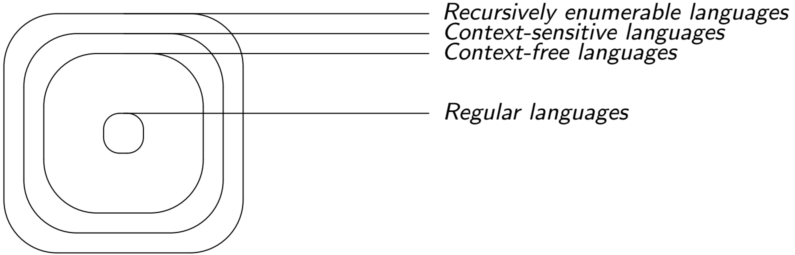
New pumping lemmas for linear and nonlinear context-free languages

Géza Horváth

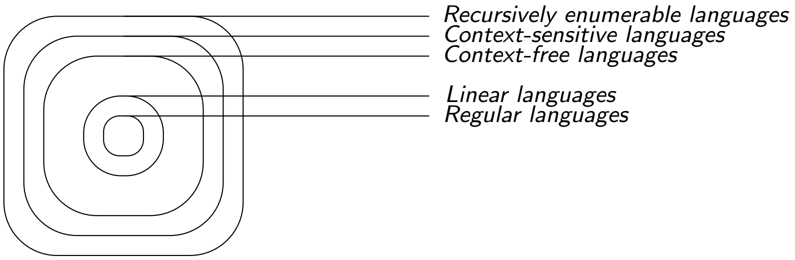
Brno, 29 March 2011.

Chomsky hierarchy



The Generative (formal) Grammar is an universal tool for creating languages.
The Chomsky Hierarchy is a containment hierarchy of classes of formal grammars.

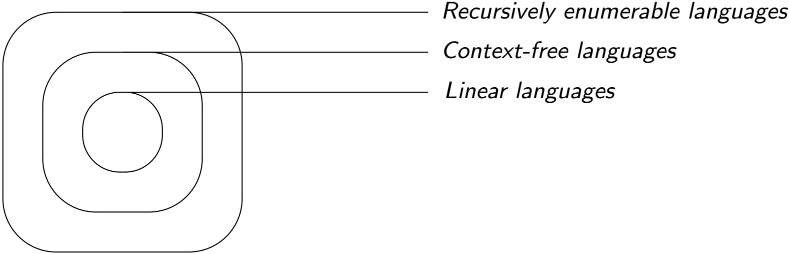
Linear languages



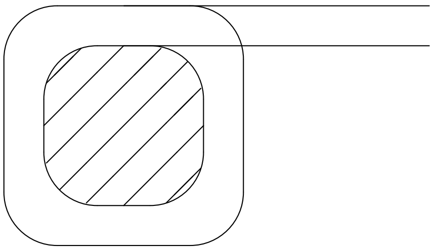
Definition

Linear languages can be generated by grammars have rules of the form $P \rightarrow a$ and $P \rightarrow aRb$, where $P, R \in V_N$, $a, b \in {V_T}^*$.

The hierarchy what we use



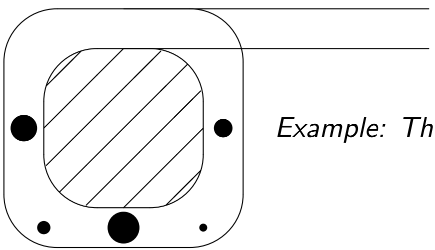
The Bar-Hillel lemma



Lemma

If a language L is context-free and infinite, then there exists integers n, m, such that any string $p \in L$, $|p| > n$ can be written as $p = uvwxy$, where $|vwx| \leq m$, $|vx| > 0$ and $uv^iwx^iy \in L$ for every integer $i \geq 0$.

