

Algoritmo de resolución de la anáfora pronominal en diálogos

Patricio Martínez-Barco

Grupo de Investigación en Procesamiento del Lenguaje y Sistemas de Información.

Departamento de Lenguajes y Sistemas Informáticos.

Universidad de Alicante

patricio@dlsi.ua.es

Resumen

En este trabajo se presenta un algoritmo para la resolución de la anáfora pronominal en diálogos para el castellano. Este algoritmo combina dos aproximaciones diferentes. Por una parte, hace uso de información lingüística para aceptar o rechazar el antecedente de la anáfora, mientras que por otra parte, usa información derivada de la estructura del diálogo para proponer dicho antecedente.

1 Introducción

Desde que Grosz y Sidner (1986) proponen su *teoría de la estructura del discurso* (como la composición de tres estructuras, la *lingüística*, la *intencional* y la *atentiva*) se origina una serie de trabajos aplicados a la resolución de la anáfora que se basan en esta idea. Entre otros, Rocha (1998) justifica la necesidad de llevar a cabo una segmentación del discurso donde cada segmento se asocia con un determinado conjunto de elementos con el fin de estudiar correctamente las relaciones anafóricas. Abundando en lo anterior, Grosz et al. (1995) definen la teoría del *centering* para relacionar los focos de atención, la elección de expresiones relacionadas y la coherencia que existe entre las intervenciones dentro de un segmento del discurso. Otros autores como Manabu y Kouji (1996) han mejorado esta teoría permitiendo su uso en oraciones complejas. Sin embargo, Walker (1997) apunta la idea de que el área de la teoría del *centering* no debería estar limitada sólo al ámbito del segmento, sino que esta teoría debería estar integrada en un modelo de la estructura global del discurso.

Entre las implementaciones que usan información sobre los focos del discurso destacamos la realizada por Azzam et al. (1998), quienes aplicaron estos métodos para la resolución de la anáfora en su sistema de

extracción de información LASIE. Sin embargo, los resultados que se obtuvieron no fueron satisfactorios puesto que no mostraron un incremento significativo sobre una aproximación basada en heurística mucho más simple. Los autores plantean que los resultados no son satisfactorios debido a la dependencia que tiene el analizador parcial sobre la búsqueda del foco, así como por la suposición teórica que mantienen las aproximaciones basadas en foco, por la cual el foco siempre debe ser un antecedente favorecido. Y ocurre que esto no siempre es cierto. Por ello, Azzam et al. (1998) sugieren una aproximación combinada. Dentro de las aproximaciones combinadas destacamos la realizada por Mitkov (1996) quien define un método que combina fuentes de información lingüística clásicas con la información de la estructura del discurso.

Con todo lo anterior, consideramos que la definición y segmentación de la estructura del discurso (en nuestro caso el diálogo) ayuda a la resolución de la anáfora como demostraremos en este trabajo. Por tanto presentaremos, por un lado, una propuesta de estructuración del diálogo que nos proporcionará el espacio adecuado de accesibilidad de la anáfora en el diálogo, y por otra parte, la información lingüística necesaria para aceptar o rechazar los antecedentes propuestos en el paso anterior.

En la siguiente sección introduciremos el marco global donde el algoritmo propuesto ha sido incluido, posteriormente definiremos el algoritmo para la resolución de la anáfora, y finalmente mostraremos una aplicación de la propuesta sobre el análisis del diálogo.

2 Un marco global para la resolución de la anáfora en diálogos en castellano.

Anteriormente hemos podido constatar que el uso exclusivo de información de la estructura del discurso (en concreto de la información del

foco) genera un alto índice de anáfora mal resuelta o no resuelta. Este fracaso es debido principalmente a la existencia de múltiples casos que no se ajustan a esta idea. Por otra parte, el uso de métodos lingüísticos clásicos sin consideración de la información que genera la estructura del discurso tampoco resulta válido para la resolución de la anáfora en un texto dialogado.

