediante un proceso de desambiguación, dichos sinónimos parciales dejan de serlo para representar unidades conceptuales diferentes (*harassing fire* / *interdiction fire*). No obstante, existen otros que mantienen su sinonimia durante el transcurso de la guerra (*prearranged fire* / *schedule fire*), los cuales caerán finalmente en desuso tras la conflagración.

8.1.4. Britanización del lenguaje especializado en el inglés americano

Nos hemos tomado el atrevimiento de acuñar este nuevo término para detallar el proceso contrario denominado *americanización* de la lengua inglesa (Baker, Hardie y McEnery 9). En 1943 se constata la existencia del empleo de unidades terminológicas diferentes por parte del US Army y del British Army y su uso a modo de sinónimos.

a) Un primer ejemplo de este proceso consiste en la caída en desuso de la unidad terminológica empleada en el inglés americano, sustituyéndola por el sinónimo británico (*agency of signal communication / signal unit, automatic supply / normal supply*) o, aún no extinguiéndose por completo, su uso se convierte en un mero hecho anecdótico (*assault fire / fire on the move*). Una variante de dicha evolución ocurre cuando el término norteamericano deja de emplearse y el británico aparece sin definición en el diccionario consultado, en el que se detalla únicamente su acrónimo (*oral order / verbal order*). Un caso diferente se aprecia cuando en nuestro corpus solo se muestra la unidad terminológica británica (*track discipline / march discipline, position of readiness / position in readiness, movable obstacle / portable obstacle,...*). Ya que dichas unidades terminológicas siguen manteniéndose en el glosario fechado en 1944, ello puede deberse a que nuestro corpus no es tan extenso como debiera o que se ha iniciado un proceso de reemplazo terminológico. Sea como fuere, ninguna de estas unidades terminológicas aparecerá en el diccionario del Ejército norteamericano publicado en 1950.

b) Una segunda variante de esta evolución la constituye la adopción de la unidad terminológica británica, pero empleada para representar una unidad conceptual diferente (*plan of attack* deja de ser sinónimo de *scheme of maneuver* y adquiere una significación propia, al igual que *forward area* con respecto a *combat zone*). De esa manera, una unidad terminológica adquiere una nueva unidad conceptual que hasta el momento carecía de representación léxica.

c) Un tercer caso acaece cuando en nuestro repositorio textual aparecen tanto la unidad terminológica británica como la norteamericana (*fire on fixed line / fixed fire, verbal order / oral order, operation orders / combat orders*). Este caso no resulta ni significativo ni concluyente, ya que puede deberse a que los autores de los artículos procedan del Imperio británico o al empleo del término del British Army en el seno del US Army. No obstante, resulta conveniente tomar en consideración el uso de ambos y tener presente el número de ocasiones que aparece en nuestro corpus. Una cifra alta implicaría que su empleo ha trascendido y es usado por los militares norteamericanos; una cifra pequeña puede significar que se haya utilizado tan solo en un artículo redactado por un oficial británico y no se use en el seno del Ejército norteamericano. Aunque no nos ha sido posible conocer ni la nacionalidad ni la identidad de los escritores de los artículos en los que aparecen dichas unidades terminológicas, la preferencia en el empleo del término británico, en

detrimento del norteamericano (3-1, 7-1 y 4-2, respectivamente), resulta un hecho más que notorio a tener en consideración en lo que respecta al uso y la evolución de la terminología.

8.2. EL TRABAJO DE EXTRACCIÓN TERMINOLÓGICO: VALORACIÓN

En términos generales, podemos constatar que la elaboración del trabajo terminológico depende fundamentalmente del terminólogo y de los objetivos que éste desee cumplir. No cabe duda que tanto las herramientas de extracción terminológica y de lingüística de corpus son de gran ayuda, pero no son determinantes. A la vista de los resultados iniciales, es el especialista el encargado de desarrollar las metodologías e interpretar los datos resultantes (Vargas Sierra 80), tales como refutar o eliminar cada candidato a término, y es en dicho proceso cuando estos programas muestran su gran utilidad: el análisis de las posibles unidades terminológicas dentro de un contexto especializado. Resulta lógico suponer que este tipo de *software* se irá perfeccionando con el paso del tiempo, pero nunca podrá reemplazar el factor humano.

No obstante, y tal como hemos reseñado con anterioridad, resulta inevitable que la labor del investigador no se limite a emplear tales programas, sino a aprender unas nociones básicas de programación para adaptar las herramientas existentes a sus propias necesidades (Anthony 155). Se trata tan solo de una mera opinión, pero estimamos que dicho criterio se convertirá en una necesidad en un futuro próximo. Si el terminólogo desea que las herramientas a las que tiene acceso faciliten su trabajo de extracción e identificación terminológicas, éste deberá ser capaz, al menos, de modificar los parámetros de búsqueda e identificación de términos y de adecuar el programa para satisfacer los objetivos en cada trabajo.

SDL MultiTerm 2009 Extract está concebido para llevar a cabo el proceso de extracción terminológica desde una concepción puramente estadística. Este enfoque, dentro del marco de la lingüística computacional y de la lingüística de corpus, ha sido empleado casi desde sus comienzos, y se fundamenta (principalmente) en técnicas de análisis subordinadas a propiedades estadísticas de la estructura del lenguaje y de ejemplos procedentes del *lenguaje real*, que no inventado o idealizado (Garside, Leech y

Sampson 16). De igual modo, dentro del marco de la terminología, el proceso estadístico está completamente automatizado y depende del programa, y no del especialista, el proceso de extracción de los resultados.

El programa informático crea una serie de relaciones de palabras, por medio de lexías, lemas o de la flexión de éstas o de morfemas, provocando unas asociaciones equívocas extraídas sin tener en consideración el contexto en el que se encuentran. Este contratiempo se agudiza a la hora de identificar relaciones multiléxicas. Si por medio de la estadística, una relación cualquiera se produce en un número significativo para el programa, queda incluida en el archivo de vaciado de términos. Esta dificultad se manifiesta más si cabe cuando se detecta un préstamo de otra lengua, puesto que, al ser una palabra extraña para el *software* (no incluida en su archivo de palabras de exclusión), será extraída automáticamente y considerada inicialmente como candidato a término.

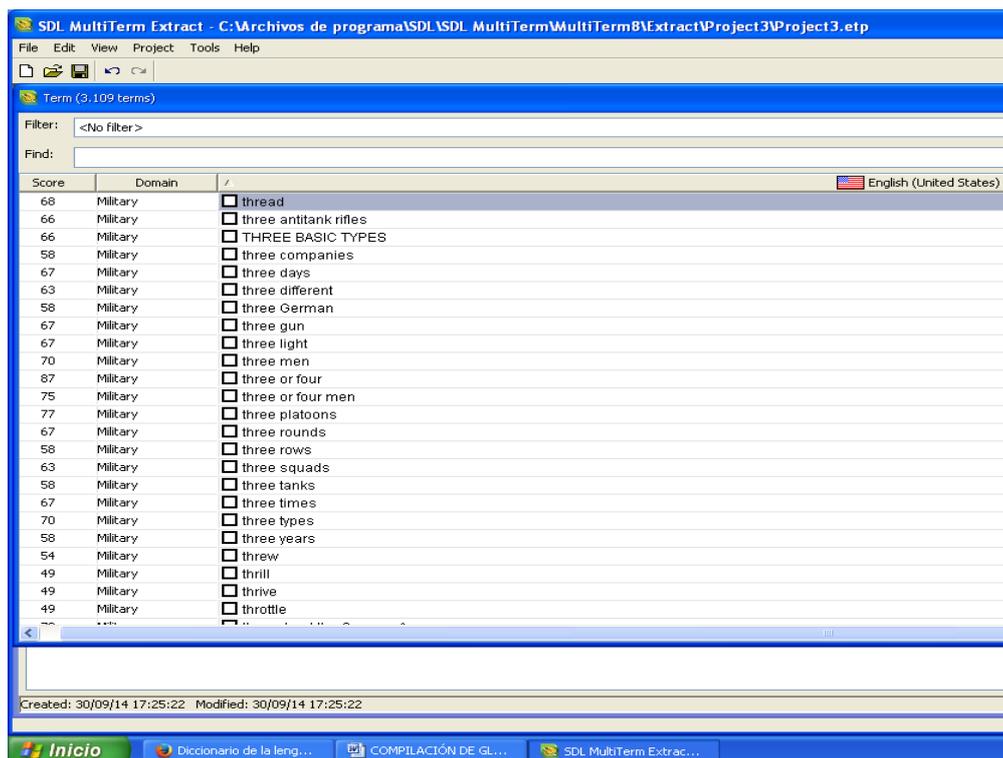


Figura 44. Inclusión en el proceso de extracción terminológica de unidades léxicas compuestas por la lexía *three*.

Otra limitante identificada radica en que a la hora de generar frases en contexto, SDL MultiTerm 2009 Extract suele incluir tanto la flexión de un verbo como todas las variaciones pertenecientes a su lema. *A priori*, este hecho debe considerarse como un avance, puesto que permite localizar un amplio número de ejemplos. No obstante, al añadir sus formas sustantivadas, provoca un alto índice de ruido cuando éstas no tienen significación en el campo de especialidad objeto de estudio. Ante ello, deberemos proceder a desechar enunciados en los que aparece una forma errónea del candidato a término sujeto a análisis que no posee relación alguna con nuestra investigación.

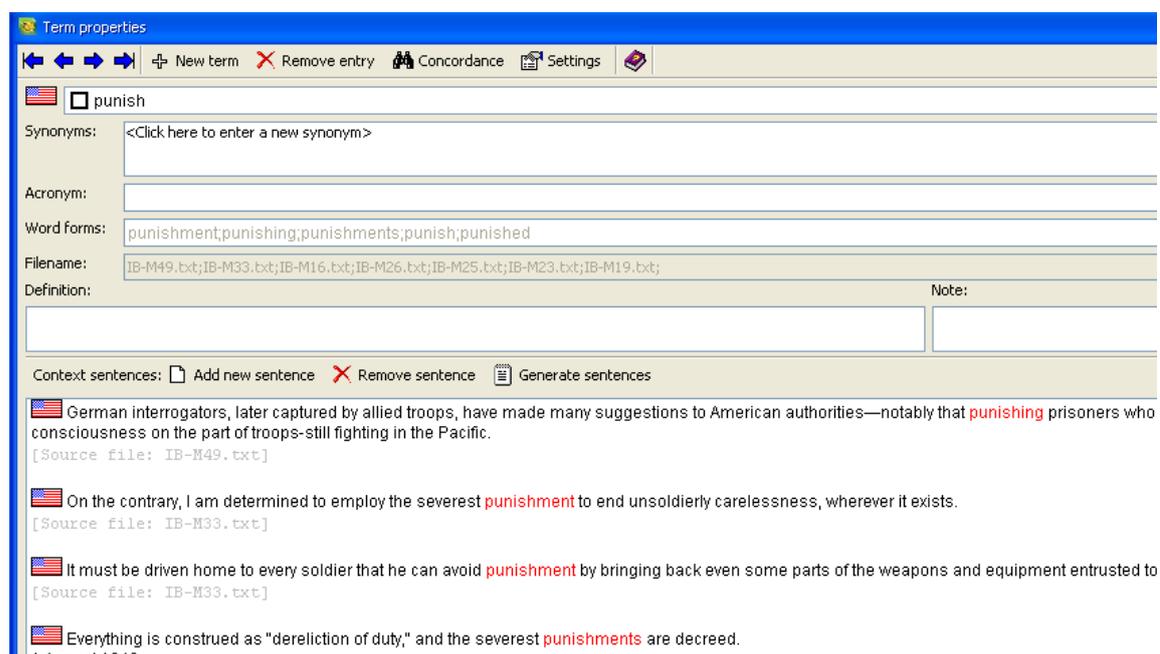


Figura 45. Lematización incorrecta del verbo *punish* al incluir su forma nominal *punishment*.

