Capítulo 3

Algoritmos de análisis sintáctico para TAG

Se describen varios de los algoritmos de análisis sintáctico que han sido propuestos para las gramáticas de adjunción de árboles desde que éstas fueron por descritas por primera vez en [93]. La mayor parte de estos algoritmos están basados en algoritmos de análisis sintáctico independientes del contexto extendidos para el tratamiento de la adjunción. Asimismo es de destacar que prácticamente todos estos algoritmos utilizan técnicas de programación dinámica, obteniendo una complejidad polinómica de orden $\mathcal{O}(n^6)$, donde n es la longitud de la cadena de entrada. La aportación de este capítulo es múltiple: se proporciona un camino evolutivo que relaciona los algoritmos de análisis de TAG más populares; se describen por vez primera algunos algoritmos de análisis de TAG, como por ejemplo, las diferentes versiones de los algoritmos de tipo Earley ascendente y las versiones propuestas de los algoritmos de tipo Earley sin la propiedad del prefijo válido; por último, se realiza una descripción conjunta de la mayor parte de los algoritmos de análisis existentes para TAG. Este capítulo está basado en [8, 9].

3.1 Introducción

A la hora de describir los algoritmos de análisis para TAG, tenemos que elegir una representación adecuada para indicar el reconocimiento parcial de los árboles elementales. En las gramáticas independientes del contexto se utilizan habitualmente producciones con punto para separar la parte de la producción que ya ha sido procesada de la que no lo ha sido aún. En los algoritmos que proceden unidireccionalmente, un solo punto es suficiente, mientras que en aquellos que proceden bidireccionalmente se necesitan dos puntos [189]. En el caso de TAG, al no ser las estructuras elementales producciones sino árboles, deberemos representar árboles con punto. Existen varias alternativas:

1. Al igual que en las gramáticas independientes del contexto se escriben producciones completas con punto, en TAG se podría escribir cada árbol elemental completo con su punto correspondiente. Un ejemplo de utilización de este tipo de notación puede encontrarse en [176].

2. Si identificamos unívocamente todo elemento en una producción independiente del contexto, para lo cual podemos utilizar subíndices correspondientes al número de la producción y a la posición que cada elemento ocupa en al producción de tal modo que la producción $k : N \to MPQ$ se representaría como $N_{k,0} \to N_{k,1}N_{k,2}N_{k,3}$, podemos representar una producción con punto simplemente indicando un elemento de la producción contiguo al

punto e indicando si este se encuentra situado a derecha o izquierda de dicho elemento. Del mismo modo, podemos indicar la posición del punto en un árbol elemental de una TAG indicando simplemente un nodo del árbol adyacente al punto y la posición relativa del punto respecto a dicho nodo: arriba-izquierda, abajo-izquierda, abajo-derecha o arriba-derecha. Este es el tipo de notación utilizada en [168].

3. Como un árbol elemental γ puede considerarse constituido por un conjunto de producciones independientes del contexto $\mathcal{P}(\gamma)$, podemos indicar la posición del punto en el árbol simplemente indicando la posición en la producción correspondiente. Este tipo de representación recibe el nombre de *multicapa* en [64].

4. Aplicar un aplanamiento a los árboles, esto es, una conversión de los mismos a cadenas de caracteres parentizadas en las cuales cada nuevo nivel de paréntesis indica un nivel adicional de profundidad en el árbol elemental. Este tipo de representación, que recibe el nombre de *plana* en [64], implica que cada árbol inicial se representa únicamente por una cadena de caracteres mientras que cada árbol auxiliar se representa por dos cadenas correspondientes a las partes situadas a la izquierda y a la derecha de la espina. Dichas cadenas pueden ser utilizadas como partes derechas de las producciones de una gramática independiente del contexto que permitiría representar de modo conciso los árboles elementales de una gramática de adjunción [64].

La primera alternativa presenta el inconveniente de que las representaciones lineales de árboles suelen dificultar su comprensión. Alternativamente podrían utilizarse representaciones pictóricas de los árboles, que si bien son incómodas a lo hora de describir los pasos deductivos del algoritmo pueden ser útiles para representar gráficamente el comportamiento del algoritmo. La segunda alternativa, aunque muy manejable y concisa tiene el inconveniente de que es necesario recurrir a las descripciones de los árboles elementales para poder ver el contexto en el que se está aplicando cada paso del algoritmo. La tercera alternativa resuelve este problema al proporcionar un contexto formado por los hermanos del nodo considerado, que si bien es limitado, proporciona al lector la información suficiente la mayor parte de las veces. La desventaja es que normalmente deberemos añadir pasos deductivos extra que realizan el recorrido del árbol en un orden determinado. La cuarta alternativa, aunque en principio supone una reducción del número de producciones independientes al contexto, en la práctica su utilización no implica generalmente una reducción del número de pasos deductivos utilizados para describir el procesamiento realizado por los algoritmos de análisis sintáctico[1] y contribuye a oscurecer la notación.

En nuestro caso hemos preferido utilizar la representación multicapa, por lo que indicaremos la posición del punto en un árbol mediante una producción $N \to \delta \bullet \nu$, donde $\delta\nu$ son los hijos de N. En el caso de los elementos del lado derecho de las producciones cuyas etiquetas pertenecen a $V_T \cup \epsilon$, puesto que no pueden tener descendientes ni son susceptibles de ser nodos de adjunción, consideraremos directamente su etiqueta a la hora de describir las producciones, en lugar del nodo propiamente dicho[2]. Adicionalmente y con el fin de simplificar los esquemas de análisis de los algoritmos más complejos, seguiremos el enfoque de [125] de considerar la producción adicional $\top \to \mathbf{R}^\alpha$ para cada árbol inicial α y las dos producciones adicionales siguientes para cada árbol auxiliar β: $\top \to \mathbf{R}^\beta$ y $\mathbf{F}^\beta \to \bot$, donde $\mathbf{R}^\beta$ y $\mathbf{F}^\beta$ se refieren a los nodos raíz y pie de β, respectivamente. Con el fin de no modificar la capacidad generativa de las gramáticas,

[1] Como ejemplo, podemos comparar el algoritmo descrito en [64] con el representado por el esquema 3.7.

[2] Con esta convención, que se adopta porque simplifica considerablemente la notación, la única forma de distinguir diferentes hojas de un árbol elemental con la misma etiqueta es referenciando los respectivos padres o la producción independiente del contexto de la que forman parte.

los nuevos nodos $\top$ y $\perp$ no pueden ser nodos de adjunción, por lo cual la nueva gramática se asemeja a una TAG *limpia*[3].

Con respecto a las restricciones de adjunción, $\beta \in \text{adj}(N^\gamma)$ si $\beta \in \boldsymbol{A}$ puede ser adjuntado en N^γ. Si $\beta = \textbf{nil}$ entonces la adjunción en el nodo N^γ es opcional. Si **nil** es el único valor que retorna la función para un nodo N^γ, entonces no es posible realizar ninguna adjunción en dicho nodo.

Ejemplo 3.1 El árbol inicial α de la figura 3.1 se representa en notación multicapa mediante el siguiente conjunto de producciones independientes del contexto:

$$\top \to \langle \alpha, 0 \rangle$$

$$\langle \alpha, 0 \rangle \to a\ \langle \alpha, 2 \rangle$$

$$\langle \alpha, 2 \rangle \to c$$

El árbol auxiliar β de la misma figura se representa mediante el siguiente conjunto de producciones:

$$\top \to \langle \beta, 0 \rangle$$

$$\langle \beta, 0 \rangle \to a\ \langle \beta, 2 \rangle$$

$$\langle \beta, 2 \rangle \to \langle \beta, 2.1 \rangle\ c$$

$$\langle \beta, 2.1 \rangle \to \perp$$

En ambos casos $\langle \alpha, g \rangle$ denota el nodo de α que ocupa la posición g según el direccionamiento de Gorn[4]. ¶

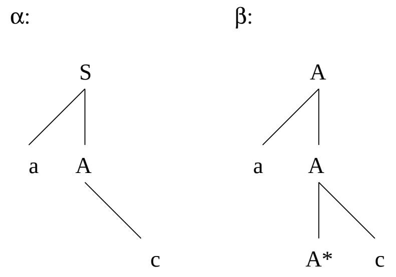

Figura 3.1: Ejemplos de árbol inicial y auxiliar

La relación $\Rightarrow$ de derivación sobre $\mathcal{P}(\gamma)$ se define como $\delta \Rightarrow \nu$ si existen $\delta', \delta'', M^\gamma, \upsilon$ tal que $\delta = \delta' M^\gamma \delta''$, $\nu = \delta' \upsilon \delta''$ y existe una producción $M^\gamma \to \upsilon \in \mathcal{P}(\gamma)$. Mediante $\overset{*}{\Rightarrow}$ denotaremos el cierre reflexivo y transitivo de $\Rightarrow$.

Sea $\mathcal{P}(\mathcal{T}) = \bigcup_{\gamma \in \boldsymbol{I} \cup \boldsymbol{A}} \mathcal{P}(\gamma)$ el conjunto de todas las producciones independientes del contexto presentes en los árboles de una gramática de adjunción de árboles $\mathcal{T}$. Con el fin de ser capaces de representar derivaciones que incluyan adjunciones extendemos la relación $\Rightarrow$ a $\mathcal{P}(\mathcal{T})$, de tal modo que $\delta \overset{*}{\Rightarrow} \nu$ si existen $\delta', \delta'', M^\gamma, \upsilon$ tal que $\delta = \delta' M^\gamma \delta''$, $\mathbf{R}^\beta \overset{*}{\Rightarrow} \upsilon_1 \mathbf{F}^\beta \upsilon_3$, $\beta \in \text{adj}(M^\gamma)$, $M^\gamma \to \upsilon_2$ y $\nu = \delta' \upsilon_1 \upsilon_2 \upsilon_3 \delta''$.

[3] Poller y Becker [149] denominan TAG *limpias* a las gramáticas de adjunción en las que los nodos raíz de los árboles elementales y los nodos pie de los árboles auxiliares tienen restricciones de adjunción nulas.

[4] En el direccionamiento de Gorn se utiliza 0 para referirse al nodo raíz, k para referirse al k-ésimo hijo del nodo raíz y $p.q$ para referirse al q-ésimo hijo del nodo con dirección p.

La mayoría de los algoritmos de análisis sintáctico diseñados para TAG se corresponden con adaptaciones de algoritmos para el análisis de gramáticas independientes del contexto y por lo tanto recibirán el mismo nombre en ambos casos. Solamente en aquellos casos en los que pueda existir confusión diferenciaremos explícitamente entre unos y otros[5]. Para describir los diferentes algoritmos de análisis sintáctico haremos uso de *esquemas de análisis*, una estructura para realizar descripciones de alto nivel de algoritmos de análisis desarrollada por Sikkel en [189] y que se describe en el apéndice A.

3.2 Algoritmo de tipo CYK

A continuación mostramos una extensión para TAG del algoritmo CYK de análisis sintáctico (ver sección B.1) basado en el algoritmo descrito en [209, 206] pero con algunas correcciones. Asumiremos que todo nodo de un árbol elemental de la gramática tiene a lo sumo dos descendientes. Este condicionante puede verse como una trasposición de la forma normal de Chomsky [85] al caso de las gramáticas de adjunción.

El reconocimiento de los nodos de un árbol elemental se realiza aplicando casi literalmente el algoritmo para el caso independiente del contexto. Sin embargo ahora es preciso definir cómo reconocer la operación de adjunción de forma totalmente ascendente. Para ello se definen ítems de la forma

$$\left\{ \begin{array}{ll} [N^\gamma, i, j \mid p, q \mid adj] \mid & N^\gamma \overset{*}{\Rightarrow} a_{i+1} \ldots a_p \, \mathbf{F}^\gamma \, a_{q+1} \ldots a_j \overset{*}{\Rightarrow} a_{i+1} \ldots a_j \quad \text{sii } (p,q) \neq (-,-) \\ & N^\gamma \overset{*}{\Rightarrow} a_{i+1} \ldots a_j \quad \text{sii } (p,q) = (-,-) \end{array} \right\}$$

que contienen un nuevo componente *adj* con respecto a los ítems definidos en [209, 206], que toma su valor del conjunto $\{\text{true}, \text{false}\}$ y que indica

- si su valor es *true* significa que el ítem es el resultado de una operación de adjunción ya totalmente realizada;
- si valor es *false* significa que el ítem no es el resultado de una adjunción.

Con ello podemos limitar a una sola la cantidad de adjunciones que se pueden realizar sobre cada nodo de un árbol elemental y podemos también asegurar el cumplimiento de las restricciones de adjunción obligatoria, ajustándonos de este modo a lo establecido en el formalismo de las gramáticas de adjunción [175].

Esquema de análisis sintáctico 3.1 El sistema de análisis $\mathbb{P}_{\text{CYK}}$ que se corresponde con el algoritmo CYK para una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\text{CYK}} = \left\{ \begin{array}{ll} [N^\gamma, i, j \mid p, q \mid adj] \mid & \text{etiqueta}(N^\gamma) \in V_N, \ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \ \ 0 \leq i \leq j, \\ & (p,q) \leq (i,j), \ adj \in \{\text{true}, \text{false}\} \end{array} \right\}$$

$$\mathcal{H}_{\text{CYK}} = \left\{ \ [a, i-1, i] \mid a = a_i, 1 \leq i \leq n \ \right\}$$

$$\mathcal{D}^{\text{Scan}}_{\text{CYK}} = \frac{[a, i, i+1]}{[N^\gamma, i, i+1 \mid -, - \mid \text{false}]} \quad N^\gamma \to a \in \mathcal{P}(\gamma)$$

[5]En el apéndice B se describen sucintamente los algoritmos de análisis sintántico para gramáticas independientes del contexto que son relevantes en este capítulo

$$\mathcal{D}^{\epsilon}_{\text{CYK}} = \frac{}{[N^\gamma, i, i \mid -, - \mid \text{false}]} \quad N^\gamma \to \epsilon \in \mathcal{P}(\gamma)$$

$$\mathcal{D}^{\text{Foot}}_{\text{CYK}} = \frac{}{[\mathbf{F}^\gamma, i, j \mid i, j \mid \text{false}]}$$

$$\mathcal{D}^{\text{LeftDom}}_{\text{CYK}} = \frac{[M^\gamma, i, k \mid p, q \mid \mathit{adj}1],\ [P^\gamma, k, j \mid -, - \mid \mathit{adj}2]}{[N^\gamma, i, j \mid p, q \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma), \\ M^\gamma \in \text{espina}(\gamma), \\ \mathit{adj}1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma), \\ \mathit{adj}2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma) \end{array}$$

$$\mathcal{D}^{\text{RightDom}}_{\text{CYK}} = \frac{[M^\gamma, i, k \mid -, - \mid \mathit{adj}1],\ [P^\gamma, k, j \mid p, q \mid \mathit{adj}2]}{[N^\gamma, i, j \mid p, q \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma), \\ P^\gamma \in \text{espina}(\gamma), \\ \mathit{adj}1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma), \\ \mathit{adj}2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma) \end{array}$$

$$\mathcal{D}^{\text{NoDom}}_{\text{CYK}} = \frac{[M^\gamma, i, k \mid -, - \mid \mathit{adj}1],\ [P^\gamma, k, j \mid -, - \mid \mathit{adj}2]}{[N^\gamma, i, j \mid -, - \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma), \\ N^\gamma \notin \text{espina}(\gamma), \\ \mathit{adj}1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma), \\ \mathit{adj}2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma) \end{array}$$

$$\mathcal{D}^{\text{Unary}}_{\text{CYK}} = \frac{[M^\gamma, i, j \mid p, q \mid \mathit{adj}]}{[N^\gamma, i, j \mid p, q] \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma \in \mathcal{P}(\gamma), \\ \mathit{adj} = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma) \end{array}$$

$$\mathcal{D}^{\text{Adj}}_{\text{CYK}} = \frac{[\mathbf{R}^\beta, i', j' \mid i, j \mid \mathit{adj}],\ [N^\gamma, i, j \mid p, q \mid \text{false}]}{[N^\gamma, i', j' \mid p, q \mid \text{true}]} \quad \beta \in \boldsymbol{A},\ \beta \in \text{adj}(N^\gamma)$$

$$\mathcal{D}_{\text{CYK}} = \mathcal{D}^{\text{Scan}}_{\text{CYK}} \cup \mathcal{D}^{\epsilon}_{\text{CYK}} \cup \mathcal{D}^{\text{Foot}}_{\text{CYK}} \cup \mathcal{D}^{\text{LeftDom}}_{\text{CYK}} \cup \mathcal{D}^{\text{RightDom}}_{\text{CYK}} \cup \mathcal{D}^{\text{NoDom}}_{\text{CYK}} \cup \mathcal{D}^{\text{Unary}}_{\text{CYK}} \cup \mathcal{D}^{\text{Adj}}_{\text{CYK}}$$

$$\mathcal{F}_{\text{CYK}} = \left\{ \ [\mathbf{R}^\alpha, 0, n \mid -, - \mid \mathit{adj}] \mid \alpha \in \boldsymbol{I},\ \mathit{adj} = \text{false sii } \mathbf{nil} \in \text{adj}(\mathbf{R}^\alpha) \ \right\}$$

donde $(p, q) \leq (i, j)$ se cumple si $i \leq p \leq q \leq j$ en el caso de que $(p, q) \neq (-, -)$ y si $i \leq j$ en el caso de que $(p, q) = (-, -)$. §

La definición de las hipótesis realizada en este sistema de análisis sintáctico se corresponde con la estándar y es la misma que se utilizará en los restantes sistemas de análisis del capítulo. Por consiguiente, no nos volveremos a referir explícitamente a ellas.

