

DİL TEORİSİ (Language Theory)

Kayıt(Note): Aşağıdakiler, noksan bir dokümandan aktarılmış olup, sadece içerdiği örneklerden yararlanılsın diyedir.

Alfabe: Boş olmayan, sonlu bir simge kümesi. Ör. $\Sigma = \{0, 1\}$, $T = \{a, b, c\}$.

Dizgi: Yan yana bir sıraya dizili simgelerden ibaret bir oluşuma bir **dizgi** denir.

Aksi belirtilmedikçe, söz konusu edilen dizgiler hep sonlu olarak kabul edilecektir.

Ör.

0100	aabab
------	-------

Bir **u** dizgisinin uzunluğu $|u|$, onu oluşturan simgelerin, tekrarlar dâhil, sayısıdır.

Ör. $u_1 = 0100$ için $|u_1| = |0100| = 4$; $u_2 = aabab$ için $|u_2| = |aabab| = 5$

ε : boş dizgi, $|\varepsilon| = 0$

Örneğin: $u = aaba$, $v = bba$

$u \cdot v = uv = aababba$ (bu işleme bitişim(**concatenation**) işlemi denir.) ve bu haliyle tek bir dizgi haline gelir.

Bitişim işleminin etkisiz elemanı; $\varepsilon \cdot u = u \cdot \varepsilon = u$

u_1, u_2, u_3 dizgileri için $w = u_1 \cdot u_2 \cdot u_3$ ise u_2 , w -nun bir **alt-dizgisidir**.

Bir dizginin en solunda bulunan bir alt-dizgiye o dizginin **ön-takısı**, en sağında bulunan bir alt-dizgiye de o dizginin **art-takısı** denir.

Her dizgi kendisinin hem alt-dizgisi, hem ön-takısı, hem de art-takısıdır.

Bir dizginin alt-dizgisi, ön-takısı veya art-takısı kendisinden farklı ise ona sırasıyla, özalt-dizgisi, özön-takısı veya özart-takısı denir.

Σ^* : Sadece Σ -nın simgelerinden oluşan bütün sonlu dizgiler kümesi.

Örnek: $\Sigma = \{0, 1\}$ burada 0 ve 1 simgedir, rakam değildir.

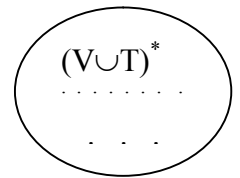
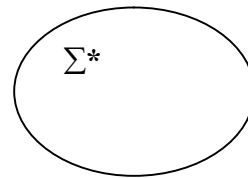
$\Sigma^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \dots\}$

burada 0 uzunlukta 1 tane, dizgi vardır.

1 uzunlukta 2 tane, dizgi vardır.

2 uzunlukta 4 tane, dizgi vardır.

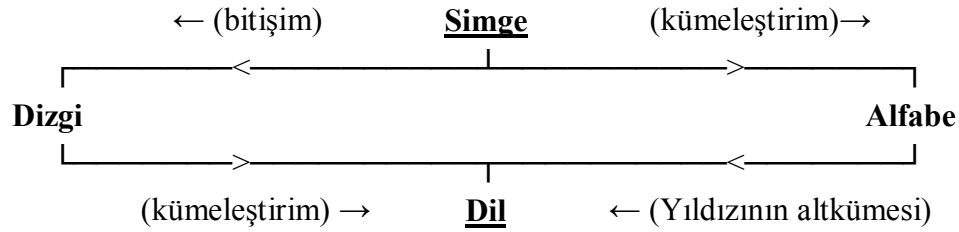
... vs.



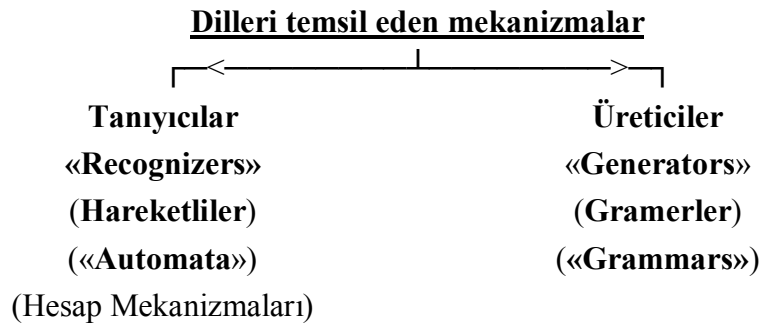
Herhangi bir dizgi, Σ^* -in bir elemanı ise onun uzunluğu bir doğal sayıdır.

Eğer $\alpha \in (V \cup T)^*$ ise, o şekilde sonlu bir dizgidir ki, içinde $(V \cup T)$ dışında hiçbir simge yoktur. ($\alpha = \varepsilon$ olabilir). Burada α -nın uzunluğu sıfır ise $\alpha = \varepsilon$ -dur. Uzunluğu 1 ise α ya V -nin bir elemanı ya da T -nin bir elemanıdır. Uzunluğu daha fazla ise o zaman da $(V \cup T)^*$ -in diğer bir elemanıdır.

L : Σ Alfabeti üzerinde bir dil demek, $L \subseteq \Sigma^*$ demektir.



Dilleri temsil için bazen küme gösterimi gibi jenerik ifadeler kullanılabilir; fakat esas üzerinde duracağımız başlıca iki türlü mekanizma vardır:



Tanıyıcılar: Belli bir dilin tanıyıcısı, o dildeki ve yalnız o dildeki dizgileri kabul eder.

Üreticiler: Dizgi üretir. Belli bir dilin üreticisi o dildeki dizgileri ve yalnız o dildeki dizgileri üretir.