Otra dificultad añadida es la producida por la puntuación (en especial, la que aparece en los acrónimos y en las abreviaturas) a la hora de seleccionar frases en contexto. El resultado es la aparición de una unidad terminológica incompleta que deberemos extraer de AntConc 3.3.5.w (Windows) e incorporarla a nuestro fichero de vaciado terminológico.

En el ejemplo mostrado en la figura 46 constatamos que en la primera frase (destacada en azul) el nombre correcto de la unidad no se muestra por completo al estar compuesta por una abreviatura que contiene un punto. El programa basado en la frecuencia estadística aplica una pauta por la que un punto confiere la finalización de una frase, siendo proclive

a errar durante el proceso de extracción y vaciado terminológicos. La denominación exacta de la unidad terminológica destacada es *2nd Med. Bomb Wing*. Además, ello tampoco hubiera permitido la identificación de *wing* como unidad táctica y administrativa de Ejército del Aire, desambiguándola así mismo de la definición que posee en el lenguaje general anglosajón.

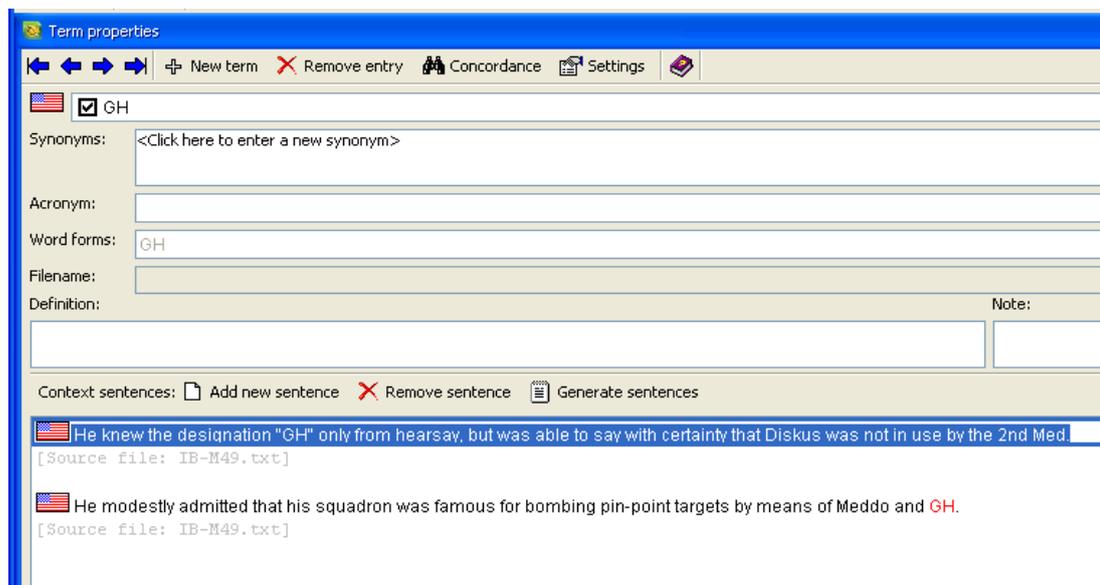


Figura 46. Interpretación errónea de un punto al extraer una frase en contexto.

Una deficiencia si cabe mayor del programa desarrollado por SDL Technologies es que, en ocasiones, la búsqueda de una unidad terminológica que incluya acrónimos o abreviaturas con puntos o guiones no proporciona resultado alguno, a pesar de conocer la existencia de éstos. Habíamos considerado la posibilidad de que, durante el proceso de conversión a formato sencillo, los espacios entre palabras pudieran aumentar o disminuir, creando una distancia excesiva entre voces (o, por el contrario, la unión de unas palabras con otras) que puede confundir al programa y provocar que no proporcione ejemplos en contexto. Lamentablemente, no se trataba de dicha causa tras analizar los resultados en AntConc 3.3.5.w (Windows) y extraer (con suma facilidad) los ejemplos desde este *software* para, de ese modo, añadirlos en la ficha terminológica.

En el caso mostrado en la figura 47 se puede comprobar que el programa SDL MultiTerm 2009 Extract afirma la inexistencia de frases en contexto para la unidad

terminológica *U.S. Army*, mientras que en AntConc 3.3.5.w (Windows) aparecen veintiséis ejemplos.

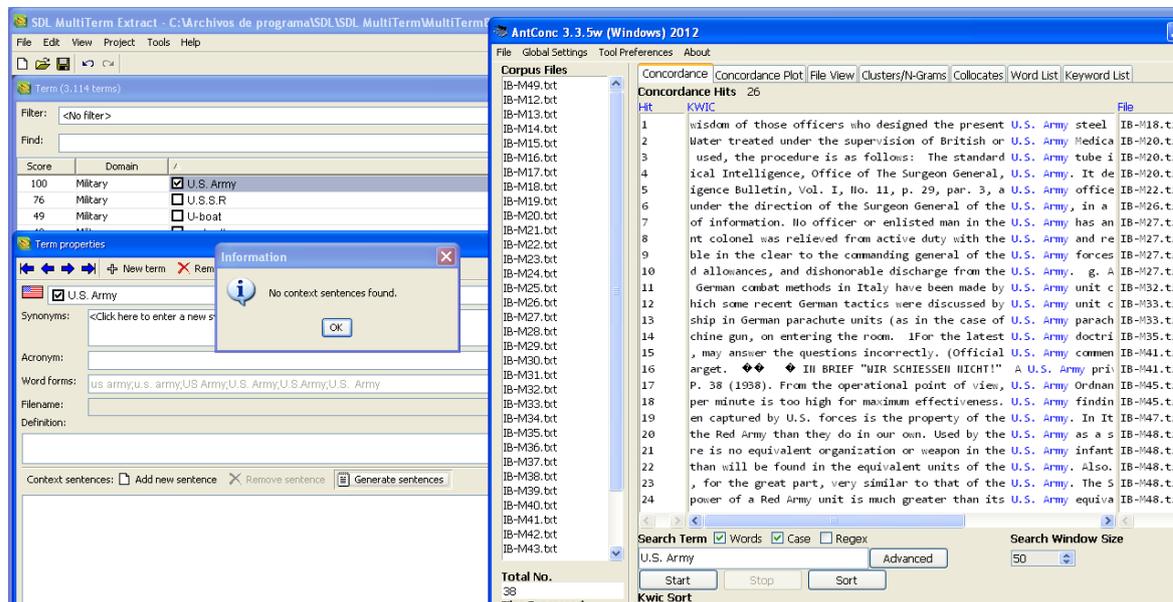


Figura 47. Comparación de resultados entre SDL MultiTerm 2009 Extract y AntConc 3.3.5.w (Windows).

Por otro lado, a la hora de realizar la búsqueda de una unidad terminológica cualquiera, SDL MultiTerm 2009 Extract no realiza distinciones en lo que se refiere a la manera en la que se encuentra ésta escrita (en mayúsculas, minúsculas o con la primera letra en mayúscula). Como es sabido, en el ámbito militar una unidad terminológica que comience con minúscula puede representar una unidad conceptual distinta (o menos genérica) de la que representa la que se escribe con mayúscula. Si a ello le añadimos el hecho de que las palabras a principio de frase se escriben con mayúsculas, la aparición de ruido se convierte en un hecho frecuente.

En el ejemplo mostrado a continuación apreciamos que SDL MultiTerm 2009 Extract no diferencia entre *army* (unidad táctica terrestre compuesta por varias divisiones o cuerpos de ejército) y *Army* (conjunto de fuerzas militares que posee una nación, a excepción de las fuerzas navales y, en algunos países, de las fuerzas aéreas). En ciertas ocasiones, dicha malinterpretación no permite la desambiguación y puede provocar la

confusión entre dos unidades terminológicas diferenciadas, pero con la misma representación léxica, y ser consideradas como una sola.

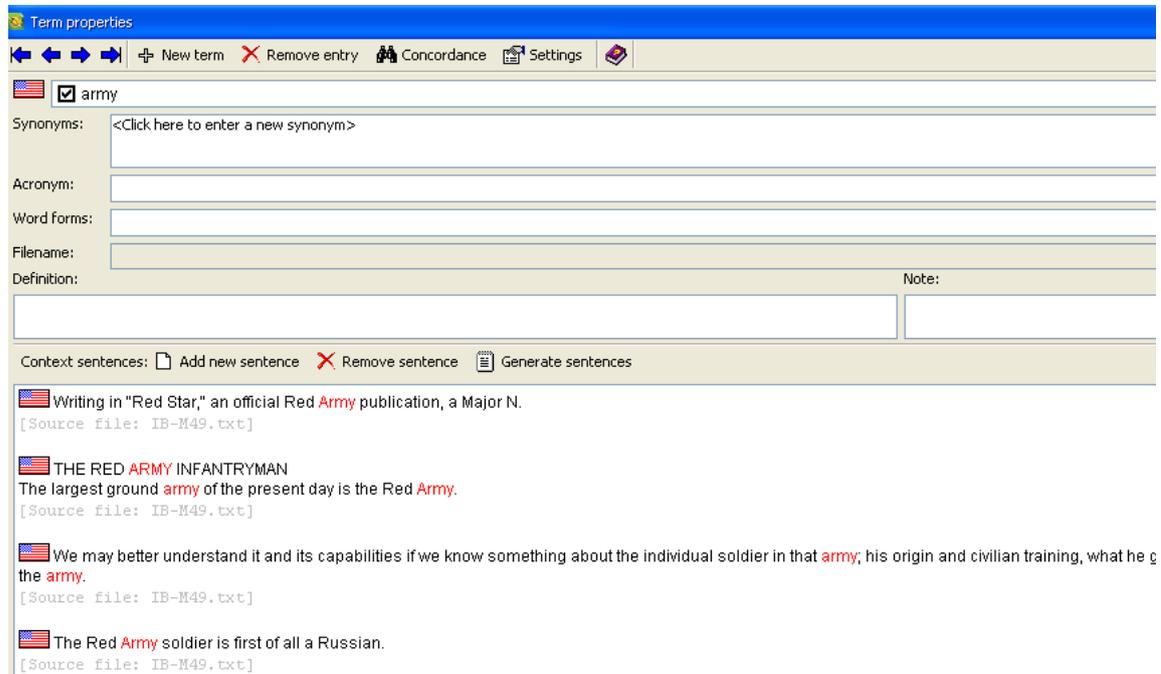


Figura 48. Interpretación equívoca en el uso de mayúsculas y minúsculas

Ya hemos destacado con anterioridad que un error que cometimos, sin ser conscientes de ello, y que debimos paliar al emplear SDL MultiTerm 2009 Extract, fue que al añadir nuevos ficheros el programa eliminaba algunas entradas previamente validadas e incluidas en el fichero de vaciado. Inicialmente, considerábamos que esta acción se ejecutaba de manera aleatoria, pero la explicación a esta circunstancia radica en que cada vez que el programa exploraba un texto eliminaba todas las palabras que encontraba y que estaban almacenadas en su archivo de palabras de exclusión. Por tanto, voces como *action* y *company* fueron de nuevo borradas de nuestro fichero. Ello nos obligó, por tanto, a volver a revisar el archivo de exclusión de términos y a rescatar dichas unidades terminológicas que, como ya hemos precisado, también poseen la condición de unidad léxica.

