UNIVERSIDAD DE GRANADA

Facultad de Traducción e Interpretación

Departamento de Traducción e Interpretación

Programa de Doctorado en Lenguas, Textos y Contextos

LOS COMPUESTOS NOMINALES EN TERMINOLOGÍA: FORMACIÓN, TRADUCCIÓN Y REPRESENTACIÓN

TESIS DOCTORAL

MELANIA CABEZAS GARCÍA

UNIVERSIDAD DE GRANADA

Programa de Doctorado en Lenguas, Textos y Contextos

LOS COMPUESTOS NOMINALES EN TERMINOLOGÍA: FORMACIÓN, TRADUCCIÓN Y REPRESENTACIÓN

TESIS DOCTORAL

MELANIA CABEZAS GARCÍA

Directoras: Pamela Faber Benítez y Pilar León Araúz

Granada, 2019

La doctoranda, **Melania Cabezas García**, y las directoras de la tesis, **Pamela Faber Benítez y Pilar León Araúz**, garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por la doctoranda bajo la dirección de las directoras de la tesis y, hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 16 de septiembre de 2019

Fdo.: Pamela Faber Benítez	Fdo.: Pilar León Araúz	Fdo.: Melania Cabezas García
Directora de la tesis	Directora de la tesis	Doctoranda

A mi madre, por darme alas.

AGRADECIMIENTOS

Quisiera dedicar unas palabras de agradecimiento a las personas que me han acompañado a lo largo de este camino. Aunque breves, espero que reflejen el calor que he recibido durante los últimos cuatro años. En primer lugar, deseo agradecer a mis directoras, Pamela Faber y Pilar León, el tiempo y esfuerzo dedicado a la orientación de este trabajo. Esta tesis no habría sido posible sin ellas, tanto por la inspiración proporcionada incluso antes de empezar este trabajo, como por las excelentes pautas que me han brindado en el ámbito académico y también en el personal. Sin duda, he tenido la suerte de aprender de las mejores maestras y amigas.

También quisiera agradecer el apoyo recibido por parte de mis compañeros del grupo de investigación LexiCon, quienes no han dudado en prestarme toda la ayuda necesaria. Gracias a Arianne Reimerink, Beatriz Sánchez, Silvia Montero, Míriam Buendía, Antonio San Martín, José Manuel Ureña, Juan Carlos Gil, Juan Rojas y Santi Chambó. Gracias por los cafés, las risas y, sobre todo, por acogerme en una familia que anima a crecer.

Mi agradecimiento también va dirigido a los miembros del Departamento de Traducción e Interpretación de la Universidad de Granada que se han interesado por el desarrollo de este trabajo y con los que he disfrutado de momentos muy buenos. Entre ellos, quisiera dar las gracias a los diferentes compañeros con los que he podido compartir agobios y risas en la Sala de Investigación. Las largas horas en aquella sala no habrían sido lo mismo sin vosotros.

Deseo agradecer el cariño y apoyo de mis amigos, tanto a los que conocí de niña y tengo la suerte de conservar hoy, como a los que me han brindado los años de estudio en Sevilla, Granada y Francia. Gracias por haber creído siempre en mí e impulsarme desde el principio.

Esta tesis tampoco habría sido posible sin el amor de mi familia. En especial, me gustaría dar las gracias a mis abuelos, por su cariño y por alimentarme muy bien durante estos años y toda la vida; a Amanda, por su apoyo constante; a Álvaro, por el tiempo que no he podido regalarle; y a mi hermano, Christian, por las risas que son tan nuestras y me han ayudado tanto. Mi agradecimiento jamás alcanzará el merecido por Javi, por ser el mejor compañero de vida que se pueda soñar. Gracias por tachar imposibles conmigo.

Por último, quisiera dar las gracias a mi padre, por la ilusión que brilla en sus ojos cada vez que mira mi camino, la misma que brillaría en los de mi madre, que estoy segura de que estaría orgullosa.

ÍNDICE

Abstract 20

1. Introduction 21
   1.1 Rationale and presentation of the research 21
   1.2 Objectives 22
   1.3 Structure and contents of the thesis 23

2. Bases lingüísticas y terminológicas 25
   2.1 La Lingüística Cognitiva 25
      2.1.1 La categorización 29
      2.1.2 La Semántica de Marcos 32
   2.2 La Terminología Basada en Marcos 35
      2.2.1 Antecedentes 35
         2.2.1.1 La Teoría General de la Terminología 35
         2.2.1.2 La Socioterminología 36
         2.2.1.3 La Teoría Comunicativa de la Terminología 37
         2.2.1.4 La Terminología Sociocognitiva 38
      2.2.2 Fundamentos 39
      2.2.3 Aplicación práctica: EcoLexicon 44

3. Los términos compuestos: el caso de los compuestos nominales 48
   3.1 Los compuestos nominales en la fraseología 49
      3.1.1 La fraseología y las unidades fraseológicas 49
      3.1.2 Las unidades fraseológicas especializadas 58
   3.2 Noción de compuesto nominal y estudios principales 70
   3.3 Tipología y formación de compuestos nominales 84
   3.4 Análisis de los compuestos nominales 96
      3.4.1 Desambiguación estructural: *bracketing* 96
      3.4.2 La semántica de los formantes 101
      3.4.3 Análisis de las relaciones semánticas internas 111

3.4.3.1 Los inventarios de relaciones semánticas 117

3.4.3.2 Las paráfrasis 128

3.4.4 Variación en los compuestos nominales 133

3.4.4.1 Variación denominativa formal 136

3.4.4.2 Variación denominativa cognitiva 141

3.4.5 Neología en los compuestos nominales 144

3.5 Traducción de compuestos nominales 152

3.6 Representación de compuestos nominales en recursos lingüísticos 169

3.6.1 Las unidades fraseológicas en recursos lingüísticos 169

3.6.2 Análisis del tratamiento de compuestos nominales en recursos terminográficos 174

3.6.2.1 *Diccionario Español de Ingeniería* 175

3.6.2.2 *Diccionario de términos médicos* 179

3.6.2.3 *Oxford Dictionary of Biochemistry and Molecular Biology* 184

3.6.2.4 *Diccionario técnico inglés-español español-inglés* 187

3.6.2.5 *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish* 192

3.6.2.6 *DiCoEnviro. Le dictionnaire fondamental de l'environnement* 193

4. Materiales y metodología 199

4.1 Materiales 199

4.1.1 Corpus de energía eólica 199

4.1.2 Sketch Engine 200

4.2 Metodología 202

4.2.1 Compilación del corpus 202

4.2.2 Extracción de compuestos nominales en inglés 205

4.2.3 Análisis de los compuestos nominales en inglés y búsqueda de equivalentes en español 211

5. Resultados y discusión 222

5.1 La formación de compuestos nominales 222

5.1.1 Análisis semántico de la formación de compuestos nominales 222

5.1.1.1 *Bracketing* 222

5.1.1.2 Asignación de categorías semánticas 237

5.1.1.3 Descodificación de la relación semántica interna 247

5.1.1.4 Asignación de roles semánticos 257

5.1.2 Los microcontextos en los compuestos nominales 264

5.1.2.1 Análisis de los microcontextos 265

5.1.2.2 Conclusiones de los microcontextos 281

5.2 La traducción de compuestos nominales 287

5.2.1 Procedimiento de búsqueda de equivalentes 287

5.2.2 Comparación de equivalentes 297

5.3 La representación de compuestos nominales: el módulo fraseológico de compuestos nominales de EcoLexicon 327

5.3.1 Plantilla de introducción de datos 327

5.3.2 Interfaz de usuario 334

5.3.2.1 Vista *CN formation/Formación de CN* 340

5.3.2.2 Vista *Equivalents/Equivalentes* 349

5.3.2.3 Vista *Morphosyntactic combinations/Combinaciones morfosintácticas* 352

5.3.2.4 Vista *Semantic combinations/Combinaciones semánticas* 356

5.3.2.5 Vista *Summary/Resumen* 363

5.3.3 Conclusiones del módulo de compuestos nominales de EcoLexicon 365

6. Conclusions 368

Bibliografía 378

Anexo 1: Extracción inicial de CN en inglés 414

ÍNDICE DE FIGURAS

Figura 1: El EVENTO MEDIOAMBIENTAL 42

Figura 2: Entrada de ANTHROPIC EROSION en EcoLexicon 45

Figura 3: Representación de *generación de energía eólica* en Google Ngram Viewer 146

Figura 4: Entrada 1 de *abolladura* en el *Diccionario Español de Ingeniería* 177

Figura 5: Entrada 2 de *abolladura* en el *Diccionario Español de Ingeniería* 177

Figura 6: Entrada de *falla* en el *Diccionario Español de Ingeniería* 178

Figura 7: Búsqueda avanzada por *tipo de lema > término complejo Y en definición > enfermedad Y en definición > pelo* en el *Diccionario de términos médicos* 183

Figura 8: Entrada de *albinismo oculocutáneo* en el *Diccionario de términos médicos* 184

Figura 9: Entrada de *antagonist* en el *Oxford Dictionary of Biochemistry and Molecular Biology* 187

Figura 10: Entrada de *multiple* en la versión impresa del *Diccionario técnico inglés-español español-inglés* 189

Figura 11: Entrada de *manifold* en la versión online del *Diccionario técnico inglés-español español-inglés* 191

Figura 12: Entrada de *polimerasa* en el *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish* 193

Figura 13: Entrada de *agriculture* en la versión de marcos de *DiCoEnviro* 197

Figura 14: Entrada de *climate* en la herramienta *NeoVisual* de *DiCoEnviro* 197

Figura 15: Interfaz principal de Sketch Engine 201

Figura 16: Resultados de la opción *Concordance* para la consulta [lemma="stall-regulated"][lemma="wind"][lemma="turbine"] 202

Figura 17: Muestra de sustantivos del corpus en inglés ordenados por frecuencia 206

Figura 18: Sistema conceptual simplificado del EVENTO DE LA ENERGÍA EÓLICA 207

Figura 19: Paráfrasis verbales de *power curve* 215

Figura 20: Parte de las relaciones de *turbine* con otros conceptos, extraídas a partir del análisis de sus CN 284

Figura 21: Variantes denominativas de *line-to-line voltage* y *phase voltage* 289

Figura 22: Elementos contextuales de *line-to-line voltage* 290

Figura 23: Consulta de elementos contextuales de *line-to-line voltage* en el corpus en español 291

Figura 24: Concordancias de la búsqueda de equivalentes de *line-to-line voltage* en español 293

Figura 25: Variantes denominativas de *line-to-line voltage* en inglés y español 294

Figura 26: Elementos contextuales de *line-to-neutral voltage* 295

Figura 27: Concordancias de la búsqueda de equivalentes de *line-to-neutral voltage* en español 295

Figura 28: Variantes denominativas de *line-to-neutral voltage* en inglés y español 296

Figura 29: Morfosintaxis de los CN en inglés 298

Figura 30: Morfosintaxis de los CN en español 301

Figura 31: Ejemplos de siglas en inglés empleadas en español 317

Figura 32: Concordancias de *wind power production*, *wind power generation* y *wind generation* cuando aluden al proceso de formación 320

Figura 33: Concordancias de *wind power production*, *wind power generation* y *wind generation* cuando aluden a la energía resultante 321

Figura 34: Concordancias de *generación eólica*, *generación de energía eólica* y *producción de energía eólica* con ambos significados 322

Figura 35: Concordancias de *sistema eólico* y *sistema de vientos* 322

Figura 36: Ejemplos de préstamos en español 323

Figura 37: Parte superior de la interfaz para acceder a las vistas del módulo de CN 335

Figura 38: Interfaz de búsqueda simple de la vista *CN formation/Formación de CN* 340

Figura 39: Interfaz de búsqueda avanzada de la vista *CN formation/Formación de CN* 341

Figura 40: Consultas posibles con los marcadores booleanos ubicados junto a las burbujas 343

Figura 41: Ejemplo de empleo de marcadores booleanos en la búsqueda avanzada de la vista *CN formation/Formación de CN* 344

Figura 42: Fragmento de la información de *generator* en la vista *CN formation/Formación de CN* del módulo de CN, con los conceptos ordenados por dimensiones 346

Figura 43: Ventana de opciones secundarias en la vista *CN formation/Formación de CN* del módulo de CN 347

Figura 44: Resultados de la opción secundaria *Usage examples* de *squirrel cage induction generator* en la vista *CN formation/Formación de CN* del módulo de CN 348

Figura 45: Fragmento de la información de *generator* en la vista *Equivalents/Equivalentes* 351

Figura 46: Interfaz de búsqueda avanzada de la vista *Morphosyntactic combinations/Combinaciones morfosintácticas* 353

Figura 47: Fragmento de la consulta *Adjective + Common noun + turbine* en la vista *Morphosyntactic combinations/Combinaciones morfosintácticas* 354

Figura 48: Resultados de la opción *Compare morphosyntactic patterns/Comparar patrones morfosintácticos* para la consulta *Common noun + common noun + common noun* 356

Figura 49: Interfaz de búsqueda avanzada de la vista *Semantic combinations/ Combinaciones semánticas* 358

Figura 50: Fragmento de la consulta avanzada *Magnitude (Attribute) + Change (Process)* en la vista *Semantic combinations/Combinaciones semánticas* 360

Figura 51: Resultados de la opción *Compare semantic patterns/Comparar patrones semánticos* para la consulta *Descriptive + wind turbine* 362

Figura 52: Interfaz de consulta de la vista *Summary/Resumen* 363

Figura 53: Resultados de la consulta *turbine* en la vista *Summary/Resumen* 364

ÍNDICE DE TABLAS

Tabla 1: Indicativos para el *bracketing* propuestos por Nakov y Hearst (2005: 19-21).....98

Tabla 2: Estructura de *qualia* relativa al concepto NOVEL (Pustejovsky 1991: 427)..... 123

Tabla 3: Ejemplos de dimensiones enfatizadas por variantes (tabla original de León Araúz 2017: 223) 143

Tabla 4: Composición de los corpus en inglés (izquierda) y en español (derecha) 200

Tabla 5: Expresión 1. CN con premodificación y el término en el núcleo 208

Tabla 6: Expresión 2. CN con premodificación y el término en el modificador 208

Tabla 7: Expresión 3. CN con postmodificación y el término en el núcleo 208

Tabla 8: Expresión 4. CN con postmodificación y el término en el modificador 209

Tabla 9: Patrones de conocimiento sinonímicos 210

Tabla 10: Ejemplo del protocolo de *bracketing* en *wind energy system* 212

Tabla 11: Fragmento del inventario de categorías semánticas de Gil Berrozpe et al. (2019) 213

Tabla 12: Ejemplo de la asignación de categorías semánticas en *generator torque control* 214

Tabla 13: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN de dos formantes 214

Tabla 14: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN de más de dos formantes..... 215

Tabla 15: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN con adjetivos denominales 216

Tabla 16: Ejemplo de consulta en CQL para obtener paráfrasis libres en CN de dos formantes 216

Tabla 17: Ejemplo de consulta en CQL para obtener paráfrasis libres en CN de más de dos formantes 217

Tabla 18: Inventario de relaciones semánticas de EcoLexicon 218

Tabla 19: Inventario de roles semánticos de EcoLexicon 219

Tabla 20: Ejemplos de la anotación con roles semánticos..... 219

Tabla 21: Ejemplos de conceptos con sus variantes denominativas en inglés y en español 220

Tabla 22: Consultas CQL para comprobar el indicador A del *bracketing* 224

Tabla 23: Consultas CQL para comprobar el indicador B del *bracketing* 226

Tabla 24: Consultas CQL para comprobar el indicador C del *bracketing* (parte 1)..... 227

Tabla 25: Consultas CQL para comprobar el indicador C del *bracketing* (parte 2) 227

Tabla 26: Ejemplo de consulta CQL para comprobar el indicador C del *bracketing* en CN de cuatro o más formantes 228

Tabla 27: CN generados por las distintas agrupaciones posibles en *wind power generation system* 228

Tabla 28: Consultas CQL para comprobar el indicador D del *bracketing* (antónimos) 230

Tabla 29: Cifras de la desambiguación de CN mediante cada indicador 231

Tabla 30: *Bracketing* de CN de 3 formantes 235

Tabla 31: *Bracketing* de CN de 4 formantes 236

Tabla 32: Combinaciones de categorías más habituales en los CN de la muestra 240

Tabla 33: Categorías de los conceptos designados por los CN de la muestra 243

Tabla 34: Relaciones identificadas mediante paráfrasis verbales, ilustradas con ejemplos 249

Tabla 35: Significados de *power system* y paráfrasis diferentes 251

Tabla 36: Relaciones identificadas mediante paráfrasis libres, ilustradas con ejemplos.. 253

Tabla 37: Relaciones codificadas por cada categoría ontológica 255

Tabla 38: Proposiciones conceptuales que subyacen en los CN de la muestra, anotadas con roles semánticos e ilustradas con ejemplos 259

Tabla 39: CN polisémicos que implicaron un cambio de relación interna y de roles semánticos 263

Tabla 40: *Slots* que abren las subcategorías de ENTITY>CREATION 266

Tabla 41: *Slots* que abren las subcategorías de ENTITY>CREATION y CN con esta categoría nuclear 271

Tabla 42: *Slots* que abre la categoría ENTITY>SYSTEM 271

Tabla 43: *Slots* que abre la categoría ENTITY>PART 274

Tabla 44: *Slots* que abre la categoría ENTITY>INFORMATION 274

Tabla 45: *Slots* que abre la categoría ENTITY>HUMAN 275

Tabla 46: *Slots* que abren las subcategorías de ATTRIBUTE 276

Tabla 47: *Slots* que abren las subcategorías de ATTRIBUTE y CN con esta categoría nuclear 278

Tabla 48: *Slots* que abre la categoría PROCESS>CHANGE 279

Tabla 49: *Slots* que abre la categoría PROCESS>MOVEMENT 280

Tabla 50: Variantes de conceptos en inglés y español, y semejanza de dimensiones 311

Tabla 51: Variantes de conceptos en inglés y español con dimensiones diferentes 313

Tabla 52: Comparación de variantes principales con dimensiones diferentes en inglés y español 314

Tabla 53: Ejemplos de variantes producidas por calco y por naturalización en español .. 324

Tabla 54: Plantilla de introducción de datos 330

Tabla 55: Organización de datos en las vistas del módulo de CN 339

Abstract

This study has focused on English and Spanish complex nominals (CNs) pertaining to wind power terminology. Such multiword terms are very frequent in general and specialized discourse, but they are a specific feature of the latter since knowledge is usually condensed in these structures. Although CNs have been widely studied, there are various aspects that require further research. These include the formation of complex structures in specialized language (e.g. *wind power forecast error*), their modification by parts of speech different from the noun (e.g. *rotating magnetic field*), their presence in or translation into other languages, and their description in linguistic resources.

With a view to analyzing CN behavior in the wind power terminology, our specific objectives were the following: (i) to study CNs and address their characteristics in specialized discourse; (ii) to analyze CNs semantically; (iii) to explore CN formation from a semantic perspective; (iv) to investigate CN translation in English and Spanish; and (v) to design a systematic proposal for CN representation in the terminological knowledge base EcoLexicon. Assuming the premises of Frame-Based Terminology (Faber et al. 2005, 2006; Faber 2012), we compiled an English-Spanish comparable corpus on wind power, which was used for term extraction, semantic analysis, and the identification of terminological correspondences.

Our study revealed that the clusters of concepts in CNs can be accessed and specified by means of semantic categories, roles and relations. These also allowed us to capture regularities in CN formation, which can be further examined by a slot-filling mechanism activated by the CN head. This triggers the so-called *microcontext*. CN translation was also found to be challenging because of the different patterns of term formation and the lack of systematicity regarding the description of CNs in terminographic resources. Along these lines, corpus techniques and guidelines for the specification of equivalences (based on an in-depth contrastive analysis) are essential. As for CN representation, it should encompass the different characteristics of these multiword terms (i.e. their morphosyntactic and semantic formation, their structure, their variants and equivalents, their position in the conceptual system, etc.). This wealth of information, together with a conceptual organization and different access paths to data, can provide a full account of specialized CNs and improve user experience. The practical application of this research is a CN section that has been devised for the phraseological module of EcoLexicon, a terminological knowledge base on the environment.

1. Introduction

1.1 Rationale and presentation of the research

Complex nominals (CNs) are multiword expressions in which a nominal head is complemented by one or several modifiers, e.g. *floating offshore wind turbine*. They are very frequent in English and Spanish, both in general and specialized language. In the latter, they hold a special status as the main type of multiword term and one of the most frequent term formation mechanisms, if not the main, thanks to their multiple possibilities of conceptual combination. Therefore, CNs are relevant in conceptual systems and contribute to specialized knowledge transmission.

In scientific and technical discourse, it is thus usual to find sequences such as the following title of a research article, which totally consists of CNs: *Inertia-Free Stand-Alone Microgrid—Part II: Inertia Control for Stabilizing DC-Link Capacitor Voltage of PMSG Wind Turbine System*. This example highlights the fact that CN analysis is an extremely complex linguistic task. In fact, different aspects of these multiword terms can be problematic, such as their identification and structural disambiguation, the semantic analysis of their constituents and the relation linking them, as well as their translation or their description in terminographic resources.

Recognizing them in texts is often difficult since CNs can be formed by three or more terms or even by general language words, which can considerably complicate the identification of the whole sequence. Additionally, ascertaining structural dependencies can be challenging in CNs composed of three terms or more. Semantic analysis can also be difficult, given the non-specification of the semantic relation between the constituents, which, along with the frequent specialization and omission of elements, can hinder the understanding of their meaning. Their transfer into other languages is not an easy task either because of the different patterns of term formation and the unsystematic treatment of CNs in terminographic resources, which also reflects the difficulty of representing them.

Certain issues (i.e. semantic relations underlying CNs, their bracketing, and structural disambiguation) have been researched in some depth. Nevertheless, many others have not been addressed in sufficient depth and are in need of further exploration. These include the analysis of highly specialized CNs, their formation and composition in English and other languages, their translation, and their description in terminographic resources with a view to enhancing knowledge acquisition.

This thesis focuses on CNs in terminology. The specialized domain analyzed was wind power, which was explored by means of a comparable corpus in English and Spanish compiled for the purposes of the study. Wind power terminology is a relatively recent area that as yet has not been sufficiently explored, especially in Spanish, despite its increasing importance. Some of the characteristics of these terms are the use of general language words, as well as the integration of different disciplines, such as Engineering and Mathematics, as reflected in the great density of formulas and technical terms from those domains.

Our methodology is based on Frame-Based Terminology[1] (Faber et al. 2005, 2006; Faber 2012), a cognitive theory of terminology that contextualizes concepts in frames or knowledge structures, and is based on corpus analysis. In the first stage of our research, we extracted English CNs by means of corpus queries based on regular expressions. The multiword terms were then semantically analyzed, which entailed: (i) their bracketing; (ii) the assignment of semantic categories; (iii) the decoding of the internal semantic relations by means of paraphrases; and (iv) semantic role labeling since these combinations have underlying propositions. Finally, correspondences were established with Spanish terms based on corpus techniques. The objective of this analysis was to gain further insights into the formation, translation, and representation of these multiword terms.

This research thus represents a comprehensive terminological proposal, which deals with a wide range of aspects of CNs, including the most problematic ones, which had previously been eluded or not fully explored. Some of them are CN formation by any number and type of elements (e.g. participles, which reveal the underlying propositions), structural disambiguation, and translation as part of the everyday work of translators and terminologists, as well as the systematic representation of CNs with a view to knowledge organization and acquisition. In addition, this thesis does not limit to the study of English CNs, but also incorporates their Spanish correspondences, which have received substantially less attention in the literature.

1.2 Objectives

The main goal of this study was to investigate CNs in terminology, specifically those found in the wind power discourse in English. Special attention was accorded to the formation of

[1] This thesis was developed as part of the CONTENT (Cognitive and Neurological Bases for Terminology-Enhanced Translation) and TOTEM (Translation-Oriented Terminology Tools for Environmental Texts) research projects of the LexiCon group, funded by the Spanish Ministry of Economy and Competitiveness. Funding was also provided by an FPU grant given by the Spanish Ministry of Education to the PhD student.

these multiword terms, their translation into Spanish, as well as their representation in terminographic resources.

For that purpose, a set of specific objectives were established:

-To study the status of the issue and address the specific characteristics of CNs in specialized discourse.

-To analyze CNs semantically, as the starting point from the following objectives.

-To explore CN formation from a semantic perspective, by means of a slot opening/filling mechanism activated by the nominal heads. This mechanism was supposed to be the basis for the regularities and productivity of CNs in term formation.

-To investigate CN translation in English and Spanish with regard to equivalence identification in corpus, correspondence analysis, and the proposal of translation correspondences. For those purposes, we assumed that CN formal and semantic structure is not always paralleled in the target language.

-To design a systematic proposal for CN representation in the terminological knowledge base EcoLexicon, which encompasses data entered by terminologists and visualization by users through different access paths.

In sum, we present an innovative proposal that aggregates the work on CNs in disciplines such as General Linguistics, Terminology, Lexicography, Computational Linguistics, and Psycholinguistics. We identified the weak points and proposed new approaches, with a view to profiling CN behavior in terminology.

1.3 Structure and contents of the thesis

After introducing the study in Section 1, the linguistic and terminological foundations upon which the research is based are presented in Section 2. More specifically, the focus is on the main aspects of Cognitive Linguistics (Section 2.1) that inspired the development of Frame-Based Terminology and, thus, our research. These aspects include categorization (Section 2.1.1) and Frame Semantics (Section 2.1.2). Frame-Based Terminology is also outlined in Section 2.2. In this sense, the theories that preceded it are summarized in Section 2.2.1: General Theory of Terminology (Section 2.2.1.1), Socioterminology (Section 2.2.1.2), Communicative Theory of Terminology (Section 2.2.1.3), and Sociocognitive Terminology (Section 2.2.1.4). Section 2.2.2 describes the main premises of Frame-Based

Terminology, whereas Section 2.2.3 presents EcoLexicon, the terminological knowledge base resulting from its application.

Section 3 explores the theoretical insights regarding CNs. After addressing the role of phraseological units (Section 3.1), the notion of CN and the main studies focusing on these multiword combinations are discussed (Section 3.2). Different types of CNs as well as what is known about their formation are then presented (Section 3.3). Next, Section 3.4 shows the different aspects emerging from their analysis, i.e. bracketing (Section 3.4.1), semantics of their constituents (Section 3.4.2), analysis of their internal semantic relations (Section 3.4.3), their variation (Section 3.4.4), and neological nature (Section 3.4.5). Finally, Section 3.5 focuses on the translation of multiword terms, and Section 3.6 explores their representation in linguistic resources.

Section 4 describes the materials used in this study (Section 4.1), as well as the methodology followed in this research (Section 4.2). The materials include a wind power corpus (Section 4.1.1) and the corpus analysis tool Sketch Engine (Section 4.1.2). As for the methodology, the corpus compilation (Section 4.2.1), English CN extraction (Section 4.2.2), their analysis, and the search for correspondences in Spanish (Section 4.2.3) are explained.

Section 5 presents the results and discussion of the thesis. Section 5.1 examines CN formation, and explains the semantic analysis carried out (Section 5.1.1) as well as our proposal of *microcontexts* (Section 5.1.2). Section 5.2 presents the results concerning CN translation, which address the procedure for correspondence identification in the corpus (Section 5.2.1) and the comparison of equivalences (Section 5.2.2). Section 5.3 provides the results that deal with the representation of CNs in terminographic resources. In particular, the template for entering the CN information in the phraseological module of EcoLexicon is given in Section 5.3.1. Whereas Section 5.3.2 describes the user interface, Section 5.3.3 sums up the conclusions derived from the CN module design. Finally, Section 6 presents the conclusions of the study as well as future lines of research.

2. Bases lingüísticas y terminológicas

2.1 La Lingüística Cognitiva

La Lingüística Cognitiva representa uno de los modelos lingüísticos más influyentes y que más investigación ha suscitado en distintas áreas, como la Psicología, la Filosofía, la Neurociencia, la Lingüística o la Terminología. Además, su aplicación ha servido de inspiración a un gran número de teorías, como la Semántica de Marcos (Fillmore 1977, 1982, 1985), la Semántica Conceptual (Jackendoff 1983, 1997), la Gramática de Construcciones (Fillmore y Kay 1987; Kay y Fillmore 1999; Goldberg 1995) o la Semántica Cognitiva (Talmy 2000). En este apartado comentamos de forma sucinta las aportaciones de la Lingüística Cognitiva que resultaron de mayor interés para nuestro trabajo. En concreto, este modelo presenta un marco muy interesante para el análisis del lenguaje especializado, ya que presta especial atención a las estructuras conceptuales y a la organización en categorías, aspectos centrales en la Terminología. Sin embargo, la investigación terminológica, por lo general, no ha incorporado modelos lingüísticos, como la Lingüística Cognitiva; por el contrario, se ha centrado en comentar descripciones y observaciones (Faber 2012).

Para empezar, el fundamento básico de la Lingüística Cognitiva es la concepción del lenguaje como reflejo de la mente humana. Así, Lakoff (1990), uno de los máximos exponentes de la Lingüística Cognitiva, señaló que los lingüistas cognitivos deben asumir dos compromisos: el compromiso de generalización (*generalisation commitment*) y el compromiso cognitivo (*cognitive commitment*).

Por una parte, según el compromiso de generalización, se deben caracterizar los principios básicos que subyacen en los diferentes aspectos del lenguaje. En este sentido, la Lingüística Cognitiva se muestra contraria a los enfoques formales, como la Gramática Generativa de Noam Chomsky, y a la separación que estos realizan de áreas como la fonología, la semántica, la pragmática o la sintaxis. Frente a la consideración de que estas se rigen por principios diferentes, desde la Lingüística Cognitiva se defiende que, aunque esta separación puede ser útil para algunos fines, los principios básicos que existen según el compromiso de generalización se dan en todas ellas. Por tanto, la Lingüística Cognitiva contradice esta teoría modular y apuesta por la interrelación (Evans y Green 2006; Evans et al. 2007).

Por ejemplo, el compromiso de generalización resulta patente en la polisemia, que según la Lingüística Cognitiva, no afecta únicamente al significado, sino a las distintas

facetas del lenguaje humano. Por su parte, las fronteras difusas también están presentes en diferentes aspectos del lenguaje, como las categorías semánticas o gramaticales. Lo mismo sucede con la metáfora, fenómeno por el cual un dominio conceptual se estructura de forma habitual en función de otro (Evans y Green 2006; Evans et al. 2007). De este modo, son frecuentes las extensiones de significado (como sucede, por ejemplo, en el CN analizado *wind penetration*), de forma que se concibe la metáfora como un fenómeno conceptual y no meramente lingüístico (Lakoff y Johnson 1980).

Los universales lingüísticos también ilustran el compromiso de generalización. Se trata de rasgos que coinciden en las diferentes lenguas, al margen de las diferencias obvias entre estas. Estos universales no deben confundirse con la noción propuesta desde teorías formales como la de Chomsky (1965), quien defiende que las personas nacemos con una gramática universal que explica la presencia de esos principios y reglas que comparten todas las lenguas. Respecto a los universales, se ha adoptado también un enfoque semántico, con la propuesta de *semantic primes* o *semantic primitives* realizada por autores como Jackendoff (1983) o Wierzbicka (1996), según la cual existe una serie de conceptos primitivos semánticos innatos que reciben sus denominaciones en cada lengua. Sin embargo, la concepción de universales lingüísticos favorecida desde la Lingüística Cognitiva dista de las aportaciones comentadas, ya que se aboga por la influencia de la cognición corporeizada en el desarrollo de universales (Evans y Green 2006; Evans et al. 2007). Esta percepción de la cognición se basa en lo experimentado por nuestro cuerpo, como comentaremos a continuación.

Junto al compromiso de generalización, el compromiso cognitivo consiste en caracterizar el lenguaje según las aportaciones de otras disciplinas acerca de la mente y el cerebro. Además de estas premisas básicas, en la Lingüística Cognitiva pueden distinguirse dos vertientes: la Semántica Cognitiva y los enfoques cognitivos de la Gramática. A pesar del papel que se otorga a la gramática, la Lingüística Cognitiva enfatiza el significado, ya que defiende que es prioritario y de él depende la gramática (Evans y Green 2006; Evans et al. 2007).

Por una parte, la Semántica Cognitiva estudia la representación del conocimiento mediante sistemas conceptuales y la construcción de significado. Por tanto, se investiga cómo el lenguaje codifica y refleja la estructura conceptual (Evans y Green 2006; Evans et al. 2007). Los diferentes ámbitos en los que se centra la Semántica Cognitiva se guían por cuatro premisas básicas:

-La estructura conceptual se basa en lo que hemos experimentado con nuestro cuerpo. Se aborda, así, la noción de *embodied experience* o experiencia corporeizada, que ocupa un papel central en la Lingüística Cognitiva. Por tanto, mediante la experiencia corporeizada, denominada *experientialism* o *experiential realism* por Lakoff y Johnson (1980), se defiende que nuestra visión del mundo depende de lo experimentado por nuestro cuerpo, ya sea de forma física o subjetiva. Así, se afirma que nuestro cerebro procesa los conceptos mediante simulaciones (Faber y León Araúz 2014). Autores como Barsalou (2003) han desarrollado teorías de cognición situada.

-La estructura semántica se corresponde con conceptos (Evans y Green 2006, Evans et al. 2007). Esto es, los significados asociados a las unidades lingüísticas aluden a estructuras conceptuales.

-La representación del significado es enciclopédica. Así, más que significado, las unidades lingüísticas tienen potencial de significado, que se especifica según el contexto (lingüístico, situacional, etc.) (Allwood 2003; Croft y Cruse 2004; San Martín 2016). Esta es una de las diferencias entre la Lingüística Cognitiva y los modelos formales, que consideran el significado tal como se presenta en los diccionarios, sin tener en cuenta cuestiones pragmáticas. No obstante, según el compromiso de generalización de la Lingüística Cognitiva, la semántica y la pragmática están interrelacionadas. Así, el contexto se concibe como fundamental para la delimitación del significado de las unidades lingüísticas, ya que además puede generar ambigüedad.

-La construcción de significado no se realiza a nivel de palabras, sino a nivel de conceptos. Las unidades lingüísticas se conciben, por tanto, como meras puertas de acceso. Uno de los principales autores que se han centrado en el dinamismo de la construcción de conocimiento ha sido Fauconnier (1994, 1997), máximo exponente de la Teoría de los Espacios Mentales, quien defiende que la construcción de conocimiento se realiza a través del lenguaje en contexto. Según Fauconnier (1994, 1997), en la construcción del conocimiento intervienen dos procesos: (i) la construcción de espacios mentales, que son «partial structures that proliferate when we think and talk, allowing a fine-grained partitioning of our discourse and knowledge structures» (Fauconnier 1997: 11) y (ii) el establecimiento de *mappings* entre los espacios mentales. Por tanto, se incide de nuevo en la relevancia del contexto y en la interrelación entre la semántica y la pragmática. De este modo, la construcción de conocimiento siempre es situada o localizada en un determinado contexto, actividad o cultura.

Por su parte, los enfoques cognitivos de la Gramática se centran en el estudio de los símbolos que comprenden el lenguaje (Evans y Green 2006; Evans et al. 2007). Estos símbolos no aluden a la gramática de forma aislada, sino como estructuras de forma-significado (*form-meaning pairings*). Ello justifica la atención que reciben combinaciones como los *idioms* y el interés de analizar los términos compuestos desde perspectivas cognitivas. Por tanto, vuelve a resultar patente el rechazo a la teoría modular, ya que se destaca la relación entre semántica y sintaxis: el significado y la gramática se conciben como dos caras de la misma moneda (Evans y Green 2006: 49). En estos modelos gramaticales se aprecian dos tendencias: el estudio de los principios cognitivos que organizan el lenguaje y la investigación más minuciosa de las unidades lingüísticas.

En la primera línea, destacan las aportaciones de Langacker (1987), quien en su teoría de la Gramática Cognitiva, propugna que el lenguaje se rige por los mismos principios que otros aspectos de la cognición humana (se aprecia, de nuevo, el compromiso de generalización). La gramática se entiende aquí en un sentido amplio para referirse a todo el sistema lingüístico (sonido, significado y morfosintaxis). Además, Langacker (1987) defiende que el uso influye en la representación gramatical en el cerebro, generando un mayor afianzamiento cuando se da un uso reiterado.

En la segunda línea, autores como Fillmore y Kay (Fillmore et al. 1988; Kay y Fillmore 1999) o Goldberg (1995) se han dedicado al análisis de las unidades del lenguaje. En estos estudios, la noción de *construcción* cobra un papel relevante. En la Lingüística Cognitiva, las construcciones se entienden como unidades básicas del lenguaje, consistentes en pares de forma-significado. Estas se han empleado en múltiples ocasiones para explicar cuestiones fraseológicas, pues desde la Lingüística Cognitiva se aboga por el significado construccional (más cercano a lo idiomático) de estas combinaciones. Frente a esta concepción, los enfoques formales adoptan una postura composicional, al considerar que el significado global constituye la suma del significado de los componentes.

En este sentido, compartimos la opinión de Taylor (2002: 550), quien no cree que las combinaciones deban clasificarse según su grado de composicionalidad, pues la composicionalidad total es poco frecuente, si no inexistente. Como veremos, aunque el significado de los CN especializados pueda inferirse de sus componentes, suele haber información implícita (como la relación interna entre los formantes) que dificulta la obtención del significado global.

Así pues, en la Gramática de Construcciones de Kay y Fillmore (Fillmore y Kay 1987; Kay y Fillmore 1999) no se incluye solo información sintáctica, sino también

semántica y pragmática. En cuanto a la composicionalidad, defienden que el significado de la combinación no se deriva de las partes sino de la propia construcción. Esta teoría y la Gramática Cognitiva (Langacker 1987) coinciden en distintas ideas, si bien presentan una diferencia formal: mientras que la Gramática de Construcciones es un enfoque formal, que adopta la Gramática Universal de Chomsky (1965), la Gramática Cognitiva se basa en el uso y, por tanto, rechaza la hipótesis de la Gramática Universal (Evans y Green 2006). No obstante, la Gramática de Construcciones, a pesar de no ser un enfoque cognitivo, comparte alguna de sus premisas (como el reconocimiento del significado idiomático y no composicional de muchas unidades), por lo que ha influido también en diversos enfoques cognitivos.

Goldberg (1995), por su parte, se apoya en estos trabajos iniciales para desarrollar un nuevo enfoque de la Gramática de Construcciones, en el que se centra en las construcciones verbo-argumento habituales y regulares de la lengua. Esta nueva propuesta, sin embargo, está basada en el uso y, por tanto, constituye un enfoque cognitivo, a diferencia de la Gramática de Construcciones de Kay y Fillmore (Fillmore y Kay 1987; Kay y Fillmore 1999).

Goldberg (1995) se posiciona a favor de la Semántica de Marcos de Fillmore (1977, 1982, 1985) como representación del significado de los verbos y utiliza roles para la caracterización de los argumentos (Sección 2.1.2). Como vemos, desde la Lingüística Cognitiva se presta especial atención a la valencia o estructura argumental. No obstante, la percepción de valencia de Goldberg (1995) difiere en cierta medida de la concepción general, ya que no parte necesariamente de un predicado, sino de una construcción oracional. Como comentaremos en la Sección 5.1.1.4, en nuestro análisis también exploramos la estructura argumental, partiendo de las proposiciones conceptuales que subyacen en los CN.

Otro de los temas centrales de la Lingüística Cognitiva es la categorización, debido a su relevancia en la estructuración conceptual, como detallaremos a continuación.

2.1.1 La categorización

La capacidad de categorizar u organizar en grupos es fundamental en la cognición humana y por ello desempeña un papel privilegiado en la Lingüística Cognitiva. La categorización se aborda tanto en la Semántica Cognitiva, donde se desarrollan categorías de conceptos, como en los enfoques cognitivos de la Gramática, donde se estudian categorías de elementos gramaticales (Hampton 1991; Keizer 2007). En este trabajo nos interesa principalmente la categorización desde el punto de vista semántico.

Con ese fin, se siguieron durante siglos las nociones clasificatorias propuestas por Aristóteles en el ámbito de la Filosofía. Según este modelo clásico, las categorías conceptuales presentan una estructura definicional, establecida según unas condiciones necesarias y suficientes que reúnen todos los miembros de la categoría. Estos requisitos son necesarios de forma individual, pero solo en conjunto pueden ser suficientes para definir una categoría (Evans y Green 2006; Evans et al. 2007). Por tanto, se entendía que, si un concepto no reunía todas las condiciones necesarias y suficientes, no formaba parte de la categoría (Hampton 1991; Keizer 2007).

Estas ideas se mantuvieron válidas hasta la década de los 70, cuando Rosch (1978) cuestionó el modelo clásico de categorización y planteó un enfoque más diluido. Para Rosch (1978), en lugar de condiciones necesarias y suficientes, las categorías se organizan tomando como referencia prototipos, que constituyen representaciones mentales del miembro más característico de una categoría, cuyos rasgos se toman como los más destacados del grupo. Por ejemplo, si pensamos en la categoría MUEBLE, es más posible que reproduzcamos la imagen de una mesa a la de un baúl. Este poder representativo de los prototipos se conoce como *typicality* y ha sido validado en diferentes estudios cognitivos. Sin embargo, no se reconocía en el modelo clásico.

Por tanto, en su Teoría de los Prototipos, Rosch (1978) defiende que estos presentan muchos rasgos en común con otros miembros de la categoría, mientras que los conceptos menos prototípicos coinciden en pocas características con el resto de miembros. Los rasgos más recurrentes entre los conceptos de una categoría se conciben como los más representativos. Así, el grado de similitud con los prototipos es el que determina la pertenencia a las categorías. Estas se organizan en círculos concéntricos alrededor del prototipo. A menor semejanza con este, más cercanía a los límites de la categoría, que son difusos. Por tanto, las categorías no se conforman según condiciones necesarias y suficientes de sus miembros, ya que en los extremos, los conceptos menos prototípicos pueden presentar también características de otras categorías.

Por otra parte, Rosch (1978) considera que el sistema de categorización presenta dos dimensiones: una vertical y otra horizontal. La dimensión vertical hace referencia al nivel de inclusión de una categoría: a mayor elevación en el eje, mayor inclusión de conceptos en la categoría. Esta dimensión alude a las relaciones jerárquicas, como la hiperonimia, pues las categorías más generales, como los hiperónimos, presentan una mayor capacidad de inclusión. Así, se distinguen tres niveles en este eje vertical, que dan lugar a categorías básicas (nivel intermedio), superordinadas (nivel superior) y

subordinadas (nivel inferior). Aunque el nivel superordinado sea el más inclusivo, el nivel óptimo para el ahorro cognitivo es el nivel básico, cuyas categorías poseen más rasgos en común (Rosch 1978).

Las categorías básicas surgen de la interacción entre la experiencia humana y el medio, de modo que se ven influidas por la cognición situada. En esta línea, aunque la organización de categorías en los niveles básico, superordinado y subordinado puede ser universal, el nivel en el que se incluye una determinada categoría o concepto puede variar según la lengua, la cultura o el nivel de conocimiento de la persona (Evans y Green 2006).

Tomando como referencia ese eje vertical, la dimensión horizontal que también señala Rosch (1978) alude a las diferencias existentes entre categorías que presentan el mismo grado de inclusión. En general, para esta autora son dos los principios que guían la formación de categorías en la mente humana. Por una parte, el principio del ahorro cognitivo, según el cual pretendemos almacenar el máximo de información ahorrando el esfuerzo cognitivo, y por otra, el principio de la estructura percibida del mundo, que atiende a los conceptos que suelen relacionarse más frecuentemente con otros.

Como se observa, la propuesta de Rosch pretende dar cuenta de la clasificación conceptual en la mente humana, señalando la existencia de modelos representativos y relaciones entre categorías, que se alejan de los límites estrictos, además de la influencia de la experiencia y el entorno en la estructuración conceptual. Con ella estuvieron de acuerdo autores como Fillmore (1982), Jackendoff (1983), Langacker (1987) o Lakoff (1987). Este último desarrolló una estructuración cognitiva de las categorías que sigue la línea de Rosch. Se trata de la teoría de los modelos cognitivos idealizados (*idealised cognitive models*) (Lakoff 1987), que son representaciones mentales de teorías sobre el mundo y se utilizan para estructurar los espacios mentales. Se les cataloga de idealizados porque, lejos de representar casos específicos, aglutinan diversas experiencias. Por tanto, se asemejan a los marcos de Fillmore (1977, 1982, 1985), que comentaremos a continuación.

Los modelos cognitivos idealizados guían procesos cognitivos como la categorización y el razonamiento. Así, entre otros aspectos, Lakoff (1987) plantea la existencia de modelos cognitivos idealizados metonímicos, en los que se toma un ejemplar como representante de la categoría (parte-todo). Junto a esta noción, que recuerda a la de prototipo, habla de las categorías radiales, que permiten explicar la inclusión de miembros en función del grado de semejanza con el prototipo.

Una de las ventajas de esta teoría, que nos interesa especialmente por su relación con los CN, es su propuesta de la composicionalidad. Para ilustrar su postura, emplearemos el clásico ejemplo de PET FISH. Mientras que, para Rosch (1978), los miembros de esa categoría comparten las características de PET y de FISH, para Lakoff (1987) esta categoría presenta rasgos propios independientes de ambas categorías, lo que da cuenta de su idiomaticidad y su falta de composicionalidad.

La categorización es, por tanto, un tema central en la Lingüística Cognitiva y también en la Terminología, ya que permite organizar el conocimiento especializado. Así, Sager (1990) y Sager y Kageura (1995) desarrollaron la primera propuesta rigurosa de categorías semánticas para la Terminología. En ella distinguen entre ENTIDADES (tienen una existencia determinada en el tiempo y en el espacio, y pueden ser materiales o inmateriales), ACTIVIDADES (son procesos, operaciones o eventos realizados con o por ENTIDADES), PROPIEDADES (son características de otros conceptos) y RELACIONES (constituyen vínculos entre ENTIDADES, ACTIVIDADES y PROPIEDADES) (Kageura 2002: 66-67).

Son múltiples las aplicaciones de la categorización en Terminología y van desde la organización conceptual (por ejemplo, en ontologías [Mairal et al. 2011]), hasta la predicción de regularidades motivadas conceptualmente, por citar solo algunos ejemplos. Es por ello que las categorías semánticas resultaron de gran utilidad en este trabajo.

2.1.2 La Semántica de Marcos

La Semántica de Marcos es una teoría desarrollada por Fillmore (1977, 1982, 1985) mediante la cual se explica el significado léxico recurriendo al conocimiento enciclopédico. Para ello, Fillmore se inspiró en la noción de marco de la Inteligencia Artificial, donde estos se conciben como estructuras de datos que representan situaciones estereotipadas, como una fiesta de cumpleaños (Minsky 1975: 212). Así, los marcos forman redes con nodos enlazados por relaciones y etiquetados por *slots* que se rellenan con los datos específicos de la situación en cuestión (Minsky 1975).

En la Semántica de Marcos, estos constituyen estructuras de conocimiento tomadas a partir de experiencias diarias que permiten la contextualización de los conceptos. En ellos se indica cómo se relacionan los conceptos implicados, de modo que el significado de una palabra no puede comprenderse si no se considera el marco en el que aparece (Fillmore 1977, 1982, 1985). A diferencia de los marcos de Minsky (1975), que eran conceptuales, los marcos de Fillmore (1977, 1982, 1985) no se plantean a nivel conceptual, sino anclados a una lengua, en concreto, la lengua inglesa.

Por tanto, para conocer el significado de elementos como *buy* o *sell*, Fillmore (1982: 116-117) señala que es necesario acceder al EVENTO COMERCIAL (COMMERCIAL EVENT), en el que se incluyen también participantes como BUYER, SELLER, GOODS y MONEY. Como se infiere de estos ejemplos, Fillmore, además de otros lingüistas cognitivos, presta especial atención a la valencia o estructura argumental, que hace alusión al modo en el que las unidades predicativas se combinan con otras para generar sentido. Esta valencia, que también mencionamos en nuestro análisis, donde se incluyen unidades predicativas, está determinada por la semántica del verbo y puede ser cuantitativa (número de participantes o argumentos requeridos) o cualitativa (tipo de argumentos). En la valencia cuantitativa, se distinguen los argumentos obligatorios, que siempre deben acompañar al predicado, y los opcionales.

En la valencia cualitativa, los argumentos suelen caracterizarse según roles semánticos (también denominados roles temáticos), que se han empleado con frecuencia en diversas vertientes de la Lingüística (Faber y Mairal 2002). Se trata de categorías funcionales independientes de la función sintáctica del argumento, por ello no deben confundirse con las funciones de sujeto, objeto, etc., como ocurre en algunos enfoques. A pesar de que existe un consenso general sobre la utilidad de los roles, son diversos los inventarios existentes, que incluyen un mayor o menor número de roles. Los ejemplos más característicos son AGENTE y PACIENTE, además de otros como RESULTADO, INSTRUMENTO, etc.

Ya en la Gramática de Casos, antecedente de la Semántica de Marcos, se utilizaban los roles semánticos para caracterizar los argumentos de los predicados. Estos ofrecen distintos beneficios, como la posibilidad de hacer generalizaciones, captar regularidades o trazar equivalencias interlingüísticas, entre otras. Por ello, se han empleado en distintos proyectos lingüísticos, como la anotación con roles que se da en WordNet (Fellbaum 1998); FrameNet (Ruppenhofer et al. 2010), que comentaremos a continuación; Verbnet (Kipper et al. 2004); o PropBank (Palmer et al. 2005).

La Semántica de Marcos guarda ciertas similitudes con otras teorías, como la Teoría de los Dominios de Langacker (1987). Según este autor, los dominios proporcionan información situacional que favorece la comprensión y el empleo de las unidades lingüísticas. No obstante, el enfoque de Langacker difiere del adoptado por Fillmore en varias cuestiones. Por una parte, Langacker (1987) considera que la comprensión de las unidades léxicas no puede realizarse considerando un solo dominio, sino necesariamente varios. Por otra, estructura los dominios de forma jerárquica y distingue, además, dos tipos: los dominios básicos, como el tiempo, y los dominios abstractos, que son más

complejos, como el amor. Asimismo, aporta una visión más conceptual que la que adopta en ocasiones Fillmore, quien a menudo se centra en el comportamiento gramatical de los participantes. Así, estas estructuras de conocimiento, entendidas como marcos o dominios, presentan distintas utilidades y aplicaciones, como su empleo en sistemas informáticos o en la Psicología, donde permiten hacer inferencias sobre situaciones complejas, elaborar predicciones sobre las consecuencias de una acción, etc. (Coulson 2001).

La aplicación práctica de la Semántica de Marcos es FrameNet (Ruppenhofer et al. 2010), una base de datos léxica de la lengua inglesa que se desarrolla en el International Computer Science Institute de Berkeley desde 1997. Esta se centra en la representación de las posibilidades combinatorias de los conceptos que evocan los marcos.

Así, en la entrada de cada marco se presenta la definición del mismo, que se complementa con ejemplos de textos reales anotados con roles, denominados *frame elements*. Estos son específicos de cada marco e identifican a los participantes en el mismo. Algunos son obligatorios (*core frame elements*) y otros, opcionales (*non-core frame elements*). Al presentarse ejemplos reales, se pueden observar las unidades léxicas que ocupan los *frame elements* y sus realizaciones sintácticas.

Por ejemplo, en el marco de Conquering, los participantes obligatorios son Conqueror y Theme, que pueden aparecer junto a participantes opcionales como Degree, Depictive, Frequency, Instrument, etc. Las unidades léxicas suelen estar representadas por verbos (*capture, conquer, fall*, etc.), aunque también pueden presentarse otras categorías gramaticales. No obstante, resulta especialmente rico en el caso de elementos predicativos. También se indican las relaciones con otros marcos; por ejemplo, el marco Agriculture se relaciona con el de Attempt_obtain_food_scenario mediante la relación *inherits from* y con los marcos de Food_gathering y Planting mediante la relación *has subframes*.

En definitiva, se observa cómo la Semántica de Marcos ha realizado grandes aportaciones en el ámbito del acceso y la estructuración del conocimiento. Por ello, su huella se aprecia en recursos terminográficos como DiCoEnviro (Sección 3.6.2.6) y otras herramientas elaboradas en la Universidad de Montreal, además de percibirse en teorías terminológicas como la Terminología Basada en Marcos (Sección 2.2.2), en la que se basa este trabajo.

2.2 La Terminología Basada en Marcos

2.2.1 Antecedentes

La Terminología Basada en Marcos es una teoría terminológica que toma algunas premisas del trabajo previo, a partir del cual propone nuevos enfoques. A continuación se resumen las principales aportaciones de las teorías que, de un modo u otro, han influido en el desarrollo de la Terminología Basada en Marcos.

2.2.1.1 La Teoría General de la Terminología

La Teoría General de la Terminología (Wüster 1968, 1979) fue la primera propuesta teórica en estudiar el área de la Terminología. Su iniciador fue Eugen Wüster, un ingeniero austriaco que desarrolló esta teoría a partir de su experiencia al elaborar *The Machine Tool. An interlingual Dictionary of Basic Concepts* (Wüster 1968), un diccionario de términos estandarizados en ocho lenguas, concebido para su uso internacional.

Wüster (1968, 1979) entendía que la función de la terminología era la de crear y estandarizar denominaciones para los conceptos. Por ello su enfoque es eminentemente prescriptivo, además de onomasiológico, ya que parte del concepto, mientras que los términos se conciben como meras etiquetas.

La visión de Wüster (1968, 1979) de la realidad terminológica es bastante idealizada, ya que simplificaba o eludía aspectos problemáticos como el contexto, la variación, la fraseología o la diacronía. Esto se produce con el ánimo de conceder a la Terminología un estatus de disciplina independiente. Es por ello también que se incide en la diferencia entre la lengua general y la lengua especializada y, con ello, en la distinción entre palabra (lengua general) y término (lengua especializada). Sin embargo, basta con acercarse a cualquier muestra de discurso especializado para percatarse de que esta idealización no es más que una quimera.

Así, desde la Teoría General de la Terminología (Wüster 1968, 1979; Felber 1984) se afirma la inexistencia de variación en la lengua especializada, sea del tipo que fuere. Esto es, se defiende la referencia unívoca y estable entre término y concepto: un concepto se denomina con un único término (se niega la variación denominativa) y un término no puede aludir a más de un concepto (se rechaza la variación conceptual). Se elimina de este modo la ambigüedad y se consigue una comunicación efectiva. Además, los aspectos gramaticales y discursivos se consideran

irrelevantes, careciendo así de interés cuestiones como el contexto, la sintaxis, el lenguaje figurado o la fraseología.

Por tanto, los principales objetivos de Wüster y de la Teoría General de la Terminología eran: (i) eliminar la ambigüedad de los lenguajes técnicos y facilitar la comunicación efectiva mediante la estandarización, (ii) convencer a los usuarios de los lenguajes especializados de la utilidad de la terminología estandarizada y (iii) otorgar a la Terminología un estatus de ciencia (Cabré 2003: 173).

Aunque los principios y la metodología de la Teoría General de la Terminología siguen vigentes en algunos ámbitos, como en el trabajo estandarizador de la Organización Internacional de Normalización (ISO), en la década de los 90 surgieron nuevos enfoques que ofrecían una visión más realista y descriptiva de la terminología, al analizar los términos en contextos reales de uso. Se trata de la Socioterminología (Gaudin 1993, 2003, 2005), la Teoría Comunicativa de la Terminología (Cabré 1993, 1999), la Terminología Sociocognitiva (Temmerman 2000, 2007) y la Terminología Basada en Marcos (Faber et al. 2005, 2006; Faber 2012), en la que nos inspiramos en este trabajo. Estos nuevos enfoques ponen el foco en cuestiones sociales, lingüísticas, comunicativas y cognitivas de la terminología, y adoptan un enfoque semasiológico, al partir del término para después analizar el concepto.

2.2.1.2 La Socioterminología

La Socioterminología (Gaudin 1993, 2003, 2005) es una teoría de la Terminología que surge al aplicar a esta disciplina los principios de la Sociolingüística. Su estudio se da en dos núcleos principales: Francia, con los seguidores de Guilbert, tales como Gambier, Guespin y Gaudin; y Quebec, con autores como Auger, Boulanger o Corbeil (Gaudin 2005).

La Socioterminología es una de las teorías que se alzaron en respuesta a la perspectiva idealizada de la Teoría General de la Terminología. De este modo, se reconoce la existencia de variación denominativa y conceptual en el lenguaje especializado, fenómenos que se estudian en sus contextos de uso. Así, el uso y los usuarios, especialmente los aspectos sociales y étnicos, son factores determinantes en la variación. Esta revela el conocimiento de los participantes, su estatus social y profesional, las relaciones de poder, así como su ubicación geográfica y temporal. Se estudia, por tanto, la diacronía y, con ella, la neología adquiere también un papel relevante (Gaudin 1993, 2003, 2005).

Se defiende que no existen límites exactos entre los dominios especializados. Asimismo, se considera que la estandarización no es más que ficción, ya que la lengua y los sistemas conceptuales no dejan de evolucionar (Gaudin 1993, 2003, 2005). Se trata, por tanto, de un enfoque descriptivo y semasiológico, en el que se recoge el uso de los términos y se parte de estos para llegar a los conceptos. Constituye un avance importante, pues inició las nuevas teorías descriptivas de la Terminología que intentaron presentar una realidad más precisa de esta disciplina.

2.2.1.3 La Teoría Comunicativa de la Terminología

La Teoría Comunicativa de la Terminología (Cabré 1993, 1999) surge en el Institut de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra, principalmente gracias a las aportaciones de Cabré, que sentó las bases de estudio que después seguirían investigadoras como Lorente, Estopà o Freixa, entre otros.

Desde esta teoría se señala el carácter de la Terminología como disciplina transdisciplinar, ya que todas las materias especializadas hacen uso de ella, e interdisciplinar, pues se sirve de diversas materias para describir su objeto de estudio. En concreto, se entiende que la Terminología integra teorías del conocimiento, de la comunicación y del lenguaje (Cabré 1993, 1999).

Así, mediante su *teoría de las puertas*, Cabré (2003) propugna que las unidades terminológicas se asemejan a un poliedro, una figura tridimensional con varias caras. De este modo, se puede acceder a ellas mediante tres puertas o dimensiones: la cognitiva, la comunicativa o la lingüística, sin que ninguna de ellas deje de estar presente en la unidad terminológica (aunque la puerta lingüística es la más habitual en el contexto de la comunicación especializada).

En su faceta cognitiva, se pretende dar cuenta del modo en que se conceptualiza la realidad. Para ello, los dominios conceptuales se entienden como conjuntos de unidades terminológicas vinculadas por relaciones de distinto tipo, en las que las unidades terminológicas constituyen nodos. En esta línea, desde la Teoría Comunicativa de la Terminología, se incide en que los conceptos pueden pertenecer a distintas disciplinas, en las que pueden conservar o modificar sus características (Cabré 1993, 1999, 2003). Se destaca, por tanto, la importancia de la multidimensionalidad, que como veremos, consiste en particularizar un mismo concepto mediante distintas perspectivas o características.

Por otra parte, en su faceta comunicativa, se atiende a los distintos tipos de situaciones que pueden darse y su influencia en la comunicación. Ello deriva en un

papel central de la variación como ámbito de estudio, a diferencia de la perspectiva inicial adoptada por Wüster y sus seguidores. Por último, en cuanto a la faceta lingüística, se defiende que los términos se comportan como palabras y adquieren su especificidad al emplearse en contextos y situaciones determinados. Así, al contrario que la Teoría General de la Terminología, se entiende que la frontera entre la lengua general y la especializada no es clara.

Por tanto, a nuestro parecer, son diversas las ventajas de esta teoría: ha contribuido a asentar la Terminología como disciplina; reconoce la importancia de la cognición, la comunicación y la lingüística en la Terminología; y contempla diferentes contextos de uso, además de fenómenos importantes como la variación o la multidimensionalidad. No obstante, presenta algunas desventajas. En concreto, no se basa en ningún modelo lingüístico; no clarifica cómo se crean las representaciones conceptuales que utilizan ni cuáles son sus características (más allá de decir que se trata de mapas conceptuales formados por nodos y relaciones); además de no especificar cómo se realiza el análisis semántico (Faber y López Rodríguez 2012).

2.2.1.4 La Terminología Sociocognitiva

La Terminología Sociocognitiva (Temmerman 2000, 2007) surge con el giro cognitivo de la Terminología, a partir del cual se otorga una mayor importancia a las estructuras conceptuales que subyacen en el lenguaje especializado. Así, esta teoría, junto a la Terminología Basada en Marcos que explicaremos a continuación, presenta un enfoque más cognitivo que la Teoría Comunicativa de la Terminología, al integrar premisas de la Lingüística Cognitiva y la Psicología.

En la Terminología Sociocognitiva (Temmerman 2000), los conceptos se denominan *unidades de comprensión* y, de este modo, su comprensión implica la organización en categorías, que se encuadran en modelos cognitivos idealizados (Lakoff 1987) y suelen presentar una estructura prototípica (Rosch 1978). Por tanto, se pone un énfasis especial en nociones cognitivas como la categorización y los prototipos. Además, se integra el enfoque cognitivo de la metáfora, que está presente en la categorización y da lugar a lexicalizaciones metafóricas (Temmerman 2000).

Se critican distintos aspectos de la Teoría General de la Terminología, en concreto el enfoque onomasiológico y prescriptivo, la existencia de fronteras bien delimitadas en los conceptos y categorías, las descripciones basadas en la intensión, la univocidad de los términos y los estudios sincrónicos (Temmerman 2000, 2007). Al contrario, desde la Terminología Sociocognitiva se adopta un enfoque semasiológico y descriptivo, y se

defiende que las fronteras entre conceptos y categorías no están bien delimitadas, pues suelen presentar una estructura prototípica. Así, la definición no debe basarse únicamente en la intensión, sino que dependerá del tipo de unidad de comprensión, el dominio especializado en el que esta se dé (se hace alusión a la multidimensionalidad) y el perfil de los usuarios. Por otra parte, se destaca la importancia de la variación y se llevan a cabo estudios diacrónicos de los términos y los conceptos.

Desde la Terminología Sociocognitiva se ha propuesto la denominada *Termontografía* (*Termontography*, en Temmerman y Kerremans 2003). Se trata de un término híbrido que resulta al unir *terminología* (el estudio del lenguaje especializado), *ontología* (representación en la que se explicita una conceptualización [Gruber 1993]) y *terminografía* (la compilación de recursos lingüísticos que recogen el léxico especializado). Constituye un enfoque multidisciplinar para la gestión y representación del conocimiento en el que se aúnan el análisis terminológico multilingüe de la Terminología Sociocognitiva y la ingeniería del conocimiento, pues se conciben las ontologías como una forma más práctica de realizar representaciones conceptuales.

Sin embargo, aunque se indica que la estructura prototípica de las unidades de comprensión se tiene en cuenta (Temmerman 2007), probablemente la dificultad de integrar este tipo de representaciones cognitivas en sistemas informáticos ha derivado en un uso principal de las ontologías en detrimento de las estructuras prototípicas y los modelos cognitivos idealizados, como señalan Faber y López Rodríguez (2012). No obstante, ello no nos parece un inconveniente, al contrario, destacamos la importancia de este enfoque por su énfasis en la representación conceptual y las perspectivas interesantes que presenta de cara a la gestión y representación del conocimiento en diferentes lenguas.

Otro enfoque interesante que integra las ontologías en el trabajo terminológico es la Ontoterminología (*Ontoterminology* en Roche et al. 2009; Roche 2012). Se trata de un modelo en el que se emplean las ontologías como medio de representación conceptual en terminología y se distingue claramente entre la definición del término (en lenguaje natural) y la definición del concepto (el lenguaje formal) (Carvalho 2018). Por tanto, constituye otra prueba de la relevancia que ha adquirido el sistema conceptual subyacente en las corrientes terminológicas actuales, como medio preciso de representar el conocimiento especializado.

2.2.2 Fundamentos

Tras comentar las teorías terminológicas iniciales, nos detendremos en la Terminología Basada en Marcos (Faber et al. 2005, 2006; Faber 2012), cuya metodología seguimos

en este trabajo. Se trata de un enfoque terminológico de corte cognitivo que surge en la Universidad de Granada y se centra en la representación del conocimiento especializado, prestando especial atención a sus aspectos semánticos y cognitivos. Así, los ejes principales de esta teoría son la organización conceptual, la naturaleza multidimensional de las unidades terminológicas, y la extracción de información semántica y sintáctica a partir de corpus multilingües (Faber 2009: 123).

De este modo, se sigue una doble metodología para acceder al sistema conceptual subyacente. Por una parte, se extrae información de corpus en varias lenguas que aborden el dominio en cuestión. Por otra, se obtiene información de recursos terminográficos y otro material de referencia, que se complementa con la ayuda de expertos del área (Faber 2009: 124).

La Terminología Basada en Marcos comparte algunas de las ideas de la Teoría Comunicativa de la Terminología (Cabré 1993, 1999) y la Terminología Sociocognitiva (Temmerman 2000, 2007), como su carácter descriptivo, al analizar los términos en los textos en los que se utilizan. Igual que estas teorías, defiende que la distinción estricta entre palabras y términos carece de sentido, ya que las fronteras entre la lengua general y la especializada no son claras y estas unidades presentan el mismo comportamiento, aunque se den en contextos diferentes. En concreto, se asumen las diferentes puertas de acceso a las unidades terminológicas, como se plantea en la Teoría Comunicativa de la Terminología, y la existencia de variación en el lenguaje especializado. También coincide con la Terminología Sociocognitiva (y con la Lingüística Cognitiva) en la existencia de prototipos en las categorías, la evolución diacrónica de los términos y los conceptos, y la utilidad de las ontologías en la representación del conocimiento especializado (López Rodríguez et al. 2010).

Sin embargo, la Terminología Basada en Marcos se diferencia de estas teorías al integrar premisas de otras corrientes de estudio, en concreto, el Modelo de la Gramática Léxica (Martín Mingorance 1989; Faber y Mairal 1999), el Lexicón Generativo (Pustejovsky 1995), la Cognición Situada (Barsalou 2003) y, principalmente, la Semántica de Marcos (Fillmore 1977, 1982, 1985), a partir de la cual se adopta la denominación de *Terminología Basada en Marcos* (Faber 2012).

Por un lado, el Modelo de la Gramática Léxica (Martín Mingorance 1989; Faber y Mairal 1999) se utiliza para la extracción y representación de relaciones semánticas y colocacionales presentes en el discurso especializado. Según este modelo, en el lexicón se distingue entre un eje paradigmático y otro sintagmático. El eje paradigmático organiza el

lexicón en una jerarquía de dominios y subdominios, e influye en el eje sintagmático, que da cuenta del potencial combinatorio de los términos. La convergencia de los dos ejes es la base de la estructura conceptual tanto en la lengua general como en la especializada (Faber 2012).

El Lexicón Generativo (Pustejovsky 1995) permite la organización y la restricción de las dimensiones conceptuales según la semántica de los conceptos (León Araúz 2009). Por su parte, la Cognición Situada (Barsalou 2003), una de las ideas claves de la Lingüística Cognitiva, subyace en la importancia que adquieren para la representación del conocimiento elementos como el contexto y las relaciones semánticas no jerárquicas (aquellas que no pertenecen a una jerarquía hiponímica o partitiva, p. ej. las que aluden al objetivo, el uso o el resultado de un concepto). Además, como demostraron Faber et al. (2014), el cerebro representa y procesa los conceptos especializados mediante simulaciones, lo que desvelaron empleando técnicas de imagen por resonancia magnética funcional. Esta idea entronca directamente con la Cognición Situada, de reciente interés en la disciplina terminológica. En general, la Terminología Basada en Marcos propugna que los conceptos deben entenderse y analizarse en contexto y este contexto adquiere la forma de un marco.

Así, los marcos son el principio estructurador del conocimiento especializado, cuya noción se toma de la Semántica de Marcos (Fillmore 1977, 1982, 1985), que a su vez parte de la idea inicial de Minsky (1975) en la Inteligencia Artifical. Como hemos comentado en la Sección 2.1.2, los marcos son estructuras de conocimiento que permiten la contextualización de los conceptos (Fillmore 1977, 1982, 1985). De este modo, se entiende que el significado depende del contexto, como ya se defendía desde el enfoque enciclopédico de la Semántica Cognitiva (Allwood 2003; Croft y Cruse 2004), y que los términos son la puerta de acceso a los conceptos. Sin embargo, a diferencia del enfoque de Fillmore (1977, 1982, 1985), las representaciones no son específicas de una lengua, sino conceptuales, de modo que se asemejan en este sentido a la propuesta de Minsky (1975).

Esta representación en forma de marcos da lugar a lo que se denomina EVENTO MEDIOAMBIENTAL (Figura 1) (Faber et al. 2005), que subyace en los textos relacionados con el medio ambiente y donde el análisis de los conceptos y sus relaciones permite adquirir el conocimiento (Faber 2012; León Araúz et al. 2012). Así, este evento, que puede particularizarse (p. ej. el EVENTO DE LA ENERGÍA EÓLICA), se compone de las acciones y procesos que tienen lugar en el medio ambiente, así como las entidades que participan en ellos (Faber 2009).

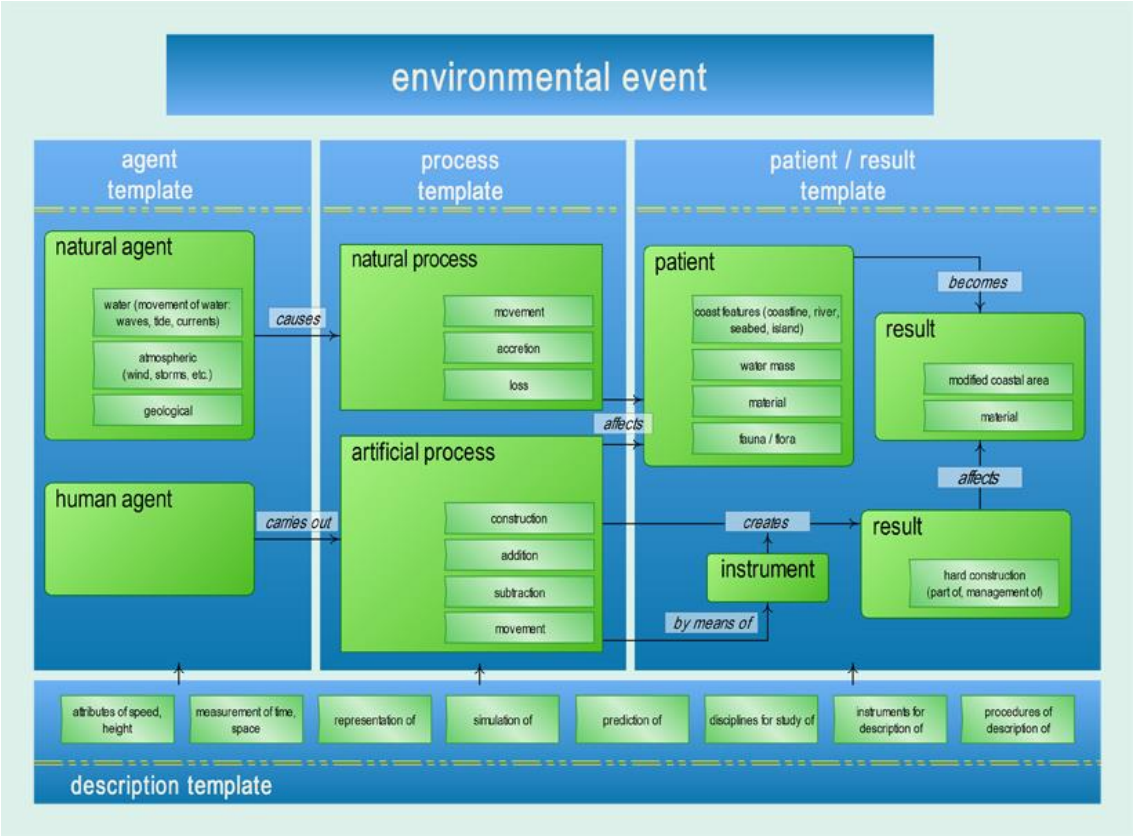

Figura 1: El EVENTO MEDIOAMBIENTAL

Como puede observarse en la Figura 1, un AGENTE (natural o humano) se encuentra en el origen de los procesos. Estos pueden ser procesos naturales, cuando los causa un AGENTE NATURAL, y pueden afectar a un PACIENTE y tener RESULTADOS. En cambio, cuando es un AGENTE ANTRÓPICO el que lleva a cabo los procesos, estos se consideran procesos artificiales y pueden realizarse utilizando algún INSTRUMENTO, además de derivar en RESULTADOS. Además, en el EVENTO MEDIOAMBIENTAL participan los atributos o propiedades de los conceptos (p. ej. SPEED, HEIGHT). Los roles semánticos empleados (AGENTE, PACIENTE, RESULTADO, etc.) se especifican con las categorías semánticas que pueden desempeñar dichos roles en el medio ambiente (p. ej. WATER, ATMOSPHERIC, GEOLOGICAL).

Sin embargo, la representación conceptual no puede basarse únicamente en la asignación de roles y categorías semánticas en un evento o marco, ya que los conceptos son dinámicos, como lo demuestra su multidimensionalidad, que puede dar lugar al énfasis de una u otra dimensión en diferentes contextos (León Araúz et al. 2012). Por tanto, las relaciones conceptuales desempeñan un papel fundamental para delimitar los conceptos según las diferentes proposiciones conceptuales de las que pueden participar y, con ellas, los diferentes contextos. Así sucede en los eventos o marcos, en los que se recogen las relaciones entre conceptos y, en esta línea, su potencial combinatorio (Faber 2009).

De este modo, en el EVENTO MEDIOAMBIENTAL intervienen relaciones jerárquicas y no jerárquicas (Faber 2015). Las relaciones jerárquicas son las más estudiadas en la Terminología y en las otras áreas que hacen uso de las relaciones semánticas, como la Inteligencia Artificial. Estas se basan en la subordinación o superordinación de los conceptos, y pueden ser genérico-específicas o hiponímicas (*type_of*, o lo que es lo mismo *is_a*, y su relación inversa, *has_type*) o bien partitivas o meronímicas (*part_of* y su inversa *has_part*). Por ejemplo, una de las relaciones jerárquicas que se aprecia en la Figura 1 es la genérico-específica: HUMAN AGENT *type_of* AGENT. En concreto, esta relación ha sido probablemente la más estudiada, debido a su importancia para la categorización y la herencia de propiedades.

Por su parte, las relaciones no jerárquicas (p. ej. NATURAL AGENT *causes* NATURAL PROCESS) se han estudiado menos, lo que, unido a la diferente granularidad de los distintos enfoques, ha desembocado en una falta de inventarios más o menos consensuados. Sin embargo, uno de los puntos fuertes de la Terminología Basada en Marcos radica en su amplia concepción y uso de las relaciones no jerárquicas (p. ej. *causes, affects, represents, has_function*, etc.), que favorecen la adquisición situada del conocimiento al reproducir los diferentes vínculos que pueden mantener los conceptos. De hecho, las relaciones no jerárquicas son fundamentales en terminología, ya que la complejidad de muchos dominios no permite su representación basada únicamente en relaciones hiponímicas y partitivas (León Araúz 2009).

En la Terminología Basada en Marcos, las relaciones, cuya capacidad para vincular dos conceptos depende de la naturaleza de los mismos (León Araúz y Faber 2012), se detectan mediante patrones de conocimiento (Hearst 1992; Meyer 2001; Bowker 2003; León Araúz et al. 2016; León Araúz y San Martín 2018; *inter alia*), que son marcadores léxico-sintácticos que se encuentran en textos reales y expresan una determinada relación entre conceptos (Meyer 2001).

Así pues, las relaciones semánticas ocupan un lugar central en la Terminología Basada en Marcos y en EcoLexicon, como veremos a continuación, ya que su visualización resulta vital para comprender cómo interactúan los conceptos en un dominio especializado. Además, su utilidad en diferentes tareas resulta innegable, pues permiten la estructuración conceptual (por ejemplo, las representaciones terminológicas o la elaboración de definiciones [Carvalho et al. 2018]), además de la comprensión de los textos (Bowker 2003). Sin embargo, no han gozado de la misma atención que otros entes conceptuales como las categorías (León Araúz 2009; León Araúz et al. 2016).

Como hemos indicado, las relaciones semánticas también son fundamentales en la elaboración de definiciones. Según la Terminología Basada en Marcos, cada categoría semántica tiene una plantilla definicional compuesta por las relaciones semánticas que activa dicha categoría (Faber et al. 2001). Estas plantillas permiten la redacción de definiciones homogéneas y la organización conceptual de los términos. Mediante su uso, los hipónimos heredan las propiedades del concepto superordinado, a las que añaden nuevas características. Así, se obtienen definiciones consistentes basadas en la estructura clásica de *genus* y *differentiae* (García de Quesada et al. 2002), cuya información se extrae del uso de corpus.

En la Terminología Basada en Marcos también se comprende el contenido visual como método para favorecer la adquisición de conocimiento, ya que la representación de los conceptos se acompaña a menudo de imágenes específicamente seleccionadas para tal fin (Reimerink et al. 2016; León Araúz y Reimerink 2016; Reimerink y León Araúz 2018). Además, en un enfoque descriptivo como este, en el que el contexto desempeña un papel crucial, no resulta extraño que campos como la variación (San Martín 2016; León Araúz 2017), la metáfora (Ureña Gómez-Moreno 2015), la fraseología (Montero Martínez 2008; Montero Martínez y Buendía Castro 2012; Buendía Castro 2013) o la traducción especializada (Montero Martínez et al. 2001; Montero Martínez y García de Quesada 2003; León Araúz y Reimerink 2018) también reciban bastante interés, contando siempre con el sistema conceptual como eje vertebrador de estas cuestiones.

En definitiva, en la Terminología Basada en Marcos, los conceptos especializados se activan como parte de un evento o marco en el que se vinculan mediante relaciones y atributos. Los marcos se conciben, por tanto, como estructuras que permiten la contextualización del conocimiento, además de la correspondencia interlingüística (Faber y Cabezas García 2019). Asimismo, la cognición situada y la combinatoria de los conceptos y los términos son muy relevantes, como se detallará en la Sección 2.2.3, y se plasman en EcoLexicon, pues los recursos de conocimiento especializado deben reflejar este comportamiento de los términos y los conceptos.

2.2.3 Aplicación práctica: EcoLexicon

El trabajo terminográfico llevado a cabo en la Terminología Basada en Marcos se aplica a la base de conocimiento terminológica EcoLexicon (http://ecolexicon.ugr.es) (Faber et al. 2011; Faber et al. 2016; San Martín et al. 2017). Se trata de un recurso desarrollado por el grupo de investigación LexiCon de la Universidad de Granada que se diseñó e implementó en 2003, y recoge la terminología del medio ambiente en español, inglés, alemán, árabe,

francés, griego moderno y ruso, siendo el inglés y el español las lenguas más presentes. Actualmente contiene 4 471 conceptos y 23 783 términos.

Entre los usuarios de EcoLexicon se cuentan especialistas de la lengua (p. ej. terminólogos, traductores), expertos del medio ambiente y el público en general. Se trata, en definitiva, de usuarios que necesiten acceder a los conceptos del medio ambiente, por ejemplo, para comprender, escribir o traducir textos especializados o semiespecializados (San Martín et al. 2017).

EcoLexicon presenta una interfaz con varios módulos que ofrecen información conceptual, lingüística y gráfica, entre los que se puede elegir para obtener los datos que resulten de mayor interés. En la Figura 2 se muestra la entrada de ANTHROPIC EROSION.

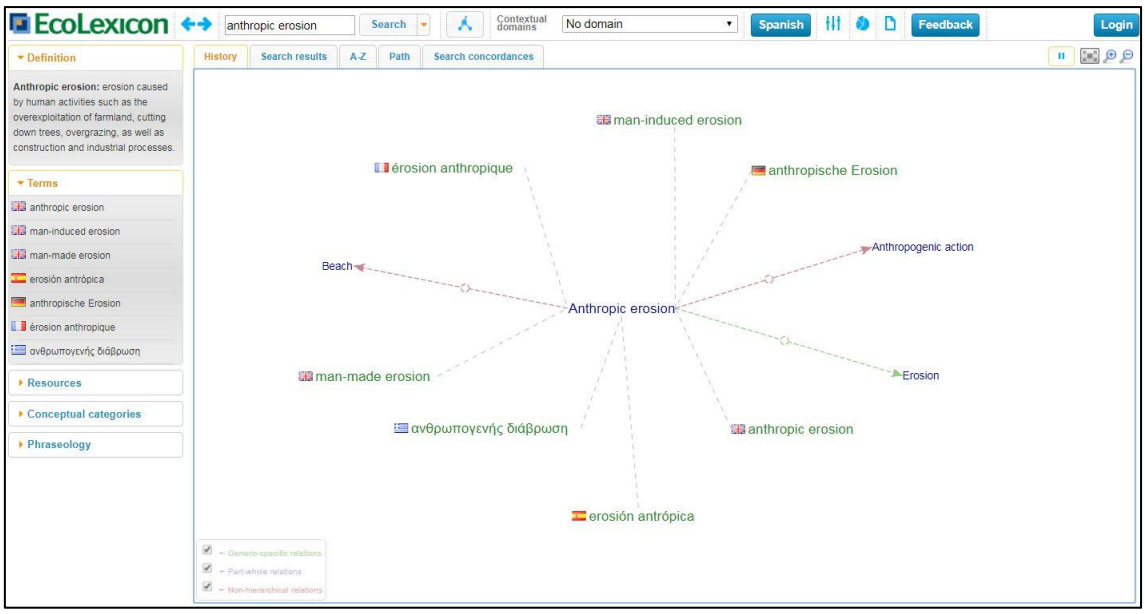

Figura 2: Entrada de ANTHROPIC EROSION en EcoLexicon

En la parte superior de la interfaz se puede acceder a la búsqueda por término o concepto. Bajo esta barra se muestran pestañas que despliegan el historial de navegación, los resultados de las consultas, una lista alfabética de los términos, el recorrido conceptual más corto entre dos conceptos y ejemplos de concordancias del término buscado. Por otro lado, a la izquierda se muestra la información del concepto, en concreto, su definición, los términos que lo designan (al hacer clic se indica la lengua, el tipo de término, ejemplos de uso, la categoría gramatical y el género), los recursos asociados (imágenes, documentos, etc.), las categorías conceptuales[2] y su fraseología[3].

[2] Este apartado está en proceso de modificación para representar de forma precisa la categoría o categorías conceptuales a las que pertenezca el concepto, ya que por el momento se presenta una mezcla de categorías y roles.

En el centro de la pantalla se muestra la red semántica en la que interviene el concepto consultado, que constituye la característica más distintiva de EcoLexicon. El concepto consultado se muestra en el centro, vinculado a otros conceptos mediante flechas que indican relaciones jerárquicas y no jerárquicas, así como atributos. Al hacer clic con el botón derecho sobre uno de los conceptos de la red, se muestra el menú contextual, que permite ajustar los resultados, por ejemplo, centrando el concepto en la red (de forma que se reorganiza la estructura conceptual) o eliminando de esta algún concepto. En la esquina superior derecha se puede modificar la visualización de la red, por ejemplo, ajustando su tamaño o pausando el movimiento, ya que la red se presenta por defecto con una leve oscilación.

Dado que la amplitud y la multidimensionalidad del dominio medioambiental dieron lugar a una sobrecarga de información en la interfaz de EcoLexicon, se tomaron diferentes medidas para paliar el exceso de información. En concreto, se permitió elegir las relaciones que querían visualizarse en la red, además de ofrecer una vista recontextualizada según restricciones contextuales al elegir un dominio determinado y mostrar diferentes puertas de acceso al sistema conceptual (a modo de red, árbol o camino).

Tras comentar los aspectos principales de la interfaz de EcoLexicon, revisaremos los productos derivados de esta base de conocimiento. En concreto, hemos señalado que EcoLexicon cuenta con un módulo fraseológico, ya que se reconoce la importancia de la fraseología en el lenguaje especializado. Este incluye actualmente colocaciones verbales, aunque en esta tesis presentamos su ampliación para representar también CN (Sección 5.3). Así, respecto a las colocaciones verbales, se presentan los verbos que suelen coocurrir con un determinado concepto, ordenados en dominios léxicos (Faber y Mairal 1999) según su significado. Una vez que esté implementada la sección de CN, será posible la conexión de ambas partes del módulo, ya que muchas colocaciones verbales tienen un equivalente en forma de CN (p. ej. *the hurricane dissipates* > *hurricane dissipation*).

Otro de los productos derivados de EcoLexicon es el *EcoLexicon English Corpus*, un corpus de 23,1 millones de palabras compuesto por textos actuales del medio ambiente en inglés. Este corpus se compiló en el grupo de investigación LexiCon para la alimentación de EcoLexicon y puede accederse a él en la opción *Search concordances* de EcoLexicon, además de estar disponible en acceso abierto en los Open Corpora de Sketch Engine. Sus

[3] El apartado de fraseología también se encuentra en construcción. Actualmente despliega los verbos que suelen coocurrir con el concepto, si bien se enriquecerá esta información y se aportará, además, un apartado de CN (Sección 5.3).

metadatos permiten realizar búsquedas basadas en factores pragmáticos, como los dominios contextuales, o el tipo de usuario o de texto. Cabe destacar que está anotado con las gramáticas de *sketches* desarrolladas por León Araúz et al. (2016), que están basadas en patrones de conocimiento y permiten extraer las principales relaciones semánticas del dominio medioambiental (genérico-específicas, parte-todo, ubicación, causa y función) (San Martín et al. 2017).

Asimismo, se ha desarrollado una herramienta web de traducción asistida destinada a facilitar la traducción de textos medioambientales en inglés y español mediante la incorporación de recursos de conocimiento y herramientas de análisis de corpus. Se trata de EcoLexiCAT (León Araúz et al. 2017; León Araúz y Reimerink 2018), que parte del software libre de traducción asistida MateCat para enriquecer el texto origen con información de EcoLexicon, BabelNet (un diccionario enciclopédico y red semántica multilingüe que se construye de forma automática) (Navigli y Ponzetto 2012) y la herramienta de análisis de corpus Sketch Engine (Kilgarriff et al. 2004, 2014).

En definitiva, como se desprende de sus diferentes funciones y los distintos productos a los que ha dado lugar, EcoLexicon constituye una herramienta terminográfica muy rica que cuenta con una base teórica bien asentada. Así, da cuenta de fenómenos relevantes en el lenguaje especializado, como son la variación, la multidimensionalidad o la fraseología. Además, sigue alimentándose con dominios adicionales del medio ambiente, así como con nuevos conceptos y términos, y funcionalidades novedosas, por lo que se trata de un recurso que, sin duda, seguirá creciendo.

3. Los términos compuestos: el caso de los compuestos nominales

Los compuestos nominales (CN), p. ej. *hydroelectric power station*, destacan por su elevada presencia en diferentes lenguas, como el inglés y el español (Nakov 2013; Hendrickx et al. 2013; Fernández Domínguez 2016). Este protagonismo del que hace gala la composición se debe a las infinitas posibilidades de combinación que ofrece la lengua para facilitar la expansión léxica y terminológica (Baldwin y Kim 2010). Así, los CN suelen consistir en la especificación de un núcleo nominal con modificadores nominales, adjetivales o preposicionales, que varían según los patrones de formación de términos de cada lengua.

Además de ser frecuentes en la lengua general, los CN destacan por su omnipresencia en el discurso especializado (Sager et al. 1980; Cabré 1993; Kageura 2002; Daille et al. 2004; Heid 2008; Faber et al. 2012; Nakov y Hearst 2013; *inter alia*), como etiquetas de los conceptos, que contribuyen a la transmisión del conocimiento científico-técnico. De tal modo, se trata de términos compuestos, los cuales entendemos como unidades fraseológicas especializadas, a pesar de la diversidad de opiniones al respecto, como comentaremos en este capítulo. Dado que los CN constituyen el principal tipo de término compuesto, es frecuente el empleo de esta denominación genérica para aludir a ellos, si bien existen otros términos compuestos, como los adjetivos compuestos (*decay-corrected*) o los verbos compuestos (*precast*) (Sager et al. 1980).

Los CN plantean problemas a diferentes niveles, como su identificación, su desambiguación estructural, el análisis semántico de sus formantes y la relación implícita que los vincula, su traducción o su representación en recursos terminográficos. Por ello, no resulta extraña la atención que han recibido desde distintas áreas, entre las que destacan: (i) la Lingüística General, que los ha estudiado como mecanismo habitual de formación léxica; (ii) la Terminología, que parte de su relevancia en los sistemas conceptuales; (iii) la Lexicografía, que ha investigado los desafíos de su representación; (iv) la Lingüística Computacional, cuyo tratamiento concibe como indispensable para el buen funcionamiento de los sistemas de procesamiento del lenguaje natural; (v) la Psicolingüística, que ha analizado la combinación conceptual que subyace en estas unidades.

Además, la investigación en CN sigue siendo un tema de actualidad, como lo demuestran los numerosos congresos y publicaciones que intentan aclarar las diferentes particularidades de este fenómeno. No obstante, aún quedan aspectos en los que profundizar, como los CN que presentan un alto grado de especialización, su formación

por cualquier número y tipo de elementos, su traducción o formación en lenguas diferentes del inglés, así como su representación en recursos terminográficos de forma que se facilite la transmisión de conocimiento, además de otras vías que sin duda seguirán surgiendo.

Este capítulo recoge los diferentes aspectos de interés en relación con este tipo de términos compuestos que constituyen los CN. En la Sección 3.1 comentamos los principales rasgos de la fraseología general y la fraseología especializada, a la que pertenecen los CN en terminología. En la Sección 3.2 desarrollamos nuestra noción de CN y presentamos la literatura principal que ha estudiado estas unidades. A continuación, en la Sección 3.3 describimos los tipos de CN y su formación. Seguidamente, en la Sección 3.4 profundizamos en el análisis de estos términos compuestos, deteniéndonos en la desambiguación estructural, la semántica de los formantes, las relaciones semánticas internas, y su frecuente variación y neología. En la Sección 3.5 abordamos la traducción de los CN, en la que pasamos revista a los problemas que plantea su transferencia y los estudios que se han realizado al respecto. Por último, en la Sección 3.6 nos detenemos en la representación de estos términos compuestos en los recursos lingüísticos en general y los de carácter especializado en particular.

3.1 Los compuestos nominales en la fraseología

La fraseología constituye un área de estudio que ha recibido un gran interés. Son muchos los trabajos que examinan estas combinaciones de palabras, principalmente en la lengua general (Vinogradov 1947; Firth 1957; Sinclair 1991; Mel'čuk et al. 1995; *inter alia*), pero también en el discurso especializado (Picht 1991; Martin 1992; Tercedor Sánchez 1999; Montero Martínez 2002; Buendía Castro 2013; *inter alia*). Como suele ocurrir, existen diferentes posturas y denominaciones respecto a la fraseología y las diferentes unidades fraseológicas. Así, en el apartado 3.1.1 comenzaremos abordando la noción de fraseología y las distintas posturas existentes al respecto, para después pasar revista a las diferentes unidades fraseológicas de la lengua general. A continuación, el apartado 3.1.2. ahonda en estas combinaciones en el lenguaje especializado, donde nos detendremos en el papel de los términos compuestos como unidades fraseológicas especializadas.

3.1.1 La fraseología y las unidades fraseológicas

La fraseología ha recibido la atención de diferentes campos de estudio desde hace décadas (Ramisch 2015; Ruiz Yepes 2017; Parra Escartín et al. 2018). Desde la Lingüística Teórica (Bally 1951[1909]; Vinogradov 1947) hasta la Psicología (Rapp 2008), pasando por la Lingüística Computacional (Smadja 1993; Sag et al. 2002) y la Lingüística de Corpus (Firth 1957; Halliday 1966; Sinclair 1991; Stubbs 2001), se han interesado por estas

combinaciones de palabras. No resulta sorprendente si consideramos que estas unidades son una parte fundamental de nuestro lenguaje, que podrían representar la mitad de nuestro vocabulario (Jackendoff 1997).

A principios del siglo XX, Bally (1951[1909]) propuso la primera clasificación de las unidades fraseológicas, siendo así pionero a la hora de abordar de manera clara estas expresiones fijas. Le siguió toda una escuela de fraseología en el bloque soviético, en la que cabe destacar la figura de Vinogradov (1947), quien estableció formalmente la Fraseología como disciplina (Cowie 1981). A partir de entonces, creció el interés en las diferentes unidades fraseológicas, especialmente a partir de la década de los 70 (Pawley 2007). Prueba de ello es la publicación de numerosas investigaciones y diccionarios centrados exclusivamente en las unidades fraseológicas (Mel'čuk et al. 1984-1999; Sinclair y Moon 1989).

La fraseología se asienta sobre la idea de que el lenguaje no se compone únicamente de palabras individuales, sino también de estructuras superiores que funcionan como un todo (Sinclair 1991; Martínez y Schmitt 2012). En este sentido, la Lingüística de Corpus ha desempeñado un papel fundamental, pues, por una parte, ha demostrado la relevancia de la fraseología en el lenguaje y, por otra, ha proporcionado las herramientas necesarias para extraer y analizar estas unidades (Granger y Meunier 2008: xix). Así, Sinclair (1991) defiende la existencia de dos principios que rigen la producción lingüística: el *open-choice principle* y el *idiom principle*. El *open-choice principle* explica la diversidad de combinaciones posibles gramaticalmente para expresar un mensaje, mientras que el *idiom principle* restringe estas posibilidades a combinaciones prefabricadas que no admiten tanta variación como las anteriores:

> A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments (Sinclair 1991: 110).

En esta línea, Sinclair (1991) sostiene que el significado no reside únicamente en el lexema, sino que se encuentra también en lo que denomina *extended units of meaning*, es decir, unidades que traspasan la barrera de la palabra. Antes, Coseriu (1981) hablaba de *discurso repetido* para referirse a las expresiones fijas y autores como Pawley y Syder (1983) y Heid (1992) coincidían en que, de las distintas formas de expresar un mensaje, solo algunas serían usadas por nativos, lo que entronca con la idea de estructuras prefabricadas sobre la que se asienta la fraseología. Desde la Gramática de Construcciones (Fillmore et al. 1988; Goldberg 1995) se defiende que gran parte de nuestra competencia lingüística reside en las *construcciones*, que son estructuras que aúnan información

sintáctica, léxica, semántica y pragmática; en otras palabras, son combinaciones de forma y significado. Se rechaza, pues, una división estricta entre sintaxis, semántica y pragmática. La noción de *lexical priming* de Hoey (2005) también es relevante, al defender que nuestro conocimiento de las palabras incluye su aparición con determinadas unidades en determinados contextos.

Como vemos, todos estos autores coinciden en contemplar la fraseología como la tendencia de las palabras a combinarse unas con otras. Estas combinaciones han recibido diferentes nombres a raíz de los distintos autores y disciplinas que las han abordado. En la Fraseología (Vinogradov 1947) se prefiere la denominación de *unidad fraseológica*, un término genérico que engloba colocaciones, locuciones, proverbios, etc. (Corpas Pastor 2013). Algunos de estos tipos de unidades fraseológicas, como las locuciones (*idioms*) o las colocaciones, se han usado indistintamente para referirse al conjunto de la clase (Ramisch 2015), aunque, como veremos a continuación, hacen referencia a diferentes conceptos. En la Lingüística Computacional, el término *colocación* (*collocation*) fue sustituido por *multiword expression* (o *multiword unit*) desde principios del 2000, tras el trabajo de Sag et al. (2002). Esta es ahora la denominación más extendida en esta disciplina (Heid 2008; Ramisch 2015). Mel'čuk et al. (1995) proponen asimismo el término *frasema*, que también goza de una gran aceptación. En este trabajo hablamos de *unidades fraseológicas*, por ser esta una denominación asentada y comprensible para las distintas disciplinas que estudian dichas unidades, que además se adapta fácilmente a diferentes lenguas. No obstante, el término *multiword expression* resulta también útil cuando abordamos los términos compuestos, pues estos suelen recibir la denominación de *multiword terms* en la literatura.

Así pues, existen diferentes tipos de combinaciones de palabras, que han recibido distintos nombres y cuya consideración como unidad fraseológica depende de la restricción del enfoque adoptado. Según la teoría de centro-periferia de la Escuela de Praga, las corrientes más estrictas que solo consideran unidades fraseológicas las expresiones fijas, como las locuciones (Zuluaga 1975; Rossenbeck 1989; García Page 2001), estarían en el centro. A medida que se va ampliando el círculo, se incluyen elementos periféricos en la lista de unidades fraseológicas. Así, existen otros muchos estudios que también consideran las colocaciones como unidades fraseológicas (Bally 1951[1909]; Gläser 1988; Hausmann 1989; *inter alia*). Finalmente, la concepción más amplia sostiene que todas las combinaciones frecuentes de palabras, incluidos los compuestos, son unidades fraseológicas (Benson et al. 1986; Pawley 2001; Ramisch 2015; Parra Escartín et al. 2018). Este trabajo se adscribe a esa concepción amplia, considerando

unidades fraseológicas las locuciones, las colocaciones y los compuestos. Excluimos, pues, las combinaciones libres, ya que nos parece necesaria la existencia de un cierto grado de fijación para considerarse unidades fraseológicas. No obstante, como veremos a continuación, la frontera entre unos tipos y otros dista de ser clara. Más bien cabría hablar de un continuo fraseológico (Fontenelle 1998) que, de mayor a menor fijación, comenzaría en un extremo por las locuciones, pasaría por las colocaciones y los compuestos en el centro y terminaría por las combinaciones libres (Ruiz Yepes 2017; Men 2018). Esta idea ya la proponía Bally (1951[1909]: 68), quien sostenía que existen extremos de máxima y mínima fijación, entre los cuales se encuentran las *series fraseológicas* o *grupos usuales*, que se corresponderían con las colocaciones.

Por lo tanto, en uno de los extremos se encuentran las locuciones o *idioms*, que son combinaciones cuyo significado no se puede deducir del significado de sus partes. Se trata de estructuras rígidas que presentan fijación semántica y sintáctica, es decir, no admiten sustitución de sus elementos ni variación sintáctica (p. ej. uso de pasiva o modificadores) (Ruiz Yepes 2017). Según la función que desempeñen, pueden ser de tipo nominal (*mosquita muerta*), adjetival (*de rompe y rasga*), adverbial (*a raudales*), verbal (*meterse en camisa de once varas*), conjuntivo (*como si*) o clausal (*salirle a alguien el tiro por la culata*) (Corpas Pastor 1996: 270). Autores como Vinogradov (1947), Cowie (1981), Benson et al. (1986) y Mel'čuk et al. (1995) proponen una clasificación más fina de las locuciones. Por un lado, distinguen aquellas combinaciones cuyo sentido no puede deducirse del significado de las partes, como *beat around the bush*. Estas son *phraseological fusions* para Vinogradov (1947), *pure idioms* para Cowie (1981), *idioms* para Benson et al. (1986) y *complete phrasemes* para Mel'čuk et al. (1995). Por otro lado, diferencian las combinaciones cuya idiomaticidad se deriva del sentido figurado de sus partes, como en *steal someone's heart*. Estas son *phraseological unities* para Vinogradov (1947), *figurative idioms* para Cowie (1981), *transitional combinations* para Benson et al. (1986) y *quasi-phrasemes* para Mel'čuk et al. (1995).

El otro extremo del continuo lo ocupan las combinaciones libres, que se suelen estudiar en contraposición a las unidades fraseológicas. Se trata de toda combinación posible siguiendo las reglas de la gramática (Cowie 1981). Son, pues, estructuras de significado transparente a las que no se aplican restricciones combinatorias. Debido a esta libertad de combinación, los formantes de la combinación libre coocurren con una frecuencia menor que la de las otras combinaciones estudiadas. Ejemplos de combinaciones libres pueden ser *coche negro* o *comprar una manzana*.

Muchos autores coinciden en hablar de las colocaciones como unidades a medio camino entre las locuciones y las combinaciones libres (Cowie 1981; Benson et al. 1986; Buendía Castro 2013; Men 2018). Se trata posiblemente de las combinaciones que más interés han suscitado en la literatura, dando lugar a diferentes corrientes fraseológicas, como veremos a continuación. En este trabajo reciben especial atención, ya que algunas de sus estructuras coinciden con las de los compuestos nominales, lo que ha generado debate en torno a la necesidad de diferenciar o no ambas formas.

Ya en los años 30 del siglo XX, el término *collocation* fue empleado en el ámbito pedagógico por Palmer para referirse al conjunto de las unidades fraseológicas. Palmer (1933: 5) entendía las colocaciones como combinaciones de dos o más palabras que deben aprenderse como un todo. Sin embargo, no especificaba las características que hacen que una combinación sea una colocación (Men 2018). No es hasta la década de los 50 que el término *collocation* aparece como tal, en *Papers in Linguistics, 1934-1951* (Firth 1957). A través de su célebre cita «you shall know a word by the company it keeps» (Firth 1957: 11), Firth abordó el tema desde una perspectiva lingüística y sentó las bases de un enfoque colocacional que tomaría la frecuencia de coaparición como elemento clave para discernir estas unidades, dejando de lado cuestiones de índole semántica o sintáctica. El uso de corpus resulta, de este modo, fundamental. En su estudio colocacional, más que en la distinción de las diferentes unidades fraseológicas, Firth se centra en la colocación como fenómeno lingüístico (Men 2018). Es por ello que utiliza el término *colocación* para referirse a combinaciones de palabras en un sentido más amplio (Nesselhauf 2004). Entre los autores que siguieron las ideas de Firth, cabe destacar la figura de Sinclair (1991), además de otros trabajos importantes como los de Halliday (1966), Moon (1998), Stubbs (2001) o Hoey (2005).

Por otra parte, se desarrolló otro enfoque colocacional que estudiaba estas unidades desde una perspectiva semántica. Así, autores como Mel'čuk et al. (1984–1999), Benson et al. (1986) o Hausmann (1989) sostienen que el fenómeno colocacional no se limita exclusivamente a la frecuencia de coaparición de sus componentes, sino que intervienen, además, cuestiones semánticas. En concreto, consideran que las colocaciones están formadas por una base autónoma semánticamente y un colocado que depende de ella. Esta idea de base y colocado también aparecía en los estudios probabilísticos comentados arriba, con la diferencia de que, en aquellos, se denomina *nodo* a la base y se considera como colocado cualquier elemento que aparezca en el *span* o ventana colocacional establecida (Sinclair 1991). En este trabajo pensamos que tanto la frecuencia

como la semántica por separado no son suficientes para identificar las colocaciones, por lo que consideramos que ambos enfoques deben usarse de manera complementaria.

En esta perspectiva semántica de las colocaciones merece especial atención la figura de Mel'čuk, quien, junto a Zholkovskij, desarrolló una de las teorías combinatorias de mayor relevancia y de gran repercusión en los diccionarios combinatorios (Montero Martínez 2002; Buendía Castro 2013): la Teoría Sentido-Texto (Mel'čuk y Zholkovskij 1970). Esta teoría se asienta sobre la idea de que toda comunicación lingüística se compone de (i) sentido, (ii) una señal escrita u oral (texto) y (iii) un vínculo entre las dos (Faber y Mairal 1999: 11). Se concede, así, un papel central a la semántica. El elemento clave de la Teoría Sentido-Texto son las funciones léxicas, que utilizan para expresar relaciones semánticas como la sinonimia, la antonimia o la hiperonimia; y relacionar los elementos de las colocaciones o los derivados sintácticos y semánticos. Por ejemplo, para disminuir *rain* podemos usar la función léxica combinada AntiMag, lo que da lugar a la colocación *light rain* (Ramisch 2015: 25-26). Sin embargo, como apunta Montero Martínez (2002), la generalidad de las funciones léxicas hace que en ocasiones se pierdan matices semánticos. No obstante, la Teoría Sentido-Texto constituye un marco muy completo y bien sustentado que se materializa en el *Dictionnaire explicatif et combinatorie du français contemporain* (Mel'čuk et al. 1984-1999) y en otros diccionarios como *DiCo*, *DiCoInfo* o *DiCoEnviro* (veáse la Sección 3.6.2.6), elaborados en la Universidad de Montreal, donde Mel'čuk continuó con sus ideas.

Además de haberse abordado desde distintos enfoques, las colocaciones también han recibido diversas denominaciones y se han tomado a menudo como término genérico para referirse a todas las unidades fraseológicas (p. ej. Palmer [1933], Firth [1957]). Así, son *phraseological combinations* para Vinogradov (1947), *restricted collocations* para Cowie (1981), *fixed combinations* o *recurrent combinations* para Benson et al. (1986) y *semi-phrasemes* para Mel'čuk et al. (1995), por nombrar solo algunos ejemplos.

Sea cual sea la denominación empleada, la mayoría de los autores coinciden en que las colocaciones son combinaciones frecuentes de dos[4] o más palabras cuyo significado es transparente y se puede deducir del significado de sus partes, es decir, son composicionales (Hausmann 1989; Lorente Casafont et al. 2017; Ruiz Yepes 2017; Men

[4] A pesar de que las colocaciones suelen estar formadas por dos palabras, también existen colocaciones más extensas, a menudo resultado de la superposición de dos colocaciones (p. ej. *prestar un servicio* y *servicio militar* > *prestar servicio militar*).

2018)[5]. Constituyen una unidad en el lexicón mental (Ruiz Yepes 2017), admiten variación formal (*prestar un servicio*, *servicios prestados*) y están formadas por una base, que conserva su significado habitual, y un colocado, que lo adquiere en función de ella (Vinogradov 1947; Cowie 1981; Mel'čuk et al. 1995; Moon 1998; Tutin y Grossman 2002). Por ejemplo, en colocaciones formadas por un verbo y un sustantivo, como *prestar servicio*, el sustantivo (*servicio*) es la base y el verbo (*prestar*) es el colocado, que adquiere su significado al combinarse con la base. Esta idea está relacionada con la selección léxica de la que hablan Mel'čuk et al. (1995), quienes sostienen que, al producir una combinación de palabras, la elección de una de ellas no es libre, sino que viene impuesta por la otra.

De ello se desprende que, en las colocaciones, uno o ambos formantes solo puede ser ocupado por determinadas palabras (Hausmann 1989), si bien no presentan tanta fijación como las locuciones. En este sentido, existe una concepción estricta, que defiende que ninguno de los elementos de una colocación se puede sustituir por un sinónimo (Pearce 2002; Ruiz Yepes 2017), y otra más amplia, que considera la permutabilidad de alguno de los elementos (Cowie 1981; Corpas Pastor 2001). A nuestro parecer, no conviene generalizar con respecto a la sustitución de los elementos de las colocaciones, ya que ello dependerá de la colocación ante la que nos encontremos. Así, por ejemplo, *pegar una paliza* puede sustituirse por *dar/meter/propinar/arrear una paliza*, mientras que *levar el ancla* o *izar la bandera* no admiten tal modificación (Corpas Pastor 2001: 92-93).

Esta variación se da en ciertos casos porque los rasgos definitorios de una de las dos formas permiten su combinación con otra forma de determinados rasgos semánticos (Buendía Castro y Faber 2015). Hablamos, así, de una relación de tipicidad en las colocaciones (Corpas Pastor 2001). Por ejemplo, *paliza* se define como 'serie de golpes dados con un palo o con cualquier otro medio o instrumento', de manera que los verbos *pegar*, *dar*, *meter*, *propinar* y *arrear* encajan en ese significado al representar la acción llevada a cabo. Como vemos, algunos de estos verbos transmiten ese significado intrínsecamente, ya que hacen referencia a los golpes en su definición (*pegar*, *propinar*, *arrear*), mientras que otros (normalmente, verbos soporte) adquieren ese sentido figurado al combinarse con el sustantivo (*dar*, *meter*). En cambio, estructuras como *dar un libro* no serían colocaciones, sino combinaciones libres, ya que no presentan una relación de tipicidad, es decir, la definición de *dar* implica entregar algo, que bien podría ser un libro, pero la definición de *libro* no implica que se trate de algo que tenga que entregarse.

[5] Las colocaciones también se entienden por algunos autores (Catford 1965; Aitchison 2003) como asociaciones mentales de palabras que no tienen por qué aparecer juntas en el discurso. Por ejemplo, hablaríamos de colocación al decir que *salt* evoca las palabras *water, sea, shaker* y *lake*, entre otras (Aitchison 2003: 86).

Así pues, podemos hablar de las colocaciones como combinaciones frecuentes de dos o más palabras, de significado transparente y semifijas. Están formadas por una base y un colocado, seleccionado por esta, entre los que existe una relación de tipicidad. Admiten variación formal y constituyen una unidad en el lexicón mental. La clasificación más extendida es la de Hausmann (1989: 1010), quien propone los siguientes tipos de colocaciones según su estructura sintáctica, válidos para las lenguas germánicas, eslavas y romances (Montero Martínez 2002):

- sustantivo + adjetivo (epíteto), p. ej. *error garrafal*[6].

- sustantivo (sujeto) + verbo, p. ej. *desatarse una tormenta*.

- verbo + sustantivo (objeto), p. ej. *hacer balance*.

- verbo + adverbio, p. ej. *negar rotundamente*.

- adjetivo + adverbio, p. ej. *políticamente correcto*.

- sustantivo + (prep.) + sustantivo, p. ej. *racimo de uvas*.

Como decíamos anteriormente, las concepciones más amplias de la fraseología también incluyen los compuestos en la lista de unidades fraseológicas (Benson et al. 1986; Pawley 2001; Ramisch 2015; Parra Escartín et al. 2018), si bien estos han recibido menos atención por parte de los estudios fraseológicos que otras unidades, como las colocaciones. Los compuestos son frecuentes en la lengua general, aunque destaca su presencia en el discurso especializado, por lo que les dedicaremos especial atención en la Sección 3.1.2. Estos se caracterizan por la unión de dos o más lexemas que constituyen una unidad de significado. Estos lexemas se pueden unir y formar una palabra, dando lugar a compuestos gráficos, como *sacacorchos*, o bien mantenerse separados, siendo entonces compuestos sintagmáticos, como *pez espada* (Castillo Carballo 2006). Aunque muchos autores han considerado compuestos únicamente a aquellos que presentan unión ortográfica (Kocourek 1982; Corpas Pastor 1996; Štekauer et al. 2012), otros muchos también aplican esta denominación a los que cuentan con un espacio entre los lexemas. Así, no es de extrañar que estas combinaciones hayan recibido diversos nombres fruto de los diferentes enfoques y estudios que se han ocupado de ellas: *compuestos* o *compuestos gráficos* para los que presentan unión ortográfica, frente a *locuciones* o *compuestos sintagmáticos* para los que se escriben con espacios; o bien *compuestos*, *compuestos nominales*, *compounds*, *nominal compounds*, *noun compounds* o *complex nominals* para

[6] Los ejemplos son nuestros.

designar al conjunto de la clase sin atender a la unión o separación gráfica. En este trabajo consideramos compuestos tanto a aquellos que presentan unión ortográfica como a los que no, si bien nos interesamos por estos últimos.

Así, los compuestos sintagmáticos cuentan con un núcleo nominal, por lo que utilizamos preferentemente la denominación de *compuestos nominales* (CN). Este núcleo está modificado por otras unidades[7], como sustantivos, adjetivos o sintagmas preposicionales. Los CN coinciden en forma con otras unidades fraseológicas, como algunos tipos de locuciones y colocaciones; en concreto, las de estructura sustantivo + adjetivo y sustantivo + (prep.) + sustantivo, como vemos en la clasificación de Hausmann (1989: 1010). Es por ello que la distinción entre unos y otros no resulta siempre evidente (Castillo Carballo 2006; Montero Martínez y Buendía Castro 2012; Buendía Castro 2013) o siquiera necesaria (Meyer y Mackintosh 1994, 1996; Corpas Pastor 1996; Montero Martínez 2002), como veremos en la Sección 3.1.2.

Para comprender las diferencias o similitudes de los compuestos con las locuciones y colocaciones nominales, es necesario partir de la diversidad de opiniones existentes respecto a la inclusión de los compuestos entre las unidades fraseológicas[8]. Por una parte, las posturas reacias a considerarlos unidades fraseológicas sostienen, principalmente, que se trata de unidades de la morfología léxica que, por lo tanto, forman parte de la gramática y no de la fraseología (García Page 2001), y destacan además su formación regular y libre (Zuluaga 1975). Por otra parte, quienes consideran que los compuestos son unidades fraseológicas (Benson et al. 1986; Pawley 2001; Ramisch 2015; Parra Escartín et al. 2018), alegan que poseen los rasgos definitorios de estas unidades. En concreto, se trata de combinaciones frecuentes de dos o más lexemas, cuyos elementos presentan una cierta fijación, menor que la que se da en las locuciones y mayor que la de las combinaciones libres. Así, hay compuestos que admiten variación, como *potencia eólica* (también llamada *potencia del viento*) y otros de estructura más estática, como *parque eólico* (**parque de viento*). Deben tratarse como una unidad y presentan una transparencia variable: desde compuestos transparentes como *reloj despertador* hasta otros más opacos como *hombre rana* (Osorio Olave y Serra Sepúlveda 2012), por lo que se encuentran a medio camino entre la opacidad de las locuciones y la transparencia de las combinaciones libres. En este trabajo compartimos estas últimas opiniones y consideramos, así, que los

[7] En los compuestos gráficos pueden intervenir también otras unidades, como verbos (p. ej. *sacacorchos*, *cortaúñas* o *limpiaparabrisas*).
[8] En la Sección 3.1.2 retomaremos el debate sobre la inclusión de los compuestos entre las unidades fraseológicas y la distinción entre colocaciones y compuestos.

compuestos son unidades fraseológicas. Estos se diferencian de las locuciones por su menor fijación e idiomaticidad (y, por lo tanto, mayor transparencia)[9].

A la luz de estas características de los compuestos, podríamos decir que se encuentran en el centro del continuo fraseológico junto con las colocaciones. Los intentos de diferenciar los compuestos de las colocaciones nominales son numerosos y subrayan principalmente la función denominativa de los compuestos (como las locuciones) frente al carácter predicativo de las colocaciones (Corpas Pastor 1996: 89; Ferrando Aramo 2002: 103). En otras palabras, la función denominativa de las colocaciones se encuentra en el núcleo, que resulta modificado por el otro elemento, mientras que en los compuestos son necesarios todos los elementos para acceder al concepto designado. A nuestro parecer, la distinción entre compuestos y colocaciones nominales no siempre es clara y el consenso respecto a esta diferenciación está lejos de alcanzarse, por lo que optamos por no diferenciar las colocaciones nominales de los compuestos y utilizamos la denominación *compuesto nominal*, ya que esta refleja su estructura, un aspecto indudable de ambas combinaciones.

Tras haber abordado las diferentes unidades fraseológicas (locuciones, colocaciones y compuestos), podemos concluir que, en su sentido más genérico, se trata de unidades léxicas formadas por dos o más palabras que coocurren frecuentemente en el discurso y presentan un grado variable de fijación, una estabilidad sintáctica y semántica, y un posible carácter idiomático (adaptado de Gläser [1988] y Montero Martínez y Buendía Castro [2017])[10].

3.1.2 Las unidades fraseológicas especializadas

Las unidades fraseológicas no solo figuran en la lengua general, sino que también forman parte del discurso especializado. A pesar de la importancia que ha cobrado la fraseología especializada en la literatura (Picht 1991; Martin 1992; Pavel 1993; Tercedor Sánchez 1999; Montero Martínez 2002; Lorente Casafont 2002; Bevilacqua 2004; Silva et al. 2004;

[9] Hablamos en todo momento de compuestos endocéntricos. No obstante, existe otro tipo de compuestos, los exocéntricos (véase la Sección 3.3), que se asemejan más a las locuciones, debido a su idiomaticidad y falta de transparencia, pues se trata de compuestos que han perdido el núcleo (Nakov 2013), como *ladybird* (Nakov y Hearst 2013) o *redskin* (Štekauer 1998). En esta línea, Nida (1966[1959]: 28) relaciona las locuciones con la noción de exocéntrico, al llamar *semantically exocentric expressions* a unidades como *to heap coals of fire on one's head*. Kocourek (1982: 148) también liga la imprevisibilidad de un sintagma (típica de las locuciones), entre otros factores, con su carácter exocéntrico.

[10] Las definiciones y clasificaciones de unidades fraseológicas propuestas desde la Lingüística Computacional (Ramisch 2015; Parra Escartín et al. 2018; *inter alia*) varían en cierta medida de la noción puramente lingüística que presentamos este trabajo, ya que intentan reflejar los problemas o aspectos que deben considerarse en el tratamiento computacional de estas combinaciones, principalmente centrados en su idiomaticidad y carácter unitario.

Aguado de Cea 2007; Buendía Castro 2013; *inter alia*), los estudios centrados en estas combinaciones especializadas son mucho menos numerosos que los dedicados a la fraseología de la lengua general (Ruiz Yepes 2017). Además, la representación de unidades fraseológicas en recursos especializados se ha abordado más desde la lexicografía computacional que desde la terminología (Tercedor Sánchez 1999), lo que ha dado lugar a simples recopilaciones de datos, términos y cifras sin estructuración conceptual subyacente (Montero Martínez 2002). Del mismo modo, solo recientemente se están concibiendo las unidades fraseológicas como merecedoras de espacios propios en los recursos terminográficos y no como meros elementos contextuales (Nkwenti-Azeh 2001; L'Homme 2012; San Martín et al. 2017). Y es que estas desempeñan un papel esencial en el lenguaje de especialidad: aportan precisión semántica a los términos, ocupan un lugar fundamental en los sistemas conceptuales y permiten la expresión según las convenciones de los diferentes discursos especializados, ya sea en lengua original o en la traducción. Igual que ocurriera con la fraseología general, los corpus especializados han permitido conocer en profundidad estas coocurrencias y explorar los modos de expresión de los expertos.

El interés en estas combinaciones especializadas surgió por cuestiones prácticas, como las dificultades que encontraban traductores y redactores técnicos al tratar con términos combinados con otras unidades o ante la necesidad de automatizar la extracción terminológica con fines lexicográficos (Aguado de Cea 2007: 54-55). Así, en los años treinta, el ingeniero y lexicógrafo especializado Schlomann comenzó a introducir unidades fraseológicas en sus diccionarios técnicos (Arntz y Picht 1995) y afirmó más tarde que la fraseología es el elemento clave para la comprensión en el ámbito técnico (Schlomann 1968: IX). Como ocurriera con la fraseología y las unidades fraseológicas de la lengua general, estas combinaciones especializadas también han recibido diversos nombres, como *specialized phraseological units*, *multi-word terminological phrases*, *terminological phrasemes* o *specialized lexical combinations* (Aguado de Cea 2007).

Pese a las distintas denominaciones y concepciones con las que se han abordado, parece haber acuerdo en que el sistema de unidades fraseológicas especializadas es más restringido que el de la lengua general, en el sentido de que presenta una menor variedad de unidades fraseológicas. En línea con Gläser (1994), consideramos que la fraseología especializada no constituye un subsistema lingüístico independiente, sino que forma parte de la fraseología general, pues a pesar de las diferencias existentes, comparte muchas de sus características. De este modo, la fraseología especializada se ha definido como el conjunto de unidades fraseológicas de contenido especializado y uso frecuente en un

ámbito de especialidad, que contienen al menos un término y presentan un cierto grado de fijación (Lorente Casafont 2002: 178; Bevilacqua 2004: 28; Aguado de Cea 2007: 54). Por lo tanto, comparten con las unidades fraseológicas generales su carácter pluriléxico y su elevada frecuencia de aparición, así como un cierto grado de fijación, estabilidad y opacidad. No obstante, es cierto que esta mayor o menor opacidad de las unidades fraseológicas especializadas se debe a su contenido especializado, ya que no suelen presentar idiomaticidad en el sentido de la fraseología general (Gläser 1994). Asimismo, las unidades fraseológicas especializadas están sujetas a una mayor variación, por ejemplo, mediante su adaptación a diferentes categorías gramaticales, como en el caso de *power generation*, *generated power* y *to generate power*. Esto se debe a que los conceptos que se combinan en las unidades fraseológicas especializadas permiten diferentes posibilidades formales siempre que se mantenga el patrón combinatorio (en este ejemplo, *power* + *generate*).

La relevancia de la combinación de conceptos en las unidades fraseológicas especializadas (Arntz y Picht 1995) llevó a autores como Martin (1992) a distinguir las *colocaciones léxicas* de la lengua general de lo que él denomina *colocaciones conceptuales*, presentes en el discurso especializado. Las colocaciones léxicas, a pesar de existir en la lengua especializada (p. ej. *to confer inmunity* [Montero Martínez 2002]), son características de la lengua general y destacan por su carácter arbitrario. Por el contrario, las colocaciones conceptuales son las más habituales en la lengua especializada y no presentan ese carácter impredecible de las combinaciones anteriores, sino que responden a combinaciones conceptuales factibles en un dominio determinado (Picht 1991; L'Homme 2000). Así, la noción de colocación conceptual de Martin (1992) parte de la estructura argumental de los conceptos, perceptible en la definición. De este modo, los conceptos abren unos *slots* que se rellenan con otros conceptos de determinados rasgos semánticos. Por ejemplo, la definición de *disease* incluye un *slot* de CAUSA, que da lugar a combinaciones como *infectious disease*. Lo mismo ocurre con *nervous system*, donde *nervous* especifica la FUNCIÓN presente en la definición de *system*; o con *etymological dictionary*, donde *etymological* rellena el *slot* de INFORMACIÓN (Martin 1992). Las aportaciones de Martin (1992) han resultado muy valiosas para el conocimiento de la fraseología especializada y constituyen la base de la noción de *microcontexto* que planteamos en esta tesis, como veremos en la Sección 5.1.2.

La dicotomía entre colocaciones léxicas y conceptuales también fue abordada por otros autores, como Heid (1994, 2001) o L'Homme (1998, 2000, 2002). Heid (1994, 2001) mantiene que las definiciones reflejan los *slots* que abren los conceptos. Estos *slots* se

rellenan con conceptos de rasgos similares, que se pueden agrupar en clases semánticas, y que generarán colocaciones conceptuales. Abundando en la distinción entre colocaciones léxicas y colocaciones conceptuales, Heid (2001: 795) analiza los adjetivos y verbos que se combinan con *respect*, *smell* y *alcohol* en el *Collins English Dictionary* y el *Oxford Advanced Learner's Dictionary*. Encuentra diferentes tipos de combinaciones, como aquellas en las que el adjetivo alude a la intensidad (*deep respect*) o a alguna cualidad del nombre (*bad smell*), así como combinaciones en las que el adjetivo posee una mayor carga conceptual (*water-soluble alcohol*). Estas últimas suelen constituir hipónimos del núcleo nominal (*water-soluble alcohol* es un tipo de *alcohol*) y son las más interesantes desde el punto de vista terminológico, ya que dan cuenta del sistema conceptual subyacente. Como vemos en estos ejemplos, la noción de colocación conceptual no se aplica únicamente a las colocaciones propiamente dichas, sino que también resulta patente en los términos compuestos. Más adelante ahondaremos en el debate sobre la distinción entre colocaciones y términos compuestos que también tiene lugar en la terminología.

L'Homme (1998: 514, 2000: 105-106) coincide en que existen diferencias entre las colocaciones de la lengua general y las del discurso especializado. Esta autora denomina *colocaciones* a las primeras, que constituyen coocurrencias léxicas restringidas de la lengua general, mientras que llama *combinaciones léxicas especializadas* a las segundas, que representan coocurrencias léxicas libres en el lenguaje especializado. L'Homme (1998, 2000) afirma que estas combinaciones léxicas especializadas suelen ser composicionales y, mediante el análisis de los colocados, se pueden establecer clases conceptuales: por ejemplo, el verbo *run* se combina con *program*, *operating system*, *routine* o *application*, que pertenecen a la categoría de PROGRAM (L'Homme 2002: 5). Estos términos de rasgos semánticos similares suelen compartir patrones colocacionales (p. ej. los miembros de la categoría PROGRAM se combinarán con el mismo tipo de verbos o adjetivos), algo que no es tan frecuente en las colocaciones de la lengua general. Los trabajos de Marie-Claude L'Homme aplican la Teoría Sentido-Texto al discurso especializado. Como comentábamos, esta teoría presta especial atención a las unidades fraseológicas al destacar la estructura argumental de las unidades de base verbal (verbos y nominalizaciones), así como las combinaciones léxicas posibles.

Las teorías de la Terminología se han interesado de manera desigual por las unidades fraseológicas especializadas. En la Teoría General de la Terminología (Wüster 1968, 1979), al considerarse el término como elemento uniforme y estático, cuyo contexto resulta irrelevante, la fraseología carecía de importancia. Con la llegada de las nuevas teorías terminológicas, que daban cuenta del uso real de la lengua desde una

perspectiva descriptiva, se reconoció la importancia del análisis de los términos en contexto. Sin embargo, desde la Socioterminología (Gaudin 1993, 2003, 2005) y la Teoría Sociocognitiva de la Terminología (Temmerman 2000, 2007) no se ha distinguido entre términos simples y estructuras de extensión superior, entre las que se encuentran las unidades fraseológicas. La Teoría Comunicativa de la Terminología (Cabré 1993, 1999), por su parte, sí se ha detenido en el estudio estas combinaciones, como lo demuestran los trabajos de Cabré et al. (1996), Lorente Casafont (2002) o Bevilacqua (2004). Esta última autora aborda la combinación conceptual que se produce en estas coocurrencias especializadas, centrándose en las propiedades y el comportamiento de lo que denomina *unidades fraseológicas especializadas eventivas*, es decir, de base verbal (Bevilacqua 2004). No obstante, el principal punto de interés de la Teoría Comunicativa de la Terminología ha sido la distinción entre unidades fraseológicas especializadas y unidades terminológicas polilexemáticas. Así, se separa la fraseología de la terminología, como muestran Cabré et al. (1996), sobre la distinción de unidades fraseológicas especializadas y unidades terminológicas polilexemáticas y su uso en la extracción terminológica automática, o Lorente Casafont (2002), sobre los tipos de verbos presentes en el discurso especializado, entre los que se encuentran los fraseológicos y los terminológicos.