Los pasos clave en el sistema de análisis $\mathbb{P}_{\text{CYK}}$ son $\mathcal{D}^{\text{Foot}}_{\text{CYK}}$ y $\mathcal{D}^{\text{Adj}}_{\text{CYK}}$, puesto que son los que definen el mecanismo de adjunción. Los demás pasos se encargan de recorrer de forma ascendente los árboles elementales, propagando la información relativa a la porción de la cadena de entrada cubierta por el pie en el caso de los árboles auxiliares. El paso deductivo $\mathcal{D}^{\text{Scan}}_{\text{CYK}}$ permite resolver la limitación planteada en [209, 206] con respecto a nodos frontera etiquetados con ϵ.

El paso deductivo $\mathcal{D}^{\text{Foot}}_{\text{CYK}}$ hace posible el análisis ascendente de los árboles auxiliares, puesto que predice todas las posibles porciones de la cadena de entrada que puede cubrir el pie. Si no fuese así, sería necesario esperar al reconocimiento total del pie antes de poder continuar el reconocimiento del árbol auxiliar, con lo cual el algoritmo de análisis no sería totalmente ascendente.

Una de las consecuencias de la utilización del paso $\mathcal{D}_{\mathrm{CYK}}^{\mathrm{Foot}}$ es que existirán múltiples análisis diferentes para cada árbol auxiliar, diferentes únicamente en las posición de la cadena de entrada que se *supone* en cada caso que cubre el pie. Pero finalmente sólo uno o unos pocos de esos árboles podrán formar parte del árbol de derivación, aquellos que hayan predicho correctamente el pie. La realización de esta comprobación le corresponde al paso $\mathcal{D}_{\mathrm{CYK}}^{\mathrm{Adj}}$. Efectivamente, cuando se ha alcanzado la raíz de un árbol auxiliar, se comprueba si existe un subárbol de un árbol elemental cuya raíz pueda ser un nodo de adjunción para dicho árbol auxiliar y que además cubra la porción de la cadena de entrada que espera ser cubierta por el pie. En caso de que así sea, se eliminará la posibilidad de realizar otras adjunciones en ese nodo y el análisis continuará ascendiendo por el nuevo árbol elemental. En la figura 3.2 se muestra gráficamente la aplicación de este paso, donde los dos árboles de la izquierda se corresponden con las descripciones de los ítems antecedentes y el árbol de la derecha con la descripción del ítem consecuente. Se utilizan líneas punteadas para indicar partes de árboles elementales que no han sido aún analizadas, mientras que las espinas se resaltan mediante líneas gruesas.

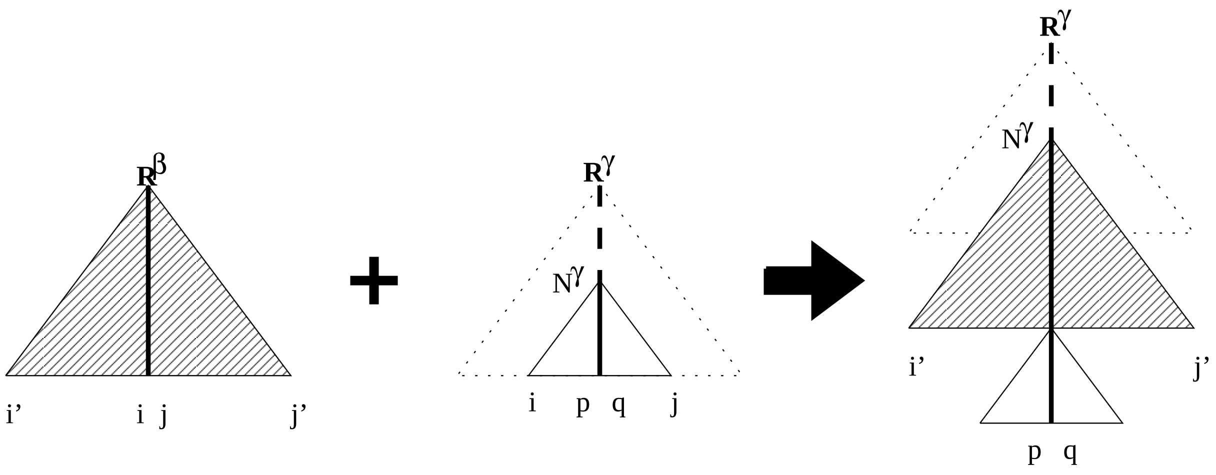

Figura 3.2: Descripción gráfica de un paso $\mathcal{D}_{\mathrm{CYK}}^{\mathrm{Adj}}$

La complejidad temporal del algoritmo con respecto a la longitud n de la cadena de entrada es $\mathcal{O}(n^6)$ y viene dada por el paso deductivo encargado de la adjunción, que debe combinar seis posiciones de la cadena de entrada. La complejidad espacial con respecto a la cadena de entrada es $\mathcal{O}(n^4)$, puesto que cada ítem almacena cuatro posiciones de la cadena de entrada.

3.3 Algoritmos de tipo Earley ascendente

El algoritmo CYK presenta un limitación muy importante: sólo es aplicable a gramáticas en las cuales un nodo no tiene más de dos descendentes. Nuestro objetivo ahora es extender el esquema **CYK** a la clase general de TAG, obteniendo lo que podríamos llamar un esquema Earley ascendente (ver sección B.2) extendido a gramáticas de adjunción de árboles. Debemos reseñar que no conocemos ninguna adaptación anterior para TAG de este algoritmo.

Como primer paso para la realización de un algoritmo de tipo Earley ascendente para gramáticas de adjunción de árboles precisamos introducir puntos en las producciones que representan los árboles elementales, por lo que los ítems que utilizaremos tendrán la forma

$$\left\{ \begin{array}{ll} [N^\gamma \to \delta \bullet \nu, i, j \mid p, q \mid adj] \mid & \delta \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_p \, \mathbf{F}^\gamma \, a_{q+1} \ldots a_j \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_j \quad \text{sii } (p,q) \neq (-,-) \\ & \delta \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_j \quad \text{sii } (p,q) = (-,-) \end{array} \right\}$$

Los ítems del nuevo esquema de análisis sintáctico, que denominaremos $\mathbf{buE_1}$, son por tanto un refinamiento de los ítems de **CYK**. Sobre los pasos deductivos aplicaremos también un *refinamiento* puesto que los pasos de tipo LeftDom, RightDom y NoDom serán separados en diferentes pasos de tipo Init y Comp, mientras que los pasos de tipo Unary y ϵ ya no serán necesarios puesto que todas las producciones serán tratadas uniformemente con independencia de su longitud. Finalmente, se realizará una extensión del dominio de las producciones permitiendo árboles con un grado arbitrario de ramificación.

Esquema de análisis sintáctico 3.2 El sistema de análisis $\mathbb{P}_{\text{buE}_1}$ que se corresponde con una primera extensión del algoritmo de Earley ascendente para TAG, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\text{buE}_1} = \left\{ \begin{array}{ll} [N^\gamma \to \delta \bullet \nu, i, j \mid p, q \mid adj] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \\ & 0 \leq i \leq j,\ (p,q) \leq (i,j),\ adj \in \{\text{true}, \text{false}\} \end{array} \right\}$$

$$\mathcal{D}^{\text{Init}}_{\text{buE}_1} = \frac{}{[N^\gamma \to \bullet\delta, i, i \mid -,- \mid \text{false}]}$$

$$\mathcal{D}^{\text{Foot}}_{\text{buE}_1} = \frac{}{[\mathbf{F}^\beta \to \bot\bullet, i, j \mid i, j \mid \text{false}]}$$

$$\mathcal{D}^{\text{Scan}}_{\text{buE}_1} = \frac{[N^\gamma \to \delta \bullet a\nu, i, j \mid p, q \mid \text{false}],\ [a, j, j+1]}{[N^\gamma \to \delta a \bullet \nu, i, j+1 \mid p, q \mid \text{false}]}$$

$$\mathcal{D}^{\text{Comp}}_{\text{buE}_1} = \frac{\begin{array}{l}[N^\gamma \to \delta \bullet M^\gamma\nu, i, k \mid p', q' \mid \text{false}], \\ [M^\gamma \to \upsilon\bullet, i, j \mid p, q \mid adj]\end{array}}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \text{false}]} \quad adj = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{Adj}}_{\text{buE}_1} = \frac{[\top \to \mathbf{R}^\beta\bullet, k, j \mid k', j' \mid \text{false}],\ [M^\gamma \to \upsilon\bullet, k', j' \mid p, q \mid \text{false}]}{[M^\gamma \to \upsilon\bullet, k, j \mid p, q \mid \text{true}]} \quad \beta \in \boldsymbol{A},\ \beta \in \text{adj}(\gamma)$$

$$\mathcal{D}_{\text{buE}_1} = \mathcal{D}^{\text{Init}}_{\text{buE}_1} \cup \mathcal{D}^{\text{Foot}}_{\text{buE}_1} \cup \mathcal{D}^{\text{Scan}}_{\text{buE}_1} \cup \mathcal{D}^{\text{Comp}}_{\text{buE}_1} \cup \mathcal{D}^{\text{Adj}}_{\text{buE}_1}$$

$$\mathcal{F}_{\text{buE}_1} = \left\{ \ [\top \to \mathbf{R}^\alpha\bullet, 0, n \mid -,- \mid false] \mid \alpha \in \boldsymbol{I} \ \right\}$$

y donde $p \cup q$ se refiere a la operación de unión de índices, función parcial de enteros a enteros definida como

$$p \cup q = \left\{ \begin{array}{ll} p & \text{si } q = - \\ q & \text{si } p = - \\ - & \text{si } p = q = - \end{array} \right.$$

donde $-$ se utiliza para indicar que el valor de uno de los parámetros está indefinido. §

Los pasos $\mathcal{D}^{\text{Init}}_{\text{buE}_1}$ son los encargados de lanzar el análisis ascendente. Los ítems generados por estos pasos son siempre válidos, puesto que no expanden ninguna porción la cadena de entrada. Tan sólo expresan la expectativa de que una determinada producción pueda ser aplicada para reconocer una porción de la cadena que comienza en una determinada posición.

Los pasos $\mathcal{D}^{\text{Foot}}_{\text{buE}_1}$ son utilizados, al igual que en el caso CYK, para predecir la porción de la cadena de entrada cubierta por el pie de un árbol auxiliar.

El algoritmo procede a reconocer ascendentemente los árboles auxiliares mediante la aplicación de pasos $\mathcal{D}^{\text{Comp}}_{\text{buE}_1}$, propagando también la información correspondiente a la porción de la cadena de entrada cubierta por el pie en el caso de los árboles auxiliares.

El paso deductivo $\mathcal{D}^{\text{Adj}}_{\text{buE}_1}$ se comporta de modo idéntico al paso homónimo del algoritmo CYK, comprobando que las predicciones respecto al pie se corresponden con las de una adjunción que se ha realizado realmente.

Proposición 3.1 $\mathbf{CYK} \stackrel{\text{ir}}{\Longrightarrow} \mathbf{CYK'} \stackrel{\text{sr}}{\Longrightarrow} \mathbf{ECYK} \stackrel{\text{ext}}{\Longrightarrow} \mathbf{buE_1}$.

Demostración:

Como primer paso definiremos el sistema de análisis $\mathbb{P}_{\text{CYK}'}$ para una gramática de adjunción $\mathcal{T}$ y una cadena de entrada $a_1 \dots a_n$.

$$\mathcal{I}_{\text{CYK}'} = \left\{ \begin{array}{ll} [N^\gamma \to \delta \bullet \nu, i, j \mid p, q \mid adj] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A} \\ & 0 \leq i \leq j,\ (p,q) \leq (k,j),\ adj \in \{\text{true}, \text{false}\} \end{array} \right\}$$

$$\mathcal{D}^{\text{Scan}}_{\text{CYK}'} = \frac{[a, i, i+1]}{[N^\gamma \to a\bullet, i, i+1 \mid -,- \mid \text{false}]}$$

$$\mathcal{D}^{\epsilon}_{\text{CYK}'} = \frac{}{[N^\gamma \to \bullet, i, i \mid -,- \mid \text{false}]} \quad N^\gamma \to \epsilon$$

$$\mathcal{D}^{\text{Foot}}_{\text{CYK}'} = \frac{}{[\mathbf{F}^\gamma \to \bot\bullet, i, j \mid i, j \mid \text{false}]}$$

$$\mathcal{D}^{\text{LeftDom}}_{\text{CYK}'} = \frac{\begin{array}{c}[M^\gamma \to \delta\bullet, i, k \mid p, q \mid adj1],\\ [P^\gamma \to \nu\bullet, k, j \mid -,- \mid adj2]\end{array}}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j \mid p, q \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma),\\ M^\gamma \in \text{espina}(\gamma),\\ adj1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma),\\ adj2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma)\end{array}$$

$$\mathcal{D}^{\text{RightDom}}_{\text{CYK}'} = \frac{\begin{array}{c}[M^\gamma \to \delta\bullet, i, k \mid -,- \mid adj1],\\ [P^\gamma \to \nu\bullet, k, j \mid p, q \mid adj2]\end{array}}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j \mid p, q \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma),\\ P^\gamma \in \text{espina}(\gamma),\\ adj1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma),\\ adj2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma)\end{array}$$

$$\mathcal{D}^{\text{NoDom}}_{\text{CYK}'} = \frac{\begin{array}{c}[M^\gamma \to \delta\bullet, i, k \mid -,- \mid adj1],\\ [P^\gamma \to \nu\bullet, k, j \mid -,- \mid adj2]\end{array}}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j \mid -,- \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma P^\gamma \in \mathcal{P}(\gamma),\\ N^\gamma \notin \text{espina}(\gamma),\\ adj1 = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma),\\ adj2 = \text{false sii } \mathbf{nil} \in \text{adj}(P^\gamma)\end{array}$$

$$\mathcal{D}^{\text{Unary}}_{\text{CYK}'} = \frac{[M^\gamma \to \delta\bullet, i, j \mid p, q \mid adj]}{[N^\gamma \to M^\gamma\bullet, i, j \mid p, q] \mid \text{false}]} \quad \begin{array}{l} N^\gamma \to M^\gamma \in \mathcal{P}(\gamma),\\ adj = \text{false sii } \mathbf{nil} \in \text{adj}(M^\gamma)\end{array}$$

$$\mathcal{D}^{\text{Adj}}_{\text{CYK}'} = \frac{[\top \to \mathbf{R}^\beta \bullet, i', j' \mid i, j \mid adj],\ [N^\gamma \to \delta \bullet, i, j \mid p, q \mid \text{false}]}{[N^\gamma \to \delta \bullet, i', j' \mid p, q \mid \text{true}]} \quad \beta \in \boldsymbol{A},\ \beta \in \text{adj}(N^\gamma)$$

$$\mathcal{D}_{\text{CYK}'} = \mathcal{D}^{\text{Scan}}_{\text{CYK}'} \cup \mathcal{D}^{\epsilon}_{\text{CYK}'} \cup \mathcal{D}^{\text{Foot}}_{\text{CYK}'} \cup \mathcal{D}^{\text{LeftDom}}_{\text{CYK}'} \cup \mathcal{D}^{\text{RightDom}}_{\text{CYK}'} \cup \mathcal{D}^{\text{NoDom}}_{\text{CYK}'} \cup \mathcal{D}^{\text{Unary}}_{\text{CYK}'} \cup \mathcal{D}^{\text{Adj}}_{\text{CYK}'}$$

$$\mathcal{F}_{\text{CYK}'} = \left\{\ [\top \to \mathbf{R}^\alpha \bullet, 0, n \mid -, - \mid adj] \mid \alpha \in \boldsymbol{I},\ adj = \text{false sii } \mathbf{nil} \in \text{adj}(\mathbf{R}^\alpha)\ \right\}$$

Para demostrar que $\mathbf{CYK} \overset{\mathbf{ir}}{\Longrightarrow} \mathbf{CYK}'$, definiremos la siguiente función

$$f([N^\gamma \to \delta \bullet, i, j \mid p, q \mid adj]) = [N^\gamma, i, j \mid p, q \mid adj]$$

de la cual se obtiene directamente que $\mathcal{I}_{\text{CYK}} = f(\mathcal{I}_{\text{CYK}'})$ y que $\Delta_{\text{CYK}} = f(\Delta_{\text{CYK}'})$ por inducción en la longitud de las secuencias de derivación. En consecuencia, $\mathbb{P}_{\text{CYK}} \overset{\mathbf{ir}}{\longrightarrow} \mathbb{P}_{\text{CYK}'}$, con lo que hemos probado lo que pretendíamos.

