

ESTUDIO DE LA INCIDENCIA DE DIFERENTES FUENTES DE INFORMACIÓN EN EL ESTABLECIMIENTO DE RELACIONES ANAFÓRICAS

Celia Rico Pérez
Departamento de Inglés
Universidad de Alicante
fing@ealiun11

Resumen

El trabajo que aquí se presenta es un estudio de corpus que tiene básicamente dos objetivos; en primer lugar, delimitar el tipo de conocimiento pertinente para la identificación de antecedentes en las relaciones anafóricas; en segundo lugar, descubrir cómo y en qué medida actúa cada uno de los tipos de conocimiento identificados.

A lo largo del estudio intentamos poner de manifiesto que la información que aporta cada tipo de conocimiento es siempre relativa y que la asignación del antecedente correcto a una expresión anafórica necesita la integración de múltiples fuentes de conocimiento en una única estrategia. Este estudio pretende, por tanto, servir de punto de partida para la creación de un modelo de interpretación de las relaciones anafóricas que coordine todas estas fuentes de información.

1. Introducción

La identificación de antecedentes de las expresiones anafóricas es un problema en el que todavía quedan muchos aspectos por resolver. Son también muchas las soluciones propuestas a este respecto pero la anáfora es un fenómeno muy complejo en el que intervienen factores muy diversos. Hay casos de relaciones anafóricas muy simples en los que la identificación del antecedente es inmediata pero existen, igualmente, otros casos donde esta identificación resulta muy difícil. Esto se debe fundamentalmente a que los criterios utilizados en un caso no son válidos para otro, o son simplemente rechazados por criterios distintos que se imponen con más fuerza. Veamos, por ejemplo, la relación anafórica que se establece en el siguiente texto:

El dolor recorría el brazo de Curt desde la mano hasta el hombro. Jess parecía no darse cuenta de que lo habían golpeado.

La relación anafórica se construye aquí entre el pronombre *lo* y su antecedente *Curt*. Sin embargo, puede ocurrir que esta relación sea incorrecta si el discurso se modifica del siguiente modo:

Curt lanzó un derechazo justo al abdomen de Jess. Fue como pegarle a un saco de sal.

El dolor recorría el brazo de Curt desde la mano hasta el hombro. Jess parecía no darse cuenta de que lo habían golpeado.

Éste es quizá un ejemplo extremo de la influencia del contexto en la identificación de antecedentes anafóricos pero sirve para mostrar cómo la interpretación de las expresiones anafóricas necesita tener en consideración aspectos y fuentes de información diversos que, en general, superan el marco de la oración.

Los supuestos básicos sobre los que se construye el presente estudio son los siguientes:

- se considera la anáfora en su dimensión discursiva, es decir, por encima de los límites de la oración.
- no existe una única fuente de información que pueda ser utilizada para todos y cada uno de los casos y tipos de anáfora.
- para los seres humanos no existen relaciones anafóricas ambigüas debido fundamentalmente a que son capaces de coordinar múltiples fuentes de conocimiento y asignar a cada una de ellas la pertinencia necesaria.

Tomando estos supuestos como punto de partida, el estudio que aquí se presenta analiza una serie de relaciones anafóricas en un corpus, con el doble objetivo de (a) descubrir el tipo de conocimiento que se necesita para el establecimiento de relaciones anafóricas y (b) delimitar la influencia que tiene cada uno de los tipos identificados.

2. Tipos de expresión anafórica analizados

Analizaremos únicamente relaciones endofóricas, es decir, aquellas que se establecen dentro del texto. Se eliminan, por tanto, las relaciones anafóricas que crean referencia a entidades extratextuales. El hecho de eliminar de nuestro análisis las relaciones exofóricas responde a la propia naturaleza del presente estudio, que se centra en el análisis empírico de datos observables. Esto es así porque solamente es posible obtener datos cuantitativos si el antecedente y la expresión anafórica son directamente observables en el texto.