En consecuencia, e independientemente de la elección que tomemos (añadir todos los textos al programa de una sola vez o de manera paulatina), resulta inevitable la aparición de un alto porcentaje de ruido, ya sea porque el programa estima que ciertos patrones de

relación entre palabras que se repiten son unidades terminológicas complejas, o debido a que el empleo de acrónimos en mayúscula o la aparición de palabras en otros idiomas sean catalogadas automáticamente como términos.

No obstante, estimamos más adecuado el empleo del archivo de exclusión de términos, y no su eliminación, debido a que es más coherente realizar una búsqueda al ser conscientes de las unidades léxicas que conforman tal fichero. Tanto si la indagación se realiza en el programa empleado para la búsqueda mediante el análisis estadístico o lingüístico, o mediante el empleo del corpus textual simple o anotado, damos como hecho contrastado que la indagación se iba a centrar, única y exclusivamente, en los vocablos que figuran en esa carpeta y en las relaciones sintácticas en las que puedan aparecer. No utilizar el archivo de exclusión de términos implicaría una búsqueda más dilatada, al desconocer qué vocablos pertenecientes al lenguaje general inglés pueden aparecer como candidatos a término y cuáles serían sus relaciones sintácticas más proclives de asociación.

Otro aspecto a destacar es el resultado que SDL MultiTerm 2009 Extract adjudica a cada entrada. Cada unidad léxico/terminológica almacenada en el archivo de vaciado terminológico recibe un porcentaje, siendo el mínimo observado el de 48 y el máximo el de 100. No obstante, esta calificación no resulta concluyente. Como norma general, el valor de un candidato a término alcanza el 48 y consigue ascender hasta el 99. Pero ello no quiere decir que el vocablo que obtenga un mayor porcentaje posea más posibilidades de catalogarse como término. Pongamos como ejemplos *misfire* (48%) y *men* (99%). A pesar de su alta puntuación, *men* pertenece al lenguaje común, mientras que *misfire* es un término militar que significa que un proyectil no ha sido disparado (a causa de la falta de ignición de la carga propulsora, fallo del percutor, etc.) o que no ha explotado (o no lo ha hecho adecuadamente) al llegar al objetivo. En consecuencia, la puntuación no nos proporciona una información relevante durante el vaciado terminológico, a no ser que sea necesario que una palabra alcance el umbral del 48% para ser incluida en el fichero por el *software*.

No obstante, no todos los términos validados alcanzan una alta puntuación. Como norma general, éstos obtienen una calificación superior a 60, pero no necesariamente alcanzan el 90 o el 100 por ciento. En el caso de *Motpulk*, *muzzle velocity*, *navigator* y

Nazi, su puntuación oscila entre el 77 y el 82 por ciento, mientras que *motorized infantry*, *mountain troops*, *moving target*, *Nazism*, *neck strap* y *neutral country* alcanzan el 100%. Ello indica que las unidades terminológicas complejas obtienen, generalmente, una alta valoración, en detrimento de las compuestas por una unidad léxica. De igual modo, el hecho de que en una entrada queden rellenos todos los campos (dicho de otro modo: que se hayan completado los campos que comprende la ficha terminológica que muestra SDL MultiTerm 2009 Extract) no implica que dicho término consiga la evaluación más alta. Por otro lado, el número de ocasiones en el que aparece un vocablo especializado en los textos influye en su valoración, pero no es del todo concluyente (Bowker y Pearson 149). De igual manera, existen múltiples voces especializadas que, al no emplearse con asiduidad, se les asignará un número relativamente bajo, pero ello no quiere decir que sean menos importantes que las que se emplean con mayor regularidad. Teniendo en cuenta el factor de la recurrencia, es necesario mantener bajo el número de repeticiones (tanto en SDL MultiTerm 2009 Extract como en AntConc 3.3.5. Windows) para no excluir tales términos de nuestro fichero de vaciado.

The screenshot shows the SDL MultiTerm Extract application window. The title bar reads "SDL MultiTerm Extract - C:\Archivos de programa\SDL\SDL MultiTerm\MultiTerm8\Extract\Project3". The menu bar includes "File", "Edit", "View", "Project", "Tools", and "Help". Below the menu bar is a toolbar with icons for file operations. The main area is titled "Term (1,308 terms)" and contains a "Filter:" field with the text "<No filter>" and a "Find:" field. Below these fields is a table with three columns: "Score", "Domain", and a list of terms with checkboxes. The table is sorted by score in descending order.

Score	Domain	Term
72	Military	<input type="checkbox"/> meeting engagement
99	Military	<input type="checkbox"/> men
100	<None>antit	<input checked="" type="checkbox"/> message center
100	Military	<input checked="" type="checkbox"/> messenger
67	Military	<input type="checkbox"/> messenger dog
100	<None>antit	<input checked="" type="checkbox"/> metal troughs
100	<None>antit	<input checked="" type="checkbox"/> meteorological message
48	Military	<input type="checkbox"/> microphone
77	Military	<input type="checkbox"/> mil
60	Military	<input type="checkbox"/> mileage indicated
95	Military	<input type="checkbox"/> miles per hour
100	<None>antit	<input checked="" type="checkbox"/> military establishment
100	<None>antit	<input checked="" type="checkbox"/> military map
100	<None>antit	<input checked="" type="checkbox"/> Military Police
100	<None>antit	<input checked="" type="checkbox"/> military service
100	Military	<input checked="" type="checkbox"/> military training
100	<None>antit	<input checked="" type="checkbox"/> mine detector
99	Military	<input checked="" type="checkbox"/> minefield
99	Military	<input type="checkbox"/> mines
82	Military	<input type="checkbox"/> mines in a row
71	Military	<input type="checkbox"/> Mines In front
64	Military	<input type="checkbox"/> miscellaneous
48	Military	<input type="checkbox"/> misfire
48	Military	<input type="checkbox"/> mist
72	Military	<input type="checkbox"/> mm gun

Figura 49. Vista del porcentaje (*Score*) que otorga el programa SDL MultiTerm 2009 Extract a los candidatos a término.

Al contrario que SDL MultiTerm 2009 Extract, AntConc 3.3.5.w (Windows) no posibilita la extracción de unidades terminológicas *per se*. No obstante, y como ya hemos destacado en el capítulo dedicado a la extracción terminológica lingüística, se trata de una herramienta muy útil que también posee la capacidad de acometer análisis estadísticos. El más simple de ellos (y mencionado con anterioridad en varias ocasiones), pero no el menos importante, es el de solicitar un mero recuento de palabras. El programa no sólo nos muestra la cantidad total de vocablos que albergan los textos, sino la frecuencia con la que aparecen, mostrándolas de manera jerarquizada (McEnery, *Introduction* 3). Como ya hemos puntualizado con anterioridad, las primeras posiciones las ocupan artículos, preposiciones, conjunciones, pronombres y verbos de uso común. Tras ellos, comienzan a aparecer sustantivos, adjetivos y verbos y, algunos, solo por su frecuencia de uso, merecen un análisis pormenorizado y son incorporados a nuestro fichero de vaciado de términos. A pesar de no ser una herramienta de extracción terminológica, su versatilidad y facilidad de uso permite su empleo para este tipo de trabajos.

Con respecto a los glosarios que hemos consultado durante nuestra investigación (a pesar de que la mayoría son denominados *diccionarios*), constatamos que su principal propósito es el descriptivo, dejando en un segundo plano el carácter lingüístico de las unidades terminológicas. Teniendo en cuenta que el nivel intelectual y formativo del personal que conforma todo un ejército en tiempos de guerra es muy heterogéneo, las definiciones se diseñaron (en la medida de lo posible) a un nivel de simplicidad tal que pudiera ser comprensible para cualquier soldado (US, *Dictionary of United States Army Terms* iv). Ello ha implicado que hechos o procedimientos altamente especializados queden definidos únicamente con el propósito de proporcionar una idea general de las mismas.

Si analizamos las unidades terminológicas procedentes de otras áreas de especialidad, apreciamos una falta de coherencia a la hora de su inclusión en los glosarios consultados. Tomando como ejemplo el campo de la automoción, constatamos que se incluyen algunos de los componentes más característicos que conforman los vehículos terrestres (*chassis, track,...*), mientras que otros no (*wheel, engine, carburetor, drive shaft...*). Lo mismo ocurre si incrementamos el grado de especialización. Dentro de la estructura conceptual representada por los vehículos propulsados por cadenas tractoras, si detallamos los elementos específicos implicados en su movimiento, aparecen unos (*track, bogie*), en

detrimento de otros (*idler wheel, drive sprocket, return roller, suspension system*). Estimamos que se podría haber usado otro método, desde la inclusión de todas las unidades terminológicas empleadas en el campo de la automoción, hasta la publicación de tan solo los términos más especializados y desconocidos para el público no experto.

Desde el punto de vista semántico, desconocemos la metodología que han seguido sus autores por la cual ciertos verbos que ostentan una significación más cercana al lenguaje general (*reduce, reorganize, retire,...*) son incluidos en el glosario, mientras que otros, poseedores de una clara connotación bélica (*repel, repulse, seize,...*), han quedado excluidos. Una posible causa pudiera ser que, con el propósito de restringir el número de unidades terminológicas, se hubiera decidido eliminar las entradas cuyo uso fuera más generalizado en el seno del lenguaje común inglés.

Otro hecho a poner de manifiesto es la inexactitud en lo que al número de palabras a nuestra disposición se refiere. El programa SDL MultiTerm 2009 Extract contabilizaba 337.616 palabras, mientras que AntConc 3.3.5.w (Windows) disminuía dicha cifra a 285.214. Debido a que el *software* de extracción terminológica no proporciona una cifra total, sino que muestra el de cada uno de los archivos, volvimos a hacer la cuenta en AntConc 3.3.5.w (Windows) con textos sin etiquetar, empleando todas las posibilidades existentes (búsqueda sin palabras, de palabras, con distinción entre mayúsculas y minúsculas y por expresiones) ordenando los resultados por frecuencia, palabras y por terminación de las palabras. A pesar de todo, el resultado seguía siendo el mismo: 285.214 palabras.

Inicialmente no podíamos explicar una diferencia de 52.402 voces y no sabíamos cuál de las dos cifras tomar como válida. Sin embargo, como se observa en la figura 50 que en la vigésimo séptima posición aparecía *panzerfa* en vez de *panzerfaust* (*hollow-charge launcher, a one-man recoilless weapon*), consideramos factible la posibilidad (entre otras) de la aparición de errores gramaticales durante el proceso de conversión de los textos a formato txt. Por tanto, y con el propósito de conocer el origen de tal anomalía, extrajimos tres párrafos del archivo IB-M32.txt que contenían un total de 203 palabras, 3 abreviaturas, 1 acrónimo y 11 cifras. Así mismo, también aparecían 3 unidades terminológicas complejas y una palabra unida a un guión, además de corchetes y comillas.

