

Gramáticas de Adjunción de Árboles: Un Enfoque Deductivo en el Análisis Sintáctico *

Víctor J. Díaz Madrigal

Depto. de Lenguajes y Sistemas Informática
E.T.S. Ingeniería Informática (Universidad de Sevilla)
Avda. Reina Mercedes s/n 41012-SEVILLA
vjdiaz@lsi.us.es

Resumen: En este resumen mostramos los aspectos más relevantes incluidos en la tesis doctoral, cuyo hilo central ha girado en torno al estudio y comparación de un conjunto significativo de estrategias de análisis para la clase de las Gramáticas de Adjunción de Árboles.

Palabras clave: Análisis Sintáctico, Analizadores Deductivos, Gramáticas de Adjunción de Árboles.

Abstract: In this abstract we show the more relevant aspects included in our dissertation whose main subject is a general and compared study of several parsing strategies for the class of Tree Adjoining Grammars.

Keywords: Parsing, Deductive Parsing, Tree Adjoining Grammars.

Las gramáticas de adjunción de árboles (TAGs, *Tree Adjoining Grammars*) (Joshi y Schabes, 97) son sistemas de reescritura de árboles que pueden considerarse una extensión de las gramáticas independientes del contexto (CFGs, *Context Free Grammars*). Dicha extensión es la consecuencia de ampliar las estructuras básicas de composición (árboles en las TAGs frente a producciones en las CFGs) y de modificar las operaciones de composición utilizadas en la derivación (la adjunción en las TAGs frente a la sustitución en las CFGs). Estos dos cambios provocan que la clase de los lenguajes reconocidos por las TAGs sea un superconjunto estricto de la clase de los lenguajes reconocidos por las CFGs, y de aquí, que el coste temporal y espacial requerido en el análisis sintáctico de TAGs sea superior al presentado por las CFGs.

Desde la perspectiva del análisis sintáctico de lenguajes naturales, las TAGs presentan tres interesantes propiedades:

- Dominio de localidad extendido: Los árboles elementales facilitan la capacidad de establecer las dependencias locales dentro de una misma estructura.

- Factorización de la recursión: La operación de adjunción, cuya aplicación requiere de árboles con capacidad de descripción recurrente, expande las dependencias locales de forma que se pueden modelar directamente dependencias no locales.
- Lexicalización: Cada árbol elemental está gobernado por un ancla (núcleo), lo que conduce a una cierta simbiosis entre el concepto de gramática y lexicón.

El enfoque deductivo (Sikkel, 97) propone una teoría inspirada en la similitud existente entre el proceso de análisis y la demostración de teoremas o fórmulas. Los beneficios de este enfoque no se restringen sólo a ofrecer un lenguaje de especificación para la definición de analizadores, denominados en este contexto esquemas de análisis, sino que otorga una teoría suficientemente rica como para poder establecer relaciones matemáticas entre distintas estrategias de análisis. Las relaciones entre los esquemas pueden ser ilustradas gráficamente mediante una red de analizadores.

Si bien el enfoque deductivo ha sido aplicado con éxito en el contexto del análisis sintáctico para CFGs, éste no es el caso para la clase de gramáticas que nos atañe, donde su aplicación ha sido tan solo parcial. A nuestro juicio, esta situación se debe funda-

* Tesis doctoral presentada en octubre de 2000 en la Facultad de Informática y Estadística de la Universidad de Sevilla por Víctor J. Díaz Madrigal y dirigida por el Dr. Miguel Toro. Autor:

mentalmente a dos factores: la propia complejidad del análisis de gramáticas TAGs y la falta de uniformidad en la presentación de los analizadores.

Teniendo en mente este escenario, podemos considerar que el objetivo del trabajo ha sido estudiar y ampliar la red de analizadores para gramáticas TAGs utilizando como filosofía el enfoque deductivo. Para la construcción de dicha red, y tras un estudio crítico de distintas alternativas, nos hemos decantado por una notación común que facilite la integración de analizadores suficientemente conocidos en la literatura. Una vez introducido los conceptos fundamentales y la motivación del trabajo, acabaremos presentando brevemente los distintos capítulos de la memoria:

El **Capítulo 1: Introducción** incluye la justificación del trabajo, mostrando además un resumen del contenido y de los resultados obtenidos.

El **Capítulo 2: Análisis sintáctico** estudia en profundidad la teoría que subyace a los analizadores deductivos enumerando las distintas relaciones matemáticas que pueden establecerse entre ellos.

El **Capítulo 3: Gramáticas de adyunción de árboles** presenta la definición formal de las TAGs junto con sus propiedades computacionales más destacadas. También se estudian los aspectos más relevantes relacionados con la utilización de estas gramáticas en el contexto lingüístico y un conjunto de formalismos estrechamente vinculados con ellas.

El **Capítulo 4: Análisis sintáctico de gramáticas de adyunción de árboles** presenta fundamentalmente un estudio crítico de distintas notaciones que pueden ser empleadas para definir esquemas para TAGs. Tras escoger una notación, y a modo de ejemplo, se introducen dos esquemas para TAGs ya conocidos en la literatura: uno basado en el método CYK y otro en una versión ascendente del método de Earley.

El **Capítulo 5: Esquemas basados en Earley sin prefijo válido** presenta un conjunto de esquemas que son una extensión para TAGs del conocido método de Earley definido originalmente para CFGs. Los principales esquemas incluidos se caracterizarán por no garantizar la propiedad del prefijo válido. Informalmente, el incumplimiento de esta propiedad conduce a que un analizador, aun siendo correcto, no detecta los errores

tan pronto como éstos suceden.

El **Capítulo 6: Esquemas basados en Earley con prefijo válido** continua el estudio de esquemas basados en el algoritmo de Earley pero ahora garantizando la propiedad del prefijo válido.

El **Capítulo 7: Esquemas basados en De Vreught y Honig** presenta principalmente dos nuevos analizadores bidireccionales basados en una extensión para TAGs del algoritmo definido originalmente por De Vreught y Honig para CFGs. También veremos la relación de estos esquemas con el esquema bidireccional dirigido por núcleos propuesto por Van Noord.

El **Capítulo 8: Estudio comparado** incluye un estudio empírico del comportamiento de los esquemas más relevantes presentados en los capítulos anteriores haciendo uso de la máquina deductiva de análisis (Shieber et al., 95). Esta máquina define un modelo computacional que permite ejecutar los esquemas de análisis tras una compilación de estos a un programa lógico.

Finalmente, el **Capítulo 9** hará un balance general del trabajo realizado incluyendo las conclusiones finales junto con una discusión de las líneas de trabajo futuras.

Bibliografía

- A. K. Joshi y Y. Schabes. Tree-adjointing grammars. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages. Vol 3: Beyond Words*, chapter 2, pp 69–123. Springer-Verlag, Berlin/Heidelberg/New York. (1997)
- S. M. Shieber, Y. Schabes y F. C. N. Pereira. Principles and implementation of deductive parsing. *Journal of Logic Programming* 24(1&2), pp 3–36. (1995)
- K. Sikkel. Parsing schemata. A framework for specification and analysis of parsing algorithms. *Texts in Theoretical Computer Science*, Springer-Verlag. (1997)