Teoría de la Computación

Clase 18: El lema de bombeo para `CFLs`

Mauro Artigiani

23 septiembre 2021

Universidad del Rosario, Bogotá

El lema de bombeo para `CFLs`

Un ejemplo de lenguaje que no sea un CFL

Hemos visto que el lenguaje

$$L = \{0^n 1^n, n \geq 0\},$$

es un lenguaje que no es regular, pero es independiente del contexto.

Un ejemplo de lenguaje que no sea un CFL

Hemos visto que el lenguaje

$$L = \{0^n 1^n, n \geq 0\},$$

es un lenguaje que no es regular, pero es independiente del contexto.

De hecho, podemos construir un autómata con pila que lo reconozca, aprovechando de la pila para tener en memoria los 0s y después compararlos con los 1s.

Un ejemplo de lenguaje que no sea un CFL

Hemos visto que el lenguaje

$$L = \{0^n 1^n, n \geq 0\},$$

es un lenguaje que no es regular, pero es independiente del contexto.

De hecho, podemos construir un autómata con pila que lo reconozca, aprovechando de la pila para tener en memoria los 0s y después compararlos con los 1s.

Consideramos ahora el lenguaje

$$B = \{0^n 1^n 2^n, n \geq 0\}.$$

Un ejemplo de lenguaje que no sea un CFL

Hemos visto que el lenguaje

$$L = \{0^n 1^n, n \geq 0\},$$

es un lenguaje que no es regular, pero es independiente del contexto.

De hecho, podemos construir un autómata con pila que lo reconozca, aprovechando de la pila para tener en memoria los 0s y después compararlos con los 1s.

Consideramos ahora el lenguaje

$$B = \{0^n 1^n 2^n, n \geq 0\}.$$

En este caso, la pila no es suficiente: si borramos los 0s para compararlos con los 1s después no tenemos memoria de cuantos eran para comparalos con los 2s.

El lema de bombeo

Como en el caso de los lenguajes regulares, necesitamos de una herramienta teórica que nos permita demostrar, de manera rigurosa lo que nuestra intuición nos dice: que B no sea un lenguaje independiente del contexto.

El lema de bombeo

Como en el caso de los lenguajes regulares, necesitamos de una herramienta teórica que nos permita demostrar, de manera rigurosa lo que nuestra intuición nos dice: que B no sea un lenguaje independiente del contexto.

Necesitamos de un lema de bombeo para lenguajes independientes del contexto.

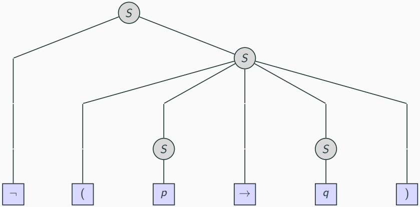
El lema de bombeo

Como en el caso de los lenguajes regulares, necesitamos de una herramienta teórica que nos permita demostrar, de manera rigurosa lo que nuestra intuición nos dice: que B no sea un lenguaje independiente del contexto.

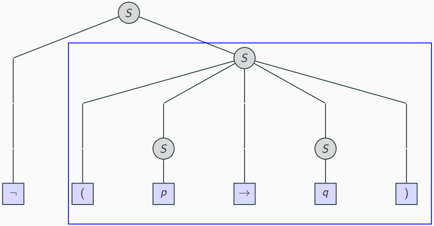
Necesitamos de un lema de bombeo para lenguajes independientes del contexto.

Esta vez el lema será un poco más complejo, pero la idea será la misma.