Definiremos ahora el sistema de análisis sintáctico $\mathbb{P}_{\text{ECYK}}$ para una gramática de adjunción de árboles $\mathcal{T}$ en la que ningún nodo puede tener más de dos descendientes y una cadena de entrada $a_1 \ldots a_n$ dada:

$$\mathcal{I}_{\text{ECYK}} = \mathcal{I}_{\text{CYK}'} = \mathcal{I}_{\text{buE}_1}$$

$$\mathcal{D}^{\text{Init}}_{\text{ECYK}} = \mathcal{D}^{\text{Init}}_{\text{buE}_1}$$

$$\mathcal{D}^{\text{Foot}}_{\text{ECYK}} = \mathcal{D}^{\text{Foot}}_{\text{buE}_1}$$

$$\mathcal{D}^{\text{Scan}}_{\text{ECYK}} = \mathcal{D}^{\text{Scan}}_{\text{buE}_1}$$

$$\mathcal{D}^{\text{Comp}}_{\text{ECYK}} = \mathcal{D}^{\text{Comp}}_{\text{buE}_1}$$

$$\mathcal{D}^{\text{Adj}}_{\text{ECYK}} = \mathcal{D}^{\text{Adj}}_{\text{buE}_1}$$

$$\mathcal{D}_{\text{ECYK}} = \mathcal{D}^{\text{Init}}_{\text{ECYK}} \cup \mathcal{D}^{\text{Foot}}_{\text{ECYK}} \cup \mathcal{D}^{\text{Scan}}_{\text{ECYK}} \cup \mathcal{D}^{\text{Comp}}_{\text{ECYK}} \cup \mathcal{D}^{\text{Adj}}_{\text{ECYK}}$$

$$\mathcal{F}_{\text{ECYK}} = \mathcal{F}_{\text{buE}_1}$$

Para demostrar que $\mathbf{CYK}' \overset{\mathbf{sr}}{\Longrightarrow} \mathbf{ECYK}$, deberemos demostrar que para todo sistema de análisis $\mathbb{P}_{\text{CYK}'}$ y $\mathbb{P}_{\text{ECYK}}$ se cumple que $\mathcal{I}_{\text{CYK}'} \subseteq \mathcal{I}_{\text{ECYK}}$ y que $\overset{*}{\vdash}_{\text{CYK}'} \subseteq \overset{*}{\vdash}_{\text{ECYK}}$. Lo primero es cierto por definición, puesto que $\mathcal{I}_{\text{CYK}'} = \mathcal{I}_{\text{ECYK}}$. Para lo segundo debemos mostrar que $\mathcal{D}_{\text{CYK}'} \subseteq \overset{*}{\vdash}_{\text{ECYK}}$. Consideremos caso por caso:

- un paso deductivo $\mathcal{D}^{\text{Scan}}_{\text{CYK}'}$ es equivalente a la secuencia de pasos deductivos constituida por la aplicación de un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$ y un paso $\mathcal{D}^{\text{Scan}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet a, i, i, \mid -, - \mid \text{false}]}$$

$$\frac{[N^\gamma \to \bullet a, i, i, \mid -, - \mid \text{false}], [a, i, i+1]}{[N^\gamma \to a \bullet, i, i+1, \mid -, - \mid \text{false}]}$$

- un paso deductivo $\mathcal{D}^{\epsilon}_{\text{CYK}'}$ es equivalente a un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet, i, i, \mid -, - \mid \text{false}]}$$

- $\mathcal{D}^{\text{Foot}}_{\text{ECYK}} = \mathcal{D}^{\text{Foot}}_{\text{CYK}'}$.

- un paso deductivo $\mathcal{D}^{\text{LeftDom}}_{\text{CYK}'}$ es equivalente a la secuencia de pasos deductivos constituida por un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$ y dos pasos $\mathcal{D}^{\text{Comp}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}],\ [M^\gamma \to \delta\bullet, i, k \mid p,q \mid adj]}{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid p,q \mid \text{false}]}$$

$$\frac{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid p,q \mid \text{false}], [P^\gamma \to \nu\bullet, k, j \mid -,- \mid adj]}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j, \mid p,q \mid \text{false}]}$$

- un paso $\mathcal{D}^{\text{RightDom}}_{\text{CYK}'}$ es equivalente a la secuencia formada por un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$ y dos pasos $\mathcal{D}^{\text{Comp}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}],\ [M^\gamma \to \delta\bullet, i, k \mid -,- \mid adj]}{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid -,- \mid \text{false}],\ [P^\gamma \to \nu\bullet, k, j \mid p,q \mid adj]}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j, \mid p,q \mid \text{false}]}$$

- un paso $\mathcal{D}^{\text{NoDom}}_{\text{CYK}'}$ es equivalente a la secuencia formada por un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$ y dos pasos $\mathcal{D}^{\text{Comp}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to \bullet M^\gamma P^\gamma, i, i, \mid -,- \mid \text{false}],\ [M^\gamma \to \delta\bullet, i, k \mid -,- \mid adj]}{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to M^\gamma \bullet P^\gamma, i, k, \mid -,- \mid \text{false}],\ [P^\gamma \to \nu\bullet, k, j \mid -,- \mid adj]}{[N^\gamma \to M^\gamma P^\gamma\bullet, i, j, \mid -,- \mid \text{false}]}$$

- un paso $\mathcal{D}^{\text{Unary}}_{\text{CYK}'}$ es equivalente a la secuencia formada por un paso $\mathcal{D}^{\text{Init}}_{\text{ECYK}}$ y un paso $\mathcal{D}^{\text{Comp}}_{\text{ECYK}}$:

$$\frac{}{[N^\gamma \to \bullet M^\gamma, i, i, \mid -,- \mid \text{false}]}$$

$$\frac{[N^\gamma \to \bullet M^\gamma, i, i, \mid -,- \mid \text{false}],\ [M^\gamma \to \delta\bullet, i, j \mid -,- \mid adj]}{[N^\gamma \to M^\gamma\bullet, i, j,\ \mid -,- \mid \text{false}]}$$

- $\mathcal{D}^{\text{Adj}}_{\text{ECYK}} = \mathcal{D}^{\text{adj}}_{\text{CYK}'}$.

El esquema de análisis sintáctico **ECYK** está definido para gramáticas de adjunción de árboles en las cuales ningún nodo puede tener más de dos descendientes mientras que el esquema de análisis $\mathbf{buE_1}$ está definido para cualquier TAG. Es fácil mostrar que $\mathbf{ECYK} \stackrel{\text{ext}}{\Longrightarrow} \mathbf{buE_1}$ puesto que $\mathbf{ECYK}(\mathcal{T}) = \mathbf{buE_1}(\mathcal{T})$ es cierto para toda gramática de adjunción de árboles, ya que por definición $\mathbb{P}_{\text{ECYK}} = \mathbb{P}_{\text{buE}_1}$. $\square$

Se puede eliminar el elemento de los ítems que indica si se ha realizado una adjunción en el nodo situado en el lado izquierdo de la producción contenida en dicho ítem, quedando definido el conjunto de ítems

$$\left\{ [N^\gamma \to \delta \bullet \nu, i, j \mid p, q] \mid \begin{array}{ll} \delta \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_p\, \mathbf{F}^\gamma\, a_{q+1} \ldots a_j \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_j & \text{sii } (p,q) \neq (-,-) \\ \delta \stackrel{*}{\Rightarrow} a_{i+1} \ldots a_j & \text{sii } (p,q) = (-,-) \end{array} \right\}$$

Para ello es necesario aplicar un filtro al esquema de análisis sintáctico $\mathbf{buE_1}$, consistente en la contracción de pasos deductivos: puesto que el ítem generado por un paso deductivo de tipo

Adj sólo podrá ser utilizado en un paso de tipo Comp para avanzar el punto en la producción que predijo el no terminal del lado izquierdo de su producción, podemos crear un nuevo tipo AdjComp de paso deductivo y eliminar los pasos de tipo Adj. Obtendremos así el esquema de análisis **buE** cuyo sistema de análisis $\mathbb{P}_{\text{buE}}$ mostramos a continuación.

Esquema de análisis sintáctico 3.3 El sistema de análisis $\mathbb{P}_{\text{buE}}$ que se corresponde con una nueva versión de la extensión del algoritmo de Earley ascendente para TAG, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\text{buE}} = \left\{ \begin{array}{ll} [N^\gamma \to \delta \bullet \nu, i, j \mid p, q] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \\ & 0 \le i \le j,\ (p,q) \le (i,j) \end{array} \right\}$$

$$\mathcal{D}^{\text{Init}}_{\text{buE}} = \frac{}{[N^\gamma \to \bullet\delta, i, i \mid -, -]}$$

$$\mathcal{D}^{\text{Foot}}_{\text{buE}} = \frac{}{[\mathbf{F}^\beta \to \bot\bullet, i, j \mid i, j]}$$

$$\mathcal{D}^{\text{Scan}}_{\text{buE}} = \frac{[N^\gamma \to \delta \bullet a\nu, i, j \mid p, q],\ [a, j, j+1]}{[N^\gamma \to \delta a \bullet \nu, i, j+1 \mid p, q]}$$

$$\mathcal{D}^{\text{Comp}}_{\text{buE}} = \frac{[N^\gamma \to \delta \bullet M^\gamma\nu, i, k \mid p, q],\ [M^\gamma \to \upsilon\bullet, k, j \mid p', q']}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p \cup p', q \cup q']} \quad \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{AdjComp}}_{\text{buE}} = \frac{\begin{array}{l} [\top \to \mathbf{R}^\beta\bullet, k, j \mid l, m], \\ [M^\gamma \to \upsilon\bullet, l, m \mid p', q'], \\ [N^\gamma \to \delta \bullet M^\gamma\nu, i, k \mid p, q], \end{array}}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p \cup p', q \cup q']} \quad \beta \in \boldsymbol{A},\ \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}_{\text{buE}} = \mathcal{D}^{\text{Init}}_{\text{buE}} \cup \mathcal{D}^{\text{Foot}}_{\text{buE}} \cup \mathcal{D}^{\text{Scan}}_{\text{buE}} \cup \mathcal{D}^{\text{Comp}}_{\text{buE}} \cup \mathcal{D}^{\text{AdjComp}}_{\text{buE}}$$

$$\mathcal{F}_{\text{buE}} = \left\{\ [\top \to \mathbf{R}^\alpha\bullet, 0, n \mid -, -] \mid \alpha \in \boldsymbol{I}\ \right\}$$

§

Como se puede observar, el paso $\mathcal{D}^{\text{AdjComp}}_{\text{buE}}$ solamente permite que un árbol auxiliar sea adjuntado en un nodo de un árbol elemental, pues una vez realizada la adjunción, el punto de la producción correspondiente al padre del nodo de adjunción se mueve a la derecha mientras que la producción del nodo de adjunción permanece sin cambios: si posteriormente otro árbol auxiliar es adjuntado en dicho nodo, representará una ambigüedad en el análisis sintáctico de la cadena de entrada, no la adjunción simultánea de dos árboles auxiliares en un mismo nodo [175]. En la figura 3.3 se muestra una representación gráfica de la aplicación de este paso deductivo para su caso más complejo, aquel en el que γ es un árbol auxiliar y el nodo de adjunción pertenece a su espina.

La complejidad temporal del esquema de análisis **buE** con respecto a la cadena de entrada es aparentemente $\mathcal{O}(n^7)$ puesto que son 7 los índices involucrados en el paso deductivo $\mathcal{D}_{\mathrm{buE}}^{\mathrm{AdjComp}}$: i, j, k, l, m y bien p y q o bien p' y q'. Sin embargo, podemos observar que los índices l y m sólo son necesarios para relacionar los dos primeros antecedentes y son irrelevantes para el resto de los ítems involucrados en el paso deductivo. Por tanto, mediante la aplicación parcial o *currificación* del paso deductivo $\mathcal{D}_{\mathrm{buE}}^{\mathrm{AdjComp}}$ la complejidad del mismo se reduce a $\mathcal{O}(n^6)$. De forma equivalente, también se podría incluir dicha aplicación parcial en el propio esquema de análisis, sustituyendo el paso $\mathcal{D}_{\mathrm{buE}}^{\mathrm{AdjComp}}$ por otros dos, uno que combine los dos primeros ítems antecedentes y genere un ítem intermedio y otro que combine dicho ítem con el tercer antecedente del paso original para obtener el ítem resultado. Sin embargo, con ello oscureceríamos la comprensión del algoritmo descrito en el esquema y estaríamos vulnerando la filosofía de los esquemas de análisis, ya que la aplicación parcial es un detalle de implementación del que nos debemos abstraer.

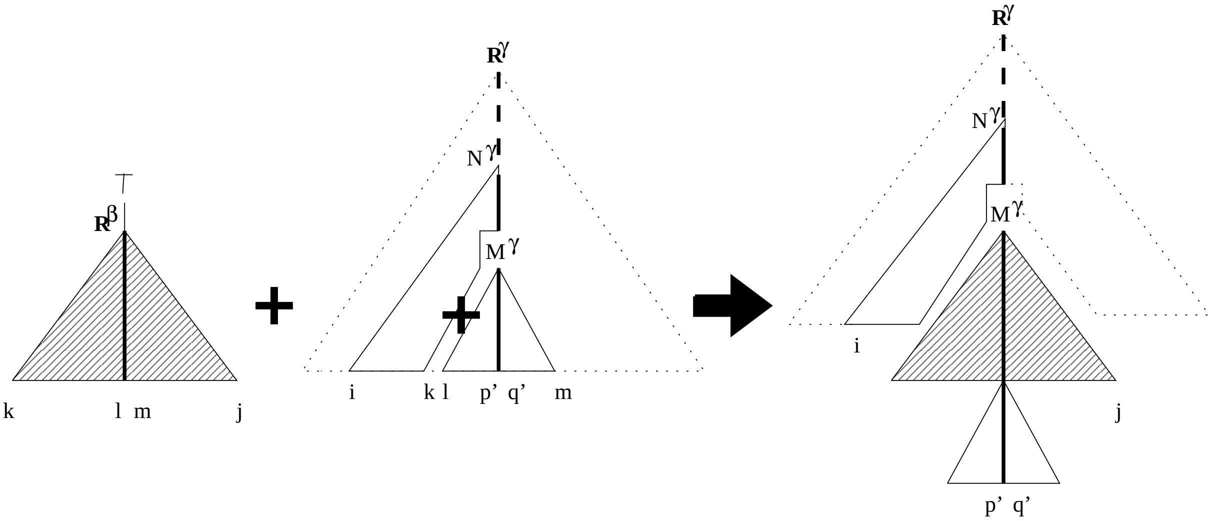

Figura 3.3: Descripción gráfica de un paso $\mathcal{D}_{\mathrm{buE}}^{\mathrm{AdjComp}}$

Proposición 3.2 $\mathbf{buE_1} \stackrel{\mathbf{df}}{\Longrightarrow} \mathbf{buE'_1} \stackrel{\mathbf{sc}}{\Longrightarrow} \mathbf{buE_2} \stackrel{\mathbf{ic}}{\Longrightarrow} \mathbf{buE}$.

