§ 1. Generating languages with phrase-structured grammars

Consider the English sentence "Jane ate an apple." We can replace the noun "Jane" by any noun phrase, the verb "ate" by any appropriate phrase, the "an" by any other article, and the "apple" by any other noun phrase — and we will still get a grammatically correct sentence. For example, we can easily get "John quickly swallowed two grapes." Roughly speaking, a natural grammar such as English grammar consists of the following:

a) A set of phrase denoting variables — \( \langle \text{sent} \rangle, \langle \text{noun} \rangle, \langle \text{verb} \rangle \ldots \)
b) A set of terminal symbols — Jane, John, ate, an, ...
c) A starting variable — \( \langle \text{sent} \rangle \)
d) A set of rewriting rules — \( \langle \text{verb} \rangle \rightarrow \langle \text{adverb} \rangle \langle \text{verb} \rangle \)
   \( \langle \text{noun phrase} \rangle \rightarrow \langle \text{noun phrase} \rangle \langle \text{conj} \rangle \langle \text{noun} \rangle \)

Def. A phrase-structured grammar (PSG) is a 4-tuple \( G = \langle V, T, S, P \rangle \) where

\( V = \mathcal{V}(G) \) is a finite set of variables,
\( T = \mathcal{T}(G) \) is a finite set of terminal symbols,
\( S = \mathcal{S}(G) \) is a finite set of starting strings, and
\( P = \mathcal{P}(G) \) is a set of productions. (see def. below)

Def. A production is an ordered pair \( \langle \alpha A \beta, \alpha \gamma \beta \rangle \) where
A is a variable in \( V \) and \( \alpha, \beta, \gamma \) are strings from \((V \cup \Sigma)^*\).

We usually write the production \( \langle \alpha \alpha \beta, \alpha \gamma \beta \rangle \) as \( \alpha \beta \to \alpha \gamma \beta \) and interpret it as meaning that \( \alpha \beta \) can be replaced by \( \alpha \gamma \beta \) when it is in the context \( \alpha \) followed by \( \beta \). If \( \alpha \) and \( \beta \) are both \( \lambda \), then \( \alpha \beta = \lambda \) and the production \( A \to \gamma \) is said to be context-free. It usually consists of a single variable — just like a natural grammar like English grammar.

**Ex.** Let \( G = \langle V, T, \Sigma, \delta, \rho \rangle \), where
\[
V = \{ A \}, \quad T = \{ a, b \}, \quad \delta = \{ A \} \quad \text{and} \quad \rho = \{ A \to aAb, A \to \lambda \}.
\]
Then \( G \) is a PSG. The set of all strings of terminal symbols that can be "generated" by \( G \) will be denoted by \( L(G) \).

We will now specify what "generate" means and use it to precisely define \( L(G) \). We often call a phrase structured grammar just a grammar.
Let $G$ be a PSG, and $\alpha A \beta \Rightarrow \alpha \gamma \beta$ be a production of $G$. Also let $\omega_1 = \gamma \alpha A \beta \gamma \in (VT)^*$ and $\omega_2 = \gamma \alpha \gamma \beta \gamma \in (VT)^*$. We say that $\omega_2$ is immediately derivable from $\omega_1$ and write $\omega_1 \Rightarrow \omega_2$, to mean that $\omega_1$ can be transformed into $\omega_2$ in $G$. If $\omega_1, \omega_2, \ldots, \omega_n$ is a sequence of strings in $(VT)^*$ such that $\omega_1 \Rightarrow \omega_2$, $\omega_2 \Rightarrow \omega_3$, \ldots, $\omega_{n-1} \Rightarrow \omega_n$, then say that $\omega_n$ is derivable from $\omega_1$, and write $\omega_1 \Rightarrow^* \omega_n$, to mean that $\omega_1$ can be transformed into $\omega_n$ in $G$ in a finite number of steps. We also say that $\omega_1 \Rightarrow^* \omega_n$ for any PSG, $G$, in zero steps.

