



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE FILOSOFÍA Y LETRAS
COLEGIO DE LETRAS HISPÁNICAS

ANÁFORAS EN LA EXPANSIÓN DE CONTEXTOS
DEFINITORIOS: UNA PROPUESTA DE ETIQUETADO

T E S I S
QUE, PARA OBTENER EL TÍTULO DE
LICENCIADA EN LENGUA Y LITERATURAS
HISPÁNICAS, PRESENTA

VALERIA AMANDA BENÍTEZ ROSETE

ASESOR: DR. GERARDO EUGENIO SIERRA MARTÍNEZ



CIUDAD UNIVERSITARIA, 2008



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A Trinidad Olvera Gómez, el amor de mi vida, el amor a la vida.
A Trinidad Rosete Olvera, me duele siempre tu inesperada ausencia.*

Agradecimientos

Esta tesis se llevó a cabo gracias al patrocinio de CONACYT, en el marco del proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos* [46832-H], dirigido por el Dr. Gerardo Sierra Martínez.

Agradezco el apoyo económico brindado por el proyecto *Métodos y herramientas computacionales para el análisis y la visualización informétrica* inscrito en el programa *Ciencia y tecnología informática* y parte del Macroproyecto de la UNAM *Tecnologías para la Universidad de la Información y la Computación*.

A mi asesor, el Dr. Gerardo Sierra por la paciencia y todas las veces que me animó a poner punto final a este trabajo; por mostrarme una nueva línea de investigación en el campo de la lingüística y por permitir a sus estudiantes ser parte de su proyecto académico.

A mis compañeros del Grupo de Ingeniería Lingüística por estos años de buenas charlas y de trabajo enriquecedor. Especialmente a César, Carlos Méndez, Adriana y Ariadna. A los que han ido a seguir la vida y los estudios fuera del GIL: Antonio, el Doc. Solares y Fonseca.

A Rodrigo Alarcón que ha sido un gurú en el mundo de los contextos definitorios, este trabajo no hubiera tomado buen rumbo sin sus observaciones y comentarios.

A mis revisores de tesis la Mtra. Margarita Palacios, el Dr. Javier Cuétara y la Dra. Adriana Ávila por acompañarme en la etapa final de mi tesis. Agradezco infinitamente sus correcciones.

Al Dr. Alfonso Medina por el placer de las charlas, por leer mi tesis entre tanto trabajo, por la sonrisa que nos da cuando tocamos a su puerta, por las correcciones de esta tesis que alargaron mis noches de desvelo.

A Yuyú, mi abuelita, mi luz, por el amor incondicional y la prudencia encantadora. Tengo miedo de separarme de ti aunque sé que siempre estás y estarás para calmar mis ansias. TE AMO.

A mi mamichi, Leonila Rosete. No hace falta más que un gesto para que sepas todo de mí, me conoces como nadie. Siempre que te necesito estás ahí, a un lado de mis miedos, acariciando mis desvelos. Gracias por enseñarme que la vida es lo que queremos que sea. Amo tus manos trabajadoras, tu carácter furioso frente a las injusticias, tus libros, las historias, tus lágrimas de mujer-madre-hija-herma-tía. Llevo siempre tu sonrisa de vida bien vivida. Gracias infinitas por apoyar cada una de mis decisiones, las buenas, las malas y las peores.

A Chofis, mi hermana. Ha habido tiempos difíciles, lágrimas estremecedoras, desencuentros y algunas fracturas del alma shhh que ya habrá tiempo de ir sanando. Ha habido también noches de chismes, de risas y charlas inolvidables, ¡Claro! y todo se debe a que eres mi polo opuesto pero complementario.

A Pili, Lauri, Paty, Cocol, Ligia y Maty, las mujeres de mi vida, las guardianas del zaguán negro que lleva el número 63. Un día me fui de casa y me descubrí sin ustedes, entonces

los domingos eran largos y llenos de saudade. Mis queridas mujeres, mi sonrisa la debo a su cariño, me han dado templanza para salir de casa y vivir el mundo que por ahí dicen que es el real, porque el mudo lindo me lo dan ustedes.

A Cocol que estuvo aquí, en cada palabra de esta tesis empapada de lágrimas de muerte.

A Cacho, Gus y Marrantain, mis tíos. Si hubiese tenido que elegir papá los escogería a los tres, sólo que no tuve que escoger, la vida me los dio. Gracias por los retoños Maty y Armando.

A mi tía Ara y a su familia por el apoyo incondicional que no cambia con los años, gracias por cada recuerdo y por los que vendrán.

A José porque tengo una Luna en el alma, la pusiste ahí muy despacio, me la entregaste llena de vida y de ganas de estar sin razones complejas ni operaciones aritméticas, se quedará para siempre, es indestructible como la del cielo, es nocturna como las mariposas que hacemos juntos. Gracias por la eternidad que hacemos juntos y por el sol de todos los días. Por fin puedes mirar mis anáforas de cerca, esos bichos que nos hicieron refunfuñar tanto. Te amo.

A Ana Aguilar, mi carnalita. Extraño tu voz, las preguntas retóricas, cada consejo de la vida, en especial los lingüísticos, aunque sin duda me matan las ganas de chismorrear. Sé que eres feliz lejos, que has trabajado duro para esto. Gracias por enseñarme amorosamente que siempre vale la pena trabajar un ching.... mucho para conseguir lo que queremos. Te amo.

A Isis, ma chiqui. Supongo que con el tiempo París se ido difuminando, he olvidado el nombre de algunas calles, no recuerdo cómo sabe el pain au chocolat y ya no sé como se siente caminar con abrigo. Cuando la memoria me traiciona vienes tú a recordarme que Paris es una fiesta y que me espera un hogar en Odeon. Gracias por apoyarme en la decisión de volver a casa, estoy aquí acabando lo que empecé. Je t'aime ma petite!

A David Sefami, Davo de mi corazón. Lágrimas, sonrisas, baile, viajes, confesiones, llamadas nocturnas es sólo un pedacito de la historia que te hace mi mejor amigo, el chico con quien lloro amargamente cuando me desencanto del mundo. Gracias por darme ánimos para seguir la tesis, el amor, la parranda. Gracias por la posibilidad eterna de crecer y seguir creciendo a tu lado. Nunca me iré sin decir te quiero...

A Cocar, carnalito porque tú sabes que yo sé que nosotros sabemos lo que no debemos saber. Quizá tienes razón, cuando uno quiere dejar de pensar no hace falta más que dejar de pensar, es sencillo pero aún no lo consigo. Gracias por recibirme en casa y cuidar de mí, fue un placer hacer la tesis cerca de los Chompetas.

A Goya, Edgar, Ele-Grillo, Nacho, Sandra, Ajax, Ingrid, Urique, Savor, Chismo, Adrián, Cocar y Davo, los tecelticanos. Gracias porque cuando me canso de la vida, están ahí para llevarme de vacaciones a la niñez. Dicen que hay un lugar de la alegría, ustedes lo han acaparado. Gracias por estos veinte años de vida juntos... busca lo más vital nomás los que has de precisar nomás y olvídate de la preocupación, si buscas lo muy esencial para vivir sin batallar y la naturaleza te lo da...

A Marcos Mohar por acompañarme las mil y una noches en que escribí esta tesis, gracias al Messenger no me sentía abrumadoramente sola. Gracias también por volverte detective de neologismos, en efecto, una parte de mi sueldo te pertenece.

A Lau, la Lunedepluie. Creo que esta tesis pudo acabar hace algunos años de no haberte conocido, en definitiva mi vida hubiera sido muy aburrida sin ti, sin tu París con aguaceros. Gracias por haber desviado mi vida porque hoy no estaría haciendo lo que hago de no haberte seguido a la aventura europea.

A Fer, el primer romiee. Seguramente la bola de la suerte dice que me haces falta, que estoy orgullosa de ti y que cuando sea grande quiero ser como tú, hacer simulaciones y vivir de millonario astrónomo viajero.

A Masaya, por su encantadora sonrisa, porque la vida nos depara todas las vidas juntas. Gracias por los martes de comida, luego miércoles y así. Infinitas gracias por escuchar mis penas y siempre cuidarme con tu cariño, te quiero al infinito.

A Juampi, porque un día hubo sueños compartidos, por cada vez que me hiciste dudar de la lingüística poniendo un poema en el alma...por la eternidad y un día más.

A Mique, por las clases de latín que tomamos juntos, por tus apapachos, por estar siempre en ese lugarcito que sólo va a ser tuyo.

A mis primos Camilo, Chucho y Edmundo porque en los tiempos difíciles puedo llegar a su mesa, comer rico y decir tonterías. Gracias por su música estremecedora y porque no hay rincón en el mundo más disfrutable que el que comparto con ustedes.

A los primos Ana Laura, Yadi, Beto y Luz, gracias porque cuando se me olvida lo Benítez recuerdo que podemos levantar la ceja como nadie. He descuidado mis genes paternos pero ustedes siempre me llevan de regreso a la verdad. Los quiero muchísimo.

A Chayo, Roberto y Rafa Mondragón que hicieron de mis últimos años en la venerable Facultad de Filosofía y Letras una verdadera delicia. Gracias por no desquermarme a pesar de las desatenciones, por fin la tesis se ha quedado atrás. Cuestión de enfoque, ahora a ustedes les queda adelante.

A Carolina Ruiz, por el futuro juntas, por cada unas de las palabras que marcaron el corazón en aquellos tiempos adoloridos. Gracias por enseñarme que la madurez no tiene que ver con la edad y que siempre habrá algo que pueda aprender contigo. Esta tesis es gracias a ti y a tu cariño.

A las chicas COLMEX: Nadi, Pam, Ana, Raquel, Uru, Glen y Are porque me han dado un espacio maravilloso para trabajar. Gracias por festejar mi cumpleaños y cambiar el mal rumbo del fin de año. De las cuatro que quedaban, esta chamaca ya se tituló...

A Nadi, por acogerme calurosamente en tu mundo, porque siempre me acompañas por café y te fumas el humo de mis cigarrillos, ha sido un placer reencontrarte después de la ENAH, después de París y después de mañana.

A Rafa Merino, el último de los Mohicanos, por los primeros años de vida unamita, por aquellos tiempos en que estábamos siempre juntitos. Gracias porque siempre me recuerdas lo mucho que te quiero.

A Sebas y Pancho, mis queridos roomies gracias por alimentar cariñosamente la lombriz de mi barriga, por recordarme a diario que la vida es más divertida que una tesis y que una chela nocturna no hace daño. Gracias por explicarme qué es una fiesta holandesa, estoy prevenida y sé que todo lo que diga podrá ser utilizado en mi contra. Gracias por el hogar que hacemos todos los días.

A Kathya porque entre nosotras las palabras sobran...

A Pavel, el inesperado, porque siempre habrá una cita no planeada para nosotros, siempre unas chelitas aguardando un nuevo lugar de encuentro. Gracias por la historia y el porvenir, por antojarme de hacer una Lunita, pero una de Valuna y de José.

Al Jefe Quinto porque me dio una infancia maravillosa, una adolescencia de soledad. Gracias por comenzar a hacer las cosas diferentes, nunca es demasiado tarde, te quiero padre.

A los amigos extraviados en el descuido: Jorge Noguez, Piña, Iván, Nuche, Bugs, Alex Bruck, Anibal, Jesi, Miquel, Ambroise, Nam, Oswaldo y cualquier otro recuerdo que no haya podido venir para ser agradecido.

A la Facultad de Filosofía y Letras por la experiencia de vida que me ha brindado, por la mejor educación. Agradezco especialmente a la Universidad Nacional Autónoma de México porque le debo más que mi formación académica, mi corazón azul y oro.

Índice

Introducción	1
1. Motivación	2
2. Objetivos de esta tesis	3
3. Organización y estructura de la tesis	4
I. Antecedentes	6
1. Proyecto “Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos”	7
1.1. Aplicaciones	8
1.2. Conceptos básicos para el proyecto	9
2. Problemática en función del criterio de extensión de contexto definitorio	16
3. Problemática en función del tipo de candidato a contextos definitorio	18
4. Corpus para la extracción automática de contextos definitorios	20
5. Discurso y contextos definitorios como entidades discursivas	22
5.1. Cohesión y coherencia textuales	25
6. Observaciones finales	26
II. Mecanismos de referencia discursiva: fenómeno de anáfora	29
1. Referencia	30
2. Expresiones referenciales	32
3. Tipos de referencia discursiva	35
4. Mecanismos de referencia discursiva	38
4.1. Repetición y sustitución	39
4.2. Deixis	41
4.3. Elipsis	42
4.4. Marcadores u operadores discursivos	43
4.5. Foricidad	43

5. Anáfora	44
5.1. El rango conceptual del término <i>anáfora</i>	45
6. Clasificación de anáfora	47
6.1. Anáfora en función del contexto en el que opera	48
6.2. Anáfora en función del tipo de elemento anafórico	49
6.2.1. Anáfora de sintagma nominal (directa e indirecta)	49
6.2.2. Anáfora pronominal	51
6.2.2.1. Pronombres	51
6.3. Anáfora en función del marco en el que ocurre	53
7. Observaciones finales	54
III. Metodología y resultados de la identificación de anáforas en contextos definitorios	56
1. Metodología para la identificación de anáforas en contextos definitorios	57
1.1. Criterios para la identificación de anáforas en contextos definitorios	57
1.1.1. Criterios de clasificación de anáfora	58
1.1.2. Extensión y expansión de contextos definitorios	59
1.1.3. Candidatos completos y candidatos incompletos	59
1.2. Datos que queremos obtener	60
2. Resultados de la identificación de anáforas en contextos definitorios	61
2.1. Conteo de candidatos completos y de candidatos incompletos	61
2.2. Tipología de expresiones (anáforas) presentes en candidatos incompletos	62
2.3. Diferencias entre pronombres y sintagmas nominales como expresiones anafóricas	70
2.4. Anáforas en candidatos completos	72
2.5. Elementos del contexto definitorio inmersos en una relación anafórica	94
2.6. Antecedente	96
2.7. Patrones recurrentes de anáforas en contextos definitorios	98
3. Observaciones finales	104

IV. Etiquetado XML para relaciones anafóricas y otros mecanismos de referencia discursiva en contextos definitorios	106
1. Corpus de contextos definitorios y etiquetado XML	107
2. Etiquetas para relaciones de referencia discursiva en contextos definitorios	110
2.1. Antecedente	111
2.2. Anáfora pronominal	116
2.3. Anáfora de sintagma nominal	119
2.4. NO Antecedente	124
3. Sintagma nominal con determinante posesivo, catáfora y otros tipos de referencia discursiva	127
3.1. Sintagma nominal con determinante posesivo (PS)	127
3.2. Catáfora (CAT)	128
3.3. Otros tipos de referencias discursivas	130
4. Problemas de etiquetado	131
4.1. Cadenas mixtas de referencia	131
4.2. Longitud del antecedente-sintagma y de la anáfora de sintagma nominal	133
4.3. Anáfora de pronombre relativo “que”	134
4.4. Anáforas y usos problemáticos	136
5. Resultados y observaciones finales	138
Conclusiones	141
Bibliografía	149
Apéndice A: Paradigma pronominal del español	157
Apéndice B: Atributos de las etiquetas XML para contextos definitorios	162

Introducción

Introducción

1. Motivación

Actualmente, en el Grupo de Ingeniería Lingüística (GIL) de la UNAM se desarrolla el proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos* [CONACYT 46832-H]. Uno de sus principales objetivos es desarrollar métodos y herramientas computacionales para la extracción automática de información.¹ El proyecto en cuestión está orientado al reconocimiento de patrones lingüísticos útiles para la extracción automática de *términos*² y *definiciones*. En esta línea de investigación han surgido varios temas de interés lingüístico que dan pie a trabajos con límites muy concretos, tal es el caso de la tesis que se desarrolla en las páginas siguientes.

La forma de esquematizar la relación que mantiene una *definición* respecto a un *término*, es considerar que ambos se configuran en una estructura llamada *contexto definitorio* (CD):³ fragmento textual donde se introduce un *término* y su correspondiente *definición* [Alarcón, 2003; Alarcón y Sierra, 2003]:

Tabla 1. Ejemplo de contexto definitorio

Un ciclón tropical es el término científico que se utiliza para definir una circulación meteorológica cerrada que se propaga sobre aguas tropicales
Término = Ciclón tropical
Definición = el término científico que se utiliza para definir una circulación meteorológica cerrada que se propaga sobre aguas tropicales

Con el ejemplo mostramos la estructura básica de un CD, más adelante en el cuerpo de este trabajo detallaremos la descripción de los elementos implicados y la pertinencia de su

¹ Los sistemas de extracción de información (SEI) se desarrollan en el área de la Inteligencia Artificial (AI) y tienen como objetivo, entre otras cosas, la obtención de manera automática de datos y conocimiento provenientes de documentos. Los conocimientos varían de acuerdo con la naturaleza de éstos y por ello los SEI se elaboran con el propósito de extraer información específica [Gonzalo, 2003].

² Un *término* es “el elemento constitutivo de cualquier nomenclatura terminológica que esté relacionada con una lengua de especialidad [...], es la denominación de un objeto, propio de una determinada área de especialidad” [Dubuc, 1992].

³ En adelante se usará CD para referir *contexto definitorio* en singular y CDs para el plural.

análisis lingüístico. Por lo pronto, damos muestra del tipo de textos con el que se trabajará en esta investigación.

En los adelantos logrados hasta ahora en la extracción automática de CDs [Alarcón 2006, Aguilar 2006] se observó que existen fenómenos lingüísticos que permiten hacer referencia a entidades del discurso que han sido antes mencionadas. Se trata de mecanismos discursivos comúnmente empleados por los hablantes para evitar repetir textualmente cierta información. Uno de los mecanismos que cumple con esta función es la *anáfora* o *referencia anafórica* [Moreno *et al.*, 1999: 205], tema central de este trabajo.

Cabe señalar que esta investigación surgió de la necesidad de estudiar ciertos fenómenos de referencia discursiva que no habían sido considerados en los trabajos previos de análisis de CDs. La tesis se ocupará, entre otras cosas, de llevar a cabo una investigación que permita apreciar la importancia del estudio de la información contextual para resolver problemas de ambigüedad y comprensión en los textos. Veremos que los CDs son estructuras discursivas que mantienen una *cohesión* y *coherencia* textuales específicas, hablamos de dos propiedades del discurso estrechamente ligadas con cierto tipo de mecanismos de referencia, entre los cuales se encuentra la *anáfora* [Renkema, 1999].

En esta tesis, entonces, se estudiarán y describirán *mecanismos de referencia discursiva*, concretamente relaciones de referencia anafórica entre el término y otras entidades lingüísticas del mismo CD. Nos interesa estudiar dichos mecanismos en la medida en que, a partir de su descripción, sea posible rastrear y delimitar la extensión de un CD. Por *extensión* entendemos la cantidad de palabras que éste comprende, es decir, el tamaño exacto del fragmento textual que contiene la definición completa de un término.

2. Objetivos de esta tesis

Uno de los objetivos principales de esta investigación es profundizar en el fenómeno de anáfora con el propósito de describir las ocurrencias en CDs y de proponer patrones recurrentes que puedan optimizar los resultados del *Extractor de contextos definitorios* (ECODE)⁴ para el español [Alarcón, 2006], una herramienta computacional que hace

⁴ <http://brangaene.upf.es/ecode/> [Alarcón, 2006]

búsquedas automáticas de CDs. Para lograr nuestro objetivo necesitamos considerar los siguientes puntos:

- Realizar una síntesis de los estudios que se han hecho sobre contextos definitorios y sobre el tema de referencia discursiva y anáforas, ello con el objetivo de plantear la problemática que da pie a esta investigación.
- Determinar el carácter discursivo de los CDs para ver de qué manera los mecanismos referenciales, específicamente las relaciones anafóricas, afectan su estructura.
- Plantear un marco teórico que permita establecer las pautas en la identificación y descripción de relaciones referenciales en CDs, concretamente las de anáfora.
- Identificar las relaciones anafóricas en CDs para describir los tipos más prominentes y explicar la incidencia de dichas relaciones en la extracción automática.

Concluido este proceso se hará una *tipología* de las anáforas, que establecen relación entre términos y otras entidades discursivas dentro de los CDs. Además, como parte de los resultados se propondrán *patrones recurrentes de la manifestación de anáforas* y se formulará un repositorio de aquellas obtenidas en el corpus delimitado para este trabajo. Finalmente se elaborará una *propuesta de etiquetado* para relaciones anafóricas en CDs.

3. Organización y estructura de la tesis

En el Capítulo I se describirán los antecedentes del proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos* y se hablará de los trabajos que se llevan a cabo en esta línea de investigación. Para ello, se definirán los conceptos básicos tanto para el proyecto como para esta tesis. También se describirá detalladamente la problemática que motivó nuestro trabajo, puntualizando en el problema de referencia discursiva y describiendo el tipo de texto (el corpus) con el que se trabajará. Finalmente, se determinará el carácter discursivo

de los CDs y la importancia de estudiar y considerar mecanismos referenciales como parte de su estructura.

En el Capítulo II se planteará un marco teórico para identificar y describir el fenómeno de anáfora en CDs. Se definirán conceptos tales como *referencia*, *referencia lingüística*, *referencia discursiva*, *expresión referencial*, *mecanismos referenciales discursivos*, *repetición*, *sustitución*, *deixis*, *elipsis*, *foricidad*, *relaciones anafóricas*, *anáfora*, *catáfora*, *expresión anafórica*, *antecedente o referente lingüístico*, *tipos de anáfora*, *pronombres*, entre los más importantes. Finalmente se explicarán los criterios de clasificación de anáfora que servirán para identificar y describir relaciones anafóricas en CDs.

En el Capítulo III se explicará la metodología y se presentarán los resultados obtenidos de la identificación de relaciones anafóricas en un total de 250 CDs. Se expondrá una tipología de las anáforas más prominentes y los patrones recurrentes que pueden considerarse para mejorar la herramienta ECODE. Se explicará además la repercusión de las relaciones referenciales en CDs, especialmente las anafóricas, puntualizando en los conceptos de *extensión* y *expansión* en la extracción automática.

En el Capítulo IV se expondrá una propuesta de etiquetado para relaciones anafóricas y otras relaciones de referencia discursiva en CDs. Las etiquetas fueron diseñadas siguiendo las pautas establecidas para el *Corpus de contextos defintorios* que se realiza en el GIL [Sierra *et al.*, 2006]. La propuesta que presentaremos es parte importante de los resultados de este trabajo, ya que en el etiquetado para CDs no habían sido considerados los mecanismos de referencia discursiva.

Por último, se plantearán las conclusiones y se hablará del trabajo que queda por hacer en el marco de la referencia discursiva en CDs, siendo que esta tesis se enmarca en un proyecto en desarrollo.

Capítulo I *Antecedentes*

Antecedentes

1. Proyecto “Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos”

En el proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos*, se lleva a cabo una búsqueda y análisis de estructuras denominadas *contextos definitorios* (CDs), previamente definidos en la “Introducción” como fragmentos textuales donde se introduce un *término* y su correspondiente *definición* [Alarcón, 2003; Sierra *et al.*, 2003], damos otro ejemplo:¹

Tabla 1-1. Ejemplo de contexto definitorio

El tipo de plus valía (grado de explotación) se define como la relación entre la plus valía y el capital variable (de salarios)
Término = tipo de plus valía (grado de explotación)
Definición = la relación entre la plus valía y el capital variable (de salarios)

La noción de *contexto definitorio* es cardinal no sólo para el proyecto, sino para esta tesis que surge de la necesidad de complementar las investigaciones que se han hecho en el GIL sobre extracción automática de CDs y de términos.² Antes de adentrarnos en el tema de relaciones referenciales, específicamente *relaciones anafóricas*, explicaremos brevemente los objetivos generales del proyecto, los antecedentes de esta investigación y algunos trabajos que se desarrollan simultáneamente. La importancia de identificar y analizar CDs se puede entender mejor a partir de sus aplicaciones.

¹ Los ejemplos de CDs se presentarán en tablas en las que se especifica el corpus de origen *Corpus Lingüístico de Ingeniería* (CLI) o *Corpus Técnico del IULA* (CT del IULA). Es importante dejar claro que muchas veces existen errores de puntuación, de ortografía o tipográficos; los CDs se transcriben tal cual; de tal manera que los errores no se corrigen con el fin de respetar el texto de origen y no alterar datos.

² Entre los trabajos más importantes podemos mencionar el de Sierra y Mc Naught [2000] sobre el desarrollo de sistemas de búsquedas onomasiológicas orientado a la terminología; el de Sierra y Alarcón [2002] respecto a la identificación de patrones recurrentes en contextos definitorios; los de Alarcón [2003] y Sierra *et al.*, [2003] enfocados en el análisis y extracción automática de contextos definitorios y el de Aguilar [2006] acerca del análisis lingüístico de definiciones en contextos definitorios.

1.1. Aplicaciones

En la actualidad existen cantidades enormes de información de diferentes campos de estudio. Debido a esto, la organización y descripción de información relevante sobre conceptos de cada área de especialidad se ha vuelto una prioridad. En gran medida la terminología, en particular la terminografía, se encarga de la elaboración de diccionarios especializados en los que se explica el significado de los términos con base en el área de conocimiento a la que pertenecen. Para facilitar esta tarea, se han diseñado herramientas computacionales que sirven para extraer términos de manera automática [Cabré *et al.* 2001]. A esto, se añaden los avances que hay en la edificación de corpus informatizados.

Principalmente en lengua inglesa ha habido un importante avance en extracción automática, no sólo de términos sino también de información útil para describir sus significados.³ En español son muy recientes los trabajos al respecto, cabe decir que esta tesis parte de una serie de investigaciones previas y otras en desarrollo que describen el proceso de construcción y los alcances de las herramientas computacionales para la extracción de información sobre definiciones de conceptos.

Como parte del proyecto en el que se enmarca esta tesis, hemos dicho que se construye un *Extractor de contextos definatorios* (ECODE) para el español [Alarcón, 2006]. Entre sus aplicaciones más importantes se pueden mencionar aquellas que conciernen al ámbito terminológico y del conocimiento especializado. El ECODE, por ejemplo, puede ser útil para elaborar bases de datos de conocimientos léxicos, para la edificación de diccionarios onomasiológicos y semasiológicos⁴, glosarios, lexicones y ontologías. Además, puede ser

³ Se destacan los trabajos concentrados en la búsqueda sistemática de contextos ricos en conocimiento (ing. *Knowledge-Rich Contexts*) por parte de Pearson [1998] y Meyer [2001], el trabajo sobre la extracción de información metalingüística por parte de Rodríguez [2004], así como el uso de enunciados definatorios (fr. *énoncés définitoires*) realizadas por Auger [1997] y Rebeyrolle [2000].

⁴ Un *diccionario semasiológico* está constituido por una lista de palabras seguidas de sus respectivos significados o definiciones, en este tipo de diccionarios las búsquedas se hacen a partir de la palabra o concepto de la que se desea obtener el significado. Un *diccionario onomasiológico*, en cambio, es aquel que permite localizar la palabra adecuada para designar un concepto o idea a partir de su descripción; en el ámbito de la terminología, las búsquedas onomasiológicas se efectúan por medio de las descripciones elaboradas por un usuario en lenguaje natural, en particular dentro de un dominio técnico o de especialización [Sierra y McNaught, 2000; Sierra *et al.*, 2005].

de gran ayuda en la traducción especializada, en la enseñanza de lenguas de cualquier campo de especialización y en los sistemas pregunta-respuesta.⁵

Si bien las aplicaciones y alcances del proyecto son de gran valor para justificar la importancia de desarrollar herramientas útiles en la extracción automática de información, en particular de CDs, no daremos más detalles al respecto. El tema que aquí nos ocupa: *relaciones referenciales discursivas* especialmente *relaciones anafóricas*, es muy extenso y se origina de problemas específicos manifestados en los resultados que hasta ahora se han logrado en la extracción automática de CDs. Dichos problemas serán explicados en este capítulo.

1.2. Conceptos básicos para el proyecto

Con base en los trabajos previos de Sierra y Alarcón [2002], Alarcón [2003, 2006], Aguilar [2006] y Rodríguez [2004], a continuación se definen los conceptos primordiales para el proyecto y por consiguiente para esta investigación:

Texto especializado: Documentos que pertenecen a un área de estudio específica; una de sus principales características es que en ellos generalmente se explican términos y conceptos asociados a ciertas ramas de conocimiento, por lo cual son idóneos para extraer CDs.

Corpus lingüístico informatizado: Un *corpus lingüístico* es un conjunto de documentos de materiales escritos y/o hablados que están compilados bajo criterios específicos. Un *corpus lingüístico especializado* corresponde a una o varias áreas de conocimiento, cumple ciertas características dependiendo del objetivo con el que está elaborado [Toruella y Listerri, 1999]. Los *corpus informatizados* son, entonces, una colección de textos en soporte informático que puede ser de dimensiones variables, abarcar algunos millones de palabras o unos cuantos documentos. La ventaja de los corpus en soporte electrónico es que pueden ser manejados con programas informáticos que proporcionan una excelente

⁵ Los *sistemas pregunta-respuesta (question- answering)* permiten recuperar información por medio de un sistema en el que un usuario formula una pregunta en lenguaje natural y, como resultado, se obtiene la información solicitada. Estos sistemas todavía tienen muchas fallas, cabe decir que son efectivos sobre todo para preguntas y respuestas puntuales que se formulan a partir de repositorios de texto.

herramienta para el trabajo de investigación. En general, los corpus sirven para hacer análisis lingüísticos de diferente índole.

Contexto definitorio (CD): Se dijo que los CDs son fragmentos textuales que presentan la definición de un término. Debe tenerse en cuenta que éstos abarcan distintas estructuras y en función de ello despliegan diversos elementos. Con base en estudios acerca de las diferencias estructurales, se ha determinado la presencia recurrente de elementos como: *término, definición, patrones verbales definitorios, marcadores reformulativos definitorios, marcadores tipográficos definitorios y patrones pragmáticos*. La identificación de dichos componentes permite distinguir patrones de comportamiento sintáctico y semántico que facilitan la obtención de datos para la búsqueda y clasificación de definiciones.

Los dos elementos mínimos e imprescindibles de un CD son el *término* y la *definición* del mismo. Se ha observado que estas unidades establecen ligas a partir de *patrones definitorios* del tipo “X se define como Y, X comprende Y, X es Y, X se denomina Y”, por dar algunos ejemplos. En los CDs también pueden aparecer marcadores discursivos (*por ejemplo, es decir, esto es, de tal manera, etc.*); así como unidades tipográficas que sirven para introducir un término o algún elemento de la definición (*viñetas, paréntesis, guiones, comillas, dos puntos, etc.*), se trata de signos que ligan al término con su definición, sustituyendo o complementando la función de las estructuras verbales⁶ empleadas típicamente en las definiciones.

Si bien los elementos básicos de un CD son el *término* y la *definición* ligados por un *patrón definitorio*, frecuentemente los CDs presentan estructuras que proporcionan información sobre el uso de términos, a estos componentes se les ha denominado *patrones pragmáticos* y más adelante serán explicados.

Conforme a las diferencias estructurales de los CDs, se muestra a continuación un diagrama, en el cual se sintetiza la relación de los componentes que los constituyen [Alarcón, 2006: 25]:

⁶ De los verbos más comunes en las definiciones tenemos, por ejemplo: *caracterizar, comprender, concebir, conocer, considerar, consistir, constar, definir, denominar, describir, entender, identificar, llamar, permitir, servir, usar, utilizar y ser*.

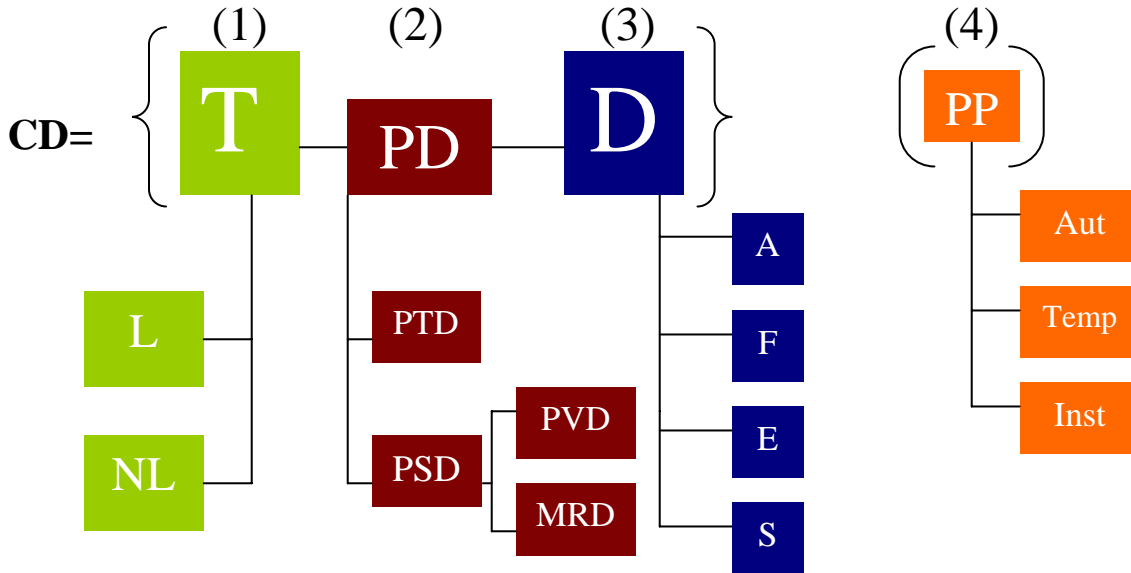


Figura 1-1. Elementos constitutivos de los CONTEXTOS DEFINITORIOS

CÓDIGOS:

- (1) **T**= *Término* / **L**= *Lingüístico* / **NL**= *No lingüístico*
- (2) **PD**= *Patrón definitorio* / **PTD**= *Patrón tipográfico definitorio* / **PSD**= *Patrón sintáctico definitorio* / **PVD**= *Patrón verbal definitorio* / **MRD**= *Marcador reformulativo discursivo*.
- (3) **D**= *Definición* / **A**= *Analítica* / **F**= *Funcional* / **E**= *Extensional* / **S**= *Sinonímica*
- (4) **PP**= *Patrón pragmático* / **Aut**= *Autoría* / **Temp**= *Temporal* / **Inst**= *Instruccional*

Siguiendo el esquema expuesto, a continuación se proporciona la definición de los elementos que pueden ser parte de la estructura de un CD:

- (1) **Término (T):** Unidad léxica especializada cuyo significado está relacionado con un área de conocimiento particular. En palabras de Dubuc [1992] es “el elemento constitutivo de cualquier nomenclatura terminológica que esté relacionada con una lengua de especialidad [...], es la denominación de un objeto, propio de una determinada área de conocimiento”. La estructura sintáctica de un término puede variar,

sin embargo, se ha observado que la categoría que lo caracteriza es la nominal, de tal manera que las unidades adjetivas, verbales y adverbiales con implicaciones terminológicas, están emparentadas con un término nominal [Cabré, 1999:139]. En el marco del proyecto se ha considerado que puede haber *términos lingüísticos* (L) y *no lingüísticos* (NL), estos últimos corresponden a fórmulas, cifras, nombres en latín, abreviaturas y símbolos que son muy comunes en los textos de especialidad. En la extracción automática no se descarta la posibilidad de recuperar definiciones que incluyan unidades que no siguen estrictamente patrones morfosintácticos pero que son igualmente definidas. En esta tesis sólo se analizan CDs con *términos lingüísticos*.

- (2) **Definición (D):** A grandes rasgos, Sager [1995] entiende por *definición* el proceso de explicar el significado de símbolos expresados lingüísticamente. En el caso de términos, se trata de la descripción lingüística de unidades relacionadas con un área de conocimiento específica: *definición especializada*. Para fines del proyecto, la definición es el elemento que aporta información del término y generalmente abarca un fragmento textual.

Existen definiciones de varios tipos, en el marco de los CDs se estudian concretamente las características de cuatro variantes relacionadas con el *modelo aristotélico o analítico de definición*. Dicho modelo considera dos constituyentes básicos: *Género próximo* (*Genus*) y *Diferencia específica*, la relación podemos entenderla como $X = Y + \text{características distintivas}$ (p.e. *una calculadora es una máquina que sirve para hacer cuentas*).

También se consideran definiciones que proporcionan información que permite clasificar al término dentro de una clase general, esto es, a partir de un hiperónimo⁷ o un merónimo.⁸ Además, encontramos definiciones que pueden establecer una relación de sinonimia, o bien, dar información acerca de la función del término. Con base en el modelo aristotélico y a partir de la presencia o ausencia del *Género próximo* y de la *Diferencia específica* se ha establecido una tipología de definiciones que comprende las *analíticas* (A), las *funcionales* (F), las *extensionales* (E) y las *sinonímicas* (S). En la

⁷ Un *hiperónimo* es aquel término general que se emplea para referirse al significado denotado por otro término más particular y con significado más preciso, p.e. *mesa-mueble* [Charaudeau, 1992:52].

⁸ La *meronimia* es la relación inclusiva parte-todo; un *merónimo* indica la parte del *holónimo* que representa al todo, p.e. *freno-automóvil* [Renkema, 1999].

figura siguiente queda esquematizada la relación que guardan estos cuatro tipos [Aguilar, 2006; Sierra *et al.*, 2008]:

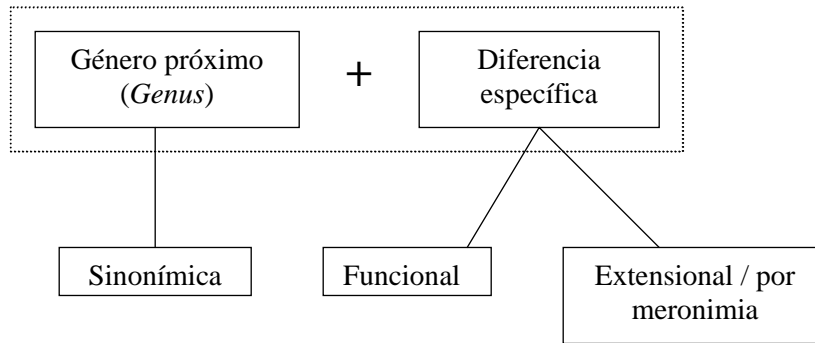


Figura 1-2. Modelo aristotélico o analítico

- a. **Definición analítica o aristotélica:** Se da cuando el *Género próximo* y la *Diferencia específica* aparecen de manera explícita dentro de una definición:

Tabla 1-2. Definición analítica o aristotélica

Lafourcae [Autor] (1980) define [PVD ₁] el perfil profesional [Término] como [PVD ₂] una especificación de habilidades, rasgos y disposiciones que orientan la construcción del plan de estudios y asuntos que definen el que hacer de los miembros de cierta profesión [Definición].
El término "sembrar partículas " [Término] se refiere a [PVD] introducir en el flujo de observación un trazador que es transportado por él mismo [Definición].

- b. **Definición sinonímica:** Se da cuando sólo se hace explícito el *Género próximo*, de tal manera que establece una equivalencia conceptual con el término que es definido. En otras palabras es la relación de similitud entre un término y un *Género próximo* referidos a un mismo concepto, ya que la diferencia específica no aparece expresada dentro de la definición. Esta relación de sinonimia se puede ejemplificar del siguiente modo:

Tabla 1-3. Definición sinonímica

Cuando por un circuito circulan corrientes de secuencia positiva, la impedancia del circuito, se denomina [PVD] impedancia [Término] a la corriente de secuencia positiva [Definición].
--

- c. **Definición funcional:** Se da cuando sólo se hace explícita la *Diferencia específica*, la cual ofrece la definición de un concepto a partir de su uso o aplicación en una situación dada. Las definiciones funcionales comparten con las extensionales el rasgo de que no expresan *Género próximo*. Este tipo de definición ofrece una descripción respecto al uso o aplicación que se da de algún objeto u entidad representada por un término. Un ejemplo es el siguiente:

Tabla 1-4. Definición funcional

El relevador auxiliar [Término] es usado para [PVD] asistir en el desarrollo de sus funciones a los relevadores de protección, como respaldo [Definición].

- d. **Definición extensional o por meronimia:** Se da cuando sólo se hace explícita la *Diferencia específica*. La definición enumera los componentes que conforman un objeto representado por el término. Esta enumeración de componentes sigue un orden basado en “relaciones inclusivas”, identificadas principalmente con la relación *merónimo-holónimo* y con la relación *hipónimo-hiperónimo*. Al igual que una definición funcional, tampoco manifiesta un *Género próximo* concreto. Un ejemplo de esto es:

Tabla 1-5. Definición extensional o por meronimia

La Terminal de contenedores [Término] cuenta con [PVD] dos muelles de atraque, el muelle del Bufadero y el muelle del Dique del Este [Definición].

- (3) **Patrones definatorios (PD):** Son los elementos que ligan un término y su definición. Se dividen en *patrones tipográficos definatorios* (PTD) y *patrones sintácticos definatorios* (PSD), estos últimos a su vez comprenden *patrones verbales definatorios* (PVD) y *marcadores reformulativos* (MR):

- a. **Patrones tipográficos definatorios (PTD):** Son signos de puntuación que operan como conectores entre el término y la definición (*dos puntos, viñetas, guiones, etc.*). Los signos de puntuación y los tipos de letra (*negritas, itálicas, subrayado*) también pueden usarse como *marcas tipográficas* (MT) para resaltar el término o parte de la definición.

- b. **Patrones verbales definatorios (PVD):** Son patrones que emplean verbos propiamente usados para introducir definiciones: *denominar, definir, describir, identificar y ser*, por enunciar algunos [Rodríguez, 1999]. Se distinguen dos tipos: *simples* y *compuestos* [Alarcón, 2003]. Cuando el verbo definatorio no se acompaña de ninguna otra partícula se trata de *patrones verbales definatorios simples* (X denomina Y, X es Y, X significa Y). En cambio, si el verbo definatorio aparece con otras partículas que pueden ser verbos auxiliares, preposiciones, adverbios o el clítico *se*, son considerados *patrones verbales definatorios compuestos* (X es entendido como Y, X se identifica como Y, X se caracteriza por Y).
- c. **Marcadores reformulativos definatorios (MRD):** Son marcadores discursivos vinculados con un proceso metalingüístico, en otras palabras, se emplean para hablar del propio lenguaje. En el caso de los CDs sirven para reinterpretar o retomar alguna parte del texto. Son estructuras del tipo: *es decir, por ejemplo, estos es, de tal manera que*, etc.
- d. **Patrones pragmáticos (PP):** Corresponden a la información relevante para entender el término dentro del texto en el cual aparece. Se han definido tres variantes: *autoría* (Aut), cuando es un autor o una institución quien define el término; *temporales* (Temp), en el caso en que se proporcione, por ejemplo, la fecha en la cual se definió o se introdujo el término; e *instruccionales* (Inst), referidos a estructuras sintácticas que dan matices diferentes para introducir la definición (*desde el punto de vista, de manera general, desde la perspectiva, científicamente, matemáticamente*, etc.).