Así pues, distinguimos dos tipos de expresiones:

- 1- las anafóricas, cuyo antecedente se puede localizar dentro del texto porque éste existe en el universo lingüístico.
- 2- las referenciales, cuyo antecedente no puede localizarse dentro del texto porque éste existe en el universo extralingüístico.

Las expresiones de las que se ocupa el presente estudio son las anafóricas y de ellas seleccionamos las relaciones anafóricas de tipo nominal, es decir, aquellas que se establecen entre una expresión anafórica y un grupo nominal que actúa como antecedente¹.

¹ Como se verá, las definiciones y ejemplos de las expresiones anafóricas están en inglés porque el corpus estudiado contiene textos en inglés. La autora hubiera deseado poder realizar este estudio sobre un corpus de textos del español pero no conocía, en el momento de llevar a cabo el estudio, un corpus del español de libre circulación entre universidades y que tuviera fines no lucrativos.

Las expresiones que mantienen este tipo de relación que son objeto de nuestro análisis son las siguientes:

a) Descripciones definidas: son frases nominales introducidas por el artículo *the*.

A shielded thermocouple was used to measure the upstream of the transpiring gas. It was exposed to a high velocity gas jet. A plug and a tube with holes in its cylindrical walls divided the chamber above the porous plug into two parts. This arrangement had the purpose to prevent heated gas to reach the thermocouple by natural convection.

[J02:0890d a J02:0940b]²

El grupo nominal *a shielded thermocouple* introduce en el discurso una nueva entidad que se constituye, desde el momento de su mención, en un "posible candidato" para referencias anafóricas posteriores. Así, seis líneas después, encontramos la descripción definida *the thermocouple* que, como expresión anafórica, establece lazos anafóricos con el citado grupo nominal.

b) Descripciones introducidas por un demostrativo: esta estructura se construye mediante el empleo de un demostrativo, *this, that, these* o *those* y una frase nominal.

The jury said it did find that many of Georgia's registration and election laws are outmoded or inadequate and often ambiguous. It recommended that Fulton legislators act to have these laws studied and revised.

[A01:0130g h A01:0160i]

En la línea 3 la descripción definida *these laws* establece una relación anafórica con el antecedente introducido en la línea 1, *many of Georgia's registration and election laws*.

c) Pronombres personales:

It took me a moment to realize what was odd about that panel: there was a gimbaled compass welded to it.

[G04:0860e a G04:0870n]

d) Pronombres posesivos:

Gray Eyes remained erect. The feathered lance was still above his head.

[N04:0980m a N040990k]

² Esta numeración corresponde a la línea del corpus de donde se ha extraído la información.

3. El corpus

Un análisis como el que se propone necesita un corpus del que extraer la información, un corpus del que puedan obtenerse datos empíricos sobre cómo se establecen los vínculos que unen una expresión anafórica concreta a su antecedente.

El corpus seleccionado para realizar el estudio que presentamos es *SUSANNE* (G. Sampson 1992). *SUSANNE* es un inventario de 128.000 palabras extraídas del *Brown Corpus of American English* (Kucera y Francis, 1967). Consta de 64 archivos, cada uno con una versión anotada de un texto de aproximadamente 2000 palabras del *Brown Corpus*. *SUSANNE* tiene alrededor de 5.3 megabytes y los textos que lo componen incluyen las siguientes categorías: artículos de prensa, biografías y memorias, textos científicos y técnicos, y literatura. Podemos decir, por tanto, que *SUSANNE* es una muestra representativa de la lengua escrita.

Las 128.000 palabras de los textos almacenados en *SUSANNE* están etiquetadas según su categoría gramatical y las oraciones están analizadas sintácticamente. Estos dos tipos de anotaciones (gramatical y sintáctica) hacen de *SUSANNE* una herramienta muy valiosa para el análisis que proponemos. Gracias a ello, la identificación de las unidades de análisis (las expresiones anafóricas) y el establecimiento de relaciones anáfora-antecedente se hace de manera más sencilla.

