Forma normale di Chomsky Ogni linguaggio context free senza $\varepsilon - produzioni$ $(A \rightarrow \varepsilon)$ è generato da una grammatica dove tutte le produzioni sono della forma

$$A \rightarrow BC$$
$$A \rightarrow a$$

dimostrazione

Costruiamo un algoritmo in grado di generare questa nuova grammatica. Prendiamo una grammatica $G = \langle V, T, P, S \rangle$ senza ε-produzioni, simboli inutili e produzioni unitarie, ed eseguiamo il seguente algoritmo:

while new $P \neq$ old P
do $\forall A \rightarrow X_1 \dots X_m \in P$ **case**:

- $m = 1$ **and** $X_1 \in T$ (perché non ci sono produzioni unitarie): **continue**
- $m = 1$ **and** $X_1 \in V$: **impossibile**
- $m \geq 2 \wedge X_i \in T$: $V = V \cup B_i$, $P = (P \setminus \{A \rightarrow X_1 \dots X_m\}) \cup \{A \rightarrow Y_1 \dots Y_m, B_i \rightarrow X_i\}$ dove $Y_i = X_i$ se $X_i \in V$ altrimenti $Y_i = B_i$ se $X_i \in T$
- $m \geq 2 \wedge X_i \in V$: $P = (P \setminus \{A \rightarrow X_1 \dots X_m\}) \cup B \rightarrow X_1 X_2, A \rightarrow B X_3 \dots X_m$

end while

Pumping lemma per grammatiche context free

Sia L un linguaggio context-free. Allora $\exists n \in \mathbb{N}$ dove $\forall z \in \mathfrak{L}$ con $|z| \geq n$ esistono $u, v, w, x, y \in T^*$ tali che:

- $z = uvwxy$
- $|vx| \geq 1$
- $|vwx| \leq n$
- $\forall i \in \mathbb{N}. uv^i wx^i y \in L$

dimostrazione

Sia $G =< V, T, P, S >$ una grammatica in forma normale di Chomsky che genera $L \setminus \varepsilon$. Possiamo assumere questo senza perdita di generalità in quanto è sempre possibile costruire una grammatica G' con un simbolo $S' \rightarrow S|\varepsilon$.
Per induzione su $i \geq 1$ si dimostra che se un albero di derivazione per una stringa $z \in T^\star$ ha tutti i cammini di lunghezza minore o uguale ad i allora $|z| \leq 2^{i-1}$
Sia $|V| = k$ $(k > 0)$ e sia $n = 2^k$. Se $z \in L$ e $|z| \geq n$ allora ogni albero di derivazione per z deve avere un cammino lungo almeno $k+1$. Pertanto, all'interno di questo cammino deve esistere un simbolo non terminale $A \in V$ che si ripete due volte. Siano allora v_1 e v_2 due vertici tali che:

- entrambi corrispondono allo stesso simbolo non terminale A
- v_1 è più vicino alla radice di v_2
- poiché $n \geq 2$ a v_1 è associata una produzione del tipo $A \rightarrow BC$
- da v_1 alla foglia la lunghezza è al più $k+1$

Sia z_1 la stringa lunga al più 2^k generata dal sottoalbero T_1 con radice in v_1. Sia w la stringa generata dal sottoalbero T_2 con radice in v_2. Possiamo scrivere $z_1 = vwx$. Notiamo che almeno una delle due stringa z o x deve essere non vuota in quanto a v_1 è associata una produzione del tipo $A \rightarrow BC$ e T_2 corrisponde quindi a B o C.

Allora abbiamo che:

$$A \to^* vAx \text{ e } A \to^* w$$

con $|vwx| \leq 2^k = n$. Quindi possiamo applicare i volte la prima regola e otteniamo:

$$A \to^* v^i w x^i$$

per ogni $i \geq 0$. Ponendo u e y pari alle stringhe tali che $z = uvwxy$ abbiamo dimostrato il pumping lemma.