Son muchos los conceptos que están en juego al momento de describir las partes de un CD (Figura 1-1); con el fin de concretar las definiciones expuestas, damos un ejemplo en el cual encontramos todos los posibles elementos:

Tabla 1-6. Ejemplo de partes de un contexto definatorio

<p>A. Richling y T. Bartkowski, entre otros; quienes consideran a la Ecología del Paisaje como: "ciencia transdisciplinar que tiene como objetivo principal la resolución del problema de la gestión y desarrollo de los territorios a escala regional y local a lo que le llaman algunos, Ecosistema Humano Total "</p>

Patrón pragmático= A. Richling y T. Bartkowski
Término= la Ecología del Paisaje
Patrón verbal definitorio= consideran a [*término definido*] como
Definición= “ciencia transdisciplinar que tiene como objetivo principal la resolución del problema de la gestión y desarrollo de los territorios a escala regional y local a lo que le llaman algunos, Ecosistema Humano Total”
Patrón tipográfico definitorio= “”

2. Problemática en función del criterio de extensión de contexto definitorio

Aunque no vamos a explicar con detalle el proceso de extracción de CDs, hay que señalar que éstos se obtienen de manera automática por medio de herramientas computacionales que utilizan *patrones verbales definitorios* como patrón de búsqueda. Esto significa que tanto el término como la definición se localizan a partir del *patrón sintáctico definitorio* que los vincula, principalmente verbos.

Hasta ahora, en el proceso de extracción automática, no se ha establecido un criterio preciso de extensión de los CDs. Por *extensión* entendemos la cantidad de palabras que comprende un CD, es decir, el tamaño exacto del fragmento textual que contiene la definición completa de un término. La dificultad para definir dicho criterio radica en que los CDs tienen distintas estructuras, no poseen el mismo número de palabras y sus elementos no se presentan en el mismo orden.⁹

La metodología del ECODE¹⁰ hasta ahora no delimita el tamaño de los CDs y por ello, en algunos casos, se pierde información importante de la definición y/o del término. En este sentido, la falta de un criterio apropiado para determinar la extensión, deriva en uno de los problemas más importantes en los resultados de la extracción automática.

Alarcón [2006] propone “el párrafo” como estructura textual para establecer la extensión de un CD. Este criterio tiene que optimizarse, ya que la definición de un término puede aparecer en varios párrafos o bien abarcar mucho menos que éste:

⁹ En los trabajos en curso de Alarcón [2006] y Aguilar [2006] se plantea que los elementos básicos de un CD (término, predicación verbal y definición) siguen ciertos patrones de ordenamiento que se manifiestan en función del tipo de predicación verbal definitoria.

¹⁰ Extractor de contextos definitorios (ECODE), <http://brangaene.upf.es/ecode/>

Tabla 1-7. Ejemplos de diferencia de longitud de contextos definitorios

Contexto definitorio que abarca más de un párrafo
<p><i>Transformadores de corriente</i></p> <p>Se conoce como <i>transformador de corriente</i> a aquél cuya función principal es cambiar el valor de la corriente de uno más o menos elevado a otro con el cual se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc.</p> <p><i>Su construcción</i>₁ es semejante a la de cualquier tipo de transformador, ya que fundamentalmente consiste de un devanado primario y un devanado secundario. <i>La capacidad de estos transformadores es</i> muy baja, se determina sumando las capacidades de los instrumentos que se van a alimentar, y puede ser 15, 30, 50, 60 y 70 VA.</p> <p><i>Estos transformadores son</i> generalmente de tamaño reducido y el aislamiento que se emplea en su construcción tiene que ser de muy buena calidad, pudiendo ser en algunos casos resinas sintéticas (compound), aceite o líquidos no inflamables (pyranol, clorextol, etc.).</p>
Contexto definitorio que abarca menos de un párrafo
<p>Por <i>mosaico</i> se entiende la existencia de más de una estirpe celular en un mismo individuo originadas después de la fecundación por algún fenómeno genético anormal (mutaciones génicas o cromosómicas , etc.), mientras que por <i>quimera</i> se entiende la aparición de líneas celulares distintas originadas a partir de diferentes fuentes de fecundación.</p>

Conforme al primer ejemplo (Tabla 1-7) podemos observar que en cada uno de los párrafos existen *patrones verbales* que introducen una definición. Si nos ceñimos al criterio de “párrafo como límite de extensión de un CD”, la extracción automática puede arrojar tres contextos definitorios distintos. El problema, en este sentido, es que el término “transformadores de corriente” engloba los significados referidos en los tres párrafos y se trata entonces de un único CD. Contrariamente, en el segundo ejemplo, el párrafo rebasa la longitud del CD, incluso existen dos definiciones de términos distintos y en consecuencia dos CDs. En el texto de este segundo ejemplo no hay relaciones discursivas que indiquen que la longitud de alguno de los CDs debe verificarse en el documento de origen, sin embargo, el párrafo tampoco aplica como criterio para determinar la extensión.

Es importante puntualizar en las diferencias estructurales de los CDs y considerar las relaciones discursivas para establecer la longitud de un CD en función de la información prominente del término y no a partir de un criterio que no resuelve por completo el problema de la extensión en la extracción automática. Se subraya la necesidad de establecer patrones de extensión de los contextos tomando en cuenta las diferentes estructuras de los

CDs. Apparently, the description and consideration of a certain type of discursive relations can provide clues to determine these patterns.

3. Problemática en función del tipo de candidato a contextos definitorios

Los resultados de búsqueda de CDs a partir de la localización de patrones verbales definitorios arrojan buenos y malos candidatos. Los *buenos candidatos* (BC) son todos aquellos contextos que contienen la definición de un término. Los *malos candidatos* (MC) son fragmentos textuales que poseen un patrón verbal que no define un término y por tanto no son CDs:

Tabla 1-8. Ejemplos de buenos y malos candidatos a contextos definitorios

Buen candidato (BC)
Se conoce como reenganche rápido a la operación de cierre de un interruptor después de una falla.
Mal candidato (MC)
Una vez descrito el escenario, pasemos a conocer las “prácticas” que tratan de llevar al alumno al mundo real de la Teoría del Buque.

In a previous stage of this investigation, the manual separation of BC and MC was performed. The MC were listed as data that should be studied to avoid forming part of the future results. The BC were placed in a repository that constitutes the corpus of the investigations that involve the project and serve for the construction of the *Corpus de Contextos Definitorios* of which will be spoken in Chapter IV. Additionally, work continues in the improvement of search patterns to perfect the results of automatic extraction. For this, in the ECODE a filtering process of exceptions in the candidates obtained, which means that through rules of exceptions, the contexts that recover noise (MC) are avoided.

The corpus of this thesis was made based on the repository of BC. As already explained, the size of the CDs is variable and many times due to structural differences and to the absence of a criterion that determines its extension, the automatic extraction throws good incomplete candidates:

Tabla 1-9. Ejemplo de candidato incompleto

Este tipo de relación se conoce matemáticamente como una relación negativa
Referencia incompleta = Este tipo de relación (sin antecedente)
Término = relación negativa

Se observa que existe un término pero su definición no está presente, en su lugar hay una expresión que apunta a cierta parte del discurso que no fue considerada en la extracción automática. Para completar la definición es necesario acudir a la parte del texto a la cual señala el sintagma nominal *este tipo de relación*. La definición completa del término la obtenemos ampliando el contexto, en este sentido se habla de *expansión* de CDs.

En oposición a los contextos incompletos existen candidatos en los que tanto el término como su definición están presentes y no es necesario recurrir al texto para verificar su extensión en el documento original:

Tabla 1-10. Ejemplo de candidato completo

El abultamiento muscular de la superficie lateral de la palma de la mano se conoce como eminencia tenar y está formada por los músculos cortos del pulgar .

En función de lo que se ha explicado con ejemplos, la extracción automática da como resultado *buenos candidatos* que pueden ser completos o incompletos, la diferencia para esta investigación es importante y se describe de acuerdo con las siguientes características:

- (1) **Candidatos completos (CC)**: Aquellos CDs en los que los elementos prominentes *término* y *definición* aparecen en el fragmento textual dado en la extracción automática.
- (2) **Candidatos incompletos (CI)**: Son los CDs en los que el *término* o la *definición* no están presentes porque se manifiesta alguna clase de fenómeno, mecanismo o relación discursiva que los afecta.

En la siguiente figura se muestra la posición de los CC y los CI en el proceso general de *extracción automática* de CDs:

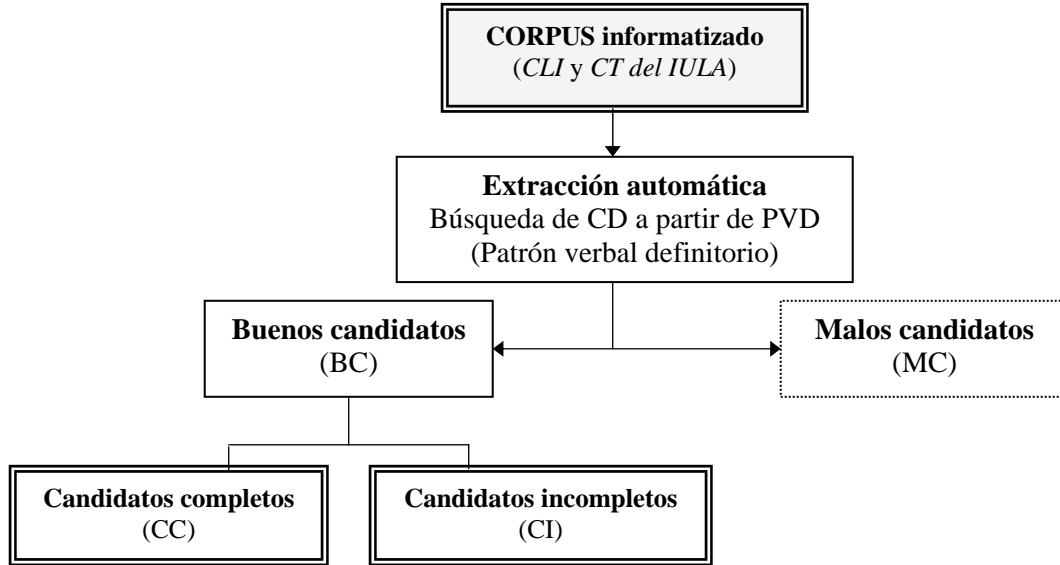


Figura 1-3. Candidatos completos y candidatos incompletos

Hemos explicado los problemas principales detectados en la extracción automática, a partir de ellos se plantea que existen estructuras lingüísticas que desencadenan fenómenos discursivos y generan recurrentemente candidatos incompletos, creemos que su descripción puede contribuir al mejoramiento de patrones de búsqueda en los trabajos de extracción automática, principalmente en el ECODE. En esta tesis se trabajará con un corpus que contiene CC y CI, esto permitirá comparar la clase de mecanismos y fenómenos discursivos que aparecen tanto en unos como en otros.

4. Corpus para la extracción automática de contextos definitorios

Los CDs fueron extraídos de dos corpus diferentes y son el resultado de dos trabajos de investigación que emplearon metodologías distintas en la extracción automática [Alarcón, 2006; Aguilar, 2006]. En ambos casos, y como ya se explicó, el patrón de búsqueda de las herramientas computacionales se basó principalmente en la identificación de *patrones verbales definitorios*.

Los corpus sujetos a la extracción automática fueron el *Corpus Lingüístico de Ingeniería* y el *Corpus Técnico del IULA*, brevemente se explica el origen de cada uno y la cantidad de CDs recuperados para esta investigación:

Corpus Lingüístico de Ingeniería (CLI) [Medina *et al.* 2004] desarrollado por el Grupo de Ingeniería Lingüística (GIL) de la UNAM. Dicho corpus está constituido por documentos tipo TXT¹¹ con un total aproximado de 500,000 palabras. El CLI es una recopilación integrada por documentos técnicos (artículos, informes, tesis, etc.) en diversas áreas temáticas de ingeniería. Al momento de iniciar esta investigación había poco menos de 200 CDs extraídos de manera automática a partir de 17 tipos de *patrones verbales definitorios*. Se eligieron 150 CDs sin atender a ningún criterio preciso, ya que la localización y estudio de *mecanismos de referencia discursiva* se efectuará en cualquier tipo de CD.

Corpus Técnico del IULA (<http://bwananet.iula.upf.edu/indexes.htm>) [Bach *et al.* 1997], desarrollado por el Instituto Universitario de Lingüística Aplicada (IULA), al cual puede accederse a través del motor de búsqueda *Bwana Net*, que entre otras cosas permite hacer consultas terminológicas. Es un corpus de documentos especializados que incluye textos en español, catalán, inglés, francés y alemán, pertenecientes a distintas áreas de conocimiento (derecho, economía, genoma, informática, medio ambiente y lenguaje general). De este corpus sólo se eligieron 100 CDs obtenidos a partir de 8 *predicaciones verbales definitorias* y se trabajó específicamente con textos en español.

Los CDs del CLI constituirán el corpus central de esta investigación, porque es el corpus principal del proyecto, sin embargo se trabajará simultáneamente con contextos del *Corpus Técnico del IULA*, esto con dos propósitos; el primero, estudiar una muestra más amplia que incluya textos de otras ramas de especialización distintas de la ingeniería; el segundo, aprovechar y explotar los adelantos que existen en la extracción automática de CDs en este corpus. La razón por la cual trabajamos con un número menor de CDs del *Corpus técnico del IULA*, es que éste es de acceso restringido y las consultas de documentos originales están limitadas.

¹¹ Son archivos de texto plano que se componen de texto sin formato, sólo caracteres. Se les conoce también como archivos de texto llano o simple (del inglés *plain text*), por no presentar formato (negritas, subrayado, cursivas, tamaño, etc.) y tipos de letra (por ejemplo, Arial, Times, Courier, etc.).

Los ejemplos que hemos expuesto dan muestra de los problemas concretos en los resultados de la extracción automática. A partir de dichos problemas, podemos anotar la importancia de estudiar ciertos fenómenos discursivos como parte de la estructura de los CDs. En este sentido, será necesario identificar las estructuras y los mecanismos discursivos que generan candidatos incompletos y para ello debemos determinar qué entendemos por *mecanismo discursivo*, qué es *discurso*, qué fenómenos son los que permiten su construcción y cómo éstos afectan la estructura de los CDs.

5. Discurso y contextos definitorios como entidades discursivas

En este apartado se explicará la noción de *discurso* y algunas de sus propiedades, puntualizando en dos conceptos: *coherencia* y *cohesión*, fenómenos que permiten que los mensajes puedan entenderse como un todo, es decir, como estructuras con sentido específico. Veremos que los CDs se enmarcan dentro de las características del discurso cuando cumplen con ciertos rasgos de textualidad.

Tradicionalmente se planteaba que la oración era la mayor unidad de estudio de la gramática, los hechos lingüísticos eran entendidos como la suma de oraciones que seguían ciertas reglas combinatorias. En los años sesentas surgen propuestas teóricas que abandonan esta perspectiva y estudian los hechos comunicativos sometidos a un universo o estructura mayor denominada *discurso*. Esta perspectiva se identifica principalmente con la *lingüística textual*¹² que define *discurso* como el escenario de todo acto comunicativo, ya sea oral o escrito.

Desde este punto de vista, la lengua no se limita a un sistema de signos que obedece reglas, sino que también implica una función comunicativa en la que la *situación*, el *contexto* y la *intención*¹³ son esenciales:

¹² La *lingüística textual* se encarga de los hechos de habla que se realizan en una determinada situación por determinado hablante (manifestaciones habladas o escritas de cualquier lengua). Estudia las *función textual* (o de *sentido*) que designa el contenido de un texto o de una de sus partes: *pregunta, respuesta, síntesis, diálogo, texto especializado, contexto definitorio*, etc. [Renkema, 1999; Casado, 1994].

¹³ El *discurso* se produce y cobra *sentido* en la *situación* concreta en que es producido, el *contexto* es el conjunto de conocimientos y creencias compartidos por los participantes de la comunicación, en tanto que la *intención* concierne a los objetivos específicos de cualquier mensaje [Renkema, 1999].

Discourse denotes the hierarchically structured, situated sequence of utterance, indexical, and illocutionary acts carried out in pursuance of some communicative goal, as integrated within a given context. [Cornish, 2002:472]

Así pues, el lenguaje no se ejerce mediante enunciados de habla aislados, sino por medio del *discurso*, en el cual dichos enunciados se suceden unos a otros y se articulan en una estructura compleja, un todo informativo.

Pese a las diferencias entre lenguaje oral y lenguaje escrito, el concepto *texto* no se reduce al discurso escrito, la lingüística textual emplea *texto* indistintamente que *discurso*, en la medida en que *texto* se entiende como una unidad del lenguaje determinada por factores pragmáticos y cognitivos que participa de una intención comunicativa concreta:

“Texto” es la unidad lingüística comunicativa fundamental, producto de la actividad verbal humana, que posee siempre carácter social; está caracterizado por su cierre semántico y comunicativo, así como por su coherencia profunda y superficial, debida a la intención (comunicativa) del hablante de crear un texto íntegro, y a su estructuración mediante dos conjuntos de reglas: las propias del nivel textual y las del sistema de la lengua. [Bernárdez, 1982: 85]

Es así que el *discurso* es una estructura que posee ciertas características y sigue algunas reglas de textualidad, entre las más importantes están: *cohesión*, *coherencia*, *intencionalidad*, *aceptabilidad*, *informatividad*, *situacionalidad* e *intertextualidad* [De Beaugrande y Dressler, 1997: 33-47]. En este trabajo se da prioridad a las dos primeras, porque el *discurso*, entre otras cosas, opera como unidad de significado gracias a que sus partes están correctamente articuladas y en consecuencia forma un *todo coherente y cohesionado*. Antes de explicar *cohesión* y *coherencia* se definen brevemente los rasgos de textualidad mencionados [Renkema, 1999]:

- (a) **Intencionalidad:** Son los objetivos informativos específicos del mensaje.
- (b) **Aceptabilidad:** Se da en función de la correcta formación semántica de las oraciones.
- (c) **Informatividad:** Se manifiesta cuando un texto cumple con la condición de dar información nueva y de conectar de manera adecuada los conocimientos previos de los participantes de la comunicación.
- (d) **Situacionalidad:** Se refiere a la situación o contexto situacional en el que se produce el discurso.

(e) **Intertextualidad:** Atañe a la organización lingüística de los textos y se plantea en función de la relación de secuencia de oraciones.

Los CDs son fragmentos textuales obtenidos de documentos especializados, estos documentos *a priori* operan como unidades discursivas, es decir, como un todo cohesionado y coherente. En este sentido, los CDs son estructuras removidas de su contexto de origen, un discurso que, en sí, cumple reglas de textualidad y de alguna manera es fragmentado. Esta característica puede generar ciertos cuestionamientos acerca del carácter discursivo de los CDs y nos lleva a preguntarnos si éstos continúan operando como discurso a pesar de ser extraídos de un universo textual mayor (tesis, artículo de divulgación, ponencia, o cualquier documento especializado). Si se trata entonces de “pedazos de texto”, ¿pueden éstos funcionar como unidades de sentido? ¿acaso no pierden ciertos rasgos de textualidad? Para responder estas preguntas se vuelve indispensable describir los mecanismos discursivos que permiten que un CD pueda tener sentido en sí mismo.

Efectivamente se podría pensar que las propiedades discursivas de los CDs se diluyen porque se trata de fragmentos supeditados a las reglas de textualidad de su documento de origen. A pesar de que los CDs guardan relación con un universo textual mayor, podemos concebirlos como *discurso* en sí mismos, ya que poseen rasgos de unidades discursivas con sentido propio:

- (a) **Intencionalidad:** Los CDs participan de una intención comunicativa concreta que consiste en proporcionar la definición de un término para que los participantes de la comunicación compartan el mismo conocimiento sobre éste.
- (b) **Aceptabilidad:** Los CDs mantienen una estructura semántico-sintáctica correctamente construida conforme a las reglas de la lengua que permiten su entera comprensión.
- (c) **Informatividad:** Los CDs dan información nueva, puesto que introducen la definición de un término.
- (d) **Situacionalidad:** Los CDs atienden a una situación comunicativa concreta, ya que son extraídos de un tipo muy específico de texto que atañe al lenguaje escrito y especializado. En este tipo de discurso las definiciones de términos son comunes, por lo cual se deberá obtener una mayor cantidad de CDs.

(e) **Intertextualidad:** Un CD no es un hecho de lengua aislado, proviene de una unidad discursiva más compleja (documentos especializados) en donde las secuencias de oraciones y estructuras sintácticas están relacionadas con otras secuencias de oraciones. Aunque se trata de fragmentos discursivos, la relación con el texto del cual son obtenidos no deja de ser importante; por esta razón se deben estudiar los mecanismos discursivos que permiten el flujo comunicativo: si en la extracción automática se obtienen candidatos incompletos porque falta algún elemento del CD, significa que la relación con alguna parte del texto no se considera y la información no se completa.

5.1. Cohesión y coherencia textuales

Además de los rasgos de textualidad enunciados, *cohesión* y *coherencia textuales* son nociones básicas para entender el funcionamiento del discurso como una *unidad de sentido*,¹⁴ ambas contribuyen a la construcción del contenido y la organización de los textos.

La *cohesión* se entiende como el conjunto de funciones lingüísticas que relacionan los elementos de un texto [De Beaugrande y Dressler, 1997]. Se trata de la *conexión* o *encadenamiento* que surge cuando un elemento del texto debe ser interpretado a través de otras partes del discurso [Renkema, 1999: 52]. La *cohesión* se logra por medio de recursos lingüísticos y *mecanismos referenciales* que relacionan acontecimientos o situaciones del discurso:

Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another. The-one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby at least potentially integrated into a text [Halliday y Hasan, 1976:6]

Si bien la *unidad de sentido* de un texto pueda plantearse en función de la relación y dependencia de sus partes, Van Dijk [1984: 83] agrega que “la conexión puede ser una condición necesaria pero no suficiente para la aceptabilidad del discurso”. El hecho de que los elementos se refieran unos a otros y se encuentren conectados, no garantiza que el texto

¹⁴ Van Dijk [1984: 135] define *sentido* como “el conocimiento que se transmite de manera efectiva mediante las expresiones que aparecen en el texto”, del tal manera que el sentido es el conocimiento adquirido en un texto determinado.

sea una unidad informativa *coherente*. Es así que *cohesión* sólo supone el *proceso referencial* entre las partes del discurso a través de la palabra y no atiende a la activación de conocimientos por medio de las expresiones referidas.

La *coherencia*, otra importante condición del texto, es concebida como una propiedad semántica [Van Dijk, 1984]. Brown y Yule [1983: 66] la explican como “aquello” que los usuarios de la lengua asumen naturalmente, mientras Renkema [1999:53] la define como la conexión que existe hacia afuera del texto y que está determinada por el conocimiento previo de los participantes en la comunicación. Se trata de un concepto difícil de delimitar, sin embargo, *coherencia* se emplea principalmente para hablar de *progresión temática*¹⁵ y continuidad de sentido de los textos.

La *coherencia* depende en gran medida de las *implicaciones* y *presuposiciones*¹⁶ efectuadas en la producción e interpretación de los textos, pone en marcha el saber enciclopédico de los hablantes, recupera el contexto de enunciación y se relaciona con el proceso total de la intención comunicativa. La coherencia, entonces, no atiende únicamente al ordenamiento de las partes del discurso, sino a su jerarquía y congruencia.

Cohesión no es *coherencia* [Carrell, 1982], pero son fenómenos discursivos que se complementan. Mientras que la *coherencia* depende de factores extratextuales relacionados con el conocimiento de mundo y el saber compartido, la *cohesión* se refiere a la conexión de las partes a través de medios lingüísticos y son estos medios lingüísticos los que nos interesa identificar en la estructura de los CDs.

6. Observaciones finales

Como ya se dijo, los CDs, específicamente los candidatos completos, mantienen ciertas propiedades discursivas o rasgos de textualidad, de tal manera que podemos concebirlos como un todo informativo, cohesionado y coherente. La *cohesión* en los CDs se manifiesta en función de la correcta articulación de sus elementos por medio de diferentes mecanismos

¹⁵ *Progresión* entendida como la condición del discurso que se refiere a la articulación de *temas* (información conocida) y *remas* (información nueva), es considerada un factor importante para la *coherencia textual* ya que permite la conexión y la interrelación de los fragmentos o partes del conjunto textual. [Soto, 2004-2005].

¹⁶ Son las creencias compartidas de los participantes en la comunicación (*background knowledge*). Todo texto implica cierta presuposición de los temas que se tratan en él, es decir, cierto conocimiento previo o saber compartido ajeno al texto de tal manera que no toda la información se enuncia explícitamente. [Sperber y Wilson, 1986]

referenciales; y la *coherencia* se expresa en la congruencia del texto y en el proceso total de la intención comunicativa.

Una vez delimitada la noción de *discurso* e identificados los CDs como unidades discursivas que mantiene cierta cohesión y coherencia textuales, nos interesa buscar las *estructuras* y *mecanismos discursivos* que en la extracción automática pueden generar candidatos incompletos como los que a continuación se exponen:

Tabla 1-11. Ejemplo de candidatos incompletos (CLI y CT del IULA)

Este consta de un banco de capacitores sumergidos en aceite en un recipiente de porcelana y conectados en serie para aumentar la resistencia de la línea de alto voltaje.
Este tipo de conexión se utiliza en capacidades no mayores de 10 MVA debido a que en casos de cortocircuito las corrientes son elevadas como consecuencia de la poca limitación.
Estos fármacos comprenden la noradrenalina , adrenalina y un gran número de fármacos de acción prolongada que tienen los mismos efectos que la adrenalina y noradrenalina .
Son un instrumento que consiste en una evaluación sistemática, periódica y objetiva del funcionamiento de la organización, de los sistemas de gestión y de los equipos puestos en marcha por la empresa para limitar y vigilar la incidencia de las actividades industriales sobre el medio ambiente.
Esta es una operación de mecanizado que consiste en torneear una o numerosas ranuras helicoidales perfiladas sobre una superficie cilíndrica a veces cónica exterior o interior.
Uno de ellos consiste en mallar tridimensionalmente el fluido y resolver en él la ecuación de Laplace ya sea por diferencias finitas o por elementos finitos.
La primera se refiere a la generación del oleaje por parte del buque y de ella se ha encargado el grupo de Hidrodinámica Numérica del Canal de Ensayos Hidrodinámicos de la Escuela Técnica Superior de Ingenieros Navales de la Universidad Politécnica de Madrid. La segunda parte consiste en el estudio de la propagación de estas olas generadas por el buque hasta las márgenes del canal de navegación.

Observamos que ciertas expresiones o estructuras del contexto guardan relación con una parte anterior del texto obtenido de la extracción automática, dichas expresiones refieren a otras anteriores: ¿Qué clase de relación se establece? ¿Qué estructuras son las que recurrentemente desencadenan CI? ¿Cómo se ven afectadas la cohesión y coherencia textuales de los CDs? No olvidemos que el patrón de búsqueda es el verbo definitorio, en este sentido se observa que dicho patrón aparece muy próximo de las estructuras que no completan su significado porque hacen referencia a otra u otras entidades del discurso. En

principio podemos decir que una estructura del tipo *Estos fármacos* (Tabla 1-11) remite a una entidad que ya ha sido antes mencionada.

En el siguiente capítulo se hace una revisión teórica general que servirá como base para describir las estructuras que generan candidatos incompletos y aquellos *mecanismos de referencia discursiva* que son más prominentes en CDs.

**Capítulo II *Mecanismos de referencia
discursiva: fenómeno de anáfora***

Mecanismos de referencia discursiva: fenómeno de anáfora

En el capítulo anterior planteamos la problemática que motivó esta investigación, se expusieron algunos ejemplos del tipo de texto con el que se trabajará y se determinó que dichos textos son *unidades discursivas* que despliegan ciertos rasgos de textualidad. Específicamente se habló de *cohesión textual*, una propiedad del discurso que consiste en que las partes de un texto estén correctamente articuladas por medio de mecanismos de referencia discursiva. Para delimitar el carácter de dichos mecanismos será preciso definir otros conceptos que permitan describir sus alcances y las implicaciones que éstos tienen en el discurso en general y en los contextos definitorios (CDs) en particular. Empezaremos delimitando la noción de *referencia*, a partir de la cual se pueden explicar los fenómenos discursivos que sirven para *señalar, referir, apuntar, o reintroducir* una parte del discurso que ha sido antes mencionada.

1. Referencia

Para describir la problemática que motivó esta investigación, en el capítulo precedente hablamos de *mecanismos o relaciones de referencia* sin especificar qué debemos entender por *referencia*, noción esencial para el estudio de la lengua como sistema de comunicación. El concepto en cuestión cobra especial importancia para nuestro trabajo, porque queremos describir cierto tipo de estructuras lingüísticas que mantienen relación con otras partes del texto a partir de *mecanismos de referencia* que, entre otras funciones, sirven para *referir* o *reintroducir* algo en el discurso sin enunciarlo de la misma manera.

Para los fines de este trabajo y antes de entrar de lleno a la problemática que nos ocupa, es preciso entender el significado del concepto *referencia* en dos sentidos distintos. En primer lugar, *referencia* es la propiedad esencial del lenguaje que emerge de la *relación denotativa* entre entidades lingüísticas y referentes extralingüísticos [Donnellan, 1966; Green, 1989; Frege, 1998]. En esta perspectiva, hablamos de que la lengua es un código especializado cuya característica principal es la transmisión de significados que *refieren* realidades, objetos, ideas y conceptos del mundo.

Desde la filosofía del lenguaje y sin ahondar en las distintas discusiones [Milner 1989; Beaugrande & Dressler, 1997; Frege, 1998;], la *referencia* se concibe como la designación de un *referente*, es decir, la vinculación semántica que se establece entre una *expresión lingüística* y una determinada *entidad* o *referente del mundo* [Frege, 1998]. En otras palabras, *referencia* es la relación entre secuencias lingüísticas y la realidad externa al acto comunicativo, constituye una propiedad designativa que se formaliza en el discurso y se concibe como la interacción *palabra-mundo* [Milner, 1989; Saiz, 2002]. En un primer sentido hablaremos de *referencia* como el vínculo entre el sistema de la lengua y el mundo extralingüístico.

En segundo lugar, nos interesa subrayar que el término *referencia*, en lingüística, se emplea también para denominar el fenómeno producido cuando ciertas unidades de la lengua aluden o *refieren* a otras construyendo *relaciones referenciales*. Hablamos de mecanismos que permiten el flujo de la comunicación, ya que en el discurso hay expresiones que hacen referencia a otras [Gundel *et al.*, 1993; Gaiffe *et al.*, 1994; Reboul, 1999].

Con la finalidad de acotar el objeto de estudio de esta investigación y con base en la diferencia que planteamos, se identificará como *referencia lingüística* a la relación referencial entre el sistema lingüístico y los referentes extralingüísticos.¹ Por otro lado, se denominará *referencia discursiva* al fenómeno que se produce cuando ciertas expresiones lingüísticas refieren a otras del mismo discurso. No se trata de propiedades excluyentes pero es importante entender que tienen alcances distintos, sobre todo si pensamos que la *referencia* primero se manifiesta en la relación lengua-mundo (*referencia lingüística*) y luego en la relación de las entidades lingüísticas entre sí (*referencia discursiva*). Es importante precisar que nuestro trabajo está enfocado a describir los fenómenos que atañen a la *referencia discursiva*.

¹ La relación entre palabras y cosas es una *relación de referencia*. Para Lyons [1970] las palabras se refieren y se relacionan con cosas existentes; entendiendo “existentes” en el sentido de existencia física, pero también ficticia y abstracta.

2. Expresiones referenciales

Dijimos previamente que la *referencia discursiva* implica cierto tipo de relación entre expresiones lingüísticas. ¿Qué clase de relación de referencia se establece entre las expresiones del discurso? ¿En qué sentido se dice que una expresión refiere a otra? ¿Cómo se ven afectados los elementos prominentes de un CD en una relación de referencia? En el siguiente CD vemos que existe un *término* (entidad nominal), ligado a un *patrón verbal* (verbo definitorio) que introduce una *definición* (fragmento textual):

Tabla 2-1. Ejemplo de contexto definitorio

Los cromosomas de eucariotas constan de largas hélices de ADN, cuyos nucleótidos se cuentan por cientos de millones.
Término = Los cromosomas de eucariotas (entidad nominal)
PVD = constan de (verbo)
Definición = largas hélices de ADN, cuyos nucleótidos se cuentan por cientos de millones (fragmento textual)

Las expresiones lingüísticas que participan de un CD (Tabla 2-1) pueden ser descritas como tipos de palabras si atendemos a la categoría gramatical (artículo, nombre, adjetivo, preposición, verbo, etc.); pero también pueden describirse como estructuras sintácticas (sintagma nominal, complementos, sintagma verbal, nexos oracionales, etc.). En este sentido, las relaciones entre los elementos del CD se dan a nivel de palabras, pero también de estructuras más complejas. Para poder describir el tipo de expresiones o estructuras que generan candidatos incompletos en la extracción automática, hemos optado por acudir al concepto *expresión referencial*.

Se han denominado *expresiones referenciales* (en inglés *referring expressions*) a las entidades lingüísticas que cumplen una función básica en la construcción del discurso, porque identifican los objetos, conceptos, ideas y realidades comunicables [Vázquez, 2004]. Se trata de estructuras que se emplean todo el tiempo en la comunicación y podemos decir que en ellas se manifiesta primordialmente la *propiedad denotativa del lenguaje* (referencia lingüística) [Grosz 1981; Hawkins 1984; Gundel 1993].

En principio, las *expresiones referenciales* se vinculan con *unidades léxicas* porque la relación de referencia con un referente extralingüístico es directa. Los *nombres propios*, por

ejemplo, poseen contenido referencial completo, tienen “denotación” pero no “connotación”² y se consideran *designadores rígidos*: estructuras que en grado mayor despliegan *referencia lingüística* [Kripke, 1972].

Para algunos autores, las *expresiones referenciales* abarcan un universo mayor que los nombres propios e incluyen otras unidades léxicas, tales como *sintagmas nominales*, *expresiones definidas e indefinidas*³ e incluso *pronombres* [Kripke, 1972; Putman, 1988; Green, 1989; Gundel *et al.*, 1993]. La amplitud del concepto deriva de la posibilidad que tenemos para referir una misma entidad de múltiples formas, por ejemplo: *enzima, una enzima, la enzima, esta enzima, aquella enzima, su enzima, esa enzima, la misma enzima, aquella, la suya, la otra, ésa, ésta*.⁴ En cada una de las estructuras existe una relación de denotación con un determinado referente extralingüístico, se trata de *expresiones* que remiten a cierta entidad del mundo [Gundel *et al.*, 1993: 275].

Además de los nombres propios, entidades nominales como los *sintagmas nominales* son también expresiones referenciales porque poseen contenido léxico propio y mantienen relación directa con un referente extralingüístico; es el caso de *términos* tales como *modelación hidrodinámica, bomba hidráulica, ecología del paisaje, esquema mexicano de privatización portuaria* [CLI, 2004], por dar algunos ejemplos.

Se ha planteado también que los *pronombres* son expresiones referenciales pero distintas de los *sintagmas nominales*, ya que son formas léxicamente vacías que exigen información semántica y sintáctica de otras partes del discurso. Los pronombres se supeditan al contexto y no a la relación directa con una entidad en el mundo; pese a esto, han sido tratados por varios autores como *expresiones referenciales* que hacen referencia a

² El *significado denotativo o referencial* es el que encontramos en los diccionarios y se identifica como una relación directa entre *referente-expresión lingüística*. En cambio, el *significado connotativo o asociativo* está vinculado a los conocimientos enciclopédicos, al conocimiento de mundo; es el significado adicional que da un sentido secundario y depende del contexto de enunciación. Es además parte de un proceso discursivo. [Beristáin, 2004:106]

³ En adelante, se utilizará el concepto *expresión* para denotar cualquier sintagma nominal actualizado por un determinante que puede ser un *demonstrativo* o un *posesivo*, además de artículo *definido* o *indefinido*. Las expresiones pueden ser *singulares* o *plurales*, en virtud del número gramatical y serán *masculinas* o *femeninas* en virtud del género. Se trata de un concepto muy cercano al *sintagma* o *frase nominal* del español y aquí se usarán indistintamente estos términos.

⁴ Gundel [1993] desde una perspectiva cognitiva propone una jerarquía (*The Givenness Hierarchy*) en la que reconoce seis formas distintas de *expresión referencial* de acuerdo con el estatus cognitivo del hablante frente al objeto de comunicación o referente. Las posibilidades para elegir distintas formas de expresiones referenciales se desprenden de la situación, del contexto de uso, de la intención del hablante y de la posición frente al objeto en cuestión.

objetos, ideas, personas y conceptos por medio de una *relación discursiva* [Donnellan, 1966; Kripke, 1972; Putman 1988; Green, 1989; Gundel *et al.*, 1993; Vázquez, 2004].

Bajo las precisiones anotadas, las *expresiones referenciales* son aquellas estructuras del discurso que denotan significado, es decir, poseen contenido léxico y sirven para nombrar entidades concretas del mundo, éstas pueden ser reales o conceptuales. Cabe aclarar que hay quienes consideran verbos y adverbios dentro de este grupo [Putman, 1988; Green, 1989], sin embargo, para los fines que persigue nuestra investigación es suficiente entender que se trata de formas lingüísticas que refieren directa o indirectamente (en el caso de los pronombres) a un referente o entidad concreta del mundo y, en principio, son estructuras ligadas al grupo nominal [Sarmiento y Esparza, 1993; Escandell, 1995].

Retomando el CD anterior (Tabla 2-1), podemos distinguir las diferentes expresiones referenciales que están en juego: no consideramos palabras sino estructuras que engloban significados prominentes del CD, es decir, los significados del *término* o de la *definición*:

Tabla 2-2. Expresiones referenciales en contexto definitorio

Los cromosomas de eucariotas constan de largas hélices de ADN, cuyos nucleótidos se cuentan por cientos de millones.
Expresión 1 =Los cromosomas de eucariotas (término)
Expresión 2 = largas hélices de ADN (definición)
Expresión 3 = cuyos nucleótidos (definición)

Localizamos, por lo menos, tres expresiones referenciales; cada una de ellas vinculada a un cierto referente en el mundo y puesta en el discurso de manera que guarda relación con otras entidades. En este sentido, las *expresiones referenciales* despliegan doble referencia: *lingüística* y *discursiva* (Apartado 1), esto quiere decir que denotan referentes extralingüísticos, mientras establecen relaciones específicas con el resto de las entidades del discurso.

3. Tipos de referencia discursiva

Veremos que los rasgos particulares de cada expresión (tipo de sintagma, determinantes,⁵ complementos, tipo de pronombre, etc.) prescriben la clase de *referencia discursiva* que se manifiesta. Un sintagma nominal con determinante definido y un pronombre, no despliegan la misma clase de relación con las entidades que los acompañan; un pronombre siempre acudirá a otra parte del discurso para completar su significado, mientras el sintagma denota significado sin recurrir a otras expresiones. Esto se expone con mayor claridad describiendo cinco tipos de *referencia discursiva* [Gaiffe *et al.*, 1994] que en adelante serán indispensables para explicar lo que ocurre en los CDs:

(1) **Referencia directa:** Ocurre a partir del significado o contenido léxico de la expresión referencial que apunta de manera directa a un referente extralingüístico; quiere decir que la expresión despliega su significado completo independientemente de las relaciones que pueda o no establecer con otras entidades.

- *El núcleo atómico es un objeto muy pequeño y denso que consta de dos tipos de nucleones* (3 expresiones referenciales distintas).

(2) **Referencia anafórica:** La expresión referencial recupera a su referente extralingüístico en otra expresión que aparece en el contexto, atañe principalmente a los pronombres.

- *Las grandes células epiteliales especializadas de la corteza externa se conocen como células "nodrizas" dado que cada una de ellas puede asociarse con grandes cantidades de linfocitos* (1 expresión referencial anafórica).

(3) **Referencia deíctica:** El referente lingüístico de la expresión se resuelve en la situación de enunciación, en un contexto específico. Se entiende por deíctico cualquier elemento que remita al contexto extralingüístico; esta clase de referencia se identifica primordialmente con adverbios y ciertos usos de los pronombres.

- *Aquí quédate, la maestra dijo ayer que la esperaras / Pásame ese de ahí*

⁵ Los determinantes son modificadores del nombre. Tienen la facultad de dar propiedades referenciales a una unidad nominal o a una secuencia. Éstos pueden ser: *artículos, cuantificadores, indefinidos, numerales, demostrativos, posesivos, relativos y exclamativos*, en todos los casos se trata de un paradigma cerrado, es decir, de un número limitado de formas [Escadell, 1995].

(4) **Referencia demostrativa (anafórica y deíctica):** Mantiene relación de carácter demostrativo con una expresión del discurso o con un referente extralingüístico. Cuando la expresión demostrativa apunta a una entidad discursiva previamente enunciada, la referencia posee un *valor anafórico*. Si la expresión señala a un referente extralingüístico que no ha sido explícitamente referido y además sólo se recupera en el momento de la enunciación, la expresión demostrativa posee *valor deíctico*.

- *Los temblores son producidos como consecuencia de la liberación repentina de energía en el interior de la Tierra. Esta energía se transmite a la superficie en forma de ondas sísmicas. (valor anafórico)*
- [Alguien frente a un gato dice] “Este gato me provocó alergia” (valor deíctico)

(5) **Referencia indirecta:** El referente se determina a partir de la referencia de otra expresión. Se identifica sobre todo con usos metafóricos e idiomáticos.

- *La mesa cinco se fue sin pagar*

Conforme a estos cinco tipos de referencia, distinguimos que un sintagma nominal con determinante definido, indefinido, cuantificador o numeral señala de manera directa a su referente extralingüístico y despliega su significado independientemente de la relación con otras entidades (*referencia directa*). Por otro lado, un pronombre expresa una *referencia anafórica*, ya que su significado se supedita a otra entidad discursiva. Un sintagma nominal con demostrativo se perfila como *referencia demostrativa*, esto quiere decir que debido a su contenido léxico posee relación directa con un determinado referente extralingüístico, pero simultáneamente está ligado a la entidad discursiva a la cual señala [Gaiffe *et al.*, 1994]. Cabe aclarar que las *expresiones referenciales* inmersas en una *referencia indirecta* (tipo 5) están sobre todo relacionadas con el *uso atributivo de la lengua*.⁶

La *referencia discursiva*, según los cinco tipos expuestos, comprende mecanismos que sirven para introducir, reintroducir, apuntar, señalar o referir una expresión referencial o un referente extralingüístico. Según la problemática planteada en el capítulo anterior, creemos que esta investigación debe centrarse en aquellas unidades léxicas, en particular *sintagmas*

⁶ Donnellan [1966] plantea una diferencia entre el *uso referencial* del lenguaje y el *uso atributivo*. El referencial alude a la identidad de la cosa y está relacionado con el significado denotativo o referencial. En el uso atributivo, en cambio, no existe una denotación directa con un referente extralingüístico, el significado de la referencia depende totalmente de su descripción.

nominales y *pronombres* que inmersos en la comunicación se vinculan entre sí a partir de una relación de referencia discursiva ¿Por qué en particular *sintagmas nominales* y *pronombres*? En el primer acercamiento al corpus (Tabla 1-11) se observó que éstas son la clase de expresiones que desencadenan candidatos incompletos; además, los elementos prominentes de un CD son el *término* y la *definición*, el primero corresponde siempre a una expresión referencial y el segundo puede abarcar más de una.