Por lo tanto, en este trabajo proponemos un marco válido para la resolución de la anáfora en diálogos. Para ello integraremos las dos aproximaciones anteriores. En primer lugar, y basándonos en la aproximación lingüística, definiremos las restricciones y las preferencias basándonos en el uso de información morfológica, léxica y sintáctica. Esta información será extraída del análisis sintáctico del corpus así como mediante la inclusión de información semántica basada en el uso de herramientas como WordNet o EuroWordNet para el castellano (método lingüístico). En segundo lugar, definiremos un sistema de tópicos del discurso (que de ahora en adelante llamaremos sistema de tópicos) capaz de generar información de la estructura del discurso válida para la alimentación de un sistema de resolución de la anáfora.

En la figura 1 se muestra una visión global del marco propuesto para la resolución de la anáfora. Esta propuesta consiste en tres

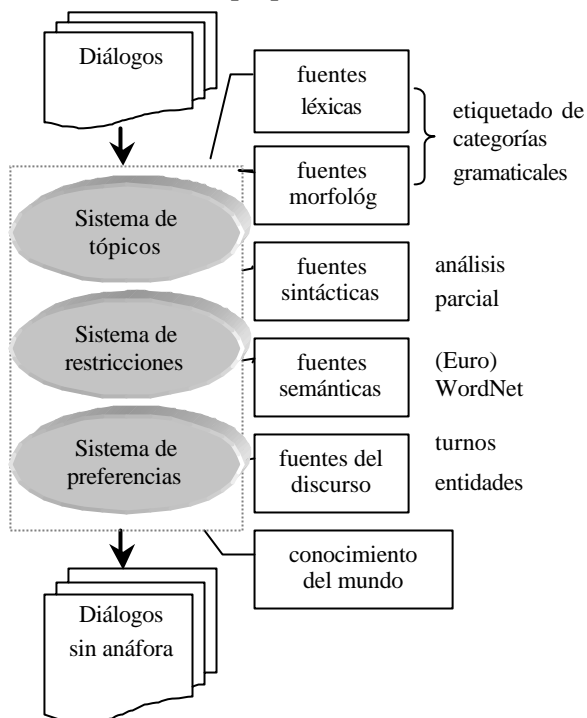


Fig. 1. Marco global para la resolución de la anáfora en diálogos.

sistemas: un sistema de restricciones, un sistema de preferencias y un sistema de tópicos. Estos tres sistemas interactúan de la siguiente forma:

?? El **sistema de restricciones** contiene las reglas necesarias para la elección de un posible antecedente para una determinada anáfora. El fallo de una de las restricciones por un antecedente causa su rechazo como posible solución de la anáfora. Un buen sistema de restricciones debe basarse siempre en una efectividad de la regla del 100%. Por este motivo, la información que se usa para la construcción de dichas reglas debe ser siempre cierta. En otro caso, los candidatos válidos podrían quedar eliminados desde el principio.

?? El **sistema de preferencias** no se usa para eliminar candidatos sino para desambiguar cuando hay varios. La información que proporciona el sistema de preferencias no tiene una efectividad del 100% aunque tiene un alto índice de veracidad. Por lo tanto, dicha información no se aplica cuando hay un único candidato, sino cuando el sistema de restricciones ha generado más de un antecedente. Por la naturaleza de ambos sistemas, deberemos aplicar en primer lugar, el sistema de restricciones, y después el sistema de preferencias.

?? El **sistema de tópicos** opera en el marco definido estableciendo los espacios de accesibilidad anafórica y proporcionando antecedentes a los sistemas de restricciones y preferencias. Intuitivamente, este sistema marca las fronteras de la resolución de la anáfora, y este espacio se define de acuerdo con el ámbito del tópico.

