

Apéndice A

Esquemas de análisis sintáctico

Los algoritmos de análisis sintáctico de esta memoria se describen utilizando *esquemas de análisis sintáctico*, una estructura para realizar descripciones de alto nivel de algoritmos de análisis desarrollada por Sikkel en [189]. En [191] se pueden encontrar una descripción más breve. En [190] se tratan los problemas inherentes a la verificación de la corrección de esquemas de análisis.

A.1. Esquemas de análisis sintáctico para CFG

Los esquemas de análisis sintáctico permiten describir cómodamente algoritmos de análisis sintáctico y estudiar fácilmente las relaciones entre diferentes algoritmos mediante el análisis de las relaciones formales entre los esquemas de análisis subyacentes.

Definición A.1 *Un sistema de análisis sintáctico para una gramática independiente del contexto \mathcal{G} y una cadena de entrada $a_1 \dots a_n$ es un triple $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$, en el que*

- \mathcal{I} es un conjunto de ítems que representan resultados intermedios del proceso de análisis;
- \mathcal{H} es un conjunto inicial de ítems denominados hipótesis que representan la cadena que va a ser analizada¹;
- $\mathcal{D} \subseteq \wp_{\text{fin}}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}$ es un conjunto de reglas de inferencia, denominadas pasos deductivos, mediante las cuales se derivan nuevos ítems a partir de los ítems existentes. Estos pasos deductivos tienen la forma $\eta_1, \dots, \eta_k \vdash \xi$ y su significado es el siguiente: si todos los antecedentes $\eta_i \in \mathcal{H} \cup \mathcal{I}$ de un paso deductivo ya existen, entonces el consecuente ξ deberá ser generado por el analizador sintáctico. Los pasos deductivos se suponen cuantificados universalmente para todas las posibles variables que aparezcan en los ítems a menos que se indique explícitamente lo contrario.

Un conjunto $\mathcal{F} \subseteq \mathcal{I}$ de ítems finales indica mediante su presencia el reconocimiento de la cadena de entrada.

Definición A.2 *Un sistema de análisis sintáctico no instanciado para una gramática independiente del contexto \mathcal{G} es un triple $\langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$ con \mathcal{H} una función que asigna un conjunto de hipótesis a cada cadena de entrada $a_1 \dots a_n \in V_T^*$, tal que $\langle \mathcal{I}, \mathcal{H}(a_1 \dots a_n), \mathcal{D} \rangle$ es un sistema de análisis sintáctico.*

¹No es necesario que $\mathcal{H} \subset \mathcal{I}$. Por el contrario, es habitual que ambos conjuntos sean disjuntos.

Definición A.3 *Un esquema de análisis para alguna subclase de gramáticas independientes del contexto es una función que asigna un sistema de análisis sintáctico no instanciado a cada gramática en dicha subclase.*

A.2. Sistemas de deducción gramatical

Es interesante reseñar la similitud de los esquemas de análisis sintáctico con los *sistemas de deducción gramatical* propuestos por Shieber, Schabes y Pereira en [188].

Definición A.4 *Un sistema de deducción gramatical para una gramática \mathcal{G} y una cadena de entrada $a_1 \dots a_n$ es una cuádrupla $\langle I, A, G, R \rangle$ en la que*

- *I es una clase de esquemas de fórmula lógica comúnmente llamados ítems que representan resultados intermedios del proceso de análisis.*
- *A es un conjunto de axiomas, esto es, ítems que representan resultados intermedios que son ciertos per se y que por tanto no precisan ser deducidos a partir de otros ítems.*
- *G es un conjunto de fórmulas meta, esto es, un conjunto de ítems que establecen si la cadena de entrada pertenece al lenguaje definido por la gramática.*
- *R es un conjunto de reglas de inferencia que permiten derivar nuevos ítems a partir de otros ítems. La forma general de tales reglas es*

$$\frac{A_1 \dots A_n}{B} \langle \text{condiciones de aplicación} \rangle$$

donde los antecedentes $A_1 \dots A_n$ antecedentes y el consecuente B son esquemas de fórmula, esto es, pueden contener metavariabes sintácticas que serán instanciadas por los términos apropiados cuando la regla sea utilizada.