En la siguiente tabla se expone una *tipología de expresiones referenciales* planteada para la lengua francesa [Gaiffe *et al.*, 1994] y adaptada al español en este trabajo con el objetivo de mostrar algunas diferencias entre las expresiones más prominentes: *sintagmas nominales* y *pronombres* [Gaiffe *et al.*, 1994]. Las expresiones se clasifican según su determinante (definido, indefinido, demostrativo, cuantificador y posesivo). Esto permite distinguir formas que denotan significado completo independientemente del contexto, de aquellas que completan su significado apoyadas en otras expresiones. Así pues, en la tipología se marcan dos características principales: (a) contenido léxico a partir del cual decimos que la expresión apunta o no directamente una entidad en el mundo (*referencia directa*); (b) expresión que según el tipo (determinante) se supedita o no al discurso para completar su significado:

Tabla 2-3. Tipología de Expresiones referenciales

	Tipo de expresión referencial	Ejemplos	Contenido referencial o léxico	Supeditada al discurso
1	Nombres propios	<i>Gómez Orea, M. Godron y G. Merriam, UICN, Colombia.</i>	SI	NO
2	Expresiones definidas (<i>SN con determinante definido</i>)	<i>El apartarrayos, el proceso de cavitación, el micro corrientómetro digital portátil.</i>	SI	NO
3	Expresiones indefinidas (<i>SN con determinante indefinido</i>)	<i>Una bobina, una trampa de línea, un relevador del tipo sencillo.</i>	SI	NO
4	Sintagmas nominales con cuantificador	<i>Algunos resultados, todos los dispositivos de arranque.</i>	SI	NO
5	Expresiones con posesivos	<i>Su actuación, su trayectoria, sus recursos naturales.</i>	SI	SI
6	Expresiones con demostrativo	<i>Este modelo, estos fluidos del ciclo, dicho patrón de búsqueda.</i>	SI	SI
7	Pronombres	<i>Yo, éste, ello, la, mí, conmigo.</i>	NO	SI

Observamos que en (1) están los *nombres propios* separados del resto por su contenido léxico inminente, característica que los hace referencial y discursivamente independientes de otras entidades para reproducir su significado, además no poseen determinante. En 2, 3, 4 se ubican expresiones con contenido léxico pero que no son designadores rígidos, de tal manera que abarcan desde *sustantivos comunes* hasta *sintagmas nominales* con determinante definido, indefinido y cuantificador. En (5,6) se localizan los *sintagmas con posesivo* y con *demonstrativo* que presentan contenido referencial propio pero se distinguen de los anteriores por estar supeditados a alguna entidad del discurso, lo cual los hace dependientes de otras expresiones. Finalmente en (7) se encuentran los *pronombres*, formas que para denotar significados forzosamente recuperan el contenido léxico de otras expresiones y están totalmente supeditados al contexto en el que aparecen.

La tipología permite sugerir, primero, que los *pronombres* se encuentran en el extremo opuesto de los *nombres propios* (designadores rígidos) por estar léxicamente vacíos. Segundo, las *expresiones con posesivo, con demostrativo y pronombres* son referencialmente dependientes de otras entidades y en consecuencia del discurso. Así pues, las *expresiones referenciales* pueden describirse en función de sus propiedades léxicas (*contenido referencial*) y de la relación de dependencia que mantienen con otras partes del discurso.

4. Mecanismos de referencia discursiva

Hemos visto que las características de las expresiones referenciales, concretamente el determinante, prescriben el tipo de referencia discursiva y por tanto el tipo de relación con otras expresiones. Como ya se expuso, los candidatos incompletos están encabezados principalmente por pronombres, sintagmas nominales con determinantes demostrativos o bien ausencia de alguna expresión (Tabla 1-11). Retomando los tipos de *referencia discursiva* propuestos en este apartado, podemos plantear la presencia de algunos fenómenos que se ven implicados, se trata de nociones tales como *repetición* o *recurrencia*, *sustitución*, *deixis*, *elipsis*, *foricidad* y *marcadores u operadores discursivos* [Van Dijk, 1983; Mederos, 1998].

Conforme se hizo la revisión teórica, se advirtió que el fenómeno de *foricidad* es

sumamente productivo para describir los problemas que presentan los candidatos incompletos. Se explican también el resto de los fenómenos, ya que todos ellos están interrelacionados y son parte de los mecanismos de referencia discursiva.

4.1. Repetición y sustitución

En general, buena parte de los mecanismos implicados en la referencia discursiva están vinculados con dos procesos: *repetición* y *sustitución*, ambos fundamentales para entender cómo una expresión referencial puede recuperar otra anterior [Dijk, 1983; Mederos, 1998]:

La *repetición* o *recurrencia* es el hecho de referir a un elemento o parte anterior del discurso por medio de la repetición de alguna entidad con contenido léxico. Distinguimos cuatro variantes principales [Dijk, 1983; Mederos, 1998; Renkema, 1999]:

(1) **Repetición léxica total** en la que se reintroduce tal cual un elemento léxico o bien se mantiene el núcleo nominal de la expresión referida, p.e. *la molécula de agua- la molécula H₂O*.

(2) **Léxica sinonímica** en la que se reitera el significado de algún elemento por medio de otro empleado como sinónimo, p.e. *lentes/gafas*. Salvando las discusiones de la semántica y la lexicología, para fines prácticos los sinónimos son las palabras que tienen un significado similar o idéntico en ciertos contextos. La relación de sinonimia se da entre dos palabras cuyas propiedades léxicas son semejantes, supone equivalencia entre dos términos semánticamente intercambiables en una situación y contexto determinados.

(3) **Repetición mediante lo designado o por meronimia** es una vinculación semántica entre una unidad léxica que denota una parte y otra que denota el correspondiente todo (relación inclusiva parte-todo), p.e. *borrego-lana / alumno-escuela*. Un *merónimo* indica la parte del *holónimo* que representa al todo. Así *cocina, recamara, sala, estudio* son merónimos de su holónimo *casa*.

(4) **Repetición por hiperonimia** se manifiesta entre elementos con significado más general e incluyente, p.e. *perro-animal*. La relación *hiperonimia-hiponimia* distingue dos elementos de distinto nivel: un *hiperónimo* que es aquel término general que se emplea para referirse

al significado denotado por un *hipónimo*, es decir, un término más particular con significado más preciso. La relación se establece en función de los rasgos semánticos que los términos comparten, se trata de una inclusión conceptual en donde *perro*, *gato*, *pájaro* son *hipónimos* de la categoría más amplia *animal* que funciona como *hiperónimo*. No es una relación parte-todo sino una relación en la que el sentido de una palabra comprende el sentido de otra [Charaudeau, 1992:52].

Podemos anotar que los mecanismos referenciales que implican *repetición* o *recurrencia* están vinculados fundamentalmente con expresiones de tipo nominal que pueden ser unidades léxicas simples o sintagmas. Se trata de un fenómeno discursivo en el cual una expresión es reintroducida o referida por medio de la misma o de otra con contenido léxico inminente.

La *sustitución*, en cambio, es el reemplazo de un elemento, palabra o parte del discurso por otra estructura o forma que funciona como sustituto. Está identificada con el uso de *proformas* que pueden ser *léxicas*, *pro-adverbios* o *pronombres* [Mederos, 1998; Saiz, 2006]:

(1) **Proformas léxicas** comprenden estructuras especializadas en la sustitución, se puede tratar de verbos o sintagmas nominales que reemplazan otras estructuras p.e. *Valeria escribe la tesis₁* y *José hace lo mismo₁* // *Chocaron en el Periférico₁*, *el hecho₁* se produjo en la *madrugada*.

(2) **Proformas adverbio** son formas que tienen valor locativo y deíctico, pueden sustituir segmentos discursivos con valor adverbial, los principales son *aquí*, *allí*, *allá*, *así*, p.e. *En el texto se han explicado algunas características, se dijo aquí que los pronombres son anáforas*.

(3) **Proformas pronombre**⁷ engloba principalmente *pronombres personales de tercera persona* (él, ella, ello, ellas, ellos), los *demonstrativos* (éste, ése, aquél, etc.), los *posesivos* (el suyo, la suya, la tuya, etc.) y los *relativos* (el cual, la cual, lo cual, los cuales, el que, etc.). En todos los casos se trata de formas con valor anafórico, a diferencia de los

⁷ El paradigma pronominal se presenta en el Apéndice A de este trabajo.

pronombres *personales de primera y segunda personas* (yo, tú, usted, nosotros, ustedes) que son de carácter deíctico (*referencia deíctica*). Más adelante explicaremos puntualmente la diferencia entre pronombres anafóricos y pronombres deícticos; ahora nos interesa subrayar que en general se trata de formas que pueden sustituir una expresión referencial o una parte del discurso.

Observamos que en el fenómeno de sustitución quedan comprendidos los pronombres, un tipo de expresión referencial que sirve para referir, a veces sustituyendo, otra entidad discursiva a la cual se supedita por completo.

Repetición y sustitución son mecanismos referenciales discursivos muy próximos, para este trabajo es importante marcar la diferencia, ya que la *repetición* genera *relación entre unidades léxicas o sintagmas nominales*, mientras la sustitución construye *relaciones* entre *proformas* y alguna otra entidad del discurso.

4.2. Deixis

Hasta ahora hemos hablado de *referencia deíctica* y de *referencia demostrativa con valor deíctico* (Apartado 3), en ambos casos está implicada la noción de *deixis* que explicaremos brevemente para entenderla con mayor precisión.

La *deixis* se refiere al lugar y tiempo de enunciación, es el factor espacio-temporal que se conoce como *situación comunicativa* [Fillmore, 1997]. Se entiende por deíctico cualquier elemento que remita al contexto extralingüístico inmerso en la comunicación; por ejemplo, los *pronombres personales* (1ª y 2ª persona), los *demostrativos* (en contexto situacional) y los *adverbios* se interpretan conforme a ciertas coordenadas de enunciación [Green, 1989]. Bajo esta perspectiva, *deixis* se entiende como una *referencia fuera del texto*, ligada a procesos que dependen del contexto situacional y remiten al acto de enunciación. La *deixis* es la referencia que se hace a la temporalidad de un evento, su locación física, o a los participantes dentro de dicha escena [Fillmore, 1997; Lyons, 1983: 228-241].

Existe, sin embargo, una distinción entre *deixis no textual* (absoluta) y *deixis textual* (relativa) [Saiz, 2006:31]. En el primer tipo, la relación deíctica entre entidad discursiva y referente extralingüístico es absoluta porque la expresión se descifra sólo en un contexto específico. Un ejemplo de esto es: *¡muévete para allá que me estás lastimando!*, podemos

anotar que los referentes extralingüísticos se interpretan únicamente en el momento de la enunciación. En el segundo tipo, en cambio, la referencia es intratextual, ya que ciertas entidades del discurso como pronombres, adverbios y verbos hacen referencia a otros elementos del mismo discurso, por ejemplo: *en el artículo se plantea que el ahorro de agua es importante y ello implica que la propuesta se concentre en dicho ahorro*, texto en el cual observamos que existe un tipo de referencia que se interpreta conforme a los elementos del mismo discurso, en este sentido se habla de *deixis textual* [Saiz, 2002].

Los pronombres son una de las principales categorías vinculadas con el fenómeno de deixis (*de persona*), pero también existe *deixis de tiempo* expresada esencialmente por los verbos y *deixis de lugar* denotada por los adverbios que remiten a las coordenadas espacio-temporales de enunciación; todas ellas construyen la *deixis del discurso* [Renkema, 1999: 102; Fillmore, 1997; Green, 1989], tema muy extenso y que sólo es pertinente mencionar en tanto atañe a ciertos mecanismos referenciales discursivos.

4.3. Elipsis

La *elipsis* puede definirse como la omisión de un elemento o una parte del discurso; es un mecanismo en el que se suprimen elementos repetidos que quedan sobrentendidos [Moreno *et al.*, 1999]. No todo lo que se sobrentiende en un texto es elipsis, ésta tiene que ver con la ausencia o supresión de algún elemento lingüístico.

Existen definiciones diversas de este concepto. Según Allen [1995] la elipsis implica el uso de oraciones incompletas sintácticamente. Para Franchini [1986] la elipsis es el fenómeno o procedimiento de la omisión de elementos sintagmáticos (nucleares) en el discurso. Covington [1994] habla de elipsis como la omisión de palabras repetidas dentro de la frase. Pese a las diferencias, todas las definiciones consideran que elipsis es un mecanismo de economía lingüística, es decir, una referencia en la cual se suprimen elementos repetidos de un texto que quedan sobreentendidos para el receptor del mensaje.

La elipsis se da en todos los niveles de la lengua, existe omisión de nombre, de pronombre, del núcleo de sintagma nominal o verbal, de adverbios y de oraciones, entre otras [Abad, 1980]. La elipsis es otro de los grandes problemas de referencia discursiva

abordados en los trabajos de extracción automática y de PLN,⁸ se trata de un fenómeno que convive con los mecanismos de *repetición* y *sustitución* en tanto que retoma “en ausencia” algo previamente dicho [Palomar *et al.*, 1995].

4.4. Marcadores u operadores discursivos

Los *marcadores del discurso*, también denominados *operadores del discurso*, *operadores pragmáticos*, *conectores* o *enlaces extraoracionales*, despliegan un muestra amplia de estructuras [Núñez, 1993]. Estos elementos sirven principalmente para conectar oraciones o partes del texto y tienen un fuerte valor referencial, sea hacia adelante o hacia atrás del discurso. Abarcan conjunciones, interjecciones, adverbios, locuciones adverbiales y prepositivas, entre otras [Cortés, 1991]. El repertorio es tan grande que resulta difícil establecer un valor general, algunos están lexicalizados, son formas fijas que en la mayoría de los casos funcionan a nivel transoracional. En seguida se enlistan algunos de los marcadores más frecuentes:

Además, asimismo, aún, todavía, encima, aparte, después, de igual forma, también, por otra parte, por otro lado, cierto, evidente, sin duda, seguro, por supuesto, bien, bueno, en efecto, por cierto, desde luego, si acaso, en consecuencia, por consiguiente, por lo tanto, así pues, de ahí que, por eso, por lo cual, por ende, en fin, por fin, por último, nada más, a pesar de todo, pese a, con todo y eso, en conclusión, a fin de cuentas, total, a condición de que, con tal (de) que, de ahí, hora bien, siquiera, para colmo, por ejemplo, tal como, en primer lugar, primero, luego, después, es decir, o sea, a saber, en otras palabras, mejor dicho, no obstante, en cambio, sin embargo, al contrario, si acaso, excepto (que si), en resumen, en una palabra, en pocas palabras, por lo que respecta, etc.

4.5. Foricidad

Muy próximo a los fenómenos de *repetición* y *sustitución*, se encuentra el *fenómeno de foricidad* conocido también como *fenómeno de anáfora*, un tema complejo que puede tener

⁸ El Procesamiento del Lenguaje Natural (PLN), es una rama de la Inteligencia Artificial en la cual se diseñan mecanismos y modelos computacionales capaces de simular el conocimiento de los humanos para poner en marcha las facultades lingüísticas y cognitivas que permiten la interacción hombre-máquina. EL PLN ha logrado avances importantes en extracción de información, minería de textos, traducción automática, sistemas pregunta-respuesta y resumidores automáticos, entre otros [Azzam, *et al.*, 1998; Mitkov, 1998; Yang *et al.*, 2003; Gelbukh y Sidorov, 2006].

distintos alcances según la teoría lingüística que lo explica. Es importante subrayar que se trata de un mecanismo de referencia discursiva que engloba procesos de sustitución y repetición, ya que se identifica tanto con la repetición de unidades nominales como con la sustitución a partir de proformas.

En las siguientes páginas se hace una revisión del tema *anáfora*, partiendo de la idea de que existen diversas posturas al respecto y con la finalidad de retomar la información que sirva para describir dicho fenómeno en los CDs.

5. Anáfora

El concepto *anáfora* fue acuñado desde la gramática griega, viene del vocablo $\alpha\nu\alpha\phi\omicron\rho\alpha$ que pasó al latín como *anapherein* ‘repetir, recordar’ (*ana-* ‘atrás, hacía atrás’ + *pherein* ‘llevar’). *Anáfora* es el término que en general se emplea para señalar el hecho de hacer referencia a algo anteriormente mencionado. En principio, dicha palabra sirvió para catalogar a los *pronombres*, formas que remiten a segmentos ya referidos en el discurso [Martínez y Álvarez, 1989].

Cabe señalar que el uso del término *anáfora* se ha extendido y no sólo considera pronombres, incluye cualquier expresión que recupera algo previamente enunciado. Además, y es importante precisarlo, *anáfora* se utiliza también para nombrar un tipo de mecanismo referencial que a grandes rasgos puede describirse como “la relación de una palabra o frase de cualquier tipo y otra palabra o frase anterior” [Peregrín, 1999:1430]. En este sentido, *anáfora* no sólo es la entidad lingüística que recupera una parte del discurso, sino que indistintamente nombra al mecanismo referencial en el que participa:

La *anáfora* se define como la relación de referencia que se establece entre una forma lingüística (*expresión anafórica*) y un objeto, una persona o una situación que ya han sido mencionados de manera implícita o explícita con anterioridad durante el proceso comunicativo y que llamaremos *antecedente* [Palomar *et al.*, 1995: 128].

Esta definición incorpora dos nociones básicas: *expresión anafórica* que es la entidad que “refiere” algo antes mencionado y *antecedente*⁹ que es ese “algo” de lo que

⁹ Generalmente se usan como sinónimos *antecedente* y *referente* en el ámbito de la referencia discursiva y del fenómeno de anáfora; sin embargo hay que aclarar que de manera estricta, para la filosofía del lenguaje, el

previamente se habló [Moreno *et al.*, 1999: 205]. Bajo esta perspectiva, el concepto *anáfora* nombra al mecanismo en el cual una entidad discursiva refiere a otra del mismo discurso: *foricidad*.

5.1. El rango conceptual del término *anáfora*

Es necesario apuntar ciertos problemas de nomenclatura respecto al tema de *anáfora*. Líneas arriba dijimos que la palabra tiene un uso amplio y nombra dos cosas distintas. Generalmente *anáfora* se usa para referir al mecanismo que relaciona un *antecedente* y una *expresión anafórica*, en este sentido funciona como sinónimo de *referencia anafórica*, *relación anafórica*, *fenómeno de anáfora* o *mecanismo de anáfora*. Con significado más restringido y preciso, la palabra *anáfora* denomina concretamente a la forma lingüística que reintroduce y refiere a una entidad enunciada en el discurso; entonces se usa como sinónimo de *expresión* o *elemento anafórico*. Así pues, *anáfora* es el mecanismo referencial pero también es el elemento implicado en dicho mecanismo. El uso indiferenciado de un mismo término para designar dos cosas distintas genera problemas de nomenclatura que causan que los conceptos tengan que ser explicados y delimitados según el uso que se les da.

El término *anáfora* únicamente nombra el mecanismo referencial hacia atrás del discurso. Hablamos del elemento que recupera y rastrea un antecedente (*anapherein* ‘repetir, recordar’). Cabe señalar que en el discurso, existen relaciones que se dan en prospectiva, hacia adelante del mensaje y cuando el *antecedente* o *referente lingüístico* aparece después de la expresión abreviada que lo refiere, p.e. *Quiero esa₁, la chamarra azul₁*. Para distinguir esta variante, tradicionalmente se ha empleado el término *catáfora*.

Atendiendo a una nomenclatura más puntual tendríamos entonces que señalar que la *anáfora* no es el mecanismo referencial tal y como se plantea en la definición que presentamos; *anáfora* es el elemento o entidad principal de una *relación* o *referencia anafórica*.

referente es la cosa, objeto, concepto o idea que está en el mundo y es ajeno al sistema de la lengua; en tanto que el *referente* en el contexto de la *foricidad* es más bien el *antecedente* o *referente lingüístico* de una expresión anafórica. En este trabajo se especifica cuando se trata de un *referente extralingüístico* o de uno del discurso (*antecedente*), ya que la *expresión anafórica* apunta a un *referente del discurso* y no directamente al objeto en el mundo.

El término *relaciones fóricas* (fenómeno de foricidad) tiene un rango mayor, ya que engloba *relaciones anafóricas* y *catafóricas*, es decir, referencias “hacia atrás” y “hacia adelante” del discurso. Así pues, las *relaciones fóricas* involucran un *elemento fórico* (anáfora o catáfora) y un *antecedente*¹⁰ o *referente lingüístico* [Moreno-Cabrera, 1991]:

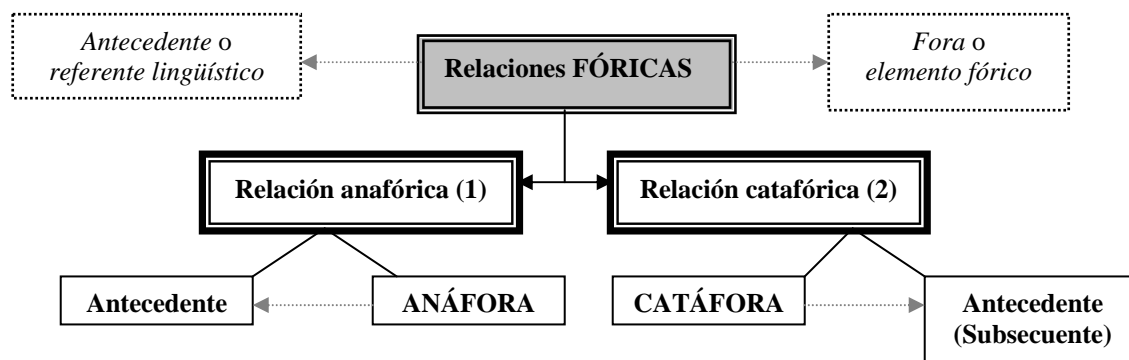


Figura 2-1. Relaciones fóricas (*foricidad*)

Tabla 2-4. Ejemplo de anáfora y catáfora (*relaciones fóricas*)

1	Anáfora	Los <i>alumnos premiados</i> ₁ serán <i>aquellos</i> ₁ con buenas calificaciones.
2	Catáfora	<i>Aquellos</i> ₁ con buenas calificaciones serán los <i>alumnos premiados</i> ₁

Otro de los problemas de nomenclatura y clasificación es la variedad de términos que existen para denominar un mismo elemento, por ejemplo: *anafor*, *anáfora*, *elemento*, *segmento*, *marca* y *expresión anafórica* son sinónimos [Gelbukh y Sidorov, 2006: 137]. En tanto que *catáfora*, *elemento*, *marca* y *expresión catafórica* se usan indistintamente. De la misma manera se emplean sin diferencia *marca*, *expresión*, *segmento* y *elemento fórico* que incluye *catáfora* y *anáfora*. Vemos así que la terminología es un problema importante en un tema que ya de suyo es muy complejo.

En la mayoría de los trabajos y estudios concentrados en la *resolución automática de anáfora*¹¹ [Liang y Wu, 1994; Azzam *et al.* 1998; Reinhart, 1983], se utiliza el término *anáfora* para designar tanto el mecanismo referencial como el elemento principal de dicho mecanismo. Esto, a pesar de que los conceptos *foricidad* y *fora* son más precisos y

¹⁰ Al hablar de *antecedente* se incurre en una imprecisión respecto a la *catáfora*, ya que en este caso la *relación fórica* no se da con un *antecedente* como tal, sino con un *subsecuente*, porque primero aparece el *elemento catafórico* y luego el *referente discursivo* al que alude [Peregrín, 1999:1430].

¹¹ La *resolución de anáfora* consiste en rastrear automáticamente la expresión a la cual una expresión anafórica hace referencia, [Moreno *et al.*, 1999: 205].

engloban relaciones “hacia atrás” o “hacia adelante” del discurso. En nuestro trabajo se usará indistintamente *anáfora* o *relación anafórica* para designar el mecanismo de referencia a un antecedente previamente enunciado; y se hablará de *relación catafórica* o *catáfora* para describir los casos específicos en que la referencia es “hacia adelante”.¹²

6. Clasificación de anáforas

Como consecuencia de los inconvenientes planteados, existen muchas definiciones, clasificaciones y distintas acepciones del término *anáfora*. No existe un consenso en la definición, ni tampoco en la descripción y clasificación de los elementos implicados en las *relaciones anafóricas*; como hemos visto, el término *anáfora* no corresponde a un significado unívoco.

La *anáfora* tiene distintos alcances según la teoría lingüística que la estudia, pero en general se describe como uno de los principales mecanismos de referencia discursiva que permite mantener la unidad de sentido de los mensajes y contribuye con la economía lingüística:

La anáfora es el mecanismo que nos permite hacer una referencia abreviada a alguna entidad o entidades, con la confianza de que el receptor del discurso sea capaz de desabreviar la referencia y por consiguiente determinar la entidad a la que se alude [Hirst, 1981 en Moreno *et al.* 1999: 204].

Veremos más adelante que los *pronombres* son caracterizados como la anáfora prototípica, porque cumplen categóricamente con las características de dicho elemento: sirven para referir o sustituir un segmento bien determinado del discurso y están totalmente relacionados con el tipo de *referencia anafórica* (ver Apartado 3). Cabe anotar que la anáfora presenta otras estructuras además de los pronombres; siendo que se trata de la referencia a una entidad concreta del discurso, engloba también la relación entre un *antecedente* y una *expresión referencial*, por ejemplo un sintagma nominal, que funcione como *expresión anafórica*.

¹² La razón por la cual no se emplearán los términos *relaciones fóricas* y *fora* es que la clasificación que presentamos en este trabajo está concentrada en la *anáfora* y no incluye la *catáfora*.

Existen varios criterios para clasificar anáforas [Olivares 1994, 1995; Moreno *et al.*, 1999; Saiz, 2002], en este trabajo nos ceñimos a tres en particular: (1) *anáfora en función del contexto en el que opera*; (2) *en función del tipo de elemento anafórico*; y (3) *en función del marco en el que ocurre* [Moreno *et al.*, 1999].

6.1. Anáfora en función del contexto en el que opera

En los trabajos de resolución automática de anáfora [Moreno *et al.*, 1999; Saiz, 2002; Olivas, 2006] se plantea que, en una relación anafórica, el *antecedente* puede ser mencionado explícita o implícitamente, quiere decir que no siempre éste ha sido enunciado y sin embargo la *expresión anafórica* lo introduce en el mensaje. Bajo esta idea se presenta una primera clasificación de *anáfora en función del contexto en que aparece* [Moreno *et al.*, 1999: 205]. El criterio se ciñe a la evocación explícita del antecedente de una expresión anafórica:

(a) **Anáfora en contexto situacional:** Es aquella que se da en un contexto preciso cuando el objeto de la comunicación no se ha mencionado explícitamente pero el emisor puede referirlo en una situación concreta. Expresiones tales como *pásamela / préstame ese/ quiero aquel*, sólo se resuelven en el momento de la enunciación. Se trata de un tipo de referencia identificada sobre todo con la lengua oral. Se habla de *anáfora en contexto situacional* cuando existe, por ejemplo, un pronombre que señala directamente a cierta entidad en el mundo y no a un antecedente en el discurso. Este planteamiento será descartado en nuestra investigación porque según lo que hemos precisado acerca de los tipos de referencia discursiva (Apartado 3), los pronombres y expresiones inmersos en usos situacionales, corresponden más bien a la manifestación de *referencia deíctica* y no precisamente *anafórica*.

(b) **Anáfora dentro del contexto lingüístico:** Es aquella que guarda relación con un referente que aparece en el mensaje. En este caso la expresión anafórica apunta a una entidad explícitamente enunciada (antecedente). Un ejemplo de estos es: *compré un videojuego₁ y mi hermano lo₁ estrenó/ La lámpara₁ que₁ alumbra la calle está fundida / Mi madre₁ me advirtió de la lluvia pero no quise escucharla₁*. La *anáfora en contexto*

lingüístico se da tanto en lengua oral como en escrita y se manifiesta como parte de una *referencia anafórica* (Apartado 3). En este trabajo el *contexto lingüístico* será un parámetro muy importante para determinar la presencia de relaciones anafóricas (incluye las catafóricas).

Tradicionalmente, para distinguir la *anáfora en contexto lingüístico* (b) de la que opera en *contexto situacional* (a) se utilizan los términos *referencia exofórica* (exóforas) que refiere entidades que no se nombran de manera explícita y *referencia endofórica* (endóforas) que señala referentes explícitamente enunciados [Peregrín, 1999]. En el marco de esta tesis se estudian únicamente *anáforas en contexto lingüístico* (*referencias endofóricas*), ya que el tipo de discurso que se analiza no está relacionado con contextos situacionales. Las denominadas “exofóras” apuntan directamente al referente extralingüístico y por tanto están más bien en el marco de la *referencia deíctica* (Apartado 3).

6.2. Anáfora en función del tipo de elemento anafórico

Para esta tesis el criterio de clasificación más importante es aquel que explica la *anáfora en función del tipo de elemento anafórico*¹³ y distingue *anáfora pronominal* de *anáfora de sintagma nominal*. El criterio básicamente separa *pronombres* del resto de *expresiones referenciales* con valor anafórico. Esta distinción es importante porque los pronombres en contexto lingüístico siempre funcionan como anáforas; en cambio, los sintagmas nominales sólo cuando refieren y reintroducen otra expresión son considerados como expresiones anafóricas [Moreno *et al.*, 1999; Saiz, 2002].

6.2.1. Anáfora de sintagma nominal (directa e indirecta)

Según el criterio descrito líneas arriba, existe la *anáfora de sintagma nominal* que se divide en *anáfora directa* y *anáfora indirecta* [Muñoz, 2001]. En ambos casos la referencia discursiva está ligada a procesos cognitivos y pragmáticos complejos que exigen

¹³ Cabe aclarar que la clasificación incluye varios tipos: *anáfora pronominal*, *anáfora de sintagma nominal* (descripciones definidas), *anáfora verbal*, *adverbios* y *complementos circunstanciales* [Moreno *et al.*, 1999; Saiz, 2006]. En este trabajo nos concentramos en dos: *anáfora pronominal* y *anáfora de sintagma nominal*, ya que estos tipos comprenden el problema de los candidatos incompletos.

conocimiento de mundo, pues se resuelven a partir de inferencias y relaciones semánticas.

Se habla de anáfora *directa* o *completa* cuando un antecedente y el sintagma nominal que opera como expresión anafórica comparten *núcleo nominal* [Woods, 1977]. En este caso el elemento anafórico recupera el contenido léxico de su antecedente, proceso que está totalmente vinculado con el fenómeno de *repetición léxica total* (Apartado 4.2):

Tabla 2-5. Ejemplo de anáfora de sintagma nominal directa

El elemento más importante del cromosoma es la molécula <i>continua de ADN</i> ₁ . Esta molécula <i>de doble cadena con forma de escalera retorcida</i> ₁ está formada por compuestos químicos enlazados.
Antecedente = la molécula continua de ADN
Anáfora de SN directa = Esta molécula de doble cadena con forma de escalera retorcida
Núcleo nominal = molécula

En el ejemplo observamos que el *antecedente* es un sintagma nominal que es reintroducido en el discurso por otro sintagma nominal que funciona como “anáfora”. A pesar de que se trata de dos estructuras con determinantes y complementos distintos, tanto el *antecedente* como la *anáfora* tiene el mismo núcleo nominal, en este sentido es que se habla de *anáfora de sintagma nominal directa*.

En contraste con esta clase existe la *anáfora indirecta*, que se da cuando un sintagma nominal (expresión anafórica) reintroduce un antecedente, pero el *núcleo nominal* de las estructuras no es el mismo:

Tabla 2-6. Ejemplos de anáfora de sintagma nominal indirecta

La <i>bicicleta</i> ₁ se averió. El <i>pedal</i> ₁ estaba atorado y la <i>cadena</i> ₁ rota.
Tenía cuatro <i>colores</i> ₁ para pintar la casa. El <i>azul</i> ₁ era el más feo.
Se cometió un <i>homicidio</i> ₁ en el parque. La policía investiga el <i>asesinato</i> ₁ .

La *anáfora indirecta*, también llamada *asociativa*, *parcial* o *infiel* [Woods, 1977] se puede dar en una *relación de meronimia* (1), *de hiperonimia-hiponimia* (2) o con un *sinónimo* (3) del antecedente. En los tres casos el núcleo nominal de los elementos implicados no coincide, sin embargo, existe un vínculo semántico que los relaciona.

El fenómeno de *anáfora indirecta* abarca, según lo que hemos dicho, la relación referencial semántica entre un sintagma nominal que funciona como expresión anafórica y

la expresión o entidad que es el antecedente. Esta clase de anáfora está vinculada con el proceso de repetición léxica y con las variantes que se plantearon previamente: *repetición por sinonimia, por meronimia y por hiperonimia* (Apartado 4.2).

Cabe señalar que la *referencia demostrativa con valor anafórico* (Apartado 3) se da entre un antecedente y un sintagma nominal con determinante demostrativo. En este caso el sintagma funciona como expresión anafórica y puede manifestarse como *anáfora directa* si repite el núcleo nominal de su antecedente, o bien, *indirecta* si la referencia se da a partir de una relación semántica. Los sintagmas con determinante demostrativo quedan contemplados dentro de la anáfora de sintagma nominal, por tanto, la *referencia anafórica* engloba la *referencia demostrativa con valor anafórico*.

6.2.2. Anáfora pronominal

Conforme al criterio de clasificación en función del tipo de elemento anafórico, el segundo tipo atañe a la *anáfora pronominal*. Como su nombre lo sugiere, incluye gran parte de los pronombres del español; el inventario es extenso y existen distintas clasificaciones. En este trabajo se sigue un enfoque tradicional en cuanto al paradigma pronominal (Apéndice A) [RAE, 1989; Martínez y Álvarez, 1989]; nuestro objetivo principal es mostrar el perfil anafórico de ciertos pronombres.

6.2.2.1. Pronombres

Para hablar del carácter anafórico de los pronombres se tiene que señalar que no todos ellos funcionan como anáforas. Tradicionalmente el pronombre se ha definido como la palabra (*proforma*) que puede sustituir al nombre [Roca-Pons, 1960:146; Lyons, 1983: 229], sin embargo, dicha característica sólo atiende a una de sus posibles funciones y a un determinado tipo de pronombres. En efecto, se trata de entidades lingüísticas que sustituyen ciertas expresiones enunciadas de manera explícita, pero también sirven para referir e introducir en la comunicación algo que no había sido antes mencionado.

En este sentido, se planteó desde la gramática griega que los *pronombres* poseen *carácter anafórico* si refieren a alguna parte del discurso (*antecedente* o *referente lingüístico*), o bien *carácter deíctico* si apuntan a objetos en el mundo (*referentes*

extralingüísticos). La diferencia entre pronombres con valor *deíctico* y con valor *anafórico* corresponde a la dicotomía *exófora vs. endófora* que ya se explicó (Apartado 6.1).

Es importante subrayar que, conforme a los tipos de *referencia discursiva* que se expusieron en un principio (Apartado 3), para esta investigación los pronombres son anáforas (*foras*) si y sólo si refieren a una entidad del discurso enunciada explícitamente, con lo cual queda sobrentendido que existen pronombres que no lo son. El pronombre *tú*, por ejemplo, no es una anáfora estrictamente hablando, se identifica más bien con “lo deíctico” porque su referente se interpreta sólo en la enunciación y en un escenario lingüístico concreto (referencia deíctica). Por otro lado, el relativo *que* se apega a “lo anafórico”, pues existe un antecedente en el discurso que lo rige y la relación referencial se constituye dentro del mensaje. Se sostiene entonces la idea de que los pronombres son “palabras que señalan y remiten a algo o a la representación de ese algo y son entonces *deícticos o anafóricos*” [RAE, 1989: 203].

Otra particularidad esencial de los miembros de esta categoría es que comprenden formas léxicamente vacías, pues no poseen contenido referencial propio, no son semánticamente autónomas y forzosamente señalan a referentes extralingüísticos cuando son deícticos o del discurso si son anafóricos. Cabe anotar que los pronombres no están totalmente desprovistos de significado, muchos poseen morfemas de género y número que inciden en la relación sintáctica que establecen con las expresiones a las cuales refieren [Martínez y Álvarez, 1989]. Por lo general, las marcas de género y número de los pronombres coinciden con el género y número de su antecedente y, aunque existen excepciones, esta tendencia es de gran ayuda en la *resolución automática de anáfora* [DuRoss, 1990].

A pesar de la amplitud del paradigma, los pronombres no son una clase abierta de palabras, poseen características morfológicas y sintácticas que permiten separarlos en subclases: *personales, posesivos, demostrativos, relativos, interrogativos, indefinidos y numerales*, según la Real Academia de la Lengua Española [RAE, 1989: 203]. No todos los gramáticos contemplan los mismos tipos, hay quienes sitúan aparte el grupo de los *reflexivos* [Martín, 1979; Peregrín, 1999], incluidos por La RAE dentro de los *personales*. En nuestro trabajo, la clasificación tradicional basta para identificar los rasgos mínimos que los inscriben dentro de una misma categoría, con el objetivo principal de rastrear su

carácter anafórico.

Cabe anotar que los *pronombres* son las formas típicas de la anáfora, se distinguen por depender semántica y sintácticamente de otras expresiones, por recuperar la información de algún segmento del discurso y por establecer relación anafórica con sus antecedentes. Esta investigación no atiende de manera especial a los pronombres como categorías gramatical, tampoco se explican las funciones que éstos pueden desempeñar en la oración; como dijimos, nuestro interés está concentrado en que se trata de expresiones referenciales que funcionan como anáforas en una relación discursiva. En este sentido, los pronombres son anáforas cuando refieren a un antecedente explicitado en el discurso y son formas deícticas cuando sólo pueden resolverse en una situación de enunciación específica (*referencia deíctica* y *referencia demostrativa con valor deíctico*).

Conforme a lo que se ha dicho podemos concluir que la referencia anafórica abarca procesos de repetición léxica, de sustitución y la referencia demostrativa con valor anafórico (Apartados 3 y 4.1). Según lo que hemos expuesto se trata de la relación entre una entidad discursiva que funciona como *antecedente* y una *anáfora* o *expresión anafórica* que lo refiere, dicha expresión, puede ser un pronombre en contexto lingüístico o un sintagma nominal.

6.3. Anáfora en función del marco en el que ocurre

La *anáfora en función del marco en el que ocurre* incluye dos tipos: intraoracional y discursiva [Moreno *et al.*, 1999]. La *intraoracional* ocurre y se resuelve dentro de una oración,¹⁴ quiere decir que tanto antecedente como expresión anafórica están en la misma estructura oracional:

¹⁴ El concepto *oración* se ha entendido de diversas maneras en los estudios de gramática española, su significado varía según se entienda como unidad conceptual o como unidad morfosintáctica. Con significado afín se han usado sinónimos tales como *proposición*, *cláusula* y *sentencia* [Lope Blanch, 1979: 37]. Para efectos de nuestro trabajo las *oraciones* son estructuras sintácticas con contenido comunicativo autónomo. En este sentido, entenderemos *oración* como una “proposición o conjunto de proposiciones con sentido completo”, cuyos elementos indispensables son un nombre y un verbo (relación sujeto-predicado) [Bello, 1951: 308]. En el marco de los CDs el indicador de una oración será una letra mayúscula y un punto final, es decir, estructuras que van de punto a punto.

Tabla 2-7. Ejemplo de anáfora intraoracional (CT del IULA)

Reacción nuclear ₁ que ₁ consiste en la división de un núcleo atómico ₂ en otros de menor masa ₂ al chocar con él ₂ determinadas partículas.
Antecedente 1 = Reacción nuclear (término, unidad nominal)
Anáfora 1 = que (pronominal)
Antecedente 2 = un núcleo atómico (parte de la definición)
Anáfora 2 = otros de menor masa (pronombre indefinido-sintagma nominal, elisión de núcleo nominal)
Anáfora 3 = con él (pronominal)

La otra variante es la anáfora *discursiva*, en la cual la expresión anafórica y el antecedente no se localizan en la misma oración. Se trata de anáforas que generan la conexión entre varios puntos de atención en el discurso, de tal manera que en un texto se puede reintroducir una entidad que ha sido enunciada varias oraciones atrás:

Tabla 2-8. Ejemplo de anáfora discursiva (CLI)

El mallado del cuerpo y los cálculos pertinentes se realizan mediante el programa Mallador30/Interfaz ₁ . Esta aplicación ₁ permite realizar cálculos de flujo alrededor de carenas Wigley con distintas relaciones de dimensiones principales y a distintas velocidades. Para ello se generan mallados de carena y superficie libre a los que se aplica el método potencial descrito.
Antecedente 1 = El mallado del cuerpo y los cálculos pertinentes se realizan mediante el programa Mallador30/Interfaz.
Anáfora 1 = Esta aplicación (referencia demostrativa con valor anafórico)
Antecedente 2 = Esta aplicación permite realizar cálculos de flujo alrededor de carenas Wigley con distintas relaciones de dimensiones principales y a distintas velocidades.
Anáfora 2 = Para ello (pronominal)

Se podría pensar que en el ámbito de los CDs, esta clase de anáfora no será común, ya que se trata de fragmentos textuales que muchas veces no comprenden más de una oración; sin embargo los candidatos incompletos presentan relaciones anafóricas con alcances transoracionales que podemos identificar con las características de la *anáfora discursiva*.

7. Observaciones finales

Llevada a cabo la revisión bibliográfica se observó que el fenómeno de anáfora abarca buena parte de los mecanismos referenciales que generan candidatos incompletos. En función de sus características trataremos de describir el tipo de expresiones y estructuras

que afectan los resultados en la extracción automática. Cabe aclarar que en esta breve revisión del tema no incluimos muchas posturas y consideraciones acerca de los mecanismos de referencia discursiva y de las relaciones anafóricas. En este sentido, se optó por retomar tres parámetros útiles para describir los alcances de la anáfora en los CDs, se trata de los criterios que consideran *anáfora en función del contexto en el que aparece*, del cual nos interesa la *anáfora en contexto lingüístico* (antecedente explícito en el discurso); *anáfora según el tipo de elemento anafórico*, que se divide en *anáfora pronominal* y *anáfora de sintagma nominal* (directa e indirecta); y *anáfora en función del marco en el que ocurre* (intraoracional y discursiva). Creemos que estos criterios de clasificación serán suficientes para rastrear las relaciones referenciales que nos interesa describir en los CDs.

Así pues, se da fin al segundo capítulo, en el cual se hizo una revisión del tema central para esta investigación con el objetivo de determinar las pautas en la identificación y descripción de *relaciones anafóricas* en un tipo de discurso denominado *contexto definitorio*.

**Capítulo III *Metodología y resultados de la
identificación de anáforas en
contextos definitorios***

Metodología y resultados de la identificación de anáforas en contextos definitorios

1. Metodología para la identificación de anáforas en contextos definitorios

Recordemos que esta tesis se concentra en la identificación y descripción de ciertos mecanismos de referencia discursiva, concretamente anáforas que están presentes en los CDs incompletos. Para establecer la tipología y determinar patrones recurrentes de relaciones anafóricas se trabajará con un corpus de CDs obtenidos de trabajos previos [Alarcón 2003; Alarcón 2006; Aguilar, 2006].