A su vez, estos tres sistemas toman la entrada de las siguientes fuentes de información:

?? **Fuentes léxicas**: Proporcionan información relativa al comportamiento de ciertas palabras o grupos de palabras en ciertas situaciones. El nombre de "léxicas" se debe a que esta información suele encontrarse en un diccionario o *lexicon*, aunque también es posible encontrar la etiqueta de cada una de las palabras en la mayoría de los corpus etiquetados con categorías gramaticales. En nuestro sistema usamos la concordancia de la raíz de algunas expresiones anafóricas con sus respectivos antecedentes. Esta información léxica, puesto que su

cumplimiento no es siempre cierto, se usa como preferencia en lugar de restricción.

- ?? **Fuentes morfológicas:** Proporcionan información relativa a la concordancia en número, género y persona entre la expresión anafórica y su antecedente. La contribución de un etiquetador de categorías gramaticales hace posible el uso de esta información como una restricción.
- ?? **Fuentes sintácticas:** La información sintáctica se extrae del texto usando el análisis parcial. Por este motivo, aplicamos la gramática (SUG) y el analizador parcial (SUPP) capaz de extraer sintagmas nominales, sintagmas preposicionales y chunks verbales del texto. Esta información se usa como restricciones c-dominio en la resolución de la anáfora y como una preferencia en forma de paralelismo sintáctico como en Ferrández (1998).
- ?? **Fuentes semánticas:** La información semántica en la resolución de la anáfora se usa habitualmente como una característica semántica que se asigna a cada una de las entidades del discurso. Esta información es válida para forzar la compatibilidad entre las características semánticas tanto de la expresión anafórica como del antecedente. En la anáfora pronominal el pronombre hereda estas características del verbo de la oración. Para conseguir dicha información proponemos el uso de WordNet, una herramienta que proporciona una red de relaciones semánticas entre palabras (en inglés) tal y como se muestra en Miller (1990). En el futuro, podremos usar EuroWordNet, una herramienta actualmente en desarrollo que nos proporcionará esta información en castellano (entre otros lenguajes), como podemos ver en Vossen (1998).
- ?? **Fuentes del discurso:** La información del discurso genera información válida para la resolución de la anáfora en corpus de diálogos. Esta información se proporciona gracias a la acción de un etiquetador de intervenciones que delimita en el texto el ámbito de actuación de cada participante, y por la herramienta que busca el posible foco del discurso. Esta herramienta genera una lista de entidades relevantes con sus pesos. La información del discurso se usa como entrada para el sistema de preferencias así como para el sistema de segmentación del discurso.

Sin embargo, al no disponer de herramientas en castellano actualmente, el sistema no hace uso de estas fuentes, quedando por el momento como una propuesta para trabajos futuros.

- ?? **Información del mundo:** Finalmente, planteamos como trabajo futuro poder contar con información del mundo capaz de simular el conocimiento que el oyente/lector tiene en ciertas situaciones. Este conocimiento permite al oyente/lector en ciertas situaciones como en el ejemplo: *El presidente de los EEUU se dirigió a los medios. El Sr. Clinton se disculpó por hacerles esperar.* En este caso, sólo el conocimiento del mundo permite al oyente relacionar al Sr. Clinton con *el Presidente de los EEUU*. La adquisición de esta información requiere el uso de una herramienta enciclopédica capaz de proporcionar este conocimiento.

2.1 Algoritmo para la resolución de la anáfora en diálogos

2.1.1 Definiciones

Antes de introducir el algoritmo para la resolución de la anáfora, se introducirán algunos conceptos para su uso posterior.

- ?? **Entidad:** unidad de información que puede ser un antecedente anafórico. Este algoritmo sólo trata la anáfora pronominal, por lo tanto, sólo un sintagma nominal puede ser un antecedente. De esta forma, sintácticamente una entidad es un sintagma nominal.
- ?? **Participante:** uno de los hablantes del diálogo.
- ?? **Turno:** un fragmento de texto donde un participante habla sin interrupción.
- ?? **Intervención:** de acuerdo con la clasificación de Gallardo (1998), una intervención es un turno donde el participante añade información. Esta información implica la ocurrencia de una o más entidades.
- ?? **Turno continuador:** un turno donde el participante no añade (usa) entidades.
- ?? **Tópico:** es la entidad más relevante en un fragmento de texto. El participante se centra en esta entidad, y se vuelve a referir a ella durante este fragmento.