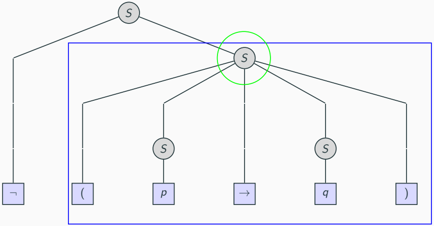
Ejemplo de bombeo



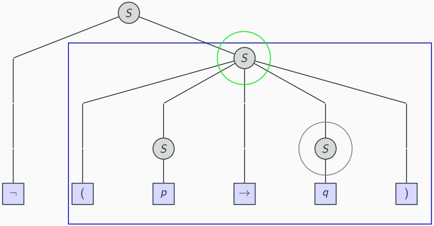
Ejemplo de bombeo



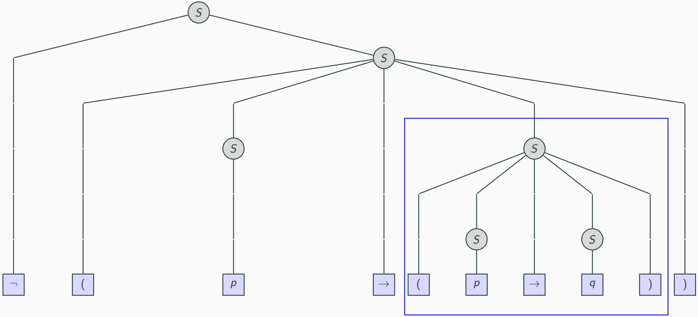
Ejemplo de bombeo



Ejemplo de bombeo



Ejemplo de bombeo



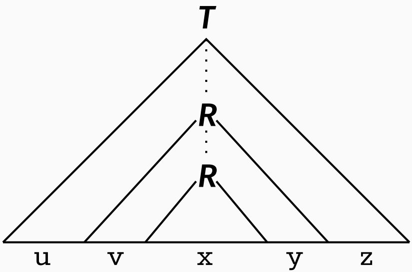
La idea atrás del lema de bombeo

Si s es una palabra de un `CFL` A, necesariamente existe un árbol de parsing que la produce. Si la palabra s es suficientemente larga, el árbol también será bastante alto.

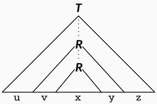
La idea atrás del lema de bombeo

Si s es una palabra de un CFL A, necesariamente existe un árbol de parsing que la produce. Si la palabra s es suficientemente larga, el árbol también será bastante alto.

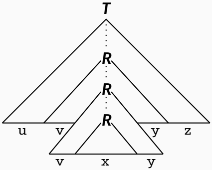
En particular, por el *principio del palomar*, si el árbol de parsing es suficientemente alto, una variable será repetida.

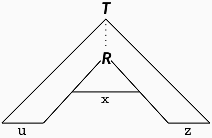


La idea atrás del lema de bombeo

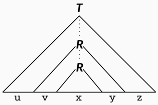


Podemos remplazar el sub-árbol bajo la segunda ocurrencia de la variable R con el sub-árbol bajo la primera ocurrencia de R misma, obteniendo todavía un árbol de parsing valido el CFL.

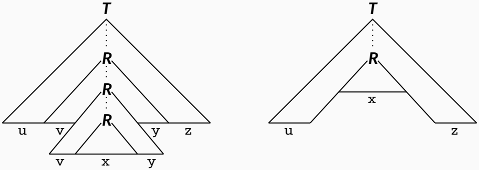




La idea atrás del lema de bombeo



Podemos remplazar el sub-árbol bajo la segunda ocurrencia de la variable R con el sub-árbol bajo la primera ocurrencia de R misma, obteniendo todavía un árbol de parsing valido el CFL.



También podemos cortar el sub-árbol bajo la segunda ocurrencia de la variable R, obteniendo todavía un árbol de parsing valido el CFL.

El lema de bombeo

Lema de bombeo para `CFLs`

Sea A un lenguaje independiente del contexto. Luego existe un número p (la *longitud de bombeo*), tal que, si s es una palabra en A de longitud por lo menos p, entonces s se puede dividir en 5 pedazos: $s = uvxyz$, tales que

1. para todos $i \geq 0$, $uv^ixy^iz \in A$;

El lema de bombeo

Lema de bombeo para `CFLs`

Sea A un lenguaje independiente del contexto. Luego existe un número p (la *longitud de bombeo*), tal que, si s es una palabra en A de longitud por lo menos p, entonces s se puede dividir en 5 pedazos: $s = uvxyz$, tales que

1. para todos $i \geq 0$, $uv^ixy^iz \in A$;
2. $|vy| > 0$;

El lema de bombeo

Lema de bombeo para `CFLs`

Sea A un lenguaje independiente del contexto. Luego existe un número p (la *longitud de bombeo*), tal que, si s es una palabra en A de longitud por lo menos p, entonces s se puede dividir en 5 pedazos: $s = uvxyz$, tales que

1. para todos $i \geq 0$, $uv^ixy^iz \in A$;
2. $|vy| > 0$;
3. $|vxy| \leq p$.