Gösterim: Herhangi bir simge veya dizgi x için,

$$x^0 = \varepsilon$$

$$x^1 = x$$

$$x^2 = xx$$

$$x^{n+1} = x^n \cdot x$$

$$L_1 \cdot L_2 = \{u_1 \cdot u_2 \mid u_1 \in L_1, u_2 \in L_2\}$$

Örnek: $L_1 = \{aa, aab, ababb\}$, $L_2 = \{b, bb\}$

$$L_1 \cdot L_2 = \{aab, aabb, aabb, aabbb, ababb, ababbb, ababbb\}$$

Örnek: $L_1 = \{aa, aab, ababb\}$, $L_2 = \{\varepsilon, b\}$

$$L_1 \cdot L_2 = \{aa, aab, aab, aabb, ababb, ababbb\}$$

$$L_2 \cdot L_1 = \{aa, aab, ababb, baa, baab, bababb\}$$

Örnek: $L_1 = \{\varepsilon, aa, ababb\}$, $L_2 = \{aab, b\}$

$$L_1 \cdot L_2 = \{aab, b, aaaab, aab, ababbaab, ababbb\}$$

$$L_1 \cdot L_2 = \{aaaab, aab, ababbaab, ababbb, b\}$$

$$|L_1 \cdot L_2| \leq |L_1| \cdot |L_2|$$

Her hangi bir L dili için

$$L^0 = \{\epsilon\} \text{ (yani boş dizgi birliği)}$$

$$L^k = L^{k-1} \cdot L \text{ (k} \geq 1 \text{ için)}$$

$$L^1 = L^0 \cdot L = \{\epsilon\} \cdot L = \{\epsilon \cdot u \mid u \in L\} = \{u \mid u \in L\} = L$$

$$L^2 = L^1 \cdot L = L \cdot L$$

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$$

Σ : Her alfabe bir dildir.alfabenin her bir elemanı bir simgedir, $a \in \Sigma$, buradaki a 1 uzunluğunda olan bir dizgidir.

Örneğin: $\Sigma = \{a, b\}$

$$\Sigma \cdot \Sigma = \{aa, ab, ba, bb\}$$

$L \subseteq \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots \cup \Sigma^n$ n bir sabit olduğuna göre, $\Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots \cup \Sigma^n$, L sonludur.

$S \rightarrow A$ ($S, A \in V$) sonlandıran bir üretim değildir.

$$\textcircled{1} S \rightarrow w_1 \quad S \Rightarrow^* w_1$$

$$\textcircled{2} S \rightarrow w_2 \quad (S \text{ den içinde } A \text{ olan bir şey üretmek mümkün değil})$$

Örnek: $L_1 = \{a^n b^m \mid n, m \geq 1\}$ bu kümenin tipik elemanı: $aa \dots a \cdot bb \dots b$

$\underbrace{\hspace{1cm}} \quad \underbrace{\hspace{1cm}}$
 n adet m adet

Bu dili kabul eden bir tanıyıcı, aa geldiğinde **ret** durumunda, aab geldiğinde **kabul** durumunda, aabb geldiğinde yine **kabul** durumunda olur. Fakat aabba geldiğinde **ret** durumuna geçer ve bunun sağına ne gelse **kabule** geçmez.

Örnek: $L_2 = \{a^n b^n \mid n \geq 1\}$ bu kümenin tipik elemanı: $aa \dots a \cdot bb \dots b$

$\underbrace{\hspace{1cm}} \quad \underbrace{\hspace{1cm}}$
 n adet n adet

Bu dili kabul eden bir tanıyıcı, aa da gelse, aab de gelse **ret** durumunda olur. Eğer aabb gelirse **kabul** durumuna geçer. Fakat aabb -nin sağına daha fazla simge gelirse **ret** durumuna geçer.

ÜRETİCİLER

Gramerler

Tanım: Bir gramer, $G = (V, T, S, P)$ şeklinde bir sıralı 4-lü olarak verilir. Burada;

V: Değişken simgeleri kümesi «set of Variable symbols» (sonlu bir simge kümesi).

T: Uç simgeleri kümesi «set of Terminal symbols» (bir alfabe).

S: Başlangıç simgesi «Start symbol» (değişkenler kümesinde özel bir eleman $S \in V$) olmaktadır.

Değişken simgeleri ve Uç simgeleri yerine sırasıyla, kısaca **Değişkenler** ve **Uçlar** denmektedir.

$V \cap T = \emptyset$ (V ile T ayrık kümeler, ortak eleman yok) olup,

$V \cup T$ Gramerin bütün simgelerinden ibaret küme olmaktadır.

$(V \cup T)^*$, bu gramerin, evrensel dizgi kümesidir. (Bu gramerin, dizgi evreni) Buna göre,

P: Üretim kuralları kümesi «set of Production rules»; $(V \cup T)^*.V.(V \cup T)^* \times (V \cup T)^*$ çapraz çarpımının boş olmayan bir sonlu altkümesidir.

Kayıt: $(V \cup T)^*.V.(V \cup T)^* \subset (V \cup T)^*$ (altküme değil, özaltküme).

Yani, $(V \cup T)^* \subseteq (V \cup T)^*.V.(V \cup T)^*$ **değil!**

Üretim kuralları yerine kısaca **Üretimler** denmektedir.

Alıştırım: $(V \cup T)^*.V.(V \cup T)^* = (V \cup T)^* - T^*$ olduğunu ispatlayınız.

P -nin her bir elemanı, (α, β) şeklinde bir sıralı ikilidir;

$\alpha \in (V \cup T)^*.V.(V \cup T)^*$ yani, Gramerin simgelerinden ibaret olan, ve en az bir değişken bulunan bir dizgi.

$\beta \in (V \cup T)^*$ yani, Gramerin simgelerinden ibaret olan bir dizgi.

P -nin böyle bir ögesi (yani üretimi) $\alpha \rightarrow \beta$ şeklinde gösterilir. Bu takdirde α -ya bu üretimin “**sol tarafı**” (kısaca “**solu**”), β -ya da bu üretimin “**sağ tarafı**” (kısaca “**sağı**”) denir.

\rightarrow işareti: “**yeniden yazılır**” veya daha kısa olarak, “**yerine**” şeklinde okunur.

Örnek: $G_1 = (V_1, T, S_1, P_1)$ öyle ki, $V_1 = \{ S_1, A \}$, $T = \{ 0, 1 \}$, $P_1 = \{ S_1 \rightarrow 0AA1, 0A1 \rightarrow \varepsilon, A \rightarrow 10 \}$.

“Doğrudan sürer” ve “Sürer” bağıntıları:

Bir $G = (V, T, S, P)$ grameri verildiğinde, P -deki üretimlere dayalı olarak $(V \cup T)^*$ (bu gramerin simgelerinden ibaret dizgiler kümesi) üzerinde tanımlanan ve özel ismi “**G de doğrudan sürer**” olan bir temel bağıntı vardır. \Rightarrow_G işaretiyle gösterilen bu bağıntı şöyle tanımlanır:

$\alpha_0, \alpha_1, \beta_0, \beta_1 \in (V \cup T)^*$ «**alfa-0, alfa-1, beta-0, beta-1 ait V bileşim T -nin yıldızı**» için

$(\alpha_0\beta_0\alpha_1, \alpha_0\beta_1\alpha_1) \in \Rightarrow_G$ «**alfa-0, beta-0, alfa-1 virgül alfa-0, beta-1, alfa-1 sıralı ikilisi ait “G de doğrudan sürer” bağıntısı**» -dır, şayet $\beta_0 \rightarrow \beta_1 \in P$ ise.

Yani herhangi $\gamma_0, \gamma_1 \in (V \cup T)^*$ için $(\gamma_0, \gamma_1) \in \Rightarrow_G$ -dır, şayet γ_0 -ın içinde herhangi bir üretimin solu bir alt-dizgi olarak yer alır da, onun yerine o üretimin sağ yazıldığında γ_1 elde edilirse.

$(\alpha_0\beta_0\alpha_1, \alpha_0\beta_1\alpha_1) \in \Rightarrow_G$ yerine daha elverişli olan

$\alpha_0\beta_0\alpha_1 \Rightarrow_G \alpha_0\beta_1\alpha_1$ « **$\alpha_0\beta_0\alpha_1$ G de doğrudan sürer $\alpha_0\beta_1\alpha_1$** » tarzı kullanılır.

Yerel bağıntılarda genelde olduğu gibi, bir $k \geq 0$ için \Rightarrow_G bağıntısının k -ıncı kuvveti \Rightarrow_G^k olarak gösterilir ve şayet $\gamma_0, \gamma_k \in (V \cup T)^*$ için $\gamma_0 \Rightarrow_G^k \gamma_k$ ise, özelde (gramerlere mahsus olarak) “ γ_0 **G de k adımda sürer** γ_k ” denir.

Her yerel bağıntının olduğu gibi, \Rightarrow_G bağıntısının da, \Rightarrow_G^* olarak gösterilen bir **yansıyıcı ve geçişli kapatma** vardır. $(V \cup T)^*$ üzerindeki bu bağıntıya “**G de sürer**” denmektedir.

G anlaşıldığı takdirde açıkça belirtilmez; \Rightarrow_G yerine sadece \Rightarrow yazılarak «**doğrudan sürer**» veya \Rightarrow_G^k yerine sadece \Rightarrow^k yazılarak “**k adımda sürer**” veya \Rightarrow_G^* yerine sadece \Rightarrow^* yazılarak “**sürer**” denir.

Kayıt: Herhangi $\gamma_1, \gamma_2 \in (V \cup T)^*$ için $\gamma_1 \Rightarrow^* \gamma_2$ \Leftrightarrow bir $k \geq 0$ için $\gamma_1 \Rightarrow^k \gamma_2$ ise.

“**Sürer**” bağıntısı özyineleyişli olarak da tanımlanabilir:

γ_0 **sürer** γ_2 şayet,

$\gamma_2 = \gamma_0$ ise veya herhangi bir $\gamma_1 \in (V \cup T)^*$ için $\gamma_0 \Rightarrow^* \gamma_1$ ve $\gamma_1 \Rightarrow \gamma_2$ ise.

Kayıt: Herhangi $\gamma \in (V \cup T)^*$ için “ **γ sıfır adımda sürer γ** ” olmakla beraber, bir $k > 0$ için “ **γ k adımda sürer γ** ” da olabilir. Ancak ikinci hal, verilen gramere bağlıdır.

Kayıt: $\gamma \Rightarrow^* \gamma$ her zaman geçerlidir. Fakat $\gamma \Rightarrow \gamma$ her zaman geçerli değildir. Genelde, bir dizginin diğerini “**doğrudan sürer**” oluşu **sürer** oluşunu gerektirir. Ancak “**sürer**” oluşu, **doğrudan sürer** oluşunu gerektirmez. Örneğin, $A \in V$ olsun; $A \Rightarrow^0 A$ (A sıfır adımda sürer A) dolayısıyla, $A \Rightarrow^* A$ -dır. Fakat $A \Rightarrow A$ dolayısıyla $A \Rightarrow^1 A$ (A bir adımda sürer A) olur, $\Leftrightarrow A \rightarrow A \in P$ ise.

Bir Gramerin Ürettiği Dil:

Tanım: $G = (V, T, S, P)$ bir gramer olsun; herhangi $\alpha \in (V \cup T)^*$ bu gramer itibarıyla bir “**cümlesel kalıp**” «**sentential form**» olur, şayet $S \Rightarrow_G^* \alpha$ ise. G gramerindeki bütün cümlesel kalıplar kümesi:

$$C(G) = \{\alpha \in (V \cup T)^* \mid S \Rightarrow_G^* \alpha\}$$

Verilen bir $G = (V, T, S, P)$ gramerinin **ürettiği dil**, değişken simgesi bulunmayan cümlesel kalıplarının kümesidir:

$$L(G) = \{w \in T^* \mid S \Rightarrow_G^* w\}; L(G) = C(G) \cap T^*$$

Kayıt: Farklı isimlenenler hariç, mümkün olan en küçük gramer, $G_k = (\{S\}, \{1\}, S, P_k)$ öyle ki, bir $\alpha \in (V \cup T)^*$ için $P_k = \{S \rightarrow \alpha\}$ olmaktadır. Eğer bir $w \in T^*$ için $\alpha = w$ ise $L(G_k) = \{w\}$, aksi halde $L(G_k) = \{\}$; yani her halükârda $|L(G_k)| \leq 1$ olur.

Örnek: Yukarıda verilen $G_1 = (V_1, T, S_1, P_1)$ öyle ki, $V_1 = \{S_1, A\}$, $T = \{0, 1\}$,

P_1 : 1. $S_1 \rightarrow 0AA1$ 2. $0A1 \rightarrow \varepsilon$ 3. $A \rightarrow 10$ için,

“Doğrudan sürer” bağıntısında kullanılan üretimlerin sayılarını da alt-indis olarak gösterelim,

$S_1 \Rightarrow_1 0AA1$, $0AA1 \Rightarrow_3 010A1$, $010A1 \Rightarrow_3 010101$
$S_1 \Rightarrow_1 0AA1$, $0AA1 \Rightarrow_3 0A101$, $0A101 \Rightarrow_3 010101$
$S_1 \Rightarrow_1 0AA1$, $0AA1 \Rightarrow_3 010A1$, $010A1 \Rightarrow_2 01$
$S_1 \Rightarrow_1 0AA1$, $0AA1 \Rightarrow_3 0A101$, $0A101 \Rightarrow_2 01$

Demek ki, $L(G_1) = \{010101, 01\}$

Bazı özel şekillerdeki Üretimler:

Bir $G = (V, T, S, P)$ gramerinde, herhangi $u, v \in T^*$; $A, B \in V$ için,

$A \rightarrow u$ şeklindeki bir üretime **bitiren üretim «terminating production»** denir.

(Bazı hatalara yol açtığından, solu birden uzun hiçbir üretime bitiren üretim denmeyecek!)

$u = \varepsilon$ olduğu takdirde bu aynı zamanda bir $(A - y)$ “silen üretim” «erasing production» olur.

$A \rightarrow uB$ şeklindeki bir üretime **sağ-doğrusal üretim «right-linear production»**,

$A \rightarrow Bu$ şeklindeki bir üretime **sol-doğrusal üretim «left-linear production»**,

$A \rightarrow uBv$ şeklindeki bir üretime **doğrusal üretim «linear production»** denir.

$A \rightarrow B$ şeklindeki bir üretime “birim üretim” denir. Böyle bir üretim, hem **sağ-doğrusal**, hem **sol-doğrusal**, hem de **doğrusal** üretim tanımına uyar.

Gramerler, üretimlerine göre sınıflanırlar. Buna göre, bazı gramer sınıfları:

Bitiren Gramerler: Üretimleri bitiren üretimlerden ibaret gramerler. Bunların ürettiği diller, **sonlu dillerdir**.

Sağ-Doğrusal Gramerler: Üretimleri sağ doğrusal ve bitiren üretimlerden ibaret gramerler. Bunların ürettiği dillere, **Sağ-Doğrusal Diller** denir.

Sol-Doğrusal Gramerler: Üretimleri sol doğrusal ve bitiren üretimlerden ibaret gramerler. Bunların ürettiği dillere, **Sol-Doğrusal Diller** denir.

Doğrusal Gramerler: Üretimleri doğrusal ve bitiren üretimlerden ibaret gramerler. Bunların ürettiği dillere, **Doğrusal Diller** denir.

Aynı dili üreten iki gramere “**eşdeğer**” denir.

Bir Doğrusal Gramer	Eşdeğer bir Doğrusal Gramer
$S \rightarrow aSb$	$S \rightarrow aA$
$S \rightarrow \varepsilon$	$A \rightarrow Sb$
	$S \rightarrow \varepsilon$

Üretilen dil: $\{a^n b^n \mid n \geq 0\}$

Ayrıca, herhangi $A \in V$; $\gamma_0, \gamma_1, \gamma_2 \in (V \cup T)^*$ için,

$A \rightarrow \gamma_0$ şeklindeki bir üretime **bağlama-duyarsız üretim**,

$\gamma_1 \rightarrow \gamma_2$ şeklinde ve $|\gamma_1| \leq |\gamma_2|$ olan bir üretime **kısaltmayan üretim** denir.

Bunlara göre de,

Bağlama-Duyarsız Gramerler **⟨Context-Free Grammars⟩**: Üretimleri **bağlama-duyarsız** üretimlerden ibaret gramerlerdir. Bunların ürettiği dillere, **Bağlama-Duyarsız Diller** denir.

Bağlama-Duyarlı Gramerler **⟨Context-Sensitive Grammars⟩**: Üretimleri **kısaltmayan** üretimler ile başlangıç simgesini silen ($S \rightarrow \varepsilon$ şeklindeki) üretimden ibaret gramerlerdir. Şu var ki, başlangıç simgesini silen üretim bulunduğu takdirde, başlangıç simgesi hiçbir üretimin sağında bulunmamalıdır. Bunların ürettiği dillere, **Bağlama-Duyarlı Diller** denir.

Kayıt: Silen üretimler dışındaki her bağlama-duyarsız üretim, aynı zamanda bir kısaltmayan üretim olmaktadır.

Gramer sınıfı	Üretimlerin olabileceği şekiller	Üretilen diller sınıfı
Bitiren Gramerler	Bitiren	Sonlu diller
Sağ-Doğrusal Gramerler (Cins-3)	Bitiren, Sağ-doğrusal	Sağ-Doğrusal Diller (Cins-3)
Sol-Doğrusal Gramerler (Cins-3)	Bitiren, Sol-doğrusal	Sol-Doğrusal Diller (Cins-3)
Doğrusal Gramerler	Bitiren, Doğrusal	Doğrusal Diller
Bağlama-Duyarsız Gramerler (Cins-2)	Bağlama-duyarsız üretimler	Bağlama-Duyarsız Diller (Cins-2)
Bağlama-Duyarlı Gramerler (Cins-1)	Kısaltmayan , başlangıç simgesini silen. Başlangıç simgesini silen üretim varsa, başlangıç simgesi hiçbir üretimin sağında bulunmaz.	Bağlama-Duyarlı Diller (Cins-1)
Gramerler (Kısıtsız) (Cins-0)	Kısıtsız	Sonlu Tanımı Olan Diller (Cins-0) (Turing Tezinden)

Bir gramerin cinsi belirtilirken, onu barındıran sınıfın bu cetvele göre en üstte olanı söylenir.

Örnek: $G_2 = (V_2, T, S_2, P_2)$ öyle ki, $V_2 = \{ S_2, B, C \}$, $T = \{ a, b, c \}$.

<p>P_2:</p> <ol style="list-style-type: none"> $S_2 \rightarrow aS_2BC$ $S_2 \rightarrow aBC$ $CB \rightarrow BC$ $aB \rightarrow ab$ $bB \rightarrow bb$ $bC \rightarrow bc$ $cC \rightarrow cc$ 	<p>Bu bir bağlama-duyarlı gramerdir.</p> <p>$L(G_2) = \{ a^n b^n c^n \mid n \geq 1 \}$ -dir.</p> <p>$S_2 \Rightarrow^* aaabbbccc$ olduğu, sağ tarafta “\Rightarrow” bağıntısı cinsinden gösterilmektedir.</p>	<p>$S_2 \Rightarrow_1 aS_2BC$</p> <p>$aS_2BC \Rightarrow_1 aaS_2BCBC$</p> <p>$aaS_2BCBC \Rightarrow_2 aaaBCBCBC$</p> <p>$aaaBCBCBC \Rightarrow_3 aaaBBCCBC$</p> <p>$aaaBBCCBC \Rightarrow_3 aaaBBCBCC$</p> <p>$aaaBBCBCC \Rightarrow_3 aaaBBBCCC$</p> <p>$aaaBBBCCC \Rightarrow_4 aaabBBCCC$</p> <p>$aaabBBCCC \Rightarrow_5 aaabbBCCC$</p> <p>$aaabbBCCC \Rightarrow_5 aaabbbCCC$</p> <p>$aaabbbCCC \Rightarrow_6 aaabbbccC$</p> <p>$aaabbbccC \Rightarrow_7 aaabbbccc$</p>
<p>Üretim-3 -ün bütün C -ler en sağa gelene kadar uygulandığını, aksi halde cümlesel kalıpta aaabbbccBC dizisinde olduğu gibi c -lerin sağında B kalacağını, ve böyle bir cümlesel kalıbın doğrudan sürer olduğu hiçbir dizginin bulunmadığını, dolayısıyla $a^*b^*c^*$ dizilimine uymayan bir uçlar dizgisinin elde edilemeyeceğini kaydediniz.</p>		

$L(G_2) = \{ a^n b^n c^n \mid n \geq 1 \}$ olduğunun ispatı için resmi bir yaklaşım:

Herhangi $w \in \{ a^n b^n c^n \mid n \geq 1 \} \wedge w \in L(G_2)$ olduğu, yani

hem $\{ a^n b^n c^n \mid n \geq 1 \} \subseteq L(G_2)$ hem de $L(G_2) \subseteq \{ a^n b^n c^n \mid n \geq 1 \}$ olduğu gösterilir.

Bunu iki aşamada yapalım:

1- Bir dizgi $\{ a^n b^n c^n \mid n \geq 1 \}$ -de varsa $L(G_2)$ -de de vardır:

İspat: $\{ a^n b^n c^n \mid n \geq 1 \}$ -deki dizgilerden herhangi birini temsilen $a^n b^n c^n$ dizgisini alalım.

$$S_2 \Rightarrow_1^{n-1} a^{n-1} S_2 (BC)^{n-1} \Rightarrow_2 a^n (BC)^{n-1} \Rightarrow_3^* a^n B^n C^n$$

$$\Rightarrow_4 a^n b B^{n-1} C^n \Rightarrow_5^{n-1} a^n b^n C^n \Rightarrow_6 a^n b^n c C^{n-1} \Rightarrow_7^{n-1} a^n b^n c^n \in L(G_2) \quad \odot$$

2- Bir dizgi $L(G_2)$ -de varsa $\{ a^n b^n c^n \mid n \geq 1 \}$ -de de vardır; veya başka bir deyişle,

bir dizgi $\{ a^n b^n c^n \mid n \geq 1 \}$ -de yoksa $L(G_2)$ -de de yoktur.

İspat: G_2 -deki (4-7 no.lu olan) bitiren üretimleri incelediğimizde, B, C değişkenlerinin sırasıyla, b ve c uçları için yer tutmak üzere kullanıldığını görüyoruz. Başlangıçta kullanılmak durumunda olan 1 ve 2 no.lu üretimler Cümlesel kalıplarda a, B, C simgelerinin eşit sayılarda olmalarını sağlamaktadır. Dolayısıyla, bir $w \in L(G_2)$ ise, eşit sayıda a -lar, b -ler ve c -lerden oluşur. Böyle bir dizginin $\{ a^n b^n c^n \mid n \geq 1 \}$ içinde olmaması, ancak a -lar, b -ler ve c -lerin diziliminin $a^*b^*c^*$ dizilimine aykırı düşmesiyle mümkündür. G_2 -deki üretimleri bu açıdan incelediğimizde, 7 no.lu üretimin kullanılabilmesi için cümlesel kalıpta c çıkmış olması gerektiği, bunun için daha önce 6 no.lu üretimin kullanılmış olması gerektiği, bunun için de, daha önce b -lerin elde edilişine 4

no.lu üretimle başlanmış olması gerektiği anlaşılmaktadır. B, C değişkenlerinin uçlara dönüşebilmesinde soldan sağa bir sıra takibi zorlanmaktadır. Dolayısıyla bir C -nin sağında kalan bir B hiçbir zaman uç simgesine dönüşemez. ☺

Örnek: $G_3 = (V_3, T, S_3, P_3)$ öyle ki,

$$V_3 = \{ S_3, B \}, T = \{ a, b, c \},$$

- P₃:**
1. $S_3 \rightarrow aS_3Bc$
 2. $S_3 \rightarrow abc$
 3. $cB \rightarrow Bc$
 4. $bB \rightarrow bb$

$$S_3 \Rightarrow_1 aS_3Bc$$

$$aS_3Bc \Rightarrow_1 aaS_3BcBc$$

$$aaS_3BcBc \Rightarrow_2 aaabcBcBc$$

$$aaabcBcBc \Rightarrow_3 aaabBcBc$$

$$aaabBcBc \Rightarrow_3 aaabBcBc$$

$$aaabBcBc \Rightarrow_3 aaabBBccc$$

$$aaabBBccc \Rightarrow_4 aaabbbccc$$

$$aaabbbccc \Rightarrow_4 aaabbbccc$$

Alıştırım: G_3 eşdeğer G_2 yani $L(G_3) = \{a^n b^n c^n \mid n \geq 1\}$ olduğunu ispatlayınız.

Örnek: Önceki G_3 gramerinin üretimlerine bir “başlangıç simgesini silen” üretim ilâve ederek şöyle bir grameri elde edelim: $G_4 = (V_3, T, S_3, P_4)$, $P_4 = P_3 \cup \{S_3 \rightarrow \varepsilon\}$. Bu başlangıç simgesini silen üretimden ötürü, $\varepsilon \in L(G_4)$ olur. buradan iki soru çıkar:

1. G_4 ile G_3 -ün cinsleri aynı olur mu?
2. $L(G_4) = L(G_3) \cup \{\varepsilon\}$ olur mu?

CK: S_3, aS_3Bc, aBc (5. $S_3 \rightarrow \varepsilon$ olursa), aS_3Bc, \dots

Örnek: Yukarıdaki G_3 gramerinden şöyle bir gramer daha elde edelim:

$G_5 = (V_5, T, S_5, P_5)$. Öyle ki, $V_5 = V_3 \cup \{S_5\}$, $P_5 = P_3 \cup \{S_5 \rightarrow \varepsilon, S_5 \rightarrow S_3\}$. Bu takdirde,

1. $L(G_5) = L(G_3) \cup \{\varepsilon\}$ olur mu?
2. G_5 ile G_3 -ün cinsleri aynı mı olur?

Bir bağlama-duyarlı dilde ε bulunabilmesi için, Bağlama-duyarlı gramer tanımına uygun bir $G = (V, T, S, P)$ gramerinde, $S \rightarrow \varepsilon$ yer almasına izin verilir. Ancak, böyle bir üretim bulunduğu takdirde, hiçbir üretimin sağında o başlangıç simgesinin bulunmaması gerekir. ε için böyle bir kuralın olması, $\gamma_1, \gamma_2 \in (V \cup T)^*$ için, $\gamma_1 \Rightarrow^* \gamma_2$ olduğunda, şayet $\gamma_1 \neq S$ ve $\gamma_2 \neq \varepsilon$ ise $|\gamma_2| \geq |\gamma_1|$ olması içindir.

Aidiyet problemi... verilen w ile $L(G)$, problem: $w \in L(G)$?

$\gamma_1, \gamma_2 \in (V \cup T)^*$ için, $\gamma_1 \Rightarrow^* \gamma_2$ olduğunda, şayet $\gamma_1 \neq S$ ve $\gamma_2 \neq \varepsilon$ ise $|\gamma_2| \geq |\gamma_1|$

$S \Rightarrow^* \gamma; S \Rightarrow^* w$?

Lema: $G = (V, T, S, P)$ bir bağlama-duyarlı gramer olsun, $\gamma_1, \gamma_2 \in (V \cup T)^*$ için eğer $\gamma_2 \neq \varepsilon$ ve $\gamma_1 \Rightarrow^* \gamma_2$ ise $|\gamma_2| \geq |\gamma_1|$ olur.

Cins-0 Gramerler: Kısıtsız gramerler. Bütün gramerleri kapsayan sınıftır. Bunların ürettiği dillere, **Cins-0 Diller** denir. Böyle diller, sonlu bir şekilde tanımlanabilen bütün dillerdir.

Kısıtsız gramerler için bazen $P \subseteq (V \cup T)^* \times (V \cup T)^*$ olarak daha da **gevşetilebilir**. Yani “üretimin solunda hiç değişken olmasa da olur” denebilir. Ne var ki, üretilen diller sınıfında farklılık olmaz.

Açıklavis: Bazen $\langle isim \rangle$ şeklinde açılı tırnaklar içinde, görünümü anlam çağrıştıran simgeler oluşturulur. Böyle olduğunda örneğin $\langle isim \rangle$, altı simgelik bir dizgiyi değil, “ $\langle \rangle$ ” şekilciğinden “ \rangle ” şekilciğine kadar bir tek simgeyi temsil eder. O zaman, şu uzunluk ifadeleri geçerli olur: $|\langle isim \rangle| = 1$; $|\langle isim \rangle \langle isim \rangle| = 2$; $|\langle isim \rangle a \langle isim \rangle b| = 4$; $|\langle isim \rangle a \langle isim \rangle b| = 2$.

Örnek: Bir harf veya bir harfin sağına bitleştirilmiş harf-rakam -lardan oluşan isimler kümesini temsil eden bir gramer: $G_{is} = (V_{is}, T_{is}, \langle isim \rangle, P_{is})$

- P_{is} :
1. $\langle isim \rangle \rightarrow \langle harf \rangle$
 2. $\langle isim \rangle \rightarrow \langle isim \rangle \langle harf-rakam \rangle$
 3. $\langle harf-rakam \rangle \rightarrow \langle harf \rangle$
 4. $\langle harf-rakam \rangle \rightarrow \langle rakam \rangle$
 - 5.-8. $\langle harf \rangle \rightarrow a/b/c/d$ (“/” kullanımı, “veya” yerine)
 - 9.-18. $\langle rakam \rangle \rightarrow 0/1/.../9$
- $\rightarrow =$

[

$\langle harf \rangle \rightarrow a$

$\langle harf \rangle \rightarrow b$

$\langle harf \rangle \rightarrow c$

$\langle harf \rangle \rightarrow d$

(Üretimler kümesinin elemanlarına, işaret amacıyla sıra sayısı verilmiştir.)

Yukarıdaki gramerde;

$V_{is} = \{\langle isim \rangle, \langle harf \rangle, \langle harf-rakam \rangle, \langle rakam \rangle\}$: Değişkenler kümesidir.

$T_{is} = \{a, b, c, d, 0, 1, \dots, 9\}$: Alfabadir.

$\langle isim \rangle$: Başlangıç Simgesidir.

P_{is} : Üretimler kümesidir. (18 elemanlı.)

Bu gramerin ürettiği dildeki simge dizgileri, aşağıdaki gibi yöntemlerle elde edilirler.

Başlangıç
simgesi

$$\begin{aligned}
 \langle isim \rangle &\Rightarrow_2 \langle isim \rangle \langle harf-rakam \rangle \Rightarrow_2 \langle isim \rangle \langle harf-rakam \rangle \langle harf-rakam \rangle \\
 &\Rightarrow_1 \langle harf \rangle \langle harf-rakam \rangle \langle harf-rakam \rangle \Rightarrow_8 d \langle harf-rakam \rangle \langle harf-rakam \rangle \\
 &\Rightarrow_4 d \langle rakam \rangle \langle harf-rakam \rangle \Rightarrow_{12} d3 \langle harf-rakam \rangle \Rightarrow_3 d3 \langle harf \rangle \Rightarrow_7 d3c \in T_{is}^*
 \end{aligned}$$

Örnek: Dil örneklerinde sık kullanılan bir tanesi şudur: $\{a^n b^n \mid n \geq 1\}$.

Bunun alfabesindeki simgeler a ile b. Bu öyle bir dil ki bir dizgi, eğer ve ancak, soldan itibaren bir veya daha çok sayıda a ve onu takip eden aynı sayıda b -den oluşuyorsa onun elemanı.

Bunu üreten bir gramer: $(\{S\}, \{a, b\}, S, \{S \rightarrow aSb, S \rightarrow ab\})$

Üretimler: **1.** $S \rightarrow aSb$, **2.** $S \rightarrow ab$. Yukarıdaki dil, bu iki üretim kuralı ile tanımlanabilir

Bu üretim kurallarından birinci kural uygulanırsa aşağıdaki gibi eşit sayıda a -lar ve b -ler üretilir.

$$S \Rightarrow_1 aSb \Rightarrow_1 aaSbb \Rightarrow_1 \dots \Rightarrow_1 aa\dots aSbb\dots b \Rightarrow_2 aa\dots aabb\dots bb$$

Bir bitiren üretim olan ikinci kuralın uygulamasıyla dile ait dizgi elde edilir.

İkinci bir dil $L_2 = \{a^n b^{2n} \mid n \geq 1\}$

1. $S \rightarrow aSbb$, 2. $S \rightarrow abb$. Şeklinde bir gramerle her bir adet a -ya karşılık iki adet b üretilir.

DİL TEORİSİ NE İŞE YARAR?

1. Öncelikle resmi dillerin teorik temellerinin araştırılmasına ve keşfedilmesine yarar.
2. Programlama dilleriyle ilgili konulara özellikle derleyici yapımına, gerçekleştirilmesine temel teşkil eder.
3. Dil teorisindeki dil tanımına göre her bir dil bir boyutlu desenler kümesidir ve dil teorisyle bir boyutlu desenleri tanıma yöntemleri geliştirilir.
4. Resmi dilleri tanımlama yöntemlerinden bir tanesi gramerler vasıtasıyladır. Gramerler sonlu veya sonsuz bir takım dilleri tanımlamakta ve tanımakta kullanılan sonlu tanımlardır.
5. Gramerlerin uygulamalarından bir tanesi de biyolojik bazı olayların modellenmesidir. Çünkü sonlu bir tanıma sahip olan bir gramer vasıtasıyla bir büyüme silsilesini temsil eden sonsuz bir kümeyi tanımlamak mümkündür. Bu küçük bir çekirdeğin içerdiği formülleriyle bir ağacın büyümesini temsil etmesine benzer.
6. Gramer tanımlarıyla aksiyomatik mantık sistemi tanımları arasında benzerlik vardır. Dolayısıyla verilen herhangi bir simge dizgisinin bir dile ait olup olmadığı probleminin ilkeleriyle verilen herhangi bir cümlemin bir mantık sistemine göre teorem olup olmadığını belirleme yöntemleri geliştirilebilir.
7. Yukarıda sayılanların dışında dil teorisinin başka uygulamaları da vardır. Bunlar teorisinin içine girildikçe meydana çıkar.

DERLEYİCİLER :

Derleyicilerde cereyan eden safhaları üç ana başlık altına yerleştirebiliriz.

1. Lexical Analysis :

Token = Gramatic Symbol (variable)

Verilen izlencedeki bu token'ların belirlenmesi

Örnek : Identifier bir Token' dır.

Ör. `if` `hız` `=` `375` `then` (5 adet uç simgesi)...

Ör. `if` `hız` `=` `3A75` `then` (1hata, 4 adet uç simgesi)...

Ör. if hrz then = 375 (5 adet uç simgesi)...

Değişken anlamsal (semantik) bir ögedir, “Identifier” gramatik bir ögedir (sentaktik)

Anlam

yazım

Değişkenler için bir değişkenin kendisi, bir de ismi olduğu gibi sabitler için de bir sabitin kendisi (veya değeri) bir de sabitin yazılışı vardır.

Lexical Analysis safhası bittiği zaman izlenice sentaktik analize hazır hale gelir.

2. Syntactic Analysis:

Burada izlencenin ilgili dilin yazım kurallarına göre çözümlemesi yapılır ve verilen dizginin ilgili dile ait olup olmadığı belirlenir. Eğer ilgili dile aitse sentaktik yapısı ortaya çıkmış olur. (Yazımsal boyut)

3. Semantic Analysis (Anlamsal karşılığın bulunması safhası):

Sentaktik analizi geçen dizgide bulunan sentaktik yapılara tekabül eden semantik yapı oluşturulur. Bu semantik yapının makine dilindeki ifadesi amaç izlenceyi oluşturur. Bu suretle verilen bir izlenice, derleyici tarafından amaç koda dönüştürülmüş olur. (Anlamsal boyut)

Bir dilin sentaktik yapısı gramerle ifade edilebildiği gibi Lexical yapılar da daha basit düzeyde olmakla birlikte gramerlerle veya eşdeğer tanım yöntemleriyle tanımlanabilirler.

Böylece dil teorisiyle derleyiciler arasındaki bağlantı kurulmuş oldu.

Pratik açıdan sonsuz kümenin anlamı elemanlarının uzunluklarına bir kısıtlama getirilmemiş bir kümedir.

Ödev: $L(G) = \{a^n b^n \mid 100 > n \geq 1\}$ ise, $G = ?$

$S_1 \rightarrow ab/aabb; S_2 \rightarrow S_1/aaS_1bb/aa S_1bb$

$S_2 \rightarrow S_1/aaS_1bb/aa S_1bb$

$/aa S_1bb S_1/$

$S \rightarrow a^{10}S_1b^{10}$

$S_1 \rightarrow a^{10}S_2b^{10}$

$S_2 \rightarrow a^{10}S_3b^{10}$

$S_3 \rightarrow a^{10}S_4b^{10}$

$S_4 \rightarrow a^{10}b^{10}/$

Program x;

begin

begin end; begin end; ...

end.

Pascal dili sonsuz bir kümedir. Dizgiler sonsuz değildir. Pascal izlenceleri kümesi sonlu değildir.

Aksiyomatik Mantık Sistemi:

Burada bir aksiyomlar kümesi vardır. Bir akıl yürütüm kuralları kümesi vardır.

Aksiyomlara akıl yürütüm kuralları uygulanarak teoremler elde edilir. Daha sonra teorem ve aksiyomlara uygulanan akıl yürütüm kurallarıyla yeni teoremler elde edilir. Böylece akıl yürütüm kurallarıyla aksiyomlar bir teoremler kümesini tanımlamış olurlar. Buradaki teoremler kümesinin dil teorisindeki benzeri(karşılığı) dildir. Gramerler de bir başlangıç simgesi vardır.

Grammerlerden Örnekler

Örnek: Öyle bir bağlama duyarsız gramer bulunuz ki,

$L(G) = \{a^n(bc)^n \mid n \geq 0\} = \{\epsilon, abc, aabc bc, \dots\}$ (yani bu dilin elemanları o şekildedir ki soldan sağa olmak üzere önce bir dizi sıfır veya daha çok sayıda a -lar gelecek, sonra aynı sayıda bc -ler gelecek.

$n = 0$ için, $a^n(bc)^n = a^0.(bc)^0 = \epsilon.\epsilon = \epsilon$

$G = (\{S\}, \{a, b, c\}, S, P)$

P: ① $S \rightarrow a S bc$

② $S \rightarrow \epsilon$ görüldüğü gibi sadece iki tane üretim vardır.

$S \Rightarrow_{\textcircled{2}} \epsilon$ elde ediyoruz $\therefore S \Rightarrow^* \epsilon \therefore \epsilon \in L(G)$.