Desde la Terminología Basada en Marcos (Faber et al. 2005, 2006; Faber 2012), en la que se enmarca esta tesis, también se ha prestado especial atención a las unidades fraseológicas especializadas. No resulta extraño, ya que la representación conceptual por medio de marcos representa un escenario ideal para entender la combinación de conceptos que se da en las unidades fraseológicas especializadas. Así, trabajos como los de Tercedor Sánchez (1999), Montero Martínez (2002) y Buendía Castro (2013) proponen un análisis de la fraseología especializada basado en la combinación conceptual, con vistas a paliar la carencia de unidades fraseológicas en los recursos terminográficos. Como parte del proyecto ONCOTERM, dedicado a la terminología oncológica, Tercedor Sánchez (1999) realiza un análisis comparativo de las unidades fraseológicas presentes en un corpus de oncología en inglés y español con el fin de proponer nuevas formas de representación útiles para la traducción. Por su parte, Montero Martínez (2002) se centra en las colocaciones y los términos compuestos (que engloba bajo la denominación de *frasemas terminológicos*, como veremos a continuación) en inglés y español en este subdominio. Realiza una propuesta de estructuración conceptual de estas unidades basada en esquemas fraseológicos compuestos por categorías y relaciones semánticas. Buendía Castro (2013) utiliza un corpus de textos medioambientales en inglés, pertenecientes al subdominio del evento extremo, para analizar las colocaciones verbales. En ellas, clasifica los verbos en dominios y subdominios léxicos (Faber y Mairal 1999) y asigna categorías y

roles semánticos a los argumentos. Nuestro estudio parte de las mismas premisas y asume la fraseología especializada como eminentemente conceptual, basada en combinaciones factibles de conceptos en un dominio, frente a la arbitrariedad e idiomaticidad de las combinaciones generales.

En nuestra opinión, la fraseología representa una parte sustancial de la terminología, pues constituye una puerta de acceso al sistema conceptual subyacente. En este sentido, Heid (1992: 535) habla de la doble vertiente que presenta el análisis terminológico por medio de las unidades fraseológicas especializadas. Por un lado, los conceptos de un dominio son indicadores de las unidades fraseológicas que pueden encontrarse en dicho campo. Por otro, las propias unidades fraseológicas permiten la descripción conceptual del dominio.

Meyer y Mackintosh (1996) también destacan de manera muy acertada la utilidad de las unidades fraseológicas en el análisis conceptual, con vistas a la creación de bases de conocimiento terminológico. Entre otros aspectos, subrayan el papel de los compuestos en las jerarquías genérico-específicas (p. ej. *CD-ROM disk* es un tipo de *disk*). Sostienen que la herencia de colocados también se da en las unidades fraseológicas, de forma que los hipónimos heredan los patrones colocacionales de los hiperónimos. Por ejemplo, «one *reads data from a CD-ROM disk* because one *reads (data) from* all computer disks» (Meyer y Mackintosh 1996: 14). Esta idea coincide con las aportaciones de L'Homme (1998, 2000) que hemos comentado previamente, relativas a los patrones colocacionales que comparten los conceptos similares, que en este caso serían más específicos.

Meyer y Mackintosh (1996: 12-13) señalan además la utilidad de los compuestos para detectar la multidimensionalidad, es decir, los casos en los que un concepto puede clasificarse desde distintas perspectivas (Bowker 1997), por ejemplo, *land/air/water vehicles* frente a *motorized/non-motorized vehicles*. Asimismo, sostienen que las unidades fraseológicas pueden revelar metáforas propias de un dominio, como «Internet is a highway», que se puede apreciar en colocaciones como *ride/travel/cruise/navigate the Internet* o compuestos como *message traffic* (Meyer y Mackintosh 1996: 5).

Asimismo, defienden que las colocaciones y los compuestos permiten identificar términos polisémicos y sinónimos. En concreto, sostienen que la combinación con diferentes tipos de colocados suele indicar cambios de significado (esto es, términos polisémicos), mientras que la coincidencia de colocados suele revelar casos de sinonimia. Así, gracias a colocaciones como *processes are born*, *processes sleep* o *to kill a process*, podemos intuir que *process* ha adquirido un significado especializado en el área de la

informática, diferente del que posee en la lengua general o en otros dominios especializados (Meyer y Mackintosh 1996: 17). Por el contrario, hay términos como *gallery* y *library* que coocurren con unidades similares, como en el caso de *gallery/library of charts*, *gallery/library of clip art* o *template gallery/library*, lo que apunta a que se trata de términos sinónimos (Meyer y Mackintosh 1996: 18). No obstante, las autoras señalan la necesidad de complementar este análisis con el uso de corpus o la consulta a expertos.

Como se desprende de estas aportaciones, Meyer y Mackintosh (1996) consideran los compuestos y las colocaciones como unidades fraseológicas especializadas, que aúnan bajo la denominación de *frasema terminológico*. Sin embargo, antes de ahondar en esa cuestión, que surge como respuesta a un profundo debate sobre la distinción entre colocaciones y compuestos, nos parece necesario abordar otro tema no menos controvertido: la consideración de los términos compuestos como unidades fraseológicas.

Los términos compuestos coinciden con las unidades fraseológicas en varios aspectos: se trata de unidades plurilexemáticas que se combinan frecuentemente, funcionan como un todo y presentan un cierto grado de fijación. Por ello, hay autores que los consideran parte de la fraseología. Sin embargo, existen otras posturas reacias a incluirlos en la lista de unidades fraseológicas especializadas. Estas defienden que, a diferencia de las unidades fraseológicas especializadas, los términos compuestos surgen ante una necesidad de denominación (Cabré y Estopà 2005). Además, propugnan que los términos compuestos representan unidades léxicas que funcionan como un constituyente autónomo en la oración (Roberts 1993; Cabré et al. 1996), mientras que las unidades fraseológicas especializadas son unidades sintácticas (ya que consideran que esta categoría está formada principalmente por estructuras verbales) (Cabré et al. 1996; Lorente Casafont et al. 1998; Aguado de Cea 2007). Asimismo, se señala la invariabilidad de los términos compuestos frente a la variación presente en las unidades fraseológicas (Cabré et al. 1996).

En este trabajo también observamos los términos compuestos como mecanismo de respuesta a las necesidades de denominación. Sin embargo, consideramos que algunos de los criterios expuestos no clarifican la distinción entre términos compuestos y unidades fraseológicas especializadas. Por ejemplo, la consideración de constituyente autónomo no siempre es evidente, ya que hay combinaciones como *radiografía del pie* que representan un sintagma nominal y, para nosotros, un constituyente autónomo, pero sin embargo no se consideran términos compuestos, ni siquiera fraseología, sino sintagmas frecuentes (Lorente Casafont et al. 1998). Por otro lado, no compartimos la invariabilidad atribuida a

los términos compuestos, ya que los resultados de esta tesis, así como de estudios precedentes (Nkwenti-Azeh 1998; Tercedor Sánchez 1999; Fernández Silva 2016, 2018; entre ellos los trabajos fundadores de la Teoría Comunicativa de la Terminología: Cabré 1993, 1999) han demostrado que la variación es una realidad en los términos, tanto simples como compuestos. Además, como afirma Cabré (1993: 145), la variación, entendida en términos de inestabilidad formal o semántica, es un indicador de neología, que también se halla presente en los términos.

Algunos autores sitúan los términos compuestos en la frontera entre la terminología y la fraseología (SanJuan et al. 2005; Sanz Vicente 2011), mientras otras voces han señalado lo borroso de los límites entre ambos grupos (Pavel 1993; Gläser 1994; Gouadec 1994). Argumentan que, a menudo, una unidad fraseológica especializada que describe un concepto va adquiriendo una mayor fijación hasta llegar a utilizarse como etiqueta para ese concepto (Pavel 1993; Gläser 1994). Esta etiqueta, que es un término compuesto, puede sufrir reducciones (a modo de nominalización, adjetivación o verbalización) (Pavel 1993) o bien mantener su carácter pluriléxico (Sager 1990). En esta línea, hay investigadores que apuestan por la inclusión de los términos compuestos entre las unidades fraseológicas especializadas, como Kocourek (1982), Meyer y Mackintosh (1996), Tercedor Sánchez (1999), Montero Martínez (2002), Buendía Castro (2013) o aquellos pertenecientes a la Lingüística Computacional, en la que se consideran *multiword expressions* todas las estructuras poliléxicas que pueden plantear problemas a un sistema, como los términos compuestos (Ramisch 2015; Parra Escartín 2018). En esta tesis también consideramos que los términos compuestos son unidades fraseológicas especializadas, pues comparten con estas su carácter pluriléxico, su frecuencia de coaparición, su funcionamiento como unidad y su cierto grado de fijación. Se diferencian de otras unidades fraseológicas especializadas, como las locuciones (p. ej. *a tenor de lo establecido*), porque representan conceptos del sistema.

Como comentábamos anteriormente, el debate sobre la consideración de los términos compuestos como unidades fraseológicas especializadas va ligado a otra cuestión también controvertida: la distinción entre colocaciones y términos compuestos. Ambas polémicas van de la mano, ya que al distinguir las unidades fraseológicas especializadas de los términos compuestos, se están distinguiendo las colocaciones (principal prototipo de unidades fraseológicas) de estos términos sintagmáticos.

Así, normalmente se afirma que los términos compuestos representan un concepto, mientras que las colocaciones representan dos (Heid 1994; Tercedor Sánchez

1999; Sager 1990; Lorente Casafont et al. 2017). A este respecto, pensamos que no es tarea fácil distinguir cuándo una combinación representa un solo concepto o más de uno. Más que tal distinción nos parece que se debería atender a la relevancia de dicha combinación en un sistema conceptual. Con ello nos referimos a unidades entendidas como colocaciones (p. ej. *enfermedad infecciosa*) que, aunque podrían considerarse como la suma de los dos conceptos implicados, deberían estar presentes en un sistema conceptual porque establecen relaciones con otros elementos del sistema. Por ejemplo, son un *tipo_de* enfermedad, *causado_por* microorganismos patógenos, que *causa_síntomas* como fiebre, diarrea, fatiga, dolores musculares o tos.

Igualmente se ha añadido que los formantes de los términos compuestos aparecen juntos, mientras que los miembros de las colocaciones pueden mostrarse separados (Heid 1994; Tercedor Sánchez 1999; Lorente Casafont et al. 2017). Esta apreciación nos parece adecuada, ya que, aunque los compuestos permiten en ocasiones la introducción[11] de otros elementos (*generación eólica* > *generación de energía eólica*), sus componentes no admiten tanta separación como la que puede darse en las colocaciones (*generar/producir potencia* > *la* potencia generada *por el aerogenerador es superior que la que se* produce *con reducidos coeficientes de cortadura*).

Se suele afirmar, también, que las colocaciones son composicionales, es decir, su significado se puede deducir del significado de sus componentes (*enfermedad infecciosa*), mientras que los términos compuestos no son composicionales, esto es, la suma de sus formantes no permite conocer su significado global (*infarto de miocardio*) (Toledo Báez y Martín Lorente 2018: 108). Sin embargo, estas autoras proponen como ejemplos de términos compuestos algunas combinaciones que podrían considerarse composicionales (p. ej. *actividad física*, *atención médica*) (Toledo Báez y Martín Lorente 2018: 111). Con ello queremos señalar que la composicionalidad no es un factor de fácil medición. Así, existen combinaciones que podríamos decir composicionales, como *energía eólica* (por lo tanto, consideradas colocaciones) y otras cuyos elementos permiten comprender a grandes rasgos el concepto, pero que requieren ahondar en el conocimiento experto para captar su significado al completo (por lo tanto, consideradas términos compuestos). Este es el caso, por ejemplo, de *stall-regulated wind turbine*. Se puede inferir que se trata de una turbina eólica que se regula y se detiene, pero faltan detalles importantes de su significado, como las altas velocidades de viento necesarias para que la turbina se detenga (Cabezas García y

[11] Cabe recordar que, en trabajos como los de Cabré et al. (1996) o Lorente Casafont et al. (2017), se alude a la fijación de los compuestos frente a la variabilidad de las unidades fraseológicas, entre las que se encuentran las colocaciones.

Faber 2017a: 153). No obstante, si tenemos en cuenta que tanto *energía eólica* como *stall-regulated wind turbine* representan conceptos particulares del dominio de la energía eólica, ambas podrían considerarse también términos compuestos.

Por otro lado, las aportaciones de Cabré et al. (1996) antes expuestas se aplican igualmente a su distinción entre colocaciones y términos compuestos. Estas autoras afirman que dicha diferenciación debe basarse en criterios cuantitativos y formales. Como hemos visto, defienden que los términos compuestos, a diferencia de las colocaciones (principalmente estructuras de base verbal), forman una unidad autónoma en la oración. Además, son unidades fijas cuyo núcleo es un término, mientras que las colocaciones especializadas cuentan con un término en el colocado. Igualmente, las combinaciones que cuenten con un verbo o una unidad deverbal constituirán siempre colocaciones (p. ej. *inoculación del virus*), pues cumplen el mismo patrón argumental del verbo (Lorente Casafont et al. 2017: 208). No compartimos este criterio de presencia de un núcleo terminológico para ser considerado término compuesto, así como la asignación de las combinaciones con un elemento verbal al grupo de las colocaciones. Con respecto a la presencia de un núcleo terminológico para ser considerado término compuesto, existen combinaciones como *surf zone* cuyo núcleo (*zone*) podría decirse general (por lo tanto, se trataría de una colocación), que designan un concepto especializado o, como preferimos afirmar en esta tesis, cuentan con un papel relevante en el sistema conceptual, por lo que también podrían considerarse términos compuestos. Por otro lado, nos parece que la distinción entre colocaciones y términos compuestos no puede utilizar el criterio de presencia o ausencia de un verbo. Como indican Sager (1990) o Gagné (2000), los compuestos responden a una combinación conceptual, en la que bien puede intervenir una forma deverbal con sus argumentos. De hecho, existen combinaciones con formas deverbales (consideradas, entonces, colocaciones), como es el caso de *horizonte de predicción*, cuya presencia en un sistema conceptual es relevante y, por tanto, también podrían considerarse términos compuestos.

Pese a que es cierto que algunos de los criterios comentados permiten delimitar la noción de término compuesto (en concreto, su designación de un concepto, la contigüidad de sus elementos y la frecuente falta de composicionalidad), estos no se aplican al conjunto de la clase. Así, la ausencia de pautas determinantes para diferenciar las colocaciones de los términos compuestos en los textos de especialidad nos lleva a cuestionar la necesidad de tal distinción. Como afirman Meyer y Mackintosh (1994, 1996) y Heid (2001), más que la clasificación de ambos fenómenos, es la relevancia que comparten en el análisis terminológico lo que resulta importante. De este modo, Meyer y

Mackintosh (1994, 1996) abogan por no distinguir las colocaciones de los términos compuestos en los textos especializados y proponen la denominación de *frasema terminológico* para englobar a ambas combinaciones.

Desde la Terminología Basada en Marcos (Faber et al. 2005, 2006; Faber 2012) también se defiende la semejanza de las colocaciones y los términos compuestos como herramientas de análisis conceptual (Montero Martínez 2002, 2008). Estos se aglutinan bajo la etiqueta de *construcciones fraseológicas*, mediante las cuales «se garantiza la coherencia en la estructuración y descripción de las unidades fraseológicas que se incluyen en la base de conocimiento terminológico EcoLexicon» (Montero Martínez y Buendía Castro 2012: 713). Las construcciones fraseológicas pueden ser relacionales, si codifican una relación, o atributivas, si aluden a atributos. En la siguiente lista se desglosan ejemplos de ambos tipos de construcciones (Montero Martínez y Buendía Castro 2012: 717):

Construcciones relacionales:

- Construcciones subcategorizantes: *spring tide* (spring tide *is_a* tide).

- Construcciones coordinantes: *luni-solar tide*.

- Construcciones partitivas: *river bank* (river bank *is_part_of* river).

- Construcciones fruto de las relaciones específicas de un dominio: *diversion channel* (*has_function*).

Construcciones atributivas:

- Construcciones con núcleo nominal: *semidiurnal tide* [TIDAL-PERIOD].

- Construcciones con núcleo adjetivo deverbal: *highly eroded* [DEGREE-OF-EROSION].

- Construcciones con núcleo prepositivo: *barco a la deriva* [DIRECTION].

Estas construcciones responden a menudo a repetidas ecuaciones semánticas. Esto llevó a plantear la existencia de *esquemas fraseológicos*, que representan generalizaciones de las combinaciones semánticas presentes en diversas construcciones (Montero Martínez et al. 2002; Montero Martínez 2008). Los rasgos semánticos transmitidos por los esquemas son heredados por sus construcciones subordinadas (Montero Martínez 2008), lo que guarda una relación directa con la herencia de colocados de la que hablaban Meyer y Mackintosh (1996) y los patrones colocacionales compartidos por conceptos similares

(L'Homme 1998, 2000). Así, el esquema fraseológico EROSION_*affects*_EARTH-CRUST incluye el esquema más específico EROSION_*affects*_RIVER-BANK. Estos pueden materializarse en construcciones de diferente forma: *to erode the river bank*, *river bank erosion*, *eroded river bank* o *erosion of the river bank* (Montero Martínez 2008: 1019). La propuesta de construcciones relacionales y atributivas, así como las generalizaciones que permiten los esquemas fraseológicos, nos parecen de vital importancia, pues recogen los principales patrones combinatorios que pueden darse en las unidades fraseológicas especializadas. Además, estas ideas resultan fundamentales para el estudio de la formación de términos, por lo que han servido de sustento a nuestro análisis terminológico (véase la Sección 5.1.1).

Así pues, respecto a la diferenciación de colocaciones nominales y compuestos, seguimos las ideas de Meyer y Mackintosh (1996) y la Terminología Basada en Marcos (Montero Martínez 2002; Montero Martínez 2008; Montero Martínez y Buendía Castro 2012) y optamos por no diferenciar estas combinaciones. Por el contrario, atendemos a un principio común: su relevancia en el sistema conceptual. De este modo, preferimos denominar a ambas *compuestos nominales* (CN), pues esta designación cuenta con una amplia aceptación y es válida en los dos casos, ya que refleja la estructura nominal sintagmática que comparten. Los CN son un tipo de términos compuestos, una etiqueta que, como hemos indicado anteriormente, también engloba otras combinaciones que no tienen un núcleo nominal (p. ej. *dose-dependent*, *highly-eroded*), por lo que se corresponden en muchos casos con las construcciones fraseológicas que acabamos de presentar (Montero Martínez y Buendía Castro 2012). Sin embargo, preferimos hablar de términos compuestos porque esta denominación refleja la designación de un concepto (*término*) y la calidad de unidad pluriléxica (*compuesto*).

Fuera de la categoría de términos compuestos quedarían, pues, las unidades fraseológicas de carácter verbal (p. ej. *consumir energía*, mientras que *energía consumida* se consideraría término compuesto). Realizamos esta distinción siguiendo a Heid (2001: 791), quien apunta que, para fines terminográficos, puede convenir establecer alguna diferenciación:

> When designing a terminological data collection, one should, however, try to establish some explicit internal guideline, e.g., classifying subtype-denoting noun-adjective collocations as (multiword) terms, and noun-verb collocations as combinatory properties of a (nominal) term.

A pesar de que esta tesis se centra en términos compuestos de carácter nominal, destacamos el papel de los verbos en el discurso especializado, ya que permiten expresar

eventos y estados, que constituyen gran parte del conocimiento general y experto (Faber y Mairal 1999). No obstante, estos han recibido menos atención que las estructuras nominales en terminología (L'Homme 1998; Aguado de Cea 2007; Buendía Castro 2013). Esto puede deberse a que, a menudo, se trata de verbos generales que adquieren un valor especializado al combinarse con términos (Aguado de Cea 2007; Buendía Castro y Faber 2015), como es el caso de *entregar potencia*. Sin embargo, también existen verbos que constituyen términos (p. ej. *acidificar*), aunque son mucho menos frecuentes. Ante la diversidad de verbos presentes en el lenguaje especializado, Lorente Casafont (2002) plantea la siguiente clasificación:

-verbos discursivos, directamente ligados a la función textual que ejercen (*describir, presentar*);

-verbos conectores, que relacionan dos conceptos (*parecer, corresponder*);

-verbos fraseológicos, que adquieren valor especializado cuando se combinan con un término, formando una unidad fraseológica especializada (*generar*, *consumir*);

-verbos-término, vinculados a un área de especialidad (*capialzar*, *eutrofizar*).

Así pues, llamaremos colocaciones a las unidades fraseológicas de estructura sustantivo + verbo, que son consideradas prototipo de colocaciones para diversos autores (Cabré et al. 1996; Lorente Casafont et al. 1998; Aguado de Cea 2007). Estas, junto a los CN, constituyen el grueso de la fraseología especializada.

3.2 Noción de compuesto nominal y estudios principales

Los compuestos nominales (CN) han recibido multitud de definiciones que, por lo general, destacan su formación por varias unidades, su núcleo nominal, su carácter unitario y su función denominativa. Esta última es una de las propiedades de los CN que goza de mayor consenso (Zimmer 1971; Downing 1977; Štekauer 1998; Fernández Domínguez 2019). Así, tanto en la lengua general como en la especializada, los CN sirven para nombrar conceptos. Esta función denominativa es la que nos lleva a considerar como CN estructuras de postmodificación preposicional (presentes en inglés, pero especialmente en español), por ejemplo, *fuente de energía*. El carácter denominativo de los CN constituye la principal distinción entre estos y otros sintagmas nominales de naturaleza descriptiva (Zimmer 1971; Downing 1977; Levi 1978). De este modo, mientras que *turbina actual* comenta un rasgo de la turbina, *turbina eólica* nombra un nuevo concepto. No obstante, la distinción entre denominación y descripción no siempre es evidente. Para ello, se han propuesto diferentes pautas, como la mayor estabilidad de los CN, perceptible en una

frecuencia superior (Fernández Domínguez 2019); la existencia de variantes monoléxicas en la misma o en otra lengua (Sager et al. 1980; Kocourek 1982; Cabré 1993; Fernández Domínguez 2019); o la presencia de definiciones del CN (Kocourek 1982).

Con respecto a la frecuencia, ya hemos mencionado que los CN suelen aparecer más veces que las combinaciones libres. Guilbert (1965) señala que, a mayor frecuencia, mayor estabilidad del CN y del concepto designado. No obstante, no siempre encontramos cifras elevadas, ya que existen distintos factores que pueden influir en estos números, entre los que destacamos la baja lexicalización que puede darse al inicio del proceso denominador de los CN (Sager et al. 1980; Sanz Vicente 2011), el carácter neológico presente en muchos CN (véase la Sección 3.4.5) o el tamaño reducido que pueda tener el corpus. En cualquier caso, podemos afirmar que la frecuencia de aparición es una de las características de los CN que debemos tener en cuenta y que, a pesar de no ser el único criterio para su diferenciación, se complementa con las otras particularidades que comentamos en este trabajo.

No obstante, a pesar del consenso sobre la función designativa de los CN, observamos que los intentos de definir estas combinaciones divergen en varios puntos, principalmente en el tipo de formantes y en la unión o separación entre estos. En cuanto al tipo de formantes, la mayor parte de los estudios destacan que los CN están formados solo por sustantivos, es decir, son series de sustantivos que funcionan como uno solo (Downing 1977; Finin 1980; Gagné 2000), p. ej. *ozone layer*. Esta concepción parte de la propensión de las lenguas germánicas como el inglés, idioma que suelen analizar estos trabajos, hacia la concatenación de sustantivos como mecanismo de formación de palabras. En este sentido, se han propuesto diferentes nombres para estos sintagmas, fruto de la proliferación de estudios al respecto: *compounds*, *noun compounds*, *noun-noun compounds*, *compound nouns* o *nominal compounds* son los principales ejemplos.

Sin embargo, a la diversidad de la nomenclatura se suma su inconsistencia, pues algunas de estas denominaciones también se han aplicado a los sintagmas nominales modificados no solo por sustantivos, sino también por otras categorías gramaticales, como los adjetivos (p. ej. *zenithal angle*). Es el caso, por ejemplo, de *nominal compound*, que Cordeiro et al. (2016: 1987) utilizan para referirse a sintagmas nominales formados por dos o más palabras que, pese a ser fundamentalmente sustantivos en inglés, pueden contar con otro tipo de formantes en otros idiomas. Sin ir más lejos, la propia lengua inglesa también es proclive a formar palabras mediante la secuencia Adj+N. Ya lo señalaba Levi (1978), en un estudio pionero al considerar que en los CN pueden intervenir

modificadores adjetivales. Le siguieron otros autores que también reconocieron la presencia de modificadores que no eran sustantivos (p. ej. adjetivos, adverbios), como Sager et al. (1980), Lauer (1995), Daille (1999, 2001, 2017), Estopà (2000), Adams (2001) o Maniez (2009, 2014). No obstante, la mayor parte de los trabajos siguen centrándose en CN en los que solo intervienen sustantivos, a pesar de la gran cantidad de CN cuyos modificadores pertenecen a otras categorías gramaticales.

Dado que nuestra tesis se inscribe en esta concepción amplia de los CN, adoptamos la denominación de *complex nominals* de Levi (1978) en nuestra producción en inglés, pues nos parece una manera acertada de nombrar CN que no están formados únicamente por sustantivos. Así, esta autora prefiere hablar de *complex nominals* para referirse a «syntactic construction(s) dominated by an N node and composed (in its simplest form) of a head noun preceded by a modifier which is either another noun or a nominal adjective» (Levi 1978: 39). Por una parte, favorece la denominación *complex* frente a *compound* porque este último suele asociarse a secuencias de sustantivos pronunciados con acentuación inicial (p. ej. *stéam boat*). Sin embargo, se ha demostrado que la acentuación no es determinante a la hora de delimitar los CN[12]. Por otra parte, utiliza *nominal* en lugar de *noun*, pues este último se ha atribuido tradicionalmente a CN con unión ortográfica formados únicamente por sustantivos (Levi 1978: 39). No obstante, como se desprende de su definición, Levi (1978) solo analiza CN con modificador adjetival si se trata de lo que ella denomina *nominal nonpredicating adjectives*, es decir, adjetivos procedentes de sustantivos (p. ej. *maternal*) que establecen relaciones semánticas. No compartimos este punto, ya que en esta tesis analizamos CN que cuentan con todo tipo de adjetivos (y otras categorías gramaticales, como participios), siempre que constituyan un término compuesto. Para referirnos en español a estas unidades preferimos la denominación de *compuesto nominal* (CN), como hemos indicado en la Sección 3.1, pues esta designación refleja la estructura nominal sintagmática de estas unidades, no está tan ligada como otras denominaciones a CN formados solo por sustantivos (p. ej. *noun compound*) y, además, cuenta con una amplia aceptación en la literatura en español.

Sin embargo, es cierto que, tanto en español como en inglés, la noción de *compuesto* no está libre de controversia, ya que unos la atribuyen únicamente a los compuestos con unión ortográfica, mientras que otros la utilizan también para compuestos con espacios entre sus formantes. Así, como comentábamos al inicio de este

[12] En este trabajo no consideramos la acentuación como un aspecto como característico de los CN, debido a la falta de consenso al respecto y a la naturaleza de nuestro estudio, en el que nos basamos en textos, que no proporcionan información sobre estas cuestiones fonológicas.

apartado, la unión o separación entre los formantes del compuesto constituye otro de los puntos discordantes entre las diferentes nociones de CN.

En la mayoría de trabajos se conciben los compuestos como la unión de dos o más palabras en una sola, por ejemplo, *breakwater* o *rompeolas* (Kocourek 1982; Corpas Pastor 1996). Sin embargo, la línea divisoria entre lo que se suele denominar compuestos gráficos y compuestos sintagmáticos no resulta fácilmente perceptible (Jespersen 1942; Downing 1977; Warren 1978; Fernández Domínguez 2019). Con frecuencia se han utilizado criterios gráficos, fonológicos y semánticos para diferenciar a unos y otros, aunque ninguna de estas pautas ha resultado ser concluyente. A esto se suma el hecho de que no todas las lenguas utilizan los mismos mecanismos de formación de palabras y, mientras unas lenguas crean compuestos gráficos con mayor facilidad (como es el caso del inglés), otras tienden a la formación de compuestos sintagmáticos (como ocurre en español) (Sanz Vicente 2011; Štekauer et al. 2012; Fernández Domínguez 2019). Por ello, siguiendo la línea de algunas de las autoras más influyentes en el campo de los compuestos, como Downing (1977), Levi (1978) y Warren (1978), denominamos *compuestos* tanto a los que presentan unión gráfica como a los que cuentan con separación entre sus formantes. Nos parece una opción unificada y sistemática que engloba las diferentes formas que puede adquirir la composición en distintas lenguas. No obstante, en esta tesis nos centramos en los compuestos sintagmáticos, en concreto en los de tipo nominal. Así, entendemos los CN como combinaciones que presentan un núcleo nominal modificado por otras unidades, como sustantivos, adjetivos o sintagmas preposicionales, que pueden aparecer unidos o separados gráficamente y constituyen una unidad léxica. En el lenguaje especializado, son términos compuestos si forman parte de un sistema conceptual.

Probablemente el único aspecto indudable de estas combinaciones sea su importancia a distintos niveles. Por un lado, se trata de unidades muy frecuentes en diferentes lenguas, especialmente en el discurso especializado, ya que constituyen el principal procedimiento de formación terminológica (Sager et al. 1980; Kocourek 1982; Fernández Domínguez 2016). Por otro lado, pueden plantear dificultades, por ejemplo al identificarlos, analizarlos sintáctica y semánticamente, traducirlos o representarlos en recursos lingüísticos. Por ello, no es de extrañar que los CN se hayan abordado desde distintas disciplinas, como la Lingüística teórica, la Terminología, la Lexicografía, la Lingüística Computacional o la Psicolingüística, dando lugar a interesantes aportaciones de las que se nutre esta tesis.

Desde la Lingüística teórica, diversos autores se han centrado en aspectos formales, gramaticales y semánticos de los CN. Uno de ellos es Jespersen (1942), quien estudia CN del inglés y analiza sus patrones de formación utilizando categorías semánticas (p. ej. MATERIAL, INSTRUMENT). Defiende que el significado de los CN no es completamente composicional, sino que es necesario extraer información del contexto (Jespersen 1942: 137). En esta línea, concluye que las relaciones semánticas que pueden darse entre el núcleo y el modificador de un CN son infinitas y, por lo tanto, no puede establecerse una clasificación fija de las mismas (Jespersen 1942: 143).

Lees (1960), por su parte, estudia la formación de los CN en inglés desde un enfoque generativo. Para ello, defiende que estos se derivan de oraciones, que concibe como estructuras puramente sintácticas. Así, habla de la función sintáctica (p. ej. sujeto, objeto) que desempeñan los elementos del CN en la oración subyacente. Esta visión despertó muchas críticas, que compartimos, y que principalmente se basaban en el carácter sintáctico de la propuesta y en su arbitrariedad (Downing 1977). Más tarde, Lees (1970) presenta una propuesta modificada, en la que adopta una visión mucho más semántica (con roles semánticos como AGENT o PATIENT), argumentando que la sintaxis solo es la forma que impone la semántica (Lees 1970: 185). Además, propone un conjunto de *generalized verbs* que subyacerían tras los CN, con el fin de dotar a su propuesta de sistematicidad. Esta concepción semántica nos parece mucho más acertada, pues consideramos que, más que la sintaxis, es la semántica la responsable de las combinaciones que se dan en los CN.

Años más tarde, en uno de los estudios más conocidos sobre CN en inglés, Downing (1977) se interesa por los patrones de creación de nuevos CN, que difieren según la categoría semántica del núcleo (p. ej. HUMAN, ANIMAL, PLANT) (Downing 1977: 831), un aspecto de gran relevancia que revisamos en nuestros términos compuestos. Destaca, además, la función denominativa de los CN y afirma que estos no son siempre composicionales, por lo que se necesitan conocimientos pragmáticos para acceder a su significado (Downing 1977: 834). Nosotros también consideramos que el contexto es fundamental para determinar la carga conceptual de un CN, si bien es cierto que no todos los autores aceptan este enfoque pragmático. En esta línea y, siguiendo a Jespersen (1942), Downing (1977) señala que no es posible realizar una clasificación fija de las relaciones semánticas presentes en los CN.

La década de los 70 resultó especialmente prolífera, pues se publicaron varios de los grandes trabajos consagrados a los CN. Además de los estudios de Lees (1970) y

Downing (1977), las aportaciones de Levi (1978) y Warren (1978) resultaron muy valiosas y sentaron las bases de investigaciones posteriores. Así, el de Levi (1978) es probablemente el estudio más conocido acerca de los CN, que ha servido de referencia para la mayoría de trabajos de todas las áreas. Desde una perspectiva generativa, Levi (1978) estudia la formación de CN en inglés. Como ya hemos mencionado, destaca por incluir CN con modificadores adjetivales. Defiende que los CN se forman a partir de un predicado subyacente, mediante un proceso de nominalización o de omisión de dicho predicado, una postura que compartimos en esta tesis. Para los casos de omisión del predicado, propone un conjunto de nueve *recoverably deletable predicates*, con los que pretende dar sistematicidad a las relaciones semánticas subyacentes en este tipo de CN. A pesar de que encontramos algunas limitaciones, como el carácter eminentemente sintáctico que adopta en muchos casos (p. ej. cuando habla de sujeto u objeto del verbo, véase la Sección 3.3), consideramos que se trata de una obra de referencia en el campo de los CN.

Asimismo, Warren (1978), en otro de los estudios principales del ámbito, explora los patrones semánticos en la formación de CN en inglés, utilizando categorías (PLACE, MATERIAL, etc.) y clases semánticas (ANIMATE, INANIMATE, etc.), que se combinan con cada relación semántica. No analiza en este trabajo los CN en los que intervienen nominalizaciones, ya que su objetivo es explorar la relación semántica implícita, que en este tipo de CN se hace explícita por medio de la nominalización. Por ejemplo, in *sunshine*, the sun *shines* (Warren 1978: 58). A nuestro parecer, el punto débil de su enfoque es la falta de sistematicidad de la que la autora hace gala, pues pretende establecer un inventario fijo de las combinaciones semánticas que se dan en los CN. Sin embargo, encontramos muchos CN que no responden a la norma y otras normas que presentan muchas acepciones.

Posteriormente, Štekauer (1998, 2005) estudia la formación de palabras en inglés desde una perspectiva semántica, en la que utiliza categorías y roles semánticos. Se detiene en los compuestos, de los que destaca su función denominativa y dentro de los cuales presta especial atención a los compuestos exocéntricos (véase la Sección 3.3). En Štekauer et al. (2012), se interesan no solo por la formación de palabras en inglés, sino también en otras lenguas menos estudiadas como el bretón, el georgiano o el hausa, por lo que constituye una gran aportación al campo de la formación léxica.

En todos los trabajos comentados hasta ahora se abordan CN pertenecientes a la lengua general, mientras que los CN del discurso especializado han recibido menos

atención, a pesar de ser este su hábitat natural. Entre los lingüistas que han examinado los CN de áreas especializadas, cabe destacar a Fernández Domínguez (2016), quien realiza un estudio contrastivo de la formación de términos en inglés y español pertenecientes a la industria del aceite de oliva. Entre los mecanismos de formación de términos, que analiza morfológica y semánticamente, destaca el papel de la composición, que se presenta de diferentes formas en inglés y en español. Nos parece muy interesante su enfoque, pues aborda un aspecto poco explorado, como es la dimensión interlingüística de la formación de términos. Además, incorpora cuestiones semánticas, que muchas veces se relegan en los estudios lingüísticos.

Desde la Terminología, que se centra en las unidades presentes en el lenguaje especializado, se ha contribuido también al conocimiento de los CN, fundamentalmente en lo que respecta a su formación y su traducción. Un ejemplo lo constituyen los trabajos de Sager, centrados en la terminología en general y en la formación de términos en particular. Así, Sager et al. (1980) estudian las características de la lengua inglesa de especialidad, en la que destacan la presencia de CN y su relevancia en los sistemas terminológicos. Plantean una serie de patrones semánticos genéricos que se observan en la formación de CN especializados (p. ej. WHOLE-PART > *pedestal body*) (Sager et al. 1980: 270). En concreto, destacan por su investigación de la relación semántica entre los formantes del CN desde el ámbito de la terminología, ya que la mayor parte de estos estudios se han realizado desde la Lingüística Computacional, como veremos más adelante. Abordan las características de los CN de forma superficial, lo que se justifica por la variedad de temas que tratan en el libro. No obstante, ello se suple con la diversidad de aspectos de los CN que plantean. Por lo tanto, constituye un primer enfoque a la riqueza y complejidad que presentan estas combinaciones.

En trabajos posteriores, Sager (1990, 1997) desarrolla lo presentado en Sager et al. (1980). Además, destacando también el papel de la composición, plantea la distinción entre formación primaria y secundaria de términos. La formación primaria acompaña al nacimiento de un concepto y es monolingüe, mientras que la formación secundaria surge tras la revisión de un término en la lengua de la formación primaria o bien mediante la transferencia de conocimiento a otra comunidad lingüística que carece de denominación para ese concepto (Sager 1990). En la Sección 3.4.5 desarrollaremos estas nociones, dada su relación con la neología.

Asimismo, nos parece necesario mencionar los trabajos de Montero Fleta (1995, 1996), quien compara sintagmas nominales equivalentes en inglés y español

en un campo especializado, el de la informática. La autora se centra en los aspectos formales básicos de estas combinaciones, entre las que incluye tanto CN como combinaciones libres (es decir, cualquier ocurrencia de un sustantivo modificado por otras unidades). No obstante, nos parece que este estudio comparativo habría constituido el marco ideal para comentar cuestiones de mayor calado en la traducción, como es la semántica.

Oster (2003, 2005, 2006) muestra un mayor interés por estos aspectos y, como Sager et al. (1980), profundiza en las relaciones semánticas entre los formantes de los CN especializados. Basándose en el campo de la cerámica, analiza CN unidos o separados gráficamente en alemán y español, para incluir las formas típicas de las dos lenguas. Propone un inventario de *esquemas relacionales* genéricos de estructura binaria (p. ej. *place-determined entity*), con el objetivo de plasmar la semántica de los formantes, frente a las tradicionales relaciones simples del tipo *causes, located_at*, etc. Destacamos sus contribuciones tanto por el interés de la combinación lingüística (poco estudiada y atractiva, dada la elevada frecuencia y características de los compuestos en alemán), así como por su planteamiento profundamente semántico que, a pesar de sumarse a los inventarios existentes de relaciones, aporta nuevas perspectivas a la semántica de estas combinaciones.

Los trabajos de Quiroz (2008) destacan por abordar concretamente sintagmas nominales extensos, es decir, formados por tres o más elementos, en inglés y español en el campo del genoma. Así, analiza la morfosintaxis y la semántica de estas unidades en las dos lenguas, considerando tanto CN como combinaciones libres. Como punto fuerte podemos citar su énfasis en la importancia y en el escaso estudio de estos sintagmas extensos. Sin embargo, nos parece una investigación puramente cuantitativa centrada en los patrones más habituales, que carece de una aplicación práctica aparente.

Por su parte, Maniez (2008, 2009, 2013, 2014) ha explorado específicamente los términos compuestos, como los CN y los adjetivos compuestos (p. ej. *HIV-infected*). Se centra principalmente en el discurso médico y sus aportaciones resultan de gran interés porque, como Quiroz (2008), pasa de la frontera de los CN formados por dos términos para trabajar con CN más extensos. Asimismo, destaca por ser uno de los principales autores en dedicarse a los CN formados por adjetivos (Maniez 2009, 2014). En su investigación deja de lado cuestiones semánticas para centrarse en la búsqueda y comparación de equivalentes de estos términos compuestos, fundamentalmente entre el inglés y el francés, aunque en algunos trabajos también incorpora el español y el italiano

(p. ej. Maniez 2014). Por lo tanto, es uno de los pocos estudiosos que se centra en profundidad en la traducción de los CN. En este sentido, es un defensor del uso de corpus y textos paralelos en internet para este fin (Maniez 2008, 2013), aspectos que compartimos en nuestro análisis de equivalentes en inglés y español.

También desde la Terminología, Sanz Vicente (2011, 2012a, 2012b) es una de las escasas autoras que abordan los CN especializados en inglés y español. Basándose en un corpus de teledetección, parte de los CN en inglés para estudiar su influencia en la formación secundaria de términos en español, por lo que concede especial importancia a la neología en la traducción. Nuestra tesis guarda similitudes con los trabajos de Sanz Vicente, ya que ambas nos centramos en el estudio de CN especializados en inglés y su traducción al español. No obstante, son muchas las diferencias existentes, que se derivan de los distintos enfoques adoptados y de la finalidad de cada estudio.

Por una parte, como hemos mencionado, Sanz Vicente (2011, 2012a, 2012b) se interesa en la influencia de la lengua inglesa en la formación secundaria de términos en español. Por ello, hace hincapié en la morfosintaxis de los equivalentes. Sin embargo, la propia autora señala que la extracción de información semántica sobre los CN no ha sido exhaustiva, ya que se ha dado preferencia a la asignación de equivalentes en español (Sanz Vicente 2011: 456). Así, su análisis semántico de los CN se limita al empleo de relaciones semánticas del tipo PROCESO-PACIENTE. Sanz Vicente (2011: 1135-1137) propone una base de datos bilingüe que consiste en un listado de los términos, partiendo del inglés o del español, con el principal equivalente, ya que cada entrada se corresponde con un término y no con un concepto. Además, en línea con su investigación, incluye la estructura morfosintáctica de los términos, los procedimientos de importación al español (p. ej. calco de expresión), algunos términos relacionados (hiperónimos, hipónimos, cohipónimos y antónimos) y los CN más extensos que se originan a partir de otro CN inferior, un aspecto que nos parece muy interesante.

Por otra parte, las líneas principales de nuestra tesis son la formación, traducción y representación de los CN en inglés y español, de manera que encontramos en la semántica un aspecto fundamental de cara al análisis de estos aspectos. Así, nos centramos más en la semántica que en la morfosintaxis y utilizamos, además de relaciones, categorías y roles semánticos. Igualmente, nos detenemos en el carácter hiponímico y multidimensional de los CN, así como en los microcontextos, que justifican la formación de CN. Todo ello desemboca en nuestra propuesta de representación de CN en el módulo fraseológico de EcoLexicon. Como veremos en la Sección 5.3, este cuenta con diferentes vistas entre las

que los usuarios pueden elegir. En ellas se ofrecen definiciones, todos los sinónimos y equivalentes recopilados (ya que nuestras entradas representan conceptos y no términos), distintos aspectos semánticos (relaciones, categorías, roles, etc.), ejemplos de uso, etc. La información de las entradas está organizada conceptualmente y se puede acceder a ellas mediante búsquedas personalizadas e hipervínculos que agilizan el movimiento en la base de datos. Por tanto, la investigación de Sanz Vicente y la que desarrollamos en esta tesis constituyen enfoques diferentes del estudio de los CN en inglés y español.

Como comentábamos, también desde la Lexicografía se ha mostrado interés en los CN. Así, sin apartarnos de la perspectiva terminológica en la que venimos ahondando, encontramos trabajos como los de Meyer y Mackintosh (1994, 1996), Montero Martínez (2002, 2008) o Pecman (2012, 2014) que estudian los CN junto a las colocaciones con vistas a su representación en recursos especializados.

Meyer y Mackintosh (1994, 1996) parten del dominio de la informática para explorar la utilidad de los frasemas terminológicos (la denominación con la que aúnan CN y colocaciones) en el análisis conceptual. Su finalidad es incluir información en lo que Meyer et al. (1992) llaman *bases de conocimiento terminológico*, que destacan por su riqueza conceptual. A pesar de que las autoras utilizan términos en inglés, su trabajo es un punto de partida muy interesante para explorar formas de análisis conceptual en otras lenguas.

En efecto, Montero Martínez (2002, 2008) sigue esta línea y también defiende la utilidad de CN y colocaciones en el análisis conceptual, con vistas a la alimentación de bases de conocimiento terminológico. Para ello, analiza estas combinaciones en inglés y español en el dominio de la Oncología (Montero Martínez 2002, Montero Martínez et al. 2002) y, más tarde, en el del medio ambiente (Montero Martínez 2008; Montero Martínez y Buendía Castro 2012, 2017). En su investigación merece especial atención la estructuración conceptual de estos frasemas terminológicos, que hemos comentado en la Sección 3.1.2. En este sentido, distingue entre construcciones relacionales y atributivas, una propuesta que adoptamos en nuestro análisis de los CN. Los trabajos de Montero Martínez se encuadran en la Terminología Basada en Marcos y plantean una sólida base semántica que, a nuestro parecer, es la clave para un estudio completo de los CN, pues esta subyace en su formación, su traducción y su representación. Por lo tanto, en esta tesis, que también se inserta en la Terminología Basada en Marcos, asumimos muchas de las ideas de esta autora, que desarrollamos en profundidad aplicadas a los CN.

Pecman (2012, 2014) investiga la fraseología especializada y la formación de términos en el dominio de las ciencias de la Tierra. Como mecanismo fundamental de formación terminológica, los CN ocupan un lugar central en sus trabajos, en los que analiza términos en inglés y en francés para luego aplicar sus resultados a la base de datos terminológica ARTES (Kübler y Pecman 2012). Uno de los ejes fundamentales de su investigación es la variación terminológica, que analiza en CN del tipo *permeability distribution* vs. *distribution of permeability* (Pecman 2012, 2014; Gledhill y Pecman 2018). En concreto, se centra en los aspectos cognitivos y retóricos que subyacen a esa variación terminológica, un aspecto muy prometedor que no abordamos en profundidad en esta tesis pero que, sin duda, es de vital importancia.

Por otra parte, la Lingüística Computacional ha sido un área especialmente fructífera en la investigación sobre CN. No resulta extraño si consideramos que los CN son un tipo de *multiword expressions* o expresiones poliléxicas, que atraen el interés de los investigadores de esta disciplina por la complejidad que plantean a los sistemas de procesamiento del lenguaje natural. En concreto, los CN presentan complicaciones a dos niveles: (i) su detección y extracción y (ii) su tratamiento automático, que pasa necesariamente por su análisis sintáctico (donde el *bracketing* suele ser el aspecto más problemático) y semántico (con la identificación de la relación semántica interna como tema predilecto en las investigaciones). Se suelen estudiar principalmente CN formados por dos sustantivos y pertenecientes a la lengua general, con la biomedicina como dominio privilegiado cuando se utilizan textos especializados. A pesar de los numerosos trabajos que abordan las expresiones poliléxicas, aquí solo comentaremos aquellos que guardan una mayor relación con esta tesis.

Así pues, hace cuarenta años, Finin (1980) se centró en la interpretación de CN en inglés, con diferente número de formantes y pertenecientes a distintos tipos de discurso. En su búsqueda por la precisión semántica, propuso el empleo de verbos específicos para plasmar la relación semántica entre los formantes de un CN, siendo así precursor de las paráfrasis verbales que más tarde plantearían Nakov y Hearst (2006, 2008) y que utilizamos en esta tesis. Además, utiliza un enfoque basado en marcos para representar conceptos y sus relaciones, y defiende la herencia de propiedades a medida que se desciende en la jerarquía conceptual (Finin 1980: 310). Por tanto, consideramos que las aportaciones de Finin (1980) constituyen un importante esfuerzo inicial hacia la interpretación sistemática de los CN.

Por su parte, Rosario (2005) subraya la importancia de los CN entre las expresiones poliléxicas, pues es en estas combinaciones donde reside el conocimiento (Rosario et al. 2002: 247). De este modo, se centra en la interpretación de CN en inglés formados por dos sustantivos y correspondientes al área de la biomedicina. Destaca por su estudio de las relaciones semánticas que se dan en los CN, con vistas a la clasificación automática de estas unidades, y por su propuesta de relaciones semánticas, que constituye uno de los pocos inventarios que incorpora relaciones específicas[13] de un dominio especializado, p. ej. en *migraine patient*, patient *is a person afflicted by a disease* (migraine) (Rosario et al. 2002). En nuestro estudio compartimos la opinión de Rosario y, además de emplear relaciones generales, consideramos que en algunos casos es necesario utilizar o acuñar relaciones específicas de cada dominio (p. ej. *uses_resource* en el campo de las energías).

También con miras a la clasificación automática de los CN, Rosario et al. (2002) plantean su propuesta de *Descent of Hierarchy*, en la que estudian si los mismos pares de categorías semánticas codifican la misma relación en los CN. Esta noción nos parece primordial, pues consideramos que subyace en el proceso cognitivo de interpretación de los CN. Así, si en *steel knife*, knife *is_made_of* steel, es muy probable que en *plastic knife*, knife *is_made_of* plastic (Rosario et al. 2002: 248). Por eso, nos inspiramos en sus ideas cuando hablamos de inferencia de relaciones semánticas en los CN. De igual forma, estos autores defienden la «estructura argumental» del núcleo de los CN, una noción que subyace en nuestra propuesta de microcontextos, que presentamos en la Sección 5.1.2. Por consiguiente, las aportaciones desde el plano computacional de Rosario et al. (2002) incorporan aspectos cognitivos de los que pueden nutrirse también áreas como la Terminología o la Traducción, que son en las que se enmarca este estudio.

Un ejemplo de interrelación entre la Lingüística Computacional, la Terminología y la Traducción lo encontramos en Daille (1999, 2001, 2017), quien aplica técnicas computacionales para investigar la traducción (Daille et al. 2004) y, sobre todo, la variación (Daille 1999, 2001, 2017) de los términos compuestos. Analiza CN del francés formados por dos o más elementos, principalmente en el dominio de la agricultura. Entre los formantes de los CN dedica especial atención a los modificadores adjetivales, que suponen una fuente importante de variación (Daille 1999, 2001, 2017). En Daille (2017) encontramos su trabajo más exhaustivo sobre la variación en los términos compuestos, en

[13] En el ámbito de la Terminología, Montero Martínez (2002) y Montero Martínez et al. (2002) también utilizan relaciones semánticas específicas del dominio médico, en concreto del subdominio de la Oncología (p. ej. *has-risk-factor* o *has-sympton*) con vistas a la representación en recursos especializados.

el que utiliza corpus de diferentes temáticas especializadas en distintas lenguas (entre ellas, el español). No obstante, a pesar de la minuciosidad del análisis, echamos en falta una visión más conceptual de la variación. Sin embargo, la propia autora reconoce que el análisis semántico que realiza es muy limitado, debido al uso de criterios puramente formales para la identificación de variantes (Daille 2017: 2). En cualquier caso, su posición nos parece especialmente interesante, pues aúna el conocimiento de la Terminología y la Traducción, enriquecido desde una perspectiva computacional.

Nakov (2008, 2013) desarrolla uno de los enfoques más completos de los CN. Analiza estas combinaciones de dos o más sustantivos en inglés general desde la óptica computacional. Su interés por la semántica interna de los CN le llevó a plantear, junto a Hearst e inspirado por Finin (1980), el que nos parece el método más acertado para alcanzar y desgranar la relación semántica interna de los CN: las paráfrasis verbales (Nakov y Hearst 2006, 2008; Nakov 2008), que veremos con más detalle en la Sección 3.4.3.2. Estas consisten en la caracterización de la semántica de los CN por medio de varios verbos. Por ejemplo, *shoe factory* puede entenderse como *a factory where shoes are produced/made/manufactured* (Nakov y Hearst 2006: 7).

Más tarde, Nakov (2013) realiza su estudio más detallado de los CN, en el que aborda los aspectos sintácticos y semánticos de estas unidades y su interés en los sistemas de procesamiento del lenguaje natural. A pesar de que no adoptamos esta perspectiva computacional, sus ideas nos parecen muy certeras y han resultado muy útiles en nuestra investigación.

Nos gustaría destacar también las aportaciones de Parra Escartín et al. (2013, 2018) quienes, a pesar de no estudiar específicamente los CN, han realizado algunos de los escasos trabajos existentes sobre las expresiones poliléxicas en español desde la Lingüística Computacional. Principalmente se centran en la traducción y la representación de expresiones poliléxicas en recursos generales y especializados (Parra Escartín et al. 2013) y en la tipología de expresiones poliléxicas en español (Parra Escartín et al. 2018).

Otra de las disciplinas que ha dedicado especial atención a los CN ha sido la Psicolingüística, que se ha interesado especialmente en la combinación conceptual que subyace en estas unidades, así como en la faceta cognitiva de su interpretación y producción. Fue desde esta disciplina desde la que, en la década de los 90, se generaron las principales teorías de combinación conceptual, centradas en los CN (Murphy 1988, 1990; Wisniewski 1996; Gagné y Shoben 1997), que retomaremos con mayor detalle en la Sección 3.3.

Así pues, una de las contribuciones más significativas realizadas desde la Psicolingüística ha sido la apertura de *slots* por parte del núcleo. En este sentido, destaca la figura de Murphy (1988, 1990), quien defiende que el modificador rellena uno de los *slots* abiertos por el núcleo (p. ej. [HABITAT] *dog* > *appartment dog*). Murphy (1988, 1990) estudia sintagmas nominales del inglés general en los que interviene un sustantivo modificado por otro sustantivo o adjetivo. Es decir, no siempre se trata de CN como los entendemos en este trabajo, con una cierta fijación y frecuencia, sino que muchos constituyen combinaciones libres de la lengua. Además, se interesa por el papel del modificador en el procesamiento de estas combinaciones. Así, defiende la mayor carga conceptual de los sustantivos y adjetivos relacionales frente a los adjetivos calificativos. Por tanto, los sintagmas nominales que cuenten con un sustantivo o un adjetivo relacional requerirán un mayor tiempo de procesamiento (Murphy 1988: 539; 1990: 259). Este estudio de los modificadores de carácter nominal o adjetival nos parece muy atractivo, ya que se enfoca desde una perspectiva menos común en terminología y traducción: su carga conceptual e influencia en el procesamiento de los CN.

Por otro lado, Wisniewski (1996, 1997) se centra en sintagmas nominales formados por dos sustantivos en inglés que crea de manera aleatoria, combinando categorías semánticas para analizar su semántica interna. Propone que los formantes de un CN pueden estar unidos por una relación (p. ej. en *robin snake*, snake *eats* robins) o una propiedad (p. ej. en *robin snake*, snake *has a red underbelly*) (Wisniewski 1996: 434).

Gagné (2000, 2001) también destaca por sus contribuciones a la combinación conceptual. Para ello, se basa en sintagmas nominales del inglés general formados por dos sustantivos, que pueden constituir CN o combinaciones libres. No se adhiere a la apertura de *slots* defendida por Murphy (1988, 1990), sino que propone que los CN se forman mediante el establecimiento de una relación semántica que vincula los dos conceptos (Gagné y Shoben 1997; Gagné 2000, 2001). Por ejemplo, *a snowman is a man* MADE OF *snow*. A pesar de coincidir con Wisniewski (1996, 1997) en la existencia de relaciones semánticas entre los formantes de un concepto combinado, no comparte la relevancia que este atribuye a las propiedades y se centra únicamente en las relaciones. A nuestro parecer, la distinción entre relaciones y propiedades en los CN es oportuna, ya que la existencia de una u otra determinará el análisis semántico de estos términos compuestos.

A partir de los estudios comentados, podemos concluir que los CN se han investigado sobre todo en la lengua general, especialmente aquellos formados por dos unidades, en su mayoría, sustantivos. El inglés ha sido la lengua por excelencia, mientras

que los CN del español han recibido mucha menos atención. Por otra parte, las teorías que explican la combinación conceptual, que supone la base de la formación de estas unidades, han obviado estos procesos cognitivos y comunicativos en el lenguaje especializado. No obstante, la combinación conceptual cobra especial relevancia en este tipo de discurso, como elemento fundamental de transmisión del conocimiento.

3.3 Tipología y formación de compuestos nominales

Los CN se dividen en dos grandes grupos: los CN endocéntricos y los CN exocéntricos (Levi 1978; Štekauer 1998; Nakov 2013). Existe, además, un tercer grupo minoritario: los CN coordinados (Warren 1978; Fernández Domínguez 2019). Esta distinción se basa en el núcleo de estas combinaciones, como veremos a continuación.

En los CN endocéntricos, uno de los formantes es el núcleo y, el otro, el modificador, que añade características al núcleo (Nakov 2013). Por ejemplo, en *wind resource*, el núcleo es *resource* y el modificador es *wind*. Aunque el CN conste de más de dos formantes, estos pueden reducirse al esquema de núcleo y modificador, como vemos en *offshore wind farm*, donde *offshore* modifica al núcleo *wind farm*. Los CN endocéntricos representan nuestra tendencia a ver los conceptos iguales a otros existentes, pero a la vez diferentes de estos (Marchand 1969). De esta manera, la estructura binaria de los CN endocéntricos representa la semejanza con otro concepto por medio del núcleo, que indica la pertenencia a una categoría, a la vez que especifica las diferencias con este por medio del modificador, que señala el criterio de subdivisión (Warren 1978; Bowker 1998; Gagné 2000; Collet 2003; Maguire et al. 2010a). Así, el nuevo concepto resultante de la combinación constituye un hipónimo del núcleo (Downing 1977; Sager 1990, 1997; Štekauer 1998). Por ejemplo, *offshore wind farm* es un tipo de *wind farm*. Los CN endocéntricos son, con diferencia, los más habituales en inglés y español, fundamentalmente en el discurso especializado, ya que son uno de los mecanismos principales para la formación de términos y la construcción de sistemas terminológicos, un aspecto que ha abordado en profundidad Sager (1990, 1997).

Por otro lado, los CN exocéntricos[14] son menos frecuentes en la lengua general y menos aún en el discurso especializado. Se caracterizan porque carecen de núcleo (Levi 1978; Warren 1978; Selkirk 1982; Štekauer 1998; Bauer 2008; Nakov 2013), como vemos en *saber tooth*, donde ninguno de los formantes designa la categoría a la que pertenece el CN (no se trata de un diente ni de un sable, sino de un felino). Ello les confiere un carácter

[14] Autores como Marchand (1969) o Selkirk (1982) realizan subdivisiones dentro de los CN exocéntricos, si bien no las compartimos en este trabajo al ser prácticamente inexistentes en los textos especializados. Preferimos, así, hablar en general de CN exocéntricos cuando abordemos CN que no presenten núcleo.

más idiomático y menos transparente, por lo que Nakov y Hearst (2013: 44) señalan que deben incluirse en los recursos lingüísticos. En su propuesta generativa, Levi (1978) sostiene que los CN exocéntricos provienen de CN endocéntricos que han perdido el núcleo mediante las diferentes transformaciones sufridas hasta adquirir su forma final. Así, por ejemplo, *birdrain* procede de una estructura antecedente del tipo **birdrain haver* (Levi 1978: 58). Štekauer (1998: 148), quien realiza un estudio minucioso de este tipo de CN, sigue esta línea y defiende que, en una primera fase, se formaría una estructura auxiliar de pleno significado que, a pesar de no emerger nunca, sirve de base para un proceso de reducción que será el que veremos en la superficie (p. ej. *saber tooth* + *tiger* > *saber tooth*). Esto le lleva a defender que los CN exocéntricos no existen en inglés, ya que son, en su origen, endocéntricos (Štekauer 1998: 154). Pese a que este procedimiento puede ser válido en muchos casos, no consideramos que sea aplicable a la totalidad de los CN exocéntricos o, en su caso, no creemos que se pueda inferir siempre con facilidad. Así, en ejemplos como *diente de león*, no es tarea sencilla reconocer esa estructura auxiliar.

Como decíamos, se suele hablar de una tercera categoría de compuestos: los coordinados. También llamados compuestos copulativos o *dvanda*, por su nombre en sánscrito, son poco frecuentes y se caracterizan porque no cuentan con un núcleo y otro elemento que lo modifique, sino que los dos formantes se encuentran al mismo nivel (Warren 1978; Nakov 2013; Fernández Domínguez 2019). Por ello, existe una mayor compactación, que a veces se consigue formando un CN unido ortográficamente (*girlfriend*) o bien utilizando un guion (*secretary-treasurer*).

Una vez abordados los diferentes tipos de CN, profundizaremos en las combinaciones que nos interesan en este trabajo: los CN endocéntricos sintagmáticos. Como hemos comentado, los CN endocéntricos poseen un núcleo nominal y están modificados por una o más unidades. La categoría gramatical y la posición de estos modificadores están determinadas por los patrones de formación de términos de cada lengua (Levi 1978; Sager 1997; Fernández Domínguez 2016). Así, en inglés los CN suelen formarse por premodificación (Nakov 2013), es decir, el núcleo se sitúa a la derecha, precedido por el modificador. Este suele ser un sustantivo (*speed ratio*), un adjetivo (*reactive power*) o un participio con valor adjetival (*rated power*). No obstante, las estructuras de postmodificación también están presentes en inglés, aunque en menor medida. En este caso, el núcleo se encuentra a la izquierda, seguido de un sintagma preposicional que lo modifica (*angle of attack*). A menudo se ha negado el carácter de compuesto de estas estructuras; sin embargo, dado que también están formadas por

varios lexemas con un núcleo nominal y desempeñan una función denominativa, en este trabajo las consideramos CN.

Frente a la yuxtaposición típica de las lenguas germánicas, como el inglés, en las lenguas romances encontramos estructuras más extensas que recurren a preposiciones para enlazar los sustantivos del CN, ya que la compactación de nombres resulta poco habitual (Sager 1997; Fernández Domínguez 2016; Cordeiro et al. 2016; Parra Escartín et al. 2018). Así, los CN en español se suelen formar por postmodificación, en la que distinguimos la postmodificación adjetival (*potencia instalada*) y la postmodificación preposicional que acabamos de comentar (*hueco de tensión*). Existen también casos de premodificación adjetival (*alta frecuencia*), aunque son poco frecuentes. A pesar de las diferencias estructurales entre el inglés y el español, hablamos de la misma realidad: la combinación conceptual tan frecuente en la lengua especializada. Prueba de ello son las cifras similares de N+N y N+SPrep obtenidas por Fernández Domínguez (2016: 67) en un corpus de inglés y español, respectivamente. No obstante, las distintas configuraciones que adopta la combinación conceptual en inglés y español son una de las causas de los problemas de traducción de los CN, como veremos en la Sección 3.5.

En cuanto al tipo de formantes, queremos recalcar la presencia de modificadores que no son sustantivos, como se ha venido afirmando tradicionalmente. Así, los CN en inglés y español cuentan frecuentemente con modificadores de tipo adjetival, entre los que se encuentran también adjetivos compuestos formados por un participio y su complemento (*área barrida por la pala*). Los adjetivos que pueden modificar a los núcleos nominales en los CN pueden ser de dos tipos: relacionales o atributivos.

Los adjetivos relacionales son los que representan un concepto y, por tanto, pueden establecer relaciones semánticas con otros elementos (*infección ocular* > infección que *afecta al* ojo). Estos adjetivos han recibido diferentes nombres en la literatura, como *relational adjectives* (Bally 1965; Maniez 2009; Daille 1999, 2001, 2017), que nos parece la designación más descriptiva; *pseudo adjectives* (Postal 1969); o *nominal non predicating adjectives* (Levi 1978). Sin embargo, los estudios coinciden en su carácter denominal, ya que suelen proceder de sustantivos o raíces griegas o latinas a las que se añade un sufijo (Levi 1978; Daille 1999, 2001, 2017; Maniez 2009).

De tal modo, para formar el adjetivo relacional, se pueden dar modificaciones en el sustantivo de base, como la modificación fonológica o gráfica (*trópico* > *tropical*), la adición de sílabas o vocales (*nombre* > *nominal*) o la modificación de la raíz con su etimología clásica (*corazón* > *cardiaco*) (Harastani et al. 2013: 315). En cuanto al sufijo,

Levi (1978) defiende que este no puede tener contenido semántico (*blood test* > **bloody test*) (Levi 1978: 152). Por el contrario, Maniez (2009: 118) señala que, en ocasiones, es precisamente gracias al sufijo del adjetivo que podemos obtener con mayor facilidad la relación semántica entre N+Adj, lo que ilustra con el ejemplo francés de *lésion fracturaire*, en el que el adjetivo implica el resultado de una fractura. A nuestro parecer, los CN con adjetivos relacionales sí pueden incluir adjetivos con sufijos cargados semánticamente, siempre que estos contribuyan a la semántica general del CN. Por ejemplo, *floral* incluye el sufijo *–al*, que indica abundancia o relación con algo, por lo que resulta adecuado en el CN *ofrenda floral* (ofrenda consistente en un conjunto de flores). En algunos casos, también se pueden añadir prefijos que expliciten el vínculo entre el núcleo y el modificador (p. ej. *tratamiento antideslizante*).

Como vemos en los ejemplos, muchos de estos adjetivos relacionales son cultismos que proceden del griego o el latín, por lo que elevan el nivel de especialización del término (p. ej. *enfermedad hepática* vs. *enfermedad del hígado*). Por ello, se dan más en las lenguas de origen latino que en las de procedencia germánica, y no resulta extraña su importancia en los campos de especialidad (Daille 1999, 2001, 2017; Maniez 2009).

Es preciso señalar que, a menudo, la designación de *relacional* o *denominal* se usa indistintamente. Sin embargo, conviene señalar las diferencias, ya que constituyen conceptos diferentes. Por una parte, los adjetivos denominales son aquellos que proceden de sustantivos (*capacidad* > *capaz*). Por otra, como hemos señalado, los adjetivos relacionales son los que pueden establecer relaciones semánticas con otros elementos (*contaminación marina* > contaminación que *afecta al* mar). Normalmente, los adjetivos relacionales son denominales, aunque no siempre es así, ya que existen también adjetivos relacionales que provienen de adverbios (*sexual attack* > *x was sexually attacked*) (Levi 1978: 252) o de verbos (*sustancia contaminante* > *sustancia que contamina*). Por tanto, no podemos olvidar que todos los adjetivos relacionales no son denominales y viceversa. Por ejemplo, en *error garrafal*, el adjetivo procede del sustantivo *garrafa* (es, por tanto, denominal), pero no existe relación semántica entre los formantes (es decir, no es un adjetivo relacional), ya que *garrafal* alude a la intensidad del error en lugar de al sustantivo subyacente.

A menudo se ha señalado que el significado de los adjetivos relacionales cambia según el sustantivo al que determinen, como vemos en *musical clock* (un reloj que produce música) o *musical criticism* (una crítica acerca de música) (Levi 1978; Kocourek 1982; Fernández Domínguez 2019). En nuestra opinión, lo que cambia no es el significado del

adjetivo, sino la relación semántica entre los formantes del CN. Esta información se puede obtener en contexto y mediante el análisis de la apertura de *slots* del núcleo, ya que el significado del núcleo determinará las relaciones que puede establecer con otros elementos (Sección 5.1.2).

Los adjetivos relacionales cuentan con las siguientes características, que permiten distinguirlos de los de tipo atributivo, que comentaremos a continuación:

a) Su ubicación en posición predicativa no suele ser posible (Levi 1978; Daille 1999, 2001; Estopà 2000), p. ej. *demanda hídrica* > **la demanda es hídrica*. Daille (1999, 2001) y Estopà (2000) añaden que, en las lenguas romances, estos adjetivos no pueden preceder al sustantivo (p. ej. **hídrica demanda*). Sin embargo, compartimos la opinión de Maniez (2009), quien señala que el reconocimiento de los adjetivos relacionales no puede basarse únicamente en este criterio, ya que algunos sí pueden aparecer detrás del verbo, p. ej. *coche eléctrico* > *el coche es eléctrico*.

b) No se pueden modificar con adverbios de grado (Lees 1960; Levi 1978; Daille 1999, 2001; Estopà 2000; Maniez 2009), p. ej. **muy hídrico.*

c) Se pueden coordinar con sustantivos y otros adjetivos relacionales, pero no con adjetivos calificativos (Levi 1978; Estopà 2000), p. ej. *recurso terrestre y marino* pero no **demanda energética e importante.* Por el contrario, Daille (1999, 2001) también rechaza la coordinación con otros adjetivos relacionales, aunque en nuestro corpus encontramos muchos ejemplos de este tipo (p. ej. *suministro eléctrico y energético*).

d) Algunos adjetivos relacionales son contables, como los sustantivos de los que provienen (Levi 1978), p. ej. *trifásico*.

e) Se les pueden asignar roles semánticos (Levi 1978). Por ejemplo, en *impacto ambiental, ambiental* es el PACIENTE. Del mismo modo, consideramos que se les pueden asignar categorías semánticas (*ambiental* es una CARACTERÍSTICA GEOGRÁFICA NATURAL).

Por otro lado, Levi (1978) señala que los adjetivos relacionales no permiten la formación de nuevos sustantivos. Sin embargo, encontramos ejemplos como *ambiental* > *ambientalización*. Por último, Daille (1999, 2001) y Estopà (2000) defienden que, cuando el núcleo está modificado por varios adjetivos, el más cercano al núcleo es relacional, p. ej. *energético* en *sector energético emergente*. Esto se debe a que es necesario determinar (es decir, nombrar el concepto) antes de calificar (Harastani et al. 2013: 316).

Como decimos, los CN también pueden incluir adjetivos atributivos (*eje horizontal*). Estos también han recibido distintas denominaciones: *bona fide adjectives* (Lees 1960), *true adjectives* (Levi 1978), *predicative adjectives* (Murphy 1988, 1990), *epithetic adjectives* (Daille 2001), *adjectifs qualifiants* (Maniez 2009) o *calificativos* (denominación más extendida), *qualitative adjectives* (Fernández Domínguez 2019) o *attributive adjectives* (Fradin 2017), como preferimos denominarlos en esta tesis debido a su función de atributos, pues añaden una propiedad al núcleo. Como hemos comentado en el ejemplo de *error garrafal*, pueden ser denominales, pero no serán relacionales.

En este sentido, Wisniewski y Love (1998) señalan que las combinaciones atributivas no presentan mayor complicación que las relacionales. Nosotros, como Gagné (2000), consideramos que las combinaciones en las que subyacen propiedades son más difíciles de interpretar que las que cuentan con relaciones, ya que el vínculo interno es más difícil de percibir, tanto en el CN como en el sistema conceptual subyacente. Esto justifica también una mayor dificultad de los CN que cuentan con adjetivos atributivos frente a los que son relacionales.

A pesar de la elevada presencia de adjetivos en los CN y su importancia a distintos niveles, como en la formación o la variación de los CN (Sección 3.4.4), la mayoría de los estudios han obviado los CN que incluyen categorías gramaticales diferentes al sustantivo. Esta nos parece una visión restrictiva porque no comprende todo el espectro de CN, oscureciendo así muchos rasgos de estos términos compuestos. Uno de los principales estudios en destacar la relevancia de los adjetivos en los CN fue el de Levi (1978), quien estudia los CN de la lengua general prestando especial atención a los que incluyen *nominal nonpredicating adjectives* (p. ej. *eye* > *ocular*), es decir, adjetivos que provienen de sustantivos y establecen relaciones semánticas. Esta autora realizó importantes contribuciones al tema y, sobre todo, reconoció la presencia e importancia de CN con adjetivos. Sin embargo, esta tesis diverge en varios puntos de su análisis.

Por una parte, no coincidimos en su visión sintáctica de la derivación de estos adjetivos (*tidal wave* proviene de la secuencia *wave be caused by tide* > *wave be tide-caused*) (Levi 1978: 131). Como hemos indicado en diferentes ocasiones, preferimos no centrarnos en elaborar este tipo de estructuras sintácticas originarias del CN, sino en el significado de los formantes y lo que estos aportan a la formación del CN. Este tipo de configuraciones sintácticas no presentan un único modo de análisis, sino que pueden adquirir muchas formas, mientras que el fondo semántico será siempre el mismo. Por otro lado, no compartimos su idea de que los adjetivos denominales y relacionales son los

únicos que pueden formar CN. Por ejemplo, según el enfoque de Levi (1978), *large wind turbine* no sería un CN, ya que el adjetivo no procede de un sustantivo ni establece una relación semántica con el núcleo. En nuestra opinión, constituye un término compuesto, pues, además de relaciones, también pueden establecerse atributos en los CN, como es el caso aquí. Del mismo modo, se podrían incluir adjetivos sin necesidad de que estos provengan de sustantivos. Además, consideramos que *large wind turbine* es un hipónimo de *wind turbine*, ya que *large* indica una particularidad que lo diferencia de su hiperónimo, y cuenta con términos opuestos que resaltan su carácter conceptual, como se observa en la taxonomía *small/medium/large wind turbine*, según la potencia del aerogenerador.

En lo que respecta a la terminología, el escaso interés recibido por los adjetivos en los CN es consecuencia del papel marginal que han desempeñado tradicionalmente los adjetivos, a menudo entendidos como simples modificadores y no como transmisores de conocimiento (Durán Muñoz 2019). Sin embargo, su papel es fundamental en los términos compuestos, ya que, como hemos indicado, pueden atribuir una propiedad al núcleo (*intestino grueso*) o establecer su relación con otro concepto (*contaminación fluvial*), contribuyendo así a la formación del concepto combinado. Además, son una fuente importante de variación, como veremos detenidamente en la Sección 3.4.4. Por ello, estudios como los de Sager et al. (1980), Daille (2001, 2017), Maniez (2009, 2014), Estopà (2000) o Alonso y Torner (2010) destacan la presencia de adjetivos en los CN especializados.

En cuanto al número de elementos de los CN, habitualmente se han estudiado CN formados por dos unidades, si bien en el discurso científico-técnico son frecuentes los CN con más de dos formantes. No obstante, la frecuencia de los CN empieza a reducirse a medida que aumenta el número de constituyentes (Barrière y Ménard 2014). Creemos que esto se debe a la mayor dificultad de procesamiento que plantean los CN más extensos y que suele desembocar en la omisión de formantes (*wind power generation system* > *wind system*), lo que explica que los CN más largos sean menos frecuentes. Por otra parte, la existencia de CN extensos se ha justificado a menudo por la concatenación de otros CN, formando lo que se denomina *compound-within-compounds* (Warren 1978; Levi 1978; Sager et al. 1980; Sager 1997), p. ej. *horizontal axis* + *wind turbine* > *horizontal-axis wind turbine*. Como Warren (1978), consideramos que su estructura se puede reducir a los dos formantes básicos: núcleo y modificador. Para ello, es necesario establecer agrupaciones internas, en un proceso que se conoce como *bracketing* (*wind* [*power plant*]) y que comentaremos en detalle en la Sección 3.4.1. No obstante, a pesar de que algunos trabajos reconocen la existencia de estas combinaciones extensas (Warren 1978; Levi 1978; Sager

et al. 1980; Sager 1997; Meyer y Mackintosh 1996; Quiroz 2008; Fernández Domínguez 2019), es necesario profundizar en este tipo de CN, principalmente en lo que concierne a su formación, su desambiguación estructural y sus implicaciones en la traducción.

Una vez que hemos conocido estos aspectos básicos de la formación de los CN, conviene señalar los criterios fundamentales para que se pueda formar un CN. La idea fundamental es que el modificador debe representar una característica distintiva del nuevo concepto (Downing 1977; Levi 1978; Warren 1978; Meyer y Mackintosh 1996). Es decir, ha de ser una característica que forme parte de las *differentiae* de ese nuevo concepto que surge de la combinación, una particularidad que lo diferencie de los otros conceptos de la categoría designada por el núcleo. De este modo se contribuye a la función clasificadora de los CN mediante la formación de hipónimos. En este sentido, el modificador no puede expresar información redundante. Levi (1978: 241) lo ilustra de manera muy clara en el siguiente ejemplo:

> It would be redundant to use a CN such as *water lake* or *heat cooker*, in view of our knowledge (or belief) that all lakes have water and all "cookers" cook by means of heat; these features are therefore predicateable from the head noun. In contrast, the familiar CNs *salt lake* and *pressure cooker* are functional CN names because lakes that have (unusual amounts of) salt and "cookers" that use not only heat but also air pressure to cook their contents are distinctive enough to constitute special subsets.

Asimismo, en el CN no se plasman aquellas características que establezcan una relación de negación, ya que estas dificultan la comprensión del concepto combinado (Zimmer 1971; Downing 1977; Levi 1978). Por ejemplo, no sería lógico hablar de *generador de electricidad* para referirse a un generador que no produce electricidad.

Por otro lado, es necesario que entre los formantes del CN exista una relación que no sea temporal, sino permanente, ya que se facilita así la interpretación por parte del receptor (Downing 1977; Levi 1978). Cuando la relación no es permanente, la estructura sintagmática no se lexicalizará (Downing 1977) y estaremos hablando de una combinación libre. Tomamos de nuevo las palabras de Levi (1978: 242) para ilustrar este concepto de temporalidad:

> If I tell you that I have seen an *aquatic bird* this morning, but I mean a chicken that had been thrown into Lake Michigan, I would be violating this principle since you would (justifiably) have assumed I meant a bird that is habitually rather than fortuitously around water, such as a duck or a seagull.

En definitiva, para que se forme un CN, es necesario que el modificador añada una característica distintiva y permanente del núcleo, que no sea redundante ni se base en la negación. Sin embargo, todo el universo semántico que se crea a partir de estos criterios de formación no resulta patente en la superficie del término compuesto, ya que la economía formal se prefiere a la claridad de expresión (Štekauer 2005). De este modo, los CN son el resultado de la condensación de información (Levi 1978; Kocourek 1982), que se consigue, principalmente, mediante la omisión de la relación semántica, pero también mediante la transformación de categorías gramaticales o la elisión de formantes de la que ya hemos hablado. Todo ello nos lleva a la idea central desde la que entendemos la formación de CN: la existencia previa de proposiciones que se condensan para dar lugar a los CN (Levi 1978; Sager et al. 1980). Esta condensación contribuye a la impersonalidad y la cohesión del texto (Kocourek 1982).

La formación de CN a partir de proposiciones subyacentes se ha defendido en trabajos anteriores, como los pertenecientes a la lingüística generativa. En esta corriente destacan los estudios de Lees (1960, 1970) (véase la Sección 3.2) y Levi (1978), quien propuso una de las teorías más influyentes sobre la formación de CN. Esta autora expone que los CN surgen a partir de la omisión o la nominalización del predicado subyacente. Así, defiende que en CN como *viral disease* se ha suprimido el verbo (virus *cause* disease), mientras que en otros como *financial analyst*, se ha nominalizado (*analyze* > *analyst*).

Levi (1978: 77) propone que, en los casos de omisión, el predicado[15] elidido es uno de los siguientes: *cause*, *have*, *make*, *be*, *use*, *for*, *in*, *about* y *from*. A estos los denomina *recoverably deletable predicates*, que abordaremos con mayor detalle en la Sección 3.4.3. Según Levi (1978: 154), cuando el predicado subyacente no pertenece a esta lista, permanece visible en la forma del CN (p. ej. *student-dominated group*). Además, cuando se omite el predicado, el modificador puede representar el sujeto o el complemento directo del verbo elidido.

Cuando se da la nominalización[16] del predicado, este figura como sustantivo en el núcleo del CN (aunque consideramos que el modificador también puede ser un verbo nominalizado). Según Levi (1978: 167), el modificador puede derivarse del sujeto o el complemento directo, aunque también hay CN con varios modificadores que cumplen ambas funciones, p. ej. *industrial water pollution*. La autora diferencia cuatro tipos de

[15] Como vemos en la lista de predicados, estos no siempre constituyen verbos, sino que se entienden más bien como relaciones semánticas internas.
[16] Importantes aportaciones respecto de la nominalización son las realizadas por Chomsky (1970) y Halliday (1994, como parte de su noción de *metáfora gramatical*).

nominalizaciones: las nominalizaciones de acción (*dream analysis*), de producto (*oil import*), de agente (*city planner*) y de paciente (*designer creation*) (Levi 1978: 168). Como señalan Evans y Green (2006) desde una perspectiva cognitiva, las nominalizaciones son el claro ejemplo de que los límites entre las categorías gramaticales son difusos.

La propuesta de Levi (1978) es la más extendida, si bien existen contribuciones similares, basadas en la presencia explícita o no del verbo. En esta línea, autores como Marchand (1969), Warren (1978) o Štekauer (1998) han hablado de compuestos verbales (también denominados sintéticos o explícitos), en los que figura el verbo y, por tanto, se corresponden con las nominalizaciones del predicado de Levi (1978). En el lado opuesto se encuentran los compuestos no verbales, también conocidos como primarios o implícitos, en los que el verbo no aparece de manera explícita y, por tanto, están relacionados con la omisión del predicado de la que habla Levi (1978) (Marchand 1969; Warren 1978; Štekauer 1998).

En nuestra opinión, todos los compuestos son verbales, pues a pesar de que el predicado no figure en la superficie, este siempre se encuentra en el origen de los CN. Así, y como hemos mencionado anteriormente, los CN se componen de un verbo subyacente y sus argumentos. Por tanto, compartimos los mecanismos de formación de CN propuestos por Levi (1978), basados en la existencia de proposiciones en los CN. No obstante, consideramos que estas son de tipo conceptual más que sintáctico. Por ello, preferimos utilizar roles semánticos (p. ej. AGENTE, PACIENTE), pues expresan de manera más acertada la relación del verbo con sus argumentos.

A pesar de que Levi (1978) realiza también aportaciones de corte semántico, como su tipología de nominalizaciones, su propuesta se caracteriza por adoptar un enfoque sintáctico, que le ha valido diversas críticas. Estas se centran principalmente en la necesidad de incorporar aspectos semánticos a su análisis. Como señala Štekauer (1998: 2), «It is the form which implements the options given by semantics. Therefore, semantics is not a mere "tag" attached to formal patterns». Por ejemplo, desde una visión meramente sintáctica, se podría decir que, en *anthropogenic pollution*, *anthropogenic* es tanto el sujeto (*el hombre contamina*) como el complemento agente del verbo (*X contaminado por el hombre*). Por tanto, afirmar que el modificador es el sujeto del verbo no sería siempre exacto, sino que dependería de la forma oracional desde la que se inicie el análisis. En cambio, al utilizar la semántica de los formantes, podríamos decir que *anthropogenic* es el AGENTE que causa la contaminación, una afirmación cierta independientemente de la estructura subyacente desde la que se parta. Asimismo, nos parece que la propuesta de

Levi (1978) sería más consistente si las relaciones semánticas se estudiaran en todos los CN en lugar de aplicarse únicamente a los formados por supresión del predicado. Del mismo modo, los roles desempeñados por la nominalización (AGENTE, INSTRUMENTO, etc.) también podrían extenderse a los CN formados por la omisión del predicado, ya que la propia autora señala que todos los CN cuentan con un verbo subyacente.

Todas estas cuestiones destacan la relevancia de la semántica en la formación de CN. Esta debe entenderse no solo desde los mecanismos de transformación que acabamos de comentar, sino también y especialmente desde la combinación conceptual de la que resultan estos términos compuestos. Desde la Psicolingüística se desarrollaron diversas teorías que explican la combinación conceptual típica de los CN.

Murphy (1988, 1990) propone una de ellas, conocida como *Schema Modification Theory* o *Concept Specialization Theory*, que distingue dos fases en la formación de conceptos combinados. En primer lugar, el modificador rellena uno de los *slots* abiertos por el núcleo. Por ejemplo, en *apartment dog*, el modificador *apartment* rellena el *slot* de HABITAT de *dog*. De este modo, el significado del núcleo determina los *slots* que este puede abrir y, con ello, los modificadores que pueden rellenar dichos *slots* (Allen 1978; Rosario et al. 2002). Su noción de apertura de *slots* es similar a la de otros autores como Allen (1978), Martin (1992) o Rosario et al. (2002), algunos de los cuales ya hemos mencionado, y constituye la base de nuestra propuesta de microcontextos. En segundo lugar, interviene nuestro conocimiento del mundo para refinar el significado del concepto combinado. Así, identificamos a este tipo de perro como más pequeño y tranquilo que uno que viva en el campo (*farm dog*) (Murphy 1988: 532-533). Por tanto, el conocimiento del mundo cobra una especial relevancia, pues además actúa como desambiguador, por ejemplo para aclarar el significado de *corporate* en *corporate lawyer* y *corporate stationery* (Murphy 1988: 536).

Por su parte, Wisniewski (1996, 1997) propone otra teoría de combinación conceptual, denominada *Dual-Process Theory*, según la cual existen dos formas de procesamiento de los conceptos combinados: (i) por un lado, el núcleo y el modificador pueden estar unidos por una relación (p. ej. en *robin snake*, snake *eats* robins) o bien (ii) el modificador transfiere una propiedad al núcleo (p. ej. en *robin snake*, snake *has a red underbelly*) (Wisniewski 1996: 434). Por tanto, justifica la combinación conceptual por medio de relaciones y propiedades (Wisniewski 1996, 1997; Wisniewski y Love 1998). Este enfoque se encuentra en sintonía con nuestro análisis semántico de los CN, ya que también consideramos que, cuando un concepto se añade a otro, puede modificarlo por

medio de una relación semántica o bien mediante propiedades atributivas, como ocurre a menudo con los adjetivos calificativos en nuestros CN.

Gagné y Shoben (1997) consideran que las propiedades no revisten tanta importancia en los CN, sino que, por el contrario, las relaciones son la base de la formación de estas combinaciones conceptuales. Así, proponen la teoría denominada *Competition Among Relations In Nominals* (*CARIN*) para explicar el trasfondo de la combinación de conceptos. En primer lugar, se selecciona una relación que vincule los dos conceptos. Una vez que se ha considerado semánticamente apropiada, se desarrolla el sentido completo de la combinación utilizando conocimiento externo. Por ejemplo, en *snowman* identificamos primero la relación: *man* MADE OF *snow*. Después, inferimos que un muñeco de nieve es frío y no caliente (Gagné y Spalding 2013: 104), en un proceso de elaboración conceptual similar al de Murphy (1988, 1990). Por otro lado, Gagné subraya el papel del modificador como indicador de la relación, porque es este el encargado de señalar las diferencias con el hiperónimo. De esta manera, las distintas relaciones que puede establecer el modificador compiten (de ahí el nombre de la teoría) y gana la más fuerte, es decir, la que facilita en mayor medida la interpretación de la combinación (Gagné y Shoben 1997; Gagné 2000; Gagné y Spalding 2013).

Compartimos con los trabajos de Gagné la importancia de la relación semántica entre los formantes del CN. Sin embargo, consideramos que el núcleo también influye en la selección de la relación, pues esta vincula dos conceptos y, por tanto, no debe determinarse de manera unilateral. Por ejemplo, para afirmar que en *wind erosion* la relación semántica es *causes*, es necesario considerar no solo las características semánticas de *wind* sino también las de *erosion*. Solo así se puede establecer el modo en que ambos conceptos encajan. En un trabajo posterior, Spalding et al. (2010) desarrollan la teoría de Gagné y Shoben (1997) y reconocen la importancia del núcleo, así como la del modificador, en la selección de la relación.

Más tarde, Maguire et al. (2010b) analizan de manera muy interesante los patrones semánticos que se dan en los CN y su utilidad en la interpretación de CN formados por dos sustantivos en inglés general. Así, en línea con Rosario et al. (2002), sostienen que los núcleos que son similares semánticamente abren *slots* también similares, que se completan con el mismo tipo de modificadores (Maguire et al. 2010b). En otras palabras, los miembros de una categoría semántica se suelen combinar de la misma forma (p. ej. AREA+ANIMAL > ANIMAL *located_at* AREA: *water mammal, freshwater fish,* etc.), lo que tiene implicaciones sobre la inferencia de relaciones semánticas, como veremos con más detalle

en la Sección 3.4.3. De este modo, las aportaciones de Maguire et al. (2010b) son claves para nuestro enfoque de los CN, ya que encontramos en las categorías semánticas una de las principales formas de hallar patrones en los CN, aplicables a su formación y su interpretación.

Como se observa, las teorías de combinación conceptual propuestas desde la Psicolingüística han estudiado unidades de la lengua general. Llama la atención que, en Terminología, un campo que se centra en los términos y los conceptos que estos designan, el único acercamiento a la combinación conceptual haya sido la noción de *colocación conceptual* propuesta por Martin (1992) y adoptada por trabajos como los de Heid (1994, 2001), L'Homme (1998, 2000, 2002), Montero Martínez (2002, 2008) o Buendía Castro (2013). Como hemos comentado en la Sección 3.1.2, Martin (1992) defiende que los conceptos presentan una estructura argumental que puede extraerse de su definición. Estos *slots* se rellenan con otros conceptos de determinadas características semánticas. Así, aplica la idea de apertura de *slots* a las combinaciones especializadas, si bien estas no constituyen necesariamente CN. Por tanto, esta tesis se nutre de las teorías de combinación conceptual de la Psicolingüística y las colocaciones conceptuales de la Terminología, así como de los mecanismos de formación de CN de Levi (1978), con vistas a explorar la formación de CN en terminología (Sección 5.1.2).

3.4 Análisis de los compuestos nominales

El correcto análisis de los CN resulta vital para la comprensión de estas unidades y sus posteriores aplicaciones, como pueden ser la traducción o la representación terminográfica. Sin embargo, este análisis no está exento de complicaciones. Algunas de ellas resultan de la especialización de los formantes, la omisión de la relación semántica entre estos, la formación de CN extensos con tres o más constituyentes o la inestabilidad que a menudo presentan estas combinaciones. Por tanto, se debe profundizar en diversos aspectos de los CN, como son el *bracketing* o análisis de las dependencias internas, la semántica de los formantes y la relación que los vincula, así como la variación y neología que suele observarse en estos términos compuestos.

3.4.1 Desambiguación estructural: *bracketing*

Los CN formados por más de dos elementos han recibido menos atención que las combinaciones binarias. Resulta sorprendente si consideramos que presentan una particularidad: su ambigüedad estructural (Lauer 1995; Girju et al. 2005; Nakov 2013; Kim y Baldwin 2013; Barrière y Ménard 2014). Así, en CN como *permanent magnet synchronous generator*, es necesario identificar las dependencias internas entre los formantes, pues de ello dependerá el correcto análisis semántico del CN y sus posteriores

aplicaciones. En un proceso que se conoce como *bracketing*[17], se agrupan los elementos que dependen entre sí, de manera que se reduce el CN a su estructura básica de núcleo y modificador[18]. Con esta configuración reducida, podremos estudiar la relación semántica principal del CN, que se establece entre los dos formantes elementales que hemos identificado. Por tanto, si no sabemos que *permanent magnet* es el modificador de *synchronous generator*, en una estructura de *bracketing* del tipo [*permanent magnet*][*synchronous generator*], el análisis y tratamiento que hagamos del CN será probablemente desafortunado.

Se debe tener en cuenta, además, que esta desambiguación deberá realizarse en cada CN, ya que las estructuras de *bracketing* no son extrapolables a otras combinaciones que presenten la misma forma. Por ejemplo, a pesar de que *offshore wind farm* y *renewable energy technology* son esquemas similares (Adj+N+N), presentan agrupaciones diferentes: *offshore* [*wind farm*] vs. [*renewable energy*] *technology*. O lo que es más, la misma secuencia léxica puede intervenir en dos tipos distintos de *bracketing*, como sucede con *wind power* en *offshore* [*wind power*] y *wind* [*power output*].

Como vemos, el *bracketing* es un proceso complejo que requiere no solo de nociones lingüísticas, sino especialmente de conocimiento del mundo y técnicas (manuales o computacionales) para su realización. Desde la Lingüística Computacional se han llevado a cabo la mayoría de los trabajos a este respecto (Lauer 1995; Girju et al. 2005; Nakov y Hearst 2005; Vadas y Curran 2011; Utsumi 2014; Barrière y Ménard 2014), pues el *bracketing* es una de las grandes dificultades que encuentran los sistemas de procesamiento del lenguaje natural. En esta disciplina se han propuesto los dos modelos principales de *bracketing*, válidos para CN de tres formantes: el modelo de adyacencia y el modelo de dependencia.

Por una parte, el modelo de adyacencia (Marcus 1980; Pustejovsky et al. 1993) compara si, en un CN del tipo p1p2p3, p2 se relaciona más con p1 o con p3. Para ello, se comprueba si hay más ocurrencias de p1p2 o, por el contrario, de p2p3. Por ejemplo, en *renewable energy technology* hay una mayor asociación entre *renewable energy* que entre *energy technology*, por lo que podemos decir que se da una estructura de *bracketing* a la izquierda: [*renewable energy*] *technology*.

---

[17] Este análisis de dependencias se presenta como *bracketing* en trabajos como el de Marsh (1984), y se extiende ampliamente con esta denominación en estudios como los de Lauer (1995), Girju et al. (2005), Nakov (2013) o Barrière y Ménard (2014).

[18] El *bracketing* también se ha aplicado a CN con unión gráfica; por ejemplo, *bookseller* > [*book*][*sell*][*er*] (Štekauer 1998: 140).

Por otra parte, el modelo de dependencia (Lauer 1995) estudia si, en un CN del tipo p1p2p3, p1 tiene una mayor vinculación con p2 o con p3. Es decir, el análisis no parte del término central, como plantea el modelo de adyacencia, sino que lo hace desde el término inicial. Así, estaremos ante una estructura de *bracketing* a la izquierda si la asociación principal es p1p2 ([*tip speed*] *ratio*), mientras que se tratará de *bracketing* a la derecha si la dependencia se da entre p1 y p3 (*mean* [*wind speed*]). No obstante, hemos observado que, a menudo, las dos combinaciones posibles presentan una frecuencia similar. Por tanto, pensamos que la desambiguación estructural que plantean estos modelos debe incorporar otros factores.

Así, además de la frecuencia, criterio por excelencia en el *bracketing*, Nakov y Hearst (2005: 19-21) proponen otros signos que pueden aclarar la configuración de los CN en inglés. Estos se basan principalmente en la búsqueda de variantes denominativas en la web que presenten los indicativos que figuran en la tercera columna de la Tabla 1, pues estos señalan la formación de un grupo interno.

CN que se debe desambiguar	**Variante denominativa**	**Indicativo del *bracketing***	***Bracketing***
cell cycle analysis	*cell-cycle analysis*	guion	[*cell cycle*] *analysis*
brain stem cell	*brain's stem cell*	genitivo sajón	*brain* [*stem cell*]
plasmodium vivax malaria	*Plasmodium vivax Malaria*	mayúscula interna	[*plasmodium vivax*] *malaria*
leukemia lymphoma cell	*leukemia/lymphoma cell*	barra interna	*leukemia* [*lymphoma cell*][19]
growth factor beta	*growth factor (beta)*	paréntesis	[*growth factor*] *beta*
tumor necrosis factor	*tumor necrosis factor (NF)*	abreviatura	*tumor* [*necrosis factor*]
health care reform	*healthcare reform*	concatenación	[*health care*] *reform*
adult male rat	*male adult rat*	cambio de orden	*adult* [*male rat*][20]
tyrosine kinase activation	*tyrosine kinases activation*	flexión interna	[*tyrosine kinase*] *activation*

Tabla 1: Indicativos para el *bracketing* propuestos por Nakov y Hearst (2005: 19-21)

Nakov y Hearst (2005) también formulan el uso de paráfrasis para identificar las dependencias internas en los CN. Por ejemplo, *health care reform* presenta un *bracketing* a la izquierda porque encontramos paráfrasis que separan los dos grupos: *reform in health care*. Como veremos en la Sección 3.4.3.2, Nakov y Hearst (2006) desarrollan esta idea y

[19] En este caso los autores afirman que hay *bracketing* a la derecha porque los términos separados por la barra se presentan como alternativas (Nakov y Hearst 2005: 20).
[20] Los autores afirman que hay *bracketing* a la derecha porque los dos primeros términos modifican de forma independiente al núcleo.

plantean la noción de *paráfrasis verbales* como medio para conocer la relación semántica interna de los CN. Los indicativos propuestos por Nakov y Hearst (2005) nos parecen de gran utilidad para el *bracketing* de CN en inglés, pero resultan de difícil aplicación a otras lenguas, como el español, que no dispone de muchos de estos mecanismos lingüísticos (como el genitivo sajón o el tipo de flexión que se presenta).

A esta propuesta se suman las claves que proporcionan Barrière y Ménard (2014), en uno de los pocos trabajos que exploran el *bracketing* de CN formados por más de tres términos. Defienden que las dependencias internas se basan en la existencia de vínculos relacionales, coordinativos o léxicos. Su gran apuesta es el uso de Wikipedia, como lista de términos y entidades nombradas, y como corpus que aúna la información de todas sus páginas (Barrière y Ménard 2014: 72).

Así pues, Barrière y Ménard (2014) propugnan que la existencia de una relación semántica entre dos de los formantes indica su dependencia. Para determinar que estos están unidos por una relación semántica, se basan en el uso de preposiciones. Por ejemplo, buscan en el corpus *n1 for n2*. En el caso de que hubiera ocurrencias, se entiende que estos dos elementos codifican una relación semántica y, por tanto, formarían un grupo del *bracketing*. Este criterio no nos parece aplicable a los CN especializados, pues sus formantes normalmente pertenecen a un sistema terminológico y es habitual que entre ellos se den relaciones semánticas. Por tanto, en CN como *offshore wind industry*, encontramos relaciones semánticas entre todos sus formantes: industry *located* offshore, industry *uses_resource* wind y wind *located* offshore.

Por otro lado, Barrière y Ménard (2014) señalan que la existencia de elementos coordinados impide su dependencia. Examinan la presencia de coordinación entre dos elementos por medio de las conjunciones *or*, *and* y *nor*. Por ejemplo, *cotton* y *polyester* son elementos coordinados, ya que encuentran ocurrencias de ambos términos con estas conjunciones. Por tanto, en el CN *cotton polyester shirt* no habría *bracketing* entre *cotton* y *polyester*. Este ejemplo es similar al cambio de orden del que hablan Nakov y Hearst (2005), como se aprecia en *adult male rat*, que presenta un *bracketing* a la derecha (Tabla 1). A pesar de que en ambos trabajos se señala que los modificadores de estos ejemplos complementan de manera independiente al núcleo (una opinión que compartimos), consideramos que estos deben agruparse. Para ello, nos basamos en que ninguno de los modificadores presenta una vinculación mayor que el otro con el núcleo, por lo que no deberían asociarse con este. Además, concebimos el *bracketing* también como elemento de agrupación interna que facilita el análisis. Así, abogamos por agrupar en

este caso los dos modificadores, que constituirían, por tanto, un CN coordinado dentro del CN más amplio.

Por último, Barrière y Ménard (2014) defienden que la existencia de vínculos léxicos entre los formantes indica su dependencia. Para ello, analizan, por una parte, el uso de determinantes y del plural, que suelen reflejar las asociaciones léxicas. Por ejemplo, para *cotton polyester shirt*, encuentran bastantes ocurrencias de *the cotton shirts*, pero muchas menos de *the cotton polyesters*, lo que sugiere que *cotton* y *polyester* modifican de forma autónoma a *shirt*. Por otra parte, se basan en la aparición de los posibles términos dependientes en un recurso léxico; Wikipedia, en su caso. No obstante, esta opción no arroja resultados en todos los casos, ya que los recursos léxicos no siempre incluyen todo lo que buscamos. Además, igual que ocurre con la frecuencia, en algunos casos podremos encontrar ocurrencias de las distintas combinaciones posibles.

A pesar de que contamos con los modelos de *bracketing* desde hace más de veinte años, la desambiguación estructural sigue planteando problemas, especialmente en los CN de más de tres formantes, que no han recibido atención más allá de trabajos como los de Vadas y Curran (2011) o Barrière y Ménard (2014). No obstante, estos CN más extensos resultan de gran interés, pues a medida que aumenta el número de formantes, se incrementa también su complicación. Asimismo, se han propuesto indicativos específicos para el *bracketing* (Nakov y Hearst 2005; Barrière y Ménard 2014), si bien estos se basan en CN en inglés, pertenecientes a menudo a la lengua general, como en el caso de Barrière y Ménard (2014). No obstante, Vadas y Curran (2011) señalan la dificultad del *bracketing* de los CN especializados. Por tanto, se hace necesaria una lista de indicativos válidos tanto para la lengua especializada como para otros idiomas.

En el discurso especializado, los CN presentan las características propias que ya hemos mencionado y que dificultan su comprensión estructural, como la existencia de CN extensos, la omisión de formantes o la especificidad de sus relaciones semánticas. Por otro lado, el español, lengua que estudiamos en este trabajo, no presenta la compactación típica del inglés y favorece en cierta medida la comprensión estructural; por ejemplo, mediante la concordancia del género o la presencia de determinantes. Sin embargo, también plantea problemas de desambiguación. Estos se deben en gran parte al uso de preposiciones, que constituyen armas de doble filo en el *bracketing*. Algunos autores consideran que facilitan esta tarea (Kocourek 1982), aunque, a nuestro parecer, la concatenación de preposiciones en estructuras del tipo *instalación de generador de electricidad* resulta cuanto menos útil.

En esta línea, faltan por conocer los tipos de *bracketing* más frecuentes según el número de formantes del CN y la lengua de la que se trate.

Como decíamos, los acercamientos al *bracketing* se han realizado principalmente desde la Lingüística Computacional, mientras que áreas como la terminología o la traducción no han mirado de cerca este fenómeno, a pesar de que tiene implicaciones de gran envergadura en sus tareas. La realidad es que el traductor o el terminólogo no recurren frecuentemente a los sistemas de procesamiento del lenguaje natural para analizar con algoritmos la estructura de los CN ante los que se encuentran. Estos profesionales necesitan soluciones manuales que puedan ejecutar ellos mismos en cualquier momento, por lo que nuestra propuesta toma premisas de la literatura precedente sobre el *bracketing* para enriquecerla con nuevas nociones que faciliten esta labor manual de los profesionales de la lengua. Así, en la Sección 5.1.1.1 presentamos nuestra propuesta de indicativos para el *bracketing*, válida para CN de cualquier extensión en inglés y español, en concreto pertenecientes al discurso especializado.

3.4.2 La semántica de los formantes

La aclaración de la estructura formal del CN va de la mano de su análisis semántico, ya que, por una parte, el *bracketing* facilita el estudio conceptual del CN (una vez eliminadas las ambigüedades estructurales) y por otra, la comprensión del significado de los formantes es imprescindible para llevar a cabo el *bracketing*.

Por tanto, para analizar conceptualmente los CN, se deben estudiar en primer lugar los rasgos semánticos de los elementos que los conforman. De este modo podremos caracterizarlos e identificar con mayor facilidad la relación semántica entre estos (Sección 3.4.3). Como señalan Smith et al. (2014: 100), refiriéndose a los CN formados por sustantivos (NNC): «The only invariant information deducible from the structure of the NNC itself is that it denotes something (conveyed by the head) that is somehow related to something else (conveyed by the modifier)». Sin embargo, la relación entre ambos conceptos no está explícita. Por ello, en este apartado profundizamos en la semántica de los elementos que constituyen los CN, pues esta nos ayudará a conocer cuál es esa relación. Además de ayudar a la comprensión de la relación interna, el significado de los formantes se encuentra en el origen mismo de los CN, ya que, como veremos, es la semántica la que dicta la combinación de conceptos que da lugar a estos términos compuestos.

Sin embargo, antes de ahondar en la semántica de los formantes, nos parece necesario abordar su contribución al significado global de la combinación. Hablamos, pues,

de la composicionalidad de los CN, uno de los grandes ejes en torno a los cuales ha girado el debate de estas combinaciones. La composicionalidad se ha definido como la medida en que los elementos de una expresión poliléxica contribuyen al significado global de la expresión (Baldwin y Kim 2010: 4). De este modo, la composicionalidad va de la mano de la idiomaticidad. Esta última hace referencia al sentido figurado de una expresión, cuando esta es idiomática y, por tanto, no composicional (*agujero negro*), o bien, a su sentido literal si se trata de una expresión poco idiomática y, por tanto, composicional (*agricultural production*).

En la mayoría de los trabajos se defiende la falta de composicionalidad de los CN (Levi 1978; Sager et al. 1980; Kocourek 1982; Cabré 1993; Grant y Bauer 2004; Utsumi 2014; Jackendoff 2016; Fernández Domínguez 2019). En línea con Hampton (1991), Coulson (2001: 142, 159) se refiere a ella como *conceptual blending* y la ilustra con el clásico ejemplo de *pet fish*. En este CN se unen dos marcos, el de *pet* y el de *fish*, de los cuales se toman algunas características que formarán el *blend* resultante (*pet fish*). En concreto, del marco de *pet* se toman las características relativas al hábitat y a su condición de animal que no se utiliza como alimento, mientras que el marco de *fish* aporta la presencia de branquias y aletas, frente a los pulmones y patas que serían más típicos de *pet* (Coulson 2001). La falta de composicionalidad de los CN constituye, con diferencia, la corriente mayoritaria, si bien otros autores hablan de CN composicionales, como es el caso de Štekauer et al. (2012) y Kageura (2015), este último sobre CN especializados. En nuestro caso, compartimos las ideas de Nakov (2013) y Cordeiro et al. (2016), quienes hablan de un continuo de idiomaticidad en los CN, que va desde CN totalmente idiomáticos como *gravy train* (ya que la idiomaticidad está especialmente presente en los CN exocéntricos) hasta otros composicionales como *access road* (Cordeiro et al. 2016: 1987).

Entre las circunstancias que afectan a la falta de composicionalidad de los CN y, por tanto, repercuten en su transparencia, destacamos el uso de alguno de los formantes con un sentido diferente al habitual (Kocourek 1982; Nakov 2013; Fernández Domínguez 2019), la omisión de la relación semántica (como veremos en la Sección 3.4.3) (Levi 1978; Grant y Bauer 2004) y la ambigüedad de preposiciones como *of* y *de* (Kocourek 1982), que son las más comunes en CN preposicionales, pero no expresan con claridad el vínculo entre los conceptos. El uso de los formantes con un sentido diferente al habitual se puede observar en ejemplos como *col de Bruselas* (*chou de Bruxelles* en Kocourek [1982: 147]) o *caja negra* (*boîte noire* en Gaudin [2005: 87]), que ni son coles que provienen de Bruselas, ni cajas de color negro. A esto se suma el uso figurado de alguno de los elementos, como vemos en *energía verde*. Por otro lado, Grant y Bauer (2004: 46) ilustran con mucha

claridad la omisión de la relación semántica con el ejemplo de *Wellington bus*. Aunque a simple vista parezca que su significado se puede extraer de sus formantes, no sabemos si se trata de «'a bus going to Wellington', 'a bus coming from Wellington' or 'a bus such as typically operates within Wellington'». Por último, la escasa precisión semántica de preposiciones como *de* y *of* se observa en CN que presentan la misma preposición pero codifican relaciones semánticas diferentes, como vemos en *producción de energía* (producción *tiene_resultado* energía) y *evaluación de impacto ambiental* (evaluación *estudia* impacto ambiental). En el caso de los CN especializados, a estos factores se suman otros, como la omisión de formantes, la elevada especialización o la longitud de muchos CN, que reducen la composicionalidad de estos términos. No obstante, esto no significa que los CN especializados suelan ser idiomáticos, ya que estos son menos frecuentes en los lenguajes de especialidad, en contraposición a la lengua general. Nuestra postura es que los CN especializados tienden más hacia la composicionalidad, aunque nunca llegan a alcanzarla plenamente, pues aun cuando no se da ninguno de los factores comentados, siempre se requiere el aporte semántico de la relación implícita.

Una vez que hemos ahondado en la contribución de los formantes al significado global del CN, cabe preguntarse cuáles son los factores que determinan que se produzca una determinada combinación de conceptos. Y es que los elementos de un CN no se combinan de manera aleatoria (Downing 1977; Levi 1978). Su coocurrencia responde a cuestiones semánticas que, después, encontrarán la forma de representación más adecuada según la lengua de la que se trate (Kageura 1997, 2002, 2015; Štekauer 1998; Maguire et al. 2010b; *inter alia*). Por tanto, entra en juego el potencial combinatorio de los conceptos que, según su propia naturaleza, se vinculan con unos conceptos y no con otros (Kageura 1997, 2002, 2015). Por ejemplo, las entidades suelen combinarse con sus características (material, color, ubicación, etc.), pero no con causas o resultados, que serían propios de los procesos (una mesa no puede tener un resultado, mientras que una inundación, sí). El sistema conceptual se manifiesta, pues, en la formación de términos compuestos (Kageura 2015), como se desprende de los resultados de nuestro estudio.

Gagné y Spalding (2013: 103) explican con el siguiente ejemplo la relevancia de la semántica en la formación de CN:

> (…) the MADE OF relation, when actually paired with a particular head noun, requires that the modifier be a material, but not just any material. It must be a material that is appropriate for the head noun. *Snow sculpture* satisfies these restrictions, but *snow hospital* does not.

Así, vemos que el núcleo desempeña un papel fundamental en la formación de CN. Como indican Rosario et al. (2002), su significado determina cuáles pueden ser los conceptos que lo acompañen para formar un CN. En la Sección 5.1.2 ahondaremos en este aspecto mediante nuestra noción de microcontextos. En cuanto a la semántica del modificador, autores como Warren (1978) señalan su mayor carga comunicativa. Esta autora se refiere al modificador como *rema*, en la tradicional dicotomía de *tema* y *rema* (por otros conocida como *tópico* y *comentario*), que se corresponderían, pues, con el núcleo y el modificador, respectivamente. El núcleo (tema) tiene una menor función comunicativa, siendo su papel el de designar la pertenencia a una categoría semántica, mientras que el modificador (rema) transmite la característica principal que distingue a este concepto de su hiperónimo y sus cohipónimos.

Las restricciones conceptuales que se dan en los CN están relacionadas también con la preferencia y la prosodia semánticas. Se trata de dos de las *extended lexical units* propuestas por Sinclair (1991), junto con la colocación y la coligación. La preferencia semántica alude, como su propio nombre indica, a la tendencia de algunas palabras a combinarse con otras de una determinada categoría o campo semántico (Sinclair 1996). Es el caso del verbo *commit*, que suele aparecer con palabras que designan crímenes o hechos inaceptables socialmente, por ejemplo, *commit suicide* (Stubbs 2001: 64). La prosodia semántica, por su parte, alude a la naturaleza positiva o negativa de los colocados de una palabra (Sinclair 1996). Por ejemplo, el verbo *tackle* aparece normalmente con elementos negativos, como en *tackle a problem* (Bednarek 2008: 130). Estos fenómenos a menudo interactúan: por ejemplo, *fraught with* suele aparecer con elementos como *danger*, *risk*, *peril* o *hazard*, que podrían pertenecer a una categoría de PELIGROS o SITUACIONES NEGATIVAS, por lo que determinan también la prosodia negativa del verbo (Morley y Partington 2009).

Como vemos en los ejemplos, la preferencia y la prosodia semánticas suelen estudiarse en *idioms* y colocaciones, mientras que su presencia en CN y, en particular, en la lengua especializada, no ha sido objeto de análisis (Stubbs 2009). Sin embargo, resultan determinantes en la combinación de conceptos que se da en los términos compuestos. Por ejemplo, en el ámbito de la Ingeniería Costera, el término *disipación* suele combinarse con elementos positivos (prosodia semántica) pertenecientes a la categoría de ENERGÍA (preferencia semántica), como vemos en *disipación de la energía del oleaje* o *disipación de la energía por fricción* (Cabezas García y Faber 2019: 12). Además, estos fenómenos pueden establecerse tanto desde el núcleo como desde el modificador. Por ejemplo, los modificadores *riesgo de inundación* se combinan normalmente con núcleos positivos que pertenecen al dominio de CAMBIO, como vemos en *adaptación al riesgo de inundación* o

reducción del riesgo de inundación (Cabezas García y Faber 2019: 13). Estos rasgos semánticos de los CN también resultan importantes en la traducción, ya que no siempre se corresponden entre lenguas.

Además, las preferencias conceptuales de las que venimos hablando en este apartado suelen atender a patrones recurrentes en torno a los cuales estructuramos el mundo (Levi 1978; Warren 1978; Kageura 2002; Rosario et al. 2002; Maguire et al. 2010b). Estos pueden trazarse mediante el uso de categorías y roles semánticos[21]. En línea con Hoey (2005: 63), consideramos que lo que se combinan entre sí, más que palabras, son categorías semánticas:

> Our focus on the way that words or word sequences form semantic associations may be insufficiently generalized and (…) instead of formulating semantic association as 'item *x* has a semantic association with semantic set Y, represented by items *a, b* and *c*' we should formulate the association thus: 'semantic set X, represented by items *p, q* and *r*, has a semantic association with semantic set Y, represented by items *a, b* and *c*.

Por ejemplo, en lugar de decir que *tarta* suele combinarse con *zanahoria* (*tarta de zanahoria*), sería más adecuado señalar que la categoría de ALIMENTOS suele coocurrir con la de sus INGREDIENTES más destacados (*pastel de carne, empanada de atún*, *flan de huevo*). Autores como Levi (1978) o Warren (1978), para la lengua general, y Rosario et al. (2002) o Kageura (1997, 2002), para la lengua especializada, han analizado la formación de CN a partir de estas combinaciones recurrentes de categorías. Así, en la lengua general son frecuentes, entre otras muchas, las combinaciones del tipo PEOPLE-OCCUPATION (*lawyer friend, student renter*) (Levi 1978: 240), denominadas ACTIVITY-ACTOR por Warren (1978: 47) (*cowboy, sportsman*). En el dominio biomédico, Rosario et al. (2002) recogen combinaciones como PATIENT-VIRUS DISEASE (*influenza patient, aids survivor*), mientras que en el ámbito de la documentación, Kageura (1997, 2002) encuentra coocurrencias como DOCUMENT-MODE OF PUBLICATION (*one-side copy, pocket book*[22]), entre otras. Como señala Levi (1978), es de esperar que estas combinaciones frecuentes no se den solo en inglés, sino que se produzcan también en el resto de lenguas. Por ello, en esta tesis comparamos la presencia de estos patrones en un dominio especializado en inglés y español.

Este tipo de generalizaciones también pueden realizarse utilizando roles semánticos, esto es, atendiendo a la estructura argumental del predicado subyacente en

[21] Como hemos señalado, no existen inventarios fijos de categorías y roles semánticos. Así, debido a la diversidad de estudios y de nomenclatura, es habitual que los mismos elementos sean considerados categorías o clases semánticas por unos y roles por otros.
[22] Traducciones literales del japonés.

los CN. Este enfoque es menos habitual que el anterior, probablemente por ser menos transparente y más difícil de aplicar, debido a la omisión frecuente del verbo o a la presencia de elementos como adjetivos cuyo rol puede resultar difícil de asignar. Sin embargo, también permite observar patrones recurrentes, como vemos en la combinación, que se da en CN del tipo *wind energy system* o *wind power plant*, en los que interviene un instrumento y el producto resultante de su acción. En esta tesis, como veremos, abogamos por el uso combinado de categorías y roles semánticos para conseguir una plena caracterización conceptual.

Otro modo habitual de estudiar los patrones de formación de CN son las relaciones semánticas entre núcleo y modificador, que veremos con detalle en la Sección 3.4.3. Destacamos las aportaciones de Sager, que se distingue por su estudio de la formación de términos compuestos. Este autor señala que, en función de la naturaleza del núcleo, el modificador puede especificar una propiedad del nuevo concepto (*waste deposit*), su uso (*insulating material*), el todo al que pertenece la parte designada por el núcleo (*pedestal body*), su material (*steel post*), etc. Este último tipo a menudo incluye el participio pasado de un verbo, que matiza la relación semántica (*rubber-sheated cable, steel-armoured conduit*) (Sager et al. 1980; Sager 1997). Como señalan los propios autores, esta lista (que incluye más tipos) no es exhaustiva. Además, observamos que algunos casos se solapan (por ejemplo, tanto *waste deposit* como *insulating material* especifican el *uso*), si bien constituye un primer acercamiento muy completo a los patrones semánticos en la formación de CN en terminología. Adams (2001) realiza una propuesta similar y señala que este tipo de clasificaciones facilita la comprensión de nuevos CN, una opinión que compartimos, ya que los términos compuestos suelen crearse según los patrones existentes.

Centrándonos en el uso de categorías y roles semánticos para caracterizar los formantes de los CN, encontramos algunos estudios que los han aplicado con distintos fines. Por ejemplo, las categorías semánticas se han utilizado principalmente para estudiar la formación de CN y sus relaciones semánticas internas.

Entre los trabajos que han explorado la formación de CN mediante el uso de categorías semánticas destacan los de Downing (1977), Levi (1978), Kageura (1997, 2002) y Štekauer et al. (2012). Downing (1977) las aplica a la creación de nuevos CN (p. ej. APPEARANCE + PLANT > *trumpet plant*). Levi (1978), por su parte, complementa el uso de categorías con el de rasgos semánticos (DEFINITE, CONCRETE, ANIMATE, HUMAN, etc.). Otro trabajo fundamental es el de Kageura (2002), que realiza un estudio minucioso de los

patrones de formación de términos compuestos en japonés, basándose en el uso de categorías y relaciones semánticas. Este autor, que explora la terminología de la documentación, destaca que las categorías no pueden entenderse de forma aislada, sino como parte de un sistema conceptual específico (como veremos en las categorías específicas del medio ambiente que utilizamos en esta tesis). Por último, Štekauer et al. (2012) reclaman el estudio de las categorías semánticas en la formación de palabras, aplicable también a la formación de términos compuestos.

Por otro lado, las categorías también favorecen el estudio de las relaciones semánticas internas, como lo demuestran Warren (1978), Adams (2001), Rosario et al. (2002) o Maguire et al. (2010b), entre otros. Estos trabajos se sustentan sobre la idea de que las características semánticas de los conceptos sirven para delimitar los conceptos que pueden vincularse mediante una relación semántica. Por ejemplo, en el dominio medioambiental, *wave* puede codificar la relación *causes* solo si pertenece a la categoría de MOVEMENT WATER (p. ej. wave *causes* beach erosion), pero no si se entiende como MODEL REPRESENTATION (León Araúz y Faber 2012). Esto es aplicable a la combinación conceptual que se da en los CN. Como vemos, el contexto resulta fundamental para la asignación de categorías semánticas.

De este modo, las categorías semánticas sirven para caracterizar las relaciones de los CN, como comenta Warren (1978). Esta autora ahonda en las diferentes relaciones que pueden establecerse en los CN de la lengua general y, para cada relación, señala las categorías y rasgos semánticos de los conceptos vinculados. Se decanta por estas combinaciones semánticas en detrimento de las de tipo sintáctico, que no considera adecuadas (p. ej. *apple core* > *apple* [subject] has *core* [object]) (Warren 1978: 256).

Esta caracterización semántica por medio de categorías facilita la inferencia de las relaciones. Como señala Adams (2001: 86), si el núcleo de un CN pertenece a la categoría de HUMAN, normalmente se complementará con su ocupación o su identidad. Por su parte, si pertenece a la de ANIMAL o PLANT, coocurrirá con conceptos que designen su apariencia, hábitat o ubicación. Los CN pertenecientes a la categoría de NATURAL OBJECT suelen especificar su material, origen o ubicación, mientras que aquellos que corresponden a la categoría de MAN-MADE OBJECT indican generalmente su función. En este sentido, Rosario et al. (2002) proponen su *Descent of Hierarchy*, que comentaremos en la Sección 3.4.3. En este modelo se aplican las categorías de la jerarquía léxica del MeSH a los formantes de CN de biomedicina para investigar si determinadas combinaciones establecen la misma relación semántica. Por ejemplo, las categorías BODY REGION y CARDIOVASCULAR SYSTEM

coocurren en CN como *scalp artery, heel capillary o limb vein* (Rosario et al. 2002: 3), que codifican la relación *located_at*. Maguire et al. (2010b) se basan en la misma idea, con el objetivo de determinar si los patrones semánticos de los CN son predecibles utilizando conocimiento estadístico.

Otra de las aplicaciones de las categorías semánticas en los CN consiste en la propia sistematización de los inventarios de categorías. Sager (1997: 35) propone las principales categorías que pueden designar los CN en terminología, en concreto, OBJETOS (*concrete breaker*), PROPIEDADES (*concrete strength*) y PROCESOS/OPERACIONES (*concrete slump test*). A estas suelen añadirse las relaciones, que se consideran otra de las categorías conceptuales genéricas. A pesar de que las tres primeras categorías son a las que suelen pertenecer los CN, consideramos que hay algunos que designan relaciones, como vemos en *adverse effect*, que puede entenderse como la relación *affects* (p. ej. X *has adverse effects on Y* > X *affects* Y). Cabe recordar que, en los CN endocéntricos, el CN es un hipónimo del núcleo y, por tanto, pertenece a la misma categoría semántica que este. También en esta línea, Grön et al. (2018) analizan las formas que suelen adquirir los CN en holandés según la categoría semántica que designan (p. ej. la categoría de FINDING suele representarse con sintagmas nominales precedidos por adjetivos).

Por último, las categorías semánticas se han aplicado también al estudio de la variación en los CN. Fernández Silva (2017) anota con categorías semánticas los formantes de los CN especializados para estudiar si estas influyen en su variación. Un estudio similar es el de Grön y Bertels (2018), que utilizan categorías semánticas para explorar la variación terminológica en el lenguaje clínico.

Por lo que respecta a los roles semánticos, los trabajos que hacen uso de estos en los CN asumen la existencia de predicados subyacentes, como es el caso de esta tesis. Sus principales aplicaciones han sido el estudio de la formación de CN, su interpretación semántica y el análisis de sus relaciones internas.

En cuanto al uso de roles para estudiar la formación de CN, cabe destacar de nuevo las aportaciones de Levi (1978), quien habla de *case relations* (AGENT, INSTRUMENT, LOCATION, etc.) y señala que, además de a los sustantivos, también se pueden atribuir a los adjetivos denominales, a los que se refiere como *nominal nonpredicating adjectives*. A pesar de los matices sintácticos de su enfoque, Levi (1978) destaca que el análisis de los CN no puede basarse en su forma superficial, sino en sus proposiciones subyacentes, en las que plantea el uso de categorías y roles semánticos. Así, por ejemplo, señala adjetivos que representan el rol de INSTRUMENT, como *microscopic* en *microscopic analysis* (Levi 1978:

27). En este sentido, aunque no aplicado al estudio de los CN, el trabajo de Gràcia Solé et al. (2000) también analiza los roles semánticos representados por adjetivos deverbales (p. ej. *vencedor*: SUJETO[23] AGENTE), que resulta de interés porque estos pueden formar parte de CN.

La interpretación semántica de los CN también se ha nutrido del uso de roles semánticos, como se observa en los trabajos de Finin (1980), Martín Arista (1997), Ferretti y Gagné (2006), Cohen et al. (2008), Güemes et al. (2016) y Jackendoff (2016). En función de los *slots* abiertos por el verbo, Finin (1980) investiga el rol más adecuado para el otro formante del CN. Por ejemplo, en *engine repair*, el verbo *repair* abre los *slots* de AGENT, OBJECT, INSTRUMENT, LOCATION, etc. De estos, identifica OBJECT (lo que se repara) como el rol que representa *engine* con mayor seguridad. Como vemos, su propuesta integra matices sintácticos, que no incorporamos en esta tesis, en la que hablamos de PATIENT en lugar de OBJECT. Martín Arista (1997), por su parte, plantea el uso de roles para desambiguar el significado de los CN. Ferretti y Gagné (2006) señalan que, para facilitar la comprensión de los sintagmas nominales, solemos generar sus proposiciones subyacentes y, en estas, es el AGENTE el que ocupa la posición de sujeto.

Cohen et al. (2008) anotan los adjetivos de los CN cuando representan argumentos del verbo subyacente. Sin embargo, solo consideran como argumentos al AGENT o INSTRUMENT (entendidos como los que realizan la acción) y al PATIENT (receptor de la acción). Por tanto, en *surgical treatment* anotan el adjetivo *surgical* como argumento 2 según su nomenclatura (esto es, INSTRUMENT), pero no etiquetan *transcriptional* en *transcriptional activation*. A nuestro parecer, este adjetivo podría anotarse como MANNER para conseguir, así, una mayor caracterización semántica. Por otro lado, Güemes et al. (2016) estudian la estructura argumental de CN del tipo verbo-nombre con unión ortográfica en español, utilizando roles semánticos (p. ej. AGENTE/PACIENTE > algo que V a N > *abrelatas*). Determinan que el procesamiento cognitivo de los CN locativos es más complejo, por lo que destacan el papel de la estructura argumental en la interpretación de los CN. Por último, Jackendoff (2016: 25) también habla de roles semánticos en los CN. Por ejemplo, señala que los verbos que se nominalizan con *-er* indican AGENT o INSTRUMENT (*hair dryer, windbreaker, aircraft carrier*, etc.).

Otra de las aplicaciones de los roles semánticos en los CN consiste, de nuevo, en analizar sus relaciones semánticas internas. Un ejemplo es el trabajo de Maroto García et

[23] La denominación de SUJETO AGENTE nos parece más ligada a la sintaxis y, por tanto, no siempre válida, dado que los agentes no ocupan necesariamente la posición de sujeto de la oración. Por tanto, preferimos el uso de la denominación AGENTE.

al. (2012) en el que, a pesar de no aplicarse exclusivamente a los CN, se anotan con roles semánticos e información sintáctica términos que denotan procesos y sus participantes. Entre estos términos se incluyen nominalizaciones, por lo que cobran especial importancia los CN. Destacan que el argumento AGENTE aparece principalmente en las formas verbales, sin nominalizar. Observamos este aspecto en nuestros CN, ya que el AGENTE suele aparecer en las paráfrasis verbales de nuestros CN, pero no siempre en la propia estructura del término compuesto. Así, por ejemplo, solemos encontrar el CN *wind farm development*, donde no se especifica el AGENTE que lleva a cabo la acción del verbo *develop*, y oraciones que expanden la información del CN, como la siguiente: «The company is investing $80 million in manufacturing facilities and developing wind farms in China», en la que sí se especifica el AGENTE de dicha proposición (*company*).

Como vemos, tanto los estudios que utilizan categorías como los que aplican roles semánticos contribuyen a conocer la semántica de los formantes de los CN. Sin embargo, son pocos los que han integrado ambos enfoques. Como hemos mencionado, esta es nuestra propuesta, que veremos en detalle en la Sección 5.1.2, pues consideramos que mediante las categorías y los roles se consigue una plena caracterización conceptual. De este modo, se pueden realizar generalizaciones y extraer resultados muy específicos, tanto en un corpus como en otros recursos lingüísticos (p. ej. bases de datos), cuando los términos están anotados.

Así, entre las escasas propuestas que combinan el uso de categorías y roles semánticos en los CN, cabe destacar la de Bevilacqua (2004), que resulta especialmente interesante por su estudio de unidades fraseológicas especializadas. Esta autora se centra en las combinaciones que cuentan con un verbo (ya sea en infinitivo, en participio o nominalizado), cuyos argumentos analiza mediante el Modelo de la Gramática Léxica (Faber y Mairal 1999). Por ejemplo, representa la combinación *acumulación del calor* del siguiente modo, introduciendo una especie de roles y categorías semánticas (Meta, Objeto desplazado, Energía), además de elementos morfosintácticos (sufijos, Sp, etc.) (Bevilacqua 2004: 202):

> [[NE(acumular)V (-ción)suf]N eventivo (Sp (del) (X2: calor))Meta / Objeto desplazado / Energía)]Proceso1.2

Por otra parte, destacamos el trabajo de Bouillon et al. (2012), quienes realizan un interesante estudio de CN en francés e italiano, que anotan con los roles de *qualia* de Pustejovsky (1995). Estos roles, que veremos con más detalle en la Sección 3.4.3, intentan reflejar la manera en que el modificador complementa al núcleo. Son cuatro: Formal (*is_a*),

Constitutive (*part_of*), Telic (*function*) y Agentive (*causes*). Además de indicar los roles de *qualia*, Bouillon et al. (2012) anotan los formantes de los CN con categorías (ARTIFACT, TIME PERIOD, etc.) y roles semánticos (AGENT, INSTRUMENT, etc.), y señalan si existe un predicado interno (p. ej. en *freight train* indican que el verbo subyacente es *transport*, lo que contribuye a especificar la relación semántica). Por tanto, nos parece un enfoque muy completo, que va en la línea del análisis que realizamos en esta tesis.