Demostración:

Como primer paso definiremos el sistema de análisis $\mathbb{P}_{\mathrm{buE}'_1}$ para una gramática de adjunción $\mathcal{T}$ y una cadena de entrada $a_1 \dots a_n$, en el cual el paso Comp ha sido desdoblado en Comp^1 y Comp^2, el primero de los cuales se encarga de avanzar el punto en aquellos casos en que el nodo sobre el que se avanza no ha sido utilizado como nodo de adjunción, mientras que el segundo avanza el punto sobre un nodo de adjunción.

$$\mathcal{I}_{\mathrm{buE}'_1} = \mathcal{I}_{\mathrm{buE}_1}$$

$$\mathcal{D}_{\mathrm{buE}'_1}^{\mathrm{Init}} = \mathcal{D}_{\mathrm{buE}_1}^{\mathrm{Init}}$$

$$\mathcal{D}_{\mathrm{buE}'_1}^{\mathrm{Foot}} = \mathcal{D}_{\mathrm{buE}_1}^{\mathrm{Foot}}$$

$$\mathcal{D}_{\mathrm{buE}'_1}^{\mathrm{Scan}} = \mathcal{D}_{\mathrm{buE}_1}^{\mathrm{Scan}}$$

$$\mathcal{D}_{\mathrm{buE}'_1}^{\mathrm{Comp}^1} = \frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{false}]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]} \quad \mathbf{nil} \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{buE}'_1}^{\mathrm{Comp}^2} = \frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{true}]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]} \quad \exists \beta \in \boldsymbol{A},\ \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}_1} = \frac{[\top \to \mathbf{R}^\beta \bullet, k, j \mid k', j' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k', j' \mid p, q \mid \mathrm{false}],}{[M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{true}]} \quad \beta \in \boldsymbol{A},\ \beta \in \mathrm{adj}(\gamma)$$

$$\mathcal{D}_{\mathrm{buE}'_1} = \mathcal{D}^{\mathrm{Init}}_{\mathrm{buE}'_1} \cup \mathcal{D}^{\mathrm{Foot}}_{\mathrm{buE}'_1} \cup \mathcal{D}^{\mathrm{Scan}}_{\mathrm{buE}'_1} \cup \mathcal{D}^{\mathrm{Comp}^1}_{\mathrm{buE}'_1} \cup \mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1} \cup \mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}'_1}$$

$$\mathcal{F}_{\mathrm{buE}'_1} = \mathcal{F}_{\mathrm{buE}_1}$$

Los pasos deductivos $\mathcal{D}^{\mathrm{Comp}^1}_{\mathrm{buE}'_1}$ y $\mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1}$ son el resultado de aplicar un filtro dinámico al paso $\mathcal{D}^{\mathrm{Comp}}_{\mathrm{buE}_1}$. El el caso de $\mathcal{D}^{\mathrm{Comp}^1}_{\mathrm{buE}'_1}$ dicho filtro consiste en comprobar si el último elemento de los ítems contiene el valor false, mientras que en el caso de $\mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1}$ consiste en comprobar que su valor es *true*.

Para demostrar que $\mathbf{buE_1} \stackrel{\mathbf{df}}{\Longrightarrow} \mathbf{buE'_1}$, deberemos demostrar que para todo esquema de análisis $\mathbb{P}_{\mathrm{buE}_1}$ y $\mathbb{P}_{\mathrm{buE}'_1}$ se cumple que $\mathcal{I}_{\mathrm{buE}_1} \supseteq \mathcal{I}_{\mathrm{buE}'_1}$ y $\vdash_{\mathrm{buE}_1} \supseteq \vdash_{\mathrm{buE}'_1}$. Lo primero se cumple puesto que por definición $\mathcal{I}_{\mathrm{buE}_1} = \mathcal{I}_{\mathrm{buE}'_1}$. Para lo segundo debemos mostrar que $\vdash_{\mathrm{buE}_1} \supseteq \mathcal{D}_{\mathrm{buE}'_1}$. Para ello sólo necesitamos considerar aquellos pasos de $\mathbb{P}_{\mathrm{buE}'_1}$ a los que se les ha aplicado el filtro dinámico, pues los restantes pasos permanecen inalterados:

- un paso $\mathcal{D}^{\mathrm{Comp}^1}_{\mathrm{buE}'_1}$ es equivalente a la aplicación del paso $\mathcal{D}^{\mathrm{Comp}}_{\mathrm{buE}_1}$:

$$\frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{false}]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]}$$

- un paso $\mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1}$ es equivalente a la aplicación del paso $\mathcal{D}^{\mathrm{Comp}}_{\mathrm{buE}_1}$:

$$\frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{true}]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]}$$

Sólo se pueden generar ítems que tengan el último componente con valor true como consecuencia de la aplicación de un paso $\mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}'_1}$ y a su vez estos ítems sólo puede ser utilizados como antecedentes en un paso $\mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1}$. Por tanto podemos juntar ambos pasos pasos en uno sólo. A continuación definimos el sistema de análisis $\mathbb{P}_{\mathrm{buE}_2}$ para una gramática de adjunción $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ que incorpora esta transformación.

$$\mathcal{I}_{\mathrm{buE}_2} = \mathcal{I}_{\mathrm{buE}'_1}$$

$$\mathcal{D}^{\mathrm{Init}}_{\mathrm{buE}_2} = \mathcal{D}^{\mathrm{Init}}_{\mathrm{buE}'_1}$$

$$\mathcal{D}^{\mathrm{Foot}}_{\mathrm{buE}_2} = \mathcal{D}^{\mathrm{Foot}}_{\mathrm{buE}'_1}$$

$$\mathcal{D}^{\mathrm{Scan}}_{\mathrm{buE}_2} = \mathcal{D}^{\mathrm{Scan}}_{\mathrm{buE}'_1}$$

$$\mathcal{D}^{\mathrm{Comp}}_{\mathrm{buE}_2} = \mathcal{D}^{\mathrm{Comp}^1}_{\mathrm{buE}'_1}$$

$$\mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}_2} = \frac{\begin{array}{c}[\top \to \mathbf{R}^\beta \bullet, k, j \mid k', j' \mid \mathrm{false}],\\ [M^\gamma \to \upsilon \bullet, k', j' \mid p, q \mid \mathrm{false}],\\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}]\end{array}}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]} \quad \beta \in \boldsymbol{A},\ \beta \in \mathrm{adj}(\gamma)$$

$$\mathcal{D}_{\mathrm{buE}_2} = \mathcal{D}^{\mathrm{Init}}_{\mathrm{buE}_2} \cup \mathcal{D}^{\mathrm{Foot}}_{\mathrm{buE}_2} \cup \mathcal{D}^{\mathrm{Scan}}_{\mathrm{buE}_2} \cup \mathcal{D}^{\mathrm{Comp}}_{\mathrm{buE}_2} \cup \mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}_2}$$

$$\mathcal{F}_{\mathrm{buE}_2} = \mathcal{F}_{\mathrm{buE}_1}$$

Para demostrar que el sistema de análisis sintáctico $\mathbb{P}_{\mathrm{buE}_2}$ es el resultado de aplicar una contracción de pasos al sistema de análisis $\mathbb{P}_{\mathrm{buE}'_1}$ tenemos que demostrar que $\mathcal{I}_{\mathrm{buE}'_1} \supseteq \mathcal{I}_{\mathrm{buE}_2}$ y que $\stackrel{*}{\vdash}_{\mathrm{buE}'_1} \supseteq \stackrel{*}{\vdash}_{\mathrm{buE}_2}$. Lo primero es cierto por definición, puesto que $\mathcal{I}_{\mathrm{buE}'_1} = \mathcal{I}_{\mathrm{buE}_2}$. Para lo segundo es suficiente con demostrar que $\overset{*}{\vdash}_{\mathrm{buE}'_1} \supseteq \mathcal{D}_{\mathrm{buE}_2}$. Puesto que los demás pasos deductivos son idénticos en ambos esquemas, debemos mostrar únicamente que $\overset{*}{\vdash}_{\mathrm{buE}'_1} \supseteq \mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}_2}$, tarea que no presenta complicación alguna puesto que como se ha mencionado anteriormente, un paso $\mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}_2}$ es equivalente a la aplicación consecutiva de un paso $\mathcal{D}^{\mathrm{Adj}}_{\mathrm{buE}'_1}$ y un paso $\mathcal{D}^{\mathrm{Comp}^2}_{\mathrm{buE}'_1}$:

$$\frac{[\top \to \mathbf{R}^\beta \bullet, k, j \mid k', j' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k', j' \mid p, q \mid \mathrm{false}]}{[M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{true}]}$$

$$\frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p', q' \mid \mathrm{false}],\ [M^\gamma \to \upsilon \bullet, k, j \mid p, q \mid \mathrm{true}]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p' \cup p, q' \cup q \mid \mathrm{false}]}$$

Finalmente observamos que todos los ítems del sistema de análisis $\mathbb{P}_{\mathrm{buE}_2}$ tienen la forma $[N^\gamma \to \delta \bullet \nu, i, j \mid p, q \mid \mathrm{false}]$. Puesto que el último elemento tiene un valor constante, su eliminación no producirá ningún efecto con respecto al conjunto de ítems que puedan ser generados. El sistema de análisis $\mathbb{P}_{\mathrm{buE}}$ se obtiene precisamente al aplicar esta transformación al sistema de análisis $\mathbb{P}_{\mathrm{buE}_2}$. Dicha transformación constituye un caso peculiar de contracción de ítems puesto que no se rompe un ítem en varios, sino que se establece una relación biyectiva entre los conjuntos $\mathcal{I}_{\mathrm{buE}_2}$ y $\mathcal{I}_{\mathrm{buE}}$. Vemos que la relación de contracción de ítems se mantiene entre los dos esquemas de análisis puesto que definiendo la función

$$f([N^\gamma \to \delta \bullet \nu, i, j \mid p, q \mid \mathrm{false}]) = [N^\gamma \to \delta \bullet \nu, i, j \mid p, q]$$

obtenemos que $f(\mathcal{I}_{\mathrm{buE}_2}) = \mathcal{I}_{\mathrm{buE}}$ y $f(\Delta_{\mathrm{buE}_2}) = \Delta_{\mathrm{buE}}$ por inducción en la longitud de las secuencias de derivación. □

3.4 La propiedad del prefijo válido en los algoritmos de análisis

Los analizadores sintácticos que satisfacen la *propiedad del prefijo válido* (VPP) garantizan que, en tanto que leen la cadena de entrada de izquierda a derecha, las subcadenas leídas son prefijos válidos del lenguaje definido por la gramática. A la propiedad del prefijo válido también se la denomina a veces propiedad de detección de errores porque implica que los errores son detectados tan pronto como es posible en una lectura de la cadena de entrada de izquierda a derecha. La carencia de la propiedad del prefijo válido no significa que los errores no sean detectados, sino simplemente que estos lo serán más tarde. En contra de la idea mantenida en ciertos momentos [168], la propiedad del prefijo válido es independiente de la propiedad de análisis *en línea* [169].

Más formalmente, un analizador sintáctico satisface la propiedad del prefijo válido si al leer la subcadena $a_1 \ldots a_k$ de la cadena de entrada $a_1 \ldots a_k a_{k+1} \ldots a_n$ garantiza que hay una cadena $b_1 \ldots b_m$, donde b_i no tiene porque formar parte de la cadena de entrada, tal que $a_1 \ldots a_k b_1 \ldots b_m$ es una cadena válida del lenguaje.

El mantenimiento de la propiedad del prefijo válido exige ir reconociendo los posibles árboles derivados de forma prefija. Este recorrido consta de dos fases, una descendente que dado un nodo expande sus nodos hijos y una ascendente que agrupa los nodos hijos para indicar el reconocimiento del nodo padre. Cuando se desea mantener la propiedad del prefijo válido estas dos fases deben actuar coordinadamente. La fase descendente debe ser además restringida, para evitar expansiones que lleven al reconocimiento de prefijos no válidos [169].

En el caso de gramáticas independientes del contexto, existen numerosos algoritmos de análisis que preservan la propiedad del prefijo válido (por ejemplo, Earley) y que muestran una complejidad en el peor caso igual a aquellos algoritmos que no la preservan (por ejemplo, CYK). Esto se debe a que la operación de sustitución puede ser aplicada de forma totalmente independiente del contexto, lo que conlleva que el conjunto de caminos de una gramática independiente del contexto sea un lenguaje regular. Como consecuencia, el mantenimiento de la propiedad del prefijo válido se puede asegurar sin tener que aplicar restricciones complejas sobre la fase descendente de los algoritmos.

En el caso de las gramáticas de adjunción, la operación de adjunción no es completamente independiente del contexto en el cual se aplica. Durante el reconocimiento de un árbol derivado en forma prefija, la expansión de un nodo puede depender de operaciones de adjunción previamente realizadas en la parte recorrida del árbol. Esta dependencia del contexto conlleva que el conjunto de caminos ya no sea un lenguaje regular sino un lenguaje independiente del contexto [230, 217]. Un algoritmo básicamente ascendente (p.ej.: de tipo CYK, algunas variantes de tipo Earley) puede simplemente hacer uso de una pila para ir almacenando las dependencias indicadas por el lenguaje que define el conjunto de caminos. Con ello se conseguiría un algoritmo de análisis correcto pero sin la propiedad del prefijo válido. Para preservar esta propiedad es necesario disponer de una fase descendente, que también tendría que disponer de una pila para satisfacer las restricciones impuestas por el lenguaje que define el conjunto de caminos. Schabes [169] argumentaba que entonces, al tener que coordinar las pilas de la fase ascendente y descendente, la complejidad del algoritmo de análisis sintáctico resultante aumentaría. Sin embargo, el algoritmo descrito por Nederhof en [125] mantiene la propiedad del prefijo válido con una complejidad $\mathcal{O}(n^6)$, igual a la de aquellos algoritmos que no la mantienen.

3.5 Algoritmos de tipo Earley sin la propiedad del prefijo válido

Mediante la aplicación de un filtrado dinámico a los esquemas de análisis sintácticos anteriores es posible obtener un esquema de análisis sintáctico de un algoritmo al estilo del de Earley pero extendido al caso de gramáticas de adjunción de árboles. Concretamente, el filtrado que se realiza es el siguiente:

- El paso deductivo $\mathcal{D}_{\text{buE}}^{\text{Init}}$ sólo contendrá producciones cuyo lado izquierdo corresponda con la raíz de un árbol inicial.
- Un conjunto de pasos deductivos predictivos controlan la generación de nuevos ítems tratando de limitarla únicamente a aquellos que puedan resultar útiles en el proceso de análisis.

Los algoritmos descritos en esta sección no preservan la propiedad del prefijo válido puesto que la fase predictiva no es lo suficientemente restrictiva como para evitar que las predicciones realizadas durante el análisis del pie de un árbol no conlleven el análisis de subárboles no válidos.

Una primera aproximación a un esquema de análisis sintáctico para un algoritmo de tipo Earley para gramáticas de adjunción consiste en transformar el esquema de análisis sintáctico **buE**. Al nuevo esquema de análisis, que presenta ciertas semejanzas con los algoritmos descritos por Schabes en [168, 169, 172], lo denominaremos **E** y su correspondiente sistema de análisis $\mathbb{P}_{\text{E}}$ se define a continuación.

Esquema de análisis sintáctico 3.4 El sistema de análisis $\mathbb{P}_{\text{E}}$ que se corresponde con una versión del algoritmo de Earley para TAG sin la propiedad del prefijo válido, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\text{E}} = \mathcal{I}_{\text{buE}}$$

$$\mathcal{D}_{\text{E}}^{\text{Init}} = \frac{}{[\top \rightarrow \bullet \mathbf{R}^{\alpha}, 0, 0 \mid -, -]} \quad \alpha \in \boldsymbol{I}$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{Scan}} = \mathcal{D}_{\mathrm{buE}}^{\mathrm{Scan}}$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}} = \frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[M^\gamma \to \bullet \upsilon, j, j \mid -, -]} \quad \mathbf{nil} \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{Comp}} = \mathcal{D}_{\mathrm{buE}}^{\mathrm{Comp}}$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}} = \frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[\top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]} \quad \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}} = \frac{[\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -]}{[M^\gamma \to \bullet \delta, k, k \mid -, -]} \quad \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}} = \frac{[M^\gamma \to \delta \bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -]}{[\mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]} \quad \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}} = \mathcal{D}_{\mathrm{buE}}^{\mathrm{AdjComp}} = \frac{\begin{array}{l}[\top \to \mathbf{R}^\beta \bullet, j, m \mid k, l],\\ [M^\gamma \to \upsilon \bullet, k, l \mid p, q],\\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p \cup p', q \cup q']} \quad \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{E}} = \mathcal{D}_{\mathrm{E}}^{\mathrm{Init}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{Scan}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{Comp}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}} \cup \mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$$

$$\mathcal{F}_{\mathrm{E}} = \mathcal{F}_{\mathrm{buE}}$$

§

El paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$, aunque es idéntico al del sistema de análisis $\mathbb{P}_{\mathrm{buE}_1}$ se repite aquí para facilitar la comprensión del algoritmo. Como se ha dicho anteriormente, el esquema de análisis **E** presenta ciertas similitudes con los algoritmos descritos por Schabes en [168, 169, 172]. Podemos establecer la siguiente relación entre los pasos deductivos de **E** y los procesos definidos en dichos algoritmos, tal y como se muestra en la tabla 3.1. La razón del cambio de nombre en las pasos encargados de la adjunción se debe a que creemos que los nombres utilizados por Schabes pueden llevar a confusión, pues se tiende a pensar que cada *Predictor* está asociado con el *Completor* correspondiente a su mismo lado, cuando no es así. Es por ello que hemos decidido emparejar los pasos deductivos por las funciones que realizan: predicción-compleción de adjunción y predicción-compleción de pie.