La selección de CDs para el corpus de esta tesis se hará en función de la disponibilidad de datos. El corpus se conformará de 250 CDs elegidos al azar de un repositorio que se edificó en una etapa previa a esta tesis. La identificación de estructuras y expresiones que desencadenan mecanismos de referencia anafórica se hará analizando cada uno de los CDs. Se trabajará con un corpus en formato electrónico con documentos TXT, es decir, archivos de texto plano que no tienen formato, sólo son caracteres. Se empleará el programa *Word* para manipular el corpus, ya que la herramienta “buscar-reemplazar” permite localizar información puntual. Con dicha herramienta podremos rastrear el paradigma pronominal del español y proceder a describir el valor anafórico de sus miembros. En el caso de la anáfora de sintagma nominal será necesario recuperar la expresiones anafóricas y sus antecedentes analizando contexto por contexto. La búsqueda de anáforas se llevará a cabo bajo los criterios que a continuación se explican.

1.1. Criterios para la identificación de anáforas en contextos definitorios

Considerando la problemática expuesta en el primer capítulo y la definición previa de algunos conceptos útiles para describir mecanismos de referencia, en adelante nos concretaremos en rastrear las expresiones anafóricas que dan como resultado candidatos incompletos. Para ello, tomaremos en cuenta tres criterios principales: (1) *criterios de clasificación de anáfora*, (2) *extensión y expansión de CDs* y (3) *diferencias entre candidatos completos y candidatos incompletos*.

1.1.1. Criterios de clasificación de anáfora

En el capítulo anterior se explicaron tres criterios que serán de gran utilidad para identificar estructuras y mecanismos referenciales que en el ECODE producen candidatos incompletos. Dichos criterios servirán como pautas para clasificar anáforas y describir sus usos; los enunciamos a continuación especificando la clase de expresiones en las que vamos a concentrarnos:

(a) **Anáfora en función del contexto en el que opera.** Se divide en *anáfora en contexto situacional*¹ y *anáfora en contexto lingüístico*. En este trabajo describiremos únicamente anáforas en contexto lingüístico, es decir, expresiones anafóricas que remiten a un antecedente enunciado explícitamente.

(b) **Anáfora en función del tipo de elemento anafórico.** Considera *anáfora pronominal* y *anáfora de sintagma nominal*. Rastreamos ambos tipos para determinar cuál de ellos es más frecuente en los CDs y cómo se manifiestan en su estructura.

(c) **Anáfora en función del marco en el que ocurre.** Toma en cuenta *anáfora intraoracional* y *anáfora discursiva*. Para efectos de este trabajo, la *oración* es la estructura sintáctica que va de punto a punto, se trata de una idea con sentido completo que implica, cuando menos, una relación sujeto-predicado. Nos interesa distinguir entre anáforas que actúan a nivel intraoracional porque tanto expresión anafórica como antecedente se ubican en la misma oración, de aquellas que tienen alcances extraoracionales, ya que el antecedente de la anáfora se localiza en otra u otras oraciones.

Una vez expuesta la clasificación básica de anáforas, insistimos en el interés por rastrear relaciones anafóricas cuyos elementos básicos son los siguientes:

(1) **Anáfora pronominal:** Pronombres con valor anafórico, los distintos tipos están supeditados al *paradigma pronominal* que se presenta completo en el Apéndice A.

¹ En el capítulo anterior dijimos que los pronombres y expresiones inmersos en usos y contextos situacionales están más cerca de la *referencia deíctica* y no precisamente *anafórica*, por ello no consideramos anáforas a las expresiones implicadas en un contexto situacional.

- (2) **Anáfora de sintagma nominal:** Sintagmas nominales que hacen referencia a alguna entidad previamente enunciada en el discurso, en principio están inmersos en procesos de repetición y de referencia demostrativa con valor anafórico (Apartado 3, Capítulo II).
- (3) **Antecedente:** Cualquier entidad del discurso que sea referida o reintroducida por una *anáfora*.

1.1.2. Extensión y expansión de contextos definitorios

Extensión y *expansión* son conceptos básicos para la identificación de anáforas en CDs. Por *extensión* entendemos la cantidad de palabras que abarca la definición de un término. En el primer capítulo explicamos que uno de los problemas en los resultados de la extracción automática es que no existe un criterio adecuado para determinar la extensión de los CDs. Se dijo también que éstos despliegan distintos elementos y entonces pueden tener dimensiones variables. Creemos que la descripción de los mecanismos referenciales, especialmente la descripción de anáforas, puede contribuir a concretar este criterio.

Por otro lado, hablaremos de *expansión* siempre que sea necesario ampliar el fragmento textual que contiene la definición de un término, sobre todo cuando dicho fragmento no funciona como unidad discursiva con sentido propio. Para fines de este trabajo, la *expansión* refiere a la pertinencia de acudir al documento de origen del CD con el objetivo de verificar la extensión de una definición. En el ECODE, queremos que la expansión se haga automáticamente y para ello se propondrán patrones recurrentes de relaciones anafóricas que sirvan de indicadores para expandir cierto tipo de contextos.

1.1.3. Candidatos completos y candidatos incompletos

Atendiendo a las características de los candidatos a CDs que forman parte de los resultados de la extracción automática (Apartado 2 y 3, Capítulo I), retomamos la diferencia entre *candidatos completos* (CC) y *candidatos incompletos* (CI). Los CC son aquellos CDs que no requieren ser expandidos porque las relaciones referenciales se resuelven en el fragmento textual dado en la extracción automática. Entendemos por *resolver* el hecho de que se encuentre explícito el *antecedente* de las *anáforas*. Los CI son fragmentos textuales

que requieren expansión puesto que presentan relaciones anafóricas que no pueden resolverse en el texto obtenido de la extracción automática.

Lo primero que se hará es separar CC y CI de un corpus de 250 CDs. Este proceso se realizará de manera manual revisando cada contexto; es decir, sin apoyo de herramientas computacionales. En seguida se rastrearán las expresiones que generan CI y se describirán conforme a los criterios de clasificación expuestos, esto en caso de que se trate de anáforas, ya que podríamos encontrar otra clase de mecanismos de referencia discursiva. Posteriormente se analizarán CC para identificar anáforas y establecer patrones recurrentes en la configuración de los CDs.

1.2. Datos que queremos obtener

En seguida se enlistan una serie de preguntas que trataremos de responder a partir de la identificación de relaciones anafóricas. Dichas preguntas se formularon atendiendo a los problemas en la extracción automática; la información que logremos responder se dará como parte de los resultados y de las conclusiones de este trabajo:

- De un corpus de 250 CDs ¿Cuántos son *candidatos completos* (CC) y cuántos son *candidatos incompletos* (CI)? ¿Las cifras resultantes son una muestra representativa de lo que ocurre en la extracción automática?
- ¿Qué tipo de estructuras y mecanismos referenciales se manifiestan en los CI? ¿Son *anáforas*?
- Si los CC son buenos candidatos ¿Cuál es la pertinencia de describir las relaciones referenciales que presentan?
- ¿Qué elementos de un CDs (*término, PVD, definición, PP, etc.*) se ven mayormente afectados por una relación anafórica?
- ¿En qué medida la cohesión y coherencia textuales de los CDs están determinadas por las relaciones anafóricas?
- A partir de los resultados de la identificación de anáforas ¿Se pueden establecer patrones recurrentes de relaciones anafóricas en CDs?

Una vez explicada la forma en que se trabajará con el corpus y habiendo definido el tipo de fenómeno que nos interesa rastrear, presentamos los resultados de esta investigación.

2. Resultados de la identificación de anáforas en contextos definitorios

Los resultados generales de este trabajo se presentan en el siguiente orden:

- Conteo de *candidatos completos* y de *candidatos incompletos*.
- Tipología de expresiones (anáforas) presentes en candidatos incompletos.
- Diferencias entre pronombres y sintagmas nominales como expresiones anafóricas.
- Anáforas en candidatos completos.
- Elementos del CD implicados en una relación anafórica.
- Antecedente.
- Patrones recurrentes de anáforas en contextos definitorios.
- Observaciones finales.

2.1. Conteo de candidatos completos y de candidatos incompletos

De los 250 CDs elegidos al azar del repositorio construido con los primeros resultados de la extracción automática, 189 son *candidatos completos* (CC) y 61 son *candidatos incompletos* (CI). En la siguiente tabla se muestran los datos que corresponden al número de CC y de CI extraídos tanto del *Corpus Lingüístico de Ingeniería* como del *Corpus Técnico del IULA*, también se presentan las cifras y porcentajes totales:

Tabla 3-1. Candidatos completos e incompletos en los corpus de origen

	CLI		CT del IULA		CORPUS TOTAL	
	CD	%	CD	%	CD	%
CC	122	81.3	67	67	189	75.6
CI	28	18.6	33	33	61	24.4
Total	150	100	100	100	250	100

El 75.6 % de contextos totales son CC y no necesitan expansión, por lo cual las relaciones de anáfora se resuelven dentro del fragmento textual. Habrá que ver de qué tipo

son dichas relaciones y qué expresiones y elementos de los CDs están implicados. Por otro lado, los CI constituyen el 24.4%, del total de contextos, un número menor al de los CC.

Los porcentajes presentados pueden ser un parámetro para describir lo que sucede a nivel macro en la extracción automática, estaríamos hablando de aproximadamente un 30% de CI, es decir, contextos que necesitan ser manipulados y completados para integrarse a los resultados finales del proyecto. En este sentido, se plantea que este 30% puede disminuir si en la extracción automática se considera expandir el contexto hasta localizar el antecedente de las anáforas que están próximas al *patrón verbal definitorio*, siendo que éste es el principal patrón de búsqueda de los CDs. Veremos si esto se refleja en los datos obtenidos del corpus usado en este trabajo.

2.2. Tipología de expresiones (anáforas) presentes en candidatos incompletos

Hemos insistido en que en la extracción automática el patrón de búsqueda es el *patrón verbal definitorio* (PVD), esto quiere decir que a partir de dicho elemento la herramienta ECODE localiza CDs. Los patrones definitorios que aparecieron en el corpus fueron:

Tabla 3-2. Patrones verbales definitorios en candidatos incompletos

PVD	CDs	PVD	CDs
caracterizar	2	denominar	2
comprender	6	describir	2
concebir	1	entender	3
conocer	13	llamar	1
considerar	2	permitir	2
consistir	16	utilizar	1
constar	7	ser	3
TOTAL= 61 CDs y 14 patrones verbales			

Lo primero que se hizo fue identificar la clase de expresiones que encabezan CI para determinar sus características y plantear una clasificación. En los 61 candidatos analizados encontramos expresiones que podemos agrupar en cuatro tipos principales. Se enlistan a continuación y en seguida se describen.

(a) Pronombres

- ⇒ (también *se le* ha **denominado**: modelo neoliberal, ortodoxo o monetarista , el mayor ejemplo de ortodoxia de los últimos treinta años , etc.)
- ⇒ *El primero* **consiste en** un vástago dentro de un solenoide o una pieza magnética atraída por un electroimán.
- ⇒ *Eso*, de reglamento, **consiste en** el aumento de la malla (espesamiento) y en la adopción de las consideradas rampas de dispersión [...]
- ⇒ *Uno de ellos* **consiste en** mallar tridimensionalmente el fluido y resolver en él la ecuación de Laplace [...]
- ⇒ *, la cual* **consiste** en un transporte de masa diferente .
- ⇒ *Esto* **puede ser** sumariamente **descrito** de la siguiente forma [...]
- ⇒ *Lo* **describirá** del siguiente modo : « Para formar una idea de si pueden los enemigos de la Corona , introducirse por el camino proyectado [...]
- ⇒ *Este* **consta de** un banco de capacitores sumergidos en aceite en un recipiente de porcelana y conectados en serie [...]
- ⇒ *que* **consta de** ocho subunidades idénticas y capaz de oxidar numerosos alcoholes además del metanol
- ⇒ *la cual* **consta de** series infinitas que contienen derivadas parciales del potencial [...]
- ⇒ *lo que* **se conoce como** condición de radiación.
- ⇒ *lo que* **permite** determinar las características anteriormente reseñadas ante la acción de las olas [...]
- ⇒ *Esta* es una operación de mecanizado *que* **consiste en** torneear una o numerosas ranuras helicoidales perfiladas sobre una superficie cilíndrica [...]
- ⇒ *Esta* es una forma de recombinación homóloga *que* **consiste en** el alineamiento incorrecto entre dos moléculas de DNA bicatenario [...]
- ⇒ *Esto* es *lo que* **se entiende por** enfoque genético de la medicina o «medicina genética».
- ⇒ *Esto* **se conoce como** respiración bronquial.
- ⇒ *el cual* **se considerará como** urbano si transcurre íntegramente por dicha área ", en cuyo caso dicha Entidad " ejercerá las competencias de ordenación y gestión de los servicios de transporte público urbano [...]
- ⇒ *de los cuales los dos primeros* **se consideran** netamente **como** microbiosy los dos últimos , como parásitos.
- ⇒ *Esta* **puede considerarse como** [...]
- ⇒ *Lo* **constituye** un conductor [...]

En este primer grupo vemos que las expresiones que encabezan CI están identificadas con cierto tipo de pronombres. De acuerdo con el paradigma pronominal expuesto en el Apéndice A, encontramos únicamente *personales* (lo, le), *demostrativos* (esto, aquellos, esta), *relativos* (la cual, lo cual, que, de los cuales) e *impersonales* (el primero). Es un número reducido de ejemplos, pero las ocurrencias nos revelan que esta clase de anáforas

pronominales son las principales disparadoras de candidatos incompletos. Tendremos que cotejarlos con las ocurrencias de pronombres en candidatos completos.

En todos los casos se trata de *anáforas en contexto lingüístico*: la expresión apunta a un antecedente omitido en la extracción automática. Podemos observar algunas anáforas discursivas, esto lo determinamos porque los candidatos comienzan con letra mayúscula y la expresión anafórica pronominal encabeza la oración, en consecuencia el antecedente está en alguna oración previa.

En los casos en que el CI comienza con minúscula sólo acudiendo al texto original podemos determinar si se trata de anáforas oracionales o son discursivas. Se expandieron manualmente los CI para localizar el antecedente del pronombre; en seguida damos un ejemplo de este trabajo:

Tabla 3-3. Ejemplos de expansión de candidatos incompletos

Candidato incompleto
(también se le ha denominado : modelo neoliberal, ortodoxo o monetarista, el mayor ejemplo de ortodoxia de los últimos treinta años, etc.)
Expansión 1
Esta economía se rige por un modelo caracterizado por el libre juego de bienes y servicios en un marco en que los factores de producción operan también con amplia libertad teniendo al mercado como su principal asignador de recursos (<i>también se le ha denominado: modelo neoliberal, ortodoxo o monetarista, el mayor ejemplo de ortodoxia de los últimos treinta años, etc.</i>)
Expansión 2
En Chile: Mercado Abierto y Estado Replegado (¿Ausente?) Chile hoy se destaca, y se le reconoce en el ámbito mundial, por ser una de las economías más abiertas del planeta y por exhibir altas tasas de crecimiento económico(2). <i>Esta economía se rige por un modelo caracterizado por el libre juego de bienes y servicios en un marco en que los factores de producción operan también con amplia libertad teniendo al mercado como su principal asignador de recursos (también se le ha denominado: modelo neoliberal, ortodoxo o monetarista, el mayor ejemplo de ortodoxia de los últimos treinta años, etc.)</i>

Para localizar el antecedente del ejemplo presentado, fue necesario ampliar el contexto, sin embargo en la primera expansión localizamos la expresión *Esta economía*; se trata de otra anáfora que hace referencia a un parte anterior del texto. Se tuvo que hacer una

segunda expansión para localizar el antecedente, en este caso, los términos *Mercado Abierto* y *Estado Replegado*.

La expansión se hizo únicamente con los candidatos incompletos del CLI, un corpus de libre acceso, contrariamente al *Corpus Técnico del IULA* que restringe las consultas. El trabajo de expansión sirvió para localizar antecedentes de pronombres y otras expresiones anafóricas. Cabe anotar que no siempre es suficiente expandir una oración, ya que el antecedente puede ser referido más de una vez por diferentes expresiones anafóricas en distintas oraciones. Así entonces, las relaciones anafóricas pueden involucrar más de dos elementos cuando una misma entidad aparece múltiples veces enunciada por otras que construyen *cadena de referencia* [Corblin, 1995].

(b) Sintagmas nominales con determinante demostrativo

- ⇒ *Este análisis* **consiste en** el cálculo de las condiciones de ma marítimo (vientos y oleaje) que se generaron por el paso de Isidoro.
- ⇒ *Este modelo* **se conoce** como modelo potencial porque con él el campo de velocidades se puede considerar como el gradiente de una función potencial [...]
- ⇒ *Este tipo de conexión* **se utiliza en** capacidades no mayores de 10 MVA debido a que en casos de cortocircuito las corrientes son elevadas [...]
- ⇒ Un aparato *de este tipo* **comprende**: una serie de captos planos como los que hemos visto un depósito de almacenamiento del agua caliente [...]
- ⇒ *Este caudal* **comprende** el agua residual doméstica y el agua de infiltración, pero excluye el agua residual industrial.
- ⇒ *Este término* **comprende** asimismo los aparatos que se deslizan sobre un colchón de aire, los artefactos flotantes [...]
- ⇒ *Estos fármacos* **comprenden** la noradrenalina, adrenalina y un gran número de fármacos de acción prolongada [...]
- ⇒ *Estas potencialidades* **comprenden** prácticamente toda la organización de la ascidia, excepto la epidermis, e incluso no es seguro que exista esta limitación [...]
- ⇒ *Estos genes* **comprenden** tanto aquellos con incremento de su función, los oncogenes, como aquellos que pierden su capacidad reguladora [...]
- ⇒ *Aquel sistema* genera una pauta de distribución que **consiste en** manchas discontinuas (la presa) encerradas por un retículo [...]
- ⇒ *Este mecanismo* **se conoce con** el nombre de « Sistema Definitivo » y tiene vigencia durante el quinquenio 198791.
- ⇒ *Este tipo de relación* **se conoce** matemáticamente **como** una relación negativa.
- ⇒ *Este modelo* **se conoce con** el nombre de ATDL [...]

- ⇒ *Este procedimiento se conoce como* punción lumbar, y se usa para extraer o tomar una muestra de líquido cefalorraquídeo o para introducir anestesia local en el espacio extradural o en el saco dural [...]
- ⇒ *Este fenómeno se conoce como* « contragolpe » y su explicación es la siguiente: [...]
- ⇒ *Esta relación se conoce como* Ley de Einthoven.
- ⇒ En su configuración actual, un ordenador *de este tipo consta de* un pequeño PC y una batería [...]
- ⇒ El filtro típico *de este tipo consta de* la ruta a la red de destino y la interfaz a través de la cual se envían las notificaciones de la ruta filtradas.
- ⇒ *Estos elementos son* parte constitutiva de los compuestos que forman la base material de la vida [...]
- ⇒ La mayoría *de estos servidores son* sistemas Compaq Proliant 2000 y 4000 con procesadores 486 o Pentium.
- ⇒ *Este modelo se conoce* como modelo potencial [...]
- ⇒ **Se utiliza esta conexión** [...]

En todos los casos se trata de sintagmas nominales con determinante demostrativo, podemos anotar que son expresiones con valor anafórico porque refieren a una parte anterior del texto [Eguren, 1999: 943]. Las ocurrencias corresponden a *anáforas discursivas* que apuntan a un antecedente en otra oración y lo vemos porque la expresión comienza con mayúscula. Esta característica es importante porque los sintagmas anafóricos demostrativos que identificamos en CI, acuden a un antecedente en alguna oración anterior. Por ello, en la expansión de CDs deberá ampliarse el contexto cuando menos una oración previa.

Conforme a los ejemplos podemos determinar cinco subtipos de sintagmas anafóricos demostrativos que generan CI:

- (1) SN demostrativo (Demostrativo + Núcleo nominal) p.e. *este análisis*.
- (2) SN demostrativo + SN preposicional p.e. *esta energía de arranque*.
- (3) SN preposicional-demostrativo p.e. la mayoría *de estos servidores* / el motor *de estas turbinas*.
- (4) SN + *de este tipo* / *de esta clase* p.e. El filtro típico *de este tipo*.
- (5) *este tipo de* / *esta clase de* + SN p.e. *Este tipo de tumor*.

En seguida mostramos la expansión de uno de los ejemplos expuestos:

Tabla 3-4. Ejemplo de expansión de candidato incompleto

Candidato incompleto
Se utiliza esta conexión en los sistemas de transmisión de las subestaciones receptoras cuya función es reducir voltajes.
Expansión
Conexión estrella -delta Se utiliza esta conexión en los sistemas de transmisión de las subestaciones receptoras cuya función es reducir voltajes.

En el contexto (Tabla 3-4) vemos que la expansión remite a un título en el que se encuentra el antecedente omitido en la extracción automática. No es necesario recuperar otra parte del texto puesto que la referencia discursiva se completa con una primera expansión. En este sentido, podemos prescribir que si en la expansión de un CD, localizamos una nueva expresión anafórica encabezando la oración, será necesario hacer una segunda expansión.

(c) Expresiones mixtas (pronombres y sintagmas nominales con demostrativo): cadenas de anáforas

- ⇒ Moses desarrolló, en 1969-1970, *este método al que llamó* « Tabulation Prediction Scheme », y que **consiste en** desarrollar un conjunto ordenado de combinaciones de variables meteorológicas relevantes [...]
- ⇒ *Esta concepción es lo que se conoce con* el nombre de materialismo histórico.
- ⇒ *Este tipo de tumor se le conoce también con* el nombre microglioma o microgliomatosis múltiple [...]
- ⇒ *Este modelo se conoce* como modelo potencial porque con él el campo de velocidades **se puede considerar como** el gradiente de una función potencial [...]
- ⇒ por lo que a *esta conexión se le conoce como* bloque o unidad.
- ⇒ *A esta precisión se le denomina* clase de precisión y se selecciona de acuerdo con la siguiente lista: [...]

Los ejemplos, aunque son pocos, muestran que podemos encontrar cadenas de anáforas o anáforas muy próximas unas de las otras. Las *cadenas de referencia* se manifiestan, entonces, con *pronombres* y *sintagmas nominales* que se encuentren en una relación anafórica. Es así que existen sintagmas nominales y pronombres que señalan al mismo *antecedente*, en cuyo caso son parte de una *cadena de referencia anafórica*. Para

efectos de corpus nos interesa identificar la reincidencia de las cadenas que puedan producir CI.

(d) Expresiones ligadas a una entidad previamente enunciada (sintagma nominal sin demostrativo, elipsis, marcadores discursivos)

- ⇒ *El primer grupo* es típico de los buques rápidos y **consiste en** olas de gran periodo , que sufren muy poca dispersión al alejarse del barco [...]
- ⇒ *El método numérico propuesto* **consiste en** transformar el problema de Navier-Stokes en un problema de Stokes solucionando este último mediante el Método del Gradiente Conjugado [...]
- ⇒ Generalmente se prefiere *otra condición* que **consiste en** asumir que la cantidad de movimiento promedio a lo largo de la longitud de la ola es cero [...]
- ⇒ En vista de ello, apareció *un nuevo método llamado* componentes simétricas, en el cual se hacen estos análisis de circuitos en forma sencilla [...]
- ⇒ *;entonces puede ser concebido como:* una categoría científica general de carácter transdisciplinario definida como: " un sistema espacio-temporal, complejo y abierto, que se origina y evoluciona justamente en la interfase naturaleza-sociedad [...]
- ⇒ No se contempla como una mera figura retórica, ni tampoco como una anomalía lingüística *,al contrario se entiende como* un proceso cognitivo que impregna nuestro lenguaje y pensamiento habitual.
- ⇒ Persisten especialmente en aguas profundas con muy poca agitación , y *sus copos son* asiento de actividad bacteriana , todo lo intensa que permite la situación de hambre crónica [...]
- ⇒ (**denominadas** , respectivamente , anestesia epidural y espinal)

Existen, conforme a este último grupo, otro tipo de construcciones que mantienen relación con la parte anterior del documento. Hasta el momento, encontramos sintagmas con determinante posesivo (*su/sus*), sintagmas nominales que guardan una relación, quizá semántica, con una entidad discursiva previa, y algunos marcadores discursivos (*entonces, al contrario*):

- ⇒ *;entonces puede ser concebido como:*
- ⇒ *al contrario se entiende como*
- ⇒ Persisten especialmente en aguas profundas con muy poca agitación , y *sus copos son*

En ciertos contextos no se recuperó la oración completa (fragmento textual de punto a punto) y esto también produjo CI. Para disminuir la cantidad de CI en la extracción automática es preciso:

- (1) Recuperar forzosamente fragmentos textuales de punto a punto, es decir oraciones completas.
- (2) Expandir el contexto cuando el fragmento textual comience con una expresión anafórica, sea pronombre o sintagma anafórico demostrativo.
- (3) Cuando en la primera expansión localizamos una nueva expresión anafórica encabezando el contexto expandido, es pertinente hacer una segunda expansión.

En los primeros acercamientos al corpus se observó que principalmente *pronombres* y sintagmas anafóricos con demostrativo generan CI. Una vez revisados todos los CDs, vemos efectivamente que buena parte de los CI están encabezados, sea por un pronombre, sea por un sintagma nominal con demostrativo; en algunos casos por ambos. Las expresiones identificadas están conectadas con un antecedente previo, esto es importante decirlo porque no se encontró ninguna ocurrencia de *catáfora* que produjera CI.

Conforme al repositorio expuesto, observamos expresiones anafóricas que cercanas al patrón verbal definitorio (PVD) producen CI. A manera de resumen podemos enlistar los siguientes tipos:

- (1) **Anáforas pronominales:** Principalmente *pronombres personales* (tercera persona), demostrativos, relativos e impersonales.
- (2) **Anáforas de sintagma nominal demostrativo:** Sintagmas anafóricos que establecen relación demostrativa con un antecedente a partir del determinante que los antecede. Pueden ser anáforas directas, si repiten núcleo nominal del antecedente; o indirectas si se vinculan con el antecedente a partir de una relación semántica.
- (3) **Expresiones mixtas:** Pronombres y sintagmas nominales que constituyen cadenas de referencia cuando apuntan a un mismo antecedente.
- (4) **Otras expresiones con valor referencial:** Sintagmas nominales que no tienen determinante demostrativo, están ligados a un antecedente por una relación de repetición léxica.

2.3. Diferencias entre pronombres y sintagmas nominales como expresiones anafóricas

Para fines prácticos, en nuestra investigación retomamos la noción de *anáfora* entendida como “la relación de una palabra o frase de cualquier tipo y otra palabra o frase anterior” [Peregrín, 1999:1430]. Bajo esta perspectiva las *anáforas* contemplan cualquier expresión referencial que rastrea un *antecedente* en el discurso, con lo cual el inventario es potencialmente infinito. De ahí que pronombres y sintagmas nominales sean los principales para esta investigación.

En la siguiente tabla se contraponen los rasgos de *pronombres* y de *sintagmas nominales* inmersos en una relación referencial; con el fin de subrayar las diferencias entre estos dos tipos de *expresiones anafóricas*:

Tabla 3-5. Diferencias entre pronombres y sintagmas nominales como expresiones anafóricas

	PRONOMBRES (CONTEXTO LINGÜÍSTICO)	SINTAGMAS NOMINALES
1	Expresión referencial.	Expresión referencial.
2	Sin contenido referencial o léxico.	Contenido referencial o léxico.
3	<i>Referencia anafórica en el discurso</i> : los pronombres no apuntan directamente a un referente extralingüístico y por tanto dependen de otra expresión para denotar significado.	<i>Referencia directa y referencia demostrativa con valor anafórico</i> : los sintagmas nominales señalan a un referente extralingüístico denotando significado independientemente de la relación con otras expresiones, pero también pueden establecer relación demostrativa con otras entidades.
4	<i>Valor anafórico</i> inminente, ya que <i>siempre</i> señalan entidades del discurso.	Su <i>valor anafórico</i> se manifiesta únicamente cuando refieren o reintroducen en el discurso una entidad previamente mencionada.
5	Implicados en el proceso de <i>sustitución</i> (proformas pronominales).	Implicadas en el proceso de <i>repetición</i> .
6	Inventario cerrado (<i>paradigma pronominal</i>).	Inventario abierto que abarca <i>unidades léxicas</i> tales como <i>sintagmas nominales con determinante demostrativo, posesivo, definido, indefinido, cuantificadores</i> , etc.

Observamos en la tabla que los *pronombres* están inmersos en una *referencia anafórica* y necesariamente acuden a determinado elemento del discurso para denotar significado; son anáforas estrictamente hablando, ya que hacen “referencia abreviada” a alguna entidad discursiva [Moreno *et al.* 1999]. Los pronombres, entonces, son elementos que sirven para recuperar y/o reproducir las características morfológicas y léxicas de sus antecedentes.

Contrariamente, los sintagmas nominales a partir de su contenido léxico despliegan *referencia directa*, dirigiéndose a un referente extralingüístico concreto. Así, cuando una entidad discursiva es reintroducida por otra expresión, en este caso un sintagma nominal, se pone en juego el significado (contenido léxico) de los elementos que participan en la relación referencial. En este sentido, los sintagmas nominales no “reproducen” ni “recuperan” como tal el contenido léxico del antecedente, ya que ellos mismos despliegan significado y establecen relación por *repetición* o *referencia demostrativa* y no por *sustitución* como ocurre con ciertos pronombres.

Conforme a los contrastes planteados en la Tabla 3-5, nos interesa subrayar que los *pronombres* en contexto lingüístico participan de una relación referencial en la que constantemente son *elementos anafóricos*. En cambio, los *sintagmas nominales* son expresiones anafóricas cuando apuntan a un antecedente (no siempre lo hacen).

El valor anafórico de los pronombres se constata con mayor precisión que el de los sintagmas nominales porque son formas que siempre remiten a una entidad del discurso. Además, se trata de un inventario cerrado que podemos rastrear con ayuda de herramientas computacionales, característica que permite determinar mejores patrones recurrentes para el ECODE.

Las anáforas de sintagma nominal (directa e indirecta), constituyen un inventario infinito de formas, ya que se dan por repetición léxica total, por meronimia, hiperonimia, sinonimia o referencia demostrativa. Para detectar automáticamente esta clase de estructuras es necesario profundizar en las implicaciones semánticas y relaciones conceptuales que establecen las anáforas con sus antecedentes. Considerando la dificultad para plantear una tipología de anáforas de sintagma nominal, nos concentraremos en este tipo de expresiones especialmente cuando están ligadas al término definido y cuando están próximas al patrón verbal definitorio (patrón de búsqueda) de un CD. Cabe aclarar que los

sintagmas demostrativos son más fáciles de rastrear de manera automática, porque podemos localizar su determinante.

En general, la *anáfora de sintagma nominal* (directa e indirecta) es muy difícil de resolver, existen varios trabajos de PLN concentrados únicamente en este tipo de relación referencial [Charolles, 1999; Lavigne-Tompss y Dubois, 1999; Kleiber, 1999; Olivas, 2006]. Ciertamente, la mayor parte de avances en resolución automática de anáfora están concentrados en la pronominal porque sus características permiten que los métodos utilicen básicamente conocimiento lingüístico (información de tipo morfológica, sintáctica y gramatical). Por el contrario, para resolver las relaciones referenciales entre sintagmas nominales (*anáfora directa e indirecta*) es necesario analizar y modelar la estructura del discurso, acudir a procesos cognitivos sumamente complejos que implican conocimiento de mundo y aspectos pragmático-semánticos del lenguaje [Cornish, 1986; Huang, 2000].

2.4. Anáforas en candidatos completos

En candidatos completos se identificaron anáforas en general. Gran parte de las anáforas pronominales se rastrearon con apoyo de la herramienta “buscar-reemplazar” del programa *Word*. Ello fue posible porque se trata de un inventario específico y algunas estructuras y sus variantes flexivas pudieron localizarse sin revisar uno por uno los CDs. Siendo que el paradigma es muy extenso y existen contrastes sustanciales entre sus miembros, los resultados se presentan conforme al paradigma del Apéndice A de este trabajo.

Las anáforas pronominales conciernen a un tipo de palabra que ha sido objeto de diversos estudios. La complejidad que conlleva la descripción de esta categoría, se constata desde su denominación: “pronombres”, es decir, formas que sustituyen al nombre [Escavy, 1987]. Cabe decir, entonces, que esta clase de expresiones cumplen otras funciones que abarcan más que la sustitución y cuando son sustitutos no siempre sustituyen nombres.

Lo pronombres conforman un paradigma muy heterogéneo; se han hechos muchos estudios que problematizan la inscripción de este tipo de palabras en una misma categoría, principalmente porque sus miembros cumplen funciones sintáctico-semánticas muy distintas. Podemos aducir que el valor anafórico es una de sus características, sin embargo

los pronombres no siempre son anáforas y si lo son, la relación que establecen con sus antecedentes es distinta de acuerdo al tipo.

Para la identificación de anáforas pronominales, en este trabajo, resolvimos emplear el inventario tradicional de pronombres, es decir, la clasificación establecida por La RAE [1989], ya que nos interesa rastrear las formas que poseen valor anafórico y no pretendemos describir y discutir las distintas posturas acerca de la función y características de esta clase de palabras.

Cabe aclarar que existen otras clasificaciones [Escavy, 1987; Martínez y Álvarez, 1989; Eguren, 1999; Picallo y Rigau, 1999; Fernández, 1999; Brucart, 1999] que integran consideraciones distintas a la concepción y descripción que plantea La RAE. Para efectos prácticos de identificación de anáforas bastó con recurrir al paradigma tradicional.

Pronombres personales (NOMINATIVO, ACUSATIVO, DATIVO, Y CASO PREPOSICIONAL).

Se buscaron todas las formas y variantes morfológicas de este paradigma.² En una primera etapa se utilizó la herramienta de *Word* “buscar-reemplazar” para localizarlos. Los resultados no fueron buenos para todo el paradigma porque las formas rastreadas no siempre correspondieron a un pronombre. Los *acusativos* (me, te, lo, la, nos, los, las) y los *dativos* (le, les, nos), por ejemplo, pueden aparecer como formas de otras entidades que no son pronombres, en cuyo caso son sílabas de otras palabras, tales como *caracterizado*, *eficiente*, *globalización*, *plataforma*, *gobiernos*, entre muchas otras.

En general del paradigma de los personales obtuvimos muchísimas palabras que no son pronombres personales, pese a esto pudimos establecer algunos patrones para localizarlos empleando una herramienta como la de *Word*.

Personales “NOMINATIVO” y de “CASO PREPOSICIONAL”. Son entidades muy específicas (*yo, tú, usted, él, ella, ello, nosotros, nosotras, ustedes, ellos, ellas*) que no comparten

² Hemos seguido la nomenclatura de la RAE [1989] que divide a los personales en *nominativos*, *acusativos*, *dativos* y *de caso preposicional*. Dicha nomenclatura viene de los casos latinos y la relación que los pronombres guardan con éstos [Fernández, 1999; Martínez y Álvarez, 1989, I]. Cabe anotar que los *personales de nominativo* también se conocen como *pronombres sujeto*, mientras los de *acusativo* son nombrados como de *objeto directo* y los de *dativo* como de *objeto indirecto*. Fernández [1999: 1229] denomina a los de *caso preposicional* como pronombres *tónicos de objeto*.

forma con otra clase de estructuras, es decir, siempre son pronombres, por lo cual la búsqueda con la herramienta de *Word* arroja muy buenos resultados para este tipo.

La diferencia formal entre *pronombres de caso nominativo* y los de *caso preposicional* es que estos últimos van precedidos por una preposición [Fernández, 1999; Martínez y Álvarez, 1989, I: 33]. Existen además *conmigo*, *contigo*, *consigo* y *sí*. En la tabla siguiente se contabilizan las ocurrencias:

Tabla 3-6. Anáforas de pronombre personal sujeto y de régimen prepositivo

P. P "NOMINATIVO"	P. P. "CASO PREPOSICIONAL"
TOTAL = 3	TOTAL = 8
yo = 0	conmigo= 0
tú= 0	contigo = 0
él = 0	con él = 2
ella = 0	por ella = 1
ello = 2	por ello = 2
nosotros =0	con ellos = 1
ustedes= 0	de ellas= 1 (cada una de)
vosotros = 0	consigo = 0
ellos= 1	sí = 0
ellas= 0	si = 1 (por si mismo)

Únicamente hubo ocurrencias de *personales nominativos* (ello, ellos) y *personales de caso preposicional* (con él, por ello, con ellos, de ellas, sí) de tercera persona. No vamos a mostrar cada contexto pero todas las anáforas hacen referencia a un antecedente explícito en el CD. Este resultado es importante porque podemos mostrar que el carácter anafórico de los personales de tercera persona es prominente. Hay que señalar que la cantidad de pronombres de caso preposicional es mayor.

Nuestro corpus es una pequeña muestra de CDs y como ya se mencionó, los resultados son indicadores de lo que puede ocurrir en una mayor cantidad de contextos. No se espera encontrar pronombres personales de primera y segunda persona (caso nominativo y caso preposicional). Por el tipo de discurso, difícilmente habrá en los CDs formas tales como *yo*, *tú*, *usted*, *ustedes*, *conmigo* y *contigo*, más cercanas a “lo deíctico” que a “lo anafórico” (Apartado 5.3, Capítulo II).

Personales de “ACUSATIVO”. La herramienta de *Word* “buscar- remplazar” nos dio malos resultados para los pronombres personales de acusativo. Esto se debe a que sus formas (*me*,

te, nos, lo, la, los, las) no corresponde únicamente a pronombres, de tal manera que la búsqueda tuvo que efectuarse contexto por contexto. En la búsqueda automática que se llevó a cabo en un principio, se consideró rastrear *formas de acusativo*³ como palabras solas, es decir, indicando en la herramienta el siguiente patrón:

- *espacio me espacio*
- *espacio te espacio*
- *espacio nos espacio*
- *espacio lo espacio*
- *espacio la espacio*
- *espacio los espacio*
- *espacio las espacio*

Con esta especificación los resultados fueron mejores, sin embargo, los *personales de acusativo* en posición preverbal (proclíticos), además, comparten forma con los artículos definidos (*lo, la, los, las*), lo cual provocó un enorme margen de error al hacer las búsquedas automáticas.

Los clíticos en posición posverbal (enclíticos) no pudieron obtenerse con el patrón “*espacio lo espacio*”. Se observó que los clíticos en posición posverbal aparecen con verbos en infinitivo, en gerundio y en imperativos afirmativos [Fernández, 1999: 1261], característica que permitió localizar estas formas empleando los siguientes patrones (por el tipo de texto no se incluyen imperativos):

- **arme / arte / arlo / arla / arlos / arlas**
- **erme / erte / erlo / erla / erlos / erlas**
- **irme / irte / irlo / irla / irlos / irlas**
- **ándome / ándote / ándolo / ándola / ándolos / ándolas**
- **éndome / éndote / éndolo / éndola / éndolos / éndolas**

En la tabla siguiente se presentan las 14 anáforas localizadas en posición de *proclítico* (formas antepuestas al verbo):

³ Atendiendo ciertas cuestiones de nomenclatura, cabe decir que los *personales de objeto directo* son también llamados *pronombres acusativos*, y los *personales de objeto indirecto* se identifican como *dativos*. Tanto los pronombres *acusativos* como los *dativos* son agrupados como “átonos” o “clíticos”; formas que cumplen con la característica de ser inacentuadas y aparecer unidas a un verbo, sea en posición preverbal (proclíticos), sea en posición posverbal (enclíticos). [Fernández, 1999: 1221; Martínez y Álvarez, 1989, I : 33]

Tabla 3-7. Anáforas de pronombre personal de objeto directo (preclítico)

CONTEXTO	ANÁFORA
inicialmente lo podemos	lo
la separación no lo permitía	lo
Como su nombre lo indica	lo
y lo vincula con	la
a la de cualquier tipo	la
Sobre la de reserva	la
es mayor que la de generación	la
en la parte superior, de la de digestión	la
una fuente exterior la transforma	la
productos pesqueros que los tienen	los
respecto a las de profundidad “	las
de las más empleadas	las
Quien las hará	las
se la llama	la

Marcamos con gris las formas *la* y *las* que no están en posición preverbal. Éstas se encuentran anteceditas por una preposición o una conjunción (*a la* de, *sobre la* de, *que la* de, *de la* de, *a las* de), seguidas de una preposición en casi todos los casos y sólo un caso seguido de un adverbio (*de las* más). En estos ejemplos se ve implicado un fenómeno de elipsis nominal, en el cual las formas *lo*, *la*, *los*, *las* pueden considerarse pronombres. Hacemos esta aclaración porque estas formas pueden alternar con artículos definidos, por ejemplo *el de*, *la de*, *los de*, en consecuencia tendríamos que describirlos como determinantes de un núcleo nominal elidido y no como pronombres. No obstante esta observación, en principio, se ha optado marcarlos como pronombres, aunque es necesario estudiar con mayor profundidad esta clase de usos para determinar su valor.

Más adelante veremos que los determinantes *la*, *lo*, *los*, *las* también pueden vincularse con anáforas de pronombre relativo que presentan artículo antepuesto, del tipo *los cuales*, *las que*, *lo que*, etc. En el subapartado que corresponde al uso anafórico de los pronombres relativos precedidos por artículo se describen estas expresiones.

A continuación se expone la tabla con las 8 anáforas de *pronombre personal acusativo* en posición posverbal (enclíticos), siempre vinculados con un verbo en infinitivo o en gerundio:

Tabla 3-8. Anáforas de pronombre personal de objeto directo (posclítico)

Verbo con anáfora	Anáfora	Tiempo verbal
Someter lo	-lo	Infinitivo
Capturar la	-la	Infinitivo
Transformar la	-la	Infinitivo
Enlatar los	-los	Infinitivo
Congelar los	-los	Infinitivo
Clasificar las	-las	Infinitivo
Dejar las	-las	Infinitivo
Considerá ndolo	-lo	Gerundio

Personales de “DATIVO”. Conforme al paradigma se buscaron las formas *me, te, le (se), nos, les*, tanto en posición preverbal (*proclítico*) como posverbal (*enclítico*); los resultados fueron los siguientes:

Tabla 3-9. Anáfora de pronombre personal de objeto indirecto

ANÁFORA EN POSICIÓN PREVERBAL	Anáf.
A lo que le llaman algunos	le
se les conoce como	les
Pues nos informan del enlace	nos
Dispositivo que nos permite proteger	nos
ANÁFORA EN POSICIÓN POSVERBAL	Anáf.
Dá ndole	-le

Para el inventario de los personales de dativo, los resultados fueron más precisos en la búsqueda automática. Esto se debe a que la forma de esta clase de palabras no alterna con otras categorías. Nótese que tres de las ocurrencias son anáforas vinculadas a un patrón verbal definitorio: *permitir, llamar y conocer*.

Pronombre “SE”. Para la RAE [1989] el pronombre *se* está incluido dentro de los personales y caracterizado como *pronombre reflejo*, se trata de una de las formas más complejas de describir para determinar su valor y función como pronombre porque comparte forma con fenómenos que no son pronombres, por ejemplo construcciones con valor recíproco [Peregrín, 1999:1430], impersonales [Fernández-Ordóñez, 1999] y voz media [Maldonado, 1992], por mencionar los más importantes. En este trabajo consideramos pronombre *se* únicamente a las formas que cumplen las características

planteadas por la REA [1989] para los reflejos, también denominados reflexivos [Martínez y Álvarez, 1989: 55; Peregrín, 1999: 1459-1472].