En resumen, el análisis semántico de los formantes de los CN engloba diferentes aspectos, principalmente su contribución al significado global del CN (esto es, la composicionalidad), las restricciones conceptuales (entre las que se encuentran la prosodia y la preferencia semánticas) y los patrones recurrentes que dan lugar a la formación de CN y que pueden estudiarse utilizando categorías y roles semánticos. Una vez aclarada la semántica de los formantes, estamos en disposición de analizar la que puede ser la característica principal de los CN: su relación semántica interna, como veremos en el siguiente apartado.

3.4.3 Análisis de las relaciones semánticas internas

En el apartado anterior hemos ahondado en la relevancia de los elementos del CN para acceder a su significado. Sin embargo, además de esto es fundamental conocer cómo interactúan estos conceptos, ya que el vínculo entre ambos no suele expresarse de manera explícita. Dicha omisión no genera grandes problemas cuando ya se conoce el concepto al que hace referencia la combinación (Warren 1978); por ejemplo, no es preciso saber que *código de barras* es un código formado por barras, ya que por lo general conocemos el referente de esta denominación. Sin embargo, la omisión de la relación interna puede resultar problemática cuando no sabemos el significado del CN. Esto cobra una especial importancia si tenemos en cuenta el papel de la composición como mecanismo de creación de nuevos términos, tanto en inglés como en español.

Así pues, la omisión del vínculo entre los formantes de los CN no solo dificulta su comprensión, sino que tiene implicaciones en otras tareas, como la búsqueda de equivalentes de traducción. Por ello, su análisis se ha convertido en uno de los ejes fundamentales en la investigación de estas unidades. Además, esta elisión da lugar a que muchos CN presenten la misma estructura, pero cuenten con relaciones semánticas diferentes, lo que complica aún más la tarea. Por ejemplo, en *wind erosion*, el viento es el causante de la erosión, mientras que en *beach erosion*, la playa no es la causante sino la entidad afectada por dicho desgaste. A pesar de contar con diversidad de estudios al respecto, la inmensa mayoría se han centrado en CN del inglés formados por dos elementos, por lo general sustantivos. No obstante, la composición adopta también otras

formas y se muestra como mecanismo frecuente en otras lenguas, entre ellas el español. De hecho, los CN con más de dos formantes plantean una mayor complicación, ya que, además de la relación principal que vincula el núcleo y el modificador tras haber realizado el *bracketing*, sus agrupaciones internas presentan relaciones adicionales. Por otra parte, las lenguas romances como el español suelen formar la composición por medio de estructuras preposicionales, que la mayoría de las veces no suponen una garantía para conocer la relación interna (p. ej. *erosión del viento* y *erosión de la playa* presentan la misma preposición pero diferentes relaciones, como hemos visto arriba). Por tanto, un estudio de las relaciones internas en CN de todo tipo y en lenguas diferentes al inglés resulta de vital importancia.

En este sentido, una de las cuestiones fundamentales que cabe plantearse es cómo se establece esta relación entre los elementos de un CN. Según Warren (1978: 68), intervienen tanto la semántica de los formantes como el contexto, nuestro conocimiento extralingüístico y nuestra noción (inconsciente) de los patrones semánticos de formación de CN. Sin embargo, no podemos olvidar que la relación interna de los CN no constituye todo el significado de la combinación. Podríamos entender el concepto combinado como si se tratara de un puzle. Además de las piezas individuales (que serían la semántica de los formantes y su relación interna), para ver la imagen del puzle necesitamos tenerlo completo. Esta imagen del puzle completo se correspondería con el concepto designado por el CN, al que también contribuye el conocimiento extralingüístico, como señala Levi (1978). De este modo, en *library book*, podemos saber que *library* CONTAINS *books*, pero ello no incluye, entre otros aspectos, el dato relativo a los códigos de barras con los que se etiquetan estos libros (Bauer y Tarasova 2013: 3).

En cuanto a la influencia del significado de los formantes en la relación que los asocia, cabría preguntarse si dicha relación viene determinada por el núcleo, el modificador o por ambos elementos. Diferentes autores se han planteado esta pregunta, con diversidad de opiniones. Por una parte, hay autores como Downing (1977) que consideran que el núcleo es el que establece la relación semántica. Por ejemplo, en *onshore wind*, el núcleo *wind* determina que la relación sea *located_at*. En el lado opuesto se encuentran Gagné y Shoben (1997), quienes afirman que es el modificador el encargado de determinar la relación. Por ejemplo, en *wood chair*, es el modificador *wood* el que sugiere la relación *made_of*. Esta postura, también defendida por Fernández Domínguez (2010), se enmarca en su teoría de *Competition Among Relations in Nominals* (Gagné y Shoben 1997), según la cual las relaciones del modificador compiten y gana la más fuerte, que suele ser la más frecuente. Estos autores señalan también que el concepto combinado

se comprenderá con mayor facilidad si la relación es una de las establecidas habitualmente por el modificador. Por ejemplo, *wood chair* (*made_of*) resulta más fácil de comprender que *wood magazine* (*about*), porque el modificador *wood* suele aparecer más con la relación *made_of* que con *about*.

Gagné (2002) continúa con el trabajo de Gagné y Shoben (1997), que modifica para incluir el papel del núcleo en la selección de la relación. En concreto, defiende que su función consiste en señalar cuál de las relaciones establecidas por el modificador es la más favorable, una idea también defendida por Fradin (2017). Por ejemplo, para analizar la relación interna de *mountain planet*, partimos de que la principal relación establecida por el modificador *mountain* es *X located_at mountain*, ya que el papel de LOCATION es una de las características más representativas de *mountain*. Sin embargo, al examinar las características de *planet*, constatamos que la relación *located_at* no es posible en este caso, ya que los planetas son muy grandes para estar ubicados en montañas (Gagné 2002: 732). Esta propuesta nos parece más acertada que las anteriores, pues consideramos que la esencia de las combinaciones conceptuales, como son los CN, es su composición por dos o más conceptos y, por tanto, ambos tienen algo que decir en el establecimiento de la relación que los vincula.

Maguire et al. (2010a) no asumen el estudio de Gagné y Shoben (1997) y señalan que la facilidad de comprensión de una combinación se debe más bien a la interacción entre el núcleo y el modificador, ya que si solo se atendiera al modificador, se estaría hablando siempre de las mismas relaciones. Por ejemplo, el 99 % de combinaciones en las que *plastic* es el modificador codifican la relación *made_of*. No obstante, esto no significa que dicha relación se dé en todos los casos, ya que es condición necesaria que el núcleo designe un objeto concreto —por ejemplo, combinaciones como *plastic texture* o *plastic recycling* no cumplen este requisito y, por tanto, no codifican esta relación— (Maguire et al. 2010a: 289).

En resumen, la propuesta de Maguire et al. (2010a) subraya el papel del potencial combinatorio de los formantes y su categoría semántica, además de otorgar un mayor énfasis al núcleo. Esta idea aparece implícita en los trabajos de Levi (1978), Murphy (1988, 1990) y Kocourek (1982), quienes sostienen que ni el núcleo ni el modificador por separado son garantías de que se establezca una determinada relación. Levi (1978: 52) lo ejemplifica con CN que cuentan con el mismo núcleo o modificador pero codifican relaciones diferentes: *electric generator* es un generador que produce electricidad, mientras que *solar generator* no es un generador que produce sol, sino que lo utiliza como

recurso energético; del mismo modo, *musical clock* es un reloj que produce música, pero *musical criticism* es una crítica sobre música. Por tanto, se establecerán distintas relaciones que dan cuenta de los microcontextos (Sección 5.1.2) y que responden a los patrones combinatorios de ambos formantes.

En otro estudio, Maguire et al. (2010b) añaden que el establecimiento de la relación dependerá no solo de la categoría semántica de ambos formantes sino también de su posición en el CN. Como ellos, consideramos que tanto el núcleo como el modificador intervienen en la codificación de la relación. Así, teniendo en cuenta sus categorías semánticas, se debe encontrar el punto en el que convergen, en el que su combinación resulta válida. Para ello, debemos considerar que las relaciones sugeridas por el modificador deben encajar en las propiedades definicionales del núcleo, pues la esencia de la composición es la especificación; es decir, formamos un CN para especificar el núcleo con alguna de sus características. Por ejemplo, en *sand beach*, la relación semántica es *made_of*, una de las principales establecidas por los MATERIALES como *sand*. Además, esta encaja en las propiedades definicionales del núcleo *beach*, pues se trata de una ENTIDAD FÍSICA, que puede especificarse según su composición. Como vemos, ni el modificador por su parte (*sand*, que podría establecer otras relaciones como *located_at* o *affects*) ni el núcleo por separado (*beach*, que también codifica las relaciones *located_at, affected_by, result_of*, etc.) son suficientes para comprender la relación de un CN. Ambos son, por tanto, fundamentales.

Por otro lado, algunos consideran que el contexto también influye en la codificación de la relación interna de los CN (Downing 1977; Warren 1978; Bauer y Tarasova 2013; Cohen y Staub 2014). Sin embargo, de manera general las relaciones de los CN se han estudiado sin tener en cuenta el contexto de estas combinaciones. Así, autores como Nakov (2013) y Ó Séaghdha y Copestake (2013) consideran que este no es necesario, ya que los formantes de los CN se desambiguan mutuamente. En esta tesis, por el contrario, consideramos que el contexto reviste una gran importancia en todo análisis semántico. En este sentido, nos parece que las proposiciones conceptuales que subyacen en los CN son un elemento muy útil, ya que ayudan a resolver posibles ambigüedades no solo a nivel de su relación interna sino también relativas a los formantes (por ejemplo, cuál es su categoría o rol semántico). Rosario et al. (2002: 5) ejemplifican ambigüedades en las relaciones semánticas de los CN en el ámbito de la biomedicina, como vemos en *kidney metabolism*, que puede ser *metabolism happening in the kidney* o *metabolism done by the kidney*. Como señala también Ó Séaghdha (2007a), este tipo de ambigüedades pueden resolverse fácilmente en contexto. A pesar de que en determinados CN pueda

darse más de una relación semántica, en nuestro estudio analizamos diferentes contextos y seleccionamos la relación más común para incluirla en el módulo fraseológico de EcoLexicon.

Así como la semántica de sus formantes y el contexto del CN influyen en el establecimiento de la relación semántica, también se ha afirmado que los patrones de combinación conceptual determinan el vínculo interno de los CN (Warren 1978) —p. ej. la combinación MATERIAL + OBJECT suele codificar la relación *made_of*—. Estos patrones de formación de términos, de los que ya hablaron autores como Levi (1978), Kageura (2002), Rosario et al. (2002) o Maguire et al. (2010b) (véase la Sección 3.4.2), son la idea principal sobre la que se sustenta la inferencia de relaciones en los CN. Numerosos trabajos han tratado de alcanzar dicha inferencia, que consiste en deducir la relación interna de nuevos CN a partir de determinados factores. Debido al papel de la composición como mecanismo de creación de nuevos términos, esta cobra una especial relevancia. Como señalan Maguire et al. (2010b), los patrones de combinación conceptual se pueden explotar para facilitar la comprensión de nuevos CN, pues los conceptos que pertenecen a las mismas categorías comparten características y esquemas de combinación similares, una idea que asumimos en esta tesis.

Así, no son pocos los trabajos que defienden que los CN formados por los mismos pares de categorías semánticas codifican la misma relación, lo que facilita la identificación de relaciones semánticas en nuevos CN (Johnston y Busa 1999; Rosario et al. 2002; Nastase et al. 2006; Ó Séaghdha y Copestake 2013; Maguire et al. 2010a, 2010b; Rallapalli y Paul 2012; Kim y Baldwin 2013). Es decir, si *irrigation hose* y *cutting jig* codifican la relación *has_function*, es muy probable que *plough disc* también establezca dicha relación, pues sus formantes son similares. Esto no significa que una determinada relación se dé únicamente entre un par de categorías, como señalan Rosario et al. (2002), sino que dicho par suele codificar habitualmente esa relación. Como vemos, se trata de un método que se aplica a CN que podríamos considerar composicionales, pues los CN idiomáticos no suelen mantener el significado de todos sus formantes y, por tanto, no son proclives a las comparaciones.

Johnston y Busa (1999) ya proponían la creación de un modelo computacional a partir de un conjunto de CN anotados con las categorías semánticas de sus formantes y su relación interna. Utilizando los ejemplos de los que dispone, dicho modelo sería capaz de deducir con qué probabilidad una determinada relación se daría en nuevos CN.

Siguiendo la misma idea de base, el de Rosario et al. (2002) constituye uno de los trabajos principales sobre las relaciones en los CN. En su teoría para la inferencia de relaciones en los CN, denominada *Descent of Hierarchy*, estos autores asignan categorías semánticas de la jerarquía léxica del MeSH a un conjunto de CN de biomedicina. Determinan que los CN que implican las mismas combinaciones de categorías suelen establecer la misma relación semántica. Así, llaman *reglas* a los pares de categorías cuyos CN establecen la misma relación. Por ejemplo, *forearm artery* y *eyelid capillary*, que están formados por las mismas categorías, codifican la relación *located_at*. Son estas reglas las que facilitan la interpretación de los CN.

Por su parte, Maguire et al. (2010b) también defienden que las combinaciones de categorías no se dan aleatoriamente, sino que siguen patrones recurrentes. De este modo, las mismas combinaciones de categorías establecerán relaciones similares. Por ejemplo, la combinación de TIME PERIOD y EVENT suele codificar la relación *during*. Esto facilitaría la interpretación simplemente gracias a nuestro conocimiento estadístico (basado en repeticiones), sin necesidad de activar representaciones detalladas de cada concepto de la combinación (Maguire et al. 2010b: 58). Estos enfoques sustentan la idea que comentábamos arriba según la cual tanto el núcleo como el modificador son importantes para el establecimiento de la relación semántica.

Gagné y Shoben (1997) y Smith et al. (2014) señalan otro de los factores que favorecen la inferencia de relaciones en los CN: la exposición a CN con los mismos formantes. Es decir, tendemos a aplicar la relación que sabemos que establece un determinado elemento en otros CN al nuevo CN ante el que nos encontramos. Por ejemplo, si *plastic chair* y *plastic bag* codifican la relación *made_of*, todo apunta a que *plastic door* también establecerá dicha relación. No obstante, cabe señalar que Gagné y Shoben (1997) aplican esta idea únicamente al modificador, es decir, no consideran que la aparición del núcleo en otros CN dé pistas de la relación ante la que nos encontramos. En la misma línea, si existen pocas ocurrencias de CN similares al nuestro, su relación será más difícil de interpretar, ya que disponemos de menos ejemplos (Gagné y Shoben 1997).

En definitiva, el vínculo entre los elementos de un CN ha sido objeto de diferentes estudios. Para abordarlo, la tendencia más extendida ha sido el uso de clasificaciones de relaciones semánticas (*causa, afecta,* etc.). Sin embargo, nuevos enfoques plantean su análisis por medio de verbos o preposiciones (p. ej. una *llave dobladora* es una llave *para* doblar), como veremos a continuación.

3.4.3.1 Los inventarios de relaciones semánticas

Con vistas a especificar la asociación entre los elementos de un CN, se han utilizado a menudo inventarios de relaciones semánticas. Estas son elementos fundamentales en terminología, pues constituyen la columna vertebral de los sistemas conceptuales. En los CN determinan cómo deben entenderse dichas combinaciones. Por ejemplo, *erosión marina* es una erosión *causada por* el mar, mientras que *código de barras* es un código *formado por* barras. Además, los CN y sus relaciones internas dan cuenta del sistema conceptual subyacente, por lo que el análisis de las categorías y relaciones semánticas de estos términos permite la reconstrucción del sistema conceptual (Sager et al. 1980).

Las relaciones semánticas (tanto las que vinculan conceptos en general como las que se aplican a los CN) pueden presentar una mayor o menor granularidad. Por ejemplo, el inventario de PropBank (Palmer et al. 2005) incluye relaciones generales, mientras que existen también taxonomías específicas de un área, como la de Rosario et al. (2002), aplicada a los CN del ámbito de la biomedicina. Cuanto más generales son estas relaciones, mayor es el número de CN al que pueden asignarse, pero también son mayores las ambigüedades sobre la relación más adecuada (Girju 2007). Lo cierto es que se han elaborado multitud de inventarios diferentes, perfilados según las aplicaciones y objetivos de cada estudio. Sin embargo, a pesar del interés que suscita la interpretación de los CN, no existe consenso sobre el número de relaciones y su especificidad, al margen de la coincidencia en relaciones generales como la hiponimia y la meronimia.

A continuación recogemos las principales taxonomías de relaciones que se centran en los CN. Como veremos, la gran mayoría se han realizado desde el ámbito de la Lingüística Computacional. Esto se debe a que la identificación de la relación interna de los CN es una de las tareas más abordadas en esta disciplina, debido a su importancia para el buen funcionamiento de los sistemas de procesamiento del lenguaje natural. Prueba del interés suscitado entre la comunidad de lingüistas computacionales son los encuentros dedicados a tal fin, como *SemEval-2007 Task 4* y *SemEval-2010 Task 8*, sobre la clasificación de relaciones semánticas entre sustantivos o *SemEval-2015 Task 17*, sobre la evaluación de métodos de extracción de taxonomías, todos de aplicación a los CN. Sin embargo, son escasos los estudios terminológicos que se hayan centrado en la relación entre los formantes de los CN, al margen de Sager et al. (1980), Pugh (1984), Sager (1997), Kageura (2002) y Oster (2003, 2005, 2006). Otros estudios pormenorizados de estos términos compuestos no han abordado explícitamente las relaciones internas, como es el caso de Sanz Vicente

(2011), debido a que la finalidad principal era la asignación de equivalentes en español (Sanz Vicente 2011: 456).

Así, en contra de autores como Jespersen (1942), que habían afirmado que las relaciones entre los formantes de los CN son infinitas, desde la Lingüística teórica Levi (1978) propone uno de los inventarios más conocidos de relaciones semánticas en los CN, después utilizado en diferentes estudios, como el de Ó Séaghdha (2007a). Así, con el objetivo de estudiar la formación de CN, diseña estas relaciones, que se aplican únicamente en los CN formados por supresión del predicado y reciben el nombre de *recoverably deletable predicates* (RDP), para aludir a que esa información omitida se puede recuperar por medio de uno de los RDP. Son nueve las relaciones o RDP propuestas: *cause, have, make, use*, *be*, *in*, *for*, *from* y *about*. Como vemos, no todos son predicados entendidos como verbos, sino que también encontramos preposiciones. Levi (1978: 161) asegura que su inventario también se puede aplicar a otras lenguas, porque los RDP son primitivos semánticos del lenguaje humano. Por tanto, introduce una idea que constituirá después toda una teoría semántica: los primitivos semánticos o *semantic primes* de Wierzbicka (1996), cuyo ámbito de aplicación no se limita a los CN, sino al lenguaje en general.

Una de las particularidades de sus relaciones semánticas es que algunas como *cause* presentan dos subtipos, ya que distingue si la causa es el núcleo (*tension headache*) o el modificador (*flu virus*). Además de esta diversidad, algunas relaciones son polisémicas. Por ejemplo, existen dos tipos de *make*: *physically producing* y *composed of* (Levi 1978: 90). A ello se suma el hecho de que muchos CN pueden entenderse según varios RDP, como es el caso de *job tension*, que puede ser *tension caused by the job, tension that the job has* o *tension on the job* (Levi 1978: 91). Independientemente de los casos concretos, Levi (1978) sostiene que la ambigüedad de los CN se reduce a doce posibilidades: nueve según los nueve RDP que pueden subyacer, a las que se añaden tres, debido a que *cause, have* y *make* pueden tener como modificador al sujeto o al objeto del verbo. Sin embargo, en su contexto de uso, estas posibilidades se reducen drásticamente, ayudadas por pistas semánticas y factores pragmáticos o enciclopédicos.

Sin embargo, la popularidad de este estudio ha provocado también reacciones contrarias. Las principales críticas se centran en el carácter extremadamente genérico de estas relaciones (Downing 1977[24]), que suscribimos en esta tesis. Prueba de ello es la relación *in*, que no distingue entre ubicación en el tiempo o en el espacio, o más notorios

[24] Downing (1977) comenta un trabajo precedente (Levi 1975), si bien este coincide en gran parte con lo expuesto en Levi (1978).

aún, los ejemplos de *fertility pills* y *headache pills*. A ambos se les asigna la relación *for*, aunque salta a la vista que el vínculo no es precisamente el mismo. Levi (1978) justifica este caso no por la ambigüedad de *for*, sino por la contraposición de los dos ejemplos, ya que su comprensión por separado no da lugar a dudas. En nuestra opinión, este caso demuestra la necesidad de relaciones más específicas, probablemente en forma de verbos con mayor contenido semántico. De lo contrario, se estaría omitiendo una parte importante de la combinatoria de estos conceptos. La propia autora concibe esta solución y propone el siguiente análisis: *pills to increase fertility* vs. *pills to decrease headache* (Levi 1978: 99).

No obstante, Levi (1978) defiende la generalidad de sus RPD, para la que alega tres razones. Por una parte, el carácter genérico de sus RDP permite hacer generalizaciones en los CN. Por otra, aunque la especificación de dichas relaciones no deja de ser posible —p. ej. *in* podría concretarse con relaciones como *inhabit* (*field mouse*), *grow-in* (*water lily*), *according-to* (*theological fallacy*), *during* (*night flight*), *found-in* (*urban park*) y *occur-in* (*industrial strike*)—, esta mayor granularidad actuaría en detrimento de la posibilidad de generalización antes señalada. Por último, defiende la generalidad de su propuesta porque otras opciones, como *x has y, x is in y* o *x is for y*, son igualmente generales y difusas. Como sabemos, este otro tipo de relaciones no serían las únicas posibles, aunque esta limitación se justifica quizás por lo temprano del estudio, que fue uno de los pioneros en el análisis de las relaciones semánticas de los CN. Sin embargo, como venimos señalando, nos parece que el carácter general de sus relaciones es uno de los puntos débiles de Levi (1978).

Otra de las críticas que ha suscitado la clasificación de Levi (1978) radica en que esta se aplica a los CN formados por supresión del predicado pero no a los que se derivan de nominalizaciones (Ó Séaghdha 2007b; Nastase et al. 2013). Esto hace que se trate de una propuesta poco intuitiva, ya que a CN como *history professor* se les asigna una relación semántica, mientras que otros muy similares como *history teacher* (entendidos como nominalizaciones) se analizan según la función de la nominalización (Ó Séaghdha 2007b: 2). Nos parece que la propuesta de Levi (1978) sería más consistente si se aplicaran los RDP a todos los CN, independientemente de su formación, ya que consideramos que las relaciones semánticas se encuentran en todos los CN. Así, por ejemplo, *water pollution*, a pesar de representar una nominalización, codifica la relación *affects* (pollution *affects* water). Además, las nominalizaciones contribuyen a especificar el papel del modificador y la relación semántica, de manera que permiten una mayor caracterización conceptual del CN: el agua se ve afectada de un determinado modo, en

concreto, sufre efectos negativos en su calidad debido a cambios químicos, físicos o biológicos, que son los que implica la contaminación.

Por otro lado, se ha señalado que hay CN que no se pueden encuadrar en ningún RDP (Downing 1977; Coulson 2001), como es el caso de *guinea pig* (Rosario y Hearst 2001: 84). En esta línea, hay otros muchos a los que se les pueden asignar varias relaciones (Downing 1977; Coulson 2001), un problema que también se observa en otros inventarios. Ó Séaghdha (2007b) señala el ejemplo de *car factory*, que puede entenderse como *a factory for producing cars* (*for*), *a factory that causes cars to be created* (*cause*), *a factory in which cars are produced* (*in*) o *a factory from which cars originate* (*from*). A pesar de que compartimos esta opinión, nos parece que el ejemplo propuesto por Ó Séaghdha (2007b) no es demasiado aclaratorio, ya que todas estas formulaciones señalan a la relación *produce*, independientemente de que se utilicen diferentes verbos y preposiciones. En cualquier caso, este autor sugiere que se deben proporcionar mejores indicaciones que guíen en la asignación de cada relación. A esto habría que añadir la relevancia del contexto para esclarecer la relación, pues consideramos que este tipo de ambigüedad sigue presente incluso en inventarios más específicos.

Centrándonos en las relaciones en particular, Rosario (2005) critica que Levi (1978) no incluya *of* entre sus RPD, pues considera que se encuentra en un buen número de CN. Nosotros, como Levi (1978), también apostamos por la exclusión de *of*, dada su extrema generalidad y su escaso potencial aclaratorio, que debe ser la base de toda relación semántica. Ó Séaghdha (2007b), por su parte, centrado en la necesidad de contar con un gran número de ejemplos para considerar la existencia de una determinada relación, cuestiona la necesidad de RDP como *cause* y *make*, al ser estas muy poco frecuentes. A nuestro parecer, estas relaciones son necesarias y extrapolables a otros ámbitos y lenguas, que sin duda aumentarán el número de ocurrencias de dichas relaciones.

A pesar de las diferentes críticas, queremos destacar la relevancia de la clasificación de Levi (1978), que fue pionera en la aplicación de las relaciones semánticas a los CN y sentó las bases para un gran número de investigaciones posteriores, que tomaron las ventajas de su propuesta e intentaron mejorar sus deficiencias.

También desde la Lingüística teórica, Warren (1978) propone otra de las taxonomías más conocidas de relaciones que se dan en los CN, entendidos como combinaciones más o menos frecuentes de la lengua general. Plantea cuatro relaciones genéricas: *Constitute and Resemblance*, *Belonging to*, *Location* y *Purpose and Activity*.

Además, propone subtipos para cada relación. Por ejemplo, *Belonging to* se compone de los subtipos *Whole-Part*, *Part-Whole* y *Size-Whole*. El primero de ellos (*Whole-Part*), por ejemplo, se desglosa a su vez en *Obj-Part* (*spoon handle*), *Group-Member* (*Peace Corps girl*) y *Obj-Geometrical outline* (*hillside*). Asimismo, comenta las ambigüedades y restricciones que presenta cada relación y, para cada una, propone paráfrasis, tanto verbales como preposicionales. Por separado analiza los CN que incluyen lo que podría entenderse como un nombre propio (*X-ray, Y-region*), ya que afirma que son los únicos a los que no se les puede asignar una relación[25].

Probablemente su aportación más interesante es su desglose de los patrones semánticos que un CN debe cumplir para encuadrarse en una determinada relación, dado que el objetivo principal del trabajo es recoger los patrones semánticos de formación de CN. Por ejemplo, en la subrelación *Activity-Actor*, el formante que represente *actor* debe ser animado y concreto y el formante que represente *activity* será una actividad, tarea o interés regular y fijo en el que participe el otro formante (Warren 1978: 210).

Por tanto, vemos que esta propuesta es de corte mucho más semántico que la de Levi (1978), aunque también presenta deficiencias. Por una parte, destacamos la generalidad de algunas relaciones, como es el caso de *Location*, que incluye tanto ubicación en el tiempo como en el espacio. Además, a pesar de que se incluyen subrelaciones, la posición de algunas no resulta muy clara, por ejemplo, *Source-Result* se entiende como una subrelación de *Constitute*. Sin embargo, los principales problemas que encontramos son dos: los solapamientos y la falta de cobertura del inventario. En cuanto a los solapamientos, cabe destacar por ejemplo el caso de *Constitute* y *Belonging to*, cuya diferencia no resulta evidente. De hecho, *Part-Whole* se incluye como subrelación en ambos, lo que generará sin duda solapamientos y asignaciones indistintas a cada grupo. Por otra parte, un gran número de CN no se pueden encuadrar en ninguna relación, algo que reconoce la propia autora. Por ejemplo, la subrelación *Part-Whole* se atribuye a algunos CN que no siguen exactamente la norma pero que no sabe cómo categorizar (p. ej. *telephone number*). Esta decisión, que responde a la intención de asignar relaciones a todos los CN, ocasiona que los patrones semánticos que propone Warren (1978) no siempre se cumplan. Sin embargo, su clasificación debe destacarse por el esfuerzo de consideración de la semántica de los formantes.

[25] Otros autores como Rosario (2005) sí incluyen estos CN en su análisis. Véanse Rojas García y Faber (2019a, 2019b) para profundizar en el estudio conceptual de las entidades nombradas, que en múltiples ocasiones presentan formas compuestas.

Sager et al. (1980) y Sager (1997) estudian la formación de términos compuestos según las relaciones semánticas entre sus formantes. Así, aunque no presentan un inventario cerrado, comentan las principales relaciones que subyacen en los CN, como son *material*, *property*, *use* o *part*, entre otras. Los trabajos de Sager destacan por ser pioneros en el estudio de las relaciones semánticas en terminología, en concreto aplicadas a los términos compuestos. Nos parecen relaciones muy acertadas y válidas para su aplicación a diferentes estudios. No obstante, consideramos que no se incluyen relaciones importantes en terminología, como *affects* o *result_of*, que implican procesos, muy presentes en los sistemas conceptuales, además de que incorporan matices sintácticos, que son poco relevantes en nuestro enfoque (p. ej. hablan de un tipo de relación que especifica el objeto del verbo).

También desde la disciplina terminológica, Pugh (1984) analiza la formación de CN en inglés, francés y español pertenecientes al ámbito de la informática. Con vistas a estudiar la unión de los formantes, utiliza también un inventario de dieciocho relaciones, entre las que se encuentran *origin, function* y *product*, entre otras.

Desde una perspectiva computacional, Vanderwende (1994) investiga la interpretación automática de secuencias de sustantivos. Para ello, extrae automáticamente un inventario de trece relaciones semánticas de un diccionario online, que aplica a las cadenas de sustantivos, a modo de preguntas: *Who/what?* (*Subject*), *Whom/*what? (*Object*), *Where?* (*Locative*), *When?* (*Time*), *Whose?* (*Possessive*), *What is it part of?* (*Whole-Part*), *What are its parts?* (*Part-Whole*), *What kind of?* (*Equative*), *How?* (*Instrument*), *What for?* (*Purpose*), *Made of what?* (*Material*), *What does it cause?* (*Causes*), *What causes it?* (*Caused-by*).

Sin alejarnos de la Lingüística Computacional, Pustejovsky (1991, 1995) propone su teoría del Lexicón Generativo, que explica cómo contribuye la composicionalidad a la semántica léxica. Así, señala que el significado léxico cuenta con los siguientes niveles de representación: *Argument Structure*, o el número y tipo de argumentos de una palabra; *Event Structure*, es decir, el tipo de evento en el que interviene el concepto en cuestión; *Qualia Structure*, esto es, las propiedades definitorias del concepto; e *Inheritance Structure* o la relación con otros conceptos del sistema (Pustejovsky 1995: 3). La estructura de *qualia* es la que nos interesa aquí, pues permite realizar generalizaciones sobre el vínculo semántico que une a los elementos de un CN, al especificar la manera en que el modificador complementa al núcleo. Como señalan Johnston y Busa (1999: 168): «the qualia structure provides the "glue" which links together the semantic contributions of

modifying nouns and the head noun in the compound». Los roles de *qualia* son cuatro: *Formal* (*is_a*), *Constitutive* (*part*), *Telic* (*function*) y *Agentive* (*cause*). Además, estos roles tienen subtipos. Por ejemplo, el rol *Formal* incluye (i) *orientation*, (ii) *magnitude*, (iii) *shape*, (iv) *dimensionality*, (v) *colour* y (vi) *position* (Pustejovsky 1995: 85-86). En la Tabla 2 se ejemplifica la propuesta de Pustejovsky (1991, 1995) con el concepto NOVEL.

novel(*x*)
Const: narrative(*x*)
Form: book(*x*), disk(*x*)
Telic: read(T,y,*x*)
Agentive: artifact(*x*), write(T,z,*x*)

Tabla 2: Estructura de *qualia* relativa al concepto NOVEL (Pustejovsky 1991: 427)

Como vemos, se entiende a modo de plantilla en la que se presentan las propiedades definitorias del concepto que, como señalamos también en nuestra noción de microcontextos (Sección 5.1.2), dan lugar a la formación de términos CN. Así, por ejemplo, a partir de estas características de *novel* se pueden formar combinaciones como *book novel* o incluso, en la actualidad, *e-book novel*. Por tanto, la de Pustejovsky (1991, 1995) nos parece una propuesta muy rica semánticamente, que atiende a factores individuales del concepto (*inheritance structure* incluye por ejemplo la categoría semántica), así como a su potencial combinatorio (los argumentos que presenta, el tipo de evento en el que participa y sus elementos definitorios, esto es, los roles de *qualia*, que darán lugar a combinaciones).

Además, el Lexicón Generativo (Pustejovsky 1991, 1995) se ha aplicado a CN en diferentes trabajos. Uno de ellos es el realizado por Johnston y Busa (1999), quienes utilizan la estructura de *qualia* para explicar la composicionalidad de los CN. Warren (2003) estudia la viabilidad del Lexicón Generativo para el análisis de los CN. Bouillon et al. (2012) proporcionan un esquema de anotación para CN en italiano y francés basado en esta teoría. Por otro lado, Bassac y Bouillon (2013) exploran la relación télica en CN del francés y el turco. Por último, Yadav et al. (2017) aplican los roles de *qualia* al análisis de las relaciones semánticas de los CN.

A pesar de basarse en la lengua general, los roles de *qualia* también pueden aplicarse al lenguaje especializado (León Araúz et al. 2012). De hecho, su relevancia y presencia son indudables también en este tipo de discurso, si bien consideramos que constituyen un conjunto limitado que, en muchos casos, no recogería la amplitud y especificidad de las relaciones que se dan en los términos compuestos. Por ejemplo, CN tan comunes en el área de las energías renovables como son *wind farm* o *solar generator* no

podrían caracterizarse mediante ninguno de los roles de *qualia*, que no recogen relaciones del tipo *use*. Además, basándonos en la Tabla 2, vemos que dicha información puede resultar muy útil para tareas como la elaboración de definiciones o el estudio de la formación de CN, pero en concreto, su aplicación a la relación interna de los CN no nos parece exhaustiva.

Otro de los enfoques computacionales a las relaciones de los CN ha sido el de Barker y Szpakowicz (1998), quienes desarrollan un sistema semiautomático para identificar las relaciones semánticas de los sintagmas nominales. A partir de la asignación manual de relaciones, dicho sistema indica las relaciones de nuevos CN, dada su similitud con los datos de los que dispone. Barker y Szpakowicz (1998) utilizan un inventario de veinte relaciones, entre las que se encuentran *Cause*, *Property* o *Equative*, entre otras.

Rosario et al. (2002) se proponen agilizar la asignación de relaciones de los CN en los sistemas de procesamiento del lenguaje natural. Para ello, asignan categorías semánticas a los formantes de CN de biomedicina utilizando la jerarquía léxica del MeSH y estudian las relaciones que se dan entre dichas categorías, obteniendo un total de treinta y ocho relaciones. Estas son específicas del ámbito de la biomedicina (p. ej. *Produce Genetically*, *Person Afflicted* o *Defect*), ya que el objetivo es trabajar con unas relaciones lo suficientemente generales para que cubran muchos CN, pero lo suficientemente específicas para que sean útiles en el análisis. Como hemos visto, la propuesta de Rosario et al. (2002), denominada *Descent of Hierarchy*, resulta de gran importancia para la inferencia de relaciones semánticas.

Por otra parte, en uno de los estudios fundamentales sobre la formación de términos compuestos en terminología, Kageura (2002) analiza las relaciones dentro de CN del área de la documentación en japonés. Para ello, se basa en la taxonomía de Pugh (1984), que modifica cuando es necesario. Así, propone seis grandes grupos en los que se aglutinan estas relaciones: *equality, functional, part/whole, internal attributes, external attributes* y *other*. Estos grupos, a su vez, incluyen subrelaciones. Por ejemplo, *internal attributes* especifica qué tipo de atributos pueden ser (*formal attributes, static manner, nature, attributed concepts* y *quantity*). Kageura (2002) se centra en las relaciones que se dan tanto dentro de los CN como dentro de los términos, a modo de afijos. Señala que estas difieren de las que se dan entre dos conceptos en general, en un sistema conceptual. No solo estudia la relación principal entre el núcleo y el modificador, sino también las relaciones adicionales cuando el CN está formado por más de dos elementos, y señala cuáles son las más frecuentes en el ámbito de la documentación. Concluye que en esta

disciplina se encuentran las mismas relaciones que en el lenguaje general (p. ej. *inclusion*, *similarity*), además de otras específicas (p. ej. *synonymy*, *disjunction*). Pese a que consideramos que estas relaciones específicas también se dan en la lengua general, las aportaciones de Kageura (2002) al estudio de las relaciones de los términos compuestos son innegables.

También desde una perspectiva terminológica, Oster (2003, 2005, 2006) profundiza en las relaciones semánticas entre los formantes de los CN especializados, pertenecientes al campo de la cerámica. Propone un inventario de catorce *esquemas relacionales* genéricos de estructura binaria (p. ej. *place-determined entity*), con el objetivo de plasmar la semántica de los formantes, frente a las tradicionales relaciones simples del tipo *causes, located_at*, etc. Algunos de estos esquemas o relaciones son *action-agent*, *zero-attribute* y *container-content*. Estos se agrupan en tres grandes categorías: *esquemas proposicionales*, *modelos esquemáticos de imagen* y *modelos metafóricos.* La propuesta de Oster (2003, 2005, 2006) destaca por su representación del papel desempeñado por los formantes del CN en la relación, un enfoque poco adoptado en las taxonomías. Además, a pesar de especificar la semántica de los formantes, se trata de relaciones de aplicación general, ya que ninguna presenta un significado demasiado restringido.

Nastase y Szpakowicz (2003) también estudian la asignación de relaciones semánticas a los CN, con el objetivo de encontrar reglas que justifiquen la existencia de una determinada relación. Así, plantean un inventario de treinta relaciones (p. ej. *detraction, property, location from, location at*) agrupadas en cinco categorías genéricas: *causality, participant, quality, spatial* y *temporality*. Dicha clasificación se ha aplicado también en otros estudios computacionales, como los de Turney (2005), Nastase et al. (2006) y Nakov y Hearst (2008), ya que se puede elegir una mayor o menor generalización de las relaciones.

Otro de los estudios computacionales a este respecto ha sido el de Girju et al. (2005), quienes pretenden profundizar en la semántica de CN formados por dos y tres elementos. Para ello, utilizan treinta y cinco relaciones, inspiradas en el inventario de Moldovan et al. (2004), entre las que se incluyen *Possession*, *Location/Space*, *Accompaniment* o *Theme*.

Con vistas a la interpretación automática de CN, en concreto a la elaboración de un esquema de anotación de relaciones en los CN, Ó Séaghdha (2007b) emplea un inventario de seis relaciones genéricas inspirado en el de Levi (1978). Estas relaciones (*be, have, in, actor, instrument* y *about*) cuentan con subrelaciones binarias, en las que el núcleo del CN

rellena un *slot* y el modificador, otro, por lo que permiten una mayor especificación. Por ejemplo, una subrelación de *have* es *property-object*. Como las relaciones binarias de Oster (2003, 2005, 2006), estas también pueden aplicarse a la mayoría de los CN.

Por último, Jackendoff (2016) se centra en el estudio de los CN desde su teoría de la Semántica Conceptual. Así, propone trece *funciones básicas* o relaciones que pueden darse en los CN: *classify*, *be*, *similar*, *kind*, *be at/in/on*, *comp*, *made*, *part*, *cause*, *make*, *serves as*, *have*, *protect*. Como vemos, alterna las típicas relaciones básicas con otras muy específicas (como *protect*), que creemos no cubrirán un gran número de CN. Encontramos de nuevo el problema de la generalidad de algunas relaciones (p. ej. *location*), un aspecto que reconoce el propio autor, quien señala que a menudo se deben expresar de forma más específica. Además, no se aprecia la diferencia entre algunas relaciones, como es el caso de *Comp* (*composed of*) y *Part*. Observamos también que, como Levi y otros autores de la tradición sintáctica, utiliza estas relaciones solo cuando no hay un predicado nominalizado, mientras que, cuando lo hay, el análisis cambia y se habla de argumentos. Como señalamos anteriormente, consideramos que las relaciones pueden aplicarse en todos los casos, por lo que deberían utilizarse siempre para conseguir una mayor coherencia.

Como vemos, de las diferentes taxonomías presentadas se desprende el interés de la Lingüística Computacional por este tema, ya que constituye la principal disciplina que ha estudiado las relaciones entre los formantes de los CN. Además, casi todos los estudios han abordado CN de la lengua general, a excepción de casos puntuales como el de Rosario et al. (2002) o Nakov y Hearst (2013), que han estudiado CN especializados, y los estudios de terminología propiamente dichos, como los de Sager et al. (1980), Pugh (1984), Sager (1997), Kageura (2002) y Oster (2003, 2005, 2006). Asimismo, observamos que el centro de estudio han sido los CN formados por dos elementos, mientras que las estructuras más extensas no han recibido demasiada atención, probablemente debido a las mayores complicaciones que presentan. Las lenguas diferentes del inglés tampoco se han estudiado en detalle. Por otro lado, cabe señalar que, a pesar de que los formantes de los CN se suelen vincular por medio de relaciones, estos a veces se relacionan mediante atributos, es decir, propiedades de uno de los formantes (p. ej. *wind speed*). Sin embargo, no todos los inventarios recogen esta posibilidad.

En cuanto a la forma que adquieren las relaciones, estas se plasman a modo de verbos (*causes*), preposiciones (*about*), una designación genérica (*origin*) o el rol semántico de los formantes (*action-agent*), aunque esta última opción, de tipo binario, ha

sido menos frecuente que las anteriores. Además, pueden partir del núcleo (*source* es la relación de *silk worm* si consideramos *worm*), del modificador (en el mismo ejemplo, otros inventarios la catalogarían como *product*, partiendo de *silk*) o revertirse en función del caso en cuestión (*causes* vs. *caused_by*). Por otra parte, la disconformidad en el número y tipo de relaciones resulta patente: podemos encontrar desde taxonomías formadas por cuatro relaciones (Warren 1978; Pustejovsky 1991, 1995) hasta otras con treinta o más relaciones (Rosario et al. 2002; Girju et al. 2005), pasando por otras con unos veinte tipos (Pugh 1984; Barker y Szpakowicz 1998).

A nuestro parecer, la adecuación de uno u otro inventario dependerá de la finalidad del estudio. Por una parte, las clasificaciones más generales, que cuentan con un menor número de relaciones, permiten realizar generalizaciones (por ejemplo, con vistas a estudiar la formación de CN o sus patrones recurrentes). Por otra, los inventarios más desglosados facilitan la interpretación de estas combinaciones, al aportar una mayor caracterización semántica. A pesar de que algunos de estos inventarios podrían resultar útiles para nuestro análisis de los CN en terminología, en esta tesis utilizamos las relaciones de EcoLexicon (además de otras específicas que diseñamos), ya que nuestro estudio se aplica a esta base de conocimiento. Como veremos, se trata de un inventario de extensión intermedia, compuesto por diecisiete relaciones.

En definitiva, las relaciones semánticas presentan ventajas, como su concisión y su capacidad de generalización (Hendrickx et al. 2013), pero también cuentan con inconvenientes. Downing (1977), por ejemplo, señala la ambigüedad y la escasa especificidad semántica de muchas relaciones, que ilustra con el ejemplo ya comentado de *headache pill* y *fertility pill*, que se entenderían según la misma relación (*purpose*) en la mayoría de inventarios, a pesar de que su significado merece ser matizado (*pill to reduce the headache* vs. *pill to increase fertility*). A esto, Nakov (2013) añade la necesidad de elegir entre un gran número de inventarios, además de la posibilidad de asignar varias relaciones en muchos casos (p. ej. *beach sand* podría entenderse como *part_of* o *located_at*) o incluso la falta de adecuación de relaciones para un determinado CN, como hemos comentado en los inventarios de Levi (1978) y Warren (1978), pero que sin duda sucede también en otras taxonomías.

A pesar de ello, los inventarios de relaciones semánticas son los más empleados en la Lingüística Computacional (Nakov 2013) y en todos los análisis conceptuales en general. No obstante, han surgido nuevas propuestas que buscan una mayor especificidad en este

vínculo por medio de paráfrasis preposicionales y, especialmente, verbales, como veremos en el siguiente apartado.

3.4.3.2 Las paráfrasis

Jespersen (1942) y, más tarde, Downing (1977) afirmaron que la asociación entre los elementos de un CN no puede caracterizarse mediante un inventario cerrado de relaciones, ya que este vínculo es mucho más rico. Afirmaciones como estas, unidas a los problemas de las relaciones tradicionales señalados por muchos, dieron lugar a propuestas alternativas que plantean el análisis de los CN por medio de paráfrasis de tipo preposicional o verbal. Como veremos, estas propuestas surgieron y se desarrollaron en el ámbito de la Lingüística Computacional.

El principal exponente de las paráfrasis preposicionales es Lauer (1995), quien investiga la utilidad del conocimiento estadístico en las tareas de procesamiento lingüístico. Así, utiliza un algoritmo que calcula cuál es la paráfrasis más adecuada para un CN, sin considerar su contexto. Aunque muy utilizadas también por otros autores, las preposiciones se habían entendido hasta ahora como relaciones semánticas, mientras que Lauer (1995) las concibe como un tipo de paráfrasis. Se inspira en el inventario de Warren (1978), a partir del cual propone ocho preposiciones que pueden vincular los formantes de los CN: *of, for, in, at, on, from, with* y *about*. Por ejemplo, *olive oil* puede entenderse como *oil from olives*.

Este enfoque generó diferentes críticas. Entre otros, Nastase et al. (2013) señalan que las preposiciones son de por sí polisémicas. Tomemos el ejemplo de *of*, probablemente la preposición que expresa el mayor número de vínculos: no es lo mismo *surface of the Earth* (*part_of*) que *effect of climate change* (*caused_by*) o *transport of sand* (*affects*). Además, Nastase et al. (2013) subrayan que muchas veces se puede asignar más de una preposición, y, otras, el empleo de distintas preposiciones no implica diferentes relaciones semánticas (p. ej. *in, at* y *on*). Por otra parte, alegan que muchos CN, en concreto, los coordinados, quedan fuera del análisis, mientras que la asignación de relaciones a otros no resulta muy certera: «should *honey bee* really be analyzed as *bee for honey*?» (Nastase et al. 2013: 18).

A pesar de la capacidad de generalización y la comodidad de esta propuesta (al tratarse de un conjunto reducido formado por preposiciones breves y fáciles de utilizar), no compartimos su enfoque. Consideramos que las preposiciones son, en su mayoría, extremadamente generales para dar cuenta de la riqueza semántica que puede existir entre los formantes de los CN. Además, el vínculo que se establezca debe ser aclaratorio y

no ambiguo, pues su objetivo es señalar cómo se unen ambos formantes. En este sentido, algunas lenguas cuentan con preposiciones que expresan claramente la relación, como es el caso de *à* en francés, que indica la función en CN como *verre à vin* (Bassac y Bouillon 2013: 110). Sin embargo, esto no se da en todas las preposiciones ni en todas las lenguas. Como veremos, en español e inglés las preposiciones *de* y *of* son las más frecuentes, pero también las menos explícitas, ya que pueden expresar diversidad de relaciones.

Otros autores plantean el uso de verbos específicos para representar la asociación entre los formantes de un CN. Uno de ellos es Lees (1970), quien propone el uso de verbos como *impel, propel, energize*, etc., de gran especialización. Levi (1978), por su parte, sugiere que su análisis de relaciones semánticas puede complementarse con el uso de verbos específicos. Por ejemplo, *automotive shop* (*repair*) *o clothing shop* (*sell*) (Levi 1978: 97, 104). Aunque su enfoque principal también es su propuesta de relaciones semánticas, Warren (1978) propone paráfrasis verbales y preposicionales para cada relación. Por ejemplo, *liquor saloon* puede entenderse como *saloon for imbibing liquor* (Warren 1978: 200). L'Homme (2002), por su parte, destaca la utilidad de los verbos para acceder al significado de las nominalizaciones. Así, sugiere buscar contextos tanto de la nominalización (p. ej. *configuration*) como del predicado subyacente (*configure*), lo que guarda una relación directa con el uso de paráfrasis verbales para conocer y especificar la relación de los CN.

Desde el ámbito de la Lingüística Computacional, Finin (1980) fue el precursor de la noción de paráfrasis verbales. Este autor plantea el uso de verbos específicos para vincular los formantes de los CN. Por ejemplo, analiza *washing machine* como *a machine which is an instrument to wash* (Finin 1980: 311). A pesar de que su propuesta dista mucho de la noción de paráfrasis, sirvió de inspiración a Nakov y Hearst (2006), quienes fueron los propulsores de esta idea, que aplicamos en el presente trabajo. Estos autores, que también se basan en FrameNet y los marcos de Fillmore (1977, 1982, 1985), proponen el uso de diferentes verbos para representar de forma específica la relación interna de los CN. De este modo, consideran que los verbos tienen un mayor potencial de representación semántica que las relaciones típicas de los inventarios. Nakov y Hearst (2006: 4) lo ejemplifican a la perfección en el siguiente fragmento:

> (…) while *malaria mosquito* can very well be characterized as CAUSE (or *cause*), further aspects of the meaning can be captured by adding some additional verbs, e.g., *carry*, *spread*, *transmit*, *be responsible for*, *be infected with*, *pass on*, etc.

Esta idea de paráfrasis verbales, después desarrollada en Nakov y Hearst (2008, 2013) y Nakov (2008, 2013), goza de gran popularidad entre la comunidad de lingüistas computacionales, como demuestran publicaciones como las de Nulty (2007), Butnariu y Veale (2008), Nakov (2013), Nakov y Hearst (2013), Hendrickx et al. (2013) o Nulty y Costello (2013), entre otros. Muchas de ellas se propusieron en encuentros científicos dedicados a este fin, como *SemEval-2010 Task 9*, sobre la interpretación semántica de CN mediante paráfrasis verbales en lugar de relaciones abstractas; o *SemEval-2013 Task 4*, sobre las paráfrasis libres de CN. La Lingüística Computacional es, sin duda, el campo más prolífero en lo que a paráfrasis se refiere.

Así, los verbos se erigen en transmisores del significado básico de los CN, del mismo modo que constituyen los cimientos para la formación de estos términos compuestos (Levi 1978). Butnariu y Veale (2008) recalcan la utilidad de los verbos como matizadores de las relaciones semánticas de los CN. Así, señalan que, frente a CN como *diamond crown*, *diamond watch*, *diamond bracelet* o *diamond throne*:

> A hat may be *crowned* with diamonds, a watch *decorated* with diamonds, a bracelet *covered* with diamonds, a throne *encrusted* with diamonds and a king *bedecked* with diamonds – each is an elaboration of a basic *covering* relation, but each adds nuances of its own that we do not want to lose in an interpretation that is maximally specific to the nouns concerned (Butnariu y Veale 2008: 83).

Como vemos, la propia semántica de los formantes determinará el tipo de verbos que se pueden utilizar en las paráfrasis. Por ejemplo, *decorate* es un verbo válido para *diamond watch*, mientras que *crown* sería más apropiado para otras combinaciones como *diamond hat*. Por otra parte, las paráfrasis son también una muestra de la importancia del contexto, porque requieren de la aparición de los formantes del CN en una oración, no aislados. Además, ya hemos señalado que para conocer la relación se requiere muchas veces del contexto, mientras que esto se resuelve con las paráfrasis, una idea que ya sugería Kocourek (1982), al señalar que la relación dentro de los CN es menos evidente que la que se observa entre los elementos de una oración.

Nakov (2013) incide en los múltiples beneficios de las paráfrasis, ya que contribuyen a la traducción automática estadística, la obtención de información o la minería de textos, entre otras aplicaciones. Por ejemplo, tras conocer por medio de paráfrasis que *prevent* es un verbo adecuado para *migraine treatment*, se puede formular la consulta "* which prevents migraines" para obtener tratamientos o medicamentos contra la migraña: *feverfew*, *Topamax*, *natural treatment*, *magnesium*, *Botox*, *Glucosamine*, etc. Si complementamos esta búsqueda con otro de los verbos posibles entre *migraine* y

treatment, como podría ser *reduce* (en una consulta del tipo "* reduces migraine"), obtenemos resultados adicionales: *lamotrigine*, *PFO closure*, *Butterbur Root*, *Clopidogrel*, *topamax*, *anticonvulsant*, *valproate*, *closure of patent foramen ovale*, *Fibromyalgia topamax*, *plant root extract*, *Petadolex*, *Antiepileptic Drug Keppra (Levetiracetam)*, *feverfew*, *Propranolol*, etc. (Nakov y Hearst 2013: 43).

Para la extracción de paráfrasis se han utilizado diferentes técnicas. La principal ha sido el uso de motores de búsqueda. Por ejemplo, Nakov y Hearst (2006) realizan consultas en la web del tipo *"noun2 that * noun1"* (pudiendo sustituirse *that* por *which* o *who*) o, a la inversa, colocando el primer sustantivo antes del segundo ("*noun1 * noun2"*). Con el asterisco se pretenden obtener elementos desconocidos, que en este caso son verbos. Nulty (2007) también busca paráfrasis en la web, de la forma *head + joining term + modifier*. Estos *términos de unión*, entendidos en el sentido de paráfrasis, son en su mayoría preposiciones, por lo que estaríamos hablando de un enfoque similar al de Lauer (1995).

No obstante, la extracción de paráfrasis de internet presenta limitaciones, ya que no resulta útil para CN nuevos y originales ni para aquellos convencionales que no expanden su significado a modo de paráfrasis (Butnariu y Veale 2008). No obstante, pensamos que la dificultad de obtener paráfrasis de estos CN de sentido más idiomático no se da únicamente en la web, sino que es extrapolable a todos los métodos de extracción de paráfrasis. Es decir, la propuesta de acceso a su significado por medio de paráfrasis no resulta válida, como tampoco lo son las relaciones semánticas. Por ejemplo, *fire rainbow* no es un tipo de *rainbow* ni de *fire*, así como tampoco puede entenderse como *rainbow made_of fire* o *rainbow caused_by fire* o, a modo de paráfrasis, *rainbow composed of/induced by/resulting from fire*. Y es que su vínculo es diferente: se trata de un fenómeno óptico atmosférico cuyo nombre se basa en su similitud con el arcoíris y que se forma debido a la presencia de cristales de hielo en algunas nubes.

Por otro lado, se han utilizado paráfrasis manuales propuestas por sujetos. Nakov y Hearst (2013) analizan la semántica de CN del dominio de la biomedicina y comparan las paráfrasis obtenidas de la web con las proporcionadas por expertos en el campo. Los resultados fueron en gran parte coincidentes, pero mostraban más discordancias cuando se trataba de CN con relaciones ambiguas. Por ejemplo, *olive oil* se puede entender como *container* (*oil that is inside the olive*), *source* (*oil that comes from olives*), *product*[26] (*oil that is produced from olives*) o *quality* (*oil that tastes like olive*) (Nakov y Hearst 2013: 13-14).

[26] Cabe señalar que las relaciones *source* y *product* se entienden en esta tesis como dos partes de la misma relación: se adoptará una u otra dependiendo de si se parte del núcleo o el modificador.

Nulty y Costello (2013) también pidieron a un grupo de sujetos que hicieran paráfrasis a partir de un conjunto de CN. Las paráfrasis más detalladas y específicas, que serían las ideales, no resultaron ser las más propuestas. Por el contrario, las del tipo *be made from*, mucho menos concretas, fueron las más utilizadas. Aunque las paráfrasis obtenidas por este medio serían útiles para nuestro trabajo, dado que los expertos pueden proporcionar información muy específica (Cabezas García y Faber 2017b, 2019), nos parece que son mucho menos asequibles que el uso de corpus o la web, pues se requiere una mayor inversión de tiempo y, además, no siempre tenemos a los expertos a nuestro alcance.

Además de las consultas en la web y a sujetos, las paráfrasis también pueden generarse a partir de corpus, aunque este enfoque ha sido menos utilizado. Nakov y Hearst (2013) destacan que estos ofrecen menos ruido que internet, pero a pesar de ello prefieren el uso de la web debido a su tamaño mucho mayor. Otra opción es la propuesta por Hashimoto et al. (2011), quienes extraen paráfrasis a partir de definiciones. Nos parece un enfoque muy interesante, ya que las definiciones cuentan con contextos ricos en conocimiento en los que suelen especificarse las relaciones mediante patrones de conocimiento, que muchas veces adquieren la forma de paráfrasis verbales (p. ej. *causado por*). En nuestra tesis combinamos el uso de corpus para la extracción de paráfrasis verbales, debido a su limpieza y especificidad, con el de la web, cuando el corpus no ha arrojado resultados. En muchas ocasiones, estas paráfrasis forman parte de contextos definitorios.

Sin embargo, hay autores como Smith et al. (2014: 138-139) que no están a favor de las paráfrasis. Alegan que, por ejemplo, *butter cookie* no es simplemente «a cookie containing butter». Es cierto, pero como comentábamos anteriormente, se necesita información pragmática adicional para captar el universo semántico del CN. Sin embargo, gracias a las paráfrasis alcanzamos una mayor especificidad que la proporcionada por las relaciones. Por otro lado, somos conscientes de que las paráfrasis no siempre son fáciles de obtener, ya que a menudo la omisión de formantes del CN complica su extracción, como comentamos con más detalle en la Sección 4. Por tanto, en esta tesis, más que considerar las paráfrasis como una alternativa a las taxonomías de relaciones, las entendemos como un método para acceder a ellas. Es decir, consideramos que las relaciones semánticas y las paráfrasis son enfoques complementarios. En primer lugar, la extracción de varias paráfrasis nos permite comprender con más precisión la asociación de conceptos de un CN, para después plasmar esta asociación mediante una relación semántica. Como mencionábamos arriba, estas relaciones favorecen la generalización, por ejemplo, sobre la formación y traducción de CN, que son nuestros objetivos principales.

3.4.4 Variación en los compuestos nominales

Un aspecto ineludible cuando se analizan CN es su frecuente variación. Sin embargo, antes de abordar el papel de este fenómeno en los CN, es necesario recordar que, durante mucho tiempo, la Teoría de General de la Terminología (Wüster 1968, 1979) negó la existencia de variación en el lenguaje especializado, como hemos comentado en la Sección 2.2.1.1. A diferencia de la lengua general, se defendía que los términos y los conceptos pertenecientes al discurso especializado eran unívocos y estables, de forma que se evitaba la ambigüedad. Así, se entendía la variación como un obstáculo para la comunicación especializada.

Sin embargo, esta visión idealista, que no representaba la realidad del lenguaje especializado, pronto encontró respuesta por parte de las nuevas teorías de la terminología, de corte comunicativo y cognitivo, que subrayaron la naturaleza cambiante de los términos y los conceptos (Cabré 1993, 1999; Tercedor Sánchez 1999; Temmerman 2000, 2007; Freixa 2002, 2006; Costa 2017; León Araúz 2017). Así, aunque menos presente que en la lengua general (Shreve 2001; Freixa 2002; Sanz Vicente 2011; Kageura 2015), la variación desempeña un papel fundamental en el lenguaje especializado, lo que se ha ido demostrando gracias al uso de corpus (Fernández Silva 2018). Esta se produce a veces con un propósito determinado (Bowker 1998; Kerremans 2017; Freixa y Fernández Silva 2017; Gledhill y Pecman 2018) y, otras, constituye un indicador del carácter novedoso de los conceptos (Cabré 1993; Picton 2011).

Desde entonces se han llevado a cabo diferentes estudios que han dado cuenta de la riqueza de la variación en el discurso especializado, como los de Tercedor Sánchez (1999, 2011), Montero Martínez (2002), Freixa (2002, 2006), Sanz Vicente (2011), Fernández Silva (2011, 2013, 2016, 2017, 2018), Drouin et al. (2017) o Daille (1999, 2001, 2017), entre otros. En gran parte de estos trabajos, en concreto los que estudian las distintas denominaciones de un concepto, se ilustra la variación con ejemplos de términos compuestos, pues, a pesar de que hay CN lexicalizados que no la admiten (p. ej. *horizonte de predicción*), lo cierto es que se trata de un fenómeno bastante habitual en los CN especializados (p. ej. *implantación/instalación/establecimiento de parques eólicos*).

La variación puede ser denominativa o conceptual. La variación denominativa, a menudo conocida como *variación terminológica* cuando se da en el discurso especializado, supone el uso de diferentes denominaciones para designar un mismo concepto (Geeraerts et al. 1994), p. ej. *wind power* y *wind energy*. Por otro lado, la variación conceptual se da cuando un término designa más de un concepto, es decir, tiene varios significados (Geeraerts et al. 1994). Es el caso de *inflammation*, que puede referirse a una función

fisiológica, una enfermedad o la parte del cuerpo que está inflamada (Gangemi et al. 2000). En esta tesis nos centramos en la variación denominativa, por ser la más habitual en los CN especializados. Como afirma Collet (2003), la mayoría, si no todos los términos compuestos, presentan diversas estructuras, por lo tanto, constituyen un ejemplo claro de este tipo de variación.

Tradicionalmente, para hablar de las causas por las que se produce la variación denominativa se han distinguido entre factores del usuario —que dan lugar a variación temporal, geográfica y social— y del uso —son el campo, el modo y el tenor— (Gregory y Carroll 1978). Sin embargo, consideramos que existen otros factores que también pueden producir variación denominativa. Así, adoptamos la clasificación de Freixa (2006), quien defiende que las causas de la variación denominativa pueden ser (i) dialectales, (ii) funcionales, (iii) discursivas, (iv) interlingüísticas o (v) cognitivas.

- Las causas dialectales se basan en el origen geográfico, cronológico y social de los hablantes. Por ejemplo, *ascensor* se traduce como *lift* en inglés británico y *elevator* en inglés americano.

- Las causas funcionales atienden al campo (área temática), el modo (canal de comunicación) y el tenor (cercanía entre el emisor y el receptor). Así, un oftalmólogo podrá hablar a un paciente de *inflamación de la córnea* y, a otro oftalmólogo, de *queratitis*.

- Las causas discursivas pueden originar cambios estilísticos (por ejemplo, para evitar la repetición o emplear formas más expresivas o creativas), así como cambios retóricos. Con respecto a este último punto, Collet (2003) sostiene que los CN contribuyen a la cohesión textual al adaptar su forma según el contexto en el que se encuentran. Freixa y Fernández Silva (2017) y Fernández Silva (2018) defienden que la variación denominativa intratextual, que estudian principalmente en términos compuestos, es uno de los principales mecanismos de cohesión gracias a su función de reiteración. Por otro lado, Fernández Silva (2016), en otro estudio sobre la variación denominativa intratextual, indica que las variantes que enfatizan diferentes perspectivas de un concepto contribuyen a la construcción de conocimiento en el texto. Por su parte, Gledhill y Pecman (2018), quienes también estudian la construcción de conocimiento en el texto, centrándose en la variación morfosintáctica, señalan que los CN en inglés del tipo N+of+N se utilizan para ofrecer contextos ricos en conocimiento, mientras que la forma N+N se utiliza para introducir información nueva.

- Las causas interlingüísticas de la variación se deben al contacto entre lenguas y se dan, entre otros casos, cuando coexiste un término local (*equipo informático*) con el préstamo adquirido de otra lengua (*hardware*).

- Las causas cognitivas, por último, se basan en las diferentes conceptualizaciones o motivaciones del usuario. Por ejemplo, mientras que *aerogenerador* y *sistema de generación eólica* enfatizan la generación (eléctrica) a partir del viento, *sistema de conversión de energía eólica* destaca la conversión de la energía del viento (en energía eléctrica).

Por tanto, las variantes denominativas resultan de estos factores, que pueden concurrir. Además, los casos de variación denominativa pueden mostrar una distancia semántica variable entre sí, como apunta Fernández Silva (2018), quien propone una clasificación de las variantes denominativas basada en su separación semántica. A pesar de que esta tipología se puede aplicar a todos los términos, se centra específicamente en los CN, como casos prototípicos donde brota este tipo de variación. Destacamos las valiosas aportaciones de esta autora a lo largo de su trayectoria, quien aborda en profundidad el tema de la variación terminológica, a menudo estudiada en CN por medio de la semántica de sus formantes, por lo que resulta de gran interés en nuestra tesis. Así, según el parecido conceptual entre las variantes, Fernández Silva (2018: 8-9) distingue entre distancia semántica mínima, media y máxima.

En los casos de distancia semántica mínima, las variantes no conllevan cambios en el contenido conceptual, son semánticamente equivalentes. Por una parte, incluye la variación formal, que puede implicar cambios gráficos (*distal estuarine system* > *DES*), morfológicos (*vínculo inseguro* > *vinculación insegura*) o morfosintácticos (*programa preventivo* > *programa de prevención*). Por otra parte, añade la sinonimia, que puede darse en todo el término (*estudio* > *investigación*) o en algún formante del CN (*vínculo inseguro* > *apego inseguro*).

La distancia semántica media se da en las variantes que implican cambios en el modificador (es decir, se asume que son CN). Señala, por un lado, las reducciones (*nanofósiles calcáreos* > *nanofósiles*) y, por otro, la adición o supresión de una característica no definitoria (*huella de terópodo* > *huella de dinosaurio terópodo*) y el uso de una característica definitoria diferente (*grupo clínico* > *TDAH group*). Las características no definitorias o las de rasgos diferentes están relacionadas con la multidimensionalidad, como veremos en la Sección 3.4.4.2, y con las características que conforman el concepto, como comentaremos al respecto de los microcontextos en la Sección 5.1.2.

Por último, en los casos de distancia semántica máxima se producen cambios en el núcleo o en ambos formantes del CN. Estas variantes son las más frecuentes en su estudio (con una presencia del 59 %) y son las que requieren un mayor esfuerzo cognitivo, ya que la categoría semántica del concepto cambia. Por una parte, puede darse un cambio de categoría, entendido no como una categoría diferente sino como el énfasis de una determinada perspectiva de conceptualización (*domo* > *estructura dómica*). Por otra, puede usarse una categoría hiperonímica (*embarazadas primigestas* > *población*) o hiponímica (*relaciones de amistad* > *díadas de amistad*). Por último, en ocasiones es posible modificar la configuración conceptual y reemplazar tanto el núcleo como el modificador por categorías que no están relacionadas taxonómicamente (*paternidad tradicional* > *modelo hegemónico de paternidad*), siendo este último tipo el que implica una mayor distancia semántica.

A raíz de esta clasificación, nos centraremos en las dos categorías principales de variación denominativa, que implican de menor a mayor distancia semántica. Nos referimos, pues, a la variación formal y la variación cognitiva (León Araúz 2017), que encuentran en los CN sus estructuras predilectas.

3.4.4.1 Variación denominativa formal

La variación denominativa formal supone modificaciones en la forma del término, sin implicaciones cognitivas, ya que no se modifica la representación del concepto. Siguiendo la tipología de variantes formales de Fernández Silva (2018), distinguimos entre cambios gráficos, morfológicos y morfosintácticos.

Los cambios gráficos se dan, por ejemplo, en los CN que permiten su fusión o separación ortográfica (*gearbox* > *gear box*), el uso de guiones (*fixed speed wind turbine* > *fixed-speed wind turbine*) o el empleo de mayúsculas en los formantes que cuentan con una sigla (*squirrel cage induction generator* > *Squirrel Cage Induction Generator* [*SCIG*]) (Sanz Vicente 2011). Este tipo de cambios ortográficos puede atender a razones geográficas, como en *synchronisation* (inglés británico) y *synchronization* (inglés americano) (Faber y León Araúz 2016), o por el contrario, al uso de términos con raíces neoclásicas que admiten diversas transliteraciones (*glycemia* vs. *glycaemia*) (Grön y Bertels 2018: 45).

Por su parte, los cambios morfológicos pueden consistir, como señala Sanz Vicente (2011), en el cambio de prefijo o sufijo (*generador síncrono* > *generador sincrónico*), la alternancia entre singular y plural en el modificador (*rango de velocidad* > *rango de velocidades*), la variación entre formas con o sin artículo (*power output of wind turbine* >

power output of the wind turbine) o la sustitución de una preposición por otra (*procedimiento de autorización* > *procedimiento para la autorización*).

Por último, los cambios morfosintácticos constituyen una fuente importante de variación y consisten en la alternancia de estructuras derivadas de los diferentes procesos de formación de términos. Este tipo de variación nos interesa especialmente, ya que los cambios morfosintácticos se dan con frecuencia en las dos lenguas que estudiamos. Por ejemplo, en inglés algunos CN del tipo N+N se pueden sustituir por sus equivalentes de la forma Adj+N (*atom bomb* > *atomic bomb*). Aunque las lenguas germánicas suelen preferir estas estructuras compactas, a menudo cuentan también con estructuras equivalentes de postmodificación (*aspirin synthesis* > *synthesis of aspirin*[27]) (Gledhill y Pecman 2018). En las lenguas romances, la variación morfosintáctica permite alternar entre la modificación adjetival y la preposicional (Daille 1999, 2001, 2017; Maniez 2009; Sanz Vicente 2011; Fernández Silva 2018), como vemos en *transporte aéreo* y *transporte por aire*.

Los adjetivos que se pueden sustituir por un sintagma preposicional suelen ser adjetivos relacionales que, como hemos visto en la Sección 3.3, representan conceptos y pueden establecer relaciones semánticas con otros elementos (*environmental degradation* > degradation *affects* environment). Las variantes que incluyen estos adjetivos contribuyen a la economía del lenguaje y elevan la especialización del discurso (Maniez 2009), ya que muchos de estos adjetivos relacionales son cultismos que proceden del griego o el latín (p. ej. *enfermedad hepática* pertenece a un nivel más elevado que *enfermedad del hígado*). Por ello, son más comunes en las lenguas romances y destacan por su importancia en los campos de especialidad, cuyos términos provienen en gran medida de las lenguas clásicas (Daille 1999, 2001, 2017; Maniez 2009).

Aunque el conocimiento etimológico juega un papel fundamental en la inferencia de estos adjetivos, los hablantes no nativos pueden tener problemas para identificar el adjetivo adecuado. Por ello, a menudo se ha defendido que las herramientas lingüísticas deberían vincular el sustantivo y el adjetivo que se deriva de este, si es el caso (Levi 1978; Daille 1999). Consistiría, por ejemplo, en relacionar *guerra* con *bélico*. Nos parece una opción interesante, pues esta información contribuiría a mejorar la producción de los usuarios.

[27] Gledhill y Pecman (2018: 30) consideran que estas alternancias son dos tipos diferentes de términos compuestos, más que variantes de la misma estructura. Sin embargo, en esta tesis consideramos que ambas formas representan CN (ya estén premodificados o postmodificados).

Por otro lado, la variación morfosintáctica plantea dificultades relacionadas con la elección de las variantes, ya que debemos pensar cuándo se puede sustituir una estructura por otra. Así, encontraremos casos en los que el sustantivo no cuente con una variante adjetival (*divorce law*) y otros en los que se den las dos posibilidades (*crime law* > *criminal law*), a las que hay que añadir la postmodificación preposicional (*law of crimes*). De nuevo, las convenciones de esta variación plantean problemas especialmente a los hablantes no nativos. Por ejemplo, de cara a la producción en inglés, los hablantes de lenguas romances, que no suelen utilizar estructuras del tipo N+N, dudan entre esta opción y sus variantes adjetivales (Adj+N), preposicionales (N+Prep+N), cuando existen. En muchos casos, la elección depende del usuario y del contexto de comunicación, ya que suele ser una cuestión de estilo o de género lingüístico (Maniez 2014; Gledhill y Pecman 2018). Sin embargo, hay ocasiones en las que, aun existiendo varias formas, una es la preferida, como sucede en *digestive system*, mucho más empleada que *digestion system* (Levi 1978: 224). En estos casos, la opción preferida se ha establecido por el uso y resulta difícil de deducir si no se dispone de información adicional (Levi 1978; Daille 1999). Por tanto, sería recomendable incluir indicaciones en los recursos lingüísticos para guiar a los hablantes no nativos en la producción textual.

Por otro lado, se suele afirmar que, cuando conviven varias formas, el proceso de lexicalización está en curso, mientras que cuando se da solo la forma compacta (N+N o N+Adj), el concepto científico emergente se ha estabilizado (Daille 2001; Maniez 2009; Gledhill y Pecman 2018). Como veremos en la Sección 3.4.5, la coexistencia de varias formas se ha relacionado a menudo con la inestabilidad del concepto y, así, con su carácter neológico (Cabré 1993; Picton 2011). De este modo, se suele decir que la forma compacta (N+N o N+Adj) goza de un mayor estatus terminológico, ya que las formas preposicionales están sujetas a una mayor variación debido a cambios en la preposición y, sobre todo, en el artículo, que puede ser definido o indefinido, o no aparecer (*instalación de aerogenerador/instalación del aerogenerador/instalación de un aerogenerador*).

A este respecto, cabe destacar los trabajos de Daille (1999, 2001, 2017), quien considera variación la modificación que pueden sufrir los CN con adjetivos relacionales y sus variantes preposicionales. Así, determina que los CN con adjetivos relacionales admiten muy poca variación sintáctica, siendo esta principalmente la coordinación (*produit agricole et produit alimentaire* > *produit agricole et alimentaire*), mientras que las variantes preposicionales admiten un grado de variación mucho mayor, por ejemplo mediante la inserción de elementos (*filtration membranaire* > *filtration frontale sur membrane*) (Daille 1999: 8). Estos resultados refuerzan la estabilidad de los CN con

adjetivos relacionales frente a las variantes preposicionales, que presentan una mayor inconsistencia. Sin embargo, como Collet (2003), consideramos estos aspectos como transformaciones contextuales más que como casos de variación, ya que no se corresponden con los dos tipos de variación: no se trata de variación denominativa porque el concepto cambia, ni de variación conceptual porque la denominación difiere.

A pesar de la relevancia de este tipo de variación en los CN, consideramos que sus características y, en particular, sus diferentes usos deberían abordarse con mayor profundidad. Por una parte, ya hemos mencionado que los CN formados por adjetivos se han ignorado durante mucho tiempo en la literatura, ya que la mayoría de los estudios se han centrado, y siguen haciéndolo, en los CN formados por sustantivos. Por otro lado, es cierto que existen algunos trabajos sobre la variación morfosintáctica, pero estos han abordado lenguas como el inglés o el francés, sin prestar demasiada atención al español. Por ejemplo, Levi (1978) estudia la alternancia entre N+N y Adj+N en inglés. Maniez (2009) analiza algunas características semánticas del núcleo en CN franceses del tipo N+Adj y N+Prep+N. Daille (1999, 2001, 2017) describe las modificaciones a las que están sujetas las variantes N+Adj y N+Prep+N en francés. Por último, Gledhill y Pecman (2018) exploran el uso de las estructuras N+N y N+Prep+N en inglés, centrándose en su función cognitiva y comunicativa. Sin embargo, se han obviado los factores discursivos que pueden declinar la balanza hacia el uso de una u otra forma, un aspecto que nos parece fundamental para la correcta producción. En definitiva, nos interesa saber cuándo se puede o debe utilizar cada forma, si es que tenemos elección. Además, lenguas como el español, en las que este tipo de variación está muy presente, también deberían estudiarse en profundidad, pues como decimos, el uso de una u otra variante puede tener implicaciones cognitivas, comunicativas o estilísticas (Cabezas García y Chambó en prensa).

En otro orden de cosas, aunque no se considera variación formal, nos parece oportuno mencionar aquí la variación léxica, que junto con la variación formal, son los cambios que entrañan menor distancia semántica (Fernández Silva 2018). Se trata de la sustitución por un sinónimo que no constituya una variante formal (*power generation*, *power production*). Este tipo de variación se da frecuentemente en los CN (Freixa 2002) y permite evitar repeticiones, aunque también puede responder a preferencias de uso o al contexto situacional de los hablantes, entre otros aspectos.

Del mismo modo, antes de abordar la variación denominativa cognitiva, cuyas variantes son las que implican la mayor separación semántica (Fernández Silva 2018), no

podemos olvidar la variación por reducción, que también está muy presente en la lengua especializada. La distancia semántica de estas variantes es media (Fernández Silva 2018) y uno de sus principales ejemplos es la sustitución del CN por sus siglas. Estas son omnipresentes en el discurso científico-técnico y a menudo se emplean más que la forma desarrollada, debido a su capacidad de aligerar los CN extensos. Además, muchas siglas y abreviaturas en inglés se incorporan al discurso especializado de otras lenguas, porque son las formas más conocidas en la comunicación científica (*DFIG*, sigla de *doubly fed induction generator*, se utiliza también en español). Autores como Freixa (2002) las han considerado un tipo de variación gráfica y, por tanto, de variación formal. Sin embargo, para nosotros constituyen más bien variantes por reducción, debido a que los cambios gráficos implican este proceso en concreto. Sin embargo, consideramos que ambas concepciones son aceptables.

Otra forma de reducción que no supone la reducción completa del CN es la sustitución por el hiperónimo (Bowker 1998), que se considerará variación si se da en el mismo texto (Fernández Silva y Kerremans 2011). En este sentido, Collet (2003) explora la variación de CN especializados en francés mediante la reducción de formantes, a la que califica de *contextual term reduction*, y en la que distingue dos tipos: la anáfora elíptica y la elisión léxica. La anáfora elíptica consiste en la sustitución por el hiperónimo que acabamos de mencionar y que solo tendrá validez dentro del mismo texto (*service mobile aéronautique par satellite* > *service*), mientras que la elisión léxica se basa en la omisión de otros formantes, sin que el CN se reduzca únicamente al núcleo y, por tanto, se considera como variante también fuera de las fronteras del texto (*antenne en reseau* > *antenne reseau*) (Collet 2003: 9-11). Un ejemplo de elisión léxica que encontramos a menudo en nuestros términos consiste en la sustitución de varios formantes por su sigla (*permanent magnet synchronous generator wind turbine system* > *PMSG wind turbine system*).

En resumen, en este apartado hemos abordado los tipos de variación denominativa formal, que no implican cambios en el concepto. Así, consideramos como variantes puramente formales las que implican cambios gráficos, morfológicos o morfosintácticos. Existen también las variantes léxicas o las que se forman por reducción, que algunos autores consideran variantes formales. A pesar de que lo que cambie no sea exactamente su estructura, las consideramos más variantes formales que cognitivas, al no implicar distintas conceptualizaciones, como sí ocurre en las modificaciones que veremos en el siguiente apartado.

3.4.4.2 Variación denominativa cognitiva

La variación denominativa también puede resultar de las diferentes conceptualizaciones y motivaciones de un mismo concepto (Freixa 2006; Faber y León Araúz 2016; Fernández Silva 2018). Así, la variación cognitiva alude a los cambios en los términos producidos por el carácter dinámico del conocimiento especializado (Barsalou 2003; Temmerman 2000, 2007; Faber 2012; Fernández Silva 2018). De este modo, entendemos este tipo de variación como un reflejo de la variación conceptual, al implicar diferencias de conceptualización, si bien se distingue de esta en que el concepto sigue siendo el mismo.

La variación denominativa cognitiva puede entrañar cambios en la configuración del concepto (*espèce capturée, produit de la pêche*), en la categoría semántica a la que pertenece (*territoire maritime, milieu marin*) o en su nivel de abstracción (*professionnel, producteur, pêcheur*), en la selección de dimensiones diferentes (*bancos marisqueiros, bancos naturais*) o distintas características de una misma dimensión (*zone de production, zone de récolte*), del mismo modo que se puede incluir una característica adicional a nivel intratextual (*activité aquacole, activité économique aquacole*), o expresar la misma información mediante un número diferente de constituyentes (*concesionario, titular de concesión*) (Fernández Silva 2013: 31-33).

Lejos de producirse por causas estilísticas o de registro, como veíamos en la variación formal, esta variación proviene de factores cognitivos, que pueden ser internos o externos, y que a menudo convergen. Los factores cognitivos internos están relacionados con la naturaleza de los conceptos y pueden entrañar diferencias de categorización o de estructuración conceptual, como hemos visto en los ejemplos anteriores. Por su parte, los factores cognitivos externos están ligados a la diversidad de contextos temporales, sociales o situacionales (Fernández Silva 2011, 2017: 16), que pueden implicar, por ejemplo, la enfatización de una dimensión o la inclusión de una característica adicional. Además, la variación puede tener consecuencias cognitivas, ya que se representan diferentes elementos del contenido del concepto por medio de las variantes y, de este modo, el receptor adquiere un mayor conocimiento del campo (Fernández Silva 2017).

A menudo se ha señalado la influencia de las diferentes disciplinas en la variación cognitiva (Bowker 1997; Shreve 2001). Como indica Fernández Silva (2013), no solo las distintas disciplinas estructuran de forma diferente una misma realidad, sino también los miembros de una misma disciplina, ya que estos tienen referencias, conocimientos, intereses y objetivos diferentes. Así, mientras que un profesor de ciencias naturales hablará probablemente de *windmill*, un ingeniero preferirá hablar de *wind turbine*. En este sentido, se ha destacado el papel fundamental que juega el punto de vista en la variación

denominativa de tipo cognitivo, como lo demuestra Fernández Silva (2013). En un interesante estudio, esta autora explora su incidencia en la variación denominativa mediante el análisis de los patrones conceptuales de CN del ámbito de la pesca. Sostiene que, según el punto de vista que adopte un especialista en un determinado contexto, este percibirá y destacará unas determinadas características como distintivas del concepto. Así, observa que el concepto PRODUCTO PESQUERO se puede denominar en seis marcos diferentes según el punto de vista adoptado: producción, comercialización, consumo, actividad pesquera, gestión de los recursos y conservación de los recursos. Además, advierte que los puntos de vista suelen variar según las disciplinas; por ejemplo, el punto de vista de la comercialización del pescado se encuentra sobre todo en textos de economía (Fernández Silva 2013: 27).

El punto de vista está relacionado directamente con la multidimensionalidad, que ocupa una posición central en la variación cognitiva. Tanto la naturaleza dinámica de los conceptos[28] como los contextos temporales, sociales o situacionales se aúnan en la multidimensionalidad, que consiste en la clasificación de los conceptos según distintas características (Bowker 1997, 1998; Kageura 1997). Cada conjunto de características constituye una dimensión y, cuando un concepto se puede organizar según distintas facetas o dimensiones, se dice que es multidimensional (Bowker 1997, 1998; Kageura 1997). Bowker (1997: 133) ilustra este fenómeno con el clásico ejemplo de *vehicle*, que puede clasificarse atendiendo a las siguientes características: (i) medio de transporte (*land/air/water vehicle*), (ii) método de propulsión (*motorized/non-motorized vehicle*) o (iii) tipo de carga (*passenger/cargo/passenger and cargo vehicle*).

Así, las variantes denominativas resultantes de la multidimensionalidad reflejan distintas facetas del mismo concepto y suelen ser CN (León Araúz 2017). En efecto, como señalan Meyer y Mackintosh (1996) estos términos compuestos constituyen el soporte ideal para el estudio de la multidimensionalidad. A continuación, reproducimos la Tabla 3 elaborada por León Araúz (2017), que representa de forma muy clara algunas de las dimensiones que pueden destacar los términos, en su mayoría CN pertenecientes al área de la Psiquiatría. Como se observa, en este caso se trata de variantes de diferentes conceptos.

[28] Como señalan Faber y León Araúz (2016), existe la multidimensionalidad intracategorial e intercategorial. La que nos ocupa aquí es la de tipo intracategorial, en la que un CN puede tener distintos hiperónimos que representan el mismo concepto pero enfatizan diferentes dimensiones o niveles de granularidad. Por otro lado, en la multidimensionalidad intercategorial, el cambio de categoría implica conceptos diferentes y constituye un tipo de variación conceptual.

Dimension	Term variant
+Discoverer	Korsakoff's psychosis
+Symptom	burning-mouth syndrome
+Cause	alcohol-induced amnestic disorder
+Body_part	teeth grinding
+Patient	boxer's dementia
+Result	bedwetting
+Intensity	mild cognitive impairment
+Time	short-term insomnia
+Location	prison psychosis

Tabla 3: Ejemplos de dimensiones enfatizadas por variantes (tabla original de León Araúz 2017: 223)

Aunque la multidimensionalidad no cobró relevancia hasta la llegada de las nuevas teorías de la terminología, que aportaron una visión mucho más realista y descriptiva que la que se tenía hasta entonces, no tardaron en surgir diferentes estudios que analizaban su papel en la variación cognitiva. Ejemplo de ello son los trabajos de Bowker (1997, 1998) y Tercedor Sánchez (2011), que destacamos por su estudio de la influencia de la multidimensionalidad en la variación denominativa, enfocados en CN en terminología.

Bowker (1998: 490) señala que la estructura de los CN de más de dos formantes puede variar según la dimensión que el emisor considere más relevante en el contexto:

> When an author uses a compound term to describe a particular concept, the structure of that term is often influenced by the dimension that is considered most significant to the discussion at hand. For example, if a particular scanner has the characteristics of having a flatbed design and a colour scanning capability, the author who is primarily concerned with the scanner's colour capability will refer to the concept as a *colour flatbed scanner,* while the author who is primarily interested in the scanner's design will refer to the same concept as a *flatbed colour scanner*.

De tal modo, se prefiere la variante que comienza por la dimensión de la que se está hablando en el contexto. Bowker (1998) concibe este énfasis de una u otra dimensión en el CN como un acto voluntario que otorga precisión al lenguaje especializado, un punto en el que coincidimos, ya que en nuestros CN también se hace uso de esta alternancia que contribuye a destacar unas determinadas características conceptuales.

Por su parte, Tercedor Sánchez (2011: 183) destaca la complementariedad de la variación terminológica y la multidimensionalidad, llegando a considerarlas dos caras de la misma moneda. Asimismo, esta autora señala que podemos acentuar diferentes aspectos de un concepto mediante sus variantes denominativas, gracias a la multidimensionalidad. De forma muy interesante, en su estudio presenta imágenes a un grupo de sujetos, que deben poner nombre a esas representaciones. Para el concepto PISTA

DE PATINAJE prefirieron con diferencia el término *skating rink* frente a *ice rink*, es decir, primó la función frente al material. Por el contrario, en el concepto BOTAS DE AGUA la dimensión prioritaria no fue la función, sino la percepción, ya que se prefirió el término *rubber boots* (Tercedor Sánchez 2011: 188). Finalmente, llega a la conclusión de que no se da un patrón claro en la variación cognitiva, sino que suelen destacarse distintas dimensiones.

Por tanto, además del punto de vista que adoptemos, la selección de una u otra dimensión vendrá determinada por el contexto y las características o preferencias del hablante. En este sentido, las distintas dimensiones suelen generar conjuntos de términos. Por ejemplo, según la velocidad de un aerogenerador, podemos hablar de *variable speed* o *fixed speed wind turbines*; según su eje existen *horizontal* y *vertical axis wind turbines*; y según la orientación del rotor, tendremos *upwind* o *downwind turbines*. Como vemos, las distintas dimensiones, reflejadas en las variantes denominativas, dibujan el contenido del concepto, de manera que resulta fundamental considerar todas las dimensiones, pues de lo contrario omitiremos una parte importante del conocimiento.

En resumen, en este apartado hemos pasado revista a los principales tipos de variación que afectan a los CN, en concreto, la variación denominativa de tipo formal y la de tipo cognitivo. La variación es un rasgo fundamental del discurso especializado, por lo que debería reflejarse en los recursos terminográficos. En este sentido, su mera representación no es suficiente, sino que abogamos por la inclusión de indicaciones que guíen al usuario hacia una producción eficiente (Faber y León Araúz 2016; Kerremans 2017). Desgraciadamente, aún queda mucho camino por recorrer hasta la plena integración de este tipo de información, aunque algunas herramientas como VariMed (Tercedor Sánchez et al. 2014), recurso de terminología médica centrado en la variación, ya han empezado su andadura en este camino.

Por otra parte, hemos señalado los factores que influyen en la variación denominativa, que son de tipo dialectal, funcional, discursivo, interlingüístico o cognitivo (Freixa 2006). Frente a esta concepción de la variación, consideramos que este fenómeno representa también la inestabilidad formal o semántica de un nuevo concepto (Cabré 1993; Picton 2011). La variación guarda, por tanto, una relación directa con la neología, como veremos en el siguiente apartado.

3.4.5 Neología en los compuestos nominales

En el análisis de los CN no podemos obviar su papel en la formación de neologismos (García Palacios y Sanz Vicente 2012). Como veremos, una de las principales formas de

crear términos nuevos consiste en la especificación de un concepto, formando así un CN. La neología está, por tanto, directamente relacionada con los CN y con la formación de términos, pues la mayoría de las veces, un término se forma debido al surgimiento de un nuevo concepto y, a su vez, estas nuevas unidades deben atender a los patrones de formación de términos de cada lengua.

En las áreas especializadas, los avances científico-técnicos conllevan la creación de términos nuevos o *neologismos* para designarlos. Por ejemplo, la frecuencia de los términos relacionados con la energía eólica aumentó considerablemente en la década de los 80, cuando este tipo de energía se desarrolló (Cabezas García y Faber 2017c). Así, la neología y la terminología están directamente relacionadas, como lo han demostrado diferentes estudios (Kageura 2002; Sanz Vicente 2011, 2012a, 2012b; Humbley y García Palacios 2012; Cabré et al. 2012; Pecman 2012, 2014; Fernández Domínguez 2016).

En su sentido más amplio, los neologismos terminológicos (también denominados *neónimos* según la noción de *néonymie* de Rondeau 1981) se entienden como términos formados recientemente en una lengua. Por tanto, el factor temporal resulta fundamental en el estudio de la neología (Rey 1976). Sin embargo, como Cabré y Nazar (2011: 3), no entendemos este criterio como el uso por primera vez de un término, sino como el fuerte aumento en su frecuencia, desde niveles muy bajos hasta valores muy altos. No obstante, el criterio diacrónico no constituye el único indicador del carácter novedoso de un término, como señala Cabré (1993: 145) en una de las contribuciones más extendidas sobre las características de los neologismos. Esta autora identifica, además, criterios lexicográficos, psicológicos y de inestabilidad sistemática, que se dan cuando el término no aparece en diccionarios, se percibe como nuevo o presenta inconsistencia formal o semántica, respectivamente.

En esta tesis consideramos que un CN es neológico cuando cumple con el criterio diacrónico y el requisito de inestabilidad, dado que la frecuente carencia de CN en los recursos terminográficos no permite considerar el criterio lexicográfico (especialmente en los CN formados por más de dos o tres elementos, que a menudo no se incluyen). Del mismo modo, la dificultad de aplicación del criterio psicológico nos lleva a eludir este factor como determinante en la neología de los CN especializados. Esta dificultad se aprecia, por ejemplo, en los casos de neología semántica, en los que un término extiende su significado, como veremos a continuación. Por tanto, entendemos que un CN es neológico cuando representa un concepto nuevo (en el sentido de que su frecuencia refleja un pico de incidencia) y su forma o significado presentan una marcada inestabilidad.

Por ejemplo, el CN *generación de energía eólica* alude a la principal actividad del ámbito de la energía eólica, que empezó a llevarse a cabo en la década de los 80, cuando se impulsó este tipo de energía (Figura 3). Su inestabilidad se observa en sus numerosas variantes, en las que cambia el núcleo (*producción de energía eólica*); se producen diferentes cambios en el modificador (*generación de energía a partir del viento*, *generación de energía eléctrica a partir del viento*); o se combinan estas variantes (*producción de energía eléctrica a partir del viento*). Así, la propia estructura de los CN es más proclive a la variación que la que presentan los términos simples. Evidentemente, esta inestabilidad hace que se reduzca la frecuencia de los CN neológicos, lo que puede revelar su carácter novedoso, ya que la inestabilidad de un término se ha considerado el paso previo a su lexicalización, cuando gozan de una mayor frecuencia (Picton 2011; Pecman 2012). Por otro lado, las aclaraciones semánticas constituyen una muestra de que la inestabilidad propia de los neologismos afecta al significado (Faber y Navajas 2016). Variación y neología van, por tanto, de la mano.

Figura 3: Representación de *generación de energía eólica* en Google Ngram Viewer[29]

El carácter neológico de los CN especializados ha sido objeto de estudio en diferentes trabajos. Por ejemplo, Kageura (2002) explora la formación de nuevos CN en japonés a partir de otros existentes, centrándose en sus categorías y relaciones semánticas internas. Sanz Vicente (2011, 2012a, 2012b) lleva a cabo un análisis contrastivo de CN en inglés y español para analizar la influencia de la lengua inglesa en la formación de neologismos en español. Esta autora hace especial hincapié en los procedimientos de importación de términos al español y en la morfosintaxis de los equivalentes. Por su parte, Pecman (2012, 2014) realiza un estudio de CN neológicos en inglés y observa que, además de la tradicional función denominativa de los neologismos, estos pueden desempeñar una función retórica (Pecman 2012), como mecanismo para destacar el carácter novedoso de

[29] Una consulta en la herramienta Google Ngram Viewer, que presenta un gráfico diacrónico de las ocurrencias de una búsqueda en Google Libros, permite detectar picos de frecuencia en los términos, lo que señala su carácter neológico (Cabezas-García y Faber 2017c).

un concepto en la ciencia, y cognitiva (Pecman 2014), para explicitar el significado del neologismo. Por otro lado, Fernández Domínguez (2016) lleva a cabo un análisis contrastivo de la formación de neologismos en inglés y español (que en gran parte constituyen CN), centrándose en sus aspectos formales y semánticos.

En cuanto a los tipos de neología, se han propuesto diferentes clasificaciones, que difieren según la perspectiva que se adopte. No obstante, estas no son contradictorias, sino complementarias. En una de las propuestas más extendidas sobre formación de términos, Sager (1990) señala que, según el contexto en el que surjan los neologismos, puede hablarse de formación primaria o secundaria de términos (antes denominadas *néonymie d'origine* y *néonymie d'appoint* por Rondeau 1981). Por otro lado, Cabré (1999) habla de neología denominativa y expresiva, según su función. La neología denominativa surge ante la necesidad de nombrar un concepto (es, por tanto, la más conocida y estudiada), mientras que la neología expresiva responde al deseo de introducir nuevas formas en el discurso. Por ejemplo, en el estudio de Pecman (2012) comentado anteriormente, la autora investiga el uso intencionado de la neología, aplicada a CN, para expresar los aspectos innovadores de una investigación. Por otra parte, según el ámbito de aplicación, Rondeau (1981) y Humbley (2006) distinguen la neología terminológica de la general. Como hemos señalado, nos interesa la neología terminológica y, en concreto, nos centraremos en la dicotomía entre formación primaria y secundaria de términos, ya que tienen implicaciones en los procesos de formación de términos y en la traducción, dos de los objetivos de esta tesis. Sin embargo, es necesario señalar que no nos proponemos estudiar el carácter neológico de los CN, simplemente destacamos que es una característica importante de los CN especializados, en la que profundizamos debido a su relación directa con la formación de términos en inglés y español.

Así, los conceptos de formación primaria y secundaria de términos se basan en el contexto en el que se crea el término. Por una parte, la formación primaria se da de forma monolingüe cuando surge un concepto. Por tanto, el nuevo término no cuenta con unidades léxicas preexistentes sobre las que basarse, al margen de las reglas de formación de términos de la lengua en cuestión o los ejemplos de términos del mismo dominio (Sager 1990, 1997; Valeontis y Mantzari 2007). Dado que el inglés constituye la *lingua franca* de la comunicación especializada, la formación primaria de términos suele tener lugar en esta lengua.

Por su parte, la formación secundaria consiste en la creación de un término para un concepto existente. Esta puede deberse a la revisión de un término en una comunidad

monolingüe (p. ej. *telephone* pasó a ser *fixed telephone* tras el invento del *mobile telephone*, como apuntan Valeontis y Mantzari 2007: 4) o a la transferencia de conocimiento a otra comunidad lingüística que no dispone de término para ese concepto (Sager 1997), dándose así un contexto de traducción. En estos casos, se da una dependencia terminológica de la lengua origen que, como decimos, es en muchas ocasiones el inglés (Humbley y García Palacios 2012: 3). De tal modo, el término origen en esta lengua puede servir de base para la formación secundaria de términos (Sager 1990; Valeontis y Mantzari 2007), como sucede muchas veces en español (García Palacios et al. 2016).

Como vemos, tanto la formación primaria como la secundaria aluden a la creación de neologismos: ya sea en el contexto de formación de un concepto o en fases posteriores, el resultado es la creación de un término nuevo. En esta línea, Sager (1997: 28) clasifica los mecanismos de formación de neologismos en tres grupos: uso de términos existentes, modificación de términos existentes o creación de nuevas formas léxicas.

El uso de términos existentes consiste en atribuir un nuevo significado a un término, por lo que no entraña cambios formales. Este procedimiento se suele conocer como neología semántica (Guilbert 1974; Rey 1976; Cabré et al. 2012)[30] y puede percibirse mediante el cambio en el tipo de unidades con las que se combina un término (Meyer y Mackintosh 1996; Cabré y Nazar 2011; Renouf 2012). Por ejemplo, *vaca* en el sentido de *vaca loca* se combina con unidades diferentes a las que acompañan a *vaca* en su primera acepción (p. ej. *enfermedad, signo, síntoma*).

La neología semántica puede producirse de diversas formas. Una de ellas es la terminologización o uso de una palabra general para denominar un concepto especializado (p. ej. la *nariz*[31] es una parte del rotor de un aerogenerador). También se puede conceder una nueva acepción propia de un área a un término de otra disciplina, lo que se conoce como *pluriterminologización* (p. ej. el término *virus mutante* de medicina ha pasado a usarse también en informática). Las metáforas constituyen asimismo un recurso útil de creación de neologismos semánticos (p. ej. *squirrel cage induction generator* es un tipo de generador cuya denominación se basa en su forma), del mismo modo que los símiles (*U-shaped valley*). Estos, muy presentes en el lenguaje especializado en inglés, suelen introducirse mediante elementos como *-style, -like, -shaped* o *–type* y a menudo constituyen el paso previo a la lexicalización del CN (Sager 1997: 29). Aunque con menor

[30] Fernández Domínguez (2016) también reconoce este fenómeno, que denomina *extensión semántica* y que contrapone a la *neología* (neología formal), de lo que se deduce que no considera la extensión semántica como un tipo de neología.
[31] Este ejemplo también podría entenderse como un símil, pues alude a la semejanza con esa parte del cuerpo.

frecuencia, en español también se forman términos con símiles, como *valle en forma de U*, que también cuenta con la variante lexicalizada *valle glaciar*. Así, es habitual la convivencia de variantes derivadas de diferentes procesos de formación de términos.

La neología semántica se contrapone a los otros dos mecanismos: la modificación de términos existentes y la creación de nuevos términos, que se entienden como neología formal (Cabré y Nazar 2011), al implicar cambios en el término. Los siguientes procedimientos permiten la creación de neologismos mediante la modificación de términos existentes: (i) derivación, (ii) formación de CN[32], (iii) conversión y (iv) compresión (Sager 1997).

La derivación consiste en la adición de prefijos o sufijos, que suelen tomarse de las lenguas clásicas, especialmente en el ámbito de las ciencias (Sager 1997: 32). Tanto en inglés como en español, los prefijos suelen determinar un concepto y contribuyen a crear conjuntos de términos (p. ej. *downwind turbine*, *upwind turbine*). Los sufijos, por su parte, permiten nombrar un concepto mediante la modificación de la categoría gramatical de un término (*acid* > *acidify*) (Sager 1997).

La formación de CN permite crear lo que se denominan *neologismos sintagmáticos* (Guilbert 1974) mediante la especificación de una o varias características distintivas del nuevo concepto. Estas se añaden como modificadores del núcleo, que constituye el hiperónimo. El CN suele representar, por tanto, un hipónimo del núcleo (p. ej. *offshore wind farm* es un tipo de *wind farm* que se desarrolló en la década de los 90, periodo en el que su uso aumentó de manera considerable). En la expresión de estas características determinantes del núcleo entra en juego su microcontexto, que resulta determinante en la formación de CN, como veremos en la Sección 5.1.2. Este procedimiento de composición representa, junto con la derivación, el principal mecanismo de formación de términos en inglés y español (Fernández Domínguez 2016). Sin duda, a este papel destacado de la composición contribuyen las numerosas posibilidades de combinación que ofrece la lengua, así como la facilidad que ofrecen los CN para indicar el aspecto novedoso de un concepto (p. ej. *teléfono móvil* se diferenció de *teléfono* porque se podía llevar a diferentes partes). Por ello, todo estudio que aborde las características de los CN no puede obviar su relación con la neología.

[32] Sager (1997) diferencia los *compounds* (formados por premodificación) de los *phrasal terms* (formados por postmodificación). Como hemos señalado en la Sección 3.2, en esta tesis consideramos ambas formas como CN.

La conversión consiste en un cambio en la categoría gramatical de un término, sin que se produzcan modificaciones en su forma (p. ej. *circle* significa *círculo* y la acción relacionada: *rodear*). Según Sager (1997), es un procedimiento habitual en inglés, pero menos frecuente en otras lenguas que requieran sufijos para cambiar la categoría gramatical, como es el caso del español. Aunque es cierto que este mecanismo entraña cambios en el término (concretamente, en su categoría gramatical), consideramos que no implica modificaciones en su forma. De este modo, lo entendemos como un caso de empleo de términos existentes y, por tanto, un neologismo semántico, ya que el significado de una de las categorías gramaticales se extiende dando lugar a otro significado con la misma forma léxica.

La compresión genera neologismos mediante la formación de *blendings* (*stagnation* + *inflation* > *stagflation*) o siglas (*levelized cost of energy* > *LCOE*), principalmente. Estas últimas constituyen la forma más típica de compresión, que a menudo se utilizan para señalar el carácter especializado de CN extensos que pueden no percibirse como términos (Sager 1997).

Por último, se pueden crear neologismos mediante el desarrollo de nuevas formas léxicas (p. ej. *byte*), aunque este procedimiento es el menos frecuente (Sager 1997; Montero Martínez et al. 2001). Como decíamos, estos nuevos términos deben respetar las reglas de formación de la lengua en cuestión, aunque es habitual la influencia de la lengua origen cuando se dan en condiciones de formación secundaria en una nueva comunidad lingüística.

Así, cuando los neologismos surgen a raíz de la transferencia de conocimiento a otra comunidad lingüística, esta influencia de la lengua origen resulta patente en el elevado recurso a préstamos y calcos.

Los préstamos consisten en la importación de formas léxicas de la lengua origen. Estos pueden implantarse sin cambios en la lengua meta, como es el caso de la sigla *WECS* del inglés *wind energy conversion system* para referirse a un *aerogenerador* en español. Como señalan Faber y Navajas (2016), estos usos son habituales cuando se supone que el receptor está familiarizado con la lengua origen y el tema en cuestión. Por otra parte, los préstamos pueden naturalizarse y adquirir una forma más propia de la lengua meta. Por ejemplo, cuando se introdujo en francés el término *FAX* del inglés, se desarrolló el término *telecopieur*, más acorde a la lengua francesa (Sager 1997: 39). Esta naturalización también puede derivar en la coexistencia de ambas formas.

Los calcos, a menudo denominados *calcos semánticos*, consisten en la traducción palabra por palabra del término en lengua origen (*permanent magnet* > *imán permanente*). A pesar de que constituyen un intento de aclimatación a la lengua meta, las reminiscencias de la lengua origen son patentes en el significado y, habitualmente, en la forma del neologismo (Montero Martínez et al. 2001), por lo que a veces se han considerado una muestra del desconocimiento del traductor o terminólogo de la lengua meta. En cuanto a los calcos de estructuras poliléxicas, como los CN, se suele distinguir entre calcos de expresión, si respetan las estructuras sintácticas de la lengua de llegada (*power system* > *sistema de potencia*), o calcos estructurales, si introducen una estructura nueva impropia en la lengua meta (*grid side* > *lado red*) (Vinay y Darbelnet 1973). En los calcos pueden darse también varias vertientes de traducción, como propone Sager (1997: 40) con los ejemplos de *lingüística informática* y *lingüística computacional*, procedentes del francés y del inglés, respectivamente.

Cuando se pretende conseguir una mayor claridad semántica, es frecuente que se combinen ambos procedimientos, como vemos en *máquina DFIG* (sigla del inglés *doubly-fed induction generator*), en el que se combinan términos en lengua origen y lengua meta, y se alcanza una mayor concisión al especificar el hiperónimo *máquina*. Otro de los recursos consiste en la alusión a otra dimensión de significado, conocido como *modulación* (Vinay y Darbelnet 1973), p. ej. *gearbox* > *caja multiplicadora*. No obstante, los límites entre los diferentes mecanismos que hemos presentado no son siempre claros y en muchos casos se dan coincidencias entre varios. Por ejemplo, *upwind turbine* se ha formado tanto por derivación (*up* + *wind* > *upwind*) como por composición (*upwind* + *turbine*), en un procedimiento que se conoce como *parasíntesis*.

Por tanto, se aprecia una gran influencia de la lengua de formación primaria, que suele ser el inglés, en la formación de neologismos en otra lengua, como lo demuestran también los ejemplos de nuestro corpus (Sección 5.2). En este sentido, cabe destacar el papel de los organismos normalizadores de la lengua, que suelen velar por el mantenimiento de esta al margen de influjos externos. Es el caso, por ejemplo, de la Real Academia Nacional de Medicina española. En su diccionario se proponen neologismos en español para evitar la implantación del término en lengua extranjera y contribuir, así, al desarrollo del español como lengua vehicular de la Medicina. Coincidimos con esta perspectiva y abogamos por la propuesta de equivalentes cuando se carece de estos en una lengua, sin dejar de reconocer y describir los usos extranjeros que estén extendidos. Así, para que la comunicación sea efectiva, es fundamental que se actualicen los recursos terminográficos con neologismos, una tarea que no siempre se lleva a cabo.

De este modo, no es extraño que el traductor especializado se encuentre con la falta de inclusión de muchos equivalentes en recursos terminográficos. Esta tendencia es habitual, además de en el caso de los neologismos, cuando se trata de términos muy específicos o usos particulares del autor[33] (en cuyo caso coincidimos en lo inapropiado de su inclusión). Así, el traductor o terminólogo deberá crear un neologismo o proponer una traducción explicativa o descriptiva (Valeontis y Mantzari 2007; Cabré et al. 2012), lo que subraya el papel de estos profesionales como desarrolladores de la terminología en las distintas lenguas para impedir el vacío de conocimiento. En cualquier caso, deberá guiarse por las reglas de formación de términos de la lengua meta (Guilbert 1974), tanto morfosintácticas como semánticas.

Por ejemplo, para trasladar el concepto GEARBOX a español deberá ser consciente de que estructuras del tipo *caja-engranajes* o *caja engranaje* no son habituales en esta lengua, en la que se prefiere la modificación con adjetivos o sintagmas preposicionales (*caja de engranajes*). Además, si desea nombrar un equivalente mediante otra dimensión de significado, es necesario que conozca el microcontexto del concepto para determinar su potencial combinatorio. Así, *gearbox* puede traducirse por *multiplicador, multiplicadora* o *caja multiplicadora*, además de *caja de engranajes*, porque se trata de un sistema formado por engranajes que se encarga de aumentar la velocidad de giro del rotor. Como se sabe, tanto el contenido conceptual como las reglas lingüísticas son esenciales en toda formación terminológica. En el siguiente apartado comentaremos cómo la neología, además del resto de características que se deben analizar de los CN, pueden tener repercusiones en la traducción.

3.5 Traducción de compuestos nominales

Como hemos señalado en la Sección 3.4.5, los CN son el principal método de formación de términos en inglés (Nakov 2013), que constituye la *lingua franca* desde la que importamos términos al español. Además de ser fundamentales en inglés, los CN también abundan en español (Fernández Domínguez 2016), por lo que deberemos hallar su traducción en inglés con vistas a la comunicación científica internacional, por ejemplo, para la publicación de artículos de investigación. Sin embargo, las diversas posibilidades de combinación de términos para formar un CN, unidas a su frecuente carácter neológico y a su alto grado de especificidad cuando se incluyen distintas características mediante varios

[33] Cuando un término es utilizado solo en un texto o por un mismo autor, se suele hablar de *referente discursivo* (Guilbert 1974). En muchas ocasiones, los CN constituyen ejemplos de este tipo, probablemente debido a las posibilidades de combinación que ofrece la propia lengua mediante la composición, que facilitan estos usos propios en mayor medida que los otros mecanismos lingüísticos.

formantes, hacen que estos no siempre figuren en los recursos terminográficos. Por tanto, el traductor o terminólogo deberá desarrollar técnicas para hallar o proponer equivalentes de estos términos compuestos en otra lengua, una tarea que entraña dificultades debido a las características propias de estas unidades.

Como en toda traducción, el establecimiento de equivalencias de los CN debe pasar por un exhaustivo análisis sintáctico-semántico, que nos guiará hacia traducciones adecuadas en lugar de intentar conformarnos con la simple correspondencia de cada formante, pues como veremos, esta técnica nos aleja en muchas ocasiones del equivalente apropiado. Por tanto, a la hora de trasladar CN a otra lengua se debe atender a las características propias de estos términos compuestos, como su formación en las distintas lenguas, el *bracketing* o análisis de dependencias necesario cuando cuentan con más de dos formantes, la semántica de sus componentes, la relación que los vincula, y la variación y neología que a menudo presentan.

Por una parte, no podemos olvidar que los patrones de formación de términos de la lengua origen y la lengua meta no siempre coinciden. Así, mientras que en inglés se prefieren CN formados por premodificación, en los que un núcleo nominal suele estar modificado por otros sustantivos o adjetivos (*wind speed*), en español son más frecuentes las estructuras con postmodificación en las que el sustantivo se complementa con un adjetivo o un sintagma preposicional (*velocidad del viento*). De este modo, los CN del español son menos compactos, siendo habituales las estructuras parafrásticas que buscan la aclaración del concepto mediante la modificación con verbos o participios que plasman la relación semántica (*rotor disc* > *área barrida por el rotor*). La formación de CN en inglés y español también difiere en el mayor número de adjetivos relacionales presentes en español, debido a su origen romance. Por tanto, al traducir CN entre estas dos lenguas, se deberán identificar los patrones morfosintácticos adecuados en cada caso (*wind farm* > *parque eólico*).

Asimismo, se debe tener en cuenta que los términos equivalentes pueden tener un número variable de componentes, además de que no siempre consistirán en una traducción lineal palabra por palabra (Shahzad et al. 2000; Daille et al. 2004[34]) (*freno aerodinámico* > *flap*). Por tanto, la búsqueda de equivalentes mediante la suma de las traducciones de sus formantes no resulta ser la mejor opción, pues aunque existan equivalentes que respeten ese paralelismo, omitiremos otras variantes que se alejen del

[34] Sin embargo, en un artículo posterior (Harastani et al. 2013), Daille defiende la composicionalidad de los CN y su uso en la búsqueda de equivalentes (esto es, mediante la unión de los equivalentes de sus formantes).

término origen. El número de formantes del CN también determina la necesidad del *bracketing*, pues aquellos CN que cuentan con más de dos elementos requieren la aclaración de sus dependencias internas. Esta desambiguación facilita la comprensión del CN, por lo que determina también su traducción, ya que si pasamos por alto este análisis o lo realizamos de forma incorrecta, podemos incurrir en equivalentes erróneos. Podría ser el caso, por ejemplo, en *offshore wind turbine*. Si no detectamos que su estructura es *offshore* [*wind turbine*], podremos proponer un equivalente como *turbina de viento marino*, que transmite un concepto diferente al original, que sería *aerogenerador marino*.

Una vez establecidas las estructuras propias de cada lengua y las dependencias internas del CN, resulta fundamental conocer los *slots* que abre el núcleo para formar CN. Como veremos en la Sección 5.1.2, este microcontexto es importante no solo para conocer la carga conceptual del CN, sino también para detectar las distintas posibilidades de combinación del núcleo, pues ya hemos señalado que los equivalentes de los CN no siempre consisten en la traducción formante por formante y, en este sentido, pueden incluir otro *slot* diferente al del término original. Por ejemplo, en Termium Plus se indica la siguiente definición para *generating station*: «unit that converts some form of energy into electrical energy». Así, vemos que el término en inglés (*generating station*) alude a la generación de energía, mientras que su equivalente en español (*central eléctrica*) especifica la energía resultante. En esta línea, en una lengua se puede omitir un *slot* mientras que en otra se puede preferir su explicitación. Por ejemplo, *aerogenerador* alude de forma específica a la generación, algo que no ocurre en su equivalente *wind turbine*. Igualmente, un mismo *slot* puede reproducirse en las dos lenguas por medio de categorías o roles conceptuales diferentes (*rated power*, *rated output* > *potencia nominal*, *potencia nominal de salida*). Por tanto, es esencial conocer en profundidad el microcontexto del núcleo, para así acceder a sus posibilidades de combinación en la lengua origen y la lengua meta.

Además, debemos tener en cuenta particularidades como el sentido diferente al habitual que pueden tener los formantes de los CN (p. ej. *green energy*) y que puede no reproducirse en la lengua meta, o la prosodia y preferencia semántica, que también puede variar entre lenguas. En este sentido, la relación semántica del CN es también fundamental para comprender cómo se vinculan los formantes, por lo que su omisión puede derivar en problemas de comprensión y, con ello, de traducción. Aunque en la mayoría de los casos los equivalentes presentan la misma relación semántica (Sanz Vicente 2011; Bouillon et al. 2012), no siempre sucede así. Como indica Sanz Vicente (2011), estas modificaciones en la relación semántica de los equivalentes suelen deberse a cambios en la categoría

gramatical de un elemento del CN, la omisión de formantes o la inclusión de nuevos elementos en el término equivalente. Asimismo, cuando el equivalente es un término simple se suprimen las relaciones semánticas internas. En esta línea, Sanz Vicente (2011) analiza, además, la forma que adquiere una misma relación en los diferentes CN que la codifican, aunque no se aprecian tendencias claras, de lo que se deduce que la existencia de una determinada relación no garantiza la forma de los equivalentes.

Por otra parte, los diferentes patrones de formación de términos del inglés y el español conllevan un mayor uso de preposiciones en esta última lengua, de forma que se explicita en cierta medida la relación. No obstante, la preposición más recurrente es *de*, que no aclara la relación semántica, ya que puede aludir a prácticamente todo tipo de vínculo. Cuando se utilizan preposiciones en los CN en inglés, la más habitual también es *of*, pero debemos barajar las distintas equivalencias posibles para no incurrir en usos sucesivos de *of* o *de* que podrían sustituirse por preposiciones más específicas. Por ejemplo, aunque tanto *procedimiento de autorización de parques eólicos* como *procedimiento para la autorización de parques eólicos* son válidas, la segunda opción evita el encadenamiento de sintagmas preposicionales con la preposición *de* y alivia el discurso. En esta línea, como mencionábamos antes, al traducir hacia el inglés es necesario identificar que los CN de tipo preposicional probablemente adopten una forma de premodificación sin preposiciones (*sistema de generación de energía eólica* > *wind power generation system*), un aspecto del que no siempre se percatan los hablantes no nativos.

Por otro lado, la variación presente en los CN también puede influir en su traducción. Las distintas lenguas presentan diferentes grados de variación en función de sus propias características (por ejemplo, en inglés se da una mayor variación gráfica debido, entre otros factores, al uso habitual de guiones). Así, es importante conocer las variantes denominativas en la lengua origen, ya que ello facilita la identificación de variantes en la lengua meta (por ejemplo, cuando se especifican diferentes *slots*). En este sentido, al trasladar el término a la lengua meta, el traductor o terminólogo debe plantearse si se debe respetar esa variación, en la medida de lo posible, o si es necesario reducir las posibilidades para alcanzar una mayor consistencia. Como señalan Fernández Silva y Kerremans (2011), hay ocasiones en las que es necesario mantener la variación en la traducción, porque esta desempeña funciones cognitivas o comunicativas. En esos casos, la existencia de variación denominativa puede generar dudas al traductor o terminólogo sobre la variante más adecuada para cada contexto, ya que los recursos terminográficos no suelen incluir indicaciones de uso.

Así, el traductor o terminólogo deberá decidir, entre otras muchas cuestiones, si traduce un CN mediante su forma extendida o su sigla, o si utiliza un préstamo o una adaptación poco empleada en la lengua meta. Asimismo, tendrá que identificar la variante preferida, si es el caso. Por ejemplo, al traducir un CN de dos formantes hacia el inglés, deberá determinar si la forma más apropiada es N+N, Adj+N o N+Prep+N. En la misma línea, deberá detectar que, en la dicotomía *digestive system* vs. *digestion system*, la primera opción es la preferida (Levi 1978: 224). A este respecto, Miyata y Kageura (2016) señalan que la aparición del término en recursos terminográficos o su mayor frecuencia frente a otras variantes son indicadores de que una variante es la más establecida. En cualquier caso, la lengua meta suele contar con un mayor grado de variación denominativa, debido a las diversas formas de traducción o formación de términos, lo que suele desembocar en la coexistencia de variantes (Sager 1997; Sanz Vicente 2011; Miyata y Kageura 2016).

Por otro lado, la variación conceptual puede influir en la comprensión del significado correcto del CN. Además, la habitual tendencia a los calcos puede derivarse en el uso de equivalentes que no transmitan el mismo significado. En esta línea, las variantes resultantes de la multidimensionalidad permiten obtener un mayor conocimiento de las distintas caras del concepto, además de que pueden plasmarse también en variantes en la lengua meta. A este respecto, Bowker (1998: 490) señala que la forma de los CN que aluden a varias dimensiones depende de la dimensión más enfatizada en el contexto inmediato, siendo esta la que aparecerá en primera posición en el CN. Por ejemplo, entre las variantes *colour flatbed scanner* y *flatbed colour scanner*, se optará por la primera cuando el color sea el tema en cuestión. Por tanto, la multidimensionalidad habitual en los CN también puede determinar la forma que adquieran los equivalentes.

Por último, el carácter neológico de muchos CN también puede influir en su traducción. En primer lugar, el traductor o terminólogo observará que los CN neológicos no siempre figuran en los recursos terminográficos. En este sentido, es posible que no exista equivalente en la lengua meta, de manera que se debe decidir entre la importación de la lengua origen mediante préstamos, la asignación de una denominación en la lengua meta (mediante la creación de una nueva designación o el uso o modificación de términos existentes) o el desarrollo de una traducción explicativa. Así, al importar un nuevo concepto o término a una lengua meta, este suele recibir la influencia de la lengua origen, lo que se aprecia en el uso frecuente de préstamos y calcos, y refleja a menudo la inestabilidad semántica o formal característica de los neologismos. Además, se debe considerar que el procedimiento por el que se ha formado el neologismo en la lengua origen no tiene por qué corresponderse con el que ha generado su equivalente en la

lengua meta. Por otro lado, si el neologismo existe en la lengua meta pero no está documentado en las herramientas terminográficas, deberá desarrollar las pautas necesarias para localizar dicho equivalente en los recursos que estén a su alcance. Como vemos, las distintas características de los CN influyen en su traducción y hacen que esta no sea siempre una tarea sencilla. En la Sección 5.2, profundizaremos en estos aspectos que nos parecen de gran interés para la traducción y analizaremos su influencia en los CN de nuestro estudio, además de la posible existencia de otros factores que repercutan en la traducción de estos términos.

Por otra parte, al hablar de la traducción de CN no podemos pasar por alto el concepto de equivalencia, siempre recurrente en los estudios de Traducción y objeto de diferentes opiniones. Esto se debe a que la noción de equivalencia desde el punto de vista de la Traducción no coincide con la que se tiene desde la Terminología (Pimentel 2013; Kerremans 2017), debido a la naturaleza y objetivos diferentes de ambas disciplinas. Para comprender las dos concepciones, hemos de distinguir entre la equivalencia a nivel de palabra, de oración o de texto.

Mientras que los terminólogos se centran en la equivalencia a nivel de término, ya que su objetivo suele ser la recopilación de términos en los recursos, en la traducción se persigue la correspondencia a nivel de oración o de texto. Esto les lleva a buscar la equivalencia funcional de su traducción (Reiss y Vermeer 1984; Nord 1997), más que una equivalencia directa del término, que sería el caso de los terminólogos (Atkins y Rundell 2008). Por tanto, la equivalencia para los traductores es más amplia que para los terminólogos y permite un rango mayor de mecanismos de traducción. Así, en el marco de la oración o del texto podemos considerar equivalentes no únicamente aquellos que transmitan el mismo significado, sino los que reproduzcan la misma función o efecto que el texto original (Catford 1965; Nida y Taber 1982). De este modo, durante el acto de traducción de un texto se puede recurrir, entre otras opciones, al uso de hiperónimos o al empleo de otros equivalentes que reflejan distintas conceptualizaciones, por ejemplo como resultado de la multidimensionalidad (Fernández Silva y Kerremans 2011). Consideramos que estas opciones no se incluirían como equivalencias en los recursos terminográficos, sino que serían válidas a nivel textual.

La complicación de la equivalencia también radica en los diferentes niveles posibles de correspondencia: forma, sentido, estilo, función, efecto, etc. (Reiss y Vermeer 1984). La literatura sobre la equivalencia interlingüística se ha centrado tradicionalmente en la equivalencia de forma, sin atender mucho a la equivalencia de significado (Buendía

Castro y Faber 2016). Sin embargo, en terminología es esta la que nos interesa, como empezaron a reflejar los trabajos de Wüster (1979), Felber (1984) o Arntz y Picht (1995). Así, las equivalencias terminológicas se establecen a nivel conceptual, de forma que se debe determinar el grado de correspondencia entre el contenido conceptual de los posibles términos equivalentes (Sanz Vicente 2011; Buendía Castro y Faber 2016).

No obstante, es necesario tener en cuenta que la equivalencia total no siempre se da, por lo que Felber (1984) y Arntz y Picht (1995) identifican distintos grados de equivalencia: equivalencia plena (coinciden todos los rasgos conceptuales), intersección (coincide la mayoría de los rasgos conceptuales), inclusión (el concepto de la lengua meta es más amplio e incluye el concepto de la lengua origen) o equivalencia nula (el concepto no existe en la lengua meta). En esta tesis nos centramos en la equivalencia de los CN a nivel de término, ya que nuestro objetivo es la inclusión de equivalentes en recursos terminográficos, cuyo ámbito de aplicación es mayor que el de una única oración o texto. Estas correspondencias se basan en la coincidencia de las características conceptuales de los posibles equivalentes, que pueden presentar diferentes grados (Felber 1984; Arntz y Picht 1995). No obstante, durante las tareas de traducción de un texto, consideramos que puede adoptarse una postura más amplia con respecto a la equivalencia, como hemos comentado.

Así, la equivalencia de las unidades fraseológicas, como los CN, se ha abordado en profundidad desde el ámbito de la Lingüística Computacional. Esta disciplina ha investigado el reconocimiento automático de términos, principalmente destinado a la extracción monolingüe, aunque también con vistas a la extracción bilingüe, debido a su importancia para mejorar los sistemas de traducción automática. No obstante, siguiendo la tendencia habitual en los estudios de CN, los trabajos se han centrado principalmente en CN de la lengua general formados por dos elementos, mientras que los CN más extensos han recibido menos atención explícita, a pesar del interés y la mayor complicación que revisten. Mientras que la Lingüística Computacional se ha centrado en la extracción, los estudios terminológicos sobre la traducción de CN se han dedicado principalmente al análisis de equivalentes.

Tradicionalmente, ha sido habitual el uso de corpus paralelos para la extracción bilingüe (Déjean y Gaussier 2002; Daille y Morin 2005). Estos consisten en un conjunto de textos originales y sus traducciones, de manera que permiten la alineación y, con ello, facilitan la identificación de equivalentes. Sin embargo, presentan diversos problemas, como su escasez, especialmente en pares de lenguas que no incluyan el inglés, o la

marcada influencia del texto origen en la lengua meta, al tratarse de traducciones. Este aspecto beneficia, por ejemplo, el análisis de errores o de las distintas posibilidades de traducción, pero dificulta el estudio de los usos propios de la lengua. Frente a los corpus paralelos, cada vez se utilizan más los corpus comparables, que consisten en conjuntos de textos del mismo tipo y temática, escritos originalmente en cada lengua, que son mucho más asequibles y permiten el análisis de los usos reales en cada lengua.

A continuación presentamos algunos de los estudios que se han llevado a cabo sobre la traducción de las unidades fraseológicas, en general, y de los CN, en particular. Se aprecian dos corrientes de estudio: la comparación y análisis de traducciones, principalmente desarrollada desde la Terminología, y la extracción de equivalentes, llevada a cabo sobre todo desde la Lingüística Computacional.

En la primera vertiente, Salager-Meyer (1985) realiza un estudio comparativo de CN especializados formados por sustantivos en inglés y en español, centrándose en sus aspectos sintácticos y semánticos. Establece las principales estructuras morfológicas del inglés y la forma que adquieren en español (p. ej. N1+N2 > N2+*de*+N1). Asimismo, determina que la traducción de los CN no se puede generalizar, ya que depende en gran medida del usuario que realice la traducción y del contexto en el que esta se produzca.

Por su parte, Montero Martínez et al. (2001) estudian el papel del traductor científico en la formación de términos en la lengua meta. Para ello, se centran en colocaciones léxicas del inglés (principalmente formadas por un verbo y un sustantivo) y determinan que estas suelen trasladarse en forma de calcos sintácticos al español. Frente al tradicional rechazo con que se perciben los calcos, los consideran un mecanismo útil para el desarrollo de la terminología en la lengua meta. En un estudio posterior, Montero Martínez y García de Quesada (2003) utilizan un corpus comparable en inglés y español para analizar CN especializados en estas dos lenguas con vistas a la traducción. Exploran las diferencias y similitudes entre los equivalentes, mediante el uso de categorías y relaciones semánticas, y determinan que los patrones semánticos suelen coincidir en ambas lenguas. Las contribuciones de estas autoras nos parecen muy interesantes, debido al enfoque semántico que plantean, que se encuentra en la línea de nuestro análisis.

Destacamos también los trabajos de Quiroz (2008), quien se centra explícitamente en combinaciones especializadas formadas por tres o más elementos en inglés y español (tanto CN como combinaciones libres). Para ello, utiliza un corpus paralelo y realiza un análisis comparativo de la morfosintaxis y la semántica de estas unidades en las dos lenguas.

Girju (2007), por su parte, habla de la traducción de unidades fraseológicas en el sentido inverso: determina que las unidades fraseológicas de la lengua meta sirven para aclarar el significado de la combinación original en inglés, que resulta complicado debido a la compactación de sustantivos habitual en esta lengua. En este sentido, señala la capacidad aclaratoria de las preposiciones que se usan en los CN en lenguas romances (italiano, francés, español, portugués y rumano), algo que también indica Ó Séaghdha (2007b). En nuestra opinión, aunque las preposiciones pueden desempeñar una función de desambiguación en algunos casos (p. ej. *para* suele indicar la función o la finalidad), lo habitual es que las preposiciones en los CN del español no aclaren la relación semántica.