Para llevar a cabo el análisis se definió un subcorpus en el que se estudiaron 674 relaciones anafóricas. Sobre esta muestra se investigó la incidencia de determinados tipos de información en el establecimiento de relaciones anafóricas, reuniendo, posteriormente, datos concluyentes con respecto a estas fuentes de información, su frecuencia, distribución, así como sobre la relación que mostraban entre ellas.

4. Fuentes de información necesarias para la identificación de antecedentes.

Para determinar los criterios mínimos que permitieran identificar los antecedentes anafóricos intentamos restringir los conocimientos necesarios a aquellos que provienen de fuentes lingüísticas y así, limitar en lo posible la utilización de información que necesitase conocimiento del mundo³.

La información que se analiza para cada expresión anafórica y su posible candidato a antecedente es la siguiente:

- información sobre rasgos gramaticales: se estudia el género, número y persona tanto del antecedente como de la expresión anafórica.
- información sobre rasgos sintácticos: se codifica la función que antecedente y anáfora tienen en la oración.

³ Esto se hizo así dado el estado de las investigaciones actuales con respecto a la formalización del conocimiento del mundo y su dificultad para integrarlo en cualquier estrategia para el tratamiento del lenguaje natural.

- información sobre rasgos semánticos: se codifican las posibles relaciones semánticas entre antecedente y expresión anafórica. Las relaciones identificadas en el corpus son las siguientes:

- sinonimia
- hiperonimia
- una parte del todo
- generalización
- suma de dos o más antecedentes
- inferencia sobre conocimiento del mundo

- información sobre rasgos pragmáticos⁴:

- distancia desde la expresión anafórica al antecedente.
- competencia o número de candidatos a antecedente
- prominencia del antecedente en el texto.

- información sobre el tipo de expresión anafórica

5. Incidencia de las diferentes fuentes de información.

Analizamos en este apartado cada una de las fuentes de información que intervienen en el establecimiento de las relaciones anafóricas. Veremos así mismo, cómo influye cada fuente, extrayendo reglas generales de comportamiento y mostrando igualmente, que los valores que asignan a los potenciales antecedentes no pueden considerarse absolutos ni totalmente definitivos.

5.1. Información gramatical.

La información gramatical extraída de cada una de las expresiones anafóricas y sus antecedentes indica que, como regla general, todas las expresiones anafóricas deben mostrar concordancia en género, número y persona con su antecedente. Existen, sin embargo, algunos casos en los que la concordancia no se mantiene:

a) nombres colectivos. Pueden aceptar una referencia en singular o en plural. Esto puede generar problemas a la hora de asignar un antecedente y, por tanto, es un rasgo que debe tenerse presente.

b) casos de generalización. En estos casos el antecedente tiene el rasgo singular y la expresión anafórica el de plural:

T1 *It was possible to make estimates of the quantum yield by observing the extent of reduction of a uranyl oxalate actinometer solution illuminated for a known time in a typical reaction cell... When typical reaction cells to which 0.3 of an atmosphere of oxygen had been added were illuminated, chlorine and phosgene were produced.*

[J06:1530n - J06:1630e]

⁴ Estos rasgos están tomados de M. Ariel (1990), quien los utiliza para desarrollar la teoría de la accesibilidad sobre los antecedentes anafóricos.

c) relación del grupo al individuo. El antecedente está en plural y la expresión anafórica en singular:

T2 *The contents of the manifold for liquid phase experiments were then mixed by shaking, redistributed to the reaction tubes, frozen down, and each tube was then sealed off.*
[J06:0740h - J06:0760n]

d) suma de varios antecedentes en singular que da como resultado una expresión anafórica en plural:

T3 *A weapons carrier took Greg, Todman, Banjo, Ferguson and Walters the two miles from the bivouac area to the strip. It was a rough long ride through the mud and the pot holes. The sky glowered down at them.*
[N15:0110e - N15:0140t]

Como podemos comprobar por los ejemplos expuestos, la formulación de una regla para la identificación de antecedentes debe dar cuenta no sólo de los casos generales donde antecedente y expresión anafórica comparten rasgos de género, número y persona, sino también de la posibilidad de que no exista tal concordancia debido a alguno de los cuatro casos expuestos. Si la regla no tuviera en cuenta todo esto, la asignación de antecedentes sería incorrecta porque o bien nunca se encontraría el antecedente (textos T1 y T2) o éste sería incorrectamente identificado (texto T3 asignaría *pot holes* como antecedente de *them*).

