

ETIQUETAJE Y DESCRIPCIÓN DE UNIDADES FRASEOLÓGICAS
ESPECIALIZADAS EN UN DICCIONARIO BILINGÜE DE
COMERCIO INTERNACIONAL

JOSE LUIS ROJAS DÍAZ

Asesor:

Gabriel A. Quiroz Herrera

Doctor en Lingüística Aplicada

Trabajo de grado para optar al título de:

Magíster en Lingüística

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE COMUNICACIONES
MAESTRÍA EN LINGÜÍSTICA
MEDELLÍN
2014



UNIVERSIDAD
DE ANTIOQUIA
1803
FACULTAD DE
COMUNICACIONES

J. L. Rojas

Agradecimientos

Es imposible dar por terminado este trabajo sin antes agradecer a todos aquellos que de una u otra forma aportaron para que esta empresa, que inició hace dos años, pudiese llegar, el día de hoy, a feliz término.

Quiero expresar mi más sincero sentimiento de gratitud a mi tutor Gabriel Ángel Quiroz no sólo por su enorme paciencia, el tiempo invertido, el material proporcionado, las charlas y discusiones que tuvimos alrededor del tema, el apoyo incondicional con el que siempre conté, el pragmatismo con el que me hacía abordar cada tema, sino también por haber sembrado desde mi pregrado, cuando era mi profesor de terminología, la necesidad de hacer este trabajo en algún momento. De la misma forma que lo hiciste en el doctorado ahora lo hago yo en mi maestría. ¡Por fin pasamos de 32736 expresiones a 198! Mil gracias por todo.

A la familia Quiroz Holguín especialmente a Mónica por toda la preocupación y las atenciones que tuvieron para conmigo en estos dos años.

A Gustavo Zapata, que no sólo me proporcionó el diccionario que luego sería la base del corpus sino que siempre tuvo palabras de aliento y comprensión, así como grandes críticas y sugerencias para este trabajo, no me alcanzará la vida para agradecerle todo lo que hizo por mí en estos dos años.

A John Jairo Giraldo, cuyo consejo y asesorías guiaron mi trabajo de grado y se ven reflejados en mi marco teórico.

A Felipe Zuluaga y a Gustavo Lopera por las innumerables horas que estuvieron a mi lado ayudando a diseñar el corpus, espero poderles retribuir de igual forma cuando llegue el momento.

A Luz Stella Castañeda por haberme aceptado en la novena cohorte de la maestría y a José Luis Orduña por el apoyo que tuvo para con los estudiantes de la misma durante el último año en que estuvo a cargo de la dirección del programa.

A Bibiana Arboleda y a Paula Bedoya, por la colaboración que recibí tanto para vincularme a UM como profesor de tiempo completo, como para sostener la beca de cátedra de la U de A.

A los profesores de la maestría, pero muy especialmente a José Luis Orduña, a Francisco Zuluaga, a Marianne Dieck, a John Jairo Gómez y a Gabriel Quiroz quienes, desde sus cátedras, me proporcionaron las bases teóricas que hoy vengo a aplicar en este trabajo.

Por último a mis compañeros y amigos del grupo de investigación TNT, de la maestría y de la facultad de comunicaciones Alejandro, Pedro, Diego, Giovanni, Antonio,

Etiquetaje y descripción de unidades fraseológicas especializadas en un diccionario bilingüe de comercio internacional

Juan Manuel, Elkin, Claudia, Vanessa, Huber, Fabio e Isabel... Gracias por tantas horas de conversaciones académicas y no tan académicas durante este tiempo.

J. L. Rojas

Yaneth, este fue un sueño que iniciamos juntos hace más de dos años y que hoy termina convertido en una realidad. Gracias por tu amor incondicional, no hubiera podido iniciar esto sin ti.

A mi padre Jairo y a mi madre Rocío.

Todo lo que soy se los debo a ambos.

Un homenaje póstumo para ti mamá y uno en vida para ti papá.

A mis hermanos Jorge y Juan.

Espero hacerlos sentir tan orgullosos como me siento yo de ustedes.

A mis sobrinas Ana Cristina y Susana.

Porque las distancias no son tan largas cuando hay amor de por medio.

A mis sobrinos Camilo, Sebastian y Santiago.

*Lo único que les puedo ofrecer es esto, espero que sea un buen ejemplo para
ustedes.*

Abreviaturas

Adj:	adjetivo
Adv:	adverbio
Conj:	conjunción
D:	determinante
N:	nombre, sustantivo
PP:	participio pasado
PPi:	participio presente
Prep:	preposición
V:	verbo
DCI:	diccionario de comercio internacional
MG:	marca gramatical
POS:	part of speech (categoría gramatical)
ESP:	español
ENG:	inglés
IMF:	International Monetary Fund
RD:	Routledge Spanish Dictionary of Business, Commerce and Finance
UT:	unidad terminológica

Índice de contenidos

1. LAS UNIDADES FRASEOLÓGICAS EN DICCIONARIOS ESPECIALIZADOS	12
1.1. INTRODUCCIÓN	13
1.2. ANTECEDENTES DE LA TESIS	14
1.3. OBJETIVOS DE LA TESIS	16
1.3.1. Objetivo general	16
1.3.2. Objetivos específicos	16
2. MARCO TEÓRICO	17
2.1. INTRODUCCIÓN	19
2.2. FRASEOLOGÍA	19
2.2.1. Origen del término	20
2.2.2 Retrospectiva histórica de las clasificaciones de las UF en lengua española	22
2.2.2.1. Casares	23
2.2.2.2. Zuluaga	24
2.2.2.3. Carneado Moré y Tristán Pérez	25
2.2.2.4. Corpas	31
2.2.2.5. Ruiz Gurillo	34
2.2.2.6. García-Page	37
2.2.3. Algunas clasificaciones de las UF en lengua inglesa	44
2.3. TERMINOLOGÍA Y FRASEOLOGÍA	49
2.3.1. Manuales de terminología	50

2.3.1.1. Felber	50
2.3.1.2. Cabré	52
2.3.1.3. Arzt-Picht	53
2.3.1.4. Fedor de Diego	53
2.3.2. Algunos otros autores que tratan la fraseología dentro de la terminología	54
2.3.2.1. Lorente	54
2.3.2.2. L'Homme	58
2.3.2.3. Bevilacqua	59
2.3.2.4. Cardero	62
2.3.3. Organismos internacionales de estandarización terminológica	63
2.4. TERMINOLOGÍA Y LEXICOGRAFÍA	64
2.5. TÉRMINOS ÚTILES RELATIVOS A LA LEXICOGRAFÍA PARA ESTE TRABAJO	66
2.5.1. Fuentes	66
2.5.2. Macroestructura	67
2.5.3. Microestructura	68
2.6. RESUMEN DE CRITERIOS Y CONCEPTOS TEÓRICOS PARA ESTE TRABAJO	69
2.6.1. Criterios de selección de UF	70
2.6.2. Definición de UFE	71
3. METODOLOGÍA GENERAL DE TRABAJO Y CONSTITUCIÓN DEL CORPUS	72
3.1. INTRODUCCIÓN	73
3.2. DESCRIPCIÓN Y SELECCIÓN DEL CORPUS INGLÉS-ESPAÑOL / ESPAÑOL-INGLÉS	73

3.3. HERRAMIENTAS Y RECURSOS	78
3.4. METODOLOGÍA DE ANÁLISIS GENERAL	88
3.5. EXTRACCIÓN DE LAS UFE Y TRATAMIENTO DE LOS DATOS	89
3.6. ASPECTOS ESTADÍSTICOS	89
4. PRESENTACIÓN Y DESCRIPCIÓN DE DATOS ESTADÍSTICOS	91
4.1. Introducción	93
4.2. DESCRIPCIÓN Y ANÁLISIS ESTADÍSTICO EN INGLÉS	93
4.2.1. Unidades con o sin POS	94
4.2.2. Descripción de marcaje de entradas con POS	95
4.2.3. Descripción de marcas simples	98
4.2.4. Descripción de la marca <i>fr</i>	104
4.2.4.1. Conteo de tokens de las expresiones marcadas como <i>fr</i>	109
4.2.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como <i>fr</i>	110
4.2.5. Descripción de las unidades sin marca de POS (subentradas)	114
4.2.5.1. Conteo de tokens en las subentradas	115
4.2.5.2. Expresiones pluriverbales que contienen comas	116
4.2.5.3. Expresiones pluriverbales que contienen guiones	119
4.2.5.4. Expresiones pluriverbales que contienen comas y guiones	121
4.2.5.5. Expresiones pluriverbales que no contienen comas o guiones	121
4.3. DESCRIPCIÓN Y ANÁLISIS ESTADÍSTICO EN ESPAÑOL	123
4.3.1. Unidades con o sin POS	124

4.3.2. Descripción de marcaje de entradas con POS	125
4.3.3. Descripción de marcas simples	126
4.3.4. Descripción de la marca <i>phr</i>	132
4.3.4.1. Conteo de tokens de las expresiones marcadas como <i>phr</i>	132
4.3.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como <i>phr</i>	132
4.3.5. Descripción de las unidades sin marca de POS (subentradas)	135
4.3.5.1. Conteo de tokens en las subentradas	135
4.3.5.2. Expresiones pluriverbales que contienen comas	136
4.3.5.3. Expresiones pluriverbales que contienen guiones	139
4.3.5.4. Expresiones pluriverbales que contienen comas y guiones	142
4.3.5.5. Expresiones pluriverbales que no contienen comas o guiones	142
5. CONCLUSIONES: RESULTADOS Y LÍNEAS DE TRABAJO FUTURO	145
5.1. VALIDACIÓN DE CRITERIOS DE UFE	146
5.1.1. Elementos deverbales en inglés	147
5.1.2. Elementos deverbales en español	148
5.1.3. Fijación de las expresiones en inglés	151
5.1.4. Fijación de las expresiones en español	161
5.2. HALLAZGOS COMPLEMENTARIOS	164
5.2.1. Evaluación formal del diccionario	164
5.2.1.1. Análisis de la macroestructura	165
5.2.1.2. Rastreo de las fuentes citadas	166

5.2.1.3. Lematización	168
5.2.1.4. Polisemia	169
5.2.1.5. Análisis de la microestructura	170
5.2.2. Errores de etiquetaje	171
5.3. CONSIDERACIONES FINALES	175
6. REFERENCIAS	177
7. ANEXOS	186
7.1. INDICE DE ESQUEMAS	187
7.2. INDICE DE FIGURAS	188
7.3. INDICE DE GRÁFICAS	189
7.4. INDICE DE TABLAS	192

1. Las unidades fraseológicas en diccionarios especializados

1. LAS UNIDADES FRASEOLÓGICAS EN DICCIONARIOS ESPECIALIZADOS	12
1.1. INTRODUCCIÓN	13
1.2. ANTECEDENTES DE LA TESIS	14
1.3. OBJETIVOS DE LA TESIS	16
1.3.1. Objetivo general	16
1.3.2. Objetivos específicos	16

1.1. Introducción

Los procesos de globalización, integración y apertura económica, la comunicación interdisciplinar y el alto ritmo de progreso tecnológico, hacen que cada día se genere más conocimiento en las distintas disciplinas de las ciencias humanas y exactas, enriqueciendo el léxico de cada uno de sus discursos.

Un gran número de usuarios entre los que podríamos incluir traductores, expertos de un área específica y aquellos otros que cumplen funciones de revisores textuales — todos ellos transmisores de conocimiento— requieren de la terminología, la fraseología, la lexicografía, entre otros para llevar a cabo su labor, permitiendo obrar como mediadores dentro de los distintos actos comunicativos orales y escritos, para lo cual se requiere, que los distintos discursos científicos y técnicos sean homogéneos y armonizados tanto como sea posible, evitando así cualquier ambigüedad en su uso (Palomar, 2004).

A pesar de contar con diferentes recursos la doctora Mercè Lorente nos indica una problemática con respecto a la búsqueda de unidades fraseológicas en diccionarios: "Los usuarios de vocabularios especializados precisan información sobre combinaciones terminológicas recurrentes, [...] que no quedan resueltas con la consulta de la mayoría de los recursos", añade además que dicha problemática se debe, ya sea a la forma como se construyen las distintas herramientas, a la carencia de trabajos descriptivos referentes a las unidades fraseológicas especializadas, y agrega una tercera que será relevante a lo largo

del presente trabajo y es la carencia de parámetros que permitan diferenciar las unidades fraseológicas generales de las especializadas (Lorente, 2001, pp. 159-160).

Partiendo de lo anterior se puede afirmar que este proyecto pretende analizar cómo se encuentran discriminadas las distintas unidades fraseológicas especializadas —en adelante UFE— en un diccionario especializado bilingüe en inglés y español.

El estudio de la fraseología en diccionarios especializados ha sido poco explorado y por tanto se desconoce cómo están representadas las UFE en estos y si son consistentes con los diccionarios generales. Para este análisis se pretende extraer todas las unidades, tanto en inglés como en español, con algún tipo de marca referente a las UFE, que se encuentren en el diccionario.

1.2. Antecedentes de la tesis

Este trabajo de grado, para optar al título de Magíster, es el resultado de una consecución de trabajos bajo la supervisión del Profesor Dr. Gabriel A. Quiroz.

El proyecto surge al encontrar una serie de problemas de etiquetaje al interior de los distintos recursos lexicográficos consultados en labores de traducción, en especial en este momento, que tiene tanto auge la traducción de textos de este tipo. La consulta de un diccionario especializado casi siempre se deriva del desconocimiento no de una unidad

lexicográfica de un solo token, sino de aquellas unidades pluriverbales que son motivo de nuestro análisis.

Es entonces dónde autores como Casares (1950), Zuluaga (1980), Moré y Tristá (1985), Corpas (1996), Lorente (2001), Sosiński (2006), García-Page (2008), aparecen a nutrir el bagaje teórico de la tradición fraseológica española.

Se hace igualmente necesario resaltar los trabajos hechos desde la tradición fraseológica inglesa, señalar algunos de sus más importantes exponentes como lo son Firth (1957), Sinclair y Moon (1989, 1995), Crystal (1995), Cowie (2001) y Mel'čuk (1988, 2001a, 2001b), debido a la característica bilingüe del diccionario.

En una primera etapa se hizo una descripción estadística de los datos encontrados en el diccionario escogido para hacer el análisis, en la cual se tuvo en cuenta variables tales como: número de entradas, subentradas, categorías gramaticales, número de tokens, marcas de uso coloquial, marcas de uso geográfico y las marcas que identifican las unidades fraseológicas en el diccionario.

1.3. Objetivos de la tesis

1.3.1. Objetivo general

- Analizar cómo se encuentran clasificadas y etiquetadas las UFE en un diccionario especializado bilingüe en inglés y español.

1.3.2. Objetivos específicos

- Revisar la macro y la microestructura del diccionario analizado.
- Extraer los patrones morfosintácticos más frecuentes de las expresiones pluriverbales encontradas en el diccionario.
- Identificar las etiquetas utilizadas para el reconocimiento de las UFE en un diccionario especializado bilingüe en inglés y español.
- Analizar cuantitativa y cualitativamente las UFE extraídas del diccionario.

2. Marco Teórico

2. MARCO TEÓRICO	17
2.1. INTRODUCCIÓN	19
2.2. FRASEOLOGÍA	19
2.2.1. Origen del término	20
2.2.2 Retrospectiva histórica de las clasificaciones de las UF en lengua española	22
2.2.2.1. Casares	23
2.2.2.2. Zuluaga	24
2.2.2.3. Carneado Moré y Tristán Pérez	25
2.2.2.4. Corpas	31
2.2.2.5. Ruiz Gurillo	34
2.2.2.6. García-Page	37
2.2.3. Algunas clasificaciones de las UF en lengua inglesa	44
2.3. TERMINOLOGÍA Y FRASEOLOGÍA	49
2.3.1. Manuales de terminología	50
2.3.1.1. Felber	50
2.3.1.2. Cabré	52
2.3.1.3. Arzt-Picht	53
2.3.1.4. Fedor de Diego	53
2.3.2. Algunos otros autores que tratan la fraseología dentro de la terminología	54
2.3.2.1. Lorente	54

2.3.2.2. L'Homme	58
2.3.2.3. Bevilacqua	59
2.3.2.4. Cardero	62
2.3.3. Organismos internacionales de estandarización terminológica	63
2.4. TERMINOLOGÍA Y LEXICOGRAFÍA	64
2.5. TÉRMINOS ÚTILES RELATIVOS A LA LEXICOGRAFÍA PARA ESTE TRABAJO	66
2.5.1. Fuentes	66
2.5.2. Macroestructura	67
2.5.3. Microestructura	68
2.6. RESUMEN DE CRITERIOS Y CONCEPTOS TEÓRICOS PARA ESTE TRABAJO	69
2.6.1. Criterios de selección de UF	70
2.6.2. Definición de UFE	71

2.1. Introducción

El contexto en el cual situaremos el presente trabajo de grado nos exige apropiarnos de varios conceptos, por tanto —en este capítulo— haremos un recorrido teórico por los campos de la lexicografía, la sintaxis y la fraseología. Este marco teórico está dividido así: definiremos el concepto de fraseología junto con la descripción de su origen, posterior a esto haremos un recuento diacrónico de las clasificaciones de las UF en lengua española e inglesa por algunos autores, luego procederemos a abordar la fraseología desde la terminología, evaluando algunos manuales de terminología y artículos que hacen referencia a ambos, por último haremos mención de los criterios que emplearemos para seleccionar y evaluar las UFE junto con la definición que emplearemos para este tipo de unidades en este trabajo.

2.2. Fraseología

La fraseología, como disciplina lingüística, no se alcanza a definir en términos de independencia o no de otra disciplina; antes bien, se ha relacionado, a menudo y durante años, con la paremiología, la lingüística aplicada, la lexicografía, la lexicología y la terminología (García-Page, 2008, p. 7). Una idea un poco más arraigada dentro de la tradición fraseológica española es definir la fraseología a partir de su objeto de estudio: la Unidad Fraseológica, pero esta propuesta conlleva un problema, pues, si se revisan aquellos autores considerados clásicos dentro de este aspecto, podremos observar que

cada uno de ellos tiene aportes de gran importancia para la disciplina, pero, a su vez, generan distancia respecto a otros que hayan tratado el mismo tema, por lo que al final cada autor plantea su propia tipología de las UF.

Ahora bien, tratar la fraseología inglesa implicaría hablar de una gran variedad de conceptos que, en vez de ser complementarios o excluyentes, tienden a la sinonimia, a tal punto que, las expresiones *lexical bundle*, *phraseme*, *collocation*, *phraseological unit*, *locution* o *modism*, con frecuencia significan cosas similares o idénticas, dependiendo de los autores que las traten. Dedicaremos más adelante un espacio prudente para la explicación de este fenómeno terminológico.

2.2.1. Origen del término

Ya el origen de la denominación de fraseología comienza a mostrar ciertas irregularidades en su desarrollo histórico. Mientras algunos teóricos afirman que el origen del término se debe a los trabajos realizados por los lingüistas estructuralistas de occidente, otros indican que esta disciplina —o subdisciplina— solo ve la luz teórica a mediados de los años 40 en los estudios lingüísticos soviéticos.

Carneado Moré nos indica que la fraseología se ve influenciada por la doctrina teórica de Bally (1905) sobre “combinaciones exactas de palabras”; sin embargo, no deja de reconocer que Vinogradov fue quien sentó los “conceptos fundamentales de la

fraseología” (1985, p. 7). Años más tarde, García-Page (2008, p. 39) ratifica lo planteado por Carneado Moré:

“La mayoría de los estudiosos está de acuerdo en proclamar a Ch. Bally como el fundador de la Fraseología en virtud de los breves pero enjundiosos apuntes sobre las combinaciones fijas expuestos sobre todo en su *Traité de stylistique française* (1909); él es también el primer autor en acuñar el término fraseología, creemos que con justicia puede atribuírsele tal distinción al ser la primera vez que lo que hoy conocemos por unidades fraseológicas recibe un estudio mínimamente científico y sistemático”.

Hecha la observación anterior, líneas más adelante, este autor nos indica también que dichos conceptos fueron tenidos en cuenta antes por el mismo Ferdinand de Saussure en sus cursos de lingüística.

Zuluaga (1980, pp. 31-37), aunque reconoce los aportes de Saussure y de Bally, tiene referencias de otros tres autores anteriores a los antes mencionados, quienes hicieron mención de algunas denominaciones —en sus respectivas lenguas— que serían las precursoras de las UF:

<i>Autor y lengua</i>	<i>Expresión</i>
Hermann Paul (alemán)	<i>isolierte Formel</i>
	<i>stehende Formel</i>
	<i>stehende Verbindung</i>
	<i>feste Verbindung</i>
	<i>Erstarrung einer Flexionsform</i>
	<i>freier Akk</i>
	<i>gebundener Akk</i>
	<i>eines Tages</i>
	<i>meines Erachtens</i>

	<i>Formeln</i>
	<i>schöpferische Gruppen</i>
Georg von der Gabelentz (alemán)	<i>stehende Redensarten</i>
	<i>gebundene Rede</i>
Michel Bréal (francés)	<i>Groupes articulés</i>
	<i>Locution</i>
	<i>Formule</i>

Tabla 1: Expresiones protogénicas de las UF anteriores a Vinogradov y Bally

2.2.2. Retrospectiva histórica de las clasificaciones de las UF en lengua española.

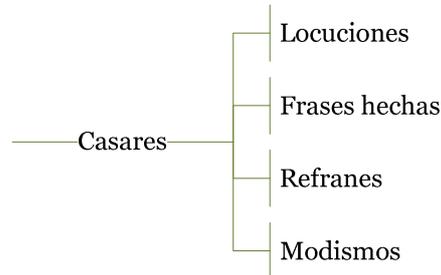
Este apartado y los siguientes son tratados con mucha diligencia por todos los estudiosos e investigadores de la fraseología. Intentaremos ofrecer alguna información adicional a lo que se puede encontrar en los trabajos más recientes, teniendo en cuenta que la fraseología tiene un origen que, en ocasiones, es considerado compartido o debatido, como lo veremos luego y que, a pesar de esto, su desarrollo ha sido independiente en cada una de las lenguas que pretendemos tratar en este trabajo de grado.

Autores como Casares (1950), Zuluaga (1980), Carneado Moré y Tristán (1985), Corpas (1996), Ruiz Gurillo (2001), García-Page (2008) son algunos de los autores más citados en lo que a fraseología, tanto clásica como contemporánea, se refiere y no pretendemos profundizar en cada una de sus concepciones, pero sí dar un pequeño esbozo de algunos de los conceptos y términos que se han empleado por cada uno de estos teóricos.

2.2.2.1. Casares

Casares (1950) es el precursor de los trabajos de fraseología en lengua española. Aunque sus conceptos hayan variado y sufrido modificaciones por parte de otros autores, no deja de ser obvio que todas estas variaciones y enfoques pragmáticos y lexicográficos tienen origen en él como lo explica Martínez (2002, pp. 139-143).

Casares nos provee de la primera clasificación de las UF en lengua española en 4 grupos, así: locuciones, frases hechas, refranes y modismos. En su texto especifica algunas características propias de cada una de estas categorías.



Esquema 1: Tipología de Casares

Este autor nos presenta, en un cuadro, cada una de las distintas UF que él clasifica dentro de su trabajo. Aun así, esta tipología carece de algunas precisiones que él mismo hace luego en su texto. A continuación, presentamos dicha ampliación de categorías de las locuciones en un esquema:

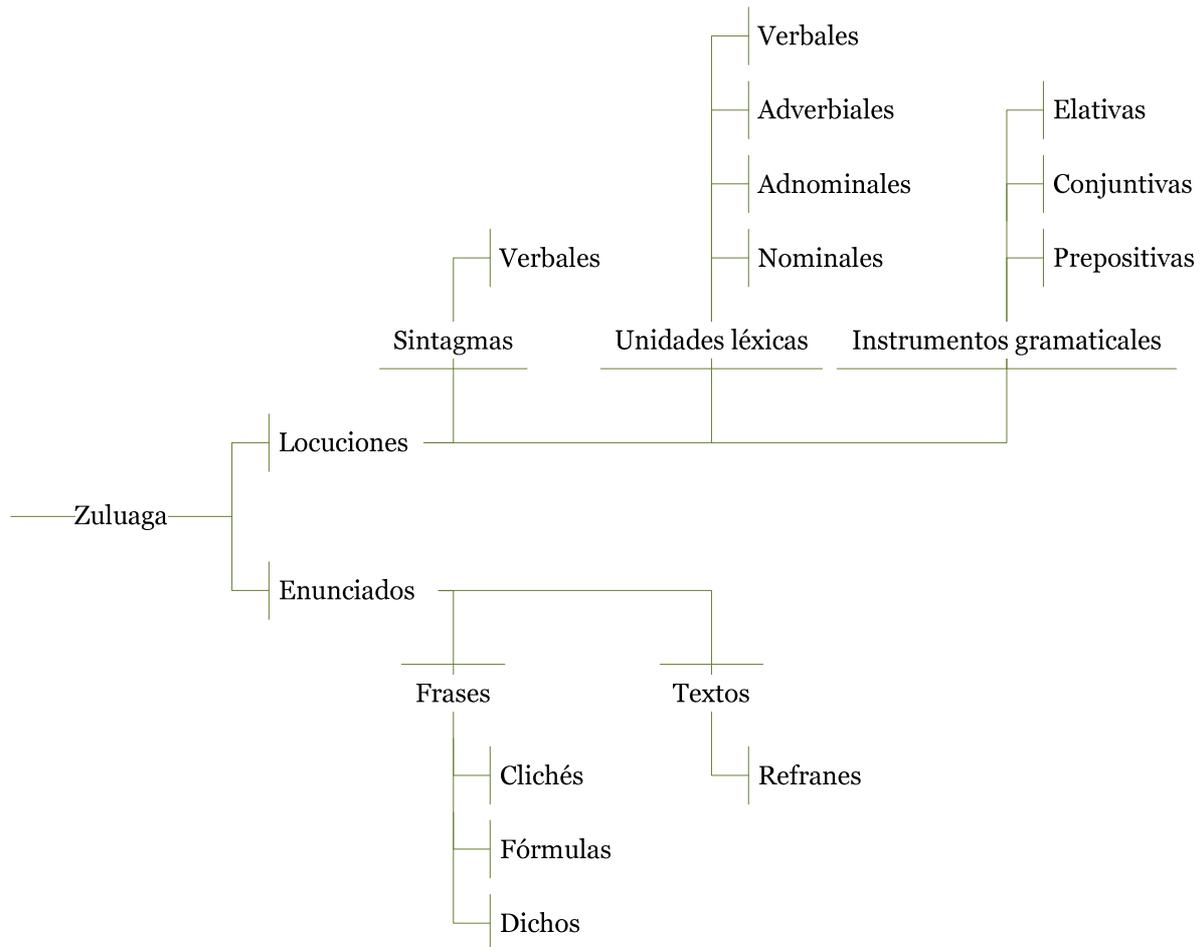


Esquema 2: Clasificación de las locuciones para Casares

2.2.2.2. Zuluaga

Su tesis doctoral —publicada luego bajo el título de “Introducción a las expresiones fijas” (1980)— es considerada uno de los primeros estudios sistemáticos de las UF en lengua española. Si bien retoma parte de la terminología usada por Casares (1950) en su estudio, nutre la conceptualización teórica de la fraseología agregando a ésta, correcciones, sugerencias y comentarios a las taxonomías antes expuestas. Finalmente,

Corpas (1996, p. 42) logra sintetizar la clasificación de Zuluaga, como observamos en el siguiente cuadro:



Esquema 3: Clasificación de las UF para Zuluaga

2.2.2.3. Carneado Moré y Tristán Pérez

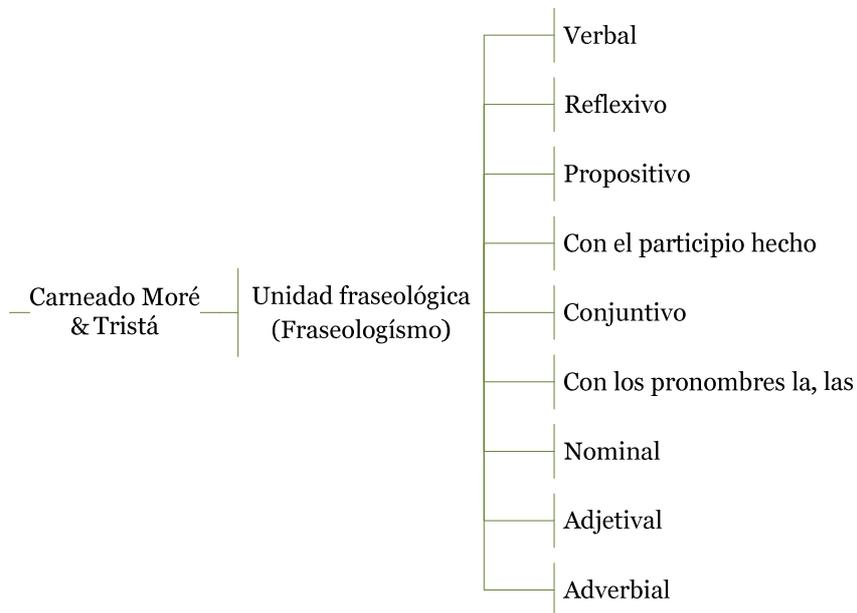
Este par de autoras cubanas, pertenecientes al Instituto de Literatura y Lingüística de La Habana, compilan parte de sus investigaciones en un libro titulado *Estudios de*

fraseología (1985). En él, las autoras plantean una clasificación alterna, tomando en buena parte los conceptos de los teóricos y lingüistas soviéticos, en especial, los trabajos de Vinogradov (1947). Corpas no escatima en elogios para describir el trabajo y la influencia de estas autoras, pues afirma que “quizás [...] sean los miembros más representativos de la escuela cubana de la fraseología” (1996, p. 47).

Estas autoras hacen una clasificación “sintáctico estructural” (Carneado Moré & Tristá, 1985), de las UF en la primera parte del libro. Corpas sintetiza esa clasificación, a pesar de esto, intentaremos dar un ejemplo de cada una de sus categorías.

A diferencia de Corpas, consideramos que la tipología ofrecida por Carneado Moré (1985, pp. 15-22) no habla de combinaciones o expresiones fraseológicas, ya que son solamente enunciadas en la voz de otros autores como Shanski: “Los proverbios, refranes clichés y otros, denominados unidades comunicativas o expresiones fraseológicas (Shanski, 1963) pueden considerarse una especie de citas, es decir entran en el uso lingüístico como unidades del folclore, como fragmentos de obras literarias, etc.” (Carneado Moré & Tristá, 1985, p. 9). Esto se confirma luego, pues en ningún momento se encuentran listadas en el apartado número II titulado “Tipos léxico-gramaticales de unidades fraseológicas”.

A continuación, sintetizaremos en una tabla la clasificación de las UF que hacen las mencionadas autoras:



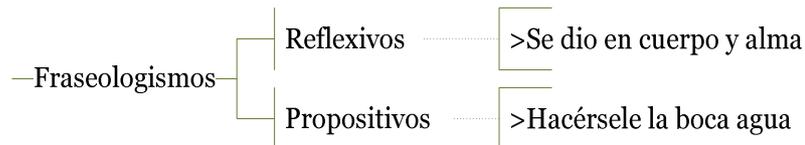
Esquema 4: Taxonomía de fraseologismos para Carneado Moré y Tristá

Algunos de estos fraseologismos tienen una subclasificación estructural. Procedemos a dar cuenta de esta con sus respectivos ejemplos. Para la siguiente figura debe entenderse V_i como verbo intransitivo y V^t como verbo transitivo.



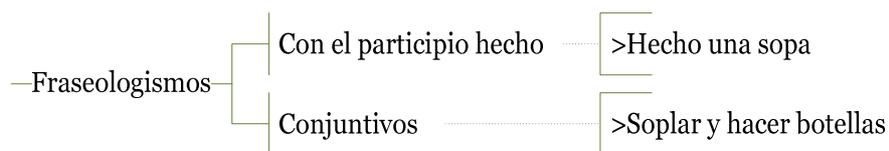
Esquema 5: Clasificación de los fraseologismos verbales según Carneado Moré y Tristá

Los fraseologismos reflexivos y propositivos suelen parecerse, aunque los últimos conforman estructuras con sujeto y predicado y los primeros solo tienen un rasgo.



Esquema 6: Clasificación de los fraseologismos reflexivos y propositivos según Carneado Moré y Tristá

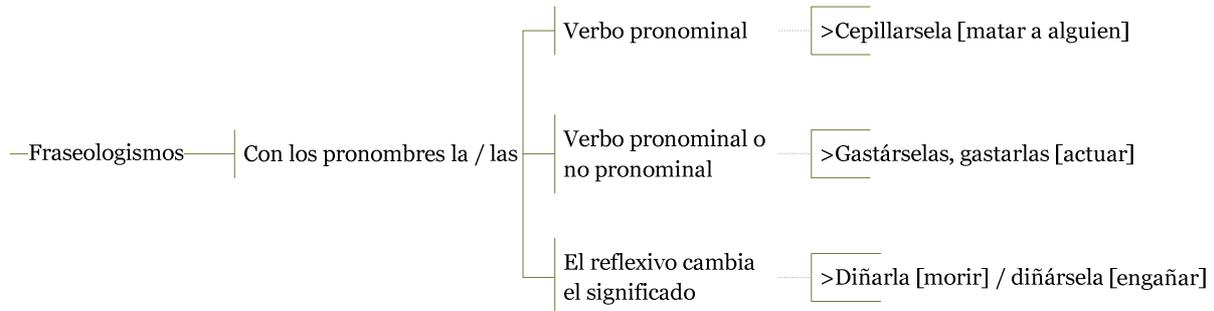
Aquellas unidades con el participio hecho, recogen lo dicho y estudiado por Casares (1950) respecto a estas unidades. Hacen especial hincapié en lo original, no solo de su estructura, sino también de su significado. Al igual que los fraseologismos conjuntivos ambos pueden llegar a cumplir una función nominal.



Esquema 7: Clasificación de los fraseologismos con el participio hecho y conjuntivos según Carneado Moré y Tristá

En lo que concierne al carácter fraseológico de las expresiones que contienen los pronombres “la” y “las”, Carneado y Tristá dicen que estas unidades no poseen relación de

número o de sentido entre el pronombre y el referente. Se exponen tres casos que resumimos en el siguiente cuadro (1985, p. 18):



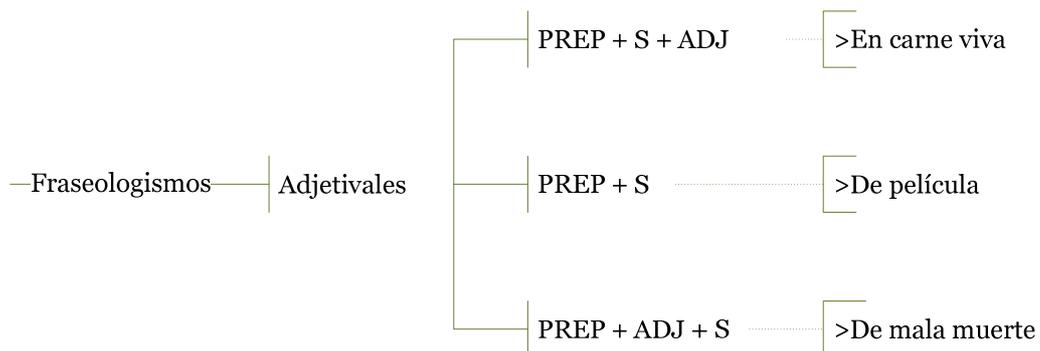
Esquema 8: Clasificación de los fraseologismos con los pronombres la/las según Carneado Moré y Tristá

Es notable que, a lo largo de la primera parte del libro de Carneado y Tristá, el enfoque para el estudio y el análisis de las unidades fraseológicas sea el criterio morfosintáctico de patrones –llamado modelo por ellas– dentro de cada una de las distintas categorías, como lo exponen con los fraseologismos nominales. Afirman, igualmente, que existen dos patrones con mayor recurrencia (1985, p. 19).



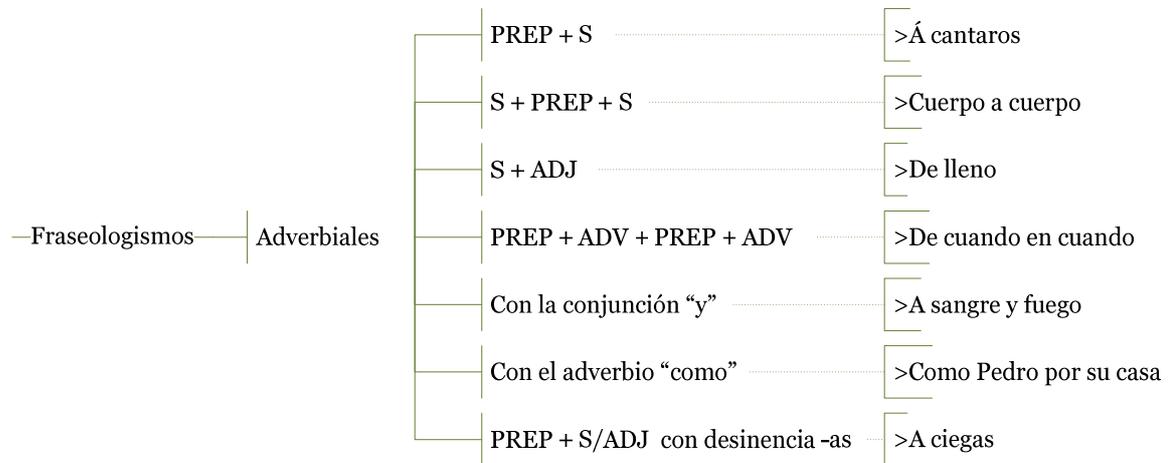
Esquema 9: Clasificación de los fraseologismos nominales según Carneado Moré y Tristá

La siguiente categoría está compuesta por los fraseologismos adjetivales, de los cuales, las autoras afirman que “la mayoría transmiten características cualitativas”, tal como se observa en la siguiente figura (Carneado Moré & Tristá, 1985, p. 20).



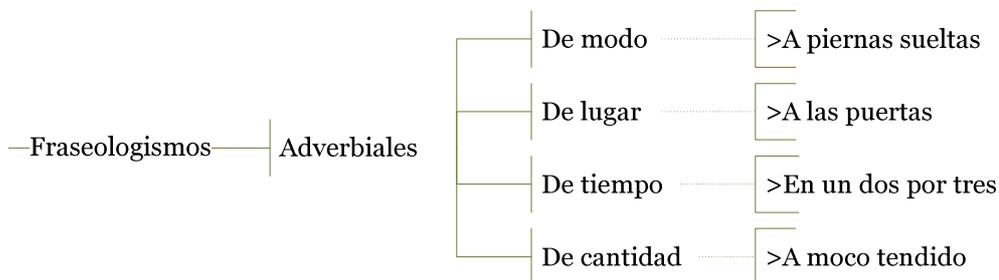
Esquema 10: Clasificación de los fraseologismos adjetivales según Carneado Moré y Tristá

Probablemente, el grupo que presentaremos a continuación sea el que las autoras tratan con mayor nivel de detalle, pues las unidades se organizan no solo por su composición morfosintáctica, sino también por su función. Tendríamos, entonces, dos cuadros que ilustran este grupo, así:



Esquema 11: Clasificación de los fraseologismos adverbiales según Carneado Moré y Tristá

De acuerdo a su referencia circunstancial estarán distribuidas así:



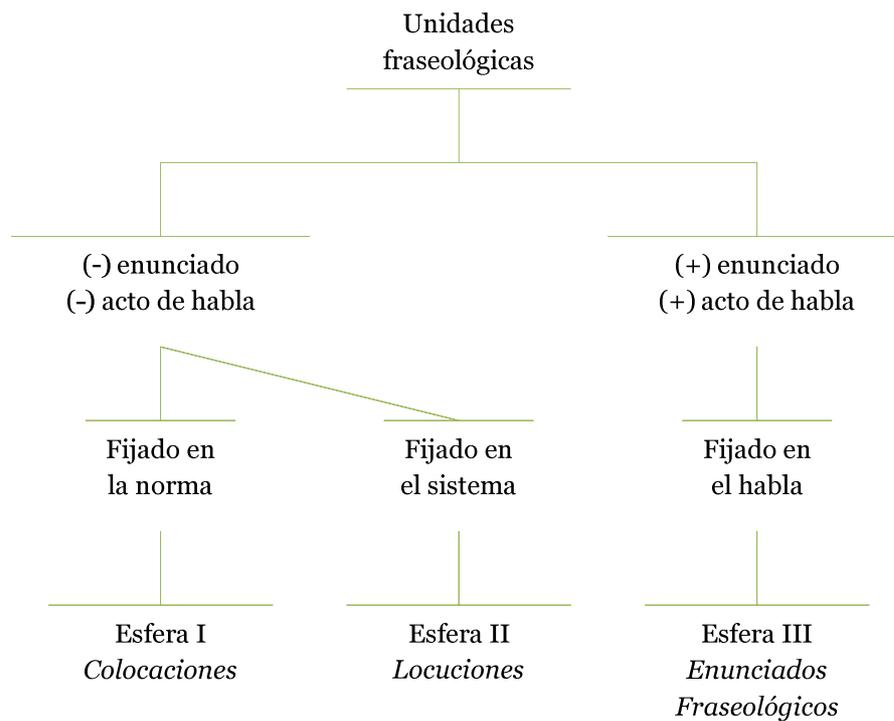
Esquema 12: Clasificación de los fraseologismos adverbiales de acuerdo a su referencia circunstancial según Carneado Moré y Tristá

2.2.2.3. Corpas

Corpas, en su *Manual de fraseología española* de 1996, aparte de ofrecernos un recuento detallado de cada uno de los autores anteriores, también nos ofrece una nueva taxonomía de las UF por medio de la inclusión del criterio de fijación al separar las

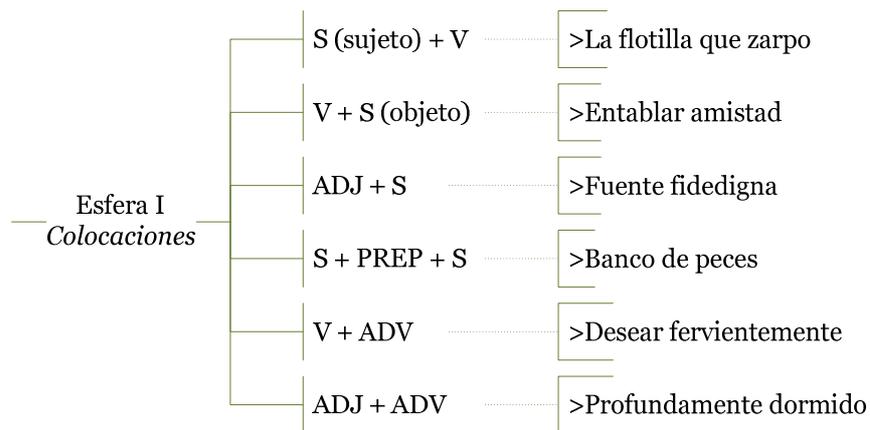
expresiones en dos grupos, según constituyen o no un acto de habla y, además, establece tres esferas de acuerdo a sus características de fijación (Corpas, 1996, p. 50).

Corpas ofrece la clasificación que aparece en el esquema 12 en el cual la autora presenta las tres esferas en las que analizan las unidades fraseológicas. Por otra parte, y como se puede observar en los esquemas 14, 15 y 16, se han ampliado cada una de las tres esferas incluyendo no sólo sus categorías sino también un ejemplo por cada una de ellas.



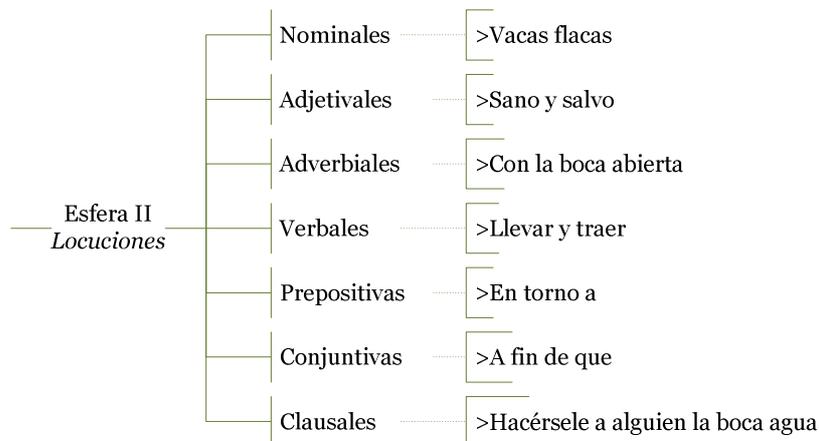
Esquema 13: Clasificación de las UF según Corpas

Los ejemplos que se muestran a continuación son tomados literalmente de los contenidos; pero, al analizarlos nos causa duda el hecho de observar cómo algunos patrones no concuerdan con los ejemplos planteados:



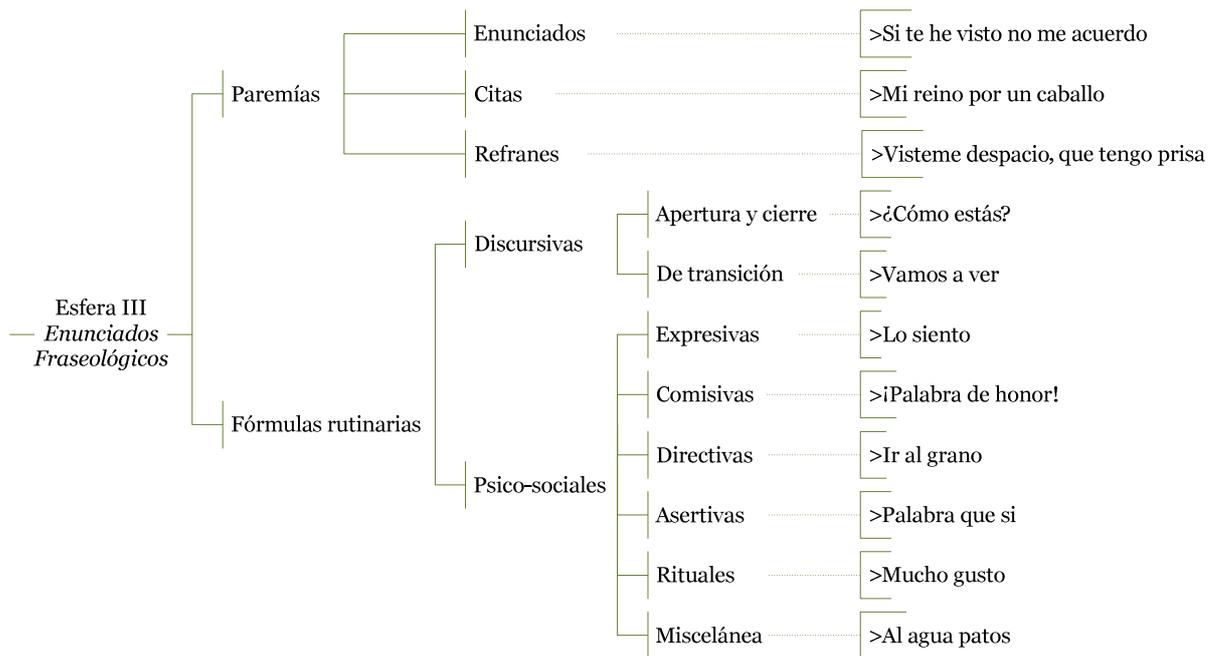
Esquema 14: Taxonomía de las colocaciones según Corpas

Las locuciones están clasificadas en 7 distintas categorías según la función que éstas cumplen, las podemos ver distribuidas así:



Esquema 15: Taxonomía de las locuciones según Corpas

Por último, los enunciados fraseológicos son categorizados en dos grandes grupos: las pemiás y las fórmulas rutinarias, a continuación se agregan a cada una de estas categorías, un ejemplo que permita una mejor comprensión de la clasificación.