Respecto al comportamiento del algoritmo, diremos que el análisis comienza creando un ítem correspondiente a una producción del nodo raíz de un árbol inicial con el punto situado en el extremo izquierdo. Posteriormente, los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}}$ y $\mathcal{D}_{\mathrm{E}}^{\mathrm{Comp}}$ se encargan de ir recorriendo el árbol de modo descendente y ascendente, respectivamente, de modo similar a como actúa el algoritmo de Earley para gramáticas independientes del contexto. Se puede predecir la adjunción de un árbol β en un nodo de un árbol elemental γ mediante la aplicación de un paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}}$,

Pasos deductivos de E	**Algoritmo de Schabes**
$\mathcal{D}_{\mathrm{E}}^{\mathrm{Init}}$	*Initial item*
$\mathcal{D}_{\mathrm{E}}^{\mathrm{Scan}}$	Scanner
$\mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}}$	Move Dot Down
$\mathcal{D}_{\mathrm{E}}^{\mathrm{Comp}}$	Move Dot Up
$\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}}$	Left Predictor
$\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$	Left Completor
$\mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}}$	Right Predictor
$\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$	Right Completor

Tabla 3.1: Relación entre $\mathcal{D}_{\mathrm{E}}$ y el algoritmo tipo Earley sin VPP de Schabes

con lo que se comienza el análisis del árbol β. Una vez alcanzado el pie de dicho árbol auxiliar, deberemos retomar el análisis de γ, concretamente del subárbol que pende del nodo de adjunción. El problema es que al no conocer en qué nodo de qué árbol elemental se ha producido la adjunción, deberemos predecir todos los posibles nodos donde esté permitida la adjunción de β, predicción realizada por un paso deductivo $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$. Es la predicción que se realiza en los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$ lo que provoca que el algoritmo no posea la propiedad del prefijo válido, puesto que se puede comenzar a analizar una parte de la cadena de entrada que no es gramatical, al predecir un subárbol que no se corresponde con el árbol desde el que se realizó la adjunción.

Una vez terminado de analizar todo el subárbol predicho por un paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$, deberemos retomar el análisis del árbol auxiliar β a partir del pie, tarea encomendada a los pasos deductivos $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}}$. Una vez terminado de analizar completamente el árbol auxiliar β, deberemos concluir la operación de adjunción aplicando un paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$. Es en estos pasos en los que se verifica que el subárbol escindido del nodo de adjunción y el árbol auxiliar han sido correctamente reconstruidos. Las adjunciones simultáneas sobre un mismo nodo [175] son evitadas por los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$ puesto que cuando se ha terminado de recorrer completamente el árbol auxiliar, se verifica que se ha analizado la parte correspondiente al subárbol del nodo de adjunción y se avanza el punto de la producción que contiene a este, sin cambiar el ítem correspondiente al nodo de adjunción. Si posteriormente otro árbol auxiliar es adjuntado en dicho nodo, representará una ambigüedad en el análisis sintáctico de la cadena de entrada pero no la adjunción simultánea de dos árboles auxiliares en un mismo nodo.

Proposición 3.3 **buE** $\stackrel{\mathbf{df}}{\Longrightarrow}$ **E**.

Demostración:

Para demostrar que el esquema de análisis sintáctico **E** es el resultado de aplicar un filtrado dinámico al esquema de análisis **buE**, debemos demostrar que $\mathcal{I}_{\mathrm{buE}} \supseteq \mathcal{I}_{\mathrm{E}}$ y que $\vdash_{\mathrm{buE}} \supseteq \vdash_{\mathrm{E}}$ para los sistemas de análisis sintáctico $\mathbb{P}_{\mathrm{buE}}$ y $\mathbb{P}_{\mathrm{E}}$. Lo primero es cierto por definición puesto que $\mathcal{I}_{\mathrm{buE}} = \mathcal{I}_{\mathrm{E}}$. Respecto a lo segundo, es suficiente con mostrar que $\vdash_{\mathrm{buE}} \supseteq \mathcal{D}_{\mathrm{E}}$.

Los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{Scan}}$, $\mathcal{D}_{\mathrm{E}}^{\mathrm{Comp}}$ y $\mathcal{D}_{\mathrm{E}}^{\mathrm{Adjcomp}}$ son idénticos a sus homónimos del sistema $\mathbb{P}_{\mathrm{buE}}$ y los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{Init}}$ generan un subconjunto de los ítems generados por $\mathcal{D}_{\mathrm{buE}}^{\mathrm{Init}}$. Respecto a los otros pasos:

- Dado un paso $\frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[M^\gamma \to \bullet \upsilon, j, j \mid -, -]} \in \mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}}$ existe un paso $\frac{}{[M^\gamma \to \bullet \upsilon, j, j \mid -, -]} \in \mathcal{D}_{\mathrm{buE}}^{\mathrm{Init}}$ y por tanto existe la inferencia

$$[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q] \vdash_{\mathrm{buE}} [M^\gamma \to \bullet \upsilon, j, j \mid -, -]$$

- Dado un paso $\frac{[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[\top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]} \in \mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}}$ existe un paso $\frac{}{[\top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]} \in \mathcal{D}_{\mathrm{buE}}^{\mathrm{Init}}$ y por tanto existe la inferencia

$$[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q] \vdash_{\mathrm{buE}} [\top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]$$

- Dado un paso $\frac{[\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -]}{[M^\gamma \to \bullet \delta, k, k \mid -, -]} \in \mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$ existe un paso $\frac{}{[M^\gamma \to \bullet \delta, k, k \mid -, -]} \in \mathcal{D}_{\mathrm{buE}}^{\mathrm{Init}}$ y por tanto existe la inferencia

$$[\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -] \vdash_{\mathrm{buE}} [M^\gamma \to \bullet \delta, k, k \mid -, -]$$

- Dado un paso $\frac{[M^\gamma \to \delta \bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -]}{[\mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]} \in \mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}}$ existe un paso $\frac{}{[\mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]} \in \mathcal{D}_{\mathrm{buE}}^{\mathrm{Foot}}$ y por tanto existe la inferencia

$$[M^\gamma \to \delta \bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -] \vdash_{\mathrm{buE}} [\mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]$$

$\square$

La complejidad temporal del esquema de análisis sintáctico **E** con respecto a la cadena de entrada es $\mathcal{O}(n^6)$ puesto que la aparente complejidad $\mathcal{O}(n^7)$ del paso deductivo $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$ puede reducirse a $\mathcal{O}(n^6)$ mediante la aplicación parcial o *currificación* de dicho paso, ya que los índices l y m sólo involucran a los dos primeros ítems antecedentes.

Con el fin de definir un esquema de análisis sintáctico que se corresponda con un algoritmo más cercano al espíritu del algoritmo de Earley, debemos fortalecer la fase predictiva de los esquemas de análisis anteriores, puesto que estos no utilizan toda la información que tienen a su disposición. En concreto:

- Los pasos deductivos $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootPred}}$ en los que se realiza la predicción del pie no comprueban que previamente se haya iniciado la adjunción del árbol β en el nodo M^γ.
- Idem para los pasos deductivos $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}}$ que finalizan el reconocimiento del pie.
- Los pasos deductivos $\mathcal{D}_{\mathrm{E}}^{\mathrm{FootComp}}$ no comprueban que el árbol auxiliar predicho para adjunción en el nodo M^γ sea el mismo que el que ha provocado el reconocimiento del subárbol enraizado en dicho nodo mediante la aplicación de los pasos FootPred.

A continuación definimos un nuevo esquema de análisis **Ear** derivado a partir de **E**, sobre el que hemos aplicado las siguientes modificaciones:

- La aplicación de un filtro dinámico a los pasos deductivos FootPred y FootComp consistente en la verificación de la existencia de los ítems que representan el comienzo de la operación de adjunción en el nodo M^γ.
- La aplicación de un refinamiento al paso AdjComp, que se divide en dos pasos $\mathrm{AdjComp}^1$ y $\mathrm{AdjComp}^2$, el primero realizando las comprobaciones pertinentes en el caso de que se haya producido la adjunción de un árbol auxiliar en un nodo de la espina de otro árbol auxiliar. En la figura 3.4 se muestra una representación gráfica de la aplicación del paso deductivo $\mathrm{AdjComp}^1$, el más complejo de los dos ya que el árbol γ en el que se realiza la adjunción es un árbol auxiliar y el nodo de adjunción pertenece a su espina. Las partes de

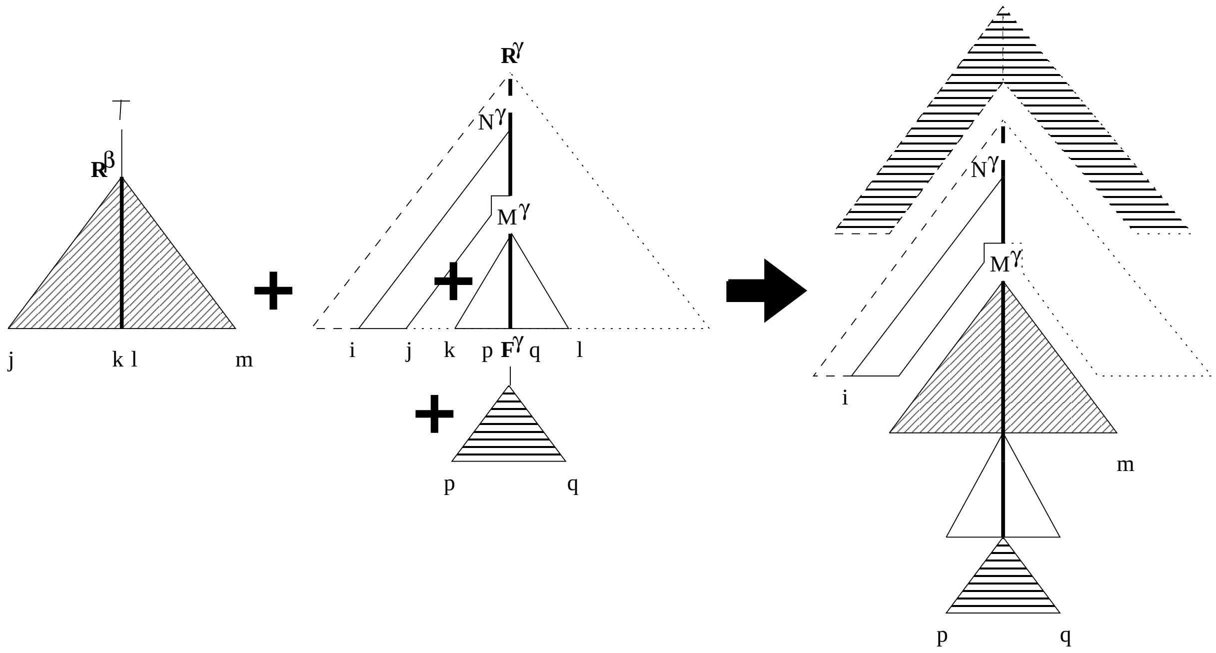

Figura 3.4: Descripción gráfica de un paso $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjComp}^1}$

los árboles involucrados que no se encuentran representados por los ítems que intervienen en el paso deductivo se muestran en línea discontinua si el algoritmo de análisis tiene que haber pasado obligatoriamente por dicha parte del árbol al menos en la fase predictiva y en línea punteada si se trata de partes que serán analizadas posteriormente.

Esquema de análisis sintáctico 3.5 El sistema de análisis $\mathbb{P}_{\mathrm{Ear}}$ que se corresponde con una versión del algoritmo de Earley sin la propiedad del prefijo válido con predicción fuerte, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\mathrm{Ear}} = \mathcal{I}_{\mathrm{buE}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Init}} = \mathcal{D}_{\mathrm{E}}^{\mathrm{Init}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Scan}} = \mathcal{D}_{\mathrm{buE}}^{\mathrm{Scan}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Pred}} = \mathcal{D}_{\mathrm{E}}^{\mathrm{Pred}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Comp}} = \mathcal{D}_{\mathrm{buE}}^{\mathrm{Comp}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjPred}} = \mathcal{D}_{\mathrm{E}}^{\mathrm{AdjPred}}$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{FootPred}} = \frac{[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -, -],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[M^\gamma \to \bullet\delta, k, k \mid -, -]} \quad \beta \in \mathrm{adj}(M^\gamma)$$

$$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{FootComp}} = \frac{\begin{array}{l}[M^\gamma \to \upsilon\bullet, k, l \mid p, q],\\ [\mathbf{F}^\beta \to \bullet\bot, k, k \mid -, -],\\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[\mathbf{F}^\beta \to \bot\bullet, k, l \mid k, l]} \quad \begin{array}{l}\beta \in \mathrm{adj}(M^\gamma),\\ p \cup p' \ y\ q \cup q' \ \textit{está definido}\end{array}$$

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

$$\mathcal{D}_{\text{Ear}}^{\text{AdjComp}^2} = \frac{\begin{array}{l}[\top \rightarrow \mathbf{R}^\beta \bullet, j, m \mid k, l],\\ [M^\gamma \rightarrow \upsilon \bullet, k, l \mid -, -],\\ [N^\gamma \rightarrow \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[N^\gamma \rightarrow \delta M^\gamma \bullet \nu, i, m \mid p', q']} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}_{\text{Ear}} = \mathcal{D}_{\text{Ear}}^{\text{Init}} \cup \mathcal{D}_{\text{Ear}}^{\text{Scan}} \cup \mathcal{D}_{\text{Ear}}^{\text{Pred}} \cup \mathcal{D}_{\text{Ear}}^{\text{Comp}} \cup \mathcal{D}_{\text{Ear}}^{\text{AdjPred}} \cup \mathcal{D}_{\text{Ear}}^{\text{FootPred}} \cup \mathcal{D}_{\text{Ear}}^{\text{FootComp}} \cup \mathcal{D}_{\text{Ear}}^{\text{AdjComp}^1} \cup \mathcal{D}_{\text{Ear}}^{\text{AdjComp}^2}$$

$$\mathcal{F}_{\text{Ear}} = \mathcal{F}_{\text{buE}}$$

§

Proposición 3.4 $\mathbf{E} \stackrel{\text{sr}}{\Longrightarrow} \mathbf{E}' \stackrel{\text{df}}{\Longrightarrow} \mathbf{Ear}$.

<u>Demostración</u>:

Como primer paso definiremos el esquema de análisis $\mathbf{E}'$ que se obtiene a partir de $\mathbf{E}$ simplemente rompiendo el conjunto de pasos deductivos $\mathcal{D}_{\text{E}}^{\text{AdjComp}}$ en dos conjuntos $\mathcal{D}_{\text{E}'}^{\text{AdjComp}^1}$ y $\mathcal{D}_{\text{E}'}^{\text{AdjComp}^2}$. El sistema de análisis $\mathbb{P}_{\text{E}'}$ sería por tanto el siguiente:

$$\mathcal{I}_{\text{E}'} = \mathcal{I}_{\text{buE}}$$

$$\mathcal{D}_{\text{E}'}^{\text{Init}} = \mathcal{D}_{\text{E}}^{\text{Init}}$$

$$\mathcal{D}_{\text{E}'}^{\text{Scan}} = \mathcal{D}_{\text{buE}}^{\text{Scan}}$$

$$\mathcal{D}_{\text{E}'}^{\text{Pred}} = \mathcal{D}_{\text{E}}^{\text{Pred}}$$

$$\mathcal{D}_{\text{E}'}^{\text{Comp}} = \mathcal{D}_{\text{buE}}^{\text{Comp}}$$

$$\mathcal{D}_{\text{E}'}^{\text{AdjPred}} = \mathcal{D}_{\text{E}}^{\text{AdjPred}}$$

$$\mathcal{D}_{\text{E}'}^{\text{FootPred}} = \mathcal{D}_{\text{E}}^{\text{FootPred}}$$

$$\mathcal{D}_{\text{E}'}^{\text{FootComp}} = \mathcal{D}_{\text{E}}^{\text{FootComp}}$$

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

$$\mathcal{D}_{\text{E}'}^{\text{AdjComp}^2} = \frac{\begin{array}{l}[\top \rightarrow \mathbf{R}^\beta \bullet, j, m \mid k, l],\\ [M^\gamma \rightarrow \upsilon \bullet, k, l \mid -, -],\\ [N^\gamma \rightarrow \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[N^\gamma \rightarrow \delta M^\gamma \bullet \nu, i, m \mid p', q']} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}_{\text{E}'} = \mathcal{D}_{\text{E}'}^{\text{Init}} \cup \mathcal{D}_{\text{E}'}^{\text{Scan}} \cup \mathcal{D}_{\text{E}'}^{\text{Pred}} \cup \mathcal{D}_{\text{E}'}^{\text{Comp}} \cup \mathcal{D}_{\text{E}'}^{\text{AdjPred}} \cup \mathcal{D}_{\text{E}'}^{\text{FootPred}} \cup \mathcal{D}_{\text{E}'}^{\text{FootComp}} \cup \mathcal{D}_{\text{E}'}^{\text{AdjComp}^1} \cup \mathcal{D}_{\text{E}'}^{\text{AdjComp}^2}$$

$$\mathcal{F}_{\mathrm{E}'} = \mathcal{F}_{\mathrm{buE}}$$

Las condiciones a verificar son que $\mathcal{I}_{\mathrm{E}} \subseteq \mathcal{I}_{\mathrm{E}'}$ y que $\overset{*}{\vdash}_{\mathrm{E}} \subseteq \overset{*}{\vdash}_{\mathrm{E}'}$. La primera condición se verifica por la propia definición de los ítems mientras que la segunda se obtiene directamente considerando que el único cambio que se ha producido en $\mathbb{P}_{\mathrm{E}'}$ es hacer explícita la incompatibilidad del par de índices (p,q) con el par (p',q') de los pasos $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$: si son p' y q' los índices que están definidos, entonces el paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$ se convierte en $\mathcal{D}_{\mathrm{E}'}^{\mathrm{AdjComp}^1}$, mientras que si son p y q los índices que están definidos, entonces el paso $\mathcal{D}_{\mathrm{E}}^{\mathrm{AdjComp}}$ se convierte en $\mathcal{D}_{\mathrm{E}'}^{\mathrm{AdjComp}^2}$.