5.2. Información sintáctica.

La identificación de la función sintáctica que antecedente y expresión anafórica tienen en la oración se utiliza en dos sentidos:

- para señalar posiciones prominentes en el discurso
- para analizar el papel que tienen las estructuras sintácticas paralelas en la identificación de antecedentes.

Con respecto al primer punto, veremos en el apartado 5.6. la influencia que tienen en la asignación de antecedentes las posiciones prominentes marcadas por determinadas funciones sintácticas. Comprobaremos entonces cómo la función sintáctica del antecedente y la expresión anafórica permiten en muchos casos resolver la ambigüedad entre varios antecedentes, gracias a que el antecedente correcto se halla en una posición más prominente.

En relación al papel que pueden tener las estructuras sintácticas paralelas, vemos en el ejemplo que se muestra a continuación cómo antecedente y anáfora tienen posiciones equivalentes en la oración y esto contribuye a una identificación inequívoca de la relación que existe entre ellos:

T4 *Their gait is impossible to convey in words. It has nothing of the proud stride of the trained runner. It is not a lope. It is not done with style or verve.*
[G04:0180d - G04:0200h]

Hay casos, sin embargo, en los que el empleo de las estructuras paralelas en la identificación de antecedentes es más problemático:

T5a *There was also a dog. Its ribs showed, it was a yellow nondescript color, it suffered from a variety of sores, hair had scabbed off its body in patches. It lay with its head on its paws.*
[G04:1020j - G04:1050j]

En T5a los antecedentes posibles a la expresión anafórica *it* pueden ser: *dog*, *color*, *hair*, *body*. Podría recurrirse a diferentes tipos de información para eliminar cada uno de los candidatos erróneos, pero la información más inmediata es la que considera que *it* sólo puede tener como antecedente a *dog* puesto que las posiciones de sujeto de las oraciones anteriores están ocupadas por el mismo tipo de expresión anafórica (*it*) y su antecedente es el mismo (*dog*). Sólo una posición de sujeto no está ocupada por el pronombre *it* y ésta se encuentra en la frase *hair had scabbed off its body in patches* donde *hair* introduce una nueva entidad en posición de sujeto. Existen, sin embargo otras fuentes de información que refuerzan la elección de *dog* como antecedente por encima de *hair*. Las fuentes de información que apuntan hacia *dog* como candidato ideal son por una parte la posición temática de *dog* y , por otra, la información semántica que se desprende de la mención de *head* (cabeza) y *paws* (patas).

Sería posible, de todos modos, deshacer la influencia de las mencionadas fuentes y forzar una referencia a *hair*, si cambiáramos el tipo de información semántica:

T5b *It lay scattered about on the ground in tufts.*

En este caso *tufts* (mechones) introduce un tipo de información semántica que fuerza la reinterpretación de la expresión anafórica *it* y asigna como antecedente *hair*.

En cualquier caso no debe olvidarse que la relación que los hablantes nativos⁵ identifican como la más natural es la que se establece entre el pronombre *it* y *dog*.

Una vez analizados algunos de los ejemplos más característicos, podemos concluir que la regla que dé cuenta de la influencia de la estructura sintáctica de las oraciones que contienen el antecedente y la expresión anafórica debe considerar siempre que los candidatos seleccionados necesitan la confirmación de otras fuentes de información, y que nunca debe señalar valores absolutos porque, como hemos visto, pueden llevar a resultados erróneos.