Esquema 16: Taxonomía de los enunciados fraseológicos según Corpas

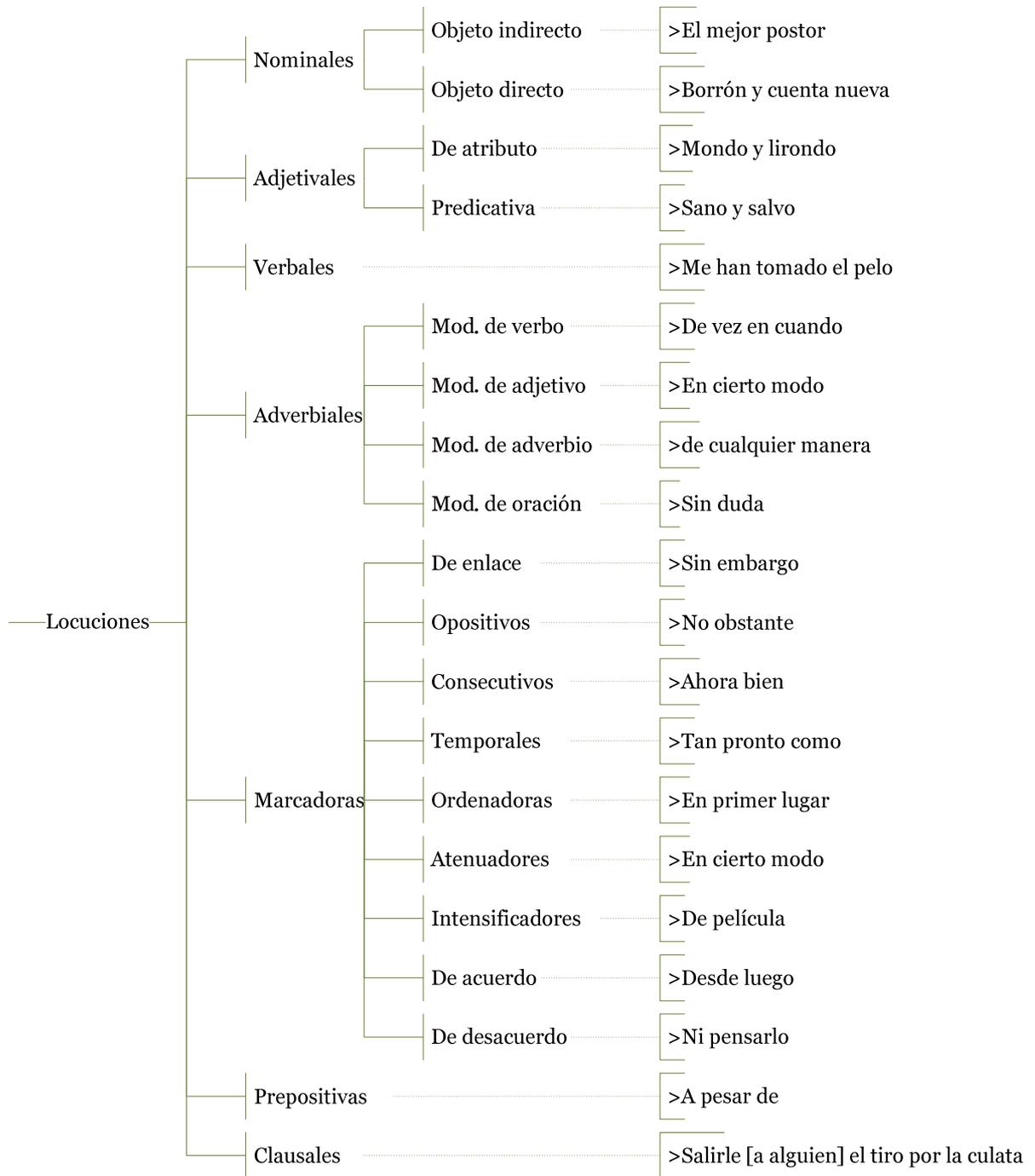
2.2.2.5. Ruiz Gurillo

Esta autora en su libro *Las locuciones del español actual* (2001) trata los llamados por otros “modismos, frases hechas, expresiones fijas etc.” (Ruiz, 2001, p. 11) En el apartado donde define el término locución, es notable el hecho de no haber una estandarización como tal del término y tan solo hace mención de las UF en la página 15,

en un pequeño fragmento luego de mencionar los trabajos de Casares (1950), Zuluaga (1980), Corpas (1996) y Penadés (2006).

Posterior a la definición y delimitación del objeto de estudio en su libro, la autora establece los rasgos generales de las locuciones a través de tres características vinculantes: la fijación, la idiomatidad y las relaciones entre ellas (fijación e idiomatidad).

Propone además 6 niveles de análisis: formal, morfológico, sintáctico, lexicográfico, pragmático e interdisciplinar, de los cuales, solo tendremos en cuenta para este trabajo el aspecto sintáctico, así como la taxonomía sugerida por la autora. Aunque en el análisis formal algunos de los sintagmas son presentados por medio de ejemplos, se debe reconocer que estos no ofrecen, en todas las ocasiones, los patrones morfosintácticos de las expresiones; estrategia que cobra mucho valor para aquellos que, en algún momento, pretendan hacer extracción o clasificación de los mismos. A continuación, ofrecemos la taxonomía de locuciones presentada por Ruiz Gurillo (2001, pp. 54-59):



Esquema 17: Taxonomía de las locuciones según Ruiz Gurillo

La autora nos presenta un cuadro donde sintetiza el trabajo de tres autores respecto a la concepción de las locuciones y si alguna de estas caracterizaciones se encuentra listada como tal dentro de sus taxonomías. Pretendemos no sólo mostrar el cuadro original, sino también complementar el mismo con la taxonomía propuesta por García-Page (2008) –y la misma Ruiz Gurillo (2001)–. Dado que Carneado Moré y Tristán Pérez (1985) no trabajan el término de locución no será incluida su taxonomía en el siguiente cuadro. Finalmente, si alguna de las clasificaciones presentadas por Ruiz Gurillo (2001) es tomada por otros autores como una subcategoría de alguna de las categorías enunciadas, no será tenida en cuenta en el cuadro.

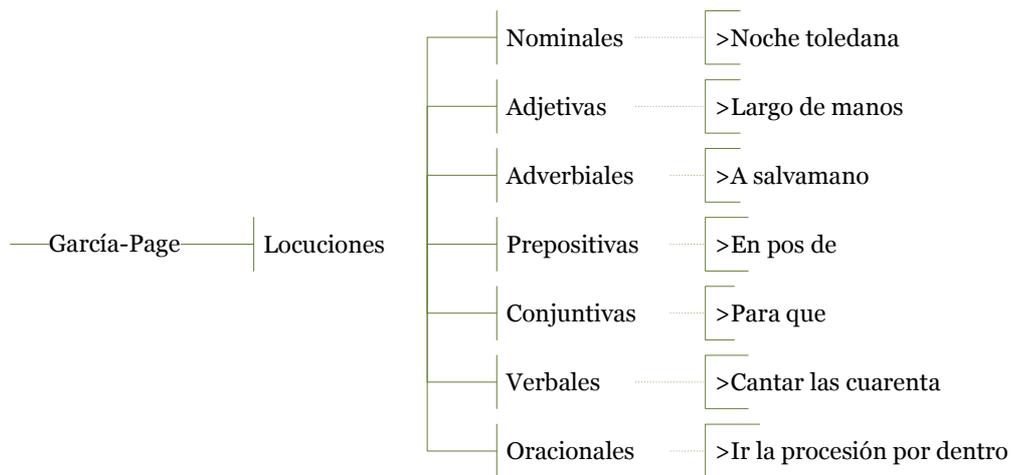
	<i>Casares</i>	<i>Zuluaga</i>	<i>Corpas</i>	<i>Ruiz Gurillo</i>	<i>García-Page</i>
<i>Nominales</i>	Sí	Sí	Sí	Sí	Sí
<i>Adjetivales</i>	Sí	Sí	Sí	Sí	Sí
<i>Verbales</i>	Sí	Sí	Sí	Sí	Sí
<i>Adverbiales</i>	Sí	Sí	Sí	Sí	Sí
<i>Prepositivas</i>	Sí	Sí	Sí	Sí	Sí
<i>Conjuntivas</i>	Sí	Sí	Sí	No	Sí
<i>Participiales</i>	Sí	No	No	No	No
<i>Pronominales</i>	Sí	No	No	No	No
<i>Exclamativas</i>	Sí	No	No	No	No
<i>Clausales</i>	No	No	Sí	Sí	No
<i>Elativas</i>	No	Sí	No	No	No
<i>Marcadoras</i>	No	No	No	Sí	No
<i>Oracionales</i>	No	No	No	No	Sí

Tabla 2: Actualización de la clasificación de las locuciones por autores de Ruiz (2001, p. 44)

2.2.2.6. García-Page

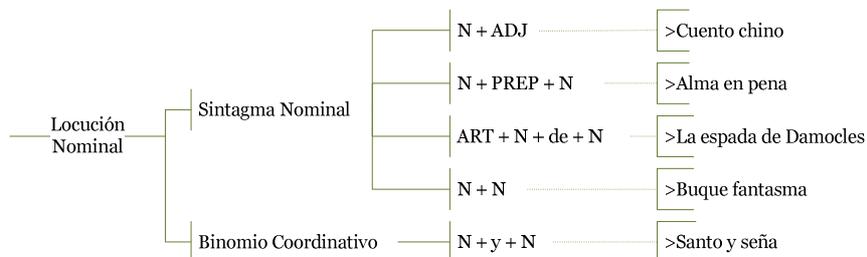
García-Page en su libro *Introducción a la fraseología española* (2008) demuestra ser, a diferencia de Corpas (1996), un teórico del enfoque estrecho de la fraseología, pues

descarta, de entrada, las colocaciones como posibles unidades de análisis dentro de las UF. En un primer vistazo, junto con los ejemplos ofrecidos, se podría ver la clasificación taxonómica de este autor resumida así:



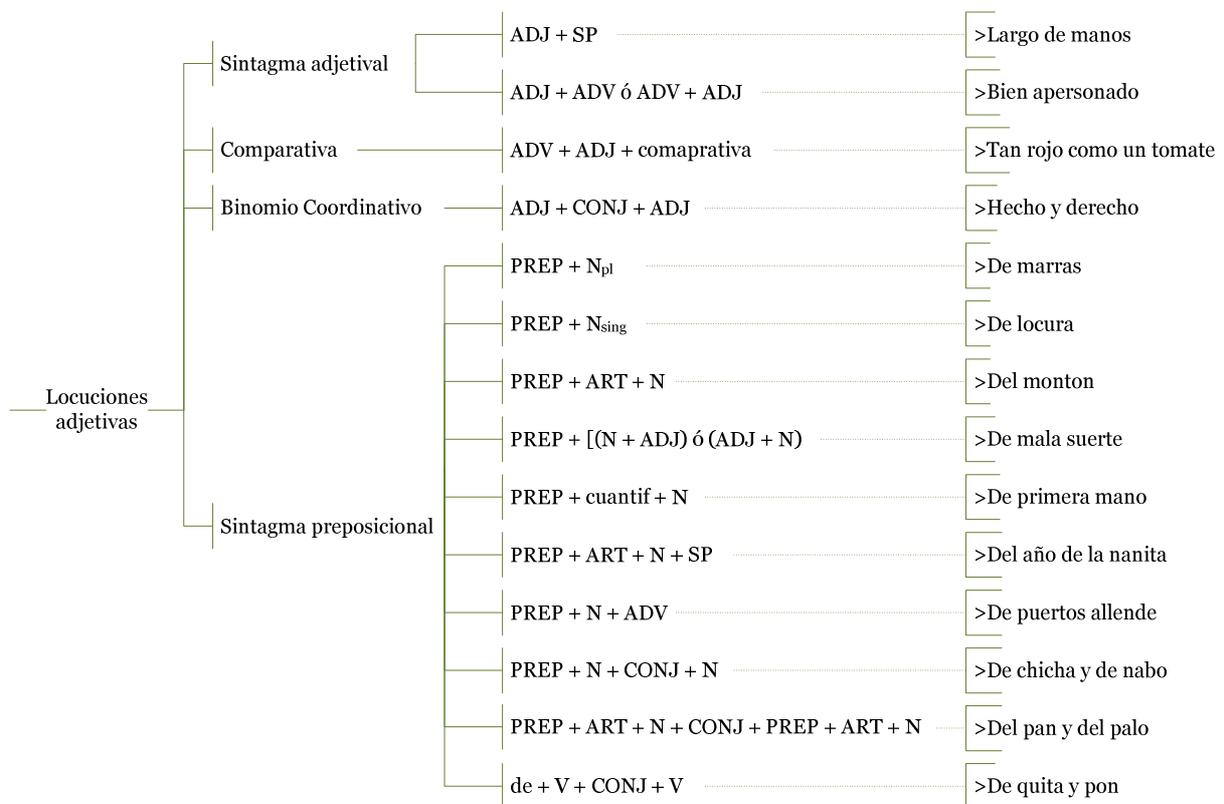
Esquema 18: Taxonomía de las locuciones según García-Page

Sin embargo, la taxonomía ofrecida por García-Page (2008) no se queda allí, sino que ofrece algunos patrones morfosintácticos típicos de cada una de las distintas locuciones; así, por ejemplo, encontramos cuatro grupos distintos dentro de las locuciones nominales.



Esquema 19: Clasificación de las locuciones nominales según García-Page

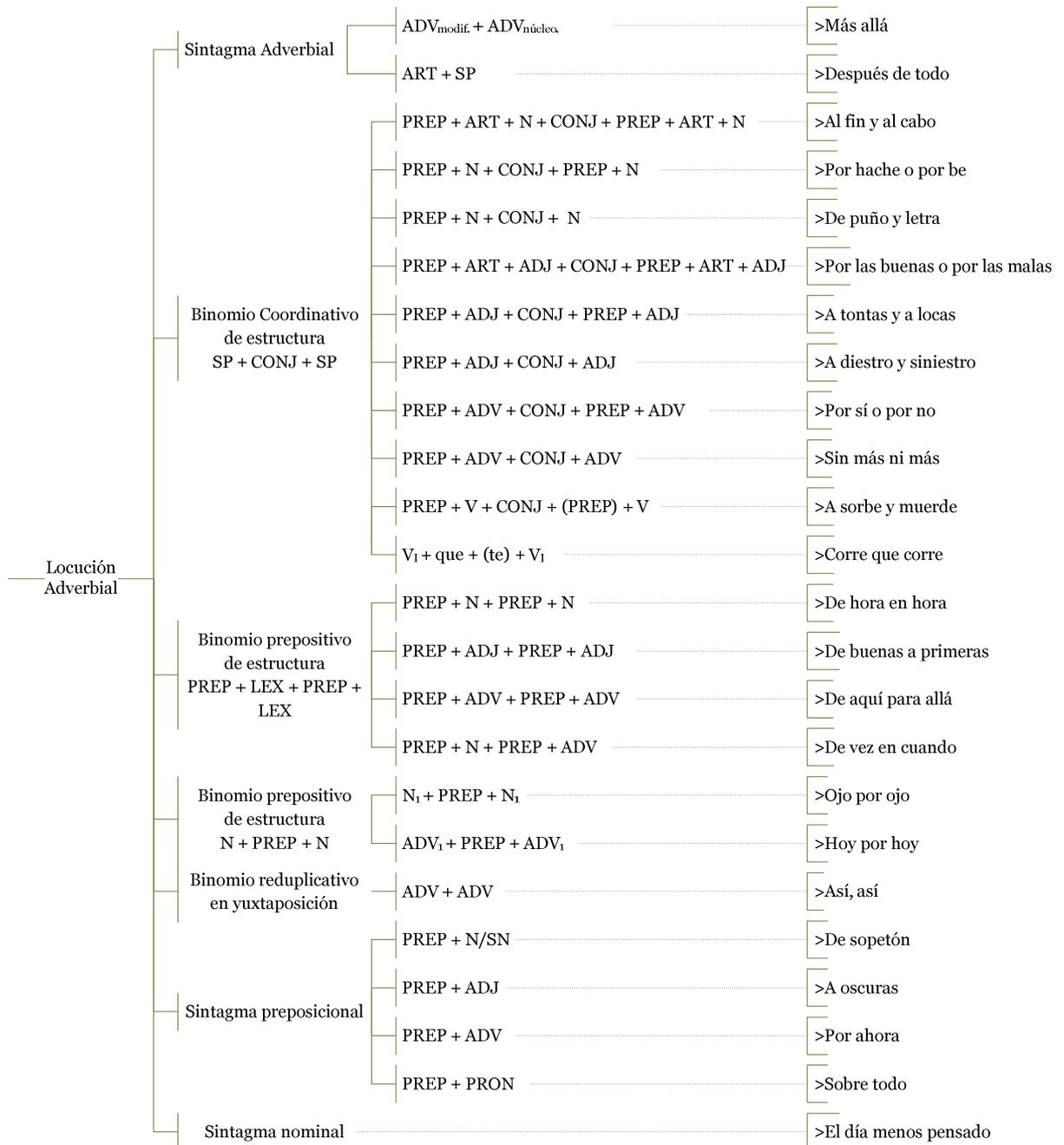
Respecto a las estructuras de las locuciones adjetivas provistas por García-Page podemos observar que se encuentran conformadas por distintos tipos de sintagmas con función adjetival:



Esquema 20: Clasificación de las locuciones adjetivas según García-Page

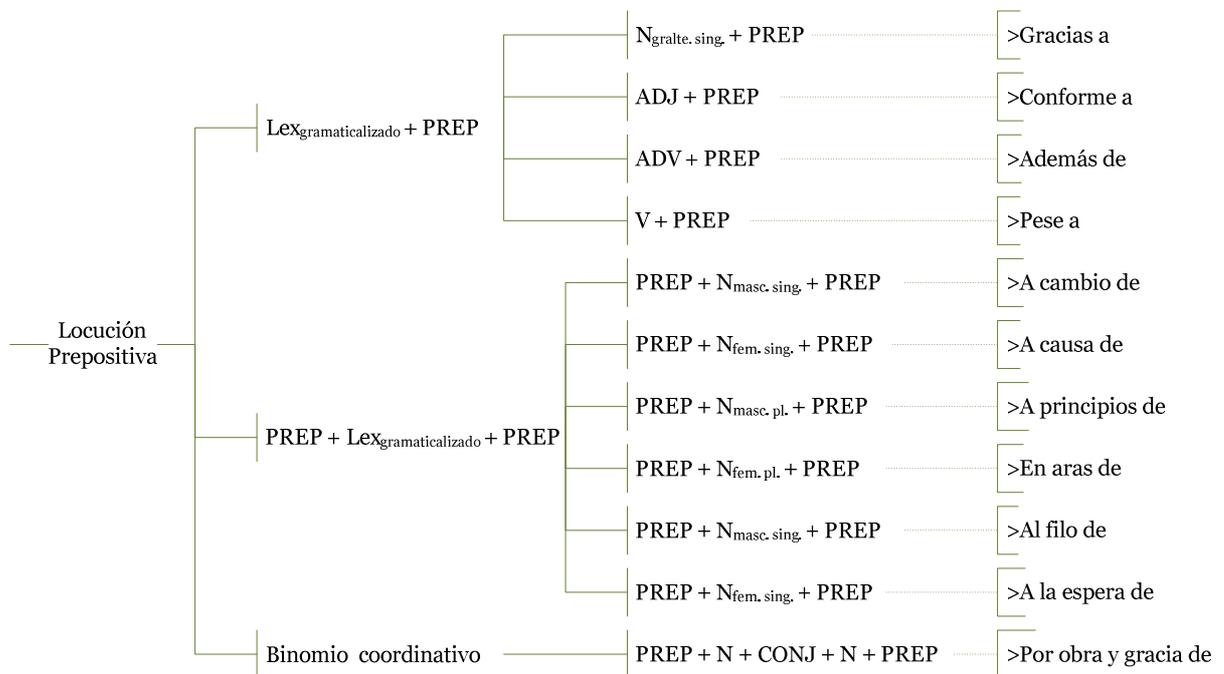
Las locuciones adverbiales tienden a tener una mayor cantidad de estructuras binomiales que el resto de las clasificaciones ofrecidas por García-Page. Es también posible que algunas locuciones con estructura adverbial tengan un giro funcional y

terminen sirviendo con otro propósito. A continuación, presentamos algunos ejemplos de estas estructuras:

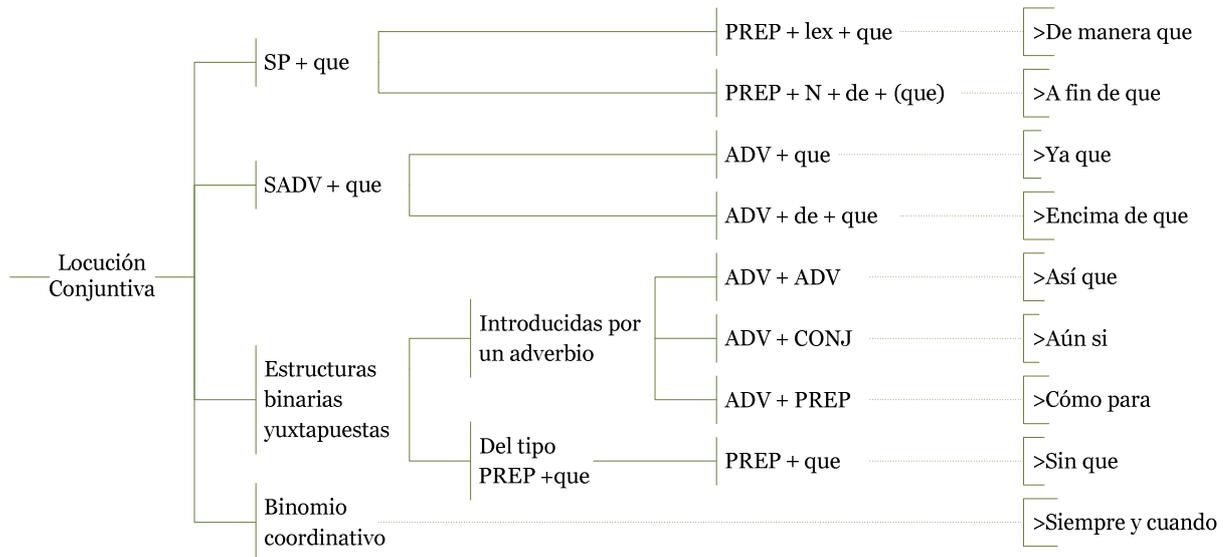


Esquema 21: Clasificación de las locuciones adverbiales según García-Page

Las locuciones prepositivas son consideradas por García-Page, junto con las locuciones conjuntivas, parte de la categoría conexiva y son las que presentan mayor fijación dentro de su taxonomía. En el capítulo 3 de su libro, el autor expone algunas de las estructuras básicas de este tipo de locuciones. Presentamos ambos cuadros a continuación:

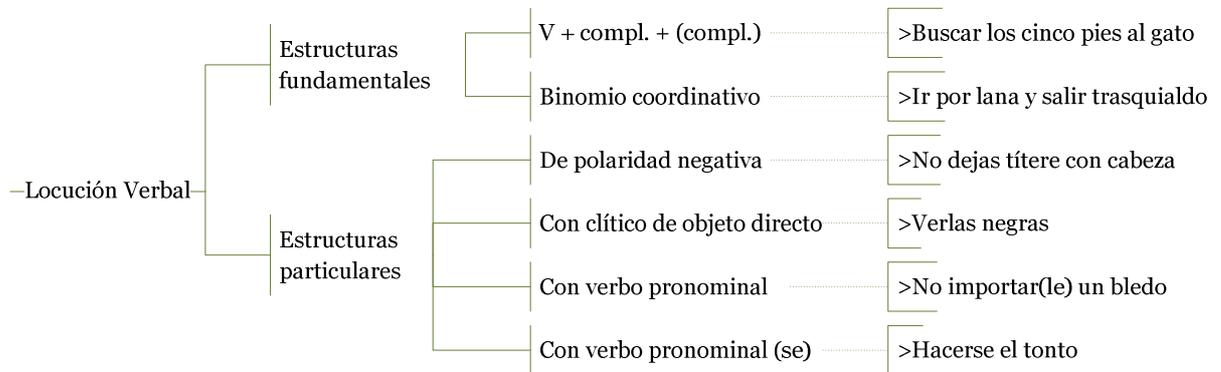


Esquema 22: Clasificación de las locuciones prepositivas según García-Page



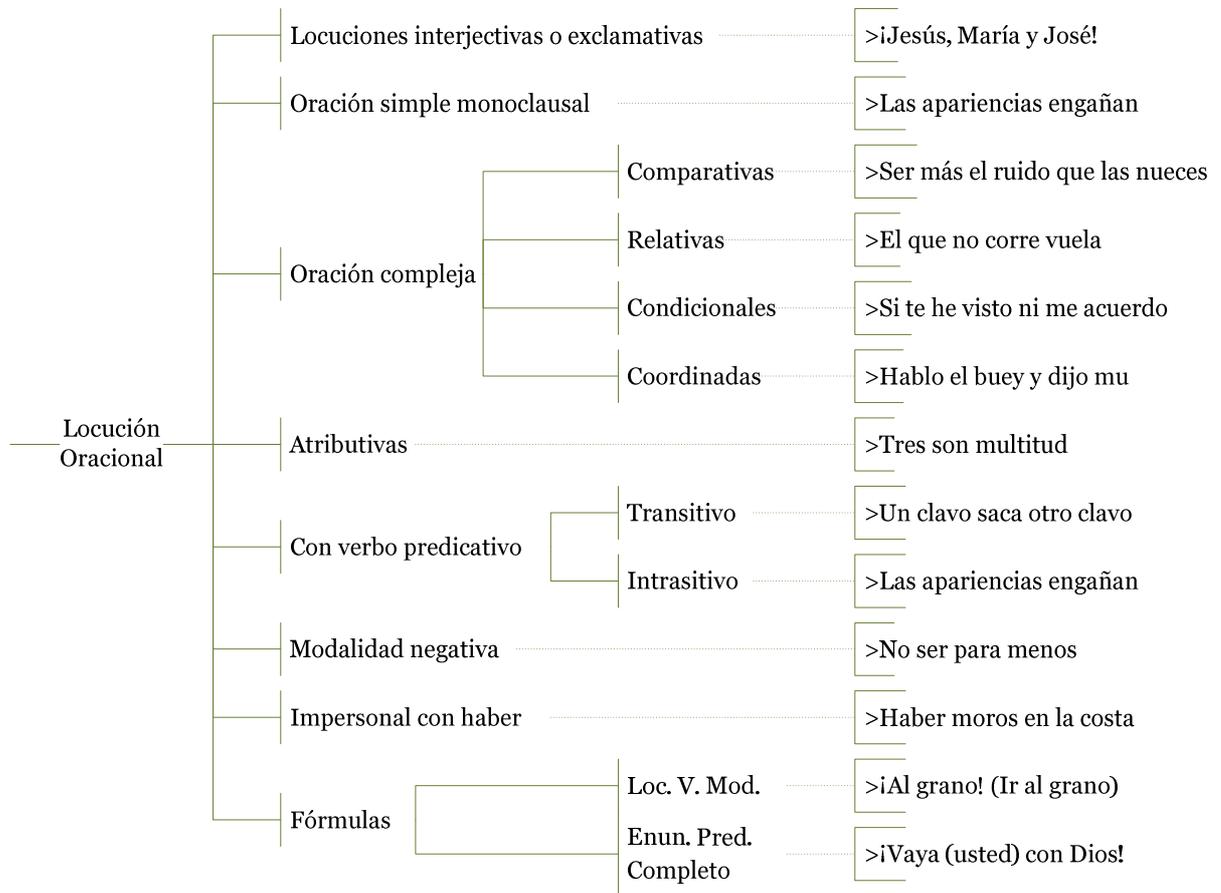
Esquema 23: Clasificación de las locuciones conjuntivas según García-Page

Las locuciones verbales ocupan, junto con las locuciones adverbiales, el primer lugar en cuanto a lo que a extensión se refiere, no tanto en lo que a subcategorías se trata sino en cuanto al número de expresiones que hacen parte de esta clasificación. García-Page (2008, pp. 134-152) trata este fenómeno y limita la extensión de las mismas al distinguir las características de las colocaciones, los predicados complejos, los sintagmas libres y los compuestos; por otra parte, a pesar de ser ésta una de las categorías más extensas, no nos ofrece, como lo hizo en ocasiones anteriores, un listado de patrones morfosintácticos que nos permita una fácil extracción semiautomática de las mismas.



Esquema 24: Clasificación de las locuciones verbales según García-Page

Por último, presentaremos el cuadro de la categoría oracional. Hemos tratado de extraer, de la mejor forma posible, las categorías explicadas por García-Page dentro de este acápite, ya que no encontramos una clasificación clara y definida. No obstante, el autor también nos ofrece algunas indicaciones sobre qué tipologías anteriores ha tomado para su propia conformación, así, entonces, utilizaremos algunas denominaciones usadas por autores anteriores.

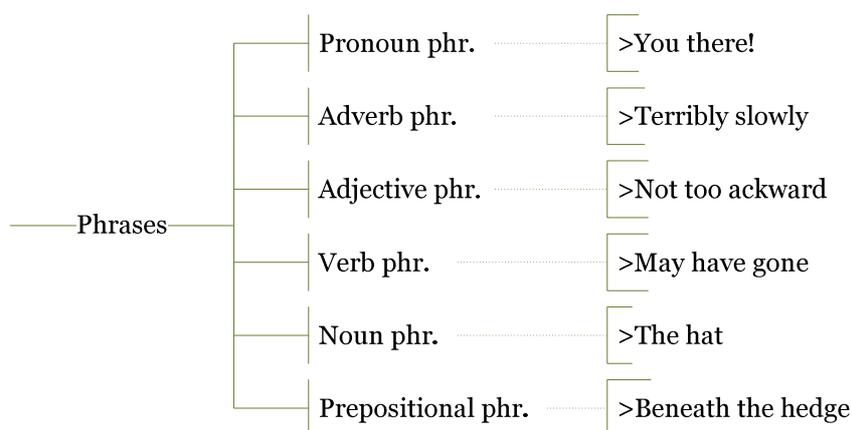


Esquema 25: Clasificación de las locuciones oracionales según García-Page

2.2.3. Algunas clasificaciones de las UF en lengua inglesa

Hablar de la fraseología en lengua inglesa conlleva una problemática aún mayor respecto a la definición del término, tal como se presenta en español. Ya hemos visto cómo el punto de partida fueron los trabajos de Charles Bally (1905) y, posteriormente, en lengua española —entre otros autores— el trabajo realizado por Casares (1950).

En lengua inglesa también encontramos varios autores que han trabajado algunos de los aspectos de la fraseología, pero, será evidente, conforme pase la elaboración de este capítulo que, cada uno de los recursos consultados puede referirse a las UF de diversas maneras. En primera instancia, al consultar *The Cambridge Encyclopedia of the English Language* (Crystal, 1995) encontramos que se tratan tanto las colocaciones —*collocations* (1995, p. 161)— como las locuciones —*phrases* (1995, p. 222)—, pero, en ningún momento, la enciclopedia hace mención alguna de términos como *locution*, *phraseology* o *phraseological unit*. Aun así, se muestra una pequeña taxonomía de *phrases* con alguna expresión que ejemplifique cada una de las categorías, ofrece también un pequeño análisis semántico que indica cómo, en algunas ocasiones, las *phrases* suelen expresar distintas características de tiempo, aspecto o modo. Finalmente, presenta el siguiente ejemplo: “*for example, an accompanying adverbial can dramatically alter the period of time to which a verb form refers: **I’m leaving tomorrow** is hours away from **I’m leaving** (said while going through the door)*” (Crystal, 1995, p. 224).



Esquema 26: Taxonomía de las *phrasal units* según Crystal

Neal Norrick, por su parte, en un capítulo del libro *Phraseology* editado por Burger (2007), habla de una posible tradición británica y otra estadounidense, ambas regidas por un acercamiento más antropológico, folclórico y literario, de algunas expresiones proverbiales y refranes. Este trabajo, aunque no ofrece ningún tipo de taxonomía, clasificación formal o sintáctica de las distintas unidades léxicas pluriverbales, sí procura un barrido histórico a través de los autores que han comprendido —para Norrick— la tradición tanto estadounidense como británica. En este punto, autores como Bloomfield (1933), Hocket (1958), Householder (1959), empiezan a aparecer en este trabajo.

En este pequeño y preliminar recuento que hace el autor da cuenta de tres etapas dentro del estudio de la fraseología inglesa: la primera es la definición del lexicón, que hace Bloomfield (1933), como “*the list of the irregularities in a language*”, pero no hace ningún análisis formal, sintáctico o funcional de las mismas. Hocket (1958), por su parte, agrupa las UF en una categoría llamada *idioms*, e incluye dentro de ésta algunos compuestos. Householder (1959) propone una gramática para este tipo de expresiones, y habla de la conveniencia de que sean inscritas dentro del lexicón (Norrick, 2007, p. 615).

Ahora bien, el verdadero inicio de la investigación de la fraseología en los Estados Unidos se da con la crítica que hace Wallace Chafe a Noam Chomsky y su GGT (Norrick, 2007, p. 616). Chomsky sostenía que el lexicón es “*simply an unordered list of all lexical formatives*” (Chomsky, 1965, p. 84) en la cual se debían incluir los *idioms*. Chafe (1968), tres años después, expone en un artículo que la idiomaticidad —una de las características de las UF— va en contra del criterio de composicionalidad de la teoría de Chomsky, el cual

explica que el significado de un compuesto lingüístico está formado por la sumatoria de los significados individuales de sus elementos compositivos y sus relaciones sintácticas.

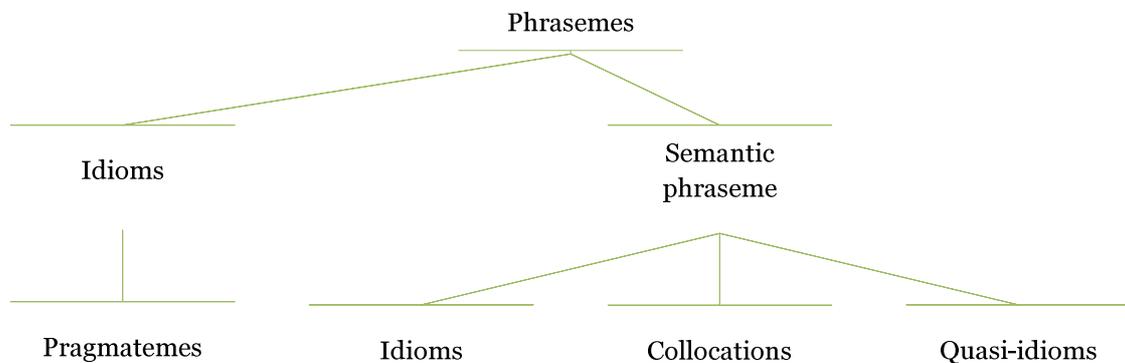
Respecto a la tradición británica, Norrick considera que sus inicios tienen en cuenta los trabajos de Firth (1957) que tratan la aparición de la noción de *collocation*, y los de Sinclair y Moon (1989, 1995) respecto a la lingüística de corpus.

Dentro de los autores consultados para este apartado sobre la fraseología inglesa hay otros dos de los que hemos tomado no solo su terminología, sino también sus taxonomías: Cowie (2001) y Mel'čuk (2001a). El primero nos presenta una tabla que permite analizar los términos relativos a las UF dadas por distintos autores que han trabajado el tema en lengua inglesa.

<i>Author</i>	<i>General category</i>	<i>Opaque, invariable unit</i>	<i>Partially motivated unit</i>	<i>Phraseological bound unit</i>
Vinogradov	Phraseological unit	Phraseological fusion	Phraseological unity	Praseological combination
Amosova	Phraseological unit	Idiom	Idiom (not differentiated)	Phraseme / Phraseoloid
Cowie	Composite	Pure idiom	Figurative Idiom	Restricted collocation
Mel'čuk	Semantic Phraseme	Idiom	Idiom(not differentiated)	Collocation
Gläser	Nomination	Idiom	Idiom(not differentiated)	Restricted collocation
Howarth	Composite unit	Pure Idiom	Figurative Idiom	Restricted collocation

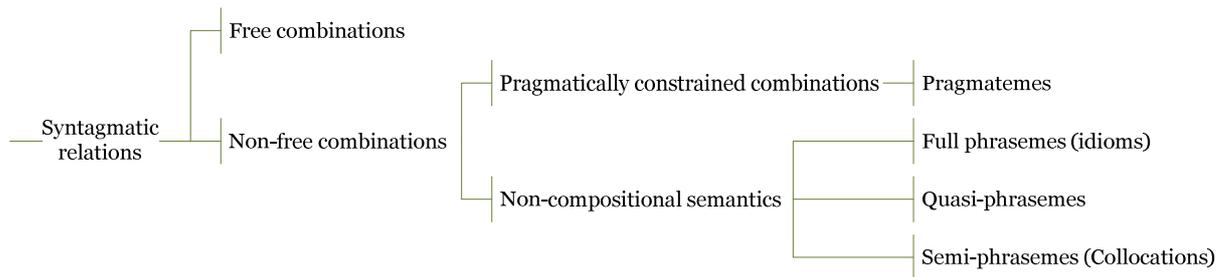
Tabla 3 Términos relacionados a las UF por autores (Cowie, 2001, p. 7)

Mel'čuk —en el mismo libro de Cowie, en la segunda parte— hace una descripción casi matemática de su taxonomía de unidades fraseológicas, pero, antes de que llegara este punto retoma el concepto saussureano del signo lingüístico, sirviéndose de éste para definir no sólo *phrase* (frasema), sino también *free phrase* (sintagma libre) y *collocation* (colocación) (Mel'čuk, 2001a, pp. 25-30). El teórico sintetiza su taxonomía en la siguiente figura:



Esquema 27: Taxonomía de *phrasemes* según Mel'čuk (2001a, p. 30)

Antunes (2006, p. 232) nos ofrece otro cuadro sobre la misma taxonomía de Mel'čuk donde tiene en cuenta las relaciones sintagmáticas de las distintas unidades. En este caso, podemos ver cómo se agrupan las distintas UF según sean frasemas, semifrasemas y cuasifrasemas, dependiendo de su nivel de fijación en la lengua.



Esquema 28: Taxonomía de Mel'cuk y sus relaciones sintagmáticas

Podemos afirmar que mientras en español se habla de un enfoque estrecho o amplio en el estudio de las UF (Carneado Moré & Tristá, 1985, p. 68), (Ruiz, 1998, p. 12), — mencionado también por: Rakotojoelima (2004, p. 25), Sosiński (2006, p. 23), Rišianová (2010, pp. 15, 18), Školnílová (2010, p. 7) y (López, 2012, p. 57)— en inglés sería impropio hacerlo —al menos para este trabajo— debido a que los autores consultados hasta el momento acuerdan en incluir las colocaciones como parte de las UF.

2.3. Terminología y fraseología

Hasta este momento hemos tratado los temas concernientes a la fraseología general en ambas lenguas, pero el objeto de estudio de este trabajo apunta a las UF especializadas. Queremos dar cuenta de los comentarios, apuntes y posibles sugerencias que se puedan hallar en algunos manuales de terminología al igual que las recomendaciones por parte de los organismos internacionales como el RITerm y el Realiter.

2.3.1. Manuales de terminología

En lo que respecta a los manuales de terminología, se han analizado autores y manuales tanto en español como en inglés, listados por aparición cronológica, nos limitamos simplemente a hacer un rastreo de palabras clave relacionadas con fraseología, —fraseología, unidades fraseológicas, locuciones, colocaciones, *phraseme*, *phraseology*, *phrase*, *phrasal*, *locution*, *collocation*, *idiom*— que nos lleven a indagar en estos recursos datos o información acerca del tema. Hemos tomado los trabajos de Felber (1984), Cabré (1993), Arnzt y Picht (1995), Fedor de Diego (1995), Pavel (2002)

2.3.1.1. Felber

Aunque tenderíamos a pensar que estudios más recientes sobre terminología dieran visiones un poco más amplias y sugerencias metodológicas sobre el tratamiento de las UF, es quizás este texto de Felber (1984) el que quizás más luces nos da sobre este fenómeno llegando al punto en que algunos de los autores ni siquiera hacen mención a éstas en ninguna de sus posibles denominaciones y tan sólo enuncian ese fenómeno.

Felber inicia su trasegar por las UF dando un pequeña taxonomía de componentes terminológicos entre los que se encontraban las *words* —*root words*, *compound words*, *derivative words*— y los *word groups* que define así: “*a word group is a linguistic expression consisting of several orthographically separated units (words)*. *Examples: headless set screw with flat point / conical teeth*” (Felber, 1984, p. 171).

Uno de los hallazgos más significativos es la inclusión dentro de este manual de terminología de signos que indicarían la presencia de un sintagma, expresión fija o UF dentro del artículo terminográfico.

“The zero and the asterisk placed superscript before a term indicate respectively that the term is either bad (°) or merely proposed (*). If the term in question is a phrase, the symbol is repeated at the end of the phrase examples: *garniture; *active part*” (Felber, 1984, p. 196)

Ya Felber nos indicaba en este año que no sólo existían recursos lexicográficos y terminográficos que contenían UF sino también la importancia de los mismos para la labor traductiva, la importancia de estos no se basa en el listado de expresiones de uso común sino de términos.

“Translation dictionaries can be complete and reliable only, if they are based on terminological vocabularies. For translation purposes also phraseological dictionaries are in existence. The items of which consist in a set of phraseological units in various languages. These phraseological units represent terminological units” (Felber, 1984, p. 201).

El autor luego define *phrase* utilizando el concepto primogenio de word group de la siguiente forma: “*Phrase is a terminological unit formed by a word group (verb group) such as "to switch off an apparatus". It is used frequently by translators*”. (Felber, 1984, p. 205). Páginas más adelante se hace mención del DICAUTOM, luego llamado

EURODICAUTOM, un banco de datos terminológico que “*was based on the idea of recording above all phraseologic units in various languages*” (Felber, 1984, p. 345)

2.3.1.2. Cabré

En su libro, la autora hace mención no sólo de la fraseología sino también de algunas de las características de las UF y afirma que todas las lenguas de especialidad poseen en mayor o en menor medida este tipo de expresiones. Define la fraseología y las UF así:

“Entre uno y otro tipo sintagmático, además, se dan una serie de construcciones que se pueden considerar a medio camino entre los términos sintagmáticos propiamente dichos y las combinaciones totalmente libres. Son combinaciones que, por un lado, aparecen con una frecuencia muy alta en el discurso de especialidad (lo que las aproxima a los sintagmas terminológicos), pero, por otro lado, no parece que correspondan a conceptos estables de un campo de especialidad, sino más bien a expresiones discursivas frecuentes en estos campos. En este caso específico hablamos de fraseología” (Cabré, 1993, p. 186).

Siguiendo estos parámetros es posible decir que Cabré no contempla la posibilidad de estudiar la fraseología desde la terminología. Estudios posteriores modificarán esta concepción.

2.3.1.3. Arnzt-Picht

A diferencia de Cabré (1993), Arnzt y Picht (1995) si consideran que tanto el contexto lingüístico como la composición léxica de las UF hacen parte del estudio de los términos y por ende de la terminología. Los autores definen la fraseología así: “se llama fraseología, al conjunto de giros de un lenguaje especializado. Pero *fraseología* —análogamente a *terminología*— se refiere también al estudio sistemático de los giros especializados”.

Los autores recopilan ejemplos del uso de la fraseología terminológica en los trabajos de Schlomann (1955) y Warner (1966) en los que no sólo se justifica la inclusión y estudio de la fraseología en los recursos terminológicos sino también algunos aspectos metodológicos a tener en cuenta al momento de organizar las expresiones fraseológicas como entradas léxicas.

2.3.1.4. Fedor de Diego

En la exploración que se hizo del libro de terminología no se halló como tal los términos fraseo o fraseología, y el único vestigio relacionado al análisis de unidades léxicas pluriverbales —sin ser estas necesariamente UF— se encuentra en el capítulo 3 en el aparte de las estrategias de formación de términos, específicamente la determinación: “en este caso el término consta de un elemento constituyente determinado (género) y uno o más elementos constituyentes determinantes del primero (características)” (Fedor de Diego, 1995, p. 55).