Para demostrar que el esquema de análisis sintáctico **Ear** es el resultado de aplicar un filtrado dinámico al esquema de análisis **E**$'$, debemos demostrar que $\mathcal{I}_{\mathrm{E}'} \supseteq \mathcal{I}_{\mathrm{Ear}}$ y que $\vdash_{\mathrm{E}'} \supseteq \vdash_{\mathrm{Ear}}$ para los sistemas de análisis sintáctico $\mathbb{P}_{\mathrm{E}'}$ y $\mathbb{P}_{\mathrm{Ear}}$. Lo primero es cierto por definición puesto que $\mathcal{I}_{\mathrm{buE}} = \mathcal{I}_{\mathrm{E}'} = \mathcal{I}_{\mathrm{Ear}}$. Respecto a lo segundo es suficiente con mostrar que $\vdash_{\mathrm{E}'} \supseteq \vdash_{\mathrm{Ear}}$.

Los pasos $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Init}}$, $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Scan}}$, $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{pred}}$, $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{Comp}}$ y $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjPred}}$ son idénticos a sus homónimos del sistema $\mathbb{P}_{\mathrm{E}'}$. Respecto a los otros pasos:

$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{FootPred}}$**:** Dado un paso $\frac{[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[M^\gamma \to \bullet\delta, k, k \mid -,-]}$ existe un paso $\frac{[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-]}{[M^\gamma \to \bullet\delta, k, k \mid -,-]} \in \mathcal{D}_{\mathrm{E}'}^{\mathrm{FootPred}}$ y por tanto existe la inferencia

$$[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q] \ \vdash_{\mathrm{E}'} \ [M^\gamma \to \bullet\delta, k, k \mid -,-]$$

$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{FootComp}}$**:** Dado un paso $\frac{[M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-], [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']}{[\mathbf{F}^\beta \to \bot\bullet, k, l \mid k, l]}$ existe un paso $\frac{[M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-]}{[\mathbf{F}^\beta \to \bot\bullet, k, l \mid k, l]} \in \mathcal{D}_{\mathrm{E}'}^{\mathrm{FootComp}}$ y por tanto existe la inferencia

$$\begin{array}{l} [M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [\mathbf{F}^\beta \to \bullet\bot, k, k \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q'] \\ \vdash_{\mathrm{E}'}\ [\mathbf{F}^\beta \to \bot\bullet, k, l \mid k, l] \end{array}$$

$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjComp}^1}$**:** Dado un paso $\frac{[\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [\mathbf{F}^\gamma \to \bot\bullet, p, q \mid p, q],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid -,-]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q]}$ existe un paso $\frac{[\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid -,-]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q]} \in \mathcal{D}_{\mathrm{E}'}^{\mathrm{AdjComp}}$ y por tanto existe la inferencia

$$\begin{array}{l} [\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid p, q],\ [\mathbf{F}^\gamma \to \bot\bullet, p, q \mid p, q],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid -,-] \\ \vdash_{\mathrm{E}'}\ [N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q] \end{array}$$

$\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjComp}^2}$**:** Dado un paso $\frac{[\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q']}$ existe un paso $\frac{[\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']}{[N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q']} \in \mathcal{D}_{\mathrm{E}'}^{\mathrm{AdjComp}}$ y por tanto existe la inferencia

$$\begin{array}{l} [\top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\ [M^\gamma \to \upsilon\bullet, k, l \mid -,-],\ [N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q'] \\ \vdash_{\mathrm{E}'}\ [N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q'] \end{array}$$

□

La complejidad temporal del esquema de análisis sintáctico **Ear** con respecto a la longitud n de la cadena de entrada es $\mathcal{O}(n^6)$ puesto que la aparente complejidad $\mathcal{O}(n^7)$ de los pasos deductivos $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjComp}^1}$ y $\mathcal{D}_{\mathrm{Ear}}^{\mathrm{AdjComp}^1}$ puede reducirse a $\mathcal{O}(n^6)$ mediante la aplicación parcial o *currificación* de dichos pasos, puesto que el índice l sólo involucra a los dos primeros ítems antecedentes en cada uno de ellos.

3.6 Algoritmos de tipo Earley con la propiedad del prefijo válido

El primer algoritmo de análisis sintáctico para TAG que satisfacía la propiedad del prefijo válido fue el descrito por Schabes y Joshi en [173] y por Schabes en [168]. La principal particularidad de dicho algoritmo es que su complejidad temporal con respecto a la cadena de entrada es $\mathcal{O}(n^7)$, tal como muestran Díaz Madrigal et al. en [65], mientras que los algoritmos sin la propiedad del prefijo válido presentan una complejidad $\mathcal{O}(n^6)$. Durante mucho tiempo cobró fuerza la opinión de que aquellos algoritmos que cumpliesen la propiedad del prefijo válido deberían tener una complejidad más alta que aquellos que no la cumpliesen. Sin embargo, Nederhof presentó en [125] un algoritmo para el análisis de TAG que cumple la propiedad del prefijo válido[6] y que presenta una complejidad temporal $\mathcal{O}(n^6)$. Veremos que dicho algoritmo es fácilmente derivable a partir del esquema de análisis **Ear** presentado en la sección anterior, correspondiente al algoritmo de tipo Earley sin la propiedad del prefijo válido.

El esquema de análisis sintáctico **Ear** describe un algoritmo que no cumple la propiedad del prefijo válido porque los pasos deductivos que se encargan de reconocer el nodo correspondiente al pie de un árbol auxiliar no pueden verificar la contigüidad de las fronteras del árbol al que pertenece el nodo de adjunción y del árbol auxiliar. Analicemos detalladamente dichos pasos:

- El paso deductivo $\mathcal{D}_{\text{Ear}}^{\text{FootPred}}$ puede verificar, mediante el ítem $[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]$, que existe un nodo M^γ en el que el árbol auxiliar β puede ser adjuntado para reconocer la cadena de entrada a partir de la posición j. También puede verificar, mediante el ítem $[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -, -]$, que se ha alcanzado el nodo pie del árbol auxiliar β. Lo que no puede verificar este paso deductivo es que el árbol al que pertenece dicho nodo pie se corresponda con la instancia de β que ha sido utilizada en la operación de adjunción que nos ocupa, pues para ello tendría que verificar que el extremo izquierdo de la frontera de la instancia de β es j, información que no es posible obtener a partir de los ítems definidos para el esquema de análisis **Ear**.

- El paso deductivo $\mathcal{D}_{\text{Ear}}^{\text{FootComp}}$ puede verificar, mediante el ítem $[N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]$, que existe un nodo M^γ en el que el árbol auxiliar β puede ser adjuntado para reconocer la cadena de entrada a partir de la posición j. También puede verificar, mediante el ítem $[\mathbf{F}^\beta \to \bullet\bot, k, k \mid -, -]$, que se ha alcanzado el nodo pie del árbol auxiliar β en la posición k de la cadena de entrada. Por último, el ítem $[M^\gamma \to \delta\bullet, k, l \mid p, q]$ permite verificar que la frontera del subárbol enraizado en M^γ comienza en la posición k de la cadena de entrada. Pero al igual que en el caso anterior y por las mismas razones, no puede verificar que el árbol al que pertenece el nodo pie se corresponde con la instancia de β que ha sido utilizada en la operación de adjunción que nos ocupa.

En consecuencia, para obtener un esquema de análisis que se corresponda con un algoritmo del tipo Earley para TAG que posea la propiedad del prefijo válido es necesario modificar la forma de los ítems para incluir un nuevo elemento, un índice que indique la posición del extremo izquierdo de la frontera del árbol al que se refieren los nodos de cada ítem que se genere. Esta operación se corresponde con la aplicación de un refinamiento de los ítems utilizados hasta el momento. Los nuevos ítems son de la forma

$$\left\{ \begin{array}{ll} [h, N^\gamma \to \delta \bullet \nu, i, j \mid p, q] \mid & \exists \alpha \in \boldsymbol{I},\ \mathbf{R}^\alpha \stackrel{*}{\Rightarrow} a_1 \ldots a_h \mathbf{R}^\gamma \mu,\ \mathbf{R}^\gamma \stackrel{*}{\Rightarrow} a_{h+1} \ldots a_i \delta \nu \upsilon \text{ y además:} \\ & \delta \stackrel{*}{\Rightarrow} a_i \ldots a_p\, \mathbf{F}^\gamma\, a_{q+1} \ldots a_j \stackrel{*}{\Rightarrow} a_i \ldots a_j \text{ sii } (p,q) \neq (-,-) \\ & \delta \stackrel{*}{\Rightarrow} a_i \ldots a_j \text{ sii } (p,q) = (-,-) \end{array} \right\}$$

con lo cual un ítem $[N^\gamma \to \delta \bullet \nu, i, j \mid p, q]$ de **Ear** se corresponde ahora con el conjunto de ítems $[h, N^\gamma \to \delta \bullet \nu, i, j \mid p, q]\ \forall h \in [0, n]$.

[6] Para una comparación experimental entre los algoritmos de Schabes y Nederhof, consultar [63].

Una vez definidos los nuevos ítems podemos pasar a describir el esquema de análisis **Earley** que corresponde a la primera versión de un algoritmo de tipo Earley para TAG que cumple la propiedad del prefijo válido. El correspondiente sistema de análisis sintáctico se define a continuación.

Esquema de análisis sintáctico 3.6 El sistema de análisis $\mathbb{P}_{\text{Earley}}$ que se corresponde con la el algoritmo de análisis sintáctico de tipo Earley para TAG que cumple la propiedad del prefijo válido, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \ldots a_n$ se define como sigue:

$$\mathcal{I}_{\text{Earley}} = \left\{ \begin{array}{ll} [h, N^\gamma \to \delta \bullet \nu, i, j \mid p, q] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \\ & 0 \leq h \leq i \leq j,\ (p,q) \leq (i,j) \end{array} \right\}$$

$$\mathcal{D}^{\text{Init}}_{\text{Earley}} = \frac{}{\vdash [0, \top \to \bullet \mathbf{R}^\alpha, 0, 0 \mid -, -]} \quad \alpha \in \boldsymbol{I}$$

$$\mathcal{D}^{\text{Scan}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet a\nu, i, j \mid p, q],\ [a, j, j+1]}{[h, N^\gamma \to \delta a \bullet \nu, i, j+1 \mid p, q]}$$

$$\mathcal{D}^{\text{Pred}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p, q]}{[h, M^\gamma \to \bullet \upsilon, j, j \mid -, -]} \quad \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{Comp}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma\nu, i, k \mid p, q],\ [h, M^\gamma \to \upsilon \bullet, k, j \mid p', q']}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p \cup p', q \cup q']} \quad \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{AdjPred}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p, q]}{[j, \top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{FootPred}}_{\text{Earley}} = \frac{[j, \mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -],\ [h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p, q]}{[h, M^\gamma \to \bullet \delta, k, k \mid -, -]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{FootComp}}_{\text{Earley}} = \frac{\begin{array}{l} [h, M^\gamma \to \delta \bullet, k, l \mid p, q], \\ [j, \mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -], \\ [h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p', q'] \end{array}}{[j, \mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]} \quad \begin{array}{l} \beta \in \text{adj}(M^\gamma), \\ p \cup p' \text{ y } q \cup q' \text{ está definido} \end{array}$$

$$\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}} = \frac{\begin{array}{l} [j, \top \to \mathbf{R}^\beta \bullet, j, m \mid k, l], \\ [h, M^\gamma \to \upsilon \bullet, k, l \mid p, q], \\ [h, \mathbf{F}^\gamma \to \bot \bullet, p, q \mid p, q], \\ [h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid -, -] \end{array}}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}} = \frac{\begin{array}{l} [j, \top \to \mathbf{R}^\beta \bullet, j, m \mid k, l], \\ [h, M^\gamma \to \upsilon \bullet, k, l \mid -, -], \\ [h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p', q'] \end{array}}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q']} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}_{\text{Earley}} = \mathcal{D}^{\text{Init}}_{\text{Earley}} \cup \mathcal{D}^{\text{Scan}}_{\text{Earley}} \cup \mathcal{D}^{\text{Pred}}_{\text{Earley}} \cup \mathcal{D}^{\text{Comp}}_{\text{Earley}} \cup \mathcal{D}^{\text{AdjPred}}_{\text{Earley}}$$
$$\cup\, \mathcal{D}^{\text{FootPred}}_{\text{Earley}} \cup \mathcal{D}^{\text{FootComp}}_{\text{Earley}} \cup \mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}} \cup \mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$$

$$\mathcal{F}_{\text{Earley}} = \left\{ \ [0, \top \to \mathbf{R}^{\alpha}\bullet, 0, n \mid -,-] \mid \alpha \in \boldsymbol{I} \ \right\}$$

§

En la figura 3.5 se muestra una representación gráfica de la aplicación del paso deductivo $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$. Es interesante comparar este nueva figura con la 3.4 correspondiente al mismo paso deductivo del algoritmo de tipo Earley sin la propiedad del prefijo válido con el fin de observar cómo se restringen los árboles candidatos en la aplicación del paso deductivo. Se puede observar que la única diferencia entre ambas radica en que el extremo izquierdo del árbol γ está explícitamente indicado por el índice h en la figura 3.5, mientras que en la figura 3.4 se consideraba universalmente cuantificado.

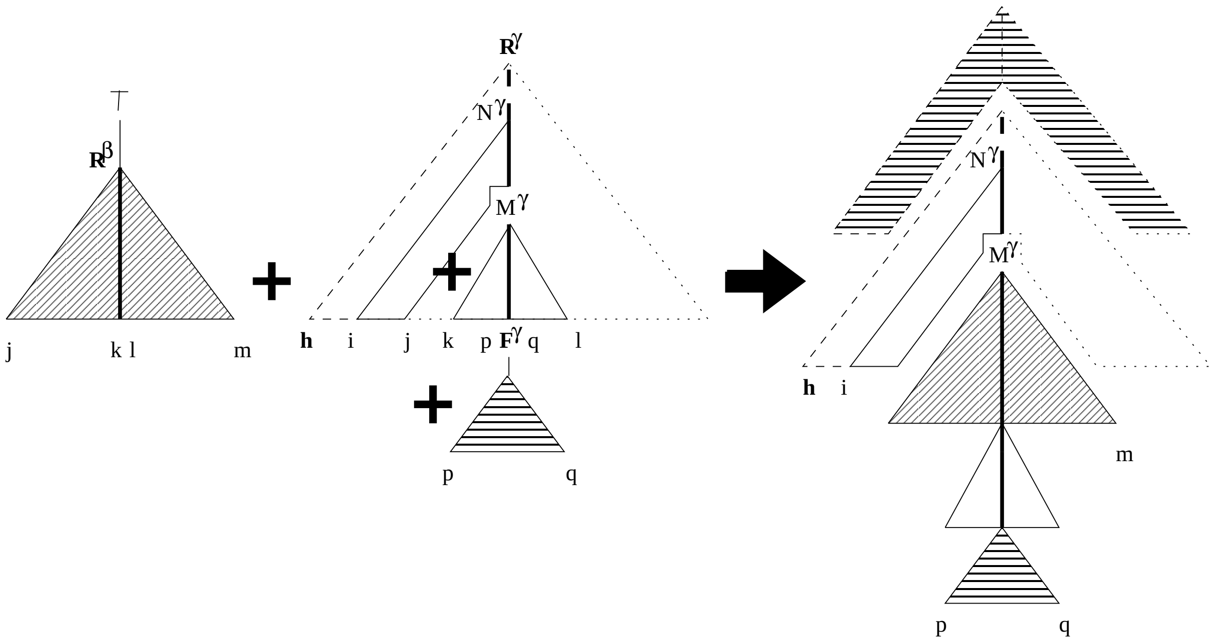

Figura 3.5: Descripción gráfica de un paso $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$

Proposición 3.5 **Ear** $\overset{\text{ir}}{\Longrightarrow}$ **Earley**.