The Bar-Hillel lemma

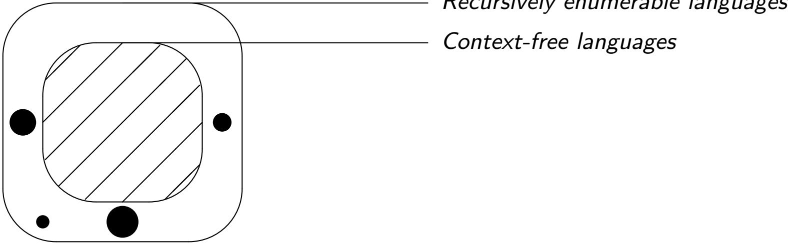


Example: The $L = (a^i b^i c^i \mid i \geq 0)$ *language is not context-free.*

Lemma

If a language L is context-free and infinite, then there exists integers n, m*, such that any string* $p \in L$, $|p| > n$ *can be written as* $p = uvwxy$*, where* $|vwx| \leq m$, $|vx| > 0$ *and* $uv^i wx^i y \in L$ *for every integer* $i \geq 0$.

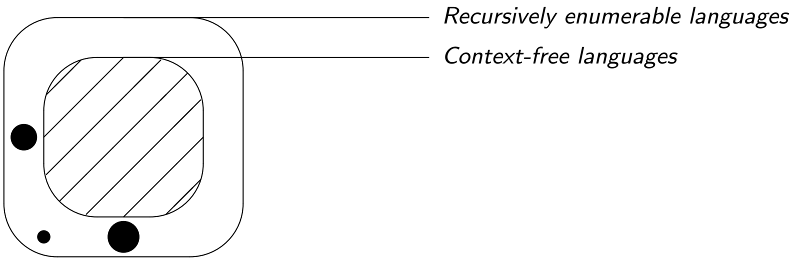
The Ogden Lemma



Lemma

All context-free language satisfies the Ogden restriction.

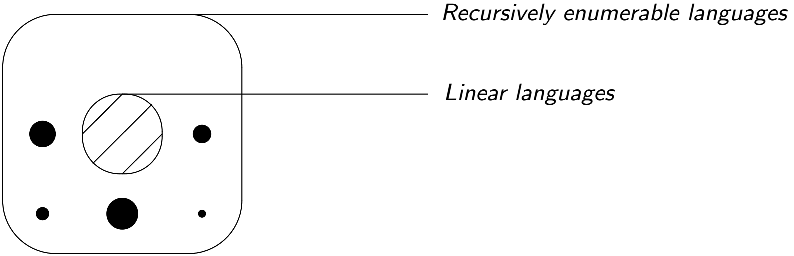
The strong Bader-Moura lemma



Lemma

All context-free language satisfies the strong Bader-Moura restriction.

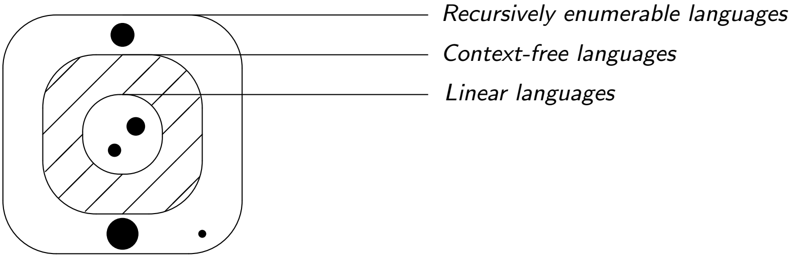
Pumping lemma for linear languages



Lemma

If a language L is linear and infinite, then there exists integer n such than any string $|p| > n$ *can be wtritten as* $p = uvwxy$*, where* $|uvxy| \leq n,\ |vx| > 0$ *and* $uv^iwx^iy \in L$ *for every integer* $i \geq 0$.