2.3.2. Otros autores que tratan la fraseología dentro de la terminología

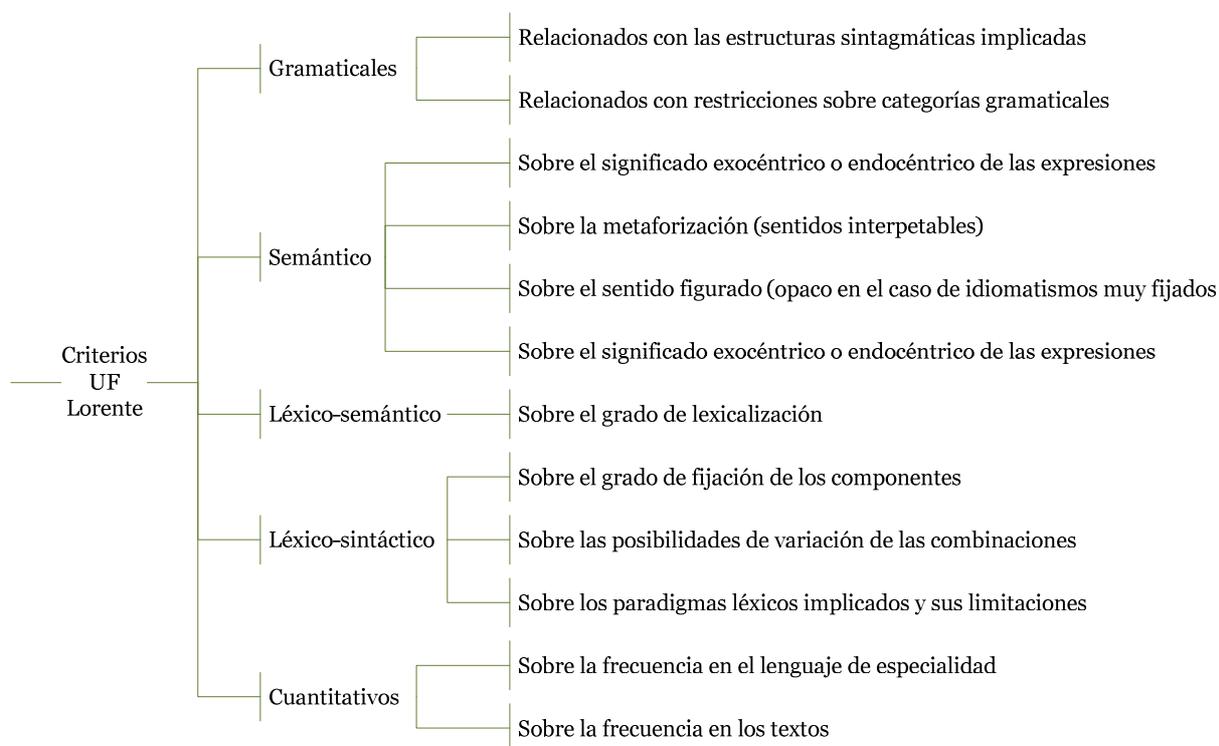
No sólo los manuales ofrecen información sobre la terminología, es importante también saber que otros autores —recientemente— han trabajado la fraseología desde un enfoque terminológico. A continuación presentamos una breve lista de autores con algunas consideraciones relacionadas.

2.3.2.1. Lorente

Para esta autora, la fraseología se puede abordar desde distintas aproximaciones, es así como establece que es debido —en primera instancia— a los traductores que se ha empezado a discutir el rol de la fraseología dentro de la terminología:

“Si bien fueron los traductores los que protagonizaron en primera instancia los debates sobre la revisión del papel de la fraseología respecto de la terminología en general y de la producción de recursos terminográficos en particular, desde otros ámbitos aplicados se han ido incorporando a este debate investigadores y profesionales que ven en la identificación del fenómeno fraseológico recurrente, y de sus relaciones con la terminología, una vía de mejora para sus aplicaciones en ingeniería lingüística, inteligencia artificial o procesamiento del lenguaje” (Lorente, 2001, p. 4).

Además, y como lo dice la misma autora en la cita anterior, uno de los grandes aportes de este estudio a la terminología, está relacionado con los extractores terminológicos automáticos y los patrones utilizados por éstos para lograr su cometido. De forma muy esquemática y sucinta, la autora nos ofrece una serie de criterios para la categorización de las UF, sintetizamos sus criterios en el siguiente esquema.



Esquema 29: Criterios para la categorización de las UF según Lorente

La autora no se limita a la descripción de la fraseología general sino que ofrece una definición y una delimitación de la misma siguiendo los preceptos de la teoría comunicativa de la terminología, —en adelante TCT— planteada por Cabré (1993), además es muy clara al afirmar que las UF no pertenecen a una lengua de especialidad en

específico sino que estas pertenecen a los textos de las lenguas de especialidad. Sin embargo nuestro único punto de divergencia es la hipótesis que las UF no migran de una lengua de especialidad a otra, pero para comprobar esto deberíamos entonces hacer un análisis de frecuencia de uso de diferentes UF en distintos argots especializados, —tanto en corpus escritos como orales— no sería extraño pensar que debido a la interacción interdisciplinar de los expertos, las UF de una lengua de especialidad no puedan migrar a otra. Teniendo todo lo antes dicho en cuenta procede a delimitar la fraseología especializada así:

“La unidades fraseológicas son expresiones lingüísticas que transmiten conocimiento especializado y que son usadas en las transmisión de información, y forman parte por lo tanto de la competencia comunicativa de los especialistas de una materia [...] así pues aunque en la TCT, distingamos claramente entre unidades léxicas y unidades fraseológicas, estas últimas también forman parte de su cuerpo de unidades de análisis” (Lorente, 2001, p. 15).

Aparte de la definición, Lorente plantea la “hipótesis categorial” modificada luego por la “hipótesis del carácter relacional del núcleo” (Lorente, Bevilacqua, & Estopà, 1998). En la primera afirma que tanto los sintagmas verbales, como los nominales que tengan sustantivos deverbales por núcleo, eran candidatos a ser UF; la modificación, por su parte, dice que los sintagmas nominales son más candidatos a convertirse en términos debido a su tendencia a la lexicalización mientras los otros tipos de sintagmas —verbales, adjetivales, adverbiales, preposicionales— no podían ser lexicalizados con tanta facilidad

y por ende eran candidatos a UF o UFE —unidades fraseológicas especializadas— como las denomina la autora (Lorente, 2001, p. 16).

Por último hacemos extracción de dos apartes del artículo de Lorente: el primero da cuenta de los criterios que comparten las UF generales y las UFE; mientras el segundo es un compendio de las diferencias entre las mismas.

“Las unidades fraseológicas especializadas o generales comparten las siguientes características:

- Son combinaciones complejas de unidades léxicas y gramaticales.
- Son estructuras sintagmáticas u oraciones completas, que responden a las reglas de la sintaxis.
- No están lexicalizadas, no forman parte del léxico (no designan).
- Aunque tienen tendencia al estereotipo (fijación reiterada en el uso), no llegan a estar plenamente fijadas, admitiendo la variación y la inserción de elementos.
- Su significado puede ser predecible, por la composicionalidad de los significados de sus integrantes, o impredecible por su carácter exocéntrico o por contener sentidos figurados” (Lorente, 2001, p. 18).

A diferencia del anterior recuento, el cuadro que presentamos a continuación es un contraste entre ambos tipos de unidades:

<i>UF de la LG</i>	<i>UFE</i>
Son combinaciones de palabras de la LG.	Son combinaciones de términos (UT).
Las UF aportan connotación, expresividad.	Tendencia mayor a la denotación.
Significado general, relacionado con el conocimiento del mundo y con la experiencia.	Significado especializado, relacionado con unos conocimientos reglados y aprehendidos voluntariamente
Aparición frecuente en el lenguaje oral, o en expresiones literarias que intentan reflejarlo.	Localización prioritaria en textos escritos, ya que la oralidad no suele ser analizada en ámbitos especializados.
Identificación con una comunidad lingüística, cultural, religiosa o con un grupo social	Identificación con una temática, con una comunidad académica, con una escuela o con un grupo profesional (argot).

Tabla 4: Características de las UF de la lengua general y las UF en lenguas de especialidad

Tanto los criterios como la definición propuesta por esta autora, son base fundamental para la consecución de las metas de nuestro trabajo, ya que nos permitirán evaluar cada una de las expresiones marcadas en nuestro corpus y sus criterios servirán como base de evaluación de las mismas expresiones.

2.3.2.2. L'Homme

La autora ha trabajado en años recientes lo que podríamos denominar fraseología especializada, y decimos “podríamos” debido a que L'Homme ha decidido acuñar un nuevo término para estas expresiones —*Combinaisons lexicales spécialisées*— (L'Homme, 1998) las cuales por sus características son más similares y comunes a las colocaciones que a las locuciones, por consiguiente su forma de clasificar las mismas podría dar pie a

una posible metodología de lematización e indización en diccionarios especializados. Observemos cómo define las combinaciones:

“Constraints related to conventions established within a given subject field makes lexeme1 prefer the company of lexeme 2 rather than that of other lexemes. Examples of specialized lexical combinations found in the literature are provided in (1):

*(1) créer un fichier: *établir un fichier (Heid & Freibott, 1991).*

*administrer un médicament: *donner un médicament (Laporte & L'Homme, 1997).*

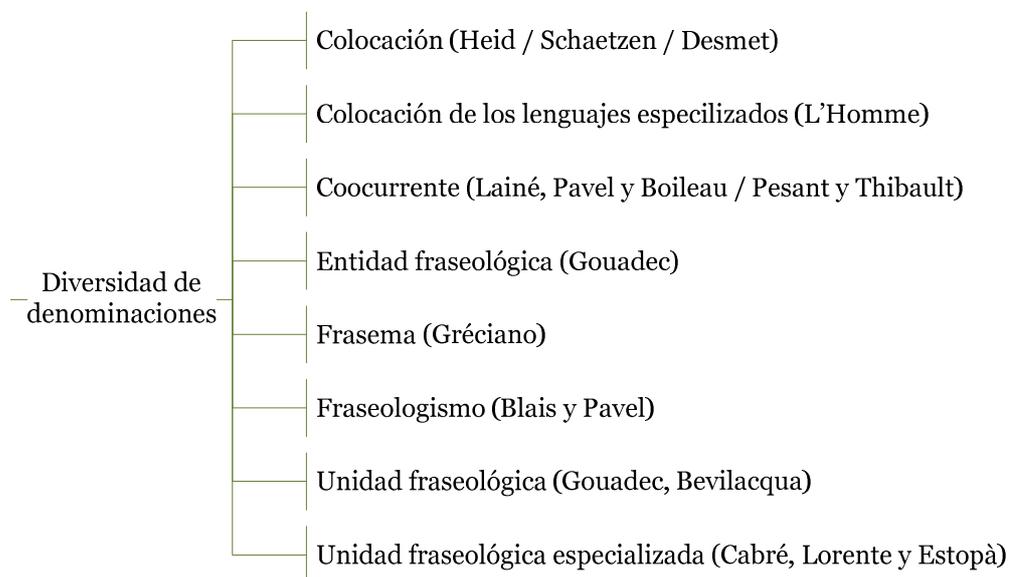
Typically, lexeme 1 is a term (defined as the keyword), a unit with special reference within a specialized subject field. In (1), for example, the terms are fichier (file) and médicament (drug, medecine). The other lexeme is often referred to as the co-occurrent” (L'Homme & Bertrand, 2000, p. 497).

Así, entonces, y según lo planteado por L'Homme, el término sería la palabra lexicográfica o lematizada en la que se listaría la expresión como tal.

2.3.2.3. Bevilacqua

Los dos trabajos consultados de esta autora son quizás los que, en lo que respecta a la percepción de las UF en lenguas especializadas, son esencialmente importantes para este trabajo en dos aspectos: el primero lo que ella misma llama diversidad de

denominaciones en cuanto a la definición de UF, y el segundo sobre los criterios de reconocimiento de una UF. Del primero, ofrece un listado de autores y trabajos de dónde se sacó dicha información, aunque resumimos a continuación ese aporte de Bevilacqua en un esquema, cabe aclarar que es tomado literalmente de la fuente (Bevilacqua, 1999, pp. 51-53) (Bevilacqua, 2004, p. 24).



Esquema 30: Denominaciones del término UF por autores

Los autores citados por Bevilacqua en ese aparte son los siguientes (Heid (1992), 1994), (1998)), de Schaetzen (1993), Desmet (1994), (L'Homme (1998), 2000)), Lainé, Pavel, and Boileau (1992), Pesant and Thibault (1998), Gouadec (1994), Gréciano (1993), Blais (1993), Pavel (1993), Bevilacqua (1996), Cabré, Lorente, and Estopà (1996).

El segundo aspecto que nos interesa abordar de esta autora está relacionado con los criterios de selección de UF en lenguas de especialidad. En su trabajo de 1999, sintetiza los criterios escogidos de la siguiente forma:

“[...] Los más significativos para aplicar a nuestro corpus. Seleccionamos, así, unidades que cumplen las siguientes condiciones:

- Unidades de carácter sintagmático.
- Que incluyan, como mínimo, una UT entre sus componentes.
- Que tengan en conjunto un valor relacional.
- Que tengan significado semicomposicional o composicional.
- Que sus componentes tengan un determinado grado de fijación interna.
- Que tengan una frecuencia relevante en los textos analizados de entre ellas.
- Que tengan un núcleo sintáctico verbal o nominal deverbal que aporta carácter relacional al conjunto.
- Que aparezcan recurrentemente en los textos del ámbito de especialidad que constituye el objeto de este trabajo” (Bevilacqua, 1999, pp. 114-115)”.

Luego, en su tesis doctoral los resume un poco más dejando estos criterios:

- “La presencia de una unidad terminológica.
- La estabilidad sintáctica, o sea, que existe un determinado grado de fijación en las estructuras morfosintácticas de las UFE.
- La estabilidad semántica, determinada por la relación establecida entre los elementos que componen la unidad.
- El índice relevante de frecuencia en los textos de un ámbito especializado.
- Su uso en un ámbito especializado”. (Bevilacqua, 2004, p. 31)

Algunos conceptos de estos criterios fueron tenidos en cuenta para el presente trabajo, pero pretendemos explicar esto posteriormente.

2.3.2.4. Cardero

En el tercer capítulo de su libro titulado *terminología y procesamiento* (2003), la autora trata el tema de la fraseología desde el ámbito de la terminología. Lo primero que se puede observar de sus comentarios es el reconocimiento de las UF como posibles unidades terminológicas y que este grupo recoge dentro de sí un gran número de posibles denominaciones y combinaciones—visto desde un enfoque ancho— “dentro del ámbito de la terminología existen diversas denominaciones para referirse a los términos mayores

que una palabra: *colocación, frasema, fraseologismo, entidad fraseológica, unidad fraseológica, unidad fraseológica especializada*, entre otras” (Cardero, 2003, p. 84).

El resto de sus apreciaciones con respecto a la fraseología y la terminología recogen los trabajos hechos —entre otros— por Lorente et al. (1998), Cabré (1999) y Bevilacqua (1999), algunas ya mencionadas en este trabajo.

2.3.3. Organismos internacionales de estandarización terminológica

La terminología —a diferencia de la traducción— ha tenido organismos internacionales que velan por la estandarización no sólo de la metodología de la gestión terminológica sino también de los recursos y las estrategias de formación de términos y neologismos. Entre los estamentos consultados tenemos el RITerm (Red iberoamericana de terminología)¹, el INFOTERM (*International Information Centre for Terminology*)² y la REALITER (Red panlatina de terminología)³.

La búsqueda se hizo empleando dos palabras clave en inglés —*phraseology* y *phrase*— y en español —fraseología y frase—, dicha búsqueda se llevó a cabo utilizando los buscadores de los propios sitios web de las distintas instituciones. Podemos encontrar que las palabras clave empleadas recogen poca o nula información sobre el tema, en la página

¹ RITerm URL: <http://www.ufrgs.br/riterm/>

² INFOTERM URL: <http://www.infoterm.info/>

³ REALITER URL: <http://www.realiter.net>

de la RITerm sólo se hayan tres entradas de ponencias de 1996, 2004 y 2006 y en su mayoría se centran en 5 autoras, 4 de ellas —Bevilacqua, Cabré, Estopà, Lorente— ya consultadas para este trabajo. En la página del INFOTERM las palabras clave no devuelven respuesta alguna desde el buscador, aun así es de resaltar que el *Terminology Manual* de Felber (1984), fue un trabajo publicado en 1984 por la UNESCO en colaboración con INFOTERM y es quizás uno de los manuales de terminología que mayores aportes hace sobre la fraseología. Por su parte, la página de la REALITER sólo da un resultado desde su buscador y tiene que ver con el proyecto ESLEE, lamentablemente el texto consultado no nos ofrece el artículo completo para poder sacar la información de este punto.

2.4. Terminología y lexicografía

Son dos los libros que fueron consultados en este aparte que sirvieron de soporte teórico para el presente trabajo, fueron escogidos estos ya que son recopilaciones de los artículos que relacionan la fraseología con la lexicografía, estos fueron: *Diccionarios y fraseología* (Alonso, 2006) y *Colocaciones y fraseología en los diccionarios* (Mellado, 2008).

Del libro de Alonso se han tomado los artículos de González (2006) y el de Penadés (2006). En el primer artículo la autora habla sobre la clasificación de las colocaciones y

las locuciones dentro del DRAE⁴, sobre las colocaciones se habla de la continua inserción de las colocaciones dentro de las UF siguiendo la taxonomía —presentada con antelación en este trabajo— de Corpas. Con respecto a las locuciones la autora afirma: “la tipología del DRAE se basa en la catalogación realizada por Casares (1950)”. (González, 2006, pp. 225-227).

Un gran acercamiento a nuestro trabajo lo hace Penadés en el segundo artículo cuando afirma que:

“La asignación de una clase a las locuciones no puede verse como una cuestión baladí, relacionada, sin más, con la reivindicación, desde la fraseología y la fraseografía, de un tratamiento idéntico para las unidades léxicas simples y para las locuciones. La razón estriba en el hecho de que considerar una locución como perteneciente a una u otra clase trae aparejado el establecimiento de lemas distintos y la redacción de definiciones diferentes para la que tal vez podría ser la misma locución. Dicho de otro modo, esta cuestión afecta al significante y al significado de una locución, o sea a la propia esencia de una unidad” (Penadés, 2006, pp. 252-253)

Para estas autoras la utilización de marcas y etiquetas a los diccionarios está ya dado por hecho, no obstante, el análisis que presentaremos en este trabajo da pie a pensar que éstas deben ser normalizadas y presentadas al inicio del recurso lexicográfico.

⁴ DRAE: diccionario de la lengua española

Del libro de Mellado hemos tomado el artículo de Ortega y González (2008), el cual recoge las distintas marcas diacríticas y diasistemáticas tanto del DRAE como DEA⁵, sin embargo ninguno de los artículos consultados da cuenta de estas mismas marcas en un entorno bilingüe y especializado que es el caso que abordaremos durante el presente trabajo de grado. Retomamos la segunda conclusión dada por los autores con respecto a los criterios fraseográficos tenidos en cuenta: “los criterios fraseográficos en ambos repertorios se presentan menos estables que los relativos al léxico simple”(2008, p. 244). Esa misma falta de rigurosidad es la que se observa en el análisis del DCI que ha servido de corpus para el presente trabajo de grado.

2.5. Términos útiles relativos a la lexicografía para este trabajo

2.5.1. Fuentes

La fuente está constituida por la bibliografía de donde se extraen los textos que harán parte del corpus, en este sentido compartimos con Porto Dapena cuando señala que fuente es:

“todo aquello capaz de proporcionar los datos o materiales necesarios para la constitución del corpus o fichero lexicográfico. En general, pueden distinguirse dos tipos fundamentales de fuentes: a) lingüísticas o primarias, representadas por toda realización concreta de la

⁵ DEA: diccionario del español actual

lengua, sea un texto oral o escrito, y b) metalingüísticos o secundarias, constituidas por todas aquellas obras –por ejemplo otros diccionarios– que se ocupan de alguna manera del léxico que va a ser estudiado por el diccionario” (Porto, 2002, p. 105).

2.5.2. Macroestructura

Hartmann y James definen la macroestructura como:

“The overall LIST structure which allows the compiler and the user to locate information in a REFERENCE WORK. The most common format in Western dictionaries is the alphabetical WORD-LIST (although there are other ways of ordering the HEADWORDS, e.g. thematically, chronologically or by frequency), which constitutes the central component.” (2002, p. 91).

En general, aunque con matices, la mayoría de autores comparten esta definición, así por ejemplo, Haensch define la macroestructura como:

“la ordenación del conjunto de materiales que forman el cuerpo del diccionario (por ej., orden alfabético u orden sistemático), conjuntamente con el prólogo, o prefacio, a veces una introducción fonética y gramatical, las instrucciones para el usuario y los posibles anexos” (1997, p. 36)

Haensch (1997, pp. 44-45) agrega los elementos formales que deben aparecer en el diccionario y cómo deben aparecer, dividiendo el diccionario en una parte introductoria, cuerpo del diccionario y los anexos. Tenemos entonces que la macroestructura es la

estructura en la cual se ordenan las entradas –que suele ser alfabéticamente para los diccionarios léxicos–, respondiendo a criterios de selección como finalidad, tamaño, prejuicios ideológicos, frecuencia de uso de los vocablos, entre otros (Porto, 2002, pp. 171-175); y que contiene elementos formales como la introducción, la bibliografía, guía al usuario, la lista de abreviaturas, el prólogo de cortesía, las marcas gramaticales usadas, y los anexos que el lexicógrafo considere pertinentes. Todo esto con el fin de que el usuario (y el compilador) encuentren fácilmente la información en el diccionario.

2.5.3. Microestructura

Entenderemos por microestructura una serie de “predicados que aportan información sobre las propiedades formales, pragmáticas y semánticas de la unidad que se somete a análisis” (Anglada, 2005, p. 148). La misma autora define cuatro tipos de informaciones que se pueden encontrar al interior de la microestructura del diccionario: etimológica –diacrónica, relacionada con el origen de la palabra–, gramatical –función, categoría y subcategoría–, particularizada –informaciones concretas de uso– y semántica –encargada de dar cuenta de la definición o el significado del vocablo–. Observamos cómo, para esta autora, la microestructura sólo la conforman las informaciones encontradas después a la entrada –lema, palabra lexicográfica o unidad de análisis–, caso contrario a lo planteado por Porto Dapena quien habla de la microestructura como la información encontrada dentro del artículo lexicográfico y la divide en dos: una parte enunciativa –la entrada–; y otra la informativa –también denominada cuerpo del artículo– (Porto, 2002, p. 182).

Lo que Anglada categoriza dentro de las informaciones particularizadas de la microestructura, Porto Dapena lo organiza por etiquetas llamadas marcas diasistemáticas –diacrónica, diatópica, diastrática, diafásica–.

Seco, en un punto medio entre Anglada y Porto Dapena, habla de dos enunciados, el primero que identifica el vocablo que se va a describir y el segundo que proporciona información sobre la entrada. El autor no sólo se limita a una simple descripción de las informaciones, también ofrece un orden normalizado (Seco, 2003, p. 25). En resumen, para este proyecto se tomarán los lineamientos de organización de la microestructura y etiquetaje sugeridos y expuestos anteriormente por Porto Dapena.

Asumiremos que la entrada es la palabra lematizada y ordenada alfabéticamente al inicio del artículo lexicográfico, que se emplea como unidad de análisis (palabra lexicográfica) y el artículo lexicográfico contendrá tanto la entrada como las informaciones posteriores a la misma.

2.6. Resumen de criterios y conceptos teóricos para este trabajo

A continuación nombraremos entonces aquellos conceptos –de índole fraseológica– que emplearemos para el siguiente trabajo. Éstos darán cuenta

principalmente de los criterios de selección de las UF en contextos especializados y las taxonomías a usar para el análisis de las UFE marcadas dentro del DCI.

2.6.1. Criterios de selección de UF

Quiroz et al. (2013), sintetizan los criterios postulados por Lorente (2001), y Aguado de Cea (2007) caracterizadores de las UFE. Hemos decidido tomar solamente los siguientes como guía para determinar si una expresión es una UFE:

1. “Son estructuras sintagmáticas con presencia de un término o UT y con tendencia a la denotación.
2. Incluyen un verbo o un elemento deverbal.
3. Tienen un alto grado de fijación sintáctica, aunque pueden permitir la inserción de otros elementos.
4. Muestran regularidad compositiva.

5. Admiten la conmutación de sus elementos, aunque hay ciertos grados. Así, se observa que el sentido especializado de un verbo está determinado por un conjunto limitado de nombres.
6. Mantienen un significado específico dentro de un dominio, por lo que adquieren una estabilidad semántica” (Quiroz et al., 2013).

2.6.2. Definición de UFE

Esta definición la ofrecen los mismos autores —Quiroz et al. (2013)— "combinación léxica que resulta del vínculo sintáctico recurrente entre un término, sea éste simple o compuesto (el núcleo de la UFE) y otros elementos lingüísticos que, al entrar en relación semántica con el término, forman una unidad léxica con significado especializado".

Esta misma definición y criterios serán evaluados tanto a las expresiones en español como en inglés. Se presentará también en ambas lenguas un listado de unidades con sus correspondientes patrones morfosintácticos.

3. Metodología general de trabajo y constitución del corpus

3. METODOLOGÍA GENERAL DE TRABAJO Y CONSTITUCIÓN DEL CORPUS	72
3.1. INTRODUCCIÓN	73
3.2. DESCRIPCIÓN Y SELECCIÓN DEL CORPUS INGLÉS-ESPAÑOL / ESPAÑOL-INGLÉS	73
3.3. HERRAMIENTAS Y RECURSOS	78
3.4. METODOLOGÍA DE ANÁLISIS GENERAL	88
3.5. EXTRACCIÓN DE LAS UFÈ Y TRATAMIENTO DE LOS DATOS	89
3.6. ASPECTOS ESTADÍSTICOS	89

3.1. Introducción

Con el fin de cumplir con los objetivos planteados y demostrar las hipótesis postuladas se utilizaron la siguiente metodología y herramientas de análisis.

3.2. Descripción y selección del corpus inglés-español / español inglés

A continuación hacemos una breve descripción de los aspectos generales del diccionario.

Nombre:	<i>Diccionario de comercio internacional – Importación y exportación, Inglés-Español/Spanish-English</i>
Autores:	Enrique Alcaráz Varó (editor) y José Castro Calvín (autor)
Editorial:	Ariel
Fecha de publicación:	9 de octubre de 2007
Páginas:	1168
Términos:	17000 (de acuerdo con la editorial)
Formato:	impreso 16 x 25 cm

Tabla 5: Características generales del diccionario

El DCI, es un diccionario bilingüe, especializado, sincrónico, descriptivo, dirigido a “especialistas y estudiosos de las diversas áreas de la economía y del comercio internacional y mediadores lingüísticos” (Alcaraz & Castro, 2007).

El criterio de selección del material se debió a varios factores, entre los que se encuentran: el prestigio y recorrido de la editorial, el área temática relacionada con eventos vinculados al proceso de globalización, así como ser un recurso lexicográfico bilingüe, de especialidad y actual. Es importante resaltar el hecho de que en nuestro quehacer cotidiano como traductores, docentes e investigadores, nos vimos en la obligación de enfrentarnos a textos y corpus de trabajo de esta misma área de especialidad. La experiencia previa en dichas labores fue lo que, en una primera etapa, motivó la selección, no sólo del objeto de estudio del presente trabajo de grado, sino también del corpus a consultar y analizar.

Dentro de los criterios que se tuvieron en cuenta para la recopilación del corpus nos basamos en algunos de los ofrecidos por Quiroz (2008, p. 74); así que optamos por la recopilación de unidades de análisis de un diccionario cuyas áreas de especificidad fueran afines, para controlar los resultados en los que, a variación horizontal, se refiere; esto se logra controlando las áreas temáticas y evitando tomar varios diccionarios para dicha labor.

Es innegable el hecho, y Quiroz lo explica también en su tesis doctoral, que “tener varios ámbitos [de análisis] abriría otras puertas, así como otras variables que no podríamos controlar de manera fiable dada la configuración del corpus que nos habíamos planteado desde un comienzo” (Quiroz, 2008, p. 75).

Ya en la estructura del diccionario, el apartado titulado “Fuentes” nos provee una lista de los acrónimos y siglas que se emplean para designar los distintos ámbitos temáticos contenidos en el mismo:

<i>Acrónimo</i>	<i>Descripción</i>
<i>CONTA / ACCTG</i>	contabilidad / accounting
<i>DER / LAW</i>	derecho / law
<i>DOC / DOCMT</i>	documentación / documentation
<i>ECON / ECON</i>	economía / economics
<i>FIN / FIN</i>	finanzas / finance
<i>GRAL / GEN</i>	general / general
<i>FISC / TAXN</i>	fiscalidad / taxation
<i>REC.HUM. / HH. RR.</i>	recursos humanos / human resources
<i>LOGÍST / LOGIST</i>	logística / logistics
<i>MERCAD / MKTG</i>	mercadotecnia / marketing
<i>ORGAN / ORGAN</i>	organización / organization
<i>SEGUR / INSCE</i>	seguros / insurance

Tabla 6: Lista de acrónimos y siglas de las marcas diatómicas empleadas en el diccionario

Luego de la recopilación del corpus, lo segmentamos y organizamos en una plantilla donde, en primera instancia, se separan por líneas tanto las entradas como las subentradas del diccionario; de igual forma, reordenamos las entradas de modo tal, que al momento de extraer los patrones morfosintácticos de las distintas unidades pluriverbales, se evitara al máximo el ruido en el orden de los componentes de las mismas.

A continuación, identificamos y catalogamos las marcas de contorno que se pudieron reconocer en cada una de las entradas y subentradas: las marcas de lematización por homonimia, los acrónimos asociados a los distintos términos y unidades pluriverbales, cada una de las categorías gramaticales que vienen asociadas a las entradas, la marca

geográfica de uso exclusivo en contextos estadounidenses (*US en la sección inglés / español*) presentes en algunos términos, los ámbitos temáticos mostrados en la tabla anterior, la marca diafásica de coloquial y por último, el primer equivalente o definición encontrada al momento de consultar.

Luego, empleamos un etiquetador automático para cada una de las entradas y subentradas extraídas del diccionario, para generar dos inventarios: uno que da cuenta de los patrones morfosintácticos de las unidades pluriverbales (2 tokens o más) y otro con las categorías gramaticales de las unidades monoléxicas (1 token), tanto de las entradas y subentradas del diccionario como de sus equivalentes. Los patrones morfosintácticos fueron organizados por medio de expresiones regulares para formar cada una de las cadenas sintácticas.

Por medio de una hoja de cálculo y la utilización de fórmulas encadenadas, incluyendo dentro de las mismas formulas lógicas (SI), de conteo (LARGO) y de texto (SUSTITUIR), fue posible realizar el conteo de cada una de las palabras presentes en cada una de las celdas consultadas. La hoja de cálculo poseería entonces 16 columnas distribuidas así: *Preposition / verb* (preposición o verbo que acompaña la palabra lexicográfica listada alfabéticamente), *Entry* (entrada o subentrada sin colocados adyacentes), *Expression to be tagged* (tanto la entrada o subentrada como sus complementos pero en orden sintáctico no lexicográfico), *Commentaries* (contorno de uso), *Homonyms* (numeración de las entradas listadas por homonimia), *Acronyms* (acrónimos, siglas, inicialismos o unidades de reducción híbridas vinculadas a las

distintas entradas y subentradas), *TT POS* (etiquetaje tomado por el etiquetador externo), *POS* (categoría gramatical de las entradas del diccionario), *Tokens* (número de palabras de cada una de las unidades léxicas listadas), *US* (marca geográfica de uso reservado para Estados Unidos), *By Topic* (marca diatópica), *Colloq* (marca diafásica de uso coloquial), *Entry 2* (referencia cruzada de las entradas o subentradas), *Equivalent / Definition* (equivalente o definición de la entrada o subentrada), *Commentaries* (contorno de uso del equivalente de la entrada o subentrada), *Lexicographic article* (informaciones restantes del artículo lexicográfico de las entradas).

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Prepositiv	Entrg	Expresión te le tagged	Commentaries	Homonym	Acronym	TT POS (ext.)	POS	Token	US	By Topic	Colloq	Entry 2	Equivalent / Definition
1	C	C				SVM	nc	1				V. consular	
2	CI	CI				NP	nc	1				V. case	
3	CA	CA				NP	nc	1				V. competent authority	
4	CABAF	CABAF				NN	nc	1				V. currency and bank adjustment factor	
5	CABEI	CABEI				NP	nc	1					Central American Bank fo
6	cable	cable				NN	nc	1		GRAL			cable
7	cable address	cable address				NNNN	nc	2		GRAL			dirección cablegráfica
8	cable broker	cable broker				NNNN	nc	2		DEP/LOGIST/COU			agente/ corredor intermed
9	cable draft	cable draft				NNNN	nc	2		DOC/FIN			giro cablegráfico
10	cable transfer	cable transfer				NNNN	nc	2		DOC/FIN			transferencia telegráfica
11	cablegram	cablegram				NN	nc	1		GRAL			cable
12	cabotage	cabotage		1		NN	nc	1		LOGST			cabotaje
13	cabotage	cabotage		2		NN	nc	1		GRAL			libre prestación de servic
14	cabotage rights	cabotage rights				NNNNS	nc	2		LOGST			derechos de cabotaje
15	CAC	CAC				NP	nc	1				V. currency adjustment charge	
16	CACM	CACM				NN	nc	1				V. Central American Common Market	
17	CAD	CAD				NN	nc	1					Canadian dollar
18	CAF	CAF				NP	nc	1					
19	calendar	calendar				NN	nc	1		GRAL		V. currency adjustment factor	calendario
20	calendar days	calendar days				NNNNS	nc	2		DEFV			días naturales
21	calendar month	calendar month				NNNN	nc	2		DEFV			mes civil
22	calendar year	calendar year				NNNN	nc	2		DEFV			año civil
23	call	call		1		VV	nc	1		GRAL			llamada
24	call	call		2		VV	nc	1		DEFV			exigencia
25	call	call		3		VV	nc	1		FINV			opción de compra de val
26	call	call		4		VV	nc	1		LOGST			escala
27	call	call		5		VV	nc	1		MERCAD			demanda de un producto
28	call	call		6		VV	nc	1		GRAL			visita
29	call	call		7		VV	nc	1		DEFV			emplazamiento
30	call a meeting	call a meeting				VVDT NN	nc	3		DEFV			convocar una reunión
31	call	call				NNJJ	nc	2		FINV			a demanda
32	at call	at call				VV INXX	nc	2		LOGST			hacer escala en...
33	at call	at call				NN INDT JJ	nc	4		LOGST			entrar en puerto
34	at call	at call				NN INDT JJJJ	nc	5		LOGST			hacer escala en un puert
35	call at an intermediate port	call at an intermediate port				NNNNS	nc	2		LOGST			gastos de escala
36	call costs or expenses	call costs or expenses			or expenses	NNNNS	nc	2		LOGST			convocatoria de licitaci
37	call for tenders	call for tenders			or tenders	NNNNS	nc	3		DEFV			convocatoria de licitaci
38	call for bids or tenders	call for bids or tenders				NNNNS	nc	3		DEFV			convocatoria de licitaci
39	call money	call money				NNNN	nc	2		FINV			dinero exigible
40	call option	call option				NNNN	nc	2		FINV			opción de compra
41	call purchase	call purchase				NNNN	nc	2		DEFV			opción de compra
42	call rate	call rate				NNNN	nc	2		FINV			opción de compra de ma
43	call station	call station				NNNN	nc	2		US	LOGST		zona de entrega y recog
44	callable	callable				JJ	nc	1		FINV			amortizable
45	calling card	calling card				NNNN	nc	2		GRAL			tarjeta de visita
46	calling surcharge	calling surcharge				NNNN	nc	2		LOGST			suplemento de escala
47	Calvo	Calvo				NP	nc	1		GRAL			Calvo

Figura 1: Compilación de corpus en una hoja de cálculo

Aunque sólo se hace necesario analizar aquellas unidades dentro del diccionario que contienen la marca *fr*, que es, hipotéticamente, la marca referente a las unidades léxicas pluriverbales de frase o frasma, se hace conveniente categorizar cada una de las entradas

y subentradas del diccionario de las cuales se podrán obtener datos para futuras investigaciones que incluyan el resto de unidades léxicas pluriverbales —mayores a 2 tokens— que no se encuentran etiquetadas dentro del recurso lexicográfico. Este análisis probablemente pueda llevarse a cabo utilizando los patrones morfosintácticos extraídos tanto de las entradas y subentradas como de cada uno de sus equivalentes.

3.3. Herramientas y recursos

Para el desarrollo del presente trabajo se utilizaron herramientas de etiquetaje léxico, tratamiento textual, expresiones regulares, compilación de datos y análisis estadístico partiendo del corpus obtenido del DCI.

Para la extracción de los patrones morfosintácticos se hizo necesario que cada una de las entradas y subentradas a tratar se encontraran en un archivo de texto plano separando cada uno de los términos.

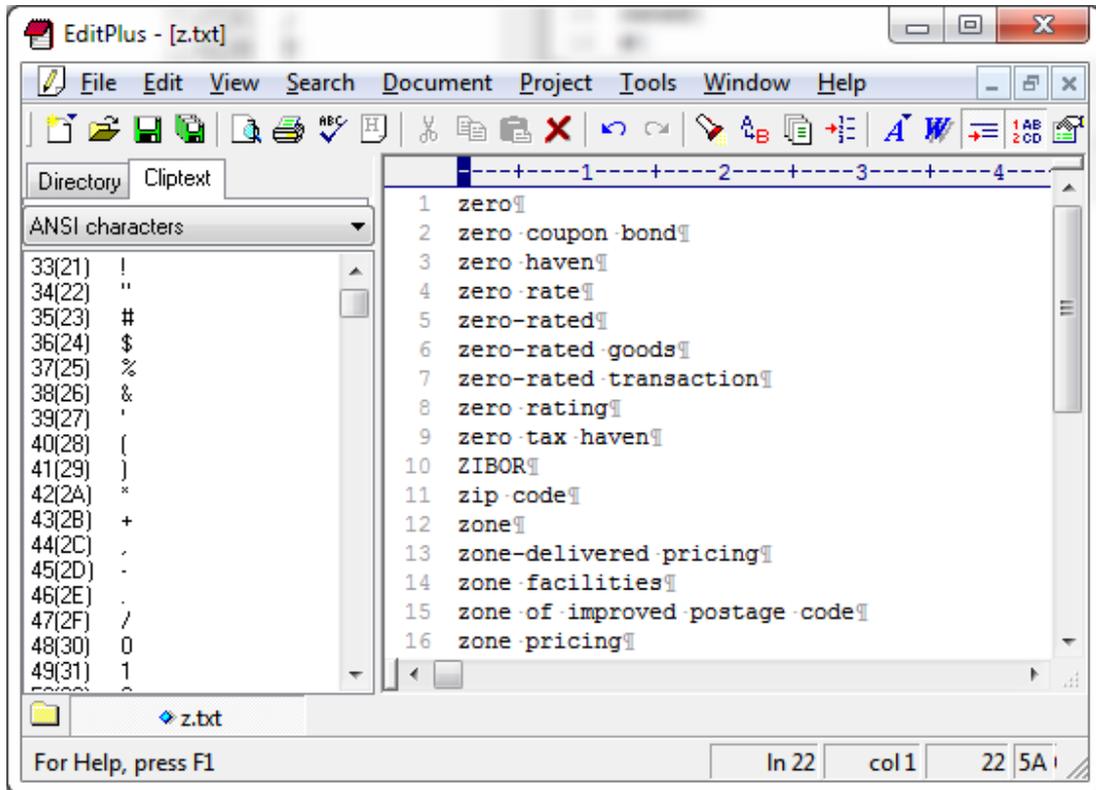


Figura 2: Procesamiento de entradas en editor de texto por expresiones regulares

La inclusión de los separadores se hizo con expresiones regulares haciendo que entre cada entrada o subentrada se incluyera un separador (#), éste permitirá luego, al momento de extraer los patrones morfosintácticos de cada unidad lexicográfica, que no se obtenga ruido de los mismos.



Figura 3: Expresiones regulares en *EditPlus*

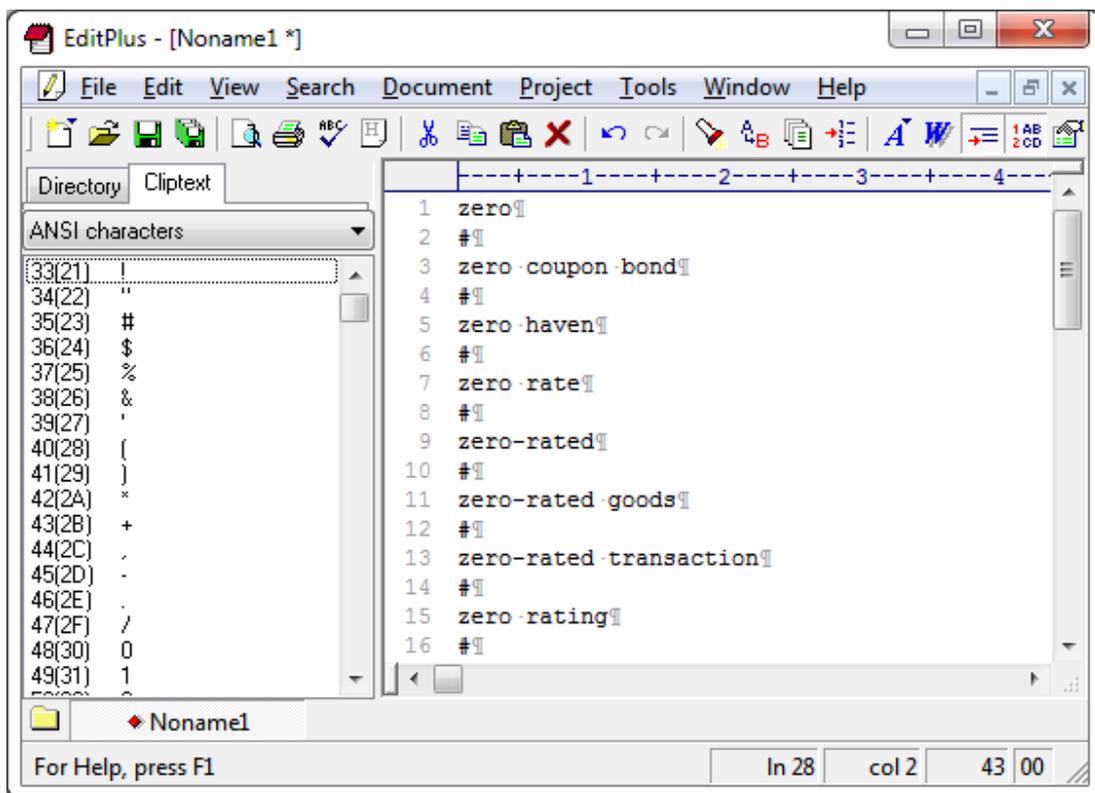


Figura 4: Entradas con separador de línea

Para tokenizar el archivo se empleó un software desarrollado por Antonio Tamayo, enmarcado en el proyecto de investigación de Recuperación semiautomática de unidades fraseológicas especializadas del área del TLC Colombia-EUA en español e inglés⁶. El tokenizador nos permitió dos cosas: la primera, separar cada una de las palabras que conforman las unidades pluriverbales, ubicando cada uno de estos elementos en una línea, la segunda, la separación de las unidades con guiones para que sean luego leídas como unidades individuales.

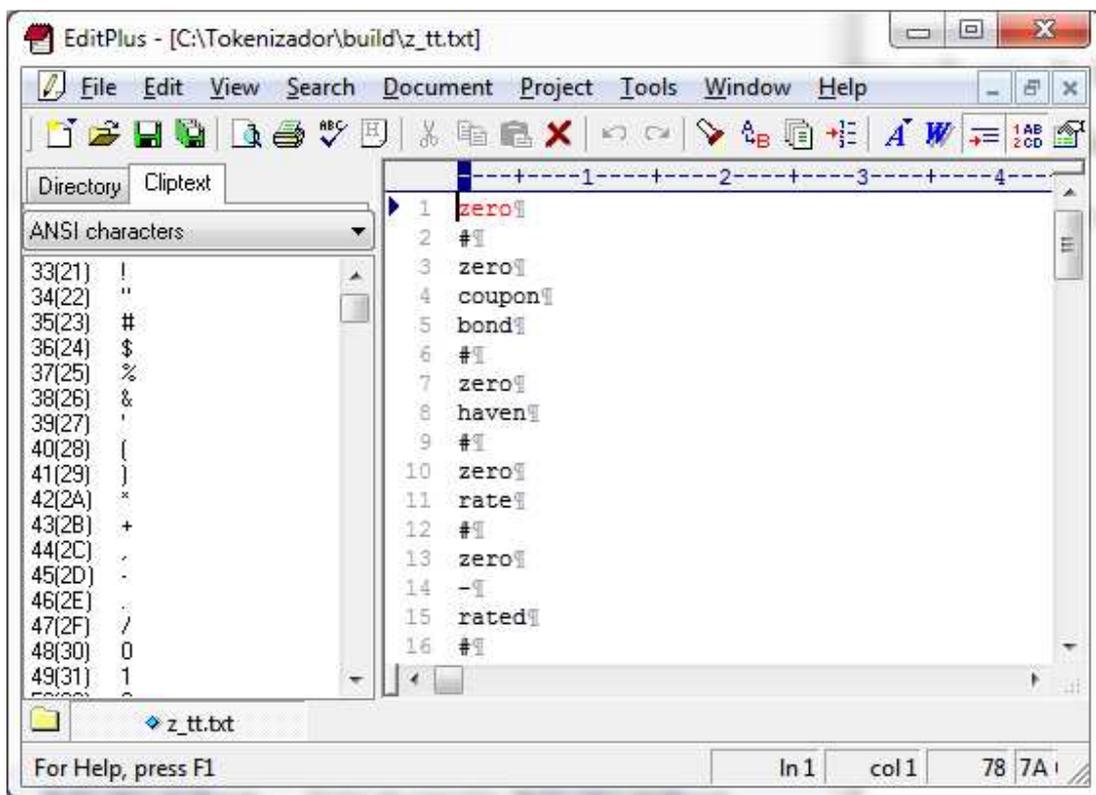


Figura 5: Tokenización de las entradas para la generación de patrones morfosintácticos

⁶ Grupo de investigación TNT (Traducción y Nuevas Tecnologías). Escuela de Idiomas, Universidad de Antioquia.

El paso a seguir implica la utilización de una herramienta desarrollada en la Universidad Stuttgart llamada *TreeTagger*⁷, ésta generará un archivo nuevo que nos ofrecerá las categorías gramaticales de cada uno de los elementos compositivos de las entradas y subentradas, lo cual luego convertiremos en patrones morfosintácticos. Se empleó para ello la interfaz gráfica⁸ para *MS Windows* desarrollada por Ciarán Ó Duibhín.