2.1.2 Algoritmo

Intuitivamente, el algoritmo de resolución de la anáfora pronominal para diálogos en castellano, figura 2, usa el sistema de tópicos para proporcionar antecedentes y los sistemas de restricciones y preferencias para escoger el antecedente apropiado.

```

Principio
Fa :=[]; Fn :=[]; Fa' :=[]; F :=[];
Turn :=1;
Mientras hay turnos
  Oración :=1;
  Mientras hay oraciones
    Fn :=relevante(np(Oración));
    Fa := Fa+Fn;
    Fn :=[];
    Si Buscaranáfora
      Buscarantecedente(Fa)
      Buscarantecedente(Fa')
      Buscarantecedente(F) por peso
    FinSi
    Si Fa<>[]
      Clase(Turno):=INTERVENCIÓN;
      Si Clase(Turno-1)=INTERVENCIÓN o
        Participante(Anterior_intervención)
        <>Participante(Turno)
        Fa' :=[];
      FinSi
    FinSi
    Oración := Oración+1;
  FinMientras
  Si Fa<>[] /* intervención */
    F :=incorporar(Fa,F);
    Fa' := Fa+Fa';
    Fa :=[];
    Anterior_intervención :=turno;
  Sino
    Clase(turno) :=CONTINUADOR;
  FinSi
  Turno :=Turno+1;
FinMientras
Fin
  
```

Fig. 2. Pseudo código del algoritmo para la resolución de la anáfora

El sistema de tópicos define tres espacios para la accesibilidad anafórica:

- El espacio de la intervención actual.
- El espacio de la intervención previa.
- El tópico relevante en el actual espacio de diálogo.

Los espacios primero y segundo se determinan directamente gracias al etiquetado de turnos que contiene el propio texto dialogado. El tercero se identificará como sigue:

Definiremos dos coeficientes. El primero (*Ciu*, coeficiente de incremento por uso) indica la importancia que adquiere una entidad que aparece en la intervención actual. El segundo (*Cdd*, coeficiente de decremento por desuso) indica la pérdida de importancia de una entidad que aparece en intervenciones previas pero no en la actual:

Ciu (Coeficiente de incremento por uso)
= 10 unidades.

Cdd (Coeficiente de decremento por desuso)
= 1 unidad.

Estos valores han sido propuestos experimentalmente y actualmente se están realizando estudios empíricos para determinar su adecuación.

El algoritmo usa tres listas diferentes para guardar las entidades que pueden ser un antecedente (de acuerdo con los espacios anteriormente definidos). La *lista de entidades locales actuales* –*Fa*– contiene una lista de entidades que aparecen en el ámbito de la intervención actual. La *lista de entidades locales anteriores* –*Fa'*– contiene una lista de entidades aparecidas en el ámbito de la última intervención. Y finalmente, la *lista de tópicos generales* –*F*– contiene una lista de entidades relevantes en el diálogo que será evaluada de acuerdo con criterios de frecuencia y distribución.

Este algoritmo analiza cada sentencia de cada turno detectando las entidades con relevancia temática en la oración y guardándolas en la lista *Fa* mediante el mecanismo *relevante*. Cuando la intervención cambia, las entidades almacenadas en la lista *Fa* se mueven a la lista *Fa'* puesto que ya forman parte de la intervención anterior, y además se incorporan a la lista de tópicos generales *F* usando el mecanismo *incorporar*. Este mecanismo se comporta de la siguiente forma: Por una parte, añade todas las entidades de *Fa* a *F* incrementando su peso con el coeficiente *Ciu*. Por otra parte el valor de *Cdd* se usa para decrementar el peso de aquellas entidades que están en *F* y no en *Fa*, es decir, aquellas que habían aparecido anteriormente pero no lo han hecho en la última intervención. Finalmente *Fa* queda vacía para almacenar las nuevas entidades que se detecten.