The sequence $(\omega_1, \omega_2, \ldots, \omega_n)$ is called a derivation of $\omega_n$ from $\omega_1$ in $G$. We often leave out the "$\Rightarrow$" from under the "$\Rightarrow^*$" when we are discussing only a single PSG.

Also, we often write a derivation $(\omega_1, \omega_2, \ldots, \omega_n)$ as $\omega_1 \Rightarrow \omega_2 \Rightarrow \cdots \Rightarrow \omega_n$.

**Def.** The language defined by a grammar $G$ (we often say "grammar" instead of PSG) is defined by

$L(G) = \{ \omega \in T^* : \omega \text{ is derivable from at least one string in } S(G) \}$

In other words, $L(G) =$ set of all terminal strings that are derivable from $S(G)$. We also say that $G$ generates the language $L(G)$. 
Notational conventions: In order to be able specify grammars more simply, we shall use the following conventions. We will denote

a) Variables by upper case Roman letters: A, B, C, ...

b) Terminal symbols by lower case Roman letters: 0, 1, a, b, s, ...

c) Strings from \((V \cup T)^*\) by \(u, v, x, w, \ldots\)

and strings from \(T^*\) by \(a, b, x, \ldots\).

We will also put an arrow \(\rightarrow\) in front of a string to indicate that it is a starting string in \(S(G)\).

Ex. 2 Let \(G_2\) be the grammar with starting strings & productions as follows:

\[
\rightarrow A, \rightarrow \lambda, \rightarrow ab, A \rightarrow aB, B \rightarrow bB, B \rightarrow ac, C \rightarrow b
\]

Then \(V(G_2) = \{A, B, C\}\), \(T(G_2) = \{a, b, 3\}\), \(S(G_2) = \{\lambda, ab, A\}\) and \(P(G) = \{A \rightarrow ab, B \rightarrow bB, B \rightarrow ac, C \rightarrow b\}\).

\[
\begin{array}{c|c}
A & A \rightarrow aB \\
\hline
ab & A \rightarrow aB \\
abB & B \rightarrow bB \\
ab^2B & B \rightarrow bB \\
ab^3B & B \rightarrow bB \\
\end{array}
\]

So \(L(G_2) = \{\lambda, ab, ab, ab^2, ab^3, \ldots\} = \{\lambda, ab^n : n \geq 0\}\).

We can write the answer in this case by using the regular expression:

\[
\lambda + ab + ab^*ab
\]

but most languages generated by PSGs cannot be described by regular expressions.
## § 2. Classification of Phrase-Structured Grammars

The phrase-structure grammars can be classified according to the type of productions that are allowed:

<table>
<thead>
<tr>
<th>TYPE</th>
<th>PRODUCTIONS ALLOWED</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSG</td>
<td>( \gamma AX \rightarrow \gamma \psi X ) with ( A \in V ), and ( \gamma, \psi, X \in (V \cup T)^* )</td>
</tr>
<tr>
<td>CSG</td>
<td>( \gamma AX \rightarrow \gamma \psi X ) with ( \gamma \neq \lambda ), and ( A \in V ) &amp; ( \gamma, \psi, X \in (V \cup T)^* )</td>
</tr>
<tr>
<td>CFG</td>
<td>( A \rightarrow \gamma ), ( A \in V ) &amp; ( \gamma \in (V \cup T)^* )</td>
</tr>
<tr>
<td>BLG</td>
<td>( A \rightarrow \alpha \beta ) or ( A \rightarrow \alpha ), with ( A \in B ), ( \alpha, \beta \in T^* )</td>
</tr>
<tr>
<td>RLG</td>
<td>( A \rightarrow \alpha B ) or ( A \rightarrow \alpha ), ( A, \beta \in V ) &amp; ( \alpha \in T^* )</td>
</tr>
<tr>
<td>LLG</td>
<td>( A \rightarrow \beta X ) or ( A \rightarrow \alpha ), ( A, \beta \in V ) &amp; ( \alpha \in T^* )</td>
</tr>
</tbody>
</table>