Fernández Silva y Kerremans (2011) estudian la variación terminológica en textos originales escritos en gallego y en sus traducciones al inglés. A pesar de que su objetivo no son los CN, los términos que estudian resultan ser en su mayoría combinaciones especializadas de este tipo. Así, establecen lo que denominan *índice interlingüístico de variación* para señalar la distancia entre los equivalentes. Nos parece un estudio muy interesante al abordarse la variación y la traducción desde una perspectiva semántica, pues se anotan con categorías semánticas tanto los CN como sus formantes.

Sanz Vicente (2011, 2012a, 2012b), como ya hemos comentado con anterioridad, se centra en la traducción al español de CN especializados del inglés. Utiliza un corpus comparable para analizar cómo ha influido el inglés en su equivalente español (especialmente, a nivel morfológico y sintáctico) y clasifica los equivalentes en español según el procedimiento de importación (préstamo, calco, paráfrasis, adaptación, etc.), por lo que se trata de un enfoque eminentemente neológico. En el corpus identifica los equivalentes del 80 % de los CN y, para el resto, propone traducciones según las técnicas principales observadas en ese 80 %, como veremos más adelante. Nos parece un trabajo muy acertado de cara a la traducción de términos compuestos, que quizás se beneficiaría de una perspectiva más semántica, aunque este enfoque se justifica por el objetivo principal del estudio: estudiar la influencia de los CN especializados del inglés en el español.

Otra investigación interesante es la de Tercedor Sánchez et al. (2013), quienes investigan el uso de calcos en la traducción de CN especializados del inglés al español. En concreto, se centran en lo que denominan *multiword cognates*, es decir, combinaciones léxicas que comparten forma y tienen un significado equivalente (p. ej. *Non-Hodgkin Lymphoma* > *linfoma no Hodgkin*). Para ello, plantean a alumnos que traduzcan CN en un periodo breve de tiempo, para después mostrarles tres opciones de traducción sin límite

de tiempo. Afirman que, en la traducción inicial, para la que no se ofrecían opciones, se produjeron más calcos, debido al escaso tiempo para traducir y a la concepción de los calcos como una opción segura, mientras que, en la segunda fase, cuando se proporcionaban tres equivalentes posibles, los sujetos optaron por otros equivalentes que no constituían calcos y que percibían como opciones mejores.

Parra Escartín et al. (2013) investigan los problemas que puede entrañar la representación de unidades fraseológicas en recursos terminográficos. Para ello, comparan las unidades fraseológicas especializadas (entre las que se encontraban CN) del inglés y del alemán con sus traducciones al español, utilizando un corpus paralelo. Determinan que las formas en español son más parafrásticas, por lo que su representación en recursos terminográficos plantea mayores problemas que las del alemán o el inglés.

Carrió Pastor y Candel Mora (2013) también emplean un corpus paralelo, esta vez de textos en inglés traducidos al español, para estudiar los equivalentes de CN especializados. Destacan la mayor longitud de los equivalentes en español, así como los casos en los que surgen diversas traducciones posibles, como muestra de la inestabilidad del concepto.

Maniez (2013) estudia cómo se trasladan al francés los adjetivos compuestos del inglés formados por un sustantivo y un participio (*decay-corrected*), propios de áreas especializadas. Se trata de uno de los pocos estudios que exploran este tipo de términos compuestos, que según el autor, suelen traducirse mediante adjetivos relacionales o calcos y cuya consulta en recursos como *Linguee* debe complementarse con la búsqueda en corpus paralelos y comparables.

Por otro lado, Salehi et al. (2014), a pesar de no dedicarse a la traducción de CN, integran este aspecto en su estudio. Estos autores proponen un método para detectar la falta de composicionalidad en unidades fraseológicas del inglés, que consiste en el uso de definiciones, sinónimos y traducciones de dichas unidades. De este modo, determinan que la presencia de los formantes o de la propia unidad fraseológica en sus definiciones, sinónimos o traducciones señala la composicionalidad de dicha combinación. A nuestro parecer, se debe tener en cuenta que, además de la suma de los formantes, el CN puede integrar rasgos adicionales de significado que, a nuestro parecer, hacen que el CN no sea composicional. De este modo, la simple aparición de los formantes en sus definiciones, sinónimos o traducciones no nos parece concluyente.

Balyan y Chatterjee (2015) se centran en la traducción de CN formados por sustantivos del inglés al hindi con vistas a su aplicación en traducción automática. Destacan que la traducción de CN de más de dos formantes no ha recibido demasiada atención, por lo que investigan CN de hasta cuatro formantes. Analizan las dependencias internas de estos términos y utilizan un algoritmo que extrae sus equivalentes basándose en las relaciones semánticas internas. Encuentran que, en hindi, las relaciones semánticas se perciben en la traducción, de modo que establecen cinco grupos de relaciones según los patrones de traducción de los CN que las codifican. Nos parece un enfoque muy intuitivo, ya que las características del hindi hacen que se encuentren regularidades, facilitándose así la traducción en función de la relación semántica del CN. Sin embargo, en los CN del español las relaciones no se plasman de forma tan regular (Sanz Vicente 2011), por lo que la traducción de estos términos debe basarse en características adicionales.

Por último, el trabajo de Fernández Domínguez (2016), que también hemos presentado con anterioridad, utiliza un corpus comparable para realizar un análisis contrastivo de la formación de neologismos en inglés y español, centrándose en sus aspectos formales y semánticos. A pesar de que el enfoque principal es el estudio de la neología, su comparación de términos en inglés y español constituye una gran contribución al estudio de la composición y sus equivalencias en inglés y español, dado que los neologismos constituyen en gran parte CN.

En general, los estudios sobre la traducción de CN suelen destacar la influencia de la lengua origen en la lengua meta. Como hemos señalado en la Sección 3.4.5, esta influencia se aprecia en los procedimientos de transferencia de los CN, entre los que abundan los préstamos y los calcos. Recordemos que los préstamos consisten en la importación de formas léxicas de la lengua origen, que pueden adaptarse o mantener su forma original. Estos son frecuentes, por ejemplo, en la importación de siglas, acrónimos y abreviaturas, que a menudo se trasladan en forma de préstamos sin adaptar. En algunos casos, se especifican mediante la anteposición del hiperónimo (Sanz Vicente 2011), p. ej. *máquina DFIG*. Los calcos, por su parte, representan la traducción palabra por palabra del término en lengua origen (*laser anemometer* > *anemómetro láser*). Los calcos de estructuras poliléxicas, como los CN, pueden ser calcos de expresión, si respetan las estructuras sintácticas de la lengua de llegada (*power system* > *sistema de potencia*) o calcos estructurales, si introducen una estructura nueva impropia en la lengua meta (*laser anemometer* > *anemómetro láser*) (Vinay y Darbelnet 1973). Como hemos señalado, Montero Martínez et al. (2001) perciben los calcos como una buena alternativa para los traductores de textos especializados, quienes mediante este mecanismo, introducen

nuevas formas de expresión en la lengua meta. Tercedor Sánchez et al. (2013), por su parte, señalan que esta técnica se utiliza como una opción segura cuando se dispone de poco tiempo para traducir.

Además de estos procedimientos, Vinay y Darbelnet (1973) proponen otros mecanismos de traducción: traducción literal, transposición, modulación, equivalencia, adaptación, expansión, reducción y compensación. En lo que respecta a la traducción de CN, un recurso muy habitual es la modulación (Vinay y Darbelnet 1973; Fernández Silva y Kerremans 2011), que permite nombrar el concepto aludiendo a otra dimensión de significado, lo que guarda una relación directa con la multidimensionalidad (p. ej. *photochemical smog, summer smog* > *smog fotoquímico, smog de verano*).

Por otro lado, cuando no existen equivalentes documentados del CN, una opción muy empleada son las paráfrasis explicativas, que constituyen un tipo de expansión, según la tipología de Vinay y Darbelnet (1973). Estas transmiten el sentido del término original incluyendo elementos que no figuraban en su forma (Sanz Vicente 2011). La expansión, por su parte, es un concepto más amplio que alude a la inclusión de elementos que estaban implícitos en el CN original (p. ej. *floating wind turbine* > *aerogenerador marino flotante*). En el lado opuesto encontramos la reducción, que consiste en la omisión de material del CN original en su equivalente (p. ej. *solar thermal collector* > *colector solar*).

Frente a la comparación de equivalentes, la otra vertiente principal que se ha investigado en la traducción de unidades fraseológicas ha sido la extracción de equivalentes. Dadas las diversas posibilidades de combinación de términos para formar un CN, su frecuente carácter neológico y el alto grado de especificidad que pueden alcanzar cuando se incluyen distintas características mediante varios formantes, es habitual que los recursos terminográficos no recojan todos los CN susceptibles de búsquedas. Por tanto, el traductor o terminólogo deberá desarrollar técnicas para hallar o proponer equivalentes de estos términos compuestos en otra lengua.

Cuando se trata de identificar equivalentes, estos especialistas suelen recurrir al uso de corpus, en los que resulta esencial manejar las técnicas de interrogación para obtener resultados específicos (Gouadec 2005). Además, Pimentel (2013) añade la importancia de la coincidencia semántica, el contexto colocacional, el conocimiento enciclopédico y el *tertium comparationis* en la búsqueda de equivalentes por parte de traductores y terminólogos. Por una parte, estos suelen analizar la coincidencia semántica mediante la elaboración de sistemas conceptuales, que tradicionalmente se confeccionaban de forma monolingüe para después proceder a su comparación. Asimismo,

el contexto colocacional les ayuda a identificar los equivalentes de términos polisémicos (p. ej. *bunch of flowers* [EN] > *bouquet* [FR]; *bunch of hair* [EN] > *touffe* [FR]), lo que subraya la importancia de su inclusión en los recursos lingüísticos (Atkins y Rundell 2008: 470). El conocimiento enciclopédico determina las búsquedas que se realizan para hallar el equivalente: desde consultas concretas cuando se intuye cuál puede ser la correspondencia, hasta búsquedas más amplias cuando no se disponen de indicios (Bergenholtz y Tarp 1995). Por último, mediante el *tertium comparationis*, los traductores o terminólogos comparan la situación de uso y los patrones colocacionales del posible equivalente con las del término original (Pimentel 2013).

Además de los terminólogos y traductores, la extracción bilingüe de términos se ha estudiado sobre todo por parte de lingüistas computacionales, debido a la importancia que reviste para los sistemas de traducción automática. Así, para la identificación de equivalentes resulta fundamental el uso de corpus, que pueden ser paralelos o comparables, como hemos comentado. En estos últimos se suele recurrir a la semántica distribucional para localizar equivalentes (Shahzad et al. 2000; Déjean y Gaussier 2002; Daille et al. 2004; Daille y Morin 2005; Sharoff et al. 2009). Esta se basa en la descripción del significado de un término mediante su distribución en un conjunto de contextos, de manera que la coincidencia de contextos indica la coincidencia semántica (como ocurre, por ejemplo, en los sinónimos) (Déjean y Gaussier 2002: 4). Así, para la identificación de equivalentes de traducción, se analiza el cotexto del término en la lengua origen, que se suele representar con vectores. Se parte de la idea de que dicho cotexto, traducido y analizado en la lengua meta, proporcionará el equivalente, ya que los términos equivalentes comparten los mismos cotextos (Shahzad et al. 2000; Déjean y Gaussier 2002; Daille et al. 2004; Daille y Morin 2005; Sharoff et al. 2009).

Uno de los estudios más conocidos es el de Shahzad et al. (2000), quienes parten de que los CN no siempre se traducen palabra por palabra, un aspecto que nos parece fundamental y que debe ser la base de toda traducción de los CN. Utilizan un corpus comparable, al que aplican un algoritmo que traduce el contexto del CN a la lengua meta para localizar los posibles equivalentes. Como hemos mencionado, esta idea basada en la semántica distribucional se ha aplicado en diferentes estudios para la extracción bilingüe de términos.

Uno de ellos es el de Déjean y Gaussier (2002), quienes también se basan en la tradicional coincidencia de elementos contextuales en un corpus comparable (inglés-alemán) para indicar una coincidencia de significado. Investigan la extracción de

equivalentes de términos simples y compuestos. Para estos últimos, buscan los elementos contextuales de cada formante por separado y utilizan como vectores solo los que coinciden entre todos los formantes del CN. Finalmente, si existe más de una equivalencia posible, se selecciona la de mayor semejanza conceptual. Trabajos como estos, a pesar de ser preliminares en la traducción de los términos compuestos, marcaron las líneas a seguir para diversos estudios que se sucederían y sirven de inspiración para la búsqueda de equivalentes de nuestra tesis (Sección 5.2.1).

Ejemplos de trabajos que siguieron la línea de la semántica distribucional son los de Daille et al. (2004) y Daille y Morin (2005), quienes presentan un método para la extracción de términos equivalentes simples y compuestos a partir de un corpus comparable en inglés, francés y español. Dicho método extrae los términos en cada corpus y agrupa las variantes denominativas, para después alinearlos con sus traducciones mediante la coincidencia de elementos contextuales de la que venimos hablando. Se trata de una propuesta muy atractiva, ya que comprende la falta de composicionalidad y las diferencias estructurales en los CN equivalentes, por lo que facilita su identificación.

En la misma línea, Sharoff et al. (2009) proponen un método computacional que utiliza un corpus comparable del inglés y el ruso para hallar equivalentes de unidades fraseológicas generales consideradas problemáticas por los traductores. Dicho modelo detecta las unidades fraseológicas en cada lengua para después alinearlas mediante la coincidencia de elementos contextuales.

Pimentel (2013), por su parte, presenta una metodología basada en la Semántica de Marcos (Fillmore 1977, 1982, 1985) para identificar equivalentes de verbos especializados. Utiliza un corpus comparable en francés y portugués, que utiliza para comparar la estructura argumental de los posibles equivalentes. Así, distingue entre equivalencias totales (cuando coincide el marco y la estructura argumental) y parciales (cuando existen discordancias, de forma que las equivalencias no son intercambiables en todos los contextos). Este trabajo sigue la estela de publicaciones anteriores (Pimentel y L'Homme 2011; L'Homme y Pimentel 2012), desarrolladas en el Observatorio de Lingüística Sentido-Texto de Montreal y aplicadas en sus diferentes herramientas terminográficas (Sección 3.6.2.6). En este caso, los resultados de Pimentel (2013) se aplican a la base de datos de terminología jurídica JuriDiCo. Nos parece una gran contribución, en línea con la propuesta que realizamos en esta tesis, ya que las proposiciones que subyacen en los CN también presentan una estructura argumental.

Un enfoque similar es el de L'Homme et al. (2018), quienes presentan una metodología para localizar traducciones de unidades fraseológicas especializadas (en concreto, colocaciones verbales) en un corpus comparable (inglés-francés). Este procedimiento consiste en traducir los verbos que se combinan con un determinado sustantivo y consultarlos en el corpus de la lengua meta. Como veremos en la Sección 5.2.1, en esta tesis planteamos un enfoque similar, que se diferencia del de L'Homme et al. (2018) en el objeto de estudio, así como en los elementos de búsqueda en el corpus de la lengua meta. Así, en lugar de centrarnos en unidades verbales, nuestro objetivo es la traducción de CN. Además, en nuestro caso no utilizamos únicamente las traducciones de los formantes del término en cuestión, sino que empleamos sus elementos contextuales, en línea con la semántica distribucional. Como estos autores, defendemos estas propuestas manuales por el valor añadido del trabajo manual de los terminólogos o traductores, ya que las contribuciones computacionales siempre requieren de la validación por parte de estos especialistas. Además, es frecuente que estos no manejen herramientas de procesamiento del lenguaje natural, por lo que necesitan técnicas manuales que puedan ejecutar.

Maniez (2008), por su parte, también propone soluciones manuales para los traductores, como el uso de la web para hallar los equivalentes de CN del inglés al francés. Para ello, se basa en el empleo del formante o formantes cuya traducción se conoce, además del uso de operadores como *. Así, para traducir CN como *gluten-sensitive enteropathy*, realiza la consulta "*entéropatie * gluten*", que arroja los siguientes resultados: *entéropathie au gluten*, *entéropathie d'intolérance au gluten*, *entéropathie induite par le gluten* y *entéropathie de sensibilité au gluten* (Maniez 2008: 163). Estas aportaciones nos parecen de gran utilidad para los traductores, gracias por una parte a la gran cantidad de información disponible en la web y, por otra, al desarrollo de técnicas manuales que puedan aplicar los traductores o terminólogos en sus tareas diarias. Sin embargo, consideramos que estos procedimientos podrían beneficiarse de búsquedas más avanzadas y específicas en corpus.

También mediante técnicas manuales, Sanz Vicente (2011: 457) identifica los equivalentes del 80 % de sus CN en inglés por medio de la generación de listas de los CN en inglés sin traducir para detectar posibles anglicismos, la búsqueda de la posible traducción al español del CN o de sus formantes, la generación de listas de los elementos que suelan coocurrir con el CN, la consulta de las traducciones documentadas de los formantes, o la búsqueda de los términos o raíces que presentan una ortografía similar en ambas lenguas. A pesar de la utilidad de las técnicas manuales para los terminólogos y

traductores y de la validez de estas técnicas para la extracción de equivalentes, nos parece que su aplicación es intuitiva y los resultados no son exhaustivos. Entre las opciones que se plantean, consideramos que la más útil es la generación de listas de los elementos que suelan coocurrir con el CN, al estar relacionada con la semántica distribucional.

Desde una perspectiva más computacional, Vintar (2010) propone un sistema de reconocimiento bilingüe de términos diseñado para la traducción entre el esloveno y el inglés, que aplica en corpus paralelos, aunque también puede utilizarse en corpus comparables si se dispone de un lexicón bilingüe. El sistema utiliza patrones morfosintácticos y clasificaciones estadísticas para proponer términos equivalentes en estas lenguas y se aplica tanto a términos simples como compuestos. Sin embargo, para la traducción de los términos compuestos se basa en la suma de los equivalentes de sus formantes, algo que no compartimos, ya que muchas veces los términos compuestos integran nuevos formantes o los suprimen, presentan un orden diferente de elementos o tienen como equivalentes términos simples, como ya hemos comentado. La propia autora reconoce esta limitación del sistema, que influye en la reducción de la precisión.

Asimismo, Bouamor et al. (2012) presentan una estrategia para detectar unidades fraseológicas equivalentes en un corpus paralelo francés-inglés del Parlamento Europeo, con vistas a mejorar el funcionamiento de los sistemas de traducción automática.

Harastani et al. (2013) exploran la traducción de CN especializados del francés en los que el núcleo está modificado por un adjetivo relacional (*cancer pulmonaire*) mediante la sustitución de dicho adjetivo por el sintagma preposicional equivalente (*cancer du poumon*). A continuación, traducen el CN de forma composicional e investigan su ocurrencia en un corpus comparable en inglés. A pesar de que obtienen una precisión del 86 %, consideramos que hay muchos CN para los que este enfoque no será válido, como aquellos que cuentan con algún formante lexicalizado y no permiten la variación a la forma preposicional (*parque eólico*, pero no **parque de viento*).

Por último, Semmar (2018) presenta un enfoque para extraer y alinear unidades fraseológicas a partir de un corpus paralelo inglés-francés. Las listas de equivalentes extraídas tras la alineación se filtran utilizando características lingüísticas de estas combinaciones. La finalidad de la extracción bilingüe es la mejora de los sistemas de traducción automática.

Como vemos, se dispone de grandes avances para la identificación de equivalentes realizados desde la Lingüística Computacional. Sin embargo, nos parece necesario aplicar

este desarrollo a la labor diaria de los terminólogos y los traductores, que a menudo no utilizan sistemas de procesamiento del lenguaje natural, lo que debe realizarse mediante un mayor conocimiento de las técnicas de interrogación de corpus. Frente a esto, también será imprescindible que estos profesionales cuenten con pautas para producir equivalentes cuando no se disponga de traducciones documentadas. En estos casos de formación secundaria de términos, ya comentamos las propuestas de Sager (1997) (Sección 3.4.5), que consisten en el uso de términos existentes, su modificación o la creación de nuevas formas léxicas. Arntz y Picht (1995), por su parte, también sugieren la creación de términos nuevos, además del recurso a préstamos o calcos, o el empleo de paráfrasis explicativas. Linder (2002) también propone pautas para la producción de términos equivalentes. En concreto, sugiere emplear términos únicamente de la lengua meta, no utilizar siempre la preposición *de*, evitar los CN demasiado largos (p. ej. mediante la omisión de elementos en las variantes denominativas a nivel textual) o recurrir a la transposición. No obstante, como señala Sanz Vicente (2011), esta propuesta no se basa en un análisis lingüístico profundo ni en la observación de patrones en un corpus.

Esto sí se observa en las indicaciones proporcionadas por esta autora (Sanz Vicente 2011), quien propone traducciones propias para los CN cuya traducción no ha conseguido identificar (el 20 % de sus términos). Estas propuestas se basan en los patrones que se observan en las traducciones encontradas, cuya existencia verifica en Internet y a la que otorga una eficiencia del 60 % (una cifra que se mejoraría si interviniese el traductor realizando ajustes estructurales). A partir de las regularidades observadas, Sanz Vicente (2011) ofrece pautas de traducción, basadas principalmente en la morfología y morfosintaxis de los equivalentes (p. ej. N+N en inglés suele traducirse por N+prep+N en español, siendo núcleo+modificador), pero también en aspectos generales percibidos en los equivalentes (p. ej. *fire* suele traducirse por *incendio* y no *fuego* en su corpus, abundan los calcos de expresión, se suele utilizar la preposición *de* para cualquier relación, etc.). Consideramos que este tipo de pautas son las que resultan de mayor utilidad de cara a la traducción de CN cuando no se dispone de equivalentes. Sin embargo, la inclusión de cuestiones semánticas beneficiaría sin duda el enfoque (si se considerara, por ejemplo, la influencia del *bracketing* o de la relación semántica interna). Del mismo modo, aunque este análisis presenta resultados excelentes para este dominio (la teledetección), puede no ser extrapolable a otros campos, por lo que consideramos que las pautas deben mantener un equilibrio entre la generalidad y la especificidad necesarias para aportar datos útiles. En cualquier caso, constituye una contribución indudable a las indicaciones tan necesarias para la transferencia de CN a otras lenguas.

En resumen, la traducción de los CN es una tarea difícil de abordar que debe realizarse desde la consideración de las características de estos términos en las dos lenguas. De este modo se podrán trazar equivalencias terminológicas que permitan alimentar los recursos especializados y facilitar, así, la labor de sus usuarios.

3.6 Representación de compuestos nominales en recursos lingüísticos

Como hemos mencionado con anterioridad, los CN se caracterizan, entre otros aspectos, por las dificultades de comprensión y traducción que plantean. Así pues, en el apartado 3.6.1. exploraremos la literatura existente sobre la representación de unidades fraseológicas, y en concreto de CN, en recursos lexicográficos y terminográficos[35]. Sin embargo, dado que la representación de CN no ha recibido demasiada atención (Cabezas García y Faber 2017a), a diferencia de otras unidades fraseológicas como las frases hechas o las colocaciones (Béjoint 1981; Louw 2006; Lew 2012; Parra Escartín et al. 2013; Buendía Castro 2013), en el apartado 3.6.2. nos centraremos exclusivamente en los recursos terminográficos y realizaremos un análisis de diferentes diccionarios y bases de datos especializados, prestando especial atención al tratamiento que reciben los CN. Nuestro objetivo es, pues, analizar la información que obtienen los usuarios cuando intentan resolver los problemas que plantean estas unidades, y detectar aciertos y errores con vistas al diseño del módulo de CN de EcoLexicon.

3.6.1 Las unidades fraseológicas en recursos lingüísticos

La representación de unidades fraseológicas plantea no pocos problemas a lexicógrafos y terminógrafos (Louw 2006). Estas dificultades radican, principalmente, en los criterios de selección, su ubicación en los recursos lingüísticos y el tratamiento microestructural que deben recibir, y desembocan en el tratamiento heterogéneo que reciben las unidades fraseológicas en general y los CN en particular (Montero Martínez y Buendía Castro 2012; Cabezas García y Gil Berrozpe 2018). Sin embargo, la mayoría de los autores coinciden en que todos estos problemas pasan por la determinación inicial de los usuarios que tendrá el recurso, ya que los diccionarios y demás herramientas lingüísticas deben responder a sus necesidades específicas (Atkins y Rundell 2008; L'Homme 2009; Bergenholtz y Tarp 2010; Nielsen 2010; Fuertes Olivera y Tarp 2014; *inter alia*).

Para definir las necesidades de los usuarios en los recursos lexicográficos o terminográficos, hay autores que apuntan que se deben determinar tanto las situaciones

[35] En este trabajo adoptamos la distinción entre lexicología/lexicografía como referidas a la lengua general y terminología/terminografía como relativas al lenguaje especializado. Véanse Bergenholtz y Tarp (2010) y Fuertes Olivera y Tarp (2014) (quienes prefieren hablar de lexicografía especializada en lugar de terminografía) para ahondar en la dicotomía *lexicografía especializada* vs. *terminografía*, que consideran (y consideramos) sinónimos.

en las que los usuarios pueden necesitar consultar un diccionario como los diferentes tipos de usuarios (Tarp 2008; Bergenholtz y Tarp 2010; López Rodríguez et al. 2012; Fuertes Olivera y Tarp 2014).

En cuanto a las situaciones en las que se hace necesario recurrir a un diccionario, Tarp (2008) distingue entre situaciones cognitivas, comunicativas, interpretativas y operacionales. Las situaciones cognitivas son aquellas relacionadas con la adquisición o mejora del conocimiento sobre un determinado tema, por ejemplo, cuando un traductor necesita documentarse para traducir un texto especializado. Por su parte, las situaciones comunicativas se dan cuando el usuario realiza una actividad lingüística, como la lectura, redacción o traducción de textos (L'Homme y Leroyer 2009). Por último, encontramos las situaciones interpretativas y operacionales. Se refieren a la habilidad de interpretar una situación mediante la observación (situación interpretativa) y responder cuando sea necesario (situación operacional), por ejemplo, al aprender el funcionamiento de una máquina (Fuertes Olivera y Tarp 2011: 148). A nuestro juicio, estas últimas podrían aunarse en una sola, pues están interrelacionadas. Si nos centramos en situaciones en las que se busque obtener respuestas en un diccionario u otra herramienta lingüística, consideramos que las situaciones más frecuentas serán las cognitivas y las comunicativas, especialmente en el caso de los traductores y los terminólogos.

Evidentemente, para determinar las necesidades de los usuarios será necesario también conocer sus diferentes perfiles. La clasificación más extendida es la que, partiendo de un determinado dominio, distingue entre: el público general o lego, que no posee conocimientos sobre el tema; el semiexperto, que está iniciado en el tema, aunque sin dominarlo; y el experto, que conoce en profundidad el dominio (Sager 1990; Pearson 1998). En el caso de los usuarios de recursos especializados, Sager (1990: 197-199) propone una clasificación más pormenorizada: (i) expertos en el tema; (ii) mediadores profesionales, como redactores técnicos, traductores e intérpretes; (iii) terminólogos y/o lexicógrafos especializados; (iv) especialistas de la información y la documentación; (v) planificadores lingüísticos; (vi) usuarios profesionales de la lengua, como editores, profesores o lingüistas; y (vii) usuarios generales de la lengua. A esta clasificación, Cabré (1999) añade los profesionales de la ingeniería lingüística y la inteligencia artificial. En el caso de los recursos terminográficos, dado que existe una gran variedad, podemos decir que el rasgo común de sus usuarios es que presentan un interés particular por los dominios especializados (Bergenholtz y Tarp 2010).

Las necesidades de los usuarios son, pues, el resultado del tipo de usuario y la situación en la que necesite utilizar una herramienta lingüística. Estas necesidades determinarán la concepción de la misma (Tarp 2008; L'Homme 2009; Bergenholtz y Tarp 2010; López Rodríguez et al. 2012). Así, el análisis de las necesidades influirá en diferentes aspectos, como el tipo de aplicación (glosario, diccionario, base de datos, etc.), la información que debe contener (definiciones, equivalentes, fonética, etc.) (Cabré 2003: 182) y, en el caso de las unidades fraseológicas, su ubicación en el recurso.

En este sentido, el lugar más conveniente para las unidades fraseológicas en los recursos lingüísticos siempre ha sido objeto de debate (Béjoint 1981; Heid 2001; Louw 2006; Pawley 2007; Lew 2012; Buendía Castro 2013; *inter alia*). Tradicionalmente, todo tipo de unidades fraseológicas se han venido agrupando como sublemas de otra entrada principal, ya que de este modo se consigue ahorrar espacio en los recursos en papel. No obstante, muchas herramientas, sobre todo digitales, conceden entradas propias a estas unidades (p. ej. EcoLexicon, *Diccionario de Términos Médicos* o DiCoEnviro, como veremos a continuación). En este sentido, compartimos la opinión de autores como Louw (2006), que abogan por un tratamiento diferente en función del tipo de unidad fraseológica del que se trate y, además, creemos que el soporte del recurso (impreso o electrónico) facilitará un tratamiento u otro.

Cuando se decide incluir las unidades fraseológicas como sublemas, surge otra cuestión: en la entrada de cuál de sus formantes deben incluirse (Lew 2012). Por una parte, se encuentran autores como Béjoint (1981), quien defiende su inclusión en la entrada del núcleo cuando se trata de CN (p. ej. *wind farm* se incluiría como sublema de *farm*). Por otro lado, otros autores prefieren ubicar las unidades fraseológicas bajo la primera palabra con contenido semántico (Louw 2006; Lew 2012) (p. ej. *wind farm* se incluiría en este caso como sublema de *wind*). Como bien indica Levi (1978), estas opciones quedarían incompletas al 50 %, ya que solo encontraríamos la unidad fraseológica en la entrada de uno de sus componentes. Además de esto, ya hemos mencionado anteriormente los problemas de identificación que pueden plantear los CN, por lo que limitar su inclusión a la entrada de uno de sus formantes reduce las posibilidades de que el usuario encuentre la información que busca. Por ello, Lew (2012) aboga por el uso de referencias cruzadas en las entradas de ambas partes. No obstante, debido a la limitación de espacio en los recursos en papel, recomienda que la información se desarrolle solo en una de las entradas (la de la primera palabra con contenido semántico).

En nuestra opinión, los CN deberían incluirse como sublemas de la primera palabra en los recursos en papel, ya que el usuario encontrará más fácilmente la información, debido a la tradición lexicográfica del orden alfabético. Consideramos que esto ocurre especialmente en el caso de CN formados por más de dos elementos. Por ejemplo, si desea buscar *horizontal axis wind turbine*, lo más común es que acuda a la letra *h*, porque no cualquier usuario identificaría *wind turbine* como el núcleo o concepto hiperónimo. No obstante, se deberían incluir referencias cruzadas en las entradas del resto de los formantes, si bien el tratamiento microestructural completo se proporcionará en la entrada del primer término. De esta forma se ahorra espacio, algo que se hace indispensable en este soporte, y se vinculan los CN con términos relacionados de algún modo. De lo contrario, su inclusión como lemas, unida al orden alfabético reinante en los recursos impresos, podría ocasionar que, en muchos casos, los CN aparecieran separados de sus formantes incluso por varias páginas. La inclusión como lemas sería una buena opción si el recurso estuviera organizado conceptualmente y pudiéramos observar las relaciones entre términos, como en el *Lexique de coocurrents. Bourse et conjoncture économique* (Cohen 2011), donde las combinaciones léxicas se clasifican según las fases del ciclo económico (Lorente Casafont et al. 2017: 216). Además, cada CN que aparezca como sublema debería mostrarse en líneas diferentes, para facilitar su identificación, y se debería evitar la sobrecarga de la entrada con una larga lista de sublemas, ya que ello dificultaría la consulta del recurso (L'Homme y Leroyer 2009).

Dado que el papel presenta otras desventajas, como el tiempo que requieren las búsquedas o la dificultad de actualización de la información, nos inclinamos por los recursos electrónicos. Estos cuentan con diferentes ventajas, como el acceso más rápido y sencillo a la información, la mayor capacidad de almacenamiento (y, por tanto, menores restricciones de espacio), la facilidad de actualización de los datos y, principalmente, las diferentes puertas de acceso a la información (Nkwenti-Azeh 2001; Bergenholtz y Tarp 2010; Kübler y Pecman 2012; Lorente Casafont et al. 2017), que, a nuestro juicio, constituyen las líneas futuras de la lexicografía y la terminografía.

Estas diferentes búsquedas incluyen, cada vez más, el acceso a la información por medio del significado. Por ejemplo, el *DELAC Dictionary of Serbian Compounds* (Krstev et al. 2006) permite buscar compuestos según sus categorías semánticas (p. ej. si buscamos la categoría +Zool obtendremos los compuestos en serbio que designan un animal). Otro buen ejemplo lo constituye el *Pattern Dictionary of English Verbs* (Hanks y Pustejovsky 1995), donde podemos buscar una lista de categorías semánticas y obtener los verbos con los que coocurren y viceversa (p. ej. la categoría COLOUR muestra verbos como *shade*, *paint*

o *dye*). Así pues, para la representación de CN en recursos electrónicos, proponemos su inclusión como lemas con entrada propia, a la que se podrá acceder desde diferentes puntos según la búsqueda que se realice. Además, sería recomendable indicar la relación con otros términos, para favorecer así la comprensión de los conceptos y la adquisición de conocimiento (L'Homme 2012; Faber et al. 2014). En la Sección 5.3 desarrollaremos con mayor detalle nuestra propuesta de inclusión de CN en recursos especializados, en concreto, en EcoLexicon.

Una vez establecida la ubicación de las unidades fraseológicas en los recursos lingüísticos, surge la controversia acerca del tratamiento microestructural que debe otorgarse a las palabras o términos en general y a las unidades fraseológicas en particular. Ya Levi (1978: 229) indicaba que la información que ofrecen muchos diccionarios es inconsistente, redundante, incompleta y/o incorrecta. Un ejemplo de ello se puede encontrar en una de las prácticas más habituales en la representación de unidades fraseológicas: su enumeración en listas, muchas veces omitiendo parte de sus formantes y prescindiendo de información adicional de tipo semántico o lingüístico (Levi 1978: 231).

En este sentido, hay autores como Lorente Casafont et al. (2017) que defienden un tratamiento microestructural diferente de las unidades fraseológicas en función de su transparencia. Como Louw (2006), pensamos que la transparencia de los conceptos no es objetivable, por lo que debería incluirse siempre un mínimo de información adicional, principalmente de tipo semántico (Levi 1978; Nkwenti-Azeh 2001; Buendía Castro y Faber 2014). En los recursos especializados, la información semántica (por ejemplo, a modo de definiciones o relaciones semánticas) resulta de gran importancia, ya que muchos de sus usuarios son terminólogos, traductores, etc. que necesitan documentarse (Nkwenti-Azeh 2001).

Teniendo en cuenta a estos usuarios mayoritarios, consideramos que estas herramientas también deben incluir información lingüística, como la sintaxis, las combinaciones de palabras (Buendía Castro y Faber 2014; Lorente Casafont et al. 2017) o la variación (Levi 1978; Stubbs 2001), que debería complementarse con información relativa a su uso. Sin embargo, ni la combinatoria de las palabras o términos (Pavel 1993; Montero Martínez y Buendía Castro 2012; Buendía Castro 2013) ni la variación (Stubbs 2001) suelen formar parte de las entradas.

Como apunta Hutcheson (2001), también nos parece necesario indicar el dominio temático al que pertenece el concepto, pues de este modo se facilita la interpretación del significado y, además, se pueden realizar consultas más específicas (en el caso de recursos

que integren los dominios temáticos en los campos de búsqueda). No obstante, la asignación de dominios no deja de ser una tarea ardua, debido a la complejidad de los conceptos, que pueden pertenecer a más de un dominio, encontrarse en la frontera de varios, etc.

Por supuesto, cuando el recurso esté dirigido a usuarios que manejen una segunda lengua, será fundamental incluir la equivalencia de las unidades fraseológicas en la lengua meta. Además, consideramos que el contexto y las notas de uso pueden resultar de gran utilidad tanto en la comprensión como en la producción lingüística. Finalmente, apostamos por la inclusión de imágenes, con las que se enriquecen las entradas, y se facilita la representación y adquisición del conocimiento (Reimerink et al. 2016). En el apartado 3.6.2. analizaremos varios recursos terminográficos y abordaremos la información que se proporciona en las entradas de cada uno de ellos.

3.6.2 Análisis del tratamiento de compuestos nominales en recursos terminográficos

En esta sección nos centramos en la representación términos compuestos en forma de CN. Para ello, elegimos seis recursos terminográficos, en los que analizamos los siguientes aspectos: (i) acceso a la información, (ii) ubicación de los CN e (iii) información de las entradas, en concreto las que describen CN. Dado que decidimos estudiar recursos de distintas especialidades para obtener un mayor panorama, no podemos comparar un mismo lema en cada recurso. Nuestro análisis se basa, pues, en entradas que sean representativas de cada herramienta.

Los recursos que analizamos son los siguientes:

- *Diccionario Español de Ingeniería* (Real Academia de Ingeniería 2014).

- *Diccionario de términos médicos* (Real Academia Nacional de Medicina 2012).

- *Oxford Dictionary of Biochemistry and Molecular Biology* (Cammack et al. 2006).

- *Diccionario técnico inglés-español español-inglés* (Beigbeder Atienza 2006).

- *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish* (Hidalgo Simón 2004).

- *DiCoEnviro. Le dictionnaire fondamental de l'environnement* (Observatoire de linguistique Sens-Texte 2019a).

Estos fueron elegidos por su interés terminográfico respecto a dos cuestiones: el tratamiento de CN y el enfoque terminográfico adoptado en el recurso. Así, seleccionamos recursos en los que, a nuestro juicio, estas cuestiones se abordaban de manera adecuada, inadecuada o diferente, dando así cuenta del distinto tratamiento que reciben los CN, así como de la variedad de recursos existentes: impresos o en formato digital, monolingües o bilingües/multilingües que incluyeran español e inglés, y de temática relacionada o de otras áreas del saber.

Como veremos a continuación, la representación de CN en herramientas especializadas sigue las tendencias comentadas en la sección 3.6.1. Así, observamos que los CN se plasman de manera muy heterogénea en los recursos terminográficos. En líneas generales, podemos decir que cuentan con entradas propias cuando se trata de herramientas digitales, mientras que, en el caso de los recursos impresos, reciben un tratamiento desigual: unas veces se presentan como sublemas (siguiendo la tradición de los diccionarios impresos) y otras, disponen de entrada propia. Observamos también que el orden alfabético sigue primando, si bien los recursos digitales empiezan a ofrecer diferentes puertas de acceso al contenido. La información que se facilita suele ser la definición o el equivalente (en el caso de recursos bilingües), además de sinónimos y breves notas de uso, en algunos casos. Muy pocas veces se explicita la relación existente con otros términos, lo que sería de gran utilidad para delimitar el marco y facilitar la adquisición de conocimiento.

Comenzaremos analizando recursos monolingües de español (*Diccionario Español de Ingeniería* y *Diccionario de términos médicos*) e inglés (*Oxford Dictionary of Biochemistry and Molecular Biology*), para después pasar a herramientas bilingües de ambos idiomas (*Diccionario técnico inglés-español español-inglés* y *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish*) y finalizar con una herramienta multilingüe (*DiCoEnviro. Le dictionnaire fondamental de l'environnement*).

3.6.2.1 Diccionario Español de Ingeniería

El *Diccionario Español de Ingeniería* es un diccionario monolingüe que incluye términos relacionados con la Ingeniería en español. Elaborado por la Real Academia de Ingeniería en 2014, se presenta en versión online (1.0) (http://diccionario.raing.es) y su consulta es abierta y gratuita. Incluye más de 50 000 términos y se actualiza periódicamente para incluir los nuevos avances tecnológicos. Como se indica en la presentación del proyecto, está dirigido a profesionales de la lengua y la ingeniería.

Con el fin de enriquecer la información y encuadrar cada entrada en el subdominio temático adecuado, se concibe la Ingeniería como eje principal del que se desprenden nueve grandes ramas: (i) Astronáutica, naval y transportes, (ii) Agroforestal, (iii) Construcción, (iv) Tecnologías de la información y las comunicaciones, (v) Seguridad y defensa, (vi) Química industrial, (vii) Energía, (viii) Ingeniería biomédica e (ix) Ingeniería general. Estas ramas se dividen, a su vez, en treinta y dos subdominios y en cada entrada se especifica la pertenencia a uno de ellos.

Para acceder a la información, se proporciona un cuadro de búsqueda por lema, entendido en el sentido de entrada y no de forma básica de una palabra o término. Sin embargo, al introducir un término en singular obtenemos las entradas del lema tanto en singular como en plural, mientras que si lo introducimos en plural solo obtenemos las entradas de este número, algo que no nos parece muy consistente. La opción que se muestra por defecto es la búsqueda por aproximación, que arroja resultados que contengan el término o fragmento introducido, lo que encontramos muy útil. No obstante, en la guía de uso se indica que la búsqueda por aproximación devuelve resultados aunque la palabra se haya escrito de manera incorrecta, si bien al probar el recurso no encontramos esta funcionalidad. La otra opción que se proporciona es la búsqueda exacta del término introducido. Además, se puede especificar el campo de conocimiento que nos interese. Se ofrece también una búsqueda avanzada por índice alfabético, que a nuestro juicio es más simple que las anteriores, en la que se indica el número de entradas existentes de cada letra.

Los CN tienen entradas propias, en las que se proporciona la misma información que en las entradas de términos simples, principalmente la marca o subdominio al que pertenece el concepto, el equivalente en inglés y la definición. Cabe destacar que no se atisba ningún tipo de organización semántica en el recurso. Por ejemplo, al realizar una búsqueda por aproximación de *potencia*, obtenemos la lista de términos en los que se incluye el elemento de búsqueda. Estos aparecen ordenados alfabéticamente, por lo tanto, es coincidencia que los CN hipónimos (p. ej. *potencia activa*) aparezcan cerca de su hiperónimo (*potencia*). Esto no ocurriría por ejemplo en inglés, ya que el núcleo se suele encontrar al final del término (p. ej. habría bastante separación entre *active power* y *power*). Por el contrario, si queremos obtener términos que guarden otra relación con *potencia* que no sea la de hiponimia (p. ej. *transformador de potencia*), tendremos que avanzar varias páginas hasta encontrarlos.

A menudo observamos que aparecen varias entradas para el mismo término, lo cual parece indicar que se trata de un término polisémico y que cada entrada corresponde a un significado. Sin embargo, no se ofrece ningún indicador que desambigüe el significado o ni siquiera que indique que existe alguna diferencia entre las entradas, simplemente se muestra el término repetido en varias líneas. Por lo tanto, es necesario acceder a cada entrada para averiguar cuál es la que nos interesa. Una vez dentro, observamos que en múltiples ocasiones cada entrada no implica un cambio de significado, sino un cambio de dominio en el que se puede encontrar el concepto. Otras veces se crean varias entradas para varias acepciones, como muestran las Figuras 4 y 5, en las que se muestran las entradas de *abolladura*.

Figura 4: Entrada 1 de *abolladura* en el *Diccionario Español de Ingeniería*

Figura 5: Entrada 2 de *abolladura* en el *Diccionario Español de Ingeniería*

Sin embargo, esta no es la regla general con la que se trata la polisemia en este diccionario, ya que encontramos otros casos en los que se incluyen varias acepciones en una misma entrada, como ocurre en el caso de *falla* (Figura 6).

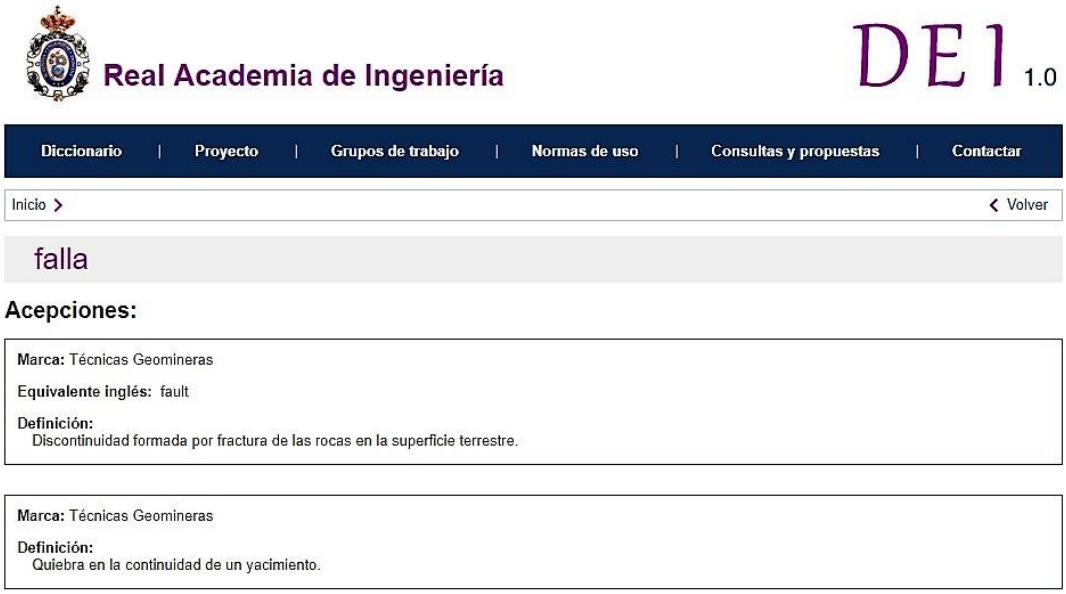

Figura 6: Entrada de *falla* en el *Diccionario Español de Ingeniería*

En la guía de uso se indica que los diferentes significados se mostrarán en la misma entrada, en recuadros diferentes, como en el caso de *falla*. Por lo tanto, creemos que el fin de realizar diferentes entradas radica en las distintas marcas temáticas o subdominios. No obstante, consideramos que en muchos casos entrañan cambios de significado (p. ej. en las Figuras 4 y 5 de *abolladura*) y, por lo tanto, se debería adoptar una postura uniforme en el tratamiento de la polisemia.

En cuanto a las traducciones, que se incluyen en la mayoría de las entradas, son un elemento muy apreciado, ya que no están presentes en muchos recursos. De hecho, la Real Academia de Ingeniería resalta la necesidad de un diccionario de términos de ingeniería en el que se pueda buscar del español al inglés y viceversa. Consideramos que esta herramienta palia algunas de estas deficiencias, pero encontramos el inconveniente de que solo se pueden hacer búsquedas desde el español y, por otra parte, que los equivalentes no se incluyen en algunas entradas (como vemos en la primera entrada de *abolladura* en la Figura 4 o en la segunda acepción de *falla* en la Figura 6). Además, observamos que la calidad de los equivalentes propuestos es a veces cuestionable. Por ejemplo, en dos de las entradas de *aerogenerador* se ofrece la misma definición: «sistema que convierte la energía cinética del viento en energía eléctrica». La marca o subdominio es prácticamente la misma: «Tecnología energética. Eléctrica» en una de las entradas y «Tecnología

energética. Renovables» en la otra. Sin embargo, son diferentes los equivalentes propuestos, cuando el concepto se presenta como el mismo. En una de las entradas se ofrece *wind turbine generator* como equivalente de *aerogenerador*, mientras que en la otra se indica que su equivalente es *windmill*. A nuestro juicio, ambos equivalentes serían válidos en ambas entradas, ya que se trata de variantes denominativas del mismo concepto.

A veces se proporciona información adicional, como observaciones, fórmulas, abreviaciones o sinónimos, lo que nos parece de gran utilidad dada la abundancia de estos elementos en la terminología de la Ingeniería. En cuanto a los sinónimos, cuando el término ha sido introducido como tal en la base de datos, este aparece sin definición y se incluye en su lugar un enlace al término principal (p. ej. en la entrada de *aeroturbina* no se proporciona información semántica, sino que se nos remite a la entrada de *turbina eólica* mediante una referencia cruzada). Ello puede deberse al deseo de distinguir un término como principal. Sin embargo, consideramos que esta práctica ralentiza las consultas y dificulta la transmisión de conocimiento, ya que serán necesarios más pasos hasta llegar a la definición. Además, dado el soporte electrónico del diccionario, nos parece innecesario este enfoque, ya que no debería haber limitaciones de espacio. La indicación de los diferentes status de los términos podría realizar mediante notas de uso. No obstante, a pesar de los inconvenientes que hemos comentado, el *Diccionario Español de Ingeniería* nos parece una herramienta muy útil, con información muy detallada que contribuye a rellenar los vacíos existentes en la terminografía de la Ingeniería.

3.6.2.2 Diccionario de términos médicos

El *Diccionario de términos médicos* es un diccionario monolingüe que incluye términos médicos en español. Fue publicado por la Real Academia Nacional de Medicina en 2012 y cuenta con unas 52 000 entradas. Está dirigido tanto a profesionales biosanitarios como a lingüistas, especialistas de otras ciencias y el público en general. Dispone de una versión impresa y otra online (dtme.ranm.es), que será la que analicemos en este estudio. Esta versión online se actualiza periódicamente para dar cuenta de la naturaleza cambiante del lenguaje médico. Se ofrece un acceso gratuito restringido a la letra A, si bien hemos podido acceder a todo el diccionario gracias a la colaboración de miembros de la institución. Nos parece un diccionario especialmente interesante por su planteamiento terminográfico y por el carácter especial que otorga a los CN en las búsquedas, como veremos a continuación.

El acceso a la información se realiza mediante el menú de búsqueda, que permite seleccionar distintos criterios, agrupados en una «Búsqueda simple» y una «Búsqueda avanzada». En la búsqueda simple podemos elegir los siguientes criterios:

-Aproximación: se trata de una búsqueda basada en la semejanza ortográfica, que muestra resultados aun cuando la palabra está mal escrita.

-Lema que contenga la palabra: extrae términos que contengan la secuencia introducida (principalmente el propio término y sus CN hipónimos). Por ejemplo, si introducimos el término *arteria*, obtenemos este término y sus CN hipónimos, como *arteria alveolar inferior* o *arteria braquial*.

-Secuencia de caracteres al inicio, en el medio o al final: por ejemplo, términos que comiencen por *hiper*, como *hiperabducción, hiperacantosis* o *hiperacetonuria*.

-Todos los campos: permite buscar un término en todo el diccionario, en cualquiera de los campos de las entradas: lema, definición, observaciones, sinónimos, nomenclatura, abreviaciones, etimología, etc. Nos parece una opción muy avanzada y poco frecuente que nos puede ayudar a conseguir resultados más específicos.

-Equivalente en inglés: permite introducir el término en inglés y obtener el equivalente exacto en español (no se muestran aquí los términos más extensos que contienen nuestra búsqueda, ya que esta información pertenece a las opciones avanzadas). A pesar de no presentarse como un diccionario bilingüe, ya que no incluye todas las equivalencias, sino que se centra en la precisión conceptual y denominativa en español, permite hacer búsquedas inversas. Por ello, nos parece una guía muy útil para, en su caso, comenzar una búsqueda bilingüe más detallada.

Por su parte, la búsqueda avanzada ofrece algunos de los criterios presentes en la búsqueda simple, además de otros más específicos y la posibilidad de combinar varias consultas con los operadores booleanos Y, O y NO, una opción que consideramos poco frecuente pero muy adecuada, ya que nos permite obtener resultados más concretos. Así pues, los criterios de la búsqueda avanzada son los siguientes: lema que contenga la palabra; secuencia de caracteres al inicio, en el medio o al final; equivalente en inglés; sigla o abreviatura; tipo de lema; categoría gramatical; lengua en etimología; *Terminologia Anatomica*; en definición; y en observaciones.

Las diferencias con la búsqueda simple son las siguientes:

-Equivalente en inglés: permite introducir un término en inglés que constituya el equivalente directo o parte de él (por ejemplo, que forme parte de un término compuesto). Esta es la opción más interesante en lo relativo a traducciones en inglés. No obstante, nos parece adecuado que, para un usuario lego, se muestre la traducción directa en la búsqueda simple y, para un usuario más avanzado, se ofrezcan en la búsqueda avanzada más términos compuestos que incluyan ese término en inglés.

-Sigla o abreviatura: muestra las siglas o abreviaturas que tienen entrada propia o que aparecen en los campos de abreviatura o nomenclatura de las entradas. Dada la abundancia de siglas en los textos especializados, como variantes de los términos compuestos, nos parece una opción muy relevante.

-Tipo de lema: tras elegir entre lemas sencillos (términos simples) o complejos (términos compuestos), obtendremos una lista por orden alfabético de los veinte primeros resultados del diccionario. Pese al inconveniente de restringir los resultados a los veinte primeros, si añadimos un marcador booleano y especificamos otro criterio de búsqueda conseguiremos enriquecer nuestra búsqueda. Por ejemplo, podemos indicar que queremos términos complejos que empiecen por *dermatitis*, y obtendremos así los CN cuyo núcleo es *dermatitis*. Nos parece fantástica esta opción, que va en la línea del trabajo terminográfico que proponemos en esta tesis, y que se ha diseñado dando cuenta de la gran abundancia e importancia de los términos compuestos en las áreas especializadas, como es la Medicina.

-Categoría gramatical: podemos indicar la categoría gramatical y, en su caso, el género que nos interese y obtendremos una lista por orden alfabético de los veinte primeros resultados del diccionario. Así, podremos obtener, por ejemplo, los verbos que aparecen descritos en el *Diccionario de términos médicos*, como *acatarrarse* o *adelgazar*.

-Lengua en etimología: permite hacer búsquedas según la lengua de origen que figura en el campo de etimología (alemán, árabe, catalán, español, francés, inglés, etc.) y obtener una lista por orden alfabético de los veinte primeros resultados del diccionario. Por ejemplo, podremos conocer términos relacionados con la medicina que proceden del árabe, como *alcohol* o *amígdala*.

-*Terminologia Anatomica*: podemos obtener los términos anatómicos del diccionario escribiendo la denominación en latín que reciben según la Terminología Anatómica Internacional. Por ejemplo, si introducimos la denominación *Nasus* de la

Terminología Anatómica Internacional, se nos mostrará la entrada correspondiente a *nariz*.

-En definición: permite la búsqueda en los campos de definición. Junto con la opción de búsqueda en observaciones, son una variante más específica de la búsqueda simple en todos los campos.

-En observaciones: permite la búsqueda en los campos de observaciones. Estas dos últimas opciones son las más parecidas a la búsqueda semántica y se pueden mejorar usando los marcadores booleanos. Por ejemplo, podemos buscar en las definiciones los términos que contengan *enfermedad* y *piel* y obtener resultados como *albinismo* o *carbunco*.

A pesar de las desventajas del orden alfabético que también adopta este diccionario, nos parece que estas opciones de búsqueda son muy acertadas y pueden dar respuesta a distintas necesidades de los usuarios. Sin embargo, pensamos que, si se indicaran de algún modo los términos relacionados, por ejemplo, mediante hipervínculos con otra tipografía en las entradas, este recurso contaría con un valor añadido. Así, además, se podrían agilizar las búsquedas.

En cuanto a los términos compuestos, estos disponen de entradas propias, a las que podemos acceder con las diferentes opciones de búsqueda que acabamos de comentar. En cada entrada, ya sea de un término simple o compuesto, se muestra el término, la etimología, números que indican las acepciones, la categoría gramatical, el equivalente en inglés de cada acepción, la definición, las observaciones de la acepción o del lema, los sinónimos (con especificación del uso: coloquial, desuso, etc.) y abreviaturas. Por lo tanto, se trata de entradas muy completas. En el caso de la sinonimia, la definición del concepto se muestra solo en la entrada del término marcado como preferente, al que remiten los sinónimos por medio de referencias cruzadas. Esta técnica puede deberse al deseo de señalar un término como principal. No obstante, como en el *Diccionario Español de Ingeniería*, pensamos que esta práctica ralentiza las consultas. En su lugar, preferimos mostrar la definición en la entrada de cada término, ya que los diferentes status de los sinónimos pueden indicarse mediante notas de uso, que de hecho son una de las ventajas de este diccionario.

En la Figura 7 se muestra un ejemplo de búsqueda avanzada en el *Diccionario de términos médicos*, en la que seleccionamos los siguientes criterios: tipo de lema > término complejo Y en definición > enfermedad Y en definición > pelo. Nuestro objetivo era

obtener términos compuestos relacionados con enfermedades y el pelo. Esta búsqueda arrojó siete resultados: *albinismo oculocutáneo, albinismo parcial, carbunco pulmonar, dermatitis seborreica del cuero cabelludo, mucinosis fulicular, quiste tricolémico* y *síndrome de Waardenburg*.

Figura 7: Búsqueda avanzada por tipo de lema > término complejo Y en definición > enfermedad Y en definición > pelo en el *Diccionario de términos médicos*

En la Figura 8 vemos la entrada de uno de ellos, *albinismo oculocutáneo*, en la que se muestra el lema en español, su equivalente principal en inglés, la definición, un sinónimo, su abreviatura y observaciones. Como se puede observar, son frecuentes en este diccionario las definiciones extensas de tipo enciclopédico, que se justifican por el carácter eminentemente conceptual y explicativo del recurso.

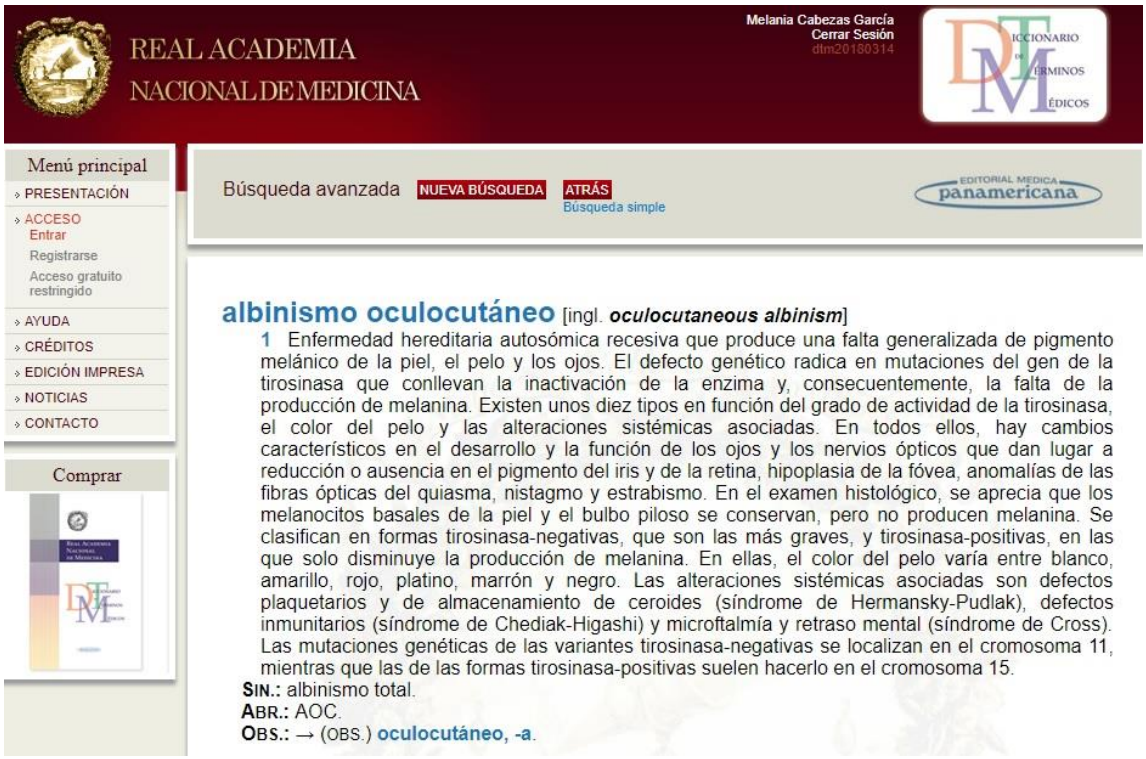

Figura 8: Entrada de *albinismo oculocutáneo* en el *Diccionario de términos médicos*

En general, el *Diccionario de términos médicos* nos parece un recurso del siglo XXI, una excepción a la norma general que destaca por su rica información y sus fantásticas opciones de búsqueda. A estas se les puede sacar partido especialmente gracias a los operadores booleanos, con los que se satisfacen las distintas necesidades que puedan tener los usuarios. Se trata, pues, de un diccionario de calidad, medido al detalle y de prestaciones excelentes.

3.6.2.3 Oxford Dictionary of Biochemistry and Molecular Biology

El *Oxford Dictionary of Biochemistry and Molecular Biology* es un diccionario monolingüe que recopila términos sobre bioquímica y biología molecular en inglés. Analizamos su segunda edición en formato impreso (Cammack et al. 2006), si bien cuenta con otra versión online, para la que se necesita una suscripción de pago. Cuenta con más de 21 000 entradas, complementadas por dos apéndices, que presentan (i) el alfabeto griego y los caracteres griegos que se usan como símbolos, y (ii) las reglas de prioridad de algunos ligandos en entidades moleculares. Está dirigido tanto a especialistas en el campo, como a profesores, estudiantes o cualquier otra persona interesada en el significado de términos de bioquímica y biología molecular.

Como es habitual, se accede a la información por orden alfabético. De este modo, los CN hipónimos no aparecen normalmente junto a su hiperónimo, ya que en inglés el núcleo suele estar a la derecha (p. ej. *adipose tissue* y su hiperónimo *tissue* están separados

por múltiples páginas). Ocurre lo mismo con los CN que contienen el término como modificador: podrán aparecer o no cerca de este término en la lista. Como ya hemos mencionado, esta práctica no facilita la adquisición de conocimiento. No obstante, esta desventaja se palia en cierto modo mediante lo que denominan «hidden entries». Estas «entradas ocultas» consisten en la inclusión de términos relacionados en la definición, que se destacan en negrita y cuyo significado puede obtenerse mediante otra breve definición o gracias al contexto en el que se insertan. Estos términos relacionados son muchas veces CN y, además de incluirse como entradas ocultas, cuentan en muchos casos con entradas propias. Nos parece una buena opción para marcar las relaciones entre términos, en el sentido de que permite relacionar los términos en contexto e identificar de qué manera están vinculados. No obstante, consideramos que cada entrada oculta debe aparecer en el orden alfabético (ya sea como lema o sublema) para facilitar su identificación, aunque no se proporcione ningún tratamiento microestructural y se remita a otra entrada más detallada.

En cuanto a los CN, estos pueden tener entradas propias o incluirse como las entradas ocultas que acabamos de comentar. En cada entrada, ya sea de un término simple o compuesto, se muestra el término en negrita y distintas acepciones cuando el término es polisémico. Destacamos la existencia habitual de definiciones muy extensas, de tipo enciclopédico, pese a que en la introducción del diccionario se menciona la brevedad de las entradas. Por otra parte, a veces se especifica el dominio al que pertenece el concepto. Además, se incluyen muchas abreviaturas y símbolos, lo cual resulta especialmente importante en el caso de los CN, ya que muchos de ellos se suelen utilizar de forma abreviada.

Otro aspecto importante y poco frecuente en los recursos lingüísticos es la inclusión de términos derivados, es decir, unidades procedentes del lema que pertenecen a otra categoría gramatical y no cuentan con entrada propia porque no requieren definición (p. ej. en la entrada de *acinus* se proporciona el adjetivo *acinar*). Esto nos parece muy adecuado, ya que la conversión a otras categorías gramaticales es muy frecuente en los textos especializados y, además, es un mecanismo recurrente de formación de CN, por lo que la identificación de categorías gramaticales derivadas puede ayudar a la comprensión de CN. Por ejemplo, el adjetivo *acinar* interviene en el CN *acinar tissue*, que no cuenta con entrada en el diccionario. Sin embargo, gracias a la inclusión de *acinar* como término relacionado con *acinus*, el usuario puede acercarse al significado del CN.

Por último, queremos destacar el tratamiento de las referencias cruzadas en este recurso. Se utilizan las referencias cruzadas para indicar (i) variantes, (ii) entradas relacionadas que amplían información, (iii) términos opuestos, o (iv) términos que se suelen confundir.

En cuanto a las variantes, tan presentes en los CN, en la guía del diccionario se indica que se eligen como términos principales los que gocen de un uso mayoritario. Tras cada término principal se muestran sus variantes, bien introducidas por *or* o precedidas de una nota de uso en cursiva (p. ej. *sometimes, formerly, Brit., an alternative name for, a former name of, a variant spelling of, a less common name for, an obsolete name for*). Pese a que algunas de estas notas de uso nos parecen adecuadas, porque aportan información de la variante (p. ej. *Brit.* o *an obsolete name for*), consideramos que otras de estas etiquetas se podrían aunar para dar más homogeneidad a la nomenclatura. Este es el caso de *or*, *an alternative name for*, *other name* o *a variant of* (entre otras), ya que no transmiten información adicional y, además, se usan de manera inconsistente en todo el diccionario.

También es preciso señalar que, en muchos casos, la identificación de variantes solo puede realizarse desde una de ellas. Por ejemplo, en la entrada de *mammotropic* se dice que es un término alternativo a *lactotropic*. Sin embargo, al consultar la entrada de *lactotropic*, no se hace alusión a la variante *mammotropic*, con lo cual no habríamos obtenido esta información de no haber accedido en primer lugar a la entrada de *mammotropic*. Ocurre así en muchos otros casos. También se alega que todas las variantes aparecen con entrada propia en el orden alfabético, a no ser que estas se encuentren cerca y se puedan identificar fácilmente. En estas entradas no se proporciona ningún tratamiento microestructural, sino que se remite a la entrada del término principal, debido a las evidentes limitaciones de espacio del papel. No obstante, observamos que este no siempre es el caso. Por ejemplo, vemos que *peroxiredoxin* tiene como variantes *thioredoxin peroxidase* y *alkyl peroxide*, pero ninguno de estos dos términos figura como entrada en el orden alfabético, de manera que el usuario no los encontrará a menos que acceda desde el término principal o haga una búsqueda directa si dispone de la versión digital del documento. No obstante, al margen de estas inconsistencias, queremos destacar la utilidad de la inclusión de variantes en los recursos lingüísticos.

Por otra parte, las referencias cruzadas que se usan para ampliar información (destacando el hiperónimo o introducidas por *see also*), avisar de términos opuestos (*compare*) o aquellos que se suelen confundir (*distinguish from*), nos parecen muy

apropiadas para la adquisición de conocimiento y la correcta comprensión de los términos, pues de nuevo aluden a términos relacionados.

En la Figura 9 mostramos la entrada de *antagonist*, en la que se pueden observar algunas de las características mencionadas, como las entradas ocultas (señaladas en negrita y muchas veces relativas a CN), las referencias cruzadas que remiten a entradas que amplían información (*see also*) y a términos opuestos (*compare*), además del adjetivo (*adj.*) y adverbio (*adv.*) derivados de este término.

antagonist **1** anything that antagonizes. **2** any agent, especially a drug or hormone, that reduces the action of another agent, the **agonist**. Many act at the same receptor as the agonist. Such antagonists may be either surmountable or insurmountable, depending on prevailing conditions. Antagonism may also result from combination with the substance being antagonized (**chemical antagonism**), or the production of an opposite effect through a different receptor (**functional antagonism** or **physiological antagonism**), or as a consequence of competition for the binding site of an intermediate that links receptor activation to the effect observed (indirect antagonism). The term 'functional antagonism' is also used to describe a less well-defined category in which the antagonist interferes with other events that follow receptor activation. **3** a muscle that works in opposition to another (the agonist). *See also* **competitive antagonism**, **noncompetitive antagonism**. *Compare* **agonist**, **inverse agonist**, **synergist**. —**antagonistic** *adj.*; **antagonistically** *adv.*

Figura 9: Entrada de *antagonist* en el *Oxford Dictionary of Biochemistry and Molecular Biology*

Así pues, el *Oxford Dictionary of Biochemistry and Molecular Biology* nos parece un recurso más enriquecido que los tradicionales diccionarios impresos, ya que no se trata de un diccionario que se limite a proporcionar definiciones, sino que, dentro del clásico orden alfabético, pretende trazar vínculos conceptuales.

3.6.2.4 Diccionario técnico inglés-español español-inglés

El *Diccionario técnico inglés-español español-inglés* (Beigbeder Atienza 2006) es un diccionario bilingüe que incluye terminología técnica en inglés y español. Analizamos su segunda edición, que dispone de una versión impresa con más de 100 000 entradas y otra online (http://www.editdiazdesantos.com/Beigbeder), actualizada en 2016, que incorpora actualizaciones y a la que se puede acceder de manera gratuita durante siete días. Dirigido a usuarios que necesiten acceder a la terminología técnica bilingüe en inglés y español, este diccionario constituye un ejemplo de práctica lexicográfica muy frecuente, que comentaremos con detalle en esta sección.

La versión impresa consta de una parte inglés-español y otra español-inglés, con entradas ordenadas alfabéticamente. Dispone, además, de un glosario de siglas y

abreviaturas en ambas lenguas. Por su parte, la versión online ofrece un cuadro de búsqueda desde cualquiera de las dos lenguas, si bien es necesario indicar la direccionalidad, ya que no es posible hacer consultas en la dirección contraria a la indicada.

Por lo que respecta a la ubicación de los CN, esta difiere en las versiones impresa y online. Por una parte, el diccionario impreso muestra los CN como sublemas del primer formante. Por ejemplo, *wind power* aparece como sublema de *wind* y no figura en la entrada de *power*. Además, como ocurre en muchos otros recursos, se omite la parte del CN que coincide con la entrada. Como vemos a continuación en la Figura 10, bajo *multiple* se presentan sangrados los CN que comienzan por este término, pero este se omite en todos ellos.

En primer lugar, no compartimos la omisión de ese primer formante del CN. Creemos que esto complica la consulta del diccionario, pues en muchos casos se generan largas listas de CN incluidos en una entrada, en los que no resulta sencillo saber qué elemento falta. Además, muchas veces los CN aparecen en columnas o incluso páginas diferentes a la de la primera palabra y el usuario obtiene solo un término y su traducción, sin saber muy bien si falta otro componente y cuál sería la forma exacta del término compuesto. Por lo tanto, apostamos por incluir la forma completa del CN o, en su defecto, sustituir el elemento omitido por algún símbolo distintivo (p. ej. ~), de manera que se facilite su identificación.

En segundo lugar, preferimos este enfoque de sublemas frente a la opción de incluir los CN como lemas ordenados alfabéticamente, sin ningún tipo de relación expresa con otros términos. Mediante el enfoque adoptado en este diccionario, se muestra al menos el vínculo con otros términos (en concreto, con uno de los términos que lo componen y con otros CN que contienen el mismo lema y que pueden estar relacionados de algún modo), aunque no se especifique cómo están vinculados. No obstante, nos parece que sería útil para el usuario poder encontrar también los CN bajo la entrada de sus otros formantes, al menos bajo la del núcleo e hiperónimo en forma de referencia cruzada. De este modo, se agruparían bajo el núcleo CN que guardan relaciones semánticas más vinculantes, ya que el núcleo es el centro de significado del CN. Así, se trataría probablemente de miembros de la misma categoría semántica que heredarían las propiedades de su hiperónimo.

En la Figura 10 se muestra la entrada de *multiple*, en la que podemos observar su extensión, señalada con una línea negra gruesa, así como la omisión de *multiple* en todos

los CN. A pesar de ello, destacamos la gran cobertura en lo que a términos compuestos se refiere.

alternator | alternador polifásico.
clocked logic | lógico de impulsos patrón multifase.
commutatorless motor | motor polifásico sin conmutador.
generator | generador polifásico.
rectifier circuit | circuito rectificador polifásico.
multiphone | multífono.
system | red telefónica con estación central única.
multiphonon | multifonon.
multiphoton photoelectric emission | emisión fotoeléctrica multifotón.
multipin | multiclavija.
connector | conectador multiclavijas (telecomunicación) | conector con varios terminales.
plug | enchufe multiclavijas.
multiplate | multiplaca | pluridisco.
camera | cámara multiplacas (fotografía).
clutch | embrague pluridisco.
friction clutch | embrague de fricción polidisco.
multiplay | reproducción múltiple.
multiple | múltiplo | cuadro múltiple manual (telefonía) | múltiple | en paralelo (electricidad).
access | acceso múltiple.
adapter | adaptador multiclavija.
antenna | antena múltiple.
antenna surveillance radar | radar de múltiples antenas de vigilancia.
axis chart | gráfica de ejes múltiples.
backscattering | retrodispersión múltiple.
band receiver | receptor omnionda | receptor de banda múltiple.
battery | batería de pilas.
bay antenna | antena de secciones múltiples.
beam interference | interferencia de haz múltiple.
box loom | telar multilanzaderas, telar de cajones.
casting | colada múltiple.
channel | canal múltiple | pluricanal.
circuit | circuito múltiple | circuito compuesto | acoplamiento en cantidad, acoplamiento en derivación, acoplamiento en paralelo.
circuit winding | devanado imbricado.
coil transformer | transformador de bobina múltiple.
conductor | conductor en haz (electricidad) | conductor múltiple.
connection | montaje múltiple, conexión múltiple | acoplamiento en cantidad, acoplamiento en derivación.
coverage | cobertura múltiple (radar).
cropping | cultivo múltiple.
crosstalk | diafonía múltiple (telecomunicación).
crucible furnace | horno multicrisoles.
currency system | sistema de cambios múltiples.
current generator | generador polimórfico.
disc clutch | embrague pluridisco.
drawing | estirado múltiple (alambres) | embutición en varias pasadas.
drop circuit | circuito multipunto.
duct conduit | canalización plurirramal | conducto multitubular (cables telefónicos - EE UU).
electrode welding | soldeo con electrodos múltiples.
fiber | haz de fibras.
field | campo múltiple (telecomunicaciones) | campo multipolo (electromagnetismo).
image | imagen múltiple (TV-fotografía).
image film | película de imágen múltiple (placa impresa).
jack | conjuntor general (telefonía).
launch rocket system | sistema múltiple de lanzamiento de cohetes.
measuring instrument | instrumento para varias medidas simultáneas.
milling | fresado en serie | fresado con tren de fresas.
modulation | modulación múltiple.
module access | acceso de módulo múltiple (computadora).
molding | moldeo múltiple.
neutron capture | captura neutrónica múltiple (nuclear).
of gearing | tren de engranajes.
on-line programming | programación múltiple en línea.
plough | arado de varias rejas.
saw | sierra alternativa de varias hojas.
scanning | exploración entrelazada, escansión entrelazada, exploración múltiple (TV).
scattering | dispersión múltiple | difracción múltiple.
series connection | conexión en series paralelas, acoplamiento en cantidad y voltaje.
shuttle loom | telar multilanzaderas.
signal unit | unidad múltiple de señalización.
sound track | pista múltiple (acústica) | pista sonora múltiple.
switch | interruptor múltiple.
switching | conmutación múltiple.
system | sistema multifilar (electricidad).
target radar | radar selector de objetivos.
television channel | canal múltiple de televisión.
track radar | radar de rastreo múltiple.
transformer | transformador en paralelo.
transmitter | transmisor múltiple.
trunk groups | formación de grupos de enlace (líneas urbanas).
tube launcher | lanzacohetes multitubo.
tuner | sintonizador múltiple.
unit capacitor | capacitor de tomas múltiples.
valve | lámpara compleja (electrónica).
way switch | conmutador multidireccional.
way system | sistema de transmisiones simultáneas.
winding | devanado múltiple.
wire antenna | antena multifilar de hilos paralelos.
wire dipole antenna | antena dipolo multifilar.
working | emisión multicanal | comunicación múltiple.
-connected | conectado en derivación, conectado en paralelo (electricidad).
-contact switch | conmutador selector.
-loop servomechanism | servomecanismo pluricircuital.
-loop system | sistema de circuitos múltiples.
-output logical network | red lógica con salidas múltiples.
-output switching circuit | circuito de conmutación de salida múltiple.
-pole | multipolar (electricidad).
-purpose tester | polímetro.
-relay system | sistema con varias estaciones repetidoras (telecomunicación).
-series battery | batería de pilas en series paralelas.
-series connection | conexiones en serie conectadas en múltiple.
-spot welding | soldeo por puntos múltiples.
-stage missile | misil de fases múltiples.
-stage rocket | cohete de etapas múltiples.
-track radar | radar de múltiples canales.
-track range | radiofaro con múltiples rumbos.
-track recording | registro de pista múltiple.
-tuned antenna | antena con sintonía múltiple.
-unit steerable antenna | antena rómbica múltiple.
-wheel bogie | bogie de varios ejes.
multiplex | telegrafía múltiple | múltiplex | transmisión múltiple | transmisión simultánea.
communication | comunicación múltiple | enlace múltiple de radio.
lap | devanado paralelo múltiple (electricidad).

Figura 10: Entrada de *multiple* en la versión impresa del *Diccionario técnico inglés-español español-inglés*

La principal diferencia entre las versiones impresa y online es la ubicación de los términos compuestos. Frente a la representación como sublemas en la versión impresa que acabamos de comentar, los CN tienen entradas propias en la versión online y no se indica ninguna relación con la entrada de alguno de sus formantes. Consideramos que, en los recursos electrónicos, los términos compuestos deben tener entradas propias, ya que no existen limitaciones de espacio. Sin embargo, como venimos indicando, creemos que sería útil indicar alguna relación entre el CN y otros términos, al menos el hiperónimo o la palabra inicial. En situaciones de transferencia interlingüística, como las que contempla este recurso, esto resulta fundamental, ya que de este modo se podrían identificar términos relacionados que podrán intervenir en el texto origen o meta.

En lo que concierne a la omisión de parte del CN, este problema se palia en la versión online, en la que se ofrece la forma completa del CN. Además, tanto en la versión impresa como en la versión online solo podemos acceder a los CN desde su primer elemento, que normalmente será el núcleo del CN en español y el último de los modificadores en inglés. Por ejemplo, para encontrar *generador síncrono*, tendremos que ir a la entrada de *generador* en el diccionario en papel o bien teclear *generador* en el diccionario online, ya que desde *síncrono* no se muestra ninguna información de los CN que incluyen este término como modificador.

En cuanto a la información que se muestra en las entradas, ambas versiones ofrecen el término en la lengua de partida y su(s) equivalente(s) en la lengua meta. Se trata, pues, de un diccionario que se centra en las equivalencias interlingüísticas, prescindiendo prácticamente de toda información semántica. El único atisbo de significado parece encontrarse en los indicadores de uso de los equivalentes. Sin embargo, en los contados casos en los que estos se incluyen, el usuario no dispone de una leyenda que indique a qué se corresponden esas notas de uso abreviadas. La Figura 11 muestra la entrada de *manifold* en la versión online del diccionario. Como podemos observar, en pocos casos se indica el dominio al que corresponde cada equivalente, solo vemos *MAR (en buques), TRA FER (de locomotora), IND PET (petroleros)* y *MAT*, que no resultan muy aclaratorios.

En este ejemplo percibimos también que se mezclan los equivalentes de *manifold* como sustantivo y como adjetivo, sin ninguna distinción de la categoría gramatical. Asimismo, observamos que, en el caso de términos polisémicos como este, varios equivalentes que expresan significados similares se presentan en distintas líneas (p. ej. *múltiple* y *variado*), incluso separados por otros equivalentes de distinto significado (p. ej.

plano de las válvulas de cubierta). Consideramos que convendría incluir juntos los términos que expresan el mismo significado, para indicar así que se pueden intercambiar. Este problema no se da en la edición en papel, en la que se agrupan los equivalentes que transmiten el mismo significado (separados por comas, mientras que los equivalentes que expresan significados diferentes se separan con el símbolo |, aunque esta técnica no se usa de manera consistente en todo el diccionario).

Figura 11: Entrada de *manifold* en la versión online del *Diccionario técnico inglés-español español-inglés*

En cuanto a las traducciones propuestas, llama la atención la ausencia de equivalentes importantes en algunas entradas. Por ejemplo, en la entrada de *aerogenerador* no se muestra *wind turbine*, que es el equivalente más aceptado y ampliamente extendido. Además de traducciones, la versión online ofrece la posibilidad de marcar un término y buscarlo en el *Diccionario de la lengua española* de la Real Academia en el caso del español o en Encyclopedia.com en el caso del inglés, además de en Wikipedia, Google, Bing o Youtube. Asimismo, se muestra el número de términos que el usuario ha consultado en el diccionario y se pueden marcar términos favoritos, lo cual puede ser útil, por ejemplo, si hay un término que se busque con mucha frecuencia.

En nuestra opinión, el punto fuerte de este diccionario es la gran cantidad de términos tanto simples como compuestos (y de todas las extensiones posibles) que incluye. Sin embargo, hay diferentes aspectos que podrían modificarse para ofrecer mejores prestaciones al usuario.

3.6.2.5 Elsevier's Dictionary of Medicine: Spanish-English English-Spanish

El *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish* (Hidalgo Simón 2004) es un diccionario en papel que recopila términos médicos en inglés y español. Cuenta con más de 28 000 entradas y está dirigido tanto a profesionales como al público general interesado en la terminología médica bilingüe en estos dos idiomas.

Dispone de dos partes: en la primera se muestran los términos del español al inglés y, en la segunda, del inglés al español. Las entradas están ordenadas alfabéticamente y corresponden a términos simples, ya que los términos compuestos se incluyen como sublemas tanto del núcleo como del modificador. Consideramos que este es el enfoque más apropiado para la inclusión de términos compuestos en recursos impresos, ya que facilita la localización de la información, agrupa términos que guardan alguna relación (sin recurrir únicamente a la linearidad del orden alfabético) y proporciona información semántica que permite trazar vínculos entre los términos. En ambas entradas se muestra el equivalente en la otra lengua, si bien la definición se suele proporcionar solo en una de las dos entradas (la del núcleo o la del modificador, indistintamente), a la que se remite mediante el indicador *see*.

Los términos se definen en inglés cuando partimos desde el español y viceversa. No se definen cuando son muy transparentes (p. ej. *frontal*) o cuando ya se ha definido un término relacionado que se encuentra cerca (p. ej. se define *cardiopatía*, pero no *cardiópata* ni *cardiopático*). No obstante, el hecho de incluir información semántica, además de equivalentes, hace de este diccionario una excepción a la norma de la tradición terminográfica bilingüe. Además, proporciona sinónimos, siglas y algunas notas de uso (p. ej. *GB, US*), que también son muy relevantes en los recursos especializados bilingües.

En cuanto a las notas de uso de los equivalentes, pensamos que deberían extenderse a un mayor número de términos. Por ejemplo, en entradas como la de *crisis* (es), para la que se proporcionan los equivalentes *crisis, disaster, catastrophe* y *emergency*, no se indican de ningún modo los diferentes contextos en los que se usan estos términos, lo que puede llevar a traducciones erróneas.

En la Figura 12 se muestra como ejemplo la entrada de *polimerasa*. En ella podemos ver su traducción al inglés y la definición en inglés entre paréntesis. Asimismo, se incluyen como sublemas dos términos compuestos que contienen el término *polimerasa*, en concreto, *polimerasa de ADN* (que lo incluye en el núcleo) y *reacción en cadena de la polimerasa* (que lo incluye en el modificador). Junto a estos CN también se

ofrece su traducción y definición, además de las siglas con las que se suelen denominar (*RCP*, en el caso de *reacción en cadena de la polimerasa*).

polimerasa polymerase (enzyme that catalyzes a polymerization, or formation of a polymer by linking other simpler compounds, or monomers)
polimerasa de ADN DNA polymerase (enzyme complex that catalyzes the incorporation of nucleotides into a DNA chain, using DNA or RNA as a template)
reacción en cadena de la polimerasa; **RCP** polymerase chain reaction; PCR (a relatively cheap and rapid technique to copy and amplify segments of DNA in the laboratory, enabling them to be sequenced)

Figura 12: Entrada de *polimerasa* en el *Elsevier's Dictionary of Medicine: Spanish-English English-Spanish*

En general, nos parece que este recurso integra información básica, dado su soporte, pero adecuada. Por lo tanto, puede cumplir las expectativas de usuarios expertos que busquen equivalentes en inglés o español, así como del público lego que requiera, además, breves explicaciones conceptuales.

3.6.2.6 DiCoEnviro. Le dictionnaire fondamental de l'environnement

DiCoEnviro es una base de datos terminológica online (http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi) que se está desarrollando en el Observatoire de linguistique Sens-Texte de la Universidad de Montreal bajo la dirección de Marie-Claude L'Homme. Contiene términos relacionados con el medio ambiente en diferentes lenguas: inglés (982 entradas), francés (1309 entradas), español (172 entradas), portugués (37 entradas) e italiano (34 entradas) (L'Homme et al. 2018). Está destinado a usuarios que deseen conocer la lengua del medio ambiente. Así, para responder a sus diversas necesidades de la mejor manera posible, se puede acceder a la información sin metalenguaje técnico, para usuarios que requieran información conceptual y combinatoria, o bien incluyéndolo, para lingüistas y usuarios interesados en el comportamiento de los términos y los conceptos. Se trata de un recurso terminográfico muy avanzado, con muchos puntos fuertes, como la integración de información conceptual y lingüística, que comentaremos a continuación.

En este sentido, se describen distintos tipos de términos: sustantivos, adjetivos, verbos y adverbios, además de algunas siglas, cuando estas se comportan como

sustantivos (p. ej. CO2). Se desmarca de los tradicionales recursos terminográficos al ofrecer la estructura argumental de los términos, mostrando cómo estos se combinan entre sí. Además, se integran algunas relaciones semánticas y se apuesta por definiciones reducidas, frente a las largas definiciones enciclopédicas que suelen ser la norma. Para describir este comportamiento lingüístico de los términos, *DiCoEnviro* se basa en la Teoría Sentido-Texto y, en concreto, en la Lexicología explicativa y combinatoria (Mel'čuk et al. 1984, 1995), aplicada en otros recursos del Observatoire de linguistique Sens-Texte, como el DiCo (Polguère 2003), el *Dicouèbe. Dictionnaire en ligne de combinatoire du français* (Jousse y Polguère 2005), el *DiCoInfo. Dictionnaire fundamental de l'informatique et de l'Internet* (Observatoire de linguistique Sens-Texte 2019b) y el *JuriDico. Dictionnaire terminologique portant sur le domaine du droit* (Pimentel 2012). Asimismo, en el plano conceptual, *DiCoEnviro* se inspira en la Semántica de Marcos (Fillmore 1977, 1982, 1985) y, en particular, en el proyecto FrameNet (Baker et al. 1998) para la anotación de contextos y la delimitación de marcos conceptuales en el dominio medioambiental.

Para acceder a la información, podemos partir desde la interfaz en cualquiera de los idiomas del recurso. Entonces, podremos elegir entre dos tipos de búsqueda: por orden alfabético (a partir de una pestaña en la que se ofrece una lista con todos los términos) o mediante opciones de búsqueda personalizada. Estas opciones personalizadas son las siguientes, que podrán combinarse del modo que se prefiera: (i) Modo (palabra, término, relación léxica o expresión); (ii) Contenido (francés, inglés, español, italiano, portugués o multilingüe); (iii) Precisión (exacto, empieza por, contiene); y (iv) el cuadro de texto libre para introducir nuestros términos. Igualmente, podremos marcar la casilla «Mostrar equivalencias» para obtener a la vez los términos correspondientes en las otras lenguas. Nos parece un planteamiento del acceso a la información muy acertado, ya que cualquier usuario podrá acceder a esta desde el punto que más les convenga.

En cuanto a los datos que obtenemos con cada opción, en la búsqueda por palabra podemos introducir un término y obtener las entradas en las que este aparezca, ya sea como término simple o como parte de un compuesto. Nos parece una opción muy útil para la identificación de CN, si bien consideramos que su combinación con el criterio de precisión exacta (si quisiéramos marcarlo) no ofrecería resultados diferentes.

La siguiente opción que encontramos es la búsqueda por término, que consiste en la búsqueda del término o términos introducidos (en su forma exacta o como parte de términos compuestos, según el criterio de precisión seleccionado). Consideramos que la

búsqueda por palabra y la búsqueda por término podrían modificarse para permitir consultas más precisas y, a su vez, simplificar el sistema de búsquedas. Por un lado, podrían nombrarse de un modo más claro para que se entienda mejor la finalidad de cada búsqueda, ya que es probable que muchos usuarios no comprendan la diferencia entre palabra y término con la que se conciben estas opciones. Por otro lado, se podrían regular las funciones de ambas consultas, pues se solapan en la mayoría de los casos. Así, por ejemplo, se podrían aunar con los criterios de precisión y reducirse a una búsqueda exacta y otra parcial. De este modo se podrían resolver ciertas discordancias, como el hecho de que la búsqueda por término y la precisión *contiene* arrojen los mismos resultados que la búsqueda por palabra.

También es posible realizar búsquedas por relación léxica, en las que se presentan relaciones como la antonimia, la sinonimia o la hiperonimia. Se muestran además otros términos relacionados, bajo la consigna «Significados relacionados». Consideramos que esta opción presentaría un valor añadido si se indicase de qué relación se trata en cada caso. Mediante esta consulta se obtienen también categorías gramaticales derivadas y relaciones paradigmáticas, es decir, colocaciones y CN (a los que también se puede acceder mediante la búsqueda por palabra).

Por último, encontramos la búsqueda por expresión, que permite obtener más información de las colocaciones verbales (p. ej. *have an impact*). En concreto, se ofrece su estructura argumental y, si se selecciona la opción de mostrar equivalencias, se mostrarán las traducciones en otros idiomas, una información muy valiosa y de vital importancia en contextos interlingüísticos.

Además de la búsqueda por modo, cuyos criterios acabamos de describir, en el cuadro «Contenido» podemos elegir el idioma que nos interese: inglés, francés, español, italiano o portugués. Resulta muy interesante la opción «multilingüe», en la que se muestra la secuencia léxica en todas las lenguas del recurso. Como detalle, consideramos que sería más conveniente que esta opción no fuera sensible a las tildes, ya que así podríamos obtener más resultados, especialmente si tenemos en cuenta que estos recursos muchas veces son utilizados por hablantes no nativos que no siempre conocen la correcta grafía de los términos. Por ejemplo, si introducimos *erosion*, aparece *erosion* (en), pero no *érosion* (fr) ni *erosión* (es), que cuentan con entradas en *DiCoEnviro*.

Se proporciona una entrada para cada concepto, diferenciadas con un subíndice, lo que facilita la identificación de los distintos conceptos. Los CN tienen entradas propias, a las que podemos acceder desde diferentes puntos, como acabamos de comentar. Cada

entrada, ya sea de un término simple o compuesto, incluye la categoría gramatical del término y su definición, que está basada en la estructura argumental del concepto. Esta estructura argumental se presenta por medio de roles semánticos (AGENT, PATIENT, DESTINATION, SOURCE y CAUSE). Así pues, en la definición aparece la realización típica de cada argumento y, si deslizamos el ratón sobre cada una de ellos, podemos ver qué tipo de rol es. Además, podemos obtener una lista de todas las realizaciones de ese argumento pulsando en el botón + verde. Pese a que alabamos la intención de reducir el tamaño de las definiciones, este tipo de definiciones compuestas solo por roles nos parecen poco explicativas en su mayoría. Por ejemplo, la definición que se propone de *sedimentation* es ~ *of particle*, que no nos parece muy aclaratoria. Otras veces encontramos que las definiciones de varias acepciones se solapan. Por ejemplo, hay varias acepciones de *pollution*: 1. *Pollution of air, water, by substance*. 2. *Pollution of air, water by human or activity, by gas.* 3. *Pollution of air, water caused by human or activity*. Consideramos que estas tres definiciones podrían aunarse en una sola, por ejemplo: *Pollution of air or water by human, activity or substance*.

Además de la categoría gramatical y la definición, a menudo se muestran imágenes, sinónimos, abreviaturas, contextos (en ocasiones anotados con roles semánticos, uno de los puntos fuertes del recurso) y relaciones léxicas, por lo que se trata de entradas muy completas. Las relaciones léxicas que se muestran en cada entrada son más exhaustivas que las que se obtienen mediante la búsqueda por relaciones léxicas, en el sentido de que incluyen más términos relacionados, además de colocaciones verbales y CN. Consideramos que esta información es muy valiosa tanto para la comprensión como para la producción de textos, ya que el usuario puede conocer verdaderamente la situación o marco en el que se inserta el concepto. A ello también contribuye la versión de *DiCoEnviro* en marcos, a la que se puede acceder desde cada entrada mediante el icono . Se trata de marcos inspirados en FrameNet y otros concebidos especialmente para *DiCoEnviro*, y constituyen un valor añadido de este recurso. Como se puede ver en la Figura 13, en esta versión de marcos se muestra la definición del concepto (*agriculture*), ejemplos contextuales, notas, los argumentos obligatorios (*Participants 1*), los argumentos opcionales o adjuntos (*Participants 2*), las unidades léxicas relacionadas con el marco en las distintas lenguas del recurso (abajo) y las relaciones con otros marcos (a la derecha).

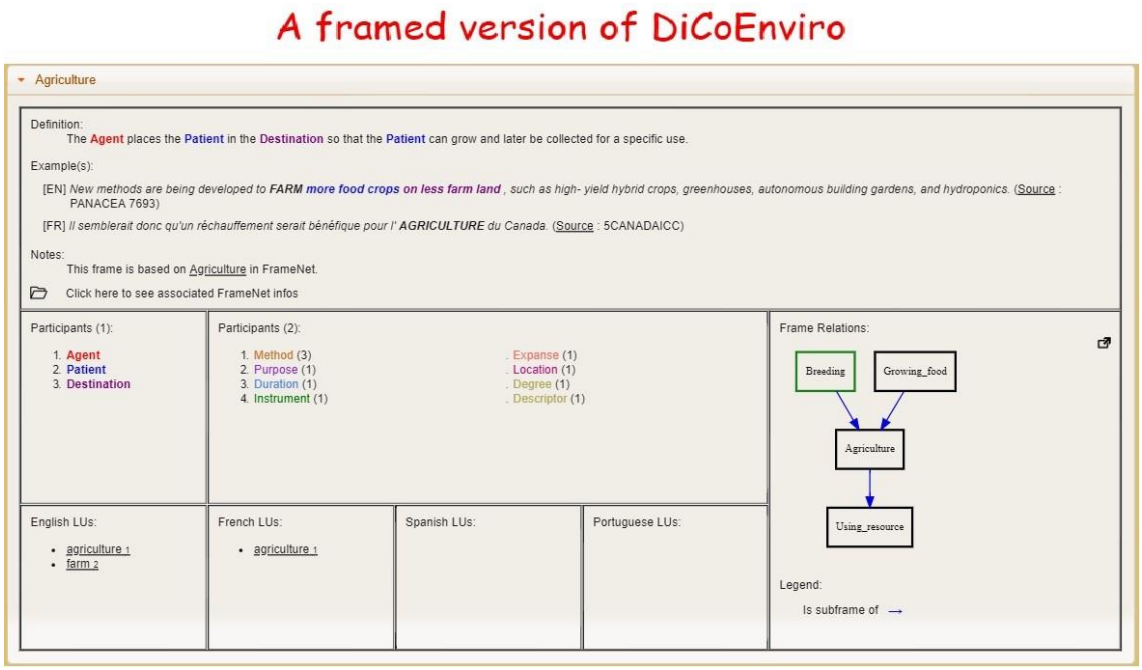

Figura 13: Entrada de *agriculture* en la versión de marcos de *DiCoEnviro*

Sin embargo, a nuestro parecer, la joya de *DiCoEnviro* es NeoVisual, una de las últimas novedades incorporadas (L'Homme et al. 2018). Se trata de una representación visual de las relaciones léxicas que aparecen en cada entrada, a la que podemos acceder mediante el icono . En la Figura 14 presentamos el ejemplo de *climate*, en el que se muestran sus argumentos obligatorios (*Arguments*), los conceptos con significado relacionado (*Related meaning*), sus combinaciones más frecuentes (*Combinations*), otras categorías gramaticales derivadas (*Word family*), sus hipónimos —que serían CN en este caso— (*Types of*) y sus equivalentes en otras lenguas (*Equivalents*).

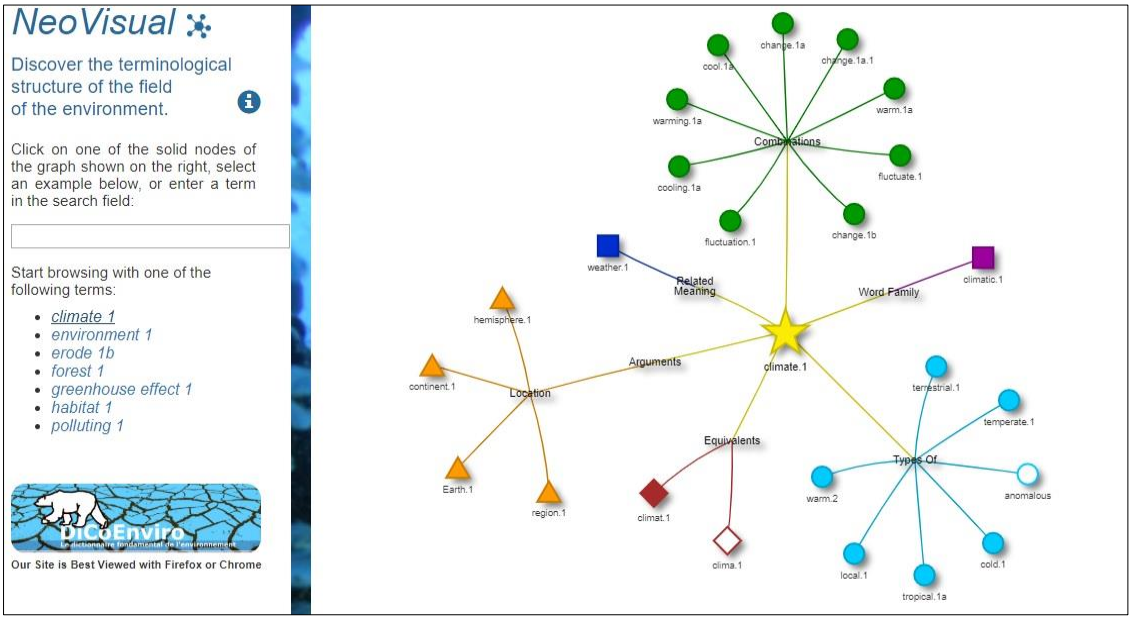

Figura 14: Entrada de *climate* en la herramienta NeoVisual de *DiCoEnviro*

Nos parece una herramienta poco común y de grandísima utilidad. Sin embargo, consideramos que algunos puntos podrían mejorarse. Por ejemplo, la relación *Others*, que se muestra en otras entradas, se podría precisar con el tipo de relación a la que hace referencia. Lo mismo ocurre con la relación *Related Meaning*: consideramos que debería especificarse cuál es la relación existente con esos términos. En cuanto a los CN, que aparecen tanto como hipónimos en el apartado *Types of* como en el apartado *Combinations*, no compartimos la práctica de omitir el núcleo, creemos que sería mejor incluir la forma completa, ya que no siempre será sencillo saber si lo que se muestra es el núcleo o el modificador de la combinación. No obstante, nos parece un gran paso que se incluya un apartado exclusivo de combinaciones (tanto nominales como verbales o adjetivales), lo que constituye una muestra de la preocupación de *DiCoEnviro* por ofrecer información completa para distintos usuarios.

En resumen, en este capítulo hemos recogido los principales aspectos en relación con los CN. En la Sección 3.1 nos hemos centrado en su papel como términos compuestos y, así, unidades fraseológicas especializadas. En la Sección 3.2 hemos descrito nuestra noción de CN y hemos presentado los principales estudios que se han realizado al respecto desde diferentes disciplinas, para después ahondar en los tipos y la formación de estos términos compuestos (Sección 3.3). El análisis pormenorizado de estas unidades ha ocupado la Sección 3.4, en la que nos hemos detenido en su desambiguación estructural (Sección 3.4.1), la semántica de sus formantes (Sección 3.4.2), el análisis de las relaciones entre estos (Sección 3.4.3), la variación habitual en los CN (Sección 3.4.4) y la neología, con la que mantienen una estrecha relación (Sección 3.4.5). Seguidamente, hemos profundizado en dos de las aplicaciones principales de dicho análisis: la traducción (Sección 3.5) y la representación de los CN en recursos lingüísticos (Sección 3.6).

4. Materiales y metodología

En este capítulo presentamos los materiales utilizados en el estudio, que consisten en un corpus de energía eólica y la herramienta de análisis de corpus Sketch Engine. A continuación, comentamos la metodología seguida para la compilación del corpus y la extracción y análisis de los términos, que sigue las premisas de la Terminología Basada en Marcos.

4.1 Materiales

4.1.1 Corpus de energía eólica

Para llevar a cabo el análisis terminológico nos servimos de un corpus que compilamos manualmente. Entendemos por *corpus* un conjunto de muestras lingüísticas seleccionadas según unos criterios específicos para utilizarse como representación de la lengua (Sinclair 1996: 4).

Se trata de un corpus comparable compuesto por textos especializados en energía eólica escritos originalmente en inglés y español, que suman aproximadamente 3 millones de palabras en cada lengua (en concreto, 3 025 237 palabras en inglés y 3 022 698 palabras en español). Este tipo de corpus no debe confundirse con los corpus paralelos (Sección 3.5), que están formados por textos originales y sus traducciones. La denominación de *corpus paralelos* se utiliza a menudo para referirse a corpus como el nuestro debido a su confusión con la noción de textos paralelos habitual de la traducción (textos de la misma temática y función).

Como señala Bowker (2015), desde la Terminología y la Traducción siempre se ha incidido en la importancia de que los textos empleados para la investigación deben estar escritos originalmente en la lengua de estudio. Esto se debe a que las traducciones pueden contener expresiones no idiomáticas, construcciones sintácticas que no son propias de la lengua o incluso errores. Así, no son pocos los autores que se han mostrado a favor de los corpus comparables, debido a los problemas que plantean los corpus paralelos. Sus desventajas se basan principalmente en su escasez, sobre todo en el ámbito científico, donde el inglés constituye la *lingua franca* (Humbley y García Palacios 2012); y en la influencia que reciben de la lengua origen (Kübler y Volanschi 2012).

Por otra parte, la energía eólica, de indudable actualidad debido a la creciente preocupación por el cuidado del medio ambiente, constituye la temática principal del corpus. Se trata de un área interdisciplinar, en la que confluyen el dominio puramente

medioambiental y la ingeniería, con tintes de otras disciplinas como las matemáticas. Todo ello tiene consecuencias en el corpus, como se observa en la abundancia de fórmulas y términos técnicos y, así, en la presencia de terminología propia de las áreas que convergen. En la Tabla 4 se muestra el tipo y número de textos que conforman los corpus en inglés y en español, cuya compilación comentaremos en la Sección 4.2.1.

Corpus en inglés		**Corpus en español**	
Tipo de texto	**N.º de textos**	**Tipo de texto**	**N.º de textos**
Artículo	113	Artículo	58
Capítulo de libro	1	Capítulo de libro	4
Tesis	8	Tesis	37
Libro	19	Libro	6
Informe	9	Informe	22
		Legislación	16
Total	**150**	**Total**	**143**

Tabla 4: Composición de los corpus en inglés (izquierda) y en español (derecha)

De la Tabla 4 se desprende que el corpus en inglés cuenta con más artículos que el español, debido a que este tipo de textos suele redactarse en inglés y no encontramos suficientes artículos en español que versaran sobre la misma temática. Lo mismo sucede con los libros en formato completo, de modo que contrarrestamos la diferencia de palabras añadiendo más capítulos de libros, tesis doctorales e informes en español, que presentaban una mayor disponibilidad en acceso abierto, del mismo modo que la legislación relativa a la energía eólica en español. A pesar de tratarse de tipos diferentes de textos, los seleccionamos porque todos abordaban la energía eólica con un elevado nivel de especialización. Además, el número de textos de cada corpus es similar: 150 textos en el corpus en inglés y 143 textos en el corpus en español. Próximamente el corpus estará disponible para su libre consulta en Open Corpora de Sketch Engine.

4.1.2 Sketch Engine

Para trabajar con el corpus, empleamos la herramienta de análisis online Sketch Engine (Kilgarriff et al. 2004, 2014). Esta permite utilizar un gran número de corpus ya incorporados, como el British National Corpus o los corpus English Web (enTenTen) para el inglés o Spanish Web (esTenTen) y EUR-Lex para el español. Igualmente, el usuario puede cargar su propio corpus o bien ampliarlo o elaborar uno nuevo de manera

automática gracias a una función que permite extraer textos de la web a partir de URL o palabras clave.

Sketch Engine ofrece una gran variedad de funcionalidades que resultan útiles para las distintas tareas para las que puede utilizarse un corpus, como se observa en el nuevo diseño[36] de su interfaz principal, lanzado en 2018 (Figura 15).

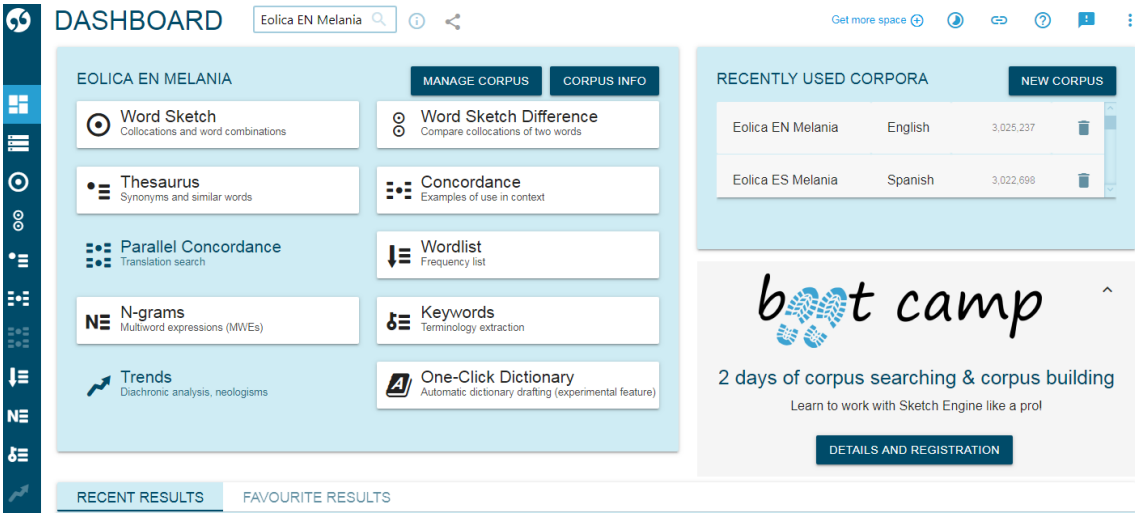

Figura 15: Interfaz principal de Sketch Engine

Destacamos las funciones de *Word Sketch*, *N-grams* y *Keywords*, que nos resultaron de utilidad al ofrecer información sobre la combinatoria de los términos. En concreto, los *word sketches* consisten en resúmenes de los principales modificadores de un término y los verbos que lo acompañan cuando este es sujeto u objeto.

La generación de términos compuestos también puede observarse mediante la opción *Keywords*, que extrae los principales términos simples (denominados *Keywords*) y compuestos (llamados *Terms*) del corpus, de modo que puede servir para obtener una panorámica de los términos más presentes en el corpus. En esta línea, la opción *N-grams* es una nueva incorporación del reciente diseño de Sketch Engine que ofrece unas prestaciones muy interesantes para la extracción de CN, ya que permite aunar consultas en Corpus Query Language (CQL) (Schulze y Christ 1996) de las secuencias que nos interesen con criterios como la frecuencia mínima o máxima deseada o la exclusión de *stop words*, o bien sustituir las consultas en CQL por parámetros como *starting with letters, ending with letters, starting with word, ending with word*, etc. Sin embargo, a pesar de lo prometedor del enfoque, el tiempo de procesamiento de la consulta es elevado y en ocasiones no llega a aportar resultados.

[36] En ocasiones podrán encontrarse en este trabajo figuras que representen el diseño anterior de Sketch Engine, al que se recurrió cuando el nuevo diseño no ofrecía las funcionalidades deseadas.

Por tanto, como veremos en la Sección 4.2.2, para la extracción de CN preferimos la opción *Concordance*. Esta fue la herramienta más empleada en nuestro estudio, ya que permite analizar las líneas de concordancia de nuestra búsqueda, a las que podemos acceder mediante consultas simples o avanzadas. Estas últimas pueden realizarse, por ejemplo, mediante expresiones regulares con el lenguaje CQL, que permite realizar búsquedas específicas basadas en lemas o etiquetas gramaticales, entre otras posibilidades. Así, empleamos la opción *Concordance* para distintas tareas, entre las que destacan la extracción de los CN, su análisis semántico y la búsqueda de equivalentes. La Figura 16 ejemplifica una de las consultas realizadas mediante la función *Concordance*, en concreto, el análisis de concordancias del CN *stall-regulated wind turbine*.

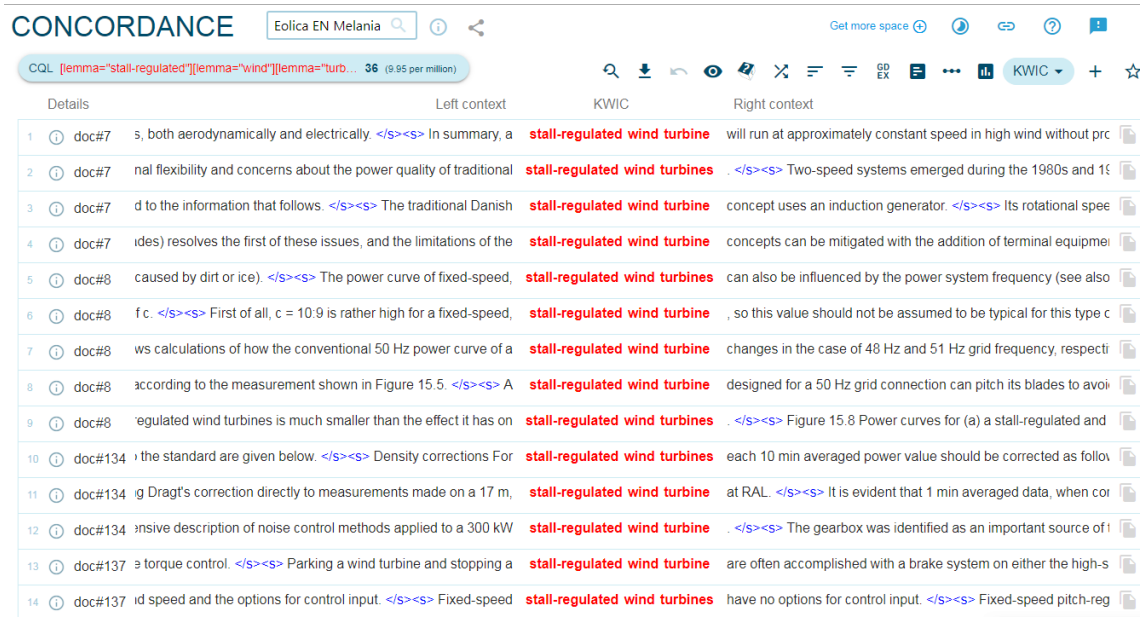

Figura 16: Resultados de la opción *Concordance* para la consulta [lemma="stall-regulated"][lemma="wind"][lemma="turbine"]

4.2 Metodología

4.2.1 Compilación del corpus

La compilación o preparación del corpus constó de tres procesos principales: (i) búsqueda y selección de textos, (ii) limpieza de textos y (iii) almacenamiento en Sketch Engine.

La búsqueda de textos se realizó de forma manual y, para ello, se consultaron principalmente bases de datos científicas, como la Web of Science y Scopus. Asimismo, nos servimos de repositorios digitales y de Google para acceder, por ejemplo, a los informes y los textos legislativos. Para la selección de los textos, nos guiamos por los tres criterios señalados por Buendía Castro (2013), en los que resume las propuestas de diversos

autores respecto a la validez y fiabilidad de las fuentes. Estos criterios son la autoridad (reputación y experiencia del autor), el contenido (cobertura temática, objetividad, actualidad y destinatarios) y el diseño de las fuentes (accesibilidad y facilidad de tratamiento) (Buendía Castro 2013: 330).

En cuanto a la autoridad, seleccionamos artículos publicados en revistas y congresos de ámbito nacional e internacional, tales como *Wind Energy* o *Ingeniería Energética*. Asimismo, se buscaron tesis doctorales realizadas en universidades ubicadas en países de habla hispana o inglesa, como la Universidad Politécnica de Madrid o la Stanford University. Los informes, por otra parte, también fueron emitidos por organizaciones de reconocida autoridad, tales como el National Renewable Energy Laboratory of the U.S. o la Asociación Empresarial Eólica. Para los libros y textos legislativos se asumió una autoría apropiada, al provenir de editoriales y organismos institucionales.

En este sentido nos gustaría realizar un apunte en torno a la lengua de redacción de los textos. Debido al estatus del inglés como *lingua franca* (Humbley y García Palacios 2012) y a la creciente movilidad internacional, no es tarea fácil conocer la lengua nativa de los autores de los textos. Dada la dificultad y escasa fiabilidad de criterios como el origen de sus nombres, decidimos guiarnos por la localización de las instituciones que publican esos trabajos, así como por la garantía de corrección lingüística en el caso de publicaciones que han pasado un proceso de revisión y edición (p. ej. en revistas o libros). No obstante, somos conscientes de que se trata de un criterio difícil de determinar.

Por lo que respecta al contenido, seleccionamos textos que versaran sobre energía eólica, para lo que comprobamos que incluyeran términos clave como *wind power*, *wind energy, wind turbine* o *wind farm* en el título o el resumen. Además, los textos presentan una temática equilibrada y representativa dentro del ámbito de la energía eólica en ambas lenguas, lo que conseguimos gracias al proceso iterativo de selección que comentaremos a continuación y a nuestra documentación sobre los diferentes temas abordados en el panorama de la energía eólica. No obstante, suscribimos las palabras de Sanz Vicente (2011: 387) respecto al equilibrio y la representatividad de los corpus:

> Por muy representativo y equilibrado que sea el muestreo, un corpus siempre será una muestra finita de una población infinita y nunca se podrá abarcar todo el conjunto de acontecimientos lingüísticos. Solo poseerá estas cualidades si responde a las necesidades conceptuales, pragmáticas y lingüísticas que se le presentan al terminógrafo, es decir, únicamente cuanto mejor responda al fin para el cual ha sido creado.

En esta línea, comprobamos también que se trataba de textos objetivos, como se observa en los distintos tipos de textos incluidos, en su mayoría producto de investigaciones científicas. Estos debían presentar un nivel de especialización elevado. La inclusión de diferentes tipologías textuales no afecta al nivel de especialización requerido, ya que los distintos textos ilustran de forma representativa el modo de comunicarse de los especialistas. Por otra parte, los textos fueron publicados en su mayoría entre la década de los 2000 y la actualidad, si bien nuestro estudio no persigue cuestiones diacrónicas.

Por último, a la hora de evaluar el diseño de los textos para su selección, atendimos a su accesibilidad (por ejemplo, la disponibilidad del texto completo) y la facilidad de su tratamiento (textos disponibles para su descarga, que no estuvieran formados por escasas cantidades de texto frente a numerosas fórmulas, etc.), de cara a la posterior fase de limpieza y preparación.

Así, comenzamos buscando textos en inglés que cumplieran con estos criterios. A partir de estos, consultamos sus palabras clave en Sketch Engine, cuyas traducciones nos sirvieron de elementos de búsqueda para los textos en español. Como ya indicamos en la Sección 4.1.1, el número de textos especializados sobre energía eólica en español es menor que el de documentos en inglés, lo que nos obligó a variar la cantidad de cada tipología textual en los dos corpus. A continuación, volvimos a incluir textos en inglés, a los que accedimos seleccionando los documentos más recientes del repositorio que estuviéramos consultando. Seguimos este proceso iterativo hasta alcanzar una cantidad similar de palabras en ambos corpus (3 millones de palabras).

Antes de almacenar los textos en Sketch Engine, los sometimos a un proceso de limpieza y preparación manual en el editor de texto plano Notepad++. Este proceso estaba encaminado a adaptar los textos para su máximo aprovechamiento en la herramienta de análisis de corpus. Así, dadas las características y el formato de los textos científicos, estos contaban con multitud de fórmulas, figuras y tablas, así como palabras rotas por guiones o números de página que interrumpían el texto, entre otros aspectos. Por tanto, la limpieza y preparación de los documentos consistió, principalmente, en la adición de puntos para diferenciar los apartados que no los tuvieran y evitar así la unión de texto que no pertenecía a la misma oración. Igualmente, suprimimos la información irrelevante para nuestro estudio (p. ej. las referencias bibliográficas, ya que suelen estar escritas en inglés y en el corpus en español no resultarían útiles, además de que aumentan el corpus por medio de información irrelevante, como los nombres de los autores) y aquella que interrumpía el texto (p. ej. los números de página o la información de tablas y fórmulas,

que sustituimos por la secuencia *xx* o una cadena similar de letras que nos sirviera como identificación de la información antes existente). En ocasiones fue necesario normalizar las tildes en español, ya que la conversión del documento a texto plano no siempre las reconocía correctamente.

Una vez preparados los textos, procedimos a su almacenamiento en Sketch Engine, en forma de dos subcorpus diferentes (inglés y español). Para poder hacer uso de las diferentes funciones que ofrece la herramienta, compilamos los corpus, lo que incluye su etiquetado morfosintáctico y la aplicación de una gramática de *sketches* para representar la combinatoria de las unidades. Así, el etiquetado morfosintáctico para el corpus en inglés se realizó con TreeTagger 2.5 y, para el corpus en español, con FreeLing 2.0. La gramática de *sketches* aplicada para el corpus en inglés fue English 3.3 for TreeTagger pipeline v2, mientras que para el corpus en español aplicamos Spanish (Freeling tagset) 2.0. Además, indicamos que cada texto que componía el corpus se identificara con una numeración diferente, para así poder evaluar la presencia de los términos en distintos textos. En general, el proceso de preparación del corpus fue eminentemente manual, lo que implica una inversión considerable de tiempo, pero a la vez supone un valor añadido y garantiza la calidad de la herramienta principal que empleamos en nuestro estudio: los corpus.

4.2.2 Extracción de compuestos nominales en inglés

La extracción automática de unidades fraseológicas, para la que a menudo se recurre a medidas de asociación (Baldwin y Kim 2010), ha sido un tema ampliamente desarrollado en la Lingüística Computacional. Sin embargo, estos trabajos se centran con frecuencia en la lengua general, pues requieren de corpus de un tamaño mucho mayor que el nuestro, que son difíciles de encontrar en la lengua especializada. Además, suelen precisar una validación manual que confirme los resultados. Por ello, a pesar de las ventajas que ofrece la extracción automática, en este trabajo llevamos a cabo un proceso de extracción manual de CN, que preferimos (a pesar de la inversión de tiempo) por sus garantías de calidad.

Así, partimos de la extracción de CN en inglés, al ser esta la lengua de la ciencia, especialmente proclive a la formación de CN. Dado que nos centramos en los CN, que tienen un núcleo nominal, el primer paso consistió en la extracción de una lista de los sustantivos del corpus, pues estos podrían intervenir como núcleo o modificador de los CN. A pesar de que puede haber otras categorías gramaticales especialmente recurrentes en el corpus (p. ej. adjetivos como *renewable*), nuestro enfoque integró estos elementos, ya que formaron parte de los CN extraídos, como veremos a continuación. Además, partiendo de los sustantivos obtuvimos mejores resultados, ya que los sustantivos son las principales unidades terminológicas que transmiten el conocimiento. Por tanto, para la extracción de

sustantivos, utilizamos la siguiente consulta en CQL: [tag="N.*"], frente a la opción de *Keywords*, que ofrece un número más reducido de resultados. Una vez ordenados por frecuencia y por lema, obtuvimos 35 240 *types* (Figura 17).

	Lemma	↓ Frequency	Frequency per million
1	wind	45,755	12,641.95
2	power	25,697	7,099.99
3	turbine	19,978	5,519.85
4	energy	15,742	4,349.46
5	system	14,342	3,962.64
6	speed	9,738	2,690.58
7	cost	8,938	2,469.54
8	farm	6,770	1,870.53
9	rotor	6,135	1,695.08
10	blade	5,956	1,645.62
11	capacity	5,953	1,644.79
12	voltage	5,950	1,643.96
13	generator	5,861	1,619.37
14	model	5,668	1,566.05

Figura 17: Muestra de sustantivos del corpus en inglés ordenados por frecuencia

A continuación, seleccionamos cinco sustantivos que, a pesar de no ser exactamente los cinco primeros elementos de la lista, son los participantes fundamentales del EVENTO DE LA ENERGÍA EÓLICA, ya que describen el proceso de recepción del viento en el aerogenerador hasta la generación de la energía. Estos sustantivos son *wind*, *power*, *turbine*, *voltage* y *generator*. Nos centramos en estos sustantivos por su representación del EVENTO DE LA ENERGÍA EÓLICA y su capacidad de generación de CN, de diferente forma y extensión, y con dificultades de traducción debido a su especificidad y su falta de representación en recursos. En la Figura 18 se presenta un sistema simplificado del EVENTO DE LA ENERGÍA EÓLICA en el que se muestran las relaciones entre estos cinco términos.

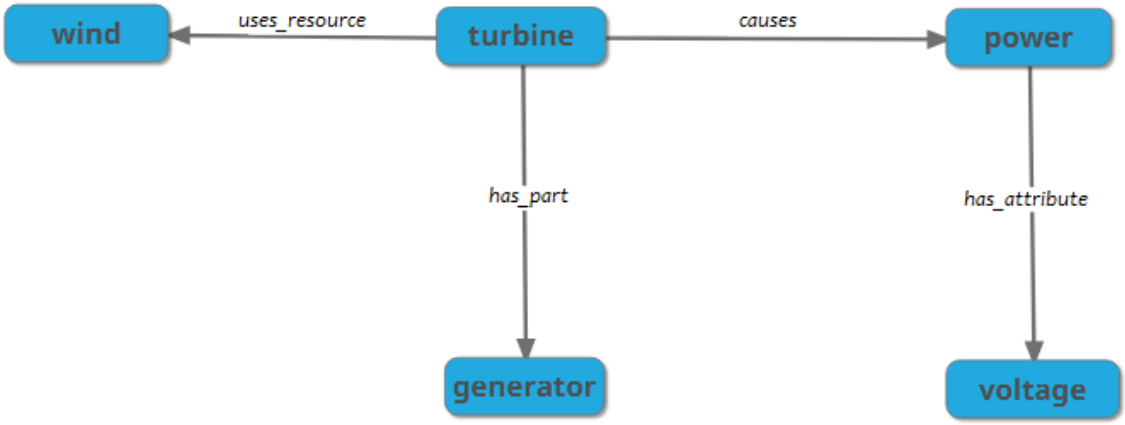

Figura 18: Sistema conceptual simplificado del EVENTO DE LA ENERGÍA EÓLICA

Una vez seleccionados los cinco términos de partida, realizamos consultas en CQL para extraer los CN que contuvieran esos términos. Elegimos este enfoque, frente a la extracción de listas de CN ordenados por frecuencia, porque el estudio de la capacidad de generación de un elemento nos parece más interesante para explorar la formación de CN y analizar cómo se relacionan entre sí en una estructura jerárquica y no jerárquica, con especial énfasis en la organización del conocimiento. El otro enfoque posible, basado en la extracción de CN ordenados por su presencia en el corpus, ofrecería CN muy frecuentes pero cuyas relaciones podrían no ser directas, pues en muchas ocasiones faltarían conceptos intermedios en el sistema conceptual. Además, no nos permitiría estudiar el potencial de formación de CN partiendo de un mismo elemento.

Por tanto, diseñamos consultas en CQL a partir de las diferentes estructuras que pueden adquirir los CN en inglés y que hemos venido analizando en artículos de investigación en los que hemos trabajado con este corpus. A pesar de que los CN formados por premodificación son los más habituales en inglés, nos interesaba analizar también la presencia y formación de otro tipo de estructuras. Así, diseñamos varias expresiones regulares, que presentamos a continuación con el ejemplo de *generator*. La primera de ellas sirve para extraer los CN formados por premodificación en inglés en los que nuestro término es el núcleo (*induction generator*) (Tabla 5), mientras que la segunda extrae CN en los que nuestro término ocupa una posición de modificador (*generator torque*) (Tabla 6). Por otra parte, diseñamos otras dos expresiones regulares para extraer CN formados por postmodificación, una para los CN en los que nuestro término es el núcleo (*generator of wind turbine*) (Tabla 7) y otra en los que es el modificador (*output of the generator*) (Tabla 8).

En la expresión 1 (Tabla 5), buscamos el lema *generator* ([lemma="generator"]), que podía ir precedido de sustantivos, adjetivos, adverbios, participios pasados o participios presentes que aparecieran una o más veces [tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,} y que no podía ir seguido de sustantivos ni adjetivos ([tag!="N.*|JJ.*"]), para evitar que el CN extraído formase parte de una secuencia mayor. Así, obtuvimos CN como *synchronous generator* o *doubly fed induction generator*.

[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,}[lemma="generator"][tag!="N.*|JJ.*"]

Tabla 5: Expresión 1. CN con premodificación y el término en el núcleo

En la expresión 2 (Tabla 6), buscamos el lema *generator* ([lemma="generator"]), que podía o no ir precedido o seguido de sustantivos, adjetivos, adverbios, participios pasados o participios presentes ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}). Necesariamente la secuencia debía acabar con un sustantivo ([tag="N.*"]), que sería el núcleo y que, de nuevo, no podía ir seguido de adjetivos ([tag!="JJ.*"]), para evitar que el CN interviniese en una cadena mayor. Extrajimos CN como *generator torque* y *generator terminal voltage*.

[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[lemma="generator"][tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[tag="N.*"][tag!="JJ.*"]

Tabla 6: Expresión 2. CN con premodificación y el término en el modificador

En la expresión 3 (Tabla 7), el núcleo ([lemma="generator"]) se encontraba en la parte izquierda del CN, si bien podía ir precedido de otras unidades ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}), en lo que sería un CN mixto formado mediante los dos procedimientos (p. ej. *renewable source of energy*). A la derecha le debía seguir necesariamente un sintagma preposicional formado por una preposición ([tag="IN"]), un artículo opcional ([tag="DT"]{0,}) y un sintagma nominal formado por un sustantivo ([tag="N.*"]) que podía ir precedido de modificadores como sustantivos, adjetivos, adverbios, participios pasados o participios presentes ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}). Esta consulta nos permitió detectar CN como *generator of wind turbine*.

[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[lemma="generator"][tag="IN"][tag="DT"]{0,}[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[tag="N.*"]

Tabla 7: Expresión 3. CN con postmodificación y el término en el núcleo

Por último, la expresión 4 (Tabla 8) permitió extraer CN formados por un sustantivo ([tag="N.*"]), que podía ir precedido de sustantivos, adjetivos, adverbios, participios pasados o participios presentes ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}). A dicho

sustantivo le seguía un sintagma preposicional formado por una preposición ([tag="IN"]), un artículo opcional ([tag="DT"]{0,}) y un sintagma nominal formado por un sustantivo ([tag="N.*"]), que no podía ir seguido de un adjetivo ([tag!="JJ.*"]), para evitar que el CN formase parte de una secuencia mayor. Además, este sustantivo podía ir precedido de modificadores, entre los que necesariamente debía figurar *generator* ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[lemma="generator"][tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}). Así, obtuvimos CN como *output of the generator*.