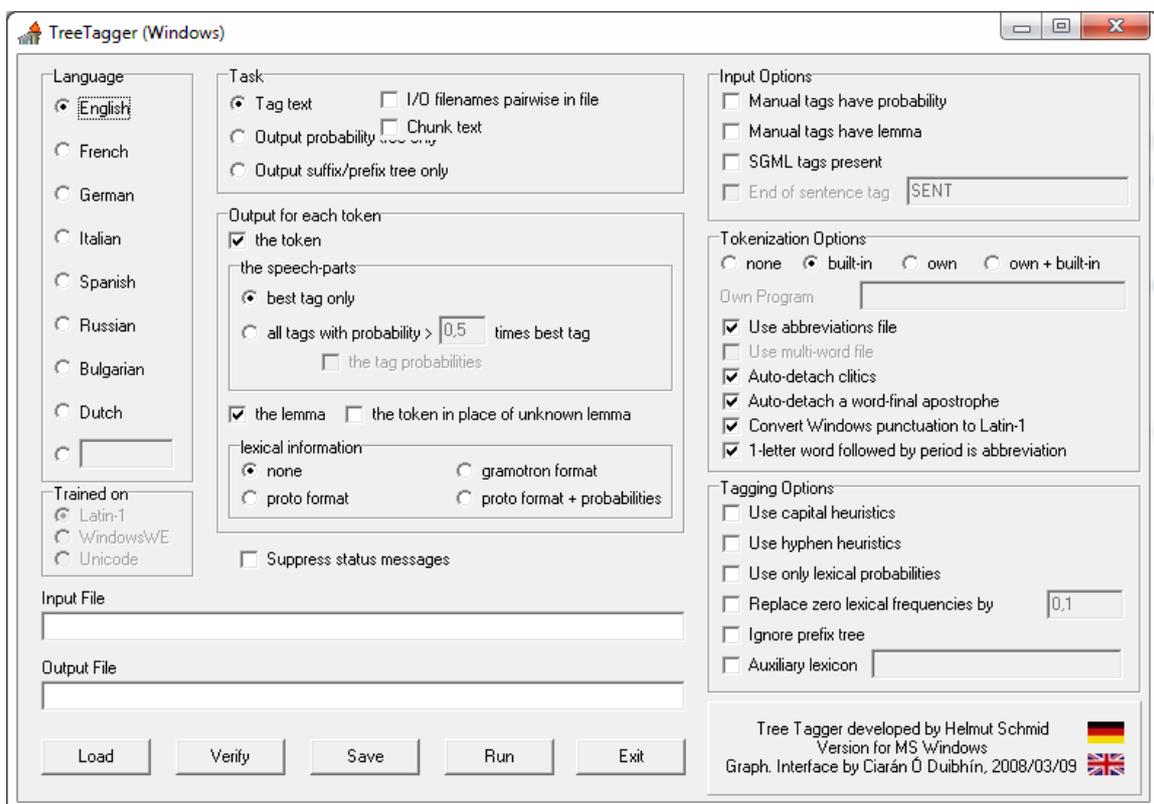


Figura 6: Interfaz gráfica de TreeTagger para Windows

⁷ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

⁸ <http://www.smo.uhi.ac.uk/~oduibhin/oideasra/interfaces/winttinterface.htm>

El archivo resultante de este proceso contendrá 3 columnas: la primera presentará la forma de cada uno de los tokens de las unidades tratadas, la segunda será la categoría gramatical de cada una de las formas y por último los lemas de esas mismas unidades.

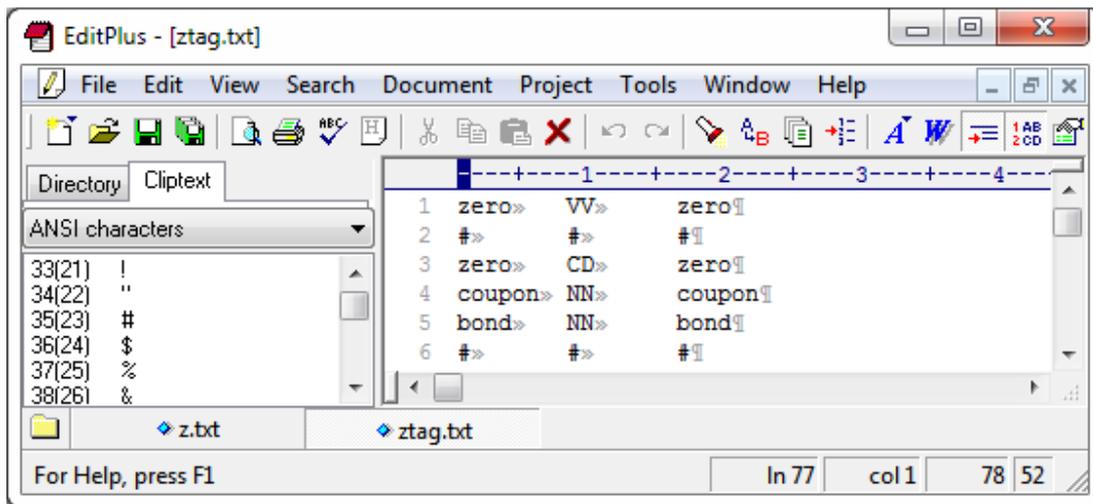


Figura 7: Resultado del etiquetaje de las palabras con TreeTagger

Es observable cómo los separadores no generan ruido al interior de las cadenas morfosintácticas, luego serán estos mismos los que permitirán conformar correctamente cada una de las mismas; sólo se empleará la segunda columna para este proceso, para ello se utilizará entonces una hoja de cálculo para separar cada una de las columnas y poderlas copiar en un archivo nuevo.

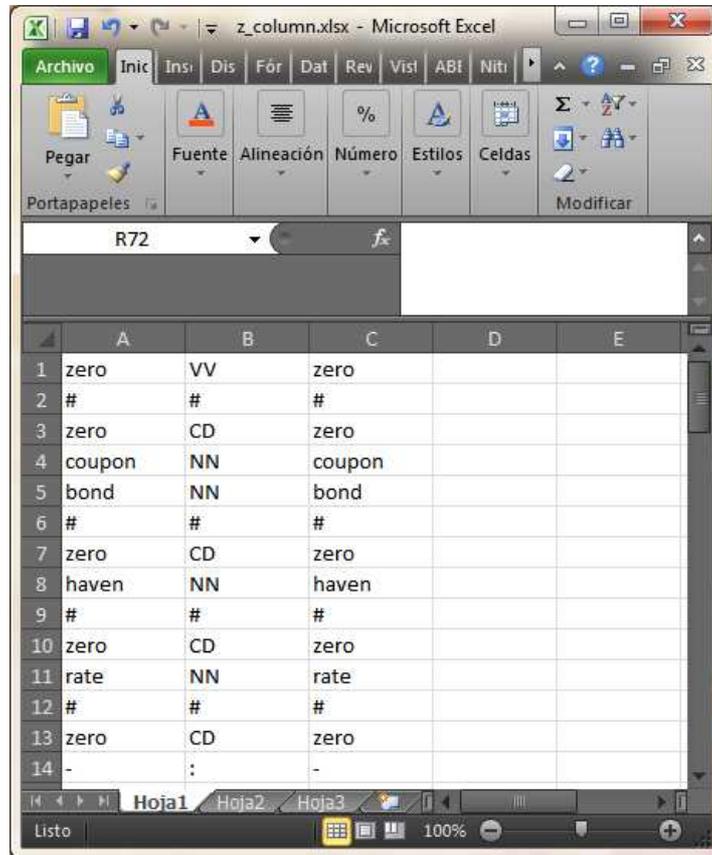


Figura 8: Selección de columna con etiquetas para patrones morfosintácticos

Este nuevo archivo se reorganiza haciendo que cada una de las categorías gramaticales quede una al lado de la otra conservando los separadores.

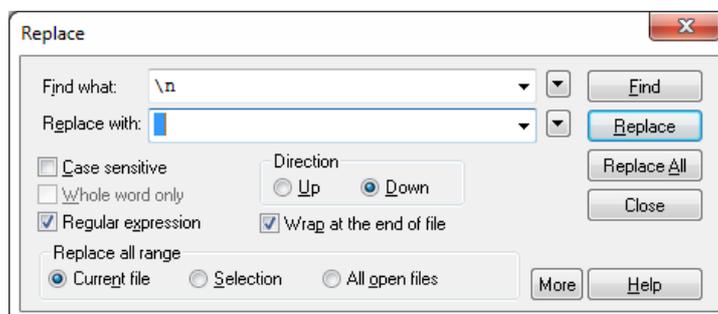


Figura 9: Eliminar espacios utilizando expresiones regulares

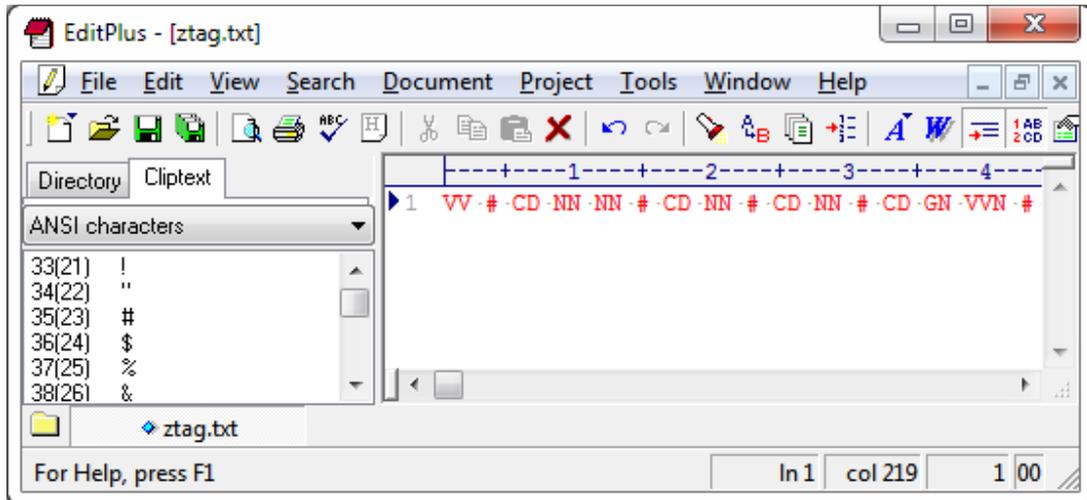


Figura 10: Cadena de patrones morfosintácticos con separador

El último paso de este proceso de extracción de patrones sería cerrar cada una de las cadenas valiéndonos de los separadores y una expresión regular.

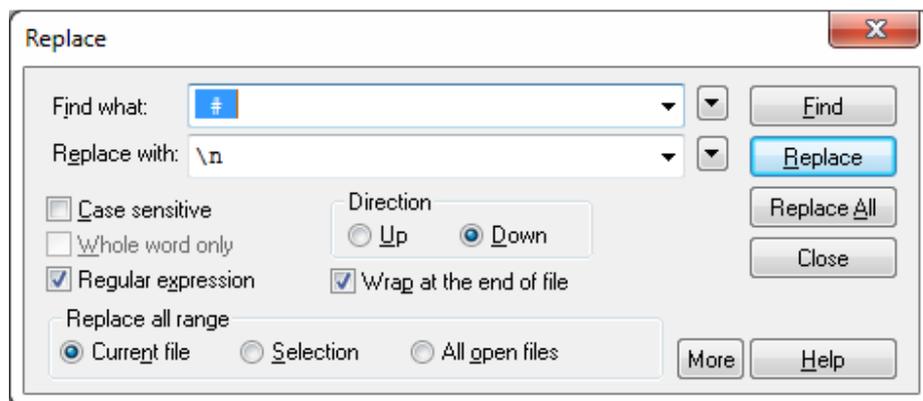


Figura 11: Expresión regular para eliminar separadores

Obteniendo al final un archivo con cada uno de los patrones morfosintácticos de las unidades tratadas, esta parte del corpus nos permitirá luego determinar el nivel de frecuencia de cada uno de los patrones morfosintácticos.

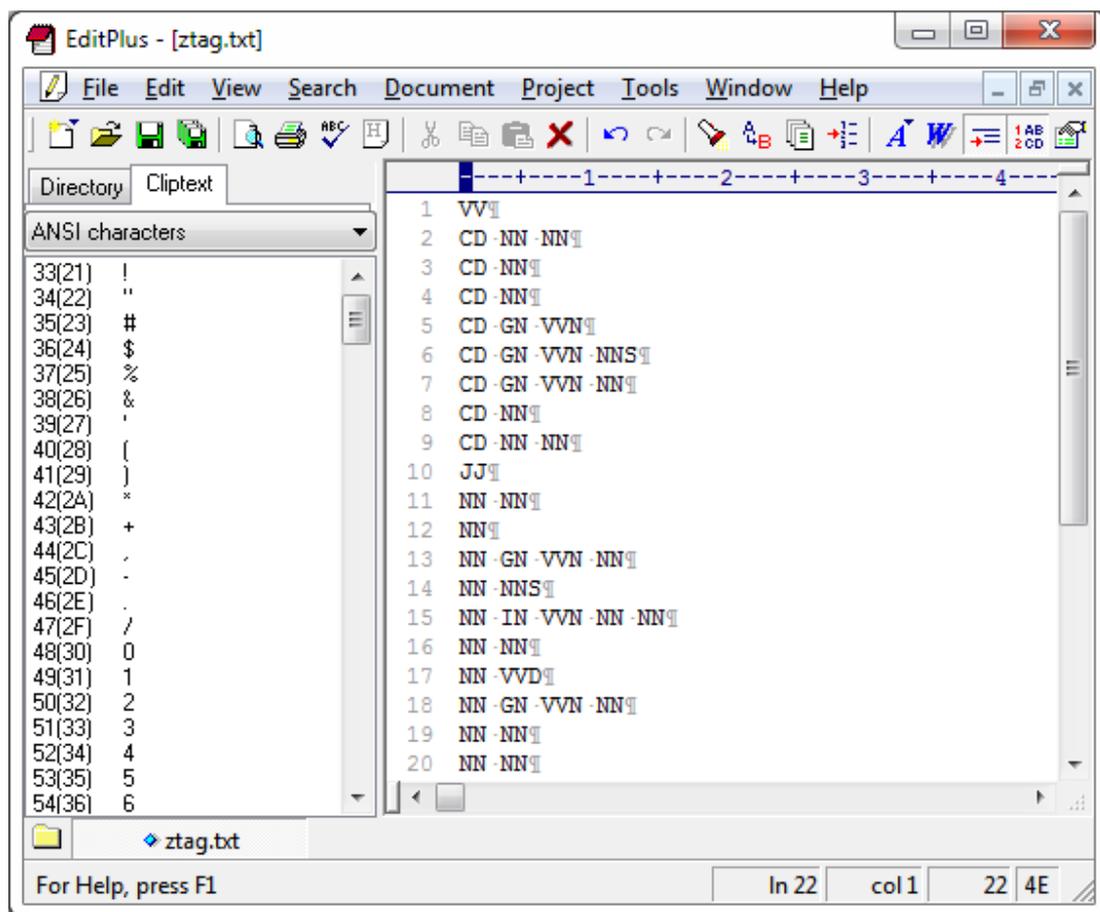


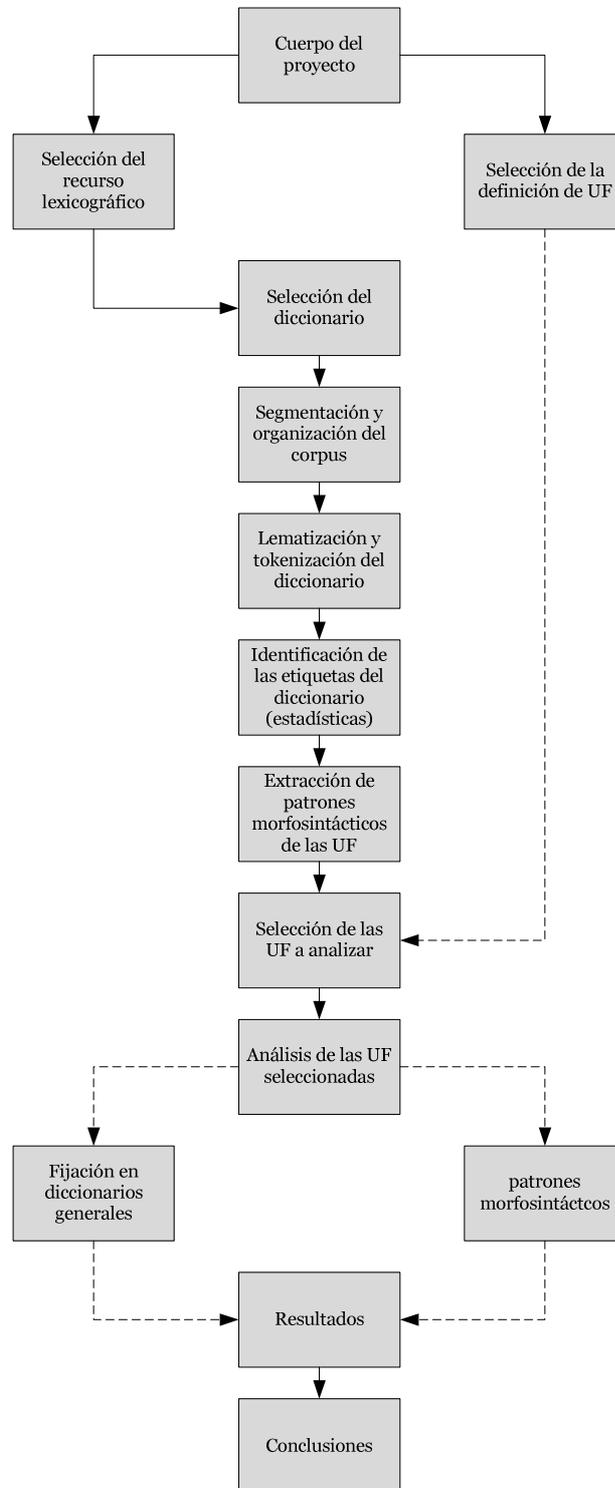
Figura 12: Resultado de los patrones morfosintácticos

El resto de procesos se llevaron a cabo dentro de la hoja de cálculo, organizando la información en cada una de las columnas, ya antes mencionadas. Todas estas informaciones permitían filtrar esta información para separar las entradas por cada uno de los factores que se pudieran categorizar, por ejemplo, utilizando este filtro, se podía

hacer que la hoja de cálculo sólo mostrara aquellas unidades que poseían la marca gramatical *fr*, esto nos permite centrar nuestro análisis en esa porción de unidades.

Es de resaltar, en todo caso, que los patrones morfosintácticos debieron ser homogenizados para permitir un análisis más claro en apartes posteriores de este trabajo, éstos ya no necesitan tener una tabla interpretativa con valores distintos para inglés y español y así hacer legibles los mismos a cualquiera, sea o no un experto en el área. Esto será visible más adelante en las estadísticas del trabajo.

3.4. Metodología de análisis general



3.5. Extracción de las UFE y tratamiento de los datos

Los errores de marcaje encontrados durante el análisis de las informaciones listadas en el corpus se encuentran en un subcapítulo en el aparte de hallazgos y conclusiones, estos, igualmente, son la introducción a las expresiones marcadas como UF.

Como ya lo hemos dicho, el corpus compilado en una tabla Excel nos permitía extraer las expresiones a analizar simplemente activando el filtro de selección permitiendo esta acción, que el programa sólo nos muestre las expresiones marcadas así en el diccionario.

Por último sólo quedaría contrastar las unidades extraídas con los criterios seleccionados en el aparte 2.6.1. del presente trabajo como validadores de las expresiones marcadas como *fr* en el diccionario, es decir, con base en estos criterios definiremos si éstas se encuentran marcadas de forma adecuada.

3.6. Aspectos estadísticos

Los resultados arrojados por las estadísticas tomadas de las distintas muestras analizadas nos permiten ser más sistemáticos y rigurosos en términos de evaluación del diccionario y por ende, del mismo corpus. El haber realizado el análisis del corpus de forma inductiva y mostrar todos los resultados de forma estadística facilitan no sólo la

descripción de los problemas sino también las características escogidas para ser observadas.

La utilización de un software estadístico permite comprobar por recuentos numéricos de frecuencias y apariciones que la información proporcionada y las posibles conclusiones sacadas de allí, puedan ser corroboradas por quienes pueden encontrar este trabajo útil para el estudio de alguno de los fenómenos tratados o afines en el presente trabajo.

La validación de algunos de los criterios de selección de unidades fraseológicas fue elaborado por medio de consultas en la web tomando de allí el número de frecuencia de apariciones de las expresiones tal cual fue hecho por Burgos (2006, p. 355) empleando el motor de búsqueda de Google para lograr este cometido.

4. Presentación y descripción de datos estadísticos

4. PRESENTACIÓN Y DESCRIPCIÓN DE DATOS ESTADÍSTICOS	91
4.1. Introducción	93
4.2. DESCRIPCIÓN Y ANÁLISIS ESTADÍSTICO EN INGLÉS	93
4.2.1. Unidades con o sin POS	94
4.2.2. Descripción de marcaje de entradas con POS	95
4.2.3. Descripción de marcas simples	98
4.2.4. Descripción de la marca <i>fr</i>	104
4.2.4.1. Conteo de tokens de las expresiones marcadas como <i>fr</i>	109
4.2.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como <i>fr</i>	110
4.2.5. Descripción de las unidades sin marca de POS (subentradas)	114
4.2.5.1. Conteo de tokens en las subentradas	115
4.2.5.2. Expresiones pluriverbales que contienen comas	116
4.2.5.3. Expresiones pluriverbales que contienen guiones	119
4.2.5.4. Expresiones pluriverbales que contienen comas y guiones	121
4.2.5.5. Expresiones pluriverbales que no contienen comas o guiones	121
4.3. DESCRIPCIÓN Y ANÁLISIS ESTADÍSTICO EN ESPAÑOL	123
4.3.1. Unidades con o sin POS	124

4.3.2. Descripción de marcaje de entradas con POS	125
4.3.3. Descripción de marcas simples	126
4.3.4. Descripción de la marca <i>phr</i>	132
4.3.4.1. Conteo de tokens de las expresiones marcadas como <i>phr</i>	132
4.3.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como <i>phr</i>	132
4.3.5. Descripción de las unidades sin marca de POS (subentradas)	135
4.3.5.1. Conteo de tokens en las subentradas	135
4.3.5.2. Expresiones pluriverbales que contienen comas	136
4.3.5.3. Expresiones pluriverbales que contienen guiones	139
4.3.5.4. Expresiones pluriverbales que contienen comas y guiones	142
4.3.5.5. Expresiones pluriverbales que no contienen comas o guiones	142

4.1. Introducción

Las estadísticas que presentamos a continuación contendrán la información numérica básica en lo que se refiere a marcaje gramatical, marcaje diasistemático y fraseológico, de igual forma hemos decidido incluir estadísticas con respecto al número de tokens tanto de las unidades etiquetadas como de las que no lo están. Por último hemos hecho un listado de los patrones morfosintácticos presentes en el corpus trabajado.

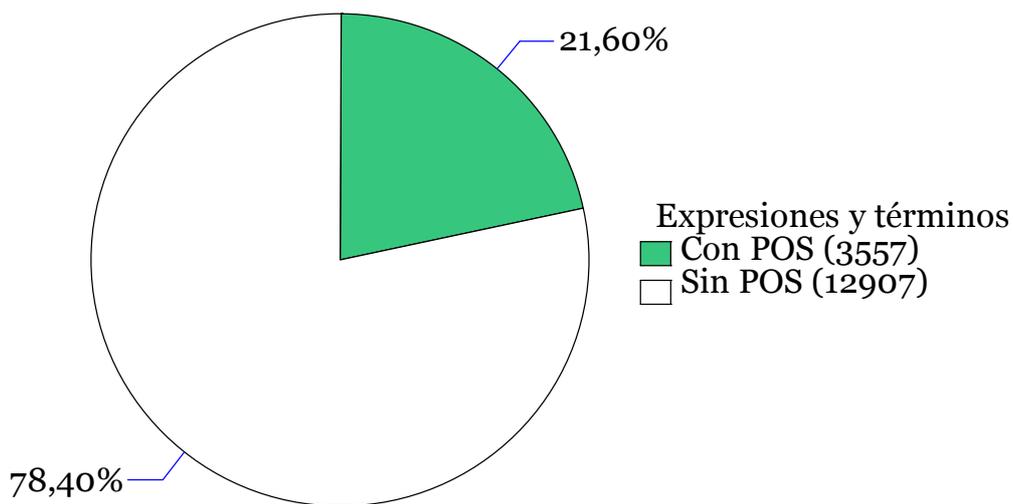
Debemos tener en cuenta que nuestro diccionario al ser un diccionario bilingüe –inglés-español–, tiene un análisis estadístico pertinente a cada uno de sus apartes, por esto hemos dividido la presentación de esta información en dos: una parte donde mostramos los datos de las entradas en inglés y una segunda parte donde se muestran los datos de las entradas en español.

4.2. Descripción y análisis estadístico en inglés

Como ya lo habíamos mencionado, hemos decidido dividir las estadísticas en inglés y español, debido a los fenómenos morfosintácticos que se percibieron durante la descripción de los hallazgos lingüísticos. Esta primera parte –en inglés– tendrá muchos más datos que la segunda, por la incorporación de la marca *fr* de algunas unidades pluriverbales que fueron sometidas a estrategias de reducción léxica –acrónimos, siglas, inicialismos o unidades híbridas–. Explicaremos esto a profundidad más adelante.

4.2.1. Unidades con o sin POS

El DCI clasifica claramente sus entradas y subentradas, mientras las primeras tienen marca gramatical, las segundas no la poseen y se restringen a tener –no todas– alguna marca diatópica. La marca gramatical es entonces un buen elemento para distinguir dentro de nuestro corpus las entradas de las subentradas del mismo. Mostramos a continuación la gráfica y su correspondiente tabla de datos:



Gráfica 1: Expresiones con y sin POS

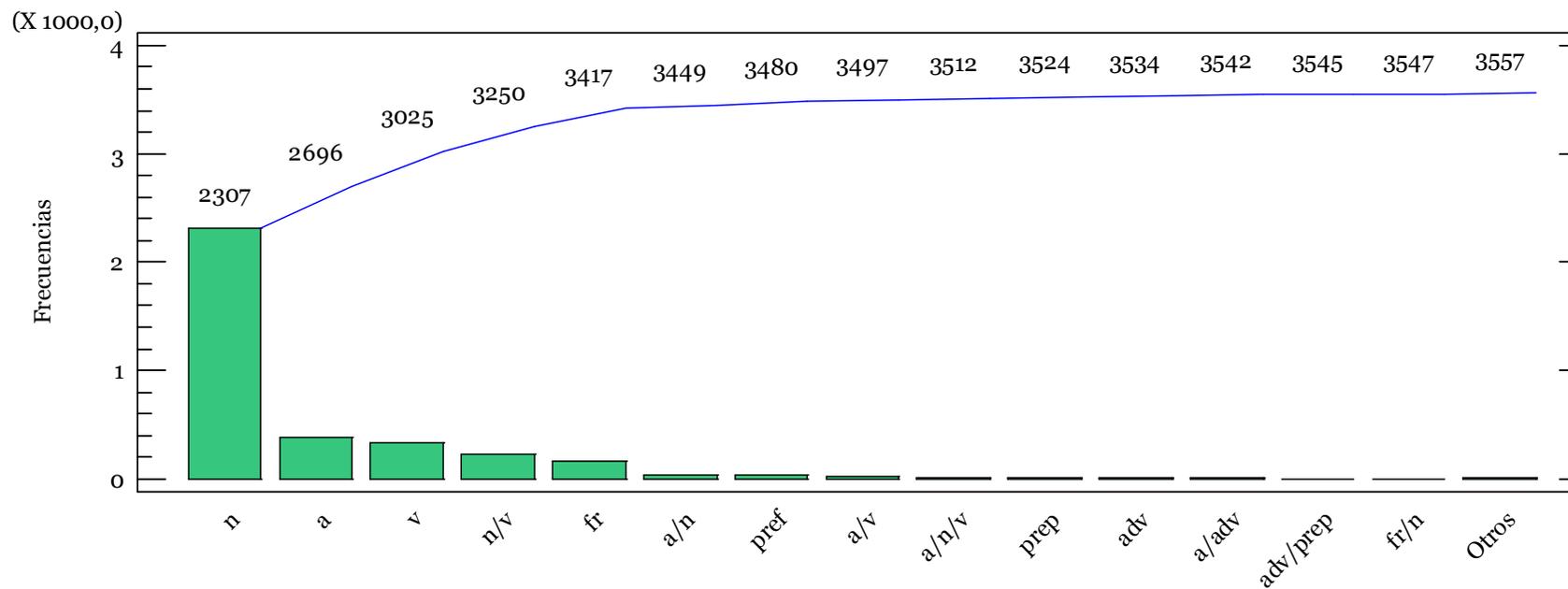
<i>Clase</i>	<i>Valor</i>	<i>Frecuencia</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. acum.</i>
1	Con POS	3557	3557	0,2160
2	Sin POS	12907	16464	1,0000

Tabla 7: Frecuencia de expresiones con y sin marca gramatical

Esta tabla muestra el número de veces que se ha presentado cada valor de POS vs NO POS así como porcentajes y estadísticas acumuladas. Por ejemplo, en 3557 filas del archivo de datos, POS vs NO POS es igual a POS. Esto representa 21,6047% de las 16464 expresiones en el corpus. Las dos columnas de la extrema derecha dan los recuentos y porcentajes acumulados, desde el inicio de la tabla hacia abajo, esas 3557 expresiones corresponderían, de igual forma, al número de entradas del DCI mientras las 12907 expresiones serían las subentradas.

4.2.2. Descripción de marcaje de entradas con POS

A continuación presentamos la gráfica y la tabla de cada una de las etiquetas gramaticales, no fueron tenidas en cuenta para la gráfica aquellas etiquetas que tuvieran menos de dos frecuencias en todo el diccionario, sin embargo, éstas mismas son descritas en la tabla que se presenta junto con la gráfica.



Gráfica 2: Frecuencia de etiquetas en entradas de 2 o más frecuencias

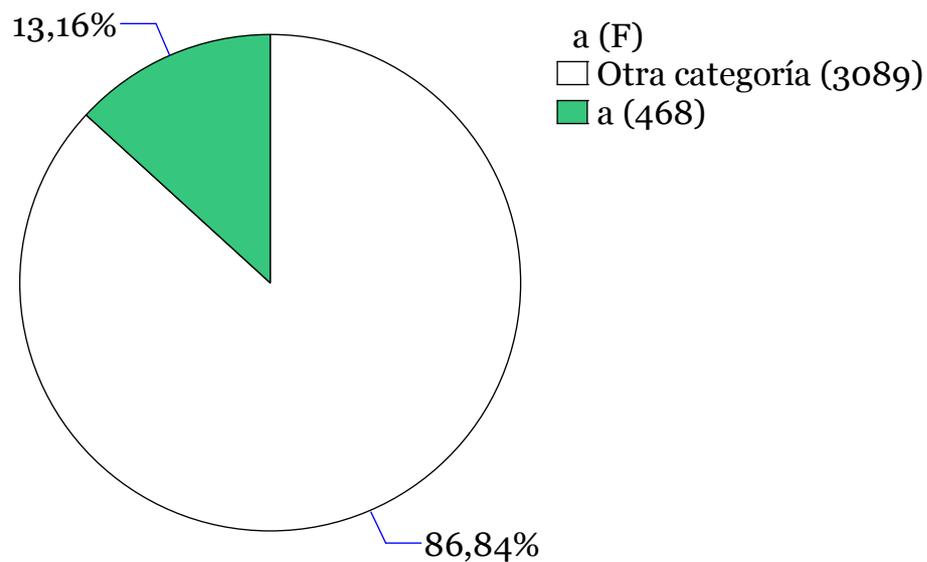
<i>Etiqueta de Clase</i>	<i>Rango</i>	<i>Recuento</i>	<i>Puntaje Acum</i>	<i>Porcentaje</i>	<i>Porcentaje Acum</i>
N	1	2307	2307	64,86	64,86
A	2	389	2696	10,94	75,79
V	3	329	3025	9,25	85,04
n/v	4	225	3250	6,33	91,37
Fr	5	167	3417	4,69	96,06
a/n	6	32	3449	0,90	96,96
Pref	7	31	3480	0,87	97,84
a/v	8	17	3497	0,48	98,31
a/n/v	9	15	3512	0,42	98,73
Prep	10	12	3524	0,34	99,07
Adv	11	10	3534	0,28	99,35
a/adv	12	8	3542	0,22	99,58
adv/prep	13	3	3545	0,08	99,66
fr/n	14	2	3547	0,06	99,72
prep/a	15	1	3548	0,03	99,75
n/a/pron	16	1	3549	0,03	99,78
Interj	17	1	3550	0,03	99,80
conj/prep/v	18	1	3551	0,03	99,83
Conj	19	1	3552	0,03	99,86
a/pron/conj	20	1	3553	0,03	99,89
a/pron	21	1	3554	0,03	99,92
a/prep/v	22	1	3555	0,03	99,94
a/adv/v	23	1	3556	0,03	99,97
a/adv/n/v	24	1	3557	0,03	100,00
Total		3557			

Tabla 8: Frecuencias y porcentajes de etiquetas gramaticales

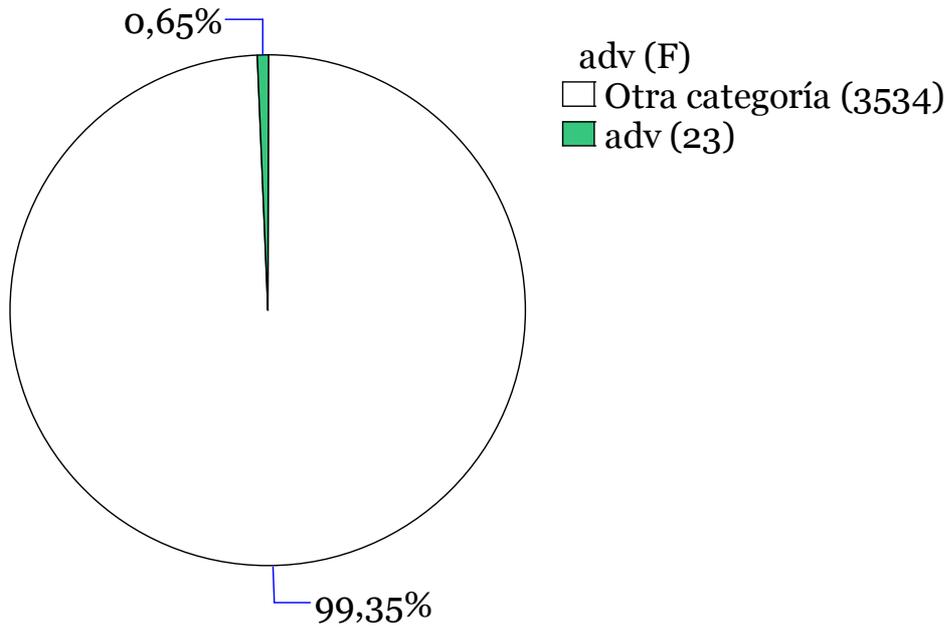
Esta tabla muestra la frecuencia de ocurrencia de los 24 valores únicos de las etiquetas unificadas, es decir, aquellos casos en los que se presentan múltiples formas, aunque ello podría identificar si son derivados unos de otros –esto no se puede comprobar debido a la inexistencia de un manual de uso del diccionario–. Hemos decidido por tanto unificar criterios, por ejemplo: al aparecer la siguiente combinación de etiquetas: “a/v/n”, “a/n/v” “n/a/v” se escogió una sola forma para contarlas –en este caso la que mayor número de frecuencia tuviera–. Las clases se ordenan de acuerdo con los recuentos, poniendo primero la clase con mayor frecuencia de ocurrencia.

4.2.3. Descripción de marcas simples

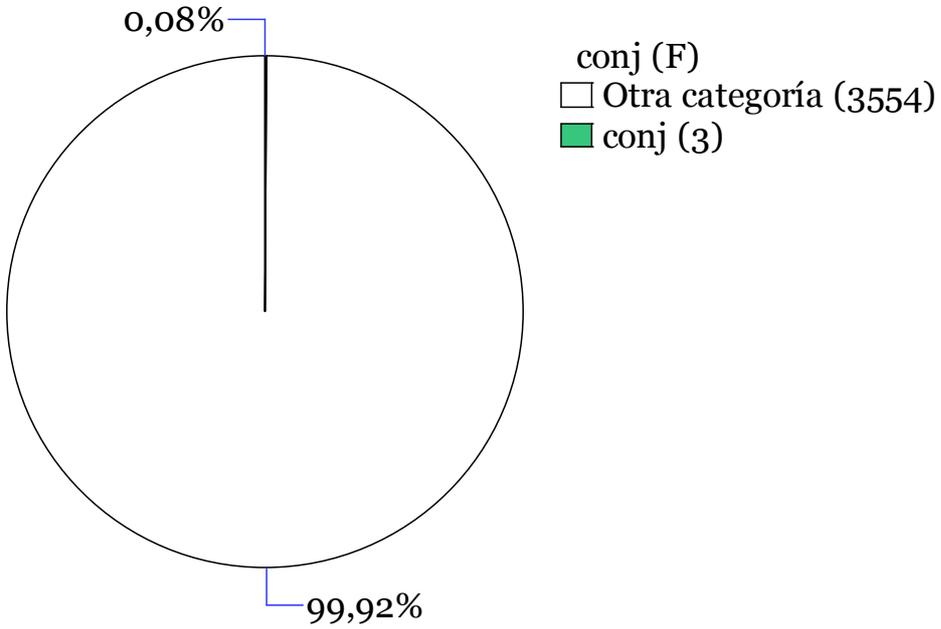
Tomamos luego la decisión de saber el número de adjetivos, adverbios, conjunciones, interjecciones, prefijos, preposiciones, pronombres, sustantivos, verbos y aquellas expresiones marcadas como *fr* que suponemos pueden ser frase, o frasema, el número de frecuencias se mide con respecto al número de entradas, es decir, aquellas expresiones que presentan marca gramatical. Los resultados se describen gráficamente a continuación:



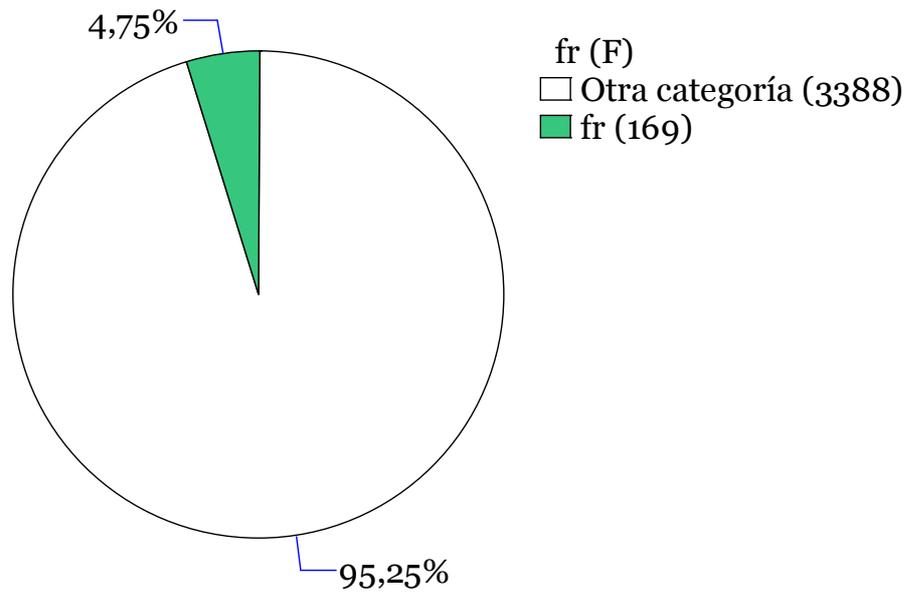
Gráfica 3: Frecuencia de adjetivos (a) en las entradas



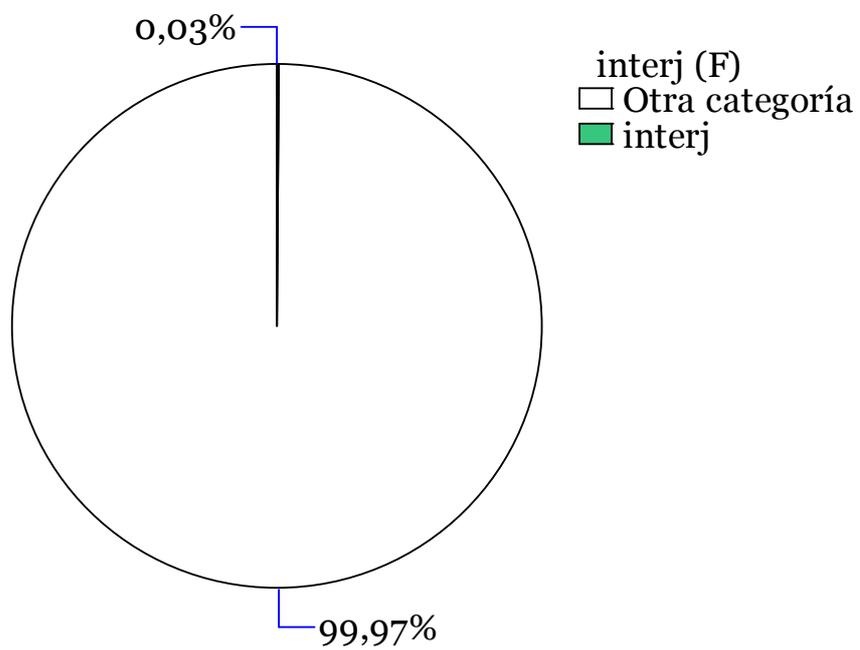
Gráfica 4: Frecuencia de adverbios (adv) en las entradas



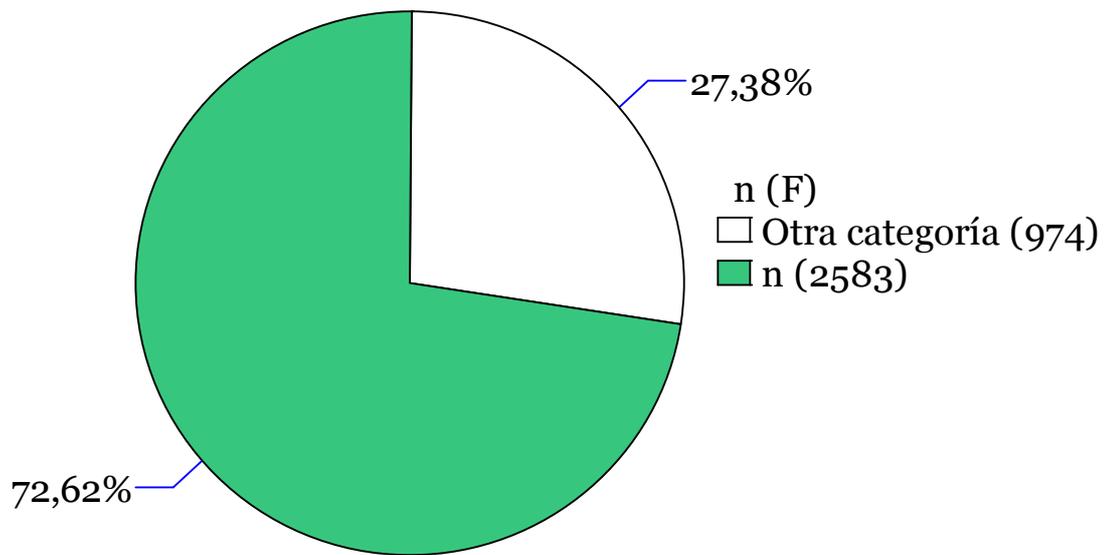
Gráfica 5: Frecuencia de conjunciones (conj) en las entradas



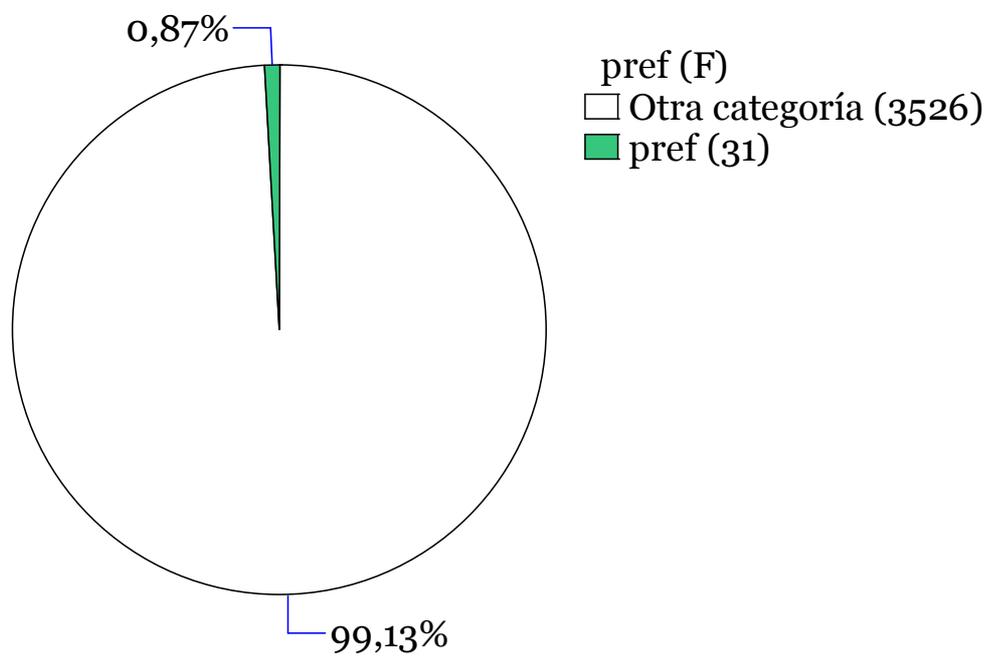
Gráfica 6: Frecuencia de entradas con la marca *fr* en las entradas



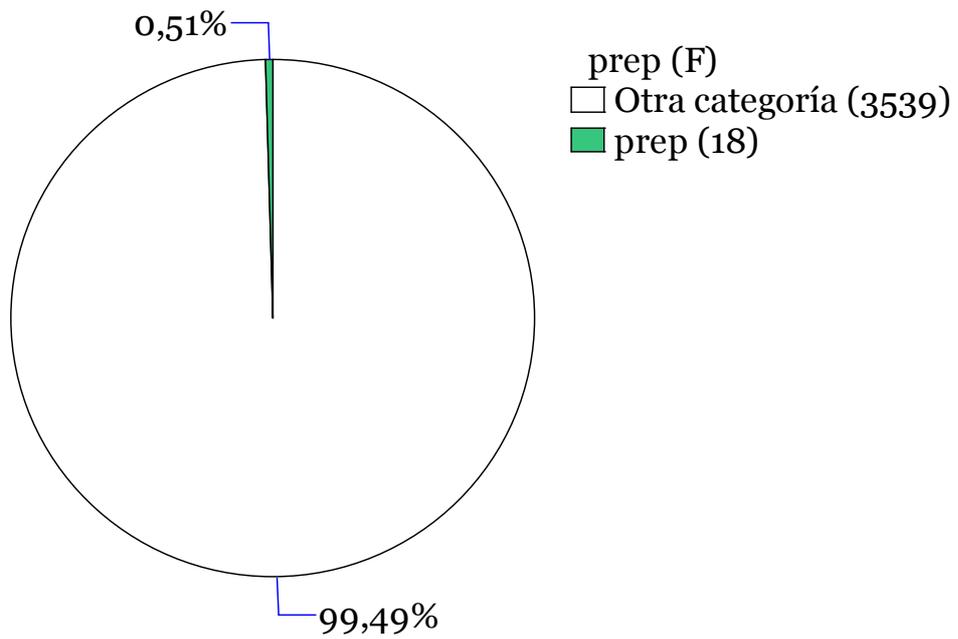
Gráfica 7: Frecuencia de interjecciones (interj) en las entradas



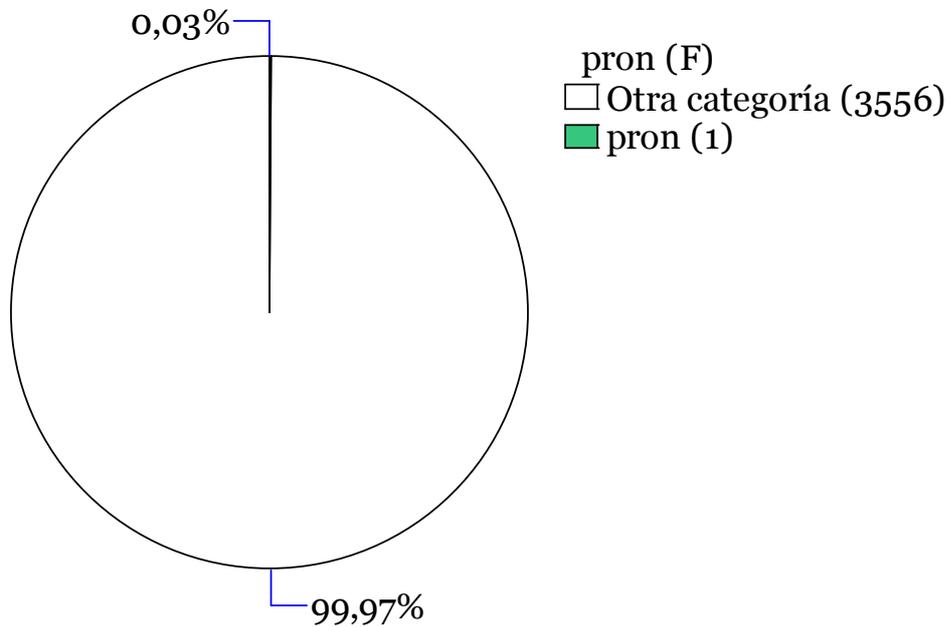
Gráfica 8: Frecuencia de sustantivos (n) en las entradas



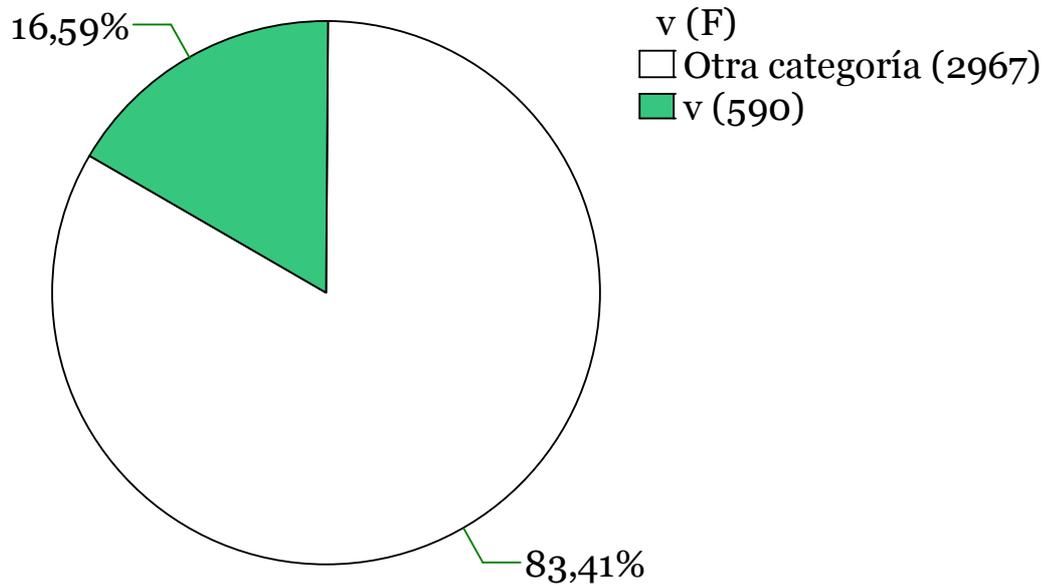
Gráfica 9: Frecuencia de prefijos (pref) en las entradas



Gráfica 10: Frecuencia de preposiciones (prep) en las entradas



Gráfica 11: Frecuencia de pronombres (pron) en las entradas



Gráfica 12: Cono de los verbos (v) en las entradas

La tabla que presentamos a continuación sintetiza las anteriores mostrando, de forma tal, que la información pueda ser validada por medio de la siguiente tabulación de datos.