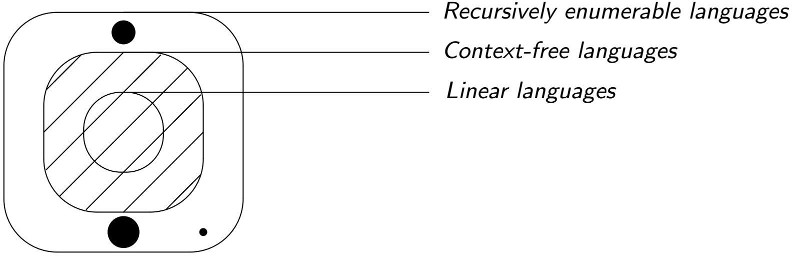
Pumping lemma for non-linear context-free languages



Lemma

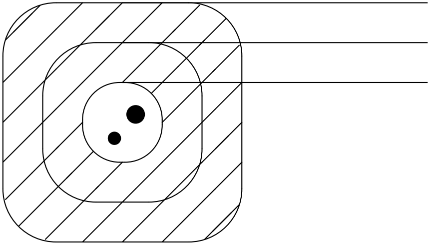
If the language L is non-linear, context-free and infinite, then there exists infinite many string $p \in L$ such that p can be written as $p = rstuvwxyz$, where $|su| > 0$, $|wy| > 0$ and $rs^i tu^i vw^j xy^j z \in L$ for every integers $i, j \geq 0$.

Classic application



Let $H \subseteq \{1^2, 2^2, 3^2, \ldots\}$ infinie set, and let
$L_H = \{a^k b^k a^l b^l \mid k, l \geq 1;\ k \in H \ vagy\ l \in H\} \cup \{a^m b^m \mid m \geq 1\}$.
1. The language L_H satisfies the Bar-Hillel condition.
2a. The language L_H does not satisfy the conditions of the pumping lemma for linear languages. $\Rightarrow$ L_H non-linear.
2b. The language L_H does not satisfy the conditions of the new pumping lemma. $\Rightarrow$ L_H not context-free.

New application

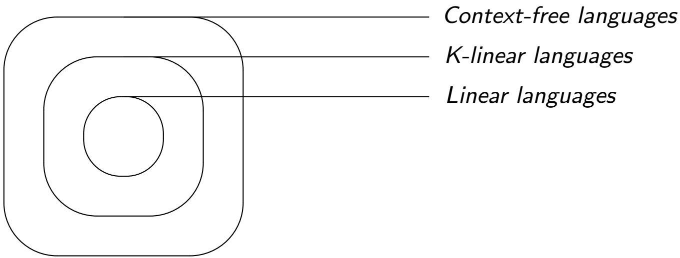


Let $L = \{a^i b^i b^i \mid i \geq 0\}$.

1. We know that the language L is context-free, because the context-free grammar
$G = (\{S, B\}, \{a, b\}, S, \{S \to aSB, S \to \lambda, B \to bb\})$ generates L.
2. The language L does not satisfy the conditions of the new pumping lemma.

$\Rightarrow$ L is linear.

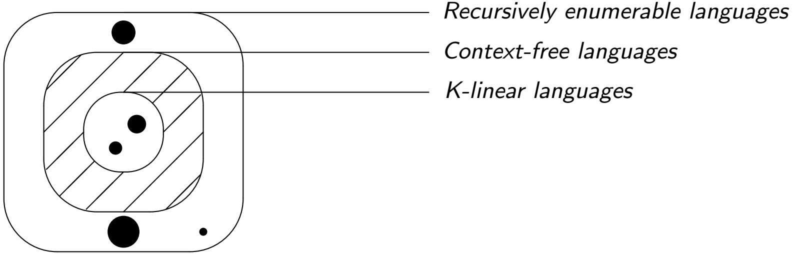
K-linear languages



Definition

A context-free grammar $G = (V_N, V_T, S, P)$ is said to be a k-linear grammar if it has the form of a linear grammar plus one additional rule of the form $S \rightarrow S_1 S_2 ... S_k$, where none of the S_i may appear on the right-hand side of any other rule and S may not appear in any other rule at all.