La relación entre los esquemas de análisis sintáctico y los sistemas de deducción gramatical queda resumida en la tabla A.1.

Esquemas de análisis sintáctico	Sistemas de deducción gramatical
ítems	esquemas de fórmula lógica
pasos deductivos	reglas de inferencia
hipótesis+ítems iniciales	axiomas
ítems finales	fórmulas meta

Tabla A.1: Relación entre esquemas de análisis y sistemas de deducción gramatical

El motivo por el cual, aun siendo prácticamente equivalentes, nos hemos decantado por la utilización de los esquemas de análisis es porque consideramos que estos están más desarrollados, por cuanto se dispone de métodos de comprobación de la corrección y la completud [190] y se han descrito las reglas que permiten derivar esquemas de análisis correctos a partir de otros que ya son conocidos como correctos [189].

A.3. Transformación de esquemas de análisis sintáctico

Se pueden establecer varias clases de relaciones entre algoritmos de análisis sintáctico mediante la definición de relaciones entre los esquemas de análisis subyacentes. Concretamente, podemos *generalizar* y *filtrar* esquemas [189]. Antes de presentar estas relaciones, debemos realizar una serie de definiciones.

Definición A.5 Dado un sistema de análisis $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$ la relación $\vdash_{\subseteq} \wp_{fin}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}$ está definida por

$$Y \vdash \xi \text{ si } (Y', \xi) \in \mathcal{D} \text{ para algún } Y' \in Y$$

Debemos precisar que existe una distinción entre el conjunto de pasos deductivos \mathcal{D} y la relación de inferencia \vdash : si $\eta_1, \dots, \eta_k \vdash \xi$ entonces también se mantiene que $\eta_1, \dots, \eta_k \xi' \vdash \xi$ para cualquier ξ' . Por consiguiente, se define \vdash como el cierre de \mathcal{D} bajo la adición de antecedentes a una inferencia.

Definición A.6 Dado un sistema de análisis $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$, una secuencia deductiva en \mathbb{P} es un par $(Y; \xi_1, \dots, \xi_j) \in \wp_{fin}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}^+$. Por cuestiones prácticas, escribiremos $Y \vdash \xi_1 \vdash \dots \vdash \xi_j$ en lugar de $(Y; \xi_1, \dots, \xi_j)$.

Definición A.7 Dado un sistema de análisis $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$, el conjunto de secuencias deductivas $\Delta \subseteq \wp_{fin}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}^+$ se define como

$$\Delta = \{(Y; \xi_1, \dots, \xi_j) \in \wp_{fin}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}^+ \mid Y \vdash \xi_1 \vdash \dots \vdash \xi_j\}$$

Definición A.8 Dado un sistema de análisis $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$, la relación $\vdash_{\subseteq}^* \wp_{fin}(\mathcal{H} \cup \mathcal{I}) \times \mathcal{I}$ se define como

$$Y \vdash_{\subseteq}^* \xi \text{ si } \xi \in Y \text{ ó } Y \vdash \dots \vdash \xi$$

A.3.1. Generalización

Existen tres tipos de relaciones de generalización:

- *Refinamiento de los ítems*: un esquema de análisis sintáctico \mathbf{B} es el resultado de aplicar un refinamiento a los ítems del esquema de análisis \mathbf{A} , denotado por $\mathbf{A} \xrightarrow{\text{ir}} \mathbf{B}$, si un ítem individual de \mathbf{A} es dividido en varios ítems en \mathbf{B} y los pasos deductivos son adaptados de acuerdo con los cambios producidos. La relación $\mathbf{A} \xrightarrow{\text{ir}} \mathbf{B}$ se mantiene si $\mathbb{P}_{\mathbf{A}} \xrightarrow{\text{ir}} \mathbb{P}_{\mathbf{B}}$ para toda gramática y toda cadena de entrada. A su vez, la relación de refinamiento $\mathbb{P}_{\mathbf{A}} \xrightarrow{\text{ir}} \mathbb{P}_{\mathbf{B}}$ entre sistemas de análisis sintáctico es válida si existe una función f de $\mathcal{I}_{\mathbf{B}}$ en $\mathcal{I}_{\mathbf{A}}$, extensible a secuencias de pasos deductivos Δ , tal que