[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[tag="N.*"][tag="IN"][tag="DT"]{0,}[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[lemma="generator"][tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{0,}[tag="N.*"][tag!="JJ.*"]

Tabla 8: Expresión 4. CN con postmodificación y el término en el modificador

Cabe señalar que, a pesar de que se recomienda establecer un *span* de ± 5 palabras para evitar ruido (Montero Martínez y García de Quesada 2004), decidimos no fijar un límite máximo de formantes en los CN, ya que nos interesaba explorar la longitud que estos pueden alcanzar.

Una vez extraídos los CN, nos centramos en los 40 CN más frecuentes generados por cada uno de los cinco términos, de manera que recogimos una muestra de 200 CN. A continuación, revisamos sus líneas de concordancias para verificar que se hubieran extraído correctamente (por ejemplo, que no formaran parte de un CN más extenso) y que se empleara en, al menos, tres[37] textos diferentes, como también indica Sanz Vicente (2011), para evitar usos propios de un autor. Por otra parte, dado que los cinco sustantivos de partida son conceptos clave del dominio, se combinan frecuentemente entre ellos para formar CN. Por tanto, la extracción de los 40 CN más frecuentes de cada sustantivo dio lugar a distintos CN duplicados en las listas de varios términos (p. ej. *wind turbine rotor* figuraba tanto entre los CN de *wind* como entre los de *turbine*). Esto se produjo en CN en los que casi siempre intervenía *wind*, al ser el término central del dominio. Descartamos, además, las combinaciones que no se incorporarían a EcoLexicon ya que no representarían conceptos del EVENTO MEDIOAMBIENTAL (*turbine type, turbine cost, turbine component*).

Nuestra lista de términos se redujo, por tanto, a 182 CN (Anexo 1), cuya frecuencia variaba entre 4821 y 3 ocurrencias. A pesar de que 3 repeticiones podrían considerarse

[37] Bowker (1998: 493) propone un número superior de textos, en concreto doce. Sin embargo, nos parece un límite muy elevado para un corpus del tamaño del nuestro, que probablemente omitiría CN interesantes. Esto puede suceder especialmente en un área con continuas innovaciones como es la energía eólica, en la que pueden surgir nuevos CN que aún no gocen de demasiada extensión.

insuficientes, no descartamos estos CN porque nos interesaba estudiar la formación y traducción de estas combinaciones, así como su relevancia en el sistema terminológico. En concreto, esta frecuencia reducida puede deberse a que denominan nuevos conceptos.

Tras realizar el análisis sintáctico-semántico de los CN, que comentaremos a continuación (Sección 4.2.3), llevamos a cabo una segunda fase de extracción de términos en inglés, en la que buscamos variantes denominativas de los CN, con especial atención a las siglas, dado su frecuente uso en el lenguaje especializado. Decidimos buscar las variantes para obtener una visión más profunda de las formas con las que conviven los CN. Así, identificamos variantes denominativas durante la fase de documentación en recursos lingüísticos y en textos sobre energía eólica, en los que encontramos patrones de conocimiento que indicaban la presencia de sinonimia (p. ej. *also known as, also referred to as, is a synonym of, can also be called*). Además, el análisis de concordancias así como las consultas concretas en CQL utilizando estos patrones de conocimiento sinonímicos (Tabla 9) también nos permitieron hallar variantes denominativas. Los patrones de conocimiento empleados fueron los siguientes:

- [tag="RB.*"]{0,}[word="known|called|referred"][tag="RB.*"]{0,}[word="to"]{0,}[word="as"]{0,} [tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,}[tag="N.*"]
- [lemma="be"]{0,}[]{0,3}[lemma="synonym"][word="of|for|to"] [tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,}[tag="N.*"]
- []{1,3}[lemma="spelling"][word="of"][]{0,3}[tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,}[tag="N.*"]

Tabla 9: Patrones de conocimiento sinonímicos

En la primera consulta buscamos las formas verbales *known, called* y *referred* ([word="known|called|referred"]), que podían ir precedidas o seguidas de adverbios ([tag="RB.*"]{0,}), además de que podían requerir la presencia de *to* o *as* ([word="to"]{0,} [word="as"]{0,}). Esta secuencia iba seguida del sustantivo que sería el sinónimo ([tag="N.*"]), al que podían preceder modificadores ([tag="N.*|JJ.*|RB.*|VVG.*| VVN.*"]{0,}). Por otro lado, en la segunda consulta buscamos el lema *synonym*, que podía ir precedido del verbo *to be* ([lemma="be"]{0,}) y un margen de hasta tres palabras ([]{0,3}), seguido de las preposiciones *of, for* o *to* ([word="of|for|to"]), además del sustantivo que designaría el sinónimo ([tag="N.*"]), que podía ir precedido de modificadores ([tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,}). Por último, la tercera consulta, que se centraba en las variantes gráficas, buscaba el lema *spelling* precedido de una a tres palabras ([]{1,3}), y seguido de la preposición *of* ([word="of"]), un margen de hasta tres palabras ([]{0,3}) y el

sustantivo que sería la variante ([tag="N.*"]), que podía ir precedido de modificadores ([tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,}).

Para la búsqueda específica de siglas utilizamos una consulta en CQL ([lemma="[A-Z]{2,}"]) que nos permitió obtener secuencias de letras mayúsculas formadas por al menos dos letras, entre las que descartamos aquellas que no constituían siglas válidas para nuestros CN (p. ej. UK, EU). Tras la inclusión de las variantes denominativas en inglés, el conjunto de términos que manejamos en esta lengua aumentó a 330 términos.

4.2.3 Análisis de los compuestos nominales en inglés y búsqueda de equivalentes en español

En este apartado presentamos el análisis de los CN, que se compuso de varias fases: (i) *bracketing*, (ii) asignación de categorías semánticas, (iii) descodificación de la relación semántica interna y (iv) asignación de roles semánticos.

Como hemos mencionado, la omisión de la relación semántica entre los formantes del CN constituye la principal dificultad para acceder al significado de estas unidades. Sin embargo, antes de descodificar dicha relación, es necesario establecer la estructura del CN en aquellos que cuentan con más de dos formantes, ya que estas dependencias internas señalan los elementos principales que se unen mediante la relación semántica. Por tanto, iniciamos el análisis de los CN con el *bracketing* o desambiguación estructural de estas unidades, con el objetivo de reducir los CN de más de dos formantes (61 de los 182 CN) a su estructura básica de núcleo y modificador. Por ejemplo, [*wind turbine*] *rotor*. Para ello, diseñamos un protocolo (Cabezas García y León Araúz 2019) que consiste en la comprobación de los siguientes indicadores:

-Indicador A: La posible agrupación aparece como un término independiente en el corpus.

-Indicador B: La posible agrupación no permite la inserción de elementos externos que modifiquen su significado.

-Indicador C: La posible agrupación forma otros CN.

-Indicador D: La posible agrupación tiene sinónimos o antónimos.

En la Sección 5.1.1.1 explicamos detalladamente la aplicación del protocolo y los resultados obtenidos. Dicho protocolo surge a partir del análisis profundo de las propiedades de los CN que realizamos en esta tesis, así como de la integración de los avances alcanzados desde la Lingüística Computacional. Se concibió con el objetivo de

ofrecer soluciones prácticas a profesionales como terminólogos o traductores, aunque también puede implementarse en sistemas de procesamiento del lenguaje natural. A modo de ejemplo, se presenta en la Tabla 10 el protocolo de *bracketing* aplicado a *wind energy system*:

wind energy system		
Indicadores	**Resultados**	***Bracketing* sugerido**
Indicador A	wind energy (1452)[38] energy system (37)	[wind energy] system
Indicador B	wind (…) energy system (2) (p. ej. wind turbine energy systems) wind energy (…) system (42) (p. ej. wind energy conversion system)	[wind energy] system
Indicador C	wind energy (1088 + 325 = 1413) (p. ej. wind energy production, offshore wind energy) energy system (7 + 118 = 125) (p. ej. energy system analysis, renewable energy system)	[wind energy] system
Indicador D	wind energy (wind power) energy system	[wind energy] system

Tabla 10: Ejemplo del protocolo de *bracketing* en *wind energy system*

Una vez aclarada la estructura de los CN, pasamos a asignarles categorías semánticas. Para ello, utilizamos un inventario de categorías desarrollado en el grupo de investigación LexiCon (Gil Berrozpe et al. 2019) a partir de los conceptos incluidos en EcoLexicon. Dicho inventario se basa en diferentes niveles de categorización (Murphy y Lassaline 1997) y en la semejanza conceptual (Hahn y Chater 1997). Para definir las categorías, se consultaron las definiciones de los conceptos así como su información contextual en el corpus de EcoLexicon. Tras determinar las características que compartían los conceptos, se establecieron manualmente las categorías y se organizaron jerárquicamente en cinco niveles. El nivel más general estaba formado por las categorías ontológicas ENTITY (objetos mentales y físicos), PROCESS (eventos que ocupan un periodo de tiempo) y ATTRIBUTE (características de las entidades o los procesos). Los siguientes

[38] Número de ocurrencias encontradas en el corpus.

niveles descendían en la jerarquía semántica, alcanzando una mayor especificidad. En la Tabla 11 se presenta un fragmento de dicho inventario.

Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
ENTITY	CREATION	ARTIFACT	CONDUIT	
			INSTRUMENT	MEASURING INSTRUMENT
				TRANSFORMING INSTRUMENT
		STRUCTURE	BUILDING	
			DEFENSE STRUCTURE	
	GEOGRAPHIC FEATURE	ARTIFICIAL GEOGRAPHIC FEATURE	ARTIFICIAL WATER BODY	
		NATURAL GEOGRAPHIC FEATURE	LANDFORM	NATURAL WATER BODY
	HUMAN	INSTITUTION		
		SPECIALIST		
PROCESS	ACTION	ANALYSIS		
		COLLECTION		
		MANAGEMENT		
	CHANGE	CHANGE IN SIZE/INTENSITY	DECREASE	
			INCREASE	
		DIVISION		
			POLLUTION	
		TRANSFORMATION	RESTORATION	
	MOVEMENT	ENERGY MOVEMENT		
		FLUID MOVEMENT	WATER MOVEMENT	
ATTRIBUTE	MEASUREMENT	MAGNITUDE	LEVEL	MEAN
	PHYSICAL ATTRIBUTE	COMPOSITION		
		SIZE		
	TIME			

Tabla 11: Fragmento del inventario de categorías semánticas de Gil Berrozpe et al. (2019)

Estas categorías se asignaron a los formantes del CN, los grupos del *bracketing* y el concepto designado por el CN mediante la consulta de sus definiciones y su información contextual en el corpus de energía eólica. En la Tabla 12 se presenta a modo de ejemplo la anotación de *generator torque control*.

generator torque control		
Elemento 3	generator	ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT
Elemento 2	torque	ENTITY>FORCE>STRESS
Elemento 1[39]	control	PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY
Bracketing	[generator torque]	ENTITY>FORCE>STRESS
	control	PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY
Concepto del CN	PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY	

Tabla 12: Ejemplo de la asignación de categorías semánticas en *generator torque control*

Cabe señalar que, tras llevar a cabo los siguientes pasos del análisis, se revisó la anotación de categorías semánticas para corregir posibles valores que se determinasen una vez conocida la relación interna, así como para anotar la categoría del concepto designado por el CN, que no se conoció hasta las fases posteriores del análisis.

Tras la asignación de categorías semánticas, analizamos la relación entre los formantes del CN. Para ello seguimos dos pasos: en primer lugar buscamos paráfrasis verbales y, cuando no se obtuvieron resultados precisos, localizamos lo que denominamos *paráfrasis libres*. En la primera fase de búsqueda de paráfrasis verbales, realizamos la siguiente consulta en CQL en la que buscamos los dos elementos o grupos principales del CN y un verbo que los vinculase (Tabla 13):

```
([lemma="power"][]{0,10}[tag="V.*"][]{0,10}[lemma="curve"]within <s/>)|([lemma="curve"][]{0,10}[tag="V.*"][]{0,10}[lemma="power"]within <s/>)
```

Tabla 13: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN de dos formantes

Como vemos en la expresión aplicada al CN *power curve*, buscamos uno de los formantes del CN ([lemma="power"]), seguido de un verbo ([tag="V.*"]) que sería el verbo de la paráfrasis que representase la relación entre los elementos del CN. Este verbo podía

[39] El elemento 1 se concibe como el núcleo del CN, ya se encuentre a la izquierda o a la derecha.

ir precedido o seguido de un espacio de hasta 10 palabras ([]{0,10}), para después finalizar con el otro elemento del CN ([lemma="curve"]). A pesar de que el margen de 10 palabras ofrece ruido en ocasiones, preferimos establecer un límite amplio para no abandonar posibles verbos que se encontrasen en oraciones más extensas, como pueden ser las subordinadas. Además, señalamos que la paráfrasis apareciera dentro de una misma oración (within <s/>), con esta secuencia que identifica las oraciones separadas por puntos, ya que no nos interesaba extraer elementos que perteneciesen a oraciones diferentes. Además, empleamos el signo | para indicar la disyunción e incluimos a continuación la misma expresión, esta vez con los términos en el orden inverso, con el objetivo de encontrar los formantes tanto delante como detrás del verbo en cuestión. Esta expresión nos permitió obtener líneas de concordancias como las siguientes (Figura 19), en las que *power* y *curve* aparecen vinculadas por verbos como *plot* y *represent*, que señalan la proposición conceptual curve *represents* power:

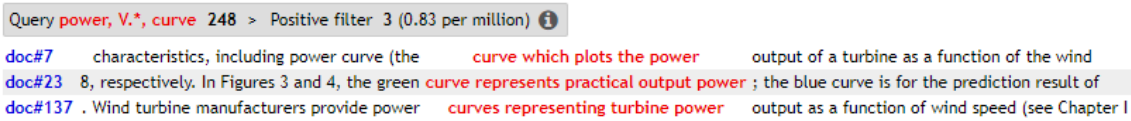

Figura 19: Paráfrasis verbales de *power curve*

En los CN formados por más de dos elementos, los elementos dependientes (conocidos gracias al *bracketing* previo) se presentaron juntos en la consulta, como muestra el siguiente ejemplo de *turbine hub height* (Tabla 14):

([lemma="turbine"][]{0,10}[tag="V.*"][]{0,10}[lemma="hub"][lemma="height"]within <s/>)| ([lemma="hub"][lemma="height"][]{0,10}[tag="V.*"][]{0,10}[lemma="turbine"]within <s/>)

Tabla 14: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN de más de dos formantes

Es preciso señalar que también realizamos búsquedas de paráfrasis verbales en aquellos CN que contenían un adjetivo denominal (es decir, un adjetivo formado a partir de un sustantivo subyacente), p. ej. *electric generator*. En esos casos, así como en los CN que contenían un verbo nominalizado (p. ej. *voltage variation*), buscamos tanto el adjetivo o sustantivo en cuestión (*electric* y *variation*) como la categoría gramatical de la que provienen (*electricity* y *vary*, respectivamente). Por ejemplo, en *electric generator* realizamos la siguiente consulta (Tabla 15):

```
([lemma="electric|electricity"][]{0,10}[tag="V.*"][]{0,10}[lemma="generator"]within <s/>)|
([lemma="generator"][]{0,10}[tag="V.*"][]{0,10}[lemma="electric|electricity"]within <s/>)
```

Tabla 15: Ejemplo de consulta en CQL para obtener paráfrasis verbales en CN con adjetivos denominales

Sin embargo, la extracción de paráfrasis verbales no está exenta de dificultades, pues en ocasiones no es posible identificar verbos que ilustren de manera clara la relación entre los conceptos que forman el CN. Esto se debe a varias razones. Una de ellas es la ausencia de contextos explicativos, ya que se asume que los receptores conocen la información, como indica Bowker (1998) sobre las dimensiones de un concepto y que también puede aplicarse a las paráfrasis. Además, puede haber elementos del CN que se hayan omitido y cuya elisión complique la extracción de paráfrasis verbales, así como adjetivos u otras unidades que se refieran a ese elemento omitido y que, por tanto, dificulten la identificación de verbos que liguen al adjetivo con el resto del CN.

En estos casos recurrimos a las paráfrasis libres, con las que nos referimos a las ocurrencias de los formantes del CN en el corpus. Estas posibilitan el análisis del contexto en el que coocurren estos elementos (tanto entre ellos como a su alrededor) y, así, la extracción de rasgos semánticos con los que dilucidar su significado. Por ejemplo, en *nominal voltage* realizamos la siguiente consulta para obtener paráfrasis libres (Tabla 16):

```
([lemma="nominal"][lemma!="voltage"][]{0,10}[lemma!="nominal"][lemma="voltage"]within
<s/>)|([lemma!="nominal"][lemma="voltage"][]{0,10}[lemma="nominal"][lemma!="voltage"]
within <s/>)
```

Tabla 16: Ejemplo de consulta en CQL para obtener paráfrasis libres en CN de dos formantes

De este modo, buscamos uno de los formantes del CN ([lemma="nominal"]), al que no podía seguir inmediatamente el otro formante para evitar ocurrencias del propio CN ([lemma!="voltage"]). A continuación introdujimos un espacio de hasta diez palabras ([]{0,10}), seguido del otro elemento del CN ([lemma="voltage"]), junto al que de nuevo restringimos la aparición del formante antes incluido ([lemma!="nominal"]). Por último, indicamos que la paráfrasis perteneciese a una misma oración (within <s/>) e introdujimos el elemento de disyunción | seguido de la misma secuencia con los formantes del CN en orden inverso. Mediante esta consulta obtuvimos paráfrasis libres como las siguientes, en las que observamos que se alude a un elemento omitido en el CN (*value*), lo que probablemente complicó la extracción de paráfrasis verbales:

It is essential that the **voltage is kept close to the nominal value**, in the entire power system.

The large-scale centralised power plants **keep the node voltages within the allowed deviation from their nominal value** and the number of dedicated voltage control devices is limited.

The voltage balance control scheme is reactivated and input capacitor **voltages quickly converge to nominal values**.

Cuando el CN estaba formado por más de dos constituyentes, realizamos la misma búsqueda, añadiendo el número de lemas necesario y limitando en cada caso la aparición de los otros formantes, como se aprecia en la consulta de *average wind speed* (Tabla 17):

```
([lemma="average"][lemma!="wind"][lemma!="speed"][]{0,10}[lemma!="average"][lemma=
"wind"][lemma="speed"]within <s/>)|([lemma!="average"][lemma="wind"][lemma="speed"]
[]{0,10}[lemma="average"][lemma!="wind"][lemma!="speed"]within <s/>)
```

Tabla 17: Ejemplo de consulta en CQL para obtener paráfrasis libres en CN de más de dos formantes

En el caso de que no pudiéramos acceder a la semántica del CN por medio de paráfrasis verbales ni paráfrasis libres, consultamos las concordancias del propio CN o utilizamos recursos lingüísticos externos. En definitiva, mediante estos procedimientos buscamos la descodificación de la relación semántica interna, que validamos cuando una determinada relación se da mayoritariamente o, en su defecto, mediante nuestro conocimiento temático adquirido gracias a la documentación. La asignación de relaciones se realizó desde el núcleo hacia el modificador, ya que la direccionalidad de las relaciones normalmente puede alternarse (*has_function* vs. *function_of*). Para ello empleamos el inventario de EcoLexicon, que se compone de las siguientes relaciones (Tabla 18), a las que añadimos la relación *uses_resource*, de elevada presencia en nuestros CN y en distintas áreas del medio ambiente. Igualmente, muchos de los formantes de nuestros CN aludían a propiedades o atributos. A pesar de que estos no constituyen relaciones semánticas, los empleamos en nuestro análisis para dar cuenta de estos vínculos (*attribute_of/ has_attribute*).

<table>
<tr><th>Relaciones genérico-específicas</th><th>Relaciones no jerárquicas</th></tr>
<tr><td>type_of</td><td rowspan="3">affects
causes
attribute_of
opposite_of
studies
measures
represents
result_of
effected_by
has_function</td></tr>
<tr><th>Relaciones de parte-todo</th></tr>
<tr><td>part_of
made_of
delimited_by
located_at
takes_place_in
phase_of</td></tr>
</table>

Tabla 18: Inventario de relaciones semánticas de EcoLexicon

Tras conocer las proposiciones conceptuales que subyacen en los CN, asignamos roles semánticos a los formantes de estos términos compuestos. Tradicionalmente, los roles semánticos se han asignado a los argumentos de un predicado. Por ello, decidimos aplicarlos también a los elementos de los CN, pues estos cuentan con proposiciones subyacentes (p. ej. en *turbine manufacturer*, *turbine* es el PACIENTE y *manufacturer* es el AGENTE). En los CN de más de dos formantes, a diferencia de la asignación de categorías semánticas, no anotamos cada formante con un rol semántico, sino que nos retrajimos a los grupos del *bracketing*. Esto se debe a que el rol de cada formante puede variar si se asigna de forma individual dentro del grupo del *bracketing* o fuera de este, en la estructura básica de núcleo y modificador, ya que se trata de dos proposiciones conceptuales diferentes, una subordinada de la otra. Nos centramos, por tanto, en los roles de los grupos del *bracketing* señalados por la relación principal. Tampoco anotamos el concepto general del CN, ya que no analizamos la relación del CN con predicados externos. Empleamos un inventario de roles semánticos que se está desarrollando en el grupo de investigación LexiCon y que se inspira en los roles propuestos en FrameNet (Fillmore 1977, 1982, 1985), la Gramática del Papel y la Referencia (Van Valin y LaPolla 1997) y VerbNet (Kipper et al. 2004) (Tabla 19). En la Tabla 20, a continuación, se presentan varios ejemplos de la anotación.

Rol semántico
AGENT
THEME
POSSESSOR
PATIENT
INSTRUMENT
LOCATION
DESCRIPTIVE

Tabla 19: Inventario de roles semánticos de EcoLexicon

CN	Grupo 2 del *bracketing*	Grupo 1 del *bracketing*	Relación semántica
AC generator	PATIENT	AGENT	generator *causes* AC
asynchronous generator	DESCRIPTIVE	THEME	generator *has_attribute* asynchronous
wind park	INSTRUMENT	AGENT	park *uses_resource* wind
DC voltage	POSSESSOR	THEME	voltage *attribute_of* DC
permanent magnet generator	THEME	POSSESSOR	generator *has_part* permanent magnet

Tabla 20: Ejemplos de la anotación con roles semánticos

Por último, tras realizar el análisis semántico, agrupamos los CN con sus variantes denominativas. Esta clasificación de variantes en conceptos fue el punto de partida para la búsqueda de términos equivalentes en español, que comentamos detalladamente en la Sección 5.2.1. A grandes rasgos, este procedimiento consistió en la aplicación de un análisis distribucional en ambas lenguas. En primer lugar, se identificaron elementos contextuales representativos de las variantes denominativas en la lengua origen. Seguidamente, la traducción de estos elementos se empleó como consulta en el corpus en español, junto a alguno de los formantes del CN cuya equivalencia se conociese o se pudiera identificar fácilmente.

Como se detallará en la Sección 5.2.1, este método permitió obtener posibles equivalencias de los términos, siendo la coincidencia de rasgos semánticos y elementos contextuales la clave para validar las correspondencias interlingüísticas[40]. En la Tabla 21 se presentan a modo de ejemplo algunos conceptos con sus variantes denominativas en inglés y en español, donde además de este procedimiento, el empleo de patrones de conocimiento sinonímicos permitió la identificación de variantes adicionales. Así, tras localizar las correspondencias en español, obtuvimos 330 términos en esta lengua que se suman a los ya citados en inglés.

Términos en inglés	Términos en español
DC generator dynamo	generador de corriente continua dinamo dínamo generador de continua
horizontal axis wind turbine HAWT	aerogenerador de eje horizontal turbina eólica de eje horizontal
DC-link voltage DC bus voltage	tensión en la etapa de continua tensión de la etapa de continua tensión de bus DC tensión de bus de continua tensión de bus de corriente continua
power density flux density	densidad de potencia densidad de flujo
wind shear	cizalladura del viento cizallamiento del viento cortante del viento

Tabla 21: Ejemplos de conceptos con sus variantes denominativas en inglés y en español

En resumen, el empleo de los materiales y la metodología indicados se hace palpable en el estudio práctico llevado a cabo, en el que nos detuvimos en la formación

[40] Al igual que en los términos en inglés, estos debían figurar en, al menos, tres textos diferentes, para garantizar que no se tratase de usos particulares de un autor o un texto.

(Sección 5.1), la traducción (Sección 5.2) y la representación (Sección 5.3) de estos términos compuestos, como se detallará a continuación.

5. Resultados y discusión

5.1 La formación de compuestos nominales

Entre los objetivos de esta tesis se encuentra el estudio de la formación de los CN en terminología. Para ello, realizamos un análisis semántico (Sección 5.1.1) que nos permitió desarrollar nuestra propuesta de microcontextos (Sección 5.1.2), con la que explicamos de manera sistemática la formación de CN en terminología.

5.1.1 Análisis semántico de la formación de compuestos nominales

Para examinar la formación de los CN, en primer lugar realizamos el *bracketing* o desambiguación estructural (Sección 5.1.1.1) y a continuación analizamos las categorías semánticas que intervienen en la generación de estos términos (Sección 5.1.1.2), la relación mediante la cual se unen (Sección 5.1.1.3) y los roles semánticos que desempeñan los formantes en la proposición conceptual subyacente (Sección 5.1.1.4). Este análisis repercute, además de en la formación, en la traducción y representación de los CN. Por tanto, en este apartado nos centramos en los CN en inglés, para después comparar su formación con la de sus equivalentes en español (Sección 5.2).

5.1.1.1 Bracketing

La desambiguación estructural o *bracketing* fue necesaria en 61 de los 182 CN, esto es, un tercio de los términos. De estos, 48 tenían tres formantes y 13 tenían cuatro formantes. Como comentamos en la Sección 3.3, es habitual que los CN más largos sean menos frecuentes, lo que puede deberse a su mayor dificultad de procesamiento, que suele desembocar en la omisión de formantes y, con ello, en la generación de variantes denominativas (p. ej. *wind power generation system* suele alternarse con variantes más reducidas, como *wind energy system* o *wind power plant*).

A pesar de que contamos con varios modelos de *bracketing* desde hace más de veinte años, este sigue planteando problemas, especialmente en los CN de más de tres formantes, cuya desambiguación estructural no ha recibido atención explícita más allá de trabajos como los de Vadas y Curran (2011) y Barrière y Ménard (2014). No obstante, estos CN más extensos resultan de gran interés, pues a medida que aumenta el número de formantes, se incrementa también su complicación, tanto en relación con su procesamiento como en lo que atañe a su desambiguación.

Como decíamos, los acercamientos al *bracketing* se han realizado principalmente desde la Lingüística Computacional, mientras que áreas como la Terminología o la Traducción no han mirado de cerca este fenómeno, a pesar de que tiene implicaciones de

gran envergadura en sus tareas. La realidad es que el traductor o el terminólogo no recurren frecuentemente a los sistemas de procesamiento del lenguaje natural para analizar con algoritmos la estructura de los CN ante los que se encuentran. Estos profesionales necesitan soluciones manuales que puedan ejecutar ellos mismos en cualquier momento, por lo que realizamos una propuesta de protocolo para el *bracketing* (Cabezas García y León Araúz 2019), principalmente basada en el uso de corpus, que toma premisas de la literatura precedente para enriquecerla con nuevas nociones que faciliten esta labor manual. Esta propuesta es válida para CN de cualquier extensión en inglés y español, en concreto pertenecientes al discurso especializado, ya que los indicativos para el *bracketing* se habían aplicado principalmente a CN del inglés de la lengua general.

Proponemos cuatro criterios, basados en las propiedades lingüísticas de los CN, que en esta tesis aplicamos a nuestros CN de más de dos formantes con el objetivo de determinar su validez en la desambiguación estructural de estos términos compuestos. Para ello, aplicamos cada indicador a todos los CN y señalamos en qué casos fueron determinantes, siguiendo un criterio cualitativo.

Nuestros indicadores son los siguientes:

- Indicador A: La posible agrupación aparece como un término independiente en el corpus.

- Indicador B: La posible agrupación no permite la inserción de elementos externos que modifiquen su significado.

- Indicador C: La posible agrupación forma otros CN.

- Indicador D: La posible agrupación tiene sinónimos o antónimos.

Como hemos mencionado, tomamos un indicador como válido basándonos en razones cualitativas, esto es, cuando los resultados para una de las agrupaciones (o para varias que no fueran contradictorias) eran superiores al resto. A pesar de que los indicadores, ya por separado, presentaron una elevada fiabilidad (por ejemplo, el indicador A fue válido en 59 de los 61 CN), consideramos que es necesario contar con un protocolo formado por varios indicadores, dado que los corpus especializados no son siempre lo suficientemente grandes o representativos y un único criterio puede no resultar determinante. Así, para confirmar una determinada estructura de *bracketing* en nuestro estudio, consideramos que debían darse al menos dos de estos indicadores, ya que

durante la aplicación de los criterios no conocíamos su fiabilidad y no pudimos guiarnos únicamente por el más fiable.

Sin embargo, como veremos, el objetivo era organizar los indicadores según su fiabilidad, de manera que, en el protocolo final, el usuario que necesite desambiguar la estructura de un CN aplique en primer lugar el criterio más fiable y, en caso de no hallar la solución, emplee los siguientes. De este modo, en muchos casos será suficiente solo con uno o varios de los indicadores. A continuación, presentamos en qué consisten y cómo se aplicaron los diferentes indicadores.

A. La posible agrupación aparece como un término independiente en el corpus

Este criterio se basa en la formación de CN largos a partir de la unión de varios CN más reducidos (p. ej. *permanent magnet* + *synchronous generator* > *permanent magnet synchronous generator*). El modelo de adyacencia desarrollado en la Lingüística Computacional, que presentamos en la Sección 3.4.1, parte de esta premisa. Además, se trata del criterio más seguido por los autores[41] para la desambiguación estructural de los CN. Aunque a menudo las distintas combinaciones aparecen en el corpus, consideramos que una frecuencia superior sirve para desambiguar.

Buscamos cada agrupación posible por medio de una consulta CQL que identificara esa posible combinación sin ningún elemento a la izquierda ni a la derecha que fuera susceptible de formar un CN más largo. Por ejemplo, en *turbine hub height* buscamos por una parte *turbine hub* y, por otra, *hub height*, sin elementos que pudieran formar CN a la izquierda ni a la derecha, en concreto sustantivos o adjetivos ([tag!="N.*|JJ.*"])[42]. De este modo, evitamos que entre los resultados de estas consultas apareciera el propio CN de partida (*turbine hub height*). Las consultas fueron las siguientes (Tabla 22):

- [tag!="N.*|JJ.*"][lemma="turbine"][lemma="hub"][tag!="N.*|JJ.*"]
- [tag!="N.*|JJ.*"][lemma="hub"][lemma="height"][tag!="N.*|JJ.*"]

Tabla 22: Consultas CQL para comprobar el indicador A del *bracketing*

En los CN con más de tres formantes, realizamos el mismo tipo de consultas, modificando el número de lemas según las posibles combinaciones, que podían estar

[41] Un ejemplo de ello son las relaciones léxicas en las que se basan Barrière y Ménard (2014).
[42] No restringimos la presencia de adverbios, participios pasados y presentes, ya que estos en general forman parte de la oración y no del término, por lo que comprobamos que producían silencio. En las consultas para extraer los candidatos a términos de la tesis sí los incluimos, a pesar del potencial ruido, con el objetivo de no perder posibles términos.

formadas por dos o más elementos. Por ejemplo, en *wound rotor induction generator*, las combinaciones posibles eran *wound rotor*, *rotor induction*, *induction generator*, *wound rotor induction* y *rotor induction generator*. Por tanto, se realizaron cinco consultas diferentes.

Cuando solo encontramos una de las combinaciones posibles, o varias que no fueran contradictorias (p. ej. [*wound rotor*][*induction generator*]), consideramos que el criterio había sido válido. Por ejemplo, en *direct drive generator*, *direct drive* aparecía 34 veces, mientras que *drive generator* no figuraba en el corpus. Por tanto, la estructura sugerida es [*direct drive*] *generator*. Igualmente, los indicadores también eran válidos cuando una de las estructuras posibles presentaba una frecuencia superior. Es el caso, por ejemplo, de *wind energy system*, donde *wind energy* tenía 1452 resultados, mientras que *energy system* presentaba solo 37. La estructura es, por tanto, [*wind energy*] *system*.

Sin embargo, cuando el corpus no es lo suficientemente grande o representativo, es posible que no se encuentre ninguna de las combinaciones posibles o que se localicen varias agrupaciones contradictorias con diferencias de frecuencia no significativas, por lo que se deben aplicar otros indicadores. Por ejemplo, en *switched reluctance generator*, no obtuvimos resultados de ninguna de las combinaciones posibles (*switched reluctance* y *reluctance generator*). Lo mismo ocurre cuando los CN presentan en su interior un verbo nominalizado, como vemos en *wind power generation system*. Este verbo nominalizado codifica la proposición subyacente de forma explícita: the system *generates* wind power. En estos casos, las diferentes estructuras son frecuentes aproximadamente en la misma medida y se tienen que aplicar otros indicadores para desambiguar el *bracketing*.

B. La posible agrupación no permite la inserción de elementos externos que modifiquen su significado

Este criterio se basa en la unidad léxica (Johnston y Busa 1999) y en la naturaleza conceptual de los CN. Aunque en algunos casos se puedan insertar elementos, estos no modifican el significado de la combinación. Por ejemplo, *large wind turbine* también puede verse en algunos casos como *large wind power turbine*, lo que no modifica la unidad léxica del CN, al tratarse únicamente de una explicitación de significado (*power*, esto es, la energía que se genera)[43]. Así, aplicamos esta propiedad de los CN a sus posibles agrupaciones internas. Por ejemplo, la posible combinación se rompe cuando se introducen elementos que modifican el significado, como vemos en *large offshore wind*

[43] De este modo, podría decirse que *large wind turbine* constituye una variante reducida de *large wind power turbine*, donde se ha omitido uno de los componentes (*power*).

turbine, donde *offshore* indica que *large* y *wind* no forman una unidad léxica en el interior del CN, inclinando así la balanza por la estructura *large* [*wind turbine*].

Para hallar las posibles inserciones de elementos entre nuestros formantes, realizamos las siguientes consultas, que ilustramos con el ejemplo de *wind turbine design*. Como vemos en la Tabla 23, en la primera consulta buscamos elementos entre *wind* y *turbine design*, mientras que en la segunda consulta exploramos los elementos entre *wind turbine* y *design*:

- [lemma="wind"][tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,}[lemma="turbine"][lemma="design"]
- [lemma="wind"][lemma="turbine"][tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,}[lemma="design"]

Tabla 23: Consultas CQL para comprobar el indicador B del *bracketing*

La secuencia [tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,} identifica las posibles inserciones de 1 o más elementos entre los formantes del CN que son susceptibles de intervenir en un CN, tales como sustantivos (N.*), adjetivos (JJ.*), adverbios (RB.*), participios pasados (VVN.*) o participios presentes (VVG.*)[44].

Consideramos que el indicador había sido válido si se daba en solo una de las posibles agrupaciones, en varias que no fueran contradictorias, o en varias contradictorias en las que una agrupación presentase una frecuencia superior. Por ejemplo, en *wind turbine design*, la distinción entre [*wind turbine*] *design* y *wind* [*turbine design*] no planteó problemas, ya que no se encontraron elementos entre *wind* y *turbine*, pero sí entre *turbine* y *design* (en concreto, 19, p. ej. *wind turbine control design*, *wind turbine blade design*, etc.). Por tanto, la estructura sugerida es [*wind turbine*] *design*. Por el contrario, si ninguna de las combinaciones posibles o varias contradictorias con una frecuencia no significativa permiten la inserción de elementos, se debe aplicar otro criterio.

C. La posible agrupación forma otros CN

Este indicador es el opuesto al que acabamos de comentar, pues indica que la agrupación representa un concepto en sí mismo y, por tanto, se puede combinar con otros núcleos y modificadores. Así, la relación con el resto del CN no es necesaria, lo que contribuye a delimitar el *bracketing*. Este indicador se basa en la apertura de *slots* a partir de los cuales se forman los CN (Sección 5.1.2). Esta propiedad señala la naturaleza conceptual del

[44] En este caso, mantuvimos los adverbios, participios pasados y participios presentes en la consulta para no perder CN que pudieran contener estas categorías gramaticales en su interior.

elemento o elementos que dan lugar a la formación de CN, que en este caso sería la agrupación seleccionada.

Realizamos las siguientes consultas, que ejemplificamos con el CN *voltage source converter*. Para la primera combinación posible (*voltage source*) (Tabla 24):

- [tag!="N.*|JJ.*"][lemma="voltage"][lemma="source"][tag="N.*|JJ.*|RB.*|VVG.*|VVN.*" & lemma!="converter"]{0,}[tag="N.*" & lemma!="converter"]
- [tag="N.*|JJ.*"]{1,}[lemma="voltage"][lemma="source"][tag!="N.*|JJ.*"]

Tabla 24: Consultas CQL para comprobar el indicador C del *bracketing* (parte 1)

Para la segunda combinación posible (*source converter*) (Tabla 25):

- [tag!="N.*|JJ.*"][lemma="source"][lemma="converter"][tag="N.*|JJ.*|RB.*|VVG.*| VVN.*"]{0,}[tag="N.*"]
- [tag="N.*|JJ.*" & lemma!="voltage"]{1,}[lemma="source"][lemma="converter"][tag!="N.*|JJ.*"]

Tabla 25: Consultas CQL para comprobar el indicador C del *bracketing* (parte 2)

En la primera consulta de la Tabla 24, buscamos *voltage source* seguido de algún elemento posible (sustantivos, adjetivos, adverbios, participios pasados o participios presentes) y un sustantivo, siempre que estos no fueran *converter* ([tag="N.*|JJ.*|RB.*|VVG.*|VVN.*" & lemma!="converter"]{0,}[tag="N.*" & lemma!= "converter"]), para extraer CN con *voltage source* y otro núcleo. Además, restringimos que a la izquierda no aparecieran sustantivos ni adjetivos, ya que podrían formar parte del CN ([tag!="N.*|JJ.*]). En la segunda consulta de la Tabla 24, buscamos *voltage source* ([lemma="voltage"][lemma="source"]) con otros modificadores nominales o adjetivales ([tag="N.*|JJ.*"]{1,}), sin que le siguieran estas categorías gramaticales ([tag!="N.*|JJ.*"]), ya que nos interesaba ver su capacidad de autonomía, sin tener que recurrir a otros núcleos. Las otras dos consultas (Tabla 25) son las mismas, aplicadas a la otra combinación posible: *source converter*. En los CN de cuatro o más formantes, también se utilizaron las mismas consultas, expandiendo o reduciendo el número de lemas según las combinaciones posibles y limitando la aparición del resto de elementos del CN que no quisiéramos que figurasen en nuestra consulta. Por ejemplo, en *vertical axis wind turbine* una de las agrupaciones posibles es *vertical axis wind*. Por tanto, la primera consulta sería la siguiente (Tabla 26):

```
[tag!="N.*|JJ.*"][lemma="vertical"][lemma="axis"][lemma="wind"][tag="N.*|JJ.*|RB.*|VVG.*|
VVN.*" & lemma!="turbine"]{0,}[tag="N.*" & lemma!="turbine"]
```

Tabla 26: Ejemplo de consulta CQL para comprobar el indicador C del *bracketing* en CN de cuatro o más formantes

Como en los casos anteriores, consideramos el indicador como válido si solo una de las agrupaciones posibles podía combinarse con otros elementos, si se cumplía en varias agrupaciones no contradictorias o si, siendo contradictorias, una de ellas presentaba una frecuencia superior. Por ejemplo, en *voltage source converter*, *voltage source* se combinaba con otros elementos formando 40 CN, tales como *voltage source inverter* o *DC voltage source*. Por el contrario, *source converter* solo formaba otros CN en 4 ocasiones, como en *current source converter*. Así, la estructura sugerida es [*voltage source*] *converter*.

Si ninguna de las agrupaciones o varias combinaciones contradictorias con una frecuencia no significativa pueden formar otros CN, se deben aplicar otros indicadores. Por ejemplo, en *wind power generation system*, todas las agrupaciones posibles pueden formar CN diferentes (Tabla 27). Además, las combinaciones que generan más CN son contradictorias entre sí (a pesar de la diferencia entre frecuencias), por lo que se deben emplear otros indicadores.

Agrupación	***Bracketing* sugerido**	**N.º de CN formados**	**Ejemplos**
wind power	[wind power] generation system	2844	*wind power sector, aggregated wind power*
power generation	wind [power generation] system	298	*power generation mix, renewable power generation*
generation system	wind power [generation system]	66	*generation system capacity, diesel generation system*
wind power generation	[wind power generation] system	73	*wind power generation capacity, renewable wind power generation*
power generation system	wind [power generation system]	5	*micro power generation system*

Tabla 27: CN generados por las distintas agrupaciones posibles en *wind power generation system*

D. La posible agrupación tiene sinónimos o antónimos

Este criterio también se basa en la naturaleza conceptual de la posible combinación, pues los conceptos pueden presentar variantes denominativas (*wind farm*, *wind power plant*), además de antónimos (*offshore wind farm* vs. *onshore wind farm*). Esta premisa está relacionada con la formación de términos mediante la apertura de *slots* (véase el indicador C), pues hay conceptos diferentes que surgen por oposición, como vemos en *offshore wind farm* y *onshore wind farm*, en los que la oposición radica en el *slot* de LOCATION (UBICACIÓN).

Como veíamos en la Sección 3.4.1, Nakov y Hearst (2005) se basan en la variación denominativa para facilitar el *bracketing*. En concreto, se centran en variantes de puntuación, como las que incluyen el genitivo posesivo del inglés, guiones, barras, siglas, etc. Por ejemplo, en *vertical axis wind turbine*, es frecuente encontrar los formantes *vertical axis* unidos por un guion (*vertical-axis*), lo que indica su agrupación. En este sentido, las variantes monoléxicas serían la prueba más fiable de la naturaleza conceptual de la posible agrupación, ya que reducen la combinación a un único lexema. Así, *wind turbine* normalmente se agrupará en el *bracketing*, ya que cuenta con la variante monoléxica *aerogenerator*, como vemos en la siguiente oración: «Wind turbines or aerogenerators transform the kinetic energy of the wind into electrical energy». Por tanto, tuvimos en cuenta las variantes denominativas de puntuación indicadas por Nakov y Hearst (2005), además de hacer búsquedas de sinónimos en CQL en las que utilizamos patrones de conocimiento que suelen expresar la sinonimia en los textos. Las consultas realizadas para obtener sinónimos fueron las presentadas en la Tabla 9 (Sección 4.2.2). Estas se aplicaron a todo el corpus, en lugar de realizar una búsqueda para cada término, ya que partimos de que la sinonimia no suele expresarse de forma explícita en los corpus especializados tanto como otros fenómenos. Algunos de los patrones sinonímicos empleados fueron *known as, called, referred to as* o *synonym of*.

Si solo una de las agrupaciones posibles cumplía con este indicador, o si se cumplía en varias combinaciones que no fueran contradictorias, consideramos válido el criterio. Por ejemplo, en el CN *doubly fed induction generator*, una de las agrupaciones posibles es *induction generator*, que presentaba la variante *asynchronous generator*, como vemos en el siguiente fragmento extraído mediante nuestras consultas: «Asynchronous generators are also called induction generators». Dado que la otra agrupación posible (*doubly fed*) también presentaba una variante (*doubly-fed*) y el resto de posibilidades no presentaron resultados, la estructura sugerida es [*doubly fed*] [*induction generator*].

Si ninguna de las agrupaciones cuenta con sinónimos, o estos se dan en varias agrupaciones contradictorias entre sí, es necesario aplicar otro criterio. Por ejemplo, en *onshore wind farm*, las dos combinaciones posibles, que son contradictorias, presentan sinónimos: *onshore wind* (*sea wind*) y *wind farm* (*wind power plant, wind power station*, etc.).

En cuanto a los antónimos, realizamos la siguiente consulta en CQL (Tabla 28), utilizando los patrones de conocimiento *contrary to* y *opposed to*, que expresan antonimia:

[lemma="be"]{0,}[]{0,3}[word="contrary|opposed"][word="to"][tag="N.*|JJ.*|RB.*|VVG.*| VVN.*"]{0,}[tag="N.*"]

Tabla 28: Consultas CQL para comprobar el indicador D del *bracketing* (antónimos)

Como vemos, buscamos esas palabras ([word="contrary|opposed"]), que podían ir precedidas del verbo *to be* y un margen de hasta tres elementos ([lemma="be"]{0,}[]{0,3}) y que iban seguidas de la preposición *to* y el sustantivo que designaría el antónimo, al que podían preceder modificadores ([word="to"][tag="N.*|JJ.*|RB.*|VVG.*|VVN.*"]{0,} [tag="N.*"]).

De nuevo, si solo una de las agrupaciones posibles cumplía con este indicador, o si se daba en más de una no contradictorias, consideramos válido el criterio. Por ejemplo, en *variable speed generator*, una de las combinaciones posibles (*variable speed*) presenta un antónimo (*fixed speed*), como vemos en el siguiente extracto del corpus: «Contrary to fixed-speed systems, variable-speed systems keep the electromagnetic torque constant». Dado que no encontramos sinónimos ni antónimos de la otra combinación posible (*speed generator*), la estructura sugerida es [*variable speed*] *generator*.

Por el contrario, si ninguna de las combinaciones tiene antónimos o sinónimos, o si estos se dan en varias agrupaciones contradictorias, es necesario aplicar otro criterio. Por ejemplo, en *generator terminal voltage*, ni *generator terminal* ni *terminal voltage* tienen antónimos ni sinónimos.

Este indicador basado en los sinónimos y antónimos ofreció buenos resultados (resultó válido en 53 de los 61 CN) pero, como hemos indicado, es difícil de encontrar en los corpus especializados a través de una consulta explícita. Por tanto, se pueden combinar las búsquedas con la consulta de recursos terminográficos externos. En nuestro caso, hicimos uso de la base de datos Termium Plus, que incluye términos especializados y destaca por recoger ampliamente la variación.

Una vez aplicados todos los indicadores a los 61 CN, obtuvimos los siguientes resultados sobre la validez de cada indicador (Tabla 29):

	N.º de CN desambiguados	N.º de CN no desambiguados
Indicador A	59	2
Indicador B	44	17
Indicador C	55	6
Indicador D	53	8

Tabla 29: Cifras de la desambiguación de CN mediante cada indicador

Como se puede observar, la fiabilidad de los indicadores resultó ser elevada, ya que presentaron desde un 72 % de capacidad de desambiguación (indicador B) hasta un 96 % (indicador A), pasando por un 86 % (indicador D) y un 90 % (indicador C). El indicador A resultó ser el más fiable, al no arrojar resultados válidos en solo 2 de los CN. Además, uno de ellos fue *wind power generation system*, que cuenta con un verbo nominalizado en su interior. Como indicamos anteriormente, esto complica la aclaración del *bracketing*, ya que el verbo es la base de la proposición subyacente y, por tanto, mantiene vínculos con todas las partes del CN, que son sus argumentos.

Por su parte, el indicador B resultó válido en 44 de los 61 CN. La ausencia de elementos insertos en los CN restantes puede justificarse por el elevado grado de fijación que es habitual en muchas combinaciones idiomáticas (p. ej. *switched reluctance generator*). Por otro lado, el indicador C permitió desambiguar 55 CN. Sin embargo, consideramos que esta cifra podría mejorarse si, en los casos conflictivos, nos detuviéramos en el tipo de CN que forman las combinaciones contradictorias, ya que en muchos casos estos son sinónimos o antónimos de otra de las estructuras y, por tanto, descartarían la combinación a partir de la cual se han formado supuestamente, como explicamos a continuación. De adoptar esta postura, este indicador sería válido en otros 4 CN, lo que elevaría su fiabilidad a 59 CN del total, es decir, un 96 %.

Es el caso de *constant-speed wind turbine, horizontal axis wind turbine, vertical axis wind turbine* y *wound rotor induction generator*. Una de las combinaciones posibles en cada CN (*speed wind turbine, axis wind turbine* y *wound rotor induction*) muestra las siguientes combinaciones: *variable* + *speed wind turbine*, *vertical/horizontal* + *axis wind turbine* y *wound rotor induction* + *machine*. Sin embargo, *variable speed* es antónimo de *constant speed*, *vertical axis* es antónimo de *horizontal axis* e *induction machine* es el término genérico de *induction generator*, que a menudo se emplea como sinónimo. Con

esto nos referimos a que, a pesar de que formen nuevos CN, estos son siempre sinónimos o antónimos de la otra combinación posible: *constant speed, horizontal axis, vertical axis* e *induction generator*, por lo que señalan estructuras de doble *bracketing*: [*constant-speed*] [*wind turbine*], [*horizontal axis*] [*wind turbine*], [*vertical axis*] [*wind turbine*] y [*wound rotor*] [*induction generator*]. No obstante, decidimos no contabilizar este tipo de casos entre los CN desambiguados, para no intervenir en los datos obtenidos en ninguno de los indicadores. Por otra parte, en los CN cuyas agrupaciones no han dado lugar a otros CN en el corpus, aludimos de nuevo a su idiomaticidad, al tratarse probablemente de formas concretas que están presentes en ese CN pero que no dan lugar a otros términos compuestos (p. ej. *line-to-neutral voltage*).

Por último, el indicador D resultó válido en 53 de los 61 CN. La ausencia de sinónimos en el resto de los CN puede responder al elevado grado de especialización de las combinaciones posibles (p. ej. *DC-link voltage*), que pueden no haber dado lugar a variación denominativa. En cuanto a los antónimos, se necesitan desarrollar nuevos patrones de conocimiento para su extracción en corpus que permitan obtener mejores resultados, pues no cabe duda de que la oposición subyace en todo sistema conceptual. Del mismo modo, estas cifras se beneficiarían de un mayor enfoque conceptual de los recursos terminológicos, que deberían integrar este tipo de relaciones.

Una vez analizada la fiabilidad de cada indicador, podemos organizar su posición en nuestro protocolo. Así, en primer lugar debería aplicarse el indicador A, es decir, la búsqueda de las posibles agrupaciones como términos independientes en el corpus. En el caso de que este indicador no arrojase resultados significativos, se aplicaría el criterio C, que consiste en la búsqueda de otros CN formados por las posibles agrupaciones. Si tampoco se obtuvieran datos determinantes, pasaríamos al indicador D, esto es, la búsqueda de sinónimos o antónimos de las posibles combinaciones. Por último, de no hallar resultados satisfactorios, se aplicaría el indicador B, mediante el que buscaríamos la inserción de elementos externos que modifiquen el significado de la posible agrupación.

Aun así, es posible, aunque poco probable (se ha dado en un único CN), que ninguno de los indicadores alumbre la estructura del CN. Ha ocurrido así en *wind power generation system*. En términos como este, ya hemos señalado la dificultad que entraña el *bracketing* al contar con una nominalización verbal en su interior. Tras observar las distintas estructuras posibles, nos percatamos de que este CN podría adoptar cualquiera de las tres alternativas: [*wind power*] [*generation system*], [*wind power generation*] *system* o *wind* [*power generation system*]. Es decir, las tres estructuras son correctas, pero

enfatizan perspectivas diferentes de la realidad, que conllevan proposiciones conceptuales diferentes, si bien todas intervienen en el mismo evento: la generación de energía a partir del viento. Así pues, para aclarar la dimensión prevalente e identificar la estructura privilegiada de *bracketing*, decidimos considerar nuestros criterios de forma más flexible: esto es, partiendo de la frecuencia superior y descartando las opciones que se superpusieran con esta.

Por ejemplo, el indicador A muestra las siguientes frecuencias para cada agrupación posible: *wind power* (2726), *power generation* (168), *generation system* (23), *wind power generation* (212) y *power generation system* (7). A pesar de que son agrupaciones que se superponen, si partimos de la que presenta, con diferencia, la mayor frecuencia (*wind power*), podemos aclarar el resto de la estructura, ya que la única combinación restante sería *generation system*, siendo por tanto el *bracketing* [*wind power*] [*generation system*]. De este modo, la proposición conceptual privilegiada sería generation system *uses_resource/causes* wind power, dado el carácter polisémico de *wind power* (energía cinética del viento o energía eléctrica producida a partir de este recurso) que comentaremos más adelante. Además, el indicador C también valida estos resultados, ya que destaca de nuevo la naturaleza conceptual de *wind power* (forma 2844 CN) y también muestra un elevado número de CN (66) formados a partir de la otra combinación complementaria, *generation system*.

Ello no significa que las otras estructuras no puedan darse: system *has_function* wind power generation o power generation system *uses_resource* wind; aunque contamos con que su presencia es menor. En próximos estudios sería interesante analizar si lo que sucede en este término es extrapolable a más CN similares, para determinar la influencia de estas formas nominalizadas en el *bracketing*.

En definitiva, normalmente bastará con aplicar el primer indicador para aclarar la estructura del CN, pues el resto de indicadores suele coincidir. En concreto, en más de la mitad de CN (35) coinciden los cuatro indicadores, lo que demuestra la fiabilidad de los criterios diseñados, mientras que en 17 CN coinciden tres indicadores y en 7 CN coinciden dos indicadores. Es decir, en 59 de los 61 CN coinciden al menos dos indicadores, lo que prueba que estos apuntan a la misma dirección. Solo encontramos cinco excepciones en las que uno de ellos arrojaba datos contradictorios al resto, si bien en cuatro de ellas se resolvió la estructura porque al menos dos de los otros indicadores coincidían en el mismo *bracketing*. Este fue el caso en *offshore wind energy*, *power electronic converter*, *wind power plant* y *wind power production*, lo que puede deberse a que los conceptos que los forman

son claves del dominio e intervienen en distintos CN (p. ej. en un corpus de energía eólica es habitual que se hable tanto de *wind power* como de *power plant*).

Llama nuestra atención el caso de *offshore wind energy*, en el que el carácter polisémico de *wind energy* (energía cinética del viento o energía eléctrica producida a partir de este recurso) desemboca en un *bracketing* variable. Es decir, encontramos ocurrencias tanto de *offshore* [*wind energy*], entendido como la energía eléctrica producida (wind energy *takes_place* offshore), como de [*offshore wind*] *energy*, entendido como la energía cinética que tiene el viento (energy *attribute_of* offshore wind). Por tanto, si por los objetivos del estudio se necesita establecer una estructura privilegiada, consideramos que una solución puede ser, como en el caso de *wind power generation system*, atender a la combinación cuyos indicadores presenten una frecuencia superior. No obstante, no se debe olvidar que, si el término es polisémico, debe valorarse si la otra opción podría ser posible. Un caso diferente es el de *wind power production*, también polisémico, cuyo cambio de significado no viene dado por el establecimiento de una u otra estructura de *bracketing*, sino por el carácter que adquiera *production* en cada contexto (proceso de producción o cantidad resultante de la producción). En esta ocasión, por tanto, no se da un *bracketing* variable.

Además de los cuatro CN mencionados, en los que al menos dos de los otros indicadores apuntaban hacia la misma estructura, *switched reluctance generator* fue el único CN en el que solo dos indicadores resultaron válidos y cada uno mostraba una estructura de *bracketing* diferente. Ante esta situación, observamos que el criterio C indicaba la estructura *switched* [*reluctance generator*]. No obstante, cuando esta agrupación se acompañaba de otro modificador, este siempre era el adjetivo *variable*, que es un sinónimo de *switched*. Por tanto, nos guiamos por el criterio D, que, en esta línea, determina que se puede hablar tanto de *switched reluctance* como de *variable reluctance*, sugiriendo por tanto la estructura [*switched reluctance*] *generator*.

Como vemos, al margen de estas excepciones, el protocolo ofrece soluciones manuales (aunque implementables automáticamente) para desambiguar CN especializados de cualquier extensión en inglés, pues resultó de utilidad para aclarar la estructura de 59 de los 61 CN[45]. En las Tablas 30 y 31 se presentan las estructuras de los CN tras haber aplicado el protocolo.

[45] Los dos CN que dieron lugar a esta cifra (59/61) fueron *wind power generation system* y *switched reluctance generator*.

CN de 3 formantes	
***Bracketing* a la izquierda**	***Bracketing* a la derecha**
[DC-link] voltage	average [wind speed]
[direct drive] generator	commercial [wind turbine]
[generator torque] control	conventional [synchronous generator]
[line-to-line] voltage	conventional [wind turbine]
[line-to-neutral] voltage	generator of [wind turbine]
[permanent magnet] generator	generator [terminal voltage]
[power electronic] converter	high [wind speed]
[power system] reliability	impact of [wind power]
[switched reluctance] generator	large [wind farm]
[three-phase] generator	large [wind turbine]
[variable speed] generator	low [wind speed]
[voltage source] converter	mean [wind speed]
[voltage source] inverter	modern [wind turbine]
[wind energy] system	offshore [wind energy][46]
[wind power] capacity	offshore [wind farm]
[wind power] fluctuation	offshore [wind power]
[wind power] generation	offshore [wind turbine]
[wind power] plant	onshore [wind farm]
[wind power] production	rated [wind speed]
[wind turbine] application	small [wind turbine]
[wind turbine] blade	turbine [hub height]
[wind turbine] design	
[wind turbine] generator	
[wind turbine] manufacturer	
[wind turbine] model	
[wind turbine] rotor	
[wind turbine] system	

Tabla 30: *Bracketing* de CN de 3 formantes

[46] A pesar de que hemos señalado que *offshore wind energy* presenta un *bracketing* variable, indicamos aquí la estructura privilegiada según los indicadores.

CN de 4 formantes		
***Bracketing* a la izquierda**	***Bracketing* a la derecha**	**Doble *bracketing***
		[constant-speed] [wind turbine]
		[doubly fed] [induction generator]
		[fixed-speed] [wind turbine]
		[horizontal axis] [wind turbine]
		[permanent magnet] [synchronous generator]
		[power output] of a [wind turbine]
		[self-excited] [induction generator]
		[squirrel cage] [induction generator]
		[stall-regulated] [wind turbine]
		[variable-speed] [wind turbine]
		[vertical axis] [wind turbine]
		[wind power] [generation system]
		[wound rotor] [induction generator]

Tabla 31: *Bracketing* de CN de 4 formantes

Aunque el conjunto de CN en los que estudiamos el *bracketing* (61) no permite realizar extrapolaciones a gran escala, se observan regularidades. Por un lado, en los CN de 3 formantes (Tabla 30), no se aprecia una preferencia clara por el *bracketing* a la izquierda o a la derecha, ya que las cifras son similares: 27 CN presentan un *bracketing* a la izquierda y 21 CN muestran un *bracketing* a la derecha. No obstante, como se observa en los ejemplos, esta última estructura suele darse cuando hay un adjetivo atributivo (*conventional, small, large, low*) que modifica de manera individual al resto del CN: *small* [*wind turbine*], *low* [*wind speed*], etc. Cabe señalar que los CN *generator of wind turbine* e *impact of wind power* presentan un *bracketing* a la derecha dada su estructura de postmodificación, si bien sus variantes formadas por premodificación (*wind turbine generator* y *wind power impact*) contarán con la estructura inversa. Es decir, cuando hablamos de *bracketing* a la izquierda o a la derecha nos referimos a la forma específica del CN que se esté analizando.

Por otro lado, en los CN de 4 formantes de nuestra muestra (Tabla 31), se prefiere la estructura de dos grupos con dos elementos cada uno, un esquema que se dio en todos los CN (13) que se componían de 4 formantes. En estudios posteriores analizaremos una muestra mayor para determinar si estos resultados pueden extrapolarse al resto de la población.

Además, se observa que hay determinadas combinaciones que habitualmente se agrupan, tales como *wind energy, wind power, wind turbine, wind farm* y *wind speed*. De esto deducimos que *wind*, el término clave del dominio, se suele asociar con el término que le sigue inmediatamente a la derecha, como se aprecia en los ejemplos analizados. Por otra parte, observamos que en inglés las preposiciones suelen aislar los grupos del *bracketing*, no siendo siempre el caso en español.

El *bracketing* de los CN, realizado gracias al protocolo que diseñamos, facilitó el análisis de su contenido semántico y, en concreto, de los microcontextos, además de intervenir en sus posteriores aplicaciones, como su traducción al español, ya que se basa en propiedades que cumplen los CN en las dos lenguas. La mayoría de las veces bastará con llevar a cabo un único paso, aunque si no se obtuvieran resultados satisfactorios, el protocolo proporciona herramientas adicionales para desambiguar la estructura de estos términos compuestos. Se trata, por tanto, de un procedimiento breve, factible y de buenos resultados. Además, podría estudiarse su aplicación a gran escala en los sistemas de procesamiento del lenguaje natural, con vistas a predecir el *bracketing* de los CN de tres o más formantes mediante técnicas de aprendizaje automático.

5.1.1.2 Asignación de categorías semánticas

Tras determinar la estructura de los CN gracias al *bracketing*, los anotamos con categorías semánticas. Estas se aplicaron tanto al concepto general designado por el CN[47] como a sus elementos individuales y a los grupos del *bracketing* (es decir, el concepto designado por la agrupación interna). Así, el empleo de categorías semánticas nos permitió especificar la semántica de los CN y profundizar en los conceptos que manejamos. Además, resultaron de gran utilidad para captar regularidades y hacer generalizaciones, como detallaremos en la Sección 5.1.2.

Habitualmente, las categorías de los conceptos nombrados por los CN coinciden con la del núcleo del CN. Cuando esto sucede, hablamos de CN endocéntricos, en los que el concepto resultante de la combinación consiste en una especificación del significado del núcleo. Por ejemplo, en *power density* tanto el núcleo *density* como el CN en general pertenecen a la categoría ATTRIBUTE>MEASUREMENT>MAGNITUDE, por lo que se trata de un CN endocéntrico que especifica el significado de *density*.

Hay otros casos en los que la categoría del núcleo y la del concepto global designado por el CN no coinciden exactamente, pues la categoría del concepto global

[47] Como hemos indicado en la Sección 4.2.3, la anotación del concepto designado por el CN se realizó una vez que se conoció la relación semántica interna.

modifica ligeramente la del núcleo para concretar el significado. Ello nos remite al diferente grado de granularidad que pueden presentar las categorías, en línea con los niveles de los que hablaba Rosch (1978). Esto se observa, por ejemplo, en los CN formados a partir del núcleo *system*, que adquieren un valor más específico una vez que entran en combinación con los otros elementos. Por ejemplo, en *wind power generation system*, el núcleo pertenece a la categoría ENTITY>SYSTEM, mientras que el concepto resultante de la combinación es mucho más específico: ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT. También sucede en algunas magnitudes que, al combinarse con su valor, adquieren una categoría más específica. Por ejemplo, el núcleo de *high voltage* pertenece a la categoría ATTRIBUTE>MEASUREMENT>MAGNITUDE, mientras que el CN se considera ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL. A pesar de que la categoría del núcleo y del CN no se correspondan exactamente, el significado del CN no deja de ser una especificación del núcleo, por lo que consideramos que se trata de CN endocéntricos.

Sin embargo, también existen CN cuyo significado se aleja de la categoría semántica del núcleo. Nos referimos a los CN exocéntricos, que a pesar de ser menos frecuentes, también pueden encontrarse en el discurso especializado. Los principales ejemplos de este tipo en nuestra muestra son *power output* y *wind penetration*. Por una parte, *output* designa la ubicación desde donde una máquina libera la energía (ATTRIBUTE>LOCATION), si bien su combinación en *power output* constituye una variante denominativa de *output power*, pues ambas aluden a la potencia que proporciona la máquina. Por tanto, el concepto resultante de la combinación se encuadra en la categoría ATTRIBUTE>MEASUREMENT>MAGNITUDE. En cuanto a *wind penetration*, el núcleo *penetration* pertenece a la categoría PROCESS>ACTION. Sin embargo, su combinación con *wind* ha dado lugar a un proceso de metaforización en el que *wind penetration* se entiende en términos de mercado como cantidad de energía producida por el viento en comparación con el total de la energía. Por tanto, el concepto global no se asemeja al del núcleo, sino que pertenece más bien a la categoría ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL.

En relación con los CN exocéntricos, también puede haber elementos internos de los CN que modifiquen su categoría semántica al combinarse con el núcleo. Un ejemplo lo encontramos en *squirrel cage induction generator*, donde *squirrel cage* modifica a *induction generator*. Por separado, *squirrel cage* alude a las cajas que suelen habilitarse para las ardillas domésticas. Por tanto, pertenece a la categoría ENTITY>CREATION>ARTIFACT> CONTAINER. Sin embargo, dentro del CN sufre también un proceso de metaforización por el que pasa a actuar como ATTRIBUTE>PHYSICAL ATTRIBUTE, pues se refiere al aspecto físico del rotor de ese tipo de generador, en comparación con ese habitáculo de las ardillas. Así,

además de anotar estos casos con su valor real dentro del CN, indicamos su categoría propia al margen del CN, para así dar cuenta de estos procesos de modificación del significado que también están presentes en los términos compuestos. En este sentido, conviene identificar correctamente este tipo de fenómenos de cara a la traducción, ya que estos puede que no se reproduzcan en la lengua meta.

Durante el análisis de las categorías semánticas observamos también la presencia de adjetivos generales que forman parte de algunos CN, ya que intervienen en los conceptos que estos designan. Para determinar si realmente contribuían al concepto o se trataba de combinaciones libres, exploramos sus relaciones con otros conceptos del sistema. Así, *apparent power* se consideró un término, ya que normalmente aparece asociado con *active power*. Otro ejemplo lo constituye *conventional generator*, que cuenta con el antónimo *inverter generator*. *Generator side* también contrasta con su antónimo *grid side* y, además, ambos designan partes específicas de los convertidores AC/DC/AC.

En esta línea, llamaron nuestra atención los adjetivos generales que adquieren un valor especializado diferente en los términos compuestos de este dominio. Es el caso de *high* en *high wind*, que podría perfectamente referirse al viento que se produce a una altura elevada, ya que *high* alude a la altura. Sin embargo, en este CN adquiere el valor de intensidad. Sucede lo mismo con adjetivos que indican tamaño, como *small* o *large* (*small wind turbine, large wind farm*). Estos, además de concretar el tamaño del aerogenerador o el parque eólico, se refieren a la potencia que estos pueden generar, algo que va ligado a sus dimensiones. De hecho, cuando interviene *size*, se suele aludir a la potencia, como se aprecia en el siguiente fragmento: «Onshore, wind turbines are typically in the 2 MW to 3 MW size range». Otro ejemplo lo encontramos en *instantaneous voltage*, donde el adjetivo *instantaneous* adquiere un significado específico en el ámbito de la electricidad para aludir a algo que se produce en un momento determinado (p. ej. *instantaneous voltage* es la tensión de un momento dado). Por tanto, es importante identificar estos adjetivos[48] generales como parte del término compuesto, ya que intervienen en el concepto. Se observa, así, que la identificación de los CN no constituye una tarea trivial.

Por otra parte, al explorar cómo se combinan las categorías semánticas en el interior de los CN, descubrimos patrones recurrentes en la formación de CN, que están directamente relacionados con los microcontextos que desarrollaremos en la

[48] Los adjetivos no constituyen siempre atributos. Así, cuando constatamos que se vinculaban con el núcleo del CN mediante una relación y no mediante un atributo, les asignamos la categoría del sustantivo subyacente. Por ejemplo, en *electric generator*, el adjetivo *electric* se anotó como PROCESS>MOVEMENT>ENERGY MOVEMENT. Este adjetivo, sin embargo, también puede actuar como atributo en CN como *electric power*, en los que se anotó como ATTRIBUTE.

Sección 5.1.2. En la Tabla 32 se presentan las combinaciones de categorías más habituales que dieron lugar a los CN de nuestro estudio.

Combinación de categorías		N.º de CN	Ejemplo
ATTRIBUTE	ENTITY>CREATION	32	*asynchronous generator*
ATTRIBUTE	ATTRIBUTE>MEASUREMENT>MAGNITUDE	27	*induced voltage*
PROCESS>MOVEMENT	ENTITY>CREATION	18	*electric generator*
ENTITY>CREATION>ARTIFACT	ATTRIBUTE	16	*grid voltage*
PROCESS>MOVEMENT	ATTRIBUTE	14	*power quality*
PROCESS>MOVEMENT	PROCESS>FORMATION	11	*wind power generation*
ATTRIBUTE	PROCESS>MOVEMENT	8	*mechanical power*
ENTITY>MATTER	ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	8	*permanent magnet generator*
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	7	*generator rotor*
ATTRIBUTE>MEASUREMENT>MAGNITUDE	PROCESS>CHANGE	7	*voltage control*
PROCESS>MOVEMENT	PROCESS>MOVEMENT>ENERGY MOVEMENT	6	*wave power*
PROCESS>MOVEMENT	ENTITY>SYSTEM	5	*power system*
ENTITY>PART>PART OF INSTRUMENT	ATTRIBUTE>MEASUREMENT>MAGNITUDE	4	*DC-link voltage*
ATTRIBUTE>MEASUREMENT>MAGNITUDE	ENTITY>INFORMATION>REPRESENTATION	4	*power spectrum*
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	ENTITY>PART	3	*turbine blade*
ENTITY>FORCE	ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	3	*induction generator*
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	PROCESS>ACTION	3	*turbine operation*
ENTITY>MATTER>SOLID MATTER>MATERIAL	ATTRIBUTE>MEASUREMENT>MAGNITUDE	2	*phase voltage*
ENTITY>SPACE>POSITION	ATTRIBUTE>MEASUREMENT>MAGNITUDE	2	*PCC voltage*
PROCESS>MOVEMENT>ENERGY MOVEMENT	PROCESS>CHANGE	2	*wind power fluctuation*
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	ENTITY>SYSTEM	2	*wind turbine system*
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	ENTITY>HUMAN	2	*turbine manufacturer*

Tabla 32: Combinaciones de categorías más habituales en los CN de la muestra

Como señalan diferentes autores (Warren 1978; Rosario et al. 2002; Maguire et al. 2010b; Smith et al. 2014; Cabezas García y León Araúz 2018), las

combinaciones similares de categorías semánticas permiten inferir el significado de nuevos CN, ya que se recurre a la comparación con las combinaciones conceptuales conocidas. En la Tabla 32 se aprecian las categorías que suelen coocurrir para dar lugar a los CN de nuestro estudio. Estas representan conceptos clave de la energía eólica y constituyen una muestra de la productividad de determinados conceptos para nombrar otros nuevos en este campo. Se recogen las combinaciones que generaron, al menos, dos CN, para centrarnos así en las más productivas. En este sentido, utilizamos diferentes niveles de jerarquización para captar regularidades.

Se observa, por tanto, que una de las preferencias de combinación es aquella por la que una entidad de creación (ENTITY>CREATION) se acompaña de un atributo (ATTRIBUTE), que suele utilizarse para especificar características de dicha entidad (*asynchronous generator*). Asimismo, las magnitudes (ATTRIBUTE>MEASUREMENT>MAGNITUDE) suelen asociarse con otro atributo (ATTRIBUTE), por ejemplo para indicar sus características (*induced voltage*). Las entidades de creación (ENTITY>CREATION) también suelen coocurrir con procesos de movimiento (PROCESS>MOVEMENT), que muchas veces utilizan para llevar a cabo su función o, al contrario, son el resultado de su acción (*electric generator*). Otra de las combinaciones más frecuentes es aquella en la que un atributo (ATTRIBUTE) se concreta con entidades de creación como los artefactos (ENTITY>CREATION>ARTIFACT) (*grid voltage*) así como con procesos de movimiento (PROCESS>MOVEMENT) (*power quality*), para indicar por ejemplo el concepto del que son atributos. Del mismo modo, los procesos de formación (PROCESS>FORMATION) coocurren a menudo con procesos de movimiento (PROCESS> MOVEMENT), por ejemplo para señalar la generación de energía (*wind power generation*). Como estos, otros pares de categorías que figuran en la Tabla 32 son recurrentes en nuestro corpus y demuestran la fertilidad de determinadas combinaciones.

En los CN de tres o más formantes, convergen varias de estas combinaciones. Gracias al *bracketing*, se determinan los vínculos conceptuales más fuertes que permiten reducir la combinación a la estructura básica de núcleo y modificador. Por ejemplo, en *wound rotor induction generator*, las categorías de cada elemento son las siguientes:

-*wound*: ATTRIBUTE>PHYSICAL ATTRIBUTE>COMPOSITION

-*rotor*: ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT

-*induction*: ENTITY>FORCE

-*generator*: ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT

Al comprobar mediante nuestro protocolo de *bracketing* que las asociaciones internas eran [*wound rotor*] [*induction generator*], establecimos que las dos categorías principales que dieron lugar al concepto resultante eran ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT y ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, coincidiendo así con una de las combinaciones recurrentes presentadas en la Tabla 32. Por tanto, se trata de una estructura de doble *bracketing* en la que la categoría nuclear de cada agrupación se concreta mediante la adición del otro elemento. Sin embargo, para profundizar en el significado de los conceptos resultantes de estas combinaciones, es necesario conocer la relación entre estos elementos (en este caso, meronímica), un aspecto en el que ahondaremos en el siguiente apartado.

De este modo, una vez que conocimos el significado global de la combinación conceptual, exploramos las categorías que suelen designar los CN de energía eólica en nuestra muestra gracias a esta anotación semántica. Como se infiere en la Tabla 33, hay una presencia mayoritaria de ENTIDADES y ATRIBUTOS en nuestra muestra (64 y 57 conceptos, respectivamente del total de 147, mientras que contamos con 26 conceptos que representaban PROCESOS). Entre las ENTIDADES destacan los instrumentos transformadores (ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT), que constituyen el grueso de la categoría con 47 conceptos. Por su parte, las magnitudes (ATTRIBUTE>MEASUREMENT>MAGNITUDE) son los principales ATRIBUTOS de nuestra muestra, expresadas por 33 conceptos. Por último, los PROCESOS más habituales fueron los que representaban un movimiento de energía (PROCESS>MOVEMENT>ENERGY MOVEMENT) y un cambio de tamaño o intensidad (PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY), con 8 y 7 conceptos respectivamente.

ENTITY			
Categorías[49]	**N.º de CN[50]**	**Ejemplos**	**N.º de conceptos**
ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	62	power converter	47
ENTITY>CREATION>ARTIFACT>CONDUIT	2	power grid	1
ENTITY>CREATION>ARTIFACT>STRUCTURE	9	power station	5
ENTITY>FORCE>STRESS	1	generator torque	1
ENTITY>HUMAN	2	wind industry	2
ENTITY>HUMAN>INSTITUTION	1	turbine manufacturer	1
ENTITY>INFORMATION>REPRESENTATION>MATHEMATICAL EXPRESSION	1	power coefficient	1

[49] Las subcategorías no se comprendieron en las categorías superiores a efectos del recuento (p. ej. las subcategorías de ATTRIBUTE>MEASUREMENT>MAGNITUDE que indican LEVEL y MEAN se contaron por separado y no se incluyeron en ATTRIBUTE>MEASUREMENT>MAGNITUDE).
[50] En el recuento de CN no se incluyeron las variantes denominativas identificadas tras la extracción inicial de CN, sino únicamente los CN seleccionados en el proceso de extracción (estos se duplicaron cuando representaban varios conceptos, para dar cuenta de las diferentes categorías posibles).

ENTITY>INFORMATION>REPRESENTATION>LINE	1	power curve	1
ENTITY>INFORMATION>REPRESENTATION>PARAMETER	1	power factor	1
ENTITY>INFORMATION>REPRESENTATION>GRAPH	1	power spectrum	1
ENTITY>PART>PART OF INSTRUMENT	3	turbine blade	3
TOTAL DE TÉRMINOS: 84 **TOTAL DE CONCEPTOS: 64**			
ATTRIBUTE			
ATTRIBUTE	2	power quality	2
ATTRIBUTE>DIRECTION	1	wind direction	1
ATTRIBUTE>MEASUREMENT>MAGNITUDE	43	wind power capacity	33
ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL	16	rated wind speed	14
ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL>MEAN	2	mean wind speed	1
ATTRIBUTE>PHYSICAL ATTRIBUTE>COMPOSITION	5	turbine design	3
ATTRIBUTE>PHYSICAL ATTRIBUTE>SIZE	1	turbine size	1
ATTRIBUTE>PHYSICAL ATTRIBUTE>STATE	2	voltage stability	2
TOTAL DE TÉRMINOS: 72 **TOTAL DE CONCEPTOS: 57**			
PROCESS			
PROCESS>ACTION	3	turbine operation	3
PROCESS>ACTIVITY	1	wind project	1
PROCESS>CHANGE	1	impact of wind power	1
PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY	9	voltage sag	7
PROCESS>FORMATION	5	power generation	2
PROCESS>MOVEMENT>ENERGY MOVEMENT	15	electric power	8
PROCESS>MOVEMENT>WIND MOVEMENT	4	wind shear	4
TOTAL DE TÉRMINOS: 38 **TOTAL DE CONCEPTOS: 26**			
194 TÉRMINOS EN TOTAL **147 CONCEPTOS EN TOTAL**			

Tabla 33: Categorías de los conceptos designados por los CN de la muestra

En estos datos se aprecia también la presencia de variación denominativa en nuestra muestra. Así pues, los 147 conceptos se designan con 194 términos. En concreto, las entidades estaban representadas por 64 conceptos y 84 términos; los atributos, por 57 conceptos y 72 términos; y los procesos, por 26 conceptos y 38 términos. Por tanto, la variación denominativa, ya se considere un signo de inestabilidad o un mecanismo cognitivo, está presente en los términos compuestos. Abordaremos este fenómeno con más detalle en la Sección 5.2.2, ya que nos interesa comparar su presencia en inglés y en los términos correspondientes en español.

La variación conceptual también formó parte de nuestros términos. Así, se identificaron diferentes términos polisémicos que, en muchos casos, requirieron la anotación con varias categorías semánticas, a excepción de aquellos casos en los que los dos conceptos pertenecían a la misma categoría. Los siguientes términos de nuestra muestra designaban más de un concepto:

-*Wind power*: puede aludir a la potencia como propiedad del viento (ATTRIBUTE>MEASUREMENT>MAGNITUDE) o a la energía eléctrica resultante de su utilización (PROCESS>MOVEMENT>ENERGY MOVEMENT). Se trata de una de las dicotomías más habituales en la variación conceptual de este dominio, originada por la polisemia de *power*, que puede referirse a la energía cinética de otro concepto o a la electricidad. El carácter polisémico del núcleo hace que codifique una relación diferente según el concepto que designe: power *attribute_of* wind (ATTRIBUTE>MEASUREMENT>MAGNITUDE) o power *effected_by* wind (PROCESS>MOVEMENT>ENERGY MOVEMENT).

-*Wind energy*: también puede designar la potencia como propiedad del viento (ATTRIBUTE>MEASUREMENT>MAGNITUDE) o a la electricidad que se obtiene de su utilización (PROCESS>MOVEMENT>ENERGY MOVEMENT). Como en el caso de *power*, la polisemia viene dada por *energy*, que se suele emplear como variante denominativa de *power*. De nuevo, esta variación conceptual del núcleo conlleva un cambio en la relación interna del CN según el concepto al que aluda: energy *attribute_of* wind (ATTRIBUTE>MEASUREMENT> MAGNITUDE) o energy *effected_by* wind (PROCESS>MOVEMENT>ENERGY MOVEMENT).

-*Wave power*: se trata de la misma dualidad de los casos anteriores, esta vez aplicada a las olas. Por tanto, puede pertenecer a las siguientes categorías y codificar las siguientes relaciones: ATTRIBUTE>MEASUREMENT>MAGNITUDE (power *attribute_of* wave) o PROCESS>MOVEMENT>ENERGY MOVEMENT (power *effected_by* wave).

-*Offshore wind power*: se da el mismo caso de polisemia, cuyos conceptos pueden aludir a las categorías ATTRIBUTE>MEASUREMENT>MAGNITUDE o PROCESS>MOVEMENT>ENERGY MOVEMENT y codificar, respectivamente, las relaciones power *attribute_of* offshore wind y wind power *takes_place* offshore.

-*Offshore wind energy*: se trata de una variante del término anterior, cuyos conceptos también pertenecen a las categorías ATTRIBUTE>MEASUREMENT>MAGNITUDE o PROCESS>MOVEMENT>ENERGY MOVEMENT y activan las relaciones energy *attribute_of* offshore wind y wind energy *takes_place* offshore. Como indicamos en la Sección 5.1.1.1, puede darse un *bracketing* variable ligado al cambio de relación (*offshore* [*wind energy*] u [*offshore wind*] *energy*).

-*Power generation*: puede aludir al proceso de generación de energía (PROCESS>FORMATION) o a la energía resultante (PROCESS>MOVEMENT>ENERGY MOVEMENT). Constituye otra de las dualidades más frecuentes en la variación conceptual de este dominio, que surge debido a la doble visión de *generation* como proceso y su resultado. A

pesar de que este énfasis en una u otra fase del proceso da lugar a variación conceptual en el CN, su relación interna no se altera (generation *has_result* power)[51].

-*Power production*: *production* se utiliza como variante de *generation*, por lo que este CN también expresa el proceso de producción de energía (PROCESS>FORMATION) y la energía resultante (PROCESS>MOVEMENT>ENERGY MOVEMENT), sin implicar un cambio en la relación interna (production *has_result* power).

-*Wind power generation*: se da el mismo caso de polisemia, cuyos conceptos pueden referirse al proceso de generación de energía eólica (PROCESS>FORMATION) y a la energía resultante (PROCESS>MOVEMENT>ENERGY MOVEMENT), codificando en ambos casos la relación *has_result* (generation *has_result* wind power).

-*Wind power production*: este CN, variante denominativa del anterior, alude a las mismas categorías: PROCESS>FORMATION (proceso de producción de energía eólica) y PROCESS>MOVEMENT>ENERGY MOVEMENT (energía resultante); y también codifica la relación *has_result*.

-*Wind generation*: este CN constituye otra variante denominativa de *wind power generation* y *wind power production*, producida por elisión léxica. Igual que en los casos anteriores, puede expresar el proceso de generación de energía eólica (PROCESS>FORMATION) y la energía resultante (PROCESS>MOVEMENT>ENERGY MOVEMENT). La relación interna en este caso no coincide con las de sus variantes debido a la omisión de *power*. Dado que se destaca el papel del viento en la generación, la proposición es generation *uses_resource* wind. Se observa aquí que las variantes denominativas pueden presentar una composición semántica diferente (p. ej. pueden formarse por categorías semánticas distintas, con relaciones diferentes, enfatizando otros aspectos, etc.).

-*Power system*: puede designar un generador (ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT) o una red eléctrica (ENTITY>CREATION>ARTIFACT> CONDUIT). En este caso, la polisemia del CN no viene originada por la polisemia del núcleo (como sucedía en CN como *wind power*), ya que este mantiene su significado. Al contrario, se produce un cambio en la relación interna del CN, que da lugar a la polisemia. Así, system *causes* power en el sentido de generador, mientras que system *affects* power en el sentido de red. Por tanto, dos conceptos pueden relacionarse mediante proposiciones conceptuales diferentes y, de este modo, aglutinarse en CN con distintos significados.

[51] Aunque la variación conceptual de *power* podría dar lugar a la relación *uses_resource* cuando *power* se entiende como propiedad del viento, nuestro análisis de las relaciones (Sección 5.1.1.3) no reveló casos de esa codificación.

-*Wind generator*: puede aludir a un generador o a la entidad superior de la que forma parte, esto es, un aerogenerador. En ambos casos, la categoría semántica es ENTITY> CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT. Por tanto, la polisemia surge de la metonimia parte-todo y no se da un cambio de relación interna (generator *uses_resource* wind).

-*Wind turbine generator*: este CN también puede referirse a un generador o al aerogenerador del que forma parte. En los dos casos, pertenecen a la categoría ENTITY> CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT. Como en *power system*, las categorías de los formantes no cambian, lo que varía en este caso es la relación interna (generator *part_of* wind turbine en el caso de generador, mientras que generator *has_type* wind turbine en el caso de aerogenerador).

-*Wind system*: puede referirse a un aerogenerador y a un sistema de vientos de la circulación atmosférica. De nuevo, la polisemia no viene originada por la variación de alguno de sus formantes, sino por el cambio de relación interna: system *uses_resource* wind (aerogenerador) y system *has_type* wind (sistema de vientos).

En definitiva, la variación conceptual de los conceptos que hemos analizado puede deberse a tres motivos. Uno de ellos es el énfasis en una u otra fase de un proceso (p. ej. *power generation*). En estos casos se produce un cambio en el concepto designado por el CN pero no en la relación interna. El segundo de los motivos puede ser el cambio en la relación interna (p. ej. *power system*). Los elementos del CN no varían su categoría semántica, sino que establecen vínculos diferentes, que originan la variación en el concepto expresado por el CN. Por último, la variación conceptual de alguno de los formantes puede entrañar el cambio en la relación interna (p. ej. *wind power*) y, con ello, la variación del concepto designado por el CN. Así, el contexto se presenta como un factor fundamental para dilucidar el significado del CN, pues la simple observación de los formantes no permitirá concretar su significado. Como comentaremos en la Sección 5.1.1.3, las paráfrasis constituyen un enfoque contextual que puede resultar muy útil para este fin.

El carácter polisémico de los CN puede tener implicaciones de diferente tipo. Por ejemplo, la codificación de diferentes relaciones internas puede suponer el cambio de los roles semánticos que desempeñan los formantes. Por otra parte, la variación conceptual también influye en la combinatoria de estos términos. Por ejemplo, *power* genera un tipo de hipónimos cuando se considera PROCESS>MOVEMENT>ENERGY MOVEMENT y otro diferente cuando se entiende como ATTRIBUTE>MEASUREMENT>MAGNITUDE, ya que cada concepto

nuclear se puede especificar de modos distintos. En el primer caso, *power* como proceso da lugar a CN como *electric power* o *renewable power*, mientras que *power* como atributo genera CN como *active power* o *maximum power*.

Así como la polisemia influye en la combinación de los conceptos para formar CN, también puede determinar su combinatoria a un nivel superior: el de la oración. Por tanto, los términos polisémicos pueden coocurrir con verbos o complementos de diferente significado. Por ejemplo, *power* en el sentido de atributo se combina con verbos como *extract* o *absorb*, mientras que *power* entendido como energía eléctrica acompaña a predicados como *store* o *consume*. Estos elementos oracionales, al compactarse, pueden dar lugar a CN (*power extraction* o *power absorption*, frente a *power storage* o *power consumption*). Además, la coincidencia de semas en una misma forma gráfica no tiene por qué reproducirse en otras lenguas. Como veremos en la Sección 5.2.2, *power* se suele traducir por *energía* o *electricidad* cuando significa PROCESS>MOVEMENT>ENERGY MOVEMENT, mientras que su equivalente habitual cuando alude a ATTRIBUTE>MEASUREMENT>MAGNITUDE es *potencia*.

En definitiva, hemos observado cómo, gracias a estas categorías, conseguimos detallar la semántica de los CN. Esto nos permitió analizar su formación y su traducción, y facilitar la inferencia de la relación semántica, además de representar los CN en el módulo fraseológico de EcoLexicon, en el que se puede acceder a ellos mediante búsquedas específicas basadas en las categorías.

5.1.1.3 Descodificación de la relación semántica interna

Aunque la semántica de los formantes de los CN es muy importante, si no se conoce el modo en que estos elementos se relacionan, falta la pieza clave del sistema. Por tanto, uno de los aspectos principales del análisis de los CN es la descodificación de su relación semántica interna. Esta relación también desempeña un papel fundamental en nuestra propuesta de los microcontextos (Sección 5.1.2), ya que es la encargada de vincular los *slots* con el concepto de partida.

En primer lugar, analizamos la relación semántica de cada CN por medio de paráfrasis verbales, que revelan la proposición subyacente y, por tanto, muestran la relación interna mediante verbos específicos. Este tipo de paráfrasis permitió detectar la

relación presente en 113 de los 190 CN[52] analizados. En la Tabla 34 se indican las relaciones obtenidas, acompañadas de ejemplos que las ilustran.

Relación[53]	Ejemplo de CN y paráfrasis verbales
affects	generator torque control (control *affects* generator torque)
	The rotor speed can be **controlled** both **acting on** the aerodynamic torque as well as on the **generator torque**. **Control** systems may dynamically **adjust** blade pitch settings and **generator torque**.
attribute_of	wind energy (energy *attribute_of* wind)
	It will be shown in Chapter 3 that the maximum power-producing potential that can be theoretically realized from the kinetic **energy contained in** the **wind** is about 60% of the available power. Naturally, wind energy developers are very interested in the **energy** that can be **extracted from** the **wind**, and how this varies by location.
has_attribute	fixed-speed wind turbine (wind turbine *has_attribute* fixed-speed)
	Wind turbines can **operate** either with a **fixed speed** or a variable speed. At the beginning of the 90s, the installed **wind turbines ran** mainly **at fixed speed**.
causes	AC generator (generator *causes* AC)
	Even if the **generator supplies alternating current (AC)** of variable frequency or direct current (DC). A **generator** that **produces AC** is called an AC generator or, more commonly, an alternator.
caused_by	generator power (power *caused_by* generator)
	They rely on back-up **generators** to **supply power** when the wind is not blowing. The wind **generators produce power** according to the wind.
effected_by	generator operation (operation *effected_by* generator)
	The induction **generator works** at the rated **operation** point. The chosen **generators** that **are in operation** have to maintain the difference between nominal and actual engaged power. Greater values of VI% are obtained when each of the DFIG, PMSG, and synchronous **generators is operating** under voltage control condition.
has_function	commercial wind turbine (wind turbine *has_function* commerce [commercial])
	The lift force **wind turbines** are most widely **used as commercial wind turbines** due to their higher power coefficient than that of drag force wind turbine. What changes will there be to **wind turbines** that **are in commercial use** (for example size, materials, technical changes, capacity factor)?
part_of	wind turbine generator (generator *part_of* wind turbine)
	A **wind turbine** can be **equipped with** any type of three-phase **generator**. A **wind turbine is composed of** an electric **generator** mounted on a suitable

[52] Esta cifra es superior a la cantidad inicial de CN extraídos, al haberse duplicado los CN que presentan varias relaciones semánticas.
[53] Expresamos la relación semántica en la dirección núcleo>modificador.

	tower (...).
has_part	permanent magnet synchronous generator (synchronous generator *has_part* permanent magnet)
	The exception is Enercon, which has long supplied direct drive generators employing a **synchronous generator** and **having** an electrical rotor with windings rather than **permanent magnets**. **Synchronous generators excited by permanent magnets**.
has_result	wind power production (production *has_result* wind power)
	It is the ratio of how much **wind power is being produced** at any specific instant. If the new EWEA targets are met, enough **wind power would be produced** to power 291 million cars in 2020 and 577 million cars in 2030.
has_type	wind resource (resource *has_type* wind)
	These conditions include power plant outages as well as predictable or uncertain variations in demand and in primary generation **resources**, **including wind**. Types of renewable energy **resources include wind**, solar, biomass, geothermal, and moving water.
located_at	offshore wind turbine (wind turbine *located* offshore)
	It has long been acknowledged that some of the design drivers for a **wind turbine installed offshore** are fundamentally different from those installed onshore. Building of foundation for **wind turbines situated offshore** requires a mix of experiences gained from on-shore wind turbines and from other areas of offshore construction (harbours, piers, oil & gas exploitation etc.).
represents	power curve (curve *represents* power)
	Power curve (the **curve** which **plots** the **power** output of a turbine as a function of the wind speed). The green **curve represents** practical output **power**.
takes_place_in	offshore wind power (wind power *takes_place* offshore)
	Wind power, especially that **located** in remote or **offshore** areas, is also very suitable to apply FFTS for long distance transmission.
uses_resource	gas turbine (turbine *uses_resource* gas)
	Some peak-load stations are gas **turbines**, usually **fuelled by gas** or (rarely) by oil. Even if one is prepared to overlook the need for the **turbines** to **run on** natural **gas**, no commercial solution has yet been demonstrated for running the generators efficiently on compressed air alone. Both **gas–fired** power plants and wind **turbines** are used to produce peak power.

Tabla 34: Relaciones identificadas mediante paráfrasis verbales, ilustradas con ejemplos

Como puede observarse, las paráfrasis verbales permitieron identificar la relación de una gran cantidad de CN, que contaban con relaciones de diferente tipo. Estas permitieron descodificar claramente relaciones como *causes* o *uses_resource*, con verbos como *supply* o *produce* en el primer caso y *fuel, run* o *fire* en el segundo. Sin embargo, a pesar de que las paráfrasis permiten descodificar muchos CN cuya relación es *attribute/has_attribute*, a menudo los verbos extraídos presentan un carácter genérico y la

relación *attribute/has_attribute* se infiere a partir del contexto. Un ejemplo lo constituye el verbo *have*, que puede expresar distintas relaciones, como *part_of/has_part* o *attribute/has_attribute*, en función de sus argumentos.

De ello se desprenden las dos principales utilidades de las paráfrasis verbales. Por una parte, los verbos obtenidos pueden emplearse para perfilar los inventarios de relaciones semánticas, en los que relaciones como *affects* podrían matizarse según los significados más específicos que pueden englobarse en estas relaciones. Por ejemplo, en nuestros CN, la relación *affects* se plasma con predicados como *control* o *adjust*, *fluctuate* o *vary*, o *transmit* y *transport*, que transmiten la relación *affects* de modos diferentes. Así, los verbos que expresan una misma relación podrían organizarse en los dominios léxicos de Faber y Mairal (1999) en función de su significado de base, adquiriendo así un mayor poder representativo de las relaciones. Como siempre que se busca una mayor granularidad, el desafío consiste en encontrar el punto medio entre esta granularidad y la cobertura de elementos.

Otra de las utilidades de las paráfrasis verbales es la obtención de información semántica y contextual adicional. Por ejemplo, las paráfrasis verbales pueden revelar otros argumentos de la proposición subyacente en el CN. Es el caso de la siguiente paráfrasis de *wind penetration*, donde se descubre dónde *penetra* o se integra la energía eólica: «Wind has already made solid steps forward, penetrating national transmission systems by as much as 10 per cent in several markets and as much as 21 per cent in Denmark». Estos nuevos argumentos que salen a la luz favorecen la comprensión del concepto, así como la identificación de nuevos CN que pueden generarse (p. ej. *system penetration*). Las paráfrasis también desvelan elementos que se han omitido en el CN. Por ejemplo, en varias paráfrasis de *floating turbine* se indica que la turbina *is installed on a floating foundation/is set on a floating column/is installed on a floating platform*. Por tanto, se aprecia que la turbina flota porque está ubicada sobre elementos de soporte como bases, columnas o plataformas, que no se materializan en el CN. Por otra parte, las paráfrasis también arrojan información sobre variantes, como sinónimos: «A generator that produces AC is called an AC generator or, more commonly, an alternator».

Sin embargo, a pesar de sus beneficios, las paráfrasis verbales también presentan complicaciones. A veces, pueden encontrarse paráfrasis verbales de un mismo CN que no indiquen la misma relación. Esto puede deberse a diferentes razones. Una de ellas es el carácter polisémico de algunos términos, que puede originar un cambio de relación en el CN, como hemos comentado en el apartado anterior. Por ejemplo, en *power system* se dan

dos tipos de paráfrasis, que señalan los dos significados que posee este CN. Como se indica en la Tabla 35, un sistema puede producir energía si se entiende como generador (system *causes* power) o puede afectar a la energía, cuando su sentido es el de red eléctrica (system *affects* power). Evidentemente, estas dos relaciones se materializan mediante paráfrasis verbales diferentes.

CN y concepto designado	Relación y paráfrasis
power system (GENERATOR)	*causes* (system *causes* power)
	The wind turbine **system** in a large-scale power system **generates** its maximum **power** according to wind speeds. A determining factor in hybrid power system design is the interdependence of the system electrical load and the **power produced by** the hybrid **system**. The device forms a two degree of freedom mechanical **system producing power** by electromagnetic induction.
power system (ELECTRICAL GRID)	*affects* (system *affects* power)
	The main function of an electrical power **system** is to **transport** electrical **power** from the generators to the loads. It is difficult to **transmit** reactive **power** long distances **in a power system** and so local reactive sources are required to support the network voltage. The difference in voltage is caused by the distribution system impedance, which consists of the distribution system resistance, R, that causes voltage changes primarily because of the real **power flowing in** the **system** and distribution system reactance, X, that causes voltage changes because of the reactive **power flowing in the system**.

Tabla 35: Significados de *power system* y paráfrasis diferentes

Igualmente, las paráfrasis verbales de un CN pueden desvelar relaciones diferentes cuando una relación es consecuencia de otra. Esto sucede, por ejemplo, en *gas turbine*, donde la relación ampliamente extendida en las paráfrasis es turbine *uses_resource* gas. No obstante, identificamos una paráfrasis en la que se hablaba de *gases released from the gas turbine*. En estas turbinas, los gases se utilizan para activar el sistema y, posteriormente, se expulsan. Por ello se pueden encontrar ambos significados, aunque la presencia significativamente superior de *gas turbine* como turbine *uses_resource* gas, frente a una única ocurrencia de la consecuencia que genera (turbine *causes* gas), indica que la primera relación (*uses_resource*) es la prioritaria y origina la segunda proposición (*causes*).

En otras ocasiones, la identificación de verbos que indican diferentes relaciones puede deberse a la presencia de ruido, ya que utilizamos un *span* amplio (±10) para

favorecer la obtención de resultados, dada la tecnicidad de muchos de nuestros CN. Así, en ocasiones los verbos extraídos que apuntan a relaciones diferentes no constituyen una relación semántica, sino que simplemente forman parte del discurso.

A pesar de los resultados obtenidos y de las utilidades de las paráfrasis verbales, estas no permitieron descodificar la relación interna en 77 de los 190 CN. A nuestro parecer, diferentes factores pudieron intervenir. Uno de ellos es el hecho de que algunas relaciones internas sean muy evidentes y no se plasmen de forma explícita con verbos. Esto sucede, entre otros casos, en *voltage sag*, donde puede sobreentenderse que es la tensión la que sufre una caída y, por tanto, se ve afectada por ella. Además, tras seguir explorando las relaciones mediante paráfrasis libres, como comentaremos a continuación, nos percatamos de que la ausencia de paráfrasis verbales también podía deberse a la frecuente omisión de elementos que se da en los CN. Por ejemplo, en CN como *instantaneous voltage*, *instantaneous* alude a un *valor* (*value*), cuya omisión probablemente complicó la extracción de paráfrasis verbales. Asimismo, otra de las circunstancias que dificultó la obtención de paráfrasis verbales pudo ser la abundancia de atributos (57 de los 77 CN no encontrados), que a menudo no se expresan mediante verbos. Esto se observa, por ejemplo, en los CN en los que interviene *average* o *mean*, que no intervienen de manera autónoma en las oraciones, complementando a verbos que señalen su carácter de atributos.

De este modo, cuando no fue posible dilucidar la relación interna del CN por medio de paráfrasis verbales, recurrimos a las paráfrasis libres. Mediante ellas, pudimos aclarar la relación de 57 de los 77 CN. En la Tabla 36 se presentan las relaciones extraídas mediante paráfrasis libres, junto a ejemplos que ilustran su utilidad.

Relación	Ejemplo de CN y paráfrasis libres
affects	voltage sag (sag *affects* voltage)
	Voltage **sag is a short duration reduction in** rms **voltage** which can be caused by a short circuit, overload or starting of electric motors. Waveform distortion (i.e., harmonics), voltage unbalance (i.e., negative phase sequence voltage), and transient **voltage variations** (**i.e.**, dips and **sags**).
attribute_of	battery voltage (voltage *attribute_of* battery)
	A passive filter to keep harmonics off the output, and a transformer to adapt the **voltage levels of battery** and grid. When designing PV modules take into account the frequent use of PV modules for charging electric **batteries, whose voltage** is 12–12.5 V.
has_attribute	instantaneous voltage (voltage *has_attribute* instantaneous)
	The immediate simulation output are the **instantaneous values of voltages**, currents and quantities derived therefrom.

	The **instantaneous value of the** onshore PCC **voltage**.
effected_by	nuclear power (power *effected_by* nuclear energy [nuclear])
	Conventional plants **generate power from fossil fuels and nuclear materials**, which use large amount of water for cooling. The above indications only include **power generated by hydropower and nuclear energy** as well as energy carriers traded on the commercial world energy markets.
has_function	reference voltage (voltage *has_function* reference)
	Furthermore, during the zero-voltage sag in the PCC, the **terminal voltage, as a reference**, will be lost and the control method might be inapplicable. Wind farms have to be able to maintain automatic regulation of reactive power, with **voltage as the reference value**.
part_of	wind turbine system (system *part_of* wind turbine)
	A DC motor that emulates the **mechanical system of the wind turbine (rotor blades, shaft)**. Clipper Windpower Liberty **wind turbine with Multi-PMG system**.
has_part	wound rotor induction generator (induction generator *has_part* wound rotor)
	The rotor of an induction generator can be designed as a so-called short-circuit rotor (squirrel cage rotor) or **as a wound rotor**.
has_type	wind turbine generator (generator *has_type* wind turbine)
	However, with the connection of small scale **generators, such as wind turbines**, the characteristics of the distribution network are changing leading to changes in the operation of these networks.
takes_place_in	offshore wind (wind *takes_place* offshore)
	Wind farms can also be located **offshore, in the water, where winds** tend to blow stronger and more steadily–and turbines can be built even bigger. The **offshore environment** can be much harsher, **with high winds**, significant wave loads, and corrosion-producing salt water.
uses_resource	wind power plant (plant *uses_resource* wind power)
	To assess the costs incurred by **wind power utilisation**, first variable and fix costs of the reference **plants** are discussed.

Tabla 36: Relaciones identificadas mediante paráfrasis libres, ilustradas con ejemplos

En la Tabla 36 se aprecian las relaciones que pudieron extraerse mediante la búsqueda de paráfrasis libres en el caso de que las paráfrasis verbales no arrojasen resultados. Al tratarse de consultas más libres, se obtuvo también más ruido, es por ello que preferimos buscar en primer lugar las paráfrasis verbales, pues además estas profundizan en el significado de la relación.

A pesar de que las paráfrasis libres no buscan verbos que codifiquen el significado de la relación de forma específica, permiten acceder al contexto compartido de los formantes del CN y, así, detectar el vínculo entre estos. Dicho vínculo se identifica, entre otras muchas posibilidades, mediante marcadores como *such as* (en la relación *has_type* de la Tabla 36), adverbios como *where* (en la relación *takes_place_in* de la Tabla 36), paréntesis (en la relación *has_part* de la Tabla 36) o el simple contexto en el que coocurren

los formantes y que permite inferir su relación. Por tanto, este análisis de las paráfrasis libres puede utilizarse también para alimentar los conjuntos de patrones de conocimiento que también se utilizan como puerta de acceso a las relaciones semánticas entre conceptos. Al tratarse de un análisis contextual como el de las paráfrasis verbales, este tipo de paráfrasis también puede revelar otros argumentos, elementos del CN que se han omitido, así como variantes.

Sin embargo, las paráfrasis libres no permitieron explicitar la relación de 20 CN de nuestra muestra. A este respecto, no debemos olvidar que las relaciones no siempre se expresan de forma manifiesta en el lenguaje pues, como hemos mencionado antes, no es extraño que, debido a la normalidad en el vínculo entre dos conceptos, este no llegue a materializarse en el corpus. Por tanto, la documentación sobre el tema jugó un papel fundamental en el conocimiento del área y de las relaciones internas de los CN.

Profundizando en el papel de las relaciones en la formación de CN, constatamos que los conceptos y relaciones del sistema conceptual subyacente pueden condensarse en estos términos compuestos. Sin embargo, no lo hacen de cualquier forma, pues las restricciones semánticas de las que venimos hablando en este trabajo también se dan, como no podía ser de otra forma, a nivel de las relaciones. Así, entre los conceptos que dan lugar a los CN debe establecerse una relación lógica, en sintonía con los conceptos que vincula. En este sentido, la Tabla 37 recoge las relaciones que codifican las categorías ontológicas de ENTIDAD (ENTITY), ATRIBUTO (ATTRIBUTE) y PROCESO (PROCESS).

Categoría ontológica	**Relación interna del CN**	**N.º de CN**
ENTITY	*has_attribute*	24
	uses_resource	19
	causes	11
	part_of	10
	affects	5
	has_part	4
	located_at	4
	represents	4
	has_function	2
	has_type	2

ATTRIBUTE	*attribute_of*	46
	has_attribute	20
	caused_by	2
	has_function	1
PROCESS	*affects*	10
	effected_by	9
	has_attribute	5
	has_result	4
	takes_place_in	3
	uses_resource	2
	caused_by	1
	has_type	1

Tabla 37: Relaciones codificadas por cada categoría ontológica

En la Tabla 37 se observa que las ENTIDADES de nuestra muestra se especifican principalmente según sus ATRIBUTOS o el recurso que utilizan para llevar a cabo su acción. Los ATRIBUTOS, por su parte, suelen generar CN en los que se concreta el concepto del que son propiedades o bien sus características propias. Por último, los PROCESOS codifican varias relaciones sin una tendencia clara. Entre ellas, las más habituales son el concepto al que afecta o, por el contrario, el concepto que permite su creación, además de ATRIBUTOS y resultados. Por tanto, de estas tres categorías destacan los ATRIBUTOS por la homogeneidad de sus relaciones.

En este sentido, se pueden resumir las principales relaciones por las que se combinan los conceptos en nuestros términos. Estas son *has_attribute* (49), *attribute_of* (46), *uses_resource* (21), *affects* (15), *causes* (11), *part_of* (10), y *effected_by* (9), mientras que el resto de relaciones intervienen en 4 o menos CN. De ello se deduce que los CN estudiados, que pueden servir como ejemplo de términos compuestos en el área de la energía eólica, se forman principalmente especificando sus características o el concepto del que son atributos. Asimismo, un buen número de CN surgen al concretar el recurso que utilizan para realizar su actividad, el concepto al que afectan o del que forman parte, los conceptos que causan y aquellos a partir de los cuales se han formado. Por tanto, utilizando la clasificación que distingue entre CN relacionales (aquellos cuyos formantes se vinculan mediante una relación semántica) y atributivos (aquellos que poseen un

atributo como enlace entre sus elementos) (Montero Martínez 2002, 2008), contamos con 96 CN atributivos y 94 CN relacionales, unas cantidades equiparadas que revelan la presencia igualitaria de estos dos tipos de vínculos en nuestros CN.

Un caso especial de relaciones semánticas son aquellas que se dan en los CN que cuentan con un verbo en su interior, ya sea nominalizado o en una forma verbal (p. ej. *stall-regulated wind turbine*). A pesar de que estos verbos permiten intuir la proposición subyacente, las paráfrasis no dejan de ser útiles, pues ofrecen una visión más profunda de su conceptualización, como el rol que desempeñan el resto de elementos o la presencia de otros argumentos.

En la Sección 5.1.1.1 señalamos el carácter problemático de este tipo de términos a la hora de establecer la estructura de *bracketing*, ya que este verbo representa el predicado de la proposición subyacente y puede mantener vínculos con diferentes elementos del CN. Ante estos CN, se presentan dos posibilidades de análisis. Una de ellas consiste en tomar el verbo como relación interna, de manera que este queda fuera de los grupos del *bracketing*. Así, en *stall-regulated wind turbine*, la proposición conceptual sería wind turbine *affected_by* stall, de modo que el participio *regulated* transmitiría la relación *affects*, matizando su significado. La otra opción de análisis consiste en incluir el verbo en el grupo del *bracketing* correspondiente, de forma que la relación interna será otra, que precisará la conceptualización subyacente. En [*stall-regulated*] [*wind turbine*], la proposición sería wind turbine *has_attribute* stall-regulated. Ambas opciones son válidas, ya que la conceptualización no varía: se alude en los dos casos a un aerogenerador que se detiene (ante fuertes velocidades de viento), para lo cual es necesario que cuente con unas determinadas características físicas, que son sus atributos (*stall-regulated*).

En este estudio adoptamos la segunda opción para mantener nuestra idea del *bracketing* como la simplificación de núcleo y modificador. No obstante, consideramos que ambas son acertadas y se encuentran, de hecho, interrelacionadas, pues constituyen un enfoque multidimensional de la misma conceptualización, que puede materializarse en proposiciones subyacentes distintas. Otro ejemplo de este caso lo constituye *wind power generation system*, que ya comentamos en la Sección 5.1.1.1, y que igualmente puede adoptar diferentes estructuras de *bracketing* y, con ello, diferentes relaciones, todas ellas válidas. Sin embargo, en nuestro estudio adoptamos el *bracketing* más frecuente y la relación que se codificaba entre esos dos grupos (generation system *uses_resource* wind power). Este tipo de CN nos parece de gran importancia, por lo que nos detendremos en su análisis en próximos estudios.

Como hemos detallado en esta sección, las relaciones semánticas son fundamentales en los CN, no solo porque se necesitan para comprender el significado de estas combinaciones, sino también porque constituyen la base para su formación.

5.1.1.4 Asignación de roles semánticos

Tras conocer las proposiciones subyacentes, exploramos los roles semánticos que desempeñan los formantes de los CN que participan en esas proposiciones. A pesar de que los roles se suelen aplicar a los argumentos de un predicado, planteamos su aplicación a los CN basándonos en las proposiciones que subyacen en estos términos. En ellas, se combinan varios conceptos entre los que se establece una relación. Por tanto, estos roles se asignan en función de la relación que vincule los conceptos y no necesariamente de la presencia explícita de un predicado (p. ej. en *power station* consideramos que *station* es el AGENTE que *causa* un PACIENTE, *power*). Así, pudimos captar regularidades según las funciones que desempeñaban los formantes de los CN, como comentaremos detenidamente en la Sección 5.1.2.

En función de las diferentes formas de combinación de los elementos de un CN, estos desempeñaron los siguientes roles semánticos, que describimos según su aplicación en este trabajo:

-AGENT: concepto que, con o sin voluntad propia, realiza una acción, es responsable de otro concepto o lo afecta de algún modo. Las relaciones que codifica en nuestros CN son, por tanto, *effects, causes, affects, has_result* y *uses_resource*.

-PATIENT: concepto que surge a partir de la acción de otro concepto, o bien, la recibe y se ve afectado de algún modo. En nuestra muestra, da lugar a CN al establecer las relaciones *effected_by*, *caused_by*, *affected_by, result_of* y *represented_by.* Como se observa, algunas de estas relaciones son las mismas que codifica AGENT, en el orden inverso, ya que conforman el par de roles más conocido.

-THEME: concepto que puede recibir la acción de otro concepto sin verse afectado, estar implicado en un evento sin ser el responsable o utilizarse cuando se acompaña de rasgos descriptivos que se quieren enfatizar. En nuestros CN codifica principalmente las relaciones *has_attribute, attribute_of* y *part_of* y, en menor medida, *has_function, has_type, takes_place_in* y *located_at*. A pesar de su semejanza con el rol de PATIENT, los distinguimos porque PATIENT sufre cambios al recibir la acción, mientras que THEME la recibe sin verse afectado.

-POSSESSOR: concepto que posee o del que forman parte otros participantes. En nuestra muestra puede establecer la relación *has_part* o *has_attribute*, según si lo que posee es un componente físico (parte) o una propiedad (atributo).

-DESCRIPTIVE: concepto que describe características de otro concepto. En los CN estudiados suele activar la relación *attribute_of* y, en menor medida, *represents, type_of* y *function_of*.

-INSTRUMENT: concepto empleado por un AGENTE para realizar una acción. En nuestra muestra codifica la relación *resource_used_by*. Cabe señalar que la asignación del rol de INSTRUMENT a un participante no implica que este deba pertenecer a la categoría semántica de INSTRUMENT, ya que se trata de inventarios diferentes que se basan en criterios distintos.

-LOCATION: concepto donde está ubicado o tiene lugar otro concepto. Así, cuando el núcleo designa una entidad ubicada en ese concepto anotado como LOCATION, la relación semántica es *located_at*, mientras que, si el núcleo alude a un proceso que se produce en esa ubicación, la relación es *takes_place_in*.

Una vez descritos los conceptos a los que puede aplicarse cada rol, mostramos en la Tabla 38 las diferentes proposiciones conceptuales que subyacen en nuestros CN, empleando estos roles semánticos. Dado que nos centramos en CN en inglés formados por premodificación, es importante señalar que la interpretación debe realizarse desde la derecha, donde se encuentra el núcleo, que marca la dirección de la relación semántica y, a continuación, se acompaña del rol del modificador (p. ej. en *line-to-neutral voltage*, voltage[THEME] *attribute_of* line-to-neutral[POSSESSOR]). Para mantener el equilibrio en el análisis, los CN formados por postmodificación (p. ej. *impact of wind power*) se anotaron en el orden inverso, como si de estructuras de premodificación se tratase, pues los CN de ese tipo constituían solo una pequeña parte de la muestra (3 CN). Los esquemas de roles también deben interpretarse desde la derecha, donde se encuentra el núcleo de los CN formados por premodificación en inglés.

Esquema de roles	Rol del modificador + relación + rol del núcleo	Ejemplo de CN	N.º de CN
POSSESSOR-THEME	POSSESSOR + *attribute_of* + THEME	line-to-neutral voltage	46
POSSESSOR-THEME	POSSESSOR + *part_of* + THEME	wind turbine blade	10
DESCRIPTIVE-THEME	DESCRIPTIVE + *has_attribute* + THEME	rated wind speed	48
DESCRIPTIVE-THEME	DESCRIPTIVE + *has_function* + THEME	generator system	3
DESCRIPTIVE-THEME	DESCRIPTIVE + *has_type* + THEME	wind resource	3
PATIENT-AGENT	PATIENT + *affects* + AGENT	power grid	15
PATIENT-AGENT	PATIENT + *causes* + AGENT	DC generator	11
PATIENT-AGENT	PATIENT + *has_result* + AGENT	wind power production	4
INSTRUMENT-AGENT	INSTRUMENT + *uses_resource* + AGENT	hydraulic turbine	21
AGENT-PATIENT	AGENT + *effected_by* + PATIENT	wave power	10
AGENT-PATIENT	AGENT + *caused_by* + PATIENT	impact of wind power	4
LOCATION-THEME	LOCATION + *located_at* + THEME	onshore wind farm	4
LOCATION-THEME	LOCATION + *takes_place_in* + THEME	offshore wind energy	3
THEME-POSSESSOR	THEME + *has_part* + POSSESSOR	three-phase generator	4
PATIENT-DESCRIPTIVE	PATIENT + *represents* + DESCRIPTIVE	power spectrum	4

Tabla 38: Proposiciones conceptuales que subyacen en los CN de la muestra, anotadas con roles semánticos e ilustradas con ejemplos

En la Tabla 38 se aprecia que los CN de nuestra muestra responden a un número limitado de esquemas. En ocasiones, estos establecen siempre la misma relación (p. ej. el esquema INSTRUMENT-AGENT siempre activa la proposición AGENT *uses_resource* INSTRUMENT), mientras que hay otros esquemas que pueden codificar varias relaciones (p. ej. el esquema PATIENT-AGENT puede activar las proposiciones AGENT *affects/causes/has_result* PATIENT). A pesar de ello, se observa que son 8 las combinaciones conceptuales posibles, a nivel de roles, que dan lugar a nuestros CN. Dichas combinaciones se describen a continuación.

El esquema POSSESSOR-THEME se asigna cuando el núcleo se concreta con el concepto del que es atributo (*attribute_of*) o componente (*part_of*). Esta combinación generó 56 CN, ya que uno de los modos más frecuentes de nombrar nuevos conceptos consiste en relacionarlos con el concepto al que pertenecen.

Por otra parte, el esquema DESCRIPTIVE-THEME se aplica cuando el núcleo se especifica con alguna de sus características, como atributos (*has_attribute*), su función (*has_function*) o un subtipo (*has_type*). Si nos centramos en la relación *has_attribute*, observamos que este esquema constituye el orden inverso del esquema anterior de POSSESSOR-THEME (es decir, en el caso anterior, el núcleo es un atributo del modificador, mientras que, en este caso, el núcleo se especifica con sus atributos). Esta combinación dio lugar a 54 CN, lo que no resulta extraño, ya que es habitual denominar nuevos conceptos especificando los rasgos que lo distinguen del resto de la clase.

El esquema PATIENT-AGENT se aplica cuando el núcleo especifica el concepto al que afecta (*affects*), al que da lugar (*causes*) o que surge como resultado cuando el núcleo designa un proceso (*has_result*). Este esquema dio lugar a 30 CN, lo que demuestra que un modo común de crear términos consiste en especificar los conceptos que generan o a los que afectan.

Otro de los esquemas habituales es el de INSTRUMENT-AGENT. En los CN formados a partir de este patrón, el núcleo se concreta con el recurso que emplea para llevar a cabo su actividad (*uses_resource*). Esta combinación también fue recurrente en nuestra muestra, dando lugar a 21 CN, ya que la energía eólica es un área basada en el empleo de un recurso, el viento, para producir energía.

El esquema AGENT-PATIENT se observa cuando el núcleo se acompaña de un modificador que especifica el AGENTE de la proposición. Esta es, por tanto, PATIENT *effected_by/caused_by* AGENT. Como puede observarse, entendemos *causes/caused_by* y *effects/effected_by* como relaciones diferentes, aunque bastante similares. Utilizamos *causes/caused_by* cuando un concepto produce otro (p. ej. en *power plant*, plant *causes* power), mientras que *effects/effected_by* se entiende como el empleo de un concepto para llevar a cabo otro (p. ej. en *wind power*, power *effected_by* wind). Este esquema se dio en 14 de los CN estudiados, pues una forma de nombrar nuevos conceptos consiste en especificar el agente a partir del cual surgen.

El esquema LOCATION-THEME, que dio lugar a 7 CN, corresponde a los CN cuyo núcleo se concreta mediante su ubicación. Como hemos indicado previamente, la relación

es *located_at* cuando el núcleo es una entidad y *takes_place_in* cuando el núcleo designa un proceso.

El esquema THEME-POSSESSOR, que generó 4 CN, se asigna a aquellos CN cuyo núcleo se especifica mediante alguno de sus componentes (*has_part*). Se trata de un esquema relacionado con el de DESCRIPTIVE-THEME, ya que también se trata de un núcleo concretado mediante algo que posee. La diferencia es que en este caso se alude a sus partes, entendidas como componentes, mientras que en el esquema DESCRIPTIVE-THEME se especifican propiedades, además de otras características. No se emplea el esquema THEME-POSSESSOR para indicar propiedades, ya que la aplicación de dicho esquema a valores de atributos (p. ej. *asynchronous*) carecería de sentido.

Finalmente, el esquema PATIENT-DESCRIPTIVE se utilizó cuando el núcleo servía para describir al modificador, entendido como PACIENTE al presentar un carácter variable en esa descripción. La relación que se establece en nuestros CN es *represents* y fueron 4 los CN generados a partir de este patrón.

Estos 8 esquemas posibles permiten delimitar el potencial combinatorio del núcleo según el rol que este desempeñe. Así, cuando el núcleo desempeña el rol de THEME, puede acompañarse de modificadores cuyo rol sea POSSESSOR (*line-to-neutral voltage*), DESCRIPTIVE (*rated wind speed*) o LOCATION (*onshore wind farm*). Por otra parte, cuando el núcleo es AGENT, puede acompañarse de modificadores que desempeñen la función de PATIENT (*power grid*) o INSTRUMENT (*hydraulic turbine*). Finalmente, en nuestra muestra cuando el rol del núcleo es PATIENT se acompaña de AGENT (*wave power*), cuando es POSSESSOR se combina con THEME (*three-phase generator*) y cuando es DESCRIPTIVE coocurre con PATIENT (*power spectrum*).

Aunque hemos señalado que en los CN no siempre interviene explícitamente un verbo, hay muchos casos en los que este forma parte expresa del término, por ejemplo, mediante nominalizaciones o formas de participio (*voltage control, self-excited induction generator*). En ocasiones, estas formas sugieren el rol de alguno de los participantes. Ocurre así en las nominalizaciones finalizadas en *-or* o *-er*, que a menudo indican el AGENTE que realiza la acción del verbo (*generator, converter*). Por otro lado, los participios pasados suelen acompañar al concepto que recibe la acción (*fixed speed*), mientras que los participios presentes a menudo coocurren con el concepto que realiza la acción (*generating unit*). En este sentido, la utilización de participios pasados o presentes dependerá de la estructura argumental de cada predicado.

Así, las formas verbales presentes en los CN se anotaron con el rol correspondiente según la proposición que subyace en el CN en cuestión. Por ejemplo, en *voltage control* señalamos a *voltage* como PATIENT y a *control*, como AGENT, ya que este *afecta* de algún modo a *voltage*. Sin embargo, debemos tener en cuenta que estas formas verbales suelen constituir el predicado subyacente que activa la estructura argumental. Así, pueden generarse otros CN en los que intervengan nuevos participantes que concreten esta estructura argumental. Uno de ellos podría ser el concepto encargado de controlar la tensión, que sería el AGENTE propiamente dicho. Este puede ser, por ejemplo, un convertidor, de forma que la proposición conceptual sería voltage *affected_by* converter, en la que el verbo *control* concreta el modo en el que debe entenderse la relación *affected_by* (voltage *controlled_by* converter). Así, la especificación de otros argumentos de la estructura da lugar a nuevos CN como *voltage converter*, en el que se ha omitido el predicado *control*, o bien *voltage control device, voltage control equipment* o *voltage control system*, en los que se dan cita el AGENTE y el PACIENTE del predicado.

Hemos aludido en varias ocasiones a la complejidad de este tipo de CN que cuentan con una nominalización verbal en su interior, lo que revierte en su desambiguación estructural y la descodificación de la relación interna, ya que el predicado puede mantener vínculos con el resto de formantes. De este modo, consideramos a estos predicados dentro del grupo del *bracketing* sugerido según nuestro protocolo y, por tanto, les asignamos el rol correspondiente en cada caso. Así, en *voltage control equipment*, la proposición sería voltage[PATIENT] *affected_by* control equipment[AGENT]. Adoptamos esta postura para respetar el esquema básico de núcleo y modificador que posibilita el resto de tareas de análisis, como la exploración de regularidades en los esquemas binarios de roles. Sin embargo, somos conscientes de que esta estructura puede desplegarse, de forma que el predicado quedaría fuera de los grupos del *bracketing* y especificaría la proposición conceptual: voltage[PATIENT] *controlled_by* equipment[AGENT].

Por otro lado, aunque solo ha sido objeto de una exploración superficial, ya que no contamos con un número suficiente de CN que cuenten con formas verbales, se observaron preferencias de combinación según el tipo de verbos que intervenían en los CN. Esto es, los verbos pertenecientes al dominio léxico de EXISTENCIA (Faber y Mairal 1999) suelen formar CN en los que interviene el PACIENTE y, en menor medida, el AGENTE o el INSTRUMENTO que emplean para la producción, como se observa en *power generation* (PACIENTE), *power generator* (PACIENTE y AGENTE) o *wind generation* (INSTRUMENTO). Por su parte, los verbos de ACCIÓN coocurren en nuestros CN con su PACIENTE (*voltage control*), al igual que los verbos de MOVIMIENTO (*wind power fluctuation*). Por último, los verbos de

CAMBIO suelen aparecer junto a su PACIENTE y, en ocasiones, su AGENTE (*voltage source inverter*). Como puede suponerse en dominios como el medio ambiente o las energías renovables, los verbos empleados pertenecen fundamentalmente a los dominios de EXISTENCIA, ACCIÓN, MOVIMIENTO y CAMBIO, y suelen acompañarse del PACIENTE que recibe la acción del verbo. En definitiva, estas preferencias que dan lugar a los CN constituyen un reflejo de la estructura argumental de los predicados.

En relación con el empleo de los roles semánticos para analizar la formación de los CN, retomamos el caso de los CN polisémicos que entrañan un cambio de la relación interna, lo que desemboca a menudo en un cambio de los roles de los participantes. A raíz de esto, un mismo CN puede surgir a partir de distintos esquemas de roles. En la Tabla 39 se presentan los CN polisémicos que implicaron un cambio de la relación y de roles, junto a las proposiciones subyacentes y los esquemas de roles que se aprecian en cada caso. Como hemos indicado anteriormente, la interpretación debe realizarse desde la derecha, dada la premodificación de estos términos.

CN	Rol del modificador + relación + rol del núcleo	Esquema de roles
wind turbine generator	POSSESSOR *part_of* THEME	POSSESSOR-THEME
	DESCRIPTIVE *has_type* THEME	DESCRIPTIVE-THEME
offshore wind power	LOCATION *takes_place_in* THEME	LOCATION-THEME
	POSSESSOR *attribute_of* THEME	POSSESSOR-THEME
wave power	AGENT *effected_by* PATIENT	AGENT-PATIENT
	POSSESSOR *attribute_of* THEME	POSSESSOR-THEME
wind power	AGENT *effected_by* PATIENT	AGENT-PATIENT
	POSSESSOR *attribute_of* THEME	POSSESSOR-THEME
offshore wind energy	LOCATION *takes_place_in* THEME	LOCATION-THEME
	POSSESSOR *attribute_of* THEME	POSSESSOR-THEME
wind energy	AGENT *effected_by* PATIENT	AGENT-PATIENT
	POSSESSOR *attribute_of* THEME	POSSESSOR-THEME
wind system	INSTRUMENT *uses_resource* AGENT	INSTRUMENT-AGENT
	DESCRIPTIVE *has_type* THEME	DESCRIPTIVE-THEME

Tabla 39: CN polisémicos que implicaron un cambio de relación interna y de roles semánticos

Así pues, *wind turbine generator* puede codificar las relaciones *part_of* y *has_type*, lo que implica el esquema POSSESSOR-THEME en el primer caso y DESCRIPTIVE-THEME en el segundo. *Offshore wind power*, como su variante *offshore wind energy*, puede activar las relaciones *takes_place_in* y *attribute_of*, con los roles LOCATION-THEME en el primer caso y POSSESSOR-THEME en el segundo. Por otro lado, *wave power*, al igual que su cohipónimo *wind power* y la variante de este, *wind energy*, puede establecer las relaciones *effected_by* y *attribute_of* y, con ello, los esquemas AGENT-PATIENT y POSSESSOR-THEME, respectivamente. Por último, *wind system* puede codificar la relación *uses_resource* y el esquema INSTRUMENT-AGENT, o bien la relación *has_type* y el esquema DESCRIPTIVE-THEME. Otro caso diferente es el de *power system*, también polisémico y con diferentes relaciones (*causes* o *affects*), pero en el que no se produce un cambio de roles, ya que ambas relaciones unen a los mismos participantes (PATIENT-AGENT).