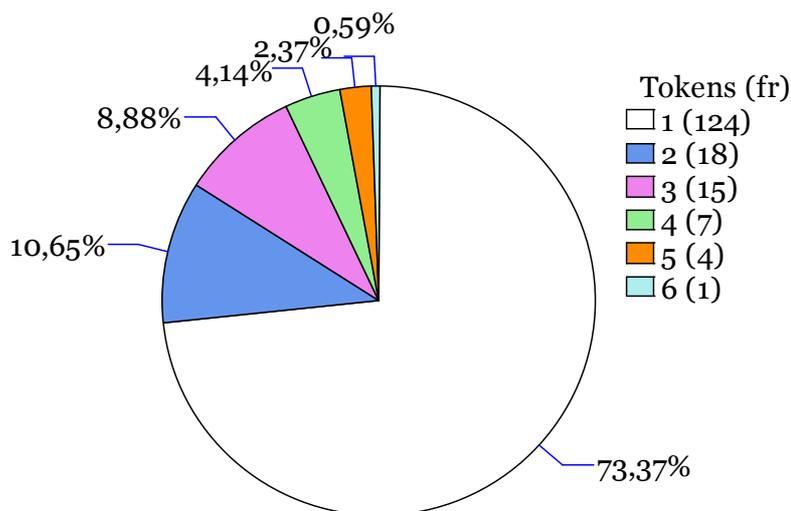
<i>Clase</i>	<i>Valor</i>	<i>Frecuencia</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Rel. acum.</i>
1	a	468	0,1204	0,1204
2	adv	23	0,0059	0,1263
3	conj	3	0,0008	0,1271
4	fr	169	0,0435	0,1706
5	interj	1	0,0003	0,1708
6	n	2583	0,6645	0,8353
7	pref	31	0,0080	0,8433
8	prep	18	0,0046	0,8480
9	pron	1	0,0003	0,8482
10	v	590	0,1518	1,0000

Tabla 9: Frecuencias para marcas gramaticales simples

4.2.4. Descripción de la marca *fr*

En este punto de la extracción de datos estadísticos fue donde tomamos la decisión de limitar nuestro corpus de análisis. Ya teníamos en inglés unos posibles candidatos a ser analizados, pero fue también este conteo de datos de frecuencias el que nos permitió observar uno de los fenómenos de inconsistencia teórico-metodológico en la creación de este recurso lexicográfico.

Se puede comprobar al ver la tabla 9 que teníamos 169 entradas marcadas como *fr*, sin embargo, y como ya lo hemos dicho, la carencia de un manual de uso del diccionario no nos permite asegurar que esta marca hiciera referencia precisamente a un frasma, UF o UFE, teníamos que analizar –en una primera instancia– la composición interna de cada una de esas unidades marcadas. Las UF tienen todas una característica en común y es la pluriverbalidad, es decir, que estas se conforman de 2 o más tokens. Contamos entonces los tokens de cada expresión marcada como *fr* y los resultados fueron los siguientes:

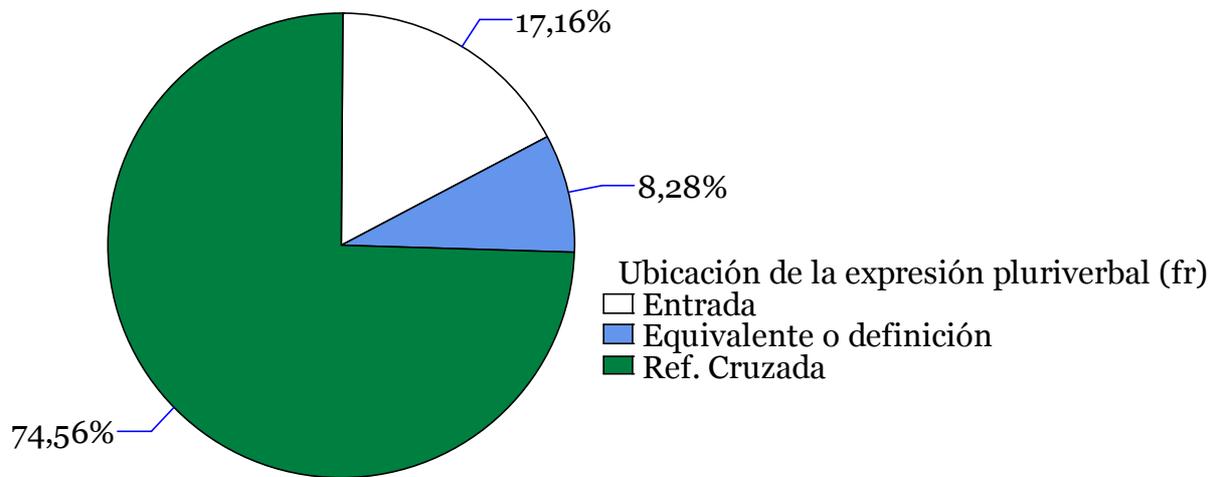


Gráfica 13: Frecuencia de las entradas marcadas como *fr* por tokens

Vemos entonces que 124 de estas unidades aparecen como estructuras de un sólo token, al verificar esta información manualmente con el texto físico evidenciamos que no era ningún error –por lo menos en el proceso de digitalización de los datos– pero sí, un error metodológico al ver que dichas unidades fueron marcadas como *fr* cuando en realidad eran unidades que habían sufrido alguna reducción léxica, en otras palabras estábamos frente a 124 acrónimos, siglas o inicialismos.

Al verificar cada una de las unidades nos dimos cuenta que al utilizar el *TreeTagger* todas las unidades recibían una marca de pertenecer a la categoría nominal –ya fuera NN o NP– es decir, ninguna era un frasema. Decidimos entonces hacer un rastreo más profundo no sólo de las unidades con un solo token sino de todas, y descubrimos que las estadísticas de la marca *fr* tenían ruido debido a que muchas otras expresiones de dos y tres tokens tenían dentro de sí una expresión pluriverbal que había sido reducida.

Sin embargo pudimos ver que las expresiones están en el diccionario pero ubicadas no sólo en la entrada, sino también en otros sectores del artículo lexicográfico ya fuera como parte de una referencia cruzada o dentro de la definición o equivalente de la entrada. La gráfica que a continuación presentamos, muestra la ubicación de cada una de las expresiones marcadas como *fr*.



Gráfica 14: Ubicación de la expresión pluriverbal completa marcada como *fr*

Después de verificar cada una de las expresiones clasificamos la información de dos formas, primero, dónde se encontraban ubicadas éstas y segundo, realizamos un nuevo conteo de tokens. Fue evidente observar que, en algunos casos, se hallaban tanto el acrónimo como la expresión pluriverbal, es decir, ambas estaban indizadas en el diccionario como entradas con su respectiva marca. Dos preguntas surgen en este punto del análisis y son: ¿qué criterio se tuvo en cuenta para la indización de estas unidades que presentaban tanto su forma completa como la reducida? ¿Por qué no se hizo lo mismo con todas?

Sin embargo, seguía una incógnita y estaba relacionada sobre la distribución de las partes compositivas de las expresiones con un solo token. Presentamos a continuación una tabla con las expresiones y su marcación original al igual que el marcaje que proporcionó el *TreeTagger* y luego una gráfica que ilustra los datos estadísticos de las mismas expresiones sin ninguna estrategia de reducción léxica.

#	Expresión	Marcaje DCI	Marcaje TreeTagger
1	a/d	<i>fr:</i>	NN
2	AOV	<i>fr:</i>	NN
3	a/s	<i>fr:</i>	NN
4	A/S	<i>fr:</i>	NN
5	ASPW	<i>fr:</i>	NN
6	ASWP	<i>fr:</i>	NN
7	BACAT	<i>fr:</i>	NN
8	BBB	<i>fr:</i>	NN
9	Bdi	<i>fr:</i>	NN
10	BENDS	<i>fr:</i>	NN
11	bi	<i>fr:</i>	NN
12	damfordet	<i>fr:</i>	NN
13	daps	<i>fr:</i>	NN
14	DBEATS	<i>fr:</i>	NN
15	DBELTS	<i>fr:</i>	NN
16	DHDATSBE	<i>fr:</i>	NN
17	DHDWTSBE	<i>fr:</i>	NN
18	D/s	<i>fr:</i>	NN
19	E&OE	<i>fr:</i>	NN
20	eoHP	<i>fr:</i>	NN
21	ets	<i>fr:</i>	NN
22	F/A	<i>fr:</i>	NN
23	fac	<i>fr:</i>	NN
24	faccop	<i>fr:</i>	NN
25	f&c	<i>fr:</i>	NN
26	fcsrcc	<i>fr:</i>	NN
27	FD	<i>fr:</i>	NN
28	F&D	<i>fr:</i>	NN
29	FDO	<i>fr:</i>	NN
30	FFI	<i>fr:</i>	NN
31	FHINC	<i>fr:</i>	NN
32	fias	<i>fr:</i>	NN
33	FILO	<i>fr:</i>	NN
34	fios	<i>fr:</i>	NN
35	fioSt	<i>fr:</i>	NN
36	fiot	<i>fr:</i>	NN
37	FIB	<i>fr:</i>	NN
38	FOC	<i>n/fr:</i>	NP
39	foc	<i>fr:</i>	NN
40	FOFI	<i>fr:</i>	NN
41	F.o.t	<i>fr:</i>	NN
42	F.o.w	<i>fr:</i>	NN
43	fos	<i>fr:</i>	NN
44	HDLTSBENDS	<i>fr:</i>	NN
45	HDWTS	<i>fr:</i>	NN
46	FHEX.	<i>fr:</i>	NP
47	H.P	<i>fr:</i>	NN
48	ipa	<i>fr:</i>	NN
49	J&WO	<i>fr:</i>	NN
50	KDCL	<i>fr:</i>	NN

#	Expresión	Marcaje DCI	Marcaje TreeTagger
51	KDF	<i>fr:</i>	NN
52	KDLCL	<i>fr:</i>	NN
53	LILO	<i>fr:</i>	NN
54	LTL	<i>fr:</i>	NN
55	M/s	<i>fr:</i>	NN
56	N/a	<i>fr:</i>	NN
57	NCR	<i>phrase:</i>	NN
58	NCV	<i>fr:</i>	NN
59	NEC	<i>fr:</i>	NN
60	nei	<i>fr:</i>	NN
61	nem	<i>fr:</i>	NN
62	nep	<i>fr:</i>	NN
63	nes	<i>fr:</i>	NN
64	Nf	<i>fr:</i>	NN
65	NIA	<i>fr:</i>	NN
66	NLM	<i>fr:</i>	NN
67	NLR	<i>fr:</i>	NN
68	NOE	<i>fr:</i>	NN
69	NOHP	<i>fr:</i>	NN
70	n.o.p.	<i>fr:</i>	NN
71	NOS	<i>fr:</i>	NN
72	NP	<i>fr:</i>	NN
73	nr	<i>fr:</i>	NN
74	nrad	<i>fr:</i>	NN
75	nral	<i>fr:</i>	NN
76	nrAs	<i>fr:</i>	NN
77	nrtb	<i>fr:</i>	NN
78	NSF	<i>fr:</i>	NN
79	NSPF	<i>fr:</i>	NN
80	NVC	<i>fr:</i>	NN
81	NVD	<i>fr:</i>	NN
82	O/o	<i>fr:</i>	NN
83	OS&D	<i>fr:</i>	NN
84	OTOR	<i>fr:</i>	NN
85	PAD	<i>fr:</i>	NN
86	P.H	<i>fr:</i>	NN
87	PKD	<i>fr:</i>	NN
88	P.P.	<i>fr/n:</i>	NN
89	R/A	<i>fr:</i>	NN
90	RCF	<i>fr:</i>	NN
91	R/D	<i>fr:</i>	NN
92	RTBA	<i>fr:</i>	NN
93	RVNX	<i>fr:</i>	NN
94	S&H/exct	<i>fr:</i>	NN
95	SATSHEX	<i>fr:</i>	NN
96	SATSHINC	<i>fr:</i>	NN
97	SC&S	<i>fr:</i>	NN
98	Shex	<i>fr:</i>	NN
99	Shinc	<i>fr:</i>	NN
100	SIT	<i>fr:</i>	NN

#	Expresión	Marcaje DCI	Marcaje TreeTagger
101	SKD	<i>fr:</i>	NN
102	SL&C	<i>fr:</i>	NN
103	SL&T	<i>fr:</i>	NN
104	SSHE	<i>fr:</i>	NN
105	SSHINC	<i>fr:</i>	NN
106	STC	<i>fr:</i>	NN
107	STW	<i>fr:</i>	NN
108	TBA	<i>fr:</i>	NN
109	TBN	<i>fr:</i>	NN
110	T/C	<i>fr:</i>	NN
111	TPND	<i>fr:</i>	NN
112	UUR	<i>fr:</i>	NN
113	wa	<i>fr:</i>	NN
114	WCCON	<i>fr:</i>	NN
115	wef	<i>fr:</i>	NN
116	WIBON	<i>fr:</i>	NN
117	WIFPON	<i>fr:</i>	NN
118	W/M	<i>fr:</i>	NN
119	WP	<i>fr:</i>	NN
120	WPD	<i>fr:</i>	NN
121	wwdshex	<i>fr:</i>	NN
122	WWD	<i>fr:</i>	NN
123	WwdFHEX	<i>fr:</i>	NN
124	wwdhinc	<i>fr:</i>	NN

Tabla 10: Expresiones marcadas como *fr* con un solo token

Nos faltaba volver a hacer el conteo, sin embargo, había expresiones que eran identificadas por dos acrónimos distintos, de éstas fueron eliminados uno de sus ejemplos para preservar la confiabilidad de la muestra dado que lo que nos importaba es observar aquellas expresiones reales de la tabla que aparece a continuación la cual contiene las entradas repetidas.

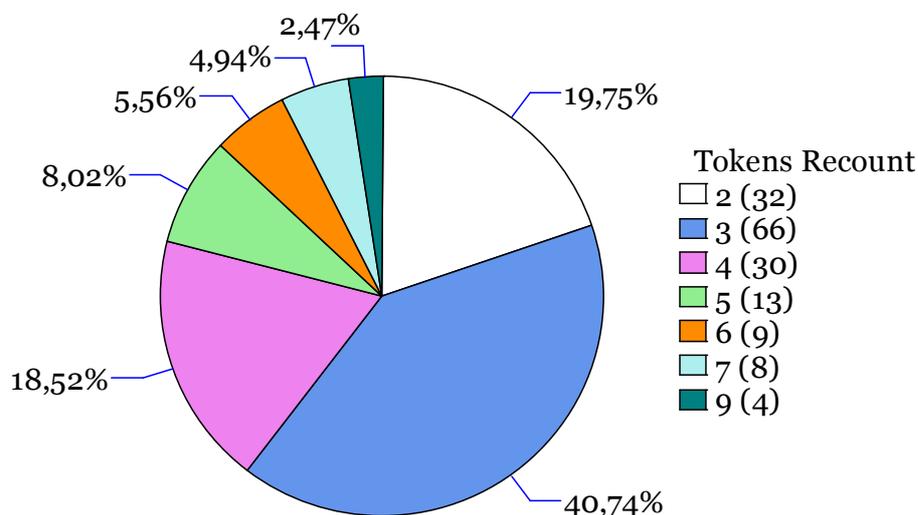
#	Expresiones	Marcaje DCI	Expresión real
1	BBB	<i>fr:</i>	before breaking bulk
2	before breaking bulk	<i>fr:</i>	before breaking bulk
3	FOC	<i>fr:</i>	flag of convenience

4	foc	<i>fr:</i>	flag of convenience
5	KDF	<i>fr:</i>	knocked down flat
6	KD flat	<i>fr:</i>	knocked down flat
7	SATSHEX	<i>fr:</i>	Saturdays, Sundays, holidays excluded
8	SSHE	<i>fr:</i>	Saturdays, Sundays, holidays excluded
9	SATSHINC	<i>fr:</i>	Saturdays, Sundays, holidays included
10	SSHINC	<i>fr:</i>	Saturdays, Sundays, holidays included
11	S&H/exct	<i>fr:</i>	sundays and holidays excepted/excluded
12	Shex	<i>fr:</i>	Sundays and holidays excepted/excluded
13	T/C	<i>fr:</i>	till countermanded, time charter
14	till countermanded	<i>fr:</i>	till countermanded

Tabla 11: Expresiones repetidas marcadas como *fr*

4.2.4.1. Conteo de tokens de las expresiones marcadas como *fr*

Al eliminar 7 de estas el número total de expresiones sería igual a 162, al hacer el conteo de tokens de las expresiones reales halladas en el diccionario los resultados fueron los siguientes:



Gráfica 15: Recuento de tokens de las expresiones reales marcadas como *fr* sin repeticiones

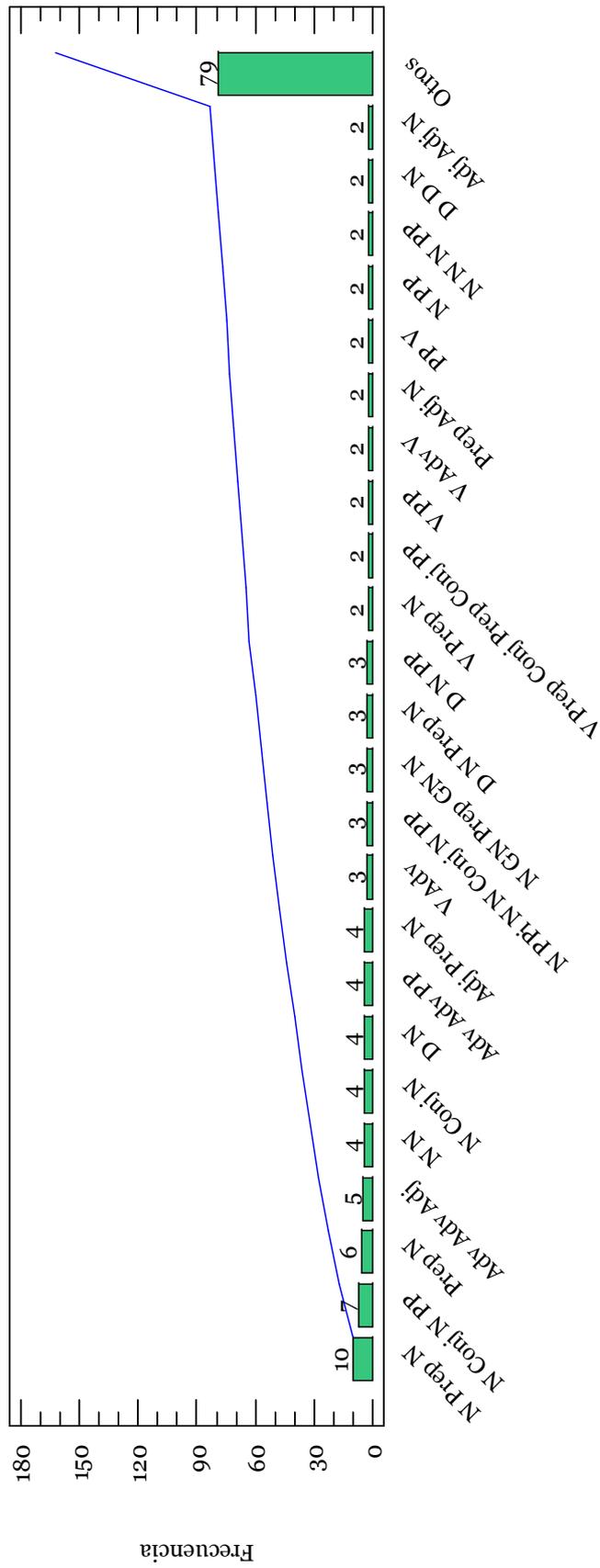
Es posible ya en este punto encontrar una unificación de criterios con los nuestros al demostrar que todas estas expresiones son en realidad pluriverbales –como lo habíamos afirmado previamente– creemos de todas maneras necesario incluir la tabla de valores en la que se pueden observar los datos del conteo.

# de tokens	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
2	32	0,1975	32	0,1975
3	66	0,4074	98	0,6049
4	30	0,1852	128	0,7901
5	13	0,0802	141	0,8704
6	9	0,0556	150	0,9259
7	8	0,0494	158	0,9753
9	4	0,0247	162	1,0000

Tabla 12: Frecuencias por número de tokens

4.2.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como *fr*

Una de las principales motivaciones de este trabajo era dar un aporte en lo que a patrones morfosintácticos se refiere. El uso de herramientas computacionales de análisis lingüístico como el *TreeTagger*, nos permiten hacer un acercamiento a este tipo de fenómenos de forma muy clara. A continuación presentamos no sólo las estadísticas de frecuencia de dichos patrones sino también la lista completa de patrones morfosintácticos de expresiones pluriverbales marcadas como *fr*, vale la pena aclarar que la gráfica mostrada sólo contiene aquellos patrones morfosintácticos que se repiten más de una vez en el corpus.



Gráfica 16: Patrones morfosintácticos más frecuentes en las expresiones marcadas como *fr*

Ampliamos la información de esta gráfica con la tabla completa de patrones morfosintácticos; incluimos en ésta aquellos patrones que solo tienen una aparición.

<i>Clase</i>	<i>Patrón</i>	<i>Frec.</i>	<i>Frec. Rel.</i>	<i>Frec. Acum.</i>	<i>Frec Rel. acum.</i>
1	Adj Adj	1	0,0062	1	0,0062
2	Adj Adj N	2	0,0123	3	0,0185
3	Adj Conj Adj	1	0,0062	4	0,0247
4	Adj N	1	0,0062	5	0,0309
5	Adj N Prep N	1	0,0062	6	0,0370
6	Adj Prep Conj Adv PP Conj PP	1	0,0062	7	0,0432
7	Adj Prep Conj PP	1	0,0062	8	0,0494
8	Adj Prep N	4	0,0247	12	0,0741
9	Adj Prep N Conj N N Conj Adj N	1	0,0062	13	0,0802
10	Adj Prep N N	1	0,0062	14	0,0864
11	Adj Prep N Prep	1	0,0062	15	0,0926
12	Adv Adv Adj	5	0,0309	20	0,1235
13	Adv Adv Adv PP	1	0,0062	21	0,1296
14	Adv Adv PP	4	0,0247	25	0,1543
15	Adv Adv PP Prep	1	0,0062	26	0,1605
16	Adv Adv V	1	0,0062	27	0,1667
17	Adv PP Adv	1	0,0062	28	0,1728
18	Adv Prep N N Prep D Adj	1	0,0062	29	0,1790
19	Conj N	1	0,0062	30	0,1852
20	D Adj	1	0,0062	31	0,1914
21	D Adj N	1	0,0062	32	0,1975
22	D Adj N N	1	0,0062	33	0,2037
23	D Adj N Prep D N	1	0,0062	34	0,2099
24	D D N	2	0,0123	36	0,2222
25	D N	4	0,0247	40	0,2469
26	D N Adj	1	0,0062	41	0,2531
27	D N N	1	0,0062	42	0,2593
28	D N N PP Conj N	1	0,0062	43	0,2654
29	D N PP	3	0,0185	46	0,2840
30	D N PPi N PP	1	0,0062	47	0,2901
31	D N Prep N	3	0,0185	50	0,3086
32	D N Prep Prep N	1	0,0062	51	0,3148
33	N Adj	1	0,0062	52	0,3210
34	N Conj Adj	1	0,0062	53	0,3272
35	N Conj N	4	0,0247	57	0,3519
36	N Conj N PP	7	0,0432	64	0,3951
37	N Conj N PP Prep	1	0,0062	65	0,4012
38	N Conj PPi Adv	1	0,0062	66	0,4074
39	N Conj V	1	0,0062	67	0,4136
40	N D N	1	0,0062	68	0,4198

41	N GN Prep GN N	3	0,0185	71	0,4383
42	N N	4	0,0247	75	0,4630
43	N N Adj GN N	1	0,0062	76	0,4691
44	N N Adj GN Prep GN N	1	0,0062	77	0,4753
45	N N Conj N PP	1	0,0062	78	0,4815
46	N N N	1	0,0062	79	0,4877
47	N N N PP	2	0,0123	81	0,5000
48	N N PP	1	0,0062	82	0,5062
49	N PP	2	0,0123	84	0,5185
50	N PPi	1	0,0062	85	0,5247
51	N PPi N	1	0,0062	86	0,5309
52	N PPi N N Conj N PP	3	0,0185	89	0,5494
53	N Pos N N Conj N	1	0,0062	90	0,5556
54	N Pos N N Conj V	1	0,0062	91	0,5617
55	N Prep	1	0,0062	92	0,5679
56	N Prep N	10	0,0617	102	0,6296
57	N Prep N Adj Prep N	1	0,0062	103	0,6358
58	N Prep N Prep	1	0,0062	104	0,6420
59	N V N Prep N	1	0,0062	105	0,6481
60	N V PP	1	0,0062	106	0,6543
61	PP Adj	1	0,0062	107	0,6605
62	PP Adv	1	0,0062	108	0,6667
63	PP Adv Adj	1	0,0062	109	0,6728
64	PP N Adv PPi	1	0,0062	110	0,6790
65	PP N Prep N	1	0,0062	111	0,6852
66	PP PP Conj PP	1	0,0062	112	0,6914
67	PP Prep	1	0,0062	113	0,6975
68	PP Prep D N N	1	0,0062	114	0,7037
69	PP Prep N	1	0,0062	115	0,7099
70	PP Prep Prep Adj Prep N N	1	0,0062	116	0,7160
71	PP Prep Prep N N	1	0,0062	117	0,7222
72	PP V	2	0,0123	119	0,7346
73	PPi Adv Adj	1	0,0062	120	0,7407
74	Prep Adj	1	0,0062	121	0,7469
75	Prep Adj Conj PP	1	0,0062	122	0,7531
76	Prep Adj N	2	0,0123	124	0,7654
77	Prep Adv Adj	1	0,0062	125	0,7716
78	Prep Adv Adv Adv PP	1	0,0062	126	0,7778
79	Prep N	6	0,0370	132	0,8148
80	Prep N Conj N	1	0,0062	133	0,8210
81	Prep N N	1	0,0062	134	0,8272
82	Prep N N Adv	1	0,0062	135	0,8333
83	Prep N PP Conj Adv	1	0,0062	136	0,8395
84	Prep N Pos N N	1	0,0062	137	0,8457
85	Prep N Prep	1	0,0062	138	0,8519
86	Prep PP	1	0,0062	139	0,8580
87	Prep Prep Adj N Conj Adv	1	0,0062	140	0,8642
88	Prep Prep N Conj Adv	1	0,0062	141	0,8704

89	V	1	0,0062	142	0,8765
90	V Adj Conj N D N PP	1	0,0062	143	0,8827
91	V Adj Conj N Prep N PP	1	0,0062	144	0,8889
92	V Adv	3	0,0185	147	0,9074
93	V Adv V	2	0,0123	149	0,9198
94	V D Adj	1	0,0062	150	0,9259
95	V D N Adj	1	0,0062	151	0,9321
96	V D N Prep Adj N PP Conj N	1	0,0062	152	0,9383
97	V D N Prep D N	1	0,0062	153	0,9444
98	V D N Prep N	1	0,0062	154	0,9506
99	V D N Prep PPI N PP Conj N	1	0,0062	155	0,9568
100	V N Adj Prep D N N	1	0,0062	156	0,9630
101	V PP	2	0,0123	158	0,9753
102	V Prep Conj Prep Conj PP	2	0,0123	160	0,9877
103	V Prep N	2	0,0123	162	1,0000

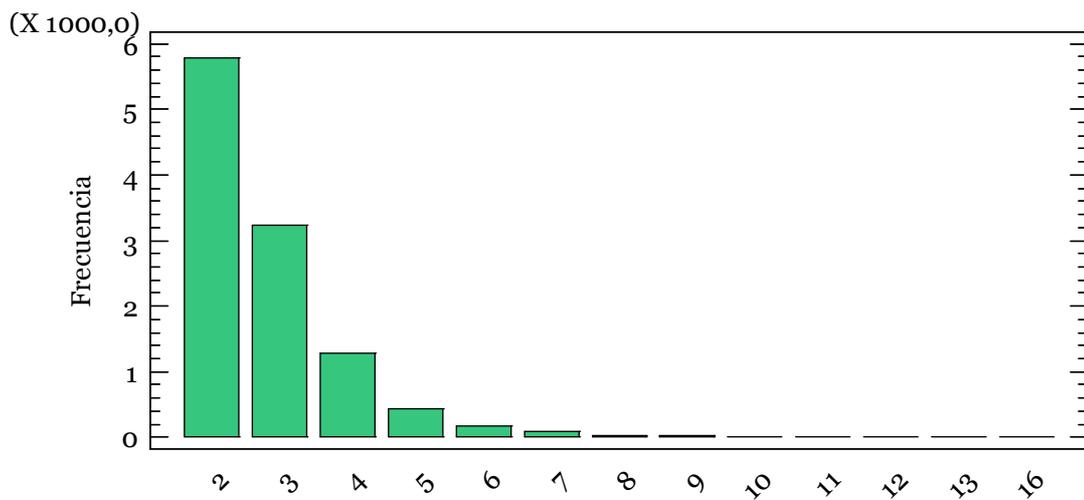
Tabla 13: Patrones morfosintácticos de las expresiones marcadas como *fr*

4.2.5. Descripción de las unidades sin marca de POS (subentradas)

Aunque el objeto de análisis del presente trabajo de grado está relacionado con el marcaje de las UFE, los datos que logramos recabar motivaron la continuación del análisis un poco más allá de las unidades marcadas. Esta será muy probablemente una de las tareas que quedará pendiente de terminar en futuros trabajos de investigación, dejando la puerta abierta al análisis de un corpus de más de 10000 términos de unidades pluriverbales, en esta tesis avanzaremos hasta ofrecer una clasificación de unidades de más de un token halladas en el DCI, ofrecemos no sólo los datos estadísticos en cuanto al número de unidades por token sino también una taxonomía morfosintáctica que podrá ser utilizada como punto de partida en proyectos por venir.

4.2.5.1. Conteo de tokens en las subentradas

En la gráfica siguiente se puede observar el número de unidades por token halladas en las subentradas del DCI.



Gráfica 17: Frecuencia de las expresiones en las subentradas por token

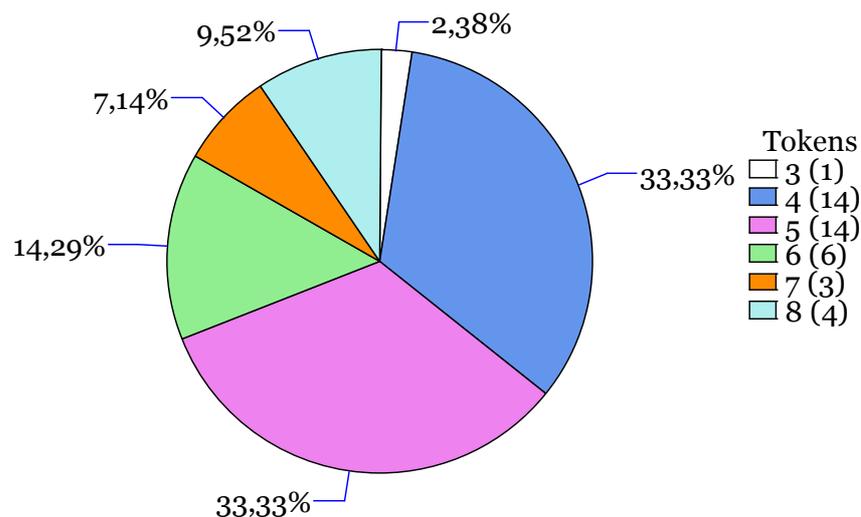
# de tokens	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
2	5780	0,5221	5780	0,5221
3	3234	0,2921	9014	0,8142
4	1288	0,1163	10302	0,9305
5	442	0,0399	10744	0,9705
6	167	0,0151	10911	0,9855
7	79	0,0071	10990	0,9927
8	31	0,0028	11021	0,9955
9	24	0,0022	11045	0,9977
10	14	0,0013	11059	0,9989
11	5	0,0005	11064	0,9994
12	5	0,0005	11069	0,9998
13	1	0,0001	11070	0,9999
16	1	0,0001	11071	1,0000

Tabla 14: Frecuencia de las subentradas por token

Sólo fueron tenidas en cuenta aquellas expresiones con más de un token ya que el propósito final de esta parte del análisis era el hallazgo de patrones morfosintácticos de unidades pluriverbales, en ese proceso se hallaron 4 tipos de expresiones: aquellas que tenían comas, como elemento compositivo, aquellas que tenían guiones, otras que no tenían ni comas ni guiones y una expresión que tenía tanto comas como guiones. Hasta este punto y después de la teoría consultada no hemos encontrado patrones morfosintácticos que agrupen al menos la primera clasificación.

4.2.5.2. Expresiones pluriverbales que contienen comas

La coma al ser un signo de puntuación no es considerado un token, sin embargo debería ser tenido en cuenta como elemento compositivo dentro de un patrón morfosintáctico porque lo ofrecen los datos recuperados del corpus donde vemos que es recurrente su uso en algunas expresiones que luego se podrían analizar con la intención de observar rasgos idiomáticos en ellas.

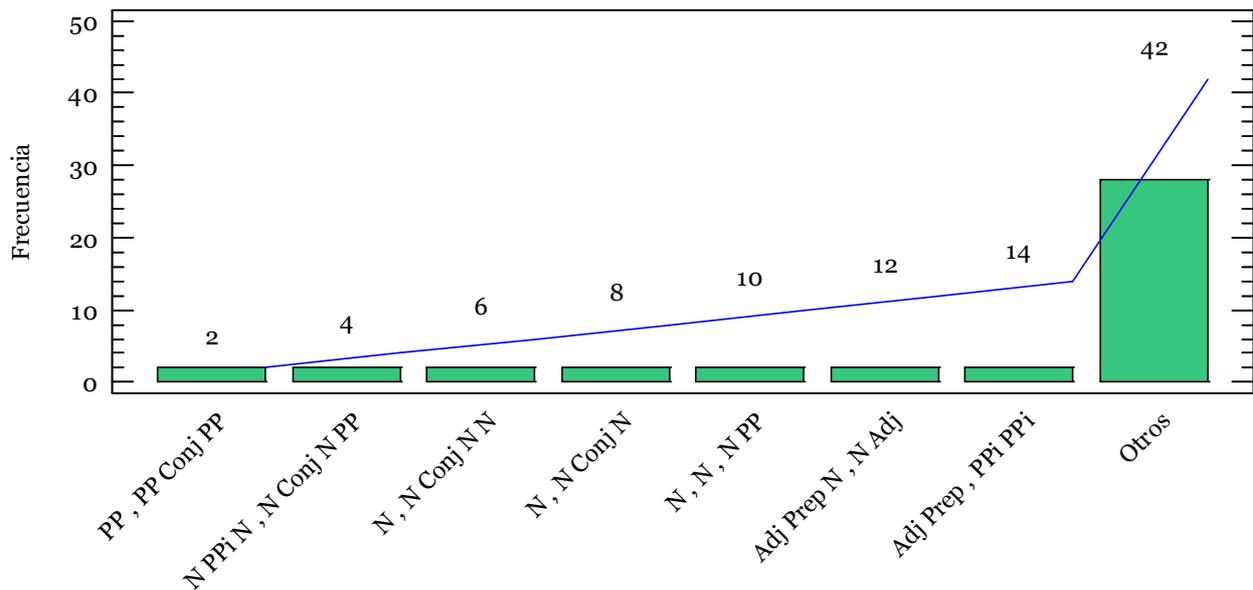


Gráfica 18: Conteo de tokens de las expresiones no marcadas con comas

Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. Acum.
3	1	0,0238	1	0,0238
4	14	0,3333	15	0,3571
5	14	0,3333	29	0,6905
6	6	0,1429	35	0,8333
7	3	0,0714	38	0,9048
8	4	0,0952	42	1,0000

Tabla 15: Frecuencia de expresiones que contienen comas entre las subentradas por tokens

La siguiente gráfica sólo muestra los patrones morfosintácticos más frecuentes, en este caso, tan sólo aquellos que tenían frecuencias superiores a 1:



Gráfica 19: Patrones morfosintácticos más recurrentes en expresiones con coma

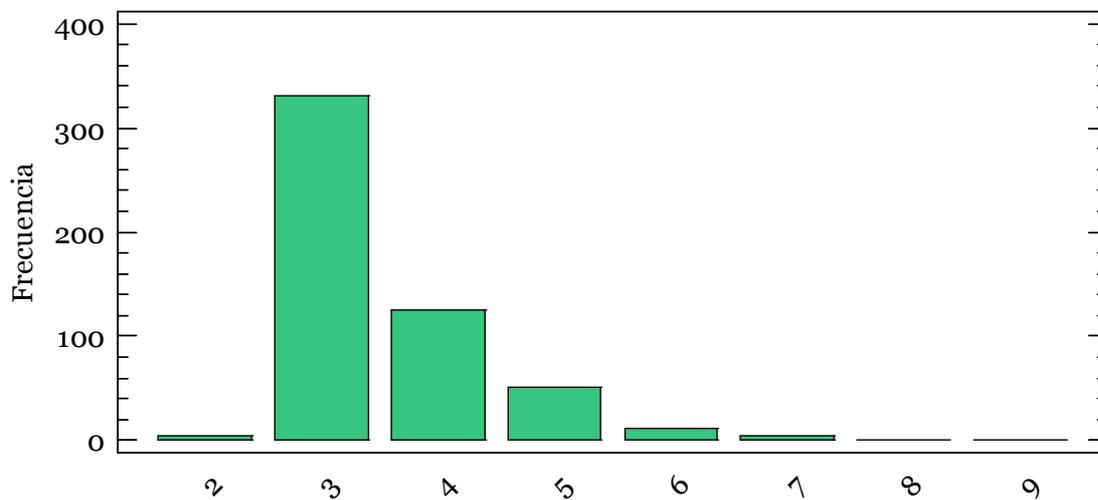
El siguiente cuadro contiene además de los antes mencionados, los patrones encontrados con una sola frecuencia:

<i>Clas e</i>	<i>Valor</i>	<i>Frec.</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. Acum.</i>
1	Adj , Adj , Adj , Adj N	1	0,0238	1	0,0238
2	Adj N , N PP	1	0,0238	2	0,0476
3	Adj N , N Prep N Pos N	1	0,0238	3	0,0714
4	Adj N , PP N	1	0,0238	4	0,0952
5	Adj Prep , N Prep	1	0,0238	5	0,1190
6	Adj Prep , PPi PPi	2	0,0476	7	0,1667
7	Adj Prep Conj PP , N Prep	1	0,0238	8	0,1905
8	Adj Prep Conj Prep , Conj PP	1	0,0238	9	0,2143
9	Adj Prep Conj Prep , PPi PPi	1	0,0238	10	0,2381
10	Adj Prep N , N , N Conj Adj N	1	0,0238	11	0,2619
11	Adj Prep N , N Adj	2	0,0476	13	0,3095
12	Adj Prep N , N N	1	0,0238	14	0,3333
13	Adj Prep N , N PP	1	0,0238	15	0,3571
14	N , N , Conj N N	1	0,0238	16	0,3810
15	N , N , N	1	0,0238	17	0,4048
16	N , N , N , Conj N	1	0,0238	18	0,4286
17	N , N , N , Conj V	1	0,0238	19	0,4524
18	N , N , N , N Conj N	1	0,0238	20	0,4762
19	N , N , N Conj N	1	0,0238	21	0,5000
20	N , N , N PP	2	0,0476	23	0,5476
21	N , N Conj N	2	0,0476	25	0,5952
22	N , N Conj N N	2	0,0476	27	0,6429
23	N , N Conj N PP	1	0,0238	28	0,6667
24	N , N Conj V	1	0,0238	29	0,6905
25	N , N XX Conj N PP Prep	1	0,0238	30	0,7143
26	N , PPi , Conj PPi N	1	0,0238	31	0,7381
27	N N Conj N Prep PPi PP ,	1	0,0238	32	0,7619
28	N N N Prep N , N , Conj N	1	0,0238	33	0,7857
29	N PPi N , N Conj N PP	2	0,0476	35	0,8333
30	N Pos N , N Conj V	1	0,0238	36	0,8571
31	N Prep N , N Conj N	1	0,0238	37	0,8810
32	PP , PP , Adj N	1	0,0238	38	0,9048
33	PP , PP Conj PP	2	0,0476	40	0,9524
34	Prep , Adj Conj N N	1	0,0238	41	0,9762
35	Prep , Adj Conj PP	1	0,0238	42	1,0000

Tabla 16: Patrones morfosintácticos de las expresiones con comas

4.2.5.3. Expresiones pluriverbales que contienen guiones

Otra posible clasificación la conformaban aquellas expresiones que contenían guiones dentro de sus elementos compositivos, al igual que con las anteriores, se hizo primero un conteo de frecuencias por token, para luego extraer los patrones morfosintácticos de las subentradas con guiones.