Se hizo una búsqueda con la herramienta de *Word* para sondear la frecuencia de esta forma (en posición de preclítico y posclítico), sin embargo la inmensa mayoría no son pronombres. En este sentido, quedan muchos estudios por hacer para precisar los parámetros para determinar el valor y función del *se*.

De las 147 ocurrencias, sólo podemos decir que una presenta el valor de anáfora como pronombre reflexivo; el resto son formas ligadas al patrón verbal definitorio (*se define, se conoce, se concibe, etc.*) y a otra clase de manifestaciones. En seguida exponemos el CD que corresponde al único caso de *se*:

Tabla 3-10. Anáfora de pronombre “se” reflejo

La respuesta de Babinski, que consiste en la extensión del dedo gordo del pie con separación del resto de los dedos en respuesta a la estimulación de la cara lateral de la planta del pie, <i>puede despertarse</i> a menudo en los neonatos ; es consecuencia del escaso control cortical de la función motora, propio del encéfalo inmaduro.

Pronombres posesivos. Los posesivos mantienen relación de pertenencia respecto a alguna expresión del discurso [Álvarez, 1989: 136]. Pueden o no llevar artículo (*el mío, los tuyos, la nuestra, las tuyas, etc.*) y tal y como sucede con el resto de los pronombres, los posesivos pueden o no ser anáforas. Cuando apuntan directamente a un referente extralingüístico son *deícticos*, pero si refieren a una entidad del discurso son *anáforas*.

Existe una gran polémica en torno al concepto “posesión” y en consecuencia en torno a los pronombres relacionados con esta propiedad [Martínez y Álvarez, 1989; Repiso, 1989]. Los posesivos, al igual que los demostrativos, están estrechamente vinculados con el grupo de los adjetivos porque pueden ser determinantes de un nombre o de un sintagma, en cuyo caso se habla de *formas inacentuadas* o *átonas* (*mi, tu, su, mis, tus, sus*); son entidades que aparecen siempre antepuestas a un nombre, y constituyen *sintagmas nominales posesivos*: *su potencia, sus mecanismos, su energía, etc.* [Martínez y Álvarez, 1989: 143]. La RAE los agrupa como pronombres con valor adjetivo, sin embargo su carácter de pronombres ha sido ampliamente cuestionado [Eguren, 1999; Picallo y Rigau, 1999: 975; Repiso, 1989: 4-14]

La RAE [1989] habla de *formas acentuadas* o *tónicas* (*mío, tuyo, suyo, mía, tuya, suya, míos, tuyos, suyos, mías, tuyas, suyas, nuestro, nuestra, nuestros, nuestras*) que generalmente se sustantivan y aparecen con un artículo definido que los antecede, aunque también se presentan solos: *la responsabilidad₁ es mía₁*. Se observó que el grupo de los *posesivos tónicos* está fuertemente relacionado con el *valor anafórico* en tanto que no sólo señala a una entidad lingüística, también la “sustituye”.

En el caso de este paradigma no se localizó ninguna anáfora de las *formas acentuadas* o *tónicas* que fueron expuestas (*mío, tuyo, suyo, mía, tuya, suya, míos, tuyos, suyos, mías, tuyas, suyas, nuestro, nuestra, nuestros, nuestras*). Contrariamente se localizaron algunos casos de las denominadas *formas inacentuadas* o *átonas*, entidades que aparecen siempre antepuestas a un nombre, sin artículo, ni variación de género (*mi, tu, su, mis, tus, sus*) [Martínez y Álvarez, 1989: 143]. La herramienta de *Word* fue muy productiva porque la forma corresponde únicamente a posesivos y se localizan sin interferencia de otras palabras. Se encontraron 29 casos con *su* y 9 con *sus*, en total 37 ocurrencias, exponemos un ejemplo para explicar la clase de fenómeno:

Tabla 3-11. Relación: antecedente-sintagma nominal con posesivo

Se conoce como transformador de corriente a aquél cuya función principal es cambiar el valor de la corriente de uno más o menos elevado a otro con el cual se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc. Su construcción es semejante a la de cualquier tipo de transformador, ya que fundamentalmente consiste de un devanado primario y un devanado secundario.
Término = transformador de corriente
PVD = Se conoce como
SN con posesivo = su construcción
Antecedente del SN con posesivo = transformador de corriente (término)

Con el ejemplo vemos que la relación referencial entre el sintagmas posesivo y el elemento del CD con el que se vincula (término), amplía el significado. En ciertos contextos encontramos esta propiedad, es decir, un sintagma posesivo (*su + nombre / sus + nombre*) cuya característica es agregar información del término definido.

Es necesario estudiar y describir con mayor precisión los sintagmas posesivos porque en la extracción automática identificamos que esta clase de expresiones pueden funcionar como términos: *su estructura se caracteriza por, sus copos son, su impedancia se define*

como, su valor se *concibe como*, por mencionar algunas. Además, encontramos casos tales como *su función es permitir*, *como su nombre lo indica*, *en el desarrollo de sus funciones*, entre otras.

En este trabajo no clasificamos los sintagmas posesivos dentro de las anáforas, no obstante reconocemos su valor referencial discursivo y la importancia de incluirlos como parte de la estructura de los CDs.

Pronombres demostrativos. Los demostrativos sirven para apuntar una entidad discursiva sin nombrarla, poseen un claro carácter anafórico pero también deíctico en tanto que sirven para “señalar” en contexto situacional: *prefiero aquéllos/ pásame ese / tomen estos* [Egueren, 1990: 941].

Los demostrativos también pueden funcionar como determinantes de un nombre, por lo cual han sido considerados parte del grupo de los adjetivos, pues además de apuntar a una entidad sea del mundo, sea del discurso, sirven para acompañar sustantivos como adyacentes nominales: *este proceso*, *aquel transformador*, *esos resultados*, *aquellos circuitos* [Martínez y Álvarez, 1989: 107]. En usos como éstos, los demostrativos abandonan su calidad estricta de pronombre y se convierten en determinantes de sustantivos y sintagmas nominales con *valor anafórico* o *deíctico*, según exista o no un antecedente lingüístico [Martínez y Álvarez, 1989: 113; Sarmiento y Esparza, 1993: 97].

Los demostrativos muestran concordancia de género y número con la entidad a la cual refieren o sustituyen; igualmente presentan concordancia cuando en calidad de determinantes acompañan a un sustantivo.

Para este paradigma, la herramienta de *Word* fue muy productiva, dado que las formas (*éste, ése, aquél, ésta, ésa, aquélla, éstos, ésos, aquéllos, éstas, ésas, aquéllas, esto, eso, aquello*) únicamente corresponden a pronombres. En los resultados, no obstante, se refleja un problema, muchas veces aunque se trate de un pronombre que debiera ir acentuado conforme a una regla ortográfica en desuso, encontramos casos en los que se omite el acento y entonces puede tratarse de un pronombre o del determinante de un sintagma demostrativo:

Tabla 3-12. Anáfora de pronombre demostrativo (sin acento ortográfico)

<p>Conexión delta abierta –delta abierta₁ Esta₁ puede considerarse como una conexión de emergencia en transformadores trifásicos, ya que si en un transformador se quema o sufre una avería cualquiera de sus fases, se puede seguir alimentando carga trifásica operando el transformador a dos fases, sólo que su capacidad disminuye a un 58.89 aproximadamente</p>
<p>Antecedente= Conexión de delta abierta –delta abierta (término)</p>
<p>Anáfora= Esta (pronombre demostrativo sin acento)</p>

Conforme al ejemplo vemos que en la búsqueda automática de pronombres demostrativos es preciso tomar en cuenta los errores ortográficos, de tal manera que para mejorar los resultados deberán considerarse las palabras que los preceden y anteceden. En el ejemplo, la anáfora está vinculada con el término definido y está muy próxima del patrón de búsqueda, es decir, el patrón verbal definitorio (Término + anáfora de demostrativo + PVD + definición).

Se localizaron tres candidatos completos con anáfora de pronombre demostrativo, el expuesto en el ejemplo anterior y dos más:

Tabla 3-13. Anáforas de pronombre demostrativo

<p>Transformadores de corriente Se conoce como transformador de corriente₁ a aquél₁ cuya función principal es cambiar el valor de la corriente de uno más o menos elevado a otro con el cual se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc. su construcción es semejante a la de cualquier tipo de transformador, ya que fundamentalmente consiste de un devanado primario y un devanado secundario .</p>
<p>Antecedente= transformador de corriente (término)</p>
<p>Anáfora= aquél (pronombre demostrativo)</p>
<p>Patrón= PVD + Término + Anáfora + Definición</p>
<p>Cuando se agotan las reservas de compuestos ricos en enlaces fosfato de alta energía (por ej., ATP o fosfocreatina) de un músculo₁, éste₁ se vuelve rígido y ya no puede ser estirado pasivamente mediante fuerzas pequeñas ; esta condición se conoce como rigor</p>
<p>Antecedente= un músculo (parte de la definición)</p>
<p>Anáfora= éste (pronombre demostrativo)</p>
<p>Patrón= Definición (Anáfora) + SN con demostrativo + PVD + Término</p>

Las ocurrencias de pronombres demostrativos en candidatos completos fueron mínimas, sin embargo no debemos perder de vista que, por su forma, son anáforas que

pueden ser fácilmente detectadas de manera automática con herramientas computacionales. Ello es de gran utilidad para mejorar patrones de búsqueda de CDs. No olvidemos que en la descripción de candidatos incompletos hubo ocurrencias de esta clase de anáforas y la gran mayoría se omitió el acento.

Pronombres relativos. Los relativos reproducen y refieren dentro de una oración los contenidos morfológicos y léxicos de otra expresión (generalmente anterior: *antecedente*). Los relativos desempeñan un doble función, la de conjunción o nexos en la subordinación sintáctica y la de *pronombre*, porque refieren o sustituyen a alguna entidad discursiva [Fernández, 1973].

Quizá lo más importante por decir sobre este grupo, es que los relativos poseen un notable carácter anafórico y nunca deíctico. Siempre apuntan en una dirección concreta, es decir, refieren especialmente a una entidad del discurso [Fernández, 1973]. Son un paradigma poco simétrico, algunos de sus miembros son totalmente invariables en género y número (*que, como, cuando, donde*); otros sólo lo son de género (*cual, quien*), en oposición de los variables tanto en género como en número (*cuanto, cuyo*). Muchas veces aparecen con artículo antepuesto (*el que, los que, los cuales*) aunque hay algunos que no lo aceptan (*cuanto, cuyo, cuya, como, quien*); cuando aparecen con artículo se sustantivan y “sustituyen” a alguna parte del discurso.

Cabe aclarar que La RAE agrupa *como, cuando, donde*, con los pronombres relativos. Sin embargo, en otros estudios son específicamente considerados *adverbios relativos* cuya función es introducir oraciones [Martínez y Álvarez, 1989; Porto, 1997; Brucart, 1999]. Debido a la gran polémica que hay para inscribirlos dentro del paradigma pronominal, para efectos de este trabajo hemos dejado fuera la descripción de *como, cuando, donde*. La razón principal es que la forma *como* cobra especial importancia en la predicación verbal definitoria de los CDs y desde los primeros estudios, se determinó tratarla como adverbio y no como pronombre.

De este grupo tenemos que destacar al pronombre *cuyo* (con variación de género y número) que es por excelencia la forma pronominal con *valor catafórico* (Apartado 4.1). Esto se debe a que está vinculado obligadamente con una entidad nominal que le sigue [Martínez y Álvarez, 1989]. *Cuyo* y sus respectivas variantes flexivas, se inscriben dentro de los adjetivos, ya que siempre aparecen en la posición de determinantes de un nombre.

Quizá lo más importante es que se trata de un pronombre que además comparte características con los posesivos. Mantiene relación con dos entidades distintas: un *antecedente* que funciona como poseedor y un *subsecuente* que representa la entidad poseída [Brucart, 1999]:

Tabla 3-14. Ejemplo con pronombre “cuyo”

Otro sistema es el llamado " de burbujeo "que **consiste en** insuflar aire a través de un tubo que llega hasta *el fondo del depósito* **cuyo nivel** se quiere medir, de forma que el aire burbujee a través del líquido saliendo por la parte inferior del tubo.⁶⁵

Los relativos son particularmente importantes para esta investigación, ya que su carácter anafórico se manifiesta en el corpus. Además, se trata del paradigma que presentó mayor número de ocurrencias (*que, cual, cuales, como, cuyo, cuya, cuyos, cuyas, quien, quienes, cuanto, cuanta, cuantos, cuantas*). En la tabla siguiente presentamos los resultados que se obtuvieron de la búsqueda de estas formas empleando la herramienta de *Word*:

Tabla 3-15. Búsqueda automática de pronombres relativos

BÚSQUEDA AUTOMÁTICA	OCURRENCIAS
que	189
cual	15
cuales	4
cuyo	4
cuya	7
cuyos	1
quien	2
quienes	2

Conforme a las cifras de la tabla, vemos que la forma *que* aparece un mayor número de veces, se trata de un pronombre relativo que además de tener valor anafórico puede desempeñar distintas funciones, por ejemplo la de nexos oracionales, o bien, puede ser parte de un marcador discursivo, además está vinculado con antecedentes de distinta clase:

Tabla 3-16. Pronombre relativo “que”

PATRÓN	EJEMPLOS	OCURR.
SN + que	La Teoría del Buque que estudia el barco	106
Verbo + que	Recordando que el proceso de cavitación	14
Preposición + Artículo definido + que	dos principios fundamentales en los que se basa la operación	17
Preposición + que	aquellos casos en que la tensión	13
Artículo definido + que	es lo que se conoce como protección	10
Adverbio + que	sólo que su capacidad disminuye	3
Adjetivo + que	La teoría de ondas es más simple que las ecuaciones	2
Conjunción + que	a manera de tapete y que tiene en uno o en ambos extremos un tubo	5
Marcadores discursivos	"ya que"	7
	"de tal manera que "	2
	"dado que"	2
	"de forma que"	2
	"tal que"	2

Nos interesa subrayar que este pronombre además de ser muy productivo, cobra especial importancia porque aparece constantemente en la estructura de definiciones analíticas (Apartado 1.2, Capítulo I). Concretamente introduce la *Diferencia específica* y la liga con *el Género próximo*:

Tabla 3-17. Ejemplo de definición analítica

El cloroplasto es un tipo de plastidio <i>que</i> se encuentra en vegetales superiores, algas y algunos protozoos
Término = El cloroplasto
Definición analítica = un tipo de plastidio <i>que</i> se encuentra en vegetales superiores, algas y algunos protozoos
Género próximo = un tipo de plastidio
Anáfora = que (liga el Género próximo y la Diferencia específica)
Diferencia específica = se encuentra en vegetales superiores, algas y algunos protozoos

El resto de ocurrencias de pronombres relativos corresponde a *quien*, *quienes*, formas localizadas con ayuda de la herramienta de *Word* y con resultados muy precisos para este tipo de pronombre:

Tabla 3-18. Pronombres relativos

P. RELATIVO "quien" "quienes"
entre otros quienes
entre otros ; quienes
identificar tareas a realizar y quien las hará
persona natural o jurídica a quien compete

Pronombres indefinidos, cuantitativos y numerales. Se trata quizá del subgrupo de pronombres más numeroso; La RAE [1989] los agrupa como *indefinidos* y *cuantitativos* en un mismo capítulo (2.8) y *numerales* en otro (2.9); los explicaremos entonces conforme a los criterios más tradicionales:

(a) Pronombres indefinidos y cuantitativos: Mantienen diferencias sustanciales con otros pronombres (*personales*, *demostrativos*, *relativos*, *posesivos* e *interrogativos*). Los *indefinidos* y los *cuantitativos* son grupos muy próximos, ya que designan a su referente de un modo indefinido, es decir, señalan entidades cuya determinación no está bien concretada. Los miembros de ambos grupos poseen cierto contenido léxico, la mayoría tiene variante de plural y algunos de femenino (*alguno*, *bastante*, *cualquiera*, *dondequiera*, *ninguno*, *poco*, *nada*, *otro*, *todo*, *uno*, etc.); otros, por ejemplo, son invariables (*algo*, *alguien*). Tienen poca estabilidad sintáctica, ya que pueden funcionar como sustantivos (*la nada*), adjetivos-determinantes (*ninguna persona*) o adverbios (*en la mañana comí poco*).

Remarcando ciertas diferencias dentro de los pronombres incluidos en este grupo podemos decir que los *cuantitativos* designan a los seres o entidades por su número, pero en una forma imprecisa, apuntando a un número indeterminado de algo (*mucho*, *todo*, *nada*, *poco*, *bastante*, *demasiado*, etc.); la RAE [1989] los caracteriza como *indefinidos* pero con la facultad de “cuantificar” en menor o mayor grado. Por otro lado, los *indefinidos* que no son *cuantitativos* incluyen formas tales como: *dondequiera*, *cualquiera*, *demás*, *mismo*, *otro*, *tanto* y sus respectivas variantes morfológicas de género y número, por enunciar los más importantes.

Tanto *indefinidos* como *indefinidos cuantitativos*, igual que los *demonstrativos* y los *posesivos*, están catalogados como entidades con valor de adjetivos cuando aparecen como adyacentes de un nombre. Por un lado, pueden actuar únicamente como pronombres sustantivos (*alguien, nadie, quienquiera, cualquiera*) mientras en otros usos se comportan como determinantes (*un coche, algún motivo, todos los días, mucho dinero, etc.*).

Estos pronombres pueden además aparecer precedidos por *lo* (*lo mucho, lo poco, lo bastante, lo mismo, etc.*) y entonces se trata de formas neutras, caracterizadas porque su antecedente abarca más de un término del contexto, es decir, los antecedentes son oraciones, locuciones o ideas expresadas en un fragmento textual.

(b) Pronombres numerales: Existen por otro lado los denominados *numerales*, caracterizados por apuntar a las entidades por medio del número exacto de aquello a lo cual refieren: *uno, una, cinco, seis, diez, cien, etc.* No siempre los numerales son usados como pronombres, pueden ser determinantes de un nombre o de un sintagma nominal y entonces son adjetivos.

Los numerales también, como los *indefinidos*, poseen cierto contenido léxico, en este caso de naturaleza numérica, lo cual se verifica en el hecho de que pueden escribirse como signos aritméticos: 1, 2, 3, 4, 5 y así sucesivamente [RAE, 1989: 237].

Este grupo se subdivide en *cardinales* (*uno, dos, tres, cuatro, cinco, seis, hasta el infinito*) y *ordinales* (*primero, segundo, tercer/o, quinto, decimocuarto, etc.*) que cuando aparecen precedidos por artículo (sea definido o neutro) están sustantivados (*el primero, los primeros, el octavo, el noventa, lo segundo, lo primero, el dos, el tres etc.*) y por tanto “sustituyen” nombres.

El listado completo de los *indefinidos*, *cuantitativos* y *numerales* se presenta con el resto del paradigma en el Apéndice A de este trabajo. La búsqueda con la herramienta Word se exponen en la siguiente tabla, incluye usos anafóricos de estos pronombres y usos como determinantes de un sintagma nominal:

Tabla 3-19. Indefinidos, cuantitativos y numerales (anáforas y determinantes)

algo, alguien, alguno (s) , alguna (s)= 12	poco, poca , pocos, pocas = 1
1: a lo que le llaman algunos = ANF	1: unos pocos millones= Det
2: en alguna forma = Det	mismo, misma, mismos, mismas = 6
3: se presentan algunos resultados = Det	1: por si mismo = ANF
4: ya que algunos pueden = ANF	2: las formas exteriores del mismo = ANF
5: de algunos productos = Det	3: por las mismas . = ANF
6: a algunos metales = Det	4: transmite la misma información = ANF
7: en el cual algunos alvéolos = Det	5: en un mismo individuo = Det
8: por algún fenómeno = Det	6: la relevancia del mismo = ANF
9: de algún universo = Det	tanto, tanta, tantos, tantas = 4
10: algo así como= ANF	1: tanto bióticos y abióticos
11: incorporar algunas prestaciones = Det	2: tanto a escala local como regional
otro, otra, otros, otras = 19	3: y, por tanto
1: entre otras cosas , = M. Discursivo	4: incluye tanto el nitrógeno orgánico como
2: entre otros quienes = ANF	mucho, mucha, muchos, muchas = 2
3: entre otros = ANF	1: en muchas enfermedades = Det
4: elevado a otro con el = ANF	2: que muchos metales = Det
5: y otro de formas finas = ANF	uno, una, unos, unas = 11
6: se tenga otra de reserva = ANF	1: corriente de uno más o menos elevado = ANF
7: y otros sobre la de reserva = ANF	2: fue uno de los dos = ANF
8: y otras cantidades = Det	3: en uno o más circuitos = ANF
9: produce cambios en otro = ANF	4: uno o en ambos extremos= ANF
10: de un lugar a otro = ANF	5: es uno de los procesos = ANF
11: equipo de otras industrias = Det	6: uno de formas llenas = ANF
12: lo vincula con otros centros poblados = Det	7: cada una de ellas = ANF
13: o arruinar otros . = ANF	8: unos pocos millones = Det
14: Según otros autores = ANF	9: hasta unos 1000 MeV = Det
15: en otros de menor = ANF	10: unas láminas = Det
16: y otros pigmentos = Det	11: consisten en unas estructuras = Det
17: Por otro lado, = M. Discursivo	todo, toda, todos, todas = 9
18: otras posibilidades. = Det	1: todos los bienes = Det
19: Por otra parte, = M. Discursivo	2: todo el dominio = Det
varios, varias = 3	3: a todos aquellos elementos = Det
1: de varios pisos = Det	4: de toda su trayectoria = Det
2: tienen varios generadores = Det	5: con toda la información = Det
3: consiste de varias chapas = Det	6: o todos ellos están = Det
bastante = 1	7: de todas las sentencias = Det
1: es bastante similar = Adv	8: comprende todo tipo de = Det
cualquiera = 2	
1: cualquiera de sus fases	
2: una realidad cualquiera	

En la tabla incluimos los usos anafóricos y también los usos de estas formas como determinantes de sintagmas nominales.

Pronombres interrogativos. Sirven como instrumentos a la función apelativa del lenguaje y apuntan a sus referentes por medio de una pregunta. La RAE [1989] los describe muy cercanos a los relativos, en este sentido hay quienes los incluyen en el mismo grupo, planteando la coexistencia de los subgrupos *relativos átonos* y *tónicos*, los últimos tradicionalmente identificados por la RAE simplemente como *interrogativos* [Martínez y Álvarez, 1989: 186]. La diferencia formal con los relativos es que los interrogativos no se usan con artículo y de manera escrita van acentuados. Todos los relativos, con excepción de *cuyo*, pueden funcionar como interrogativos (*acentuados*).

Otro contraste importante de los *relativos* y los *interrogativos* es que éstos últimos señalan a una entidad que aún no ha aparecido en el discurso, lo cual los provee de valor más bien catafórico (Apartado 4.1, Capítulo II), son entonces *catáforas* cuando hacen referencia a una entidad discursiva que será enunciada después de la referencia. De este paradigma no encontramos ocurrencias en CDs.

Sintagmas nominales anafóricos. Las anáforas de sintagma nominal fueron rastreadas una a una en los candidatos incompletos; sólo consideramos las anáforas que se vinculan con el término definido, porque el inventario es extenso y nos interesan las expresiones que puedan desencadenar CI. Los sintagmas nominales anafóricos pueden tener varios tipos de determinante [Sarmiento y Esparza, 1993:24]. Además, pueden presentar complementos y modificadores, del tipo preposicionales o adjetivos principalmente; despliegan género masculino o femenino, y número singular o plural. Conforme a las ocurrencias en el corpus describimos los componentes básicos de los sintagmas nominales anafóricos [Escadell, 1995]:

- (1) **Núcleo:** Unidad léxica que confiere las propiedades categoriales al sintagma.

- (2) **Determinantes:** Son modificadores que confieren propiedades referenciales al sintagma; en otras palabras son los que convierten a una unidad o secuencia en un constituyente con referencia propia. En el capítulo anterior presentamos una tipología de *expresiones referenciales* clasificadas conforme a su determinante, en ella podemos observar las implicaciones del determinante en la relación referencial con otras expresiones (Apartado 3, Capítulo II). Los determinantes inciden, además, en las

relaciones sintácticas con otras expresiones. En palabras de Escandell [1995], los determinantes “fijan la referencia del sintagma nominal [...] seleccionan e individualizan un subconjunto de entidades denotadas por el nombre y sus modificadores para referirse a él como una entidad [...] indican la relación o no entre el referente y los conocimientos de los hablantes”. Los determinantes anteceden al núcleo nominal y comprenden: *artículos definidos, artículos indefinidos, demostrativos, posesivos, numerales, cuantificadores, distributivos e indefinidos*, un inventario que comparte muchas formas con los pronombres; en consecuencia cuando hicimos búsquedas automáticas de pronombres, muchas veces obtuvimos determinantes (Determinante + Nombre).

(3) **Modificadores.** Dentro de los modificadores se agrupan distintos tipos de estructuras con diferencias categoriales y de comportamiento sintáctico. Su función es añadir significado al núcleo nominal de un sintagma; en el corpus encontramos los tipos siguientes:

- a. *Sintagmas adjetivos.* Principalmente adjetivos o estructuras con las características de adjetivo, su posición dentro del sintagma nominal puede ser variable, esto es, a veces anteceden al núcleo nominal y otras veces lo suceden, p.e. *una circulación meteorológica cerrada*.
- b. *Oración de relativo.* Aparecen siempre introducidas por una anáfora de pronombre relativo (con función de adverbio o adjetivo). A esta clase de estructuras se les ha otorgado valor adjetivo, sin embargo se agrupan aparte porque modifican núcleos nominales y agregan información a partir de una oración, p.e. *una circulación meteorológica cerrada que se propaga sobre aguas tropicales*.
- c. *Sintagmas preposicionales.* Modificadores del nombre introducidos por una preposición. No describiremos sus funciones ni la variedad de esta clase de complementos, únicamente queremos precisar su forma: son estructuras nominales que modifican a una unidad nominal, están precedidas por una

preposición, pueden presentar sintagmas adjetivos, oraciones de relativo y otros sintagmas preposicionales, p.e. la reglamentación del transporte de mercancías peligrosas.

Así pues, los sintagmas anafóricos despliegan estructuras que varían según el determinante y los complementos. En CC se localizaron dos tipos de expresión anafórica que establecen relación con el término, sea reintroduciendo el núcleo nominal de su antecedente, o bien estableciendo referencia demostrativa con valor anafórico:

(1) **Sintagmas anafóricos que reintroducen núcleo nominal (anáfora directa):** La relación anafórica se da por repetición léxica total, ya que se reintroduce el núcleo de una expresión previa (antecedente). La expresión anafórica puede ser:

a. *Sintagma nominal* que comparte núcleo nominal con su antecedente, la expresión anafórica presenta distintos determinantes y puede presentar modificadores diversos, además dicha expresión puede estar implicada en procesos de coordinación oracional (conjunciones “y”, “e”, “o”, “u”), por ejemplo:

Tabla 3-20. Ejemplo de anáfora de sintagma nominal

M. Godron y G. Merriam entre otros quienes consideran a [PVD] la Ecología del Paisaje [término-antecedente] como: "la ecología de los sistemas móviles y heterogéneos, [anáfora] estudiando entonces la influencia de la estructura del paisaje sobre los procesos ecológicos, tanto a escala local como regional" (Merriam, G. 1993).

b. *Sintagma preposicional* que modifica a otro sintagma, en este caso el sintagma preposicional (modificador) comparte núcleo nominal con un antecedente. La relación anafórica se da con el modificador de un sintagma nominal y un antecedente:

Tabla 3-21. Ejemplo de anáfora de sintagma preposicional

Se define como [PVD] tensión de contacto [término-antecedente] al valor de la tensión que se presenta entre las masas metálicas conectadas a tierra y el terreno circunvecino [anáfora], que puede eventualmente, en alguna forma, entrar en contacto con una persona.

(2) **Sintagmas anafórico demostrativos:** Son expresiones que establece una relación anafórica a partir de una referencia demostrativa. Los sintagmas son demostrativos por su determinante, pueden ser nominales o bien preposicionales (modificadores), p.e., *la energía continua de estos transformadores*. En el corpus se localizaron dos ocurrencias principales:

a. *Sintagma demostrativo que reintroduce un núcleo nominal*

Tabla 3-22. Ejemplo de anáfora de sintagma demostrativo

La frontera en mar abierto **es una línea** ficticia que separa una porción del mar del dominio de análisis [definición-antecedente], **esta línea** puede en ocasiones ser recta o bien, puede tomar cualquier forma, esto dependerá del dominio requerido a analizar.

b. *Sintagmas demostrativo que entabla una relación semántica con su antecedente y no recupera un núcleo nominal*

Tabla 3-23. Ejemplo de anáfora de sintagma demostrativo

Cuando se agotan las reservas de compuestos ricos en enlaces fosfato de alta energía (por ej., ATP o fosfocreatina) de un músculo, éste se vuelve rígido y ya no puede ser estirado pasivamente mediante fuerzas pequeñas; **esta condición** se conoce como rigor [término].

Otras estructuras con valor anafórico (Sintagmas distributivos). Por último, encontramos construcciones del tipo *los unos... los otros, ni lo uno ni lo otro, tanto...como, tanto una como otra, tan... como, tanto...cuanto, tal como, tales como, los dos primeros, una de las dos*; locuciones con valor referencial que serán igualmente consideradas en nuestro trabajo y las llamaremos *sintagmas distributivos* [Saiz, 2002]. A continuación un ejemplo:

Tabla 3-24. Ejemplo de anáfora de sintagma distributivo

Existen varios *fenómenos de desastre*₁, *tales como*₁ *temblores*₂, *diluvios*₂, *derrumbes*₂, *tanto unos como otros*₂ son muy comunes en la naturaleza.

Para fines de nuestro trabajo, consideramos sintagmas distributivos a aquellos que presentan las formas: *tanto...como, tal...como, tales como...*, *ni...ni...*, etc. Sin embargo,

aquellos sintagmas distributivos que presenten pronombres indefinidos, cuantitativos o numerales: *tanto el primero como el segundo, tal como el otro, tales como los primeros, los unos y los otros, tanto unos como otros, ni uno ni otro*, serán tratados como anáforas pronominales y no como sintagmas distributivos puesto que no hay un núcleo nominal, en su lugar hay un pronombre y le damos prioridad al uso anafórico pronominal.

Otros mecanismos referenciales en contextos definitorios. Como vimos en el Capítulo II, los mecanismos que contribuyen a la cohesión textual implican procesos de sustitución (*anáforas pronominales*) y de repetición (*anáfora de sintagma nominal*), así como *deixis*, *elipsis* y *marcadores discursivos*. Son fenómenos que brevemente describimos en el capítulo precedente; ahora sólo se darán ejemplos de CDs que dan cuenta de su funcionamiento:

Tabla 3-25. Ejemplo de deixis, elipsis y marcadores discursivos en contextos definitorios

Deixis
En última instancia , el problema conceptual que aquí se plantea es que los puertos difícilmente pueden seguir siendo considerados como una estructura o núcleo de actividades por si mismo,
Elipsis (pronominal)
Las Barras Bravas en el fútbol pueden considerarse a grandes rasgos como grupos que inician los cantos en los estadios, ⊙ portan la mayor cantidad de banderas y ⊙ promueven ataques organizados contra Barras Bravas o hinchas rivales”.
Marcador discursivo
Una computadora procesa informacion y almacena datos, pero su aspecto mas importante es el de la simulacion, lo cual consiste en representar una realidad cualquiera mediante un modelo, es decir, un conjunto de datos organizados de tal manera que constituyen una representacion teorica de la realidad.

En esta investigación, como hemos venido diciendo, nos concentramos en los fenómenos referenciales que involucran *anáforas*. No obstante, es importante señalar que la manifestación de otra clase de referencias discursivas es muy frecuente. En el futuro será fundamental describir y estudiar sus ocurrencias para evaluar la *cohesión* y *coherencia* de los CDs y obtener mejores resultados en la extracción automática.

Elipsis. En el Capítulo II se habló de elipsis como un mecanismo referencial muy próximo a la repetición y a la sustitución. Dicho mecanismo convive con la anáfora en tanto que retoma “en ausencia” algo dicho previamente, en cuyo caso puede tratarse como *anáfora* \odot (marca cero). Tanto *anáfora* como *elipsis* contribuyen a la economía lingüística, la diferencia es que la anáfora es una sustitución y la elipsis una omisión [Moreno *et. al.*1999].

El español es una lengua que por sus características morfológicas recurre frecuentemente a la elipsis pronominal, la más tratada en los trabajos de resolución automática porque hecha mano de la información gramatical. Para efectos de corpus y en este trabajo, la elipsis no se considera como anáfora \odot (cero) sino como otro más de los mecanismos referenciales que participan en la construcción de la *cohesión* textual, por ello no describiremos las ocurrencias que implican elipsis.

La identificación de anáforas en CC se llevó a cabo con la finalidad de rastrear los usos más prominentes, se trata de fragmentos textuales que no requieren ser expandidos pero que internamente también presentan relaciones anafóricas que serán consideradas para plantear patrones recurrentes.

Dimos prioridad a las anáforas pronominales porque se trata de un inventario específico que pudo rastrearse con ayuda de herramientas computacionales. Ello repercute en que para mejorar patrones de búsqueda que incluyan anáforas, será necesario precisar patrones que ayuden a localizarlas.

Cabe señalar que las anáforas de sintagma nominal con determinante *demonstrativo* y aquellas con determinantes *posesivos*, *indefinidos*, *cuantitativos* y *numerales* también pudieron rastrearse buscando pronombres, formas que cuando preceden a un sustantivo, operan como determinantes. En la figura siguiente resumimos la clasificación de anáforas más recurrentes en CC:

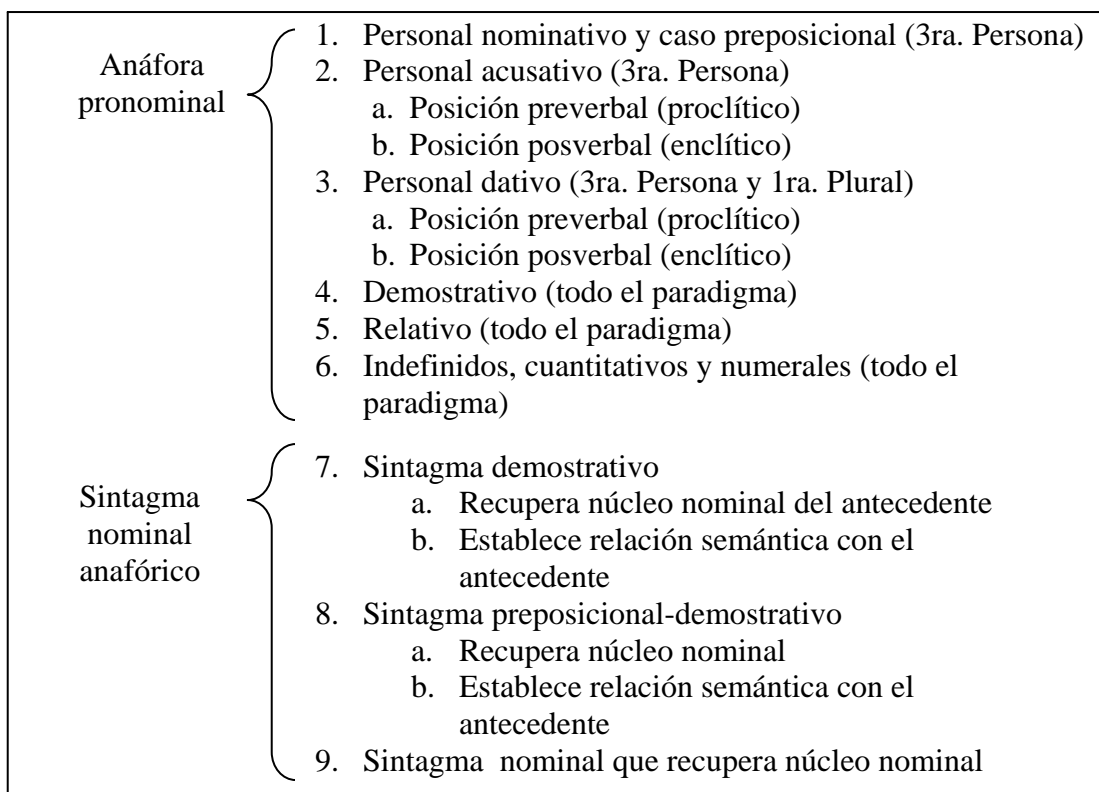


Figura 3-1. Anáforas en candidatos completos

2.5. Elementos de contexto definitorio inmersos en una relación anafórica

Si se atiende a las características de un CD en función de sus elementos, se observa que algunos de ellos son susceptibles a formar parte de una relación referencial discursiva, cuando operan como *anáfora* o como *antecedentes* de una cadena de referencias. A continuación se expone una tabla que muestra cuáles son los elementos de un CD que pueden estar implicados en una relación anafórica, según el tipo de estructura sintáctica que los caracteriza. La tabla se hizo con base en el las ocurrencias rastreadas en el corpus:

Tabla 3-26. Elementos de contexto definitorios como candidatos a una relación referencial

Elementos del CD	Estructuras sintácticas características	Candidato a relación referencial
Término	Unidades léxicas simples o complejas (expresiones referenciales): <i>Sintagmas nominales, verbales, adjetivos y adverbiales.</i>	Sí
Definición	Estructuras sintácticas diversas (generalmente oraciones y fragmentos discursivos pero también puede tratarse de unidades léxicas). La complejidad depende del tipo y la extensión de la definición.	Sí
PVD	<i>Formas verbales simples</i> (verbo definitorio en forma simple) o <i>compuestas</i> (verbo definitorio + auxiliar + clítico <i>se</i> + nexos-adverbios).	Sí
PTD	Signos de puntuación que funcionan como nexos entre el término y la definición. Marcas tipográficas (tipos de letra, itálicas, negritas) que sirven para distinguir el término o parte de la definición. No son estructuras sintácticas.	No
Marcadores reformulativos	Estructuras fijas (lexicalizadas): <i>por ejemplo, es decir, siendo así, estos es</i> etc. Corresponden a algunos marcadores discursivos mencionados en el primer capítulo (Apartado 5.3.4).	No
Patrones pragmáticos	Diversas estructuras sintácticas que abarcan oraciones y unidades léxicas simples o compuestas (fechas, nombres propios, marcadores discursivos)	Sí

El *término*, la *definición*, el *patrón verbal definitorio* (PVD) y algunos *patrones pragmáticos* (excepto marcadores discursivos) son los cuatro elementos de un CD susceptibles a formar parte de una relación referencial, esto está determinado por el tipo de estructuras sintácticas que presentan.

No se espera, por ejemplo, que un *marcador reformulativo* esté en relación anafórica, muchas veces se trata de formas lexicalizadas que son parte de la información contextual del CD y cumplen una función similar a la de las anáforas, conectar las partes del discurso.

El *término* y la *definición*, en cambio, son la información prominente del CD, aparecen una y otra vez en los documentos; por su estructura sintáctico-semántica son candidatos totalmente idóneos para conformar una cadena de referencias. Nos interesa principalmente las anáforas que afectan directamente al *término*, en función de esto se determinarán los patrones recurrentes.

2.6. Antecedente

Los elementos en una relación anafórica son *anáfora* y *antecedente*. En los apartados anteriores expusimos una tipología de anáforas, describimos sus características y las ocurrencias más prominentes sin describir la relación con el *antecedente*, otro elemento constante en las relaciones anafóricas.

A grandes rasgos se trata de entidades discursivas previamente mencionadas. Hasta ahora no hemos descrito ningún tipo de estructura identificada con el elemento *antecedente*. Se trata de cualquier entidad del discurso referida o reintroducida por una *anáfora*. En consecuencia, hablamos de un amplio inventario de estructuras que pueden ir desde unidades léxicas, tales como expresiones referenciales, hasta un fragmento de texto ¿Por qué? Muchas veces la referencia condensa, resume y abarca una idea expresada en una oración o en un fragmento discursivo.

El *antecedente* no siempre corresponde a una expresión referencial particular, puede tratarse de un sintagma nominal con complementos, de uno verbal, de uno adverbial, de una oración o de un fragmento discursivo. En la siguiente figura exponemos la tipología de antecedentes recuperados en el corpus, en seguida se proporcionan ejemplos:

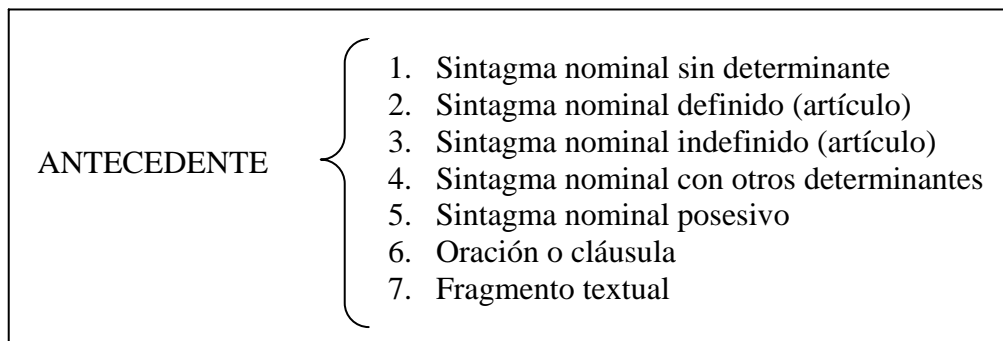


Figura 3-2. Tipología de antecedentes

(1) *Sintagma nominal sin determinante*

Tabla 3-27. Ejemplo de antecedente

Conexión delta abierta -delta abierta Esta puede considerarse como una conexión de emergencia en transformadores trifásicos, ya que si en un transformador se quema o sufre una avería cualquiera de sus fases, se puede seguir alimentando carga trifásica operando el transformador a dos fases, sólo que su capacidad disminuye a un 58.89 aproximadamente .

<p>Antecedente= Conexión delta abierta -delta abierta Anáfora1= Esta Anáfora2= una conexión de emergencia en transformadores trifásicos</p>

(2) *Sintagma nominal definido (artículo)*

Tabla 3-28. Ejemplo de antecedente

<p>La técnica percutánea, en la que se emplea un introductor con la ayuda de una guía, a veces se conoce como técnica de Seldinger.</p>
<p>Antecedente= La técnica percutánea Anáfora 2= en la que Anáfora 3= técnica de Seldinger</p>

(3) *Sintagma nominal indefinido (artículo)*

Tabla 3-29. Ejemplo de antecedente

<p>Un campamento es un satélite residencial de una ciudad multifuncional que atiende una gran parte de las necesidades del campamento y lo vincula con otros centros poblados.</p>
<p>Antecedente= Un campamento Anáfora1= una gran parte de las necesidades del campamento Anáfora 2= lo</p>

(4) *Sintagma nominal con otros determinantes*⁴

Tabla 3-30. Ejemplo de antecedente

<p>El etileno es el principal agente inductor de la maduración de frutas y verduras y puede causar la maduración prematura de algunos productos o arruinar otros.</p>
<p>Antecedente= algunos productos Anáfora= otros</p>

(5) *Sintagma nominal posesivo*

Tabla 3-31. Ejemplo de antecedente

<p>En general, se considera como río principal el que tiene más caudal y su cauce es más ancho, o bien, a veces, el que tiene mayor longitud.</p>
<p>Antecedente= su cauce Anáfora= el que</p>

⁴ Se trata de *pronombres indefinidos, cuantificadores y numerales* (Apéndice B) que cuando preceden un sintagma nominal funcionan como determinantes, son del tipo: *ningún, todos, algún, pocos, muchos*, etc. Son llamados también determinantes cuantitativos [Rigau, 1981].