El lema de bombeo

Lema de bombeo para `CFLs`

Sea A un lenguaje independiente del contexto. Luego existe un número p (la *longitud de bombeo*), tal que, si s es una palabra en A de longitud por lo menos p, entonces s se puede dividir en 5 pedazos: $s = uvxyz$, tales que

1. para todos $i \geq 0$, $uv^ixy^iz \in A$;
2. $|vy| > 0$;
3. $|vxy| \leq p$.

La segunda condición nos asegura que por lo menos una entre v y y no sea la palabra vacía y que entonces las palabras del primer punto sean de verdad palabras distintas.

El lema de bombeo

Lema de bombeo para CFLs

Sea A un lenguaje independiente del contexto. Luego existe un número p (la *longitud de bombeo*), tal que, si s es una palabra en A de longitud por lo menos p, entonces s se puede dividir en 5 pedazos: $s = uvxyz$, tales que

1. para todos $i \geq 0$, $uv^ixy^iz \in A$;
2. $|vy| > 0$;
3. $|vxy| \leq p$.

La segunda condición nos asegura que por lo menos una entre v y y no sea la palabra vacía y que entonces las palabras del primer punto sean de verdad palabras distintas. La última condición es técnica y nos ayudará cuando utilizaremos el lema de bombeo.

Demostración del Lema de bombeo

Sea G una gramática independiente del contexto para el CFL A.

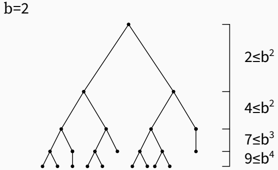
Demostración del Lema de bombeo

Sea G una gramática independiente del contexto para el $\texttt{CFL}$ A. Llamemos b el máximo número de símbolos (variables y terminales) en la derecha de las reglas de G.

Demostración del Lema de bombeo

Sea G una gramática independiente del contexto para el `CFL` A. Llamemos b el máximo número de símbolos (variables y terminales) en la derecha de las reglas de G.

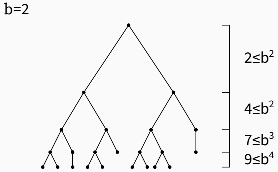
Esto nos asegura que cada variable tendrá abajo de ella, en cualquier árbol de parsing, a lo máximo b hijos. En particular, la variable inicial tendrá máximo b hijos.



Demostración del Lema de bombeo

Sea G una gramática independiente del contexto para el CFL A. Llamemos b el máximo número de símbolos (variables y terminales) en la derecha de las reglas de G.

Esto nos asegura que cada variable tendrá abajo de ella, en cualquier árbol de parsing, a lo máximo b hijos. En particular, la variable inicial tendrá máximo b hijos.



Además tendremos máximo b^2 hojas a distancia 2 de la variable inicial. En general, al nivel h de cualquier árbol de parsing estarán máximo b^h hojas.

Demostración del Lema de bombeo

Esto nos dice que, si un árbol de parsing tiene altura menor de h, entonces la palabra tiene longitud menor que b^h:

$$\text{Altura}(\tau) \leq h \iff |w| \leq b^h$$

Demostración del Lema de bombeo

Esto nos dice que, si un árbol de parsing tiene altura menor de h, entonces la palabra tiene longitud menor que b^h:

$$\text{Altura}(\tau) \leq h \iff |w| \leq b^h$$

Equivalentemente, si una palabra tiene $b^h + 1$ o más letras, necesariamente su árbol de parsing será por lo menos alto $h + 1$:

$$\text{Altura}(\tau) > h \iff |w| > b^h$$

Demostración del Lema de bombeo

Supongamos que G tenga $|V|$ variables y definamos la longitud de bombeo $p = b^{|V|+1}$.

Demostración del Lema de bombeo

Supongamos que G tenga $|V|$ variables y definamos la longitud de bombeo $p = b^{|V|+1}$.

Tomamos una palabra $s \in A$ con longitud $|s| \geq p$. Sea τ un árbol de parsing para s. Si existe más de un árbol, elegimos el que tiene el menor número de vertices.

Demostración del Lema de bombeo

Supongamos que G tenga $|V|$ variables y definamos la longitud de bombeo $p = b^{|V|+1}$.

Tomamos una palabra $s \in A$ con longitud $|s| \geq p$. Sea τ un árbol de parsing para s. Si existe más de un árbol, elegimos el que tiene el menor número de vertices.

Dado que $|s| \geq b^{|V|+1} \geq b^{|V|} + 1$, τ tiene altura por lo menos $|V| + 1$.