Gráfica 20: Conteo de tokens de las expresiones no marcadas con guiones

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	2	4	0,0076	4	0,0076
2	3	331	0,6257	335	0,6333
3	4	125	0,2363	460	0,8696
4	5	51	0,0964	511	0,9660
5	6	12	0,0227	523	0,9887
6	7	4	0,0076	527	0,9962
7	8	1	0,0019	528	0,9981
8	9	1	0,0019	529	1,0000

Tabla 17: Conteo de frecuencias por tokens para expresiones con guiones

Fueron tantos los patrones morfosintácticos encontrados en esta clasificación que graficar el fenómeno fue imposible, es por esto que solamente presentaremos estos por medio de la siguiente tabla:

<i>Etiqueta de Clase</i>	<i>Rango</i>	<i>Recuento</i>	<i>Puntaje Acum</i>	<i>Porcentaje</i>	<i>Porcentaje Acum</i>
N GN N N	1	87	87	16,45	16,45
Adj GN N N	2	71	158	13,42	29,87
N GN PP N	3	29	187	5,48	35,35
Adj GN PP N	4	17	204	3,21	38,56
N GN Adj N	5	16	220	3,02	41,59
N GN N N N	6	14	234	2,65	44,23
N GN Prep GN N N	7	13	247	2,46	46,69
D GN N N	8	13	260	2,46	49,15
Adj GN Adj N	9	12	272	2,27	51,42
N GN PPi N	10	11	283	2,08	53,50
Prep GN N N	11	10	293	1,89	55,39
Adj GN N N N	12	10	303	1,89	57,28
N GN Prep N	13	7	310	1,32	58,60
N Prep Adj GN N	14	6	316	1,13	59,74
N GN PPi N N	15	6	322	1,13	60,87
Adj GN Adj N N	16	6	328	1,13	62,00
N GN N N N N	17	5	333	0,95	62,95
V GN Prep N	18	4	337	0,76	63,71
N GN PP N N	19	4	341	0,76	64,46
N GN N	20	4	345	0,76	65,22
N GN Adv Adj	21	4	349	0,76	65,97
Adj GN Adj N Prep N	22	4	353	0,76	66,73
Prep GN D N N	23	3	356	0,57	67,30
N N GN N N	24	3	359	0,57	67,86
N N GN N	25	3	362	0,57	68,43
N GN Adj N N	26	3	365	0,57	69,00
D GN N N N	27	3	368	0,57	69,57
Adv GN Prep GN N N	28	3	371	0,57	70,13
Adv Adj N GN N N	29	3	374	0,57	70,70
Adj N GN N N	30	3	377	0,57	71,27
Adj GN PPi N	31	3	380	0,57	71,83
Adj GN PP N N	32	3	383	0,57	72,40
Adj GN N N Prep N	33	3	386	0,57	72,97
Adj GN Adj Adj	34	3	389	0,57	73,53
Adj Conj Adj GN PP N	35	3	392	0,57	74,10
Prep GN N N Prep N	36	2	394	0,38	74,48
N Prep N GN Adj N	37	2	396	0,38	74,86

N Prep D GN D	38	2	398	0,38	75,24
N N GN Adv	39	2	400	0,38	75,61
N GN V N	40	2	402	0,38	75,99
N GN Prep N N	41	2	404	0,38	76,37
N GN PPi Adj N	42	2	406	0,38	76,75
N GN PP Adj	43	2	408	0,38	77,13
N GN N Prep N N	44	2	410	0,38	77,50
N GN N Prep N	45	2	412	0,38	77,88
N GN N GN Prep N	46	2	414	0,38	78,26
N GN Adv V	47	2	416	0,38	78,64
D GN PP N	48	2	418	0,38	79,02
Adj GN PP Adj	49	2	420	0,38	79,40
Adj GN N PPI	50	2	422	0,38	79,77
Otros	51	107	529	20,23	100,00
Total		529			

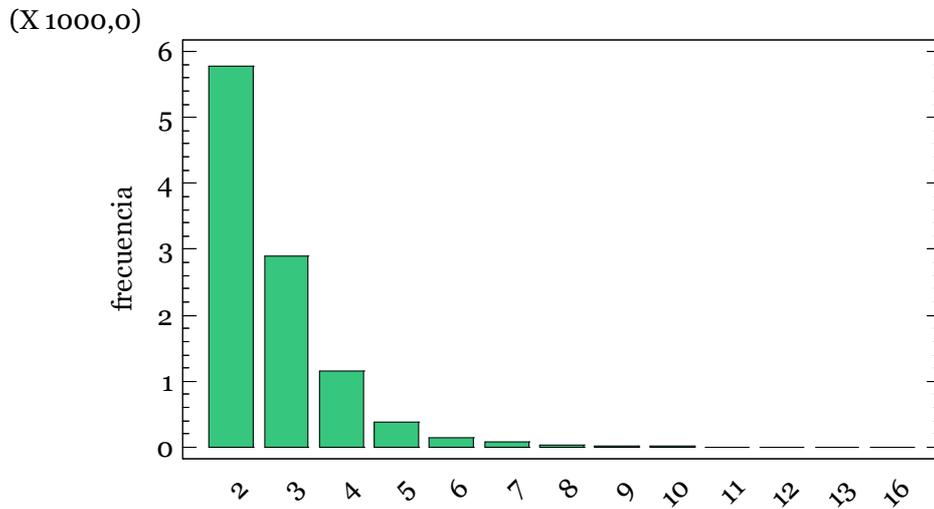
Tabla 18: Patrones morfosintácticos de las expresiones con guiones

4.2.5.4. Expresiones pluriverbales que contienen comas y guiones

Éste fue un caso excepcional, entre todos los hallazgos que pudieron hacerse tan sólo encontramos un caso en el que se integraban tanto comas como la utilización de guiones: *bulk-in, bag-out*, tiene por definición en español carga a granel y descarga en sacos y posee tanto coma, como un par de guiones, su patrón morfosintáctico es el siguiente N GN Prep ,N GN Prep.

4.2.5.5. Expresiones pluriverbales que no contienen comas o guiones

Quizás el grupo más grande de expresiones es el que viene a continuación, los niveles de frecuencia de los patrones morfosintácticos de esta categoría es una de las más altas dentro de todo el corpus, eso se ve reflejado en el gran número de unidades que lo conforman, veamos:



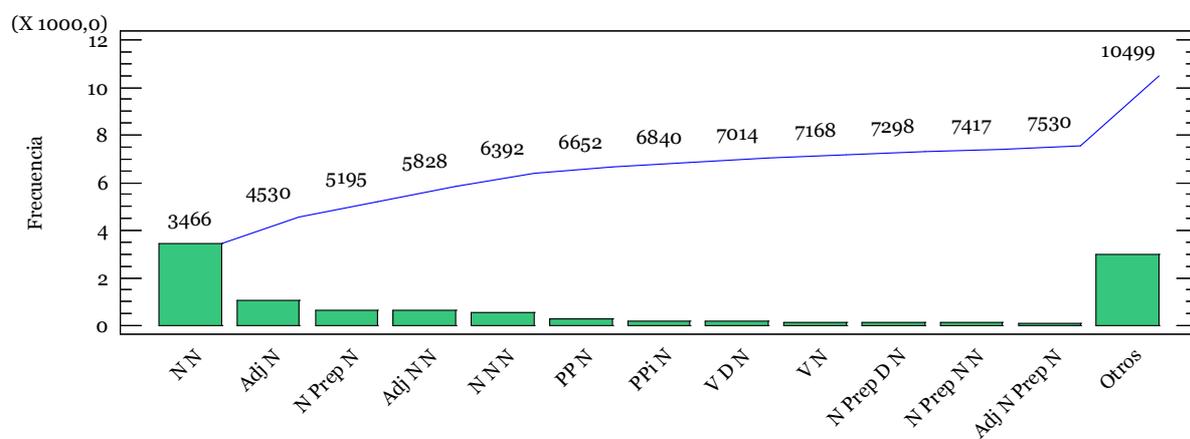
Gráfica 21: Frecuencias por token de las expresiones no marcadas sin comas o guiones

# de tokens	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
2	5776	0,5501	5776	0,5501
3	2902	0,2764	8678	0,8266
4	1148	0,1093	9826	0,9359
5	377	0,0359	10203	0,9718
6	149	0,0142	10352	0,9860
7	72	0,0069	10424	0,9929
8	26	0,0025	10450	0,9953
9	23	0,0022	10473	0,9975
10	14	0,0013	10487	0,9989
11	5	0,0005	10492	0,9993
12	5	0,0005	10497	0,9998
13	1	0,0001	10498	0,9999
16	1	0,0001	10499	1,0000

Tabla 19: Frecuencias por tokens de unidades con y sin comas

Vemos que de acuerdo a los datos de la tabla tenemos estructuras que van desde los 2 tokens hasta los 16 tokens, muchas de estas expresiones aunque pueden ser compuestos o sintagmas extensos, podrían también albergar algún patrón morfosintáctico recurrente.

A nivel metodológico podríamos considerar éste el primer filtro de selección de las UFE reduciendo el número de candidatos por la presencia de un patrón recurrente. La gráfica que presentamos a continuación sólo ilustra los patrones morfosintácticos que tuvieron mínimo un 1% de frecuencia entre todas las unidades ya que la lista total de patrones morfosintácticos son un poco menos de 400.



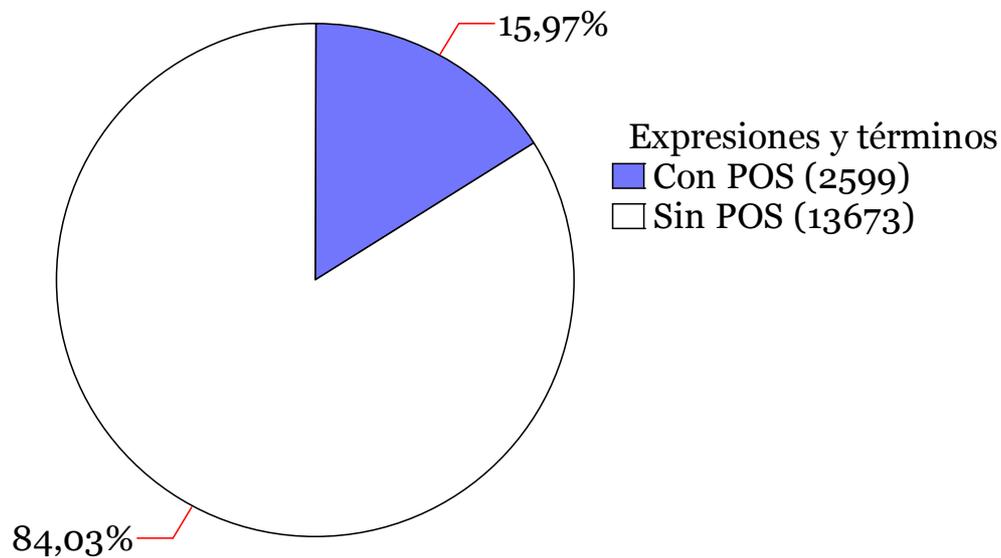
Gráfica 22: Patrones morfosintácticos más comunes en expresiones sin comas o guiones

4.3. Descripción y análisis estadístico en español

La presentación de los datos será consistente con lo hecho con las unidades en inglés, salvo que en español no tendremos que hacer claridad con respecto a algunos fenómenos que se presentaron en inglés más no en español.

4.3.1. Unidades con o sin POS

El primer hallazgo que hicimos fue comprobar el número de unidades que poseían o no marca, como lo habíamos dicho antes éste es un indicador del número de entradas y subentradas en el diccionario. Las siguientes son las gráficas y tablas correspondientes a los hallazgos estadísticos.



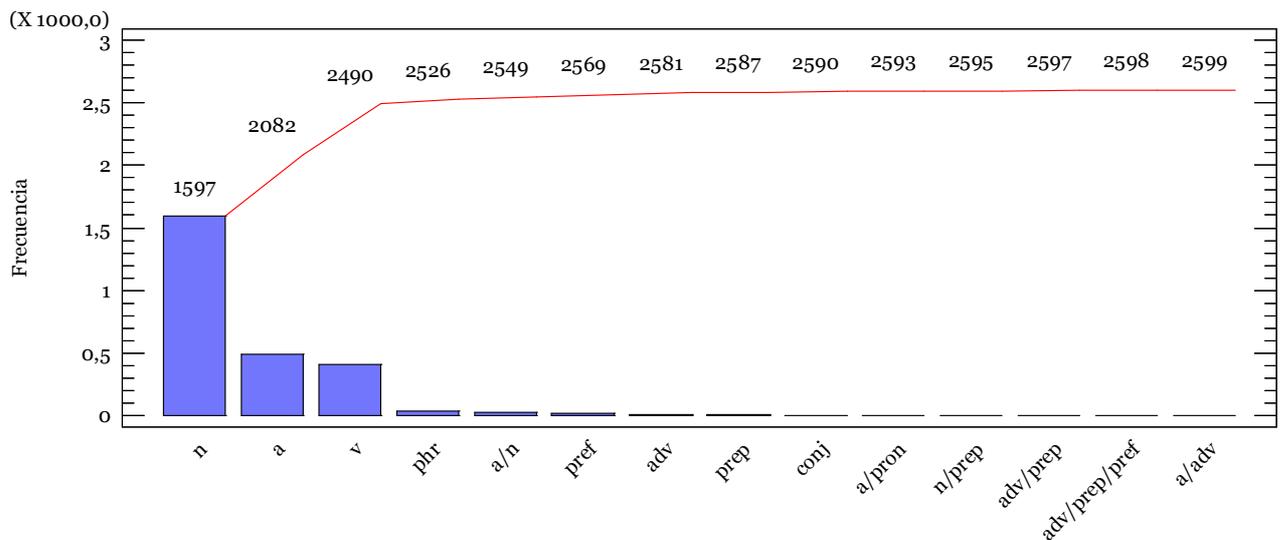
Gráfica 23: Expresiones con y sin POS en español

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	Con POS	2599	0,1597	2599	0,1597
2	Sin POS	13673	0,8403	16272	1,0000

Tabla 20: Frecuencia de expresiones con y sin marca gramatical

4.3.2. Descripción de marcaje de entradas con POS

A continuación presentamos el listado de marcas que presenta el diccionario en su segunda parte.



Gráfica 24: Conteo de etiquetas en entradas de 2 o más frecuencias en español

Las marcas presentadas a diferencia de las vistas en inglés poseían muchas menos variantes, no era tan común encontrar cambios en el orden de estas es decir entre a/n y n/a. A continuación mostramos los datos hallados en el diccionario.

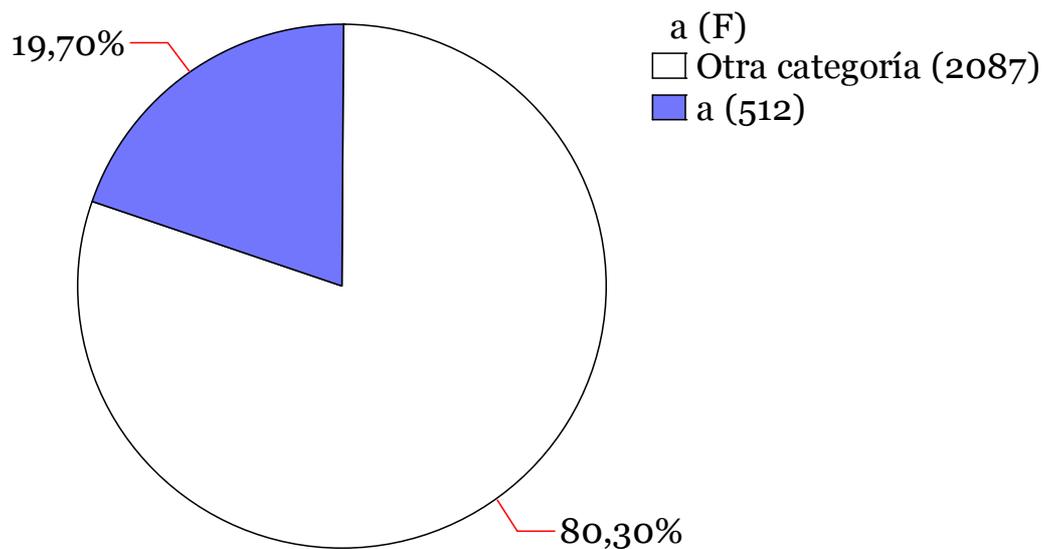
<i>Etiqueta de Clase</i>	<i>Recuento</i>	<i>Puntaje Ponderado</i>	<i>Puntaje Acum</i>	<i>Porcentaje</i>	<i>Porcentaje Acum</i>
N	1597	1597	1597	61,45	61,45
A	485	485	2082	18,66	80,11
V	408	408	2490	15,70	95,81
Phr	36	36	2526	1,39	97,19
a/n	23	23	2549	0,88	98,08
Pref	20	20	2569	0,77	98,85
Adv	12	12	2581	0,46	99,31
Prep	6	6	2587	0,23	99,54
Conj	3	3	2590	0,12	99,65
a/pron	3	3	2593	0,12	99,77

n/prep	2	2	2595	0,08	99,85
adv/prep	2	2	2597	0,08	99,92
adv/prep/pref	1	1	2598	0,04	99,96
a/adv	1	1	2599	0,04	100,00
Total	2599	2599			

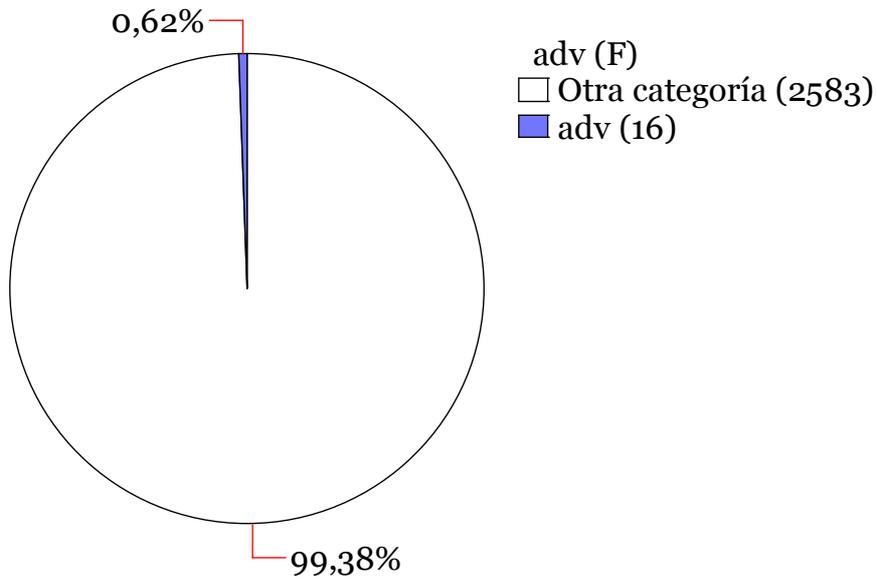
Tabla 21: Frecuencias y porcentajes de etiquetas gramaticales en español

4.3.3. Descripción de marcas simples

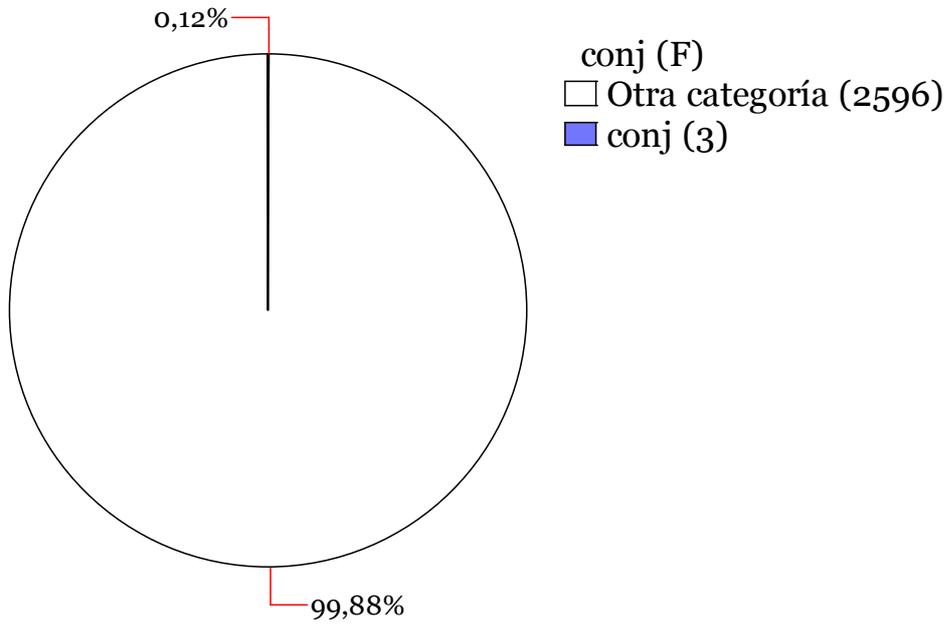
Las marcas simples encontradas en esta parte del diccionario las mostramos a continuación con sus gráficas independientes.



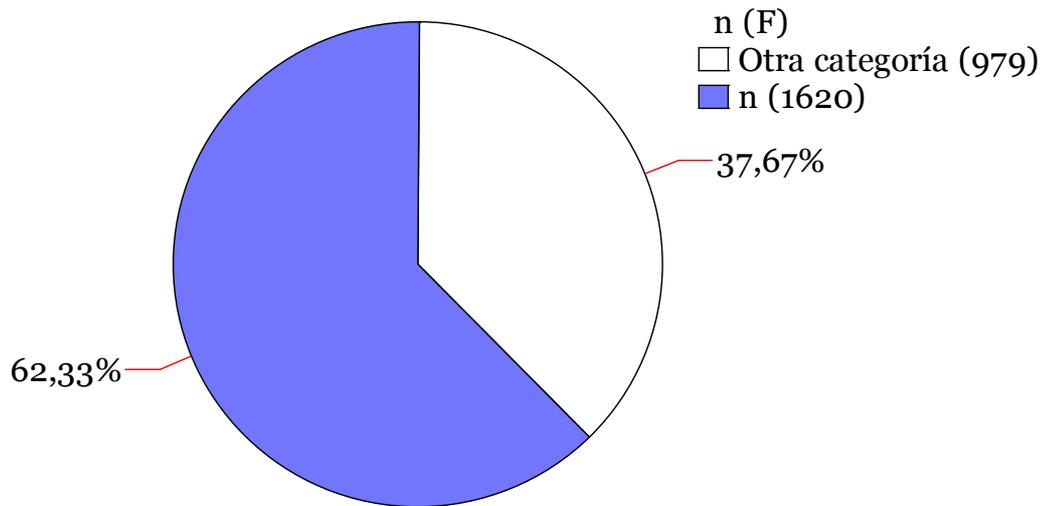
Gráfica 25: Frecuencia de adjetivos (a) en las entradas



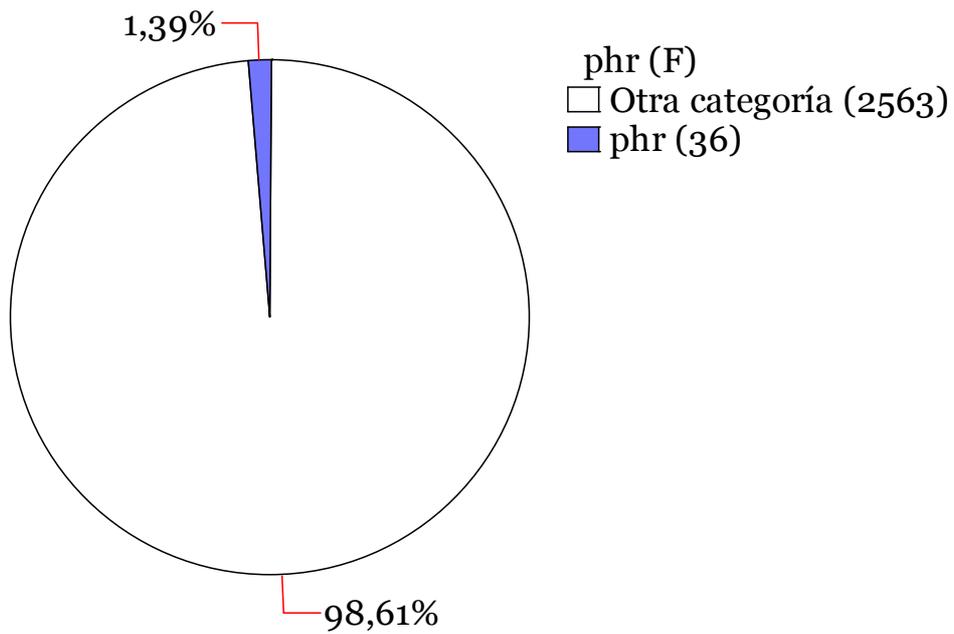
Gráfica 26: Frecuencia de adverbios (adv) en las entradas



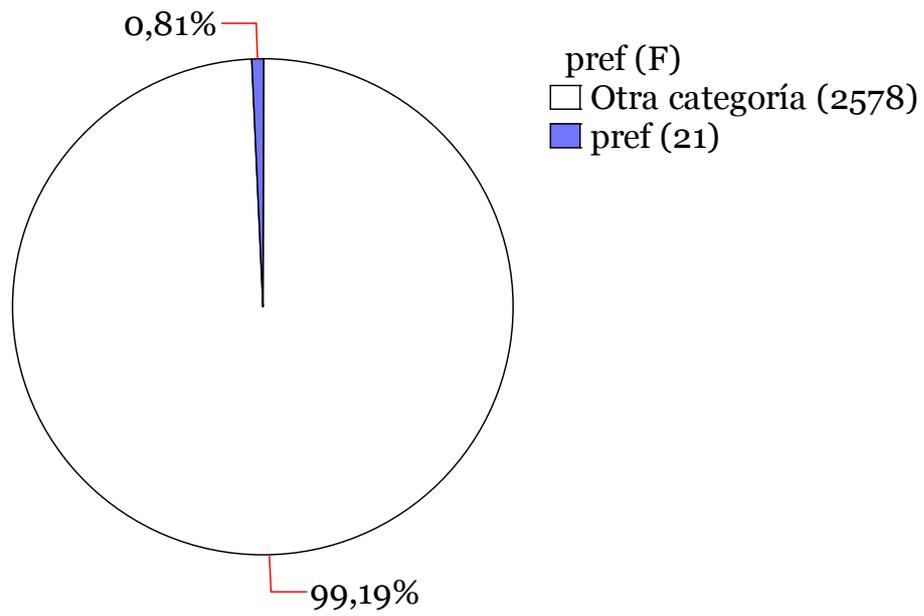
Gráfica 27: Frecuencia de conjunciones (conj) en las entradas



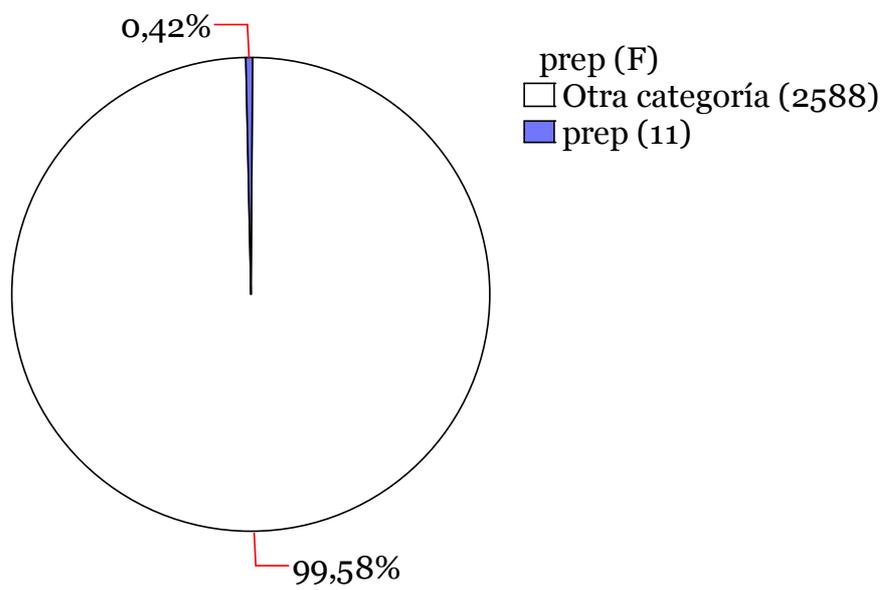
Gráfica 28: Frecuencia de sustantivos (n) en las entradas



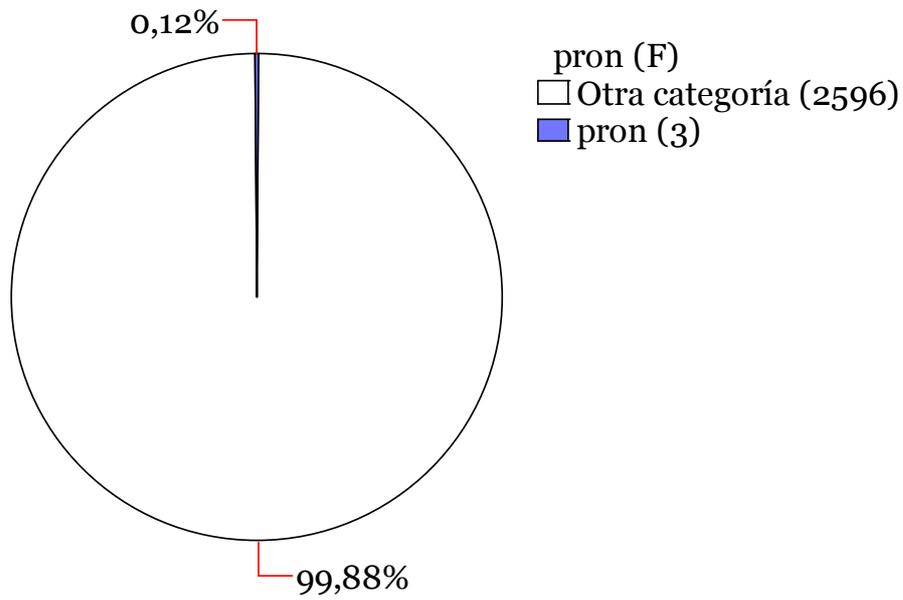
Gráfica 29: Frecuencia de expresiones marcadas como (phr) en las entradas



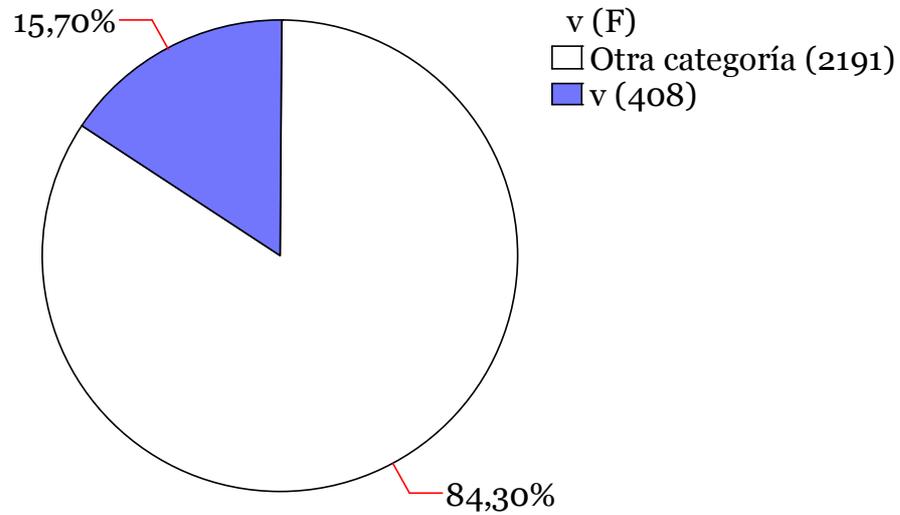
Gráfica 30: Frecuencia de prefijos (pref) en las entradas



Gráfica 31: Frecuencia de preposiciones (prep) en las entradas



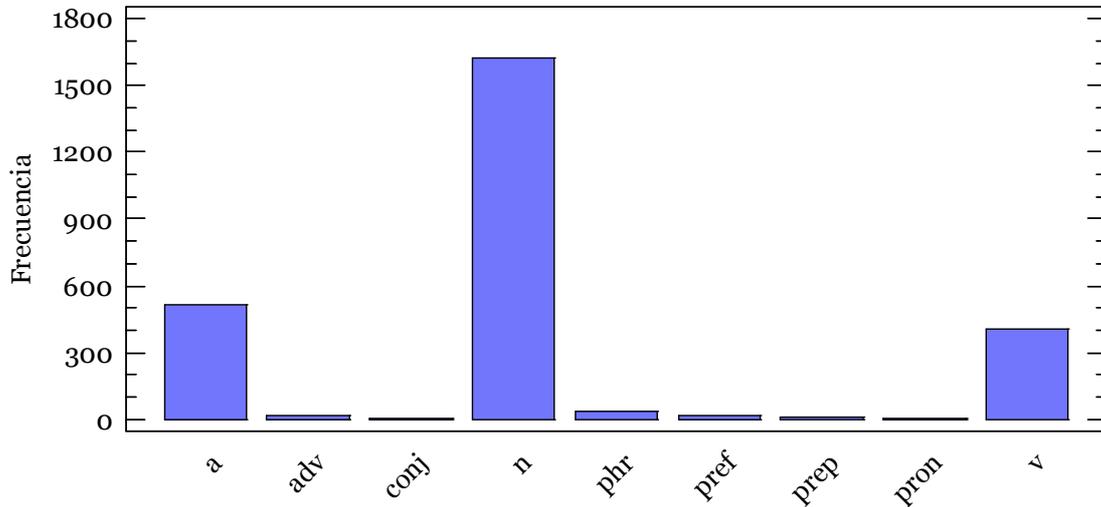
Gráfica 32: Frecuencia de pronombres (pron) en las entradas



Gráfica 33: Frecuencia de verbos (v) en las entradas

A diferencia de la parte de inglés, en español no se hallaron intejecciones; se debe entender de igual forma que las marcas gramaticales en este espacio se encuentran en

inglés debido a que, al momento de ser consultadas, su equivalente y definición harían parte de ese paradigma de la lengua inglesa.



Gráfica 34: Frecuencias para marcas gramaticales simples en español

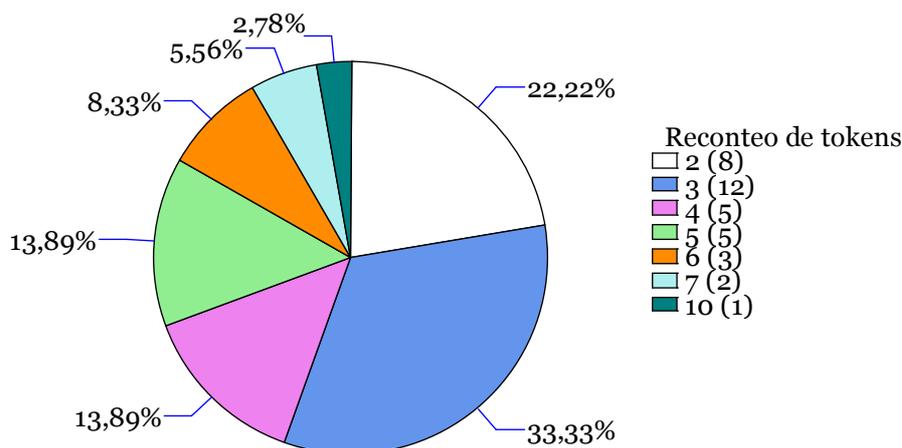
Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	a	512	0,1947	512	0,1947
2	adv	16	0,0061	528	0,2008
3	conj	3	0,0011	531	0,2019
4	n	1620	0,6160	2151	0,8179
5	phr	36	0,0137	2187	0,8316
6	pref	21	0,0080	2208	0,8395
7	prep	11	0,0042	2219	0,8437
8	pron	3	0,0011	2222	0,8449
9	v	408	0,1551	2630	1,0000

Tabla 22: Frecuencias para marcas gramaticales simples en español

4.3.4. Descripción de la marca *phr*

La marca *phr*, en esta parte del diccionario al igual que el resto de las etiquetas gramaticales, tenía una gran homogeneidad, tanto es que ninguna de las entradas –a diferencia de la primera parte– tenían expresiones de un solo token, esto evitó en gran medida tener que hacer un recuento.

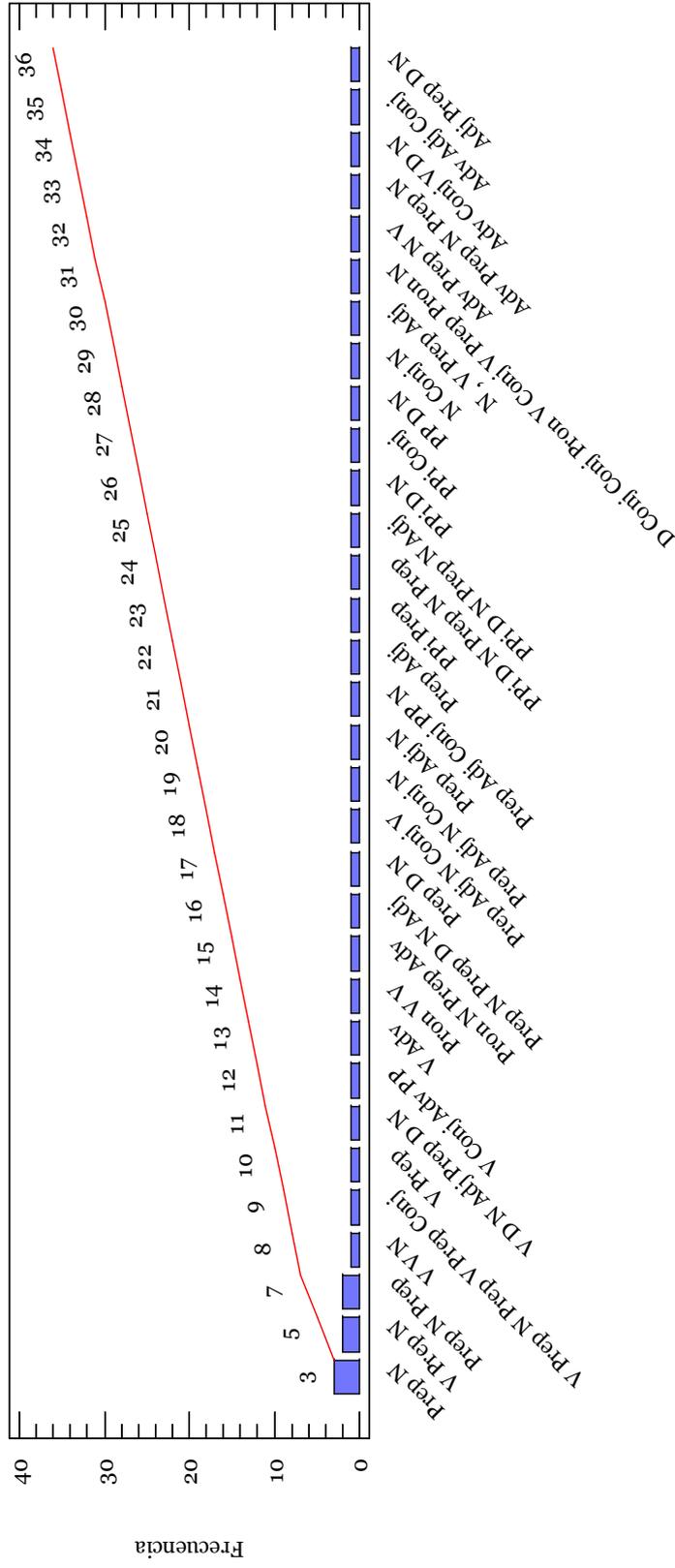
4.3.4.1. Conteo de tokens de las expresiones marcadas como *phr*



Gráfica 35: Frecuencia de las expresiones marcadas como *phr* sin repeticiones

4.3.4.2. Identificación y presentación de patrones morfosintácticos de expresiones marcadas como *phr*.

Los patrones morfosintácticos extraídos de las 36 unidades marcadas como *phr* se presentan a continuación tanto de forma gráfica, como en una tabla, que permite analizar su frecuencia de aparición en las entradas.



Gráfica 36: Patrones morfosintácticos en las expresiones marcadas como *phr*

Sin embargo, hablar de patrones morfosintácticos es un poco ambiguo ya que se esperaría encontrar repetición en alguno de ellos, es sin duda también entendible no encontrar repeticiones dado que sólo se hallan 36 expresiones entre las 2599 expresiones marcadas.

Este es el listado completo que se tomó para hacer la gráfica:

<i>Clase</i>	<i>Patrón</i>	<i>Frec.</i>	<i>Frecuencia Relativa</i>	<i>Frec. Acum.</i>	<i>Frec. Rel. acum.</i>
1	Adj Prep D N	1	0,0278	1	0,0278
2	Adv Adj Conj	1	0,0278	2	0,0556
3	Adv Conj V D N	1	0,0278	3	0,0833
4	Adv Prep N Prep N	1	0,0278	4	0,1111
5	Adv Prep N V	1	0,0278	5	0,1389
6	D Conj Conj Pron V Conj V Prep Pron N	1	0,0278	6	0,1667
7	N , V Prep Adj	1	0,0278	7	0,1944
8	N Conj N	1	0,0278	8	0,2222
9	PP D N	1	0,0278	9	0,2500
10	PPi Conj	1	0,0278	10	0,2778
11	PPi D N	1	0,0278	11	0,3056
12	PPi D N Prep N Adj	1	0,0278	12	0,3333
13	PPi D N Prep N Prep	1	0,0278	13	0,3611
14	PPi Prep	1	0,0278	14	0,3889
15	Prep Adj	1	0,0278	15	0,4167
16	Prep Adj Conj PP N	1	0,0278	16	0,4444
17	Prep Adj N	1	0,0278	17	0,4722
18	Prep Adj N Conj N	1	0,0278	18	0,5000
19	Prep Adj N Conj V	1	0,0278	19	0,5278
20	Prep D N	1	0,0278	20	0,5556
21	Prep N	3	0,0833	23	0,6389
22	Prep N Prep	2	0,0556	25	0,6944
23	Prep N Prep D N Adj	1	0,0278	26	0,7222
24	Pron N Prep Adv	1	0,0278	27	0,7500
25	Pron V V	1	0,0278	28	0,7778
26	V Adv	1	0,0278	29	0,8056

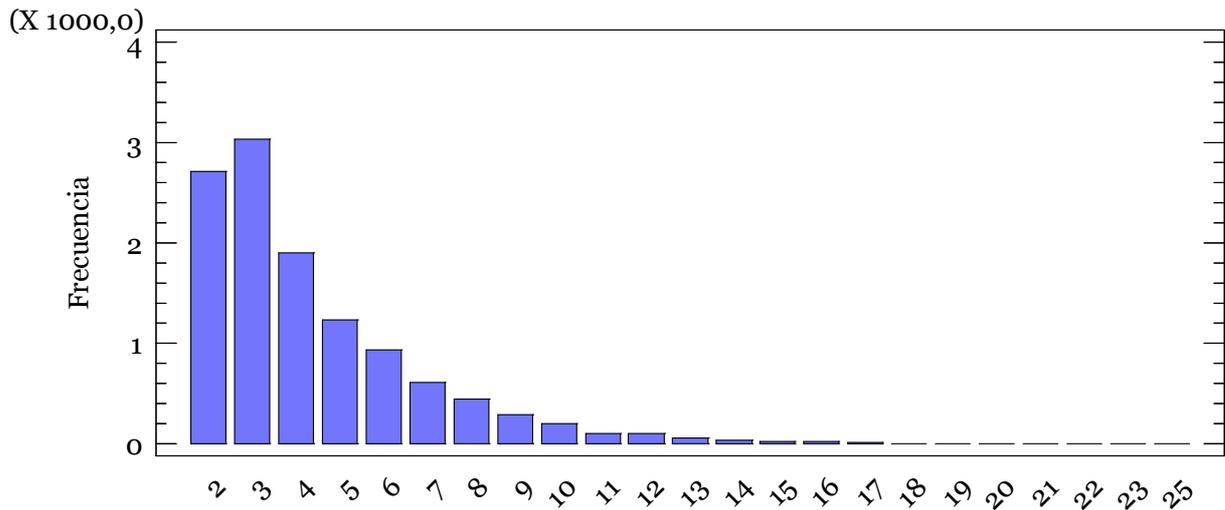
27	V Conj Adv PP	1	0,0278	30	0,8333
28	V D N Adj Prep D N	1	0,0278	31	0,8611
29	V Prep	1	0,0278	32	0,8889
30	V Prep N	2	0,0556	34	0,9444
31	V Prep N Prep V Prep Conj	1	0,0278	35	0,9722
32	V V N	1	0,0278	36	1,0000

Tabla 23: Patrones morfosintácticos en las expresiones marcadas como *phr*

4.3.5. Descripción de las unidades sin marca de POS en las subentradas

Las expresiones que no poseían marca de ningún tipo son presentadas a continuación en términos estadísticos.

