Capítulo 12

Conclusiones

El trabajo descrito en esta tesis representa una aportación significativa y original al análisis sintáctico de los lenguajes de adjunción de árboles y por extensión al procesamiento del lenguaje natural, a la inteligencia artificial, y a la teoría de autómatas y lenguajes formales.

El panorama al comenzar el trabajo que daría lugar a esta memoria consistía en un grupo disperso de algoritmos para el análisis sintáctico de gramáticas de adjunción de árboles y apenas un par de algoritmos para el análisis sintáctico de las gramáticas lineales de índices. En esta memoria se ha mostrado que es posible establecer un camino evolutivo continuo en el que se sitúan los algoritmos de análisis sintáctico que incorporan las estrategias de análisis más importantes, tanto para el caso de las gramáticas de adjunción de árboles como para el caso de las gramáticas lineales de índices. Los diferentes algoritmos se han definido con esquemas de análisis sintáctico, de tal modo que los algoritmos más complejos se derivan a partir de los menos complejos aplicando una secuencia de transformaciones simples.

En el caso de las gramáticas lineales de índices el resultado es doblemente interesante, pues si bien se ha esgrimido a su favor su adecuación como formalismo intermedio para el análisis de gramáticas de adjunción de árboles, lo cierto es que numerosas estrategias de análisis para estas últimas no se hallaban incorporadas a ningún algoritmo de análisis sintáctico para gramáticas lineales de índices. En consecuencia, era necesario sacrificar la estrategia de análisis si se optaba por este enfoque, lo que limitaba enormemente su aplicación práctica. Con el trabajo desarrollado en esta memoria hemos salvado ese obstáculo definiendo algoritmos de análisis sintáctico para gramáticas lineales de índices que incorporan la versión equivalente de las estrategias de análisis más populares para gramáticas de adjunción de árboles.

En el caso de las gramáticas independientes del contexto es posible optar por un diseño modular en el cual se separa la definición y la ejecución de una determinada estrategia de análisis. En particular, es posible definir un algoritmo de análisis sintáctico como un conjunto de transiciones de un autómata a pila, probablemente no determinista, el cual puede ser interpretado eficientemente mediante las técnicas de tabulación disponibles. Este enfoque presenta ventajas evidentes, entre la cuales cabe citar la simplificación de las pruebas de corrección de los algoritmos, los cuales son más fáciles de comprender y, al ser ejecutados en un entorno homogéneo, son fácilmente comparables. En esta memoria hemos adaptado este enfoque a los lenguajes de adjunción de árboles, proporcionando modelos de autómata con los que describir los algoritmos de análisis y técnicas de tabulación con las que pueden ser ejecutados eficientemente.

El primer modelo de autómata considerado ha sido el de los autómatas a pila embebidos, una extensión de los autómatas a pila que utiliza como estructura de almacenamiento una pila de pilas. Hemos definido una versión de este tipo de autómatas en la cual se simplifica la forma de las transiciones y se elimina el control de estado finito, manteniendo la potencia expresiva. La nueva formulación permite diseñar una técnica de tabulación que, si bien no es general, permite ejecutar eficientemente los autómatas a pila embebidos obtenidos a partir de los esquemas de compilación que incorporan las estrategias de análisis sintáctico presentes en la literatura.

En el caso de las gramáticas de adjunción de árboles, los autómatas a pila embebidos permiten reconocer las adjunciones de modo descendente. Si deseamos reconocer las adjunciones de manera ascendente, debemos recurrir a los autómatas a pila embebidos ascendentes, los cuales han sido completamente reformulados con el fin de lograr una definición formal consistente, hecho que no se había conseguido hasta el momento. Se han definido versiones con estados y sin estados, esta última más simple y apta para la aplicación de técnicas de tabulación.

Hemos considerado también autómatas con una estructura de almacenamiento diferente, una pila en la que cada uno de los elementos que la componen tiene asociado una pila de índices. En este caso hemos seguido dos aproximaciones diferentes. La primera aproximación parte de la restricción de los autómatas lógicos a pila para adaptarlos al reconocimiento de las gramáticas lineales de índices, obteniéndose diferentes tipos de autómatas especializados en diversas estrategias de análisis según el tipo de transiciones permitidos. Este enfoque presenta el inconveniente de que es preciso añadir ciertas restricciones a la combinación de transiciones para garantizar que la clase de los lenguajes aceptados es equivalente a la clase de los lenguajes de adjunción de árboles.