### Examples

1. \( \rightarrow A, A \rightarrow \alpha A, A \rightarrow 10B, B \rightarrow \lambda \)  
   \( \text{RLG} \)

2. \( \rightarrow B, B \rightarrow Ba, B \rightarrow Cbb, C \rightarrow a \)  
   \( \text{LLG} \)

3. \( \rightarrow C, C \rightarrow \alpha S, C \rightarrow \emptyset \)  
   \( \text{BLG} \)

4. \( \rightarrow D, D \rightarrow \alpha D, D \rightarrow 6b, D \rightarrow \lambda \)  
   \( \text{BLG} \)

5. \( \rightarrow A, A \rightarrow AB, B \rightarrow bB, B \rightarrow aC, C \rightarrow \lambda \)  
   \( \text{CFG} \)

6. \( \rightarrow B, B \rightarrow bC, bC \rightarrow bb, C \rightarrow \emptyset, D \rightarrow \alpha \)  
   \( \text{CSG} \)

7. \( \rightarrow C, C \rightarrow bCD, bC \rightarrow bb, C \rightarrow D, D \rightarrow \lambda \)  
   \( \text{PSG} \)

**Def**  
A language \( L \), is said to be phrase-structured if we can find a phrase-structured grammar (PSG) \( G \), such that \( L(G) = L \).
We define context-sensitive, context-free, bilinear, right-linear and left-linear languages in a similar ways by using CSG, CFG, BLG, RLG and LLG instead of PSG. This leads to a strict hierarchy of classes of languages.

Let $V = \{a, b, c\}$. Then $L(V)$ is as shown below:

- $FNL = \text{class of finite languages}$
- $RGL = \text{class of regular languages}$
- $BLL = \text{class of bi-linear languages}$
The languages $L_i (i=0,1,2,\ldots,6)$ show that the classes are all non-empty and that the hierarchy is strict. Unfortunately we do not have the machinery to describe $L_5$ and $L_6$ as yet. Below we will specify $L_0, \ldots, L_4$ and find grammars of the appropriate type for them.

$L_0 = \{a, ab\}$

$G_0 : \rightarrow A, A \rightarrow a, A \rightarrow ab.$

$L_1 = \{a^n : n > 1\}$

$G_1 : \rightarrow B, B \rightarrow aB, B \rightarrow a$.

$L_2 = \{a^n b^n : n > 1\}$

$G_2 : \rightarrow C, C \rightarrow aCb, C \rightarrow ab.$

$L_3 = \{a^n b^n a^k b^k : n > 1, k > 1\}$

$G_3 : \rightarrow A, A \rightarrow BC, B \rightarrow aBb, C \rightarrow aCb, B \rightarrow ab,$

$L_4 = \{a^n b^n c^n : n > 1\}$. The CSG for $L_4$ is a little bit complicated.

$G_4 : \rightarrow A, A \rightarrow abC, A \rightarrow aABC, bB \rightarrow bb, bC \rightarrow bc,$

$cC \rightarrow cc, CB \rightarrow Xb, XB \rightarrow XY, XY \rightarrow BY, BY \rightarrow BC$,

-equiv to $CB \rightarrow BC$ but this is not a legal production.

Two derivations in $G_4$:

$\rightarrow A \Rightarrow abC \Rightarrow abc$

$\rightarrow A \Rightarrow aABC \Rightarrow aabCBC \Rightarrow aabXBC \Rightarrow aabXYC \Rightarrow aabBYC$

$\Rightarrow aabbCC \Rightarrow aabbCC \Rightarrow aabBC \Rightarrow aabCC \Rightarrow aabBCC \Rightarrow aabBC$.
We will now find grammars of various types which generate certain specific languages.

Ex. 1a) Find an RLG which generates the language
\[ L_1 = \{ a^k (ba)^k : k \geq 0 \} \]

\[ G_a : \rightarrow A, \quad A \rightarrow aA, \quad A \rightarrow B, \quad B \rightarrow baB, \quad B \rightarrow ab \]

Ex. 1b) Find an LLG which generates the language
\[ L_1 = \{ a^k (ba)^k : k \geq 0 \} \]

\[ G_b : \rightarrow B, \quad B \rightarrow Cab, \quad C \rightarrow C ba, \quad C \rightarrow Dc, \]
\[ D \rightarrow aD, \quad D \rightarrow \lambda \]

Ex. 2 Find a CFG which generates the language
\[ L_2 = \{ a^n b^k : k \geq 2n, \ n \geq 0 \} \]

\[ G_2 : \rightarrow C, \quad C \rightarrow aCbb, \quad C \rightarrow D, \]
\[ D \rightarrow Db, \quad D \rightarrow \lambda \]

Ex. 3 Find a CFG which generates

\[ L_3 = \{ a^n b^k : n \geq 0, \ 0 < k \leq 2n \} \]

\[ G_3 : \rightarrow D, \quad D \rightarrow aDBB, \quad D \rightarrow \lambda, \]
\[ B \rightarrow b, \quad B \rightarrow \lambda, \]

Note: \[ G_2' : \rightarrow C, \quad C \rightarrow aCbb, \quad C \rightarrow bC, \quad C \rightarrow \lambda \] does not generate \( L_2 \),
but \[ G_2'' : \rightarrow C, \quad C \rightarrow aCbb, \quad C \rightarrow Cb, \quad C \rightarrow \lambda \] generates \( L_2 \).
§ 3. Parsing & Ambiguity in CFGs

PSGs are the most general grammars that we consider. CSGs are almost as complicated as PSGs but they are needed to understand the grammars of natural languages. For the rest of the chapter we shall only consider CFGs because they are a lot simpler and are exactly the kind of grammars that we need in Computer Science for designing Programming languages. (RLGs are very simple CFGs.)

Q: Suppose G is a CFG and y is a terminal string. How can we tell if y ∈ L(G)? And if y ∈ L(G), how can we find if there is essentially more than one derivation of y in G?

A: If G is of a CFG of special format, then there are parsing algorithms that can tell us whether or not y ∈ L(G). The parsing algorithms is basically a breadth-first search algorithm and after a certain point, if we do not find a derivation of y, then it means there is none. But the CFG has to be of this special format; otherwise the algorithm won’t work. (Parsing is the process of understanding how a terminal string was generated.) So we will now indicate what these special formats are. They are called normal forms of CFGs.

Def: Two grammars G₁ & G₂ are equivalent if L(G₁) = L(G₂).
Def. Let $G$ be a CFG. An unreachable production of $G$ is one which involves a variable that cannot be reached from the starting string of $G$.

(To be more specific, $B \rightarrow \emptyset$ is unreachable if none of the strings that are derivable in $G$ includes the variable $B$.)

Ex. 1 $G_1: \rightarrow A, A \rightarrow aC, B \rightarrow bB, C \rightarrow Cb, C \rightarrow \lambda, B \rightarrow a$

In $G_1$, $B \rightarrow bB$ and $B \rightarrow a$ are unreachable prod.

An equivalent CFG $G'_1$ can be obtained by deleting these two unreachable productions

$G'_1: \rightarrow A, A \rightarrow aC, C \rightarrow Cb, C \rightarrow \lambda$

Def. A non-terminating production is one which involves a variable which does not eventually terminate. (Again to be more specific, $C \rightarrow \emptyset$ is non-terminating if there is no terminal string that can be derived by starting with $C \rightarrow \emptyset$.)

Ex. 2 $G_2: \rightarrow B, B \rightarrow aC, D \rightarrow bD, B \rightarrow E6, E \rightarrow aE, E \rightarrow \lambda$

Here $B \rightarrow aC$ and $D \rightarrow bD$ are non-terminating productions. ($D \rightarrow bD$ is also unreachable.)

An equiv. CFG $G'_2$ is as shown below:

$G'_2: \rightarrow B, B \rightarrow E6, E \rightarrow aE, E \rightarrow \lambda$

Def. A useless production in a CFG is one that is unreachable or one that is non-terminating.
Def. A unit production is one of the form $B \rightarrow C$
where $B$ and $C$ are variables.

Ex.3 $G_3: \rightarrow A, A \rightarrow ab, B \rightarrow C, C \rightarrowCb, C \rightarrow \lambda$
Here $B \rightarrow C$ is a unit production. We can
eliminate $B \rightarrow C$ to get an equivalent CFG
$G'_3: \rightarrow A, A \rightarrow ab, B \rightarrow Cb, B \rightarrow \lambda, C \rightarrowCb, C \rightarrow \lambda$

Def. A \( \lambda \)-production is one of the form $B \rightarrow \lambda$
where $B$ is a variable.

Ex.4 $G_4: \rightarrow A, A \rightarrow bA, A \rightarrow ab, B \rightarrow Bab, B \rightarrow \lambda$
Here $B \rightarrow \lambda$ is a \( \lambda \)-production. We can
eliminate $B \rightarrow \lambda$ to get an equivalent CFG
$G'_4: \rightarrow A, A \rightarrow bA, A \rightarrow ab, A \rightarrow a, B \rightarrow Bab, B \rightarrow ab$

Def. A terminal production in a CFG is a production
of the form $B \rightarrow \alpha$ where $\alpha \in \Sigma^*$
A CFG $G$ is increasing if whenever $A \rightarrow \psi \in G$,
then $|\psi| > 1$. It is strictly increasing if
it is increasing and $A \rightarrow \psi \in G$ implies $|\psi| > 2$.

The parsing algorithms will work for any strictly increasing
CFG. We can also prove the following.

Normal Form Theorem for CFG: Any CFG $G$ is equivalent to
one in which $\Delta(G) = \{ A, \lambda \}$ and all productions
are of the form $B \rightarrow CD$ or $B \rightarrow a$ with $\psi \in \Sigma^*$
and $a \in \Sigma$. 

Consider the grammar $G$ below.

$$
G : \rightarrow A, \quad A \rightarrow BC, \quad B \rightarrow aBb, \quad B \rightarrow \lambda, \quad C \rightarrow aCb, \quad C \rightarrow \lambda
$$

Let $\psi = abab$. Then

(a) $A \rightarrow BC \rightarrow aBbC \rightarrow abC \rightarrow abab$ and $A \rightarrow BC \rightarrow BaCb \rightarrow Bab \rightarrow aBbab \rightarrow abab$

are both derivations of $\psi$ in $G$.

Although (a) & (b) look different, they are essentially the same derivation. (a) is called a left-most derivation because the left-most variable is replaced at each step. (b) is called a right-most derivation for a similar reason. We can even have other derivations of $\psi$ such as:

(c) $A \rightarrow BC \rightarrow B aCb \rightarrow aBb aCb \rightarrow aBbab \rightarrow abab$

which is neither left-most nor right-most.

We can see that all three derivations of $\psi$ are essentially the same. By drawing the derivation tree of each of the derivations (a), (b) & (c),

For each of the derivations (a), (b) & (c) we will end up with the same tree on the right.

The string derived is $a \lambda b a \lambda b = abab$. 

Consider the sentence, "They are flying planes." This sentence can be parsed in two different ways and each way has a different meaning.

They are flying planes

So we say that the sentence is ambiguous.

Def: A CFG $G$ is ambiguous if it generates a string $y$ which has at least two non-identical (different) derivation trees. (Equivalently, we can say that $G$ is ambiguous if it generates a string $y$ which has at least 2 leftmost derivations.) A CFG is unambiguous if it is not ambiguous.