Además, este análisis permite confirmar que en los términos compuestos también intervienen conceptos que pueden desempeñar roles diferentes debido a su carácter multidimensional (León Araúz 2009). Ello se observa, por ejemplo, en *wave* y *wind* que, sin variar sus características conceptuales, pueden actuar como AGENT o POSSESSOR, en función de las características que se enfaticen: AGENT cuando se concibe como concepto que da lugar a otro y POSSESSOR cuando se muestra como un concepto con unas determinadas características. Por tanto, constatamos de nuevo que la multidimensionalidad no puede ignorarse en el estudio de los CN.

En definitiva, la anotación con roles semánticos abre una vía interesante de análisis de los CN, pues estos determinan las combinaciones que pueden dar lugar a los CN. Así, se observaron patrones recurrentes de roles que suelen generar CN. Estas restricciones constituyen el punto de partida de los microcontextos, como veremos a continuación.

5.1.2 Los microcontextos en los compuestos nominales

Una vez realizado el análisis semántico de los CN, ahondaremos en el procedimiento de formación de estos términos compuestos. Para ello, partimos de una idea formulada desde las teorías de combinación conceptual de la Psicolingüística (Murphy 1988, 1990; Maguire et al. 2010b), la Lingüística Computacional (Rosario et al. 2002) y la fraseología especializada (Martin 1992; Heid 1994, 2001; L'Homme 1998, 2000, 2002; Montero Martínez 2002, 2008). En concreto, se defiende que las unidades similares semánticamente abren unos *slots* que se rellenan por otras unidades de contenido semántico también similar, dando lugar a combinaciones especializadas. Estos *slots* son una especie de estructura argumental, que puede percibirse en la definición del concepto de partida.

Sin embargo, esta idea no siempre se ha aplicado a CN especializados, ya que en muchas ocasiones el objeto de estudio han sido las combinaciones libres de la lengua general o las colocaciones verbales especializadas. Por tanto, en esta tesis adoptamos dicha hipótesis para evaluar su incidencia en la formación de CN especializados de cualquier número y tipo de formantes en el campo de la energía eólica. Así, nos interesan, además de los CN de dos elementos, aquellos que están formados por tres o más constituyentes, pues representan combinaciones conceptuales más complejas. Este análisis nos permitió desarrollar nuestra noción de *microcontextos*, que detallaremos tras el análisis y se presenta como una propuesta sistemática que da cuenta de la productividad en la formación de CN especializados, diferenciando claramente entre roles y categorías semánticas de los *slots*, y las relaciones que los vinculan.

5.1.2.1 Análisis de los microcontextos

Nuestro análisis partió de la agrupación de unidades similares semánticamente, para lo que nos basamos en las categorías semánticas de los núcleos de los CN, que denominamos *categorías nucleares*. Por ejemplo, los términos *turbine, generator, converter* y *rotor*, entre otros, pertenecen a la categoría nuclear de ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT. En esta línea, el término *grid* también pertenece a la categoría nuclear de ENTITY>CREATION, en concreto, ENTITY>CREATION>ARTIFACT>CONDUIT; mientras que los términos *plant* o *farm* se engloban en la categoría de ENTITY>CREATION>STRUCTURE. En otras palabras, las tres subcategorías presentadas coinciden en la categoría nuclear general de ENTITY>CREATION, en la que especifican conceptos más concretos. Así, exploramos si las tres subcategorías que se incluyen en esa categoría general presentaban los mismos patrones de combinación, para lo que comparamos las categorías y roles semánticos con los que coocurren para formar CN, así como las relaciones que vinculan a estos elementos. Por ejemplo, la proposición subyacente en *AC generator* es generator *causes* AC, de modo que entendemos que *AC* es el PACIENTE, igual que sucede con *voltage source* en *voltage source converter* (converter *affects* voltage source) o *power* en *power plant* (plant *causes power*). Así, la especificación del PACIENTE es una de las formas más productivas de crear CN cuyos núcleos pertenezcan a la categoría de ENTITY>CREATION.

Realizamos este análisis en las diferentes categorías nucleares generales que agrupaban varias subcategorías, para así poder evaluar si sus patrones de combinación coincidían. Por el contrario, no comentaremos los *slots* de aquellas categorías nucleares en las que solo se agrupa una categoría o incluso un CN, pues no permiten explorar nuestra hipótesis.

Abordamos en primer lugar la categoría ontológica ENTITY, en la que comentaremos los *slots* que abren las categorías ENTITY>CREATION, ENTITY>SYSTEM, ENTITY>PART, ENTITY>INFORMATION y ENTITY>HUMAN de nuestra muestra.

- **ENTITY>CREATION**

La categoría ENTITY>CREATION es una de las categorías principales en nuestro estudio, pues engloba a un gran número de CN, cuyos núcleos pertenecen a las siguientes subcategorías:

ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT (*AC generator*, *wound rotor induction generator*, etc.)

ENTITY>CREATION>ARTIFACT>CONDUIT (*power grid*)

ENTITY>CREATION>STRUCTURE (*wind power plant*, *offshore wind farm*, etc.)

En la Tabla 40 se muestran los *slots* que abren estas tres subcategorías. En la columna de la izquierda se indican los diferentes *slots*, que se designan según el rol que desempeñen en la proposición conceptual del CN, a continuación se presentan las categorías semánticas que pueden ocuparlos, la relación semántica que codifica el núcleo con esos *slots* y, por último, algunos CN que cumplen con dicho patrón de combinación.

ENTITY>CREATION			
***Slot* (rol)**	**Categoría**	**Relación**	**Ejemplos**
PATIENT	PROCESS>MOVEMENT>ENERGY MOVEMENT	*causes* *affects*	AC generator voltage source converter
INSTRUMENT	ENTITY>MATTER ENTITY>FORCE PROCESS>MOVEMENT	*uses_resource*	steam generator wind farm
DESCRIPTIVE	ATTRIBUTE ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	*has_attribute* *has_function* *has_type*	vertical axis wind turbine large wind farm
LOCATION	ATTRIBUTE>LOCATION	*located_at*	offshore wind turbine onshore wind farm
POSSESSOR	ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	*part_of*	generator rotor wind turbine generator
THEME (PART)	ENTITY>MATTER>SOLID MATTER>MATERIAL ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	*has_part*	permanent magnet generator wound rotor induction generator

Tabla 40: *Slots* que abren las subcategorías de ENTITY>CREATION

Como se observa, la categoría ENTITY>CREATION abre diferentes *slots*: PATIENT, INSTRUMENT, DESCRIPTIVE, LOCATION, POSSESSOR y THEME (PART). Por una parte, el *slot* de PATIENT alude al concepto que sufre la acción del núcleo. En todos los casos, la categoría[54] que ocupa este *slot* es PROCESS>MOVEMENT>ENERGY MOVEMENT, como se aprecia en los ejemplos *AC generator* y *voltage source converter*. Sin embargo, este paciente puede adquirir dos conceptualizaciones según la relación que codifique el núcleo. Así, PATIENT se entiende como el concepto que se convierte o transporta cuando esa es la acción que realiza la categoría ENTITY>CREATION y, por tanto, la relación interna es *affects* (converter *affects* voltage source). Este es el caso en los CN generados con los núcleos *converter, inverter* y *grid* de nuestra muestra. Por el contrario, el paciente puede entenderse también como el concepto resultante de la actividad de ENTITY>CREATION cuando este enfatiza la fase de producción y, así, codifica la relación *causes*. Este patrón es habitual en los CN cuyo núcleo es *generator*, *plant* o *station* (generator *causes* AC). En cualquier caso, observamos cómo la categoría ENTITY>CREATION puede especificarse indicando el PACIENTE o concepto que sufre la acción. Este *slot* coincide en las tres subcategorías nucleares que comprende la categoría general ENTITY>CREATION.

Por su parte, el *slot* de INSTRUMENT permite especificar el concepto que emplea ENTITY>CREATION para realizar su actividad. Son dos los tipos de conceptos que suelen rellenar este *slot*. Por un lado, los que aluden al recurso externo que se utiliza (p. ej. *gas turbine, diesel generator*), que pertenecen a las categorías de ENTITY>MATTER o PROCESS>MOVEMENT. Por otro lado, los conceptos que se refieren a la fuerza o los materiales empleados para generar movimiento (p. ej. *induction generator, impulse turbine*), que pertenecen a las categorías de ENTITY>FORCE y ENTITY>MATTER. Como se desprende de las diferentes categorías, los recursos empleados pueden ser de diferente tipo: fluidos, sustancias químicas, gases, etc. (ENTITY>MATTER), además de procesos como el viento o la inducción, lo que determina las características de ENTITY>CREATION para adaptarse al recurso que utilice.

Uno de los casos particulares en los que se rellena este *slot* es el adjetivo *renewable*, que se utiliza como sustantivo para referirse a las energías renovables (p. ej. *renewable generator*). La categoría semántica en este caso también es PROCESS>MOVEMENT, pues alude concretamente a un movimiento de energía. Así, los CN en los que se rellena este *slot* con *renewable* presentan un carácter más general que aquellos que especifican el

[54] Las categorías que rellenan los *slots* se presentan con diferentes niveles de jerarquización para captar regularidades en cada caso. Es decir, cuando se obtuvieron coincidencias en niveles específicos de la jerarquía conceptual, se muestra ese nivel; de lo contrario, se muestra el nivel superior en el que varias subcategorías coincidieron.

recurso concreto que se emplea (*wind, water*, etc.). Como se desprende de estos términos, la relación que codifica el núcleo es *uses_resource*, que hemos añadido al inventario de EcoLexicon dada su elevada presencia en el ámbito. Este *slot* coincide en todas las subcategorías de ENTITY>CREATION, salvo en la subcategoría ENTITY>CREATION>ARTIFACT> CONDUIT, que solo incluye un término en nuestra muestra (*power grid*) y se centra en el concepto que transporta, en lugar de en el recurso empleado para el transporte.

Cabe destacar que la presencia del modificador *power* en los *slots* de PATIENT e INSTRUMENT está originada por la polisemia de *power*, que puede referirse a la energía eléctrica generada (PATIENT) o a la energía cinética del viento empleada como recurso de generación (INSTRUMENT).

El *slot* DESCRIPTIVE permite incluir rasgos descriptivos de ENTITY>CREATION. Estos rasgos se designan con la categoría de ATTRIBUTE y pueden ser de diferentes tipos: atributos físicos, magnitudes, etc. La relación semántica interna es, por tanto, *has_attribute* (p. ej. en *vertical axis wind turbine*, wind turbine *has_attribute* vertical axis). Sin embargo, existen dos excepciones en nuestra muestra en las que se codifica otra relación. Por una parte, en *commercial wind turbine* se codifica la relación *has_function* (el aerogenerador está destinado a un uso comercial), aunque *commercial* no deja de ser un rasgo descriptivo del núcleo y, por ello, recibe el rol DESCRIPTIVE. Del mismo modo, *wind turbine generator* codifica la relación *has_type*, dado que el aerogenerador es un tipo de instrumento generador. En este caso cambia además la categoría semántica que rellena el *slot*, pues no se trata de un atributo, sino de la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT. A pesar de ello, el rol de *wind turbine* en este CN es el de DESCRIPTIVE, por lo que se incluye en este patrón de combinación. En definitiva, se trata de dos conceptos que no constituyen atributos, sino rasgos descriptivos de otro tipo: funciones y subtipos.

Este *slot* se rellena especialmente en la subcategoría nuclear ENTITY>CREATION> ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, aunque también en la subcategoría ENTITY>CREATION>STRUCTURE. Por el contrario, en nuestra muestra, la categoría ENTITY>CREATION>ARTIFACT>CONDUIT no abre este *slot*, lo que puede justificarse de nuevo por la existencia de un único concepto en esta subcategoría. Sería interesante evaluar otros conceptos de esa subcategoría para determinar si suele especificarse en función de otros criterios o si, por el contrario, la ampliación de la muestra permite identificar este *slot* en dicha subcategoría nuclear.

Por otra parte, el *slot* de LOCATION alude a la ubicación de ENTITY>CREATION. Como en el *slot* DESCRIPTIVE, se da en las subcategorías ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT y ENTITY>CREATION>STRUCTURE. No obstante, consideramos, de nuevo, que su análisis en otros CN de la subcategoría ENTITY>CREATION>ARTIFACT>CONDUIT revelaría su presencia como *slot* de esa categoría nuclear, pues todos los conceptos de la categoría ENTITY>CREATION deben ubicarse físicamente en algún lugar. Este *slot* se rellena con la categoría ATTRIBUTE>LOCATION y, por tanto, el núcleo codifica la relación *located_at* (en *offshore wind turbine*, wind turbine *located* offshore). La presencia superior de CN que expliciten la ubicación *offshore* frente a su antónimo *onshore* no indica que haya más aerogeneradores y parques eólicos situados en el mar, sino que esta característica debe hacerse explícita para aludir a esta ubicación. Por el contrario, si se habla de *wind turbine* o *wind farm* en términos generales, se asume que se encuentran en tierra, esto es, *onshore*, de manera que no es necesario especificar esta ubicación.

El *slot* de POSSESSOR permite indicar el concepto del que forma parte la categoría nuclear. En nuestra muestra, la subcategoría de ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT abre este *slot*, que se rellena con conceptos de la misma categoría, de forma que se combinan dos conceptos de características similares (instrumentos transformadores). La relación que codifica el núcleo es *part_of* (p. ej. en *generator rotor*, rotor *part_of* generator), por lo que se trata de CN que designan los componentes de un instrumento transformador encargados de la transformación (p. ej. *generator rotor*, *wind turbine generator*).

Por último, en nuestros CN la categoría nuclear ENTITY>CREATION abre también el *slot* de THEME (PART), que se establece cuando la categoría nuclear actúa como POSSESSOR y permite indicar sus partes o características de estas. Para facilitar su comprensión, en este *slot* concretamos el rol de THEME indicando que se refiere a las partes de la categoría nuclear: THEME (PART). En nuestros CN, es la subcategoría ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT la que abre este *slot*, que se rellena con las categorías semánticas ENTITY>MATTER>SOLID MATTER>MATERIAL (para elementos que forman parte de la ENTIDAD, p. ej. *permanent magnet* en *permanent magnet generator*) y ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT (para sus componentes encargados de la transformación, p. ej. *wound rotor* en *wound rotor induction generator*). En todos los casos, la relación que activa el núcleo es *has_part*.

Tras analizar los diferentes *slots* que abren las subcategorías de ENTITY>CREATION, se aprecia que se trata de una categoría nuclear con un elevado potencial de producción de

CN, como se infiere de los múltiples CN que cumplen con estos patrones de combinación y que se muestran seguidamente en la Tabla 41. Además, varios de estos *slots* coinciden en las diferentes subcategorías de ENTITY>CREATION, lo que indica que estas presentan patrones de combinación similares. En concreto, nos referimos al *slot* de PATIENT, que se rellena en las tres subcategorías, así como a los *slots* de INSTRUMENT, DESCRIPTIVE y LOCATION, también presentes en la práctica totalidad de las subcategorías. Por tanto, determinamos que estos *slots* constituyen la «estructura argumental» o *microcontexto* de la categoría nuclear ENTITY>CREATION.

A este respecto, es necesario recordar que nuestro estudio describe los CN que forman parte de la muestra analizada; es decir, el hecho de que un *slot* no figure en alguna de las subcategorías de ENTITY>CREATION no significa necesariamente que esa combinación conceptual no sea posible, sino que no se ha producido en los CN que estudiamos. Por tanto, consideramos que la ampliación de la muestra en próximos trabajos podrá confirmar la presencia de determinados *slots* en subcategorías que no los han establecido en este estudio, además de revelar nuevos *slots* posibles. En este sentido, en nuestra extracción de CN de la categoría ENTITY>CREATION se aprecia un elevado número de CN de la subcategoría ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT y, por tanto, mayores posibilidades de que los *slots* se den en esta subcategoría. Sin embargo, esto no implica que esos *slots* sean específicos de dicha subcategoría, como hemos señalado, sino que esta cuenta con una mayor presencia en nuestro corpus.

En la Tabla 41 se presentan todos los CN cuyo núcleo pertenece a la categoría ENTITY>CREATION, agrupados según el *slot* que rellenan. No obstante, se observará que los CN de más de dos formantes pueden combinar varios *slots* (en *large wind farm*, *farm* rellena primero el *slot* de INSTRUMENT con *wind* y, después, el de DESCRIPTIVE con *large*), una propiedad en la que profundizaremos tras el análisis de la apertura de *slots*.

ENTITY>CREATION		
Slot	**Todos los CN**	
PATIENT	AC generator DC generator electric generator electrical generator electricity generator power generator	voltage source converter voltage source inverter power converter power grid power plant power station
INSTRUMENT	diesel generator renewable generator steam generator wind generator induction generator power electronic converter gas turbine	wind turbine impulse turbine reaction turbine wind farm wind park wind power plant wind plant

	hydraulic turbine	
DESCRIPTIVE	asynchronous generator synchronous generator variable speed generator conventional generator conventional synchronous generator squirrel cage induction generator doubly fed induction generator self-excited induction generator direct drive generator distributed generator switched reluctance generator wind turbine generator constant-speed wind turbine	fixed-speed wind turbine variable-speed wind turbine conventional wind turbine downwind turbine floating turbine horizontal axis wind turbine large wind turbine modern wind turbine small wind turbine vertical axis wind turbine stall-regulated wind turbine commercial wind turbine large wind farm
LOCATION	offshore turbine offshore wind turbine	offshore wind farm onshore wind farm
POSSESSOR	generator rotor turbine rotor wind turbine rotor	turbine generator wind turbine generator generator of wind turbine
THEME (PART)	permanent magnet generator permanent magnet synchronous generator	three-phase generator wound rotor induction generator

Tabla 41: *Slots* que abren las subcategorías de ENTITY>CREATION y CN con esta categoría nuclear

- **ENTITY>SYSTEM**

A continuación, presentamos los *slots* que conforman el microcontexto de la categoría nuclear ENTITY>SYSTEM, que genera CN como *wind power generation system* o *wind turbine system* (Tabla 42). A pesar de que esta categoría no cuenta con subcategorías, como en el resto de ejemplos de nuestro análisis, resulta interesante por su paralelismo con la categoría anterior de ENTITY>CREATION, como comentaremos a continuación.

ENTITY>SYSTEM			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
PATIENT	PROCESS>MOVEMENT>ENERGY MOVEMENT	*causes* *affects*	power system power system
INSTRUMENT	PROCESS>MOVEMENT	*uses_resource*	wind power generation system wind energy system wind system
DESCRIPTIVE	ENTITY>CREATION>ARTIFACT >INSTRUMENT> TRANSFORMING INSTRUMENT PROCESS>MOVEMENT>WIND MOVEMENT	*has_function* *has_type*	generator system wind system
POSSESSOR	ENTITY>CREATION>ARTIFACT >INSTRUMENT>TRANSFORMIN G INSTRUMENT	*part_of*	wind turbine system

Tabla 42: *Slots* que abre la categoría ENTITY>SYSTEM

La categoría ENTITY>SYSTEM abre los *slots* de PATIENT, INSTRUMENT, DESCRIPTIVE y POSSESSOR. El *slot* de PATIENT alude al concepto que sufre la acción de ENTITY>SYSTEM, que se rellena con la categoría PROCESS>MOVEMENT>ENERGY MOVEMENT, ya que se da en el CN *power system*. Se trata de un término polisémico que puede codificar dos relaciones diferentes y, con ello, dos concepciones diferentes del *slot* de PATIENT. Por una parte, se da la proposición system *causes* power, de manera que PATIENT es la energía eléctrica resultante de la acción de *power system*, que se entiende como un instrumento transformador (es una variante de *generador*). En el otro sentido de *power system*, la proposición conceptual es system *affects* power, ya que *power system* se entiende como un conducto (es una variante de *red eléctrica*) y, por tanto, transporta al PACIENTE, entendiéndose este como el proceso sujeto a la acción del núcleo.

Otro de los *slots* que abre ENTITY>SYSTEM es el de INSTRUMENT, que permite especificar el recurso que emplea el sistema para llevar a cabo su actividad. Este *slot* se rellena con la categoría PROCESS>MOVEMENT, que puede aludir al viento o, por metonimia, a su energía cinética. La relación que codifica es, como viene siendo habitual cuando se abre este *slot*, *uses_resource*.

Por otro lado, mediante el *slot* de DESCRIPTIVE se pueden precisar características del núcleo. Dado el carácter genérico de la categoría nuclear, las categorías que rellenan este *slot* adquieren valores más específicos que permiten delimitar el concepto al que alude el CN. Por ejemplo, en *generator system*, el *slot* de DESCRIPTIVE se rellena con la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, que especifica la acción para la que está destinado el concepto del núcleo (system *has_function* generator) y, por tanto, indica que *generator system* es una variante de *generator*. Por su parte, en *wind system* el *slot* de DESCRIPTIVE se rellena con la categoría PROCESS>MOVEMENT>WIND MOVEMENT, que indica el tipo de sistema del que se trata (system *has_type* wind). Hablamos de la circulación atmosférica, que constituye por tanto un sistema de índole diferente a los distintos *slots* que componen este microcontexto. De este modo, el CN *wind system* en este último sentido puede considerarse idiomático, ya que su significado no puede deducirse de la suma de sus formantes, y así no encaja en el microcontexto de la categoría nuclear ENTITY>SYSTEM.

Por último, el *slot* de POSSESSOR indica el concepto al que pertenece el núcleo. Por ejemplo, en *wind turbine system*, la proposición conceptual es system *part_of* wind turbine. La categoría con la que se rellena este *slot* en nuestra muestra es ENTITY>CREATION> ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT.

De este análisis se desprende que los *slots* del microcontexto de ENTITY>SYSTEM coinciden con los que establecen las subcategorías de ENTITY>CREATION. Esto no resulta extraño si consideramos que el carácter genérico de ENTITY>SYSTEM se precisa en su combinación con los otros elementos del CN y en la relación interna entre ellos. De este modo, ENTITY>SYSTEM adquiere el valor de ENTITY>CREATION en todos los CN de nuestra muestra, salvo en el empleo idiomático de *wind system*. Por tanto, la categoría nuclear ENTITY>SYSTEM establece sus significados específicos en función de los valores con los que rellene su microcontexto.

- **ENTITY>PART**

La categoría nuclear ENTITY>PART incluye los conceptos de las siguientes subcategorías:

ENTITY>PART (*generator side*[55])

ENTITY>PART>PART OF INSTRUMENT (*turbine blade*, *wind turbine blade*)

En la Tabla 43 se presenta el *slot* que abren estas dos subcategorías en los CN analizados. Como se observa, presentan patrones de combinación similares, pues ambas activan el *slot* de POSSESSOR que, como no es de extrañar, se refiere al concepto del que forma parte la categoría nuclear ENTITY>PART. En nuestros términos, esta entidad pertenece a la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT y, dado que los CN que cumplen este patrón designan la pertenencia a otro concepto, la relación que codifica el núcleo es *part_of*. En este sentido, cabe esperar que, en una jerarquía conceptual todo-parte (p. ej. parque eólico, aerogenerador y sus componentes), la entidad superior (parque eólico) se especifique en función de sus componentes inmediatamente inferiores (aerogeneradores) y, estos, según sus componentes (partes de los aerogeneradores). Ello justificaría que el *slot* de POSSESSOR de la categoría nuclear ENTITY>PART se rellene con la categoría ENTITY>CREATION> ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, a la que pertenecen los aerogeneradores.

ENTITY>PART			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
POSSESSOR	ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT	*part_of*	generator side turbine blade wind turbine blade

[55] El concepto designado por el núcleo *side*, independientemente de su modificador, pertenece a la categoría ENTITY>PART en lugar de a su nivel más específico ENTITY>PART>PART OF INSTRUMENT porque puede constituir una parte de otra entidad que no sea necesariamente un instrumento. No obstante, al combinarse en el CN *generator side*, este concepto combinado adquiere la categoría ENTITY>PART> PART OF INSTRUMENT.

Tabla 43: *Slots* que abre la categoría ENTITY>PART

- **ENTITY>INFORMATION**

La categoría nuclear ENTITY>INFORMATION genera CN cuyos núcleos pertenecen a las siguientes subcategorías:

ENTITY>INFORMATION>REPRESENTATION>MATHEMATICAL EXPRESSION (*power coefficient*)

ENTITY>INFORMATION>REPRESENTATION>LINE (*power curve*)

ENTITY>INFORMATION>REPRESENTATION>GRAPH (*power spectrum*)

ENTITY>INFORMATION>PARAMETER (*power factor*)

En la Tabla 44 se muestra la apertura de *slots* de esta categoría nuclear que, como en el caso anterior de ENTITY>PART, destaca por su homogeneidad. En los CN analizados, ENTITY>INFORMATION activa el *slot* de PATIENT, que se rellena con la categoría ATTRIBUTE>MEASUREMENT>MAGNITUDE. El núcleo establece la relación *represents* (p. ej. en *power coefficient*, coefficient *represents* power). Conviene señalar que nombramos a este *slot* PATIENT en lugar de THEME porque el núcleo indica que puede sufrir cambios (pueden ser distintos coeficientes o factores, así como una curva o un espectro con diferentes valores). Por tanto, se observa un claro ejemplo de uniformidad en los microcontextos, en los que las diferentes subcategorías de ENTITY>INFORMATION exhiben el mismo patrón de combinación.

ENTITY>INFORMATION			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
PATIENT	ATTRIBUTE>MEASUREMENT>MAGNITUDE	*represents*	power coefficient power curve power spectrum power factor

Tabla 44: *Slots* que abre la categoría ENTITY>INFORMATION

- **ENTITY>HUMAN**

En la categoría nuclear ENTITY>HUMAN se agrupan los CN cuyos núcleos pertenecen a las siguientes subcategorías:

ENTITY>HUMAN (*turbine manufacturer*, *wind turbine manufacturer*)

ENTITY>HUMAN>INSTITUTION (*wind industry*)

En la Tabla 45 puede observarse que, en nuestros CN, esta categoría nuclear abre, por una parte, el *slot* de PATIENT para aludir al concepto producido por esa categoría nuclear. Así, el *slot* se rellena con la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT> TRANSFORMING INSTRUMENT y el núcleo codifica la relación *causes* (manufacturer *causes* turbine). Observamos también que el núcleo explicita la acción codificada en la relación (*manufacture*), mediante el procedimiento de nominalización del predicado habitual en los CN. Por tanto, este *slot* muestra la concepción de ENTITY>HUMAN en nuestro corpus de energía eólica como un agente que produce los instrumentos necesarios para la generación de energía eólica.

Por otra parte, ENTITY>HUMAN en el CN *wind industry* abre el *slot* de INSTRUMENT, en un paralelismo con los instrumentos transformadores que utilizan el viento. De este modo, rellena el *slot* con la categoría PROCESS>MOVEMENT>WIND MOVEMENT y codifica la relación *uses_resource*. A la vista de este microcontexto de ENTITY>HUMAN, no debemos olvidar que estos *slots* se activan por parte de dicha categoría nuclear en un corpus de energía eólica como el que utilizamos en este estudio, aunque evidentemente la categoría ENTITY>HUMAN es lo suficientemente amplia como para rellenar sus *slots* con otros valores y establecer *slots* diferentes en las distintas áreas en las que pueda analizarse.

ENTITY>HUMAN			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
PATIENT	ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	*causes*	turbine manufacturer wind turbine manufacturer
INSTRUMENT	PROCESS>MOVEMENT>WIND MOVEMENT	*uses_resource*	wind industry

Tabla 45: *Slots* que abre la categoría ENTITY>HUMAN

- **ATTRIBUTE**

Tras haber detallado la apertura de *slots* por parte de las ENTIDADES, nos detendremos a continuación en la combinatoria de los ATRIBUTOS. Así, en la categoría genérica de ATTRIBUTE se comprende gran parte de nuestros CN, cuyos núcleos pertenecen a las siguientes subcategorías:

ATTRIBUTE (*power quality*, *power system reliability*)

ATTRIBUTE>PHYSICAL ATTRIBUTE>COMPOSITION (*generator design*, *wind turbine model*, etc.)

ATTRIBUTE>PHYSICAL ATTRIBUTE>STATE (*voltage stability*, *wind condition*)

ATTRIBUTE>PHYSICAL ATTRIBUTE>SIZE (*turbine size*)

ATTRIBUTE>MEASUREMENT>MAGNITUDE (*generator terminal voltage*, *apparent power*, etc.)

ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL (*voltage level*, *turbine hub height*)

ATTRIBUTE>LOCATION (*power output*)

ATTRIBUTE>DIRECTION (*wind direction*)

La Tabla 46 recoge los *slots* que abren estas subcategorías: POSSESSOR, DESCRIPTIVE y AGENT.

ATTRIBUTE			
***Slot* (rol)**	**Categoría**	**Relación**	**Ejemplos**
POSSESSOR	ENTITY>CREATION>ARTIFACT ENTITY>PART>PART OF INSTRUMENT ENTITY>MATTER>SOLID MATTER>MATERIAL ENTITY>SPACE>POSITION ENTITY>SYSTEM PROCESS>MOVEMENT ATTRIBUTE	*attribute_of* *has_attribute*	turbine hub height line-to-line voltage
DESCRIPTIVE	ATTRIBUTE ENTITY>INFORMATION>PARAMETER	*has_attribute* *has_function*	instantaneous voltage mean wind speed
AGENT	ENTITY>CREATION>ARTIFACT	*caused_by*	generator power power output of a wind turbine

Tabla 46: *Slots* que abren las subcategorías de ATTRIBUTE

El *slot* de POSSESSOR es la combinación por excelencia de esta categoría nuclear, lo que no resulta extraño, ya que es habitual que los atributos acompañen a los conceptos de los que son propiedades. Así, este *slot* se rellena con diferentes categorías, ya que podría decirse que cualquier concepto puede tener atributos. En nuestra muestra, estas categorías son las siguientes:

ENTITY>CREATION>ARTIFACT

ENTITY>PART>PART OF INSTRUMENT

ENTITY>MATTER>SOLID MATTER>MATERIAL

ENTITY>SPACE>POSITION

ENTITY>SYSTEM

PROCESS>MOVEMENT

ATTRIBUTE

Así, la categoría nuclear ATTRIBUTE activa la relación *attribute_of* con todos sus modificadores que rellenan el rol de POSSESSOR en nuestros CN. Se trata, por tanto, de un *slot* bastante homogéneo, a pesar de las diferentes categorías semánticas que pueden rellenarlo. Además, las diferentes subcategorías comprendidas en ATTRIBUTE abren este *slot*, excepto la subcategoría de ATTRIBUTE>LOCATION. Esto puede justificarse por el término incluido en esta subcategoría, *power output*, un CN idiomático cuyo sentido puede inferirse a partir del orden inverso de los elementos (output *has_attribute* power). Es decir, a pesar de que el núcleo presente la categoría ATTRIBUTE>LOCATION, el concepto del CN no alude a un lugar, sino a una magnitud: la potencia de salida (se trata, así, de una variante de *output power*).

Por otro lado, el *slot* de DESCRIPTIVE permite especificar valores del atributo. Se rellena por regla general con la categoría ATTRIBUTE, que comprende atributos de diferentes tipos (origen, tiempo, medida, etc.), aunque en el caso del CN *reference voltage* la categoría que lo ocupa es ENTITY>INFORMATION>PARAMETER. El núcleo activa la relación *has_attribute* en todos los casos, excepto en *reference voltage*, cuya relación es *has_function*. El *slot* de DESCRIPTIVE se da en las subcategorías ATTRIBUTE>MEASUREMENT> MAGNITUDE y ATTRIBUTE>LOCATION, aunque su estudio en un conjunto superior de CN probablemente revelaría su presencia en las otras subcategorías, pues es habitual que los atributos se especifiquen con sus valores.

Por último, el *slot* de AGENT indica el concepto que ha dado lugar al atributo. Cabe señalar que el atributo no se crea en sí, pues no es un concepto independiente, sino que se nombra al atributo por metonimia, refiriéndose a ese atributo y al concepto que lo posee (p. ej. *generator power* alude a la potencia producida por un generador, en otras palabras, a la potencia de la energía producida por el generador). Este *slot* se rellena con la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT en los CN analizados, cuyo núcleo codifica la relación *caused_by* (power *caused_by* generator). En nuestra muestra, este *slot* lo abrió la subcategoría ATTRIBUTE>MEASUREMENT>MAGNITUDE y, dadas las restricciones semánticas de la proposición subyacente (ATTRIBUTE *caused_by* AGENT), no creemos que pueda pertenecer al microcontexto de la categoría general ATTRIBUTE, pues habrá ATRIBUTOS que no resulten de la acción de un AGENTE.

En definitiva, el microcontexto de la categoría nuclear ATTRIBUTE se compone de los *slots* de POSSESSOR (concepto que posee el ATRIBUTO) y DESCRIPTIVE (valor del ATRIBUTO), que

son habituales en varias subcategorías. En la Tabla 47 se aprecia lo prolífero de la categoría nuclear ATTRIBUTE, que genera un elevado número de CN que cumplen con el patrón de combinación indicado.

ATTRIBUTE		
Slot	**Todos los CN**	
POSSESSOR	power quality	direct voltage
	power system reliability	generator terminal voltage
	generator design	generator voltage
	turbine design	grid voltage
	turbine model	input voltage
	wind turbine design	line voltage
	wind turbine model	network voltage
	voltage stability	output voltage
	wind condition	PCC voltage
	turbine size	voltage at the PCC
	power density	phase voltage
	generator speed	rotor voltage
	wind speed	stator voltage
	wind velocity	system voltage
	wind power capacity	terminal voltage
	voltage level	line-to-line voltage
	turbine hub height	line-to-neutral voltage
	wind direction	offshore wind power
	AC voltage	output power
	battery voltage	wave power
	bus voltage	wind power
	DC voltage	offshore wind energy
	DC-link voltage	wind energy
DESCRIPTIVE	nominal voltage	rated power
	reference voltage	reactive power
	instantaneous voltage	real power
	high voltage	maximum power
	low voltage	average wind speed
	medium voltage	high wind speed
	induced voltage	low wind speed
	rated voltage	mean wind speed
	harmonic voltage	rated wind speed
	active power	power output
	apparent power	
AGENT	generator power	power output of a wind turbine

Tabla 47: *Slots* que abren las subcategorías de ATTRIBUTE y CN con esta categoría nuclear

Finalmente, nos detendremos en la combinatoria de la categoría ontológica PROCESS, en la que exploramos los *slots* que abren las categorías nucleares generales PROCESS>CHANGE y PROCESS>MOVEMENT.

- **PROCESS>CHANGE**

La categoría nuclear PROCESS>CHANGE comprende las subcategorías PROCESS>CHANGE, que produce CN como *impact of wind power*, y PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY, a

partir de la cual se generan CN como *generator torque control* o *voltage drop*. La Tabla 48 recoge los *slots* que abre esta categoría nuclear.

PROCESS>CHANGE			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
PATIENT	ENTITY>FORCE>STRESS ATTRIBUTE>MEASUREMENT>MAGNITUDE PROCESS>MOVEMENT>ENERGY MOVEMENT	*affects*	generator torque control voltage control voltage dip voltage drop voltage fluctuation voltage sag voltage variation power fluctuation wind power fluctuation
AGENT	PROCESS>MOVEMENT>ENERGY MOVEMENT	*caused_by*	impact of wind power

Tabla 48: *Slots* que abre la categoría PROCESS>CHANGE

Esta categoría nuclear abre dos *slots*, que permiten identificar el AGENTE y el PACIENTE del proceso. En nuestros CN, cada *slot* se establece por parte de una de las subcategorías: PROCESS>CHANGE rellena el *slot* de AGENT y PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY, el de PATIENT. Sin embargo, esto puede deberse al reducido número de formas que componen las subcategorías en cuestión, en concreto la de PROCESS>CHANGE, que solo cuenta con un CN en nuestra muestra. No obstante, consideramos que ambos conforman el microcontexto de la categoría nuclear PROCESS>CHANGE, a la vista de otros términos de nuestro corpus que rellenan esos *slots* (p. ej. *impact* activa el *slot* de AGENT en nuestra muestra, aunque en CN como *environmental impact* se observa que también puede abrir el *slot* de PATIENT).

Así pues, en los CN con los que trabajamos, el *slot* de PATIENT se rellena con las categorías ENTITY>FORCE>STRESS, ATTRIBUTE>MEASUREMENT>MAGNITUDE y PROCESS> MOVEMENT>ENERGY MOVEMENT, con las que establece la relación *affects* (p. ej. control *affects* generator torque). Como puede observarse en el núcleo de los CN, las nominalizaciones expresan el modo concreto en que el PACIENTE resulta afectado (se controla, fluctúa, cae, etc.). Por su parte, el *slot* de AGENT se rellena en nuestra muestra con la categoría PROCESS>MOVEMENT>ENERGY MOVEMENT. Así, el proceso resulta causado por (*caused_by*) este AGENTE (impact *caused_by* wind power). En resumen, la categoría nuclear de PROCESS>CHANGE puede especificarse según el AGENTE que realiza el proceso y el PACIENTE que lo sufre.

- **PROCESS>MOVEMENT**

Por último, en la categoría nuclear PROCESS>MOVEMENT se comprenden las subcategorías PROCESS>MOVEMENT>ENERGY MOVEMENT, que genera CN como *electric power* o *wind energy*, y PROCESS>MOVEMENT>WIND MOVEMENT, que da lugar a CN del tipo de *offshore wind* o *wind shear*. En la Tabla 49 se indican los *slots* que abren estas subcategorías.

PROCESS>MOVEMENT			
***Slot* (rol)**	**Categoría**	**Relación**	**Todos los CN**
DESCRIPTIVE	ATTRIBUTE	*has_attribute*	electric power electrical power mechanical power generated power high wind
LOCATION	ATTRIBUTE>LOCATION	*takes_place_in*	offshore wind power offshore wind energy offshore wind
AGENT	PROCESS>MOVEMENT	*effected_by*	nuclear power wave power wind power wind energy renewable power
PATIENT	PROCESS>MOVEMENT>WIND MOVEMENT	*affects*	wind shear

Tabla 49: *Slots* que abre la categoría PROCESS>MOVEMENT

Las subcategorías de PROCESS>MOVEMENT abren los *slots* de DESCRIPTIVE, LOCATION, AGENT y PATIENT. Por una parte, el *slot* de DESCRIPTIVE se da en las dos subcategorías y permite especificar características del proceso. Así pues, se rellena con la categoría ATTRIBUTE (de diferente tipo: tiempo, medida, etc.). En todos los casos, el núcleo codifica la relación *has_attribute*.

Por otro lado, el *slot* de LOCATION también se activó en las dos subcategorías y se utiliza para indicar dónde tiene lugar el proceso. Por tanto, la relación que codifica el núcleo es *takes_place_in* (wind power *takes_place* offshore) y el *slot* se rellena con la categoría ATTRIBUTE>LOCATION. Como indicamos anteriormente, la mayor presencia de la ubicación *offshore* en los CN no significa una mayor existencia real, sino la necesidad de especificar esta característica distintiva.

El *slot* de AGENT, en cambio, no forma parte del microcontexto común de PROCESS>MOVEMENT, ya que solo se da en la subcategoría PROCESS>MOVEMENT>ENERGY

MOVEMENT, algo que no resulta extraño si consideramos que en la generación de la energía sí pueden intervenir agentes, algo que no sucede en el caso del viento. Así, este *slot* se rellena con la categoría PROCESS>MOVEMENT y el núcleo codifica la relación *effected_by* (power *effected_by* wind).

Finalmente, el *slot* de PATIENT también acompaña en nuestra muestra a solo una de las subcategorías de PROCESS>MOVEMENT, en concreto, la subcategoría PROCESS>MOVEMENT> WIND MOVEMENT. Este *slot* lo ocupa la categoría PROCESS>MOVEMENT>WIND MOVEMENT y, en los CN en los que se rellena, la relación interna es *affects* (shear *affects* wind). Como en casos anteriores, sería interesante comprobar en próximos estudios si, a diferencia del *slot* de AGENT, que es exclusivo de una subcategoría, este *slot* de PATIENT puede generar CN también en la otra subcategoría nuclear (PROCESS>MOVEMENT>ENERGY MOVEMENT). Por el momento, determinamos que el microcontexto de la categoría nuclear general PROCESS>MOVEMENT está formado por el *slot* de DESCRIPTIVE y el *slot* de LOCATION, que es el patrón que coincide en las dos subcategorías.

5.1.2.2 Conclusiones de los microcontextos

Los CN son términos compuestos en los que convergen dos o más conceptos. Ahora bien, estos no se combinan de forma aleatoria, sino que su unión responde a restricciones semánticas. Así, el análisis de la combinatoria de las categorías nucleares (las categorías que ocupan la posición del núcleo en los CN) reveló que los conceptos o términos que pertenecen a una determinada categoría nuclear presentan patrones de combinación similares. Esto es, se combinan con conceptos de determinados roles y categorías semánticas, con los que se vinculan mediante ciertas relaciones.

Esta similitud en la combinatoria de los términos compuestos nos llevó a delimitar nuestra noción de *microcontexto*, que da cuenta de la productividad en la formación de los CN. Los microcontextos constituyen una especie de estructura argumental del núcleo. Es decir, las categorías semánticas nucleares abren una serie de *slots* en función de su significado. Estos *slots* determinan los CN que pueden formarse para especificar esa categoría nuclear, en la que pueden englobarse diferentes términos o incluso diferentes subcategorías más específicas, como hemos presentado en nuestro análisis.

Estos *slots* se nombran en función de los roles semánticos que puedan desempeñar sus conceptos en la proposición subyacente en cada CN (p. ej. PATIENT, INSTRUMENT). No se utilizan las categorías para delimitar los *slots*, ya que una misma categoría puede guardar relaciones diferentes con el núcleo. Para ilustrar la importancia de los roles en el establecimiento de los *slots*, se muestran las proposiciones conceptuales que subyacen en

los siguientes CN, pertenecientes a la categoría nuclear ENTITY>CREATION: *AC generator, voltage source converter, steam generator, vertical axis wind turbine* y *onshore wind farm*.

-*AC generator*: generator *causes* AC

-*voltage source converter*: converter *affects* voltage source

-*steam generator*: generator *uses_resource* steam

-*vertical axis wind turbine*: wind turbine *has_attribute* vertical axis

-*onshore wind farm*: wind farm *located* onshore.

Del análisis de las proposiciones subyacentes se infiere que los modificadores que acompañan al núcleo desempeñan los siguientes roles semánticos en cada caso: PATIENT (*AC* en *AC generator* y *voltage source* en *voltage source converter*), INSTRUMENT (*steam* en *steam generator*), DESCRIPTIVE (*vertical axis* en *vertical axis wind turbine*) y LOCATION (*onshore* en *onshore wind farm*). Por tanto, estos roles se utilizan para nombrar los *slots* que conforman el microcontexto de ENTITY>CREATION, ya que son los *slots* que coinciden en las diferentes subcategorías que incluye. Así, los términos de la categoría ENTITY>CREATION pueden especificarse en función del PACIENTE que reciba su acción, el INSTRUMENTO que utilice, sus rasgos descriptivos (DESCRIPTIVE) o su UBICACIÓN. Estos *slots* constituyen su microcontexto.

La productividad y sistematicidad de los microcontextos se aprecia en el hecho de que estos *slots* coinciden en los diferentes términos o subcategorías comprendidos en una categoría nuclear. Por ejemplo, esos cuatro *slots* han permitido la generación de 57 CN de la categoría ENTITY>CREATION en nuestra muestra.

En esta línea, los *slots* se rellenan con determinadas categorías semánticas y suelen establecer las mismas relaciones. Por ejemplo, el *slot* de PATIENT que abre ENTITY>CREATION se rellena con la categoría PROCESS>MOVEMENT>ENERGY MOVEMENT y la relación interna de los CN es *causes* o *affects* (generator *causes* AC, converter *affects* voltage source). El *slot* de INSTRUMENT, por su parte, se rellena con las categorías ENTITY>MATTER, ENTITY>FORCE y PROCESS>MOVEMENT y la relación que codifica el núcleo es *uses_resource* en todos los casos (en *steam generator*, generator *uses_resource* steam; en *wind farm*, farm *uses_resource* wind, etc.). El *slot* de DESCRIPTIVE se rellena con las categorías ATTRIBUTE y ENTITY> CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, y suele establecer la relación *has_attribute* (en *vertical axis wind turbine*, wind turbine *has_attribute* vertical axis; en *large wind farm*, wind farm *has_attribute* large, etc.). Por último, el *slot* de LOCATION se

rellena con la categoría ATTRIBUTE>LOCATION y la relación interna es *located_at* (en *offshore wind turbine*, wind turbine *located* offshore; en *onshore wind farm*, wind farm *located* onshore, etc.).

En definitiva, con un limitado número de roles, categorías y relaciones, se ha generado un gran número de términos compuestos, lo que confirma la hipótesis que sostiene que los conceptos similares presentan esquemas de combinación semejantes. Sin embargo, ello no implica que estas sean las únicas formas en las que un núcleo pueda especificarse, pero en ese caso no estaríamos hablando de generalizaciones a nivel de categoría nuclear que puedan aplicarse a los diferentes conceptos incluidos.

Por tanto, nuestra propuesta de microcontextos integra de forma sistemática los tres elementos básicos con los que se caracterizan las proposiciones conceptuales que subyacen en los CN: roles, categorías y relaciones semánticas. Así, a diferencia de los postulados de Gagné y Shoben (1997) y Gagné (2000, 2001), no consideramos que los CN se formen únicamente a partir del establecimiento de relaciones entre dos conceptos, sino que interviene un proceso más complejo en el que interactúan los roles y las categorías semánticas de los *slots*, así como las relaciones que los unen y que indican cómo debe entenderse la combinación conceptual. Se trata, por tanto, de un enfoque renovado del clásico procedimiento de *slot filling* que refleja el complejo mecanismo conceptual que da lugar a los términos compuestos. En definitiva, los microcontextos delimitan el poder combinatorio de las categorías nucleares y, por tanto, pueden entenderse como perfiles semánticos de la categoría en cuestión.

A pesar de que nuestro análisis reveló algunos CN que escapaban a estos esquemas de combinación debido a su carácter idiomático (p. ej. *wind system* en uno de sus significados o *power output*), en general se observa que la productividad de los microcontextos presenta implicaciones conceptuales a distintos niveles.

Una de sus utilidades es la recreación del sistema conceptual subyacente. Dado que, para acceder a los microcontextos, se analizan las proposiciones conceptuales que subyacen en los CN, se obtiene una vista mucho más profunda de los conceptos que se relacionan en el sistema conceptual en el que nos encontremos. Por ejemplo, analizando cómo se rellenan los *slots* que abre el núcleo *turbine*, detectamos algunas de las relaciones que activa con otros conceptos (Figura 20). Algunos de los CN que permiten este análisis son *gas turbine, hydraulic turbine, wind turbine* (relación *uses_resource*), *fixed-speed wind turbine, downwind turbine* (relación *has_attribute*), *onshore turbine, offshore turbine* (relación *located_at*), *turbine rotor* y *wind turbine generator* (relación *has_part*).

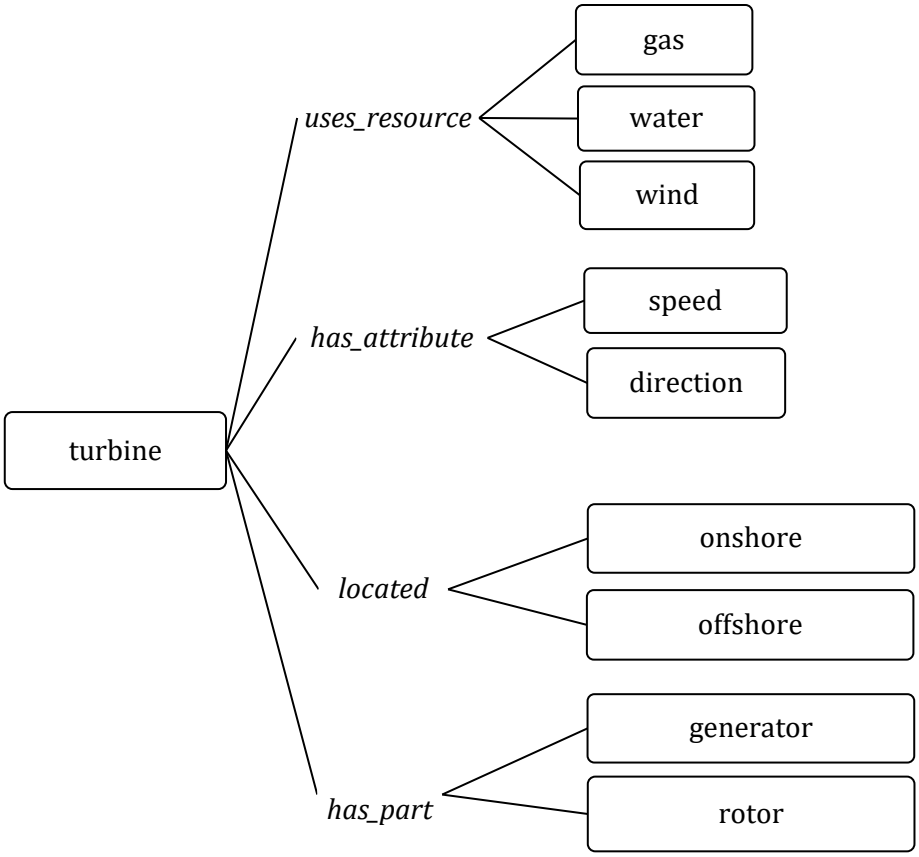

Figura 20: Parte de las relaciones de *turbine* con otros conceptos, extraídas a partir del análisis de sus CN

En la Figura 20 se hace patente una de las principales complejidades de los CN: la omisión de su relación interna, que da lugar a CN de forma similar pero de significado distante. Sin alejarnos de la utilidad de los microcontextos para reconstruir el sistema conceptual, constatamos también que el estudio de los diferentes valores que ocupen los *slots* puede revelar conceptos relacionados. Por ejemplo, cuando el *slot* INSTRUMENT de la categoría nuclear ENTITY>CREATION se rellena con valores como *diesel, steam* o *wind*, se forman cohipónimos como *diesel generator, steam generator* o *wind generator*.

Esta propiedad da lugar a un gran número de conjuntos de términos, como la dicotomía entre *synchronous generator* y *asynchronous generator* cuando se especifica el *slot* DESCRIPTIVE de ENTITY>CREATION o los términos *constant-speed wind turbine, fixed-speed wind turbine* y *variable speed wind turbine* que también surgen al rellenar este *slot*. Otras de las formaciones habituales son, por ejemplo, la dualidad entre *offshore* y *onshore* cuando se especifica el *slot* de LOCATION (*offshore/onshore wind turbine*, *offshore/onshore wind farm*), los términos *AC voltage* y *DC voltage*, que completan el *slot* de POSSESSOR abierto por la categoría nuclear ATTRIBUTE, o los CN *high voltage, medium voltage, low*

voltage, que especifican el rol DESCRIPTIVE de esta última categoría nuclear. Estos son ejemplos de la estrecha relación entre los microcontextos y la productividad de los CN.

Al inicio mencionamos la utilidad de las definiciones para delimitar los *slots* que abren los conceptos. Por ejemplo, la definición de *coefficient* en el diccionario Merriam-Webster es «a number that serves as a measure of some property or characteristic (as of a substance, device, or process)». De esta definición puede deducirse que *coefficient* se podrá especificar en función de la propiedad o característica que mide o representa (p. ej. *power coefficient, lift coefficient, drag coefficient*).

Sin embargo, al realizar el estudio de los microcontextos observamos una doble vertiente en esta relación que existe entre las definiciones y los microcontextos. Así, además de poder inferirse los microcontextos a partir de las definiciones de los conceptos, el análisis de los CN que se generan a partir de un núcleo permite también delimitar las características que podrán formar parte de su definición. Por ejemplo, algunos de los CN que se forman a partir de *plant* en el ámbito de la energía eléctrica son *power plant, wind power plant, wind plant, hydropower plant* o *thermal power plant*, en los que el núcleo *plant* pertenece a la categoría de ENTITY>CREATION>STRUCTURE y se acompaña de los *slots* de PATIENT (*power*) e INSTRUMENT (*wind, hydro-, thermal*). Por tanto, su definición podrá elaborarse utilizando estos datos, por ejemplo, del siguiente modo:

plant: structure that converts some form of energy into electrical energy.

Por tanto, los microcontextos y las definiciones son entes conceptuales directamente relacionados, hasta tal punto que la productividad de los términos compuestos puede ser una herramienta útil en tareas terminográficas como la confección de definiciones.

Por otra parte, los CN pueden surgir al rellenar un *slot* o varios de estos a la vez. Esto sucede a menudo en los CN de tres o más formantes, como *wound rotor induction generator*, donde el núcleo *generator* activa en primer lugar el *slot* de INSTRUMENT (*induction*) y, seguidamente, el *slot* de THEME (PART) (*wound rotor*). Este tipo de CN que permiten explicitar varios *slots* del microcontexto son, necesariamente, más extensos y complejos conceptualmente, pues su significado global se deriva de un mayor número de componentes. Como suele suceder en los sintagmas, el rasgo más importante del microcontexto sigue inmediatamente al núcleo y el resto se van alejando de este a medida que son menos relevantes en su microcontexto.

Para que el núcleo pueda explicitar varios *slots* en un CN, debe darse una estructura de *bracketing* a la derecha o de doble *bracketing*, en el caso de los CN de cuatro o más formantes. De este modo, el núcleo podrá indicar en primer lugar un elemento del microcontexto, unirse a él y, a continuación, indicar el resto de *slots*. Por ejemplo, en *rated* [*wind speed*], *speed* rellena primero el *slot* de POSSESSOR (*wind*) y, después, el *slot* de DESCRIPTIVE (*rated*). Esto se observa también en variantes en las que solo se rellena el segundo *slot* y que responden a la elisión léxica (*rated speed*). Otro ejemplo se observa en [*stall-regulated*] [*wind turbine*], donde *turbine* completa en primer lugar el *slot* de INSTRUMENT (*wind*) y, después, el *slot* de DESCRIPTIVE (*stall-regulated*). De nuevo, pueden darse variantes reducidas en las que solo se completa el segundo *slot* (*stall-regulated turbine*).

Por el contrario, cuando se da una estructura de *bracketing* a la izquierda, esos elementos se agrupan y, por lo general, no permiten su activación individual por parte del núcleo. Por ejemplo, en [*switched reluctance*] *generator*, el núcleo *generator* activa el *slot* de DESCRIPTIVE (*reluctance*) que, a su vez, especifica también sus rasgos mediante el rol DESCRIPTIVE (*switched*). Sin embargo, es habitual el recurso a mecanismos como la metonimia, que hacen que en estas estructuras de *bracketing* a la izquierda también se puedan crear variantes por elisión léxica, tomando el todo por la parte ([*wind energy*] *system*, *wind system*). Así, observamos que la alusión a diferentes *slots* es frecuente entre las variantes denominativas de un mismo concepto (p. ej. *power plant, generating plant* y *electric power plant*).

La inclusión de varios *slots* en los CN recuerda a la multidimensionalidad, pues la especificación de varias características implica diferentes perspectivas desde las que puede abordarse el concepto nuclear. Por ejemplo, según el recurso que utilicen, las turbinas pueden ser *wind turbines, hydraulic turbines, gas turbines,* etc. Según rasgos descriptivos como el eje de rotación o el mecanismo de regulación del movimiento, se distingue entre *horizontal axis turbines* y *vertical axis turbines*, o *variable-pitch turbines* y *stall-regulated turbines*, respectivamente. Estas dimensiones, además de otras, pueden converger en CN como *stall-regulated horizontal axis wind turbine*, de una especificidad semántica y exigencia cognitiva elevadas, lo que deriva en un menor uso de estas secuencias en su forma completa. No obstante, son habituales entre los especialistas, pues permiten designar conceptos neológicos mediante la aglutinación de sus rasgos. Sin embargo, debido a su longitud, suelen emplearlas en forma de siglas, como comentaremos más adelante. Por tanto, la multidimensionalidad encuentra en los CN un escenario propicio para desenvolverse.

Por último, como hemos indicado a lo largo del análisis, nos gustaría subrayar que nuestro estudio de los microcontextos describe la muestra analizada, por lo que su ampliación en próximos estudios podrá determinar si existen nuevos *slots* que intervengan en el microcontexto de una categoría nuclear, así como nuevas categorías que puedan rellenarlos. En este sentido, el estudio de los microcontextos de las categorías nucleares en disciplinas distintas de la energía eólica entrañará, sin duda, la intervención de categorías semánticas diferentes en los *slots*, aunque sería interesante explorar si los distintos *slots* se mantienen.

Así, tras haber profundizado en la formación de los CN mediante su análisis semántico y, con ello, en sus microcontextos, en la Sección 5.2.2 a continuación ahondaremos en un aspecto poco abordado hasta el momento: las implicaciones de los *slots* del microcontexto en la traducción. Después, plasmaremos los resultados del análisis de la formación y la traducción de los CN en nuestra propuesta de representación de estos términos (Sección 5.3).

5.2 La traducción de compuestos nominales

Para favorecer la comunicación internacional, en áreas como la investigación o el comercio (por solo citar algunas), resulta fundamental traducir los textos científico-técnicos y, con ellos, trasladar los CN que los integren a la lengua meta. Sin embargo, las diversas posibilidades de combinación que pueden dar lugar a un CN, unidas a su frecuente carácter neológico y al elevado grado de especificidad que a menudo presentan, hacen que estos no siempre figuren en los recursos terminográficos o no se incluyan de forma satisfactoria para el usuario. Por tanto, resulta esencial manejar las técnicas para identificar o producir términos equivalentes, lo que reviste dificultad debido a las características propias de los CN en las distintas lenguas. En la Sección 5.2.1 planteamos un procedimiento de búsqueda de equivalentes de los CN en corpus para, a continuación, en la Sección 5.2.2, realizar un estudio comparativo de estos y presentar pautas para su producción cuando no sea posible identificarlos mediante otros recursos.

5.2.1 Procedimiento de búsqueda de equivalentes

Ante la necesidad de traducción de un CN, lógicamente una de las tareas iniciales consistirá en la consulta de recursos lingüísticos en busca del término o términos correspondientes en la lengua meta. Sin embargo, como sabemos, no es extraño que esta búsqueda resulte infructuosa, ya que los CN a menudo no se incluyen en los recursos terminográficos o lo hacen de forma poco sistemática. Sea por la razón que fuere, se deberán emplear entonces otras herramientas para identificar las equivalencias.

Una de las más utilizadas es internet, dada su facilidad y rapidez de acceso, donde se pueden realizar consultas de diferente tipo hasta dar con el término buscado. Sin duda, la red ofrece todo un horizonte de posibilidades, aunque también presenta desventajas, como la escasa fiabilidad de muchas fuentes, además de la ingente cantidad de información o la volatilidad de la misma. Esto no significa que no pueda emplearse; al contrario, consideramos que constituye un valioso recurso complementario. No obstante, la actividad terminológica debe reposar sobre una muestra bien delimitada como es un corpus. Además, estas colecciones de texto permiten realizar consultas mucho más específicas con las que la búsqueda de equivalentes será sin duda más efectiva.

También es cierto que se dispone de grandes avances para la identificación de equivalentes realizados desde la Lingüística Computacional. Sin embargo, los lingüistas, como terminólogos y traductores, a menudo no utilizan sistemas de procesamiento del lenguaje natural. Por tanto, nos parece necesario desarrollar un método de identificación de equivalentes de CN en corpus, basado en las características de estos términos compuestos, que puedan aplicar los especialistas lingüísticos en su labor diaria. Se trata de un procedimiento que probablemente subyace en la búsqueda de equivalentes que estos realizan. No obstante, consideramos necesario plasmarlo paso por paso, para indicar su utilidad en la traducción de términos compuestos, además de contribuir a desarrollar las técnicas de interrogación de corpus de cara a los traductores y terminólogos. Además, en línea con L'Homme et al. (2018), defendemos los enfoques manuales como el que proponemos por el valor añadido que supone el trabajo manual de los terminólogos o traductores, ya que las contribuciones computacionales siempre requieren de la validación por parte de estos especialistas.

Por tanto, el procedimiento que aquí presentamos se aplica a las necesidades concretas de un lingüista, al ser este el perfil que generalmente hace uso de los corpus. Es decir, no planteamos este procedimiento con vistas a la alimentación de recursos a gran escala, ya que el proceso realizado en ese caso podría ser diferente. Como muchas otras técnicas de localización de equivalentes, el proceso que presentamos se basa en la semántica distribucional, según la cual el significado de un término puede percibirse a partir de su distribución en un conjunto de contextos. De esta manera, la coincidencia de contextos indica la coincidencia semántica, tanto entre variantes en una misma lengua como entre términos correspondientes en dos lenguas. Así, en este trabajo las correspondencias se basan en la coincidencia de las características conceptuales de los posibles equivalentes, ya que los términos estudiados se integrarán posteriormente en EcoLexicon.

Para ello, partimos de la agrupación de variantes denominativas en inglés. Este paso es fundamental no solo para conocer los distintos términos de partida, sino también para facilitar la identificación de variantes en la lengua meta (por ejemplo, la especificación de diferentes dimensiones puede plasmarse también en la otra lengua). En la Figura 21 se observan dos conceptos en inglés con sus respectivas variantes denominativas. Utilizaremos estos términos para ilustrar el proceso de búsqueda, dada la especialización y similitud de los términos de ambos conceptos, que puede generar confusión en la comprensión y en la identificación del equivalente en cada caso. En concreto, se observa que *line voltage* y *phase voltage* no son sinónimos, aunque esta opción podría ser perfectamente posible, ya que *line-to-line voltage* y *phase-to-phase voltage* sí resultan ser variantes denominativas (de *line voltage*, para ser exactos). Por tanto, será necesario evaluar la situación en español, para así identificar los equivalentes correctos en cada caso.

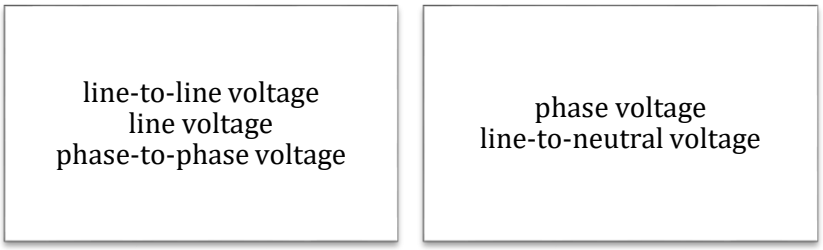

Figura 21: Variantes denominativas de *line-to-line voltage* y *phase voltage*

De este modo, para la agrupación de variantes, el paso inicial e indiscutible es el análisis del significado de los términos, que puede realizarse de distintos modos. En nuestro caso, recurrimos al análisis contextual con paráfrasis y a la documentación en recursos externos. Así, conseguimos elucidar el concepto en cada caso. Por una parte, *line-to-line voltage* alude a la tensión entre los conductores de un circuito polifásico. De este modo, cabe esperar una relación directa con *line-to-neutral voltage*, término con el que se denomina a la tensión entre un conductor y el neutro en los circuitos polifásicos.

Una vez aclarado el significado, el siguiente paso consiste en la identificación de elementos contextuales característicos del término, cuyas traducciones se emplearán como consultas en el corpus en español. La idea de base es que la coincidencia de elementos contextuales constituye el puente entre las denominaciones en la lengua origen y la lengua meta. Estas pistas contextuales pueden ser componentes del significado del concepto, como partes de la definición, así como traducciones literales de los formantes, elementos que formen parte de la combinatoria del término, cantidades específicas, nombres de leyes, nombres propios, conceptos con los que guarde una relación directa,

etc. En definitiva, cualquier tipo de términos o datos que coocurran de manera habitual con el término en lengua origen.

Estos elementos deben aparecer en diferentes textos del corpus, de lo contrario podrían referirse a cuestiones propias de un único texto. En cuanto al número de elementos apropiado, no hay una cantidad fija de unidades que se deban incluir. Por una parte, si se indica un mayor número habrá más posibilidades de hallar ocurrencias, pero también más ruido; por el contrario, si se incluye un número limitado podemos generar silencio. Por tanto, la clave está en localizar los elementos que guardan una relación fuerte con el CN, para garantizar la obtención de ocurrencias, así como aquellos cuyo vínculo es más distintivo o específico, para evitar obtener datos irrelevantes. Las consultas, por supuesto, se ven influidas por el conocimiento enciclopédico del que se disponga, que determinará el número mayor o menor de elementos que se indiquen y la tipología de los mismos.

De este modo, tras analizar las concordancias de *line-to-line voltage* y sus variantes denominativas, elegimos *line, VL, VLL, amplitude* y *three-phase* como elementos contextuales. Estos aparecían repetidas veces junto a las variantes del concepto y guardaban relaciones directas y características con este: *line* forma parte del término y podría intervenir en una traducción composicional (esto es, literal), *VL* y *VLL* son los símbolos con los que se identifica este tipo de tensión, *amplitude* es un valor máximo que a menudo se asocia con el término y *three-phase* alude a los sistemas trifásicos en los que se mide este tipo de tensión. En la Figura 22 se pueden observar concordancias en las que estos términos coocurren con *line-to-line voltage*.

e is boosted to a level higher than the *amplitude* of the grid line-to-line voltage in order to achieve full control of the grid current. The p
The *line* feeding a Y-connected three-phase generator has a line-to-line voltage of 480 V. What is the line-to-neutral voltage? B.5.5 Simpl
ages in *three-phase* systems may be line-to-neutral, VLN, or line-to-line voltages , VLL. They may also be described as line voltages (VLL)
hase currents are equal, the neutral current is zero, and the line-to-line voltage , *VLL* , is fi3p multiplied by the line-to-neutral voltage,
nced, Y-connected *three-phase* system (VLN) is equal to the line-to-line voltage (VLL) divided by the square root of 3. The line feeding a

Figura 22: Elementos contextuales de *line-to-line voltage*

Tras este análisis en la lengua origen, el siguiente paso consiste en la traducción de los elementos contextuales, que utilizamos después en la consulta en el corpus español. Para ello, empleamos recursos terminológicos como *Termium Plus* o *IATE*, dado su carácter plurilingüe y la cobertura de léxico especializado. De este modo, los elementos contextuales que utilizamos en español fueron *línea* (*line*), *VL, VLL* (al emplearse frecuentemente los mismos símbolos en ambas lenguas), *amplitud* (*amplitude*), *trifásico* (*three-phase*) y *fase* (ya que *trifásico* también puede nombrarse como *de tres fases*).

Para realizar la consulta en el corpus en español, partimos de que se conoce o se ha podido identificar previamente parte del CN (ya sea el núcleo o algún modificador), por ejemplo, mediante su aparición en recursos terminográficos o su intervención en otros CN. Así, nuestro procedimiento se basa en la búsqueda combinada de este formante y los elementos contextuales. En concreto, dichos elementos contextuales deberán aparecer junto a la parte del CN cuya traducción se conoce. Sin embargo, la simple coocurrencia de dichos términos no es suficiente: estos deben relacionarse del mismo modo que en la lengua original. Así, se consigue acotar la búsqueda, de forma que esta no consista únicamente en la indicación de diversos elementos contextuales, ya que estos pueden coocurrir en distintos tipos de contextos expresando diferentes relaciones. De este modo, se espera que, junto al formante del CN que ya conocemos, encontremos el resto de la traducción. En el caso de *line-to-line voltage*, partimos de que *voltage* se traduce en español por *tensión* o *voltaje*. Así, la consulta introducida en Sketch Engine se presenta en la Figura 23.

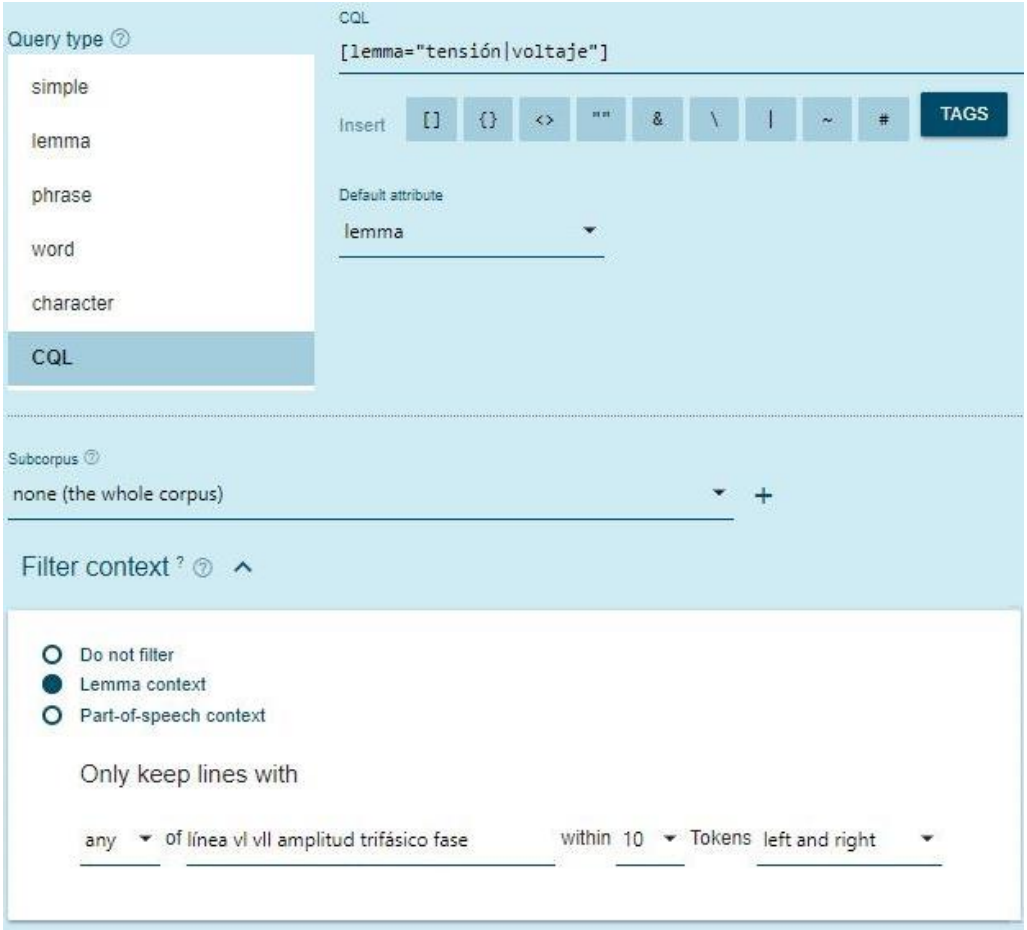

Figura 23: Consulta de elementos contextuales de *line-to-line voltage* en el corpus en español

Para integrar las búsquedas, combinamos dos modos de consulta en la opción *Concordance* de Sketch Engine: la búsqueda simple o en CQL y el filtro de contexto. Por una

parte, en la búsqueda simple o en CQL podremos introducir la parte del CN que se conoce, que aparecerá en el centro de las concordancias si utilizamos el modo de visualización *Key Word In Context* (KWIC). También puede utilizarse, a la inversa, para insertar aquí los elementos contextuales, aunque nos parece más efectivo que se facilite la visualización del posible equivalente frente a la agrupación de los elementos contextuales. Podremos utilizar la búsqueda simple si deseamos introducir uno o varios términos consecutivos, mientras que la búsqueda en CQL ofrece opciones avanzadas, al poder insertar, por ejemplo, disyunciones para que se busque un término u otro, sin que sea necesaria la presencia de ambos. En nuestra consulta hicimos uso de esta función, en la que buscamos *tensión* o *voltaje* en forma de lemas, para obtener resultados tanto en singular como en plural.

El filtro de contexto, por su parte, se utiliza para indicar los elementos contextuales. Para ello, seleccionamos la opción *Lemma context*, con la que se buscan los lemas introducidos en la barra de consulta del filtro de contexto. Al escribirse separados por espacios se identifican como lemas diferentes. Cabe señalar que, tanto en esta barra como en la superior, se pueden realizar búsquedas truncadas, cuando se desean obtener grafías diferentes para alguna parte del término (p. ej. *regul.* > regular, regulador, regulación*, etc.). Esta opción resultó muy útil, ya que es común que una determinada categoría gramatical en la lengua origen se plasme en forma de otra categoría en la lengua meta.

En el filtro de contexto tenemos la opción de seleccionar que aparezcan *all* (todas), *any* (alguna) o *none* (ninguna) de las unidades introducidas. Marcamos *any* (alguna) para indicar que solo es necesaria la presencia de alguna de ellas en el contexto de *tensión* o *voltaje*. Del mismo modo, esta ventana contextual puede abarcar 1, 2, 3, 4, 5, 7, 10 o 15 *tokens* a la izquierda (*left*), a la derecha (*right*) o a ambos lados (*left and right*) del término de partida (*tensión* o *voltaje*). En nuestro caso, la ventana contextual fue de 10 elementos a ambos lados del término. En la Figura 24 se muestran algunas de las concordancias en las que se identificaron posibles equivalentes.

. En circuitos *trifásicos* se expresa generalmente por la	tensión	entre fases. Tensión nominal del circuito de control Vnc:
;encilla para medir la profundidad, es de acuerdo a la mínima	tensión	entre *fases* (de línea) existente, de modo que en un cortocircuitc
del valor de la resistencia del crowbar. Si el valor pico de la	tensión	de *línea* supera el valor de la tensión de bus Vbus, la corriente
l crowbar se representa el valor máximo de la *amplitud* de la	tensión	de línea en bornes del rotor. Figura 3.6 Máxima tensión del
está inhabilitado es de 6000A. La siguiente figura muestra la	tensión	entre *fases* a la salida del convertidor de rotor y en ella se puede
lan como sigue. En la Tabla 3.1 se muestran los valores de las	tensiones	, tanto de *fase* como de línea, para cada posible combinación de
niveles para las tensiones de *fase* y 5 posibles niveles para las	tensiones	de línea. Tabla 3.1.- Combinaciones de estados del NPC. Las
de estados del NPC. Las ecuaciones de tensión fase-neutro y	tensión	de *línea* son igualmente válidas para el convertidor binivel,
obtener una señal de salida del convertidor entre una	tensión	de *línea* de VL = (400/3) ÷ 400 V. 3.6.2. Modelo del convertidor la
y considerando que al trabajar al 50 % de la frecuencia la	tensión	de *línea* será la mitad del nominal Vl = 200 V y el de fase Vf = 115
(5.50), se despejan las tensiones de *fase* en función de las	tensiones	de línea. Dada una determinada tensión en la etapa de continua,
de tensión en un sistema bifásico es igual a la magnitud de	tensión	de *línea* del sistema trifásico equilibrado. Por otra parte, el
una red *trifásica* equilibrada de tensión sinusoidal con	tensión	de línea (VL) = 400 V, y frecuencia de red (fs) = 50 Hz. El eje d se
de tensión). Utilizando las fórmulas 4.2 y 4.3, siendo U, la	tensión	entre *fases* , y considerando la máxima tensión de defecto posibl
mo se calculó en el tema dedicado al generador síncrono. Xs,	tensión	entre *fases* en vacío de la máquina o fuerza electromotriz interna
tensión viene dada por (3.32). Y finalmente el valor rms de la	tensión	*línea* a línea incluyendo todos los armónicos es el siguiente. El
las tensiones línea a neutro (punto medio del bus CC) y de las	tensiones	de *línea* a línea se muestran en la Fig. 3-38 a) y b) respectivamen
muestran en la Fig. 3-38 a) y b) respectivamente. Aunque las	tensiones	de *línea* a línea y de línea a neutro, son diferentes, éstas poseen
1.1. Tabla 1.1: Valor eficaz de los armónicos presentes en la	tensión	de *línea* . Sobremodulación (ma > 1.0). La sobremodulación se
fasor de tensión de red a paritr de alguna composición de las	tensiones	de *línea* . Suponiendo secuencia directa, la representación

Figura 24: Concordancias de la búsqueda de equivalentes de *line-to-line voltage* en español

El análisis distribucional en español arrojó varios CN que podrían ser equivalentes de *line-to-line voltage*: *tensión entre fases, tensión de línea, tensión de fase, tensión línea a línea* y *tensión de línea a línea*. Como vemos, se trata de formas similares, que bien podrían corresponderse con *line-to-line voltage*, pero también con el término *line-to-neutral voltage* o quizás con otros términos. Por tanto, mediante un profundo análisis contextual y gracias a la presencia de concordancias desambiguadoras como las siguientes —además de la coocurrencia con los términos en inglés que también se dio en otros ejemplos analizados—, pudimos diferenciar las denominaciones que no correspondían al concepto en cuestión:

-*Se muestran los valores de las* ***tensiones****, tanto* ***de fase*** *como* ***de línea****.*

-*(…) niveles para las* ***tensiones de fase*** *y 5 posibles niveles para las* ***tensiones de línea****.*

-*Las ecuaciones de* ***tensión fase-neutro*** *y* ***tensión de línea*** *(…).*

-*Se despejan las* ***tensiones de fase*** *en función de las* ***tensiones de línea****.*

-*(…) de las* ***tensiones de línea a neutro*** *(punto medio del bus CC) y de las* ***tensiones de línea a línea****.*

-*Aunque las* ***tensiones de línea a línea*** *y* ***de línea a neutro*** *son diferentes.*

De este modo, constatamos que los términos *tensión de fase, tensión fase-neutro* y *tensión de línea a neutro* no son equivalentes de *line-to-line voltage*, sino probablemente de *line-to-neutral voltage*, lo que comprobaremos a continuación. Así, los equivalentes de *line-*

to-line voltage hallados mediante este procedimiento se redujeron a *tensión entre fases, tensión de línea, tensión línea a línea* y *tensión de línea a línea*. Su equivalencia se comprobó mediante la coincidencia de elementos contextuales relacionados del mismo modo que en la lengua origen, además del análisis del concepto designado. Así, en el corpus en español se utiliza también el símbolo VL, se indica su presencia en sistemas trifásicos y las variantes coocurren con el término *amplitud* (si bien este no constituye una coocurrencia específica de este caso, ya que suele emplearse siempre que se habla de tensión). Además, algunas concordancias revelan la variación denominativa presente entre estos términos, como se aprecia en *tensión entre fases (de línea)*.

Antes de validar estas equivalencias, fue necesario verificar que no se trataba de usos propios de un autor. Para ello, consultamos las concordancias de los términos obtenidos, esta vez sin elementos contextuales. De este modo, también pudimos analizar más ocurrencias para confirmar que se trataba del mismo concepto e identificar nuevas posibles variantes[56]. En concreto, localizamos una nueva denominación, *tensión compuesta*, y detectamos que los términos *tensión línea a línea* y *tensión de línea a línea* constituyen usos propios de un autor, ya que todas sus ocurrencias se encontraban en el mismo texto. Por tanto, los términos empleados para denominar *line-to-line voltage* en inglés y español que identificamos en nuestro corpus son los que se presentan en la Figura 25.

line-to-line voltage
line voltage
phase-to-phase voltage
——

tensión de línea
tensión entre fases
tensión compuesta

Figura 25: Variantes denominativas de *line-to-line voltage* en inglés y español

Tras clarificar las correspondencias en el caso de *line-to-line voltage* y predecir algunos de los equivalentes posibles del término relacionado *line-to-neutral voltage* (*tensión de fase, tensión fase-neutro* y *tensión de línea a neutro*), aplicamos el procedimiento a este concepto. Nuestro objetivo era verificar que dichas denominaciones se correspondían con el concepto en cuestión, además de identificar otros términos

[56] Al igual que en el análisis de términos en inglés, el empleo de patrones de conocimiento sinonímicos (p. ej. *también conocido como, denominado,* etc.) fue un complemento a la identificación de variantes en español.

posibles. En este caso, los elementos contextuales seleccionados fueron *phase, current, neutral* y *three-phase*. Por una parte, *phase* y *neutral* son partes importantes del significado y pueden intervenir en el término equivalente. *Current* suele coocurrir con *phase voltage*, ya que a menudo se habla de *phase voltages and currents* y, además, mantienen una relación directa, pues por ejemplo, la reducción de la tensión implica la reducción de la corriente. Por último, *three-phase* indica el tipo de sistemas en el que se mide esta tensión. En la Figura 26 se muestran algunas concordancias de *line-to-neutral voltage* y sus variantes con estos elementos contextuales.

is therefore based on measurements of three instantaneous	phase voltages	and *currents* , which are followed by an analytical dete
-0.95, respectively, as per Fig. 12. VA and igA are the grid-side	phase voltage	and the injected *current* , respectively. B. Experimental
phase current is automatically limited by reducing the applied	phase voltage	. Therefore, the current regulator can be eliminated. As
0. Part b) shows the line-to-line voltages, part c) is one of the	phase voltages	with respect to *neutral* point N of the transformer wind
orm. B.5.4 Three-Phase Power. Show that the magnitude of the	line-to-neutral voltage	in a balanced, Y connected three *phase* system (VLN) i
generator, the average power can be expressed as [17-20]. The	phase voltage	and *phase* current can be represented by (2.27) (2.28) (

Figura 26: Elementos contextuales de *line-to-neutral voltage*

Los elementos contextuales que empleamos en español fueron, por tanto, *fase* (*phase*), *corriente* (*current*), *neutro* (*neutral*) y *trifásico* (*three-phase*). Como en el ejemplo de *line-to-line voltage*, también se parte de la traducción de *voltage* por *tensión* o *voltaje* en español. Tras seguir el procedimiento indicado, obtuvimos los siguientes CN que podrían ser equivalentes de *line-to-neutral voltage* (Figura 27): *tensión de fase, tensión de fase-punto medio, tensión fase-punto medio, tensión fase-neutro, tensión fase neutro* y *tensión simple*.

ponderada propuesta. La figura A.6 muestra la evolución de las	tensiones	de *fase* en el tiempo, donde perturbaciones equilibradas y
ecuación. Cada combinación de interruptores origina una	tensión	de *fase* - punto medio distinta. En la Fig 3.2 también se muestra
fase respecto al *neutro* se pueden considerar como la resta de la	tensión	fase-punto medio con la tensión del neutro respecto al punto
de comparativa, en las siguientes figuras se representan las	tensiones	de *fase* y de línea obtenidas a la salida de ambos convertidores.
posibles de los estados de *fase* del NPC, junto con sus	tensiones	fase-neutro (Tabla 3.1). Si se asigna a cada combinación su
expresiones. Siendo (P) la potencia activa, (!v□ f) la	tensión	máxima de *fase* , (!□if) la corriente máxima de fase, (f) el ángu
tensión en un estado k cualquiera es la siguiente. Los fasores de	tensión	*fase* neutro se muestran en la figura 5.3. Dividen al hexágono en
e consigue una tensión eficaz de aproximadamente el 100 % de la	tensión	*fase* - neutro de la red con bastantes más armónicos, [70], [8].
generador trifásico a la red, considerando que la tensión U es la	tensión	de *fase* o simple, será la siguiente. Por otra parte por
las tensiones U, E, representaran tensiones de línea en lugar de	tensiones	de *fase* , la fórmula anterior se transforma en la potencia activa
calculada de la siguiente manera, donde ϕ es el desfase entre la	tensión	y la *corriente* de la misma fase. La potencia reactiva del sistema
a la frecuencia de conmutación y a sus múltiplos, presentes en la	tensión	de *fase* (referida al punto N), no se mantienen en las tensiones c

Figura 27: Concordancias de la búsqueda de equivalentes de *line-to-neutral voltage* en español

Gracias al análisis contextual y conceptual, además de la presencia de concordancias desambiguadoras, comprobamos que los términos que descartamos en el análisis anterior de *line-to-line voltage* corresponden efectivamente a *line-to-neutral voltage* y desechamos algunos de los equivalentes posibles, en concreto, *tensión de fase-punto medio* y *tensión fase-punto medio*. Como se observa en el siguiente extracto, se trata de un concepto diferente al expresado por *line-to-neutral voltage*:

-*(…) fase respecto al neutro se pueden considerar como la resta de la tensión fase-punto medio con la tensión del neutro respecto al punto (…)*.

Por tanto, redujimos las posibilidades a *tensión de fase, tensión fase-neutro, tensión fase neutro* y *tensión simple*. Estos se consideraron equivalentes válidos porque expresan el mismo concepto que los términos en inglés y guardan la misma relación con los elementos contextuales. De este modo, se indica, por ejemplo, su relación con el punto neutro y las fases, además de su coocurrencia con el término *corriente*. Además, la variación denominativa se presentó de forma explícita en concordancias como la siguiente:

-*(…) considerando que la tensión U es la tensión de fase o simple*.

Para verificar que estos términos se empleaban en textos diferentes, además de comprobar que remitían al mismo concepto y detectar posibles variantes, analizamos sus concordancias sin emplear la búsqueda contextual. Gracias a ello detectamos que el término *tensión fase neutro* solo se empleaba en un texto, por lo que lo descartamos. Por tanto, las variantes de *line-to-neutral voltage* en inglés y en español se presentan en la Figura 28.

phase voltage
line-to-neutral voltage
——
tensión de fase
tensión fase-neutro
tensión simple

Figura 28: Variantes denominativas de *line-to-neutral voltage* en inglés y español

Como se desprende del análisis de estos dos conceptos, los CN a menudo presentan formas complejas y similares a las de otros términos, aunque los conceptos que expresan pueden ser diferentes. Por tanto, el procedimiento de trasvase entre comunidades lingüísticas debe tener en cuenta estos factores y apoyarse en un análisis conceptual y contextual.

En ello se inspira el procedimiento presentado, que como puede observarse, nos guía hacia equivalentes de los CN que también presentan una forma compuesta. Aunque esto puede percibirse como una limitación, lo cierto es que la mayoría de CN del inglés también adoptan formas compuestas en español, y viceversa, como comentaremos en la

Sección 5.2.2. Sin embargo, una de las ventajas de este procedimiento es que permite identificar correspondencias que no presenten el mismo número de formantes que el término en la lengua origen, ya que a menudo no se dan traducciones lineales en las que cada elemento se sustituya por otro en la lengua meta (p. ej. *line-to-neutral voltage* > *tensión simple*).

Del mismo modo, a pesar de que sea necesario conocer la traducción de alguno de los formantes del CN, este suele ser el caso, ya que habitualmente se conoce la traducción de alguno de ellos, pero se ignora la forma completa del término en la lengua meta. En cualquier caso, este procedimiento debe complementarse con el análisis de concordancias de las correspondencias halladas, que permitirá identificar posibles variantes que no se hubieran obtenido debido a las características de estas búsquedas, como variantes monoléxicas o aquellas que no incluyan el formante del que partimos. Así, a pesar de que a veces se genera ruido, con concordancias en las que no se obtiene la traducción del CN, observamos también que este procedimiento permite obtener equivalentes válidos en casos en los que los recursos terminográficos no aportan resultados. En definitiva, dicho proceso constituye una muestra de que los corpus son herramientas de gran valor que pueden ofrecernos todas las respuestas. Para ello, simplemente hay que saber cómo hacerles las preguntas.

5.2.2 Comparación de equivalentes

La traducción de los CN del inglés al español permitió explorar las características de estos términos en la lengua origen y la lengua meta, identificando así las diferencias y similitudes entre ellos. A continuación, desarrollamos el análisis comparativo realizado, deteniéndonos en los principales aspectos de interés.

En primer lugar, las diferencias en la formación morfosintáctica de los CN en inglés y en español resultan patentes. Los CN del inglés se forman mediante la premodificación del núcleo, que se encuentra a la derecha y al que se anteponen los modificadores. Solo 8 de los CN del inglés se formaron por postmodificación (p. ej. *impact of wind power*). Sin embargo, no podemos olvidar que los CN formados por premodificación en los que interviene más de un sustantivo a menudo pueden alternarse con su variante resultante de la postmodificación (p. ej. *rotor blade* > *blade of the rotor*). No obstante, se suele afirmar que esta forma preposicional posee un menor nivel de especialización. Lo mismo sucede en español, donde la modificación con adjetivos se concibe como más especializada que sus variantes con sintagmas preposicionales (p. ej. *energía eólica marina* > *energía del viento en el mar*). Sin embargo, en esta tesis no atendimos al mayor o menor nivel de especialización de las variantes, sino que nuestro estudio se guio por su presencia en el

corpus especializado y en el sistema conceptual. Sean cuales fueren los modificadores en español, se observó que estos suelen ubicarse a la derecha del núcleo, es decir, se dan CN formados por postmodificación, a diferencia del inglés. Fueron únicamente 8 los CN del español en los que intervino la premodificación (p. ej. *alta tensión*).

Evidentemente, la preferencia por una u otra estructura de los CN está determinada por las características propias de cada lengua, que también influyen en el tipo de constituyentes que forman los CN. En la Figura 29 se aprecian los elementos que intervinieron en los CN en inglés. En estos datos se incluyen, además de los CN obtenidos en nuestra extracción inicial, las variantes denominativas detectadas en la etapa posterior, que nos interesan para el análisis contrastivo de la formación de CN en cada lengua, así como para el estudio de su traducción. Se elevó, por tanto, el número de términos analizados a 330 términos en inglés y 330 términos en español. Sin embargo, en las Figuras 29 y 30, no recogimos las variantes monoléxicas, pues los datos presentados se centran en la formación contrastiva de CN en cada lengua.

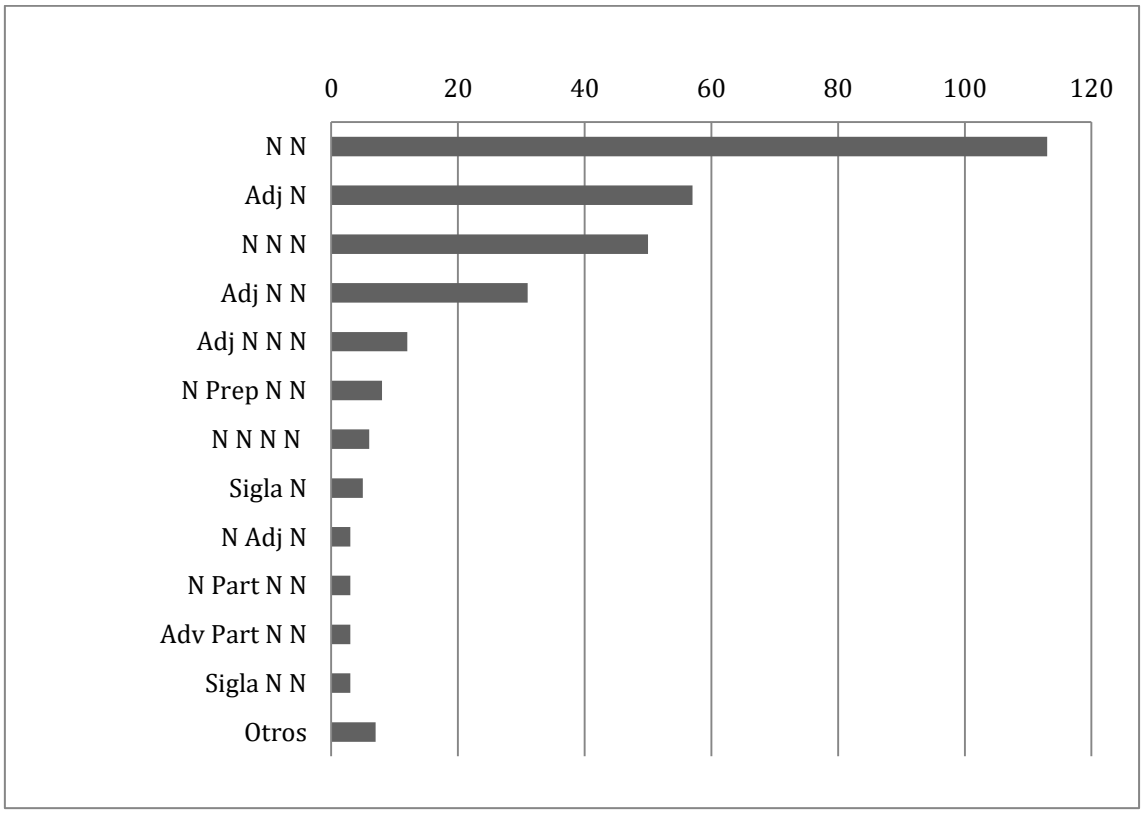

Figura 29: Morfosintaxis de los CN en inglés

Como se observa en la Figura 29, los CN estudiados estaban formados mayoritariamente por dos sustantivos (N N, p. ej. *flux density*). Las siguientes estructuras más habituales fueron aquellas en las que un adjetivo modifica a un sustantivo (Adj N, p. ej. *nuclear power*), dos sustantivos modifican al núcleo nominal (N N N, p. ej. *wind turbine*

generator) o un adjetivo y un sustantivo lo complementan (Adj N N, p. ej. *nominal wind speed*). Otras estructuras también presentes fueron aquellas en las que el CN está formado por un adjetivo y tres sustantivos (Adj N N N, p. ej. *variable speed wind turbine*), el núcleo está postmodificado por una preposición y dos sustantivos (N Prep N N, p. ej. *generator of wind turbine*) o aquellas en las que se agrupan cuatro sustantivos (N N N N, p. ej. *wind turbine power output*). Asimismo, encontramos CN cuyo núcleo nominal se modifica con una sigla (Sigla N, p. ej. *PCC voltage*), además de términos en los que no solo hay un sustantivo, sino dos precedidos de la sigla (Sigla N N, p. ej. *DC link voltage*). Por último, algunos CN se formaron con un sustantivo, un adjetivo y el núcleo nominal (N Adj N, p. ej. *power spectral density*); un sustantivo, un participio[57] y dos sustantivos (N Part N N, p. ej. *stall-regulated wind turbine*); o un adverbio, un participio y dos sustantivos (Adv Part N N, p. ej. *doubly fed induction generator*).

Al margen de estas formas más habituales, hay otras que fueron menos productivas en nuestra muestra y solo generaron un CN en cada caso (categoría *Otros* en la Figura 29): Adj Adj N (*conventional synchronous generator*), Adj Adj N N (*floating offshore wind turbine*), Adj N Adj N (*permanent magnet synchronous generator*), Adj N Part N N (*passive stall-regulated wind turbine*), N Prep Adj N (*energy from renewable sources*), N Prep Art Sigla (*voltage at the PCC*) y N N Prep Art N N (*power output of a wind turbine*). La escasa presencia de estas formas puede deberse a la longitud de los CN que generan (de tres, cuatro y hasta cinco formantes), así como a la concreción de las estructuras, difícilmente adaptables a otros CN.

De estos datos se desprende que, lejos de los estudios tradicionales que se han centrado en los CN en inglés formados únicamente por sustantivos, este mecanismo de formación de términos integra otras categorías gramaticales. Prueba de ello es la amplia presencia de adjetivos y, en menor medida, de otras categorías como los participios y los adverbios. En esta línea, también se generan CN modificando el núcleo con una sigla, lo que constituye una forma de compactar CN de mayor extensión. En estos casos, el CN se vuelve más opaco, pues se necesita conocer el concepto al que refiere la sigla.

Cuando intervienen participios también se transmite una imagen de mayor tecnicismo, ya que estos generalmente conllevan una mayor extensión del CN. Sin embargo, a nuestro parecer, estos términos contribuyen a la comprensión del concepto

[57] Se consideraron como participios solo aquellos casos en los que este se acompañaba de algún complemento (p. ej. *doubly-fed induction generator*). De lo contrario, se consideraron adjetivos, ya que desempeñan la misma función y, de este modo, pudimos extraer más regularidades (p. ej. *installed capacity*).

gracias a la explicitación del verbo subyacente (p. ej. *stall-regulated wind turbine*, *doubly-fed induction generator*).

En general, resulta patente el encadenamiento de los formantes sin recurrir a elementos de conexión como preposiciones y artículos, que son minoritarios en nuestra muestra, lo que va ligado a la práctica omnipresencia de la premodificación en inglés. Así, al traducir hacia el inglés se debe tener en cuenta la posibilidad de que los CN de tipo preposicional en español adopten una estructura de premodificación sin preposiciones (*potencia de salida de un aerogenerador* > *wind turbine power output*), algo en lo que no siempre será fácil reparar. Cuando se da la postmodificación, la forma más habitual es la de N Prep N N (*impact of wind energy*), siendo *of* la preposición más productiva.

Este análisis también revela el número[58] de formantes que suelen constituir los CN estudiados: 176 CN estaban formados por dos elementos (*wind blade*), 97 CN tenían tres formantes (*power output curve*), 27 CN presentaban cuatro elementos (*large-scale wind farm*) y, por último, 1 CN estaba formado por cinco constituyentes (*passive stall-regulated wind turbine*). Como indicamos anteriormente, es común que los CN de menor extensión sean más habituales, ya que a medida que aumenta el número de formantes, disminuye la presencia de estos términos más extensos, de mayor exigencia cognitiva. No obstante, como veremos a continuación, en muchas ocasiones estos términos reducidos constituyen variantes de términos más largos en los que se han omitido formantes. Además, estos términos de mayor extensión se enmascaran frecuentemente bajo siglas, lo que ofrece una imagen distorsionada de la presencia real de dichos términos extensos, que no dejan de ser una parte fundamental del lenguaje especializado.

Por su parte, los CN en español que figuraron entre los equivalentes de los CN del inglés presentaban la siguiente composición morfosintáctica (Figura 30).

[58] En el cómputo de formantes no se incluyen los nexos, como preposiciones y artículos.

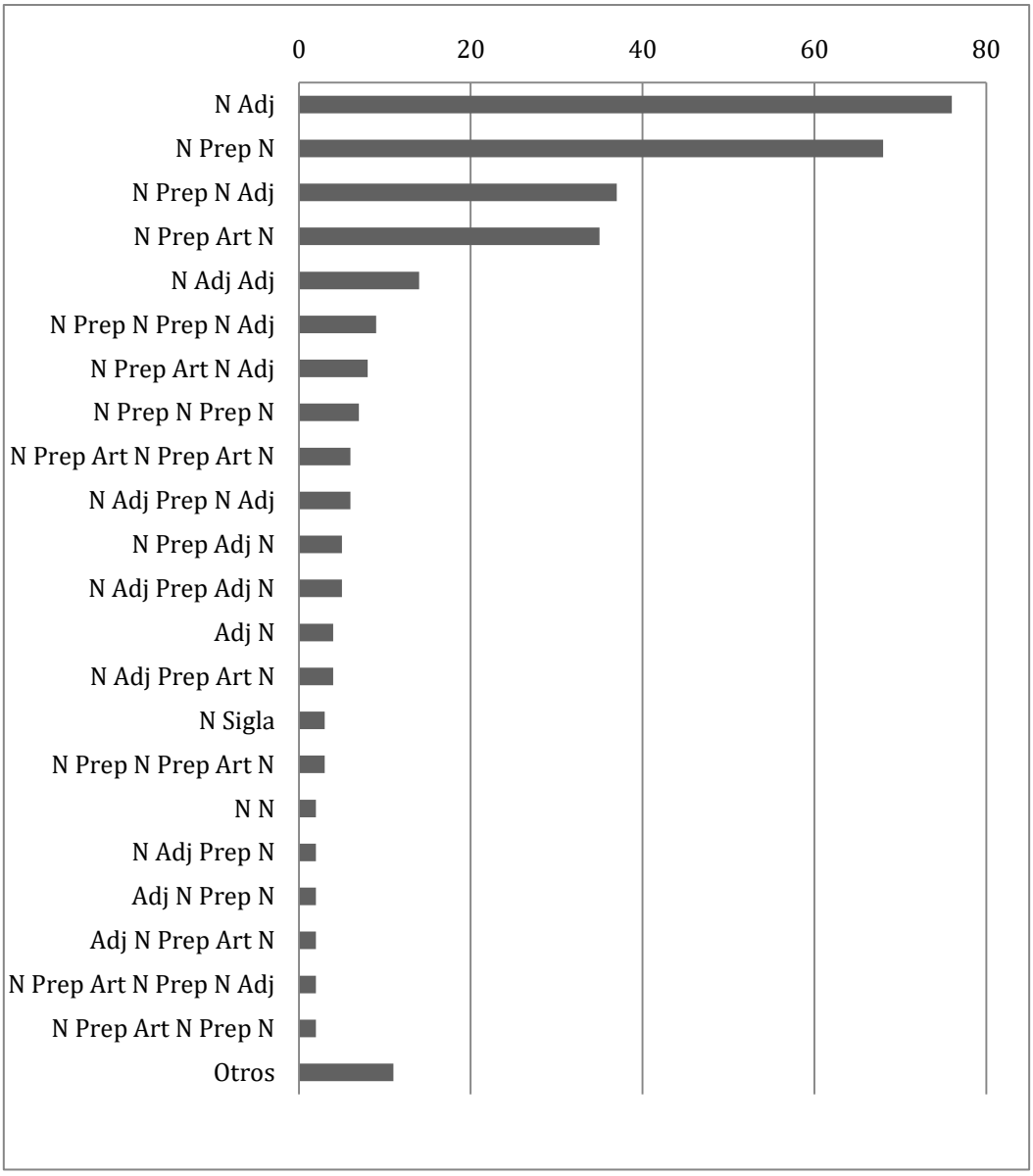

Figura 30: Morfosintaxis de los CN en español

Las dos estructuras principales de los CN en español son aquellas en las que un sustantivo está postmodificado por un adjetivo (N Adj, p. ej. *generación eólica*) o por una preposición y un sustantivo (N Prep N, p. ej. *curva de potencia*). Les siguen otras dos estructuras también bastante frecuentes, en las que el núcleo nominal está seguido de una preposición, un sustantivo y un adjetivo (N Prep N Adj, p. ej. *generador de corriente alterna*), o bien de una preposición, un artículo y un sustantivo (N Prep Art N, p. ej. *par del generador*). El resto de estructuras habituales en español, aunque ven su frecuencia reducida, son las siguientes:

-N Adj Adj, p. ej. *turbina eólica flotante*

-N Prep N Prep N Adj, p. ej. *aerogenerador de paso de pala fijo*

-N Prep Art N Adj, p. ej. *industria de la energía eólica*

-N Prep N Prep N, p. ej. *inversor en fuente de tensión*

-N Prep Art N Prep Art N, p. ej. *control del par del generador*

-N Adj Prep N Adj, p. ej. *generador síncrono de imanes permanentes*

-N Prep Adj N, p. ej. *aerogenerador de baja potencia*

-N Adj Prep Adj N, p. ej. *generador asíncrono de doble alimentación*

-Adj N, p. ej. *baja tensión*

-N Adj Prep Art N, p. ej. *par electromagnético del generador*

-N Sigla, p. ej. *generador DFIG*

-N Prep N Prep Art N, p. ej. *tensión en bornes del generador*

-N N, p. ej. *generador multipolo*

-N Adj Prep N, p. ej. *densidad espectral de potencia*

-Adj N Prep N, p. ej. *baja velocidad de viento*

-Adj N Prep Art N, p. ej. *alta velocidad del viento*

-N Prep Art N Prep N Adj, p. ej. *tensión en el punto de acoplamiento común*

-N Prep Art N Prep N, p. ej. *tensión en la etapa de continua*.

Además, hubo otras formas que también dieron lugar a CN en español, aunque fueron menos productivas y solo generaron un CN en cada caso (categoría de *Otros* en la Figura 30): N N N (*tensión fase-neutro*), N Prep N Prep N Prep N (*generador de inducción de jaula de ardilla*), N Adj Prep N Prep N (*generador asíncrono de jaula de ardilla*), N Prep N Adj Prep N Adj (*aerogenerador de regulación pasiva por pérdida aerodinámica*), N Part Prep N Adj (*aerogenerador regulado por pérdida aerodinámica*), N Prep N Adv Part (*generador de inducción doblemente alimentado*), N Adj Adv Part (*generador asíncrono doblemente alimentado*), N Prep N Prep Adj N (*generador de inducción de doble alimentación*), N Prep Art N Prep N Prep Art N (*tensión en el punto de conexión a la red*), N Prep Art Sigla (*tensión en el PCC*) y N Prep N Sigla (*tensión de bus DC*). Como sucedía en inglés, el reducido empleo de estas formas puede resultar de su escasa extrapolación a diferentes conceptos y de la longitud de los CN a los que dan lugar.

En concreto, contamos con 189 CN en español en los que intervenían dos formantes (*turbina de reacción*), frente a 176 CN de la misma extensión en inglés. Por su parte, 94 CN en español estaban formados por tres elementos (*central de generación eléctrica*), frente a los 97 CN del inglés. Por último, 29 CN en español se constituían de cuatro formantes (*generador de inducción de rotor devanado*), a diferencia de los 27 CN del inglés, y 1 CN tenía cinco elementos (*aerogenerador de regulación pasiva por pérdida aerodinámica*), la misma cifra encontrada en inglés. Por tanto, contamos con cantidades similares en las dos lenguas, en las que la frecuencia de los CN se reduce a medida que aumentan los formantes.

En las diferentes estructuras morfosintácticas de los CN en español se observa que, como en inglés, estos también incorporan diferentes categorías gramaticales entre sus modificadores: principalmente sustantivos y adjetivos, pero también otros elementos como participios. Cuando el CN tiene dos formantes, encontramos cifras similares entre la modificación adjetival (N Adj) y preposicional (N Prep N). A medida que intervienen más elementos, tampoco se observa una preferencia por una u otra de estas formas, sino que se aprecia su convivencia en la modificación de los CN: *generador asíncrono de doble alimentación*, *densidad espectral de potencia*, etc.

En este sentido, es frecuente que la yuxtaposición de sustantivos propia del inglés (p. ej. N N N) se traslade al español mediante modificadores adjetivales o preposicionales. Un ejemplo clásico es el de *wind*, traducido por el sintagma preposicional *de/l viento* o el adjetivo *eólico*. Sin embargo, debemos considerar que este adjetivo no constituye siempre un equivalente válido de *wind*, ya que su empleo alude al viento como recurso aprovechable (un uso particular que no suele indicarse en los recursos lingüísticos). Así, *eólico* no sería un equivalente válido de *wind* en CN como *wind system* (en su significado relacionado con la circulación atmosférica) o *wind speed*, que se trasladan como *sistema de vientos* y *velocidad del viento*, respectivamente, pues no implican el aprovechamiento del viento. En cambio, cuando *eólico* es un equivalente adecuado, suele poder alternarse con su variante morfosintáctica *de/l viento*, a excepción de algunos casos fijos como *parque eólico*.

Respecto a los adjetivos, estos se colocan tras el sustantivo en español, a excepción de algunos como *bajo, medio* y *alto*, que por lo general lo preceden. Ejemplo de ello son CN como *aerogenerador de baja potencia, baja tensión, baja velocidad de viento, alta tensión, alta velocidad de viento* y *media tensión*. Recordemos que los adjetivos que se colocan delante de los sustantivos en español adquieren un matiz literario y enfático, por lo que no

se recomienda abusar de este recurso. Sin embargo, los ejemplos comentados constituyen formas fijas que, aunque puedan presentar variantes con el adjetivo pospuesto (*baja velocidad de viento, velocidad de viento baja*), son mucho menos frecuentes. Además, la posición del adjetivo es determinante, ya que su ubicación detrás del sustantivo puede indicar otro significado. Es el caso, por ejemplo, de *average voltage*, que se traduce por *tensión media*, a diferencia de la forma *media tensión*, cuyo equivalente es *medium voltage* y transmite un significado diferente. Cuando el CN tiene más de dos formantes, la posición inicial del adjetivo indica su relación con el núcleo y no únicamente con el modificador, por lo que facilita el *bracketing* (p. ej. en *baja velocidad de viento*, lo bajo es la velocidad, no el viento).

En esta línea, destaca la presencia de nexos en los CN en español, mucho más elevada que en sus homólogos ingleses. Así, es habitual el recurso a preposiciones y artículos, lo que genera un mayor grado de variación, al ser frecuente la alternancia entre formas preposicionales y adjetivales (*energía del viento*, *energía eólica*), como hemos señalado. El empleo de nexos también genera variación a raíz del empleo o la exclusión de los artículos (*tensión de red*, *tensión de la red*). Estos aspectos derivan en un mayor grado de variación morfológica y morfosintáctica de los CN en español frente a los CN del inglés.

Respecto a los nexos, la preposición de mayor uso en los CN en español es *de*, al igual que sucede con *of* en inglés. Este enlace es el más frecuente debido a su validez para representar diferentes relaciones semánticas, lo que lo convierte en una opción segura. Sin embargo, ello conlleva también una escasa concreción semántica de dicha preposición, ya que puede aplicarse a prácticamente todo tipo de vínculos: *aerogenerador de eje horizontal* (*has_attribute*), *central de generación eléctrica* (*has_function*), *convertidor de potencia* (*affects*), etc. En algunos casos, intervienen otras preposiciones más específicas semánticamente, si bien su uso es minoritario (p. ej. *tensión* ***en*** *continua, aerogenerador de regulación* ***por*** *pérdida aerodinámica* o *tensión* ***entre*** *fases*). Además, los CN que incluyen estas preposiciones suelen contar con variantes que incorporan la preposición *de* (*tensión en la etapa de continua, tensión de la etapa de continua*).

Tras *de*, la preposición más recurrente es *en*, que puede indicar tiempo o lugar. Sin embargo, encontramos un caso particular en el que el empleo de la preposición *en* no expresa este significado. Se trata de los términos *convertidor en fuente de tensión* e *inversor en fuente de tensión*, equivalentes de los términos ingleses *voltage source converter* y *voltage source inverter*. En estos términos, en lugar de expresar tiempo o lugar, el empleo de la preposición *en* alude al concepto afectado por el convertidor, es decir, la fuente de

tensión. Por tanto, se trata de un tecnicismo, de elevada opacidad semántica, que puede confundir al usuario y debe someterse a un profundo análisis conceptual antes de incurrir en posibles traducciones inapropiadas (p. ej. calco de la preposición al traducir hacia el inglés o, al contrario, empleo de otra preposición o estructuras diferentes de la preferida en español).

En los CN de más de dos elementos en español, podemos encontrar el empleo de la preposición *de* para enlazar las diferentes partes, así como el recurso a otras preposiciones. Esta nos parece una mejor opción para favorecer la comprensión del concepto y aliviar el discurso, evitando así concatenaciones de la preposición *de*. Algunos ejemplos son *tensión en bornes del generador* (y no *tensión de bornes del generador*), *tensión en el punto de conexión a la red* (en lugar de *tensión del punto de conexión de la red*) y *tensión en la etapa de continua* (frente a *tensión de la etapa de continua*). Como hemos señalado, es habitual que estos CN convivan con variantes en las que se emplee la preposición *de*, generando así un encadenamiento de esta preposición. A nuestro parecer, esta opción no clarifica la relación semántica y satura el discurso.

En la estructura de los CN en español también salta a la vista la influencia del inglés, que resulta patente en el abundante recurso a los calcos. Los calcos de estructuras poliléxicas pueden ser calcos de expresión, cuando se respetan las estructuras sintácticas de la lengua meta, o calcos estructurales, cuando se adoptan estructuras impropias de la lengua meta. Estos últimos están relacionados directamente con la formación morfosintáctica de los CN y se encontraron en dos vertientes principales. Por una parte, los CN del español en los que se yuxtaponen dos sustantivos no constituyen formas habituales de esta lengua, de modo que representan calcos estructurales. Por ejemplo, *generador multipolo* (*multipole generator*) y *generador diésel* (*diesel generator*).

Por otro lado, los CN que cuentan con un participio en inglés presentan dos formas de traducción en español. Una de ellas es el calco, que constituye el otro ejemplo común de calco estructural junto a la yuxtaposición de sustantivos. Mediante esta forma, se reproduce el participio en español junto a los complementos que lo acompañen: *aerogenerador regulado por pérdida aerodinámica* (*stall-regulated wind turbine*), *generador de inducción doblemente alimentado* (*doubly-fed induction generator*) o *generador de inducción autoexcitado* (*self-excited induction generator*). Frente a esta técnica, suele darse también otra forma de traducción de estos CN, consistente en la adaptación a una forma más natural en español, en concreto, la nominalización del predicado subyacente: *aerogenerador de regulación por pérdida aerodinámica* (*stall-*

regulated wind turbine) o *generador de inducción de doble alimentación* (*doubly-fed induction generator*). No obstante, la tendencia más habitual es el calco, al tratarse de la opción más fácil y segura.

Aunque no constituyan calcos, cabe señalar que los CN del inglés normalmente adquieren la forma de CN en español. También encontramos variantes monoléxicas, que suelen reproducirse en las dos lenguas: *alternator* > *alternador*, *generator* > *generador*, *dynamo* > *dínamo/dinamo*, etc. En concreto, se dieron 7 términos simples en inglés y 8 en español, además de 22 siglas empleadas como términos independientes en inglés y 9 en el caso del español, lo que indica un mayor recurso a las siglas en la lengua inglesa. Esto no resulta extraño si consideramos la naturaleza sintética de esta lengua, frente a la tendencia a la explicación del español.

Para la generación de términos simples también se da en ocasiones la derivación con elementos clásicos que, aunque existe en las dos lenguas, es más propia del español que del inglés. Algunos ejemplos son *aerogenerador* (término que ha primado frente a la variante *turbina eólica*, a diferencia de lo que sucede en inglés, donde se prefiere hablar de *wind turbine*), *hydraulic turbine* o *energía undimotriz*. En ocasiones estos términos conviven con variantes formadas mediante otros procedimientos, como vemos en *energía undimotriz* y *energía olamotriz* o *energía de las olas*. No obstante, la derivación con elementos clásicos no constituyó un recurso habitual de formación ni traducción en nuestros términos, en los que se empleó en mayor medida la composición sintagmática.

Además de la formación morfosintáctica, el *bracketing* también desempeña un papel clave en el traslado de los CN del inglés al español. Así, la desambiguación estructural del CN en la lengua origen es fundamental para detectar o producir el equivalente correcto atendiendo a distintas cuestiones. Una de ellas es la ubicación de la preposición que probablemente intervenga en la traducción al español. Por ejemplo, una vez que conocemos que la estructura de *permanent magnet synchronous generator* es [*permanent magnet*] [*synchronous generator*], podremos detectar correctamente el equivalente *generador síncrono* ***de*** *imanes permanentes* y no incurrir en posibles traducciones erróneas como **generador permanente síncrono con imanes*, que podría surgir al confundir la dependencia de *permanent* (que modifica a *magnet* y no a *generator* en la posible estructura *permanent* [*magnet synchronous generator*]).

Otro ejemplo destacable lo constituye *power electronic converter*. Gracias al protocolo de *bracketing*, detectamos que *electronic* no es un adjetivo que modifique al núcleo, sino que alude al sustantivo *electronics*, siendo la estructura del CN [*power*

electronic] *converter*. A partir de este esquema, se identifica el equivalente *convertidor* ***de*** *electrónica de potencia*.

Por otra parte, el *bracketing* determina la posición de los adjetivos o elementos con función adjetival (p. ej. participios) que pueden intervenir en los CN, ya que las preferencias de ubicación varían en inglés y en español. En inglés, hay muchos CN en los que un adjetivo precede a varios elementos, sin modificar necesariamente al que le sigue, sino más bien al núcleo o a otro elemento intermedio. Un ejemplo de ello es *rated wind speed*, donde *rated* no modifica a *wind* (en una estructura del tipo [*rated wind*] *speed*) sino al núcleo *speed* (siendo la estructura correcta *rated* [*wind speed*]). Cuando esto sucede en español, dicho adjetivo suele introducirse en el CN inmediatamente después del elemento al que modifica, para aclarar así la estructura (*velocidad* ***nominal*** *del viento* y no *velocidad del viento* ***nominal***).

Observamos esta tendencia en los CN estudiados en español, entre los que también figuran ejemplos como *aerogenerador de regulación* ***pasiva*** *por pérdida aerodinámica* (***passive*** *stall-regulated wind turbine*), *velocidad* ***media*** *del viento* (***mean*** *wind speed*) o *densidad* ***espectral*** *de potencia* (*power spectrum*). En este último caso, el adjetivo no figura explícitamente en el CN en inglés, sino que se obtiene al desglosar el núcleo en forma de CN en español (*spectrum* > *densidad espectral*). Sin embargo, a veces conviven las dos variantes: aquella en la que el adjetivo sigue inmediatamente al elemento al que modifica (*velocidad* ***media*** *del viento*) y su variante en la que el adjetivo se coloca al final del CN (*velocidad del viento* ***media***). No obstante, la primera forma suele ser la más utilizada, al facilitar la comprensión del concepto y generar un CN más natural en español. En general, estos casos de *bracketing* presentan una mayor facilidad en español, porque la posición y la concordancia del adjetivo pueden desambiguar, mientras que en inglés son más complejos de resolver, al colocarse frecuentemente el adjetivo al inicio de toda la cadena del CN (p. ej. ***passive*** *stall-regulated wind turbine*).

En este sentido, tanto en inglés como en español, los diferentes modificadores se ordenan según su contribución al concepto global: a mayor importancia, mayor cercanía al núcleo. Por ejemplo, en *parque eólico marino*, prima el recurso (*eólico*) y, después, la ubicación (*marino*). A este respecto, observamos que la concatenación de dos modificadores adjetivales en español no es muy usual, probablemente debido a la sobrecarga del discurso, prefiriéndose así su alternancia con sintagmas preposicionales (*sistema de generación eólica* en lugar de *sistema generador eólico*).

Por último, el *bracketing* determina la concordancia de los adjetivos o elementos con función adjetival que pueden intervenir en los CN en español, tanto en género como en número. De este modo, al conocer a qué elemento modifica el adjetivo, podremos concordarlo de la manera adecuada, como se observa en los siguientes ejemplos: *generador de reluctancia* ***conmutada*** ([*switched reluctance*] *generator*), *aerogenerador de regulación* ***pasiva*** *por pérdida aerodinámica* ([*passive stall-regulated*] [*wind turbine*]) o *generador de imanes* ***permanentes*** ([*permanent magnet*] *generator*). Así, un mal *bracketing* puede ocasionar equivalentes erróneos en los que el adjetivo se concuerde de forma inapropiada (p. ej. **generador de reluctancia conmutado* o **generador conmutado de reluctancia*).

En relación con la posición del adjetivo que hemos mencionado arriba, cuando este muestra un género o número diferente al del resto de formantes del CN, puede colocarse al final del CN, ya que esta concordancia resuelve los problemas de desambiguación estructural. Algunos ejemplos de ello son *generador de inducción autoexcitado* o *generador de inducción doblemente alimentado*, donde el género masculino de *autoexcitado* y *alimentado* indica que estos modifican a *generador*. En cambio, si el adjetivo se coloca al final y este presenta una forma válida para ambos géneros o números (p. ej. *velocidad del viento* ***nominal***), la desambiguación se hace complicada. Por ello, lo habitual es que el adjetivo se ubique detrás del elemento al que modifica (*velocidad* ***nominal*** *del viento*).

Por otra parte, al analizar el papel de la semántica de los CN en su traducción encontramos CN cuya traducción no presenta grandes dificultades debido a la transparencia de los formantes y a la recurrencia de los CN (p. ej. *wind direction, synchronous generator*), frente a otros de mayor complejidad por su opacidad y nivel técnico, a veces generado por los conceptos que lo componen (p. ej. *line-to-line voltage*) o por la extensión del CN (p. ej. *passive stall-regulated wind turbine*). Evidentemente, en el grado de dificultad intervienen distintos factores, como el conocimiento del que se disponga o la labor de documentación que se haya realizado, aunque resulta innegable que hay un amplio espectro de CN en lo que a complejidad semántica y, con ello, dificultad de traducción se refiere.

En cuanto a la semántica de los formantes en ambas lenguas, resulta patente la abundancia de lo que denominamos traducciones *composicionales* o lineales, que consisten en la traducción literal de los formantes del CN y no suelen implicar grandes cambios en lo que a categorías, roles y relaciones se refiere (p. ej. *nuclear energy* > *energía nuclear*). No

obstante, también se encuentran muchos casos que escapan a estas traducciones composicionales.

Esto puede producirse por diversos motivos, como el empleo de equivalentes de mayor o menor extensión que el término original (p. ej. *distributed generator* > *unidad de generación distribuida, AC generator* > *alternador* o *power spectrum* > *densidad espectral de potencia*). También puede suceder que, aunque la traducción literal de los formantes exista, se utilicen constituyentes de diferente carga semántica, como ocurre en *impulse turbine*, que suele denominarse *turbina de acción* en español, en lugar de *turbina de impulso*. Este ejemplo está relacionado con el papel de la multidimensionalidad en la formación de CN, en la que nos detendremos a continuación.

Asimismo, cuando el equivalente del CN de origen constituye también un CN, el núcleo suele consistir en una traducción literal, a excepción de algunos casos en los que se recurre a un término de carácter más general que se especifica en combinación con los modificadores. Algunos ejemplos de ello son *distributed generator* > *unidad de generación distribuida* o *wind turbine* > *sistema eólico* o *máquina eólica*, y constituyen ejemplos de variación cognitiva, como comentaremos más adelante. Por tanto, estas traducciones no composicionales pueden implicar desigualdades en la semántica de los formantes en ambas lenguas. Como se infiere de estos ejemplos, la traducción de los CN no puede limitarse, por tanto, a un traslado lineal, sino que debe basarse en un procedimiento de traducción conceptual y contextual como el presentado en la Sección 5.2.1.

El empleo de adjetivos de la lengua general en el lenguaje especializado en inglés también se da en español. Así pues, adjetivos generales como *conventional* y *apparent* que intervenían en los CN también figuran en sus equivalentes en español (*generador síncrono convencional, potencia aparente*). Aunque estos adjetivos formen parte del término, retienen su significado de la lengua general, a diferencia de otros como *large, small* o *instantaneous*, que adquieren un valor especializado diferente al que poseen en la lengua general.

Por una parte, *large* y *small* no solo aluden al tamaño del aerogenerador o parque eólico, sino también a la potencia que estos pueden generar, algo que va ligado a su tamaño. En los CN en inglés no se explicita este sentido, mientras que esto sí sucede en sus equivalentes en español, en los que además suelen preferirse las variantes que aluden a la potencia: *aerogenerador de gran potencia* (*large wind turbine*), *aerogenerador de baja potencia* (*small wind turbine*). Sin embargo, este paralelismo entre tamaño y capacidad sí se contempla en inglés cuando no intervienen estos adjetivos, sino el sustantivo al que

hacen referencia (*turbine size, turbine capacity* > *tamaño de la turbina, capacidad de la turbina*). Por su parte, el adjetivo *instantaneous* también reproduce su significado específico en español mediante el equivalente directo *instantáneo* (*instantaneous voltage* > *tensión instantánea*). En definitiva, conviene comprobar si estos significados específicos se repiten en la otra lengua, para evitar caer en el calco, que puede no ser válido.

Uno de los aspectos más interesantes al contrastar la semántica de los formantes en ambas lenguas es la multidimensionalidad, que nos conecta directamente con el estudio contrastivo de la variación en los términos estudiados. A menudo, las variantes de un concepto enfatizan dimensiones cognitivas diferentes, que pueden reproducirse o no en la lengua meta. En los CN estudiados, lo habitual es que estas dimensiones se plasmen en ambas lenguas. En la Tabla 50 se presentan algunos ejemplos de conceptos con sus variantes en inglés (izquierda) y sus equivalentes en español (derecha), en los que se aprecia que las dos lenguas transmiten las mismas dimensiones.

Variantes de los CN en inglés y dimensiones	**Variantes de los equivalentes en español y dimensiones**
TAMAÑO/CAPACIDAD large wind turbine large-scale wind turbine **FUNCIÓN** commercial wind turbine utility-scale wind turbine	**TAMAÑO/CAPACIDAD** aerogenerador de gran potencia aerogenerador de gran tamaño **FUNCIÓN** aerogenerador comercial
MECANISMO DE REGULACIÓN stall-regulated wind turbine stall-controlled wind turbine passive stall-regulated wind turbine **COMPOSICIÓN** fixed-pitch wind turbine	**MECANISMO DE REGULACIÓN** aerogenerador de regulación por pérdida aerodinámica aerogenerador de regulación pasiva por pérdida aerodinámica aerogenerador regulado por pérdida aerodinámica **COMPOSICIÓN** aerogenerador de paso fijo aerogenerador de paso de pala fijo
RECURSO EMPLEADO induction generator **MOVIMIENTO DEL ROTOR Y EL ESTÁTOR** asynchronous generator	**RECURSO EMPLEADO** generador de inducción **MOVIMIENTO DEL ROTOR Y EL ESTÁTOR** generador asíncrono
RECURSO EMPLEADO wind turbine WT windmill wind system wind energy system wind power system aerogenerator wind generator wind turbine system	**RECURSO EMPLEADO** turbina eólica turbina de viento molino de viento sistema eólico aeroturbina máquina eólica aerogenerador sistema de energía eólica sistema de conversión de energía eólica

wind energy conversion system wind power generation system wind turbine generator **FUNCIÓN** -[CONVERSIÓN] wind energy conversion system WECS -[GENERACIÓN] wind power generation system aerogenerator wind generator wind turbine generator **ENERGÍA PRODUCIDA** wind energy system wind energy conversion system wind power generation system wind power system	sistema de generación eólica generador eólico **FUNCIÓN** -[CONVERSIÓN] sistema de conversión de energía eólica -[GENERACIÓN] sistema de generación eólica aerogenerador generador eólico **ENERGÍA PRODUCIDA** sistema de conversión de energía eólica sistema de energía eólica
RECURSO EMPLEADO wind rotor **ENTIDAD A LA QUE PERTENECE** wind turbine rotor	**RECURSO EMPLEADO** rotor eólico **ENTIDAD A LA QUE PERTENECE** rotor del aerogenerador
RECURSO EMPLEADO gas turbine **PROCESO INTERNO** combustion turbine	**RECURSO EMPLEADO** turbina de gas **PROCESO INTERNO** turbina de combustión
CAPACIDAD DE TRABAJO active power **PERCEPCIÓN** real power **CANTIDAD** rated power rated capacity power rating nominal power rated output nominal capacity **REQUISITO TÉCNICO** installed capacity installed power	**CAPACIDAD DE TRABAJO** potencia activa **PERCEPCIÓN** potencia real **CANTIDAD** potencia nominal capacidad nominal **REQUISITO TÉCNICO** capacidad instalada potencia instalada
UBICACIÓN output power power output **PERCEPCIÓN** useful power	**UBICACIÓN** potencia de salida **PERCEPCIÓN** potencia útil

Tabla 50: Variantes de conceptos en inglés y español, y semejanza de dimensiones

Como puede observarse en la Tabla 50, cada fila presenta un concepto con sus variantes en inglés y en español. Los términos con los que se denominan estos conceptos aluden a diferentes dimensiones, lo que da lugar a variación denominativa cognitiva. Estas

perspectivas se repiten también en sus denominaciones en español. Destacamos el ejemplo de *wind turbine*, por presentar un elevado grado de variación, en la que hace referencia a tres dimensiones diferentes.