Demostración:

Para demostrar que el esquema de análisis **Earley** es derivable del esquema de análisis **Ear** mediante refinamiento de los ítems definiremos la siguiente función:

$$f([h, N^{\gamma} \to \delta \bullet \nu, i, j \mid p, q \mid adj]) = [N^{\gamma} \to \delta \bullet \nu, i, j \mid p, q \mid adj]$$

de la cual se obtiene directamente que $\mathcal{I}_{\text{Ear}} = f(\mathcal{I}_{\text{Earley}})$ y que $\Delta_{\text{Ear}} = f(\Delta_{\text{Earley}})$ por inducción en la longitud de las secuencias de derivación. En consecuencia, $\mathbb{P}_{\text{Ear}} \xrightarrow{\text{ir}} \mathbb{P}_{\text{Earley}}$, con lo que hemos probado lo que pretendíamos. $\square$

Un aspecto interesante a tener en cuenta es que el ítem $[h, \mathbf{F}^\gamma \to \bot\bullet, p, q \mid p, q]$ es redundante en el paso $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$, puesto que su existencia viene implícitamente determinada por la existencia del ítem $[h, M^\gamma \to \delta\bullet, k, l \mid p, q]$, ya que en otro caso este último sería inconsistente y por consiguiente algoritmo sería incorrecto. La finalidad de su presencia en dicho conjunto de pasos deductivos, así como en $\mathcal{D}^{\text{AdjComp}^1}_{\text{Ear}}$, es facilitar la transición hacia el esquema de análisis **Nederhof**. Si prescindimos de dicho ítem los pasos deductivos $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ se podrían fundir en uno solo:

$$\mathcal{D}^{\text{AdjComp}}_{\text{Earley}} = \frac{\begin{array}{l}[j, \top \to \mathbf{R}^\beta\bullet, j, m \mid k, l],\\ [h, M^\gamma \to \upsilon\bullet, k, l \mid p, q],\\ [h, N^\gamma \to \delta \bullet M^\gamma\nu, i, j \mid p', q']\end{array}}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p \cup p', q \cup q']} \quad \beta \in \text{adj}(M^\gamma)$$

Por idénticas razones, los pasos $\mathcal{D}^{\text{AdjComp}^1}_{\text{Ear}}$ y $\mathcal{D}^{\text{AdjComp}^1}_{\text{Ear}}$ del esquema **Ear** podrían fundirse en un nuevo paso $\mathcal{D}^{\text{AdjComp}}_{\text{Ear}}$:

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

En consecuencia, en lugar de la evolución $\mathbf{E} \overset{\mathbf{sr}}{\Longrightarrow}\overset{\mathbf{df}}{\Longrightarrow} \mathbf{Ear} \overset{\mathbf{ir}}{\Longrightarrow} \mathbf{Earley}$ podríamos haber definido una línea evolutiva $\mathbf{E} \overset{\mathbf{df}}{\Longrightarrow} \mathbf{Ear}' \overset{\mathbf{ir}}{\Longrightarrow} \mathbf{Earley}' \overset{\mathbf{sr}}{\Longrightarrow}\overset{\mathbf{df}}{\Longrightarrow} \mathbf{Earley}$, donde $\mathbf{Ear}'$ y $\mathbf{Earley}'$ son como **Ear** y **Earley**, respectivamente, excepto por la sustitución de los pasos AdjComp^1 y AdjComp^2 por AdjComp. Este resultado viene a mostrar una vez más que existen varios caminos para transformar un esquema de análisis sintáctico en otro, tal como establece Sikkel en [189] para el caso de los algoritmos de análisis de gramáticas independientes del contexto y que nosotros mostramos aquí para el caso de las gramáticas de adjunción de árboles.

El algoritmo descrito por el esquema **Earley** presenta una complejidad temporal de $\mathcal{O}(n^7)$. Aunque aparentemente la utilización de 8 índices con respecto a la cadena de entrada en los pasos deductivos $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ hace pensar en una complejidad $\mathcal{O}(n^8)$, la utilización de aplicación parcial o *currificación* en dichos pasos reduce la complejidad hasta $\mathcal{O}(n^7)$. En el caso de $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ la aplicación parcial sobre los dos primeros ítems involucra combinar 7 índices, de donde resulta la complejidad $\mathcal{O}(n^7)$ con respecto a la cadena de entrada, pero únicamente los 5 índices j, m, h, p y q forman parte del resultado intermedio puesto que son los únicos que se necesitarán en posteriores aplicaciones parciales. La siguiente aplicación parcial combina el ítem intermedio con el tercer ítem del paso deductivo, operación que involucra únicamente a los 5 ítems mencionados anteriormente, que son conservados en el resultado intermedio producido. Por último, la aplicación parcial con el cuarto ítem involucra la combinación de los 6 índices h, i, j, m, p, q. El caso de $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ es análogo al de $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$.

El aumento de la complejidad de $\mathcal{O}(n^6)$ a $\mathcal{O}(n^7)$ se debe al índice adicional incorporado en los ítems y que indica la posición del extremo izquierdo de la frontera del árbol que se está considerando. La inclusión en los ítems de este índice había surgido por la necesidad de controlar que se están utilizando los árboles correctos para reconocer el pie de un árbol auxiliar. En consecuencia, dicho índice sólo es de real utilidad en los pasos FootPred y FootComp. El resto de los pasos deductivos únicamente propagan el valor de dicho índice. Por consiguiente, en caso de que sea necesario estos últimos pasos deductivos pueden ser refinados, dividiéndolos en varios pasos con el fin de generar ítems intermedios carentes de dicho índice. Esta técnica, si se aplica convenientemente, puede llegar a reducir la complejidad de los algoritmos de análisis. Evidentemente, hay que verificar que los ítems intermedios portan la información necesaria para que el resultado de la composición de los pasos deductivos obtenidos mediante el refinamiento de uno dado sea equivalente al resultado obtenido por aplicar directamente el paso deductivo sin refinar.

En el caso completo del esquema de análisis **Earley**, para reducir la complejidad a $\mathcal{O}(n^6)$ es suficiente con hacer uso de la *propiedad de independencia del contexto de TAG* [214]. Básicamente, lo que dicha propiedad establece es que cada operación de adjunción es independiente de la previa o posterior aplicación de cualquier otra operación de adjunción. Una consecuencia de esta propiedad es que si en un nodo M^γ de un árbol γ está permitida la adjunción de un árbol auxiliar β y se cumplen las tres condiciones siguientes:

1. la parte izquierda de la frontera de $\gamma - M^\gamma$ se extiende desde la posición h hasta la posición j de la cadena de entrada, donde $\gamma - M^\gamma$ denota el resultado de escindir el subárbol enraizado por M^γ de γ;

2. la frontera del árbol β se expande desde la posición j hasta la posición m de la cadena de entrada con una discontinuidad en el pie desde la posición k hasta la l;

3. la frontera del subárbol enraizado por M^γ abarca precisamente desde la posición k hasta la posición l;

como resultado de la adjunción de β en M^γ la frontera del subárbol enraizado por este último nodo se expande desde las posición j hasta la m sin discontinuidad y dicho subárbol puede ser insertado en todo árbol $\beta - M^\gamma$ cuya frontera izquierda finalice en la posición j independientemente de la posición h en la que se sitúe el extremo izquierdo de dicha frontera. Precisamente, los algoritmos de tipo Earley para TAG que no cumplen la propiedad del prefijo válido hacen uso de esta propiedad para verificar la corrección de la adjunción realizada en los pasos AdjComp. Como se recordará de la sección precedente, los ítems de los esquemas de análisis de dichos algoritmos no incorporan el índice h del extremo izquierdo.

En consecuencia, los ítems que utilizaremos en el esquema de análisis sintáctico **Nederhof** correspondiente al algoritmo de tipo Earley para TAG presentado por Nederhof en [125] que preserva la propiedad del prefijo válido con una complejidad $\mathcal{O}(n^6)$, son de dos tipos: los definidos para el esquema **Earley** y los pseudo-ítem que definimos a continuación

$$\left\{ \begin{array}{lll} [[N^\gamma \to \delta \bullet \nu, i, j \mid p, q]] \mid & \delta \stackrel{*}{\Rightarrow} a_i \dots a_p \, \mathbf{F}^\gamma \, a_{q+1} \dots a_j \stackrel{*}{\Rightarrow} a_i \dots a_j & \text{sii } (p,q) \neq (-,-) \\ & \delta \stackrel{*}{\Rightarrow} a_i \dots a_j & \text{sii } (p,q) = (-,-) \end{array} \right\}$$

Ahora podemos definir el esquema de análisis **Nederhof** cuyo sistema de análisis mostramos a continuación.

Esquema de análisis sintáctico 3.7 El sistema de análisis $\mathbb{P}_{\text{Nederhof}}$ que se corresponde con la el algoritmo de análisis sintáctico de tipo Earley para TAG que cumple la propiedad del prefijo válido y posee una complejidad $\mathcal{O}(n^6)$, dada una gramática de adjunción de árboles $\mathcal{T}$ y una cadena de entrada $a_1 \dots a_n$ se define como sigue:

$$\mathcal{I}^{(1)}_{\text{Nederhof}} = \mathcal{I}_{\text{Earley}} = \left\{ \begin{array}{ll} [h, N^\gamma \to \delta \bullet \nu, i, j \mid p, q] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \\ & 0 \leq h \leq i \leq j,\ (p,q) \leq (i,j) \end{array} \right\}$$

$$\mathcal{I}^{(2)}_{\text{Nederhof}} = \left\{ \begin{array}{ll} [[N^\gamma \to \delta \bullet \nu, i, j \mid p, q]] \mid & N^\gamma \to \delta\nu \in \mathcal{P}(\gamma),\ \gamma \in \boldsymbol{I} \cup \boldsymbol{A}, \\ & 0 \leq h \leq i \leq j,\ (p,q) \leq (i,j) \end{array} \right\}$$

$$\mathcal{I}_{\text{Nederhof}} = \mathcal{I}^{(1)}_{\text{Nederhof}} \cup \mathcal{I}^{(2)}_{\text{Nederhof}}$$

$$\mathcal{D}^{\text{Init}}_{\text{Nederhof}} = \mathcal{D}^{\text{Init}}_{\text{Earley}} = \frac{}{\vdash [0, \top \to \bullet \mathbf{R}^\alpha, 0, 0 \mid -, -]} \quad \alpha \in \boldsymbol{I}$$

$$\mathcal{D}^{\text{Scan}}_{\text{Nederhof}} = \mathcal{D}^{\text{Scan}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet a\nu, i, j \mid p, q],\ [a, j, j+1]}{[h, N^\gamma \to \delta a \bullet \nu, i, j+1 \mid p, q]}$$

$$\mathcal{D}^{\text{Pred}}_{\text{Nederhof}} = \mathcal{D}^{\text{Pred}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[h, M^\gamma \to \bullet \upsilon, j, j \mid -, -]} \quad \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{Comp}}_{\text{Nederhof}} = \mathcal{D}^{\text{Comp}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma \nu, i, k \mid p, q],\ [h, M^\gamma \to \upsilon \bullet, k, j \mid p', q']}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, j \mid p \cup p', q \cup q']} \quad \mathbf{nil} \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{AdjPred}}_{\text{Nederhof}} = \mathcal{D}^{\text{AdjPred}}_{\text{Earley}} = \frac{[h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[j, \top \to \bullet \mathbf{R}^\beta, j, j \mid -, -]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{FootPred}}_{\text{Nederhof}} = \mathcal{D}^{\text{FootPred}}_{\text{Earley}} = \frac{[j, \mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -],\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p, q]}{[h, M^\gamma \to \bullet \delta, k, k \mid -, -]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{FootComp}}_{\text{Nederhof}} = \mathcal{D}^{\text{FootComp}}_{\text{Earley}} = \frac{\begin{array}{l}[h, M^\gamma \to \upsilon \bullet, k, l \mid p, q],\\ [j, \mathbf{F}^\beta \to \bullet \bot, k, k \mid -, -],\\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[j, \mathbf{F}^\beta \to \bot \bullet, k, l \mid k, l]} \quad \begin{array}{l}\beta \in \text{adj}(M^\gamma),\\ p \cup p' \ y\ q \cup q' \ est\acute{a}\ definido\end{array}$$

$$\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}} = \frac{\begin{array}{l}[j, \top \to \mathbf{R}^\beta \bullet, j, m \mid k, l],\\ [h, M^\gamma \to \upsilon \bullet, k, l \mid p, q],\end{array}}{[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]]} \quad \beta \in \text{adj}(M^\gamma)$$

$$\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}} = \frac{\begin{array}{l}[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]],\\ [h, \mathbf{F}^\gamma \to \bot \bullet, p, q \mid p, q],\\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid -, -]\end{array}}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q]}$$

$$\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}} = \frac{\begin{array}{l}[[M^\gamma \to \upsilon \bullet, j, m \mid -, -]],\\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']\end{array}}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q']}$$

$$\begin{aligned}\mathcal{D}_{\text{Nederhof}} = \ & \mathcal{D}^{\text{Init}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{Scan}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{Pred}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{Comp}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{AdjPred}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{FootPred}}_{\text{Nederhof}} \\ & \cup \mathcal{D}^{\text{FootComp}}_{\text{Nederhof}} \cup \mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}} \cup \mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}} \cup \mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}\end{aligned}$$

$$\mathcal{F}_{\text{Nederhof}} = \mathcal{F}_{\text{Earley}} = \left\{ \ [0, \top \to \mathbf{R}^\alpha \bullet, 0, n \mid -, -] \mid \alpha \in \boldsymbol{I} \ \right\}$$

§

Obsérvese que se ha aplicado un refinamiento al paso deductivo $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ para obtener los pasos $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$. Análogamente, el paso deductivo $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ ha sido refinado en los pasos $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$. Para garantizar la corrección se verifica que:

- La aplicación consecutiva de $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ (resp.$\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$) es equivalente a la aplicación de $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ (resp. $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$). Es fácil comprobar que ambos utilizan la misma información: todos los antecedentes de $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$ (resp. $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$) son utilizados por pasos del esquema **Nederhof** y toda la información presente en los antecedentes de los pasos del esquema **Nederhof** es utilizada por el paso $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$ (resp. $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$), puesto que el ítem intermedio no crea nueva información, sino que simplemente es un "resumen" de información contenida en los otros antecedentes. Es también fácil comprobar que en ambos casos se genera la misma información, puesto que los ítems generados (excluyendo pseudo-ítems) son idénticos en ambos casos.

- El paso deductivo $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$ (resp. $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$) solo puede ser aplicado si previamente se ha aplicado el paso $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$. Se puede verificar fácilmente puesto que el paso $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ genera un ítem intermedio y los únicos pasos que toman un ítem intermedio como antecedente son $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Nederhof}}$.

En la figura 3.5 se muestra una representación gráfica de la aplicación de los pasos deductivos $\mathcal{D}^{\text{AdjComp}^0}_{\text{Nederhof}}$ y $\mathcal{D}^{\text{AdjComp}^1}_{\text{Nederhof}}$.

La complejidad del algoritmo descrito por el esquema de análisis **Nederhof** es $\mathcal{O}(n^6)$, puesto que en la combinación de los ítems de cualquier paso intervienen activamente a los sumo 6 índices con respecto a la cadena de entrada.

En [64] se describe una versión de este algoritmo en la que se utiliza una representación plana de los árboles elementales en lugar de la representación multicapa que se ha utilizado aquí.

Proposición 3.6 Earley $\stackrel{\text{sr}}{\Longrightarrow}$ Nederhof.