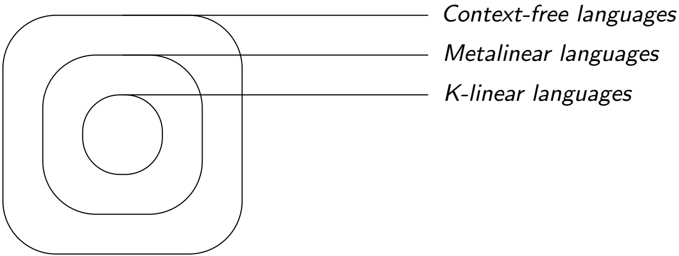
Pumping lemma for not k-linear context-free languages



Theorem

Given a context-free language L which does not belong to any k-linear language for a given positive integer k. There exist infinite many words $w \in L$ *which admit a factorization* $w = uv_0w_0x_0y_0 \ldots v_kw_kx_ky_k$ *satisfying* $uv_0^{i_0}w_0x_0^{i_0}y_0 \ldots v_k^{i_k}w_kx_k^{i_k}y_k \in L$ *for all integer* $i_0, \ldots, i_k \geq 0$ *and* $|v_jx_j| \neq 0$ *for all* $0 \leq j \leq k$.

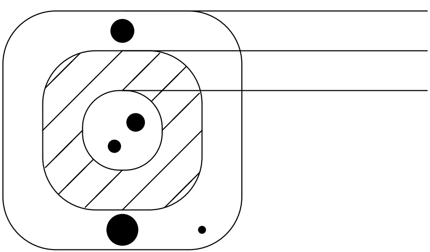
Metalinear languages



Definition

A context-free language is said to be metalinear if it is a k-linear language for some $k \geq 1$.

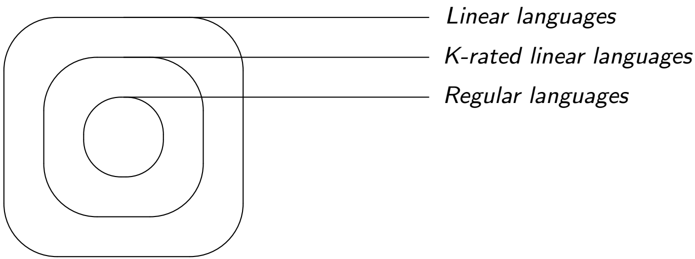
Pumping lemma for not metalinear context-free languages



Proposition

Given a context-free language L which is not in the class of metalinear languages. For all integers $k \geq 1$ there exist infinite many words $w \in L$ which admit a factorization $w = uv_0w_0x_0y_0 \ldots v_kw_kx_ky_k$ satisfying $uv_0^{i_0}w_0x_0^{i_0}y_0 \ldots v_k^{i_k}w_kx_k^{i_k}y_k \in L$ for all integer $i_0, \ldots, i_k \geq 0$ and $|v_jx_j| \neq 0$ for all $0 \leq j \leq k$.

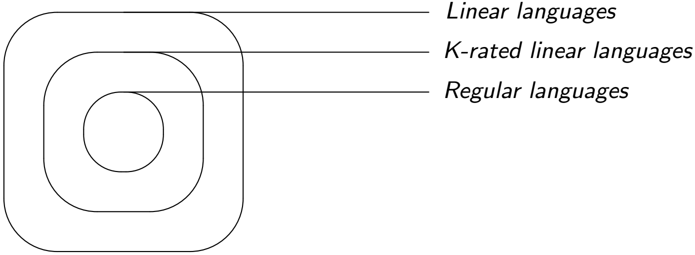
K-rated linear languages



Definition

A context-free grammar is said to be a k-rated linear grammar if it has the form of a linear grammar and there exists a rational number k such that for each rule of the form $A \to vBw$ *the* $|w|/|v| = k$.

