

Resolución computacional de la anáfora en diálogos: Estructura del discurso y conocimiento lingüístico*

Patricio Martínez-Barco

Universidad de Alicante

Campus de Sant Vicent del Raspeig

patricio@dlsi.ua.es

Resumen: En este trabajo se ha desarrollado un sistema para la resolución computacional de la anáfora generada en el uso del lenguaje natural por pronombres (personales y demostrativos) y anáforas adjetivas en los diálogos. El método propuesto se basa en el uso de conocimiento lingüístico (información léxica, morfológica y sintáctica) y de conocimiento estructural (la estructura del diálogo y el espacio de accesibilidad anafórica estructural que genera), ambos independientes del dominio.

Palabras clave: fenómenos lingüísticos, resolución de la anáfora, diálogos, estructura del discurso

Abstract: In this work, a computational resolution system for anaphora has been developed to solve the anaphora generated because of using natural language with pronouns (personal and demonstrative pronouns) and adjectival anaphora, both in dialogue discourse. The proposed method is based on the use of linguistic knowledge (lexical, morphological and syntactic information), and structural knowledge (dialogue structure and the anaphoric accessibility space it generates), both domain independent.

Keywords: linguistic phenomena, anaphora resolution, dialogues, discourse structure

1 Introducción

Esta Tesis fue dirigida por Dr. Manuel Palomar Sanz, y presentada en la Universidad de Alicante con fecha 8 de marzo de 2001. El tribunal estuvo formado por: Dra. Dña. Lidia Moreno Boronat, Dr. D. Antonio Ferrández Rodríguez, Dra. Dña. Nuria Castell Ariño, Dra. Dña. Ana García Serrano, y Dr. D. José Gabriel de Amores Carredano. La calificación obtenida fue de Sobresaliente *cum laude*.

2 Resumen de la Tesis Doctoral

En este trabajo se ha desarrollado un sistema para la resolución computacional de la anáfora generada en el uso del lenguaje natural por pronombres (personales y demostrativos) y anáforas adjetivas en los diálogos. El método propuesto se basa en el uso de conocimiento lingüístico (información léxica, morfológica y sintáctica) y de conocimiento estructural (la estructura del diálogo y el es-

pacio de accesibilidad anafórica estructural que genera), ambos independientes del dominio.

La estructura del diálogo en la conciencia del oyente, construida mediante pares adyacentes y sus relaciones, y mediante la definición de los tópicos del discurso, influye decisivamente en la toma de decisiones para resolver la anáfora (Martínez-Barco y Palomar, 2000). Por este motivo se ha propuesto un esquema de anotación que sustituye el conocimiento que sobre la estructura del diálogo dispondría un oyente, para que pueda ser aplicado por un sistema automático de resolución de anáfora.

Gracias al uso de información no dependiente del dominio, este sistema de resolución de la anáfora para diálogos al que hemos denominado ARIADNA (Anaphora Resolution In Automatic Dialogue systems with Natural language Analysis) está preparado para ser aplicado a sistemas que trabajen sobre cualquier dominio, y de una forma muy especial para los sistemas de diálogo que constituyen su aplicación más inmediata.

Así este trabajo presenta las siguientes

* Esta investigación ha sido parcialmente financiada por el Ministerio de Ciencia y Tecnología a través del proyecto núm. TIC2000-0664-C02-01/02

aportaciones:

- La definición de un espacio de accesibilidad anafórica estructural, garantizado mediante un estudio empírico, en el que se demuestra que en los diálogos en español, los antecedentes de la anáfora pronominal y adjetiva se encuentran en este espacio. Este espacio se basa en los estudios realizados por Fox (1987), quien se centra en la creencia de que existe una fuerte relación entre el flujo de información en un texto, la estructura del texto y el uso de la anáfora. De hecho es sabido que aquellos antecedentes que se encuentran en el “foco” o en “la consciencia del oyente” pueden ser anaforizados.
- La propuesta y definición de un esquema de anotación para diálogos en español con el objetivo de obtener el espacio de accesibilidad anafórica. Este esquema está basado en la estructuración de diálogos en español propuesta por Gallardo (1996) quien se basó a su vez en las teorías desarrolladas por Sacks *et al.* (1974)
- La definición de un conjunto de restricciones morfológicas y sintácticas que definen la no correferencialidad entre la anáfora y los candidatos a antecedente, un conjunto de preferencias lingüísticas basadas en la estructura sintáctica parcial generada por el analizador sintáctico, y un conjunto de preferencias estructurales basados en la propia estructura del diálogo.
- El desarrollo de un algoritmo de resolución de la anáfora basado en los conjuntos de restricciones y preferencias lingüísticas y estructurales definidos. Este algoritmo aplica el conjunto de restricciones como un filtro lingüístico que descarta los posibles antecedentes, y el conjunto de preferencias como un modo de ordenar los candidatos para devolver el mejor de ellos. La construcción del sistema de resolución de la anáfora al que hemos denominado ARIADNA que integra el algoritmo de resolución de la anáfora y el espacio de accesibilidad anafórica estructural. La implementación del sistema se ha realizado en el lenguaje de programación lógico Prolog.

- La definición de un conjunto de herramientas interfaz que integran la información que proviene de las distintas fuentes de conocimiento usadas por ARIADNA con el módulo de resolución anafórica, es decir, el núcleo de ARIADNA. La existencia de estas herramientas proporciona una total independencia y facilidad de integración con cualquier sistema de diálogo o aplicación final a la que se destine ARIADNA. Sobre el sistema final, se ha realizado un proceso de entrenamiento con el objeto de obtener de un modo empírico y cuantitativo el mejor conjunto de preferencias, y una evaluación del sistema sobre un corpus independiente obteniendo como resultado se obtuvo que el 81,37% de las anáforas tratadas fueron resueltas correctamente.

Uno de los trabajos más importantes publicados a raíz de esta Tesis puede ser consultado en Palomar y Martínez-Barco (2001).

Bibliografía

- Fox, B. 1987. *Discourse Structure and Anaphora*. Written and conversational English. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.
- Gallardo, B. 1996. *Análisis conversacional y pragmática del receptor*. Colección Sinapsis. Ediciones Episteme, S.L., Valencia.
- Martínez-Barco, P. y M. Palomar. 2000. Dialogue structure influence over anaphora resolution. En O. Cairo L.E. Sucar, y F.J. Cantu, editores, *MICAI 2000: Advances in Artificial Intelligence*, volumen 1793 de *Lecture Notes in Artificial Intelligence*, páginas 515–525, Acapulco, México. Springer-Verlag.
- Palomar, M. y P. Martínez-Barco. 2001. Computational approach to anaphora resolution in Spanish dialogues. *Journal of Research in Artificial Intelligence*, 5:263–287.
- Sacks, H., E. Schegloff, y G. Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4):696–735.