(6) Oración o cláusula

Tabla 3-32. Ejemplo de antecedente

La energía es materia , y ello permite afirmar que la energía lo es todo, algo así como la esencia última e irreductible del universo.
Antecedente = La energía es materia Anáfora1 = ello Anáfora2 = la energía Anáfora3 = lo

(7) Fragmento textual (mayor a una oración)

Tabla 3-33. Ejemplo de antecedente

El tratamiento térmico de temple es la operación que consiste en calentar el acero hasta la temperatura adecuada, enfriando luego el metal lo más rápido posible en el medio que corresponda al tipo de acero, ello cambia naturalmente la estructura del metal, dándole una condición dura y quebradiza
Antecedente = calentar el acero hasta la temperatura adecuada, enfriando luego el metal lo más rápido posible en el medio que corresponda al tipo de acero Anáfora = ello

El inventario de antecedentes es inmenso, esta tipología se hizo conforme a las ocurrencias que conciernen al término definido o parte importante de la definición. En los resultados no contemplamos la relación antecedente-anáfora describiendo la estructura de ambos. No obstante, notamos que hay cierta relación entre el tipo de anáforas y el tipo de antecedentes que las rigen. Anáforas tales como *por ello, para ello, ello, de ello, lo cual, por lo cual* etc. están vinculadas a antecedentes que abarcan una oración o un fragmento textual. En cambio, anáforas como *la cual, el cuales, aquellos, esos, este*, etc. se identifican con antecedente que son expresiones sintagmáticas con las que establecen concordancia de género y número. Como trabajo a futuro es importante describir el tipo de antecedente en función del tipo de anáfora.

2.7. Patrones recurrentes de anáforas

Siendo que el patrón de búsqueda es el verbo definitorio, los patrones recurrentes de anáforas en CDs se plantearon en función de dicho elemento y del tipo de predicaciones

verbales que aparecieron en el corpus. Los patrones que vamos a presentar consideran las estructuras que pueden generar CI, en cuyo caso es preciso expandir el contexto.

Partimos de tres estructuras básicas de CD atendiendo al orden de sus elementos [Alarcón, 2003; Alarcón 2006]:

- Término + Patrón verbal definitorio + Definición (T + PVD + D)

T *se define como / se entiende como/ se concibe como* **D**

- Definición + Patrón verbal definitorio + Término (D + PVD + T)

D *se define como / es conocido como / es entendido como* **T**

- Patrón verbal definitorio + Término + Definición (PVD + T + D)

Se conoce como / se define como / se entiende por **T D**

Estos patrones de CDs se identifican con ciertos verbos definitorios y con cierta clase de definiciones [Aguilar, 2006]. Quiere decir que el tipo de patrón verbal incide en el tipo de definición y en el orden de los elementos del CD.

Como resultado final de la identificación de anáforas en CDs presentamos patrones recurrentes que pueden mejorar los resultados en la extracción automática. Cuando hablamos de patrones hablamos de estructuras que serán incluidas en la metodología de la extracción automática para que la herramienta considere anáforas como elementos constitutivos de los CDs.

Dadas las características de los pronombres y su recurrencia en el corpus empleado en este trabajo, principalmente pudimos establecer patrones para anáfora pronominal y únicamente consideramos sintagmas anafóricos demostrativos, ya que el determinante es un indicador para anáforas de este tipo.

Por el momento, no consideramos anáforas de sintagma nominal que establecen relación con el antecedente a partir de una relación semántica. Hemos diseñado patrones para expresiones referenciales que pueden localizarse con herramientas computacionales.

Anáfora + PVD + D: En este patrón se enmarcan los CDs que presentan una anáfora en lugar del término definido. Para recuperar el término es necesario expandir el fragmento textual. En la tabla siguiente se exponen patrones que en la extracción automática pueden indicar la necesidad de expandir cuando menos una oración:

Tabla 3-34. Anáfora + Patrón verbal definitorio + Definición

TIPO ANÁFORA	ANÁFORA (NO TÉRMINO)	PVD	DEFINICIÓN
Personal	(y) ello /ellos/ ella / ellas	- está(n) caracterizado-a(s) por	Fragmento textual
Demostrativo	éste / ésta / esto / éstas /	- se caracteriza(n) en (que)	
Sint. Posesivo	su + (nombre) sus + (nombre)	- se ha(n) caracterizado en (que) / por (que)	
Relativo	SN + que	- puede(n) ser concebido-a (s) como	
	lo cual / la cual / los cuales / las cuales	- debe(n) ser concebido-a (s) como	
	cuyo (s) + nombre cuya (s) + nombre	- puede(n) considerarse como	
Indefinido/ cuantificador	algunos / algunas	- se considera(n) como	
	mucho (s) / mucha(s)	- se conoce(n) como	
	ninguno / ninguna	- se define(n) como	
	otro (s)/otra (s)	- es / son definido(s) como	
	pocos / pocas	- puede(n) definirse como	
	todos / todas	- se ha(n) definido como	
	unos /unas	- comprende(n)	
	varios / varias	- debe(n) comprender	
	otros/ otras	- puede(n) comprender	
SN anafórico	Sintagma demostrativo (con determinante <i>este, esta, estos, estas, ese, esos, esas, aquel, aquellos, aquellas, dicho, dicha, dichos, dichas</i>)	- consiste(n) de	
		- debe(n) consistir de/en	
		- puede(n) consistir de/en	
		- es / son descrito(s) como	
		- se describe(n) como	
		- consta(n) de	
		- puede(n) constar de	
		- se entiende(n) como	
		- permite(n)	
		- puede(n) permitir	
- sirve(n) para			
- puede(n) servir para			
- se usa (n) para			
- puede(n) usarse para			
- es/son usado(s) para			
- se utiliza(s) para			
- puede(n) ser utilizado para			
- es /son			

Se incluyeron todos los patrones verbales que aparecieron en el corpus y se colocaron las anáforas que pueden parecer próximas y producir candidatos incompletos. Cabe aclarar que las anáforas pronominales no pueden combinarse con todos los tipos de predicación verbal.

Anáfora + PVD + D: Este patrón posee el mismo orden de elementos que el anterior, la diferencia es que el patrón verbal definitorio está en participio, dicha conjugación no admite el mismo tipo de anáfora que los patrones presentados en la tabla anterior:

Tabla 3-35. Anáfora + Patrón verbal definitorio (participio) + Definición

TIPO ANÁFORA	ANÁFORA (NO TÉRMINO)	PVD	DEFINICIÓN
Demostrativo	aquel, aquello(s), aquella (s)	- caracterizado-a (s) por - conocido-a (s) como - definido-a (s) como - entendido-a (s) como - concebido-a (s) como - considerado-a (s) como	Fragmento textual
Sint. Posesivo	su + (nombre) sus + (nombre)		
Indefinido/ cuantificador	algunos / algunas		
	mucho (s) / mucha(s)		
	ninguno / ninguna		
	otro (s)/otra (s)		
	pocos / pocas		
	todos / todas		
	unos /unas		
	varios / varias		
	otros/ otras		
	el primero el segundo los primeros los segundos		
SN anafórico	Sintagma demostrativo (determinante <i>este, esta, estos, estas, esos, esas aquel, aquellos, aquellas</i>)		

Anáfora + PVD + T: Son CDs que presentan una anáfora en la posición de la definición. Quiere decir que el término está después de la PVD y para localizar la definición completa es necesario recurrir al texto que la precede. Nuevamente nos concentramos en anáforas pronominales y anáforas de sintagma nominal demostrativo:

Tabla 3-36. Anáfora + Patrón verbal definitorio + Término

TIPO DE ANÁFORA	ANÁFORA (NO DEFINICIÓN)	PVD	TÉRMINO
Personal	(y) ello	- se debe concebir como	Unidad léxica
Demostrativo	éste / ésta / esto / éstas /aquello	- se considera(n) como	
Sint. posesivo	su + (nombre) sus + (nombre)	- se conoce(n) como - se define(n) como	
Relativo	SN + que lo cual / la cual / los cuales / las cuales cuyo (s) + (nombre) cuya (s) + (nombre)	- es/son definido(s) como - puede(n) definirse como - se ha(n) definido como - es / son descrito(s) como - se describe(n) como - se entiende(n) como	
Sintagma anafórico	Sintagma demostrativo (<i>este, esta, estos, estas, ese, esos, esas, aquel, aquellos, aquellas, dicho, dicha, dichos, dichas</i>)	- es/son usado(s) para - es /son - es/son denominado, a(s) - es/son llamado, a(s)	

PVD + T + D: Este patrón considera las anáforas que aparecen entre el término y la definición cuando la PVD encabeza el CD.

Tabla 3-37. Patrón verbal definitorio + Término + Definición

PVD	TÉRMINO	NEXO	ANÁFORA CON EL TÉRMINO	DEFINICIÓN
Se concibe (n) como Se define (n) como Se considera(n) como Se conoce (n) como Se entiende (n) como Se denomina(n) Se considera(n) Se define(n)	Unidad léxica	a	el/ la/los/las (que) (que cuyo/cuya /cuyos/cuyas) ese/eso/ esos/esas (que) (cuyo/cuya /cuyos/cuyas) aquel/ aquello / aquella/ aquellas (que) (cuyo/cuya /cuyos/cuyas) todo aquello /toda aquella/ todos aquellos / todas aquellas (que) cualquiera (que) (cuyo/cuya /cuyos/cuyas) el mismo / la misma/ los mismos/ las mismas (que)	Fragmento textual

Anáfora en el Patrón Verbal Definitorio: Incluye anáforas ligadas al patrón verbal definitorio. Son pronombres personales de acusativo y de dativo, de tercera persona proclíticos y enclíticos. Se plantean ocurrencias sólo con los verbos que permiten esta clase de anáforas:

Tabla 3-38. Anáforas en el patrón verbal definitorio (personal acusativo)

PVD	TIPO ANÁFORA	EJEMPLO
caracterizar comprender concebir conocer considerar denominar definir describir llamar utilizar usar	PVD [+Anáfora] Enclítico (posición posverbal) Personal-acusativo (-lo / -la / -los / -las)	caracterizarlo / caracterizándolo
		comprenderlo / comprendiéndolo
		concebirlo / concibiéndolo
		conocerla/ conociéndola
		considérala/ considerándola
		denominarlo / denominándolo
		definirlos / definiéndolos
		describirlos / describiéndolos
		llamarlas / llamándolas
		utilizarlos / utilizándolos
	usarlos / usándolos	
	[Anáfora +] PVD Proclítico (posición posverbal) Personal-acusativo (-lo / -la / -los / -las)	lo caracterizan como
		lo comprenden como
		la conciben como
		las consideran como
		las denominan como
		los definen como
		los describen como

Tabla 3-39. Anáforas en el patrón verbal definitorio (personal dativo)

PVD	TIPO ANÁFORA	EJEMPLO
conocer considerar denominar definir llamar utilizar usar	[Anáfora +] PVD Proclítico (posición posverbal) Se + le + verbo + prep./ como Se + les + verbo + prep./ como	se le conoce como
		se les considera como
		se le denomina
		se le define
		se les llama
		se les utiliza
		se le usa
permitir llamar denominar	PVD [+Anáfora] Enclítico (posición posverbal) Personal-dativo (-le/ -les)	permitirles llamarle denominarles

Los patrones expuestos servirán para indicar aquellos casos en que el candidato debe expandirse en la extracción automática, la expansión deberá hacerse cuando menos una oración, ya que el término o la definición no aparecen. En los patrones únicamente se incluyeron anáforas que guardan relación con el término o con la definición completa.

3. Observaciones finales

A manera de conclusión de este capítulo hacemos una serie de observaciones finales que se lograron de la identificación de anáforas en CDs.

De un total de 250 CDs, 30% son candidatos incompletos (CI) que requieren expansión (manual por el momento) porque presentan una relación referencial incompleta. Esta cifra es representativa porque da muestra de lo que puede suceder en un corpus de mayor número de CDs.

Gran parte de los CI están encabezados por anáforas, principalmente pronombres y sintagmas demostrativos, la mayoría de dichas anáforas son discursivas y por ello el candidato debe expandirse, ya que su antecedente está en alguna oración anterior.

Se rastrearon anáforas en candidatos completos (CC) para identificar las ocurrencias más frecuentes en CDs. Los resultados fueron mejores para los pronombres porque éstos pudieron localizarse con ayuda de una herramienta computacional y su descripción permitió proponer nuevos patrones de búsqueda.

En el corpus destacaron los pronombres relativos por el número de ocurrencias y su inminente valor anafórico, además encontramos frecuentemente sintagmas demostrativos que establecen referencia anafórica con antecedentes que suelen ser el *término* o la *definición*.

También se localizaron anáforas dentro del patrón verbal definitorio, en todos los casos son anáforas pronominales y oracionales, por lo cual su antecedente está en la misma oración; son expresiones que no pueden ignorarse, ya que siempre apuntan al término definido.

En este sentido, la identificación de anáforas en CC nos permitió establecer una tipología de las estructuras que desencadenan relaciones referenciales anafóricas que en la extracción automática pueden producir CI.

Se describieron las diferencias entre pronombres y sintagmas nominales como expresiones anafóricas para puntualizar en las características que deben considerarse en trabajos dirigidos a la localización y resolución automática de anáforas.

Pudimos determinar que, por su estructura, tanto el término como la definiciones son elementos susceptibles a formar parte de una relación anafórica, esto porque se trata de información prominente que se repite una y otra vez en los textos de especialidad. Así pues, la anáfora es un mecanismo que permite reintroducir en el discurso información que es importante destacar.

A partir de la identificación de anáforas pudimos establecer patrones recurrentes de anáforas en CDs, dichos patrones se integrarán a la metodología de la extracción automática para mejorar los resultados y para que el 30% de CI se reduzca. Así pues damos fin a este tercer capítulo y concluimos este trabajo con una propuesta de etiquetado de relaciones anafóricas que se expone en las páginas siguientes.

Capítulo IV *Etiquetado XML para relaciones anafóricas y otros mecanismos de referencia discursiva en contextos definatorios*

Etiquetado XML para relaciones anafóricas y otros mecanismos de referencia discursiva en contextos definitorios

1. Corpus de contextos definitorios y etiquetado XML

A partir de la noción de *contexto definitorio* (CD), el Grupo de Ingeniería Lingüística lleva a cabo la edificación de un *Corpus de Contextos Definitorios* (CCDs)¹ que se construye de los CDs obtenidos de los trabajos en curso de extracción automática. El corpus es una herramienta que tiene como objetivo auxiliar en trabajos de terminología, lexicografía, en la edificación de diccionarios electrónicos onomasiológicos y semasiológicos, en la elaboración de bancos terminológicos y en el diseño de ontologías, entre los más importantes; además de servir como corpus principal para las investigaciones que se llevan a cabo en torno al proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos*.

En una primera etapa, el corpus era sólo un repositorio de CDs, conforme se lograron avances en la identificación y descripción de sus elementos, se formularon herramientas que permiten manipularlos y clasificarlos. Esto quiere decir que, con base en los resultados en la extracción de CDs, se diseñó un sistema de etiquetas en XML² para facilitar la identificación de sus partes. Dicho sistema permite, entre otras cosas, visualizar los CDs en una página WEB y recuperar información puntual como *términos*, *patrones definitorios* o *definiciones*, por dar algunos ejemplos.

¹ Sierra *et. al.*, 2006.

² Viene de las siglas *eXtensible Markup Language* (lenguaje de marcas extensible). Se trata de un conjunto de reglas para definir etiquetas semánticas que sirven en la organización de documentos. XML no es un lenguaje en particular, sino más bien un estándar que permite diseñar lenguajes para fines determinados, por ejemplo, dividir un documento, distinguir sus partes, recuperar cierto tipo de información y visualizarla en una página WEB. En otras palabras XML es un metalenguaje que determina la estructura de un documento con base en etiquetas que no están predefinidas, sino que se diseñan conforme a las necesidades específicas de los usuarios para expresar información estructurada, es decir, información que se compone de partes bien definidas.

Las etiquetas³ sirven para anotar cada una de las partes de los CDs descritas en el Apartado 1.2 del Capítulo I, por tanto, existen marcas para *término*, *definición*, *patrón verbal definitorio*, *marcadores reformulativos*, *patrones tipográficos definitorios* y *patrones pragmáticos*. El sistema de etiquetado también permite especificar características particulares de cada elemento; por ejemplo, para *patrones verbales definitorios* se define el lema del verbo principal, los argumentos sintácticos y la posición del clítico *se*; para la definición se anota uno de los cinco tipos posibles; y cuando existen patrones pragmáticos se determina la clase.

Los elementos del CD están codificados por una etiqueta de apertura y otra de cierre conforme a las reglas de XML. El texto correspondiente se encuentra contenido entre dichas etiquetas. Para agregar los valores o características específicas de cada elemento se utilizan *atributos* indicados entre comillas en la etiqueta de apertura. En la siguiente tabla se exponen las once etiquetas diseñadas para las partes de los CDs, en el Apéndice C de esta tesis se explican con detalle tanto código como etiquetas:

Tabla 4-1. Etiquetas XML para contextos definitorios

CÓDIGOS	ELEMENTO CD	ETIQUETAS
CD	contexto definitorio	<CD num=""> </CD>
TERM	término	<TERM term="" tipterm=""> </TERM>
DEF	definición	<DEF tipdef=""> </DEF>
PVD	patrón verbal definitorio	<PVD> </PVD>
	verbo definitorio	<VD lema="" args="" mdo=""> </VD>
	marca de clítico "se"	<SEmarc tipsemarc=""> </SEmarc>
	verbo auxiliar	<Vaux> </Vaux>
	nexo	<NX tipnx=""> </NX>
MRD	marcador reformulatorio discursivo	<MRD> </MRD>
PTD	patrón tipográfico definitorio	<PTD pdef="" mt=""> </PTD>
PP	patrón pragmático	<PP tippp=""> </PP>

³ Una etiqueta consiste en una marca hecha en el documento que señala una porción de éste como un elemento o pedazo de información con un sentido claro y definido. Las etiquetas tienen la forma <contexto>, donde "contexto", por ejemplo, es el nombre de la información que se desea señalar.

Los primeros contextos fueron etiquetados de manera totalmente manual, es decir, se redactaba cada etiqueta. Al tratarse de un proceso que resultaba lento, se optó por manipular los textos a través de macros⁴ en *Word*. Esto hace que el etiquetado sea semiautomático, ya que con el cursor se señala el elemento y se presiona un botón en la barra de herramientas que aplica la etiqueta correspondiente. La barra de acceso es la siguiente:

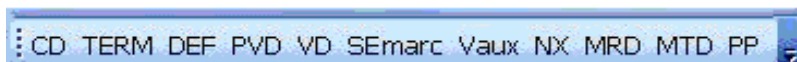


Figura 4-1. Barra de herramientas de macros CONTEXTOS

En el siguiente ejemplo mostramos un CD tal y como se obtiene del documento TXT, es decir de un archivo de texto plano que no tiene formato (tipo de letra, negritas, itálicas, etc.). En seguida se proporciona el ejemplo del mismo CD con etiquetas:

Tabla 4-2. Ejemplo de contexto definitorio

CD sin etiquetas
Reenganche rápido. Se conoce como reenganche rápido a la operación de cierre de un interruptor después de una falla.
CD con etiquetas
<CD num=""> Reenganche rápido. <PVD> <SEmarc tipsemarc=""> Se </SEmarc> <VD lema="" args="" mdo=""> conoce </VD> <NX tipnx=""> como </NX> </PVD> <TERM tipterm="" stterm=""> reenganche rápido </TERM> <DEF tipdef=""> a la operación de cierre de un interruptor después de una falla. </DEF> </CD>
Término = reenganche rápido Patrón verbal definitorio compuesto = Se conoce como Definición = a la operación de cierre de un interruptor después de una falla.

Una vez que se colocan las marcas correspondientes a cada elemento del CD, se integran manualmente los atributos para señalar sus particularidades. Antes mencionamos que para el *patrón verbal definitorio* se especifica el lema, el modo y los argumentos

⁴ Función que permite automatizar y agilizar tareas repetitivas. Por ejemplo, cuando en un procesador de textos debe realizarse una misma secuencia de comandos varias veces, resulta práctico disponer de una macro para simplificar en un solo paso (combinación de teclas, opción o clic del mouse) varios comandos individuales.

sintácticos del verbo definatorio, también se anota la posición del clítico *se* y el tipo de nexos que lo liga con la definición. Empleando el ejemplo anterior, vemos los atributos que se agregan a cada parte del CD:

Tabla 4-3. Ejemplo de contexto definatorio etiquetado con atributos

<pre><CD num="13"> Reenganche rápido. <PVD> <SEmarc tipsemarc="proc"> Se </SEmarc> <VD lema="conocer" args="3" mdo="fin"> conoce </VD><NX tipnx="adv"> como </NX></PVD> <TERM tipterm="L" stterm="fn"> reenganche rápido </TERM> <DEF tipdef="GD"> a la operación de cierre de un interruptor después de una falla. </DEF> </CD></pre>
<p>Término lingüístico (frase nominal)= reenganche rápido Patrón verbal definatorio= se conoce como Lema= conocer Argumentos del verbo principal= 3 Modo= Forma finita (conjugada) Nexo con la definición= como Definición Analítica= a la operación de cierre de un interruptor después de una falla</p>

Los CDs que se van incluyendo en el *Corpus de Contextos Definatorios* son previamente etiquetados con el fin de crear una base de datos estructurada. Los resultados últimos de esta tesis consisten en una propuesta de etiquetado para relaciones de referencia discursiva, en especial *relaciones anafóricas*, ya que este fenómeno es muy frecuente en los CD y creemos que debe ser incluido en el etiquetado general.

2. Etiquetas para relaciones de referencia discursiva en contextos definatorios

En una etapa posterior a la identificación de anáforas en el corpus, se clasificaron los elementos prominentes en las relaciones anafóricas y se procedió al diseño de etiquetas XML siguiendo los patrones de formación de aquellas establecidas para el *Corpus de Contextos Definatorios*.

Las cuatro etiquetas principales para relaciones referenciales discursivas son: *antecedente*, *anáfora pronominal*, *anáfora de sintagma nominal* y *no- antecedente*; para cada una se especifican ciertos rasgos:

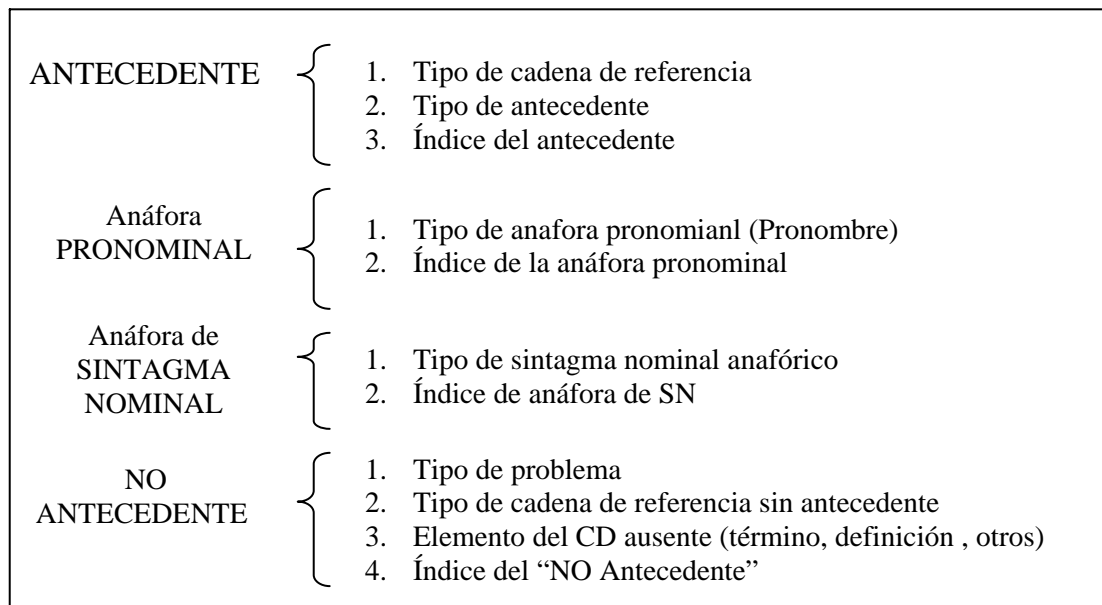


Figura 4-2. Elementos etiquetados en una relación anafórica

Además de los elementos y rasgos que se muestran en la figura anterior, se diseñaron etiquetas para *sintagmas nominales con determinante posesivo*, *catáforas* y otro tipo de referencias en el discurso, tales como *marcadores discursivos*, *marcas de deixis* y *nexos*, entre los más importantes. Cada una de las etiquetas posee ciertos atributos dependiendo de las características del elemento marcado; el diseño y descripción de dichas etiquetas se presenta como parte de los resultados de este trabajo y se detalla a continuación.

2.1. Antecedente (ANT)

El *antecedente* corresponde al componente que encabeza la cadena de referencias, es decir, la primera enunciación de aquello que será referido nuevamente en el contexto. Tanto las *anáforas pronominales* como *anáforas de sintagma nominal* apuntan a un *antecedente*, por lo cual existe una etiqueta para este elemento y se usa de forma indistinta para ambos tipos de expresión anafórica. La marca de antecedente está constituida por una etiqueta de apertura, una de cierre y los respectivos atributos que, como mencionamos, se especifican únicamente en la etiqueta de apertura. El antecedente se coloca entre dichas etiquetas y los

valores de los atributos se indican entre las comillas. Los atributos permiten marcar diferentes características para un mismo elemento. La etiqueta desglosada es la siguiente:

Tabla 4-4. Etiqueta para antecedente

Etiqueta de apertura	<ANT cadena="" tipant="" idant="">
Etiqueta de cierre	</ANT>
Atributos	cadena="" tipant="" idant=""
Completa	<ANT cadena="" tipant="" idant=""> </ANT>

El texto correspondiente queda comprendido entre la etiqueta de apertura y la de cierre. Con base en los resultados presentados en el Capítulos III, se observó que el *antecedente* puede ser de tipos diversos: encontramos *sintagmas nominales* precedidos por determinantes de distintas clases (definido, indefinido, posesivo, demostrativo, etc.), de género masculino o femenino y de número singular o plural. Los sintagmas además pueden ser modificados por complementos preposicionales o estar implicados en un fenómeno de coordinación oracional.

Localizamos con menor frecuencia antecedentes con estructura de sintagma verbal, adjetivo o adverbial y ejemplos en los que el antecedente corresponde únicamente a la parte de algún sintagma; otros, por ejemplo, abarcan una oración completa o una idea expresada en un fragmento textual. El antecedente debe ser etiquetado contemplando el determinante, los complementos y las relaciones de coordinación en el caso de los sintagmas. Se trata entonces de estructuras de distinta clase, por lo cual se establecieron tres atributos para describir sus rasgos:

Atributo 1 del ANTECEDENTE (cadena): La *cadena de referencias* nos permite precisar en la etiqueta del antecedente el tipo de relación referencial que éste desencadena. Existen cuatro clases de cadenas⁵ en función de los elementos que las constituyen; se plantearon los siguientes códigos para marcar sus valores:

⁵ En el Capítulo III dijimos que cuando una entidad (antecedente) aparece múltiples veces enunciada por otras, se construye una *cadena de referencias* [Corblin, 1995]. Para fines prácticos de etiquetado se considera también *cadena de referencias* a la relación entre antecedente y una única anáfora.

- **PR= Cadena de anáforas pronominales** en la cual el antecedente está exclusivamente relacionado con uno o varios *pronombres*.
- **SN= Cadena de anáforas de sintagma nominal** que se caracteriza porque el antecedente está ligado a uno o más *sintagmas nominales* que funcionan como expresiones anafóricas.
- **PS= Cadena de posesivo** atañe a la relación entre un antecedente y un sintagma nominal con determinante posesivo (relación *poseedor-poseído*).
- **MX= Cadena mixta** es aquella en la cual el antecedente pertenece tanto a *anáforas pronominales* como a *anáforas de sintagma nominal* y a expresiones con *posesivo*.

Es así que lo primero que se marca en la etiqueta del antecedente es el tipo de cadena que encabeza; los siguientes ejemplos exponen los cuatro posibles tipos:

- `<ANT cadena="PR" tipant="" idant=""> antecedente </ANT>`
- `<ANT cadena="SN" tipant="" idant=""> antecedente </ANT>`
- `<ANT cadena="PS" tipant="" idant=""> antecedente </ANT>`
- `<ANT cadena="MX" tipant="" idant=""> antecedente </ANT>`

Atributo 2 del ANTECEDENTE (tipant): El *tipo de antecedente* se marca conforme a dos propiedades: la primera considera la estructura sintáctica del antecedente y la segunda codifica específicamente el elemento del CD que funciona como antecedente en una relación de referencia.

Para efecto de los contextos y conforme a los resultados obtenidos en el corpus, se determinaron diez tipos de antecedente, éstos se especifican con un código (letra mayúscula) en la primera posición del *atributo 2*:

Tabla 4-5. Tipos de antecedente según la estructura sintáctica

VALOR	CÓDIGO
Sintagma nominal sin determinante	N
Sintagma nominal definido (artículo)	D
Sintagma nominal indefinido (artículo)	I
Sintagma nominal posesivo	P
Sintagma nominal con otros determinantes ⁶	S
Sintagma verbal	V
Sintagma adjetivo	A
Sintagma adverbial	R
Oración o cláusula	C
Fragmento textual (mayor a una oración)	F

En el caso de los antecedentes-sintagma nominal, los códigos están establecidos en función del tipo de determinante⁷ que los precede. Los sintagmas con uno o varios complementos son marcados como parte del antecedente. Los sintagmas con conjunciones disyuntivas (“o”, “y”, “e”) que despliegan una relación de coordinación, son contenidos en la misma etiqueta. Para dejar claras las especificaciones y diferencias señaladas, en los siguientes ejemplos se muestran construcciones que en el corpus aparecieron como antecedentes. Como ya dijimos, el código para marcar el tipo de estructura sintáctica del antecedente se coloca en primera posición del *atributo 2*:

- `<ANT cadena="" tipant="N_" idant="">` Instituciones relacionadas con la actividad
`</ANT>`
- `<ANT cadena="" tipant="D_" idant="">` el gradiente de una función potencia
`</ANT>`
- `<ANT cadena="" tipant="S_" idant="">` todos los bienes muebles e inmuebles
`</ANT>`
- `<ANT cadena="" tipant="I_" idant="">` una herramienta más de diseño en las formas hidrodinámicas del buque
`</ANT>`

⁶ Se trata de *pronombres indefinidos* (Apéndice A) que cuando preceden un sintagma nominal funcionan como determinantes, son del tipo: *ningún, todos, algún, pocos, muchos*, etc. Son llamados también determinantes cuantitativos [Rigau, 1981].

⁷ En los códigos del antecedente no se consideran las marcas de número y género de los sintagmas nominales; sin embargo, para no perder esta información se ha considerado incluir el determinante dentro de la etiqueta. El material anotado puede servir para trabajos futuros de procesamiento automático de anáforas y para ello la concordancia morfológica es de gran ayuda al momento de rastrear antecedentes.

En la segunda posición del *atributo 2 (tipant)* se señala el elemento del CD que es afectado en la relación de referencia. Esto quiere decir que las *anáforas* pueden referir a antecedentes que son el *término*, la *definición* u otro elemento del CD, principalmente *patrones pragmáticos* de autoría (Aut) y *temporales* (Temp). En el corpus se manifestaron cuatro tipos, para los cuales se plantean los siguientes códigos:

Tabla 4-6. Códigos de antecedente según el elemento del CD afectado

ELEMENTO DEL CD	CÓDIGOS
Antecedente= Término definido	T
Antecedente= Definición	D
Antecedente= Una parte de la definición	F
Antecedente= Otra parte del CD	X

En el ejemplo siguiente se expone un CD con los códigos correspondientes al *atributo 2*. Se trata de un antecedente etiquetado como *sintagma nominal indefinido* que abarca la *definición* del término:

Tabla 4-7. Ejemplo de “atributo 2” del antecedente

La frontera cerrada se considera como <ANT cadena=“PR” tipant=“ID” idant=“1” > una pared lateral sólida </ANT> por la cual ₍₁₎ no puede fluir ni penetrar el agua.
Antecedente = una pared lateral / I = SN con determinante indefinido / D = Definición
Anáfora pronominal = por la cual

Atributo 3 del ANTECEDENTE (idant): El *índice del antecedente* sirve para relacionar los elementos implicados en una relación de referencia discursiva. En los trabajos de anáfora, tradicionalmente se emplea un subíndice (un número, una letra o un símbolo) que sirve para destacar que en un texto dos o más entidades del discurso están *coindexados* en una relación de referencia [Castañeda, 1967; Van Hoek, 1995]:

Tabla 4-8. Ejemplo de elementos coindexados

la meiosis consta de <u>dos mitosis</u> ₁ : <u>las cuales</u> ₁ a su vez constan de profase, metafase, anafase y telofase.
--

En el caso de los CDs, el subíndice es sustituido por un índice (*id*), es decir, un número que sirve para rastrear los elementos de una cadena. Dicho número será el mismo que el de los índices de las anáforas pertenecientes a la misma cadena; además este número distingue una cadena de otra.

Una vez explicados los tres atributos del antecedente, y como resumen de la información contenida en la etiqueta, se proporciona un ejemplo que incluye tipo de cadena de referencias (*atributo 1*); tipo de antecedente según la estructura sintáctica y según el elemento del CD afectado (*atributo 2*); y, finalmente, el índice que liga los elementos de una cadena anafórica (*atributo 3*):

Tabla 4-9. Antecedente etiquetado

<ANT cadena="PR" tipant="FD" idant="1" > El conjunto de aparatos y sistemas puestos al servicio del sistema eléctrico, que vigilan que se cumpla adecuadamente el propósito para el que fue creado </ANT>, es lo que ₍₁₎ se conoce como protección.
Atributo 1 (cadena="PR") = Antecedente de cadena de anáfora pronominal.
Atributo 2 (tipant="FD") = Idea expresada en un fragmento textual (F) que corresponde a toda la definición (D).
Atributo 3 (idant="1") = Índice "1" que identifica la cadena y coincide forzosamente con la anáfora a la que pertenece.

2.2. Anáfora pronominal (PR)

La etiqueta de *anáfora pronominal* se diseñó con base en el paradigma pronominal expuesto en el Apéndice A y siguiendo los patrones de formación de las etiquetas generales de CDs. La marca de anáfora pronominal consta entonces de una etiqueta de apertura, una de cierre y, en este caso, dos atributos:

Tabla 4-10. Etiqueta para anáforas pronominales

Etiqueta de apertura	<PR tipron="" idpron="">
Etiqueta de cierre	</PR>
Atributos	tipron="" idpron=""
Completa	<PR tipron="" idpron=""> </PR>

Atributo 1 de ANÁFORA PRONOMINAL (tipron): El tipo de anáfora pronominal se codifica a partir de dos características de los pronombres: *tipo* y *presencia de artículo* y/o

preposición. Cada uno de dichos valores se explican primero y posteriormente se presentan los códigos en una tabla:

- (1) El tipo de anáfora según el pronombre supone el siguiente inventario: *personal, demostrativo, posesivo, relativo, indefinido, cuantitativos, numera, y reflexivo*. Existen marcas para cada uno, pero además se establecieron códigos para pronombres personales clíticos que conciernen únicamente a aquellos en función de acusativo (objeto directo) y dativo (objeto indirecto): *lo, la, los, las, le, les, se*. También se asignó una marca para reflexivos (*me, te, se, nos*) de acuerdo con la posición que ocupan respecto al verbo. La distinción se hace con el objetivo de marcar pronombres personales de dativo, acusativo y reflexivos distinguiendo la posición preverbal (proclítico) y posverbal (enclítico). Se trata de casos tales como *definiéndolo vs. lo define, considerándolas vs. las consideran, concibiéndolos vs. los conciben*, etc. En total son diez códigos para especificar los tipos de pronombre.

- (2) Los pronombres *relativos, posesivos, personales* (caso preposicional), *demonstrativos, numerales e impersonales* pueden aparecer precedidos por artículo y/o preposición, en cuyo caso se marca la presencia de estos elementos. Por ejemplo, ocurrencias tales como *en el cual, de los cuales, por la suya, lo que, el cual, los mismos, con los primeros, de ésta, para ello*, por señalar algunos. El objetivo de tomar en consideración la presencia de artículo y preposición en las formas anafóricas pronominales, es que en trabajos futuros se podrán estudiar anáforas en función de las relaciones sintácticas que entablan con sus antecedentes, de tal manera que es importante anotar esta clase de ocurrencias, tomando en consideración que la herramienta de etiquetado permite recuperar información puntual.

A continuación se exponen en una tabla los códigos del *atributo 1* de la anáfora pronominal, como ya se dijo, éstos corresponde al *tipo* (posición 1) y *uso de artículo-preposición* antepuestos (posición 2):

Tabla 4-11. Valores y códigos del “atributo 2”: Tipo de anáfora

POS.	ATRIBUTO	VALOR	CÓDIGO
1	Tipo	Personal (nominativo y caso preposicional)	P
		Personal acusativo y dativo (proclítico)	C
		Personal acusativo y dativo (enclítico)	E
		Demostrativo	D
		Posesivo	X
		Relativo	R
		Indefinido y cuantitativo	I
		Numeral	N
		Reflexivo (proclítico-posición preverbal)	F
		Reflexivo (enclítico-posición posverbal)	V
2	Preposición y artículo	Artículo	A
		Preposición	P
		Preposición y artículo	Y

En los contextos que exponemos a continuación se aprecian tipos distintos de anáforas pronominales etiquetadas; se resalta el *atributo 1* para mostrar sus características. Los valores para tipo de pronombre y marca de preposición y/o artículo, se codifican en posiciones diferentes. Es necesario llenar las dos posiciones con sus respectivos caracteres para no confundir el orden. Cuando el pronombre no está precedido de artículo y/o de preposición se coloca obligatoriamente el valor 0 (cero). En el caso de los personales y reflexivos en posición posverbal (enclíticos), la etiqueta abarca únicamente el pronombre y no el verbo con el cual aparecen:

Tabla 4-12. Ejemplos de “atributo 1” de anáfora

Anáfora de pronombre personal
Por consiguiente, el viento es una forma indirecta de la energía solar y el molino o el aerogenerador son artificios para capturar <PR tipron=“T0” idpron=“”> la </PR> y transformarla en energía útil para el hombre
“ capturar-la ” (tipron=“T0”)= personal enclítico (T), sin artículo ni preposición (0).
“ trasnformar-la ” (tipron=“T0”)= personal enclítico (T), sin artículo ni preposición (0).
Anáfora de pronombre relativo
El tanque Imhoff consta de un depósito <PR tipron=“RY” idpron=“”> en el que </PR> se separa la zona de decantación, que se ubica en la parte superior, de la de digestión, situada [*] en la inferior.
“ en el que ” (tipron=“RY”)= relativo (R) con artículo y preposición en posición antepuesta.

Anáfora de pronombre reflexivo
La respuesta de Babinski, que consiste en la extensión del dedo gordo del pie con separación del resto de los dedos en respuesta a la estimulación de la cara lateral de la planta del pie, puede despertar <PR tipron="V0" idpron=""> se </PR> a menudo en los neonatos; es consecuencia del escaso control cortical de la función motora, propio del encéfalo inmaduro.
"despertar-se" (tipron="V0")= reflexivo en posición posverbal (V), sin artículo ni preposición antepuestas (0).

Atributo 2 de ANÁFORA (idpron): El índice de la anáfora pronominal tiene la misma función que la del índice del antecedente (idant), es esencial que ambos números coincidan porque es el código que rastrea todos los elementos de una cadena.

En el ejemplo siguiente (Tabla 4-13) se despliega el resultado final del etiquetado de una cadena de anáforas y su respectivo antecedente:

Tabla 4-13. Ejemplo de CD con anáforas pronominales etiquetadas

Ejemplo de CD con cadena de anáforas etiquetadas
Un genoma es el número total de cromosomas, o sea todo el D.N.A..n.a (ácido desoxirribonucleico) de un organismo, incluido <ANT cadena="PR" tipant="PX" idant="1"> sus genes </ANT> , <PR tipron="RA" idpron="1"> los cuales </PR> llevan la información para la elaboración de todas <ANT cadena="PR" tipant="DX" idant="2"> las proteínas </ANT> requeridas por el organismo, y <PR tipron="RA" idpron="2"> las que </PR> determinan el aspecto, el funcionamiento, el metabolismo, la resistencia a infecciones y otras enfermedades, y también algunos de sus procederés .

Explicadas y mostradas las etiquetas en el ejemplo, se puede apreciar el tipo de información que es pertinente codificar para cada elemento inmerso en una relación de anáfora pronominal.

2.3. Anáfora de sintagma nominal (SN)

Según lo establecido en el Capítulo III, la anáfora de sintagma nominal reintroduce un elemento del discurso por medio de la repetición de un núcleo nominal o bien a partir de una relación semántica con el antecedente; en ambos casos puede tratarse de una referencia demostrativa. La marca diseñada para anotar esta entidad consta igualmente de etiquetas de apertura y cierre. Se consideran dos atributos para precisar las particularidades de las anáforas de sintagma nominal:

Tabla 4-14. Etiqueta para sintagma nominal anafórico

Etiqueta de apertura	<SN tipsn="" idsn="">
Etiqueta de cierre	</SN>
Atributos	tipsn="" idsn=""
Completa	<SN tipsn="" idsn=""> </SN>

Atributo 1 de ANÁFORA DE SINTAGMA NOMINAL (tipSN): El *tipo de anáfora de sintagma nominal* se etiqueta conforme a ciertas características identificadas en el corpus. La información que se codifica se estableció en función del determinante y la estructura del sintagma. Para el *atributo 1* se ocupan dos posiciones mostradas en la tabla siguiente:

Tabla 4-15. Códigos para tipos de anáfora de sintagma nominal

Posición	Atributo	Valor	Código
1	Tipo de anáfora de sintagma nominal	Sintagma nominal (sin complementos)	N
		Sintagma nominal con complemento preposicional o complemento preposicional ⁸	P
		Sintagma nominal con modificador oracional	M
		Sintagma-distributivo	V
2	Determinante	Definido (artículo)	D
		Indefinido (artículo)	I
		Demostrativo	T
		Cuantificador e Indefinido (pronombre)	C
		Numeral (ordinal y cardinal)	N

A simple vista, quizá no son claras las diferencias estructurales que se toman en cuenta para determinar el tipo de *anáfora de sintagma nominal* (primera posición del atributo). A continuación damos ejemplos para dejar cada caso explicado.