La segunda aproximación parte de los autómatas a pila y de la extensión de las técnicas de tabulación propuestas para estos últimos, dando lugar al concepto de autómata lineal de índices, con tres variantes diferentes: Los autómatas lineales de índices orientados a la derecha, adecuados para estrategias en las cuales las adjunciones se reconocen de manera ascendente, los autómatas lineales de índices orientados a la izquierda, aptos para estrategias de análisis en las que las adjunciones se tratan de forma descendente, y los autómatas lineales de índices fuertemente dirigidos, capaces de incorporar estrategias de análisis en las cuales las adjunciones se tratan de manera ascendente, descendente o de ambas maneras. Esta última variante es completamente original, como también lo son su técnica de interpretación tabular y la correspondiente a la variante orientada a la izquierda.

El último modelo de autómata definido hace uso de una estructura de almacenamiento diferente. En este caso, se conserva la pila de los autómatas a pila, que pasa a denominarse pila maestra pues es la encargada de dirigir el proceso de análisis, y se añade una nueva pila, que recibe el nombre de pila auxiliar pues es utilizada para almacenar elementos auxiliares o índices que restringen las transiciones aplicables en un momento dado. Hemos descrito dos versiones diferentes de este tipo de autómatas, los autómatas con dos pilas fuertemente dirigidos, aptos para describir estrategias de análisis arbitrarias, y los autómatas con dos pilas ascendentes, adecuados para describir estrategias de análisis en las cuales las adjunciones de procesan ascendentemente.

Finalmente, hemos analizado conjuntamente todos estos modelos de autómata, los cuales se pueden dividir en tres grandes grupos: la familia de los autómatas generales, de la que forman parte los autómatas lineales de índices fuertemente dirigidos y los autómatas con dos pilas fuertemente dirigidos; la familia de los autómatas descendentes, en la que se encuadran los autómatas a pila embebidos y los autómatas lineales de índices orientados a la izquierda; y la familia de los autómatas ascendentes, en la que se enmarcan los autómatas a pila embebidos ascendentes, los autómatas lineales de índices orientados a la derecha y los autómatas con dos pilas ascendentes.

12.1 Trabajo futuro

El trabajo de investigación mostrado en esta memoria no termina en sí mismo sino que abre las puertas a numerosas líneas de trabajo futuro.

Un aspecto que presenta gran interés es el estudio del comportamiento práctico de los diferentes modelos de autómata y el desarrollo de esquemas de compilación y de técnicas de tabulación optimizadas que tengan en cuenta las construcciones específicas que aparecen en las gramáticas de las lenguas naturales. Premisa indispensable para emprender este estudio es la disponibilidad de una gramática de adjunción de árboles de amplia cobertura, desgraciadamente inexistente en el caso de las lenguas peninsulares.

En los últimos años se están dedicando esfuerzos importantes a la definición de formalismos gramaticales que preservan, al menos en parte, las características que hacen de las gramáticas de adjunción de árboles un formalismo adecuado para la descripción sintáctica del lenguaje natural pero que presenten una menor complejidad computacional, aunque ello conlleve un pérdida parcial de la sensibilidad al contexto. En este ámbito, resulta de interés la definición de modelos de autómata especializados en el análisis de dichos formalismos, entre los que cabe destacar las gramáticas de adjunción de árboles que restringen las adjunciones realizables en la espina [167] y las gramáticas de inserción de árboles [179].

Una vez que se dispone de mecanismos operacionales que permiten analizar eficientemente la estructura sintagmática de las frases, resulta inevitable avanzar hacia el siguiente nivel de análisis del lenguaje, tratando de capturar los aspectos semánticos del mismo. La extensión natural de las gramáticas de adjunción de árboles que permite capturar, al menos en parte, la semántica del lenguaje, son las gramáticas de adjunción de árboles con estructuras de rasgos [210]. En esta dirección, ya se han dado los primeros pasos para incorporar el tratamiento de las estructuras de rasgos a los modelos de autómata propuestos en esta memoria.

Ya finalmente, apartándonos del procesamiento del lenguaje natural, una línea de investigación futura consiste en la aplicación al campo de la programación lógica de los principios subyacentes a las técnicas de tabulación presentadas para los modelos de autómata definidos en esta memoria. Las gramáticas lineales de índices pueden verse como un tipo especial de programas lógicos en el que los argumentos de los predicados tienen un único argumento en la forma de una pila. La forma particular en que se manipulan los predicados y sus argumentos, nos ha permitido colapsar las derivaciones en una forma muy compacta. Tenemos la intuición de que un tipo similar de compactación es posible para ciertos predicados de programas lógicos, en particular para aquellos cuyos argumentos se construyen a partir de los valores de otros argumentos tal que dichos valores son poco dependientes del contexto en el que son evaluados.

El sino más hermoso de una teoría física es el de señalar el camino para establecer otra más amplia, en cuyo seno pervive como caso límite.

Albert Einstein [70, página 72]