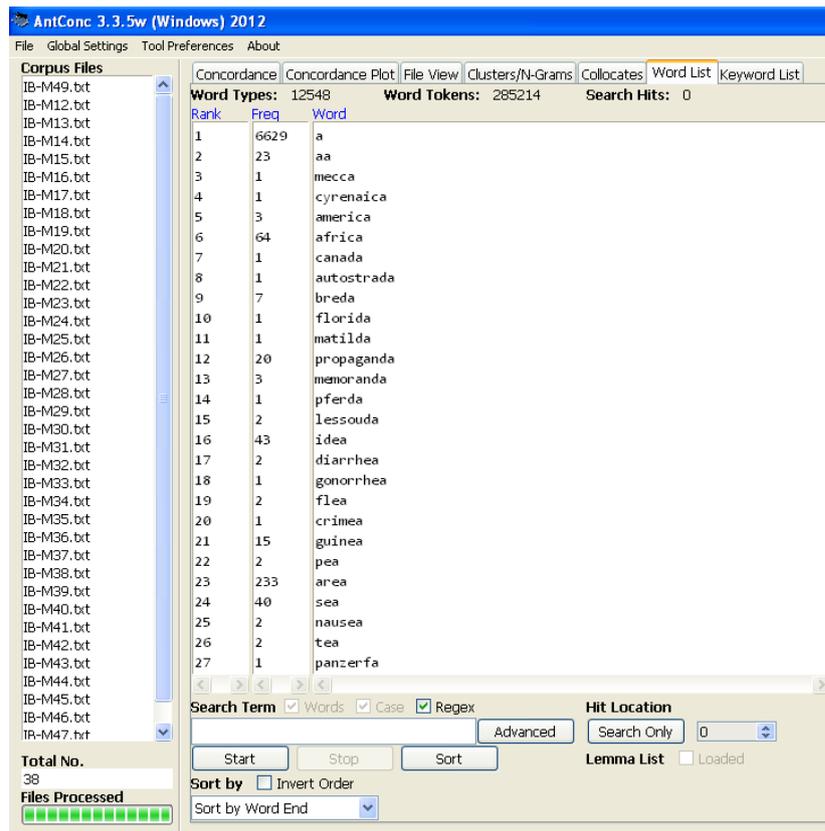


Figura 50. Resultado del recuento de palabras con el programa AntConc 3.3.5.w (Windows).

Seguidamente, analizamos el texto, cuya denominación es Test.txt, por medio de AntConc 3.3.5.w (Windows). A la hora de solicitar un recuento de palabras, dicha indagación nos proporcionó un total de 212: nueve más que en nuestro cómputo. Al analizar los resultados apreciamos que el programa contabilizó cada vocábulo como tal, independientemente de que se trate o forme parte de una unidad multiléxica o terminológica compleja unida con guiones (*machine-gun position*, *trip-wire belt*, *zig-zagged approach trench*), considerando como voz todo acrónimo que estuviera compuesto por cualquier tipo de puntuación (como *fig.*, *pp.* y *vol.* Por el contrario, *U.S.* se calculaba como dos palabras: *u* y *s*) y la cifra II (*Vol. II*) era tomada como una palabra (ii). Por tanto, con este balance obtuvimos 9 palabras más; una cifra total de 212, del todo inexacta.

Con el propósito de realizar un estudio comparativo, realizamos el mismo proceso con SDL MultiTerm 2009 Extract. Para ello, creamos un nuevo proyecto y cargamos el mismo archivo Test.txt. El resultado del recuento fue de 256 palabras: una cifra también errónea. La suma total de voces, acrónimos, abreviaturas y cifras ascendía a 221, por lo que ello

quiere decir que el *software* había contabilizado como vocablos todos los signos ortográficos (40), a excepción de los guiones (5).

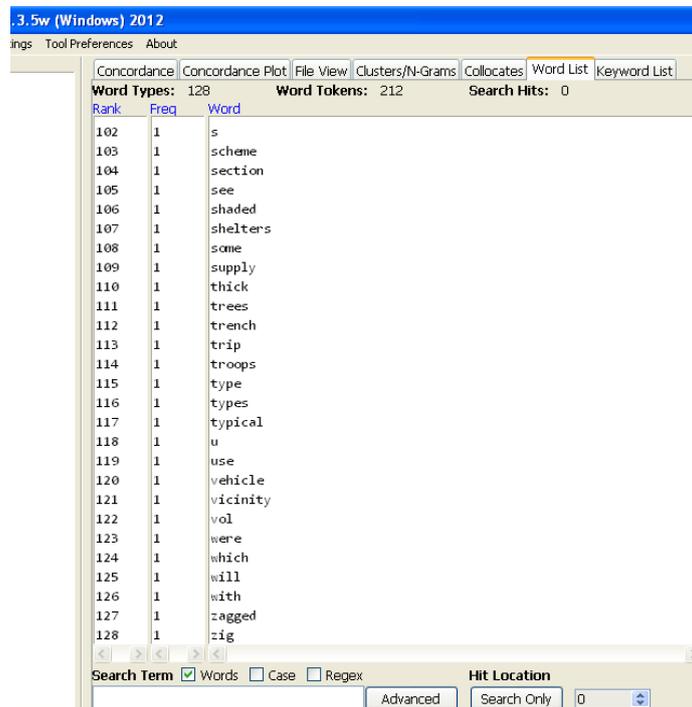


Figura 51. Recuento del archivo Test.txt con AntConc 3.3.5.w (Windows).

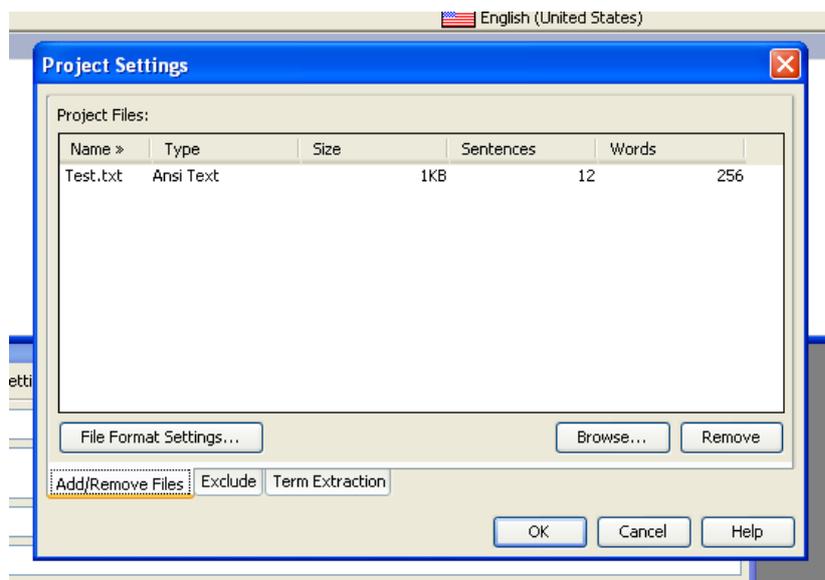


Figura 52. Resultado de la contabilización de voces del archivo Test.txt con SDL MultiTerm 2009 Extract.

Un hecho evidente es que una misma investigación realizada en el seno de la lingüística de corpus con diferentes herramientas tiende a no coincidir, siendo uno de los errores más evidentes el recuento y la frecuencia de aparición de palabras. Para ello, es necesario determinar qué considera como palabra el programa que empleamos (Anthony 149). El diseñador de AntConc 3.3.5.w (Windows) realizó un estudio comparativo del discurso inaugural de la presidencia de los Estados Unidos de 2009 con su *software*, WordSmith Tools y MonoConc Pro. El resultado más relevante fue que AntConc 3.3.5.w (Windows) consideró que la aparición de apóstrofes en una palabra implicaba la existencia de dos, mientras que las otras herramientas estimaron que cada apóstrofe era parte integrante de cada vocablo (*Ibid*). En consecuencia, *we'll* fue computado erróneamente como una palabra por WordSmith Tools y MonoConc Pro, y como dos por AntConc 3.3.5.w (Windows).

Una de las deficiencias más significativas de las herramientas a disposición de la lingüística de corpus (en particular, las herramientas de la tercera y cuarta generaciones) es que no muestran su configuración y, por tanto, desconocemos qué consideran qué es una palabra. Como respuesta a éste y otros problemas algunos expertos animan a los investigadores a aprender a programar y a desarrollar sus propias herramientas de análisis conforme a sus propios criterios y necesidades (Biber et al 256). El empleo de herramientas actualmente en el mercado obliga al investigador a limitarse a los parámetros que ha creado el diseñador para un programa en particular, mientras que si el lingüista idea sus propias herramientas puede controlar, en toda su amplitud, su trabajo de investigación (Gries 11-2). En definitiva, y debido a los diferentes resultados obtenidos, únicamente podemos estimar la cifra total de palabras que nos proporcionó AntConc 3.3.5.w (285.214) como la más cercana a la realidad, aunque no podemos asegurar con total exactitud que sea del todo correcta.

9. CONCLUSIONES

La primera deducción a la que hemos llegado al concluir este estudio es que el trabajo de compilación y extracción terminológicas es una labor costosa, puesto que implica un importante esfuerzo tanto de tiempo como de recursos. En consecuencia, el fruto de dicho trabajo debe ser reutilizado para que otros usuarios puedan tomar provecho de éste y, si procede, emplearse como base para futuras investigaciones.

El terminólogo se encuentra frente a una dicotomía: la necesidad de emplear herramientas para agilizar su labor, y la pertinencia de almacenar los datos resultantes en un entorno de fácil uso para su posterior empleo o consulta por terceros. Para ello, resulta de vital importancia la creación de programas informáticos que agilicen dicha labor. Uno de los objetivos a alcanzar sería el diseño de un *software* capaz de realizar dicho proceso tanto mediante una base estadística como lingüística. Pensamos que la fusión de ambas modalidades ocasionaría una mayor productividad a la hora de identificar, con una mayor exactitud, los posibles candidatos a término.

Desde el enfoque lingüístico, estimamos de gran ayuda la posibilidad de modificar ciertos parámetros del programa, como los de seleccionar los patrones lingüísticos específicos por los que se asocian las unidades terminológicas complejas en cada disciplina antes de iniciar el proceso de vaciado terminológico. Estos patrones no solo varían entre lenguas, sino también entre disciplinas. Así mismo, consideramos indispensable que el programa identifique correctamente los signos de puntuación, en particular lo que se refiere a los puntos y los acentos. De igual manera, ya hemos constatado que SDL MultiTerm 2009 Extract no identifica los acrónimos (puesto que considera cada punto como punto seguido o aparte), hecho que provoca un exceso de silencio, aunque sí reconoce los signos ortográficos, en particular, la tilde y la diéresis. Por el contrario, AntConc 3.3.5.w (Windows) posibilita la búsqueda de acrónimos, al no dividir las palabras o las sílabas cada vez que encuentra un punto. Al mismo tiempo, AntConc 3.3.5.w (Windows) sí reconoce los signos ortográficos, aunque debemos tener la precaución de seleccionar un sistema de codificación de caracteres compatible con las lenguas empleadas en nuestro trabajo.

Se trata de un asunto de importancia relevante ya que, en muchos tipos de textos especializados, la unidad terminológica se detalla al inicio de la obra, empleándose posteriormente su acrónimo o su abreviatura, perdiéndose una información vital si ésta no se identifica, la cual debe recuperarse por otros medios. Con respecto a la acentuación, si el programa es incapaz de detectarla, se produce un elevado índice de silencio que, en particular en las lenguas romances, representaría un contratiempo de grandes magnitudes, que haría necesaria otra metodología para reconocer todas las unidades terminológicas en las que se emplea este signo ortográfico.

Otro factor que puede mejorar el trabajo terminológico sería la posibilidad de modificar el archivo de exclusión de términos. Debido a que el número de palabras es finito, muchas de éstas tienen la particularidad de poseer, al mismo tiempo, el estatus de unidad léxica y unidad terminológica. No obstante, y dependiendo de la disciplina, las palabras (o un buen número de ellas) que poseen dicha característica son de sobra conocidas por los expertos, pudiendo quedar eliminadas del mencionado fichero para obtener un mayor número de candidatos a término durante el proceso de extracción terminológica. De igual modo, sería adecuado contar con la posibilidad de incluir las palabras que estimemos oportuno a dicho fichero con el mismo propósito. En nuestro campo de estudio, unidades léxicas tales como *action*, *chief*, *company*, *field*, *general*, *landing* o *training* poseen dicha peculiaridad. Ello no solo implica su ausencia en el fichero de vaciado, sino también sus asociaciones con otras por las que generan unidades terminológicas complejas, las cuales tampoco serían detectadas en dicho proceso.