Los sintagmas nominales comúnmente están precedidos por un determinante que se codifica en las tres posiciones restantes del mismo atributo (determinante, género y número). No hay que perder de vista que las anáforas de sintagma nominal reintroducen una parte del discurso por medio de la repetición de un núcleo nominal o a partir de una relación semántica con el antecedente:

⁸ El sintagma anafórico puede ser un complemento preposicional, esto quiere decir que la relación con el antecedente se puede dar con el núcleo del sintagma preposicional, p.e. “de este modelo”. Se resolvió etiquetar los sintagmas preposicionales anafóricos incluyendo el sintagma nominal que lo precede: <SN tipSN="" idsn=""> la estructura de este modelo </SN> y entonces se usa el mismo código para sintagmas anafóricos con complemento preposicional y para sintagmas preposicionales anafóricos.

Tabla 4-16. Tipo de sintagmas nominales

Tipo	Núcleo nominal del antecedente	Anáfora de sintagma nominal
Sintagma nominal sin complementos (N)	modelo	Este modelo
Sintagma nominal con complemento preposicional o sintagma preposicional (P)	voltaje/ restablecimiento	voltaje de restablecimiento/ voltaje de restablecimiento / contra el de restablecimiento
Sintagma nominal con modificador oracional (M)	entramados	los diafragmas o entramados verticales <i>que conforman los muros</i>
Sintagma-distributivo (V)	(no reintroduce núcleo nominal)	tanto el nitrógeno orgánico como el amoníaco

Para expresiones del tipo *dicho transformador*, el determinante del sintagma es un participio y se codificará como demostrativo, por similitud con el valor y la función de los demostrativos *este, esta, ese, aquel*, etc. Los sintagmas con modificador oracional incluyen principalmente las expresiones que están modificadas por una oración de relativo. En todos los casos, si el sintagma está implicado en un proceso de coordinación oracional, dentro de la etiqueta queda marcada la relación de coordinación.

Encontramos construcciones del tipo *tanto...como, tan... como, tanto...cuanto, tal como, tales como*, un tipo de locuciones con valor referencial que serán consideradas y etiquetadas como anáforas de sintagma nominal. A continuación un ejemplo de esta clase de ocurrencias:

Tabla 4-17. Ejemplo de anáfora de sintagma-distributivo

Existen varios *fenómenos de desastre*₁, *tales como*₁ *temblores*₂, *diluvios*₂, *derrumbes*₂, tanto unos como otros son muy comunes en la naturaleza.

La expresión *tanto unos como otros* se etiqueta como anáfora pronominal de pronombre indefinido puesto que presenta *unos, otros* (Apéndice A); se trata de una expresión que implica un sintagma distributivo, pero para marcar la referencia damos prioridad a los pronombres indefinidos, cuantitativos y numerales.

El código “V” en la primera posición del atributo identifica entonces *sintagmas distributivos*. Para estos casos, el valor del determinante en la etiqueta es “0” (cero):

Tabla 4-18. Ejemplo de sintagma distributivo etiquetado

Ya en los 80's la Estrategia Mundial de la Conservación de la UICN planteó que el desarrollo en relación con la naturaleza se debía concebir como " la modificación de la Biosfera y <ANT cadena="SN" tipant="DF" idant="1"> la aplicación de los recursos humanos y financiero </ANT> , <SN tipSN="V0" idsn="1"> tanto bióticos y abióticos </SN>, a la satisfacción de las necesidades humanas y al mejoramiento de la calidad de vida
Antecedente (tipant="DF") = la aplicación de los recursos humanos Anáfora de SN distributivo (tipsn="V0") = tanto bióticos y abiótico

Con la variante de *sintagma distributivo*, en las etiquetas quedan incluidas aquellas construcciones sintácticas que poseen valor anafórico; no están agrupadas con los pronombres y se codifican con la etiqueta empleada para anáforas de sintagma nominal que se dan a partir de una relación semántica.

El etiquetado de anáforas de sintagma nominal (SN) incluye los sintagmas nominales con determinante demostrativo. En el corpus observamos que existen referencias demostrativas en las que se repite el núcleo nominal del antecedente y otras en las que la relación es semántica, tal es el caso del ejemplo siguiente:

Tabla 4-19. Ejemplo de sintagma nominal con demostrativo

Según otros autores la sustentabilidad es el estado o calidad de la vida, en la cual las aspiraciones humanas son satisfechas manteniendo la integridad ecológica. Esta definición , lleva implícito el hecho de que nuestras acciones actuales deben permitir la interacción con el medio ambiente y que las aspiraciones humanas se mantengan por mucho tiempo (Mooney, F. 1993).
Sintagma nominal con determinare demostrativo = Esta definición Antecedente = la sustentabilidad es el estado o calidad de la vida, en la cual las aspiraciones humanas son satisfechas manteniendo la integridad ecológica

Muchas veces en los resultados de la extracción automática, los sintagmas de este tipo aparecen cerca del patrón verbal definitorio (patrón de búsqueda), y frecuentemente están en el lugar del término. Las etiquetas permiten marcar el valor demostrativo de este tipo de sintagmas:

Tabla 4-20. Ejemplo de SN con determinante demostrativo etiquetado

<p>Según otros autores <ANT cadena="EC" tipant="FD" idant="1"> la sustentabilidad es el estado o calidad de la vida, en la cual las aspiraciones humanas son satisfechas manteniendo la integridad ecológica </ANT> <SN tipsn="ND" idsn="1"> Esta definición </SN>, lleva implícito el hecho de que nuestras acciones actuales deben permitir la interacción con el medio ambiente y que las aspiraciones humanas se mantengan por mucho tiempo (Mooney, F. 1993).</p>
--

Se observó que en ciertos casos el antecedente de un sintagmas demostrativos anafórico no recupera núcleo nominal; corresponde entonces a una idea expresada en un pedazo de texto, el cual se etiqueta abarcando todo el fragmento, ya que los atributos permiten codificarlo de este modo.

Atributo 2 de ANÁFORA DE SINTAGMA NOMINAL (idsn): El *índice de la anáfora de sintagma nominal* es necesariamente el mismo que el de su antecedente (idant) y siempre tiene que coincidir con el resto de los elementos de una cadena, sean expresiones pronominales o sintagmáticas. En seguida se presenta un ejemplo de CD con una cadena de anáforas de sintagma nominal y los respectivos elementos etiquetados. Primero se expone el contexto sin etiquetas, marcamos con negritas los núcleos nominales que constituyen la referencia anafórica, en seguida se presenta el resultado del etiquetado:

Tabla 4-21. Ejemplos CD con anáforas de sintagma nominal

<p>Las Barras Bravas en el fútbol pueden considerarse a grandes rasgos como grupos que inician los cantos en los estadios, portan la mayor cantidad de banderas y promueven ataques organizados contra Barras Bravas o hinchas rivales "</p>
<p><ANT cadena="SN" tipant="DT" idant="1"> Las Barras Bravas </ANT> en el fútbol pueden considerarse a grandes rasgos como grupos que inician los cantos en los estadios, portan la mayor cantidad de banderas y promueven ataques organizados <SN tipsn="P0" idsn="1"> contra Barras Bravas o hinchas rivales</SN>".</p>
<p>Antecedente= Las Barras Bravas SN= contra Barras Bravas o hinchas rivales (Sintagma preposicional/coordinación oracional) Núcleo nominal= Barras Bravas</p>

Se observó que la repetición léxica que genera la anáfora de sintagma nominal es muy productiva en los CDs, sin embargo, no es preciso marcar todas las referencias de este tipo.

Nos interesa etiquetar únicamente las anáforas de sintagma nominal que tienen como antecedente el término, ya que se trata de información prominente que se repite en el fragmento textual con la finalidad de ampliar la definición.

2.4. NO Antecedente (NOANT)

La etiqueta “NO Antecedente” se planteó con el objetivo de indicar la ausencia del antecedente de alguna anáfora. La etiqueta se usará principalmente para señalar que un CD está incompleto y es necesario entonces expandir el fragmento textual para completar la relación referencial. La etiqueta se coloca al inicio del CD y se diseñó conforme al resto:

Tabla 4-22. Etiqueta para No Antecedente

E. de apertura	<NOANT prob="" cad="" elem="" idnant="">
E. de cierre	</NOANT>
Atributos	prob="" cad="" elem="" idnant=""
Completa	<NOANT prob="" cad="" elem="" idnant=""> </NOANT>

Atributo 1 de NO ANTECEDENTE (prob). La ausencia de antecedente puede darse en función de distintos problemas, se consideran dos principales: *expansión*, que se refiere a contextos incompletos que deben expandirse para que una relación anafórica se complete, y *ambigüedad*, que se manifiesta cuando no puede identificarse con precisión el antecedente de la cadena, ya que en el contexto existe más de un candidato que puede funcionar como tal. Se incluye el código *otros* en caso de que se presente algún otro fenómeno por el cual el antecedente no esté en el CD. Se marcará respectivamente “expansion”, “ambigüedad” y “otros” como códigos para indicar el problema de ausencia de antecedente:

Tabla 4-23. Ejemplo de No Antecedente

Ejemplo de ausencia de antecedente
<code><NOANT prob="expansion" cad="" elem="" idnant="1"> </NOANT> <SN tipSN="ND" idsn="1"> Estos genes </SN> comprenden tanto aquellos con incremento de su función, los oncogenes , como aquellos que pierden su capacidad reguladora de la inhibición del crecimiento celular .</code>
Ejemplo de ambigüedad de antecedente
<code><NOANT prob="ambigüedad" cad="" elem="" idnant="1"> </NOANT> ARTICULO 8.- Se considerarán Socios Institucionales a los titulares de las siguientes Instituciones relacionadas con la actividad: Dependencias del Ejecutivo Federal, empresas de participación federal mayoritaria, las instituciones de educación superior, y <PR for="anafora" tipron="RNIX" idpron="1"> que </PR> convengan previamente su afiliación con la AMIP</code>
<p>Anáfora= (y) que</p> <p>Posible antecedente 1= los titulares de las siguientes Instituciones relacionadas con la actividad</p> <p>Posible antecedente 2= Dependencias del Ejecutivo Federal, empresas de participación federal mayoritaria, las instituciones de educación superior</p> <p>Posible antecedente 3= las instituciones de educación superior</p>

En el primer caso es muy claro que no hay antecedente de la anáfora “estos genes” y es necesario expandir el contexto para identificarlo. En cambio, en el segundo CD de la Tabla 4-23, la anáfora de relativo *que* liga una idea que puede referirse a tres partes diferentes del contexto, de tal manera que en procesos de resolución automática de anáforas sería muy difícil identificar el antecedente correcto. En gran medida, los problemas de ambigüedad de antecedente se dan porque éste no se encuentra inmediatamente antes de la anáfora, de modo que se trasponen otras estructuras que pueden cumplir dicha función. En el caso expuesto, la anáfora tiene tres candidatos a antecedente y como la cadena no se completa correctamente, se resolvió marcar “No Antecedente” y marcar el fenómeno como *problema de ambigüedad*.

Atributo 2 de NO ANTECEDENTE (cad). Se señala el *tipo de cadena de referencias* que no posee antecedente. En este atributo podemos señalar la ausencia de antecedente de alguna anáfora, de tal manera que se especifica el elemento referencial afectado; para ello se emplean los mismos códigos del *atributo 1* del antecedente:

- **PR**= Cadena de anáforas pronominales (**cad**="PR")
- **SN**= Cadena de anáforas de sintagma nominal (**cad**="SN")
- **PS**= Cadena de posesivo (**cad**="PS")
- **MX**= Cadena mixta (**cad**="MX")

Atributo 3 de NO ANTECEDENTE (elem). Se especifica el antecedente ausente en función del *elemento del CD* que debe localizarse en la expansión del contexto. Conforme a los resultados se plantearon tres códigos:

Tabla 4-24. Códigos de atributo 3: "elemento CD"

Término ausente	T
Definición ausente	D
Otro elemento del CD ausente	X

Este código permite señalar al antecedente ausente a partir de características concretas; por ejemplo, cuando la referencia apunta a un término omitido, se rastrea una unidad nominal; en cambio cuando no hay definición y la anáfora apunta a esta parte del CD, se busca una oración o un fragmento textual para completar el contexto:

Tabla 4-25. Ejemplo de elemento ausente en CD

Término definido ausente
<NOANT prob="expansion" cad="SN" elem ="T" idnant="1"> </NOANT> <SN tipsn="ND" idsn="1"> Estos elementos </SN> son parte constitutiva de los compuestos que forman la base material de la vida; los organismos los recaban de su entorno .
NOANT = Término SN = Estos elementos Definición = parte constitutiva de los compuestos que forman la base material de la vida ; los organismos los recaban de su entorno .
Definición ausente
<NOANT prob="expansion" cad="PR" elem ="D" idnant="1"> </NOANT> <PR tipron="D0" idpron="1"> Esto </PR> es <PR tipron="RA" idpron="1"> lo que </PR> se entiende como enfoque genético de la medicina o « medicina genética ».
NOANT = Definición PR1 = Esto PR2 = lo que

Atributo 4 de NO ANTECEDENTE (idnant) El índice es el mismo que el de la anáfora o cadena de anáforas sin antecedente en el contexto; es la marca que relaciona la anáfora con la etiqueta “No Antecedente”. El número se codifica igual que el índice de los elementos hasta ahora explicados.

3. Sintagma nominal con determinante posesivo, catáfora y otros tipos de referencia discursiva

Además de las etiquetas principales, se diseñaron etiquetas para codificar relaciones entre un sintagma nominal posesivo, casos de catáforas de pronombre relativo y otro tipo de referencias discursivas. Esto se hizo con el fin de considerar otros mecanismos referenciales que no son anáforas pero que aparecen en la estructura de los CDs.

3.1. Sintagma nominal con determinante posesivo (PS)

Se codifican sintagmas nominales con determinante posesivo, ya que se observó que éstos agregan información a la definición de los términos. En otras palabras, lo que se codifica es la *relación poseedor-poseído*, en donde la etiqueta antecedente marca específicamente al “poseedor” y otra etiqueta señala lo “poseído”.

Se diseñó esta marca porque los sintagmas nominales con determinante posesivo no son pronombres como tal y tampoco sintagmas nominales que reintroducen el núcleo nominal del antecedente. No obstante, esta clase de sintagmas despliegan cierta deixis textual y mantienen una relación semántica con su antecedente. Los sintagmas posesivos se etiquetan cuando el antecedente es el término o forma parte importante de la definición, ya que la información que introducen es prominente.

La estructura de la etiqueta es similar al resto, contiene únicamente un atributo índice (idpos) que coincide con el de su antecedente. Cabe recordar que en las etiquetas de *antecedente* y de *NO antecedente* hay un código para cadenas de posesivos destinado a marcar este tipo de relación referencial:

Tabla 4-26. Etiqueta para sintagma nominal con determinante posesivo

Etiqueta de apertura	<PS idpos="">
Etiqueta de cierre	</PS>
Atributo	idpos=""
Completa	<PS idpos=""> </PS>

En el siguiente caso veremos que la definición del término se extiende a partir de un sintagma nominal con determinante posesivo, de ahí la importancia de registrar estas estructuras cuando el antecedente es el término definido:

Tabla 4-27. Ejemplo de sintagma nominal posesivo

Se conoce como <ANT cadena="PS" tipant="NT" idant="1"> transformador de corriente </ANT> a aquél cuya función principal es cambiar el valor de la corriente de uno más o menos elevado a otro con el cual se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc. <PS idpos="1"> Su construcción </PS> es semejante a la de cualquier tipo de transformador, ya que fundamentalmente consiste de un devanado primario y un devanado secundario. La capacidad de estos transformadores es muy baja, se determina sumando las capacidades de los instrumentos que se van a alimentar, y puede ser 15, 30, 50, 60 y 70 VA.

3.2. Catáfora (CAT)

En el corpus encontramos algunos casos de *catáfora* (Apartado 5.1, Capítulo II), es decir elementos fóricos que aparecen antes que su *referente lingüístico* (“antecedente”). Siendo que primero aparece el *elemento catafórico* y luego el *referente discursivo* al que alude, el término *antecedente* es poco preciso para la *catáfora*, ya que en este caso la *relación fórica* no se da con un *antecedente* como tal, sino con un *subsecuente* [Peregrín, 1999:1430]. Hemos diseñado una etiqueta para los casos de catáfora con el objetivo de distinguirlas de las anáforas.

Dadas las ocurrencias de catáforas en el corpus, específicamente los casos del pronombre relativo *cuyo*, *cuya*, *cuyos*, *cuyas* se diseñó una etiqueta para marcar esta clase de relación referencial:

Tabla 4-28. Etiqueta para catáfora

Etiqueta de apertura	<CAT>
Etiqueta de cierre	</CAT>
Completa	<CAT> </CAT>

Vimos en el capítulo precedente que los relativos *cuyo*, *cuya*, *cuyos*, *cuyas* son pronombres que aparecen en posición de determinante, concuerdan en género y número con el nombre o núcleo nominal al cual anteceden [Picallo y Rigau, 1999]. Es importante anotar que los relativos de este tipo mantienen relación con dos entidades discursivas distintas: la que anteceden (puede ser un sintagma nominal o una oración) y aquella que les sigue (obligatoriamente un nombre o el núcleo de un sintagma nominal). La relación que queda marcada en la etiqueta es con el *subsecuente*.

Los únicos casos de catáfora que aparecieron en el corpus, como ya mencionamos, fueron casos con relativos *cuyo*, *cuyas*, *cuyos*, *cuyas*. Siendo que el subsecuente está obligatoriamente después del relativo, la etiqueta abarca tanto el pronombre como el nombre o núcleo nominal que les sigue:

Tabla 4-29. Ejemplos de catáforas de relativo etiquetadas

La resistencia de tierra R, <CAT> cuyo valor en ohms </CAT> se define por medio de la relación entre la tensión y la corriente de tierra o sea $R = V / I \cdot 4$.
Cada disparador consistió de una playa de amortiguamiento, <CAT> cuyo perfil </CAT> es de pendiente uniforme 1:4, construidos con grava de tezontle (material poroso) de 1 cm de diámetro, para producir una reflexión menor al 10%.
d) Relevador verificador. Es <CAT> cuya función </CAT> es verificar las condiciones del sistema de fuerza con respecto a límites prescritos.
En la cabecera se prolonga en <i>un</i> canalillo de trimado <CAT> cuyas paredes laterales </CAT> están constituidas por cristales de 12 mm de espesor.

3.3. Otros tipos de referencia discursiva (OREF)

En el proceso de etiquetado se detectaron algunas formas de referencia discursiva que no son anáforas ni catáforas ni sintagmas posesivos. Se diseñó una etiqueta muy sencilla y sin atributos para marcar todas aquellas estructuras que deberán ser estudiadas como parte de los mecanismos discursivos en los CD, ya que contribuyen a la cohesión y coherencia de los contextos. Se muestran a continuación algunos ejemplos y la etiqueta que se les consigna:

- <OREF> *ya que* </OREF>
- <OREF> *sólo que* </OREF>
- <OREF> *además* </OREF>
- <OREF> *sino también* </OREF>
- <OREF> *dado que* </OREF>
- <OREF> *de tal manera que* </OREF>
- <OREF> *de tal forma que* </OREF>
- <OREF> *de esta forma* </OREF>
- <OREF> *en este sentido* </OREF>

Cabe señalar que buena parte de estas estructuras fueron rastreadas al momento de hacer búsqueda automática del pronombre relativo *que* y de los determinantes demostrativos *esta, este*. Fue necesario diferenciarlas porque, en este caso, no cumplen una función de pronombre con valor anafórico ni de determinante de sintagma nominal anafórico. A pesar de que dichos resultados no conciernen a anáforas, se resolvió marcar esta clase de construcciones dada la frecuencia en los contextos. La etiqueta “<OREF></OREF>” revela que los mecanismos de referencia, cohesión y coherencia discursivas son amplísimos y todavía queda mucho por describir en lo que a CDs respecta.

4. Problemas de etiquetado

Como se mencionó, las etiquetas fueron diseñadas en una etapa posterior a la identificación de anáforas; sin embargo, fueron modificadas constantemente en el proceso de etiquetado hasta llegar a la versión que se ha expuesto en este documento. Al aplicar las etiquetas se detectaron problemas recurrentes, algunos pudieron resolverse agregando atributos, pero otros siguen presentándose al colocar las etiquetas y no son sencillos de resolver puesto que implican fenómenos de orden pragmático-semántico. De manera breve expondremos las complicaciones más comunes del etiquetado de relaciones anafóricas en CDs, planteamos las soluciones con ejemplos como los que hasta ahora se han expuesto.

4.1. Cadenas mixtas de referencia

Las cadenas mixtas contienen expresiones anafóricas pronominales y de sintagma nominal, además pueden aparecer sintagmas posesivos ligados al mismo antecedente. En el siguiente observamos esta variante:

Tabla 4-30. Ejemplos CD con cadena mixta (CLI)

Se conoce como transformador de corriente **a aquél cuya** función principal es cambiar el valor de la corriente **de uno** más o menos elevado **a otro con el cual** se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc.

Su construcción es semejante a la de **cualquier tipo de transformador**, ya que fundamentalmente consiste de un devanado primario y un devanado secundario. La capacidad **de estos transformadores** es muy baja, se determina sumando las capacidades de los instrumentos que se van a alimentar, y puede ser 15, 30, 50, 60 y 70 VA.

Estos transformadores son generalmente de tamaño reducido y el aislamiento que se emplea en **su construcción** tiene que ser de muy buena calidad, pudiendo ser en algunos casos resinas sintéticas (compound), aceite o líquidos no inflamables (pyranol, clorextol, etc.).

El contexto abarca más de un párrafo. Se trata de la definición de un término a partir de varios patrones verbales, por lo cual la extracción automática puede producir más de un CD. No obstante, en casos como el expuesto (Tabla 4-30), consideramos que se trata de un único contexto, en el cual las relaciones anafóricas sirven para agregar información sobre

un mismo término y por tanto no deben etiquetarse distintos CDs. En seguida se muestran las marcas para esquematizar el fragmento textual de varios párrafos como un CD único:

Tabla 4-31. Ejemplo de CD con cadena mixta etiquetada

<CD> Se conoce como <ANT cadena="MX" tipant="NT" idant="1"> transformador de corriente </ANT> <PR tipron="DP" idpron="1"> a aquél </PR> <CAT> cuya función </CAT> principal es cambiar el valor de la corriente <PR tipron="IP" idpron="1"> de uno más o menos elevado </PR> <PR tipron="IP" idpron="1"> a otro </PR> <PR tipron="RY" idpron="1"> con el cual </PR> se puedan alimentar instrumentos de medición, control o protección, como ampérmetros, wáttmetros, instrumentos registradores, relevadores de sobrecorriente, etc.

<PS idps="1"> Su construcción </PS> es semejante a la de cualquier tipo de transformador ya que fundamentalmente consiste de un devanado primario y un devanado secundario. <SN tipsn="CD" idsn="1"> La capacidad de estos transformadores </SN>, es muy baja, se determina sumando las capacidades de los instrumentos que se van a alimentar, y puede ser 15, 30, 50, 60 y 70 VA.

<SN tipsn="ND" idsn="1"> Estos transformadores </SN> son generalmente de tamaño reducido y el aislamiento que se emplea en <PS idps="1"> su construcción </PS> tiene que ser de muy buena calidad, pudiendo ser en algunos casos resinas sintéticas (compound), aceite o líquidos no inflamables (pyranol, clorextol, etc.).</CD>

Este mismo ejemplo sirve para describir el etiquetado de expresiones anafóricas tales como *de uno más o menos elevado a otro con el cual* que en principio puede segmentarse como tres expresiones distintas: (1) *de uno más o menos elevado* (2) *a otro* (3) *con el cual*. En los tres casos la referencia se da con el término *transformador de corriente* y por tanto forman parte de la misma cadena de referencia. La expresión (1) puede etiquetarse como anáfora de pronombre indefinido o bien como anáfora de sintagma nominal considerando el núcleo nominal elidido. En este trabajo se ha decidido etiquetar todas aquellas expresiones que impliquen un pronombre impersonal, cuantitativo o numeral (Apéndice B) como anáforas pronominales aun cuando exista elipsis de núcleo nominal:

- <PR tipron="IP" idpron="1"> *de uno más o menos elevado* </PR>
(Anáfora de pronombre impersonal (I), precedido por preposición (P); índice "1")
- <PR tipron="IP" idpron="1"> *a otro* </PR>
(Anáfora de pronombre impersonal (I), precedido por preposición (P); índice "1")

En el caso de los contextos incompletos, se acude al documento original para recuperar el antecedente ausente. Muchas veces el resultado son textos como el anterior, es decir, fragmentos textuales de mayor extensión que un párrafo. De ahí la importancia de resolver las relaciones anafóricas cercanas a una predicación verbal definitoria.

El etiquetado de relaciones discursivas en esta clase de contextos es necesario para optimizar los resultados de la extracción automática, ya que si las relaciones referenciales son resueltas, las anáforas no deberán aparecer en el lugar del término o de la definición y entonces puede disminuir la cantidad de contextos incompletos (CI).

4.2. Longitud del antecedente-sintagma y de la anáfora de sintagma nominal

Hemos visto en los ejemplos algunos de los antecedentes y algunos sintagmas anafóricos que aparecen en el corpus. Muchas veces al momento de etiquetar es difícil determinar su extensión, sucede sobre todo cuando el sintagma, sea un antecedente o una anáfora, tiene uno o mas complementos preposicionales, o bien, cuando aparecen conjunciones disyuntivas o complementos oracionales.

Pudimos observar que un sintagma anafórico puede ser preposicional y reintroducir el núcleo nominal de una expresión previa. Esto quiere decir que la relación anafórica se da por medio del núcleo de un sintagma preposicional:

Tabla 4-32. Extensión de sintagma anafórico

Cuando la tensión de una secuencia dada se aplica a una pieza o aparato, circula una pequeña corriente bien definida de la misma secuencia . Los aparatos se pueden caracterizar como poseedores de una impedancia definida de esta secuencia ; nombres especiales se han dado a esta impedancia, a saber: impedancia de secuencia positiva, impedancia de secuencia negativa e impedancia de secuencia cero.
Cuando <code><ANT cadena="SN" tipant="DX" idant="1"></code> la tensión de una secuencia dada <code></ANT></code> se aplica a una pieza o aparato, circula <code><SN tipsn="IX" idsn="1"></code> una pequeña corriente bien definida de la misma secuencia <code></SN></code> . Los aparatos se pueden caracterizar como <code><SN tipsn="ND" idsn="1"></code> poseedores de una impedancia definida de esta secuencia <code></SN></code> ; nombres especiales se han dado a esta impedancia, a saber: impedancia de secuencia positiva, impedancia de secuencia negativa e impedancia de secuencia cero.

Se resolvió etiquetar el sintagma anafórico abarcando el determinante que lo modifica y los complementos adyacentes que pueden ser adjetivos, preposicionales y oracionales de

tipo relativo (por ejemplo, los introducidos por el pronombre *que*). La anáfora es el complemento preposicional pero se ha optado por etiquetar toda la expresión para recuperar el sentido completo; así entonces, el determinante que se codifica es el de la expresión completa.

En el ejemplo precedente (Tabla 4-32) observamos que el *antecedente* es un sintagma con complemento preposicional. Los sintagmas anafóricos recuperan el núcleo nominal de el complemento preposicional y no del sintagma nominal principal. De la misma manera que con los sintagmas anafóricos, se resolvió etiquetar el antecedente incluyendo al sintagma principal y sus modificadores y complementos. En este sentido planteamos que la extensión de un sintagma nominal, sea antecedente o anáfora, se marcará abarcando estructuras complejas.

4.3. Anáfora de pronombre relativo “que”

Conforme a los resultados expuestos en el Capítulo III, el pronombre relativo *que* es uno de los más frecuentes en los CDs. El etiquetado de este tipo de anáfora puede parecer muy simple porque generalmente el antecedente de un relativo *que* se encuentra inmediatamente antes. No obstante, debido a la cantidad de oraciones relativas en los contextos, se manifestó cierta dificultad para etiquetar el pronombre *que*.

Esto significa que los sintagmas nominales modificados por oraciones relativas son muy frecuentes y el nexo que liga estas estructuras es un pronombre muy productivo, de tal manera que su carácter de anáfora debe ser marcado independientemente de si son parte de un antecedente o de sintagma nominal anafórico. En un caso como “esta compresión y descompresión del aire *que* está alojado en las cavidades de las rocas costeras” es un sintagma anafórico al cual le corresponde una etiqueta que abarca los modificadores oracionales. En la expresión también debe quedar marcada la anáfora *que* con antecedente inmediato *compresión y descompresión del aire*.

En un principio los *que* de oraciones relativas se etiquetaban como el resto de las anáforas pronominales, aun cuando éstos podían estar dentro de la etiqueta de un antecedente o de una anáfora de sintagma anafórico con modificador oracional. Hablamos de estructuras tales como “este dispositivo *que* nos permite proteger las instalaciones contra sobretensiones de origen atmosférico” se etiquetaba como *anáfora de sintagma nominal* y

además se colocaba la etiqueta de *anáfora pronominal* para indicar la presencia del pronombre relativo *que* operando como nexos entre el sintagma nominal con determinante demostrativo y la oración de relativo. Dada la frecuencia de casos como éste, se implementó una última etiqueta para evitar confusiones en la anotación de antecedentes y sintagmas anafóricos modificados por oraciones relativas. La nueva marca simplificó mucho la anotación general de anáforas, ya que por medio de una etiqueta distinta se codifica esta clase de ocurrencias.

Es importante insistir en que la etiqueta anota exclusivamente los *que* de oraciones relativas con antecedente nominal inmediato, mientras el resto de los casos con *que* se indican con la etiqueta “Anáfora pronominal”. Para dejar claras las especificaciones expuestas, en el siguiente CD se aprecia el uso de relativos etiquetados:

Tabla 4-33. Etiqueta para pronombre “que” en oraciones relativas

Etiqueta completa: <RELQ> SN+relativo <i>que</i> </RELQ>
La corriente eléctrica generada por el corazón se conduce a través de cables al aparato de registro, que consta fundamentalmente de un amplificador que magnifica las señales eléctricas y de un galvanómetro que mueve una aguda inscriptora, que se desplaza más o menos según la magnitud del potencial eléctrico (diferencia de potencial) que genera el paciente.
La corriente eléctrica generada por el corazón se conduce a través de cables <RELQ> al aparato de registro, que </RELQ> consta fundamentalmente de <RELQ> un amplificador que </RELQ> magnifica las señales eléctricas y <RELQ> de un galvanómetro que </RELQ> mueve <RELQ> una aguda inscriptora, que </RELQ> se desplaza más o menos según la magnitud <RELQ> del potencial eléctrico (diferencia de potencial) que </RELQ> genera el paciente.

Si para estos casos se usara la etiqueta de anáfora pronominal (PR), tendría que colocarse una marca para cada *que* y otra para su antecedente nominal. En cambio, con la marca <RELQ> </RELQ> se codifica el antecedente y la anáfora en la misma etiqueta, esto con el objetivo de no agregar otra etiqueta para el antecedente siendo que éste se encuentra inmediatamente antes del pronombre-nexo relativo. Así pues, el valor anafórico de los nexos de oraciones relativas se etiqueta como un fenómeno referencial aparte de los pronombres relativo *el que, para la que, de lo que, en la que*, etiquetados estrictamente como anáforas pronominales.

Con esta última etiqueta se evita entonces, interferencia entre los *que* precedidos de artículo y/o preposición y aquellos que funcionan como nexos oracionales con antecedente inmediato. En el siguiente CD se expone el etiquetado de diferentes ocurrencias del *que*; en este caso el pronombre queda marcado dentro de la etiqueta que corresponde al antecedente de otra anáfora del contexto:

Tabla 4-34. Ejemplo de “que” etiquetado

Albuferas : cuando los materiales que son transportados en suspensión en las corrientes marinas, chocan con un saliente de la costa, estos materiales son sedimentados y dan lugar a lo que se conoce como cordón litoral.
Albuferas : <ANT cadena=“PR” tipant=“FD” idant=“1”> cuando <RELQ> los materiales que </RELQ> son transportados en suspensión en las corrientes marinas, chocan con un saliente de la costa, estos materiales son sedimentados y dan lugar </ANT> <PR tipron=“RY” idpron=“1”> a lo que </PR> se conoce como cordón litoral.

La marca <RELQ> </RELQ> puede ser muy productiva en especial para el tipo de *definición analítica*, ya que comúnmente la *diferencia* es introducida por un pronombre *que*. Por ejemplo en un caso como “la mitocondria es un órgano *que* regula la respiración celular”, la etiqueta abarca <RELQ> un órgano *que* </RELQ> y entonces se puede rastrear la *diferencia* de la definición a partir de la anáfora.

4.4. Anáforas y usos problemáticos

Uno de los problemas más frecuentes en el marcaje de anáforas es que los pronombres aparecen juntos, unos seguidos de otros. Se trata de casos tales como *entre otros quienes, a quienes les pertenece, a lo cual se le agrega, aquellos cuyos, tanto unos como otros*, usos gramaticalmente correctos que se separan y etiquetan como distintas expresiones anafóricas. Además de estos casos, comúnmente nos enfrentamos a usos “problemáticos” de pronombres que provocan dificultades al etiquetar.

Los resultados de este trabajo, entre otras cosas, mostraron que las anáforas son expresiones referenciales muy productivas que evitan la repetición de cierta información. No obstante, el uso desmesurado, en especial de las anáforas pronominales, incide en ciertos problemas de comprensión de los textos. Quiere decir que las anáforas en ocasiones no están puestas en el discurso de modo que podamos rastrear su antecedente, y entonces

nos enfrentamos a problemas de ambigüedad y por tanto de coherencia y cohesión de los textos.

En este apartado exponemos dos problemas comunes al momento de aplicar las etiquetas: pronombres aglomerados y usos problemáticos. El siguiente ejemplo es un CD con muchos problemas para aplicar etiquetas debido al uso de anáforas:

Tabla 4-35. Ejemplo de usos “anómalos” de anáforas

<p>Las zonas costeras, en un alto porcentaje, se han caracterizado por tener importantes actividades agropecuarias, por lo que a cortas distancias disponen de alimentos, a los que es necesario adicionar los productos pesqueros que los tienen a la mano y al no necesitar enlatarlos ni congelarlos su costo es inferior a un 50 % de lo que cuestan en el altiplano.</p>
<p>Término= Las zonas costeras PVD= se han caracterizado por Definición= tener importantes actividades agropecuarias, por lo que a cortas distancias disponen de alimentos, a los que es necesario adicionar los productos pesqueros que los tienen a la mano y al no necesitar enlatarlos ni congelarlos su costo es inferior a un 50 % de lo que cuestan en el altiplano.</p>
<p>Las zonas costeras, en un alto porcentaje, <ANT cadena="PR" tipant="FD" idant="1"> se han caracterizado por tener importantes actividades agropecuarias </ANT> <PR tipron="RY" idpron="1"> por lo que </PR> a cortas distancias disponen de <ANT cadena="PR" tipant="NF" idant="2"> alimentos </ANT>, <PR tipron="RY" idpron="2"> a los que </PR> es necesario adicionar <ANT cadena="MX" tipant="DF" idant="3"> </RELQ> los productos pesqueros </ANT> que </RELQ> <PR tipron="R0" idpron="3"> los </PR> tienen a la mano y al no necesitar enlatar<PR tipron="T0" idpron="3"> los </PR> ni congelar<PR tipron="T0" idpron="3"> los </PR> <POS idpos="3"> su costo </POS> es inferior a un 50 % <PR tipron="RY" idpron=""> de lo que </PR> cuestan en el altiplano.</p>
<p>Antecedente 1= se han caracterizado por tener importantes actividades agropecuarias Antecedente 2= alimentos Antecedente 3= los productos pesqueros</p>

El ejemplo revela diversos problemas de etiquetado, entre los más evidentes está la dificultad para determinar la extensión de los antecedentes que en este caso son tres. Dado que las relaciones de referencia se dan con distintas partes del contexto y hay muchas anáforas, se genera ambigüedad de antecedentes. Nótese que en el caso de la oración “los productos pesqueros *que los* tienen a la mano”, el pronombre personal proclítico podría eliminarse sin alterar el significado, sin embargo, tenemos que marcarlo con su etiqueta correspondiente, a pesar de que se trate de un problema de redacción.

Se resolvió etiquetar esta clase de anáforas “problemáticas” siempre y cuando se logre identificar el antecedente, de lo contrario se coloca la etiqueta “No Antecedente” y el respectivo atributo que señala el problema “ambigüedad”. Las anáforas del tipo *entre otros quienes*, se marcan como dos anáforas distintas, por ejemplo, en este caso, se trata de una *anáfora de indefinido* correspondiente a *entre otros* y una de *relativo* que atañe a *quienes*.

La realidad es que en esta tesis y en el proyecto en general se trabaja con contextos como el expuesto en la Tabla 4-35. Si bien se trata de lengua escrita, en muchos casos se trata también de textos que no han sido sometidos a una revisión rigurosa de normas estandarizadas de gramática y redacción. Debemos estar conscientes de que las herramientas computacionales para extracción automática de CDs se aplican a lengua viva (“*parole*”), es decir, a la manifestación escrita de la lengua de cada autor, por lo cual encontraremos siempre casos difíciles de etiquetar. Entre mejor sea la descripción de los fenómenos referenciales discursivos, será más fácil resolver y evitar los problemas de etiquetado.

Por otro lado, la identificación y clasificación de los problemas de etiquetado será de gran ayuda, ya que se continúa con el proceso de etiquetado y de manera simultánea se detectan las dificultades para anotar anáforas, con el fin de analizar en un futuro los problemas más frecuentes para que las herramientas computacionales puedan solucionarlos automáticamente.

5. Resultados y observaciones finales

Las etiquetas codifican buena parte de los fenómenos de referencia anafórica y registran las particularidades predominantes. Con las etiquetas será posible rastrear información específica en el *Corpus de contextos definitorios*. Como dijimos, en el proceso de etiquetado se descubrieron fenómenos que no se habían descrito ni considerado en la identificación de relaciones discursivas en los CDs. El diseño y modificación persistente de las etiquetas implicó un esfuerzo por abarcar, en la medida de lo posible, todos los rasgos y problemas de las relaciones anafóricas.

Conforme a las descripciones que se presentaron en este capítulo, quedan establecidas y explicadas etiquetas XML para *antecedente*, *anáfora pronominal*, *anáfora de sintagma*

nominal, no antecedente, pronombre “que” de oraciones relativas, catáfora de pronombre relativo, sintagmas nominales con determinante posesivo y otras referencias discursivas:

Tabla 4-36. Etiquetas XML para anáforas y otros mecanismos de referencia discursiva

ELEMENTO	ETIQUETAS
Antecedente	<ANT cadena="" tipant="" idant=""></ANT>
Anáfora pronominal	<PR tipron="" idpron=""></PR>
Relativo <i>que</i>	<RELQ> </RELQ>
Relativo <i>cuyo</i>	<CAT> </CAT>
A. de sintagma nominal	<SN tipsn="" idsn=""></SN>
Ausencia de antecedente	<NOANT prob="" elem="" cad="" idnant=""> </NOANT>
Sintagma posesivo	<PS idps=""> </PS>
Otras referencias	<OREF> </OREF>

Estamos conscientes de que el proceso de etiquetado es muy laborioso y puede parecer complicado, sobre todo al momento de colocar atributos y revisar los valores y códigos correspondientes. No obstante, las etiquetas son necesarias para ordenar y clasificar la información, de tal manera que podemos hablar de información estructurada que podrá explotarse en otras investigaciones. Un corpus anotado es un corpus que otros pueden aprovechar para objetivos particulares. Para facilitar el trabajo de etiquetado se diseñaron macros en *Word* que simplifican el proceso de etiquetado. De tal manera que con el cursor se marca el elemento que se va a codificar y se presiona un botón en la barra de herramientas para colocar la etiqueta. Los valores de los atributos se colocan manualmente consultando los códigos en las tablas.

Si bien esta tesis es un trabajo básicamente descriptivo, enfocado a problemas concretos de la extracción automática de CDs, la resolución automática de anáforas es una más de las motivaciones de esta investigación. En este sentido, sería ideal que el ECODE pudiera localizar los antecedentes de anáforas que afectan términos y definiciones, hablamos entonces de trabajo en el campo de la resolución automática de anáfora y del Procesamiento de Leguaje Natural [Reinhart, 1983; Azzam *et al.* 1998; Charolles, 1999; Kleiber, 1999; Gelbukh y Sidorov, 2006].

Este último capítulo expuso una propuestas acerca de la manera de marcar relaciones en el discurso, con el fin de complementar los trabajos previos de etiquetado de artes de los

CDs. Los resultados del etiquetado XML de anáforas y otras relaciones de referencia se presentan en un CD-ROM adjunto a este documento.

Conclusiones

Conclusiones

Esta investigación consistió en un trabajo principalmente descriptivo, concentrado en identificar cierta clase de expresiones referenciales que se manifiestan en la estructura de los *contextos definitorios* (CDs), un tipo de fragmento textual que se distingue por introducir la *definición* de un *término* relacionado con un área de conocimiento particular.

No olvidemos que esta tesis se enmarca en el proyecto *Extracción de conceptos en textos de especialidad a través del reconocimiento de patrones lingüísticos y metalingüísticos*, cuyo objetivo es desarrollar métodos y herramientas computacionales para la extracción automática de información, concretamente la extracción de CDs. Cada uno de los resultados logrados en esta investigación sirve para complementar los trabajos previos que se han hecho sobre descripción de CDs, con el fin de mejorar la herramienta de extracción automática ECODE (*Extractor de contextos definitorios*).

En los primeros estudios sobre CDs, no se tomó en cuenta la presencia de ciertos mecanismos referenciales discursivos que constantemente aparecen en su estructura. Uno de los objetivos del trabajo expuesto en estas páginas, fue detectar y describir los fenómenos y expresiones que producen candidatos incompletos en el ECODE. Un candidato incompleto es un CD en el cual el *término* o la *definición* no están presentes porque se manifiesta alguna clase de fenómeno, mecanismo o relación discursiva que los afecta.

Para hablar de mecanismos referenciales, en primer lugar, tuvimos que determinar el carácter discursivo de los CDs. A partir de su descripción pudimos establecer que, efectivamente, los CDs son discurso en sí mismos, ya que participan de una intención comunicativa concreta y despliegan rasgos de textualidad tales como *aceptabilidad*, *informatividad*, *intencionalidad*, *situacionalidad* e *intertextualidad* [De Beaugrande y Dressler, 1997].

Los CDs, además, cumplen con características de entidades con sentido propio, de ahí que podemos precisar que constantemente acuden a mecanismos de la lengua, concretamente mecanismos de referencia discursiva que consolidan la *cohesión* y *coherencia textuales* que los circunscriben como discurso.

El corpus de este trabajo se conformó de 250 CDs que constituían los primeros resultados del proyecto. Aproximadamente el 30% fueron candidatos incompleto puesto que presentaban estructuras que hacían algún tipo de referencia al texto del cual fueron obtenidos. Dichas referencias son el fenómeno discursivo que nos abocamos a describir.