$$\mathcal{I}_{\mathbf{A}} = f(\mathcal{I}_{\mathbf{B}})$$

$$\Delta_{\mathbf{A}} = f(\Delta_{\mathbf{B}})$$

- *Refinamiento de los pasos*: un esquema de análisis sintáctico \mathbf{B} es el resultado de aplicar un refinamiento a los pasos del esquema de análisis \mathbf{A} , denotado por $\mathbf{A} \xrightarrow{\text{sr}} \mathbf{B}$, si un paso deductivo individual en \mathbf{A} es dividido en varios pasos deductivos en \mathbf{B} , teniendo en cuenta que se pueden introducir nuevos ítems siempre que sea necesario almacenar nuevos resultados intermedios. La relación $\mathbf{A} \xrightarrow{\text{sr}} \mathbf{B}$ se cumple si $\mathbb{P}_{\mathbf{A}} \xrightarrow{\text{sr}} \mathbb{P}_{\mathbf{B}}$ para toda gramática

y toda cadena de entrada. A su vez, la relación de refinamiento $\mathbb{P}_A \xrightarrow{\text{sr}} \mathbb{P}_B$ entre sistemas de análisis sintáctico es válida si

$$\begin{aligned} \mathcal{I}_A &\subseteq \mathcal{I}_B \\ \vdash_A^* &\subseteq \vdash_B^* \end{aligned}$$

donde la relación \vdash^* en $\wp_{fin}(\mathcal{I} \cup \mathcal{H}) \times \mathcal{I}$ para un sistema de análisis sintáctico $\mathbb{P} = \langle \mathcal{I}, \mathcal{H}, \mathcal{D} \rangle$ se define como $Y \vdash^* \xi$ si $\xi \in Y$ ó $Y \vdash \dots \vdash \xi$.

- *Extensión:* un esquema de análisis \mathbf{B} es una extensión del esquema \mathbf{A} , denotado por $\mathbf{A} \xrightarrow{\text{ext}} \mathbf{B}$, si está definido para una clase más amplia de gramáticas. Esta relación es válida si $\mathbf{A}(\mathcal{G}) = \mathbf{B}(\mathcal{G})$ para toda gramática \mathcal{G} para la cual \mathbf{A} está definida.

Proposición A.1 Cada una de las relaciones $\xrightarrow{\text{ir}}$, $\xrightarrow{\text{sr}}$ y $\xrightarrow{\text{ext}}$ posee las propiedades reflexiva y transitiva.

Proposición A.2 Si $\mathbf{A} \xrightarrow{\text{ir}} \mathbf{B}$ entonces la corrección y completud de \mathbf{B} implica la corrección y completud de \mathbf{A} , pero no necesariamente a la inversa.

Proposición A.3 La relación $\xrightarrow{\text{sr}}$ preserva la corrección y completud: si $\mathbf{A} \xrightarrow{\text{sr}} \mathbf{B}$ entonces la corrección y completud de \mathbf{A} implica la corrección y completud de \mathbf{B} .

Proposición A.4 Si $\mathbf{A} \xrightarrow{\text{sr}} \mathbf{B} \xrightarrow{\text{ir}} \mathbf{C}$ entonces existe un esquema \mathbf{D} tal que $\mathbf{A} \xrightarrow{\text{ir}} \mathbf{D} \xrightarrow{\text{sr}} \mathbf{C}$.