4.3.5.1. Conteo de tokens en las subentradas



Gráfica 37: Frecuencia de expresiones en las subentradas por token

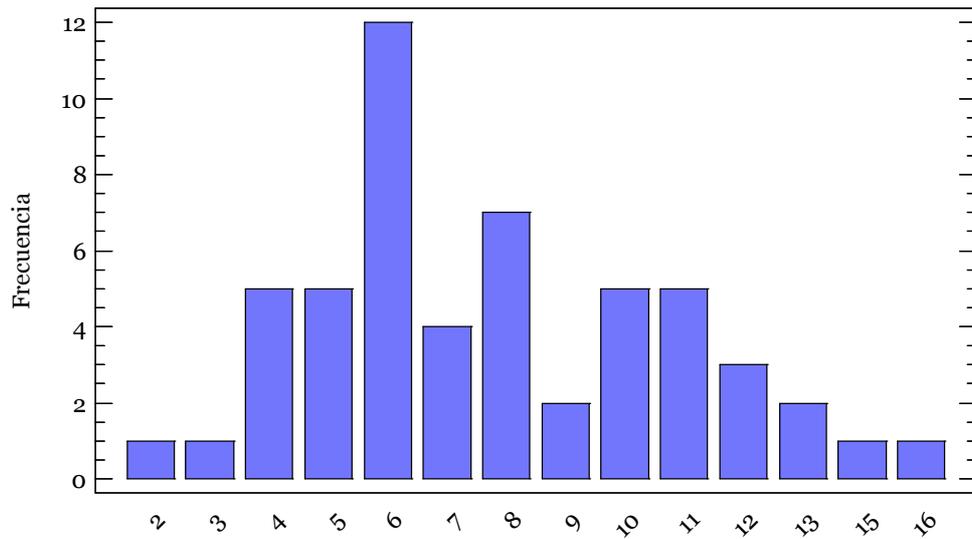
<i># de tokens</i>	<i>Frecuencia</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. acum.</i>
2	2710	0,2316	2710	0,2316
3	3032	0,2591	5742	0,4907
4	1896	0,1620	7638	0,6527
5	1235	0,1055	8873	0,7582
6	933	0,0797	9806	0,8380
7	609	0,0520	10415	0,8900
8	442	0,0378	10857	0,9278
9	286	0,0244	11143	0,9522
10	197	0,0168	11340	0,9691
11	100	0,0085	11440	0,9776
12	98	0,0084	11538	0,9860
13	58	0,0050	11596	0,9909
14	37	0,0032	11633	0,9941
15	19	0,0016	11652	0,9957
16	20	0,0017	11672	0,9974
17	12	0,0010	11684	0,9985
18	4	0,0003	11688	0,9988
19	2	0,0002	11690	0,9990
20	4	0,0003	11694	0,9993
21	2	0,0002	11696	0,9995
22	3	0,0003	11699	0,9997
23	2	0,0002	11701	0,9999
25	1	0,0001	11702	1,0000

Tabla 24: Frecuencia de las subentradas por token

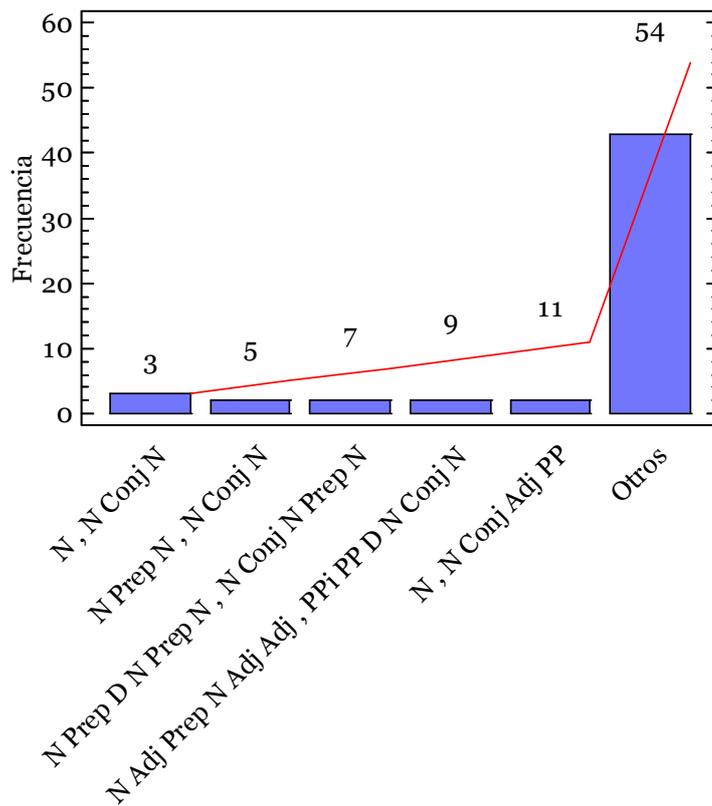
Igual que en la primera parte de la presentación de los datos estadísticos dividiremos los datos hallados en 4 grupos, dependiendo si las expresiones contienen o no guiones y comas.

4.3.5.2. Expresiones pluriverbales que contienen comas

El primer grupo a tener en cuenta dentro de estos tipos de unidades lo conforman aquellas unidades que contienen comas dentro de su composición morfosintáctica:



Gráfica 38: Conteo de tokens de expresiones con coma en español



Gráfica 39: Patrones morfosintácticos más recurrentes en expresiones con coma

El conteo de frecuencias de los distintos patrones morfosintácticos puede ser analizado con mayor facilidad a través de la siguiente tabla:

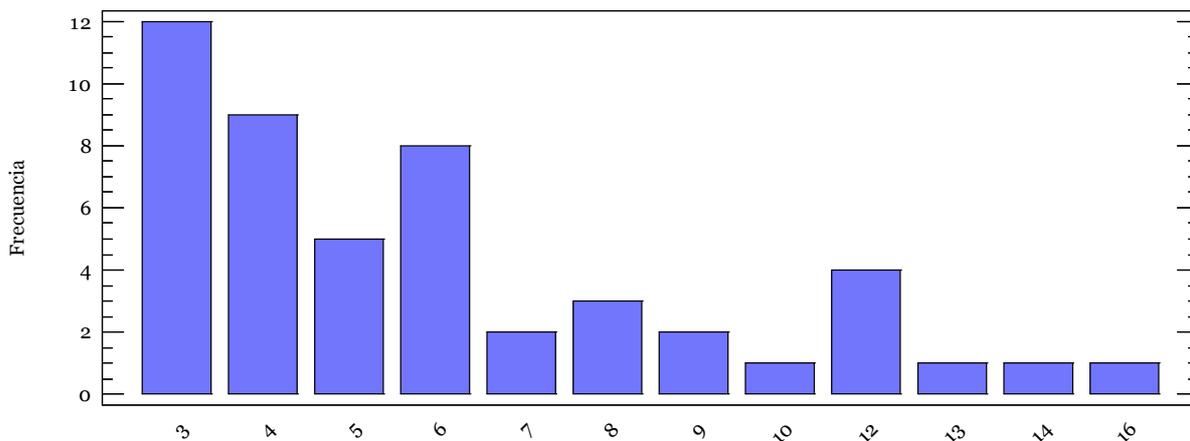
<i>Clase</i>	<i>Patrón</i>	<i>Frec.</i>	<i>Frec. Relativa</i>	<i>Frec. Acum.</i>	<i>Frec. Rel. acum.</i>
1	Adj , V Prep Adj	1	0,0185	1	0,0185
2	Adj N , PPi D N Prep D N N Prep N	1	0,0185	2	0,0370
3	Adj Prep N , N , N Conj N Adj	1	0,0185	3	0,0556
4	Adj Prep N , N Conj N	1	0,0185	4	0,0741
5	Adj Prep N Prep N , PP Conj PP	1	0,0185	5	0,0926
6	N , Adj , V , N Conj N	1	0,0185	6	0,1111
7	N , Adj , V Conj N	1	0,0185	7	0,1296
8	N , Adj , V Conj V	1	0,0185	8	0,1481
9	N , Adj Conj V	1	0,0185	9	0,1667
10	N , Adj Conj V Conj N Adj	1	0,0185	10	0,1852
11	N , Adj Conj V PP Prep N	1	0,0185	11	0,2037
12	N , Conj V Conj V	1	0,0185	12	0,2222
13	N , N , N Conj N , N	1	0,0185	13	0,2407
14	N , N , N Prep D N	1	0,0185	14	0,2593
15	N , N , PP Conj N Prep D N	1	0,0185	15	0,2778
16	N , N Conj Adj PP	2	0,0370	17	0,3148
17	N , N Conj N	3	0,0556	20	0,3704
18	N , N Conj N Prep D N N Prep N	1	0,0185	21	0,3889
19	N , Prep	1	0,0185	22	0,4074
20	N , Prep PP	1	0,0185	23	0,4259
21	N , V Conj V Prep D N PP Prep D N	1	0,0185	24	0,4444
22	N Adj , Adj , Adj Conj Adj	1	0,0185	25	0,4630
23	N Adj , Adj Conj Prep N	1	0,0185	26	0,4815
24	N Adj , PP Conj Prep N	1	0,0185	27	0,5000
25	N Adj Conj V N , Adj Conj Adj Adj Adj	1	0,0185	28	0,5185
26	N Adj PP D N Prep N , D N Conj D N	1	0,0185	29	0,5370
27	N Adj Prep N Adj Adj , PPi PP D N Conj N	2	0,0370	31	0,5741
28	N Conj Pron V N , V Conj V	1	0,0185	32	0,5926
29	N N Prep D N , D N Conj D N	1	0,0185	33	0,6111
30	N N Prep N , N N Conj N	1	0,0185	34	0,6296
31	N PP , Conj Adv V D	1	0,0185	35	0,6481

32	N Prep D N , D V Conj Pron N	1	0,0185	36	0,6667
33	N Prep D N Prep N , N Conj N Prep N	2	0,0370	38	0,7037
34	N Prep N , Adj Prep N Prep N	1	0,0185	39	0,7222
35	N Prep N , N , N Conj N	1	0,0185	40	0,7407
36	N Prep N , N Conj N	2	0,0370	42	0,7778
37	N Prep N , N Conj N Prep D N N Prep N	1	0,0185	43	0,7963
38	N Prep N Adj , PP D D Conj D Adj	1	0,0185	44	0,8148
39	N Prep N Conj N , Conj Adv V Prep D	1	0,0185	45	0,8333
40	N Prep N Conj N , PP Conj PP	1	0,0185	46	0,8519
41	N Prep N Conj PP , PPi D N Prep N Prep N	1	0,0185	47	0,8704
42	N Prep N PP Prep D N Adj , Conj PP Conj	1	0,0185	48	0,8889
43	N Prep V , N Conj V	1	0,0185	49	0,9074
44	PP , PP Prep N Conj PP	1	0,0185	50	0,9259
45	Prep D N , Prep N Conj Prep N , N Conj A	1	0,0185	51	0,9444
46	Prep N , V D N Prep N Prep N Conj N	1	0,0185	52	0,9630
47	V , V Conj Adj PP Prep XX	1	0,0185	53	0,9815
48	V D N PP Conj V Prep D , Conj	1	0,0185	54	1,0000

Tabla 25: Patrones morfosintácticos de expresiones con coma en español

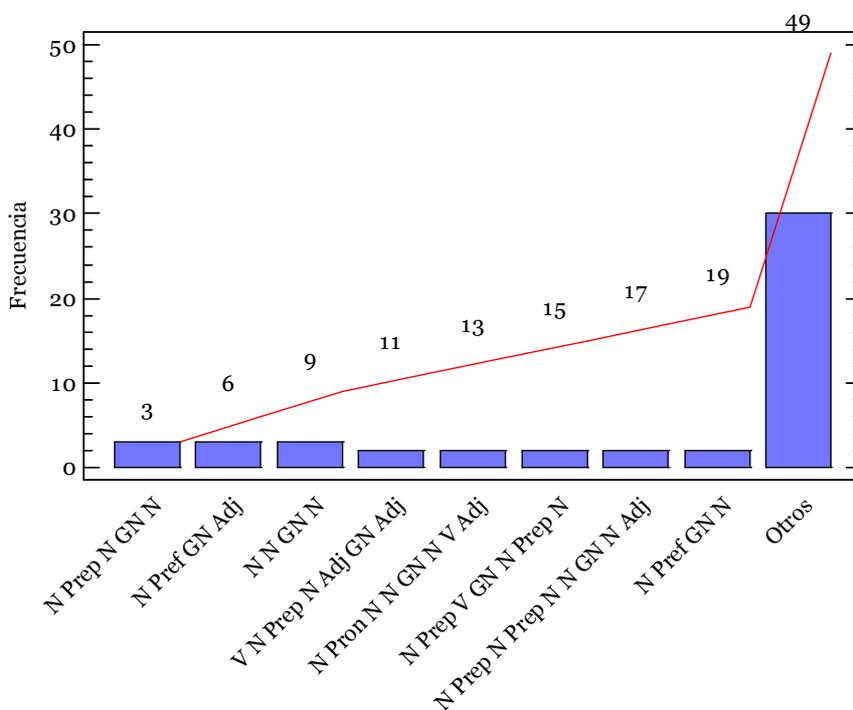
4.3.5.3. Expresiones pluriverbales que contienen guiones

El segundo grupo de expresiones los conforman aquellas que poseen guiones, a continuación presentamos el conteo de tokens, seguido del análisis morfosintáctico.



Gráfica 40: Conteo de tokens de expresiones con guiones en español

Los patrones morfosintácticos extraídos de estas expresiones se muestran a continuación, la gráfica sólo contiene los más frecuentes, la tabla la lista completa:



Gráfica 41: Gráfica de patrones morfosintácticos frecuentes de expresiones con guiones

<i>Clase</i>	<i>Valor</i>	<i>Frec.</i>	<i>Frec. Rel.</i>	<i>Frec. Acum.</i>	<i>Frec. Rel. acum.</i>
1	N Adj GN Adj	1	0,0204	1	0,0204
2	N Adj PP Prep D N Prep V GN N	1	0,0204	2	0,0408
3	N Adj Prep N GN N Prep D N Conj D N Adj	1	0,0204	3	0,0612
4	N Adv Adj N GN Adj	1	0,0204	4	0,0816
5	N GN N Adj Adj	1	0,0204	5	0,1020
6	N GN N Prep D N Adj	1	0,0204	6	0,1224
7	N N GN Adj	1	0,0204	7	0,1429
8	N N GN Adj Prep N Conj D N	1	0,0204	8	0,1633
9	N N GN Adj Prep N N	1	0,0204	9	0,1837
10	N N GN N	3	0,0612	12	0,2449
11	N N GN Prep	1	0,0204	13	0,2653
12	N N N GN N	1	0,0204	14	0,2857
13	N N Prep D N GN N	1	0,0204	15	0,3061
14	N PP Prep V Prep N N PP Prep D Pref GN N	1	0,0204	16	0,3265
15	N Pref GN Adj	3	0,0612	19	0,3878
16	N Pref GN N	2	0,0408	21	0,4286
17	N Pref GN PP	1	0,0204	22	0,4490
18	N Prep D N GN N Conj N Prep D N Prep N A	1	0,0204	23	0,4694
19	N Prep D N N Prep D N Prep N GN N Adj	1	0,0204	24	0,4898
20	N Prep D N N Prep N Conj N Prep D N Prep	1	0,0204	25	0,5102
21	N Prep D N PP Prep D V GN N Prep D N	1	0,0204	26	0,5306
22	N Prep D V GN N Prep N Prep N Prep Adv P	1	0,0204	27	0,5510
23	N Prep N Adj GN Adj	1	0,0204	28	0,5714
24	N Prep N Adj GN N	1	0,0204	29	0,5918
25	N Prep N GN N	3	0,0612	32	0,6531
26	N Prep N N GN N	1	0,0204	33	0,6735
27	N Prep N Prep N GN N	1	0,0204	34	0,6939
28	N Prep N Prep N N GN N Adj	2	0,0408	36	0,7347
29	N Prep N Prep V GN N Adj Prep N	1	0,0204	37	0,7551
30	N Prep Pref GN Adj	1	0,0204	38	0,7755
31	N Prep Pref GN N	1	0,0204	39	0,7959
32	N Prep Pref GN V	1	0,0204	40	0,8163
33	N Prep V GN N	1	0,0204	41	0,8367

34	N Prep V GN N Prep N	2	0,0408	43	0,8776
35	N Prep V GN N Prep N Prep N Prep PP	1	0,0204	44	0,8980
36	N Pron N N GN N V Adj	2	0,0408	46	0,9388
37	Pref GN N PP Conj PP	1	0,0204	47	0,9592
38	V N Prep N Adj GN Adj	2	0,0408	49	1,0000

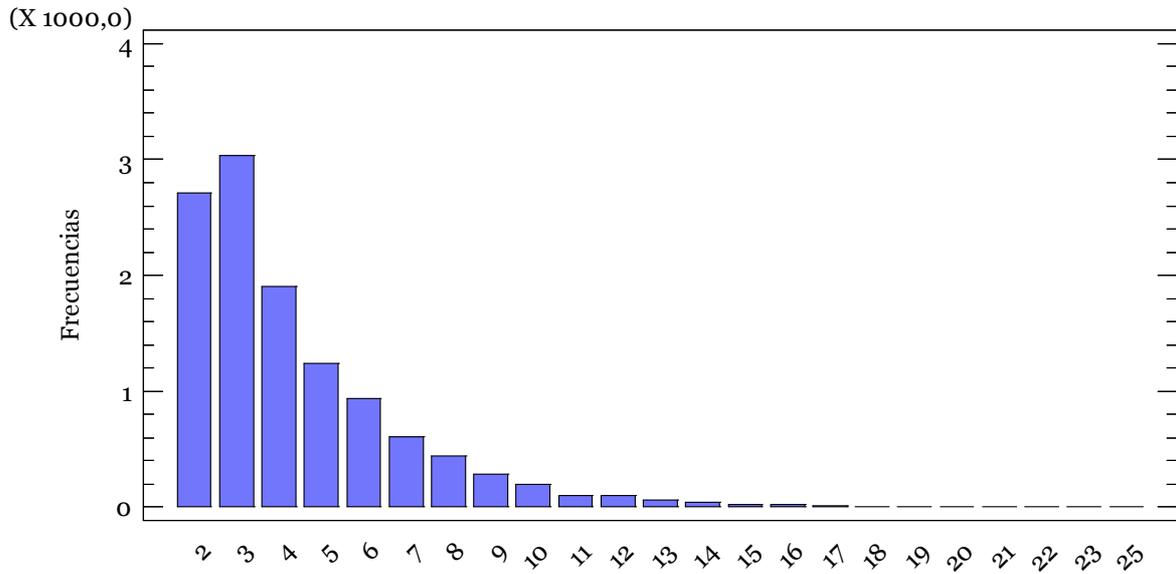
Tabla 26: Patrones morfosintácticos de expresiones con guion en español

4.3.5.4. Expresiones pluriverbales que contienen comas y guiones

Igual que en la primera parte del diccionario la segunda se caracterizó por tener también una sola expresión en la que se pudieran encontrar tanto guiones como comas: “arrendamiento de la nave aérea con tripulación, fuel-oil, incluidos” el patrón que encontramos acá es: N Prep D N Adj Prep N , N GN N , PP, una expresión de 10 tokens.

4.3.5.5. Expresiones pluriverbales que no contienen comas o guiones

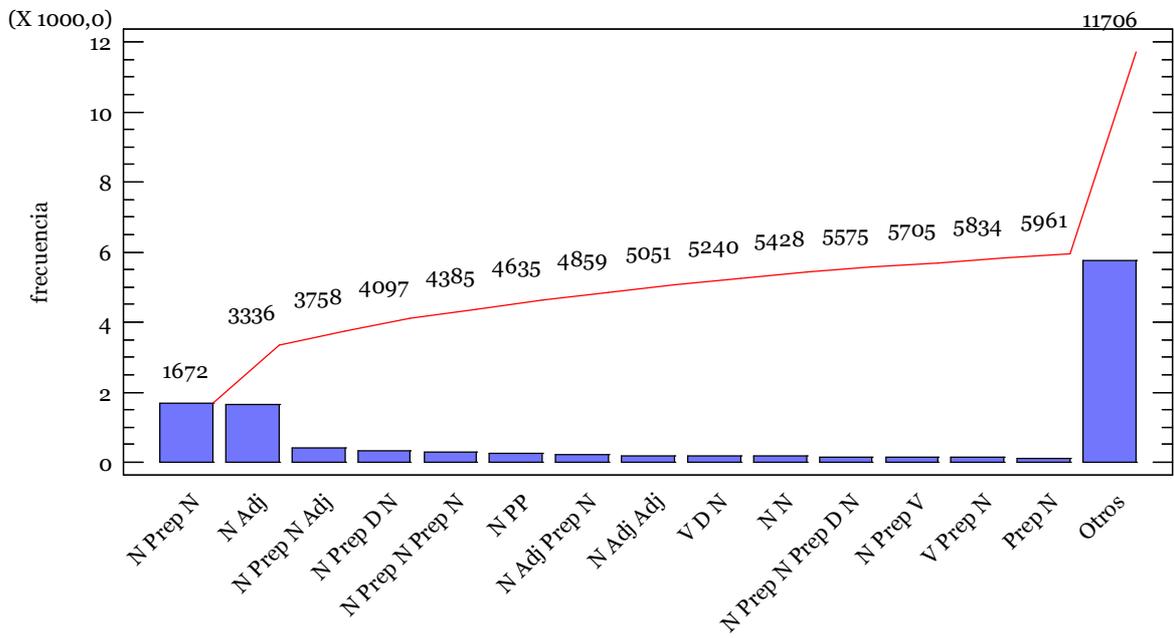
Como lo afirmamos páginas atrás este grupo es, en ambos idiomas, el que más expresiones contiene y por ende una gran fuente de patrones morfosintácticos para analizar, uno de los posibles estudios que se podría decantar de estos hallazgos tiene que ver con la comparación de los patrones encontrados en las expresiones marcadas como *fr* contra los patrones hallados en este grupo de expresiones. El conteo de tokens y la extracción de patrones morfosintácticos se muestra a continuación junto con sus respectivas gráficas:



Gráfica 42: Frecuencia de expresiones sin comas y guiones en español por token

<i>Valor</i>	<i>Frecuencia</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. acum.</i>
2	2712	0,2317	2712	0,2317
3	3034	0,2592	5746	0,4909
4	1898	0,1621	7644	0,6530
5	1235	0,1055	8879	0,7585
6	935	0,0799	9814	0,8384
7	606	0,0518	10420	0,8901
8	442	0,0378	10862	0,9279
9	286	0,0244	11148	0,9523
10	196	0,0167	11344	0,9691
11	101	0,0086	11445	0,9777
12	97	0,0083	11542	0,9860
13	58	0,0050	11600	0,9909
14	37	0,0032	11637	0,9941
15	19	0,0016	11656	0,9957
16	20	0,0017	11676	0,9974
17	12	0,0010	11688	0,9985
18	4	0,0003	11692	0,9988
19	2	0,0002	11694	0,9990
20	4	0,0003	11698	0,9993
21	2	0,0002	11700	0,9995
22	3	0,0003	11703	0,9997
23	2	0,0002	11705	0,9999
25	1	0,0001	11706	1,0000

Tabla 27: Frecuencia de unidades con y sin comas por token



Gráfica 43: Patrones morfosintácticos (>1%) en expresiones sin comas o guiones

5. Conclusiones: resultados y líneas de trabajo futuro

5. CONCLUSIONES: RESULTADOS Y LÍNEAS DE TRABAJO FUTURO	145
5.1. VALIDACIÓN DE CRITERIOS DE UFE	146
5.1.1. Elementos deverbales en inglés	147
5.1.2. Elementos deverbales en español	148
5.1.3. Fijación de las expresiones en inglés	151
5.1.4. Fijación de las expresiones en español	161
5.2. HALLAZGOS COMPLEMENTARIOS	164
5.2.1. Evaluación formal del diccionario	164
5.2.1.1. Análisis de la macroestructura	165
5.2.1.2. Rastreo de las fuentes citadas	166
5.2.1.3. Lematización	168
5.2.1.4. Polisemia	169
5.2.1.5. Análisis de la microestructura	170
5.2.2. Errores de etiquetaje	171
5.3. CONSIDERACIONES FINALES	175

5.1. Validación de criterios de UFE

Ya en los apartes 2.6.1. y 3.5. de este trabajo hemos indicado no sólo los criterios a tener en cuenta en la validación, sino también la forma en que éstos serán empleados. A continuación presentamos los hallazgos encontrados durante esta etapa de análisis del corpus.

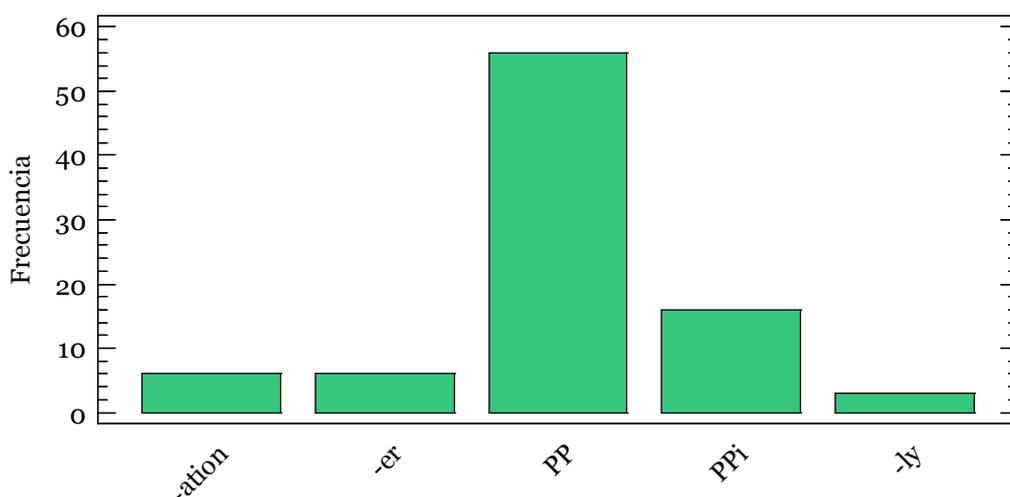
De los criterios ofrecidos por Quiroz et al. (2013) dos de ellos, el 1 (Son estructuras sintagmáticas con presencia de un término o UT y con tendencia a la denotación) y el 6 (Mantienen un significado específico dentro de un dominio, por lo que adquieren una estabilidad semántica) serían cumplidos por todas las expresiones al ser estas tomadas de un diccionario especializado y por ende todas ellas estarían en presencia de un término y poseen un significado determinado dentro de los dominios especializados en que las agrupa el diccionario.

Ahora bien, aunque estos dos criterios pueden ser fácilmente analizados, el número 5 (Admiten la conmutación de sus elementos, aunque hay ciertos grados. Así, se observa que el sentido especializado de un verbo está determinado por un conjunto limitado de nombres) no será tenido en cuenta en el presente trabajo debido a la imposibilidad –por el tiempo de duración de la maestría– de crear un corpus de contraste distinto a un diccionario donde se pudiera evaluar la conmutabilidad de los elementos compositivos de las expresiones seleccionadas marcadas como *fr.*

5.1.1. Elementos deverbales en inglés

El segundo criterio (Incluyen un verbo o un elemento deverbal) nos indica que las expresiones deben tener ya por núcleo o en alguno de sus componentes adyacentes un elemento deverbal. Quiroz (2008, pp. 119-238) en su tesis doctoral nos ofrece un listado de los sufijos o derivados deverbales más frecuentes tanto en inglés como en español. Los mismos fueron tenidos en cuenta en el presente aparte para hacer el análisis

Se encontraron tres categorías deverbales: sustantivos (terminados en “-ation” y “-er”), participios verbales (PP y PPI) y adverbios (terminados en -ly), siendo los segundos –los participios– los más productivos dentro de las estadísticas como lo podemos ver en la siguiente gráfica:

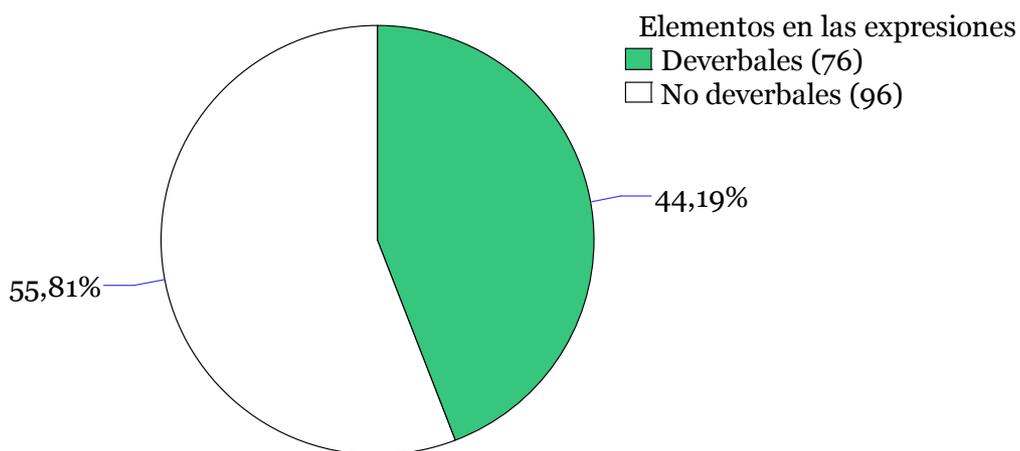


Gráfica 44: Frecuencias de elementos deverbales en inglés

<i>Valor</i>	<i>Frecuencia</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. acum.</i>
-ation	6	0,0690	6	0,0690
-er	6	0,0690	12	0,1379
PP	56	0,6437	68	0,7816
PPi	16	0,1839	84	0,9655
-ly	3	0,0345	87	1,0000

Tabla 28: Descripción de elementos deverbales en inglés

Ahora bien, no todas las expresiones contienen un sólo elemento verbal, por eso nos parece conveniente presentar el siguiente gráfico que ilustra esta situación:

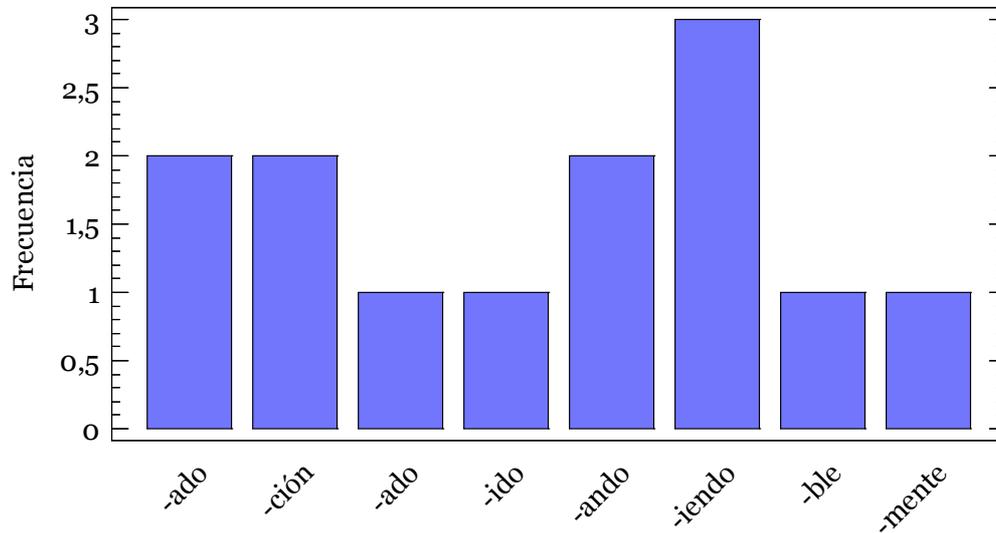


Gráfica 45: Expresiones con o sin elementos deverbales en inglés

5.1.2. Elementos deverbales en español

A diferencia del inglés se encontraron 4 categorías deverbales: sustantivos (terminados en “-ado” y “-ción”), participios verbales (terminados en -ado, -ido, -ando e -iendo), adverbios (terminados en -mente) y los adjetivos. Los participios

fueron de nuevo los más productivos pero la muestra es demasiado pequeña y los participios eran el 50% de las terminaciones deverbales, he aquí la gráfica:

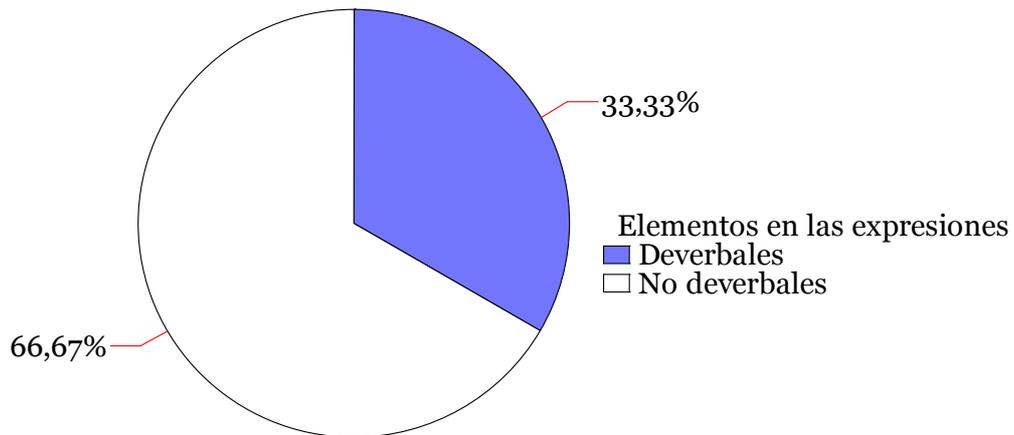


Gráfica 46: Frecuencias de elementos deverbales en español

Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
-ado	2	0,1538	2	0,1538
-ción	2	0,1538	4	0,3077
-ado	1	0,0769	5	0,3846
-ido	1	0,0769	6	0,4615
-ando	2	0,1538	8	0,6154
-iendo	3	0,2308	11	0,8462
-ble	1	0,0769	12	0,9231
-mente	1	0,0769	13	1,0000

Tabla 29: Descripción de elementos deverbales en español

Igual que en inglés mostramos la relación existente entre elementos que contengan elementos deverbales y aquellos que no:



Gráfica 47: Expresión con o sin elementos deverbales en español

La pregunta ahora sería qué hacer entonces con estos elementos que no cumplen este criterio, tendríamos que tener en cuenta que hemos seleccionado 5 de los 6 criterios postulados inicialmente y que dos de ellos ya fueron comprobados, el cumplimiento de uno sólo de los criterios restantes sería la mayoría absoluta y podríamos partir de ese punto para determinar qué expresiones son UFE y qué otras no lo son.

En este trabajo de maestría solo daremos cuenta de la evaluación de estos criterios y de los datos estadísticos arrojados de la muestra, sin embargo, no encontramos un número de criterios a cumplir por una expresión para ser denominada como UFE. Este interrogante queda abierto como una posible extensión al presente trabajo.

5.1.3. Fijación de las expresiones en inglés

Los últimos dos criterios a evaluar –el 3 y el 4– serán los de fijación sintáctica y regularidad compositiva. Para ambos hemos diseñado un par de tablas que dan cuenta de dos aspectos: su frecuencia en la web –por medio de una búsqueda en *Google*– y una posterior confirmación de esto corroborando los hallazgos en un corpus de especialidad como lo el *Ngram viewer*⁹ de *Google*.

En la siguiente tabla podemos evidenciar el número de apariciones, el cual se lista por número de frecuencias de la menor a la mayor:

#	Expresión	Google	Ngram Viewer
1	served by a liner vessel	1	No
2	dispatch payable both ends on laytime saved	2	No
3	find an application admissible	2	No
4	dispatch payable both ends all time saved	4	No
5	flag of convenience free of charge	4	No
6	CIF price free-at-frontier	5	No
7	fast as can customs of the port	5	No
8	half dispatch laytime saved both ends	5	No
9	weather working days Sundays and holidays and holidays included	6	No
10	weather working days Friday and holidays excluded	7	No
11	dispatch half demurrage on actual time saved both ends	8	No
12	handtruck here	117	No
13	cost and freight cleared	169	No
14	per quickest route	289	No
15	no risk until on board	401	No
16	strapped corded and sealed	708	No
17	weather working days Sundays and holidays excluded	1370	No

⁹ <https://books.google.com/ngrams>

18	dispatch half demurrage on working time saved both ends	1600	No
19	dispatch money payable at half demurrage rate	1640	No
20	knocked down in less than carload lots	1720	No
21	cost insurance and freight landed	2710	No
22	half dispatch working time saved	3670	No
23	before breaking bulk	4170	Si
24	knocked down in carload lots	4270	No
25	hours purposes	5220	No
26	free in liner out	6190	No
27	under usual reserve	6220	No
28	for declaration purposes only	7110	No
29	cost and freight landed	7960	No
30	barge aboard catamaran	8920	No
31	order notify bill of lading	9830	Si
32	payment against documents	13200	Si
33	free astray	15800	Si
34	requisite instructions for Customs	16300	No
35	till countermanded	16500	Si
36	picked port	17300	No
37	pier-to-house	17400	No
38	free of capture and seizure riots and civil commotions	24700	No
39	liner in liner out	25400	No
40	no risk after landing	26700	No
41	storage in transit	27200	Si
42	Sundays and holidays excepted	27300	Si
43	whether Customs cleared or not	27500	No
44	shipper's load and tally	27700	No
45	Saturdays Sundays holidays excluded	28500	No
46	whether in free pratique or not	31700	No
47	foreign first	32600	Si
48	house-to-pier	35100	No
49	franco domicile	35700	Si
50	no risk after shipment	36900	No
51	affect the operation of the agreement	38200	No
52	jettison and washing overboard	48900	Si
53	no value declaration	50800	No
54	free in and out stowed and trimmed	51800	No
55	Fridays and Holidays included	54300	No
56	no carbon required	54600	Si

57	whether in berth or not	57300	Si
58	released value not exceeding	57500	Si
59	weather permitting day	59600	No
60	refer to acceptor	62200	Si
61	on truck or rail	62600	Si
62	smurf money	70000	No
63	free on truck	71000	Si
64	prepare the goods for shipment	71700	Si
65	received for shipment	76400	Si
66	free on wagon	87700	Si
67	acid with care	89300	Si
68	free on car	91900	No
69	damages for detention	94400	SI
70	including particular average	98600	No
71	freight and insurance paid to	105000	No
72	weather working day	116000	Si
73	exploit a patent	127000	Si
74	Fridays and Holidays excepted	127000	No
75	not legally marked	135000	Si
76	see overleaf	153000	Si
77	after sight	155000	Si
78	free in and out and stowed	156000	No
79	freight and demurrage	158000	Si
80	free on steamer	175000	Si
81	free in and out and trimmed	176000	No
82	partially knocked down	184000	Si
83	delivery and redelivery	190000	Si
84	not elsewhere mentioned	191000	Si
85	free in and stowed	198000	No
86	non sufficient funds	202000	No
87	no risk after discharge	232000	No
88	shipper's load and count	243000	Si
89	knocked down flat	250000	Si
90	refer to drawer	268000	Si
91	do not fold	316000	Si
92	theft pilferage non-delivery	323000	No
93	both inclusive	356000	Si
94	estimated time of sailing	358000	No
95	to starboard	361000	Si

Etiquetaje y descripción de unidades fraseológicas en un diccionario bilingüe de comercio internacional

96	no commercial value	365000	Si
97	after date	390000	Si
98	any safe world port	391000	No
99	no value declared	397000	No
100	do not bend	412000	Si
101	semi-knocked down	435000	Si
102	no protest	441000	Si
103	cloud on title	488000	Si
104	months after sight	489000	Si
105	less than truck load	524000	Si
106	station to station	557000	Si
107	Sundays and holidays included	575000	Si
108	no advice	604000	Si
109	not elsewhere included	607000	Si
110	any one voyage	634000	Si
111	fast as can	641000	Si
112	Saturdays Sundays holidays included	646000	No
113	no funds	658000	Si
114	days after sight	677000	Si
115	any safe port in the world	719000	Si
116	said to contain	753000	Si
117	not otherwise enumerated	772000	Si
118	sun protection required	822000	No
119	not otherwise herein provided	847000	Si
120	non delivery	885000	Si
121	on truck	891000	Si
122	free and clear	892000	Si
123	not elsewhere classified	923000	No
124	charges pre-paid	945000	No
125	not elsewhere provided	963000	Si
126	ad valorem	1050000	Si
127	over short and damaged	1090000	No
128	said to weigh	1110000	Si
129	at sight	1250000	Si
130	at arm's length	1260000	Si
131	postage and packing	1440000	Si
132	weather permitting	1490000	Si
133	carriage paid	1550000	Si
134	except as otherwise herein provided	1630000	Si

135	not otherwise specified	1940000	Si
136	null and void	2660000	Si
137	with average	3040000	Si
138	leakage and breakage	3090000	No
139	not specifically provided for	3220000	Si
140	days all purposes	3280000	No
141	nearest international airport	3340000	Si
142	any one vessel	3920000	Si
143	No license required	4640000	Si
144	no risk	5000000	Si
145	charges paid	5280000	Si
146	both dates inclusive	5380000	Si
147	errors and omissions excepted	5900000	Si
148	both ends	6910000	Si
149	rate to be agreed	7350000	Si
150	port to port	9330000	Si
151	weight and measurement	12600000	Si
152	not elsewhere specified	13600000	Si
153	en route	15200000	Si
154	door-to-door	17800000	Si
155	not otherwise provided	20700000	Si
156	for further instructions	20900000	Si
157	on board	36300000	Si
158	order of	44400000	Si
159	said to	49300000	Si
160	to be nominated	59000000	Si
161	with effect from	115000000	Si
162	to be advised	314000000	Si

Tabla 30: Frecuencias en *Google* y en el *Ngram Viewer* de las expresiones en inglés

Los primeros once casos presentan una diferencia significativa de frecuencias entre ellos y el caso doce y siguientes siendo el primero de ellos un caso de *happax legomena* total, no obstante las otras diez expresiones deberían ser tratadas de forma similar llegando a considerar la posibilidad de que estos datos den pruebas de su

falta de fijación en la lengua de especialidad. El 58.6 % de las expresiones aparecen listadas en el *Ngram Viewer*.

La regularidad composicional la comprobamos empleado dos recursos electrónicos especializados bilingües para este propósito: *IMF Terminology*, *Routledge Spanish Dictionary of Business* y *Terminological Dictionary English – Spanish. Spanish – English*, buscando cada una de las expresiones en los recursos con los siguientes resultados:

#	Expression	Routledge (parcial)	IMF	Term. Dict
1	served by a liner vessel	No	No	No
2	dispatch payable both ends on laytime saved	Parcial	No	No
3	find an application admissible	No	No	No
4	dispatch payable both ends all time saved	No	No	No
5	flag of convenience free of charge	Parcial	No	No
6	CIF price free-at-frontier	No	No	No
7	fast as can customs of the port	No	No	No
8	half dispatch laytime saved both ends	Parcial	No	No
9	weather working days Sundays and holidays included	Parcial	No	No
10	weather working days Friday and holidays excluded	Parcial	No	No
11	dispatch half demurrage on actual time saved both ends	Parcial	No	No
12	handtruck here	No	No	No
13	cost and freight cleared	No	No	No
14	per quickest route	No	No	No
15	no risk until on board	No	No	No
16	strapped corded and sealed	No	No	No
17	weather working days Sundays and holidays excluded	Parcial	No	No
18	dispatch half demurrage on working time saved both ends	Parcial	No	No
19	dispatch money payable at half demurrage rate	Parcial	No	No

20	knocked down in less than carload lots	No	No	No
21	cost insurance and freight landed	No	No	No
22	half dispatch working time saved	No	No	No
23	before breaking bulk	Parcial	No	No
24	knocked down in carload lots	No	No	No
25	hours purposes	Parcial	No	No
26	free in liner out	Parcial	No	No
27	under usual reserve	No	No	No
28	for declaration purposes only	No	No	No
29	cost and freight landed	No	No	No
30	barge aboard catamaran	Parcial	No	No
31	order notify bill of lading	Parcial	No	No
32	payment against documents	No	No	No
33	free astray	No	No	No
34	requisite instructions for Customs	No	No	No
35	till countermanded	No	No	No
36	picked port	No	No	No
37	pier-to-house	Si	No	No
38	free of capture and seizure riots and civil commotions	Parcial	No	No
39	liner in liner out	No	No	No
40	no risk after landing	No	No	No
41	storage in transit	No	No	No
42	Sundays and holidays excepted	Parcial	No	No
43	whether Customs cleared or not	No	No	No
44	shipper's load and tally	No	No	No
45	Saturdays Sundays holidays excluded	Parcial	No	No
46	whether in free pratique or not	No	No	No
47	foreign first	No	No	No
48	house-to-pier	No	No	No
49	franco domicile	No	No	No
50	no risk after shipment	No	No	No
51	affect the operation of the agreement	No	No	No
52	jettison and washing overboard	No	No	No
53	no value declaration	No	No	No
54	free in and out stowed and trimmed	Parcial	No	No
55	Fridays and Holidays included	Si	No	No
56	no carbon required	No	No	No
57	whether in berth or not	No	No	No
58	released value not exceeding	No	No	No

Etiquetaje y descripción de unidades fraseológicas en un diccionario bilingüe de comercio internacional

59	weather permitting day	No	No	No
60	refer to acceptor	Si	No	No
61	on truck or rail	No	No	No
62	smurf money	No	No	No
63	free on truck	Si	No	No
64	prepare the goods for shipment	No	No	No
65	received for shipment	No	No	No
66	free on wagon	Parcial	No	No
67	acid with care	Parcial	No	No
68	free on car	Parcial	No	No
69	damages for detention	No	No	No
70	including particular average	No	No	No
71	freight and insurance paid to	Parcial	No	No
72	weather working day	No	No	No
73	exploit a patent	Parcial	No	No
74	Fridays and Holidays excepted	Si	No	No
75	not legally marked	No	No	No
76	see overleaf	No	No	No
77	after sight	Parcial	No	No
78	free in and out and stowed	Parcial	No	No
79	freight and demurrage	Si	No	No
80	free on steamer	No	No	No
81	free in and out and trimmed	Parcial	No	No
82	partially knocked down	Parcial	No	No
83	delivery and redelivery	Si	No	No
84	not elsewhere mentioned	Parcial	No	No
85	free in and stowed	Si	No	No
86	non sufficient funds	Si	No	No
87	no risk after discharge	No	No	No
88	shipper's load and count	No	No	No
89	knocked down flat	No	No	No
90	refer to drawer	Parcial	No	No
91	do not fold	No	No	No
92	theft pilferage non-delivery	No	No	No
93	both inclusive	No	No	No
94	estimated time of sailing	Si	No	No
95	to starboard	No	No	No
96	no commercial value	Si	No	No
97	after date	Si	No	No

98	any safe world port	No	No	No
99	no value declared	Si	No	No
100	do not bend	No	No	No
101	semi–knocked down	Parcial	No	No
102	no protest	No	No	No
103	cloud on title	No	No	No
104	months after sight	Si	No	No
105	less than truck load	No	No	No
106	station to station	No	No	No
107	Sundays and holidays included	Parcial	No	No
108	no advice	Si	No	No
109	not elsewhere included	Parcial	No	No
110	any one voyage	Parcial	No	No
111	fast as can	No	No	No
112	Saturdays Sundays holidays included	Parcial	No	No
113	no funds	No	No	No
114	days after sight	Si	No	No
115	any safe port in the world	No	No	No
116	said to contain	No	No	No
117	not otherwise enumerated	No	No	No
118	sun protection required	No	No	No
119	not otherwise herein provided	Parcial	No	No
120	non delivery	No	No	Si
121	on truck	Parcial	No	No
122	free and clear	Si	No	No
123	not elsewhere classified	No	Si	No
124	charges pre–paid	Parcial	No	No
125	not elsewhere provided	Parcial	No	No
126	ad valorem	Si	No	No
127	over short and damaged	No	No	No
128	said to weigh	No	No	No
129	at sight	Parcial	No	Si
130	at arm’s length	Si	No	No
131	postage and packing	Si	No	No
132	weather permitting	No	No	No
133	carriage paid	Si	No	No
134	except as otherwise herein provided	Parcial	No	No
135	not otherwise specified	No	No	No
136	null and void	No	No	No

137	with average	No	No	No
138	leakage and breakage	Si	No	No
139	not specifically provided for	No	No	No
140	days all purposes	No	No	No
141	nearest international airport	Parcial	No	No
142	any one vessel	Si	No	No
143	No license required	No	No	No
144	no risk	No	No	No
145	charges paid	No	No	No
146	both dates inclusive	Si	No	No
147	errors and omissions excepted	No	No	No
148	both ends	No	No	No
149	rate to be agreed	No	No	No
150	port to port	Si	No	No
151	weight and measurement	No	No	No
152	not elsewhere specified	Si	Si	No
153	en route	No	No	No
154	door-to-door	Parcial	No	Si
155	not otherwise provided	No	No	No
156	for further instructions	Si	No	No
157	on board	Parcial	No	No
158	order of	Si	No	No
159	said to	No	No	No
160	to be nominated	No	No	No
161	with effect from	No	No	No
162	to be advised	No	No	No

Tabla 31: Aparición de expresiones en inglés en otros recursos lexicográficos en inglés

Solo para tomar el primer diccionario que fue el que mayores resultados arrojó nos encontramos con que el 48,3% de las expresiones aparecen en este diccionario. Es notable que a nivel de fijación y regularidad composicional estamos rondando un poco menos del 50%. En los otros dos recursos: el *IMF Terminology* y el *Terminological Dictionary English* tan sólo aparecen 2 expresiones (1.2%) en el primero y 3 (1.8%) en el segundo.