Lo anteriormente comentado se refiere al trabajo de extracción terminológica. En lo que concierne al almacenamiento y uso de la terminología, resulta primordial la posibilidad de exportar la información almacenada en nuestro fichero de vaciado terminológico a una base de datos o glosario para que se pueda consultar con facilidad. En nuestro caso, el empleo de programas que no permiten una estructuración adecuada de la ficha terminológica según el estudio, la imposibilidad de su lógica ordenación, diseñada a nuestras propias necesidades, y la incompatibilidad de estos programas con otros incluidos en el mismo paquete informático entre sí, han hecho necesaria la creación del glosario en Microsoft Word. A pesar de que se han realizado mejoras en el paquete informático empleado (SDL Trados Studio 2009), sus actualizaciones dejan de ser las más idóneas; en particular, en el caso de la exportación de los datos resultantes de una investigación.

Por otro lado, el trabajo terminológico está íntimamente relacionado con el empleo del corpus; no se trata de un mero repositorio de ejemplos en contexto, sino de una metodología por la que analizar eficazmente el comportamiento y la evolución de las unidades terminológicas. Partiendo de la premisa de que la comunicación escrita es el medio más importante de divulgación científica, sería desacertado no emplear la lingüística de corpus en el estudio de la terminología. Independientemente de que el estudio sea sincrónico o diacrónico, monolingüe, bilingüe o plurilingüe, el corpus proporciona ejemplos reales y representativos del uso de la terminología durante la transmisión del conocimiento.

La lingüística de corpus se encuentra presente en los bancos terminológicos. Estos bancos son, desde nuestro punto de vista, el medio más eficaz para presentar y divulgar la terminología. Si nos referimos a su actualización, los bancos de datos terminológicos proporcionan una información fiable y abundante del empleo de las unidades terminológicas, condición de la que adolecen los glosarios tradicionales. La adición de nuevos textos al corpus, su análisis, la incorporación de información extralingüística y su etiquetado gramatical proporcionan una información de incalculable valor para constatar el uso y la evolución de la terminología. El empleo de ambas metodologías (o disciplinas) en el estudio de las representaciones conceptuales de la terminología constituye una mejora exponencial en el trabajo terminológico. No obstante, existen dos aspectos que deben puntualizarse: la necesidad de mejorar sus herramientas y la relevancia del factor humano en dicho proceso.

Si bien es cierto que las nuevas tecnologías deben aportar mejoras a la hora de facilitar dicho trabajo, la labor del terminólogo sigue siendo crucial para mejorar su calidad; en particular, para decidir qué candidatos a término obtienen la condición de unidad terminológica y asignar información relevante sobre éste, además de sus ejemplos de uso. El factor humano sigue siendo primordial en esta labor, aunque consideramos que el especialista debe poseer tanto formación técnica como lingüística y no solo una de ellas.

En lo que se refiere a nuestro trabajo, consideramos que los resultados emanados de este estudio terminológico de la disciplina militar han resultado del todo fructíferos y alentadores. No sólo hemos ampliado nuestro conocimiento, sino que hemos descubierto unidades terminológicas desconocidas incluso para los expertos que nos han auxiliado en

esta labor. Hemos de reiterar que la aparición de unidades terminológicas en nuestro corpus que no habían sido registradas en los glosarios consultados destaca la vertiente investigativa de nuestra labor. En este aspecto, el descubrimiento de éstas (tanto las procedentes estrictamente del campo bélico como del histórico) enriquece nuestro conocimiento, al igual que la identificación de la variación conceptual que sufren ciertos términos, la profusión de sinónimos (*machine carbine, machine pistol, submachine gun, Tommy gun, zipper gun,...*), y la aparición de nuevas unidades conceptuales, representadas por nuevas unidades terminológicas o por otras ya existentes que sufren una modificación, al representar una unidad conceptual del todo diferente.

La aportación de definiciones a las nuevas representaciones conceptuales da fe de la originalidad de nuestro trabajo. La definición, el nexo de unión entre unidad conceptual y unidad terminológica (Sager 52), ha sido siempre elaborada mediante el estudio analítico del término. Ya hemos puntualizado que, en innumerables ocasiones, hemos procedido a emplear parte de una definición ya existente de un concepto perteneciente a la misma estructura conceptual de la unidad terminológica objeto de análisis, o que posee unas propiedades, relaciones, cualidades o características similares a las que deseamos ejemplificar. Con ello formulamos la explicación siguiendo las pautas que emplearon los creadores de los glosarios usados como referencia en este trabajo.

De igual modo, puntualizar que uno de los resultados de nuestra investigación ha sido el de cuestionar la metodología por la que se habían compilado la mayor parte de los glosarios consultados y editados durante los años cuarenta del pasado siglo. Si por un lado, se define un tipo de arma, como es la ametralladora (*machine gun*), incluyendo la entrada de dos de sus versiones, la ligera y la pesada (*light machine gun, heavy machine gun*), en esta relación genérica (que une los conceptos pertenecientes a una misma categoría, cuyo término genérico o matriz sería *gun*) podríamos ampliar la estructura arbórea subordinando más conceptos. Por tanto, sería factible añadir otro estrato relacionado con el objetivo a batir, si es terrestre o aéreo, e incluir *antiaircraft machine gun* (de la misma manera que aparece *antiaircraft gun*). No obstante, este grado de especificidad hubiera quedado incompleto, al no existir el término *antiground machine gun, air-to-ground machine gun* u otro similar que situar al otro extremo de la nueva rama conceptual que hubiéramos deseado crear en esta estructura arbórea. De la misma forma, y en referencia a su tamaño, peso y calibre, no se ha incluido a la ametralladora de tipo

medio (*medium machine gun*), a pesar de que las frases en contexto demuestran la existencia y uso de dicha unidad terminológica.

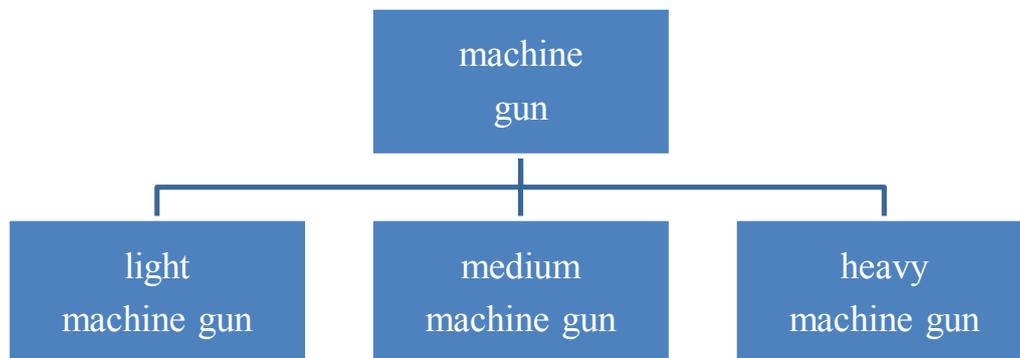


Figura 53. Relación genérica de los tipos de ametralladora según su peso.

Por otro lado, en la relación que se origina con el término *grenade* como concepto genérico, estimamos que es factible la inclusión de más estratos subordinados dentro de su representación jerárquica. La primera relación subordinada podría ser *hand grenade*, *rifle grenade* y *frangible grenade*, aunque Merino (336) también incluye *mortar grenade*. Así mismo, Merino añade *drill grenade* y *practice grenade*, con las que se podrían crear entradas adicionales en esta primera estructura arbórea. Creando un nuevo estrato de especificidad, es posible dividir las granadas de mano por su forma, incluyendo *stick grenade* y *egg grenade*. De igual forma, sería admisible distinguir entre tipos de carga, y subordinar una nueva rama con *fragmentation hand grenade*, *heat hand grenade* y *smoke hand grenade*. Sin embargo, tanto el primer como el segundo término aparecen con el concepto genérico *grenade* (*heat grenade* solo aparece en la obra de Merino). Entendemos que los tipos de carga explosiva, al estar incluidos en el concepto jerárquicamente superior (granada), deben considerarse como asimilados en los términos más estrechos, como a granada de mano.

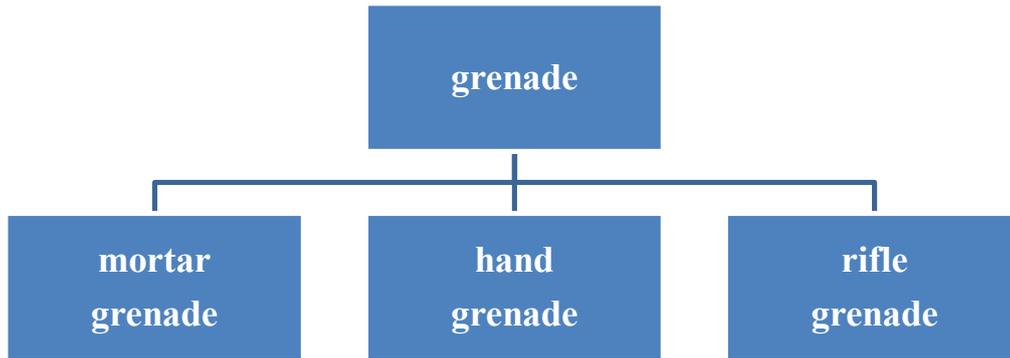


Figura 54. Relación genérica de los tipos de granada según su medio de lanzamiento.

Teniendo en cuenta todas estas consideraciones, realizamos una estructuración conceptual correspondiente a granada. Atendiendo al tipo de carga explosiva y a su forma, llevamos a cabo una estructuración, previamente aprobada por un experto en la disciplina, que difiere de las que han seguido otros a la hora de compilar algunos glosarios consultados y que a continuación detallamos:

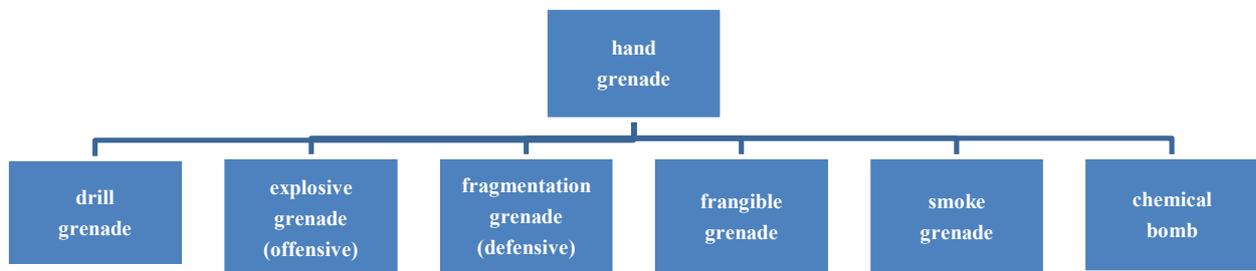


Figura 55. Estructuración conceptual de las granadas de mano atendiendo al tipo de carga explosiva.

