# ۱. گرامرهای مستقل از متن و فرم‌های نرمال

## ۱.۱

برای هر کدام از زبان‌های زیر، گرامر مستقل از متن بنویسید.

**پاسخ** a

$L_a = \{a^n b^m \mid n \neq 2m\}$

$S \to aaSb \mid A \mid B$

$A \to aA \mid a$

$B \to bB \mid b$

**پاسخ** b

$L_b = \{\omega \in \{a,b\}^* \mid \forall\, v \in \mathsf{Pref}(\omega)\ .\ n_a(v) \geq n_b(v)\}$

$S \to aSb \mid SS \mid aS \mid \epsilon$

**پاسخ** c

ابتدا بهتر است که گرامری برای زبان ساده‌تر زیر بنویسیم:

$L'_c = \{\omega \in \{a,b\}^* \mid n_a(\omega) = 2n_b(\omega)\}$

$S \to aSab \mid aaSb \mid abSa \mid aSba \mid baSa \mid bSaa \mid \epsilon$

حال آن را به نحوی تغییر می‌دهیم که رشته‌های زبان مورد نظر را تولید کند:

$$L_c = \{\omega \in \{a, b\}^* \mid n_a(\omega) = 2n_b(\omega) + 1\}$$

$$S \to AaA$$

$$A \to aAab \mid aaAb \mid abAa \mid aAba \mid baAa \mid bAaa \mid \epsilon$$

## ۲.۱

الف) برای هر کدام از زبان‌های زیر، یک گرامر ساده بنویسید.

**پاسخ a**

$$a)\ L = \{a^n b^n \mid n \geq 1\}$$

$$S \to aD$$

$$D \to aDB \mid b$$

$$B \to b$$

**پاسخ b**

$$b)\ L = \{a^n b^{n+1} \mid n \geq 2\}$$

$$S \to aA$$

$$A \to aD$$

$$D \to aDB \mid bBB$$

$$B \to b$$

ب) کدام یک از گرامرهای بخش الف مبهم [1] هستند؟ به یک نتیجه‌گیری کلی درخصوص گرامرهای ساده برسید.

### پاسخ (محمدامین شریفی)

ب) هردو گرامر داده شده غیر مبهم هستند. بصورت کلی می‌توان گفت که گرامرهای ساده غیر مبهم هستند (در صورتی که $\epsilon$ در گرامر نباشد). با برهان خلف فرض کنید یک رشته $\omega$ دو $leftmost\,derivation$ متفاوت داشته باشد. اولین مرحله‌ای که این دو اشتقاق قاعده‌های متفاوتی را اعمال می‌کنند را در نظر بگیرید. یعنی هر دو اشتقاق به رشته $\omega' A \cdots$ رسیده‌اند که $\omega'$ تنها شامل پایانه‌ها است و $A$ یک متغیر است. در این مرحله دو اشتقاق با دو قاعده متفاوت رشته‌های $\omega' a x \cdots$ و $\omega' a' x' \cdots$ را ایجاد می‌کنند. دقت کنید که طبق فرض گفته‌شده در مورد گرامرهای ساده $a$ و $a'$ نمی‌توانند یکی باشند (چون از قواعد متفاوت تولید شده‌اند)که یعنی رشته تولید شده توسط دو اشتقاق یکی نیست. تناقض حاصله نشان می‌دهد که گرامرهای ساده غیر مبهم هستند.

### پاسخ آلترناتیو (محسن دهقان‌کار)

ب)

هیچ کدام از گرامر های بالا مبهم نیستند. گرامر های تولید شده، هر دو از نوع LL1 هستند و گرامر های LL1 مبهم نیستند. گرامرهای ساده، همه از نوع LL1 هستند.

در واقع برای این گرامر ها به راحتی می توان جدول پارس LL1 را تولید کرد. زیرا از هر زوج مرتب (A, a) حداکثر یک مورد داریم و این دقیقا یک خانه از این جدول را تولید می کند. مثلا جدول تولید شده برای مورد ب)، این گونه خواهد بود:

| | a | b |
|---|---|---|
| S | S → aABDD | |
| A | A → a | |
| B | B → aBD | B → b |
| D | | D → b |

[1] Ambigious

## ۳.۱

گرامر زیر را در نظر بگیرید:

$$E \rightarrow E \ + \ E \mid E \ * \ E \mid (E) \mid id$$

الف) با اولویت دادن به عمل‌گر ضرب نسبت به جمع، یک گرامر غیرمبهم برای این زبان طراحی کنید و به کمک استقراء، عدم ابهام آن را اثبات کنید.

**پاسخ الف – بخش گرامر غیرمبهم**

$E \rightarrow E + T \mid T$

$E \rightarrow T * F \mid F$

$F \rightarrow id \mid (E)$

**پاسخ الف – بخش اثبات به کمک استقراء**

باید نشان دهیم که به ازای هر رشته‌ی $\omega$ که از هر کدام از متغیرهای $E$ و $T$ و $F$ قابل اشتقاق است، برای $\omega$ تنها یک اشتقاق (راست یا چپ) از آن متغیر وجود دارد. اثبات به کمک استقرای ریاضی روی طول $\omega$ صورت می‌گیرد. در گام پایه، به ازای $\omega = a$ بدیهی است که از هر کدام از متغیرها، تنها یک اشتقاق برای $a$ (که $a \in Terminals$) قابل تعریف است.

در گام استقرایی، باید نشان دهیم که اگر به ازای هر رشته‌ی $\gamma$ به طول کوچک‌تر مساوی $k$ تنها یک اشتقاق قابل تعریف باشد، برای هر رشته‌ی $\omega$ به طول $k + 1$ نیز این موضوع صدق می‌کند.

ابتدا، حالتی را در نظر می‌گیریم که در $\omega$ دست‌کم یک‌بار $+$ خارج از پرانتز ظاهر شده باشد. چنانچه بیشتر از یکبار آمده باشد، به فرض راست‌ترین را در نظر می‌گیریم. از آنجا که $+$ در رشته‌های اشتقاق‌یافته از $T$ و $F$ تنها در داخل پرانتزها به وجود می‌آید، هر اشتقاق چپ $\omega$ باید با قاعده‌ی $E \Rightarrow E + T$ آغاز شده باشد. پس:

$$E \Rightarrow E + T \overset{*}{\Rightarrow} \gamma + T \overset{*}{\Rightarrow} \gamma + z$$

در بالا، فرض کرده‌ایم که $\gamma$ از $E$ و $z$ از $T$ مشتق شده‌اند. از آنجا که رشته‌ی مربوطه به طول $k + 1$ فرض شده بود، پس $z$ و $\gamma$ هرکدام طول کوچک‌تر مساوی $k$ و در نتیجه، تنها یک اشتقاق دارند. پس برای $\omega$ نیز تنها یک اشتقاق وجود دارد.

حالت بعدی، مربوط به زمانی است که همه‌ی +ها در داخل پرانتزها هستند، ولی دست‌کم یک $*$ خارج از پرانتز وجود دارد. درست مشابه حالت قبلی، اینجا نیز در همه‌ی رشته‌هایی که از $F$ مشتق می‌شوند، $*$ها تنها در داخل پرانتز ظاهر می‌گردد. همه‌ی رشته‌هایی که از $E$ مشتق می‌شوند و دارای ویژگی‌های مذکور هستند، باید با $E \Rightarrow T \Rightarrow T * F$ آغاز شوند (دقت داشته باشید که +ها صرفا در پرانتز مورد تایید هستند و در نتیجه، استفاده از قاعده‌ی $E \rightarrow E + T$ مجاز نیست)؛ و همه‌ی آن‌ها که از $T$ مشتق می‌شوند و دارای ویژگی‌های بالا هستند، باید با $T \Rightarrow T * F$ آغاز شوند. باز هم همانند بالا:

$$T * F \overset{*}{\Rightarrow} \gamma * F \overset{*}{\Rightarrow} \gamma * z$$

و اثبات به مانند قبل ادامه می‌یابد.

در حالت پایانی که با بررسی آن، تمامی حالت‌ها تحت پوشش قرار گرفته‌اند، همه‌ی $*$ها و +ها صرفا در داخل پرانتز ظاهر می‌شوند. در این حالت، رشته‌ی $\omega$ می‌تواند از هر سه‌ی متغیرها اشتقاق یابد. برای $E$، فرآیند اشتقاق باید صرفاً با $E \Rightarrow T \Rightarrow F \Rightarrow (E)$ آغاز گردد. برای $T$ و $F$ هم به مانند قبلی، اما با حذف به ترتیب یک و دو مرحله‌ی نخست صورت می‌گیرد. داریم:

$$(E) \overset{*}{\Rightarrow} (\gamma)$$

به مانند حالت‌های گذشته، و با توجه به فرض استقراء، اثبات این حالت و نیز کل قضیه، به اتمام می‌رسد.

**پاسخ آلترناتیو (محمدامین شریفی)**

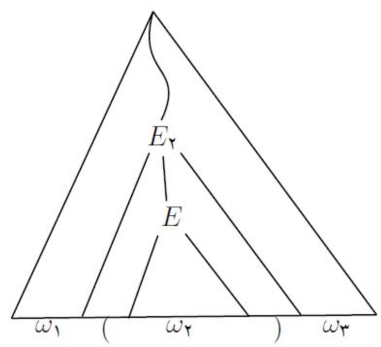
الف) با اضافه کردن یک متغیر جدید این کار را انجام می‌دهیم.

$$\begin{cases} E \rightarrow E_1 + E \mid E_1 \\ E_1 \rightarrow E_2 \times E_1 \mid E_2 \\ E_2 \rightarrow (E) \mid id \end{cases}$$