Como se dijo a lo largo de este trabajo, el lenguaje apela a diversos procedimientos que garantizan la cohesión de los textos en general y de los CDs en particular. Se hizo una revisión bibliográfica con el fin de determinar el tipo de estructuras lingüísticas y los mecanismos referenciales que repercuten en los resultados de la extracción automática.

En la extracción automática, no existe un criterio de extensión de CDs y como resultado el ECODE arroja candidatos incompletos que no funcionan como entidades discursivas con sentido propio. Para reducir el número de esta clase de candidatos es necesario que la herramienta expanda el contexto hasta completar el CD. La expansión permite fijar la extensión del CD, la cual está determinada por la presencia del término y su correspondiente definición.

Conforme se hizo la revisión bibliográfica observamos, y posteriormente comprobamos en el corpus, que *la anáfora* es un fenómeno lingüístico importante en la construcción de los textos en general, y por tanto también de los CDs. La anáfora es un mecanismo que permite hacer una referencia abreviada a una entidad previamente enunciada en el discurso; sirve para reintroducir en la comunicación un asunto, tema, concepto del que se seguirá hablando. La referencia anafórica, entonces, supone el punto de inflexión entre el pasado y el futuro de un texto.

Considerando que el tema de anáfora es muy extenso y se estudia desde distintas teorías y escuelas lingüísticas, para describir las ocurrencias en el corpus, nos ceñimos a tres criterios de clasificación específicos [Moreno *et al.*, 1999: 205-213]. Dichos criterios explican la anáfora en virtud del *tipo de expresión anafórica*, según el *contexto en el que opera* y conforme al *marco en el que ocurre* (se basa en la proximidad con el antecedente).

Se determinó que *pronombres y sintagmas nominales anafóricos* son el tipo de expresiones que principalmente dan como resultado contextos incompletos en el ECODE. Las anáforas son, entonces, pronombres y sintagmas nominales que refieren o reintroducen una entidad que funciona como antecedente.

Las anáforas en CDs constantemente remiten a un antecedente expresado explícitamente, son denominadas *anáforas en contexto lingüístico*. Podemos anotar que por el tipo de discurso, difícilmente encontraremos *anáforas en contexto situacional*, es decir, expresiones que se resuelven en una situación de enunciación determinada, ya que el elemento anafórico remite al contexto extralingüístico.

Considerando el marco en el cual la anáfora ocurre, se identificaron *anáforas discursivas* y *anáforas intraoracionales*. Las intraoracionales son aquellas en las que el antecedente está en la misma oración que la expresión anafórica. Las discursivas, en cambio, remiten a un antecedente que se encuentra en una o varias oraciones previas. Un buen porcentaje de los candidatos incompletos están encabezados por anáforas discursivas. Esto significa que el antecedente de la anáfora está en alguna oración previa que no fue recuperada por la herramienta.

En trabajos futuros será preciso estudiar la relación que existe entre el *tipo de expresión anafórica* (sea pronominal o de sintagma nominal) y el *tipo de anáfora según el marco en el que opera* (puede ser oracional o discursiva). La distinción es importante porque se busca que el ECODE haga expansiones automáticas. En este sentido, si podemos determinar el vínculo que se entabla entre pronombres y sintagmas anafóricos con respecto a la distancia respecto a su antecedente, se podrá especificar qué tanto debe expandirse un CD.

Otro de los objetivos de este trabajo fue ahondar en el fenómeno de anáfora para proponer patrones recurrentes que permitan determinar un criterio de extensión de los CDs. Conforme a esto, se observó que el *patrón verbal definitorio* (patrón de búsqueda) es muy importante para determinar la extensión del fragmento textual que contiene la definición. En el corpus vimos que cualquier expresión referencial cercana al patrón de búsqueda, puede producir un candidato incompleto cuando ésta entabla relación con una parte anterior del texto. Es importante no perder de vista que para el ECODE el *patrón verbal definitorio* es el principal indicador de la configuración de un CD.

Determinamos que la expansión de los CDs es necesaria cuando una anáfora está próxima al patrón verbal definitorio. En este sentido, la extensión de un CDs puede establecerse en función de las relaciones referenciales que el término y la definición instauran entre ellos o con otras partes del CD, e incluso con el texto del cual se extraen.

Para que la expansión del candidato se haga automáticamente debe tomarse en cuenta que la extensión de los CDs es variable. En principio, un CD debe expandirse (1) cuando una anáfora afecta al *patrón verbal definitorio*; (2) cuando el fragmento textual comienza con una anáfora (pronombre y/o sintagma nominal anafórico); y (3) cuando la oración que sigue al fragmento textual, comienza con una anáfora.

Los mecanismos de referencia en CDs inciden en la cohesión textual, permiten el flujo comunicativo, la progresión y continuidad temática a partir de la *repetición*, *sustitución* y *omisión* (elipsis) de determinados elementos del CD. A lo largo de esta tesis describimos específicamente los fenómenos de repetición léxica y de sustitución, pero será necesario en trabajos futuros estudiar la ocurrencia de otros mecanismos de referencia tales como la elipsis y los marcadores discursivos.

Las anáforas que describimos son aquellas que afectan al término o a la definición. El *término* y la *definición* son la información prominente del CD, y por su estructura sintáctico-semántica son candidatos totalmente idóneos para participar en una relación anafórica, ya que muchas veces, su significado se reitera y amplía acudiendo a mecanismos de referencia discursiva.

La descripción para formular patrones de búsqueda fue más productiva para *anáforas pronominales*. Esto se debe a que los pronombres constituyen un paradigma limitado que podemos rastrear con ayuda de herramientas computacionales. Entre las ventajas que nos proporcionó el uso de herramientas computacionales para manejar el corpus, podemos mencionar que, en el caso específico de esta investigación, localizamos anáforas de manera automática.

Con la herramienta de *Word*, identificamos *anáforas pronominales* que por sus características formales arrojan muy buenos resultados en la localización automática. Observamos, no obstante, que algunos pronombres como los *personales nominativos* (*yo, tú, él, nosotros*, etc.), los de *caso preposicional*, los *demonstrativos* (*éste, ése, aquél*, etc.), los *relativos*, (*cual, que, cuyo*, etc.) y los *indefinidos* (*algunos, muchos, pocos, ninguno*, etc.) producen mejores resultados con respecto a los personales de acusativo (*la, lo, los, las*, etc.) y dativo (*me, te, le, les*, etc.). La diferencia en los resultados se debe a que los *personales de acusativo* y *de dativo* comparten forma con otras estructuras como *artículos definidos* (*la, lo, los, las*).

En trabajos futuros será necesario hacer una descripción más precisa, no sólo de la forma de los pronombres sino también de su función y las relaciones sintácticas que presentan con otros elementos de la oración. Esto permitirá afinar los patrones de búsqueda.

Por otro lado, los sintagmas anafóricos, con excepción de los sintagmas demostrativos, no pudieron ser localizados automáticamente. Para encontrarlos hubo que revisar contexto por contexto, siendo que mantienen una relación semántica con su antecedente. En un futuro tendremos que estudiar con mayor detenimiento el tipo de relación semántica entre un sintagma anafórico y su antecedente (anáfora indirecta).

Cuando hablamos de localización automática de anáforas (específicamente de pronombres) damos por supuesto que las anáforas sirven como indicadores para expandir contextos y producir un menor número de candidatos incompletos. Cabe aclarar que las anáforas en sí mismas no son indicadores de la presencia de CDs en los textos (CLI y Corpus Técnico del IULA). Sin embargo, las anáforas que afectan o están cercanas al patrón de búsqueda (patrón verbal definitorio) indican que la herramienta puede producir un candidato incompleto.

La descripción de las anáforas contribuyó al mejoramiento de patrones de búsqueda de CDs y al diseño de etiquetas que incluyen y marcan relaciones discursivas. Finalmente podemos subrayar la importancia de estudiar relaciones anafóricas como parte de la estructura de los CDs y concluir que:

- I. Los CDs mantienen una cohesión y coherencia textuales específicos, a pesar de que se trata de fragmentos discursivos extraídos de un universo textual mayor. Funcionan como entidades con sentido propio.
- II. Es importante establecer un criterio de extensión de los CD en la herramienta ECODE para que la información del *término* y la *definición* no se pierda.
- III. La extensión de un CD está en gran medida determinada por el tipo de relaciones referenciales que presenta, ya que la definición de un término se agota cuando se resuelven las relaciones referenciales que lo afectan.

- IV. A través del reconocimiento y análisis de anáforas fue posible proponer criterios de extensión de los CDs que permiten señalar la pertinencia de expandir en la extracción automática ciertos candidatos para recuperar información prominente de la definición y/o del término.
- V. Los CDs deben expandirse cuando una expresión anafórica está cercana al patrón verbal definitorio (patrón de búsqueda), ya que su antecedente se encuentra en una parte previa del texto y esta característica puede producir candidatos incompletos.
- VI. El mejoramiento de patrones de búsqueda que incluyen relaciones de anáforas podrá optimizar los resultados en la extracción automática, es decir, disminuirá la cantidad de candidatos incompletos.

A partir de esta investigación, nos percatamos de la complejidad del tema de anáfora, tanto como para describir sus características como para comprender las teorías diversas que la explican. Para fines prácticos, adoptamos una visión teórica-descriptiva que permitió proponer soluciones a problemas específicos de la extracción automática.

Es importante anotar que este trabajo es el preámbulo de una investigación que seguirá desarrollándose. Quedan muchos estudios por hacer respecto al tema de anáforas en CDs. Además, debemos atender otro tipo de mecanismos referenciales que conviven con la anáfora y que igualmente inciden en la estructura de los CDs, es el caso específico de la elipsis.

Ciertamente, en esta tesis sólo consideramos anáforas pronominales y de sintagma nominal, aunque estamos conscientes de que en el campo de la resolución automática de anáfora se han tratado otros tipos [Moreno *et al.* 1999] que también tenemos que tomar en cuenta.

No era asunto de este trabajo adoptar una visión del tema enfocada a la resolución automática de anáforas, sin embargo, ésta es una importante línea de investigación en el campo del Procesamiento del Lenguaje Natural. Para lograr avances en este sentido, como labor a futuro es necesario analizar la relación anáfora- antecedente. Es decir, describir el vínculo entre el tipo de anáfora y el tipo de antecedente al que apunta.

Nuestro principal objetivo estuvo abocado a mejorar la herramienta de extracción automática ECODE y logramos resultados importantes al plantear los primeros patrones recurrentes de anáforas que afectan al patrón de búsqueda. Cabe señalar que la descripción de patrones seguirá siendo cardinal, en la medida que dicha descripción sea más precisa, los resultados serán mejores.

Estamos conscientes de que la anáfora y sus implicaciones tanto en el discurso como en las herramientas computacionales que la identifican, involucran mucho más que un trabajo descriptivo como el que se presentó, sin embargo, pudimos dimensionar la magnitud de los problemas que tenemos que tratar en investigaciones futuras.

Como resultado final, se expuso una propuesta de etiquetado para complementar las etiquetas XML que ya existen en el *Corpus de Contextos definitorios*. El etiquetado permite ordenar y estructurar la información de los CDs.

Las etiquetas fueron cambiando conforme se iban integrando al corpus. La ventaja del sistema de etiquetas XML es que podemos modificarlas tantas veces sea necesario. Esto permitió agregar o cambiar marcas cuando se presentaban problemas para etiquetar.

Es importante decir que el etiquetado de relaciones referenciales en CDs no ha concluido y podrán efectuarse cambios en tanto aparezcan usos problemáticos o casos difíciles de marcar. El objetivo final es que el conjunto de etiquetas abarque, en la medida de lo posible, las características potenciales de las relaciones anafóricas y otras relaciones de referencia discursiva.

Por último, cabe destacar que el etiquetado que presentamos servirá, entre otras cosas, para recuperar información puntual de relaciones anafóricas, lo cual permite continuar trabajando en la descripción de anáforas. Queda mucho por hacer en lo que a referencia discursiva atañe; esta tesis fue un acercamiento al tema anáfora con fines muy precisos: detectar expresiones que generan candidatos incompletos en el ECODE.

Bibliografía

Bibliografía

- ABAD, A. (1980). *Introducción a la lingüística*. Madrid: Alhambra.
- AGUILAR, C. (2006). "Análisis lingüístico de definiciones situadas en contextos definitorios". Proyecto de tesis doctoral. Universidad Nacional Autónoma de México.
- AUGUER, A. (1997). *Repèrage des énoncés d'intérêt définitoire dans les bases de données textuelles*. Tesis de doctorado. Université de Neuchâtel, Suisse.
- ALARCÓN, R. (2003). *Análisis lingüístico de contextos definitorios en textos de especialidad*. Tesis de licenciatura. Universidad Nacional Autónoma de México, México.
- ALARCÓN, R. (2006). *Extracción automática de contextos definitorios en corpus especializados. Propuesta para el desarrollo de un ECODE (Extractor de contextos definitorios)*. Proyecto de tesis doctoral. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona.
- ALARCÓN, R. y SIERRA, G. (2003). "El rol de las predicaciones verbales en la extracción automática de conceptos". En *Estudios de Lingüística Aplicada* (pp. 129-144), Núm. 38, Centro de Enseñanza en Lenguas Extranjeras, Universidad Nacional Autónoma de México.
- ALLEN, J. (1995). *Natural Language Understanding*. London: Benjamin Cummings Publishing Company.
- AZZAM, S., HUMPHREYS K. y GAIZAUSKAS, R. (1998). "Extending a Simple Coreference Algorithm with a Focusing Mechanism". En *Proceedings of the Second Colloquium on Discourse Anaphora and Anaphor Resolution* (pp. 15-27), Lancaster: Department of Computer Science, University of Sheffield.
- BACH, C., SAURÍ, R., VIVALDI, J. y CABRÉ, M.T. (1997). *El corpus del l'IULA descripció. Informe 17*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- BELLO, A. (1951). *Gramática de la lengua castellana*. En *Obras completas*. Caracas.
- BERISTÁIN, H. (2004). *Diccionario de retórica y poética* (8ª ed.). México: Porrúa.
- BERNÁRDEZ, E. (1982). *Introducción a la lingüística del texto*. Madrid: Espasa- Calpe.
- BEAUGRANDE DE, R. y DRESSLER, W. (1997). *Introducción a la lingüística de texto*, BONILLA, S. (Ed.). Barcelona: Ariel.

- BROWNS, G. y YULE, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- BRUCART, J. M. (1999). "La estructura del sintagma nominal: las oraciones de relativo". En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.393-522). Madrid: Espasa Calpe.
- CABRÉ, M. (1999). *La terminología representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- CABRÉ, M.T., ESTOPA, R. y VIVLADI, J. (2001). "Automatic term detection. A review of current systems". En BOURIGAULT, D., JACQUEMIN, C., L'HOMME, M.C. (Eds.), *Recent advances in Computational Terminology* (pp. 53-87). Amsterdam: John Benjamin's.
- CARRELL, P. L. (1982). "Cohesion is not coherence". En *Tesol Quartely*, 16 (4): 28-30.
- CASADO, M. (1994). *Introducción a la gramática del texto del español*. Madrid: Arco/Libros.
- CASTAÑEDA, H. (1967). "Omniscience and Indexical Reference". En *The Journal of Philosophy*, 64, (7): 203-210.
- CHARAUDEAU, P. (1992). *Grammaire du sens et de l'expression*. Paris: Hachette-Éducation.
- CHAROLLES, M. (1999). "Associative anaphora and its interpretation". En *Journal of Pragmatics*, 31: 311-326.
- COVINGTON, M. (1994). *Natural Language Processing for Prolog Programmers*. New Jersey: Prentice Hall.
- CORBLIEN, F.(1995). *Les formes de reprise dans le discours : Anaphores et chaînes de Référence*. Rennes: Presses universitaires de Rennes.
- CORNISH, F. (1986). "Anaphora and semantic interpretation". En *Lingua*, 68: 241-256.
- CORNISH, F. (2002). "Anaphora: lexico-textual structure, or means for utterance integration within a discourse? A critique of the Functional Grammar account". En *Linguistics*, 40 (3): 469-493.
- CORTÉS, L. (1991). *Sobre conectores, expletivos y muletillas en el español hablado*. Málaga: Ágora.
- DIJK, T.A. VAN (1983). *La ciencia del texto*. Barcelona: Paidós.

- DIJK, T.A. VAN (1984). *Texto y contexto*. Madrid: Cátedra.
- DONNELLAN, K. (1966). "Reference and Definite Descriptions". En *The Philosophical Review*, 75: 281-304.
- DUBUC, R. (1992). *Manuel pratique de terminologie*. Canadá: Linguattech éditeur.
- DUROSS, E. (1990). "Anaphora in natural language processing and information retrieval". En *Informarion Processing & Management* 26 (1): 39-52. School of Information Studies, Syracuse University, Syracuse, New York.
- EGUREN, L.J. (1999). "Pronombres y adverbios demostrativos. Las relaciones deícticas". En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.929-972). Madrid: Espasa Calpe.
- ESCANDELL, M.V. (1995). *Los complementos del nombre*. Madrid: Arco/Libros.
- ESCAVI, R. (1987). *El pronombre. Categorías y funciones pronominales en la teoría gramatical*. Murcia: Secretariado de Publicaciones de la Universidad de Murcia.
- FERNÁNDEZ, S. (1987). *La nueva gramática académica: el camino hacia el esbozo*. Madrid: Paraninfo.
- FERNÁNDEZ, S. (1990). "El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos". En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.1208-1269). Madrid: Espasa Calpe.
- FERNÁNDEZ-ORDÓÑEZ, I. (1999). "Leísmo, laísmo y loísmo". En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.1377-1399). Madrid: Espasa Calpe.
- FILLMORE, C. (1997). *Lectures on deixis*. Stanford, California: CSLI Publications.
- FRANCHINI, E. (1986). "Las condiciones gramaticales de la coordinación copulativa en español". En *Romanica Helvetica*, 12.
- FREGE, G. (1998). *Ensayos de semántica y filosofía de la lógica*. VALDÉS, L. M. (Ed.). Madrid: Tecnos.
- GAIFFE, B., REBOUL, A., y ROMARY, L. (1994). "Références et gestion du dialogue". En *Actes du colloque TALN'94*, Marseille: 134-150.
- GELBUKH, A. y SIDOROV, G. (2006). *Procesamiento automático del español con enfoque en recursos léxicos grandes*. México, DF: Centro de investigaciones en computación, Instituto Politécnico Nacional.

- GONZALO A. J. (2003). "Recuperación y extracción de información". En MARTÍ, MA. A. (Coord), *Las tecnologías del lenguaje* (pp. 157-192). Barcelona: Editorial UOC.
- GREEN, G. M. (1989). *Pragmatics and Natural Language Understanding*. New Jersey: LEA.
- GROSZ, B. J. (1981). "Focusing and description in natural language Dialogues". En JOSHI A., SAG K., y WEBBER, B. (Eds.), *Elements of Discourse*. Cambridge: Cambridge University Press.
- GUNDEL, J., HEDBERG, N. y ZACHARSKI, R. (1993). "Cognitive Status and the Form of Referring Expressions in Discourse". En *Language*, 69 (2): 274-307.
- HAWKINS, J. (1984). "A note on referent identifiability and co-presence". En *Journal of Pragmatics*, 8: 649-659.
- HOEK, H. VAN (1995). "Conceptual Reference Points: A Cognitive Grammar Account of Pronominal Anaphora Constraints". En *Language*, 71 (2): 310-340.
- HUANG, Y. (2000). "Discourse anaphora: Four theoretical models". En *Journal of Pragmatics*, 32: 151-176.
- KLEIBER, G., SCHNEDECKER, C. y UJMA, L. (1994). "Anaphore associative: d'une conception l'autre". En C. SCHNEDECKER y ALII (Eds.), *L'anaphore associative. Aspects linguistiques, psycholinguistiques et automatiques* (pp. 5-64). Paris: Klincksieck.
- KRIPKE, S. (1972). *Naming and Necessity*. Cambridge, Massachusetts: Harvard University Press.
- LAVIGNE-TOMPSS, F. y DUBOIS, D. (1999). "Context effects and associative anaphora in reading". En *Journal of Pragmatics*, 31: 399-415.
- LIANG, T. y WU D-S. (2004). "Automatic Pronominal Anaphora Resolution". En *Computational Linguistics and Chinese Language Processing*, 9 (1): 21-40.
- LYONS, J. (1983). *Lenguaje, significado y contexto*. Barcelona: Paidós.
- LOPE BLANCH, J.M. (1979). *El concepto de oración en la lingüística española*. México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Filológicas.
- MALDONADO, R. (1992). *Middle voice: the case of Spanish se*. Tesis doctoral. Universidad de California.

- MARTÍN, M.A. (1979). *Las construcciones pronominales en español: paradigma y desviaciones*. Madrid: Gredos.
- MARTÍNEZ, J.A. y ÁLVAREZ, M. A. (1989). *El pronombre* (2 Vol.). Madrid: Arco.
- MEDEROS, H. (1988). *Procedimientos de cohesión en el español actual*. Santa Cruz de Tenerife: Publicaciones del Cabildo Insular de Tenerife.
- MEDINA, A. y SIERRA, G. (2004). "Criteria for the Construction of a Corpus for a Mexican Spanish Dictionary of Sexuality". En *Proceedings of 11th Euralex International Congress*. Lorient, France: Université de Bretagne-Sud, July 6-10.
- MEYER, I. (2001). "Extracting a knowledge-rich contexts for terminography: A conceptual and methodological framework". En BOURIGAU, D. JAQUEMIN, C. y L'HOMME, M.C. (Eds.), *Recent Advances in Computational Terminology* (pp.279-302). Amsterdam: Benjamin's.
- MILNER, J.C. (1989). *Introduction à une science du langage*. Paris: Seuil.
- MITKOV, R. (1998). "Robust Pronoun Resolution with Limited Knowledge". En *Proceedings of the 18th International Conference on Computational Linguistics (COLING'98)/ACL'98*. Conference: 869-875.
- MORENO-CABRERA, J. C. (1991). *Curso universitario de lingüística general*, Vol. 1. Madrid: Síntesis.
- MORENO, L., PALOMAR, M., MOLINA, A. y FERRÁNDEZ, A. (1999). *Introducción al Procesamiento del Lenguaje Natural*. Alicante: Servicio de Publicaciones de la Universidad de Alicante.
- MUÑOZ, R. (2001). *Tratamiento y resolución de las descripciones definidas y su aplicación en sistemas de extracción de información*. Tesis Doctoral. Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, España.
- NÚÑEZ, L. (1993). *Teoría y práctica de la construcción del texto. Investigación sobre gramaticalidad, coherencia y transparencia de la elocución*. Barcelona: Ariel.
- OLIVARES, M. A. (1994). "La anáfora en el discurso francés: aspectos teórico-prácticos". En *Actas del Ier Congreso de Lingüística general*, Valencia: 118-125.
- OLIVAS, O. (2006). *Sistema de construcción de redes semánticas con detección de anáfora*. Tesis de maestría. Instituto Politécnico Nacional, México.
- PALOMAR, M., FERRÁNDEZ, A. y MORENO, L. (1995). "Aportaciones a la Resolución de la elipsis en la coordinación". En *Procesamiento del Lenguaje Natural*, 16.
- PEARSON, J. (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamin's.

- PEREGRÍN, C. (1999). “Pronombre reflexivos y recíprocos”. En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.1427-1514). Madrid: Espasa Calpe.
- PICALLO, M. C., y RIGAU, G. (1999). “El posesivo y las relaciones posesivas”. En BOSQUES I. Y DE MONTE, V. (Coord.), *Gramática descriptiva de la lengua española*, Tomo 1: Sintaxis básica de las clases de palabras (pp.973- 1023). Madrid: Espasa Calpe.
- PORTO, J-A. (1997). *Oraciones de relativo*. Madrid: Arco/Libros.
- PUTMAN, H. (1988). *Representation and Reality*. Cambridge, Massachusetts: The MIT Press.
- REAL ACADEMIA ESPAÑOLA (1989). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- REBEYROLLE, J. (2000). *Forme et fonction de la définition en discours*. Tesis de doctorado. Université Toulouse II - Le Mirail, France.
- REBOUL, A. (1999). “Reference, agreement, evolving reference and the theory of mental representations”. En COENE, M. (Ed.), *Traiani Augusti vestigia pressa sequamur. Studia linguistica in honorem L. Tasmowki* (pp. 601-616). Padova: Unipress.
- REINHART, T. (1983). *Anaphora and semantic interpretation*. Chicago: University press.
- RENKEMA, J. (1999). *Introducción a los estudios sobre el discurso*, MELÓN, M.L. (Trad.). Barcelona: Gedisa.
- REPISO, S. (1989). *Los posesivos*. Salamanca: Publicaciones del Colegio de España (Colección Problemas fundamentales del Español).
- RIGAU, G. (1981). *Gramàtica del discurs*. Barcelona: Universitat Autònoma de Barcelona.
- ROCA- PONS, J. (1960). *Introducción a la gramática*. Barcelona: Vergara.
- RODRÍGUEZ, C. (2004). *Metalinguistic Information Extraction from specialized texts to enrich computational lexicons*. Tesis de doctorado. Departament de Traducció i Filologia, Universitat Pompeu Fabra, Barcelona, España.
- SAGER, J. C. (1993). *Curso práctico sobre el procesamiento de la terminología*, CHUMILLAS M. L. (Trad.). Madrid: Edición Pirámide.
- SAGER, J. C. y NDI-KIMBI, A. (1995). “The conceptual structure of terminological definition and their linguistic realisations: A report on research in progress”. En *Terminology*, 2(1): 61-85.

- SAIZ, M. (2002). *Influencia y aplicación de papeles sintácticos e información semántica en la resolución de la anáfora pronominal en español*. Tesis doctoral. Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante.
- SARMIENTO, R. y ESPARZA M.A. (1993). *El pronombre*. Madrid: Editorial Sociedad General Española de Librería.
- SIERRA, G. y MCNAUGHT, J. (2000). "Design of an onomasiological search system: A concept-oriented tool for terminology". En *Terminology*, 1(6): 1-34.
- SIERRA, G. y ALARCÓN, R. (2002). "Identification of recurrent patterns to extract definitory contexts". En *Lecture Notes in Computer Science 2276*: 436-438.
- SIERRA, G., REYES, A. y AGUILAR, C. (2005). "Hacia el desarrollo de una base de conocimiento léxico para diccionarios onomasiológicos". En *Memorias del XIV Congreso Internacional del ALFAL*. Monterrey: Universidad Autónoma de Nuevo León, (formato en CD).
- SIERRA, G., ALARCÓN, R., AGUILAR, C., BARRÓN, A., BENÍTEZ, V. y BACA, I. (2006). "Corpus de contextos definitorios como una herramienta para la lexicografía y la terminología". Ponencia presentada en el *IX Encuentro Internacional de Lingüística en el Noroeste*, Hermosillo, México, del 15-17 de noviembre.
- SIERRA, G., ALARCÓN R., AGUILAR C. y BACH, C. (2008). "Definitional verbal patterns for semantic relation extraction". En *Terminology: Special Issue on Pattern Based approaches to Semantic Relations* (En prensa).
- SOTO, G. (2004-2005). "La estructuración jerárquica de la información en el discurso científico escrito: segmento de orientación y núcleo informativo". En *Lenguas Modernas*, 30: 7-24.
- SPERBER, D. y WILSON, D. (1986). *Relevance. Communication and Cognition*. Oxford: Basil Blackwell.
- TORUELLA, J. y LLISTERRI, J. (1999): "Diseño de corpus textuales y orales", En BLECUA J.M. (Ed.), *Filología e informática: Nuevas tecnologías en los estudios filológicos*. Barcelona: Editorial Milenio y Universidad Autónoma de Barcelona.
- VÁZQUEZ, V. (2004). "Algunas reflexiones sobre el cálculo de la distancia referencial". En *DELTA*, 20 (1): 27-47.
- YANG, X., ZHOU, G., SU, J. y LIM TAN, C. (2003). "Coreference Resolution Using Competition Learning Approach". En *Computational Linguistics* (pp. 176-183). Proceedings of the 41 Annual Meeting of the Association.
- WOODS, W. (1977). "Explorations in natural language question answering". En ZAMPOLLI, A. (Ed.), *Linguistic Structures Processing*. Nueva York: Elsevier.

Apéndice A

Paradigma pronominal del español

A continuación se exponen algunas tablas con el paradigma pronominal que se utiliza en este trabajo [RAE, 1985]. El inventario se presenta con el fin de rastrear cualquier pronombre empleado como anáfora en contexto lingüístico. Los pronombres marcados con gris tienen un fuerte valor anafórico, en tanto que el resto se identifican con un carácter más bien deíctico

I. Pronombres personales

Tabla B-1. Paradigma del pronombre personal

Núm	Pers	Nominativo	Acusativo	Dativo	Caso preposicional
		Sujeto	O. Directo	O. Indirecto	Tónicos de objeto
		Tónicos	Átonos	Átonos	Tónicos
Sg.	1	<i>yo</i>	<i>me</i>	<i>me</i>	<i>mí, conmigo</i>
	2	<i>tú / vos</i>	<i>te</i>	<i>te</i>	<i>ti, contigo</i>
		<i>usted</i>	--	--	<i>usted</i>
	3	<i>él / ella / ello</i>	<i>lo (le) / la / lo</i>	<i>le / (se)</i>	<i>él / ella / ello // sí, consigo</i>
Pl.	1	<i>nosotros/as</i>	<i>nos</i>	<i>nos</i>	<i>nosotros</i>
	2	<i>vosotros/as</i>	<i>os</i>	<i>os</i>	<i>vosotros</i>
		<i>ustedes</i>	--	--	<i>ustedes</i>
	3	<i>ellos/ellas</i>	<i>los / las</i>	<i>les / se</i>	<i>ellos/ ellas // sí, consigo</i>

Tabla 2-B. Ejemplos de anáfora de pronombre personal

Pronombre personal	<i>Edgar</i> ₁ y <i>Gloría</i> viven en Tepepan. <i>Él</i> ₁ todavía no conoce bien la colonia.
Neutro pronominal	<i>Ana</i> y <i>Emilio</i> se besaron ₁ . <i>Ello</i> ₁ no implica que sean novios.
Pronombre personal complemento (acusativo)	El profesor me regaló <i>un libro</i> ₁ muy interesante, no <i>lo</i> ₁ traje porque no creí necesitarlo ₁ .
Pronombre personal complemento (dativo)	Dejé un mensaje para <i>los alumnos</i> ₁ porque no voy a <i>darles</i> ₁ las calificaciones.
Pronombre de caso preposicional	<i>Sofía</i> ₁ quiere que sus amigos asistan a su cumpleaños, <i>para ella</i> ₁ es importante.
Pronombre omitido¹	<i>Alicia</i> ₁ estaba muy enferma y ⊙ tuvo ₁ que ir al doctor.

¹ Es un caso de *elipsis* pronominal, en este trabajo no es tratada como “problema de anáfora” sino como *elipsis*, otro más de los mecanismos referenciales discursivos. Sin embargo cabe señalar que se tratad e dos mecanismos referenciales muy próximos.

II. Pronombres demostrativos

Tabla 3-B. Paradigma pronombre demostrativo y anáfora de pronombre demostrativo

	Masculino	Femenino	Anáfora de pronombre demostrativo
Singular	<i>éste</i>	<i>ésta</i>	
	<i>ése</i>	<i>ésa</i>	
	<i>aquél</i>	<i>aquella</i>	
Plural	<i>éstos</i>	<i>éstas</i>	<i>Con ese te verías mejor que con aquel (uso deíctico)</i>
	<i>esos</i>	<i>esas</i>	
	<i>aquéllos</i>	<i>aquellas</i>	
	Neutros: <i>esto, eso, aquello</i>		

III. Pronombres posesivos

Tabla 4-B. Paradigma de pronombre posesivo y anáfora de pronombre posesivo

Pers	Singular		Plural	
1	<i>mío/ mía</i>	<i>nuestro/nuestra</i>	<i>míos/mías</i>	<i>nuestros/nuestras</i>
2	<i>tuyo/tuya</i>	<i>vuestro/vuestra</i>	<i>tuyos/ tuyas</i>	<i>vuestros/vuestras</i>
3	<i>suyo/suya</i>	<i>suyo/suyo</i>	<i>suyos/suyas</i>	<i>suyos/suyas</i>
Anáfora de pronombre posesivo				
<i>La ventaja₁ en el concurso es nuestra₁ (uso anafórico)</i>				
<i>Mira los míos // Entrégame el tuyo (uso deíctico)</i>				

IV. Pronombres reflexivos

Tabla 5-B. Paradigma de pronombre reflexivo y anáfora de pronombre reflexivo

Pers	Singular	Plural	Anáfora de pronombre reflexivo
1	<i>me</i>	<i>nos</i>	
2	<i>te</i>	<i>se/ os</i>	<i>Ustedes₁ se₁ bañaron antes de ir a la playa.</i>
3	<i>se</i>	<i>se</i>	<i>El barco₁ se₁ hundió en el Atlántico</i>

V. Pronombres relativos

Tabla 6-B. Paradigma de pronombre relativo y anáfora de pronombre relativo

Funciones sintáctica	Pronombre
Sustantivo	<i>quien, quienes</i>
Sustantivo y adjetivo	<i>que, cual, cuales</i>
Adjetivo	<i>cuyo, cuya, cuyos, cuyas</i>
Sustantivo, adjetivo y adverbio	<i>cuanto, cuanta, cuantos, cuantas</i>
Adverbio (sin genero ni número)	<i>donde, adonde, cuando, como</i>

Anáfora de pronombre relativo
<i>El pastel₁ que₁ estaba en la mesa sabía delicioso, fue el que₁ tuvo más éxito.</i>

VI. Pronombres interrogativos

Tabla 7-B. Paradigma de pronombre interrogativo y anáfora de pronombre interrogativo

Funciones sintáctica	Pronombre
Sustantivo	<i>quién, quiénes</i>
Sustantivo y adjetivo	<i>qué, cuál, cuáles</i>
Sustantivo, adjetivo y adverbio	<i>cuánto, cuánta, cuántos, cuántas</i>
Adverbio (sin género ni número)	<i>dónde, cuándo, cómo</i>
Anáfora de pronombre interrogativo	
¿Según el texto revisado <i>cuáles₁</i> son los <i>elementos de un contexto definitorio₁</i> ?	

VII. Pronombres indefinidos, cuantitativos y numerales

Tabla 8-B. Paradigma de pronombre indefinido y cuantitativos

Pronombre indefinidos				
<i>algo</i>	<i>misma</i>	<i>ninguna</i>	<i>poco</i>	<i>todo</i>
<i>alguien</i>	<i>mismas</i>	<i>ningunas</i>	<i>pocos</i>	<i>todos</i>
<i>alguna</i>	<i>mismo</i>	<i>ninguno</i>	<i>quienquiera</i>	<i>una</i>
<i>algunas</i>	<i>mismos</i>	<i>ningunos</i>	<i>quienesquieran</i>	<i>unas</i>
<i>algunos</i>	<i>mucha</i>	<i>otra</i>	<i>tanta</i>	<i>uno</i>
<i>bastante</i>	<i>muchas</i>	<i>otras</i>	<i>tantas</i>	<i>unos</i>
<i>bastantes</i>	<i>mucho</i>	<i>otro</i>	<i>tanto</i>	<i>varias</i>
<i>cualesquiera</i>	<i>muchos</i>	<i>otros</i>	<i>tantos</i>	<i>varios</i>
<i>cualquiera</i>	<i>nada</i>	<i>poca</i>	<i>toda</i>	
<i>demás</i>	<i>nadie</i>	<i>pocas</i>	<i>todas</i>	
Anáfora de pronombre indefinido y cuantitativo				
En la manifestación estaban <i>las mujeres inconformes₁</i> , me parece que fueron <i>muchas₁</i> .				
En la junta dieron <i>la información pertinente para hacer un informe₁</i> y como fue <i>mucha₁</i> no <i>toda₁</i> se entendió.				

Pronombres numerales: Distinguen a las entidades por medio del número exacto de aquello a lo que se alude: *uno, una, cinco, seis, diez, cien*, etc. No siempre los numerales son usados como pronombres, pueden ser determinantes de un sintagma nominal (uso adjetivo), de tal forma que cuando son pronombres sustituyen alguna entidad del discurso:

Tabla 9-B. Ejemplos de anáfora de pronombre numeral

En la farmacia había <i>cuatro</i> <i>medicinas para la gripe</i> ₁ pero sólo compré <i>dos</i> ₁ .
Vinieron <i>las amigas de mi abuela</i> ₁ , <i>una</i> ₁ era un poco presumida.

Se expuso de manera muy general el paradigma pronominal del español. En algunos casos se señaló la función que desempeñan en la oración con el fin de mostrar la flexibilidad sintáctico-pragmática de estas formas y la heterogeneidad de la categoría.

Existen pronombres mayormente identificados con el carácter anafórico, mientras otros poseen un valor principalmente deíctico. En general, los pronombres poseen ambas cualidades; la distinción existe en función del uso que se les da: son anáforas cuando refieren a un antecedente explicitado en el discurso y son deícticos cuando se resuelven en una situación de enunciación determinada.

Apéndice B

Atributos de las etiquetas XML para contextos definatorios

Etiquetas para partes de CDs:

CD	contexto definatorio
TERM	término
DEF	definición
PVD	patrón verbal definatorio
VD	verbo definatorio
SEmarc	marca de clítico “se”
Vaux	verbo auxiliar
NX	nexo
MRD	marcador reformulatorio discursivo
PTD	patrón tipográfico definatorio
PP	patrón pragmático

CÓDIGOS	ETIQUETAS	
CD		<CD num= “”> </CD>
TERM		<TERM term= “” tipterm= “”> </TERM>
DEF		<DEF tipdef= “”></DEF>
PVD		<PVD> </PVD>
	VD	<VD lema= “” args= “” mdo= “”> </VD>
	SEmarc	<SEmarc tipsemarc= “”></SEmarc>
	Vaux	<Vaux> </Vaux>
	NX	<NX tipnx= “”> </NX>
MRD		<MRD> </MRD>
PTD		<PTD pdef= “” mt= “”> </PTD>
PP		<PP tippp= “”> </PP>

Atributos

ETIQUETA	ATRIBUTOS	FUNCIÓN
CD	<ul style="list-style-type: none"> num= 1,2,3,4 num= número 	Indica los elementos que constituyen al CD. Dentro de ellos se encuentran el término, su definición, la predicación verbal y relaciones de correferencia.
TERM	<ul style="list-style-type: none"> term= L / NL tipterm= fn / fn Y fprep / fv Y fn/ term= <i>término</i> L= <i>lingüístico</i> NL= <i>No lingüístico</i> tipterm= <i>tipo de término</i> fn= <i>frase nominal</i> fn Y fprep = <i>frase nominal y frase prepositiva</i> fv Y fn= <i>frase verbal y frase nominal</i> 	En sus atributos se marca si se trata de un término lingüístico (L) o de uno no lingüístico (NL). Se toman en cuenta tres tipos de término: <i>fn</i> (frase nominal), <i>fn Y fprep</i> (frase nominal seguida de frase prepositiva) y <i>fv Y fn</i> (frase verbal seguida de frase nominal).
DEF	<ul style="list-style-type: none"> tipdef= A / F / E / S A= <i>analítica</i> F= <i>funcional</i> E= <i>extensional</i> S= <i>sinonímica</i> 	Contiene la definición del término. En ella se debe omitir cualquier texto complementario que de manera estricta no forme parte de dicha definición. Existen cuatro tipos: analítica, funcional, extensional y sinonímica.
PVD	Sin atributos	Contiene todos los componentes de una PVD: VD, clítico <i>se</i> , verbo auxiliar, verbo definitorio y nexos.
VD	<ul style="list-style-type: none"> lema= “definir” “constar” args= 2/ 3 mdo= inf / ger /part /fin lema= <i>verbo en infinitivo</i> argumentos= <i>valencia verbal</i> modo verbal= <i>inf, part, ger, fin</i> inf= <i>infinitivo</i> part= <i>participio</i> ger= <i>gerundio</i> fin= <i>conjugado</i> 	Contiene al verbo definitorio. Se marcan tres atributos: 1) atributo <i>lema</i> en el cual se indica el verbo en infinitivo; 2) el verbo tiene una valencia verbal, es decir, puede tener dos o tres argumentos o participantes; 3) se marcará el modo del VD en cuatro variantes: infinitivo (inf), participio (part), gerundio (ger), verbo conjugado (fin).
SEmarc	<ul style="list-style-type: none"> semarc= enc /prec semarc= <i>marca del clítico “SE”</i> enc= <i>enclítico</i> proc= <i>proclítico</i> 	Se indica su posición respecto al verbo. El atributo distingue entre <i>enclítico</i> (<i>enc</i>) cuando <i>se</i> es parte de la morfología verbal y está en posición final y <i>proclítico</i> (<i>proc</i>) cuando el clítico está en posición preverbal.
Vaux	Sin atributos	Contiene cualquier verbo auxiliar dentro de la PVD (p. e., <i>se puede considerar como</i> , <i>se ha</i> definido, <i>se debe</i> concebir como...)
NX	<ul style="list-style-type: none"> tipnx= Adv /Prep tipnx= <i>tipo de nexos</i> Adv= <i>adverbio</i> Prep= <i>preposición</i> 	Señala la función que cumple un adverbio o preposición entre el verbo y la definición.

<p>PP</p>	<ul style="list-style-type: none"> • tipp= Aut /Temp /Inst <p>tipp= <i>tipo de patrón pragmático</i> Aut= <i>Autoría</i> Temp= <i>Temporalidad</i> Inst= <i>Istruccional</i></p>	<p>Dan información sobre el uso de los términos. Los tres patrones considerados en este rubro son: <i>Autoría</i> (Aut) ; <i>Patrones pragmáticos temporales</i> (Temp) y <i>Patrones instruccionales</i> (Inst).</p>
<p>MRD</p>	<p>Sin atributos</p>	<p>Abarcan estructuras sintácticas con la función de explicar el propio lenguaje. Son frases que retoman algún elemento discursivo para reintroducirlo al discurso de otra forma, p.e.: <i>es decir, por ejemplo, esto es</i>, etc.</p>
<p>PTD</p>	<ul style="list-style-type: none"> • mdef= dp / viñ / par / gui / cl /otr • mt = neg / curs / subr, otr <p>mdef= <i>marcadores definitorio</i> dp= <i>dos puntos</i> viñ = <i>viñeta</i> par = <i>paréntesis</i> gui= <i>guiones</i> cll= <i>comillas</i> otr= <i>otros</i> mt= <i>marcadores tipográficos</i></p>	<p>Señala cualquier signo de puntuación o <i>marcadores tipográficos definitorios</i> (MTD). Se distingue en dos tipos: 1) <i>marcadores definitorios</i> (mdef): unen a un término con su definición, sustituyendo o complementando la función de la PVD. En los atributos se señalan como mdef= dp, viñ, par, gui, cl, otr 2) <i>marcadores tipográficos</i> (mt): indicación de negritas, cursivas, subrayado y otras marcas que dan prominencia al término definido o a la definición, este caso se marca mt= neg,curs,subr,otr.</p>