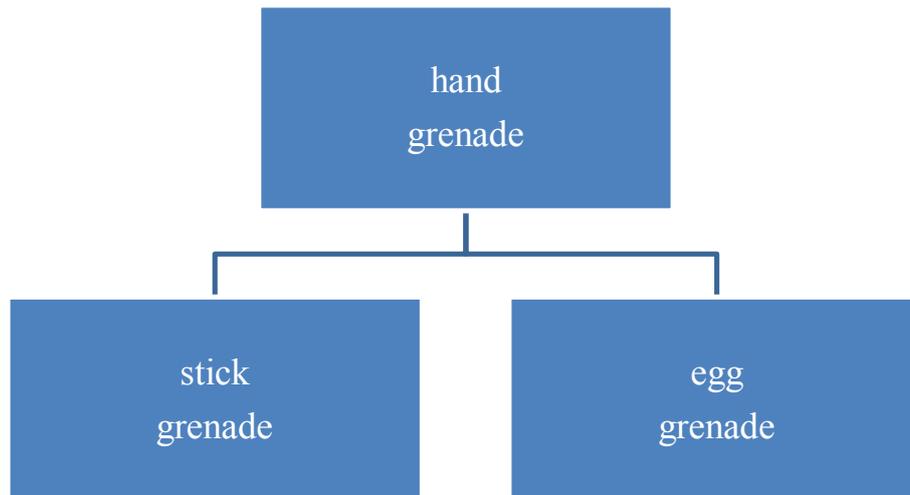


Figura 56. Subdivisión de los tipos de granada de mano atendiendo a su forma.

Ello no quiere decir que nuestra estructuración del conocimiento sea más acertada que la de otros autores, sino que pone de manifiesto la diversidad de enfoques por los que un terminólogo puede llevar a cabo su trabajo.

Todo lo mencionado anteriormente no deja de ser una reflexión sobre la dificultad de compilar un glosario y el tipo de relación existente entre términos: genérica, partitivas, polivalentes y complejas (Sager 147). En lo que concierne a la agrupación en áreas conceptuales dentro de un área de especialidad, procedimos a englobar los términos en subdominios, siguiendo parcialmente el enfoque de *abajo a arriba*. Ello nos sirvió para estructurar el conocimiento de tal manera que nos permitió una fácil clasificación de las unidades terminológicas agrupadas en torno a ciertas subáreas conceptuales. Sobre dichas subáreas procedimos a desarrollar sus correspondientes estructuras arbóreas y los términos situados en un sustrato inferior recibieron parte de la unidad conceptual de la unidad terminológica ubicada sobre ellos.

A pesar de analizar la terminología durante un periodo temporal determinado, juzgamos que esta labor posee, de igual modo, un carácter diacrónico. Como hemos mencionado con anterioridad, la aparición de sinónimos representa un claro ejemplo de la falta de normalización de la disciplina y de su rápida evolución durante esa franja de tiempo. Los textos empleados en este trabajo demuestran una cierta disparidad entre la

definición y el uso de un buen número de unidades terminológicas. Ya hemos constatado que muchos de ellos aumentan su campo de especificidad, mientras que otros lo delimitan, incrementando su intensidad. Por otro lado, los préstamos procedentes de otras lenguas no solo corroboran la aparición de nuevas unidades conceptuales, sino la necesidad de su adecuación a las reglas gramaticales de la lengua inglesa. En definitiva, existen multitud de datos susceptibles de análisis y sería necesario un estudio más pormenorizado de los resultados obtenidos.

Por otro lado, ya hemos detallado que, con respecto al estatus de nuestro corpus, lo hemos definido como un corpus cerrado una vez finalizada nuestra labor. No obstante, nuestro interés es continuar con esta línea de investigación y considerarlo un corpus abierto. Hemos mencionado con anterioridad que tuvimos la fortuna de encontrar todos los números de la revista *Intelligence Bulletin* en formato original. Tanto dicho hallazgo como la práctica obtenida durante esta tesis en lo que respecta al cambio de formato de los textos como el empleo de las herramientas de extracción terminológica y de lingüística de corpus, nos ha llevado a tomar la decisión de realizar un análisis completo del vocabulario terminológico militar empleado durante la Segunda Guerra Mundial.

Estimamos que la adición de todos estos textos hará que nuestro corpus llegue a albergar no menos de tres millones de palabras y que, con ello, estaremos en condiciones de realizar un estudio pormenorizado del uso de todas las unidades terminológicas incluidas en los glosarios editados en 1943 y 1944, al igual de aquéllas que fueron incorporadas a dicho campo del saber con posterioridad. En particular, nuestro interés se centrará en la evolución y el dinamismo que poseen las unidades terminológicas que, con el propósito de incrementar su carácter unívoco, identifican diferentes unidades conceptuales a través de este breve periodo de tiempo.

Finalmente, y a modo de conclusión, podemos añadir que todo trabajo terminológico se encuentra restringido a las necesidades, a los objetivos y a los medios a disposición del investigador y toda alteración de los parámetros establecidos inicialmente supondría una pérdida en lo que se refiere al resultado final del trabajo diseñado originalmente. Ello no quiere decir que sea ilícito alterar los objetivos de una investigación debido a ciertos condicionamientos que aparecen durante su proceso; el objetivo final es siempre lograr el fin estipulado. En este caso, el *fin* puede justificar los *medios*.

10. BIBLIOGRAFÍA

- Aarts, John. "Comments on S. Greenbaum Contribution." *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*. Ed. J. Svartvik. Berlin: de Gruyter, 1992. 180-83. Print.
- . "Intuition-based and Observation-based Grammars." *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. Ed. K. Aijmer, and B. Altenberg. London: Longman, 1991. 44-62. Print.
- Abercrombie, D. "Pseudo-procedures in Linguistics." *Studies in Phonetics and Linguistics*. Ed. D. Abercrombie. Oxford: Oxford UP, 1965. 114-19. Print.
- Adamo, Giovanni. "La terminología en la teoría y la práctica de la traducción." *Hieronymus* 9-10 (2000). Centro Virtual Cervantes, 2000. 83-96. Print.
- Aguilar, Lourdes. *Lexicología y terminología aplicadas a la traducción: curso práctico de introducción*. Barcelona: U Autónoma de Barcelona P, 2001. Print.
- Alexander, L.G. *Essay and Letter Writing*. Harlow: Longman, 1982. Print.
- Alsop, Sian, and Hilary Nesi. "Issues in the Development of the British Academic Written English (BAWE) Corpus." *Corpora* 4/1 (2009): 71-83. Print.
- "American Speech." *Duke University Press*, n.d. Web. 15 Nov. 2014.
- Anderson, Wendy, and John Corbett. *Exploring English with Online Corpora: An Introduction*. New York: Palgrave Macmillan, 2009. Print.
- Andrew, Thomas. *The German Army 1939-45 (5). Western Front 1943-45*. Oxford: Osprey, 2000. Print.
- Anthony, Laurence. "A Critical Look at Software Tools in Corpus Linguistics." *Linguistic Research* 30/2 (2013): 141-61. Print.
- . "Welcome." *Laurence Anthony's Homepage*, n.d. Web. 19 Dec. 2013.
- Antia, Bassey. "Wüster and (Applied) Linguistics Research." *LSP & Professional Communication* 2/1 (2002): 102-110. Print.
- Arnold, Douglas et al. *Machine Translation: an Introductory Guide*. London: NCC Blackwell, 1994. Print.
- Atkins, Sue, Jeremy Clear, and Nicholas Ostler. "Corpus Design Criteria." *Literary and Linguistic Computing* 7/1 (1992): 1-16. Print.

- Baker, Paul. "Diachronic Variation." *Sociolinguistics and Corpus Linguistics*. Edimburgo: Edinburgh UP, 2010. Print.
- . "Flamboyant, Predatory, Self-Confessed Homosexual: Discourse Prosodies in the British Tabloid Press." *Public Discourse of Gay Men*. New York: Routledge, 2005. Print.
- Baker, Paul, Andrew Hardie, and Tony McEnery. *A Glossary of Corpus Linguistics*. Edimburgo: Edinburgh UP, 2006. Print.
- Benavent, Paloma y Sara Parrilla. "Análisis de la extracción automática de términos con el programa informático Extraterm." *Jornades de Foment de la Investigació, Universitat Jaume I* (2008): 1-10. Print.
- Berry, R. "The Seven Sins of Pedagogic Grammar." *Language Analysis, Description and Pedagogy*. Ed. R. Berry, B. Asker, K. Hyland y M. Lam. Hong Kong: The Hong Kong University of Science and Technology, 1999. 29-40. Print.
- Biber, Douglas. "Representativeness in Corpus Design." *Current Issues in Computational Linguistics: in Honour of Don Walker*. Ed. A. Zampolli, N. Calzolari, and M. Palmer. *Linguistica Computazionale IX.X*. Pisa: Giardini, 1994. n. pag. Print.
- Biber, D., S. Conrad, and R. Reppen. *Corpus Linguistics*. Cambridge: Cambridge UP, 1998.
- Biber, D., and E. Finegan. "On the Exploitation of Computerized Corpora in Variation Studies." *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. Ed. K. Aijmer and B. Altenberg. London: Longman, 1991. 204-20. Print.
- "Biblioteca Virtual del Ministerio de Defensa." *Biblioteca Virtual de Defensa*. N.p., n.d. Web. 30 Nov. 2013.
- Blasucci, N. *Macchiavelli, Opere Letterarie*. Milano: Nantia, 1964. Print.
- Boas, F. *Race, Language and Culture*. New York: Macmillan, 1940. Print.
- Bod, R. *Beyond Grammar. An Experience-based Theory of Language*. Stanford: CSLI, 1998. Print.
- Bowker, Lynne. *Computer-Aided Translation technology: A Practical Introduction*. Ottawa: U of Ottawa P, 2002. Print.
- Bowker, Lynne, and Jennifer Pearson. *Working with Specialized Language. A Practical Guide to Using Corpora*. London: Routledge, 2002. Print.
- Bréal, M. *Essai de Sémantique*. Paris: Hachette. 1897. Print.
- Brenes, Patricia. "Create your First Corpus and Analyze it with AntConc (and Related Links to Explore!) - In My Own Terms." *In My Own Terms*. Organic Themes –

WordPress Hosting, 2015. Web. 16 June 2015.

Brezina, Vaclav, and Dana Gablasova. "Is there a Core General Vocabulary? Introducing the New General Service List." *Applied Linguistics*, (2013): 1-23. Print.

Brook, O'Donnell, et al. "Exploring text-initial words, clusters and concgrams in a newspaper corpus." *Corpus Linguistics and Linguistic Theory*, 8-1 (2012): 73-101. Print.

Buchstaller, Isabelle. "Quotations across the generations: A multivariate analysis of speech and thought introducers across 5 decades of Tyneside speech." *Corpus Linguistics and Linguistic Theory* 7-1 (2011): 59-92. Print.

Budin, G. "A critical evaluation of the state-of-the-art of terminology theory." *ITF Journal* 12.1-2 (2001): 7-23. Print.

Cabré, M. Teresa. *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Gerona: Documenta Universitaria, 1999. Print.

---. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Empúries, 1993. Print.

---. "Terminologie et Linguistique: la théorie des portes." *Terminologies nouvelles. Terminologie et diversité culturelle* 21 (2000): 10-25. Print.

---. "Theories of Terminology: Their Description, Prescription and Explanation." *Terminology* 9/2 (2003): 169-199. Print.

Cabré, M. Teresa, y Rosa Estopà. "Formar en terminología: una experiencia docente (I parte)." *Institut Universitari de Lingüística Aplicada* (2000): 1-23. Print.

Cabré, M. Teresa, Rosa Estopà, y Jordi Vivaldi. "Automatic term detection. A review of current systems." *Recent Advances in Computational Technology*. Ed. Didier Burigault, Christian Jacquemin y Marie-Claude L'Homme. Amsterdam: Benjamins, 2001. 71-105. Print.