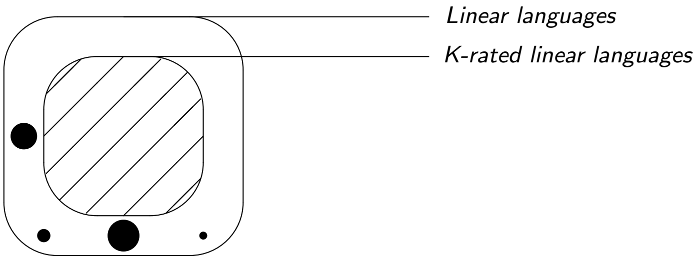
Normal form for k-rated linear grammars



Lemma

Every k-rated ($k = g/h$) linear grammar has an equivalent one in which for every rule of the form $A \to vBw : |w| = g, |v| = h$ such that g and h are relatively primes and for all rules of the form $A \to u$ with $u \in V^* : |u| < g + h$ holds.

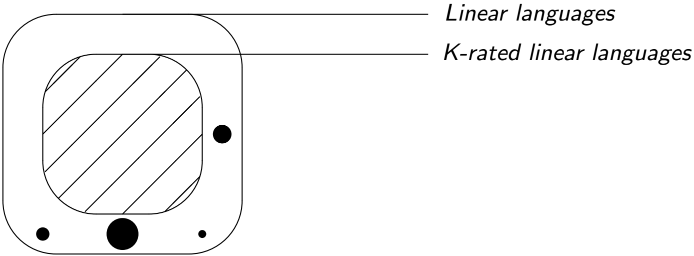
Pumping lemma for k-rated linear languages



Theorem

Let L be a k-rated linear language. (k=g/h) Then there exists an integer n such that any word $p \in L$, $|p| \geq n$ *can be written as* $p = uvwxy$, *satisfying* $uv^iwx^iy \in L$ *for all integer* $i \geq 0$, $0 < |u|, |v| \leq n(h/(g+h))$, $0 < |x|, |y| \leq n(g/(g+h))$, $|x|/|v| = |y|/|u| = g/h = k$.

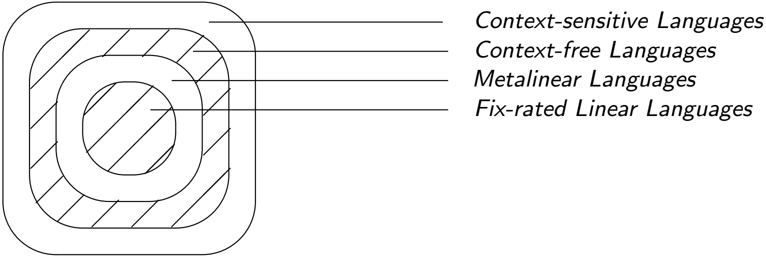
Another pumping lemma for k-rated linear languages



Theorem

Let L be a k-rated linear language. (k=g/h) Then there exists an integer n such that any word $p \in L$, $|p| \geq n$ can be written as $p = uvwxy$, satisfying $uv^iwx^iy \in L$ for every integer $i \geq 0$, $0 < |v| \leq n(h/(g+h))$, $0 < |x| \leq n(g/(g+h))$, $0 < |w| \leq n$, $|x|/|v| = |y|/|u| = g/h = k$.

Summary



The target language classes of the new iteration lemmas

Bibliography

- Amar, V. and Putzolu, G.R.: On a Family of Linear Grammars, (1964).
- Amar, V. and Putzolu, G.R.: Generalizations of Regular Events, (1965).
- Bar-Hillel, Y., Perles, M., Shamir, E.: On formal properties of simple phrase structure grammars, (1961).
- Hopcroft, J. E., Ullman, J. D.: Introduction to Automata Theory, languages, and Computation, (1979).
- Horváth, G.: New Pumping Lemma for Non-Linear Context-Free Languages, (2006).
- Nagy, B.: On $5' \rightarrow 3'$ sensing Watson-Crick finite automata, (2008).