5.3. Información semántica.

Consideramos como regla general que siempre debe existir consistencia semántica en la relación anafórica. Así pues, definimos tres tipos de consistencia semántica:

- consistencia entre el antecedente y la expresión anafórica.
- consistencia entre el verbo y la expresión anafórica.
- consistencia dentro de la oración en la que se encuentra la expresión anafórica.

En el primer caso, la consistencia semántica entre antecedente y expresión anafórica se mantiene mediante el empleo de sinónimos, hiperónimos, relaciones del todo a una parte, del grupo al individuo y de generalización. Este tipo de consistencia semántica se manifiesta normalmente en aquellas expresiones anafóricas que tienen mayor carga léxica (descripciones definidas y descripciones introducidas por un demostrativo):

T6 *Reverend Jason got an Indian, the Canadians the other. I saw the clergyman kneel*

⁵ Agradecemos, en este sentido, la colaboración del profesor Bryn Moody del departamento de inglés de la Universidad de Alicante, quien actuó como informante en los casos más problemáticos.

for a moment.

[N04:1440i - N04:1450p]

En el caso de la consistencia entre el verbo y la expresión anafórica mostramos a continuación un ejemplo claro del efecto que el cambio de un verbo puede tener sobre la asignación de antecedentes:

T7a *The jury recommended that Fulton legislators act to have these laws studied and revised to the end of modernizing and improving them.*

[A01:0150g - A01:0170c]

En este ejemplo la identificación de *these laws* (estas leyes) como antecedente de *them* se hace gracias a la necesidad de mantener la consistencia semántica con los verbos *improve* (mejorar) y *modernize* (modernizar). Por el contrario, si en lugar de estos verbo introducimos el verbo *fulfil* (cumplir) y construimos la oración como:

T7b *The jury recommended that Fulton legislators act to have these laws studied and revised to the end of fulfilling the task assigned to them.*

el antecedente del pronombre debe ser entonces *Fulton legislators*.

Veamos ahora el caso de consistencia semántica dentro de la oración que contiene a la expresión anafórica. Estos son los casos más complejos porque necesitan el conocimiento del mundo para la identificación del antecedente y este tipo de conocimiento es difícil de codificar.

Así, por ejemplo, en T8 sólo puede identificarse como antecedente de la expresión *its operating procedures* (procedimientos operativos) a *the court* (el tribunal) si sabemos que *the jury* (el jurado) no tiene *operating procedures*.

T8 *The jury said it found the court has incorporated into its operating procedures...*

[A01:0460a - A01:0470c]

De cualquier modo, las relaciones anafóricas identificadas en el corpus en las que el criterio semántico de inferencia es el definitivo para identificar el antecedente son muy poco numerosas (alrededor de un 6% del total de expresiones analizadas).

Por último, señalaremos que la influencia de la regla de consistencia semántica es definitiva para la selección inequívoca del antecedente en una gran mayoría de casos, sobre todo cuando la expresión anafórica tiene gran carga léxica. Hay también, sin embargo, casos donde la consistencia semántica se neutraliza y no aporta ninguna información concluyente. Es entonces cuando deben actuar otras fuentes de información.

T9 *The arc started to strike the anode holder instead of the porous graphite plug and that it became highly unstable.*

[J02:1290k - J02:1310f]

En T9 los posibles antecedentes de *it* son *arc*, *anode holder* y *porous graphite plug*. Los tres candidatos mantienen consistencia semántica con la expresión anafórica por lo que no es éste un criterio que ofrezca una solución definitiva para este ejemplo. La única clave que podríamos tener sería considerar la posición temática de *arc*.

5.4 Información sobre la distancia.

Se trata, en este caso, de medir la distancia que existe entre la expresión anafórica y el antecedente con el cual se establece la relación anafórica. Para ello se considera el número de oraciones que separan la expresión anafórica de su antecedente. Se muestra en la tabla 1 la distribución de las expresiones anafóricas según la distancia a la que se encuentra el antecedente.