Demostración:

Para demostrar que el esquema de análisis **Nederhof** puede ser obtenido mediante un refinamiento de los pasos deductivos del esquema **Earley** debemos probar que para todo sistema de análisis $\mathbb{P}_{\text{Earley}}$ y $\mathbb{P}_{\text{Nederhof}}$ se cumple $\mathbb{P}_{\text{Earley}} \stackrel{\text{sr}}{\Longrightarrow} \mathbb{P}_{\text{Nederhof}}$. Ello conlleva demostrar que $\mathcal{I}_{\text{Earley}} \subseteq \mathcal{I}_{\text{Nederhof}}$ y que $\stackrel{*}{\vdash}_{\text{Earley}} \subseteq \stackrel{*}{\vdash}_{\text{Nederhof}}$. Lo primero se obtiene directamente puesto que $\mathcal{I}_{\text{Earley}} \subseteq \mathcal{I}_{\text{Nederhof}}$ por definición de los sistemas de análisis. Lo segundo se obtiene demostrando que $\mathcal{D}_{\text{Earley}} \subseteq \stackrel{*}{\vdash}_{\text{Nederhof}}$. Los únicos pasos deductivos de $\mathbb{P}_{\text{Earley}}$ que no se han incorporado directamente en $\mathbb{P}_{\text{Nederhof}}$ son $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ y $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ y para ellos se cumple que:

- Un paso $\mathcal{D}^{\text{AdjComp}^1}_{\text{Earley}}$ es equivalente a la aplicación de un paso $\mathcal{D}^{\text{Comp}^0}_{\text{Nederhof}}$ seguido de la aplicación de un paso $\mathcal{D}^{\text{Comp}^1}_{\text{Nederhof}}$:

$$\frac{[j, \top \to \mathbf{R}^\beta \bullet, j, m \mid k, l],\ [h, M^\gamma \to \upsilon \bullet, k, l \mid p, q],}{[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]]}$$

$$\frac{[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]],\ [h, \mathbf{F}^\gamma \to \bot \bullet, p, q \mid p, q],\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid -, -]}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p, q]}$$

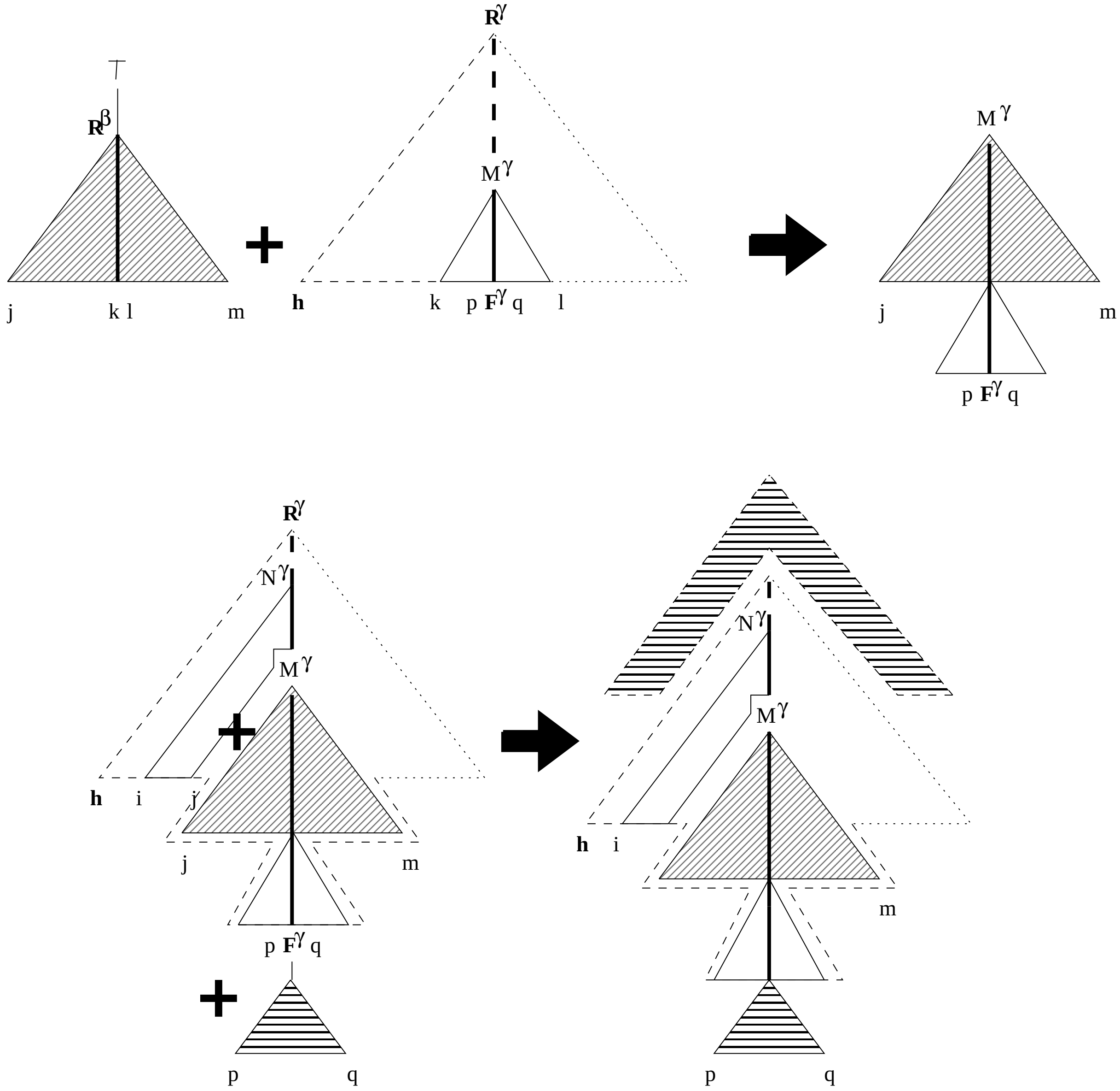

Figura 3.6: Descripción gráfica de la aplicación consecutiva de los pasos $\mathcal{D}^{\mathrm{AdjComp}^0}_{\mathrm{Nederhof}}$ y $\mathcal{D}^{\mathrm{AdjComp1}}_{\mathrm{Nederhof}}$

- Un paso $\mathcal{D}^{\text{AdjComp}^2}_{\text{Earley}}$ es equivalente a la aplicación de un paso $\mathcal{D}^{\text{Comp}^0}_{\text{Nederhof}}$ seguido de la aplicación de un paso $\mathcal{D}^{\text{Comp}^2}_{\text{Nederhof}}$:

$$\frac{[j, \top \to \mathbf{R}^\beta \bullet, j, m \mid k, l],\ [h, M^\gamma \to \upsilon \bullet, k, l \mid p, q],}{[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]]}$$

$$\frac{[[M^\gamma \to \upsilon \bullet, j, m \mid p, q]],\ [h, N^\gamma \to \delta \bullet M^\gamma \nu, i, j \mid p', q']}{[h, N^\gamma \to \delta M^\gamma \bullet \nu, i, m \mid p', q']}$$

□

3.7 Análisis sintáctico de TAG lexicalizadas

Las gramáticas lexicalizadas poseen una propiedad muy interesante desde el punto de vista del análisis sintáctico: son finitamente ambiguas. Puesto que cada componente de la gramática (en el caso de TAG, cada árbol elemental) está asociado con un componente léxico, solamente un conjunto finito de tales estructuras pueden ser utilizadas para el análisis de una cadena de entrada dada y además solamente existe un número finito de combinaciones de dichas estructuras. En resumen, las gramáticas lexicalizadas impiden la aparición de análisis cíclicos.

El análisis de gramáticas lexicalizadas puede realizarse en dos fases, una primera en la cual se seleccionan todas las estructuras relevantes para la cadena de entrada que se pretende analizar y una segunda en la cual se aplica un algoritmo de análisis sintáctico que combine dichas estructuras. Este tipo de procesamiento se corresponde con un análisis fuera de línea. Las gramáticas lexicalizadas también pueden ser analizadas en línea, de tal modo que según se va avanzado en la lectura de la cadena de entrada se proporcionen las estructuras elementales correspondientes. Algunos autores [174, 168] sugieren que en análisis fuera de línea está mejor adaptado a este tipo de gramáticas puesto que las estructuras seleccionadas en la primera fase posibilitan al analizador sintáctico la utilización de información ascendente no local[7], restringiendo de este modo las posibilidades de combinación de las estructuras e incluso el número de estructuras a considerar. En efecto, al actuar de este modo, el analizador sintáctico solamente considerará aquellas estructuras relevantes para la cadena a analizar, por lo que se podría decir que trabaja sobre una subgramática relevante para la cadena de entrada.

Los beneficios que se obtiene de la lexicalización dependen del algoritmo de análisis que se vaya a aplicar. Los algoritmos puramente ascendentes, del tipo CYK, únicamente se benefician de la reducción del número de estructuras a considerar durante el proceso de análisis. Un algoritmo puramente descendente, basado en una exploración en profundidad con retroceso, conseguiría mayores beneficios, puesto que al ser las gramáticas lexicalizadas finitamente ambiguas el espacio de búsqueda es finito y por lo tanto el análisis terminará en todos los casos[8]. Loa algoritmos mixtos que utilizan información ascendente y descendente, como por ejemplo los algoritmos de tipo Earley, se ven también beneficiados por la lexicalización. Una primera ventaja surge del hecho de que ningún árbol elemental tiene la cadena vacía por frontera, lo cual significa que una adjunción no puede ser predicha y completada sin avanzar en el reconocimiento de la cadena de entrada. Por tanto, al terminación está asegurada. Adicionalmente, la

[7] Esta información puede incluso no estar acotada con respecto a la distancia [168], de tal modo que no se puede imitar su efecto mediante la utilización de un número limitado de símbolos de preanálisis. Esta características se puede aplicar por ejemplo al reconocimiento de frases hechas con expresiones arbitrarias intercaladas.

[8] Un analizador sintáctico puramente descendente puede no terminar para una gramática no lexicalizada puesto que puede intentar repetir indefinidamente el análisis del mismo conjunto de estructuras sin avanzar en el reconocimiento de la cadena de entrada.

utilización de una estrategia de dos fases permite que la selección de estructuras de acuerdo a los componentes léxicos presentes en la cadena de entrada ayude al analizador sintáctico en la tarea de filtrar las predicciones y/o compleciones para la adjunción y la sustitución. Resultados experimentales realizados Schabes y Joshi [174, 168] muestran que la estrategia de dos fases aumenta considerablemente la eficiencia de un algoritmo de análisis sintáctico de tipo Earley para TAG.

Todos los algoritmos mostrados en este capítulo pueden ser fácilmente adaptados a gramáticas de adjunción de árboles lexicalizadas. Para ello sólo es preciso incluir un paso deductivo para tratar la sustitución de un árbol en un nodo de sustitución. Puesto que dicho operación es independiente del contexto, no afecta a la complejidad espacial ni temporal de los algoritmos.

3.8 El bosque de análisis

Los algoritmos mostrados hasta el momento, tal y como han sido descritos, son realmente reconocedores y no analizadores sintácticos, puesto que no construyen árboles de derivación. Sin embargo, cada uno de los pasos deductivos contiene la información necesaria para generar la parte correspondiente de un árbol de derivación y, puesto que todos los algoritmos recorren todas las posibles derivaciones, se pueden reconstruir todos los posibles árboles de derivación.

Puesto que estamos tratando con analizadores no deterministas se trata de construir una estructura, denominada *bosque de análisis* [30, 215] que permita representar todas las derivaciones de un forma compacta, compartiendo subderivaciones comunes, y que permita extraer cada una de las derivaciones en tiempo lineal con respecto al tamaño del bosque de análisis. El problema de la construcción del bosque de análisis para TAG ha sido estudiado con anterioridad por Vijay-Shanker y Weir en [215], que han propuesto dos posibles soluciones: la utilización de gramáticas independientes del contexto y la utilización de gramáticas lineales de índices. Cualquiera de las soluciones es aplicable a los algoritmos de análisis sintáctico mostrados en este capítulo.

3.8.1 Gramáticas independientes del contexto como bosque de análisis

Es posible representar el bosque compartido mediante una gramática independiente del contexto que capture la independencia al contexto de la operación de adjunción. Los no-terminales de la gramática serán de la forma $\langle tb, N^\gamma, i, j, p, q\rangle$ donde $tb \in \{\top, \bot\}$ se utiliza para indicar si el no-terminal representa al nodo N^γ antes ($\bot$) o después ($\top$) de una adjunción. Es interesante observar que los no-terminales son casi idénticos a los ítems utilizados en el esquema de análisis **CYK**. Mediante una pequeña modificación en los pasos deductivos[9] sería posible hacer $adj =$ true siempre que $tb = \top$ y que $adj =$ false siempre que $tb = \bot$. Puesto que los ítems de los restantes esquemas son un refinamiento de los ítems de **CYK** la información necesaria para los no-terminales se puede obtener directamente a partir de los ítems.

Respecto a la forma de las producciones, a modo de ejemplo, mostramos la producción correspondiente a la adjunción del árbol auxiliar β en el nodo N^γ:

$$\langle \top, N^\gamma, i, j, r, s\rangle \to \langle \top, \mathbf{R}^\beta, i, j, p, q\rangle \; \langle \bot, N^\gamma, p, q, r, s\rangle$$

la cual se corresponde con el paso deductivo de adjunción del esquema de análisis **CYK**. Al igual que ocurría con los no-terminales, las producciones del bosque de análisis se pueden obtener directamente a partir de los pasos deductivos en los diferentes esquemas de análisis.

[9] Esencialmente la adición de un paso deductivo $\mathcal{D}_{\mathrm{CYK}}^{\mathrm{NoAdj}} = \frac{[N^\gamma, i, j \mid p, q \mid \mathrm{false}]}{[N^\gamma, i, j \mid p, q \mid \mathrm{true}]}$ es aplicable siempre que la realización de una adjunción sobre N^γ sea opcional.

El número de producciones es $\mathcal{O}(n^6)$ y la construcción de la gramática tienen una complejidad temporal $\mathcal{O}(n^6)$, por lo que la complejidad temporal de los algoritmos permanece inalterable, aunque la complejidad espacial aumenta de $\mathcal{O}(n^4)$ ó $\mathcal{O}(n^5)$ a $\mathcal{O}(n^6)$.

Un aspecto interesante a destacar es que aunque el bosque de análisis construido de esta forma codifica las derivaciones para una cadena de entrada dada, el lenguaje derivado por la gramática independiente del contexto no es importante. Lo que importa es que el lenguaje generado es no vacío si la cadena pertenece a la TAG original y en tal caso las derivaciones para la TAG original puede ser obtenidas en tiempo lineal a partir de las derivaciones de la gramática independiente del contexto que codifica el bosque compartido, siempre que esta haya sido podada para eliminar los símbolos inútiles.

3.8.2 Gramáticas lineales de índices como bosque de análisis

Se puede representar el bosque compartido mediante una gramática lineal de índices utilizando la transformación de TAG en LIG definida en [214]. A modo de ejemplo, las siguiente producciones representan la adjunción del árbol auxiliar β en el nodo N^γ:

$$\langle \top, i, j\rangle[\circ\circ N^\gamma] \rightarrow \langle \top, i, j\rangle[\circ\circ N^\gamma \mathbf{R}^\beta]$$
$$\langle \bot, p, q\rangle[\circ\circ N^\gamma \mathbf{F}^\beta] \rightarrow \langle \bot, p, q\rangle[\circ\circ N^\gamma]$$

La primera producción representa el final de la adjunción mientras que la segunda representa el reconocimiento del nodo pie del árbol auxiliar.

La información contenida en los no-terminales y producciones de la gramática lineal de índices puede obtenerse directamente a partir de los ítems y pasos deductivos de los esquemas de análisis sintáctico.

El tamaño de la gramática es $\mathcal{O}(n^3)$ y el tiempo necesario para construirla es $\mathcal{O}(n^6)$, por lo que las complejidades temporal y espacial de los algoritmos no se ven afectadas.

El inconveniente de esta representación estriba en que no se pueden extraer directamente los árboles de derivación individuales, sino que es preciso construir una estructura auxiliar que toma la forma de un autómata finito que reconoce las pilas de índices asociadas a cada terminal. Una vez construido dicho autómata finito, cada uno de los árboles de derivación puede ser extraído con una complejidad temporal que tiene por cota inferior $\mathcal{O}(n^4)$ y que en el caso de las gramáticas de adjunción de árboles lexicalizadas tiene como cota superior $\mathcal{O}(n^5)$. Puesto que el tamaño del bosque compartido es $\mathcal{O}(n^3)$, no se asegura que en todos los casos la recuperación de los árboles de derivación individuales se pueda realizar en tiempo lineal con respecto al tamaño del bosque compartido.

3.9 Otros algoritmos de análisis sintáctico para TAG

Presentamos a continuación un conjunto de algoritmos de análisis sintáctico de TAG que si bien no están en el camino principal de la evolución de los algoritmos de análisis mostrado anteriormente, presentan interés bien por constituir caminos laterales del camino principal de evolución, bien por haber constituido hitos importantes aunque posteriormente hayan quedado relegados, o bien por constituir ejemplos singulares de aplicación a TAG de ciertas áreas del análisis sintáctico, como la incrementalidad o el paralelismo.

3.9.1 El algoritmo de Lang

Lang describe en [106] un algoritmo tabular de análisis de TAG, con el objetivo principal de mostrar que técnicas de análisis muy generales pueden ser utilizadas para desarrollar un analizador sintáctico de tipo Earley para TAG. Concretamente, Lang utiliza las siguientes técnicas: