دانشکده‌ی علوم ریاضی

نظریه زبان‌ها و ماشین‌ها — ۱۶ آبان ۹۱

جلسه‌ی ۱۶: گرامر مستقل از متن
(درخت تجزیه، معادل بودن تعریف‌ها، گرامر مبهم و زبان ذاتاً مبهم)

مدرّس: دکتر شهرام خزائی — *نگارندگان: مهسا افتخاری حصاری، مریم رنجبر*

مطالبی که در این جلسه به آن‌ها خواهیم پرداخت:

۱. درخت تجزیه[۱]

۲. معادل بودن تعریف‌ها

۳. گرامر مبهم و زبان ذاتاً مبهم[۲]

۱ درخت تجزیه

درخت تجزیه یا درخت اشتقاق، درخت مرتبی است که در آن گره‌های داخلی معرف متغیرها و برگ‌های آن متعلق به $T \cup \{\varepsilon\}$ است.

تعریف ۱ *فرض کنید $G = (V, T, P, S)$ یک گرامر مستقل از متن باشد. یک درخت مرتب، درخت اشتقاقی برای G نامیده می‌شود، اگر و فقط اگر دارای خواص زیر باشد:*

۱. *برچسب ریشه، متغیر شروع، یعنی S، است.*

۲. *هر برگ دارای برچسبی از $T \cup \{\varepsilon\}$ است.*

۳. *هر گره داخلی دارای برچسبی از V است.*

۴. *اگر گره‌ای دارای برچسب ε باشد و فرزندان آن دارای برچسب‌های $X_1, X_2, ... X_n$ باشند؛ آن‌گاه P باید شامل قانونی به شکل زیر باشد:*

$$A \to X_1 X_2 ... X_n$$

۵. *برگی با برچسب ε فاقد همزاد است. یعنی گره‌ای با فرزند ε نمی‌تواند دارای فرزند دیگری باشد.*

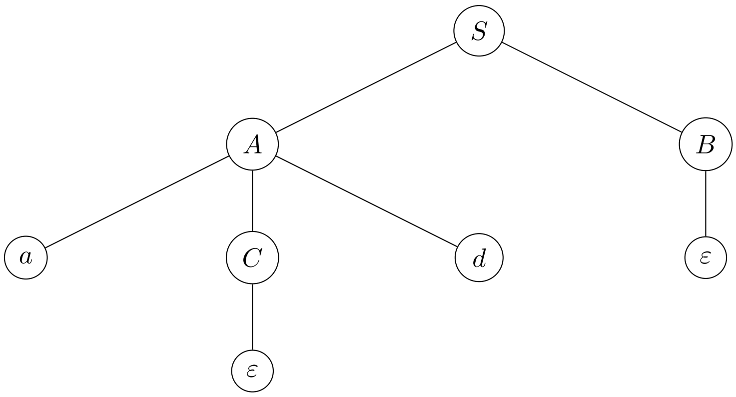
تعریف ۲ *رشته‌ای از پایانه‌ها را که با خواندن برگ‌های درخت از چپ به راست و با حذف ε ها به دست می‌آید، محصول درخت[۳] تجزیه می‌نامیم.*

مثال ۱ *گرامری را در نظر بگیرید که شامل قوانین تولید زیر باشد:*

$$\begin{array}{lll} S & \to & AB \\ A & \to & aCd \\ B & \to & \varepsilon \\ C & \to & \varepsilon \end{array}$$

درخت زیر، یک درخت تجزیه برای این گرامر با محصول $a\varepsilon d\varepsilon = ad$ است.

---
[۱] Parse tree
[۲] Ambiguous grammer
[۳] Yield



سؤال ۱ *آیا هر رشته تنها یک درخت تجزیه یکتا دارد؟*

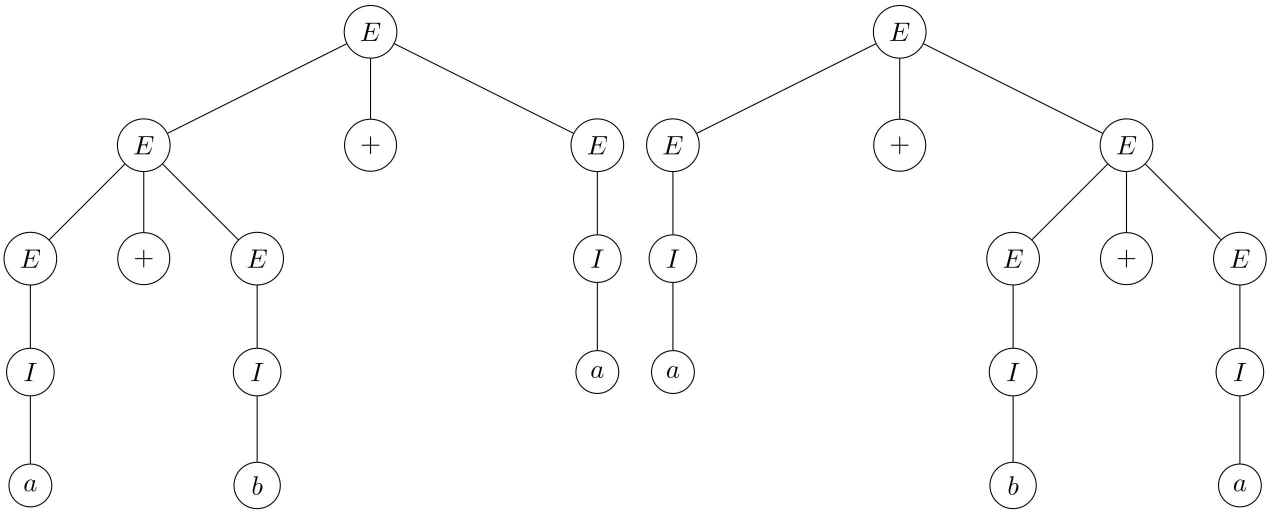
خیر؛ مثال زیر را در نظر بگیرید.

مثال ۲ *گرامری را در نظر بگیرید که شامل قوانین تولید زیر باشد:*

$$E \to E + E|E|I$$

$$I \to a|b$$

واضح است زبان فوق رشته $w = a + b + a$ *را می‌پذیرد. این رشته دارای دو درخت تجزیه متفاوت زیر است:*



در بخش سوم این گونه مثال‌ها را بیش‌تر بررسی می‌کنیم.

۲ معادل بودن تعریف‌ها

در جلسه قبل با روش‌های مختلفی برای تعریف زبان یک گرامر مستقل از متن آشنا شدیم؛ سه روش اشتقاق مکرر، درخت تجزیه و استنتاج بازگشتی معرفی شدند. در این بخش ثابت می‌کنیم این سه تعریف معادل‌اند.

قضیه ۱ *فرض کنید $G = (V, T, P, S)$ یک گرامر مستقل از متن باشد؛ آن‌گاه عبارت‌های زیر معادل‌اند.*

۱. *می‌توان با استنتاج بازگشتی*[۴] *نشان داد که متغیر A رشته w را تولید می‌کند.*

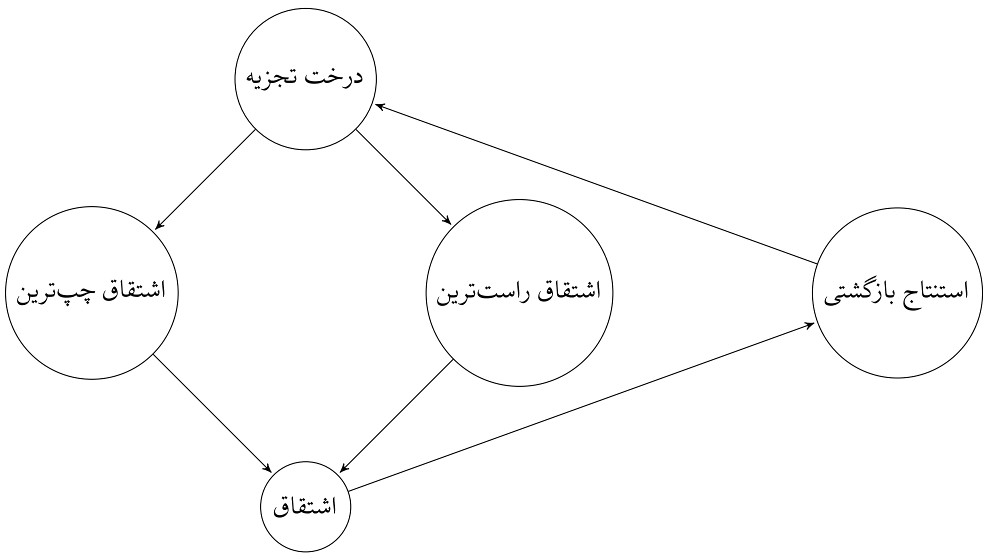
۲. *رشته w اشتقاق مکرری از متغیر A است؛ یعنی: $A \overset{*}{\Rightarrow} w$*

۳. *رشته w اشتقاق چپ‌ترین*[۵] *مکرری از متغیر A است؛ یعنی: $A \underset{lm}{\overset{*}{\Rightarrow}} w$*

۴. *رشته w اشتقاق راست‌ترین*[۶] *مکرری از متغیر A است؛ یعنی: $A \underset{rm}{\overset{*}{\Rightarrow}} w$*

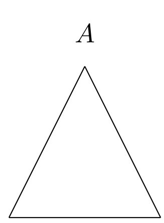
۵. *یک درخت تجزیه با ریشه A و محصول w وجود دارد.*

روند اثبات در نمودار زیر نشان داده شده است:



از استنتاج بازگشتی به درخت تجزیه

برهان. با استقرا روی تعداد مراحل استنتاج نشان می‌دهیم که اگر در n مرحله بتوان استنتاج کرد که متغیر A رشته w را تولید می‌کند، آن‌گاه یک درخت تجزیه با ریشه A و محصول w وجود دارد.
پایه: اگر طی $n = ۱$ مرحله بتوان استنتاج کرد که متغیر A رشته w را تولید می‌کند، می‌توان نتیجه گرفت $A \to w \in P$ (P مجموعه قوانین تولید است).
بنابراین برای چنین قانون تولیدی یک درخت تجزیه به صورت زیر وجود دارد:



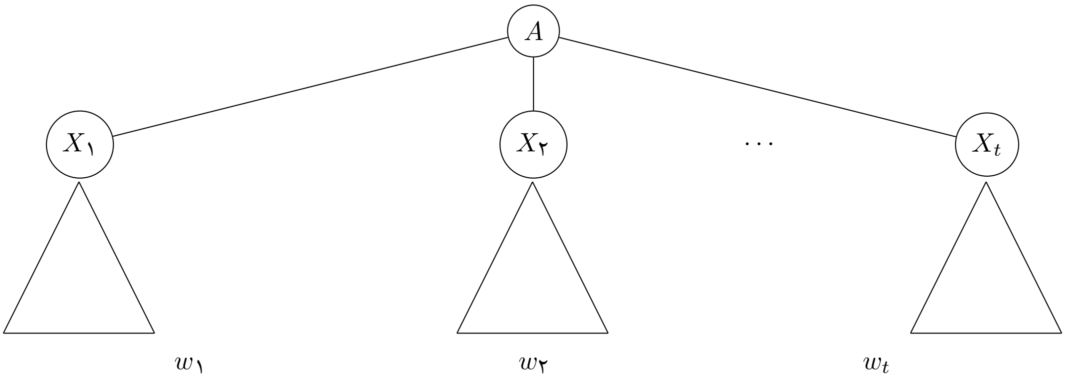
[۴] Recursive inference
[۵] Leftmost derivation
[۶] Rightmost derivation

w

استقرا: فرض کنید در $n + ۱$ مرحله استنتاج کرده‌ایم که متغیر A رشته w را تولید می‌کند. می‌توان نتیجه گرفت در آخرین مرحله استنتاج از قانون تولیدی به صورت زیر برای استنتاج رشته $w = w_۱ w_۲ w_۳ \cdots w_t$ استفاده شده است:

$$A \to X_۱ X_۲ X_۳ \cdots X_t, \forall i : X_i \in V \cup T$$

که اگر X_i یک متغیر باشد، یعنی $X_i \in V$، حداکثر در n مرحله استنتاج کرده‌ایم که متغیر X_i رشته w_i را تولید می‌کند و اگر X_i پایانه باشد، یعنی $X_i \in T$، آن‌گاه $X_i = w_i$. بنابراین طبق فرض استقرا درخت تجزیه‌ای مانند زیر وجود دارد که به ازا $X_i \in T$ باید w_i و X_i را منطبق برهم تصور کنیم.



■

از درخت تجزیه به چپ‌ترین اشتقاق

برهان. اگر یک درخت تجزیه با ریشه A و محصول w داشته باشیم، می‌خواهیم نشان دهیم که یک اشتقاق چپ‌ترین مکرر به صورت زیر وجود دارد:

$$A \underset{lm}{\overset{*}{\Rightarrow}} w$$

این بار روی ارتفاع درخت استقرا می‌زنیم؛
پایه: درخت تجزیه‌ای با ریشه A، ارتفاع $h = ۱$ و محصول w را در نظر بگیرید؛ پس $A \to w$ یکی از قوانین تولید است. طبق تعریف، w یک اشتقاق چپ‌ترین مکرر از A است.
استقرا: فرض کنید w محصول درخت تجزیه‌ای با ریشه A باشد که دارای ارتفاع $h + ۱$ است. چون درخت تجزیه است، پس اولین سطح آن یک قانون تولید است، فرض کنید این قانون تولید به صورت زیر باشد:

$$A \to X_۱ X_۲ X_۳ ... X_t, \forall i : X_i \in T \cup V$$

اگر $X_i \in V$، درخت تجزیه‌ای با ریشه X_i و محصول w_i، زیر درختی با ارتفاع حداکثر h از درخت اصلی است و اگر $X_i \in T$ تعریف می‌کنیم $X_i = w_i$ که $w = w_۱ ... w_t$. بنابر فرض استقرا داریم:

$$X_i \underset{lm}{\overset{*}{\Rightarrow}} w_i$$

می‌خواهیم ثابت کنیم:

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_۱ w_۲ ... w_t$$

کافی است لم زیر را ثابت کنیم تا اثبات تمام شود، اگر لم را درست فرض کنیم و در آن $i = t$ قرار دهیم به حکم مطلوب می‌رسیم.

لم ۲ برای $i = 0, 1, ... t$ داریم:

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 ... w_t X_{i+1} ... X_t$$

اثبات با استقرا روی i صورت می‌گیرد.
پایه:

$$i = 0 : A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 ... X_t$$

که قانون تولید است.
استقرا: فرض کنید $A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 ... w_i X_{i+1} ... X_t$ را داریم، می‌خواهیم $A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 ... w_{i+1} X_{i+2} ... X_t$ را ثابت کنیم.
ادامه اثبات را بر عهده دانشجو می‌گذاریم.

■

به طور مشابه می‌توان برای اشتقاق راست‌ترین نیز نشان داد.

۳ گرامر مبهم

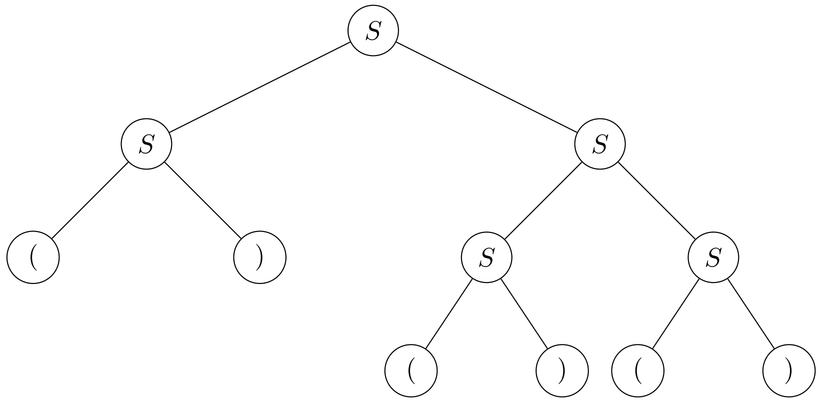
تعریف ۳ گرامر مستقل از متن $G = (V, T, P, S)$ را یک گرامر مبهم می‌گوییم اگر رشته $w \in L(G)$ وجود داشته باشد که دارای دو درخت تجزیه متفاوت باشد.