Puesto que el turno continuador no introduce entidades no se considera como intervención y por lo tanto se omite. Así, cuando un turno finaliza podremos determinar si se ha encontrado algún elemento temático (una entidad) de tal forma que lo identifique como una intervención, o si por el contrario no se encuentra ninguno con lo que se tomará como un turno continuador. Si se tratara de una intervención, la primera oración que introduzca nuevos elementos temáticos provocará que la lista *Fa'* con las entidades relevantes de la intervención anterior se vacíe totalmente. Esto es debido a que consideramos que las entidades que aparecen en la intervención anterior sólo tienen importancia local mientras la intervención actual aún no ha introducido

ninguna entidad. A partir de este instante ya sólo serán consideradas como tópicos generales de acuerdo con el valor de peso que tienen en la lista F . Sin embargo, si se trata de un turno continuador, Fa' no se vacía puesto que sus entidades siguen teniendo importancia local (ya que la intervención anterior sigue vigente).

Conforme se analiza cada oración, el mecanismo *buscar-anáfora* se pone en marcha intentando detectar la existencia de una anáfora pronominal. Si el resultado es afirmativo, se pone en marcha el mecanismo *buscar-antecedente* usando el sistema de restricciones y preferencias de acuerdo con los trabajos de Ferrández et al. (1998). Este mecanismo comenzará buscando el antecedente en la lista Fa . Si no se encuentra ninguna entidad, el mecanismo la buscará en la lista de entidades locales anteriores Fa' , y como último recurso la buscará en F .

2.1.3 Un ejemplo

El siguiente ejemplo es un fragmento de un diálogo en castellano donde se puede comprobar el funcionamiento del algoritmo mediante una traza. Las etiquetas Tnn nos introducen cada turno mientras $\langle Hn \rangle$ marca el participante que está hablando en cada momento.

T01: $\langle H1 \rangle$ Buenos días.
T02: $\langle H2 \rangle$ Buenos días. ¿Qué deseas?.
T03: $\langle H1 \rangle$ Quiero **manzanas**.
T04: $\langle H2 \rangle$ ¿De qué clase **las** quieres?.
T05: $\langle H1 \rangle$ No importa si son buenas.
T06: $\langle H2 \rangle$ **Éstas las** he recibido **esta mañana**.
Son muy buenas.
T07: $\langle H1 \rangle$ Entonces dame de **esas**.
T08: $\langle H2 \rangle$ ¿Cuántas quieres?.
T09: $\langle H1 \rangle$ Media docena.
T10: $\langle H2 \rangle$ Muy bien, ¿qué más quieres?.
T11: $\langle H1 \rangle$ ¿Tienes **limones**?
T12: $\langle H2 \rangle$ Sí, **los** tengo en **una caja** por aquí...
Aquí están. ¿**Los** quieres muy verdes?.
T13: $\langle H1 \rangle$ No... Son para hacer **limonada**. A **mis hijos les** encanta.
T14: $\langle H2 \rangle$ ¿Cuánto quieres?.
T15: $\langle H1 \rangle$ **Un kilo**.
T16: $\langle H2 \rangle$ **Éstos** te **los** dejo a **buen precio**.
Muy bien, ¿algo más?.
T17: $\langle H1 \rangle$ Nada más, gracias. ¿Cuánto es todo?.
T18: $\langle H2 \rangle$ 350.
T19: $\langle H1 \rangle$ Ahí tienes. Hasta luego.
T20: $\langle H2 \rangle$ Hasta luego.