Distancia	Expresión anafórica			
	Descripción definida	Descripción demostrativo	Pronombre personal	Pronombre posesivo
0	15	2	83	77
1	94	34	126	34
2	24	6	16	4
3	105	3	44	7
TOTAL	238	45	269	122

- 0 = el antecedente se encuentra en la misma oración,
- 1 = el antecedente se encuentra una oración antes,
- 2 = el antecedente se encuentra dos oraciones antes,
- 3 = el antecedente se encuentra más de dos oraciones antes

Los porcentajes obtenidos aparecen en el gráfico 1.

Como puede observarse, las expresiones anafóricas que admiten mayor distancia con su antecedente son las descripciones definidas (44.1% casos con valor 3), seguidas de las descripciones introducidas por un demostrativo (75.5% casos con valor 1), los pronombres personales (46.8% casos con valor 1) y, por último, los pronombres posesivos (63.1% casos con valor 0).

El hecho de que descripciones definidas y descripciones con demostrativo sean las expresiones anafóricas que permiten mayor distancia con su antecedente se debe a que son precisamente éstas las expresiones que tienen mayor carga léxica y, por tanto, el antecedente es también más fácilmente identificable, aunque la distancia que lo separa

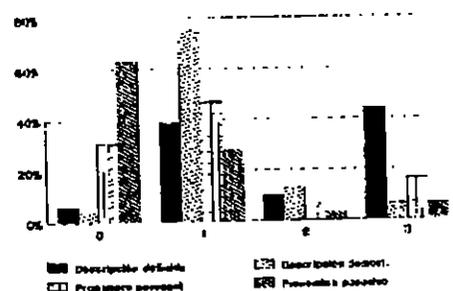


Gráfico 1

de su anáfora sea muy grande⁶.

La conclusión que puede extraerse a la vista de los datos es que mientras los pronombres permiten, en general, muy poca distancia con su antecedente, las descripciones admiten hasta más de dos frases de distancia. Puede entonces considerarse el rasgo de distancia como el mecanismo de control que impide que el antecedente de una expresión anafórica se busque más lejos de lo que realmente le corresponde.

Ocurre, sin embargo, que en un porcentaje relativamente elevado de casos (16.3%) los pronombres personales identifican su antecedente a más de dos oraciones de distancia. Esto es posible cuando entre la expresión anafórica y el antecedente no se encuentra ningún otro antecedente que pueda competir⁷.

Así pues, podemos asignar dos usos al rasgo de distancia:

- en el caso de los pronombres identifica como antecedente correcto aquel que se encuentre más próximo, siempre que no exista ningún otro criterio que impida esta identificación.
- en el caso de las descripciones el antecedente puede buscarse mucho más lejos.

Conviene recordar que la distancia, como el resto de los criterios de identificación aquí expuestos, no ofrece nunca valores absolutos para identificar antecedentes y que su papel es siempre relativo a otros criterios que puedan ejercer mayor influencia. Lo que sí es cierto, es que la distancia evita, en muchas ocasiones, realizar una búsqueda demasiado exhaustiva.

5.5. Rasgo de competencia

Este rasgo permite decidir sobre el comportamiento de las expresiones anafóricas dependiendo de si tienen un único antecedente, dos antecedentes o más de dos antecedentes. Actúa junto al rasgo de distancia y permite asignar antecedentes en los casos más simples. Estos son casos donde la información que proviene de otras fuentes no es pertinente:

T10 *She had been sentenced to 180 years in prison, but former Gov. Stratton commuted her term to 75 years.*
[A20:1660e - A20:1670m]

Este tipo de casos muestra el paradigma de la situación ideal en la que el antecedente se identifica sin problemas.

⁶ M. Ariel explica este hecho por la accesibilidad del antecedente de manera que cuanto más accesible sea el antecedente mayor distancia se le permite.

⁷ Veremos en el siguiente apartado la influencia del criterio de competencia.

5.6 Rasgo de prominencia.

Mencionábamos en 5.2. la importancia que tiene en la identificación de antecedentes el análisis de las posiciones prominentes⁸ que pueden tener las entidades discursivas a lo largo del texto. La asignación del rasgo de prominencia a una entidad se realiza dependiendo de dos factores:

- la función sintáctica que la entidad tenga en la oración.
- la repetición que la entidad muestre a lo largo del texto.