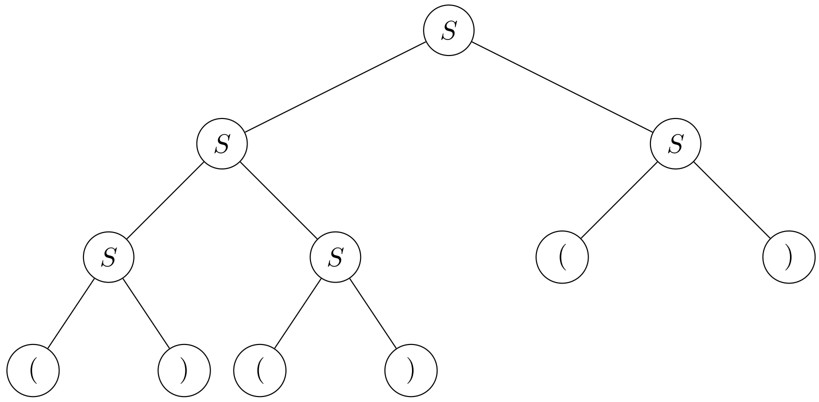
تعریف ۴ زبان‌های ذاتا مبهم به زبان‌هایی می‌گویند که هر گرامری برای آن در نظر بگیریم باز هم مبهم است.

مثال ۳ پرانتزهای متعادل: یک عبارت جبری معتبر بدون عملگرها و عملوندهایش (تنها تشکیل شده از پرانتزها)، یک رشته پرانتزهای متعادل نامیده می‌شود.
به طور دقیق‌تر زبان پرانتزهای متعادل، مجموعه تمام رشته‌هایی از پرانتزهای باز و بسته است که دو شرط زیر را داشته باشد:
۱) تعداد پرانتزهای باز و بسته برابر باشند؛ ۲) هم‌چنین در هر پیشوند آن تعداد پرانتزهای باز، بیشتر یا مساوی تعداد پرانتزهای بسته باشد.
می‌توان نشان داد، زبان گرامر زیر، زبان پرانتزهای متعادل است:

$$S \rightarrow ()|SS|(S)$$

این گرامر مبهم است؛ به طور مثال می‌توان رشته $w = ()()()$ را در نظر گرفت؛ دو درخت تجزیه متفاوت زیر برای این رشته رسم شده‌اند:





سؤال ۲ *آیا می‌توان گرامر زبان مثال۳ را رفع ابهام کرد؟*
پاسخ این سوال مثبت است؛
زبان گرامر زیر را در نظر بگیرید؛ این گرامر بدون ابهام است:

$$
\begin{array}{lcl}
B & \to & \varepsilon \\
R & \to &) \\
R & \to & (RR \\
B & \to & (RB
\end{array}
$$

می‌توان نشان داد (با حذفε) زبان گرامر فوق، که در آن B متغیر شروع است، زبان پرانتزهای متعادل می‌باشد. دقت کنید متغیر B زبان رشته‌های متعادل را تولید می‌کند و متغیرR زبان رشته‌هایی را تولید می‌کند که با اضافه کردن یک پرانتز باز به اول آن (سمت چپ) به رشته متعادل برسیم.

سؤال ۳ *گرامر بدون ابهام برای زبان زیر را بیابید.*

$$
\begin{array}{lcl}
E & \to & I|E+E|E*E|(E) \\
I & \to & a|b|Ia|Ib|I0|I1
\end{array}
$$