$S \Rightarrow_{\textcircled{1}} aSbc$ diyoruz (çünkü böyle bir üretim var)

$a S bc \Rightarrow_{\textcircled{2}} a \epsilon bc = abc$ (2 nolu üretime göre)

$\gamma_1 \quad \gamma_2 \quad \gamma_1 \quad \gamma_2$

① ① ②

$S \Rightarrow a S bc \Rightarrow aa S bc bc \Rightarrow aa \epsilon bc bc$

Tanım: Bir w dizgisinin “Aksi” w^A , özyinelemeli olarak şöyle tanımlanır:

$w^A = \begin{cases} \epsilon, & \text{şayet } w = \epsilon \text{ ise;} \\ u^A x, & \text{şayet herhangi bir } x \text{ simgesi ve } u \text{ dizgisi için } w = xu \text{ ise.} \end{cases}$

Örnek: $G: \textcircled{1} S \rightarrow a S a$

② $S \rightarrow b S b$

③ $S \rightarrow \epsilon$

$L(G) = \{\epsilon, aa, bb, abba, baab, aaaa, bbbb, \dots\} = \{u u^A \mid u \in \{a,b\}^*\} = \{u \in \{a,b\}^* \mid u = u^A \text{ ve } |u| \text{ çift}\}$

$u u^A = (u u^A)^A$

Bu dilin elemanları “Palindrom”lardır. Yani sağdan sola dizilişi ile soldan sağa dizilişi aynı olan dizgilerdir.

Örnek: $G: \textcircled{1} S \rightarrow a S a$

$\textcircled{2} S \rightarrow b S b$

$\textcircled{3} S \rightarrow a$

$\textcircled{4} S \rightarrow b$

$\textcircled{5} S \rightarrow \varepsilon$

Örnek: DYCK LANGUAGE (DENGELİ Parantezler DİLİ)

$G_{D1}: S \rightarrow SS$ (Burada alfabe sağ ve sol yuvarlak parantezden oluşur)

$S \rightarrow (S)$ $T = \{(,)\}$

$S \rightarrow ()$

Grameri bu üç üretimden oluşan dil acaba nasıl bir dil olur. Dengeli parantezlerden oluşan bütün dizgileri bu üç üretimle elde etmek mümkün mü?

$G_{D2}: S \rightarrow SS$ (Burada alfabe yuvarlak ve köşeli parantezlerden oluşur)

$S \rightarrow (S)$ $T = \{(,), [,]\}$

$S \rightarrow [S]$

$S \rightarrow ()$ İki ayrı cins parantezin dengelendiği bütün dizgiler amaçlanıyor?

$S \rightarrow []$

Örnek: $G = \{V, T, S, P\}$, $V = \{A, B, S\}$, $T = \{a, b\}$,

$P = \{ S \rightarrow Aba,$

$A \rightarrow BB,$

$B \rightarrow ab,$

~~$AB \rightarrow b$~~ $\}. L(G) = ?$

Örnek: $G = \{V, T, S, P\}$, $V = \{S\}$, $T = \{0, 1\}$,

$P = \{ S \rightarrow 11S,$

$S \rightarrow 0\}.$

$L(G) = ?$

– $L(G) = \{0, 110, 11110, 1111110, \dots\}$

Örnek: $G = \{V, T, S, P\}$

$V = \{A, B, C, S\}$, $T = \{a, b, c\}$,

$cbab \in L(G)$? veya $S \Rightarrow^* cbab$? sorusu!

Aidiyet problemi için iki yaklaşım: Üstten-aşağı ayrıştırım, Altan-yukarı ayrıştırım:

G: $S \rightarrow AB$

$A \rightarrow Ca$

$B \rightarrow Ba$

$B \rightarrow Cb$

$B \rightarrow b$

$C \rightarrow cb$

$C \rightarrow b$

$cbab \in L(G)?$

Üstten-aşağı ayrıştırım:

$S \Rightarrow AB$

$S \Rightarrow AB \Rightarrow CaB$

$S \Rightarrow AB \Rightarrow CaB \Rightarrow cbaB$

$S \Rightarrow AB \Rightarrow CaB \Rightarrow cbaB \Rightarrow cbab$

Altan-yukarı ayrıştırım:

$Cab \Rightarrow cbab$

$Ab \Rightarrow Cab \Rightarrow cbab$

$AB \Rightarrow Ab \Rightarrow Cab \Rightarrow cbab$

$S \Rightarrow AB \Rightarrow Ab \Rightarrow Cab \Rightarrow cbab$

Örnek:

$G_{S2} = (\{A, B, C, D, S\}, \{0, 1\}, S, P)$ ve	P:	1. $S \rightarrow C0A$	4. $A \rightarrow 0B$	7. $B \rightarrow 1C0$	10. $D \rightarrow 0C$
		2. $S \rightarrow 0C1$	5. $A \rightarrow 0$	8. $C \rightarrow 0CS$	11. $D \rightarrow 0CS$
		3. $S \rightarrow 0B1$	6. $B \rightarrow A1$	9. $C \rightarrow 0D$	olarak veriliyor

$S \Rightarrow 0B1 \Rightarrow 0A11 \Rightarrow 0011$

Soru:

$G_{S1} = (\{A, B, S\}, \{0, 1\}, S, P)$ ve	P:	1. $S \rightarrow 0A$	4. $A \rightarrow 0A$	7. $B \rightarrow 1B$	
		2. $S \rightarrow 1B$	5. $A \rightarrow 0S$	8. $B \rightarrow 1$	
		3. $S \rightarrow B0$	6. $A \rightarrow B1$	9. $B \rightarrow 0$	olarak veriliyor

Buna göre şu ifadeleri, altlarına “geçerli” veya “geçersiz” yazarak değerlendiriniz. (1 yanlış 1 doğruyu götürür.)

$S \Rightarrow^* \varepsilon$	$(S, 0S) \in \Rightarrow$	$(S, 0S) \in \Rightarrow^*$	$0S \Rightarrow^* 0S$	$0S \Rightarrow^* 00S$	$0S \Rightarrow 0S$	$0S \Rightarrow 00S$
geçersiz	geçersiz	geçersiz	geçerli	geçersiz	geçersiz	geçersiz

Halbuki...	$(S, 0A) \in \Rightarrow$	$(S, 0A) \in \Rightarrow^*$
	geçerli	geçerli

Alıştırım: $L(G) = \{a^n \mid 100 \geq n \geq 1\}$ ise, $G = ?$