Las funciones sintácticas que marcan mayor prominencia en inglés se ordenan del siguiente modo: en primer lugar el sujeto, a continuación objeto directo, objeto indirecto, otros complementos subcategorizados y en último lugar los circunstanciales regidos por el verbo.

Con respecto al criterio de repetición o recurrencia apuntaremos que en muchas ocasiones es este criterio el que permite la identificación de un antecedente que, a pesar de no estar en posición sintáctica prominente, consigue, por repetición, hacerse prominente:

T11 ... *he knew Donovan would keep his back to the strip... Greg rumbled down the rough metal strip... and the clean roar of the engine that took him off the rough strip... the metal strip they had taken off from...* [N15:0330a - N15:0480m]

En T11 desde la primera mención de *the strip* hasta la última hay once oraciones y dos párrafos de distancia y aún así es posible identificar la última mención con la primera, precisamente gracias a la repetición de la entidad a lo largo del texto.

La regla general sobre el rasgo de prominencia predice que las expresiones anafóricas encuentran su antecedente en posiciones prominentes siempre que los criterios de concordancia gramatical y consistencia semántica lo confirmen. Así ocurre en T12, donde la regla de prominencia selecciona correctamente *this temperature* como antecedente de *it* a pesar de que los antecedentes posibles son tres: *environmental temperature*, *anode holder* y *radiation*.

T12 *This temperature was taken as environmental temperature to which the anode holder was exposed as far as radiation is concerned. It is sufficiently small compared with the surface temperature of the anode holder.* [J02:1610c - J02:1690a]

Sin embargo, como hemos visto en otros casos la regla tiene una aplicación relativa y necesita la confirmación de otras fuentes de información, que incluso pueden rechazar el candidato propuesto por el rasgo de prominencia.

T13 *The mean temperature of this region was approximated by the temperature measured halfway between the edge of the hot spot and the rim of the plug. The mean temperature of the surface was then computed according to the following relation: <formula> where x is the fraction of the plug area covered by the hot spot.* [J02:1410h - J02:1470f]

⁸ Grosz y Sidner (1986) proponen un modelo de interpretación anafórica basado en la localización del foco del discurso y en la creación de un mecanismo de selección y actualización del mismo. El rasgo de prominencia, tal como se interpreta en este estudio, comparte con este modelo la idea de que las entidades focalizadas son candidatos con más probabilidades de ser antecedentes.

En este ejemplo la posición prominente es rechazada como antecedente en favor de otra menos prominente debido al criterio de consistencia semántica entre antecedente y expresión anafórica.

5.7. Información sobre el tipo expresión anafórica.

Analizamos, a continuación, la relación que existe entre el tipo de expresión anafórica utilizada y la incidencia de las fuentes de información. Es importante conocer esta relación porque según el tipo de expresión anafórica que estemos tratando, la información necesaria para la identificación del antecedente será diferente también.

Mostramos en el gráfico 2 el grado de incidencia de cada una de las fuentes de información según la expresión anafórica de que se trate.

Como puede observarse existe una clara distribución del tipo de información entre descripciones y pronombres. Para las descripciones, el criterio que mayor incidencia tiene es el semántico y dentro de él, la sinonimia. En los pronombres el criterio con mayor incidencia es la distancia, seguido de la información que aporta el rasgo de prominencia.

La influencia de la información semántica en la identificación de antecedentes de las descripciones se explica precisamente por las condiciones en las que se usan este tipo de expresiones anafóricas:

- si hay más de un posible antecedente;
- si hay distancia excesiva entre antecedente y anáfora;
- si la posición del antecedente no es prominente y no ha tenido ningún refuerzo a lo largo del discurso.