Ex. 1: $G: \rightarrow S, S \rightarrow AB, A \rightarrow a, B \rightarrow BC, B \rightarrow BCD, B \rightarrow b, C \rightarrow CD, C \rightarrow a, D \rightarrow c$

(a) $S$

(b) $S$

So $G$ is an ambiguous CFG.
We can also see this by writing down the equivalent leftmost derivations.

(a) \[ S \Rightarrow AB \Rightarrow aB \Rightarrow aBCD \Rightarrow abCD \Rightarrow abaD \Rightarrow abc \]
(b) \[ S \Rightarrow AB \Rightarrow aB \Rightarrow aBC \Rightarrow abC \Rightarrow abCD \Rightarrow abaD \Rightarrow abc \]

Now it would be nice if we can find an unambiguous CFG which is equivalent to any ambiguous CFG, but unfortunately, this is not so.

Ex. 2 Let \[ L_2 = \{a^n b^n c^k : n, k \geq 0\} \cup \{a^n b^k c^k : n, k \geq 0\} \]
Then the CFG \( G_2 \) below will generate \( L_2 \).

\( G_2 : \)
- \[ A \Rightarrow B, \quad B \Rightarrow Be, \quad B \Rightarrow C, \quad C \Rightarrow aCb, \quad C \Rightarrow \] 
- \[ A \Rightarrow D, \quad D \Rightarrow aD, \quad D \Rightarrow E, \quad E \Rightarrow bEc, \quad E \Rightarrow \] 

Now \( G_2 \) is ambiguous because \( abc \) has two different (non-identical) derivation trees.

(a) \[ A \Rightarrow B \Rightarrow Bc \Rightarrow Cc \Rightarrow aCb \Rightarrow a bc \]
(b) \[ A \Rightarrow D \Rightarrow aD \Rightarrow aE \Rightarrow abEc \Rightarrow abc \]

But it can be shown that there is no unambiguous CFG which can generated \( L_2 \). In other words, every CFG which generates \( L_2 \) is ambiguous.

Def. A context-free language (CFL) \( L \) is inherently ambiguous if there is no unambiguous CFG which can generate \( L \).
It can be shown that any ambiguous RLG $G$ is equivalent to an unambiguous RLG $G'$ and we will see this in Ch.4.

We cannot say what DCFL are at this stage because it depends on theory of machines known as DPDA.

§4. Other Properties of CFG's and RLG's

It is easy to show that any CFG is equivalent to one in which the set of starting strings consist of just one variable. From now on we shall assume this.

Given CFGs $G_1$ and $G_2$, how can we find CFGs which will generate

(a) $L(G_1) \cup L(G_2)$?
(b) $L(G_1) \cdot L(G_2)$?
(c) $L(G)^*$?

Let $G_1 = \langle V_1, T, \{S\}, \rho \rangle$ & $G_2 = \langle V_2, T, \{S_2\}, \rho \rangle$.

First we must rename the variables in $V_2$ to get $V_2'$ to ensure that $V_1 \cap V_2' = \emptyset$. Also let $\rho'$ be $\rho_2$ with the renamed variables and $S$ be a new variable which is not already in $V_1 \cup V_2'$.
(a) \( G_a = \langle V, uv, \{S\}, T, \{S\}, \{S\}, \{S\}, \phi \rangle \) where
\[ \phi = \{ S \to S_1, S \to S_2 \} \cup \phi \cup \phi'. \]
Then \( L(G_a) = L(G_1) \cup L(G_2). \)

(b) \( G_b = \langle V, uv, \{S\}, T, \{S\}, \{S\}, \{S\}, \phi \rangle \)
where \( \phi = \{ S \to S_1, S \to S_2 \} \cup \phi \cup \phi'. \)
Then \( L(G_b) = L(G_1), L(G_2). \)

(c) \( G_c = \langle V, uv, \{S\}, T, \{S\}, \{S\}, \{S\}, \phi \rangle \)
where \( \phi = \{ S \to \lambda, S \to SS_1 \} \cup \phi \).
Then \( L(G_c) = L(G_1)^* \).

