

**Aportaciones a la resolución de la elipsis en lenguaje natural
utilizando técnicas incrementales**

Tesis doctoral. Universidad Politécnica de Valencia. Septiembre 1996

Autor: Manuel Palomar (e-mail: mpalomar@dlsi.ua.es)

Directora: Lidia Moreno

El propósito principal de esta Tesis ha sido el tratamiento y resolución de la elipsis que se produce en la coordinación, como uno de los problemas principales destacados en el análisis de sentencias para la obtención de la representación de su significado.

Mucho se ha debatido en los últimos años sobre el apropiado nivel de procesamiento del lenguaje que trate la resolución de la elipsis. Y se ha debatido desde dos puntos de vista; la aproximación sintáctica y la semántica. La aproximación sintáctica, Lappin, McCord, Haik, Fiengo y May, consiste en la reconstrucción sintáctica del constituyente incompleto a partir de su antecedente. Mientras que la aproximación semántica, Dalrymple, Kehler, Gawron y Peters, consiste en identificar un significado para el constituyente incompleto y asignárselo directamente; es decir, la información se recupera en la representación semántica.

Para esta resolución, nosotros hemos adoptado el punto de vista semántico, donde se toma la forma lógica del antecedente para la reconstrucción del constituyente incompleto o elidido. Consideramos que no es necesario una reconstrucción sintáctica de los elementos elididos, ya que la reconstrucción semántica proporciona los mecanismos adecuados para una vez identificado el antecedente del constituyente elidido asignarle su forma lógica, por lo que resuelve de forma más eficiente el problema.

El problema de la resolución de la elipsis lo hemos dividido en dos tareas: la primera será determinar la estructura paralela entre el constituyente incompleto y su antecedente, y por otro lado la formación de

la relación implícita. La división del problema de la elipsis en dos partes permite una mejor descripción de la formación de la relación y una mejor caracterización del problema de determinación del paralelismo. Por lo que desde nuestro punto de vista, para determinar el paralelismo es necesario criterios sintácticos semánticos y posteriormente para la formación de la relación se pueden adoptar criterios semánticos. Con ello, aportamos un nuevo método de resolución que incorpora una adaptación de recientes trabajos sobre la elipsis, donde resalta la idea de estructuras paralelas, pero mientras estas aproximaciones acentúan el paralelismo semántico, nosotros usamos tanto el paralelismo sintáctico como el semántico.

Este nuevo método está basado en las técnicas incrementales de evaluación de programas Datalog como herramienta para la definición e implementación de gramáticas de contexto libre; Y está desarrollado mediante un nuevo tipo de Gramáticas de Cláusulas Definidas (DCGs), las Gramáticas Datalog (DLG), basadas en la teoría de bases de datos.

El nuevo formalismo que presentamos es una extensión de las Gramáticas Datalog, presentadas por V.Dahl, P.Tarau, L.Moreno y M.Palomar, el cual mejora la eficiencia de las tradicionales Gramáticas de Cláusulas Definidas (DCG) mediante apropiados mecanismos de compilación (como por ejemplo la resolución-OLDT). Esta extensión mejora el tratamiento y resolución de las estructuras coordinadas y plantea por primera vez una implementación en Prolog de tal gramática, utilizando el algoritmo de evaluación semi-naive, que nos permite "congelar" el proceso de análisis de las oraciones para poderlo tratar metagramaticalmente.

Por último, hemos adaptado el mecanismo de comprobación semántica de tipos (IRSAS) a nuestro método, que permite resolver gran número de casos de coordinación y de elipsis.