En todos estos casos la identificación del antecedente es más problemática y sólo la utilización del criterio de consistencia semántica permite seleccionar el antecedente correcto. En los pronombres, sin embargo, el criterio con más influencia es el de distancia. Este rasgo de distancia actúa cuando ningún otro rasgo se impone con fuerza y selecciona como antecedente aquel que se encuentra a menos distancia de la expresión anafórica y mantiene la concordancia gramatical. Sin embargo, cuando hay más de un posible antecedente el rasgo de distancia no es válido y es entonces cuando actúan rasgos como la prominencia o la información del verbo. Lo que ocurre realmente con los pronombres es que como expresiones anafóricas incorporan poca información sobre su potencial antecedente y, por tanto, cuando hay varios antecedentes posibles, necesitan otras fuentes de información que ayuden a la identificación del antecedente correcto.

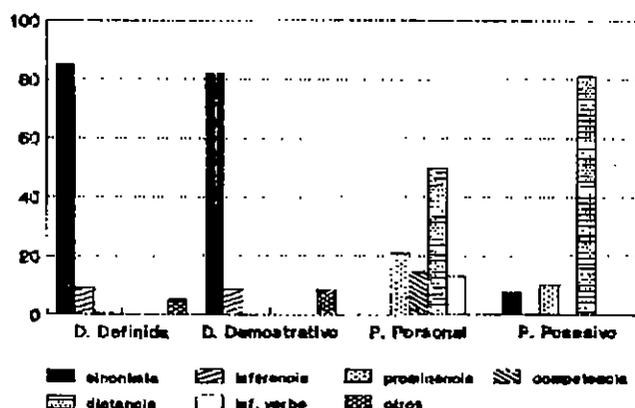


Gráfico 2

6. Conclusiones

El análisis de corpus presentado en esta comunicación nos ha permitido descubrir las fuentes de información necesarias para la identificación de antecedentes anafóricos así como su incidencia en las diferentes expresiones anafóricas. Los datos extraídos le dan a este estudio un carácter explicativo y predictivo al mismo tiempo. Gracias a este estudio podemos saber cuáles son las claves para la asignación de antecedentes anafóricos, cómo actúan estas claves, en qué casos y con qué incidencia, y de este modo, explicar con detalle los mecanismos para el establecimiento de relaciones anafóricas. Del mismo modo, es posible predecir el tipo de expresión anafórica que se utilizará para determinados antecedentes dependiendo del tipo de información con la que se cuente en cada caso.

Como hemos podido comprobar a lo largo del presente estudio, la tarea de asignación de antecedentes a las expresiones anafóricas es verdaderamente compleja. En efecto, para llegar a la identificación correcta de antecedentes son necesarias muchas y diversas fuentes de conocimiento y en cualquier caso, ninguna de ellas aporta información totalmente concluyente sobre el antecedente. En la valoración de los posibles candidatos a antecedente siempre actúan varios tipos de conocimiento y no puede decirse que exista un único criterio que supere en eficacia a los demás. Así pues, sólo de la coordinación de todas las fuentes de información es posible una aproximación real al problema de la anáfora. Por tanto, cualquier modelo que quiera explicar la anáfora en toda su extensión deberá tener en cuenta todas estas fuentes de conocimiento y conseguir integrarlas en una única estrategia. Sólo la conjunción de todos los rasgos estudiados permitirá una aproximación correcta al establecimiento de relaciones anafóricas. Esto es lo que proponen autores como Carter (1990) y, tras el estudio aquí presentado, pensamos que esta es la estrategia correcta.

7. Referencias

Ariel, M. (1990): *Accessing Noun-Phrase Antecedents*. Londres: Routledge.

Carter, D. (1990): 'Control Issues in Anaphor Resolution', en *Journal of Semantics* 7(4), pp. 435-454.

Kucera, H. y Francis, W. (1967): *Computational Analysis of Present-Day American English*, Providence: Brown University Press.

Grosz, B. y Sidner, C. (1986): 'Attentions, Intentions and the Structure of Discourse', en *Computational Linguistics* 12(3), pp. 175-204.

Sampson, G. (1992): The Susanne Corpus, release 1.