حال اثبات می‌کنیم هر رشته‌ای از پایانه و متغیرهای این گرامر که توسط گرامر تولید شود یک درخت پارس دارد. این کار را با استقرا قوی روی طول رشته‌ها اثبات می‌کنیم. پایه استقرا برای رشته‌های به طول ۱ که $E$ ، $E_1$ ، $E_2$ و $id$ هستند واضح است. حال فرض کنید حکم برای تمام رشته‌های به طول کمتر از $n$ برقرار باشد و آنرا برای رشته $\omega$ به طول $n$ ثابت می‌کنیم. در این قسمت ۲ حالت داریم.

i) رشته $\omega$ شامل پرانتز است.

با در نظر گرفتن پرانتز بندی این رشته می‌توان دو پرانتز پیدا کرد که در بین آنها هیچ پرانتز دیگری نباشد. بطور معادل $\omega = \omega_1(\omega_2)\omega_3$ که $\omega_2$ شامل هیچ پرانتزی نیست (این کار با در نظر گرفتن درخت پارس رشته $\omega$ و در نظر گرفتن پایین ترین جایی که قاعده $E_2 \rightarrow (E)$ استفاده شده است قابل انجام است). حال با توجه به فرض‌های ما می‌توان گفت که درخت پارس $\omega$ بصورت زیر است.




چون طول رشته‌های $\omega_1 E_2 \omega_3$ و $\omega_2$ از $n$ کمتر است ، طبق فرض استقرا درخت پارس آنها نیز یکتا است و درخت پارس $\omega$ نیز یکتا است که حکم را ثابت می‌کند.

ii) رشته $\omega$ شامل پرانتز نیست. در این صورت می‌توانیم فرض کنیم گرامر قاعده $E_2 \rightarrow (E)$ را ندارد.

در این حالت رشته شامل حداقل یک علامت بعلاوه یا ضرب هست زیرا اگر اینطور نباشد طول آن ۱ است. حال دو حالت داریم.

در حالت اول رشته $\omega$ شامل حداقل یک علامت بعلاوه است. اولین علامت بعلاوه در $\omega$ را در نظر بگیرید.

$$\omega = x + y$$

که رشته $x$ هیچ علامت بعلاوه‌ای ندارد. دقت کنید که چون با متغیر $E_1$ هیچ علامت بعلاوه‌ای تولید نمی‌شود اشتقاق $\omega$ بصورت زیر است.

$$E \Rightarrow E_1 + E, E_1 \overset{*}{\Rightarrow} x, E \overset{*}{\Rightarrow} y$$

چون طول رشته $x$ و $y$ از $n$ کمتر است درخت اشتقاق آنها نیز یکتا است که حکم را در این قسمت ثابت می‌کند. در حالت دوم نیز رشته شامل علامت بعلاوه نیست و تنها علامت ضرب در میان متتغیرها و idها هست. در این حالت نیز اگر اولین علامت ضرب را بگیرید دقیقا مانند حالت قبلی قاعده استفاده شده بدست می‌آید و اثبات می‌شود که این گرامر غیر مبهم است.

ب) زبان گرامر بخش الف را به نحوی تغییر می‌دهیم که هیچ کدام از رشته‌های زبان، پرانتزهای زاید نداشته باشند. پرانتزهای زاید را پرانتزهایی تعریف می‌کنیم که حذف آن‌ها، تاثیری بر حاصل عبارت نمی‌گذارد. برای این زبان جدید یک گرامر غیرمبهم طراحی کنید.

**پاسخ ب**

در عبارت‌هایی که تنها عملگرهای جمع و ضرب را شامل می‌شوند، پرانتزهای زاید در سه حالت بروز می‌یابند. حالت نخست، زمانی است که زوج پرانتز، کل عبارت را در بر گرفته باشد. حالت دیگر، زمانی است که زوج پرانتز، یک عملیات ضرب را دربر گرفته باشد که با توجه به اولویت ذاتی این عملگر، زاید است. حالت آخر، در شرایطی بروز می‌یابد که پرانتز، یک عبارت جمع را در برگرفته باشد، اما بیرون از پرانتز نیز صرفا عملگر جمع باشد که با توجه به خاصیت شرکت‌پذیری، زاید است. گرامر زیر، هر سه‌ی این حالت‌ها را حذف و شرایط موردنظر را ارضاء می‌کند. توصیه می‌گردد که گرامر جدید را با قبلی مقایسه و با تولید رشته‌های مختلف به کمک این گرامر، ارضای خاصیت‌های بالا را بررسی نمایید.

$E_1 \to E_1 + T_1 \mid T_1$

$T_1 \to T_2 * F \mid id$

$F \to id \mid (E_2)$

$E_2 \to E_1 + T_1$

$T_2 \to T_2 * F \mid F$