Example:
\( G_1: S \to aA, A \to bA, A \to aA, B \to bB, B \to \lambda \)
\( G_2: S \to bA, A \to aA, A \to bC, C \to bA, C \to \lambda \)

Then \( V_1 = \{ S, A, B \} \) & \( V_2 = \{ S_2, A, C \} \). We rename the \( A \) in \( V_2 \) as \( A' \) to get \( V_2' = \{ S_2, A', C \} \).

(a) \( G_a: S \to S_1, S_1 \to aA, A \to bA, A \to aA, B \to bB, B \to \lambda, S \to S_2, S_2 \to bA', A \to aA', A' \to bC, C \to bA, C \to \lambda \).
Then \( L(G_a) = L(G_1) \cup L(G_2). \)

(b) \( G_b: S \to S_1, S_2 \to bA', A \to aA', A' \to bC, C \to bA, C \to \lambda \).
Then \( L(G_b) = L(G_1), L(G_2). \)

(c) \( G_c: S \to \lambda, S \to SS_1, S_1 \to aA, A \to bA, A \to aA, B \to bB, B \to \lambda \).
Then \( L(G_c) = L(G)^* \).

Question:
Given RLGs \( G_1 \) & \( G_2 \), how can we find RLGs which will generate
(a) \( L(G_1) \cup L(G_2) \)?
(b) \( L(G_1), L(G_2) \)?
(c) \( L(G)^* \)?
This is the same question as the previous one except that CFGs are replaced by RLGs. The solution to part (a) is the same because $S \to S_1 \& S \to S_2$ are allowed in a RLG — but the solutions to parts (b) & (c) have to be changed because $S \to SS_2 \& S \to SS_1$ are not allowed in a RLG.

2(6) Put $G_b = \langle \{v, u, s, t, \}, \{\}, \{s, t\}, P \rangle$ where $P = \{S \to S_1 \cup P''\}$ Here $P''$ consists of the same productions as $P$, except that all the terminal productions will have an $S_2$ added to their right-hand side. (If $A \to a$ or $A \to \lambda$ is in $P$, then $A \to aS_2$ & $A \to \lambda S_2 \in P''$.)

(c) Put $G_c = \langle \{v, u, s, t\}, \{\}, \{s, t\}, P \rangle$ where $P = \{S \to \lambda, S \to S_1 \cup P'''\}$ Here $P'''$ consists of the same productions as $P$, except all terminal productions will have $S$ added to their right-hand side.

**Ex. 2** $G_1 : S_1 \to aA, A \to cA, A \to aAB, B \to ba, B \to \lambda$
$G_2 : S_2 \to bA, A \to abA, A \to bC, C \to b, C \to \lambda$

**RLG for (a)** $G_a : S \to S_1, S_1 \to aA, A \to cA, A \to aAB, B \to ba, B \to \lambda$,
$L(G_1) \cup L(G_2)$ $S \to S_2, S_2 \to bA', A \to abA', A' \to bC, C \to b, C \to \lambda$.

**RLG for (b)** $G_b : S \to S_1, S_1 \to aA, A \to cA, A \to aAB, B \to baS_2, B \to S_2$,
$L(G_1) \cup L(G_2)$ $S_2 \to bA', A \to abA', A' \to bC, C \to b, C \to \lambda$.

**RLG for (c)** $G_c : S \to \lambda \& S_1 \to aA, A \to cA, A \to aAB, B \to baS, B \to S$,
$L(G_1)$ $S \to S_1$. Then $L(G_c) = L(G_1)^*$.

Notice how we had to change the variable $A$ in $G_2$ into $A'$ to avoid mixing it up with the $A$ from $G_1$.

END OF CH. 2