Así, las variantes de *aerogenerador* pueden aludir al recurso utilizado (el viento), como vemos en *wind turbine, windmill, wind system, turbina de viento, molino de viento, aeroturbina*, etc. En lugar del recurso empleado, estas pueden referirse a la función del sistema. En concreto, dicha función presenta dos caras, la conversión y la generación, de modo que la variación surge también al especificar distintas características de una misma dimensión (*wind energy conversion system, wind power generation system, sistema de conversión de energía eólica, sistema de generación eólica*). Por último, el foco puede estar en la energía producida (energía eólica): *wind power system*, *sistema de energía eólica*. En concreto, el modificador *energía eólica*, al ser polisémico, puede entenderse como el recurso empleado o la energía producida.

Como hemos señalado en varias ocasiones a lo largo de este trabajo, la multidimensionalidad encuentra en los CN su terreno más propicio, ya que estas estructuras permiten también la especificación de varias dimensiones en un mismo término. Por ejemplo, en *aerogenerator* y *aerogenerador* se explicita la función de generación y el recurso empleado (*aero-*), igual que sucede en *wind generator* y *generador eólico*. Por su parte, en *wind energy conversion system* y *sistema de conversión de energía eólica* se indica la función de conversión y el recurso que emplea o la energía que se produce (*wind energy, energía eólica*). La especificación de distintas dimensiones conlleva también la alusión a diferentes *slots* del microcontexto del núcleo, como el PACIENTE (***wind energy*** *conversion system*) o el INSTRUMENTO (***aero****generador*). Por tanto, la inclusión de varias dimensiones en un mismo término suele redundar en una mayor extensión de los CN.

Sin embargo, a pesar de que las dimensiones suelen ser paralelas en inglés y en español, también pueden producirse desigualdades. Por ejemplo, en una lengua puede ser posible la generación de CN en los que convergen varias dimensiones, mientras que en otra quizás se prefiera su especificación por separado en diferentes variantes. Un ejemplo de ello lo encontramos en *central eléctrica*, que cuenta con variantes en las que se explicita tanto la función como la energía producida (*central de generación eléctrica, instalación de producción de energía eléctrica*), mientras que en inglés las dimensiones suelen indicarse de forma individual (*power plant, generating station*). Del mismo modo, los términos de

una lengua pueden incluir dimensiones diferentes a las que se indican en la otra lengua, como sucedió en algunos de nuestros términos (Tabla 51).

Variantes de los CN en inglés y dimensiones	Variantes de los equivalentes en español y dimensiones
ACCIONAMIENTO direct-drive generator **COMPOSICIÓN** multipole generator	**ACCIONAMIENTO** generador de accionamiento directo **TRANSMISIÓN** generador de transmisión directa **COMPOSICIÓN** generador multipolo
UBICACIÓN line-to-line voltage line voltage phase-to-phase voltage	**UBICACIÓN** tensión de línea tensión entre fases **COMPOSICIÓN** tensión compuesta
UBICACIÓN phase voltage line-to-neutral voltage	**UBICACIÓN** tensión de fase tensión fase-neutro **COMPOSICIÓN** tensión simple

Tabla 51: Variantes de conceptos en inglés y español con dimensiones diferentes

Como se aprecia en el primer ejemplo de la Tabla 51, la alusión al modo de transmisión (*generador de transmisión directa*) no se realiza en inglés, donde las variantes se forman en función del accionamiento y la composición del generador. En el segundo y tercer ejemplo, los CN del inglés surgen al indicar la ubicación donde se mide la tensión (p. ej. *line-to-line voltage, phase voltage*). En cambio, el español añade otra dimensión posible desde la que enfocar el concepto: la composición de la tensión según el número de elementos en los que se mida, de manera que también surgen las denominaciones *tensión compuesta* y *tensión simple*, respectivamente.

En esta línea, es común que la variante principal[59] de cada lengua enfatice la misma perspectiva, de modo que el término más extendido en ambas lenguas destaque la misma información, si bien hay casos en los que las distintas lenguas se decantan por dimensiones diferentes (Tabla 52).

[59] Para seleccionar la variante principal atendimos a la mayor frecuencia en el corpus. Sin embargo, el orden del resto de variantes no atiende a criterios específicos, pues estos se estudiarán en próximos trabajos.

Términos en inglés	Términos en español
wind turbine rotor (variante principal) wind rotor	rotor eólico (variante principal) rotor del aerogenerador
small wind turbine (variante principal) SWT small wind system small-scale wind turbine	aerogenerador de baja potencia (variante principal) aerogenerador de pequeña potencia aerogenerador pequeño aerogenerador de pequeño tamaño turbina eólica de baja potencia
large wind turbine (variante principal) commercial wind turbine utility-scale wind turbine large-scale wind turbine	aerogenerador de gran potencia (variante principal) aerogenerador de gran tamaño aerogenerador comercial
stall-regulated wind turbine (variante principal) stall-controlled wind turbine fixed-pitch wind turbine passive stall-regulated wind turbine	aerogenerador de paso fijo (variante principal) aerogenerador de regulación por pérdida aerodinámica aerogenerador de regulación pasiva por pérdida aerodinámica aerogenerador regulado por pérdida aerodinámica aerogenerador de paso de pala fijo

Tabla 52: Comparación de variantes principales con dimensiones diferentes en inglés y español

Por una parte, *wind turbine rotor* alude a la entidad de la que forma parte el rotor, mientras que la variante principal en español señala el recurso empleado (*wind rotor*). En *small wind turbine* y *large wind turbine* se destaca el tamaño del aerogenerador, que alude de forma implícita a su potencia, si bien esta es la información preferida en español (*aerogenerador de baja potencia, aerogenerador de gran potencia*). Por último, en *stall-regulated wind turbine* se subraya el mecanismo de regulación, mientras que en español se prefiere la variante que indica la composición que permite dicha regulación (*aerogenerador de paso fijo*).

Hasta ahora nos hemos sumergido en el análisis contrastivo de la variación denominativa cognitiva, en concreto aquella que entraña cambios en la selección de las dimensiones. Sin embargo, este tipo de variación también puede producirse cuando se modifica la configuración del concepto, como sucede en las variantes *generated power* y *power generation*, que también se dan en español (*energía generada, generación de energía*). Asimismo, hablamos de variación denominativa cognitiva cuando la categoría semántica del concepto sufre cambios, por ejemplo, derivados del empleo de un núcleo

más general, como comentamos anteriormente (*wind turbine, wind power generation system* > *aerogenerador, sistema de generación eólica*).

Otro tipo de variantes que también podríamos catalogar de cognitivas son aquellas en las que se recurre a la metonimia; es decir, se percibe el mismo concepto de distintos modos, aludiendo a la parte por el todo. Un ejemplo son las variantes *turbine hub height, turbine rotor height* y *turbine tower height*, cuyos equivalentes también utilizan este recurso (*altura del buje del aerogenerador, altura del rotor del aerogenerador, altura de la torre del aerogenerador*). En estos términos, se juega con la alternancia entre *torre* (*tower*), *rotor* (*rotor*) y *buje* (*hub*) para indicar la altura del centro del rotor de la turbina (el buje) con respecto al suelo.

Además de cognitiva, la variación denominativa también puede ser formal, lo que depende en gran medida de las características de la lengua, por lo que se deben cuidar los calcos inapropiados. Por ejemplo, las variantes pueden derivarse de cambios gráficos, como el empleo del guion, muy habitual en los CN estudiados en inglés (p. ej. *variable speed generator, variable-speed generator*). No obstante, el abundante recurso a los calcos para trasladar los términos al español ocasiona errores lingüísticos relacionados con estos elementos formales. Por ejemplo, al calcar el término inglés *line-to-neutral voltage*, se obtiene en español el término *tensión fase-neutro*, ampliamente extendido. Sin embargo, este equivalente constituye un calco estructural, ya que en español no es común modificar un sustantivo con otros dos sustantivos unidos por un guion.

Del mismo modo, el empleo de los guiones en español difiere de su uso en inglés y, por tanto, no es apropiado en el término *generador de inducción auto-excitado*. Dicho equivalente surge al calcar la estructura del inglés *self-excited induction generator*, aunque en español los prefijos no deben seguirse de guiones a no ser que se enlacen con una mayúscula, una sigla o un número. Dado que no es el caso en este término, la escritura adecuada debe ser *generador de inducción autoexcitado*. Otro error ortográfico surge al adoptar la grafía *diesel* del inglés, sin incluir la tilde necesaria en español. De este modo, la traducción correcta de *diesel generator* es *generador diésel* y no **generador diesel*. Un caso similar es el de los CN que incluyen el término *stator* en inglés. La tilde en español es opcional (*estátor, estator*), si bien abunda el uso de la variante que no la utiliza.

Además de los cambios gráficos, las variantes formales también pueden deberse a cambios morfológicos, como la alternancia entre modificadores en singular y en plural que a menudo se da en español. En este sentido, se deben detectar los casos en los que el modificador en inglés alude a un plural, ya que este no se refleja explícitamente en inglés,

pero puede materializarse en español (*generador de imanes permanentes, fabricante de aerogeneradores, tensión en bornes, energía de las olas*, etc.). No obstante, hay excepciones como *generador multipolo*, cuyo modificador debería adoptar una forma plural, pero mantiene el singular como resultado de un calco del inglés *multipole generator*.

La variación morfológica también surge a raíz de los cambios en las preposiciones que intervienen en los CN, o como consecuencia del empleo o la supresión de artículos. Ambas características son más propias del español que del inglés, debido a las estructuras de postmodificación, que implican el uso de preposiciones y artículos. Por último, la variación formal puede ser de tipo morfosintáctico, como la alternancia entre sintagmas preposicionales y adjetivos relacionales (*viento en el mar, viento marino*), que hemos comentado al desarrollar la formación morfosintáctica de los términos en ambas lenguas.

Por su parte, la variación léxica, que consiste en la sustitución de elementos de los CN por sinónimos que no constituyen variantes formales, se produce de manera similar en los CN de ambas lenguas. Además, muchos de los sinónimos de una lengua encuentran su equivalente directo en la otra: *production/generation* > *producción/generación*, *fixed-speed wind turbine/constant-speed wind turbine* > *aerogenerador de velocidad fija/constante*, *converter/inverter* > *convertidor/inversor*, *voltage control/voltage regulation* > *control de tensión/regulación de tensión* o *fluctuation/variation* > *fluctuación/variación*. También constituyen variantes léxicas las que se derivan del uso geográfico. Estas no destacaron en nuestra muestra, a excepción del término *voltage sag*, que se utiliza en inglés americano, y su variante *voltage dip*, propia del inglés británico.

Otro tipo de variantes denominativas son las variantes por reducción, cuya presencia fue muy elevada en los CN analizados en ambas lenguas. Este tipo de variantes surgen mediante diferentes mecanismos, entre los que se encuentra la sustitución del CN por su sigla. A este respecto, cabe señalar el amplio uso de las siglas inglesas en español, que a menudo se prefieren frente a las siglas españolas. Algunos ejemplos de siglas del inglés de reconocido uso en español son *DFIG* (del inglés *doubly-fed induction generator*, en detrimento de las siglas *GIDA* y *GADA* del español), *SCIG* (*squirrel cage induction generator*), *PMSG* (*permanent magnet synchronous generator*), *VSC* (*voltage source converter*), *VSI* (*voltage source inverter*) o *PSD* (*power spectral density*), todas ellas más empleadas que las posibles siglas españolas. En la Figura 31 se presentan ejemplos en los que se refleja el uso de estos términos reducidos en el corpus en español. Recordemos que, a pesar de constituir préstamos, al tratarse de siglas no se deben escribir en cursiva.

Query dfig|scig|pmsg|vsc|vsi|psd 288 > Positive filter 10 (2.76 per million)

doc#12	. Figura 5 Velocidad de referencia aplicada al	PMSG	, con carga resistiva. Resultados de
doc#24	de viento en ese rango. La integral de la	PSD	(f) de la velocidad de viento para todas las
doc#32	menor potencia que los requeridos en el caso del	PMSG	. La plataforma desarrollada en este trabajo se
doc#61	de operación es más complejo que el de un	SCIG	, se considera que este tipo de generador tiene
doc#66	, se incluye una delta de Dirac en la	PSD	en f=0 Hz. 2.2 Caracterización del
doc#70	a través de un convertidor, como en el caso del	DFIG	, como podemos ver en la Figura 1.9. Figura: 1.9.
doc#102	se presenta un análisis del comportamiento del	DFIG	ante un hueco de tensión, En el apartado D se
doc#106	, conectado a la red de transporte a través de un	VSC	y un transformador en serie (ver Fig. 3-19). El
doc#106	implementar la etapa de potencia de un	VSC	en MT. Es por esto que el STATCOM suele ser
doc#121	se podría conectar al bus de continua del	VSC	un convertidor DC-DC bidireccional para

Figura 31: Ejemplos de siglas en inglés empleadas en español

El uso de siglas en inglés está más extendido que el de sus adaptaciones al español porque se trata de los términos más empleados en la comunicación experta internacional, que permiten acceder de forma rápida y concisa al concepto. No obstante, también hay casos en los que la sigla en español prima sobre la versión inglesa, como sucede principalmente en AT (*alta tensión*), MT (*media tensión*) y BT (*baja tensión*). Del mismo modo, hay siglas muy extendidas en inglés que, en español, prefieren adoptar su forma extensa. Algunos ejemplos de ello son *AC generator* > *generador de corriente alterna*, *DC generator* > *generador de corriente continua*, *DC voltage* > *tensión continua* y *PCC* > *punto de conexión a la red*. Por tanto, en la posible sustitución de un CN por su sigla, conviene estudiar detenidamente el empleo de estas variantes reducidas en la lengua origen y la lengua meta.

Otro ejemplo de variación por reducción lo constituye la sustitución del CN por su hiperónimo, en lo que se conoce como *anáfora elíptica*. Este fenómeno es bastante común en las dos lenguas (p. ej. *wind turbine* > *turbine*, *potencia máxima* > *potencia*), si bien no recogimos estas variantes ya que, para remitir al mismo concepto que el CN, precisan de su uso en el mismo texto. Por tanto, no serían apropiadas para su inclusión en un recurso terminográfico.

Por último, la reducción léxica también es una fuente importante de variación en inglés y español. Este tipo de variación engloba diferentes disminuciones del término, sin llegar a reducirlo únicamente al núcleo. Así, lo más frecuente es que se supriman formantes intermedios, como se observa en los siguientes ejemplos: *wind power plant/wind plant* > *planta de energía eólica/planta eólica* o *generator electromagnetic torque/generator torque* > *par electromagnético del generador/par del generador*.

También pueden suprimirse elementos del inicio o el final del CN, como el núcleo o el último modificador. Por ejemplo, en la sustitución de *direct current* por *direct* se ha

suprimido el núcleo, algo que también encontramos en español (*corriente continua* > *continua*) y en los CN que integran este CN (*DC voltage/direct voltage* > *tensión DC/tensión continua*). Por su parte, en el paso de *tensión de la red eléctrica* a *tensión de red* se ha elidido el último modificador (además del artículo), una variación que también se produce en inglés (*electric grid voltage/grid voltage*).

Estas reducciones léxicas pueden dar lugar a variantes cuyo significado no es composicional, como se observa en *renewable generator* (*generador renovable*) y *wind generation* (*generación eólica*). En el primer caso, lo renovable no es el generador, sino la fuente que utiliza y la energía que produce; mientras que, en el segundo, el carácter de *eólico* no pertenece directamente a la generación, sino a la energía resultante.

En general, se omite la información más evidente, mientras que se mantienen las diferencias, al ser esta la información que distingue al CN del resto de elementos de su categoría. Por ejemplo, cuando se sustituye *wind power plant* por *wind plant* no se pierde información fundamental, ya que se sobreentiende que una central genera energía, mientras que el rasgo distintivo es el uso del viento para la generación. De este modo, se consiguen términos más manejables, dada su menor extensión. Por tanto, se trata de un recurso habitual que encontramos de forma paralela en inglés y español.

Sin embargo, aunque ambas formas sean posibles en las dos lenguas, puede suceder que en una de ellas se prefiera la reducción léxica, mientras que en la otra esté más extendida la forma extensa del término. Esto se observa, por ejemplo, en las siguientes variantes principales en inglés y español: *wind penetration* > *penetración de la energía eólica*, *wind turbine application* > *aplicación eólica* o *line-to-line voltage* > *tensión de línea*. En definitiva, las aparentes similitudes entre estas lenguas no deben confundirnos en cuanto a las preferencias de uso que pueden ser propias de cada una de ellas.

Otro modo de generar variantes por reducción léxica consiste en utilizar una parte del CN en su forma completa y, otra, a modo de sigla (p. ej. *direct current voltage/DC voltage*). Cuando esto sucede en español, las siglas que intervienen suelen ser las del inglés (*tensión DC*). Un tipo de formación parecido a este consiste en utilizar una sigla y precederla de un núcleo que especifica la clase (p. ej. *generador DFIG*). Como vemos, de nuevo se prefieren en español las siglas inglesas. La diferencia con el caso anterior es que, en este tipo de términos, el núcleo forma parte de la sigla, pero se explicita para aclarar la pertenencia a una categoría. La sigla, por su parte, es el término más extendido y contribuye a especificar el núcleo. Por tanto, se trata de una variante más transparente que el recurso directo a la sigla (*generador DFIG* en lugar de *DFIG*). A pesar de que este

tipo de composición se da en inglés y en español, no siempre existen correspondencias en las dos lenguas (p. ej. mientras que *AC generator* se da ampliamente en inglés, en español esta forma no es habitual y se prefiere hablar de *generador de corriente alterna*).

En definitiva, la variación denominativa constituye un fenómeno de elevada presencia en la mayoría de conceptos de la muestra. Sin embargo, de nuestros ejemplos se desprende que, así como existen correspondencias, también pueden darse desigualdades entre las dos lenguas, por lo que el estudio contrastivo de estos mecanismos es esencial.

La variación conceptual, por su parte, también debe tenerse en cuenta en la traducción de CN entre el inglés y el español. Así, el análisis de los CN polisémicos en inglés y su traducción al español reveló que la polisemia no siempre se reproduce en la lengua meta. Por una parte, diversos términos contaban con equivalentes en español que también presentaban una naturaleza polisémica. Por ejemplo, *wind power* y *wind energy* pueden traducirse por *energía eólica* y *energía del viento*. En ambas lenguas, estos términos pueden referirse a una propiedad del viento o a la electricidad resultante de su utilización. Lo mismo sucede con *wind generator* y *wind turbine generator*, que en español pueden traducirse por *generador eólico* y, en las dos lenguas, aluden tanto a un generador como a la entidad superior de la que forman parte: un aerogenerador.

Hay otros casos en los que la polisemia del inglés también se percibe en español, donde además, uno de los sentidos presenta otras variantes denominativas. Por ejemplo, *wind power production, wind power generation* y *wind generation* son variantes que pueden indicar el proceso de generación de energía eólica y la energía resultante. En español, los términos *generación eólica* y *producción de energía eólica*, entre otros similares, también expresan estos dos significados. Además, en el sentido de energía eólica resultante pueden emplearse otros términos como *energía eólica producida* o *electricidad de origen eólico*.

Por su parte, *wave power* se traduce por *energía de las olas* y *energía undimotriz*, términos que se utilizan para designar una propiedad de las olas o la energía resultante de su aprovechamiento. En este último sentido también se utiliza el término *energía olamotriz*. Lo mismo sucede en *offshore wind power* y *offshore wind energy*, que se traducen por *energía eólica marina* y *energía eólica offshore* para indicar una propiedad del viento que sopla en el mar o bien la electricidad producida al utilizarla. En cambio, cuando nos referimos a la propiedad del viento en el mar también podemos emplear los términos *energía del viento en el mar* y *energía eólica en el mar*. Por último, *power generation* y *power production* pueden traducirse por *generación de energía* o *producción*

de energía para indicar el proceso de generación y la energía resultante de dicho proceso. Además, para este último sentido se utilizan también términos como *energía producida* o *electricidad generada*.

Por otra parte, es posible que la polisemia no se reproduzca en la lengua meta. Un ejemplo lo encontramos en el término *wind system*, que puede referirse a un aerogenerador o a un sistema de vientos de la circulación atmosférica. En el primer sentido recibe diferentes variantes, entre ellas *sistema eólico*, mientras que en el segundo caso el equivalente es *sistema de vientos*. De este modo, observamos cómo la variación morfosintáctica puede entrañar también cambios de sentido.

Por tanto, cuando nos encontremos ante un término polisémico en la lengua origen, debemos llevar a cabo comprobaciones para cerciorarnos de si la equivalencia es también polisémica. El análisis contextual ayuda a revelar el sentido ante el cual nos encontremos. De este modo, mediante el procedimiento de traducción indicado en la Sección 5.2.1, conseguiríamos descubrir el equivalente adecuado, sin caer en posibles calcos inapropiados. Por ejemplo, como hemos indicado, *wind power production, wind power generation* y *wind generation* pueden referirse al proceso de formación (Figura 32) o a la energía resultante (Figura 33).

Query wind, power, production|generation 1,594 > Positive filter 7 > Positive filter 6 (1.66 per million)

doc#8	will describe the power system and large-scale wind power production . We will then look at the effects of wind power
doc#27	wind power consumptive capacity, improve wind power generation efficiency, and reduce thermal power output as
doc#59	Rate (%) of Installed Wind Capacity 1999-2012. Wind power generation is a fast-growing industry. It is one of the
doc#71	and market conditions. Development of wind power generation has been popular in Denmark since the 1970's
doc#80	identified as the second most suitable site for wind power generation . EWT DW52-500 with 50 m hub height and EWT
doc#83	. Abstract. An internal gear ring for wind power generation is an important basic part of transmission

Figura 32: Concordancias de *wind power production*, *wind power generation* y *wind generation* cuando aluden al proceso de formación

Como se aprecia en la Figura 32, estos términos se refieren al proceso de formación cuando coocurren con elementos como la envergadura de los proyectos (*large-scale wind power production*), su eficiencia (*wind power generation efficiency*), su carácter industrial (*wind power generation is a fast-growind industry*), el lugar donde puede producirse (*site for wind power generation*) o los instrumentos que intervienen en el proceso (*an internal gear ring for wind power generation*).

Query wind, power, production|generation 1,594 > Positive filter 11 (3.04 per million)

doc#7	produced by the wind farm. As the penetration of	wind power generation	increases (in terms of the overall energy mix),
doc#7	or high winds. High penetration levels of	wind power production	affect the operation of the transmission
doc#7	operating cost increases by about €1 -4/MWh of	wind generation	. This is typically 10 per cent or less of the
doc#7	. The other curves show increasing levels of	wind power production	: the 150-500 MW curve shows a situation with low
doc#7	of the demand coverage. In practice, almost all	wind generation	, amounting to 95 per cent of the total
doc#8	power system in order to be able to sell the	wind power production	: . Wind power requirement 1 (WP1): similar to
doc#8	that there will be a certain amount of	wind power production	available during a peak load situation, which
doc#8	(PW) and minimum consumption (PD), and zero	wind power production	(PW) and maximum consumption (PD). This
doc#8	Zeeland and Jutland) in January 2000. Average	wind power production	in January was 687MW and in 2000 the yearly
doc#45	for the 2020 plant mix. The critical point for	wind generation	in 2015 is 2,994MW and in 2020 is 3,705MW,
doc#45	three test years, the optimal penetration of	wind generation	exceeded 30% of electricity generated in this

Figura 33: Concordancias de *wind power production*, *wind power generation* y *wind generation* cuando aluden a la energía resultante

Por otro lado, cuando estos términos aluden a la energía resultante del proceso de generación, se relaciona con conceptos como la penetración, que alude a la proporción de un tipo de energía dentro del total generado (*penetration of wind power generation, high penetration levels of wind power production, the optimal penetration of wind generation exceded 30 % of electricity generated*), lo que puede hacerse con ella (*sell the wind power production*), su cantidad (*increasing levels of wind power production, amount of wind power production, zero wind power production*) y, especialmente, las unidades con las que se mide (*MWh of wind generation, average wind power production in January was 687 MW, the critical point for wind generation in 2015 is 2,994 MW*).

Al consultar sus traducciones literales en el corpus en español (Figura 34), constatamos que estos términos también presentan un carácter polisémico. Esto se desprende de su combinación con elementos como sus unidades de medida (*MWh, kWh, 39.000 MW, 5.481 MW*) o su porcentaje del total de la energía eléctrica (*lo que representa la generación eólica en el total de generación eléctrica*) cuando el sentido es el de energía generada. Por el contrario, cuando se subraya el proceso de formación, se indican las tecnologías e instalaciones que intervienen en dicho proceso (*tecnologías de generación eólica, instalaciones para la generación eólica, centrales de generación eólica*), su carácter de actividad (*actividades de generación eólica*) o el lugar donde puede producirse (*emplazamientos óptimos para la producción de energía eólica*), entre otros aspectos.

Query generación|producción, de, energía, eólico 810 > Positive filter 11 (3.04 per million)

doc#9	que se ha pasado de un LCOE estimado para la	generación eólica	de 200 €/MWh en los años 80 a los niveles actuales
doc#11	es si lo comparamos con lo que representa la	generación eólica	en el total de generación eléctrica de cada país
doc#14	o ensayo de nuevos prototipos y tecnologías de	generación eólica	. ——La investigación del recurso eólico. Como
doc#14	a las actividades e instalaciones para la	generación eólica	de energía eléctrica en cualquiera de sus
doc#14	de regular el tratamiento de las actividades de	generación eólica	a través de sus instrumentos de gestión,
doc#17	. A finales de 2003, la capacidad mundial de	generación de energía eólica	había alcanzado unos 39.000 MW, de los que
doc#19	eléctrico presente en instalaciones de	generación de energía eólica	ante las perturbaciones más habituales. En
doc#19	eléctrico presente, tanto en centrales de	generación eólica	como en entornos industriales, frente a dichas
doc#23	eólica. Los emplazamientos óptimos para la	producción de energía eólica	no son abundantes por lo que la fabricación de
doc#23	un factor de correlación de 0,99. Donde: E es la	producción de energía eólica	(kWh), x es la densidad del aire húmedo (kg/m3),
doc#24	, Castilla y León lidera la presencia de	generación eólica	con 5.481 MW, seguida por Castilla-La Mancha

Figura 34: Concordancias de *generación eólica*, *generación de energía eólica* y *producción de energía eólica* con ambos significados

Frente a este ejemplo, en el caso de *wind system*, donde la polisemia no se reproduce en español, no se aprecia la coincidencia de significados en las formas *sistema eólico* ni *sistema de vientos*, sino más bien el empleo de cada forma para aludir a un significado diferente (Figura 35). De este modo, *sistema eólico* se entiende como aerogenerador, al relacionarse con conceptos como la *potencia,* la *red*, la *energía*, su *diseño* o su *instalación*; mientras que *sistema de vientos* coocurre con conceptos como *estacional, local, global* o *mistral*.

Query sistema, de, eólico|viento 173 > Positive filter 10 (2.76 per million)

doc#16	a prevalecer las bajas presiones cálidas. Los	sistemas de vientos	estacionales asociados a estos patrones de
doc#16	y en Zonda en los Andes. Otros ejemplos de	sistemas de viento	locales son el Mistral, que sopla a lo largo del
doc#24	las fluctuaciones de potencia activa de	sistemas eólicos	conectados a red. Con el objetivo de establecer
doc#27	la información de los lugares y el diseño del	sistema eólico	usando las herramientas estadísticas e
doc#31	se traducen en variaciones de presión. De los	sistemas de vientos	globales, uno de los más importantes es el de los
doc#31	de suministrar energía a la red con pequeños	sistemas eólicos	. Esto es aplicable en los casos en que exista una
doc#48	se genera en la región por la instalación de un	sistema eólico	. • Estimar, a partir de un estudio de opinión, el
doc#69	de mar-tierra y la brisa valle-montaña. Estos	sistemas de vientos	locales se desarrollan como resultado del
doc#70	de doble devanado en el estátor aplicado en	sistemas eólicos	de velocidad variable. ACRÓNIMOS,
doc#88	con la red. 2.1.5. Modelo de la conexión a red. El	sistema eólico	se conecta a la red eléctrica a través del

Figura 35: Concordancias de *sistema eólico* y *sistema de vientos*

En resumen, constatamos que, a pesar de que la polisemia puede darse por igual en la lengua origen y la lengua meta, es fundamental cerciorarse de ello mediante un análisis detallado para así evitar caer en usos inapropiados.

En el análisis comparativo de los CN y sus equivalentes en español resulta indiscutible el influjo del inglés en la formación de términos en español. Ello evoca la idea de neología, al tratarse probablemente de conceptos que se denominaron inicialmente en inglés para seguir la tendencia de la comunicación internacional y que requirieron después su traducción al español para suplir un vacío denominativo. Aunque la tesis de la neología requiere un estudio diacrónico (Cabezas García y Faber 2017c) que no se llevó a

cabo en este trabajo al no figurar entre sus objetivos, son diversos los factores que apuntan al carácter neológico de los términos y que observamos en nuestra muestra.

Uno de ellos es el empleo de préstamos o formas léxicas del inglés. Estas se importaron sin cambios, ya que se presupone que el lector está familiarizado con la lengua inglesa y el campo temático. Los préstamos del inglés cumplían dos funciones en nuestro corpus. Por una parte, desempeñaban una función aclaratoria, cuando se indicaba el término en español y, a modo de aclaración, el término o su sigla en inglés, al ser esta la referencia más extendida y que favorece la comprensión del concepto. En la Figura 36 se presentan algunos ejemplos, en los que se utiliza el término *stall* y las siglas *HAWT* y *VAWT* para aclarar los conceptos.

Query hawt|vawt|stall 84 > Positive filter 8 (2.21 per million)

doc#16	aerogeneradores por pérdida aerodinámica, "	stall	controlled". Los aerogeneradores de
doc#23	con turbinas eólicas de eje horizontal (	HAWT	). Sin embargo, es necesario desarrollar
doc#23	sentido, los aerogeneradores de eje Vertical (	VAWT	) son claramente más ventajosos en comparación
doc#74	eje horizontal (Horizontal Axis Wind Turbine,	HAWT	). Descripción de códigos para el estudio
doc#90	pueden clasificar en turbinas de eje vertical (	VAWT	) y turbinas de eje horizontal (HAWT). Entre las
doc#92	de 10º,"pitch", para ajustar mejor el perfil de	stall	(pérdida) en la zona de altas velocidades del
doc#106	aerodinámica o control pasivo (en inglés "	Stall	Control"). □ Control de potencia por variación
doc#107	en tierra firme. Pérdida aerodinámica (	stall	): Separación física de la capa límite fluida (

Figura 36: Ejemplos de préstamos en español

Asimismo, los préstamos pueden ejercer una función de denominación, cuando se utilizan como términos en español, sin necesidad de recurrir a otros términos para clarificar el uso extranjero. Este caso ya se comentó al respecto de las siglas (p. ej. *PMSG*), aunque también se da en términos en su forma completa, que intervienen en los CN en español. Los principales ejemplos se encuentran en la dicotomía *offshore/onshore* (p. ej. *aerogenerador offshore, parque eólico onshore*). Estos suelen intervenir como modificadores en los CN, en los que funcionan a modo de *differentiae*. Aunque el corpus no permite observar su tipografía, los préstamos deben escribirse en cursiva (a excepción de las siglas), ya que no son formas léxicas del español. Cabe señalar que, aunque ambas opciones sean válidas, se prefiere la variante que incluye el término en español (p. ej. *parque eólico terrestre* en lugar de *parque eólico onshore*).

Los calcos, de elevada presencia en los CN estudiados, también constituyen una prueba de la influencia del inglés. Además de los calcos estructurales, en los que ya nos hemos detenido, estos pueden ser calcos de expresión, como *aerogenerador de gran potencia*. Como sucede cuando intervienen participios en los CN, a menudo conviven las variantes surgidas a partir de calcos con otras formas naturalizadas (Tabla 53).

Término en inglés	Variante en español producida a partir del calco	Variante naturalizada en español
small wind turbine	aerogenerador de pequeña potencia	aerogenerador de baja potencia
line-to-neutral voltage	tensión fase-neutro	tensión de fase tensión simple
generator torque	torque del generador	par del generador

Tabla 53: Ejemplos de variantes producidas por calco y por naturalización en español

Como se observa en la Tabla 53, en la que se presentan algunos ejemplos de la convivencia de estas variantes, *aerogenerador de pequeña potencia* constituye un calco de expresión del inglés, ya que en español la potencia no se califica con adjetivos de tamaño (*pequeña*), sino de altura (*baja*). Se constata, así, la importancia de la fraseología en el empleo idiomático del lenguaje especializado. Por su parte, *tensión fase-neutro* constituye un calco estructural, ya que la estructura N N-N no es propia del español. Por último, *torque del generador* representa un calco de expresión del inglés al emplear el núcleo *torque*, de grafía idéntica al inglés, que cuenta con un equivalente ampliamente extendido en español (*par*).

En general se prefieren las variantes naturalizadas en español, aunque el calco es bastante utilizado al percibirse como una opción segura al traducir. En este sentido, las distintas posibilidades de traducción redundan también en un mayor grado de variación. El empleo de calcos derivó también en la extensión de metáforas como las empleadas en *squirrel cage induction generation* y *wind penetration* al español, donde se utilizan los equivalentes *generador de inducción de jaula de ardilla* y *penetración eólica*, respectivamente. No obstante, como venimos comentando, la reproducción de estas formas en la lengua meta debe pasar por una comprobación exhaustiva.

Cabe señalar que los calcos a menudo derivan en errores ortográficos en español, como el empleo de guiones ya comentado. Asimismo, pueden generar traducciones que, a priori, parecen posibles en español, ya que los elementos formantes existen y además intervienen en otros CN similares, pero que no son adecuadas. Un ejemplo de ello es el uso de **tensión terminal* como traducción de *terminal voltage*. El adjetivo *terminal* figura en español y, además, existen términos válidos de forma similar, como *tensión nominal*. Sin embargo, el CN **tensión terminal* es un calco inapropiado, que solo encontramos en ocurrencias del mismo texto en nuestro corpus, lo que apunta a usos propios (incorrectos) de un autor. En su lugar, puede emplearse como equivalente el término *tensión en bornes*.

En esta línea, la dificultad para hallar equivalentes documentados también puede ser un indicio del carácter neológico de un término. Ello genera un elevado uso de paráfrasis explicativas para nombrar el concepto en español, como se observa en *aerogenerador de regulación pasiva por pérdida aerodinámica* o *aerogenerador de paso de pala fijo*. Estas paráfrasis resultan patentes, además, en expresiones como *los de regulación por cambio del ángulo de paso* o *los de paso fijo*, que aluden a dos subtipos por medio de explicaciones. Dichas estructuras van ligadas también a una elevada inestabilidad, como resultado de la convivencia de variantes producidas por calcos con sus formas naturalizadas, el uso inestable de preposiciones y artículos, etc. En definitiva, todos ellos, factores que apuntan hacia el carácter neológico de muchos de los CN estudiados.

La neología es precisamente uno de los motivos que pueden llevar a la necesidad de producir un equivalente. Cuando no se consigue identificarlos, el traductor o terminólogo deberá realizar una propuesta de denominación del concepto en cuestión. A la vista del análisis comentado en esta sección, presentamos a continuación unas pautas generales para la traducción de CN en inglés y español en el área de la energía eólica.

Estos términos compuestos suelen trasladarse en forma de CN en ambas lenguas, que frecuentemente no superan los dos o tres constituyentes. Los CN de mayor extensión también son posibles, pero menos habituales. Frente a esto, se debe atender a las estructuras de modificación de cada lengua: la premodificación habitual del inglés, en la que al núcleo nominal le preceden sustantivos o adjetivos; y la postmodificación propia del español, donde el núcleo va seguido de adjetivos o sintagmas preposicionales. Estos modificadores pueden combinarse entre sí en cada lengua, de forma que se generen, por ejemplo, CN del tipo N Adj Sprep en español (p. ej. *par electromagnético del generador*). Además, pueden intervenir otros modificadores como participios o siglas en las que se condensen varios modificadores (p. ej. *PCC voltage*).

En esta línea, cuando el CN en inglés está formado por una sucesión de sustantivos, estos suelen adquirir la forma de adjetivos o sintagmas preposicionales al trasladarse al español (*generator torque control* > *control del par del generador*). Debe adoptarse la perspectiva inversa cuando la traducción se realiza hacia el inglés (*proyecto de energía eólica* > *wind power project*). Por tanto, en español se permiten estructuras de tipo más explicativo que en la lengua inglesa. Cuando se emplean sintagmas preposicionales en estas lenguas, las preposiciones suelen ser *de* y *of*. También es posible el uso de otras preposiciones más específicas, aunque las anteriores suelen ser la opción más recurrente en todos los casos. En los CN en los que intervenga más de una preposición, se puede usar

de en ambos casos o bien alternarla con el empleo de otra preposición (p. ej. *tensión en el punto de conexión a la red*).

En cuanto al uso de adjetivos, estos suelen seguir inmediatamente al elemento al que modifican, por ejemplo, insertándose tras el núcleo (p. ej. *velocidad media del viento*). Este uso es diferente en inglés, donde el adjetivo no siempre aparece directamente junto al elemento modificado (p. ej. *rated wind speed*). En español, los adjetivos deben concordar en género y número con el sustantivo al que complementan. Por tanto, dicha concordancia permite relacionar a ambos elementos, por lo que no es necesario que aparezcan juntos cuando el resto de formantes presenta un género o número diferente (p. ej. *generador de inducción autoexcitado*). No obstante, la aparición conjunta es la opción más frecuente. Respecto a la concordancia de número, se debe prestar atención al uso de los plurales tanto en adjetivos como en sustantivos, ya que es común que la marca del plural se refleje en español, pero no en inglés (p. ej. *terminal voltage* > *tensión en bornes*). Asimismo, algunos adjetivos habituales del dominio presentan un uso particular. Es el caso de *bajo, medio* y *alto*, que a menudo se colocan delante del sustantivo, y *large* y *small*, que explicitan la alusión a la potencia en español (p. ej. *aerogenerador de baja potencia*).

Por otra parte, cuando se emplean siglas en español, estas suelen adoptarse en forma de préstamos del inglés, a excepción de *AT* (*alta tensión*), *MT* (*media tensión*) y *BT* (*baja tensión*), que prefieren su forma en español. También hay otros préstamos del inglés de amplio uso en español, como son *offshore* y *onshore*. Al tratarse de formas propias de otra lengua, deben escribirse en cursiva, a excepción de las siglas. Del mismo modo, los calcos del inglés están bien aceptados en el campo de la energía eólica en español, aunque estos suelen convivir con variantes naturalizadas más extendidas. Aunque estos datos pretenden orientar en la producción de CN en inglés y, especialmente, en español, la labor de trasvase lingüístico en la que nos centramos no puede prescindir de un profundo análisis semántico y contextual en ambas lenguas.

En este capítulo nos hemos detenido en la traducción de los CN. Dado que estos no siempre figuran en los recursos terminográficos, o no se incluye toda la información que el traductor o terminólogo pueda requerir, resulta fundamental el desarrollo de técnicas de corpus para identificar los equivalentes de estos términos compuestos. Con tal fin, desglosamos un procedimiento de búsqueda basado en la semántica distribucional que puede seguirse para localizar equivalentes de los CN en un corpus. Además de estudiar su identificación, también comparamos estos términos compuestos con sus equivalentes, lo que reveló un notable influjo del inglés en los términos en español, además de semejanzas

y diferencias en la formación terminológica en las dos lenguas. Basándonos en estas observaciones, propusimos unas guías para la traducción de CN en inglés y en español.

5.3 La representación de compuestos nominales: el módulo fraseológico de compuestos nominales de EcoLexicon

La base de conocimiento terminológica (BCT) EcoLexicon reconoce el papel fundamental de la fraseología en el lenguaje especializado por medio de un módulo fraseológico en el que se recogen colocaciones verbales del dominio del medio ambiente (Sección 2.2.3). El estudio llevado a cabo en esta tesis cuenta entre sus objetivos la remodelación de dicho módulo fraseológico, con vistas a incluir también los CN. Presentamos en esta sección el nuevo módulo de CN, que junto con el apartado de colocaciones verbales, conformará el módulo fraseológico de EcoLexicon.

Para el diseño de este módulo, que se encuentra en proceso de implementación en EcoLexicon, tuvimos en cuenta el tratamiento de los CN en otros recursos terminográficos (Sección 3.6.2), tanto aquellos aspectos que desde nuestro punto de vista constituyen aciertos, como los puntos que convendría evitar. Asimismo, quisimos plasmar las cuatro líneas principales de estos términos compuestos, que pueden resultar de mayor interés para los usuarios de una base de conocimiento terminológica multilingüe como es EcoLexicon. Estos aspectos son (i) la productividad en la formación de CN, (ii) su traducción, (iii) su morfosintaxis y (iv) su semántica. En estas vertientes se basan las cuatro vistas del módulo de CN, además de la vista resumen, y por tanto, la información que debe consignarse en la plantilla de introducción de datos. Así, siguiendo las nuevas tendencias en lexicografía y terminografía, se diseñaron distintas vistas para ofrecer diferentes puntos de acceso a la información y satisfacer los intereses variados de los usuarios de EcoLexicon, que van desde traductores e intérpretes profesionales hasta estudiantes de traducción, terminólogos, redactores técnicos o especialistas del medio ambiente.

Así pues, una vez determinada la información que se quería consignar en el módulo, su diseño se estructuró en dos ejes principales: la plantilla de introducción de datos del terminólogo (Sección 5.3.1) y la interfaz de usuario (Sección 5.3.2).

5.3.1 Plantilla de introducción de datos

Tras delimitar el tipo de datos que recibirá el usuario en el módulo de CN, elaboramos la plantilla que emplearán los terminólogos de EcoLexicon para introducir dicha información fraseológica de los CN. Esta plantilla estará disponible en el acceso privado de EcoLexicon, en las entradas de los términos compuestos.

Dado que en el módulo de CN se desgranan los aspectos morfosintácticos y semánticos de cada CN, decidimos que, en la plantilla, debían consignarse las dependencias internas o *bracketing*, así como categorías semánticas, roles semánticos, dimensiones, categorías gramaticales, relaciones semánticas internas y ejemplos de uso, información en la que nos detendremos a continuación.

Estos datos se agrupan según los elementos del CN a los que se refieren: CN completo, constituyentes individuales o grupos del *bracketing*. En la Tabla 54 se presenta la plantilla rellena con la información de *permanent magnet synchronous generator*, que emplearemos para describir el formulario.

Categoría de datos 1.er nivel	Ejemplo	Categoría de datos 2.º nivel	Ejemplo	Tipo de datos y procedencia
1. Compuesto nominal	permanent magnet synchronous generator			Obligatorio. Lista de selección de términos ya presentes en la BCT.
2. Concepto	GENERADOR SÍNCRONO DE IMANES PERMANENTES			Obligatorio (se extrae del concepto).
3. Categoría semántica	ENTITY>CREATION>ARTIFACT> INSTRUMENT> TRANSFORMING INSTRUMENT			Obligatorio (se extrae del concepto).
4. Dimensión	ROTOR AND STATOR SPEED			Opcional. Texto libre con lista acumulativa.
4. Dimensión	EXCITATION			Opcional. Texto libre con lista acumulativa.
[Botón: Añadir dimensión]				
5. Elemento 1 (núcleo)	generator			Obligatorio. Lista de selección de cada palabra que conforme el término.
		5.1 Categoría semántica	ENTITY>CREATION>ARTIFACT> INSTRUMENT> TRANSFORMING INSTRUMENT	Obligatorio. Heredada del concepto al que este término pertenezca (si figura en la BCT). Si no está incluido, lista de selección de categorías semánticas.
		5.2 Categoría gramatical	nombre común	Obligatorio. Heredada del término si figura en la BCT. Si no está incluido, lista de selección de categorías gramaticales.
		5.3 Rol semántico	-	Opcional. Lista de selección de roles semánticos.
5. Nexo 1	-			Opcional. Lista de selección de cada palabra que conforme el término.
		5.1 Categoría gramatical	-	Obligatorio si el anterior se ha rellenado. Lista de selección de categorías gramaticales.
[Botón: Añadir nexo]				
5. Elemento 2 (modificador)	synchronous			Obligatorio. Lista de selección de cada palabra que conforme el término.

		5.1 Categoría semántica	ATTRIBUTE>TIME	Ídem
		5.2 Categoría gramatical	adjetivo	Ídem
		5.3 Rol semántico	-	Ídem
[Botón: Añadir elemento/nexo]				
5. Elemento 3 (modificador)	magnet			Opcional. Lista de selección de cada palabra que conforme el término.
		5.1 Categoría semántica	ENTITY>MATTER>SOLID MATTER>MATERIAL>MINERAL	Ídem
		5.2 Categoría gramatical	nombre común	Ídem
		5.3 Rol semántico	-	Ídem
[Botón: Añadir elemento/nexo]				
5. Elemento 4 (modificador)	permanent			Opcional. Lista de selección de cada palabra que conforme el término.
		5.1 Categoría semántica	ATTRIBUTE>TIME	Ídem
		5.2 Categoría gramatical	adjetivo	Ídem
		5.3 Rol semántico	-	Ídem
[Botón: Añadir elemento/nexo]				
6. *Bracketing* 1	synchronous generator			Opcional. Texto libre.
		6.1 Categoría semántica	ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT	Obligatorio si se ha rellenado el campo 6. Heredada del concepto al que este término pertenezca (si figura en la BCT). Si no está incluido, lista de selección de categorías semánticas.
		6.2 Categoría gramatical	SN	Obligatorio si se ha rellenado el campo 6. Heredada del término si figura en la BCT. Si no está incluido, lista de selección de categorías gramaticales.
		6.3 Rol semántico	POSSESSOR	Obligatorio si se ha rellenado el campo 6. Lista de selección de roles semánticos.
		6.4 Relación semántica	generator ***has_attribute*** synchronous	Opcional. Consta de tres campos: texto libre + lista de selección con las relaciones de EcoLexicon, con posibilidad de añadir nuevas + texto libre.
		[Botón: Añadir relación]		
		[Botón: Extraer información]		
6. *Bracketing* 2	permanent magnet			Opcional. Texto libre.
		6.1 Categoría semántica	ENTITY>MATTER>SOLID MATTER>MATERI	Ídem

			AL>MINERAL	
		6.2 Categoría gramatical	SN	Ídem
		6.3 Rol semántico	THEME	Ídem
[BOTÓN: Añadir *bracketing*]		**6.4 Relación semántica**	magnet ***has_attribute*** permanent	Opcional. Consta de tres campos: texto libre + lista de selección con las relaciones de EcoLexicon, con posibilidad de añadir nuevas + texto libre.
		[Botón: Añadir relación]		
		[Botón: Extraer información]		
7. Relación semántica principal	synchronous generator ***has_part*** permanent magnet			Obligatorio. Consta de tres campos: lista de selección con los grupos del *bracketing* o los elementos si solo hay dos formantes + lista de selección con las relaciones de EcoLexicon, con posibilidad de añadir nuevas + lista de selección con los grupos del *bracketing* o los elementos si solo hay dos formantes.
8. Ejemplos de uso	The shaft is connected via a brake to the permanent magnet synchronous generator, PMSG, the electrical output from the PMSG is connected to an inverter which performs an AC to DC conversion.			Obligatorio. Texto libre. (Acceso a Sketch Engine en la visualización).

Tabla 54: Plantilla de introducción de datos

En la primera columna se muestran las categorías de datos del primer nivel. Se trata de ocho tipos de datos: 1. Compuesto nominal, 2. Concepto, 3. Categoría semántica, 4. Dimensión, 5. Elementos individuales del CN (Elemento 1, Elemento 2, etc. además de los posibles elementos que funcionen como nexos), 6. Agrupaciones internas o grupos del *bracketing* (*Bracketing 1* y *Bracketing 2*), 7. Relación semántica principal y 8. Ejemplos de uso. Como se observará, la plantilla se distribuyó de forma que se introdujera la parte del CN a la que se refiere la información (Compuesto nominal, Elemento o *Bracketing*) y, a continuación, sus datos. De este modo, algunos de estos campos de primer nivel se desglosan en un segundo nivel de datos, como veremos a continuación.

En el campo *1. Compuesto nominal* se debe introducir el CN en cuestión. Se trata de una lista de selección que se despliega según los caracteres que se van introduciendo y muestra los CN ya presentes en EcoLexicon. Así, la idea es que se consigne información

fraseológica de términos que figuren en la base de conocimiento. Por supuesto, se trata de un campo obligatorio.

Los siguientes campos se refieren a la forma completa del término compuesto. Por una parte, los campos *2. Concepto* y *3. Categoría semántica* son obligatorios y se extraen del concepto en cuestión una vez indicado el CN, no es necesario introducirlos. Los nombres de los conceptos se indican en español a modo de designación del concepto expresado, cuando se introduce el concepto en la base de conocimiento. Ocurre igual con la categoría semántica del concepto, si bien esta se adapta al idioma en cuestión. El campo *4. Dimensión* es opcional, dado que no se trata de un aspecto que siempre sea claro, y permite elegir una dimensión de la lista así como introducir texto libre. Esto es, se mostrará una lista acumulativa con las dimensiones ya introducidas en la base de conocimiento, aunque también se podrá introducir una dimensión nueva en el instante, que se añadirá a la lista acumulativa. Este campo se podrá volver a añadir si el CN evoca más de una dimensión, haciendo clic en el botón *Añadir dimensión*. Es el caso de *permanent magnet synchronous generator*, que alude a la velocidad de giro del rotor y el estátor (*synchronous*) y también a su modo de excitación (*permanent magnets*). Además, cabe señalar que la dimensión es un dato relativo al término y no al concepto, pues las diferentes variantes de un mismo concepto pueden enfatizar dimensiones diferentes (p. ej. *gas turbine* alude al recurso empleado, mientras que *combustion turbine* indica el proceso que tiene lugar en su interior).

Tras estos campos se muestran los relativos a cada constituyente individual del CN: *5. Elemento 1 (núcleo)*, *5. Elemento 2 (modificador)*, etc. El Elemento 1 se concibe como el núcleo, ya se encuentre a la izquierda o a la derecha del CN. Se trata de un campo obligatorio que despliega una lista de selección con las palabras que conforman el CN, para evitar posibles erratas si se permitiera texto libre. Cada elemento establece un segundo nivel de datos: *5.1 Categoría semántica, 5.2 Categoría gramatical* y *5.3 Rol semántico*.

El campo *5.1 Categoría semántica* es obligatorio y su información se extrae del concepto al que este término pertenezca, si está incluido en EcoLexicon. En los términos polisémicos, se indicarán las categorías posibles al terminólogo, quien elegirá la correcta en cada caso. Si, por el contrario, el concepto no figura en la base de conocimiento, se mostrará una lista de selección con las categorías semánticas. El campo *5.2 Categoría gramatical* también es obligatorio. La categoría gramatical del término se extrae automáticamente si este figura en EcoLexicon. De lo contrario, se mostrará una lista de selección con las categorías gramaticales: nombre común, nombre propio, adjetivo,

adverbio, verbo, preposición, artículo, SN, participio pasado y participio presente. A continuación, el campo *5.3 Rol semántico* es opcional, para permitir que, en CN de más de dos formantes, se indiquen los roles de los grupos y no de los formantes individuales, pues es habitual que los roles de cada elemento no guarden relación con los roles de los grupos, que son los principales. Sin embargo, en los CN de dos formantes se indicará que deben consignarse los roles de cada elemento. Los roles se pueden seleccionar de una lista de selección que incluye el inventario presentado en la Sección 4.2.3. A diferencia de las categorías semánticas y gramaticales, los roles no pueden extraerse de los conceptos ya presentes en la base de conocimiento, ya que estos dependen de cada proposición conceptual.

Al igual que el Elemento 1 (núcleo), los siguientes formantes del CN despliegan los mismos campos en los que se escribe el formante, esta vez junto a la etiqueta de Elemento 2 (modificador), Elemento 3 (modificador), etc., además de sus categorías semántica y gramatical, y su rol semántico. De estos elementos, el único obligatorio es el Elemento 2; de lo contrario, no nos encontraríamos ante un término compuesto. Los elementos adicionales pueden añadirse pulsando sobre el botón *Añadir elemento/nexo* que aparece después del segundo elemento.

Entre ellos también puede haber nexos, tales como preposiciones o artículos. Por tanto, se muestra el campo *5. Nexo 1* como opcional. En *permanent magnet synchronous generator* no existe ningún nexo, pero en un CN como *impact of wind power*, debería consignarse *of* en ese campo. Para ello, se despliega una lista de selección de cada palabra que conforme el término. En un segundo nivel se muestra el campo de su categoría gramatical, de obligada cumplimentación si se ha rellenado el Nexo 1 con algún valor. Así, se desplegará la siguiente lista de selección: preposición, artículo, artículo contraído, signo (p. ej. la barra en CN como *output/input ratio*). El botón *Añadir nexo* que se encuentra tras el primer nexo, así como el botón *Añadir elemento/nexo* que aparece tras cada elemento o nexo, permiten añadir otros nexos (p. ej. para *generador de inducción de jaula de ardilla*). Estos serán también opcionales. Como en el Nexo 1, su categoría gramatical será obligatoria si se ha indicado algún valor en el nexo.

Tras consignar la información de cada elemento, en los CN de más de dos formantes se deberán introducir los datos relativos al *bracketing*. Así, los campos *6. Bracketing 1* y *6. Bracketing 2* se refieren a los dos grupos del *bracketing*, que se establecen con el objetivo de simplificar la estructura de estos CN extensos a su forma básica de núcleo y modificador. Como ocurre con la numeración de los elementos, el *bracketing* 1 es

la agrupación que designa al núcleo, mientras que el *bracketing* 2 se entiende como modificador. Los campos *6. Bracketing 1* y *6. Bracketing 2* son opcionales, ya que solo deberán introducirse en CN de más de dos constituyentes. Se trata de campos de texto libre en los que se activará un control de erratas que comprueba que lo que se escribe coincide con parte de la cadena del CN.

A continuación se muestran los campos de segundo nivel relativos a cada grupo del *bracketing*: *6.1 Categoría semántica, 6.2 Categoría gramatical, 6.3 Rol semántico* y *6.4 Relación semántica*. La categoría semántica, gramatical y el rol semántico son obligatorios si se ha rellenado el campo 6, es decir, si se ha indicado que existe *bracketing*. Las categorías semántica y gramatical funcionan como en los elementos individuales, esto es, se heredan del concepto o término en caso de que este figure en la base de conocimiento, o bien se despliega una lista de selección de categorías gramaticales o semánticas si el concepto o término no está incluido en EcoLexicon. El rol semántico se indica mediante una lista de selección.

La novedad de los campos de *bracketing* es el punto *6.4 Relación semántica*, que permite la identificación de la relación entre los elementos que conforman el *bracketing*, es decir, no se trata de la relación principal del CN. Este campo es opcional, pues en CN de tres elementos, por ejemplo, uno de los grupos del *bracketing* puede estar formado por un único elemento, p. ej. *small* en *small* [*wind turbine*]. Si se desea indicar una relación dentro de un grupo del *bracketing*, se utilizarán tres campos. Uno de ellos es un campo de texto libre en el que se activará un control de erratas que compruebe que lo que se escribe coincide con parte de la cadena del CN; otro es una lista de selección con las relaciones de EcoLexicon (con una única direccionalidad), a las que se podrán añadir nuevas relaciones que surjan en los CN; y el último consiste, de nuevo, en un campo de texto libre con control de erratas. Esto nos permitirá introducir proposiciones conceptuales del tipo término + relación + término.

Bajo la relación semántica del *bracketing* se muestra el botón *Añadir relación*, con el que podremos introducir una segunda relación dentro de un grupo del *bracketing*, cuando estos agrupen tres o más elementos, p. ej. [*offshore wind farm*] *design* (farm *uses_resource* wind y wind farm *located* offshore). Por otra parte, es posible que uno de los grupos del *bracketing* esté formado por un único elemento (*rated* [*wind speed*]). Así, para ahorrar tiempo en la cumplimentación de la plantilla, se presentará junto a cada grupo el botón *Extraer información*, que abrirá una lista de selección en la que se podrá elegir el elemento individual cuya información se quiere extraer (*rated*). Además, bajo el

bracketing 2, el botón *Añadir bracketing* nos permitirá consignar un grupo adicional del *bracketing* para CN cuya estructura no encaje en el esquema binario de núcleo y modificador.

Por último, tras haber indicado la información propia de los elementos o los grupos del *bracketing*, se debe indicar la relación principal que subyace en el CN. Se trata de un campo obligatorio, que despliega tres listas de selección: una en la que se muestran los grupos del *bracketing* (que deberán elegirse en los CN de tres o más elementos) o los elementos del CN (en el caso de los CN de dos constituyentes); otra en la que se muestran las relaciones de EcoLexicon, a las que se podrán añadir nuevas que figuren en los CN; y una nueva lista de selección con los grupos del *bracketing* o los elementos del CN. Como vemos en la Tabla 54, la relación principal en *permanent magnet synchronous generator* es synchronous generator *has_part* permanent magnet.

También se deberán indicar ejemplos de uso del CN, que se incluirán en un campo de texto libre obligatorio. Dado que esperamos poner el corpus en acceso abierto en Sketch Engine, se ha habilitado un botón que permitirá al usuario obtener ejemplos de uso en dicha herramienta. Sin embargo, por el momento se incluirán manualmente los ejemplos, a la espera de la disponibilidad del corpus en Sketch Engine.

Como se ha observado, esta plantilla recoge los datos de los CN que se mostrarán en las diferentes vistas del módulo. Para garantizar su validez, la plantilla se probó con CN en inglés y en español de diferente forma y extensión. No obstante, la futura incorporación de información fraseológica en otras lenguas puede entrañar su modificación. A continuación presentamos la interfaz de usuario del módulo de CN, en la que se aprecia cómo toman forma los datos consignados.

5.3.2 Interfaz de usuario

El usuario de EcoLexicon podrá acceder al módulo fraseológico mediante dos rutas. Por un lado, si se encuentra en la entrada de un concepto podrá hacer clic en un término para acceder al módulo fraseológico. Por otra parte, en la interfaz general de EcoLexicon, podrá seleccionar la pestaña *Phraseological module/Módulo fraseológico*[60], que estará ubicada junto a la opción *Search concordances/Buscar concordancias*. En ambos casos, tras indicar si desea acceder al módulo de colocaciones verbales o al de CN y seleccionar esta segunda opción, se le mostrará por defecto la interfaz de búsqueda de la vista *Summary/Resumen*.

[60] La nomenclatura del módulo de CN se adapta en función del idioma elegido en la interfaz (por el momento, inglés o español), que puede modificarse siempre que se desee.

Esta cuenta con un cajetín de texto libre[61] en el que se puede insertar el término o términos que se deseen en cualquiera de los idiomas (por el momento, inglés y español).

Como veremos en la Sección 5.3.2.5, los resultados de esta vista constituyen un resumen de las cuatro vistas específicas. Esta se muestra como vista inicial para ofrecer al usuario una panorámica de la información que encontrará en cada apartado (formación de CN, términos equivalentes e información morfosintáctica y semántica). A partir de ahí, podrá continuar su búsqueda seleccionando una vista específica.

Se puede acceder a ellas en la parte superior de la interfaz, donde se presentan cinco pestañas que dan acceso a las diferentes vistas del módulo (Figura 37) y entre las que se puede cambiar cuando se desee: (i) *CN formation/Formación de CN*, (ii) *Equivalents/Equivalentes*, (iii) *Morphosyntactic combinations/Combinaciones morfosintácticas*, (iv) *Semantic combinations/Combinaciones semánticas* y (v) *Summary/Resumen*.

Figura 37: Parte superior de la interfaz para acceder a las vistas del módulo de CN

Antes de profundizar en cada vista, nos detendremos en la organización de los resultados que se presentarán al usuario. Este observará que, en todas las vistas, los CN se estructuran en dos grupos. En primer lugar se muestran los CN que cuentan con el término de búsqueda en el núcleo y, en segundo lugar, los que lo incluyen entre sus modificadores, pues las diferentes vistas permiten búsquedas de los CN generados a partir de uno o más términos. En caso de que en la búsqueda no se incluya ningún término, sino una combinación de categorías semánticas o gramaticales, así como roles o relaciones, no se dará esta organización en núcleo y modificador, sino un orden alfabético, dado que no se habrá incluido ningún término de partida[62].

Así pues, cuando se incluyen uno o varios términos de consulta, el sistema identificará los CN que tienen esa secuencia (en el mismo orden y sin elementos de por medio) como núcleo si esta se ha consignado como Elemento 1 o *Bracketing* 1 en la

[61] Los cuadros de texto libre del sistema identifican cualquier ocurrencia del lema introducido (p. ej. si se escribe un adjetivo en femenino singular, se buscan sus ocurrencias en cualquier género o número).
[62] El sistema establecerá los apartados de núcleo y modificador cuando se incluya un término de búsqueda en el campo de texto libre. De lo contrario, no mostrará estos dos enunciados.

plantilla de introducción de datos. En ese caso, presentará el CN en cuestión en el primer grupo de CN, que se denominará *el término o términos + as a CN head/como núcleo* (p. ej. *Generator as a CN head/Generator como núcleo*). Por el contrario, si la secuencia introducida no constituye el Elemento 1 o el *Bracketing* 1 de la plantilla, se entenderá que forma parte de los modificadores del CN y, por tanto, figurará bajo el apartado *el término o términos + as a CN modifier/como modificador* (p. ej. *Generator as a CN modifier/Generator como modificador*).

Al margen de la organización en núcleo y modificador, las diferentes vistas presentan alteraciones en la organización de los resultados. Así, las vistas *CN formation/Formación de CN* y *Equivalents/Equivalentes* muestran los CN agrupados por conceptos, en los que se incluyen también otras variantes denominativas del mismo concepto a pesar de que no respondan a los criterios indicados (por ejemplo, siglas o términos simples). No se extraen, sin embargo, los conceptos en los que ninguna variante cumpla con los requisitos de búsqueda. La agrupación conceptual de estas dos vistas se plantea con el objetivo de mostrar los *slots* que abre un término para denominar un mismo concepto, así como las diferentes variantes que puede adquirir un CN. Esta organización por conceptos es, además, la base para el establecimiento de equivalencias interlingüísticas.

No obstante, esta agrupación de conceptos puede entrar en conflicto con la estructuración según núcleo y modificador. Así, puede suceder que un concepto se agrupe en la sección de núcleo atendiendo a la forma de una de sus variantes y a la vez presente otra variante en la que ese término no figure en el núcleo, sino entre los modificadores (p. ej. *generator* en las variantes *electric generator* y *generator system*). Dado que estamos a favor de la organización según núcleo y modificador por sus beneficios para observar el potencial de formación de CN y su cercanía a la estructuración semántica, la solución que adoptamos en esos casos fue la duplicación del concepto en el grupo de núcleo y el de modificador.

Por el contrario, en las vistas *Morphosyntactic combinations/Combinaciones morfosintácticas, Semantic combinations/Combinaciones semánticas* y *Summary/Resumen* no se agrupan las variantes en conceptos, pues más que en el concepto designado por el CN, estas vistas se centran en la caracterización morfosintáctica y semántica de los CN. En el caso de la vista *Summary/Resumen* no se agrupan las variantes dado que los resultados solo incluyen algunos CN a modo de ejemplo. Por tanto, a diferencia de las vistas

anteriores en las que las variantes sí se agrupan por conceptos, en estos casos solo se extraen variantes que cumplan con los requisitos de consulta indicados.

Así pues, la organización conceptual de las secciones *CN formation/Formación de CN* y *Equivalents/Equivalentes* facilita la agrupación de los hipónimos. Es decir, los conceptos subordinados denominados por algún CN que reúna los requisitos de búsqueda se presentan sangrados bajo su hiperónimo. De este modo, si buscamos por ejemplo *generator + X*, uno de los CN que obtendremos será *generator voltage*, bajo el que se encontrará su hipónimo *generator terminal voltage*, que también cumple con la consulta introducida. Para ello, el sistema buscará los conceptos que establezcan la relación *type_of* con cada concepto extraído y que cuenten, además, con variantes denominativas que cumplan las condiciones de búsqueda (p. ej. *generator + X*), pues no nos interesa extraer todos los hipónimos si estos no responden a nuestra consulta. De este modo, presentar los hipónimos bajo sus hiperónimos contribuye a relacionar los conceptos y obtener una mejor imagen del sistema conceptual.

En estas dos vistas, también se propone una organización por dimensiones, que se presenta únicamente en el grupo de CN cuyo núcleo es la secuencia introducida (p. ej. CN cuyo núcleo es *generator*). Es decir, los diferentes conceptos extraídos se organizan por dimensiones, según la dimensión que enfatice el término principal de cada concepto[63], que se habrá consignado en la plantilla de introducción de datos. Por ejemplo, en función de la velocidad a la que giren el rotor y el estátor, los generadores pueden ser *synchronous generator* o *asynchronous generator*; según el recurso que utilicen para generar movimiento, pueden ser *wind generator*, *diesel generator*, *steam generator*, etc.

Para ofrecer esta organización, el sistema detecta la dimensión que enfatiza el término principal de cada concepto (p. ej. ROTOR AND STATOR SPEED, RESOURCE). Seguidamente, agrupa los conceptos cuya variante principal enfatiza la misma dimensión (*wind generator*, *diesel generator* y *steam generator* aluden a la dimensión RESOURCE) y los ordena alfabéticamente según esa variante principal. Las distintas dimensiones se presentan también por orden alfabético. Aunque habría sido interesante presentar las dimensiones y los conceptos por orden de relevancia, la complejidad de estas tareas nos llevó a posponerlas a fases posteriores de perfeccionamiento de la herramienta.

[63] A pesar de que utilizamos la dimensión de la variante principal para aludir a la dimensión enfatizada por el concepto, somos conscientes de que las distintas variantes pueden evocar dimensiones diferentes. En estudios posteriores profundizaremos en este aspecto para ofrecer una solución más específica.

También es preciso señalar que los CN que explicitan varias dimensiones, como suele ser habitual en los CN extensos, se incluyen como hipónimo de cada uno de sus hiperónimos, para mostrar las diferentes jerarquías conceptuales que pueden establecerse. Por ejemplo, *permanent magnet synchronous generator* es hipónimo tanto de *permanent magnet generator* (según el tipo de excitación) como de *synchronous generator* (según la velocidad de giro del rotor y el estátor). A pesar de que esta organización puede generar repeticiones, nos parece que la indicación de los diferentes hiperónimos constituye un valor añadido que favorece la comprensión de los conceptos.

Por otra parte, dado que las dimensiones se establecen a nivel de término, decidimos presentar también esta organización dimensional en la vista *Semantic combinations/Combinaciones semánticas*, pues nos parecía un enfoque interesante relacionar la semántica de los formantes del CN con la dimensión enfatizada en cada caso. De nuevo, las dimensiones y los términos se organizan alfabéticamente.

Esta estructuración en dimensiones nos permitió detectar nuevos CN o términos simples que no habíamos extraído dada nuestra metodología, pero que serían necesarios de cara a la representación completa de las dimensiones en EcoLexicon. Por ejemplo, uno de los CN con los que trabajamos es *downwind turbine*, que alude a un tipo de aerogenerador definido según la posición del rotor (a sotavento). Durante su análisis, descubrimos que cuenta con un antónimo, *upwind turbine*, que no figuraba entre nuestros términos, pero cuya inclusión sería necesaria para representar por completo la dimensión. Se espera también que los próximos trabajos en los que hagamos uso de este corpus, así como de otros relacionados con el medio ambiente, permitan ampliar las dimensiones representadas.

Por su parte, el grupo de CN que incluye la secuencia entre sus modificadores no presenta una organización por dimensiones, sino un orden alfabético, dado que la variedad de núcleos no posibilita esta estructuración dimensional (p. ej. *generator torque, generator speed, generator terminal voltage*). En la Tabla 55 se resume la organización de la información en las diferentes vistas. Como se desprende de este enfoque, la apuesta por la organización según la posición del término (núcleo o modificador), la jerarquía de conceptos y las dimensiones activadas representa el modo en que los conceptos se almacenan en el cerebro y, por tanto, facilita la adquisición de conocimiento.

Vista	Organización
CN formation/Formación de CN	-CN con el término en el núcleo Términos agrupados en conceptos, ordenados por dimensiones y, en cada dimensión, ordenados alfabéticamente. Hipónimos sangrados bajo cada hiperónimo. -CN con el término en el modificador Términos agrupados en conceptos y ordenados alfabéticamente. Hipónimos sangrados bajo cada hiperónimo.
Equivalents/Equivalentes	
Semantic combinations/Combinaciones semánticas	-CN con el término en el núcleo Términos ordenados por dimensiones y, en cada dimensión, ordenados alfabéticamente. -CN con el término en el modificador Términos ordenados alfabéticamente.
Morphosyntactic combinations/Combinaciones morfosintácticas	-CN con el término en el núcleo Términos ordenados alfabéticamente. -CN con el término en el modificador Términos ordenados alfabéticamente.
Summary/Resumen	
*Orden alfabético si no se introduce ningún término de partida.	

Tabla 55: Organización de datos en las vistas del módulo de CN

Además de la organización de la información, otro aspecto que comparten todas las vistas es el cuadro de búsqueda directa, que se encuentra en la parte superior derecha de la interfaz. En él se pueden introducir uno o más términos, que se resaltarán en los resultados obtenidos, para así poder localizarlos con mayor facilidad. Estos elementos de búsqueda no tienen por qué contar con una entrada propia en EcoLexicon y tampoco es necesario que aparezcan juntos en los CN de la lista. Por ejemplo, si buscamos *permanent generator*, se señalarían estos elementos dentro de los CN en los que figuran:

permanent magnet generator

permanent magnet synchronous generator

Esta funcionalidad puede resultar interesante, por ejemplo, para la correcta identificación de CN, dado que no es extraño que se identifiquen secuencias incompletas como CN (p. ej. se podría pensar en *magnet generator* como el CN, cuando lo correcto sería *permanent magnet generator*).

En los siguientes apartados detallamos el proceso de búsqueda y obtención de resultados, aplicado a cada vista. Como veremos, en cada una de ellas se puede elegir entre realizar una búsqueda simple o avanzada, si bien ambas posibilidades presentan el mismo tipo de información. Los métodos de consulta coinciden en gran medida en las diferentes vistas, por lo que los desglosamos más detenidamente en la vista *CN formation/Formación de CN* (Sección 5.3.2.1), al ser el primer apartado que se describe.

5.3.2.1 Vista CN formation/Formación de CN

Esta vista muestra la formación de CN a partir de uno o varios términos (*power*: *wind power system, power output*, *wind power,* etc.). Como en cada vista, el usuario podrá elegir entre realizar una búsqueda simple o avanzada, haciendo clic en la opción que le interese (Figura 38). Si selecciona la búsqueda simple, se mostrará un cajetín de texto libre en el que podrá introducir uno o varios términos en cualquiera de los idiomas (por el momento, inglés y español), p. ej. *generator* (Figura 38). Esta búsqueda arrojará todos los CN que incluyen *generator* entre sus formantes, como veremos a continuación.

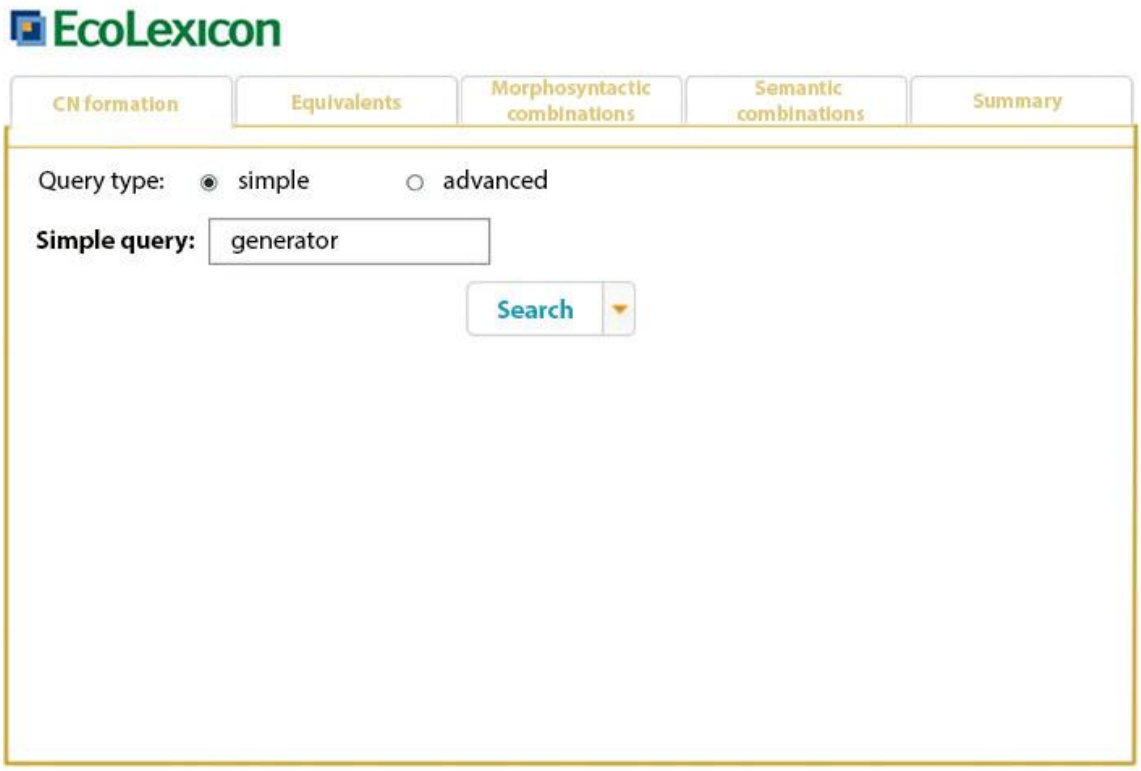

Figura 38: Interfaz de búsqueda simple de la vista *CN formation/Formación de CN*

Si en la fase de consulta el usuario desea cambiar y realizar una búsqueda avanzada, podrá hacer clic en esa opción en la parte superior de la interfaz. Entonces, en la búsqueda avanzada se presentan dos burbujas: una de texto libre en la que podremos introducir uno o varios términos y otra denominada *Any term(s)/Cualquier término(s)*. Ambas podrán añadirse tantas veces como se desee, pulsando en el botón + inferior, además de arrastrarse y colocarse en el lugar que se prefiera, diseñándose así una consulta a medida. Del mismo modo, las nuevas burbujas añadidas podrán eliminarse en el símbolo X que se mostrará en la esquina superior derecha. Así, por ejemplo, podremos centrarnos en obtener los CN cuyo núcleo es *generator* (Figura 39).

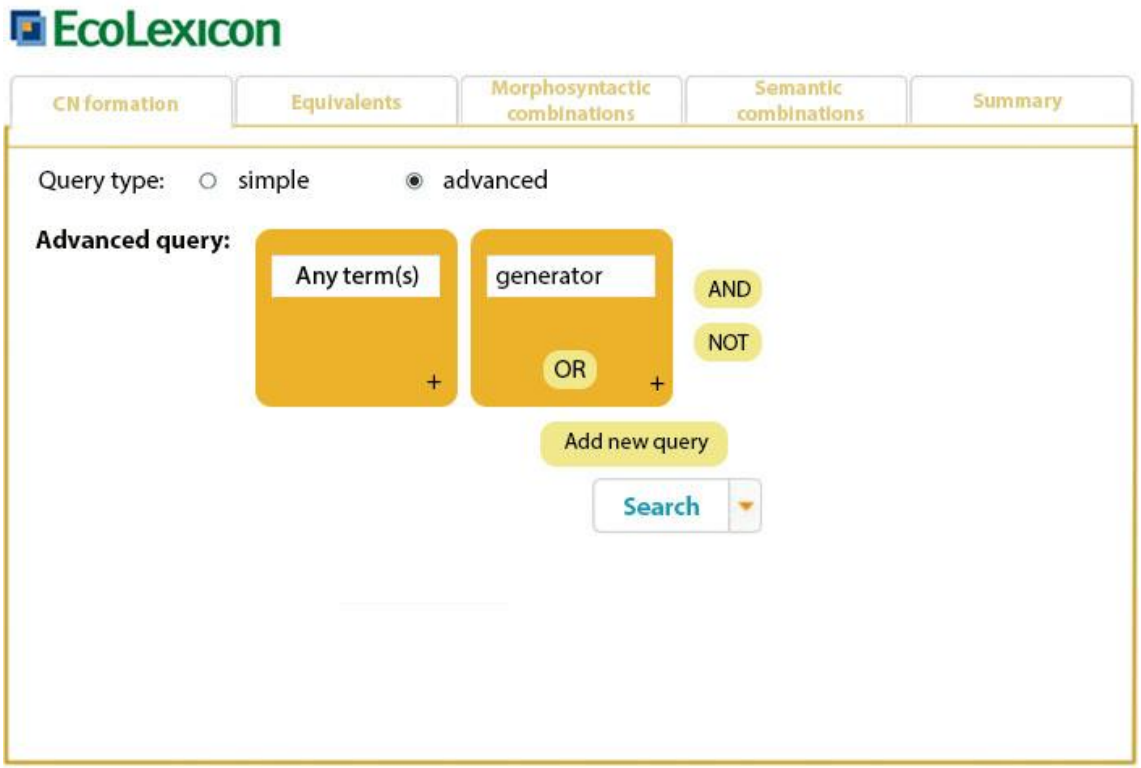

Figura 39: Interfaz de búsqueda avanzada de la vista *CN formation/Formación de CN*

Como vemos, la función de la búsqueda avanzada es permitir especificar la posición del término, aunque también ofrece la posibilidad de indicar diferentes posiciones (p. ej. los CN en los que *generator* interviene como núcleo y también como modificador). Para ello, se debe hacer clic en el botón *Add new query/Añadir nueva consulta* que figura en la parte inferior derecha y despliega una línea adicional de búsqueda con las dos burbujas.

La burbuja *Any term(s)* mostrará cualquier número de elementos que acompañen al término en esa posición. Tomamos esta decisión considerando que el usuario puede sobreentender que se mostrarán uno o más elementos, de manera que preferimos esta

opción, a pesar de que se puedan ofrecer más resultados de los esperados. La otra opción consistiría en la especificación del número de formantes mediante la repetición de la burbuja, lo que omitiría información que podría resultar de interés, como términos de una extensión diferente. Por tanto, la posibilidad de añadir más de una vez la burbuja *Any term(s)* se plantea con el objetivo de que pueda colocarse a ambos lados del término, ya que su duplicación para representar dos elementos contiguos arrojará los mismos resultados que si apareciese una única vez. Lo mismo sucede con la burbuja de texto libre, en la que se pueden insertar varios términos, que se buscarán sin ningún elemento de por medio. Su duplicación se plantea, por tanto, con vistas a incluir varios términos separados entre sí.

Como se aprecia también en la Figura 39, podremos combinar nuestra búsqueda con marcadores booleanos. Estos permiten dos tipos de búsquedas diferentes, pues se incluyen tanto dentro de la burbuja de texto libre en la que podremos introducir los términos como fuera de las burbujas. Dentro de la burbuja se puede usar el marcador OR, que permite buscar y comparar los CN generados por varias unidades, por lo que abre un campo de texto libre. Por ejemplo, un generador y un motor son máquinas eléctricas que pueden convertir ya sea la energía mecánica en energía eléctrica, en el caso del generador, o al contrario, en el caso del motor. Por tanto, nos puede interesar comparar los CN formados por *generator* y *motor* para explorar si esta diferencia de significado se plasma en su combinatoria. No incluimos el marcador AND dentro de la burbuja ya que arrojaría los mismos resultados que OR, ni tampoco NOT, pues no tiene sentido hacer una restricción a nivel de un formante, sino más bien a nivel de todo el CN, como veremos más adelante. En la Figura 40 se desglosan las consultas que permite el otro conjunto de marcadores booleanos: los que se encuentran junto a las burbujas y permiten hacer restricciones a nivel de todo el CN.

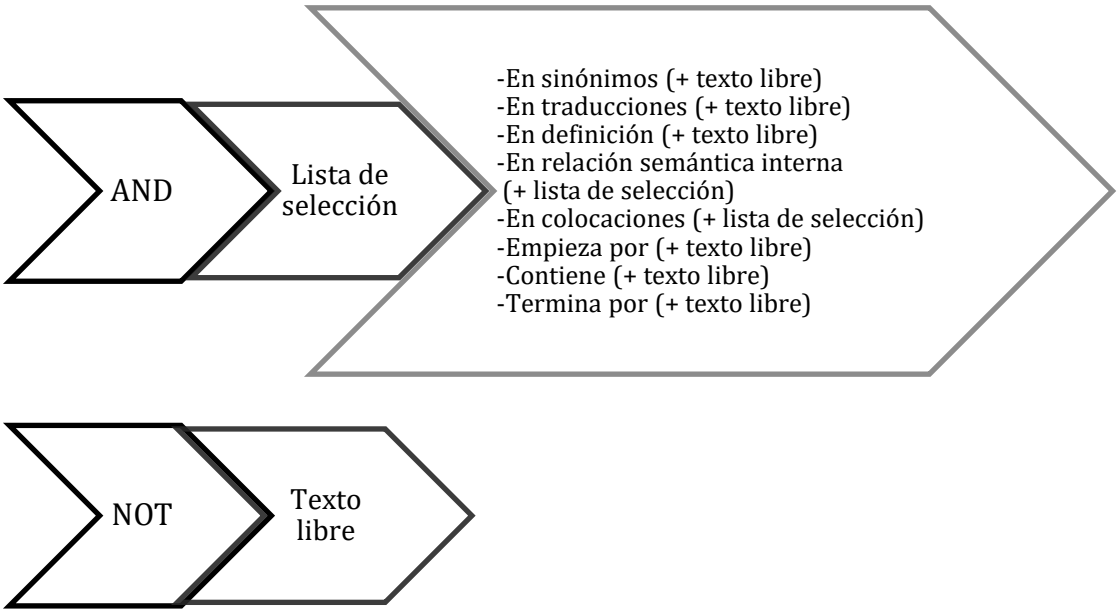

Figura 40: Consultas posibles con los marcadores booleanos ubicados junto a las burbujas

Como se observa en la Figura 39, junto a las burbujas también se presentan los marcadores AND y NOT, que son los marcadores que resultan funcionales para aplicarse a la consulta completa (por ejemplo, OR no tendría cabida en este caso, pues desempeñaría la misma función que el botón *Add new query/Añadir nueva consulta*). Estos, al seleccionarse, se señalan en un tono más oscuro (Figura 41). Tras AND se activa una lista de selección elaborada según la información del módulo fraseológico que podría servir como campo de búsqueda: *in synonyms/en sinónimos*, *in translations/en traducciones*[64], *in definition/en definición*, *in internal semantic relation/en relación semántica interna*[65], *in collocations/en colocaciones*[66], *starts with/empieza por*, *contains/contiene* y *ends with/termina por*. Se ha excluido, por tanto, la información que a nuestro juicio no sería objeto de una búsqueda específica o que ya se ofrecería en las otras opciones, como sería el caso de las búsquedas *in usage examples/en ejemplos de uso*, *in notes/en notas* e *in concept entry in EcoLexicon/en la entrada de concepto en EcoLexicon*. Así, una consulta podría ser la siguiente: *generador + Cualquier término(s) AND en definición: agua*, para obtener generadores que guarden alguna relación con el agua.

[64] Las búsquedas *en sinónimos* y *en traducciones* se diferencian en que *en sinónimos* detecta la lengua de consulta y busca entre sus variantes, mientras que *en traducciones* busca las variantes en las otras lenguas.
[65] Despliega una lista de selección de las relaciones identificadas en el interior de los CN.
[66] Como hemos comentado, esta funcionalidad está aún por determinar, ya que la conexión entre el módulo de CN y el de colocaciones verbales todavía no se ha hecho efectiva. No obstante, se espera que el modo de restricción podría basarse en una lista de selección con los dominios léxicos de Faber y Mairal (1999).

Una vez seleccionado un elemento de la lista de selección que despliega AND se abren tres posibilidades: (i) aparece el campo de texto libre o la lista de selección que acompañe al elemento elegido (p. ej. *Any term(s) + generator AND in definition: induction*, como puede observarse en la Figura 41), (ii) el operador NOT, al que seguirá un campo de texto libre (p. ej. *generador + Cualquier término(s) AND en definición NOT inducción*) o (iii) el operador OR, que desplegará de nuevo una lista de selección y el campo que acompañe al elemento que se haya elegido de la lista (p. ej. *generador + Cualquier término(s) AND en definición OR en sinónimos inducción*).

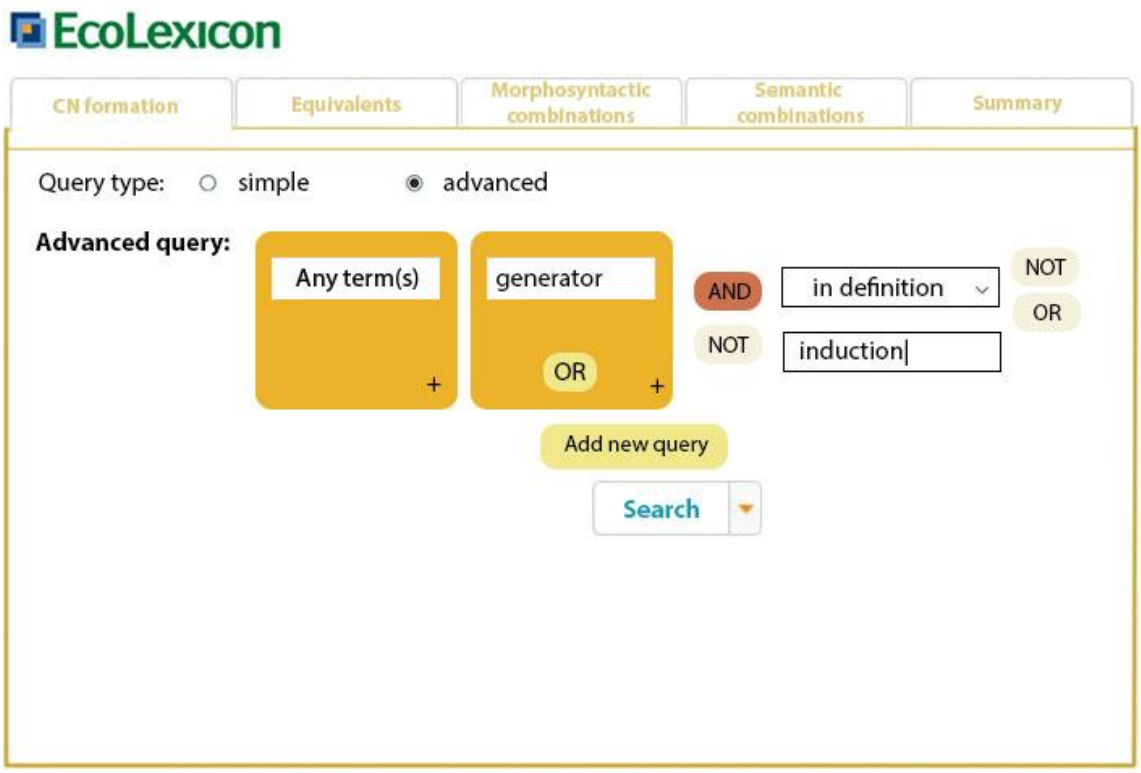

Figura 41: Ejemplo de empleo de marcadores booleanos en la búsqueda avanzada de la vista *CN formation/Formación de CN*

Por otra parte, NOT activa un campo de texto libre que nos permitirá hacer búsquedas como la siguiente: *generador + Cualquier término(s) NOT imán permanente*, para extraer generadores que no incluyan el lema *imán permanente* entre sus formantes. Esta función se diferencia del uso de NOT como marcador secundario de AND, que hemos comentado arriba, en que en este caso el uso de NOT indica que la secuencia introducida no figure como parte del CN en cuestión.

En todos los casos, a los campos de texto libre podrán seguirlos, de nuevo, los operadores NOT y OR, como en la consulta *generador + Cualquier término(s) AND en definición inducción OR reacción*. Por su parte, a las listas de selección podrá seguirles el

operador OR, que permitirá indicar otra opción de la lista, ya que la inclusión de NOT no tendría sentido dado el carácter restrictivo de la propia lista de selección.

De este modo, si restringimos la búsqueda por medio de los marcadores booleanos, se presentará el CN o los CN que hayamos conseguido por medio de nuestra búsqueda específica, seguido de sus hipónimos, en el caso de que los tenga y estos respeten las restricciones introducidas. Como se observa, se pone a disposición del usuario una gran variedad de consultas que buscan dar respuesta a sus posibles intereses.

Una vez descritos los procedimientos disponibles de búsqueda, nos detendremos a continuación en los resultados que se obtienen. Así, tras pulsar el botón *Search/Buscar*, el usuario recibe una lista de los CN que se forman con el término introducido como núcleo o modificador, si ha utilizado una búsqueda simple, o en una determinada posición, si ha empleado una búsqueda avanzada. En todos los casos, estos CN se muestran junto con sus variantes denominativas y su definición.

Como hemos indicado, en primer lugar se presentan los conceptos designados por CN que contienen el término como núcleo (*Generator as a CN head/Generator como núcleo*). Estos aparecen ordenados por dimensiones. A continuación se muestran los conceptos denominados por CN que incluyen el término entre sus modificadores (*Generator as a CN modifier/Generator como modificador*). En este caso, no se muestra una organización por dimensiones, sino que se da un orden alfabético. Bajo cada concepto se presentan sus hipónimos, siempre que cumplan con la consulta realizada. Tanto el apartado de *Generator as a CN head/Generator como núcleo* como el de *Generator as a CN modifier/Generator como modificador* se mostrarán vacíos si se ha indicado en la búsqueda avanzada que el término aparezca en la otra posición. Por ejemplo, si se quieren conocer los CN que tienen *generator* como núcleo, no se mostrarán obviamente los que tengan *generator* como modificador (a no ser que constituyan una variante de otro término principal en el que *generator* sea el núcleo). Además, la búsqueda se puede modificar en cualquier momento haciendo clic en el botón superior *Modify query/Modificar consulta*. La Figura 42 muestra un ejemplo de los resultados que se obtienen en la vista *CN formation/Formación de CN*.

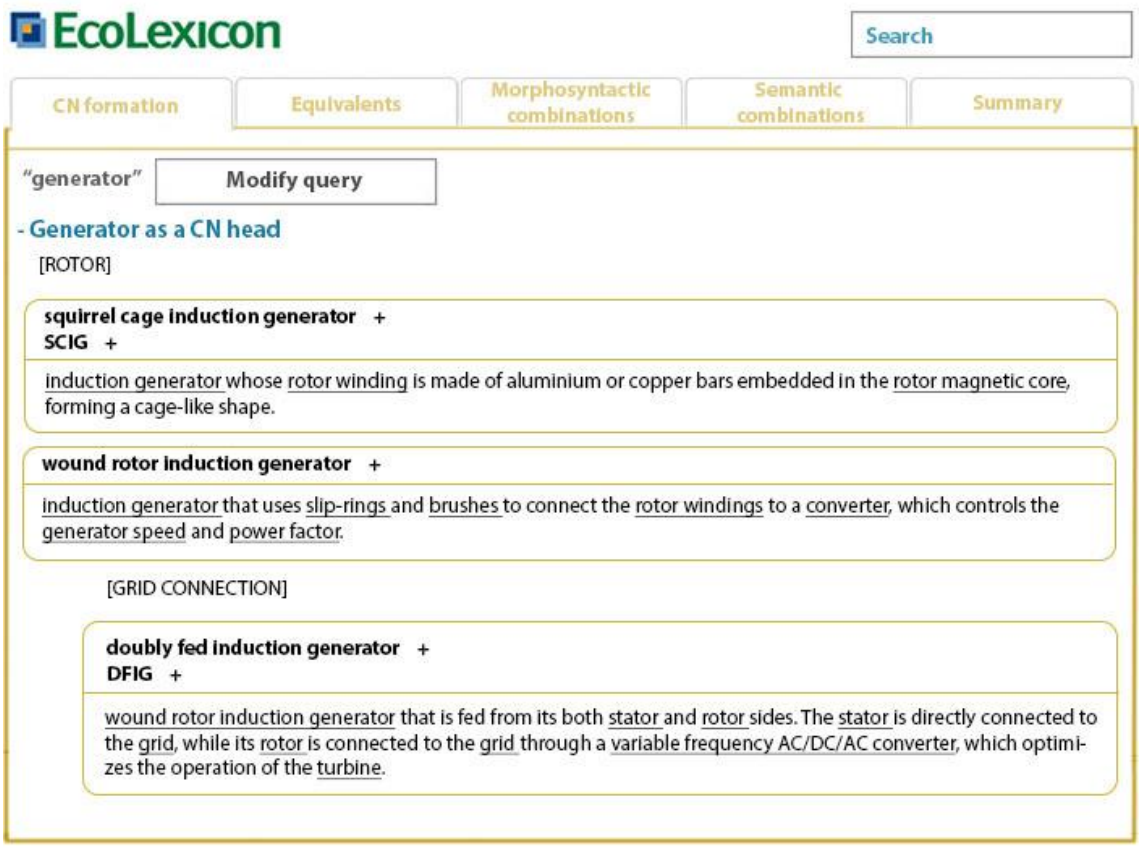

Figura 42: Fragmento de la información de *generator* en la vista *CN formation/Formación de CN* del módulo de CN, con los conceptos ordenados por dimensiones

Como se aprecia en la Figura 42, las dimensiones (p. ej. *ROTOR* y *GRID CONNECTION*) preceden a los conceptos, cuyos hipónimos se presentan sangrados (*doubly fed induction generator* es hipónimo de *wound rotor induction generator*). Dentro de cada concepto se indica en primer lugar el término principal, que ha sido consignado como tal en el formulario de introducción de términos en EcoLexicon. A continuación le siguen sus variantes denominativas. Entre estas a menudo figuran siglas. Consideramos que estas formas reducidas deben incluirse, dado su elevado uso en el lenguaje de los especialistas (muchas veces sin la forma extensa) y su opacidad semántica, que pueden derivar en su consulta en recursos terminográficos.

En la definición se señalan con hipervínculos los términos incluidos en EcoLexicon, que despliegan una ventana flotante con la definición de cada término, además de que facilitan un acceso rápido a sus entradas.

Junto a cada CN podemos hacer clic en el signo + para acceder a sus opciones secundarias, que permiten ampliar la información. Estas opciones secundarias varían ligeramente en cada vista, como iremos comentando. En este caso, las opciones que se presentan son (i) *internal semantic relation/relación semántica interna*, (ii) *usage*

examples/ejemplos de uso, (iii) *verb collocations/colocaciones verbales*, (iv) *notes/notas* y (v) *concept entry in EcoLexicon/entrada de concepto en EcoLexicon*. Se trata de opciones a las que se puede acceder en cada término, ya que podrá haber información que no se introduzca a nivel conceptual, sino que pertenezca a cada variante, como puede ser la relación semántica interna. Al hacer clic en la opción en cuestión, se abre una ventana que presenta la información elegida. En la Figura 43 se muestran las diferentes opciones secundarias que se despliegan y en la Figura 44 se presentan los resultados de una de ellas: los ejemplos de uso.

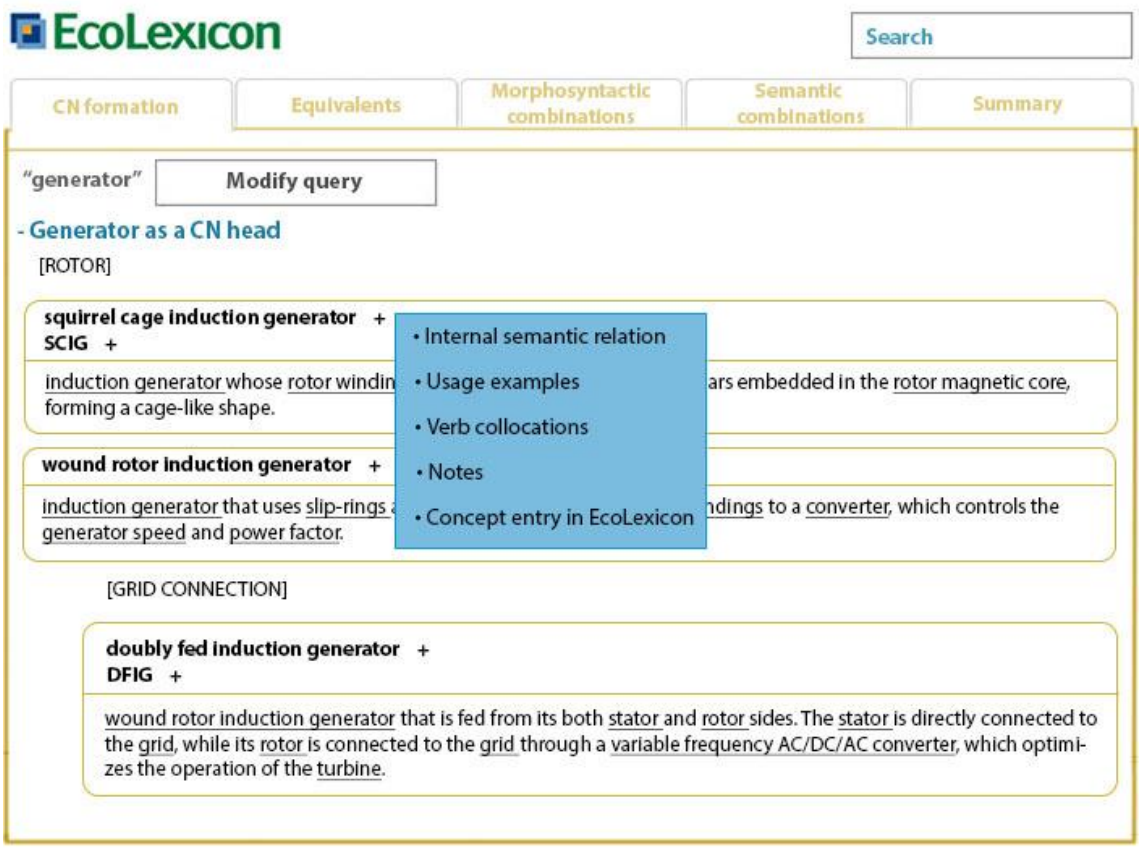

Figura 43: Ventana de opciones secundarias en la vista *CN formation/Formación de CN* del módulo de CN

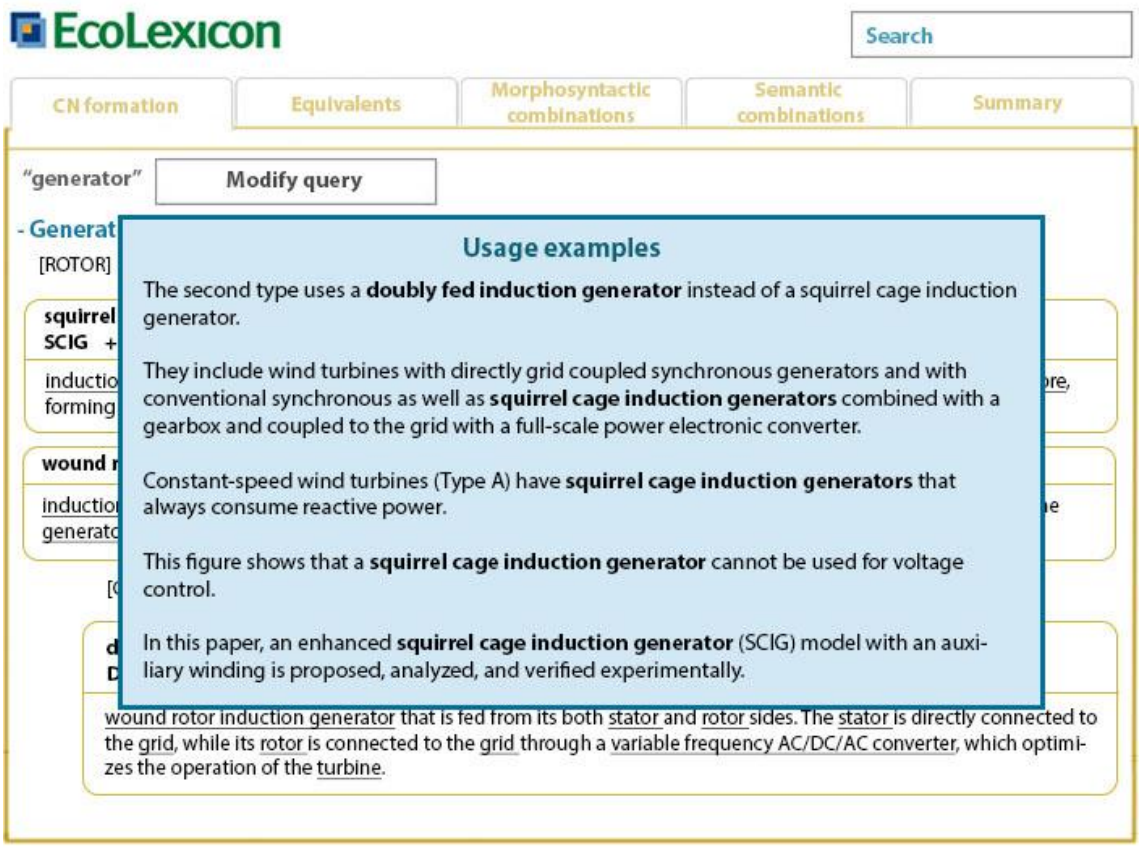

Figura 44: Resultados de la opción secundaria *Usage examples* de *squirrel cage induction generator* en la vista *CN formation/Formación de CN* del módulo de CN

Decidimos incluir la relación semántica interna porque su elisión constituye una de las principales dificultades de estas unidades fraseológicas. Por ejemplo, en *direct drive generator*, *direct drive* alude a la caja multiplicadora de la que se prescinde en el generador, lo que hace que su accionamiento sea directo y, por tanto, se establezca la relación de *atributo* entre *direct drive* y *generator*. Los ejemplos de uso, por su parte, enmarcan el CN en contextos reales.

En cuanto a las colocaciones verbales, se presenta la conexión entre el CN y la colocación o colocaciones verbales correspondientes que figuran en la otra sección del módulo fraseológico, pues muchas de ellas tienen un equivalente nominal en forma de CN (*voltage fluctuates* > *voltage fluctuation*). Se trata de una funcionalidad que se desarrollará en próximos estudios y que tendrá como punto de partida los microcontextos. Es decir, se asociarán los CN y las colocaciones verbales en los que se dé una coincidencia de predicados y estructura argumental. Para ello, se utilizará la anotación de los microcontextos, en la que los predicados se caracterizarán con los dominios léxicos de Faber y Mairal (1999) y se contará con la anotación de los argumentos con categorías y roles semánticos. Por ejemplo, el CN *erosion control structure* y la colocación verbal *a*

structure controls erosion siguen el mismo patrón fraseológico: a CONSTRUCTION[INSTRUMENT] *controls*[ACTION] an ENVIRONMENTAL PROCESS[PATIENT], que también genera otras colocaciones verbales y CN más específicos, como *a groyne controls beach erosion* o *beach erosion control*.

En los CN cuyo predicado esté omitido (p. ej. *power system*), las paráfrasis verbales facilitan la identificación de dicho predicado, que puede materializarse con diferentes verbos, si bien todos se agrupan bajo el mismo significado genérico representado por el dominio léxico. Por ejemplo, *power system* puede aludir a un generador o a la red eléctrica. Cuando tiene el sentido de *generador*, los predicados subyacentes indican el dominio de ACTION (a system *generates/produces* power). Por el contrario, cuando se refiere a la red eléctrica, los predicados pertenecen al dominio de MOVEMENT (a system *transports/transmits* power). En definitiva, la comparación de unidades fraseológicas equivalentes se concibe como un recurso adicional para conseguir riqueza en la expresión.

Otra de las opciones secundarias son las notas[67], en las que se proporciona información de interés relativa a los conceptos y a los términos, como el uso extendido de una sigla en inglés en la terminología en español (p. ej. *DFIG*, del inglés *doubly fed induction generator*, se utiliza más en español que la sigla *GIDA*, de su equivalente *generador de inducción doblemente alimentado*). Por último, la opción *entrada de concepto en EcoLexicon* permite acceder a la entrada completa de la base de conocimiento, en la que se proporciona la definición del concepto, los términos que lo designan en las diferentes lenguas incluidas, su red conceptual, imágenes, etc.

En definitiva, la vista *CN formation/Formación de CN* se centra en el papel de los CN como uno de los principales mecanismos de formación de términos. Así, permite analizar el potencial de formación de CN de un término, así como las variantes denominativas de los CN y las distintas dimensiones que estas pueden evocar, adquiriéndose así una perspectiva diferente sobre parte del sistema conceptual.

5.3.2.2 Vista Equivalents/Equivalentes

Esta vista muestra las variantes de los CN en inglés y español. Por el momento, se limita únicamente a estas dos lenguas, cuyos términos compuestos se han estudiado en esta tesis. No obstante, se irá ampliando progresivamente para incorporar las otras lenguas incluidas en EcoLexicon.

[67] La información que se muestre en las notas no es específica de cada vista, ya que se extrae de las notas indicadas para el concepto y el término en EcoLexicon.

Como veremos, funciona en gran medida como la vista anterior. Así, en la búsqueda simple se nos presenta también un cajetín de texto libre en el que podremos introducir nuestra consulta en cualquiera de los idiomas, p. ej. *generator*. Esta consulta detectará todos los CN que contienen *generator* en cualquiera de sus formantes.

También es posible realizar una búsqueda avanzada que presenta las mismas características que la vista *CN formation/Formación de CN*. Así, se muestra una burbuja de texto libre, en la que podremos introducir uno o varios términos, y otra denominada *Any term(s)/Cualquier término(s)*. Ambas podrán añadirse tantas veces como sea necesario, pulsando en el botón + inferior, y ubicarse en la posición deseada, para llevar a cabo consultas específicas. Igualmente, las nuevas burbujas añadidas podrán eliminarse haciendo clic en el símbolo X de su esquina superior derecha. La burbuja *Any term(s)/Cualquier término(s)* recupera cualquier número de elementos, como hemos comentado en la vista anterior. Se ofrece su duplicación para ubicarlas tanto a la izquierda como a la derecha del término. Lo mismo ocurre con la burbuja de texto libre, cuya duplicación se plantea para incluir varios términos separados entre sí, ya que la introducción de varios términos seguidos puede realizarse dentro de una misma burbuja. Como vemos, la búsqueda avanzada se plantea con el objetivo de permitir búsquedas específicas, si bien el usuario también podrá hacer clic en el botón *Add new query/Añadir nueva consulta* para desplegar una nueva línea de búsqueda con las dos burbujas e indicar, así, varias posibilidades.

Igual que en la vista *CN formation/Formación de CN*, podremos combinar nuestra consulta con marcadores booleanos, que funcionan como hemos descrito en la Sección 5.3.2.1. De nuevo, estos se presentan tanto dentro de la burbuja de texto libre en la que podremos introducir los términos como fuera de las burbujas. Una consulta posible sería *aerogenerador + Cualquier término(s) AND en traducciones stall*, que nos presentaría el CN *aerogenerador de regulación por pérdida aerodinámica* y su equivalente en inglés *stall-regulated wind turbine*. Al restringir la consulta con estos marcadores, se presentarán el CN o los CN resultantes, seguidos de sus hipónimos, en el caso de que los tengan y estos respeten las condiciones indicadas. En la Figura 45 se muestra un ejemplo de la información específica que se ofrece en esta vista.

Figura 45: Fragmento de la información de *generator* en la vista *Equivalents/Equivalentes*

Se presentan, en primer lugar, los CN que resultan de la búsqueda en lengua origen junto con sus variantes denominativas (EN) y, seguidamente, sus equivalentes en la lengua meta (ES). El primer término que figura en cada caso es el término principal, que se ha consignado como tal en su introducción en EcoLexicon. Como hemos señalado, en primer lugar aparecen los conceptos designados por algún CN que contenga el término como núcleo (*Generator as a CN head/Generator como núcleo*). Estos se muestran ordenados por dimensiones (p. ej. *EXCITATION* y *ROTOR AND STATOR SPEED*). A continuación figurarían los conceptos denominados por algún CN que cuente con el término entre sus modificadores (*Generator as a CN modifier/Generator como modificador*). Estos se ordenan alfabéticamente. Bajo cada concepto se presentan sangrados sus hipónimos, siempre que alguna de sus variantes cumpla con los requisitos indicados (p. ej. *permanent magnet synchronous generator* es hipónimo de *permanent magnet generator*).

Junto a cada término se incluye el signo +, que muestra las opciones secundarias. Como en la vista *CN formation/Formación de CN*, estas opciones son (i) *internal semantic relation/relación semántica interna*, (ii) *definition/definición*, (iii) *usage examples/ejemplos de uso*, (iv) *verb collocations/colocaciones verbales*, (v) *notes/notas* y (vi) *concept entry in EcoLexicon/entrada de concepto en EcoLexicon*. La única incorporación es la opción

definition/definición, que en la vista *CN formation/Formación de CN* se mostraba como elemento principal de la interfaz, mientras que en esta vista priman los equivalentes interlingüísticos.

En resumen, la vista *Equivalents/Equivalentes* se concibe con el objetivo de explorar la formación de términos compuestos desde una perspectiva interlingüística, que reviste un interés indudable si se desea difundir el conocimiento en distintas lenguas. Para ello, la comparación de términos compuestos formados a partir del mismo elemento, así como la observación de sus variantes denominativas y las dimensiones activadas, pueden ofrecer una nueva óptica de cara a la comunicación interlingüística.

5.3.2.3 Vista Morphosyntactic combinations/Combinaciones morfosintácticas

A pesar de que el análisis morfosintáctico no ha figurado entre los objetivos principales de esta tesis, somos conscientes de su interés, por lo que desarrollamos esta vista, que muestra la formación de los CN desde un punto de vista morfosintáctico. La búsqueda simple presenta los mismos elementos que en las vistas anteriores. De este modo, se podrán introducir uno o varios términos en cualquier idioma en el cajetín de texto libre que se proporciona. El resultado que se obtendrá serán todos los CN que se forman con la secuencia introducida, junto a su información morfosintáctica, como veremos a continuación.

La búsqueda avanzada, por su parte, se estructura en torno a dos elementos de consulta, que se presentan en sendas burbujas y favorecen la realización de búsquedas basadas en la morfosintaxis de los términos: una burbuja de texto libre y otra burbuja denominada *Part of speech/Categoría gramatical*. La burbuja de texto libre permite introducir uno o varios términos, mientras que la burbuja *Part of speech/Categoría gramatical* despliega una lista de selección con las categorías gramaticales que pueden seleccionarse, que son aquellas que pueden intervenir como elementos de un CN: nombre común, nombre propio, adjetivo, adverbio, preposición, artículo, participio pasado y participio presente.

Ambas burbujas pueden añadirse cuantas veces se desee y colocarse en el lugar que interese. Igualmente, las nuevas burbujas añadidas pueden eliminarse. No obstante, como hemos comentado en las vistas anteriores, la posibilidad de duplicar la burbuja de texto libre se hace con el objetivo de ubicar los términos separados entre sí, ya que si se desean introducir dos términos de búsqueda consecutivos, estos pueden insertarse en una única burbuja de texto libre. En cambio, la burbuja de *Part of speech/Categoría gramatical* debe duplicarse para incluir dos categorías contiguas, pues se trata de una lista de

selección que no permite marcar varios valores. Además, las burbujas pueden desempeñar una función neutra, esto es, pueden actuar como cualquier término y cualquier categoría gramatical si no se introduce ningún término y no se selecciona ninguna categoría de la lista. Para aclarar funcionalidades como esta, que pueden no resultar claras para los usuarios, se elaborará próximamente una guía de uso en la que se explicarán todas estas cuestiones.

En la Figura 46 se ejemplifica una posible consulta avanzada, en la que se buscan CN en inglés formados por un adjetivo, un nombre común y el núcleo *turbine*. Obtendríamos, por tanto, CN como *offshore wind turbine* o *large wind turbine*.

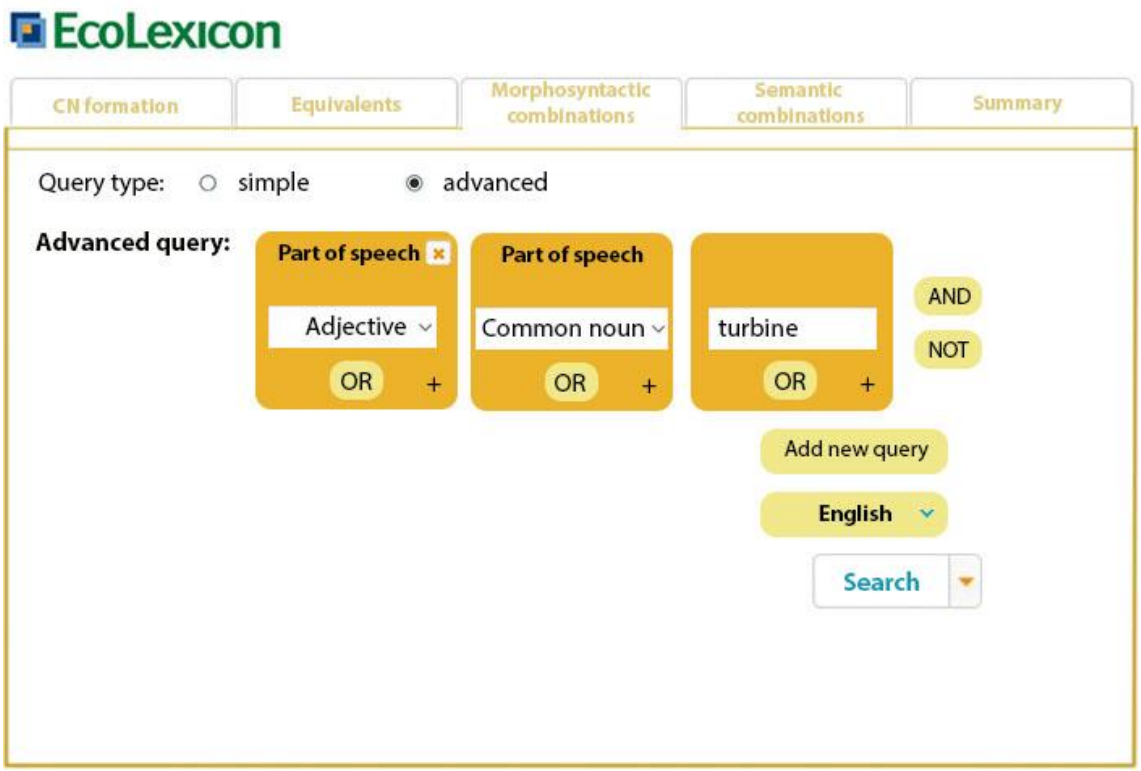

Figura 46: Interfaz de búsqueda avanzada de la vista *Morphosyntactic combinations/Combinaciones morfosintácticas*

En la Figura 46 se observa también el botón *Add new query/Añadir nueva consulta* en la parte inferior derecha. Como hemos señalado en las vistas anteriores, este botón posibilita la realización de búsquedas más complejas, ya que despliega una nueva línea de burbujas. De este modo, nos podría interesar, por ejemplo, combinar la consulta de Adjetivo + Nombre común + *turbine* con la búsqueda Adjetivo + *hydraulic turbine* y obtener los resultados de ambas consultas.

Asimismo, podemos hacer uso de los marcadores booleanos para indicar condiciones adicionales a la consulta. Dentro de la burbuja de texto libre se puede utilizar

el operador OR para permitir varias opciones en una misma posición: Participio pasado + *voltage* OR *power* (*induced voltage, rated voltage, rated power, generated power*). Por su parte, en la burbuja de *Part of Speech/Categoría gramatical* se incorpora el marcador OR, para indicar la disyunción en las categorías gramaticales. Estos marcadores activan el tipo de campo de la burbuja en la que se encuentren (texto libre o lista de selección). El marcador AND no se ha empleado, pues buscaría elementos que desempeñasen dos categorías gramaticales a la vez, así como NOT, ya que al elegir una categoría, evidentemente se descartan las restantes. Fuera de las burbujas se incluyen los marcadores AND y NOT, que funcionan del modo descrito en la Sección 5.3.2.1. Por ejemplo, una consulta posible sería *viento* + Adjetivo AND en traducciones *offshore*, que ofrecería CN como *viento marino* tras haber detectado su equivalente en inglés *offshore wind*. Al restringir los resultados, se muestran aquellos CN que cumplen las condiciones, en el habitual orden de núcleo y modificador. Tras elegir la lengua[68] en la que desea realizarse la consulta, se obtienen los resultados (Figura 47).

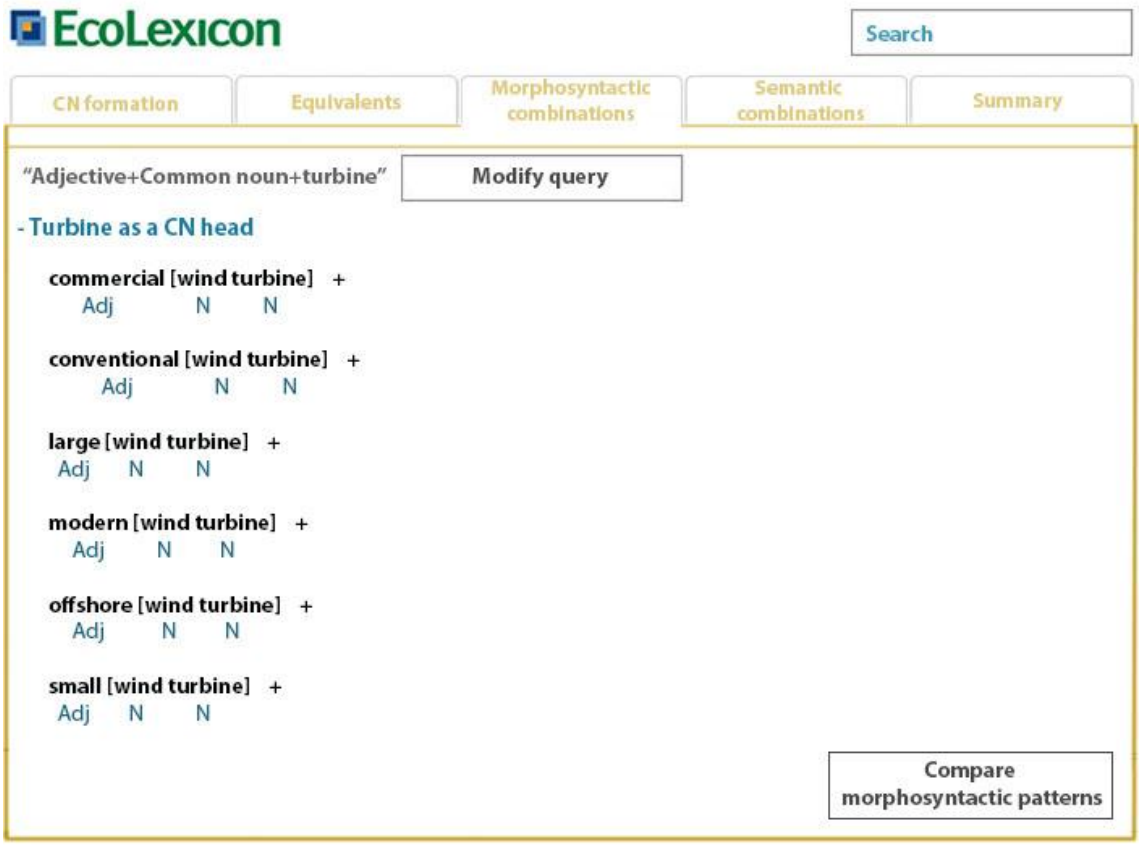

Figura 47: Fragmento de la consulta *Adjective + Common noun + turbine* en la vista *Morphosyntactic combinations/Combinaciones morfosintácticas*

[68] Tanto en esta vista como en la de *Semantic combinations/Combinaciones semánticas*, se incluye la indicación de la lengua de consulta, ya que se pueden realizar búsquedas en las que no se inserte ningún término y, por tanto, el sistema no pueda identificar la lengua en cuestión.

En esta vista se presentan los CN anotados con la categoría gramatical de sus formantes y la indicación del *bracketing*. De nuevo, los resultados se organizan según el núcleo o el modificador, cuando se ha introducido un término de consulta. Recordemos que, en esta vista morfosintáctica, los CN no se organizan por dimensiones, las variantes no se agrupan en conceptos y, por tanto, los hipónimos no se agrupan bajo sus hiperónimos, pues se trata de información de carácter semántico que no concuerda con los datos proporcionados en esta vista. Así, esta vista se centra en la anotación morfosintáctica a nivel de término.

Junto a cada CN se podrá acceder a sus opciones secundarias (signo +): (i) *definition/definición*, (ii) *usage examples/ejemplos de uso*, (iii) *verb collocations/colocaciones verbales*, (iv) *notes/notas* y (v) *concept entry in EcoLexicon/entrada de concepto en EcoLexicon*. Por otra parte, se ha descartado la opción que permitía consultar la relación semántica interna, pues ese dato semántico no resulta relevante en esta vista.

Una incorporación de esta vista es el botón *Compare morphosyntactic patterns/Comparar patrones morfosintácticos* que se muestra en la parte inferior derecha de la interfaz y que también está disponible como *Compare semantic patterns/Comparar patrones semánticos* en la vista de *Semantic combinations/Combinaciones semánticas*. Este botón nos permitirá comparar el patrón morfosintáctico que hayamos introducido en inglés con el de su equivalente principal en español. Es decir, se mostrarán los mismos resultados que responden a nuestros criterios de búsqueda, acompañados de sus variantes principales en la otra lengua. Estas estarán anotadas con la categoría gramatical de cada formante, de forma que se pueden comparar los patrones morfosintácticos de formación de términos en ambas lenguas.

Decidimos realizar la comparativa partiendo de los resultados iniciales de la consulta (p. ej. los CN en inglés), en lugar de utilizar la misma consulta en la otra lengua, porque esta opción no arrojaría resultados, debido a las diferencias estructurales entre las lenguas. Así, por ejemplo, la consulta nombre común + nombre común + nombre común ofrecerá resultados en inglés, pero no en español, mientras que si, a partir de esos resultados en inglés, comparamos la morfosintaxis de sus equivalentes, obtendremos un enfoque más interesante (p. ej. la compactación de sustantivos en inglés frente a las estructuras preposicionales en español) (Figura 48). La inclusión de la variante principal en la lengua meta se realiza para presentar la forma más habitual en español del término

buscado en inglés, además de por motivos de espacio. En el caso de términos polisémicos, se presentan las variantes denominativas principales de cada concepto en español.

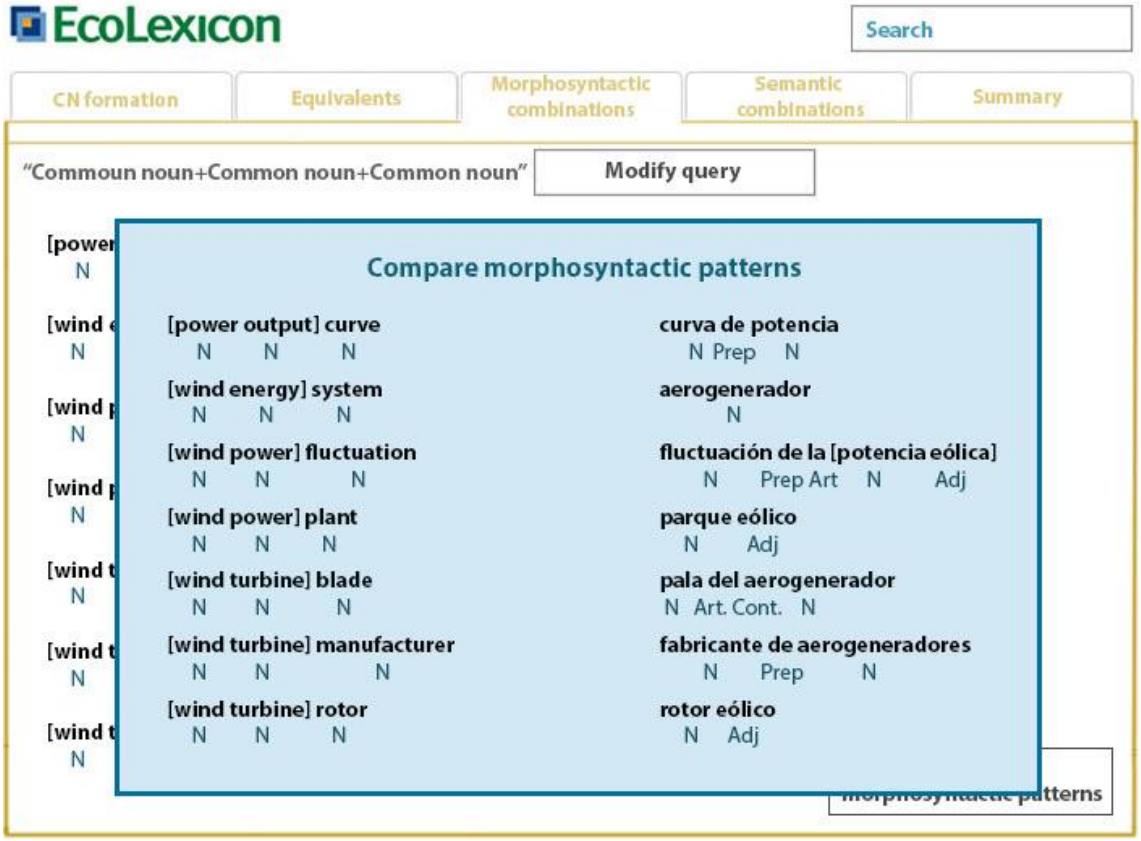

Figura 48: Resultados de la opción *Compare morphosyntactic patterns/Comparar patrones morfosintácticos* para la consulta *Common noun + common noun + common noun*

En definitiva, la vista *Morphosyntactic combinations/Combinaciones morfosintácticas* permite ahondar en la morfología de la formación de CN, así como realizar búsquedas específicas basadas en las categorías gramaticales o comparar este tipo de información en varias lenguas. Por tanto, puede resultar de utilidad para diferentes análisis gramaticales.

5.3.2.4 Vista Semantic combinations/Combinaciones semánticas

Esta vista se centra en la formación de CN desde una perspectiva semántica. La búsqueda simple se realiza del mismo modo que en las vistas anteriores: mediante un cajetín de texto libre en el que se pueden insertar uno o varios términos en cualquier idioma. Se mostrarán los CN que se forman con el término o términos introducidos, junto a su anotación semántica, como detallaremos a continuación.

La búsqueda avanzada, por su parte, se organiza en torno a cuatro elementos de consulta, que permiten la realización de búsquedas según la semántica de los términos.

Como en el resto de vistas, estos elementos se muestran en cuatro burbujas, que podrán duplicarse y colocarse en la posición que se desee.