A.3.2. Filtrado

Mediante la generalización se pueden obtener mejoras cualitativas de un esquema de análisis, puesto que se obtiene un control más fino sobre los ítems y los pasos deductivos. En cambio, si lo que se desea es mejorar cuantitativamente un esquema de análisis, entonces deberemos reducir el número de ítems y pasos deductivos. Esto es posible mediante la realización de uno de los siguientes tipos de relaciones de filtrado:

- *Filtrado estático:* las partes redundantes de un esquema son descartadas sin más. La relación $\mathbf{A} \xrightarrow{\text{sf}} \mathbf{B}$ se mantiene si $\mathbb{P}_A \xrightarrow{\text{sf}} \mathbb{P}_B$ para toda gramática y cadena de entrada. A su vez, la relación de refinamiento $\mathbb{P}_A \xrightarrow{\text{sf}} \mathbb{P}_B$ entre sistemas de análisis sintáctico es válida si

$$\begin{aligned} \mathcal{I}_A &\supseteq \mathcal{I}_B \\ \mathcal{D}_A &\supseteq \mathcal{D}_B \end{aligned}$$

- *Filtrado dinámico:* la validez de algunos ítems puede hacerse depender a su vez de la validez de otros ítems. Por consiguiente, en este caso se aplica información acerca del contexto para determinar si se puede aplicar un determinado paso deductivo. La relación $\mathbf{A} \xrightarrow{\text{df}} \mathbf{B}$ se mantiene si $\mathbb{P}_A \xrightarrow{\text{df}} \mathbb{P}_B$ para toda gramática y cadena de entrada. A su vez, la relación de refinamiento $\mathbb{P}_A \xrightarrow{\text{df}} \mathbb{P}_B$ entre sistemas de análisis sintáctico es válida si

$$\begin{aligned} \mathcal{I}_A &\supseteq \mathcal{I}_B \\ \vdash_A &\supseteq \vdash_B \end{aligned}$$

- *contracción de los ítems*: Varios ítems son reemplazados por un único ítem del esquema de análisis. La relación $\mathbf{A} \xrightarrow{\text{ic}} \mathbf{B}$ se mantiene si $\mathbb{P}_A \xrightarrow{\text{ic}} \mathbb{P}_B$ para toda gramática y cadena de entrada, para la cual debe existir una función f de \mathcal{I}_A en \mathcal{I}_B , extensible a secuencias de pasos deductivos Δ , tal que

$$\mathcal{I}_B = f(\mathcal{I}_A)$$

$$\Delta_B = f(\Delta_A)$$

- *Contracción de pasos*: secuencias de pasos deductivos son reemplazadas por un único paso deductivo. La relación $\mathbf{A} \xrightarrow{\text{sc}} \mathbf{B}$ se mantiene si $\mathbb{P}_A \xrightarrow{\text{sc}} \mathbb{P}_B$ para toda gramática y cadena de entrada. A su vez, la relación de refinamiento $\mathbb{P}_A \xrightarrow{\text{sc}} \mathbb{P}_B$ entre sistemas de análisis sintáctico es válida si

$$\mathcal{I}_A \supseteq \mathcal{I}_B$$

$$\vdash_A^* \supseteq \vdash_B^*$$

Proposición A.5 $\xrightarrow{\text{sf}} \subseteq \xrightarrow{\text{df}} \subseteq \xrightarrow{\text{sc}}$.

Proposición A.6 Cada una de las relaciones $\xrightarrow{\text{sf}}$, $\xrightarrow{\text{df}}$, $\xrightarrow{\text{ic}}$ y $\xrightarrow{\text{sc}}$ posee las propiedades reflexiva y transitiva.

Proposición A.7 Las relaciones $\xrightarrow{\text{sf}}$, $\xrightarrow{\text{df}}$ y $\xrightarrow{\text{sc}}$ preservan la corrección, aunque no necesariamente la completud.

Proposición A.8 La relación $\xrightarrow{\text{ic}}$ preserva la corrección y completud.

Proposición A.9 La relación $\mathbf{A} \xrightarrow{\text{ir}} \mathbf{B}$ es válida si y sólo si $\mathbf{B} \xrightarrow{\text{ic}} \mathbf{A}$.

Proposición A.10 La relación $\mathbf{A} \xrightarrow{\text{sr}} \mathbf{B}$ es válida si y sólo si $\mathbf{B} \xrightarrow{\text{sc}} \mathbf{A}$.