Demostración del Lema de bombeo

Supongamos que G tenga $|V|$ variables y definamos la longitud de bombeo $p = b^{|V|+1}$.

Tomamos una palabra $s \in A$ con longitud $|s| \geq p$. Sea τ un árbol de parsing para s. Si existe más de un árbol, elegimos el que tiene el menor número de vertices.

Dado que $|s| \geq b^{|V|+1} \geq b^{|V|} + 1$, τ tiene altura por lo menos $|V| + 1$. Esto implica que el camino más largo en el árbol tiene longitud $|V| + 1$ o más, que está compuesto por $|V| + 2$ vertices o más, todos variables menos el último (que es un terminal).

Demostración del Lema de bombeo

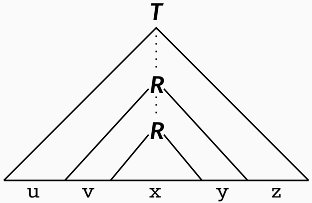
Dado que G tiene $|V|$ variables y nuestro camino tiene por lo menos $|V| + 1$ vértices que son variables, por el *principio del palomar* hay por lo menos una repetición.

Demostración del Lema de bombeo

Dado que G tiene $|V|$ variables y nuestro camino tiene por lo menos $|V| + 1$ vértices que son variables, por el *principio del palomar* hay por lo menos una repetición. Llamemos R la variable que se repite entre las últimas $|V| + 1$ variables en el camino.

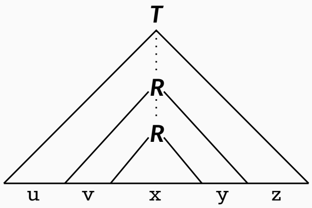
Demostración del Lema de bombeo

Dado que G tiene $|V|$ variables y nuestro camino tiene por lo menos $|V|+1$ vértices que son variables, por el *principio del palomar* hay por lo menos una repetición. Llamemos R la variable que se repite entre las últimas $|V|+1$ variables en el camino.



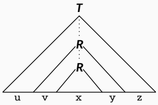
Demostración del Lema de bombeo

Dado que G tiene $|V|$ variables y nuestro camino tiene por lo menos $|V| + 1$ vértices que son variables, por el *principio del palomar* hay por lo menos una repetición. Llamemos R la variable que se repite entre las últimas $|V| + 1$ variables en el camino.

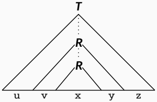


De la primera R empieza un sub-árbol que genera vxy. De la segunda, un sub-árbol más pequeño que genera x.

Demostración del Lema de bombeo

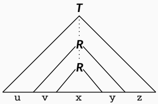


Demostración del Lema de bombeo

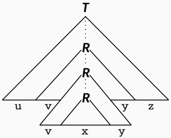


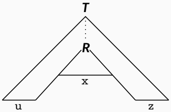
Si remplazamos el sub-árbol más pequeño por el sub-árbol más grande obtenemos todas las palabras de la forma uv^ixy^iz, con $i > 1$.

Demostración del Lema de bombeo

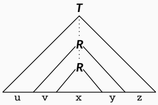


Si remplazamos el sub-árbol más pequeño por el sub-árbol más grande obtenemos todas las palabras de la forma uv^ixy^iz, con $i > 1$. Si remplazamos el sub-árbol más grande con el más pequeño, obtenemos uxz.

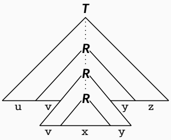


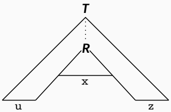


Demostración del Lema de bombeo



Si remplazamos el sub-árbol más pequeño por el sub-árbol más grande obtenemos todas las palabras de la forma uv^ixy^iz, con $i > 1$. Si remplazamos el sub-árbol más grande con el más pequeño, obtenemos uxz.





Esto nos dice que todas la palabras $uv^ixy^iz \in A$, por $i \geq 0$. Hemos demostrado la primera condición del lema de bombeo para `CFLs`.

Demostración del Lema de bombeo

Vamos a ver que la descomposición de s en $uvxyz$ satisface los otros dos requerimientos del Lema.