Cabré, M. Teresa, y J. Feliu, eds. *Terminología y cognición: II Simposio Internacional de Verano de Terminología (13-16 de julio 1999)*. Barcelona: Institut Universitari de Lingüística Aplicada, 2001. Print.

Chovanec, Jan. "Neznalost neomlouvá aneb k. tzv. demystifikaci právního jazyka." *ToP (tlumocení-preklad)* 16.3 (2005): 9-11. Print.

Clear, J. "From Firth principles: collocation tools for the study of collocation." *Text and Technology: In Honour of John Sinclair*. Ed. M. Baker, G. Francis, and E. Tognini-Bonelli. Amsterdam: Benjamins. 1993. 271-92. Print.

- . "Trawling the Language. Monitor Corpora." *ZuriLEX 1986 Proceedings. Papers read at the Euralex International Congress, 9-14 September 1986*. Ed. Mary Snell-Hornby. Amsterdam: Benjamins, 1987. Print.
- Chomsky, Noam. *The Independence of Grammar*. na, 1962. n. pag. Print.
- . *Language and Mind*. New York: Harcourt Brace, 1968. Print.
- . *Modular Approaches to the Study of the Mind*, San Diego: California State UP, 1984. Print.
- Condamines, Anne. "Corpus Analysis and Conceptual Relation Patterns." *Terminology* 8/1 (2002): 141-62. Print.
- Condamines, Anne, and Anne Rebeyrolle. "CTKB: A Corpus-Based Approach to a Terminological Knowledge Base." *First Workshop on Computational Terminology*. Ed. Didier Bourigault, and Chistian Jacquemin. Montreal: COMPUTERM, 1998. 29-35. Print.
- "Corpus4u." XenForum Ltd, n.d. Web. 15 Apr. 2014.
- Cova Morillo-Velarde, M^a Elena. "Lost in XLIFF Translation. XLIFF Effect on Translation Process." Diss. U de Sevilla, 2010. Print.
- Davies, Mark. "Corpus.byu.edu." *CORPORA: 1.9 Billion*. Brigham Young University, 2013. Web. 30 April 2013.
- . "Expanding Horizons in Historical Linguistics with the 400-Million Word Corpus of Historical American English." *Corpora* 7/2 (2012): 121-57. Print.
- "Dictionary, Encyclopedia and Thesaurus." *The Free Dictionary*. Farlex Inc., n.d. Web. 22 Feb. 2014.
- Economic and Social Research Council. "ESRC Centre for Corpus Approaches to Social Science (CASS)." *Corpus Approaches to Social Science*. The University Centre for Computer Corpus Research on Language (UCREL), and the Academy of Social Sciences, n.d. Web. 10 Dec. 2013.
- "Ejército de Tierra." Departamento de Comunicación del Ejército de Tierra, n.d. Web. 26 Feb. 2014.
- Elkin, Peter L. *Terminology and Terminological Systems*. London: Springer, 2012. Print.
- "Enciclopedia Universal." *Los Diccionarios y las Enciclopedias sobre el Académico*. Academic, n.d. Web. 21 Sept. 2013.
- "English Dictionary." *Collins Dictionaries*. HarperCollins Publishers, n.d. Web. 31 Jan. 2014.
- Esselink, B. *A Practical Guide to Localization*. Amsterdam: Benjamins, 2000. Print.

- Fernández Sierra, Luisa. "Trados como herramienta de traducción asistida por ordenador. Un recorrido por su historia y su evolución." *Panace@* X/29 (2009): 29-37. Print.
- Firth, JR. *Papers in Linguistics 1934-1951*. London: Oxford UP, 1957. Print.
- . "A synopsis of linguistic theory." *Selected Papers of J.R. Firth 1952-59*. Ed. F. R. Palmer. London: Longman, 1968. 168-205. Print.
- Francis, W. Nelson, and Henry Kucera. *Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English for use with Digital Computers*. Providence: Brown UP, 1964. Print.
- Fries, C. *American English Grammar*. New York: Appleton Century, 1940. Print.
- Gabrielatos, Costas. "Corpora and Language Teaching: Just a Fling or Wedding Bells?" *TESL-EJ* 8/4 (2005): 1-39. Print.
- García-Pelayo y Gros, Ramón. *Larousse. Gran diccionario moderno, español-inglés*. París: Larousse, 1983. Print.
- Garside, Roger, Geoffrey Leech, and Tony McEnery. *Corpus Annotation*. Harlow: Longman, 1997. Print.
- Garside, Roger, Geoffrey Leech, and Geoffrey Sampson. *The Computational Analysis of English: a Corpus-Based Approach*. Harlow: Longman, 1987. Print.
- Gaudin, F. *Pour une socioterminologie: Des problèmes pratiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen, 1993. Print.
- Ghadessy, Mohsen, and Yanjie Gao. "Small Corpora and Translation. Comparing Thematic Organization in two Languages." *Small Corpus Studies and ELT: Theory and Practice*. Ed. Mohsen Ghadessy, Alex Henry, and Robert L. Roseberry. Philadelphia: Benjamins, 2001: 335-355. Print.
- "Guide to Translation Memory (TM)." *Guide to Translation Memory*. Globalization Partners International, n.d. Web. 13 March 2014.
- Government of Canada. "TERMIUM Plus." *Government of Canada, Public Works and Government Services Canada, Translation Bureau*, n.d. Web. 24 June. 2013.
- . "WeBiText." *Canada.ca Web Services*. Terminotix, n.d. Web. 14 May. 2013.
- Gries, S. Th. "What is corpus linguistics?" *Language and Linguistics Compass* 3 (2009): 1-17. Print.
- Halliday, Michael A.K. "Language as System and Language as Instance: the Corpus as a Theoretical Construct." *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*. Berlin: Mouton de Gruyter, 1992. 61-77. Print.

- Halliday, Michael A.K., and Z. James. "Quantitative Studies and Probabilities in Grammar." *Data, Description, Discourse. Papers on the English Language in Honour of John Sinclair*. Ed. M. Hoey. Indiana: Indiana U, 1993. 1-25. Print.
- . "A Quantitative Study of Polarity and Primary Tense in the English Finite Clause." *Techniques of Description: Spoken and Written Discourse*. Ed. John. M. Sinclair, M. Hoey, and G. Fox. London: Routledge, 1993. 32-66. Print.
- Hanks, Patrick. "The Impact of Corpora on Dictionaries." *Contemporary Corpus Linguistics*. Ed. Paul Baker. London: Bloomsburg, 2009. 214-236. Print.
- Hardie, Andrew. "CQPweb Main Page." UCREL, 2013. Web. 24 Feb. 2015.
- "Hemeroteca de ABC - Abc.es." *Diario ABC, S.L.*, 2008. Web. 9 June 2013.
- Hockett, Charles F. "Two Models of Grammatical Description." *Word* 10 (1954): 210–34. Print.
- . *The State of the Art*. The Haag: Mouton, 1967. Print.
- Hunston, S., and G. Francis. *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins, 2000. Print.
- "ICAME Journal." *De Gruyter Online Academic Publishing*. Walter De Gruyter GmbH, n.d. 2014. Web. 21 Nov. 2014.
- "Intelligence Bulletin." *Internet Archive: Digital Library of Free Books, Movies, Music & Wayback Machine*. N.p., n.d. Web. 2 May 2015.
- "Intelligence Bulletin Series." *LoneSentry. Photographs, Documents, and Research on World War II*. N.p., 2003. Web. 5 Nov. 2014.
- "International Information Centre for Terminology." INFOTERM, n.d. Web. 27 Nov. 2014.
- Johansson, S. "Mens Sana in Corpore Sano: On the Role of Corpora in Linguistic Research." *European Messenger* IV.2 (1995): 19-25. Print.
- Kageura, Kyo. *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Amsterdam: Benjamins, 2002. Print
- . "On the Study of Dynamic of Terminology: A Proposal of a Theoretical Framework." *National Center for Science Information Systems Publications* (1999): 1-10. Print.
- Kageura, Kyo, and Takeshi Abekawa. "Modelling and Exploring the Network Structure of Terminology Using the Potts Spin Glass Model." *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics* (2007): 236-245. Print.

- Kenny, Dorothy. "Creatures of Habit? What Translators Usually Do with Words." *Meta* XLIII/2 (1998): 1-9. Print.
- Kilgariff, Adam. "Sketch Engine | Home Page." *Sketch Engine*. N.p., 2013. Web. 15 March 2014.
- Kokourec, R. *La langue française de la technique et de la science*. Wiesbaden: Brandstetter, 1982. Print.
- Kytö, Merja, Anna-Britta Stenström, and Ilka Mindt eds. "ICAME Journal." N.p., n.d. Web. 20 Mar. 2014.
- "Larousse.fr : Encyclopédie et Dictionnaires Gratuits en Ligne." Editions Larousse, n.d. Web. 17 Feb. 2014.
- Laurén, C, J. Myking, and H. Picht. *Terminologie unter der Lupe*. Vienna: TermNet, 1998. Print.
- Laurén, C., and H. Picht, eds. "Vergleich der Terminologischen Schulen." *Ausgewählte Texte zur Terminologie*. Vienna: TermNet, 1993. 493-539. Print.
- Laviosa, Sara. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose." *Meta* XLIII/4 (1998): 557-570. Print.
- Leech, Geoffrey. "Corpora and theories of linguistic performance." *Directions in Corpus Linguistics: Proceedings of the Nobel symposium 82*. Ed. J. Svartvik. Berlin: Mouton de Gruyter, 1992. 105-22. Print.
- . "An Introduction to Corpus Linguistics." UCREL Summer School in Corpus Linguistics. Lancaster U, Lancaster. 15-18 July 2013. Lecture.
- . "The State of the Art in Corpus Linguistics." *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 1991. 8-29. Print.
- . "Teaching and Language Corpora: a Convergence." *Teaching and Language Corpora*. Ed. A. Wichmann, S. Fligelstone, A. McEnery, and G. Knowles. London: Longman, 1997. Print.
- "Linguee | Dictionary for German, French, Spanish, and More." *Linguee.com*, n.d. Web. 1 May 2013.
- McEnery, Tony. "An Introduction to Corpus Linguistics." UCREL Summer School in Corpus Linguistics. Lancaster U, Lancaster. 15-18 July 2013. Lecture.
- . "'So you Recorded Swearing'. Bad Language in Present-Day English." *Swearing in English: Bad Language, Purity and Power, 1586 to the Present*. New York: Routledge, 2005. 29-57. Print.