El algoritmo comienza considerando los turnos T01 y T02 como turnos continuadores debido a que no contienen ninguna entidad. De esta forma, las tres listas permanecen vacías hasta el turno T03. Al procesar T03, *manzanas* es considerado como una entidad, y por lo tanto se introduce en la lista de entidades locales actuales Fa :

$$Fa=[manzanas]$$

T03 se considera por tanto una intervención. Cuando esta intervención finaliza, las entidades de Fa se incorporan a la lista de tópicos generales F con su peso correspondiente a la vez que se guardan en la lista de entidades locales anteriores Fa' . Tras esta operación las listas quedan de la siguiente forma:

$$Fa=[]$$

$$Fa'=[manzanas]$$

$$F=[(manzanas,10)]$$

En T04, tras la incorporación de *clase* como entidad aparecida, se detecta la primera anáfora con la aparición del pronombre *las*. En este punto el estado de las listas es:

$$Fa=[clase]$$

$$Fa'=[manzanas]$$

$$F=[(manzanas, 10)]$$

Para resolver la anáfora, el algoritmo busca, en primer lugar, un antecedente en la lista de entidades locales actuales (Fa). En este caso se encuentra la entidad *clase*, sin embargo, tras la intervención del sistema de restricciones se rechaza puesto que no cumple las características morfológicas requeridas (se busca un antecedente en plural). De esta forma, al no contener más candidatos la lista Fa , el algoritmo pasará a buscar en la lista de entidades locales anteriores Fa' encontrando en este caso *manzanas*. Tras la aplicación de las restricciones, *manzanas* se propone como antecedente para el pronombre y se incluye como entidad en Fa :

$$Fa=[clase, manzanas]$$

Tras procesar la primera oración de T04, Fa' se vacía:

$$Fa'=[]$$

Y al terminar T04 las listas quedan de la siguiente forma:

$$Fa=[]$$

$$Fa'=[clase, manzanas]$$

$F=[(manzanas, 20), (clase, 10)]$

Puesto que T05 es un turno continuador las listas no se modifican. Tras T06, el peso de *manzanas* se incrementa gracias a nuevas ocurrencias de esta instancia (introducidas por los pronombres), mientras que la entidad *clase* pierde peso por lo contrario:

$Fa=[]$

$Fa'=[manzanas, esta mañana]$

$F=[(manzanas, 40), (esta mañana, 10), (clase, 9)]$

Destacamos que las anáforas producidas por los pronombres *éstos* y *los* en el turno T16 sólo pueden ser resueltas usando las entidades de la lista de tópicos generales F puesto que las listas locales Fa y Fa' no contienen el antecedente correcto. El contenido de las listas en este momento es:

$Fa=[]$

$Fa'=[un kilo]$

$F=[(manzanas,45), (limones,39), (un kilo, 10), (caja,8), (media docena,6), (esta mañana, 4), (clase,3)]$

Así se cumple que, por un lado, la lista Fa está vacía, el sistema de restricciones rechaza el posible antecedente de la lista Fa' , y finalmente, tanto el sistema de restricciones como el de preferencias actúan sobre la lista F devolviendo *limones* como antecedente válido.

2.2 Realimentación entre el sistema de tópicos y la propia resolución de la anáfora.

El algoritmo propuesto incluye un mecanismo de detección de la anáfora (*buscaranáfora*) y otro para su resolución (*buscarantecedente*). Este último usa la información del sistema de tópicos para determinar el antecedente más relevante en la lista de tópicos generales. La información de los tópicos se une a la información que se obtiene de otras fuentes para obtener el mejor antecedente. De la misma forma, como podemos ver en la figura 3, una vez que la anáfora queda resuelta, el antecedente ocupa el espacio que ha dejado la expresión anafórica. De esta forma, este antecedente verá incrementado su peso en la lista de tópicos generales.

Así podemos afirmar que la estructura del discurso influye sobre la resolución de la anáfora de la misma forma que la resolución de la anáfora influye sobre la estructura del discurso.