Demostración del Lema de bombeo

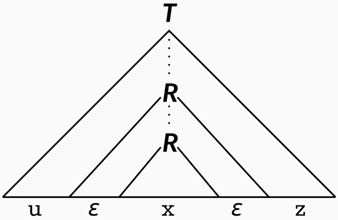
Vamos a ver que la descomposición de s en $uvxyz$ satisface los otros dos requerimientos del Lema.

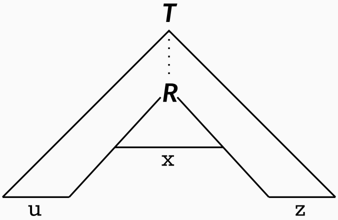
Si ambas v y y fueran la palabra vacía, el árbol de parsing obtenido remplazando el sub-árbol más pequeño en lugar del más grande sería un árbol de parsing para la palabra $uxz = uvxyz = s$ con menos vértices de τ, lo que sería una contradicción con nuestra elección de τ mismo.

Demostración del Lema de bombeo

Vamos a ver que la descomposición de s en $uvxyz$ satisface los otros dos requerimientos del Lema.

Si ambas v y y fueran la palabra vacía, el árbol de parsing obtenido remplazando el sub-árbol más pequeño en lugar del más grande sería un árbol de parsing para la palabra $uxz = uvxyz = s$ con menos vértices de τ, lo que sería una contradicción con nuestra elección de τ mismo.





Demostración del Lema de bombeo

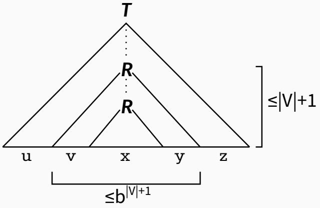
Finalmente, la penúltima ocurrencia de R en τ genera la palabra vxy. Sabemos que la última y la penúltima ocurrencia de R son en las últimas $|V| + 1$ variables del camino (que es el más largo en τ).

Demostración del Lema de bombeo

Finalmente, la penúltima ocurrencia de R en τ genera la palabra vxy. Sabemos que la última y la penúltima ocurrencia de R son en las últimas $|V|+1$ variables del camino (que es el más largo en τ). Entonces el sub-árbol bajo la primera ocurrencia de R tiene altura menor o igual a $|V|+1$.

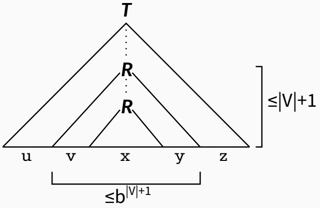
Demostración del Lema de bombeo

Finalmente, la penúltima ocurrencia de R en τ genera la palabra vxy. Sabemos que la última y la penúltima ocurrencia de R son en las últimas $|V|+1$ variables del camino (que es el más largo en τ). Entonces el sub-árbol bajo la primera ocurrencia de R tiene altura menor o igual a $|V|+1$. Por definición de b, este sub-árbol genera palabras de longitud menor o igual a $b^{|V|+1} = p$.



Demostración del Lema de bombeo

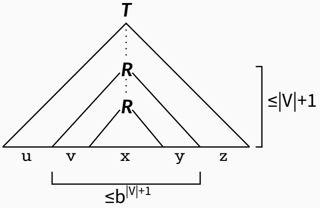
Finalmente, la penúltima ocurrencia de R en τ genera la palabra vxy. Sabemos que la última y la penúltima ocurrencia de R son en las últimas $|V|+1$ variables del camino (que es el más largo en τ). Entonces el sub-árbol bajo la primera ocurrencia de R tiene altura menor o igual a $|V|+1$. Por definición de b, este sub-árbol genera palabras de longitud menor o igual a $b^{|V|+1} = p$.



Entonces $|vxy| \leq p$.

Demostración del Lema de bombeo

Finalmente, la penúltima ocurrencia de R en τ genera la palabra vxy. Sabemos que la última y la penúltima ocurrencia de R son en las últimas $|V|+1$ variables del camino (que es el más largo en τ). Entonces el sub-árbol bajo la primera ocurrencia de R tiene altura menor o igual a $|V|+1$. Por definición de b, este sub-árbol genera palabras de longitud menor o igual a $b^{|V|+1} = p$.



Entonces $|vxy| \leq p$. Esto demuestra la tercera condición del lema de bombeo para `CFLs` y acaba la demostración.

Resumen

Resumen

Hoy aprendimos:

- Qué no todos los lenguajes son independientes del contexto;
- Una herramienta teórica para demostrar que un lenguaje no es independiente del contexto.