- McEnery, Tony, and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge UP, 2012. Print.
- McEnery, Tony, and Andrew Wilson. "Early Corpus Linguistics and the Chomskyan Revolution." *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh UP, 1996. 1-26. Print.
- McNab, Chris. *The Great Book of Guns*. London: Salamander, 2004. Print.
- Meyer, Ingrid, and Kristen Mackintosh. "When terms move into our everyday lives: an overview of de-terminologization." Ed. Kyo Kageura and Marie-Claude L'Homme. *Terminology* 6.1. (2000): 111-38. Print.
- Merino, José. *Diccionario militar técnico español-inglés, inglés-español*. Madrid: Ministerio de Defensa, 1953. Print.
- "Military Definitions." *GlobalSecurity.org*. N.p., n.d. Web. 25 June 2015.
- "Military History." *LinkedIn*. LinkedIn, 5 May 2003. Web. 9 Nov. 2014.
- "Military Lexicon." New Systems Training Office, n.d. Web. 22 Nov. 2013.
- "Military-Ranks.org." *United States Army Ranks in Order*. N.p., n.d. Web. 25 Feb. 2015.
- "Military Terms and Definitions." Rubicon Planning, n.d. Web. 10 Oct. 2014.
- Miller, Christopher E. *A Glossary of Terms and Concepts in Peace and Conflict Studies*. San Jose: University for Peace, 2005.
- Mindt, D. "Syntactic Evidence for Semantic Distinctions in English." *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London: Longman, 1991. 183-96. Print.
- Modern Language Association. *MLA Handbook for writers of Research Papers*. 7th ed. New York: MLA, 2013. Print.
- Mollin, Sandra. "The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts." *Corpora* 2/2 (2007): 187-210. Print.
- Munday, Jeremy. "A Computer-Assisted Approach to the Analysis of Translation Shifts." *Meta* XLIII/4 (1998): 1-16. Print.
- Muñoz, Pablo. "Algo más que Traducir | Blog sobre Traducción y Localización." *Algo más que Traducir*. Minimum Theme en Genesis Framework, n.d. Web. 15 Feb. 2014.
- Myking, J. "Against Precriptivism? The Sociocritical Challenge to Terminology." *ITF Journal* 12.1-2 (2001): 49-64. Print.
- NATO. "Homepage." NATO, n.d. Web. 31 Jan. 2014.

- O’Keeffe, Anne, and Steve Walsh. “Applying Corpus Linguistics and Conversation Analysis in the Investigation of Small Group Teaching in Higher Education.” *Corpus Linguistics and Linguistic Theory* 8/1 (2012): 159-181. Print.
- Olohan, Maeve. “Spelling out the Optionals in Translation: a Comparative Study.” *Centre for Translation and Intercultural Studies, UMIST* (2001): 1-10. Print.
- “Oxford Dictionaries - Dictionary, Thesaurus, & Grammar.” Oxford University Press, n.d. Web. 31 Jan. 2014.
- Palmer, F.R. *Selected Papers of J.R. Firth 1952-59*. London: 1968, Longman. Print.
- Parker, Sybil P. *McGraw-Hill Dictionary of Scientific & Technical Terms*. New York: McGraw-Hill, 2003. Print.
- Picht, Heribert, and Klaus-Dirk Schmitz, eds. *Terminology und Wissensordnung. Ausgewählte Schriften auf dem Gesamtwerk von Eugen Wüster*. Vienna: TermNet, 2001. Print.
- Quah, C.K. *Translation and Technology*. Houndmills: Palgrave Macmillan, 2006. Print.
- “Ranks.” *British Army Website*. Crown, n.d. Web. 28 July 2014.
- “Real Academia Española.” *Real Academia Española*, n.d. Web. 24 Sept. 2014.
- Rey, A. “La terminologie, entre l’expérience du réel et la maîtrise des signes.” Séminaire de Terminologie Théorique. Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada, Barcelona. 15 July 1999. Lecture.
- Römer, U. *Progressives, Patterns, Pedagogy: A Corpus-Driven Approach to English Progressive Forms, Functions, Contents, and Didactics*. Amsterdam: Benjamins, 2005. Print.
- Rosset, Edward R. *Guide to Prepositions English to Spanish*. Irún: Stanley, 2003. Print.
- Sageder, Dagmar. “Terminology Today: A Science, an Art or a Practice? Some Aspects on Terminology and its Development.” *Brno Studies in English* 36/1 (2010): 123-134. Print.
- Sager, Juan C. *A Practical Course in Terminology Processing*. Amsterdam: Benjamins, 1990. Print.
- Santamaría Pérez, Isabel. *La terminología como disciplina: definición, funciones y aplicaciones*. Alicante: EPA, 2006. Print.
- Schmied, Josef, and Hildegard Schäffler. “New Approaches to Contrastive Linguistics and Translation Studies: Bilingual Corpora.” *Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on*

- Computerized Corpora*. Ed. Carol E. Percy, Charles F. Meyer, and Ian Lancashire. Amsterdam: Rodopi, 1996. 1-28. Print.
- Schmitz, Klaus-Dirk. "Applied Principles of Terminology Work." Terminology Summer School. Cologne U of Applied Languages, Cologne. 6-10 July 2009. Lecture.
- "SDL: Language Translation & Content Management Company." RSS. SDL, n.d. Web. 20 Feb. 2013.
- Semino, Elena, and M. Short. "A Corpus-Based Approach to the Study of the Discourse Presentation in Written Narratives." *Corpus Stylistics: Speech, Writing and Thought Presentation*. New York: Routledge, 2004. 1-18. Print.
- "Servlet Platform Proc." Universitat Oberta de Catalunya, n.d. Web. 24 Dec. 2013.
- Sinclair, John M. *Collins Cobuild. English Language Dictionary*. London: Collins, 1987. Print.
- . *Corpus Concordance Collocation*. Oxford: Oxford UP, 1991. Print.
- . "Corpus Linguistics at the Millennium." *New Horizons in CALL*. Ed. Kohn et al. Szombathely: Bersenyi Dániel College, 1997. Print.
- . "The Search for Units of Meaning." *TEXTUSIX 1* (1996): 75-106. Print.
- Sinclair, John M., S. Jones, R. Daley, and R. Krishnamurthy. *English Collocational Studies: The OSTI Report*. London: Continuum, 2004. Print.
- Smith, Neil, y Deirdre Wilson. *La lingüística moderna. Los resultados de la revolución de Chomsky*. Barcelona: Anagrama, 1983. Print.
- Steiner, George. *Después de Babel. Aspectos del lenguaje y la traducción*. Madrid: F.C.E., 1981. Print.
- Stubbs, M. "British Traditions in Text Analysis." *Text and Technology: in honour of John Sinclair*. Ed. M. Baker, G. Francis, and E. Tognini-Bonelli. Amsterdam: Benjamins, 1993. 1-33. Print.
- . *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Cambridge: Blackwell, 1996. Print.
- Swan, Michael. *Practical English Usage*. Oxford: Oxford UP, 1986. Print.
- Temmerman, Rita. *Towards new Ways of Terminology Description. The Sociocognitive-Approach*. Amsterdam: Benjamins, 2000. Print.
- Teubert, W. "My Version of Corpus Linguistics." *International Journal of Corpus Linguistics*. 10.1 (2005): 1-13. Print.

- Toft, Bertha. "Review of "Towards New Ways of Terminology Description; the sociocognitive approach" by Rita Temmerman 2000." *Terminology* 7.1 (2001): 125-28. Print.
- Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Amsterdam: Benjamins, 2001. Print.
- "Translation, Localization, Globalization, Internationalization Tools." *Translator's Translation, Localization, Globalization, Internationalization Language Tools (Localisation, Globalisation)*. N.p., n.d. Web. 24 Feb. 2016.
- "UCREL Home Page, Lancaster UK." *UCREL Home Page, Lancaster UK*. Lancaster University, 1993. Web. 17 Jan. 2014.
- United States, Department of the Army. *Dictionary of United States Army*. Washington: Military Publications, 1983. Print.
- , ---. *The Modular Force*. Washington: Army Knowledge Online, 2008. Print.
- , Department of Defense. *Department of Defense Dictionary of Military Terms and Associated Terms*. Washington: Joint Publication 1-02, 2010. Print.
- , ---. *Dictionary of Military and Associated Terms*. Washington: Department of Defense, 2005. Print.
- , ---. "JEL - Joint Electronic Library." Joint Chiefs of Staff, n.d. Web. 19 Jan. 2014.
- , Joint Chiefs of Staff. *Dictionary of United States Military Terms*. Washington: GPO, 1950. Print.
- , War Department. *British Military Terminology*. Washington: GPO, 1943. Print.
- , ---. *Dictionary of United States Army Terms*. Washington: GPO, 1944. Print.
- , ---. *German Military Abbreviations*. Washington: GPO, 1943. Print.
- Vargas Sierra, Chelo. "La tecnología de corpus en el contexto profesional y académico de la traducción y la terminología: panorama actual, recursos y perspectivas." *Tecnología, traducción y cultura*. Ed. M.A. Candel Mora y E. Ortega Arjonilla. Valencia: Tirant Humanidades, 2012. 67-99. Print.
- Ward, Keith. *World War II German Field Weapons & Equipment*. Solihull: Helion, 2014. Print.
- "WeBiText." Yahoo! Search, n.d. Web. 24 Dec. 2013.
- "WordSmith Tools Home Page." Lexical Analysis Software, n.d. Web. 30 Aug. 2013.
- Wordreference. "English to French, Italian, German & Spanish Dictionary – WordReference.com." *Wordreference.com*. Wordreference, n.d. Web. 1 May 2013.

- Wüster, E. "Die Allgemeine Terminologielehre – Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und der Sachwissenschaften." *Linguistics* 119 (1974): 61-106. Print.
- . *Bibliography of Monolingual Scientific and Technical Glossaries*. Paris: Unesco, 1955. Print.
- Yerger, Mark C. *German Cross in Silver. Holders of the SS and Police*. San Jose: Bender, 2002. Print.

11. APÉNDICE – SISTEMA DE ETIQUETADO
GRAMATICAL UCREL CLAWS 7

APPGE	possessive pronoun, pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a, an, every)
BCL	before-clause marker (e.g. in order (that),in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these,those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)

EX	existential there
FO	Formula
FU	unclassified Word
FW	foreign Word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)
JK	Catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number, neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40-50, 1770-1827)
MD	ordinal number (e.g. first, second, next, last)
MF	fraction, neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	Following noun of title (e.g. M.A.)
NNB	Preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun, neutral for number (e.g. dozen, hundred)

NNO2	numeral noun, plural (e.g. hundreds, thousands)
NNT1	temporal noun, singular (e.g. day, week, year)
NNT2	temporal noun, plural (e.g. days, weeks, years)
NNU	unit of measurement, neutral for number (e.g. in, cc)
NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)
NP	proper noun, neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London, Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)
NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	Subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)

PPX1	singular reflexive personal pronoun (e.g. yourself, itself)
PPX2	plural reflexive personal pronoun (e.g. yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g. else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)
RGQ	wh-degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g. alongside, forward)
RP	prep. adverb, particle (e.g. about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh-general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g. better, longer)
RRT	superlative general adverb (e.g. best, longest)
RT	quasi-nominal adverb of time (e.g. now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g. oh, yes, um)
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	Were
VBDZ	Was
VBG	Being
VBI	be, infinitive (To be or not... It will be ..)
VBM	Am
VBN	Been
VBR	Are

VBZ	Is
VD0	do, base form (finite)
VDD	Did
VDG	Doing
VDI	do, infinitive (I may do... To do...)
VDN	Done
VDZ	Does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	Having
VHI	have, infinitive
VHN	had (past participle)
VHZ	Has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g. A,b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)

NOTE: "DITTO TAGS"

Any of the tags listed above may in theory be modified by the addition of a pair of numbers to it: eg. **DD21**, **DD22** This signifies that the tag occurs as part of a sequence of similar tags, representing a sequence of words which for grammatical purposes are treated as a single unit. For example the expression *in terms of* is treated as a single preposition, receiving the tags:

in_II31 terms_II32 of_II33

The first of the two digits indicates the number of words/tags in the sequence, and the second digit the position of each word within that sequence.

Such *ditto tags* are not included in the lexicon, but are assigned automatically by a program called **IDIOMTAG** which looks for a range of multi-word sequences included in the **idiomlist**. The following sample entries from the idiomlist show that syntactic ambiguity is taken into account, and also that, depending on the context, ditto tags may or may not be required for a particular word sequence:

at_RR21 length_RR22
a_DD21/RR21 lot_DD22/RR22
in_CS21/II that_CS22/DD1