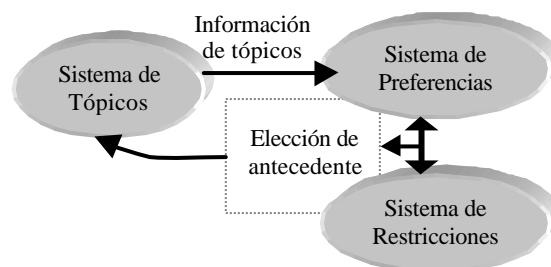


Fig. 3 Realimentación entre los sistemas de restricciones y preferencias y el sistema de tópicos.

3 Análisis sintáctico del diálogo: Una aplicación de la propuesta.

En esta sección mostraremos una aplicación del marco definido, en la tarea de análisis sintáctico del diálogo. Como se muestra en la figura 4, el marco global tiene una doble interacción con el analizador sintáctico.

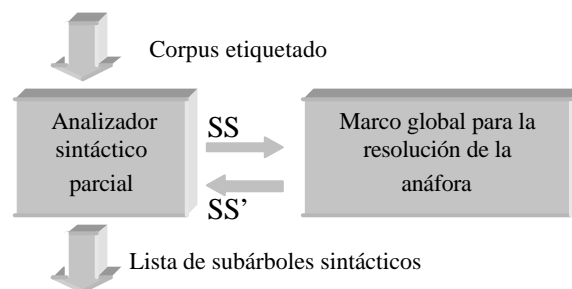


Fig. 4. Interaction between partial parser module and the global framework for anaphora resolution in dialogues

Este analizador sintáctico parcial SUPP (Slot Unification Partial Parser) fue presentado por Martínez-Barco et al. (1998) y utiliza una gramática SUG (Slot Unification Grammar) parcial para extraer del texto de entrada los sintagmas nominales, sintagmas nominales coordinados, sintagmas preposicionales y chunks verbales que se encuentren.

SUPP construye una estructura de huecos (Slot Structure, SS) para cada constituyente con la información morfológica, léxica y sintáctica. El algoritmo para la resolución de la anáfora procesa cada SS y extrae la información necesaria para resolver la anáfora devolviendo una nueva estructura de huecos (SS'). Si este algoritmo detecta que la estructura SS es una expresión anafórica, la resolverá e incluirá la información morfológica, léxica y sintáctica del antecedente en su SS que ahora se llamará SS' .

Finalmente, SUPP genera una lista de salida con los subárboles sintácticos donde la anáfora

ha sido resuelta. Esta lista podrá ser usada por módulos posteriores.

4 Evaluación

El algoritmo de resolución de la anáfora depende de la eficacia del analizador parcial. En nuestro caso, el analizador parcial SUPP ya fue evaluado en Palomar et al. (1999) obteniendo una precisión del 95% y una cobertura del 94% en sintagmas nominales simples, 80% y 79% respectivamente para sintagmas nominales complejos, 93'5% y 95% en chunks verbales, así como 95% y 94% en sintagmas preposicionales.

Por otra parte, el algoritmo de resolución de anáfora expuesto anteriormente ha sido implementado usando LPA-Prolog y posteriormente evaluado sobre 10 diálogos independientes en castellano etiquetados morfológica y léxicamente por el etiquetador de categorías gramaticales de Xerox conteniendo una media de 700 palabras cada uno. Dicha evaluación muestra unos resultados de precisión¹ del 77% y cobertura² del 75% en la resolución de anáfora pronominal.

5 Conclusión

En este artículo hemos presentado un algoritmo para la resolución de la anáfora pronominal en castellano. Este algoritmo combina información lingüística con la información de la estructura del diálogo. Así mismo, consideramos que en los diálogos, a diferencia de los monólogos, es posible definir tres espacios de accesibilidad anafórica para proporcionar antecedentes a los sistemas de restricciones y preferencias.