A.4. Esquemas de análisis sintáctico para autómatas a pila

El método de interpretación dinámica para autómatas a pila propuesto en [104, 107] puede verse como un sistema de inferencia en el cual las transiciones del autómata juegan el papel de las reglas de inferencia [52]. Puesto que los esquemas de análisis sintáctico también son sistemas de inferencia es posible utilizar este formalismo para describir interpretaciones de autómatas a pila en programación dinámica. Abusando del formalismo, utilizaremos la misma notación de los pasos deductivos para describir las transiciones del autómata. En la página 429 puede encontrarse un ejemplo de una descripción de autómata a pila utilizando la estructura de los esquemas de análisis sintáctico.

A.5. Esquemas de análisis sintáctico para TAG y LIG

Los esquemas de análisis sintáctico fueron diseñados para representar algoritmos de análisis sintáctico de gramáticas independientes del contexto. Sin embargo, no resulta complicado extenderlos a formalismos gramaticales que son a su vez extensiones de las gramáticas independientes al contexto, como es el caso de las gramáticas de adjunción de árboles (TAG) y de las gramáticas lineales de índices (LIG). Los mecanismos operaciones de los esquemas de análisis seguirán siendo válidos aunque debemos resaltar que:

- los ítems serán más complejos que los utilizados en el caso de gramáticas independientes del contexto, puesto que los resultados intermedios de los algoritmos de análisis sintáctico deberán tomar nota de ciertas dependencias contextuales;
- las pruebas de validez y corrección tendrán que realizarse de acuerdo a las características propias del formalismo gramatical utilizado en cada momento.

El aumento del tamaño de los ítems, consecuencia del aumento de la información contextual utilizada, puede hacer difícil la lectura de los algoritmos si utilizamos la notación mostrada anteriormente para los pasos deductivos. Aunque ello no presenta inconvenientes desde el punto de vista meramente formal, atenta en cierta medida contra la intención de los esquemas de análisis de proporcionar una visión clara del funcionamiento de los diferentes algoritmos de tal modo que aquellos que no estén interesados en detalles acerca de la corrección y la validez puedan comprender sin demasiada dificultad las bases que rigen el comportamiento de cada algoritmo. Es por ello que en el caso de TAG y LIG hemos decidido utilizar la siguiente notación

$$\frac{I_1, I_2, \dots, I_n}{I_{n+1}} \langle \text{condición} \rangle$$

para los pasos deductivos en lugar de la clásica

$$I_1, I_2, \dots, I_n \vdash I_{n+1} \mid \langle \text{condición} \rangle$$

donde $I_1 \dots I_n$ son los antecedentes, I_{n+1} es el consecuente y $\langle \text{condición} \rangle$ es el filtro dinámico aplicado al paso deductivo.

A.6. Análisis de complejidad

A.6.1. Complejidad espacial

El análisis de la complejidad espacial es sencillo, pues para calcular el espacio total necesario para almacenar todos los posibles ítems es suficiente con multiplicar las cotas superiores de los distintos componentes. Por ejemplo, dado un ítem $[h, B^\gamma \rightarrow \delta \bullet \nu, k, l \mid p, q]$ tenemos que:

- Los componentes h, k, l, p y q son 5 posiciones de la cadena de entrada y por tanto su cota es la longitud n de la cadena de entrada.
- El componente $B^\gamma \rightarrow \delta \bullet \nu$ depende del tamaño de la gramática y por tanto su cota superior es dicho tamaño $|G|$.

Por tanto, la complejidad espacial viene dada por $\mathcal{O}(|G|n^5)$.