Por una parte, se proporciona una burbuja de texto libre en la que se pueden introducir uno o varios términos. Esta puede duplicarse para buscar varios términos separados por otra de las burbujas. Además, se presenta la burbuja *Semantic relation/Relación semántica*, que despliega una lista de selección con las relaciones recogidas en el sistema. Estas se presentan en una única dirección, ya que se busca la simplificación (p. ej. *causes* pero no *caused_by*). La burbuja *Semantic category/Categoría semántica* también ofrece una lista de selección con el inventario de categorías semánticas empleado en EcoLexicon. Por último, la burbuja *Semantic role/Rol semántico* despliega en una lista de selección los roles semánticos empleados en el módulo fraseológico. En cualquier caso, si el usuario desea indicar un valor neutro de estas burbujas (esto es, cualquier término, o cualquier relación, categoría o rol semántico), bastará con no introducir ni elegir ningún valor en la burbuja en cuestión.

A diferencia de las vistas anteriores, donde todas las burbujas eran necesarias para obtener resultados, en este caso puede que el usuario solo desee utilizar algunas de las burbujas que se proponen (p. ej. un término + una categoría semántica). Para ello, todas las burbujas presentan el símbolo X en su esquina superior derecha, que permite eliminar aquellas burbujas que no interesen en la consulta. Por otra parte, al igual que en la vista *Morphosyntactic combinations/Combinaciones morfosintácticas*, se debe indicar la lengua de consulta. En la Figura 49 se muestra la interfaz de búsqueda avanzada de la vista *Semantic combinations/Combinaciones semánticas* con las diferentes burbujas.

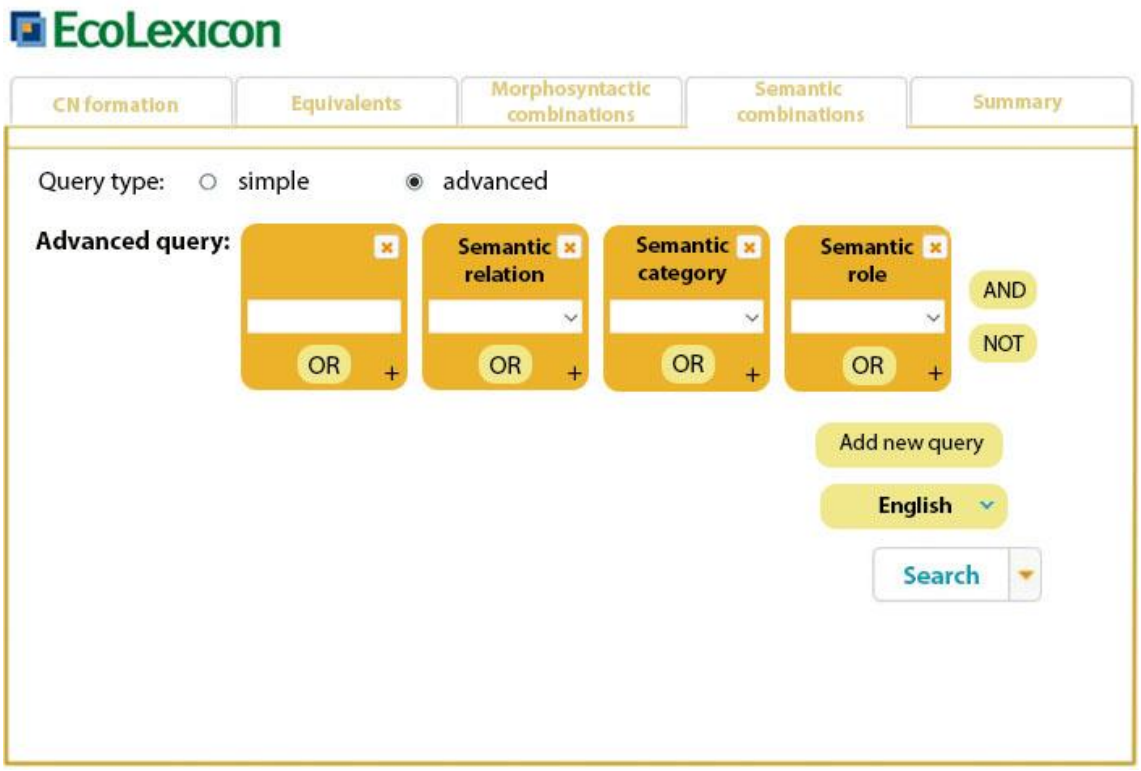

Figura 49: Interfaz de búsqueda avanzada de la vista *Semantic combinations/Combinaciones semánticas*

De tal modo, la combinación de las distintas burbujas ofrece un gran abanico de consultas posibles para los usuarios interesados en la particularización semántica. Algunas de ellas podrían ser las siguientes:

-*wind* (término) + STRUCTURE (categoría): *wind farm, wind park, wind plant*, etc.

-INSTRUMENT (rol) + *turbine* (término): *gas turbine, hydraulic turbine, wind turbine*, etc.

-AGENT (rol) + PROCESS (categoría): *wind power, wave power, wind penetration*, etc.

-PROCESS (categoría) + *uses_resource* (relación) + INSTRUMENT (categoría): *impulse turbine, reaction turbine, induction generator*, etc.

La posición de las burbujas determinará el tipo de CN extraídos. Es decir, no se obtendrán los mismos resultados con la consulta *generator* + categoría semántica que con la búsqueda categoría semántica + *generator*. Sucede igual cuando nuestra consulta se compone de una relación semántica y los dos conceptos que vincula, que se pueden buscar mediante término, categoría o rol semántico. Por ejemplo, si queremos conocer los CN que designan procesos causados por el término *generator*, tendremos que plantear la consulta respetando el sentido de la relación: *generator* + *causes* + PROCESS. Sin embargo, en este

caso, no significa que los CN resultantes deban plasmar ese orden de la proposición conceptual en su estructura.

Es preciso señalar que la burbuja de relaciones semánticas no funciona del mismo modo cuando solo se introduce uno de los elementos que vincula (p. ej. el término *turbine* y la relación *type_of*). Dado que la disposición en estos casos puede generar dudas sobre el tipo de resultados que se obtendrá (hiperónimos o hipónimos de *turbine*), en este tipo de consulta se presentarán todos los CN que se forman con el término *turbine* y la relación *type_of*, en cualquiera de sus direcciones: *turbine type_of X* (que mostraría CN que indiquen el hiperónimo de *turbine*) y *turbine has_type X* (que presentaría CN que aludan a hipónimos de *turbine*). Optamos por esta solución dado que comprende las dos direcciones que pueda estar buscando el usuario, enmarcando así los CN en un sistema más amplio. Además, no habrá confusión con la relación, ya que la dirección se indica de forma explícita, partiendo desde el núcleo.

Igual que en el resto de vistas, el botón *Add new query/Añadir nueva consulta* proporciona una nueva línea de consulta para aunar dos búsquedas, como vemos en el siguiente ejemplo:

ENTITY (categoría) + voltage: *battery voltage, bus voltage, DC-link voltage*, etc.

POSSESSOR (rol) + voltage: *AC voltage, battery voltage, bus voltage, DC voltage*, etc.

También se incluyen en esta vista marcadores booleanos, tanto dentro de las burbujas, para permitir varias opciones en la misma posición, como fuera, para aplicar la restricción a todo el CN. Dentro de la burbuja de texto libre se puede utilizar el marcador OR. Por ejemplo, *offshore OR onshore + STRUCTURE (categoría)*, que recuperaría CN como *offshore wind farm* y *onshore wind farm*. El resto de burbujas también permiten el uso del marcador OR, para señalar categorías, relaciones o roles alternativos, p. ej. *offshore + STRUCTURE OR INSTRUMENT (categorías)*, que muestra CN como *offshore wind farm, offshore turbine* y *offshore wind turbine*. El operador AND no se incluye, dado que nos parece poco probable que un elemento desempeñe dos categorías semánticas o dos roles, así como que se den dos relaciones en un mismo CN, si a esto añadimos las otras restricciones que se empleen en la consulta. Como hemos comentado sobre la burbuja de categoría gramatical de la vista anterior, el operador NOT tampoco se incluye en estas burbujas al tratarse de listas de selección, en las que la elección de un valor ya implica la exclusión del resto.

Los marcadores AND y NOT que figuran fuera de las consultas realizan las mismas funciones que se han comentado en las vistas anteriores. Así, se podría buscar, por

ejemplo, *DESCRIPTIVE (rol) + THEME (rol) AND in translations tensión* y obtener CN como *induced voltage* y *rated voltage*, pero no *generated power*, que presenta los roles indicados pero se ha descartado al no incluir el término *tensión* en sus equivalentes. Si empleamos los marcadores booleanos, se nos mostrarán los CN que cumplen las restricciones, organizados según núcleo y modificador si se ha introducido algún término en la burbuja de texto libre o, de lo contrario, por orden alfabético. La Figura 50 muestra un ejemplo de los resultados que se obtienen en la vista *Semantic combinations/Combinaciones semánticas*, en concreto al buscar las categorías semánticas MAGNITUDE (ATTRIBUTE) + CHANGE (PROCESS).

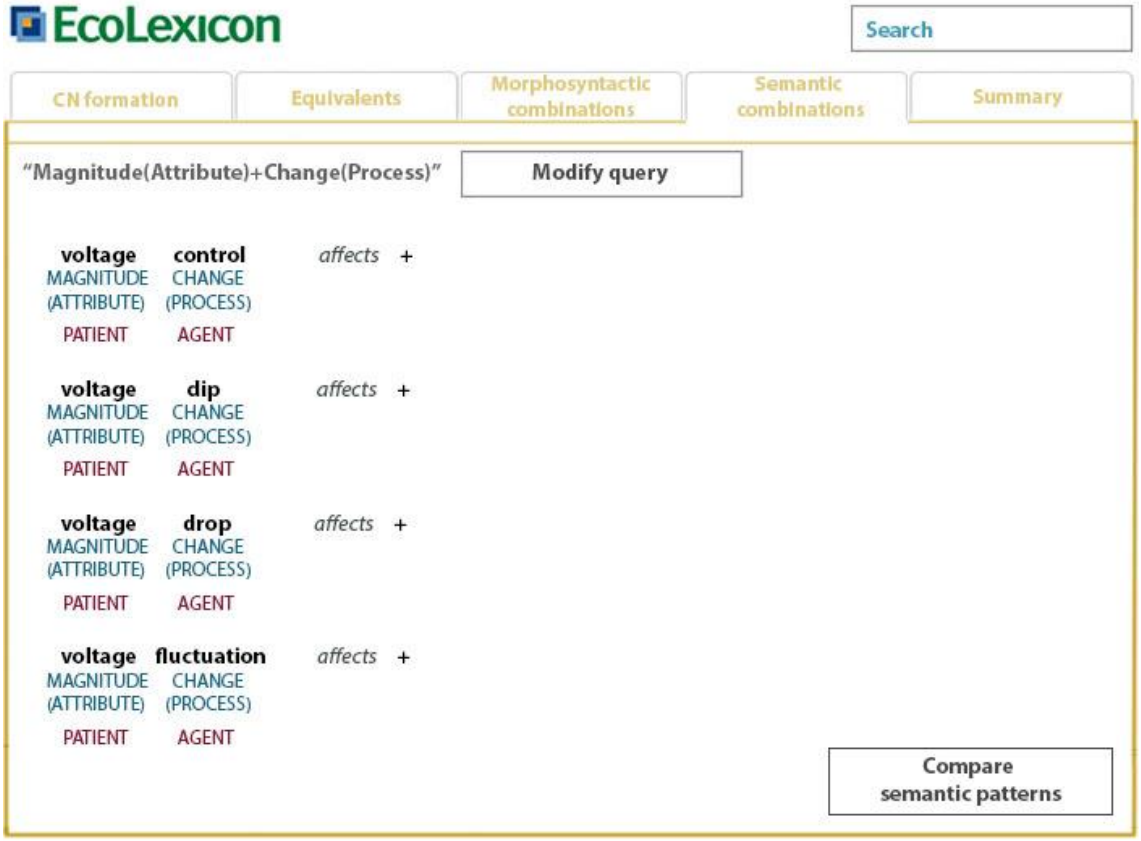

Figura 50: Fragmento de la consulta avanzada *Magnitude (Attribute) + Change (Process)* en la vista *Semantic combinations/Combinaciones semánticas*

En los CN resultantes de la búsqueda se indica su estructura de *bracketing* cuando estos tienen más de dos formantes, así como la categoría[69] y el rol semántico de los elementos o los grupos del *bracketing*, en su caso. Por ejemplo, en *voltage control* observamos que *voltage* pertenece a la categoría de MAGNITUDE (ATTRIBUTE) y *control* a la de CHANGE (PROCESS). En esta combinación, *control* desempeña el rol de AGENT, pues incide

[69] Se muestra la categoría más específica anotada y, entre paréntesis, la categoría genérica (ENTIDAD, PROCESO o ATRIBUTO).

sobre *voltage* (PATIENT). Junto a cada CN, se muestra también su relación semántica interna, con la dirección adecuada partiendo desde el núcleo hacia el modificador (p. ej. en *voltage control,* control *affects* voltage). En esta vista, de carácter conceptual, decidimos indicar la categoría semántica y el rol del *bracketing* (para los CN de más de dos formantes), en lugar de la categoría y el rol de cada constituyente, ya que consideramos que cuando se hace una búsqueda basada en la semántica, se piensa en los principales conceptos que componen el CN, que serían los que marca el *bracketing*. Así, un CN como *constant-speed wind turbine* se define mejor mediante la combinación de los conceptos ATTRIBUTE + ENTITY (*constant-speed* + *wind turbine*) que por medio de los conceptos individuales ATTRIBUTE + ATTRIBUTE + PROCESS + ENTITY (*constant* + *speed* + *wind* + *turbine*). No obstante, esta información individual no carece de interés, por lo que se presenta en las opciones secundarias, como comentaremos a continuación.

Como hemos indicado, si se ha introducido algún término en la burbuja de texto libre, los CN se organizarán según los que presenten esta secuencia como núcleo o como modificador. En ese caso, los CN que cuenten con ese núcleo se organizarán en dimensiones, mientras que los que incluyan el término entre sus modificadores se ordenarán alfabéticamente. En esta vista, como sucede con las combinaciones morfosintácticas, las variantes no se agrupan en conceptos, ya que lo que interesa es la caracterización semántica de cada CN. Por tanto, los hipónimos no aparecen sangrados bajo su hiperónimo.

Se puede hacer clic en el símbolo + que aparece junto a cada término para acceder a sus opciones secundarias, que son las siguientes: (i) *definition/definición*, (ii) *additional semantic information/información semántica adicional*, (iii) *usage examples/ejemplos de uso*, (iv) *verb collocations/colocaciones verbales*, (v) *notes/notas* y (vi) *concept entry in EcoLexicon/entrada de concepto en EcoLexicon*. A diferencia de las vistas anteriores, el acceso a la relación semántica no se proporciona como opción secundaria, ya que forma parte de la información principal de esta vista. Además, se ofrece la opción *additional semantic information/información semántica adicional*, que muestra la categoría semántica del concepto y de los formantes del CN, además de todas las relaciones internas, tanto la primaria como la que figura entre los elementos del *bracketing* (en los CN de más de dos formantes).

Por último, al igual que en la vista *Morphosyntactic combinations/Combinaciones morfosintácticas*, se muestra el botón *Compare semantic patterns/Comparar patrones semánticos* en la parte inferior derecha de esta vista. Esta opción parte de la lista de CN

obtenidos a través de nuestra consulta para mostrar su variante principal en la otra lengua, provista de su anotación semántica (Figura 51). Así, podremos comparar el patrón semántico de nuestros resultados con el de sus equivalentes en la otra lengua y analizar, de este modo, las diferencias que pueden darse a nivel de categorías, roles y relaciones semánticas. Gracias a ello se podrá obtener un enfoque interlingüístico de fenómenos tan habituales como la variación o la multidimensionalidad, o incluso contrastar los componentes semánticos de la formación de los CN en ambas lenguas (p. ej. *small wind turbine* frente a su equivalente *aerogenerador de baja potencia*).

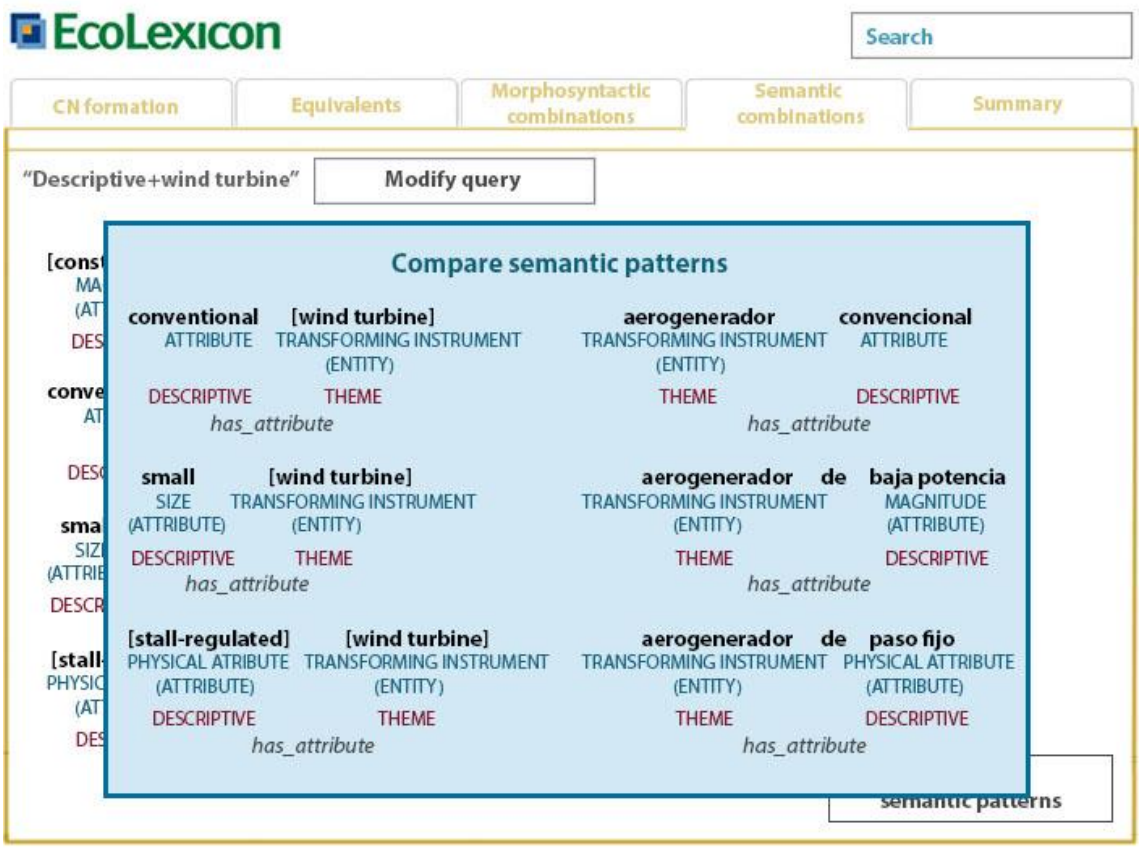

Figura 51: Resultados de la opción *Compare semantic patterns/Comparar patrones semánticos* para la consulta *Descriptive + wind turbine*

En resumen, la vista *Semantic combinations/Combinaciones semánticas* se plantea partiendo del papel clave de la semántica en la formación de CN. En estos términos compuestos se unen varios conceptos, vinculados por una relación semántica. Por tanto, el análisis profundo de esos conceptos por medio de categorías semánticas y roles semánticos, así como la identificación de la relación interna, permiten profundizar en el sistema conceptual en el que se enmarcan estos términos. Además, mediante las consultas específicas, se puede partir de los elementos del sistema conceptual para explorar cómo se materializan en los CN, una perspectiva poco abordada hasta el momento.

5.3.2.5 Vista Summary/Resumen

Esta vista sintetiza la información de las cuatro vistas específicas. A pesar de ser la primera que obtiene el usuario, la describimos en último lugar, pues sus características se comprenden mejor una vez que se han comentado las cuatro vistas particulares.

Se puede acceder a ella mediante la entrada de un concepto o a través de la pestaña *Phraseological module/Módulo fraseológico* de la interfaz de EcoLexicon. En ambos casos se debe seleccionar si se desea consultar el módulo de colocaciones verbales o el de CN. Al marcar esta última opción, se nos deriva directamente a la interfaz de búsqueda de la vista *Summary/Resumen*, que presenta un cajetín de texto libre en el que se podrán introducir uno o varios términos en cualquiera de los idiomas (por el momento, inglés y español) (Figura 52). Además, si nos encontramos en otra de las vistas, podemos acceder a esta o cualquiera de las otras opciones haciendo clic en su pestaña de la parte superior de la interfaz.

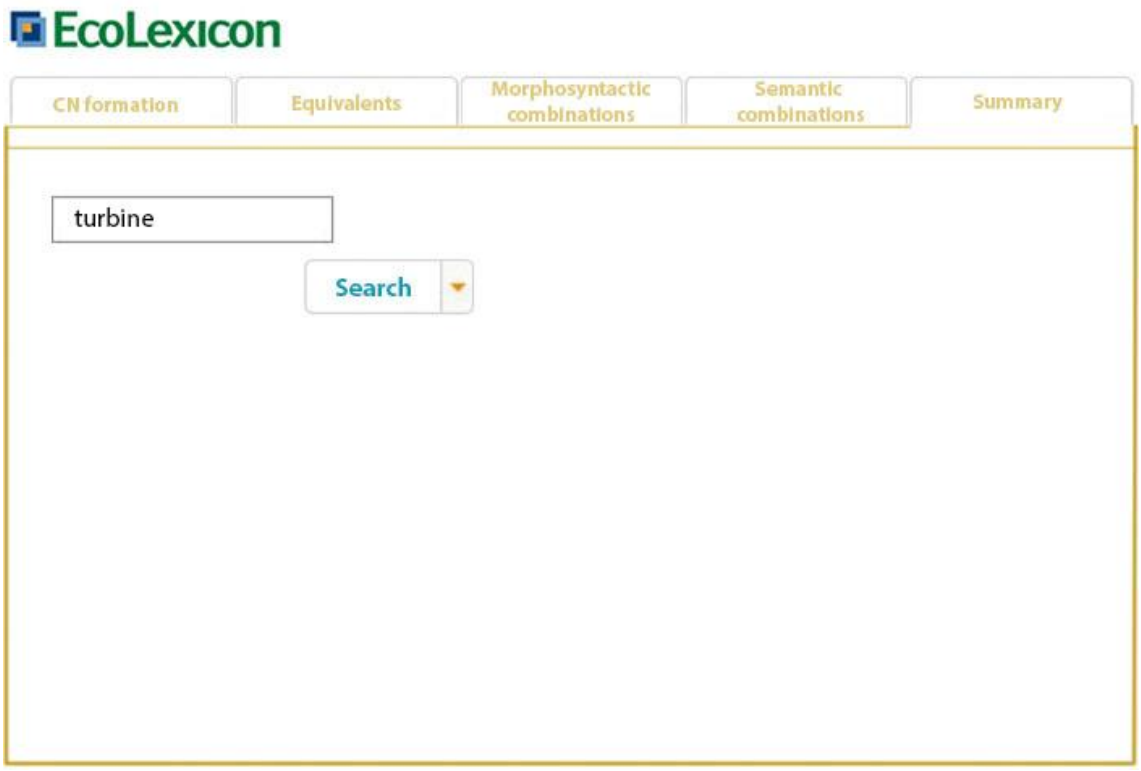

Figura 52: Interfaz de consulta de la vista *Summary/Resumen*

Los resultados de esta vista constituyen un resumen de lo que se obtiene en las vistas *CN formation/Formación de CN, Equivalents/Equivalentes, Morphosyntactic combinations/Combinaciones morfosintácticas* y *Semantic combinations/Combinaciones semánticas*. De este modo, los resultados se distribuyen en cuatro burbujas con estos nombres. La idea subyacente en esta vista es que el usuario obtenga una panorámica

general del módulo de CN y, así, pueda continuar su búsqueda en consonancia. En la Figura 53 se muestran los resultados de la búsqueda *turbine* en la vista *Summary/ Resumen*.

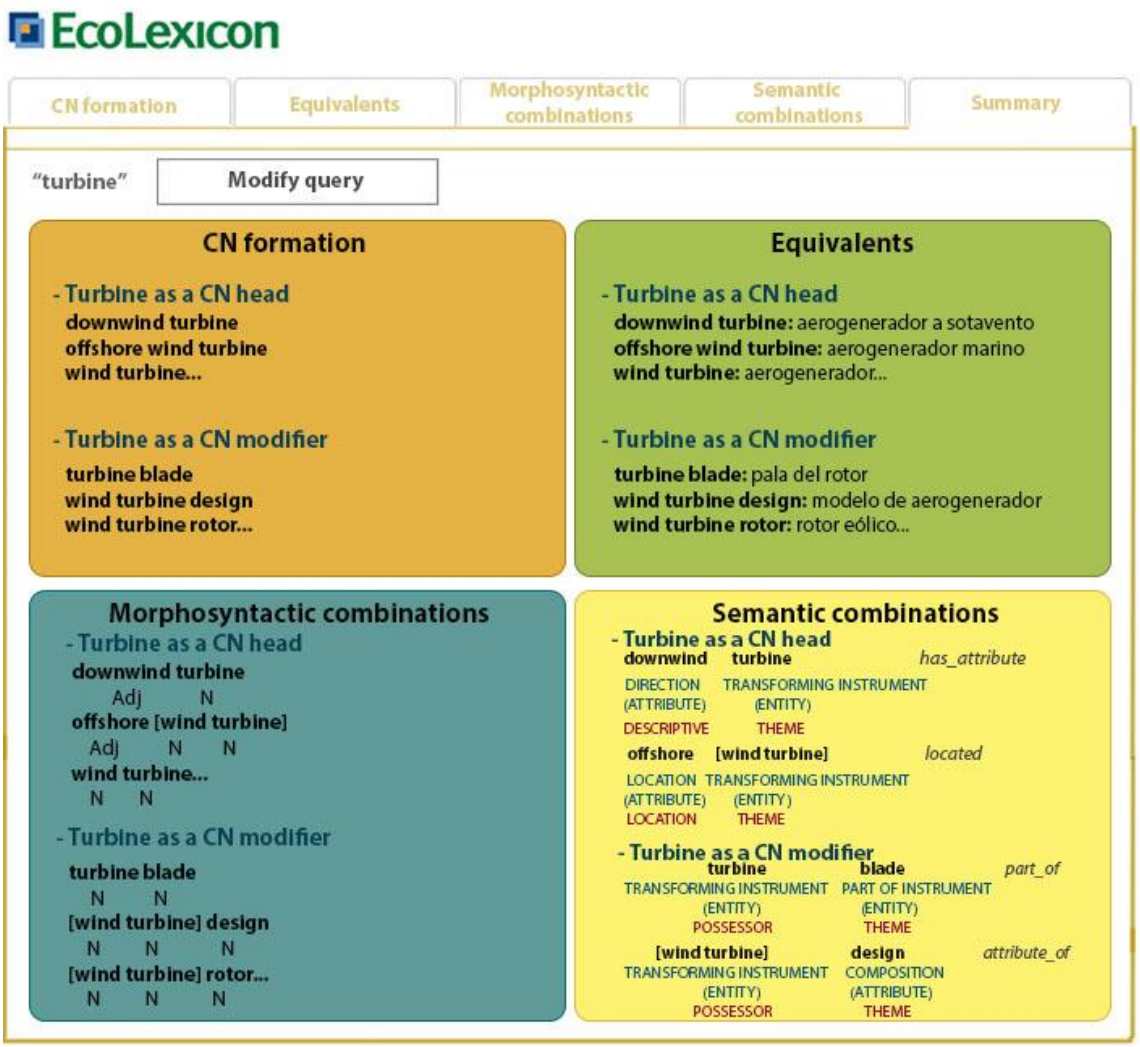

Figura 53: Resultados de la consulta *turbine* en la vista *Summary/Resumen*

Como vemos en la Figura 53, la burbuja *CN formation/Formación de CN* presenta algunos de los CN que se forman con *turbine*. Estos ejemplos se muestran también en la burbuja *Equivalents/Equivalentes* junto con su principal equivalente en español. Como vemos, en estas burbujas no se adopta la organización de variantes en conceptos habitual de las vistas que desarrollan estos dos tipos de información, dado el carácter simplificado de esta vista. Ello justifica que, en la burbuja de *Equivalents/Equivalentes*, solo se muestre el equivalente principal de cada CN. Por su parte, la burbuja *Morphosyntactic combinations/Combinaciones morfosintácticas* muestra el *bracketing* y la anotación con categorías gramaticales de los formantes de los CN presentados como ejemplo. Por último, la burbuja *Semantic combinations/Combinaciones semánticas* indica la anotación de los

elementos de estos CN con categorías semánticas (azul), roles semánticos (rojo) y la relación semántica interna (gris, a la derecha).

Como ocurre en las vistas desarrolladas, se plantea una estructuración según la posición del término en los CN: núcleo o modificador, aunque, dado el carácter reducido de la vista, no se adopta la organización de variantes, hipónimos ni dimensiones.

Cabe señalar que, en el caso de que se introduzca un CN como elemento de búsqueda, el propio CN no figurará en las burbujas de *CN formation/Formación de CN* y *Equivalents/Equivalentes*, dado que el objetivo de estas es mostrar los CN que se generan a partir de él. Sin embargo, dicho CN aparecerá en las burbujas de *Morphosyntactic combinations/Combinaciones morfosintácticas* y *Semantic combinations/Combinaciones semánticas*, ya que puede resultar interesante ver la anotación de sus formantes. En ese caso, aparecerá al principio, fuera de los grupos de núcleo y modificador.

En esta línea, si el término introducido como consulta es un CN que no genera otros CN, las dos primeras burbujas no proporcionarán evidentemente ninguna información, si bien se mostrará la anotación morfosintáctica y semántica de los formantes de dicho término en las burbujas correspondientes. No ocurre así cuando el elemento de búsqueda es un término simple, pues su anotación morfosintáctica y semántica ya está disponible en su entrada de concepto en EcoLexicon.

Como hemos mencionado, la vista *Summary/Resumen* no pretende ser exhaustiva, sino ofrecer una síntesis de la formación de CN desde diferentes perspectivas, así como servir de guía de la información que se obtendrá en las diferentes vistas.

5.3.3 Conclusiones del módulo de compuestos nominales de EcoLexicon

El módulo de CN que acabamos de presentar pretende ser de utilidad para un perfil variado de usuarios, que van desde traductores e intérpretes, terminólogos y redactores técnicos hasta estudiantes o especialistas del medio ambiente. Estos pueden no ser hablantes nativos de español o inglés, una situación que está a la orden del día en la comunicación científica. Podrán hacer uso del módulo fraseológico de CN para solventar situaciones cognitivas o comunicativas, en las que requieran de información conceptual o lingüística en varios idiomas.

De este modo, el recurso que planteamos se beneficia de la inclusión de diferentes tipos de datos: anotación morfosintáctica y semántica, potencial combinatorio de los términos, equivalencias interlingüísticas, información lingüística y de uso, variación,

multidimensionalidad, etc. En definitiva, información de utilidad tanto para la comprensión como para la producción.

En este sentido, en una evaluación previa en la que se valoraron las opiniones de los usuarios de EcoLexicon, López Rodríguez et al. (2012) señalaron que los usuarios identificaron las relaciones conceptuales, los equivalentes, los sinónimos, la fraseología, los contextos y las definiciones como la información más útil del recurso. Además, los usuarios prefirieron la organización conceptual frente al orden alfabético, pues esta contribuye a la adquisición de conocimiento. Otras investigaciones (Durán Muñoz 2010; León Araúz y Reimerink 2018) obtuvieron resultados similares y señalaron la utilidad de incluir CN en recursos de conocimiento, junto a sus siglas. León Araúz y Reimerink (2018) destacan, también, que las opciones de búsqueda se deberían mejorar en los recursos terminográficos.

Nos guiamos por estas opiniones de los usuarios y, como hemos comentado, dotamos al recurso de este tipo de datos. Estos se organizaron conceptualmente siempre que el tipo de información lo permitiera, pues la estructuración conceptual no está exenta de dificultades (p. ej. ya hemos comentado que la organización por dimensiones entra en conflicto con la organización por conceptos). En cuanto a las opciones de búsqueda, al tratarse de un recurso digital, pudimos ofrecer diferentes puertas de acceso a la información, que cada usuario podrá elegir en función de sus necesidades. Entre ellas destaca la búsqueda de combinaciones semánticas, que permite realizar consultas basadas en el significado. En este sentido, ofrecimos tanto búsquedas simples como avanzadas, con las que se pueden hacer consultas a medida.

En general, pretendimos adoptar un enfoque sistemático para así paliar el tratamiento heterogéneo que reciben los CN en los recursos lingüísticos, además de problemas específicos como la omisión de formantes del CN, la presentación de varios CN en una misma línea o la falta de información semántica de CN presumiblemente transparentes. De este modo, esperamos que el recurso sea beneficioso para los usuarios de EcoLexicon, algo que evaluaremos en próximos trabajos. Asimismo, en el futuro nos gustaría implementar nuevas posibilidades de búsqueda, como las palabras truncadas (*generat.** para obtener diferentes terminaciones de esta raíz), así como incorporar el resto de lenguas de EcoLexicon y realizar la conexión con el módulo de colocaciones verbales. Además, desarrollaremos una guía de uso en la que se detallarán las diferentes funcionalidades de la herramienta. Igualmente, sería interesante asignar el dominio al que pertenece cada término, una tarea que presenta complicaciones, ya que los conceptos

pueden pertenecer a más de un dominio o encontrarse en la frontera entre varios, entre otros aspectos. En definitiva, el diseño de un recurso terminográfico es una labor compleja que necesita de su utilización por parte de los usuarios para mejorarse continuamente.

6. Conclusions

This PhD thesis focused on CNs in terminology. In particular, it targeted the formation, translation, and representation of these multiword terms. CNs are common in different languages, such as English and Spanish. They are essential in general language, but particularly in specialized discourse, where they are one of the main term formation mechanisms, thus playing a significant role in conceptual systems. This is due to their exceptional contribution to lexical and terminological expansion by means of concept combination.

From a wider phraseological perspective, we assumed CNs to be specialized phraseological units, despite those that consider that they fall outside the scope of phraseology. In CNs, two or more concepts combine to form a unit of meaning. From our point of view, they are phraseological units because they are multiword expressions, whose components combine frequently and must not be treated as independent parts, but rather as a whole. Following Meyer and Mackintosh (1994, 1996), we did not distinguish between CNs and noun collocations, because the difference is not always clear or even necessary. Instead, this study was based on their relevance in the conceptual system and referred to both of them as *complex nominals* or *compuestos nominales*, a denomination that highlights the structure of these combinations.

However, in spite of the large body of research dealing with CNs, there are several aspects that required further investigation, such as the characteristics of highly specialized CNs, their formation by any number and type of constituents, their translation and presence in languages other than English, and their representation in terminographic tools.

With a view to unveiling the ins and outs of CNs in specialized texts, we followed the methodology of Frame-Based Terminology (Faber et al. 2005, 2006; Faber 2012). The terminological analysis in this theory is mostly based on the use of specialized corpora. Along these lines, a comparable corpus in English and Spanish composed of specialized texts on wind power was manually compiled. The authority as well as the contents and design of the resources were the guarantee of the reliability and validity of the texts (Buendía Castro 2013). Wind power is an attractive domain because of its increasing presence as a result of environmental awareness, and the lack of studies on these terms in this specialized field, especially in Spanish.

The choice of compiling a comparable instead of a parallel corpus was motivated by the added value of texts originally written in a language, which avoids non-idiomatic or inappropriate expressions, as well as possible mistakes. Additionally, parallel corpora are scarce and may not include the languages and domain in question. The corpus analysis tool Sketch Engine (Kilgarriff et al. 2004, 2014) was used, which provides users with a wide range of possibilities for concordance and combinatorial potential analysis, among other questions.

We thus started from five of the key terms of the wind power domain (i.e. *wind, power, turbine, voltage*, and *generator*) and extracted English CNs formed from these terms. Neither the number and type of constituents, nor the form of these multiword terms were restricted, with a view to analyzing the whole spectrum of CNs in the wind power domain. For that purpose, corpus queries based on regular expressions were designed, which were a reflection of the different forms and elements that English CNs can acquire.

We focused on a set of English CNs, which were then semantically analyzed. First, bracketing or structural disambiguation was necessary in CNs consisting of three terms or more. For example, the structure of *generator terminal voltage* was *generator* [*terminal voltage*]. Once the internal dependencies had been ascertained, semantic categories were assigned to the CNs. This annotation concerned: (i) the individual elements of CNs; (ii) the groups resulting from bracketing; and (iii) the concept derived from the combination. An inventory of semantic categories developed in the LexiCon research group from the concepts in EcoLexicon was used, which includes different levels of categorization. For instance, in *voltage stability*, *voltage* is an ATTRIBUTE>MEASUREMENT>MAGNITUDE and *stability* is an ATTRIBUTE>PHYSICAL ATTRIBUTE>STATE. The concept designated by *voltage stability* is also an ATTRIBUTE>PHYSICAL ATTRIBUTE>STATE.

The semantic relations that bind CN constituents were examined. To this end, verb paraphrases (Nakov and Hearst 2006, 2008) were queried in the corpus through the use of regular expressions (e.g. a *power curve* is a curve that *represents/plots* power). When no results were retrieved, free paraphrases (i.e. occurrences of CN constituents that reveal the internal relation) were searched for; in other cases, CN concordances were analyzed. The relation inventory in EcoLexicon was used to identify the relations (e.g. in *power curve*, curve *represents* power).

After unveiling the conceptual propositions underlying CNs, semantic roles were assigned to their constituents, which reflect the participants in the proposition.

Traditionally, the arguments of a predicate have been the units labeled with semantic roles. However, we decided to assign them to CN constituents because they have concealed propositions (although, strictly speaking, they are not predicates). For example, in *turbine manufacturer*, *turbine* is the PATIENT and *manufacturer* is the AGENT.

CNs were then grouped with their denominative variants, which had been identified during documentation and by means of knowledge pattern-based queries in the corpus. These conceptual clusters were the starting point for the search for Spanish equivalents, which was carried out by means of a distributional analysis in both corpora. This study allowed the examination of the three main aspects of CNs: their formation, their translation, and their representation.

However, before dealing with these practical questions, one of our specific objectives was to address the specific characteristics of CNs in specialized discourse (Cabezas García and Faber 2018). In this regard, the following conclusions were drawn:

-Not surprisingly, English CNs have received more attention than these combinations in other languages, such as Spanish. The major focus has been on two-term CNs formed by nouns. However, longer CNs also exist and can pose different challenges. In addition, modifiers that belong to other parts of speech, such as adjectives, are also important because they can also reflect relations and attributes. General language has been the most popular type of discourse for the study of CNs, whereas further investigation on specialized CNs should be carried out, which takes into account their important role in conceptual systems. In fact, in Martin's (1992) classification, specialized CNs are *conceptual collocations*, which respond to conceptual combinations that are feasible in a particular domain and are predictable. This is in direct contrast to *lexical collocations*, which are typical in general language and are arbitrary. Furthermore, psycholinguistic theories that explain conceptual combination, which is the basis for CN formation, have failed to account for these cognitive and communicative processes in specialized language.

-Although bracketing is the first step towards subsequent semantic tasks (e.g. relation analysis), the development of clues that facilitate CN structural disambiguation still needs to be further explored in specialized discourse as well as in languages other than English. In particular, such a guide should be easily applicable in areas such as terminology or translation since most research has been carried out in the domain of Computational Linguistics.

-CN semantics is not fully compositional. The use of a constituent with a special sense (Kocourek 1982; Nakov 2013; Fernández Domínguez 2019), as well as relation omission (Levi 1978; Grant and Bauer 2004), and the ambiguity of prepositions such as *of* and *de* (Kocourek 1982) can affect CN compositionality. With a view to describing CN conceptual content, semantic categories have been used in the literature. These allow us to capture regularities on CNs, which is also possible by means of semantic role labeling, although this type of inventory has been less frequently applied in CNs. Accordingly, semantic categories and roles facilitate relation inference in new CNs (Rosario et al. 2002; Maguire et al. 2010b; *inter alia*). For instance, CNs formed by the combination MATERIAL + OBJECT usually codifies the *made_of* relation. Ascertaining CN internal relations has been one of the main lines of research, although its application to all types of CNs in different languages needs to be further developed.

-Variation (Fernández Silva 2018) and neology (Sanz Vicente 2011; Humbley and García Palacios 2012) can usually be found in specialized CNs. In this respect, the language of primary term formation, which is often English, influences neology formation in Spanish. CN translation becomes thus essential since English is the *lingua franca* (Humbley and García Palacios 2012) and their CNs must be properly transferred to the target language. For that purpose, techniques for enhancing equivalent identification in corpus should be developed, especially given the unsystematic treatment of CNs in terminographic resources. These should focus on translators and terminologists, because most advances have emerged in the computational area, whose techniques are not easily employed by translators and terminologists in their everyday tasks. Together with the need for translation, the systematic description of CNs in terminographic resources is also a necessity. Nevertheless, representation of other types of phraseological units has been the focus although the difficulties posed by CNs in terminology, as well as their essential role for knowledge transfer highlight the need for addressing their description.

After exploring this issue, our objective was to analyze CNs semantically (Cabezas García and Faber 2017b; Cabezas García and Gil Berrozpe 2018; Cabezas García and León Araúz 2019). Based on this, the following conclusions were reached:

-Specialized CN bracketing can be solved by a new protocol, which integrates the following indicators: (a) the possible cluster has an independent, term status in the corpus; (b) the cluster does not allow the insertion of external elements modifying its meaning; (c) the cluster forms other CNs; and (d) the cluster has synonyms or antonyms. The final protocol was derived from the classification of these indicators, based on their

reliability (a, c, d, b), which facilitates bracketing identification without the need to follow the whole protocol.

-Access to the conceptual combinations in CNs can be achieved using semantic categories. The category designated by the CN usually coincided with that of the head, although it can also be more specific or even completely different (e.g. in *wind penetration*, the head belongs to the PROCESS>ACTION category, whereas the CN designates the ATTRIBUTE>MEASUREMENT>MAGNITUDE>LEVEL category). General language words were also found in specialized CNs, such as the adjective in *apparent power*. These can even acquire a specific value that is different from its general meaning (e.g. in *small wind turbine*, *small* alludes not only to the size but also to the power capacity). The main contribution of semantic categories is the uncovering of regular patterns in the formation of wind power CNs, which highlights the combinatorial potential of concepts (e.g. the combination ATTRIBUTE + ENTITY>CREATION gives rise to several CNs, such as *asynchronous generator*, *large wind farm* or *constant-speed wind turbine*). Semantic categories also revealed conceptual and denominative variation in CNs.

-Verbal and free paraphrases proved to be useful for the decoding of internal relations in CNs. Our study revealed the potential of these paraphrases. On the one hand, the verbs obtained can be used to achieve finer-grained relations (e.g. the relation *affects* was transmitted by means of different predicates, such as *control* or *adjust*, *fluctuate* or *vary*, or *transmit* and *transport*, which specify how a concept can affect another). Additionally, free paraphrases may reveal new knowledge patterns, which facilitate relation identification in texts. Furthermore, both types of paraphrase offer additional semantic and contextual information. Combination restrictions were also found at the relational level. In other words, relations in CNs cannot be randomly assigned since they depend on the semantic content of the concepts combined.

-Semantic roles captured regularities in CNs based on the functions of their constituents in the underlying propositions. The CNs studied were found to be derived from a limited number of role schemas (e.g. the POSSESSOR-THEME schema generated CNs such as *line-to-neutral voltage* or *wind turbine blade*). These can encode the same or different relations. Along these same lines, the combinatorial potential of the head is revealed by these relation schemas, which indicate the argument structure of this concept.

This analysis was the basis for other specific objectives. One of them was to explore CN formation from a semantic perspective by means of a slot opening/filling

mechanism activated by the nominal heads (Cabezas García and Faber 2018; Faber and Cabezas García 2019). This gave rise to the following conclusions:

-In CNs, two or more concepts converge. However, this does not happen by chance. On the contrary, the combinatorial potential of CN heads revealed that concepts belonging to a specific nuclear category (those found in the CN head) exhibit similar combination patterns. An example can be found in *turbine, generator, converter* and *rotor*, among others, which belong to the ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT nuclear category.

-In particular, the open slots are filled by similar concepts, thus giving rise to specialized CNs. These slots can be defined according to the semantic roles of the concepts that occupy them, and are usually the same for the different concepts or subcategories included in the nuclear category. Specifically, the slots are filled by similar semantic categories, which are also bound to the head by similar semantic relations. For example, one of the slots opened by ENTITY>CREATION is PATIENT, which is filled by the PROCESS>MOVEMENT>ENERGY MOVEMENT category and can encode the *causes* or *affects* relation. Some of the CNs resulting from this pattern include *AC generator* (generator *causes* AC) and *voltage source converter* (converter *affects* voltage source). Therefore, this approach clearly and systematically integrates semantic roles, categories and relations, which are the three main elements that characterize the underlying conceptual propositions in CNs.

-These slots determine the CNs that can be formed from a particular head. In our approach, this is conceived as the argument structure or *microcontext* of the nuclear category or concept, which are a sort of semantic profile that reflects combinatorial potential. Microcontexts are thus at the origin of CN productivity. A CN can also include more than one slot, which results in longer and more cognitively complex combinations. This is directly related to multidimensionality since the inclusion of several slots often entails allusion to different perspectives (e.g. *stall-regulated horizontal axis wind turbine*).

-In addition to serving as an explanation for CN formation, microcontexts are also useful for recreating the underlying conceptual system in CNs. Along these lines, they can be used for crafting definitions, which at the same time, are like a conceptual snapshot where slots can be identified.

CN formation is directly linked to their transference to other languages. For this reason, another of our objectives was to investigate CN translation in English and Spanish

(Cabezas García and Faber 2018) with regard to equivalence identification in corpus, correspondence analysis, and equivalence proposal. We reached the next conclusions:

-Since users such as translators or terminologists may not be able to find CNs in terminographic resources, they should master corpus query techniques for equivalent identification. We presented a procedure that describes the manual application of the premises of distributional semantics in a comparable corpus. We thus assumed that a term meaning can be grasped from its distribution across different contexts and that context match shows semantic coincidence. We observed that CN equivalents can be found in corpus by: (i) identifying contextual elements of a concept in the source language; (ii) using the translated contextual items in the target corpus, together with a part of the CN whose equivalence is known; and (iii) selecting CN equivalents that are related to contextual elements in the same way as the original language. This procedure has the advantage of showing equivalences that are not linear or word-to-word translations. However, it is true that the search is directed to complex forms in the target language, which on the other hand are usually the trend when translating CNs into Spanish. In order not to limit the results to these forms, this initial query was complemented by the concordance analysis of the equivalent or equivalents found, which can reveal additional variants. Therefore, finding CN equivalents in corpus is possible by means of a deep contextual and conceptual analysis.

-The comparison of English CNs with their Spanish equivalents confirmed that term formation patterns are different in these languages. First, morphosyntactic disparities were evident, which mostly entailed the premodification by nouns in English against the postmodification by adjectives and prepositional phrases in Spanish. Furthermore, the bracketing of English CNs was essential to determine the location of the prepositions that may be used in Spanish CNs, as well as the position and agreement of adjectives in this language.

-Linear or word-to-word translations of CNs into Spanish were the norm. However, this was not always the case since CNs often included a different number of constituents, which can have a different semantic load. Denominative as well as conceptual variation were usually parallel in both languages, although differences were also found, which highlight the importance of a contrastive study of these mechanisms.

-Finally, the English footprint was evident in Spanish correspondences. This influence was retrieved in the form of calques, which usually coexist alongside a more natural form in Spanish (e.g. *aerogenerador regulado por pérdida aerodinámica* vs.

aerogenerador de regulación por pérdida aerodinámica). Loan words were also frequent, for example in the form of English abbreviations widely extended in Spanish (even more than the correspondent abbreviation in this language), e.g. *PSD* from *power spectral density*. Neology comes inevitably to mind since these are probably concepts that were initially named in English to facilitate international communication and later required translation into Spanish. Very often, the role of English as the language of international exchange led to the widespread use of many English specialized terms in other languages. Neology can also lead to the absence of CNs in terminographic resources. In those cases, translators or terminologists should propose an equivalent denomination. To this end, contrastive studies such as the one in this thesis can provide useful guidelines, thanks to the general picture of CN formation in both languages.

Our last specific objective was to design a systematic proposal for CN representation in the terminological knowledge base EcoLexicon (Cabezas García and Faber 2018; Cabezas García and Gil Berrozpe 2018). This resulted in the following conclusions:

-Developing a linguistic resource entails more than deciding which data will be shown to users. The most complex part of creating such a resource is the design of the internal system, the search protocol, and the template to be filled in by terminologists. This included different types of data, such as semantic or morphosyntactic information.

-As for visualization, different views were proposed so as to offer several access paths to information and satisfy the various user needs. The views were: (i) *CN formation/Formación de CN* (CN formation from one or more terms); (ii) *Equivalents/Equivalentes* (CN translation between English and Spanish); (iii) *Morphosyntactic combinations/Combinaciones morfosintácticas* (morphosyntactic annotation of CN constituents); (iv) *Semantic combinations/Combinaciones semánticas* (semantic categories, roles and relations that are present in CNs); and (v) *Summary/Resumen* (general picture of the information in the CN module).

-To enhance user experience, simple as well as advanced queries were devised. The latter facilitate customized searches by means of different query elements that differ in each view (e.g. terms, parts of speech, or semantic categories, roles and relations). In addition, Boolean operators can complement these queries.

-Data were shown by means of an enriched structure. Although this structure differs in each view of the module, when possible a conceptual organization was adopted.

For example, CNs were ranged according to the position of the term entered (if any), i.e. head or modifier of the CN. Additionally, the grouping of term variants in concepts was also specified, as well as the relation of hyponyms with their hypernyms, and the organization according to the different dimensions emphasized by terms. Our commitment to conceptual organization is in consonance with the way concepts are stored in the brain, and thus facilitates knowledge acquisition.

-We provided a wide range of information, which can be centered, based on user preferences. Additional details can also be retrieved by means of secondary options that are tailored to each view. In general, several aspects of the CN module mark the difference. Some of them are the customized queries that allow the use of conceptual and morphosyntactic elements to direct the search, the conceptual organization, and the interlinguistic comparison of morphosyntactic and semantic patterns of CNs. However, it is undeniable that the complexity of the design process may reveal shortcomings or mistakes when fully implemented. For this reason, a revision and evaluation process is essential, and improvements will undoubtedly be carried out.

Along these lines, the work developed in this thesis opens new lines of research that have not as yet been addressed because of the characteristics of the study, in some cases, and the limitations encountered, in other. The main lines for future work include the following:

-Exploration of a different specialized domain to observe CN behavior and confirm whether the same tendencies are found.

-Use of a parallel English-Spanish corpus to analyze the leanings and possible mistakes in CN translation.

-Study of the usage context of denominative variants. Although this was not carried out in this research, it is our aim to include the different uses of term variants in the CN module.

-Design of the connection between the CN and the verb collocation modules. This task could not be achieved because of the in-depth study necessary for the computational implementation in EcoLexicon, which falls outside the scope of this work. However, it is our plan to develop this connection because of the added value to the phraseological module and its usefulness for rich textual productions.

-Development of a user guide for the CN module. Time constraints hindered the completion of this task, although it is important for the effective use of the different functionalities of the module.

-User evaluation of the CN module. Although this was not one of our objectives, it is certainly relevant for the improvement of the resource.

Since this thesis focused on CNs, it addressed an essential part of specialized discourse. In particular, it investigated how knowledge is disseminated by means of these multiword terms, which were studied in their multiple facets: morphosyntactic and semantic formation, translation, and representation. It is thus a thorough, comprehensive proposal from which the area of Terminology, as well as terminology users, are expected to benefit. In summary, our aim was to show the exceptional potential of these terms, which have the power to condense knowledge.

Bibliografía

Adams, V. 2001. *Complex Words in English*. Abingdon, Nueva York: Taylor & Francis.

Aguado de Cea, G. 2007. «La fraseología en las lenguas especializadas». En *Las lenguas profesionales y académicas*, E. Alcaraz Varó et al. (eds.), 53-65. Madrid: Ariel.

Aitchison, J. 2003. *Words in the Mind: An Introduction to the Mental Lexicon* (3.ª ed.). Oxford: Black-well.

Allen, M. R. 1978. *Morphological Investigations*. Tesis doctoral. University of Connecticut.

Allwood, J. 2003. «Meaning potentials and context: Some consequences for the analysis of variation in meaning». En *Cognitive Approaches to Lexical Semantics*, H. Cuyckens et al. (eds.), 29-66. Berlín: Mouton de Gruyter.

Alonso, A. y Torner, S. 2010. «Adjectives and collocations in specialized texts: lexicographical implications». En *Proceedings of the XIV EURALEX International Congress*, A. Dykstra y T. Schoonheim (eds.), 872-881. Leeuwarden: Fryske Akademy.

Arntz, R. y Picht, H. 1995. *Introducción a la Terminología*. Madrid: Fundación Germán Sánchez Ruipérez.

Atkins, B. T. S. y Rundell, M. 2008. *The Oxford Guide to Practical Lexicography*. Oxford, Nueva York: Oxford University Press.

Baker, C. F., Fillmore, C. J. y Lowe, J. B. 1998. «The Berkeley FrameNet Project». En *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL '98) and 17th International Conference on Computational Linguistics (COLING '98)*, 86–90. ACL: Montreal.

Baldwin, T. y Kim, S. N. 2010. «Multiword Expressions». En *Handbook of Natural Language Processing, Second Edition*, N. Indurkhya y F. J. Damerau (eds.), 267-292. Boca Ratón: CRC Press.

Bally, C. 1951[1909]. *Traité de stylistique française*. Heidelberg: C. Winter.

Bally, C. 1965. *Linguistique génerale et lingustique française.* Berna: Francke Verlag.

Balyan, R. y Chatterjee, N. 2015. Translating noun compounds using semantic relations. *Computer Speech and Language 32*, 91-108.

Barker, K. y Szpakowicz, S. 1998. «Semi-automatic recognition of noun modifier relationships». En *Proceedings of the 17th International Conference on Computational Linguistics, ICCL'98*, 96–102. Montreal: ACL.

Barrière, C. y Ménard, P. A. 2014. «Multiword noun compound bracketing using Wikipedia». En *Proceedings of the First Workshop on Computational Approaches to Compound Analysis (ComAComA)*, 72-80. Dublín: ACL, Dublin City University.

Barsalou, L. 2003. Situated simulation in the human conceptual system. *Language and Cognitive Processes 5*(6), 513–562.

Bassac, C. y Bouillon, P. 2013. «The telic relationship in compounds». En *Advances in Generative Lexicon Theory*, J. Pustejovsky et al. (eds.), 109–126. Dordrecht: Springer.

Bauer, L. 2008. «Les composés exocentriques de l'anglais». En *La composition dans une perspective typologique*, D. Amiot (ed.), 35-47. Arras: Artois Presses Université.

Bauer, L. y Tarasova, E. 2013. The meaning link in nominal compounds. *Skase Journal of Theoretical Linguistics 10*(3), 2-18.

Bednarek, M. 2008. Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory 4*, 119–139.

Beigbeder Atienza, F. 2006. *Diccionario técnico inglés-español español-inglés*, 2.ª ed. Madrid: Díaz de Santos.

Béjoint, H. 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics 2*(3), 207-222.

Benson, M., Benson, E. e Ilson, R. F. 1986. *Lexicographic Description of English. Studies in Language Companion Series* 14. Ámsterdam, Filadelfia: John Benjamins.

Bergenholtz, H. y Tarp, S. 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Ámsterdam, Filadelfia: John Benjamins.

Bergenholtz, H. y Tarp, S. 2010. «LSP Lexicography or Terminography? The Lexicographer's Point of View». En *Specialized Dictionaries for Learners*, P. A. Fuertes Olivera (ed.), 27-37. Berlín, Nueva York: Mouton de Gruyter.

Bevilacqua, C. R. 2004. *Unidades fraseológicas eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Tesis doctoral. Barcelona: Universitat Pompeu Fabra.

Bouamor, D., Semmar, N. y Zweigenbaum, P. 2012. «Identifying bilingual Multi-Word Expressions for Statistical Machine Translation». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, N. Calzolari et al. (eds.), 674-679. Estambul: ELRA.

Bouillon, P., Jezek, E., Melloni, C. y Picton, A. 2012. «Annotating qualia relations in Italian and French complex nominals». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, N. Calzolari et al. (eds.), 1527–1532. Estambul: ELRA.

Bowker, L. 1997. «Multidimensional Classification of Concepts and Terms». En *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*, S. E. Wright y G. Budin (eds.), 133-143. Ámsterdam, Filadelfia: John Benjamins.

Bowker, L. 1998. Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta 43*(4), 631-651.

Bowker, L. 2003. Lexical knowledge patterns, semantic relations, and language varieties: Exploring the possibilities for refining information retrieval in an international context. *Cataloging and Classification Quarterly*, *37*(1/2), 153-171.

Bowker, L. 2015. «Terminology and translation». En *Handbook of Terminology*, H. Kockaert y F. Steurs (eds.), 304-323. Ámsterdam, Filadelfia: John Benjamins.

Buendía Castro, M. 2013. *Phraseology in specialized language and its representation in environmental knowledge resources*. Tesis doctoral. Granada: Universidad de Granada.

Buendía Castro, M. y Faber, P. 2014. Collocation Dictionaries: A Comparative Analysis. *MonTI. Monografías de Traducción e Interpretación 6*, 203-235.

Buendía Castro, M. y Faber, P. 2015. Phraseological units in English-Spanish legal dictionaries: a comparative study. *Fachsprache 37*(3-4), 161-175.

Buendía Castro, M. y Faber, P. 2016. «Phraseological correspondence in English and Spanish specialized texts». En *Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives = Fraseología*

computacional y basada en corpus: perspectivas monolingües y multilingües, G. Corpas Pastor (ed.), 391–398. Ginebra: Tradulex.

Butnariu, C. y Veale, T. 2008. «A concept-centered approach to noun-compound interpretation». En *Proceedings of the 22nd International Conference on Computational Linguistics. COLING '08*, 81–88. Mánchester: COLING.

Cabezas García, M. y Chambó, S. En prensa. Multi-word term variation: Prepositional and adjectival complex nominals in Spanish. *RESLA (Revista Española de Lingüística Aplicada*).

Cabezas García, M. y Faber, P. 2017a. «A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources». En *Computational and Corpus-Based Phraseology*, R. Mitkov (ed.), Lecture Notes in Computer Science 10596, 145-159. Cham: Springer.

Cabezas García, M. y Faber, P. 2017b. «Exploring the semantics of multi-word terms by means of paraphrases». En *Temas actuales de terminología y estudios sobre el léxico*, M. A. Candel Mora y C. Vargas Sierra (eds.), Interlingua 172, 193-217. Granada: Comares.

Cabezas García, M. y Faber, P. 2017c. The role of micro-contexts in noun compound formation. *Neologica*. *La néologie en terminologie 11*, 101-118.

Cabezas García, M. y Faber, P. 2018. Phraseology in Specialized Resources: an Approach to Complex Nominals. *Lexicography 5*(1), 55-83.

Cabezas García, M. y Faber, P. 2019. Semantic prosody and semantic preference in multi-word terms. *Fachsprache 41*(1-2), 2-21.

Cabezas García, M. y Gil Berrozpe, J. C. 2018. «Semantic-based Retrieval of Complex Nominals in Terminographic Resources». En *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, 269-281. Liubliana: EURALEX.

Cabezas García, M. y León Araúz, P. 2018. «Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study». En *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al. (eds.), 2511-2518. Miyazaki: ELRA.

Cabezas García, M. y León Araúz, P. 2019. On the structural disambiguation of multi-word terms. En *Computational and Corpus-Based Phraseology*, G. Corpas Pastor y R. Mitkov (eds.). Springer.

Cabré, M. T. 1993. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida, Empúries.

Cabré, M. T. 1999. *La terminología: representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Cabré, M. T. 2003. Theories of terminology: their description, prescription and explanation. *Terminology 9*(2), 163-199.

Cabré, M. T., Estopà, R. y Lorente Casafont, M. 1996. «Terminología y Fraseología». En *Actas del V Simposio de Terminología Iberoamericana*, 67-81. Ciudad de México: Colegio de México.

Cabré, M. T. y Estopà, R. 2005. «Unidades de conocimiento especializado, caracterización y tipología». En *Coneixement, llenguatge i discurs especialitzat*, M. T. Cabré y C. Bach (eds.), 69-94. Barcelona: IULA, Documenta Universitaria.

Cabré, M. T., Estopà, R. y Vargas Sierra, C. 2012. Neology in specialized communication. *Terminology 18*(1), 1-8.

Cabré, M. T. y Nazar, R. 2011. Towards a new approach to the study of neology. *Cahiers de lexicologie 6*, 63-80.

Cammack, R., Atwood, T., Campbell, T., Parish, H., Smith, A., Vella, F. y Stirling, J. 2006. *Oxford Dictionary of Biochemistry and Molecular Biology (2nd ed.)*. Nueva York: Oxford University Press.

Carrió Pastor, M. L. y Candel Mora, M. A. 2013. Variation in the translation patterns of English complex noun phrases into Spanish in a specific domain. *Languages in Contrast 13*(1), 28-45.

Carvalho, S. 2018. *A terminological approach to knowledge organization within the scope of endometriosis: the EndoTerm project*. Tesis doctoral. Lisboa: Universidade Nova de Lisboa.

Carvalho, S., Costa, R. y Roche, C. 2018. «The Role of Conceptual Relations in the Drafting of Natural Language Definitions: an Example from the Biomedical Domain». En

En *Proceedings of the LREC 2018 Workshop "Globalex 2018 – Lexicography & WordNets"*, I. Kerneman y S. Krek (eds.), 10-16. Miyazaki: Globalex.

Castillo Carballo, M. A. 2006. *Las colocaciones y la fraseología*. Madrid: Liceus.

Catford, J. C. 1965. *A Linguistic Theory of Translation*. Oxford: Oxford University Press.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. 1970. «Remarks on Nominalization». En *Readings in English Transformational Grammar*, R. Jacobs y P. Rosenbaum (eds.), 184-221. Waltham: Ginn.

Cohen, A. y Staub, A. 2014. Online processing of novel noun–noun compounds: Eye movement evidence. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY 67*(1), 147-165.

Cohen, B. 2011. *Lexique de cooccurrents : bourse et conjoncture économique*. Montreal: Linguatech.

Cohen, K., Palmer, M. y Hunter, L. 2008. Nominalization and Alternations in Biomedical Language. *PLoS ONE 3*(9), e3158.

Collet, T. 2003. A two-level grammar of the reduction processes of French complex terms in discourse. *Terminology 9*(1), 1-27.

Cordeiro, S., Ramisch, C., Idiart, M. y Villavicencio, A. 2016. «Predicting the compositionality of nominal compounds: Giving word embeddings a hard time». En *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016*, 1986-1997. Berlín: ACL.

Corpas Pastor, G. 1996. *Manual de fraseología española*. Madrid: Gredos.

Corpas Pastor, G. 2001. En torno al concepto de colocación. *EUSKERA – XLVI 2001*(1), 89-108.

Corpas Pastor, G. 2013. «Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas». En *Fraseopragmática*, I. Olza Moreno y E. Manero Richard (eds.), 335-374. Berlín: Frank & Timme.

Coseriu, E. 1981. *Lecciones de lingüística general*. Madrid: Gredos.

Costa, R. 2017. Les normes en terminologie. Que faire des synonymes ? *Cahiers de lexicologie : Revue internationale de lexicologie et lexicographie 110*, 45-57.

Coulson, S. 2001. *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Nueva York: Cambridge University Press.

Cowie, A. P. 1981. The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics 2*(3), 223-235.

Croft, W. y Cruse, D.A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Daille, B. 1999. «Identification des adjectifs relationnels en corpus». En *Proceedings of TALN 1999*, P. Amsili (ed.), 105-114. Cargèse: ATALA.

Daille, B. 2001. «Qualitative Terminology Extraction: Identifying Relational Adjectives». En *Recent Advances in Computational Terminology*, D. Bourigault et al. (eds.), 149-166. Ámsterdam: John Benjamins.

Daille, B. 2017. *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Ámsterdam: John Benjamins.

Daille, B., Dufour-Kowalski, S. y Morin, E. 2004. «French-English Multi-word Term Alignment Based on Lexical Context Analysis». En *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004),* M. T. Lino et al. (eds.), 919-922. Lisboa: ELRA.

Daille, B. y Morin, E. 2005. «French-English Terminology Extraction from Comparable Corpora». En *IJCNLP 2005*, R. Dale et al. (eds.), Lecture Notes in Computer Science 3651, 707-718. Berlín, Heidelberg: Springer.

Déjean, H. y Gaussier, E. 2002. Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables. *Lexicometrica 2002*, 1-22.

Downing, P. 1977. On the creation and use of English compound nouns. *Language 53*, 810-842.

Drouin, P., Francœur, A., Humbley, J. y Picton, A. 2017. *Multiple Perspectives on Terminological Variation*. Ámsterdam: John Benjamins.

Durán Muñoz, I. 2010. «Specialized lexicographical resources: A survey of translators' needs». En *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of ELEX 2009. Cahiers du Cental 7*, S. Grandger y M. Paquot (eds.), 55-66. Louvain-la-Neuve: Presses Universitaires de Louvain.

Durán Muñoz, I. 2019. Adjectives and their Keyness. A Corpus-based Analysis in English Tourism. *Corpora 14*(3).

Estopà, R. 2000. Los adjetivos en las unidades terminológicas poliléxicas: un análisis morfosemántico. *Organon 14*(28-29), 233-246.

Evans, V. y Green. M. 2006. *Cognitive Linguistics: An Introduction*. Edimburgo: Edinburgh University Press.

Evans, V., Bergen, B. K. y Zinken, J. 2007. The cognitive linguistics enterprise: an overview. En *The cognitive linguistics reader*, V. Evans et al. (eds.), 2-36. Londres: Equinox.

Faber, P. 2009. The cognitive shift in terminology and specialized translation. *MonTI. Monografías de Traducción e Interpretación* 1, 107–134.

Faber, P. 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlín, Boston: De Gruyter Mouton.

Faber, P. 2015. «Frames as a framework for terminology». En *Handbook of Terminology*, H. Kockaert y F. Steurs (eds.), 14-33. Ámsterdam, Filadelfia: John Benjamins.

Faber, P. y Cabezas García, M. 2019. Specialized Knowledge Representation: From Terms to Frames. *Research in Language 17*(2), 197-211.

Faber, P. y León Araúz, P. 2014. «Specialized knowledge dynamics: From cognition to culture-bound terminology». En *Dynamics and Terminology. An interdisciplinary perspective on monolingual and multilingual culture-bound communication*, R. Temmerman y M. Van Campenhoudt (eds.), Terminology and Lexicography Research and Practice 16, 135-158. Ámsterdam, Filadelfia: John Benjamins.

Faber, P. y León Araúz, P. 2016. Specialized knowledge representation and the parameterization of context. *Frontiers in Psychology 7*(196), 1-20.

Faber, P., León Araúz, P. y Reimerink, A. 2016. «EcoLexicon: new features and challenges». En *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, I. Kernerman et al. (eds.), 73-80. Portorož: Globalex.

Faber, P., León Araúz, P. y Reimerink, A. 2011. «Knowledge representation in EcoLexicon». En *Technological Innovation in the Teaching and Processing of LSPs:*

Proceedings of TISLID, N. Talaván et al. (eds.), 367-385. Madrid: Universidad Nacional de Educación a Distancia.

Faber, P., López Rodríguez, C. I. y Tercedor Sánchez, M. 2001. Utilización de técnicas de corpus en la representación del conocimiento médico. *Terminology* 7(2), 167–198.

Faber, P. y López Rodríguez, C. I. 2012. «Terminology and specialized language». En *A cognitive linguistics view of terminology and specialized language*, P. Faber (ed.), 20, 9-31. Berlín, Boston: De Gruyter Mouton.

Faber, P. y Mairal, R. 1999. *Constructing a lexicon of English verbs*. Berlín: Mouton de Gruyter.

Faber, P. y Mairal, R. 2002. «Reconsidering lexical representations in Functional Grammar». En *New Perspectives on Argument Structure in Functional Grammar*, R. Mairal y M. J. Pérez Quintero (eds.), 39-94. Berlín: Mouton de Gruyter.

Faber, P., Márquez Linares, C. y Vega Expósito, M. 2005. Framing Terminology: A process-oriented approach. *Meta: Journal des traducteurs 50*(4).

Faber, P., Montero Martínez, S., Castro Prieto, M. R., Senso Ruiz, J., Prieto Velasco, J. A., León Arauz, P., Márquez Linares, C. y Vega Expósito, M. 2006. Process-oriented terminology management in the domain of Coastal Engineering. *Terminology 12*(2), 189-213.

Faber, P. y Navajas, R. 2016. «Secondary Term Formation and Term Stability in Genetics: An English-Spanish Corpus Study». En *The Language of Science and the Science of Language*, E. Petkova et al. (eds.), 262-279. Sofía: Bulgarian Academy of Sciences.

Faber, P., Verdejo, J., León Araúz, P., Reimerink, A. y Guzmán, G. 2014. Neural Substrates of Specialized Knowledge Representation: An fMRI study. *Revue française de linguistique appliquée 19*(1), 15-32.

Fauconnier, G. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, Nueva York: Cambridge University Press.

Fauconnier, G. 1997. *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fellbaum, C. A. 1998. Semantic Network of English: The Mother of All WordNets. *Computers and the Humanities 32*, 209–220.

Fernández Domínguez, J. 2010. N+N Compounding in English: Semantic Categories and the Weight of Modifiers. *Brno Studies in English 36*(1), 47-76.

Fernández Domínguez, J. 2016. A morphosemantic investigation of term formation processes in English and Spanish. *Languages in Contrast 16*(1), 54–83.

Fernández Domínguez, J. 2019. «Compounds and multi-word expressions in Spanish». En *Complex Lexical Units. Compounds and Multi-Word Expressions*, B. Schlücker (ed.), 189-219. Berlín, Boston: De Gruyter.

Fernández Silva, S. 2011. *Variación terminológica y cognición. Factores cognitivos en la denominación del concepto especializado*. Tesis doctoral. Barcelona: Universitat Pompeu Fabra.

Fernández Silva, S. 2013. Variación denominativa y punto de vista. *Revista Debate Terminológico 9*, 11-37.

Fernández Silva, S. 2016. The Cognitive and Rhetorical Role of Term Variation and its Contribution to Knowledge Construction in Research Articles. *Terminology 22*(1), 52–79.

Fernández Silva, S. 2017. La estructura cognitiva de la variación terminológica: el papel de la clase conceptual en la denominación de los conceptos especializados. *Terminàlia 15*(2017), 15-26.

Fernández Silva, S. 2018. The cognitive and communicative functions of term variation in research articles: a comparative study in Psychology and Geology. *Applied Linguistics 2018*: 1-23.

Fernández Silva, S. y Kerremans, K. 2011. Terminological variation in source texts and translations: a pilot study. *Meta. Journal des Traducteurs 56*(2), 318-335.

Ferrando Aramo, V. 2002. «Colocaciones y compuestos sintagmáticos: dos fenómenos léxicos colindantes». En *Léxico y gramática: [Selección de ponencias e comunicacións presentadas no Congreso Internacional de Lingüística "Léxico & Gramática"]*, M. González Pereira et al. (eds.), 99-107. Lugo: Tris Tram.

Ferretti, T. R. y Gagné, C. L. 2006. The recovery of thematic role structure during noun-noun interpretation. *Psychonomic Bulletin and Review 13*(3), 423-428.

Fillmore, C. J. 1977. Scenes-and-frames semantics. *Linguistic Structures Processing 5*, 55-81.

Fillmore, C. J. 1982. «Frame semantics». En *Linguistics in the Morning Calm*, The Linguistic Society of Korea (ed.), 111–137. Hanshin.

Fillmore, C. J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica 6*(2), 222–254.

Fillmore, C. J. y Kay, P. 1987. The goals of Construction Grammar. *Berkeley Cognitive Science Report 50*. University of California at Berkeley.

Fillmore, C. J., Kay, P. y O'Connor, M. C. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language 64*(3), 501-538.

Finin, T. 1980. *The semantic interpretation of compound nominals*. Tesis doctoral. Urbana: University of Illinois.

Firth, J. R. 1957. *Papers in Linguistics 1934-1951*. Londres: Oxford University Press.

Fontenelle, T. 1998. «Discovering Significant Lexical Functions in Dictionary Entries». En *Phraseology: Theory, Analysis, and Applications*, A. P. Cowie (ed.), 189-207. Oxford: Clarendon Press.

Fradin, B. 2017. The multifaceted nature of denominal adjectives. *Word Structure 10*(1), 27-53.

Freixa, J. 2002. *La variació terminològica. Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient*. Tesis doctoral. Barcelona: Universitat de Barcelona.

Freixa, J. 2006. Causes of Denominative Variation in Terminology: A typology proposal. *Terminology 12*(1), 51-77.

Freixa, J. y Fernández Silva, S. 2017. «Terminological variation and the unsaturability of concepts». En *Multiple Perspectives on Terminological Variation*, P. Drouin et al. (eds.), 155-181. Ámsterdam: John Benjamins.

Fuertes Olivera, P. A. y Tarp, S. 2011. Lexicography for the third millennium: Cognitive-oriented specialised dictionaries for learners. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos 21*, 141-161.

Fuertes Olivera, P. A. y Tarp, S. 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography and Terminography*. Lexicographica Series Maior 146. Berlín, Boston: De Gruyter.

Gagné, C. L. 2000. Relational-Based Combinations Versus Property-Based Combinations: A Test of the CARIN Theory and the Dual-Process Theory of Conceptual Combination. *Journal of Memory and Language 42*, 365-389.

Gagné, C. L. 2001. Relation and Lexical Priming during the Interpretation of Noun-Noun Combinations. *Journal of Experimental Psychology: Learning Memory and Cognition 27*(1), 236-254.

Gagné, C. L. 2002. Lexical and Relational Influences on the Processing of Novel Compounds. *Brain and Language 81*, 723-735.

Gagné, C. L. y Shoben, E. J. 1997. Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning Memory and Cognition 23*(1), 71-87.

Gagné, C. L. y Spalding, T. L. 2013. Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun-noun compounds. *Psychology of Learning and Motivation 59*, 97-130.

Gangemi, A., Pisanelli, D. M. y Steve, G. 2000. «Understanding Systematic Conceptual Structures in Polysemous Medical Terms». En *Proceedings of the AMIA Symposium 2000*, M. J. Overhage (ed.), 285-289. Filadelfia: Hanley & Belfus.

García de Quesada, M., Olivera Fuertes, P. A. y Montero Martínez, S. 2002. Propuesta de estructura definicional terminográfica en OntoTerm®. *Terminology 8*(1), 57-90.

García Page, M. 2001. ¿Son las expresiones fijas expresiones fijas? *Moenia: Revista lucense de Lingüística & Literatura 7*, 165-196.

García Palacios, J. y Sanz Vicente, L. 2012. El papel de la traducción en la formación secundaria de términos sintagmáticos. *Neologica 6*, 95-109.

García Palacios, J., De Sterck, G., Linder, D., Maroto, N., Sánchez Ibáñez, M. y Torres del Rey, J. 2016. *La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones*. Frankfurt: Peter Lang.

Gaudin, F. 1993. *Pour une socioterminologie: des problèmes sémantiques aux pratiques institutionnelles*. Ruan: Publications de l'Université de Rouen.

Gaudin, F. 2003. *Socioterminologie: une approche sociolinguistique de la terminologie*. Bruselas: Duculot.

Gaudin, F. 2005. La socioterminologie. *Langages 157*(1), 81-93.

Geeraerts, D., Grondelaers, S. y Bakema, P. 1994. *The structure of lexical variation. Meaning, naming, and context*. Berlín, Nueva York: Mouton de Gruyter.

Gil Berrozpe, J. C., León Araúz, P. y Faber, P. 2019. «Ontological Knowledge Enhancement in EcoLexicon». En *Proceedings of the 6th International Conference on Electronic Lexicography in the 21st Century (eLex 2019)*. Sintra: Lexical Computing CZ s.r.o.

Girju, R. 2007. «Improving the Interpretation of Noun Phrases with Cross-linguistic Information». En *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 568-575. Praga: ACL.

Girju, R., Moldovan, D., Tatu, M. y Antohe, D. 2005. On the semantics of noun compounds. *Computer Speech & Language 19*(4), 479–496.

Gläser, R. 1988. «The grading of idiomaticity as a presupposition for a taxonomy of idioms». En *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*, W. Hüllen y R. Schulze (eds.), 264-279. Tübingen: Max Niemeyer.

Gläser, R. 1994. Relations between Phraseology and Terminology with Special Reference to English. *ALFA* 7/8, 41-60.

Gledhill, C. y Pecman, M. 2018. On Alternating Pre-Modified and Post-Modified Nominals Such As Aspirin Synthesis Versus Synthesis of Aspirin: Rhetorical and Cognitive Packing in English Science Writing. *Fachsprache 40*(1-2), 26-48.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Gouadec, D. 1994. «Nature et traitement des entités phraséologiques. Terminologie et phraséologie. Acteurs et amenageurs». En *Actes du deuxième Université d'Automne en Terminologie*, 164-193. París: La Maison du Dictionnaire.

Gouadec, D. 2005. Terminologie, traduction et rédaction spécialisées. *Langages 157*(1), 14-24.

Gràcia Solé, L., Cabré Castellví, M. T., Varela Ortega, S. y Azkarate Villar, M. 2000. *Configuración morfológica y estructura argumental: léxico y diccionario*. Bilbao: Universidad del País Vasco.

Granger, S. y Meunier, F. 2008. *Phraseology: An interdisciplinary perspective*. Ámsterdam, Filadelfia: John Benjamins.

Grant, L. y Bauer, L. 2004. Criteria for Re-defining Idioms: Are we Barking up the Wrong Tree? *Applied Linguistics 25*(1), 38-61.

Gregory, M. y Carroll, S. 1978. *Language and Situation: Language Varieties and their Social Contexts*. Londres, Henley, Boston: Routledge & Kegan Paul.

Grön, L. y Bertels, A. 2018. Clinical sublanguages: Vocabulary structure and its impact on term weighting. *Terminology 24*(1), 41-65.

Grön, L., Bertels, A. y Heylen, C. 2018. «The Interplay of Form and Meaning in Complex Medical Terms: Evidence from a Clinical Corpus». En *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, 18–29. Santa Fe: ACL.

Gruber, T. R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition 5*(2), 199-220.

Güemes, M. M., Gattei, C., Tabullo, A. y Wainselboim, A. 2016. El papel de la estructura argumental en el procesamiento de los compuestos verbo-nombre del español: Evidencias experimentales. *Interdisciplinaria 33*(2), 1-17.

Guilbert, L. 1965. *La formation du vocabulaire de l'aviation*. París: Larousse.

Guilbert, L. 1974. Grammaire générative et néologie lexicale. *Langages 36*, 34-44.

Hahn, U., y Chater, N. 1997. «Concepts and similarity». En *Studies in Cognition: Knowledge, Concepts, and Categories*, K. Lamberts y D. Shanks (eds.), 43-92. Cambridge, Londres: MIT Press.

Halliday, M. A. K. 1966. Patterns in words. *The Listener LXXV*(1920), 53-55.

Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. 2.ª ed. Londres: Arnold.

Hampton, J. 1991. «The combination of prototype concept». En *The Psychology of Word Meanings*, P. Schwanenflugel (ed.), 91-116. Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates.

Hanks, P. y Pustejovsky, J. 1995. A Pattern Dictionary for Natural Language Processing. *Revue française de linguistique appliquée 10*(2), 63-82.

Harastani, R., Daille, B. y Morin, E. 2013. «Identification, alignement, et traductions des adjectifs relationnels en corpus comparables». En *Vingtième conférence du Traitement Automatique du Langage Naturel 2013 (TALN 2013)*, 313-326. Sables d'Olonne: ATALA.

Hashimoto, C., Torisawa, K., De Saeger, S., Kazama, J. y Kurohashi, S. 2011. «Extracting paraphrases from definition sentences on the Web». En *ACL-HLT 2011 – Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 1087-1097. Portland: ACL.

Hausmann, F. J. 1989. «Le dictionnaire de collocations». En *Wörterbücher / Dictionaries / Dictionnaires — Ein internationals Handbuch zur Lexikographie / An International Enyclopedia of Lexicography / Enyclopédie internationale de lexicographie*, F. J. Hausmann et al. (eds.), 1010-1019. Berlín, Nueva York: Walter de Gruyter.

Hearst, M. A. 1992. «Automatic acquisition of hyponyms from large text corpora». En *Proceedings of the 14th International Conference on Computational Linguistics (COLING-1992)*, 539-545. Nantes: COLING.

Heid, U. 1992. Décrire les collocations. *Terminologie et Traduction 2*(3), 523-548.

Heid, U. 1994. «On Ways Words Work Together – Topics in Lexical Combinatorics». En *Euralex 1994: Proceedings*, W. Martin et al. (eds.), 226-257. Ámsterdam: EURALEX.

Heid, U. 2001. «Collocations in sublanguage texts: Extraction from corpora». En *Handbook of Terminology Management. Volume 2: Applications-Oriented Terminology Management*, S. E. Wright y G. Budin (eds.), 788-808. Ámsterdam, Filadelfia: John Benjamins.

Heid, U. 2008. «Computational phraseology: An overview». En *Phraseology: An interdisciplinary perspective*, S. Granger y F. Meunier (eds.), 337-360. Ámsterdam, Filadelfia: John Benjamins.

Hendrickx, I., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Szpakowicz, S. y Veale, T. 2013. «SemEval-2013 task 4: Free paraphrases of noun compounds». En *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 138-143. Atlanta: ACL.

Hidalgo Simón, A. 2004. *Elsevier's dictionary of medicine Spanish–English English–Spanish*. Ámsterdam: Elsevier.

Hoey, M. 2005. *Lexical Priming. A New Theory of Words and Language*. Abingdon: Routledge.

Humbley, J. 2006. «La néologie : interface entre ancien et nouveau». En *Langues et cultures: une histoire d'interface*, R. Greenstein (ed.), 91-103. París: Publications de la Sorbonne.

Humbley, J. y García Palacios, J. 2012. Neology and Terminological Dependency. *Terminology 18*(1), 59-85.

Hutcheson, H. 2001. «Practical Considerations for a Term Bank». En *Handbook of Terminology Management. Volume 2: Application-Oriented Terminology Management*, S. E. Wright y G. Budin (eds.), 666-676. Ámsterdam, Filadelfia: John Benjamins.

Jackendoff, R. 1983. *Semantics and Cognition*. Cambridge: MIT Press.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge: MIT Press.

Jackendoff, R. 2016. English noun-noun compounds in Conceptual Semantics. En *The Semantics of Compounding*, P. Ten Hacken (ed.), 15-37. Cambridge: Cambridge University Press.

Jespersen, O. 1942. *A modern English grammar: On historical principles*. Copenhage: Munksgaard.

Johnston, M. y Busa, F. 1999. «Qualia structure and the compositional interpretation of compounds». En *Breadth and depth of semantic lexicons*, E. Viegas (ed.), 167-187. Dordrecht: Springer.

Jousse, A. L. y Polguère, A. 2005. *Le DiCo et sa version DiCouèbe. Document descriptif et manuel d'utilisation. Document technique*. Montreal: Département de linguistique et de traduction, Université de Montréal.

Kageura, K. 1997. A preliminary investigation of the nature of frequency distributions of constituent elements of terms in terminology. *Terminology 4*(2), 199–223.

Kageura, K. 2002. *The dynamics of terminology: A descriptive theory of term formation and terminological growth*. Ámsterdam: John Benjamins.

Kageura, K. 2015. «Terminology and lexicography». En *Handbook of Terminology. Volume 1*, H. J. Kockaert y F. Steurs (eds.), 45-59. Ámsterdam, Filadelfia: John Benjamins.

Kay, P. y Fillmore, C. J. 1999. Grammatical constructions and linguistic generalizations: The *what's X doing Y?* construction. *Language 75*(1), 1-33.

Keizer, E. 2007. *The English noun phrase: The nature of linguistic categorization*. Cambridge: Cambridge University Press.

Kerremans, K. 2017. «Towards a resource of semantically and contextually structured term variants and their translations». *Multiple Perspectives on Terminological Variation*, P. Drouin et al. (eds.), 83-108. Ámsterdam: John Benjamins.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. y Suchomel, V. 2014. The Sketch Engine: ten years on. *Lexicography 1*(1), 7–36.

Kilgarriff, A., Rychlý, P., Smrz, P. y Tugwell, D. 2004. «The Sketch Engine». En *Proceedings of the 11th EURALEX International Congress,* G. Williams y S. Vessier (eds.), 105-115. Lorient: EURALEX.

Kim, S. N. y Baldwin, T. 2013. A lexical semantic approach to interpreting and bracketing English noun compounds. *Natural Language Engineering 19*(3), 385-407.

Kipper, K., Snyder, B. y Palmer, M. 2004. «Extending a verb-lexicon using a semantically annotated corpus». En *Proceedings of the 4th International Conference on Language Resources and Evaluation* (*LREC 2004*), 1557-1560. Lisboa: ELRA.

Kocourek, R. 1982. *La langue française de la technique et de la science: Vers une linguistique de la langue savante*. Wiesbaden: Brandstetter.

Krstev, C., Vitas, D. y Savary, A. 2006. «Prerequisites for a Comprehensive Dictionary of Serbian Compounds». En *Advances in Natural Language Processing*, T. Salakoski et al. (eds.), Lecture Notes in Computer Science 4139, 552-563. Berlín, Heidelberg: Springer.

Kübler, N. y Pecman, M. 2012. «The ARTES bilingual LSP dictionary: from collocation to higher order phraseology». En *Electronic Lexicography*, S. Granger y M. Paquot (eds.), 187-209. Oxford: Oxford University Press.

Kübler, N. y Volanschi. A. 2012. «Semantic Prosody and Specialised Translation, or How a Lexico-grammatical Theory of Language can Help with Specialised Translation». En *Corpus-Informed Research and Learning in ESP: Issues and Applications*, A. Boulton et al. (eds.), 103-134. Ámsterdam, Filadelfia: John Benjamins.

L'Homme, M. C. 1998. «Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale». En *Euralex 1998: Proceedings*, 513-522. Lieja: EURALEX.

L'Homme, M. C. 2000. Understanding Specialized Lexical Combinations. *Terminology 6*(1), 89-110.

L'Homme, M. C. 2002. «What can Verbs and Adjectives Tell us about Terms?». En *Proceedings of the 6th Terminology and Knowledge Engineering International Conference (TKE 2002)*, 1-6. Le Chesnay: INRIA.

L'Homme, M. C. 2009. «A methodology for describing collocations in a specialized dictionary». En *Lexicography in the 21st Century*, S. Nielsen y S. Tarp (eds.), 237-256. Ámsterdam: John Benjamins.

L'Homme, M. C. 2012. Adding syntactico-semantic information to specialized dictionaries: an application of the FrameNet methodology. *Lexicographica 28*(1), 233-252.

L'Homme, M. C. y Leroyer, P. 2009. Combining the semantics of collocations with situation-driven search paths in specialized dictionaries. *Terminology 15*(2), 258–283.

L'Homme, M. C. y Pimentel, J. 2012. «Capturing syntactico-semantic regularities among terms: An application of the FrameNet methodology to terminology». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 262–268. Estambul: ELRA.

L'Homme, M. C., Robichaud, B. y Prévil, N. 2018. «Browsing the Terminological Structure of a Specialized Domain: A Method Based on Lexical Functions and their Classification». En *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al. (eds.), 3079-3086. Miyazaki: ELRA.

Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Lakoff, G. 1990. The Invariance Hypothesis: is abstract reason based on image-schemas? *Cognitive Linguistics 1*(1), 39-74.

Lakoff, G. y Johnson, M. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. 1987. *Foundations of Cognitive Grammar, vol. I*. Stanford: Stanford University Press.

Lauer, M. 1995. *Designing Statistical Language Learners: Experiments on Noun Compounds*. Tesis doctoral. Sydney: Macquarie University.

Lees, R. B. 1960. *The Grammar of English Nominalizations.* Bloomington: Indiana University Press/La Haya: Mouton.

Lees, R. B. 1970. «Problems in the Grammatical Analysis of English Nominal Compounds». En *Progress in Linguistics*, M. Bierwisch y K. E. Heidolph (eds.), 174-186. La Haya: Mouton.

León Araúz, P. 2009. *Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura*. Tesis doctoral. Granada: Universidad de Granada.

León Araúz, P. 2017. «Term and Concept Variation in Specialized Knowledge Dynamics». En *Multiple Perspectives on Terminological Variation*, P. Drouin et al. (eds.), 213-258. Ámsterdam: John Benjamins.

León Araúz, P. y Faber, P. 2012. «Causality in the Specialized Domain of the Environment». En *Proceedings of the Workshop "Semantic Relations-II. Enhancing Resources and Applications" (LREC'12)*, V. B. Mititelu et al. (eds), 10-17. Estambul: ELRA.

León Araúz, P., Faber, P. y Montero Martínez, S. 2012. «Specialized Language Semantics». En *A cognitive linguistics view of terminology and specialized language*, P. Faber (ed.), 20, 95-175. Berlín, Boston: De Gruyter Mouton.

León Araúz, P. y Reimerink, A. 2016. «Evaluation of EcoLexicon Images». En *Joint Second Workshop on Language and Ontology & Terminology and Knowledge Structures (LangOnto2 + TermiKS) in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, F. Khan et al. (eds.), 16-22. Portorož.

León Araúz, P., Reimerink, A. y Faber, P. 2017. «EcoLexiCAT: a Terminology-enhanced Translation Tool for Texts on the Environment». En *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, I. Kosem et al. (eds), 321-341. Brno: Lexical Computing.

León Araúz, P. y Reimerink, A. 2018. «Evaluating EcoLexiCAT: A terminology-enhanced CAT tool». En *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al. (eds.), 2374–2381. Miyazaki: ELRA.

León Araúz, P., San Martín, A. y Faber, P. 2016. «Pattern-based word sketches for the extraction of semantic relations». En *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, 73–82. Osaka: ACL.

León Araúz, P. y San Martín, A. 2018. «The EcoLexicon Semantic Sketch Grammar: from Knowledge Patterns to Word Sketches». En *Proceedings of the LREC 2018 Workshop "Globalex 2018 – Lexicography & WordNets"*, I. Kerneman y S. Krek (eds.), 94-99. Miyazaki: Globalex.

Levi, J. 1975. *The Syntax and Semantics of Non-predicating Adjectives in English*. Tesis doctoral. University of California.

Levi, J. 1978. *The Syntax and Semantics of Complex Nominals*. Nueva York: Academic Press.

Lew, R. 2012. The Role of Syntactic Class, Frequency, and Word Order in Looking up English Multi-Word Expressions. *Lexikos 22*, 243-260.

Linder, D. 2002. «Translating noun clusters and 'nounspeak' in specialized computer text». En *Translating Science*, J. Chabas et al. (eds.), 261-270. Barcelona: Universitat Pompeu Fabra.

López Rodríguez, C. I., Faber, P., León Araúz, P., Prieto Velasco, J. A. y Tercedor Sánchez, M. 2010. La Terminología basada en marcos y su aplicación a las Ciencias Ambientales: los proyectos MARCOCOSTA y ECOSISTEMA. *Arena Romanistica 7* (10), 52–74.

López Rodríguez, C. I., Buendía Castro, M. y García Aragón, A. 2012. User needs to the test: Evaluating a terminological knowledge base on the environment by trainee translators. *JoSTrans. The Journal of Specialized Translation 18*, 57-76.

Lorente Casafont, M., Bevilacqua, C. R. y Estopá, R. 1998. «El análisis de la fraseología especializada mediante elementos de la lingüística actual». En *Actas del VI Simposio Iberoamericano de Terminología: Terminología, desarrollo e identidad nacional*, 647-666. La Habana: RITerm.

Lorente Casafont, M. 2002. «Terminología y fraseología especializada: del léxico a la sintaxis». En *Panorama actual de la terminología*, G. Guerrero y F. Pérez Ramos (eds.), 159-180. Granada: Comares.

Lorente Casafont, M., Martínez Salom, M. A., Santamaría Pérez, I. y Vargas Sierra, C. 2017. «Specialized collocations in specialized dictionaries». En *Collocations and other lexical combinations in Spanish: theoretical, lexicographical and applied perspectives*, S. Torner y E. Bernal (eds.), 200-222. Londres, Nueva York: Routledge.

Louw, P. 2006. Inclusion Strategies for Multi-word Units in Monolingual Dictionaries. *Lexikos 16*, 95-103.

Maguire, P., Maguire, R. y Cater, A. W. S. 2010a. The Influence of Interactional Semantic Patterns on the Interpretation of Noun-Noun Compounds. *Journal of Experimental Psychology: Learning Memory and Cognition 36*(2), 288-297.

Maguire, P., Wisniewski, E. J. y Storms, G. 2010b. A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory 6*(1), 49-73.

Mairal, R., Periñán, C. y Samaniego, E. 2011. «Using ontologies for terminological knowledge representation: a preliminary discussion». *En Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10*, N. Talaván et al. (eds.), 267-280. Madrid: Universidad Nacional de Educación a Distancia.

Maniez, F. 2008. Using the Web and computer corpora as language resources for the translation of complex noun phrases in medical research articles. *Panace@ 8*(26), 162-167.

Maniez, F. 2009. L'adjectif dénominal en langue de specialité: étude du domaine de la médicine. *Revue française de linguistique appliquée 14*, 117-130.

Maniez, F. 2013. The translation into French of adjectives formed with a noun and a past participle in English-language medical articles. *Panace@ 14*(38), 240-247.

Maniez, F. 2014. «Implantation of English Terms Including the -ING Morpheme in French, Spanish and Italian: A Corpus-Based Study of the Debates of the European

Parliament». En *La néologie en langue de spécialité : détection, implantation et circulation des nouveaux termes*, P. Dury et al. (eds.), 189-201. Lyon: Travaux du CRTT.

Marchand, H. 1969. *The categories and types of present-day English word-formation*. Múnich: C. H. Beck.

Marcus, M. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge: MIT Press.

Maroto García, N., L'Homme, M. C. y Alcina, A. 2012. «Semantic relations established by specialized processes expressed by nouns and verbs: identification in a corpus by means of syntactico-semantic annotation». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, N. Calzolari et al. (eds.), 3814-3819. Estambul: ELRA.

Marsh, E. 1984. «A computational analysis of complex noun phrases in navy messages». En *Proceedings of the 10th International Conference on Computational Linguistics*, 505-508. Standford: ACL.

Martín Arista, F. J. 1997. La representación subyacente de los compuestos nominales en una gramática funcional del inglés. *Atlantis: Revista de la Asociación Española de Estudios Anglo-Norteamericanos 19*(2), 169-175.

Martín Mingorance, L. 1989. «Functional grammar and lexematics». En *Meaning and lexicography*, J. Tomaszczyk y B. Lewandowska (eds.), 227–253. Ámsterdam: John Benjamins.

Martin, W. 1992. Remarks on Collocations in Sublanguages. *Terminologie et Traduction 2*(3), 157-164.

Martínez, R. y Schmitt, N. 2012. A Phrasal Expressions List. *Applied Linguistics 33*(3), 299-320.

Mel'čuk, I. y Zholkovskij, A. 1970. Towards a Functioning Meaning-Text Model of Language. *Linguistics 57*, 10-47.

Mel'čuk, I. et al. 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain*. *Recherches lexico-sémantiques I, II, III, IV*. Montreal: Les Presses de l'Université de Montréal.

Mel'čuk, I., Clas, A. y Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Men, H. 2018. *Vocabulary Increase and Collocation Learning. A Corpus-Based Cross-sectional Study of Chinese Learners of English*. Springer/Shanghai Jiao Tong University Press 2018.

Meyer, I. 2001. «Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework». En *Recent advances in computational terminology*, D. Bourigault et al. (eds.), 279–302. Ámsterdam: John Benjamins.

Meyer, I., Skuce, D., Bowker, L. y Eck, K. 1992. «Towards a new generation of terminological resources: an experiment in building a terminological knowledge base». En *Proceedings of the 14th Conference on Computational Linguistics (COLING '92)*, 956–960. Nantes: ACL.

Meyer, I. y Mackintosh, K. 1994. «Phraseme Analysis and Concept Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon». En *Euralex 1994: Proceedings*, W. Martin et al. (eds.), 339-348. Ámsterdam: EURALEX.

Meyer, I. y Mackintosh, K. 1996. Refining the terminographer's concept-analysis methods: How can phraseology help? *Terminology 3*(1), 1-26.

Minsky, M. 1975. «A framework for representing knowledge». En *The Psychology of Computer Vision*, P. H. Winston (ed.), 211-277. Nueva York: McGraw-Hill.

Miyata, R. y Kageura, K. 2016. «Constructing and Evaluating Controlled Bilingual Terminologies». En *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, 83-93. Osaka: ACL.

Moldovan, D., Badulescu, A., Tatu, M., Antohe, D. y Girju, R. 2004. «Models for the semantic classification of nonnominalized noun phrases». En *Proceedings of the HLT Computational Lexical Semantics workshop*, 60-67. Boston: ACL.

Montero Fleta, M. B. 1995. *La estructura del grupo nominal complejo en el inglés científico escrito: sus componentes premodificadores y sus correspondencias en español*. Tesis doctoral. Valencia: Universitat de València.

Montero Fleta, M. B. 1996. Technical Communication: Complex Nominals Used to Express new Concepts in Scientific English - Causes and Ambiguity in Meaning. *ESPecialist 17*(1), 57-72.

Montero Martínez, S. 2002. *Estructuración conceptual y formalización terminográfica de frasemas en el subdominio de la oncología*. Tesis doctoral. Valladolid: Universidad de Valladolid.

Montero Martínez, S. 2008. «A constructional Approach to Terminological Phrasemes». En *Proceedings of the XIII EURALEX International Congress*, E. Bernal y J. DeCesaris (eds.), 1015-1022. Barcelona: Institut Universitari de Lingüística Aplicada/Documenta Universitaria.

Montero Martínez, S. y Buendía Castro, M. 2012. «La sistematización en el tratamiento de las construcciones fraseológicas: el caso del medio ambiente». En *Empiricism and Analytical Tools for 21st Century Applied Linguistics*, I. Elorza et al. (eds.), 711-724. Salamanca: Ediciones de la Universidad de Salamanca.

Montero Martínez, S. y Buendía Castro, M. 2017. Clasificación semántica de colocaciones verbales para la adquisición y codificación de conocimiento experto: El caso de los riesgos naturales. *Revista Española de Lingüística Aplicada 30*(1), 240-272.

Montero Martínez, S., Fuertes Olivera, P. A. y García de Quesada, M. 2001. The Translator as' Language Planner': Syntactic Calquing in an English-Spanish Technical Translation of Chemical Engineering. *Meta: Journal des traducteurs 46*(4), 687-698.

Montero Martínez, S., García de Quesada, M. y Fuertes Olivera, P. A. 2002. Terminological phrasemes in OntoTerm: A new theoretical and practical approach. *Terminology 8*(2), 177-206.

Montero Martínez, S. y García de Quesada, M. 2003. Terminological analysis for translation. *Perspectives: Studies in Translatology 11*(4), 293-314.

Montero Martínez, S. y García de Quesada, M. 2004. Designing a corpus-based grammar for pragmatic terminographic definitions. *Journal of Pragmatics 36*(2), 265-291.

Moon, R. 1998. *Fixed Expressions and Idioms in English*. *A Corpus-Based Approach*. Oxford: Clarendon Press.

Morley, J. y Partington, A. 2009. A few frequently asked questions about semantic — or evaluative — prosody. *International Journal of Corpus Linguistics 14*, 139-158.

Murphy, G. L. 1988. Comprehending complex concepts. *Cognitive Science 12*, 529-562.

Murphy, G. L. 1990. Noun phrase interpretation and conceptual combination. *Journal of Memory and Language 29*, 259-288.

Murphy, G. L. y Lassaline, M. E. 1997. «Hierarchical Structure in Concepts and Basic Level of Categorization». En *Studies in Cognition: Knowledge, Concepts, and Categories*, K. Lamberts y D. Shanks (eds.), 93-131. Cambridge, Londres: MIT Press.

Nakov, P. 2008. «Paraphrasing Verbs for Noun Compound Interpretation». En *Proceedings of the LREC 2008 Workshop: Towards a Shared Task for Multiword Expressions (MWE 2008)*, 46-49. Marrakech.

Nakov, P. 2013. On the interpretation of noun compounds: syntax, semantics, and entailment. *Natural Language Engineering 19*, 291-330.

Nakov, P. y Hearst, M. 2005. «Search engine statistics beyond then-gram: application to noun compound bracketing». En *Proceedings of the Ninth Conference on Computational Natural Language Learning, CoNLL '05*, 17-24. Ann Arbor: ACL.

Nakov, P. y Hearst, M. 2006. «Using Verbs to Characterize Noun-Noun Relations». En *Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2006*, J. Euzenat y J. Domingue (eds.), 233-244. Berlín: Springer.

Nakov, P. y Hearst, M. 2008. «Solving relational similarity problems using the web as a corpus». En *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, 452-460. Columbus: ACL.

Nakov, P. y Hearst, M. 2013. Semantic interpretation of noun compounds using verbal and other paraphrases. *ACM Transactions on Speech and Language Processing 10*(3), 1-51.

Nastase, V., Nakov, P., Ó Séaghdha, D. y Szpakowicz, S. 2013. *Semantic relations between nominals*. Morgan & Claypool.

Nastase, V., Sayyad-Shirabad, J., Sokolova, M. y Szpakowicz, S. 2006. «Learning Noun-Modifier Semantic Relations with Corpus-based and WordNet-based Features». En *Proceedings of the 21st National Conference on Artificial Intelligence*, A. Cohn (ed.), 781-786. Boston: AAAI Press.

Nastase, V. y Szpakowicz, S. 2003. «Exploring noun-modifier semantic relations». En *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, J. Geertzen et al. (eds.), 285-301. Tilburg.

Navigli, R. y Ponzetto, S. P. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence 193*, 217-250.

Nesselhauf, N. 2004. *Collocations in a Learner Corpus*. Ámsterdam, Filadelfia: John Benjamins.

Nida, E. A. 1966[1959]. Principles of Translation as Exemplified by Bible Translating. *The Bible Translator 10*(4), 148-164.

Nida, E. A. y Taber, C. R. 1982. *The Theory and Practice of Translation*. Leiden: The United Bible Societies.

Nielsen, S. 2010. «Specialised Translation Dictionaries for Learners». En *Specialised Dictionaries for Learners*, P. A. Fuertes Olivera (ed.), 69-82. Berlín, Nueva York: De Gruyter.

Nkwenti-Azeh, B. 1998. «Information Mediation: The Interface between Terminology and Translation». En *Encyclopedia of Library and Information Science.* Vol. 62 (Supplement 25), A. Kent (ed.), 157-170. Nueva York, Basilea, Hong Kong: Marcel Dekker, Inc.

Nkwenti-Azeh, B. 2001. «User-Specific Terminological Data Retrieval». En *Handbook of Terminology Management. Volume 2: Application-Oriented Terminology Management*, S. E. Wright y G. Budin (eds.), 600-612. Ámsterdam, Filadelfia: John Benjamins.

Nord, C. 1997. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Mánchester: St. Jerome Publishing.

Nulty, P. 2007. «Semantic Classification of Noun Phrases Using Web Counts and Learning Algorithms». En *Proceedings of the ACL 2007 Student Research Workshop*, 79-84. Praga: ACL.

Nulty, P. y Costello, F. 2013. General and specific paraphrases of semantic relations between nouns. *Natural Language Engineering 19*, 357-384

Ó Séaghdha, D. 2007a. «Designing and evaluating a semantic annotation scheme for compound nouns». En *Proceedings of Corpus Linguistics 2007*, 1-17. Birmingham.

Ó Séaghdha, D. 2007b. «Annotating and Learning Compound Noun Semantics». En *Proceedings of the ACL 2007 Student Research Workshop*, 73-78. Praga: ACL.

Ó Séaghdha, D. y Copestake, A. 2013. Interpreting compound nouns with kernel methods. *Natural Language Engineering 19*(3), 331-356.

Observatoire de linguistique Sens-Texte. 2019a. *Dictionnaire fondamental de l'environnement. DiCoEnviro*.

Observatoire de linguistique Sens-Texte. 2019b. *Dictionnaire fondamental de l'informatique et de l'Internet. DiCoInfo.*

Osorio Olave, G. y Serra Sepúlveda, S. 2012. Colocaciones, compuestos sintagmáticos y locuciones nominales: hacia un intento de delimitación conceptual. *Lenguas Modernas 39*, 103-116.

Oster, U. 2003. Los términos de la cerámica en alemán y español. Análisis semántico orientado a la traducción de los compuestos nominales alemanes. Tesis doctoral. Castellón: Universitat Jaume I.

Oster, U. 2005. *Las relaciones semánticas de términos polilexemáticos*. Frankfurt am Main: Peter Lang.

Oster, U. 2006. Classifying domain-specific intraterm relations: a schema-based approach. *Terminology 12*(1), 1-17.

Palmer, H. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.

Palmer, M., Gildea, D. y Kingsbury, P. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics 31*(1), 71-106.

Parra Escartín, C. 2018. ¿Cómo ha evolucionado la traducción automática en los últimos años? *La Linterna del Traductor 16*, 20-28.

Parra Escartín, C., Losnegaard, G. S., Samdal, G. I. L. y Patiño García, P. 2013. «Representing Multiword Expressions in Lexical and Terminological Resources: An Analysis for Natural Language Processing Purposes». En *Proceedings of the eLex 2013 Conference*, 338-357. Liubliana, Tallín: Trojina, Institute for Applied Slovene Studies, Eesti Keele Instituut.

Parra Escartín, C., Nevado Llopis, A. y Sánchez Martínez, E. 2018. «Spanish multiword expressions: Looking for a taxonomy». En *Multiword expressions: Insights from a Multi-lingual perspective*, M. Sailer y S. Markantonatou (eds.), 271-323. Berlín: Language Science Press.

Pavel, S. 1993. «Neology and phraseology as terminology-in-the-making». En *Terminology: applications in interdisciplinary communication*, H. B. Sonneveld y K. L. Loening (eds.), 21-34. Ámsterdam: John Benjamins.

Pawley, A. 2001. Phraseology. Linguistics and the Dictionary. *International Journal of Lexicography 14*(2), 122-134.

Pawley, A. 2007. «Developments in the study of formulaic language since 1970: A personal view». En *Phraseology and Culture in English*, P. Skandera (ed.), 3-45. Berlín: Mouton de Gruyter.

Pawley, A. y Syder, F. H. 1983. «Two puzzles for linguistic theory: nativelike selection and nativelike fluency». En *Language and Communication*, J. C. Richards y R. W. Schmidt (eds.), 191-227. Londres: Longman.

Pearce, D. 2002. «A comparative evaluation of collocation extraction techniques». En *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, 1530-1536. Las Palmas: ELRA.

Pearson, J. 1998. *Terms in Context.* Ámsterdam, Filadelfia: John Benjamins.

Pecman, M. 2012. Tentativeness in Term Formation. A Study of Neology as a Rhetorical Device in Scientific Papers. *Terminology 18*(1), 27-58.

Pecman, M. 2014. Variation as a cognitive device: how scientists construct knowledge through term formation. *Terminology 20*(1), 1-24.

Picht, H. 1991. Fraseología LSP desde el Punto de Vista Terminológico. *Sendebar 2*, 91-105 (traducido del alemán por Marián Hens Córdoba).

Picton, A. 2011. Picturing short-period diachronic phenomena in specialised corpora: A textual terminology description of the dynamics of knowledge in space technologies. *Terminology 17*(1), 134-156.

Pimentel, J. 2012. *JuriDiCO: Users Guide*. Montreal: Université de Montréal.

Pimentel, J. 2013. Methodological bases for assigning equivalents: a contribution. *Terminology 19*(2), 237-257.

Pimentel, J. y L'Homme, M. C. 2011. «Annotation syntaxico-sémantique de contextes spécialisés : application à la terminographie bilingue». En *Passeurs de mots, Passeurs d'espoir. Actes des 8èmes Journées du Réseau LTT – Lexicologie, Terminologie et Traduction*, M. Van Campenhoudt et al. (eds.), 651-670. París: Éditions des archives contemporaines.

Polguère, A. 2003. Étiquetage sémantique des lexies dans la base de données DiCo. *Traitement Automatique des Langues 44*(2), 39-68.

Postal, P. M. 1969. «Anaphoric islands». En *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, R. I. Binnick et al. (eds.), 209-239. Chicago: University of Chicago.

Pugh, J. 1984. «Contrastive Analysis of Noun Compound Terms in English, French, and Spanish within a Restricted, Specialized Domain». En *Proceedings LeXeter '93*, R. R. K. Hartmann (ed.), 395-400. Tübingen: Niemeyer.

Pustejovsky, J. 1991. The Generative Lexicon. *Computational Linguistics 17*(4), 409-441.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge: MIT Press.

Pustejovsky, J., Anick, P. y Bergler, S. 1993. Lexical semantic techniques for corpus analysis. *Computational Linguistics 19*(2), 331-358.

Quiroz, G. A. 2008. *Los sintagmas nominales extensos especializados en inglés y en español: descripción y clasificación en un corpus de genoma*. Tesis doctoral. Barcelona: Universitat Pompeu Fabra.

Rallapalli, S. y Paul, S. 2012. «A hybrid approach for the interpretation of nominal compounds using ontology». En *26th Pacific Asia Conference on Language, Information and Computation*, 554-563. Bali: Faculty of Computer Science, Universitas Indonesia.

Ramisch, C. 2015. *Multiword Expressions Acquisition: A Generic and Open Framework*. Cham: Springer.

Rapp, R. 2008. «The computation of associative responses to multiword stimuli». En *Proceedings of the COLING 2008 Workshop on cognitive aspects of the lexicon (COGALEX 2008)*, 102-109. Mánchester.

Real Academia de Ingeniería. 2014. *Diccionario Español de Ingeniería*.

Real Academia Nacional de Medicina. 2012. *Diccionario de términos médicos*. Madrid: Panamericana.

Reimerink, A., León Araúz, P. y Faber, P. 2016. Image selection and annotation for an environmental knowledge base. *Language Resources and Evaluation 50*(3), 443-474.

Reimerink, A. y León Araúz, P. 2018. «Manzanilla: An Image Annotation Tool for TKB Building». En *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al. (eds.), 1053-1059. Miyazaki: ELRA.

Reiss, K. y Vermeer, H. J. 1984. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.

Renouf, A. 2012. Defining neology to meet the needs of the translator: a corpus-based perspective. *Neologica 6*, 17-41.

Rey, A. 1976. Néologisme : un pseudo-concept ? *Cahiers de lexicologie 28*, 3-17.

Roberts, R. P. 1993. Phraseology: The State of the Art. *L'Actualité Terminologique / Terminology Update 26*(2), 4-8.

Roche, C. 2012. «Ontoterminology: How to unify terminology and ontology into a single paradigm». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, N. Calzolari et al. (eds.), 2626-2630. Estambul: ELRA.

Roche, C., Calberg-Challot, M., Damas, L. y Rouard, P. 2009. «Ontoterminology: A new paradigm for terminology». En *International Conference on Knowledge Engineering and Ontology Development*, 321-326. Madeira.

Rojas García, J. y Faber, P. 2019a. Extraction of terms for the construction of semantic frames for named bays. *Argentinian Journal of Applied Linguistics 7*(1), 27-57.

Rojas García, J. y Faber, P. 2019b. Extraction of Terms related to Named Rivers. *Languages 4*(46), 1-11.

Rondeau, G. 1981. *Introduction à la terminologie*. Chicoutimi: Gaëtan Morin.

Rosario, B. 2005. Extraction of semantic relations from bioscience text. Tesis doctoral. Berkeley: University of California.

Rosario, B. y Hearst, M. 2001. «Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy». En *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, 82-90. Pittsburgh.

Rosario, B., Hearst, M. y Fillmore, C. J. 2002. «The descent of hierarchy, and selection in Relational Semantics». En *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, P. Isabelle (ed.), 247-254. Filadelfia: ACL.

Rosch, E. 1978. «Principles of Categorization». En *Cognition and Categorization*, E. Rosch y B. Lloyd (eds.), 27-48. Hillsdale: Lawrence Erlbaum Associates.

Rossenbeck, K. 1989. «Lexikologische Und Lexikographische Probleme: Fachsprachlicher Phraseologie Aus Kontrastiver Sicht». En *Translation and Lexicography*, M. Snell-Hornby y E. Pöhl (eds.), 197-210. Ámsterdam, Filadelfia: John Benjamins.

Ruiz Yepes, G. 2017. «Hybrid Methods for the Extraction and Comparison of Multilingual Collocations in Languages for Specific Purposes». En *Proceedings of EUROPHRAS 2017*, R. Mitkov (ed.), 11-18. Londres: Tradulex.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R. y Scheffczyk, J. 2010. *FrameNet II: Extended theory and practice*.

Sag, I. A., Baldwind, T., Bond, F., Copestake, A. y Flickinger, D. 2002. «Multiword expressions: A pain in the neck for NLP». En *Computational Linguistics and Intelligent Text Processing. Third International Conference, CICLing 2002*, A. Gelbukh (ed.), 1-15. Berlín, Heidelberg: Springer.

Sager, J. C. 1990. *A practical course in terminology processing*. Ámsterdam: John Benjamins.

Sager, J. C. 1997. «Term Formation». En *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*, S. E. Wright y G. Budin (eds.), 25-41. Ámsterdam, Filadelfia: John Benjamins.

Sager, J. C., Dungworth, D. y McDonald P. F. 1980. *English special languages. Principles and practice in science and technology*. Wiesbaden: Brandstetter Verlag.

Sager, J. C. y Kageura, K. 1995. «Concept classes and conceptual structures: their role and necessity in Terminology». En *Symposyum on French Language and Linguistics, ALFA,* 191-216.

Salager-Meyer, F. 1985. Syntax and Semantics of Compound Nominal Phrases in Medical English Literature: A Comparative Study with Spanish. *English for Specific Purposes 95*, 6-12.

Salehi, B., Cook, P. y Baldwin, T. 2014. «Detecting non-compositional MWE components using Wiktionary». En *Proceedings of EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing*, 1792-1797. Doha: ACL.

San Martín, A. 2016. *La representación de la variación contextual mediante definiciones terminológicas flexibles*. Tesis doctoral: Universidad de Granada.

San Martín, A., Cabezas García, M., Buendía Castro, M., Sánchez Cárdenas, B., León Araúz, P. y Faber, P. 2017. Recent advances in EcoLexicon. *Dictionaries: Journal of the Dictionary Society of North America 38*(1), 96–115.

SanJuan, E., Dowdall, J., Ibekwe-SanJuan, F. y Rinaldi, F. 2005. A symbolic approach to automatic multiword term structuring. *Computer Speech and Language 19*(4), 524-542.

Sanz Vicente, L. 2011. *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español*. Tesis doctoral. Salamanca: Universidad de Salamanca.

Sanz Vicente, L. 2012a. Approaching Secondary Term Formation through the Analysis of Multiword units: An English–Spanish Contrastive Study. *Terminology 18*(1), 105-127.

Sanz Vicente, L. 2012b. «Searching for patterns in the transfer of multiword units: a corpus-based contrastive study on secondary term formation». En *Proceedings of CHAT 2012. The 2nd Workshop on the Creation, Harmonization and Application of Terminology Resources. Co-located with TKE 2012*, T. Gornostav (ed.), 11-18. Linköping: Linköping University Electronic Press.

Schlomann, A. 1968. *Illustrierte technische Wörterbücher in sechs Sprachen: Deutsch, Englisch, Russisch, Französisch, Italienisch, Spanisch*. Múnich: Oldenbourg.

Schulze, B. M. y Christ, O. 1996. *The CQP user's manual*. Stuttgart: Universität Stuttgart.

Selkirk, E. O. 1982. *The Syntax of Words*. Cambridge: MIT Press.

Semmar, N. 2018. «A Hybrid Approach for Automatic Extraction of Bilingual Multiword Expressions from Parallel Corpora». En *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al. (eds.), 311-318. Miyazaki: ELRA.

Shahzad, I. Ohtake, K., Masuyama, S. y Yamamoto, K. 2000. Identifying Translations of Compound Nouns Using Non-aligned Corpora. En *Proceedings of the Workshop MAL'99*, 108-113.

Sharoff, S., Babych, B. y Hartley, A. 2009. 'Irrefragable answer' using comparable corpora to retrieve translation equivalents. *Language Resources and Evaluation 43*, 15-25.

Shreve, G. 2001. «Terminological Aspects of Text Production». En *Handbook of Terminology Management. Volume 2: Applications-Oriented Terminology Management*, S. E. Wright y G. Budin (eds.), 772-787. Ámsterdam, Filadelfia: John Benjamins.

Silva, R., Costa, R. y Ferreira, F. 2004. Entre langue générale et langue de spécialité : une question de collocations. *Éla. Études de linguistique appliquée 135*(3), 347-359.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. 1996. The search for units of meaning. *TEXTUS: English Studies in Italy 9*, 75-106.

Sinclair, J. y Moon, R. 1989. *Collins COBUILD Dictionary of Phrasal Verbs*. Londres: Harper Collins.

Smadja, F. 1993. Retrieving Collocations from Text: Xtract. *Computational Linguistics 19*(1), 143-177.

Smith, V., Barratt, D. y Zlatev, J. 2014. Unpacking noun-noun compounds: Interpreting novel and conventional food names in isolation and on food labels. *Cognitive Linguistics 25*(1), 99-147.

Spalding, T. L., Gagné, C. L., Mullaly, A. C. y Ji, H. 2010. Relation-based interpretations of noun–noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft 17*, 283-315.

Štekauer, P. 1998. *An onomasiological theory of English word-formation*. Ámsterdam: John Benjamins.

Štekauer, P. 2005. *Meaning Predictability in Word Formation*. Ámsterdam, Filadelfia: John Benjamins.

Štekauer, P., Valera, S. y Kőrtvélyessy, L. 2012. *Word-Formation in the World's Languages.* Cambridge: Cambridge University Press.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Stubbs, M. 2009. The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics 30*(1), 115-137.

Talmy, L. 2000. *Toward a cognitive semantics*, *Vol. 1.* Cambridge: MIT Press.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.

Taylor, J. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Temmerman, R. 2000. *Towards New Ways of Terminology Description: The sociocognitive approach*. Ámsterdam, Filadelfia: John Benjamins.

Temmerman, R. 2007. Approaches to terminology. Now that the dust has settled. *Synaps 20*, 27-36.

Temmerman, R. y Kerremans, K. 2003. «Termontography: Ontology building and the sociocognitive approach to terminology description». En *Proceedings of CIL 17*, E. Hajicová (eds.), 1-10. Matfyzpress.

Tercedor Sánchez, M. 1999. *La Fraseología en el Lenguaje Biomédico: Análisis desde las Necesidades del Traductor*. Tesis doctoral. Granada: Universidad de Granada.

Tercedor Sánchez, M. 2011. The cognitive dynamics of terminological variation. *Terminology 17*(2), 181-197.

Tercedor Sánchez, M., López Rodríguez, C. I. y Alarcón Navío, E. 2013. Identifying features of translation through multiword lexical units. *Belgian Journal of Linguistics 27*(1), 87-109.

Tercedor Sánchez, M., López Rodríguez, C. I. y Prieto Velasco, J. A. 2014. También los pacientes hacen terminología: retos del proyecto VariMed. *Panace@ 25*(39), 95-103.

Toledo Báez, M. C. y Martín Lorente, R. 2018. Colocaciones, locuciones y compuestos sintagmáticos bilingües (español-francés) sobre diabetes en el corpus comparable Cordiabicom. *Panace@ 19*(47), 106-114.

Turney, P. D. 2005. «Measuring semantic similarity by latent relational analysis». En *Proceedings of the 19th International Joint Conferences on Artificial Intelligence (IJCAI-05)*, 1136-1141. Edimburgo: Morgan Kaufman.

Tutin, A. y Grossman, F. 2002. Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de Linguistique appliquée 7*(1), 7-25.

Ureña Gómez-Moreno, J. M. 2015. Refining the understanding of novel metaphor in specialised language discourse. *Terminology 22*(1), 1-29.

Utsumi, A. 2014. A semantic space approach to the computational semantics of noun compounds. *Natural Language Engineering 20*, 185-234.

Vadas, D. y Curran, J. R. 2011. Parsing Noun Phrases in the Penn Treebank. *Computational Linguistics 37*(4), 753-809.

Valeontis, K. y Mantzari, E. 2007. «The linguistic dimension of terminology: principles and methods of term formation». En *1st Athens International Conference on Translation and Interpretation. Translation: Between Art and Social Science*, 1-13. Atenas.

Van Valin, R. D. y LaPolla, R. J. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.

Vanderwende, L. 1994. «Algorithm for automatic interpretation of noun sequences». En *Proceedings of the 15th Conference on Computational Linguistics (COLING 1994)*, M. Nagao y Y. Wilks (eds.), 782-788. Kyoto: ACL.

Vinay, J. P. y Darbelnet, J. 1973. *Stylistique Comparée du français et de l'anglais*. París: Didier.

Vinogradov, V. V. 1947. *Ob osnovnij Tipaj fraseologicheskij iedinits v ruskom yazike.* Moscú.

Vintar, S. 2010. Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology 16*(2),141-158.

Warren, B. 1978. *Semantic Patterns of Noun-Noun Compounds*. Göteborg: Acta Universitatis Gothoburgensis.

Warren, B. 2003. «The Role of Links and/or Qualia in Modifier-Head Constructions». En *Polysemy – Flexible Patterns of Meaning in Mind and Language*, B. Todd et al. (eds.), 233-251. Berlín, Nueva York: Mouton de Gruyter.

Wierzbicka, A. 1996. *Semantics: primes and universals*. Oxford, Nueva York: Oxford University Press.

Wisniewski, E. J. 1996. Construal and similarity in conceptual combination. *Journal of Memory and Language 35*, 434-453.

Wisniewski, E. J. 1997. When concepts combine. *Psychonomic Bulletin & Review 4*, 167-183.

Wisniewski, E. J. y Love, B. C. 1998. Relations versus properties in conceptual combination. *Journal of Memory and Language 38*(2), 177-202.

Wüster, E. 1968. *The Machine Tool. An Interlingual Dictionary of Basic Concepts.* Londres: Technical Press.

Wüster, E. 1979. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Viena, Nueva York: Springer in Komm.

Yadav, P., Jezek, E., Bouillon, P., Callahan, T. J., Bada, M., Hunter, L. E. y Cohen, K. 2017. Semantic relations in compound nouns: Perspectives from inter-annotator agreement. *Studies in Health Technology and Informatics 245*, 644–648.

Zimmer, K. E. 1971. Some General Observations about Nominal Compounds. *Working Papers on Language Universals 5*, 1-21.

Zuluaga, A. 1975. La fijación fraseológica. *Thesaurus XXX*(2), 225-248.

Anexo 1: Extracción inicial de CN en inglés

CN	Frecuencia	CN	Frecuencia
wind turbine	4821	fixed-speed wind turbine	68
wind power	2726	nuclear power	68
wind farm	2670	power spectrum	67
wind speed	2294	wind turbine manufacturer	67
wind energy	1452	power system reliability	64
power system	819	wind turbine generator	64
offshore wind farm	718	modern wind turbine	63
reactive power	664	impact of wind power	61
wind generation	634	voltage sag	61
power output	475	voltage fluctuation	60
offshore wind	418	voltage variation	60
induction generator	366	electric generator	58
synchronous generator	349	turbine operation	55
wind resource	336	real power	54
rated power	321	wind turbine application	54
wind project	285	vertical axis wind turbine	53
wind power plant	271	apparent power	52
power factor	246	permanent magnet generator	52
power plant	244	gas turbine	49
wind direction	244	renewable power	49
power production	216	DC-link voltage	48
wind power generation	212	impulse turbine	48
active power	209	constant-speed wind turbine	47
wind power production	209	voltage dip	47
power curve	199	electrical generator	45
wind shear	199	high voltage	45
offshore wind turbine	195	conventional wind turbine	43

output power	192	squirrel cage induction generator	43
high wind	184	generator rotor	41
power generation	177	voltage stability	41
rated wind speed	165	generator power	40
terminal voltage	164	DC voltage	39
wind plant	158	rotor voltage	39
wind industry	157	stator voltage	39
wind system	156	wind turbine model	39
mean wind speed	155	nominal voltage	32
high wind speed	148	network voltage	31
variable-speed wind turbine	148	input voltage	30
wind energy system	148	generator side	29
small wind turbine	146	low voltage	28
large wind farm	142	power output of a wind turbine	27
wind penetration	140	turbine hub height	27
power coefficient	137	downwind turbine	26
average wind speed	131	generator voltage	26
wind power capacity	131	permanent magnet synchronous generator	26
wind turbine rotor	129	voltage source converter	26
power quality	127	hydraulic turbine	25
turbine blade	127	AC voltage	24
wind turbine blade	123	commercial wind turbine	24
wind velocity	123	rated voltage	24
electrical power	122	turbine model	24
onshore wind farm	122	floating turbine	23
voltage level	122	phase voltage	22

generator speed	121	reaction turbine	22
voltage control	121	stall-regulated wind turbine	22
diesel generator	120	wound rotor induction generator	22
wind park	118	DC generator	21
offshore wind energy	115	system voltage	21
grid voltage	113	line-to-line voltage	20
offshore wind power	112	direct drive generator	19
power converter	112	induced voltage	19
mechanical power	106	medium voltage	19
large wind turbine	104	PCC voltage	19
power density	104	generator operation	18
wind condition	104	turbine generator	17
wind generator	104	conventional synchronous generator	16
maximum power	101	switched reluctance generator	16
low wind speed	100	battery voltage	15
generator torque	99	electricity generator	15
power electronic converter	99	harmonic voltage	15
turbine rotor	98	instantaneous voltage	15
voltage drop	98	voltage source inverter	15
wind power generation system	94	bus voltage	14
horizontal axis wind turbine	93	distributed generator	14
generated power	92	AC generator	13
offshore turbine	92	generator design	13
wind turbine design	89	steam generator	12
electric power	86	generator torque control	11
turbine design	85	power generator	11

turbine manufacturer	85	reference voltage	11
wind turbine system	82	generator terminal voltage	10
asynchronous generator	80	line-to-neutral voltage	10
doubly fed induction generator	80	voltage at the PCC	10
wind power fluctuation	75	direct voltage	9
output voltage	73	line voltage	9
turbine size	72	variable speed generator	8
conventional generator	71	renewable generator	6
power fluctuation	71	generator system	5
power station	71	self-excited induction generator	4
power grid	70	three-phase generator	4
wave power	70	generator of wind turbine	3