Valdría la pena contrastar estas expresiones con diccionarios generales de lengua, de esta forma podríamos corroborar o descartar la hipótesis de que algunas de estas expresiones que se hallan en diccionarios especializados hayan tenido su génesis en la lengua general y poco a poco hayan migrado a las lenguas de especialidad con significados específicos. Esta será otra posibilidad de ampliación del presente trabajo.

5.1.4. Fijación de las expresiones en español

En español al igual que en inglés se hizo la misma búsqueda, primero en Internet y posteriormente en los diccionarios. Estos son los resultados:

#	<i>Expresión</i>	<i>Google</i>	<i>Ngram Viewer</i>
1	antes de fraccionar la carga	0	No
2	crystal, manipúlese con cuidado	0	No
3	remítase al aceptante	0	No
4	remítase el importe junto con el pedido	0	No
5	tal y como se comunicó o notificó en su momento	0	No
6	sólo en avión carguero	4	No
7	corriendo los gastos por cuenta propia	5	No
8	esté o no atracado	7	No
9	únicamente a efectos de declaración	7	No
10	deducidos una vez	403	No
11	a prueba de las inclemencias meteorológicas	1360	No
12	ábrase aquí	1660	No
13	corriendo cierto riesgo	17700	No
14	sírvanse acusar recibo	31900	No
15	corriendo los gastos por cuenta de	62300	No
16	nos complace comunicarle	72300	No
17	véase al dorso	73600	No
18	este lado hacia arriba	78100	No
19	a estribor	165000	Si

20	entra en vigor a partir de que	233300	No
21	accesible en todo momento	247000	Si
22	en buena y debida forma	292000	Si
23	en buen estado y condiciones	407000	No
24	a cuestras	746000	Si
25	en buen fin	1160000	Si
26	en buenas condiciones para navegar	1400000	Si
27	alegando que	1710000	Si
28	toma y daca	1990000	Si
29	contando con	2460000	Si
30	basándose en	2700000	Si
31	a caballo	3390000	Si
32	en contraposición a	8190000	Si
33	tan rápido como	17700000	Si
34	a través	67900000	Si
35	a la par	153000000	Si
36	en virtud de	299000000	Si

Tabla 32: Frecuencias en *Google* y en el *Ngram Viewer* de las expresiones en español

Si en inglés los resultados de *happax* dejaban ciertas dudas con respecto a los criterios que emplearon inicialmente quienes hicieron el diccionario, encontrar que 5 expresiones de estas no se hallaban en toda la red, ya se convierte en un factor delicado a tener en cuenta, como base para una crítica en la rigurosidad de la creación de este recurso lexicográfico. El 44,4% de las expresiones antes mencionadas aparecen en el *Ngram Viewer*, tanto en inglés como en español las apariciones en esta plataforma eran mayores en la medida que sus apariciones en el motor de búsqueda de Google también fueran de una alta frecuencia.

#	Expresión	Routledge (parcial)	IMF	Terminological dictionary
1	ábrase aquí	No	No	No

2	accesible en todo momento	Parcial	No	No
3	alegando que	No	No	No
4	antes de fraccionar la carga	No	No	No
5	basándose en	No	No	No
6	en buen estado y condiciones	No	No	No
7	en buen fin	No	No	No
8	en buena y debida forma	No	No	No
9	en buenas condiciones para navegar	No	No	No
10	a caballo	No	No	No
11	contando con	No	No	No
12	en contraposición a	No	No	No
13	corriendo cierto riesgo	No	No	No
14	corriendo los gastos por cuenta de	No	No	No
15	corriendo los gastos por cuenta propia	No	No	No
16	crystal, manipúlese con cuidado	No	No	No
17	a cuestras	No	No	No
18	deducidos una vez	No	No	No
19	entra en vigor a partir de que	No	No	No
20	este lado hacia arriba	No	No	No
21	esté o no atracado	No	No	No
22	a estribor	No	No	No
23	a prueba de las inclemencias meteorológicas	No	No	No
24	nos complace comunicarle	No	No	No
25	a la par	Parcial	No	No
26	remítase al aceptante	No	No	No
27	remítase el importe junto con el pedido	No	No	No
28	sírvanse acusar recibo	No	No	No
29	sólo en avión carguero	No	No	No
30	tal y como se comunicó o notificó en su momento	No	No	No
31	tan rápido como	Parcial	No	No
32	toma y daca	No	No	No
33	a través	Parcial	No	No
34	únicamente a efectos de declaración	No	No	No
35	véase al dorso	No	No	No
36	en virtud de	Parcial	No	No

Tabla 33: Aparición de expresiones en inglés en otros recursos lexicográficos en español

La consulta en los diccionarios, antes que aclarar el panorama, nos vuelve a dejar la sensación de toparnos con una prueba más de la falta de rigurosidad de la elaboración de este recurso, tan solo el 13,9 % de las expresiones aparecen repetidas parcialmente en los recursos de contraste, así pues quedan expuestos los datos y resultados de las expresiones marcadas con *fr* y sus criterios de selección.

5.2. Hallazgos complementarios

Aunque el objetivo principal de este proyecto era dar cuenta del marcaje de las UFE en algunas etapas de la realización de éste, nos topamos con un gran número de gazapos y errores metodológicos que hemos querido presentar a modo de crítica y reflexión con respecto a la elaboración de recursos lexicográficos especializados, para expertos y mediadores lingüísticos.

5.2.1. Evaluación formal del diccionario

A continuación se explorarán algunos detalles referentes a la macroestructura y la microestructura del diccionario.

5.2.1.1. Análisis de la macroestructura

Uno de los aspectos más notorios del *DCI*, con respecto a la macroestructura, es la carencia de prefacio, de introducción y de manual de uso. Apenas hay un prólogo de cortesía titulado “Fuentes donde se listan las abreviaturas empleadas”, los agradecimientos y las fuentes consultadas.

Sin estos datos no se tiene información sobre la metodología usada en el diseño y elaboración del *DCI*, ni mucho menos información sobre los criterios que se siguieron para la selección terminológica, lo que, como señala Fuentes (2005, p. 77), son indicios para calificar la fiabilidad y la calidad de un diccionario.

Sobre las abreviaturas empleadas en el *DCI*, aparecen las marcas diasistemáticas –diatópicas y diatécnicas– abreviadas en ambas lenguas, seguidas de su forma compleja entre paréntesis. Luego aparecen algunos términos relacionados con el área, a modo de ejemplo, sin aportar mayor información lingüística sobre el área temática.

5.2.1.2. Rastreo de las fuentes citadas

En relación a las fuentes empleadas para la extracción de términos y definiciones, decidimos rastrear los recursos electrónicos que permitieran hacer una descripción objetiva y sistemática.

Sobre las fuentes consultadas que aparecen en Internet, 22 en total, realizamos la consulta con el motor de búsqueda de *Google* y con la aplicación *Wayback Machine* de *Internet Archive*¹⁰, la cual permite examinar un gran archivo de páginas web almacenadas desde 1996. Cabe anotar que dichas fuentes no cumplen con ningún estándar para citar una página web, lo que de entrada pone en duda la fiabilidad de la fuente, sin embargo, se hizo la búsqueda de las fuentes tanto en la web actual como en *Internet Archive* para verificar la situación actual de estas fuentes; al final, solo hallamos 4.

<i>Fuente</i>	<i>Existe hoy en día</i>	<i>Aparece en IA</i>
1. http://www.agecon-ag.ohio-state.edu/ae601/glossary.htm	No	No
2. http://www.anani.co...ssary/trade-finance.htm	No	No
3. http://business-in-asia.com/glossary1.htm	No	Sí
4. http://www.ato.gov.au/content.asp?doc=content/corporate/glossary.htm	No	No
5. http://www.bizmove.com/export/m7q.htm	Sí	Sí
6. http://www.exportsa.sa.gov.au/home/glossary.htm	No	Sí

¹⁰ <http://archive.org/web/web.php>

7.	http://www.globalsources.com/TNLLIST/TRADE/GLOSSARY/TERMS1.htm	Sí	No
8.	http://www.geocities.com/mauduit/glossary.html	No	No
9.	http://www.itds.gov/xp/itds/home.xml	Sí	No
10.	http://www.nextlinx.com/tools_gloss.html	No	Sí
11.	http://www.ntia.doc.gov/lexcon.txt	No	Sí
12.	http://www.otal.com/incoterm.htm	No	No
13.	http://www.pbb.com/wnew/glossary-a.html	No	No
14.	http://www.meetworldtrade.com/en/help/GlossaryA.jsp	No	No
15.	http://members.tripod.com/jblairjr/Glossary/a.html	No	Sí
16.	http://www.seaschott.com/intlterm.html	Sí	Sí
17.	http://sugarweb.co.uk/glossary/alphabetical/a.html	No	Sí
18.	http://www.tradeport.org/ts/refs/gloss/a.html	No	Sí
19.	http://usinfo.state.gov/products/pubs/trade/glossac.htm	No	Sí
20.	http://www.weddispatchers.com/glossary.html	No	No
21.	http://worldcargoalliance.com/Library/globaldata/Acronyms/a.html	No	Sí
22.	http://www.yellowfreight.com/serviceguide/fyi/glossary.html	No	Sí

Tabla 34: Rastreo de fuentes electrónicas

Como vemos, el *DCI* muestra inconsistencias en la forma cómo presenta su macroestructura. La falta de fuentes fácilmente verificables, de instrucciones de uso, de listados de marcas lexicográficas –gramaticales y diasistemáticas–, y la presentación de citas mal elaboradas, entre otros, son aspectos que, si bien no descalifican el diccionario, son indicio de la mala planeación en la elaboración del recurso y dejan muchos interrogantes sobre su elaboración.

5.2.1.3. Lematización

Siguiendo la corriente tradicional de organización, los autores evitan presentar las entradas pluriverbales en la mayor parte del diccionario –aunque aparezcan unas pocas sueltas sin ningún criterio de selección aparente¹¹–, y optan por dejarlas como subentradas; sin embargo, no son claras las reglas que rigen este ordenamiento. Por otra parte, algunas subentradas son tratadas de manera inconsistente, al igual que las derivaciones de algunas entradas; por ejemplo: *operate* presenta *operating* como subentrada, mientras que *tie* y *tying* aparecen como entradas independientes. Este tratamiento es repetitivo a lo largo del diccionario; de nuevo, la carencia de algún tipo de manual de uso se hace evidente, esto sin contar aquellos casos en los que un error –posiblemente mecanográfico– separa una subentrada del artículo lexicográfico convirtiéndola en un entrada como en *mark* y *marking*. Ilustramos a continuación los casos presentados.

tie¹ n/v: GRAL atar, inmovilizar, vincular. [Exp. **tie²** (GRAL empatar), **tie³** (GRAL empate; equivale a *draw*⁵), **tie-in buying** (GRAL operación de compra vinculada; es el sistema por el cual –*whereby*– al comprador de un producto/ser-

¹¹ Es el caso de *cease to be covered by the arrangements* que aparece como entrada, marcada como verbo sin dejar claro qué criterios se usaron para esta elección y por qué no entran otras como *deposit goods in a warehouse*.

tying a: vinculante. [Exp: **tying arrangement**
US (DER acuerdo de lealtad/vinculación,
acuerdo de vinculación condicional, acuerdo
condicional de venta; alude al acuerdo en vir-
tud del cual *–whereby–* el vendedor se com-

operate v: GRAL operar, funcionar. [Exp: **oper-
ate**² (GRAL obrar, proceder), **operate**³ (GRAL
poner en marcha, explotar; equivale a *run*¹),
operate⁴ (LOGÍST pilotar), **operating** (CONTA
operativo; de explotación), **operating differ-
ential subsidy**, **ODS US** (LOGÍST/FIN subven-

Figura 13: Inconsistencia en la lematización de las entradas

mark n/v: GRAL marca, señal; marcar, señalar ◊
*The merchandise was not properly marked
with the foreign country of origin; V. shipping
marks.* [Exp: **mark of origin** (LOGÍST marca
de origen ◊ *Sometimes the marks of origin are
affixed at the time of importation;* equivale a
origin mark), **marked cheque** (FIN cheque
conformado; equivale a *certified cheque, cer-
tified check* en EE.UU.),
marking (LOGÍST marcado ◊ *Marking is intend-
ed mainly to identify shipments; V. cargo*

Figura 14: Separación incorrecta de una subentrada en un artículo lexicográfico

5.2.1.4. Polisemia

En los casos en que se presenta una entrada con dos o más acepciones, se marcan con un número en superíndice. Como en otros aspectos, aquí también se

encuentran errores de marcación, por ejemplo, enunciar una variación sin presentar la siguiente o marcar dos subentradas con el mismo superíndice en el mismo artículo.

set *n/v*: GRAL juego, conjunto ∨ *A full set of original clean on-board bills of lading was required under the irrevocable documentary credit*; fijar, señalar; V. *full set*. [Exp: **set**² (GRAL lote: equivale a *batch, lot, pack*²,

Figura 15: Error de marcaje polisémico

sorio), **sancionar**¹ (LAW sanction, fine, penalize; S. *castigar, multar*), **sancionar**³ (LAW adopt, approve, sanction; S. *aceptar, aprobar, adoptar, asumir; autorizar ratificar*), **sancio-**

Figura 16: Inconsistencia en el consecutivo de la serie polisémica

5.2.1.5. Análisis de la microestructura

Uno de los aspectos fundamentales de la microestructura es la coherencia y adecuación en la organización de los elementos. Dicha coherencia está definida por la metodología empleada en el diseño del diccionario e influye directamente en el producto final.

Las entradas del *DCI* se presentan con un lema en negrita, seguido de la marca gramatical en itálica, las marcas diasistemáticas –diafásica, diatópica, diatécnica– y su equivalente, referencia cruzada, o definición. En algunas entradas aparece un corto contexto, en itálica, marcado con un rombo.

5.2.2. Errores de etiquetaje

Las marcas lexicográficas, al igual que otros elementos del diccionario, se constituyen en una estructura propia y tienen una función definida en el mismo (Fajardo, 1996), además de su utilidad para la organización de la microestructura de la entrada (Barahona, 2005, p. 30). En el caso del *DCI*, pensado para “profesionales y estudiosos de las diversas áreas de la economía [...] (y) de la traducción” dichas marcas toman mayor importancia, pues son un referente casi obligado en el caso de los traductores, aunque, para un usuario común, esta información no entorpece la comprensión del artículo.

Las entradas del *DCI* se organizan alfabéticamente, seguidas de una marca gramatical (en adelante MG) poco rigurosa y en algunos casos ambigua en su sistematización, que –como indicamos anteriormente– no se reseña en la introducción.

Nuestro análisis estadístico presenta 24 distintas combinaciones de MG en el *DCI*. Las gráficas presentadas a continuación son el conteo de las distintas etiquetas gramaticales –simples y combinadas– del diccionario.

Seguida de la marca gramatical y la marca diatécnica, aparece la diastrática *coloq*, sin embargo, no existe ninguna otra como formal o informal. Ahora bien, la aparición de una expresión como ésta no implica encontrar 4 marcas distintas para el *DCI* y no encontrar ni una sola en español.

Expresión	Marca diafásica
box ²	<i>coloq</i>
drawback ²	<i>coll</i>
Greenback	<i>coloq</i>
Harmonized System	<i>colloq</i>
kite ²	<i>coloq</i>
Lloyd's Registry	<i>coloq</i>
reefer ¹	<i>colloq</i>
sales reps	<i>coll</i>
seven sisters	<i>coll</i>
smurf money	<i>col</i>
smurfing	<i>col</i>
swaptions	<i>colloq</i>

Tabla 35: Irregularidad en el uso de la marca diastrática (español-inglés)

La marca diatécnica quizás fue la más empleada tanto en entradas como en subentradas. Las marcas empleadas en el diccionario se sintetizan en la tabla 36.

Sin embargo, estas etiquetas no estuvieron exentas de posibles correcciones. Mientras en la primera parte del diccionario –inglés-español– su uso fue prolijo, sin permitir ambigüedades, la segunda parte tuvo algunas inconsistencias:

<i>Expresión</i>	<i>Marca original</i>	<i>Corrección</i>
asignación de recursos	CONT	ACCTG
cantidad consignada	CONT	ACCTG
conocimiento de embarque combinado	DOC/LOGIST	DOCMT/LOGIST
conocimiento de embarque mixto	DOC/LOGIST	DOCMT/LOGIST
fidelidad	LAWW/MKTG	LAW/MKTG
orden de compra	GEB	-
derechos de salvamento	MAR	-
derechos por salvamento	MAR	-
descuento en el precio facturado por defecto o calidad deficiente del producto	MARK	-

Tabla 36: Inconsistencias en las marcas diatélicas empleadas en el DCI

Del análisis de las MG se pueden sacar las siguientes conclusiones:

a) ambigüedad MG combinatoria.

<i>Entrada</i>	<i>Etiqueta</i>
base	<i>a/n/v:</i>
conglomerate	<i>a/v/n:</i>
duplicate	<i>n/a/v:</i>

Tabla 37: Ejemplo de entradas con MG combinada y sus respectivas variantes (inglés-español)

<i>Entrada</i>	<i>Etiqueta</i>
afiliado	<i>a/n:</i>
futuro	<i>n/a:</i>

Tabla 38: Ejemplo de entradas con MG combinada y sus respectivas variantes (español-inglés)

Aunque podemos observar que este evento ocurre en ambas lenguas las mayores fluctuaciones se presentaron en la primera sección del diccionario –inglés-español–, los ejemplos que presentamos a continuación ilustran esta ambigüedad. Hemos decidido solo tomar un caso y este fue el reseñado en la tabla 37

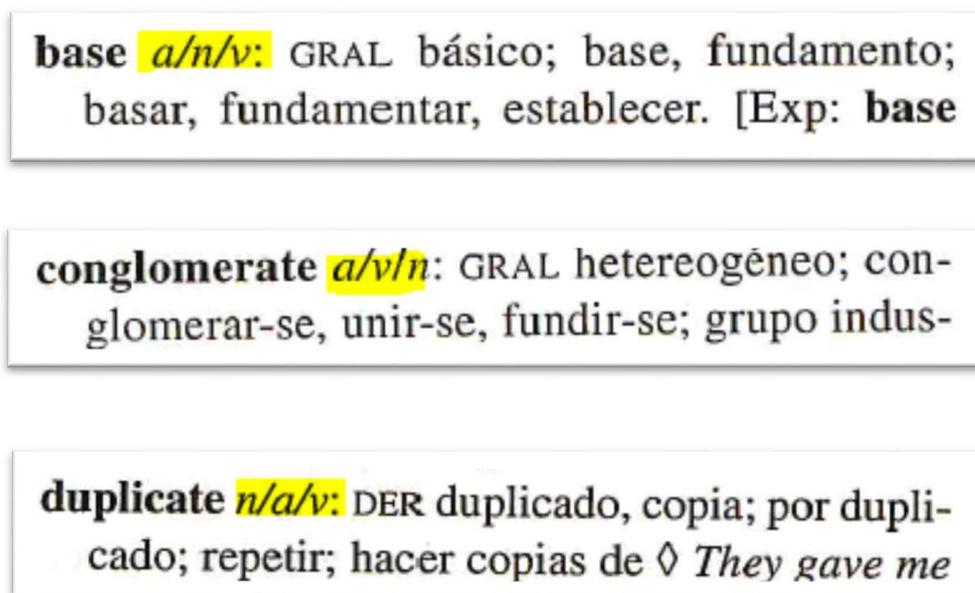


Figura 17: Variantes de MG combinadas

b) **falta de claridad de la MG.** Aunque algunas MG son transparentes por su uso (v=verbo, n=sustantivo, etc.), otras no. Las abreviaturas *fr*, *phrase* y *phr* no son claras, ya que sólo la expresión *phrase* puede indicar un sintagma, una locución o una colocación en lengua inglesa. Esto no ocurre solamente en las marcas anteriores; veamos la siguiente recopilación.

Ejemplos de lemas mal marcados	Marcado como	Categoría	Equivalente
Canc	v.	Abrev.	Ver. <i>cancel.</i>

cease to be covered by the arrangements	v.	Sintagma	dejar de estar amparado por un régimen aduanero
convene a meeting	v.	Colocación	convocar una reunión
perfect the sight	v.	Unidades fraseológicas	rellenar debidamente en aduana los documentos correspondientes a la descripción de las mercancías importadas
a/d	a.	Inicialismo	<i>after date</i>
Alien	a.	Adjetivo, sustantivo,	Súbdito extranjero

Tabla 39: Comparación entre marcas del diccionario y categorías reales

Por último, deseamos llamar la atención sobre una de las inconsistencias más notorias en lo que a marcaje se refiere, es debatible la forma cómo se desee indizar las unidades en el diccionario; sin embargo, sea cual sea la opción *fr*, *phr* o *phrase* las expresiones marcadas de este modo deben ser, por ende, expresiones pluriverbales.

5.3. Consideraciones finales

Es innegable la gran cantidad de datos que pueden ser tomados de este diccionario por parte de los expertos de las áreas implicadas, sin embargo, estas imprecisiones metodológicas en la construcción de este recurso lexicográfico, no pasan por alto si el usuario que desea consultarlo es un lingüista en búsqueda de información lexicográfica, o de un mediador lingüístico –como los traductores e intérpretes– que sirva para extraer información sintáctica y de uso de las expresiones allí contenidas.

La carencia total de un manual de uso e incluso de un posible manual metodológico del proceso de creación del recurso lexicográfico, son dos elementos con los que la editorial queda en deuda con sus compradores y usuarios del diccionario evaluado, esto sin hablar por supuesto del marcaje general y en especial el de las UFE. Una posible solución a esta problemática, sería repensar la forma cómo se generan los diccionarios, más aún cuando en lengua española la literatura metalexigráfica es prolifera no sólo en cantidad de autores, libros y manuales sino también en la calidad de la información suministrada por los ellos.

Entendemos que es imposible unificar el marcaje y el etiquetaje de los distintos diccionarios y recursos afines, debido a la gran cantidad de editoriales y responsables de su concepción, pero la falta de rigurosidad es algo que debería preocupar a las editoriales y sus miembros, por el impacto que esto puede causar en la imagen sobre el usuario final de la herramienta de consulta.

Reiteramos que al momento de finalizar este trabajo de grado no hemos hallado dentro de la literatura consultada, un determinado número de criterios a cumplir para que la expresión a evaluar sea o no considerada una UF o UFE.

6. Referencias

- Aguado de Cea, G. (2007). La fraseología en las lenguas de especialidad. *Las lenguas profesionales y académicas* (pp. 53-65). Barcelona: Ariel.
- Alcaraz, E., & Castro, J. (2007). *Diccionario de Comercio Internacional: Importación Y Exportación: Inglés-Español, Spanish-English*. Barcelona: Editorial Ariel, S.A.
- Alonso, M. (2006). *Diccionarios y fraseología*. La Coruña: Universidade da Coruña.
- Anglada, E. (2005). *Lexicografía española*. Barcelona, España: Publicacions i Edicions de la Universitat de Barcelona.
- Antunes, S. (2006). Typologies of MultiWord Expression Revisited: A Corpus-driven Approach *Spoken Language Corpus and Linguistic Informatics* (pp. 227-244). Amsterdam/philadelphia: John Benjamins Publishing Company.
- Arnzt, R., & Picht, H. (1995). *Introducción a la terminología*. Madrid: Fundación Germán Sánchez Ruipérez.
- Bally, C. (1905). *Précis de stylistique*. Genève: Eggiman.
- Barahona, A. (2005). La marcación en la lexicografía actual: *Káñina, Rev. Artes y Letras* (Vol. XXIX, pp. 29-52): Univ. Costa Rica.

Bevilacqua, C. (1996). A fraseología jurídico-ambiental [*Dissertação de Mestrado*].

Porto Alegre: Universidade Federal do Rio Grande do Sul.

Bevilacqua, C. (1999). *Unidades fraseológicas especializadas: estado de la cuestión y perspectivas*. Pompeu Fabra, Barcelona.

Bevilacqua, C. (2004). *Unidades fraseológicas especializadas eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Pompeu Fabra, Barcelona.

Blais, E. (1993). La phraséologie. Une hypothèse de travail. *Terminologies nouvelles*, 10, 43-49.

Bloomfield, L. (1933). *Language*: New York.

Burger, H. (2007). *Phraseology: an International Handbook of Contemporary Research*. New York: Walter de Gruyter.

Burgos, D. (2006). Concept and Usage Based Approach for Highly Specialized Technical Term Translation. *Linguistic Insights: Insights into Specialized Translation*, 46.

Cabré, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida.

Cabré, M. T. (1999). *La terminología. Representación y comunicación*. Barcelona: Instituto universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

- Cabré, M. T., Lorente, M., & Estopà, R. (1996). Terminología y fraseología. *Actas del V Simposio de Terminología Iberoamericana*, 37-81.
- Cardero, A. (2003). *Terminología y procesamiento*. México: UNAM.
- Carneado Moré, Z. V., & Tristá, A. M. (1985). *Estudios de fraseología*. La Habana: Academia de Ciencias de Cuba, Instituto de Literatura y Lingüística.
- Casares, J. (1950). *Introducción a la lexicografía española*. Madrid: CSIC.
- Corpas, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- Cowie, A. (2001). Theoretical Perspectives: The Russian Tradition. *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Crystal, D. (1995) *The Cambridge Encyclopedia of the English Language* (First ed., pp. 489).
- Chafe, W. (1968). Idiomaticity as an Anomaly in the Chomskyan Paradigm. *Foundations of Language* 4, 109-127.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- de Schaetzen, C. (1993). Un accès rapide aux collocations *Terminologies nouvelles* (Vol. 10, pp. 103-114). Bruselas: RINT.
- Desmet, I. (1994). Propositions pour la recherche en phraséologie contrastive *Banque des mots* (Vol. N° special 6, pp. 45-59). Paris: Presses Universitaires de France.

Fajardo, A. (1996). Las marcas lexicográficas: concepto y aplicación práctica en la lexicografía española. *Revista Lexicográfica, III*(1996-1997), 31-57.

Fedor de Diego, A. (1995). *Terminología: teoría y practica*. Caracas: Equinoccio.

Felber, H. (1984). *Terminology Manual*. Paris: Unesco.

Firth, J. (1957). *Papers in linguistics, 1934-1951*: Oxford University Press.

Fuentes, M. T. (2005). Algunos apuntes sobre la evaluación de diccionarios. *Revista de Lexicografía*(2004-2005), 69-80.

García-Page, M. (2008). *Introducción a la fraseología española*. Barcelona: Anthropos.

González, M. I. (2006). La definición lexicográfica de las unidades fraseológicas: la aplicación de modelos formales. In M. Alonso (Ed.), *Diccionarios y fraseología* (pp. 221-233). La Coruña: Universidade da Coruña.

Gouadec, D. (1994). Extraction, description, gestion et exploitation des entités phraséologiques. *Terminologies nouvelles, 10*, 83-91.

Gréciano, G. (1993). Vers une modélisation phraséologique: acquis et projets d'EUROPHRAS. *Terminologies nouvelles, 10*, 16-22.

Haensch, G. (1997). *Los diccionarios del español en el siglo XXI*. Salamanca: Ediciones Universidad de Salamanca.

Hartmann, R. R. K., & James, G. (Eds.). (2002). New York: Routledge.

Heid, U. (1992). Décrire les collocations: deux approches lexicographiques et leur application dans un outil informatisé. *Terminologie et traduction* (Vol. 2/3, pp. 523-548). Bruselas: Commission des Communautés Européennes, Service de Traduction.

Heid, U. (1994). On Ways Words Work Together - Topics in Lexical Combinatorics. In W. Martin (Ed.), *Euralex'94 Proceedings* (pp. 226-257). Amsterdam.

Heid, U. (1998). Towards a corpus-based dictionary of German noun-verb collocations. In T. Fontenelle (Ed.), *Euralex'98 Proceedings* (Vol. II, pp. 513-522). Liège: Université de Liège.

Heid, U., & Freibott, G. (1991). Collocations dans une base de donnée terminologique et lexicale. *Meta*, 1(36), 77-91.

Hockett, C. F. (1958). *A Course in Modern Linguistics*. New York.

Householder, F. W. (1959). On Linguistic Primes *Word* 15 (pp. 231-239).

L'Homme, M.-C. (1998). Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale. In T. Fontenelle (Ed.), *Euralex'98 Proceedings* (Vol. II, pp. 513-522). Liège: Université de Liège.

L'Homme, M.-C. (2000). Understanding Specialized Lexical Combinations. *Terminology*, 6(1), 89-110.

L'Homme, M.-C., & Bertrand, C. (2000). Specialized Lexical Combinations: Should They Be Described as Collocations or in Terms of Selectional Restrictions?

Proceedings. Ninth Euralex International Congress (pp. 497-506). Stuttgart: Stuttgart.

Lainé, C., Pavel, S., & Boileau, M. (1992). La phraséologie - nouvelle dimension de la recherche terminologique. *Travaux du module canadien du RINT L'Actualité terminologique* (Vol. 25/3, pp. 5-9). Canada: Bureau de Traduction.

Laporte, I., & L'Homme, M.-C. (1997). Exemple d'application dans le domaine de la pharmacologie cardiovasculaire. *Terminologies nouvelles*(16), 95-101.

López, X. (2012). *Fraseología española de origen latino y motivo grecorromano*. Universitat de Lleida, Lleida.

Lorente, M. (2001). Terminología y fraseología especializada: del léxico a la sintaxis. *Terminología*, 159-180.

Lorente, M., Bevilacqua, C., & Estopà, R. (1998). El análisis de la fraseología especializada mediante elementos de la lingüística actual *Actas del VI Simposio Iberoamericano de Terminología*. Cuba.

Martínez, J. (2002). La fraseología en J. Casares. *Estudios de Lingüística Universidad de Alicante*, 139-189.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. New York: SUNY Publ.

Mel'čuk, I. (2001a). Collocations and Lexical Functions *Phraseology: Theory, Analysis, and Applications*. (pp. 23-53). Oxford: Oxford University Press.

Mel'čuk, I. (2001b). Fraseología y diccionario en la lingüística moderna *Isabel Uzcanga Vivar et al. (eds.) Presencia y renovación de la lingüística francesa* (pp. 267 - 310). Salamanca: Universidad de Salamanca.

Mellado, C. (2008). *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang.

Norrick, N. (2007). English Phraseology. In H. Burger (Ed.), *Phraseology* (Vol. II, pp. 615-619). New York: Walter de Gruyter.

Ortega, G., & Gonzáles, M. I. (2008). La técnica fraseográfica: el DRAE (2001) frente al DEA (1999). In C. Mellado (Ed.), *Colocaciones y fraseología en los diccionarios* (pp. 232-245). Frankfurt Am Main: Peter Lang.

Palomar, V. (2004). La importancia de la normalización terminológica *Las palabras del traductor* (pp. 67 - 76). Bruselas: ESLEtRA.

Pavel, S. (1993). La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques. *Terminologies nouvelles*, 10, 67-82.

Pavel, S. (2002). *Handbook of Terminology* Retrieved from <http://www.bt-tb.tpsgc-pwgsc.gc.ca/btb.php?pdf=/publications/documents/termino-eng.pdf&cont=100>

Penadés, I. (2006). La información gramatical sobre la clasificación de las locuciones en los diccionarios. In M. Alonso (Ed.), *Diccionarios y fraseología* (pp. 249-259). La Coruña: Universidade da Coruña.

- Pesant, G., & Thibault, E. (1998). Pour une combinatoire phraséologique de la publicité des droits. *Meta: Journal des Traducteurs*, 43/2, 328-331.
- Porto, J. (2002). *Manual de Técnica Lexicográfica*. Madrid: Arco Libros, S.L.
- Quiroz, G. (2008). *Los sintagmas nominales extensos especializados en inglés y en español: descripción y clasificación de un corpus de genoma*. (Tesis Doctoral), Pompeu Fabra, Barcelona.
- Quiroz, G., Mira, G., & Patiño, P. (2013). *De la fraseología general a la especializada. Criterios de identificación en Español*. Manuscrito sin publicar.
- Rakotojoelimaria, A. (2004). *Esbozo de un diccionario de locuciones español-malgache*. Universidad de Alcalá, Alcalá de Henares.
- Rišianová, H. (2010). *Fraseología en el argot español*. Brno.
- Ruiz, L. (1998). *La fraseología del español coloquial*. Barcelona: Ariel.
- Ruiz, L. (2001). *Las locuciones en el español actual*. Madrid: Arco Libros.
- Schlomann, A. (Ed.) (1955) *Elektrotechnik und Elektrochemie: Band II de Illustrierte technische Wörterbücher: In sechs Sprachen; Deutsch, Englisch, Russisch, Französisch, Italienisch, Spanisch / Hrsg. von Alfred Schlomann*. Berlin: Technische Wörterbücher-Verlag.
- Seco, M. (2003). *Estudios de Lexicografía Española*. Madrid, España: Gredos.

Sinclair, J., & Moon, R. (Eds.). (1989) Dictionary of Phrasal Verbs. London: Collins Cobuild.

Sinclair, J., & Moon, R. (Eds.). (1995) Dictionary of Idioms. London: Collins, Cobuild.

Školnílová, P. (2010). *Las colocaciones léxicas en el español actual*. Brno.

Sosiński, M. (2006). *Fraseología comparada del Polaco y del Español: su tratamiento en los diccionarios bilingües*. Universidad de Granada, Granada.

Vinogradov, V. V. (1947). *Ob osnovnij Tipaj fraseologičeskij iedinits v ruskom yazike*. Moskva.

Warner, A. (Ed.) (1966) Internationale Angleichung fachsprachlicher Wendungen der Elektrotechnik. - Versuch einer Aufstellung phraseologischer Grundsätze für die Technik. Berlin: VDE.

Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Frankfurt am Main: Lang.

7. Anexos

7. ANEXOS	186
7.1. ÍNDICE DE ESQUEMAS	187
7.2. ÍNDICE DE FIGURAS	188
7.3. ÍNDICE DE GRÁFICAS	189
7.4. ÍNDICE DE TABLAS	192

7.1. Índice de esquemas

Esquema 1:	Tipología de Casares	23
Esquema 2:	Clasificación de las locuciones para Casares	24
Esquema 3:	Clasificación de las UF para Zuluaga	25
Esquema 4:	Taxonomía de fraseologismos para Carneado Moré y Tristá	27
Esquema 5:	Clasificación de los fraseologismos verbales según Carneado Moré y Tristá	27
Esquema 6:	Clasificación de los fraseologismos reflexivos y propositivos según Carneado Moré y Tristá	28
Esquema 7:	Clasificación de los fraseologismos con el participio hecho y conjuntivos según Carneado Moré y Tristá	28
Esquema 8:	Clasificación de los fraseologismos con los pronombres la/las según Carneado Moré y Tristá	29
Esquema 9:	Clasificación de los fraseologismos nominales según Carneado Moré y Tristá	29
Esquema 10:	Clasificación de los fraseologismos adjetivales según Carneado Moré y Tristá	30
Esquema 11:	Clasificación de los fraseologismos adverbiales según Carneado Moré y Tristá	31
Esquema 12:	Clasificación de los fraseologismos adverbiales de acuerdo a su referencia circunstancial según Carneado Moré y Tristá	31
Esquema 13:	Clasificación de las UF según Corpas	32
Esquema 14:	Taxonomía de las colocaciones según Corpas	33

Esquema 15:	Taxonomía de las locuciones según Corpas	33
Esquema 16:	Taxonomía de los enunciados fraseológicos según Corpas	34
Esquema 17:	Taxonomía de las locuciones según Ruiz Gurillo	36
Esquema 18:	Taxonomía de las locuciones según García-Page	38
Esquema 19:	Clasificación de las locuciones nominales según García-Page	38
Esquema 20:	Clasificación de las locuciones adjetivas según García-Page	39
Esquema 21:	Clasificación de las locuciones adverbiales según García-Page	40
Esquema 22:	Clasificación de las locuciones prepositivas según García-Page	41
Esquema 23:	Clasificación de las locuciones conjuntivas según García-Page	42
Esquema 24:	Clasificación de las locuciones verbales según García-Page	43
Esquema 25:	Clasificación de las locuciones oracionales según García-Page	44
Esquema 26:	Taxonomía de las <i>phrasal units</i> según Crystal	45
Esquema 27:	Taxonomía de <i>phraseemes</i> según Mel'cuk	48
Esquema 28:	Taxonomía de Mel'cuk y sus relaciones sintagmáticas	49
Esquema 29:	Criterios para la categorización de las UF según Lorente	55
Esquema 30:	Denominaciones del término UF por autores	60

7.2. Índice de figuras

Figura 1:	Compilación de corpus en una hoja de cálculo	77
Figura 2:	Procesamiento de entradas en editor de texto por expresiones regulares	79
Figura 3:	Expresiones regulares en <i>EditPlus</i>	80

Figura 4:	Entradas con separador de línea	80
Figura 5:	Tokenización de las entradas para la generación de patrones morfosintácticos	81
Figura 6:	Interfaz gráfica de TreeTagger para Windows	82
Figura 7:	Resultado del etiquetaje de las palabras con TreeTagger	83
Figura 8:	Selección de columna con etiquetas para patrones morfosintácticos	84
Figura 9:	Eliminar espacios utilizando expresiones regulares	84
Figura 10:	Cadena de patrones morfosintácticos con separador	85
Figura 11:	Expresión regular para eliminar separadores	85
Figura 12:	Resultado de los patrones morfosintácticos	86
Figura 13:	Inconsistencia en la lematización de las entradas	169
Figura 14:	Separación incorrecta de una subentrada en un artículo lexicográfico	169
Figura 15:	Error de marcaje polisémico	170
Figura 16:	Inconsistencia en el consecutivo de la serie polisémica	170
Figura 17:	Variantes de MG combinadas	174

7.3. Índice de gráficas

Gráfica 1:	Expresiones con y sin POS	94
Gráfica 2:	Frecuencia de etiquetas en entradas de 2 o más frecuencias	96
Gráfica 3:	Frecuencia de adjetivos (a) en las entradas	98

Gráfica 4:	Frecuencia de adverbios (adv) en las entradas	99
Gráfica 5:	Frecuencia de conjunciones (conj) en las entradas	99
Gráfica 6:	Frecuencia de entradas con la marca <i>fr</i> en las entradas	100
Gráfica 7:	Frecuencia de interjecciones (interj) en las entradas	100
Gráfica 8:	Frecuencia de sustantivos (n) en las entradas	101
Gráfica 9:	Frecuencia de prefijos (pref) en las entradas	101
Gráfica 10:	Frecuencia de preposiciones (prep) en las entradas	102
Gráfica 11:	Frecuencia de pronombres (pron) en las entradas	102
Gráfica 12:	Conteo de los verbos (v) en las entradas	103
Gráfica 13:	Frecuencia de las entradas marcadas como <i>fr</i> por tokens	104
Gráfica 14:	Ubicación de la expresión pluriverbal completa marcada como <i>fr</i>	106
Gráfica 15:	Reconteo de tokens de las expresiones reales marcadas como <i>fr</i> sin repeticiones	109
Gráfica 16:	Patrones morfosintácticos más frecuentes en las expresiones marcadas como <i>fr</i>	111
Gráfica 17:	Frecuencia de las expresiones en las subentradas por token	115
Gráfica 18:	Conteo de tokens de las expresiones no marcadas con comas	116
Gráfica 19:	Patrones morfosintácticos más recurrentes en expresiones con coma	117
Gráfica 20:	Conteo de tokens de las expresiones no marcadas con guiones	119
Gráfica 21:	Frecuencias por token de las expresiones no marcadas sin comas o guiones	122
Gráfica 22:	Patrones morfosintácticos más comunes en expresiones sin comas o guiones	123

Gráfica 23:	Expresiones con y sin POS en español	124
Gráfica 24:	Conteo de etiquetas en entradas de 2 o más frecuencias en español	125
Gráfica 25:	Frecuencia de adjetivos (a) en las entradas	126
Gráfica 26:	Frecuencia de adverbios (adv) en las entradas	127
Gráfica 27:	Frecuencia de conjunciones (conj) en las entradas	127
Gráfica 28:	Frecuencia de sustantivos (n) en las entradas	128
Gráfica 29:	Frecuencia de expresiones marcadas como (<i>phr</i>) en las entradas	128
Gráfica 30:	Frecuencia de prefijos (pref) en las entradas	129
Gráfica 31:	Frecuencia de preposiciones (prep) en las entradas	129
Gráfica 32:	Frecuencia de pronombres (pron) en las entradas	130
Gráfica 33:	Frecuencia de verbos (v) en las entradas	130
Gráfica 34:	Frecuencias para marcas gramaticales simples en español	131
Gráfica 35:	Frecuencia de las expresiones marcadas como <i>phr</i> sin repeticiones	132
Gráfica 36:	Patrones morfosintácticos en las expresiones marcadas como <i>phr</i>	133
Gráfica 37:	Frecuencia de expresiones en las subentradas por token	135
Gráfica 38:	Conteo de tokens de expresiones con coma en español	137
Gráfica 39:	Patrones morfosintácticos más recurrentes en expresiones con coma	137
Gráfica 40:	Conteo de tokens de expresiones con guiones en español	140
Gráfica 41:	Gráfica de patrones morfosintácticos frecuentes de expresiones con guiones	140

Gráfica 42:	Frecuencia de expresiones sin comas y guiones en español por token	143
Gráfica 43:	Patrones morfosintácticos (>1%) en expresiones sin comas o guiones	144
Gráfica 44:	Frecuencias de elementos deverbales en inglés	147
Gráfica 45:	Expresiones con o sin elementos deverbales en inglés	148
Gráfica 46:	Frecuencias de elementos deverbales en español	149
Gráfica 47:	Expresión con o sin elementos deverbales en español	150

7.4. Índice de tablas

Tabla 1:	Expresiones protogénicas de las UF anteriores a Vinogradov y Bally	22
Tabla 2:	Actualización de la clasificación de las locuciones por autores de Ruiz	
Tabla 3:	Términos relacionados a las UF por autores	47
Tabla 4:	Características de las UF de la lengua general y las UF en lenguas de especialidad	58
Tabla 5:	Características generales del diccionario	73
Tabla 6:	Lista de acrónimos y siglas de las marcas diatécnicas empleadas en el diccionario	75
Tabla 7:	Frecuencia de expresiones con y sin marca gramatical	94
Tabla 8:	Frecuencias y porcentajes de etiquetas gramaticales	97
Tabla 9:	Frecuencias para marcas gramaticales simples	102

Tabla 10:	Expresiones marcadas como <i>fr</i> con un solo token	108
Tabla 11:	Expresiones repetidas marcadas como <i>fr</i>	109
Tabla 12:	Frecuencias por número de tokens	110
Tabla 13:	Patrones morfosintácticos de las expresiones marcadas como <i>fr</i>	114
Tabla 14:	Frecuencia de expresiones que contienen comas entre las subentradas por tokens	115
Tabla 15:	Frecuencia de expresiones que contienen comas entre las subentradas por tokens	117
Tabla 16:	Patrones morfosintácticos de las expresiones con comas	118
Tabla 17:	Conteo de frecuencias por tokens para expresiones con guiones	119
Tabla 18:	Patrones morfosintácticos de las expresiones con guiones	121
Tabla 19:	Frecuencias por tokens de unidades con y sin comas	122
Tabla 20:	Frecuencia de expresiones con y sin marca gramatical	124
Tabla 21:	Frecuencias y porcentajes de etiquetas gramaticales en español	126
Tabla 22:	Frecuencias para marcas gramaticales simples en español	131
Tabla 23:	Patrones morfosintácticos en las expresiones marcadas como <i>phr</i>	135
Tabla 24:	Frecuencia de las subentradas por token	136
Tabla 25:	Patrones morfosintácticos de expresiones con coma en español	139
Tabla 26:	Patrones morfosintácticos de expresiones con guion en español	142
Tabla 27:	Frecuencia de unidades con y sin comas por token	143
Tabla 28:	Descripción de elementos deverbales en inglés	148
Tabla 29:	Descripción de elementos deverbales en español	149
Tabla 30:	Frecuencias en <i>Google</i> y en el <i>Ngram Viewer</i> de las expresiones en inglés	155

Tabla 31:	Aparición de expresiones en inglés en otros recursos lexicográficos en inglés	160
Tabla 32:	Frecuencias en <i>Google</i> y en el <i>Ngram Viewer</i> de las expresiones en español	162
Tabla 33:	Aparición de expresiones en inglés en otros recursos lexicográficos en español	163
Tabla 34:	Rastreo de fuentes electrónicas	167
Tabla 35:	Irregularidad en el uso de la marca diastrática (español-inglés)	172
Tabla 36:	Inconsistencias en las marcas diatécnicas empleadas en el DCI	173
Tabla 37:	Ejemplo de entradas con MG combinada y sus respectivas variantes (inglés-español)	173
Tabla 38:	Ejemplo de entradas con MG combinada y sus respectivas variantes (español-inglés)	173
Tabla 39:	Comparación entre marcas del diccionario y categorías reales	175