Finalmente, concluimos que nuestro algoritmo ha sido evaluado satisfactoriamente como muestran los prometedores resultados del 77% de precisión, teniendo en cuenta que el sistema no usa información semántica. Así mismo, hemos estimado que eliminando los errores introducidos por la limitación del juego de etiquetas usado, y obviando aquellos errores que se producen por arrastre de otros, podríamos alcanzar valores cercanos al 85% y

¹ Definimos precisión como el cociente entre el número de anáforas resueltas correctamente y el número de anáforas detectadas.

² Definimos cobertura como el cociente entre el número de anáforas resueltas correctamente y el número real de anáforas existentes.

83% en precisión y cobertura respectivamente.

6 Agradecimientos

Este artículo ha sido subvencionado por la Comisión Interministerial de Ciencia y Tecnología con el proyecto número TIC97 - 0671-C02 - 02.

Agradecemos al Profesor Fernando Sánchez León su labor de etiquetado de los corpus de prueba, así como al Laboratorio de Lingüística Computacional de la Universidad Autónoma de Madrid por brindarnos su corpus oral.

7 Referencias

- Azzam S., Humphrey K. and Gaizauskas R. (1998) Evaluating a Focus-Based Approach to Anaphora Resolution. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL*, Montreal, Canada. pp 74-78
- Ferrández, A. (1998) *Aproximación computacional al tratamiento de la anáfora pronominal y de tipo adjetivo mediante gramáticas de unificación de huecos*. PhD Thesis. Universidad de Alicante.
- Ferrández A., Palomar M. and Moreno L. (1998) Anaphor resolution in unrestricted texts with partial parsing. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL*, (Montreal, Canada). pp 385-391.
- Gallardo B. (1998) *Comentario de textos conversacionales*. Colección Comentario de Textos. Arco Libros, S.L., Madrid.
- Grosz B and Sidner C. (1986) Attentions, intentions and the structure of discourse. *Computational Linguistics*, **12**, pp 175-204.
- Grosz B., Joshi A. and Weinstein, S. (1995) Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**. pp 203-225.
- Manabu O. and Kouji T. (1996) Zero Pronoun Resolution in Japanese Discourse Based on Centering Theory. In *Proceedings of the 16th International Conference on Computational Linguistics. COLING'96, II*, (Copenhagen, Denmark), pp 871-876.
- Martínez-Barco P., Peral J., Ferrández A., Moreno L. and Palomar M. (1998) Analizador Parcial SUPP. In *Proceedings of the 6th Iberoamerican Conference on Artificial Intelligence, IBERAMIA '98* (Lisbon, Portugal). pp 329-341.
- Miller G.A. (1990) Wordnet: an on-line lexical database. *International Journal of Lexicography*, **4**. pp 235-312.

- Mitkov R. (1996) Anaphor resolution: a combination of linguistic and statistical approaches. *In Proceedings of the Discourse Anaphora And Resolution Colloquium, DAARC96.*
- Palomar M., Ferrández A., Moreno L., Saiz-Noeda M., Muñoz R., Martínez-Barco P., Peral J. and Navarro B. (1999) A Robust Partial Parsing Strategy based on the Slot Unification Grammars. *Traitement Automatique des Langues Naturelles . 6th annual conference on NLP, TALN99.* Cargèse, Corsica, France. July, 1999., Bergen, Norway. Acceptance pending.
- Rocha M.A.E. (1998) *A corpus-based study of anaphora in dialogues in English and Portuguese.* PhD Thesis. University of Sussex.
- Vossen P. (1998) EuroWordNet: Building a Multilingual Database with Wordnets for European Languages. In: K. Choukri, D. Fry, M. Nilsson (eds), *The ELRA Newsletter*, **3**, ISSN: 1026-8200.
- Walker M.A. (1997) Centering, Anaphora Resolution, and Discourse Structure. *Centering In Discourse*, eds. Walker, M.A.; Joshi, A.K.; Prince, E.F., Oxford University Press,.