A.6.2. Complejidad temporal

Nos referiremos únicamente a la complejidad con respecto a la cadena de entrada, que es la medida usualmente utilizada en la literatura. Dado un algoritmo de análisis sintáctico definido por un esquema de análisis, su complejidad viene dada por el paso deductivo \mathcal{D} más complejo. En una primera aproximación, la complejidad de \mathcal{D} vendrá dada por $\mathcal{O}(n^p)$, donde p es el máximo número de posiciones de la cadena de entrada utilizadas por los antecedentes de \mathcal{D} . Sin embargo, un análisis más refinado nos lleva a excluir aquellas posiciones que ocurren una sola vez en \mathcal{D} , a las que llamaremos *sin-importancia* [125]. Esto es así puesto que se podría aplicar un paso intermedio implícito $I \vdash I'$ que al ser aplicado redujese un ítem I con q posiciones a otro ítem

I' con q' posiciones, con $q' \leq q$, mediante la omisión de las posiciones sin-importancia. El ítem I' tomaría entonces el lugar de I en la parte antecedente de \mathcal{D} .

Por ejemplo, el paso deductivo

$$\mathcal{D} = \frac{\begin{array}{l} [h, B^\gamma \rightarrow \delta \bullet, k, l \mid p, q], \\ [j, \mathbf{F}^\beta \rightarrow \bullet \perp, k, k \mid -, -], \\ [h, A^\gamma \rightarrow \delta \bullet B^\gamma \nu, i, j \mid p', q'] \end{array}}{[j, \mathbf{F}^\beta \rightarrow \perp \bullet, k, l \mid k, l]}$$

que tiene 9 posiciones de la cadena de entrada, presenta un complejidad $\mathcal{O}(n^4)$ puesto que i, p, q, p' y q' son posiciones sin-importancia ya que aparecen una sola vez en \mathcal{D} . En consecuencia, para aplicar este paso es suficiente con 4 bucles anidados, cada uno tomando uno de las posiciones h, k, l y j como variable de control.

En ciertos casos la complejidad de un paso deductivo se puede reducir mediante *aplicación parcial*. Un paso deductivo no tiene porqué ser aplicado necesariamente en una sola operación, sino que se puede convertir en una secuencia de pasos intermedios, en cada uno de los cuales se toma un subconjunto de los antecedentes del paso original y se obtiene un resultado en la forma de un ítem intermedio, encargado de transmitir la información entre pasos intermedios. El resultado del paso deductivo original será igual al resultado del último paso deductivo intermedio. Esta transformación es correcta si todos los ítems antecedentes, así como las condiciones que les afectan, son tomados en cuenta en al menos un paso intermedio. Por ejemplo, el paso deductivo

$$\mathcal{D} = \frac{\begin{array}{l} [\top \rightarrow \mathbf{R}^{\beta \bullet}, j, m \mid k, l], \\ [M^\gamma \rightarrow \delta \bullet, k, l \mid -, -], \\ [N^\gamma \rightarrow \delta \bullet M^\gamma \nu, i, j \mid p, q] \end{array}}{[N^\gamma \rightarrow \delta M^\gamma \bullet \nu, i, m \mid p, q]} \quad \beta \in \text{adj}(M^\gamma)$$

presenta a priori una complejidad $\mathcal{O}(n^7)$ puesto que las 7 posiciones j, m, k, l, i, p y q son utilizadas para determinar la validez de los ítems a combinar. Sin embargo, dicho paso deductivo es equivalente a la aplicación consecutiva de los siguientes pasos \mathcal{D}^1 y \mathcal{D}^2 , el primero de complejidad $\mathcal{O}(n^4)$ y el segundo de complejidad $\mathcal{O}(n^5)$, por lo que la complejidad conjunta de ambos es $\mathcal{O}(n^5)$. El ítem intermedio producido se distingue por la utilización de dobles corchetes:

$$\mathcal{D}^1 = \frac{\begin{array}{l} [\top \rightarrow \mathbf{R}^{\beta \bullet}, j, m \mid k, l], \\ [M^\gamma \rightarrow \delta \bullet, k, l \mid -, -], \end{array}}{[[\beta, N^\gamma \rightarrow \delta M^\gamma \bullet \nu, j, m]]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^2 = \frac{[[\beta, N^\gamma \rightarrow \delta M^\gamma \bullet \nu, j, m]]}{[N^\gamma \rightarrow \delta \bullet M^\gamma \nu, i, j \mid p, q]} \quad \beta \in \text{adj}(M^\gamma)$$