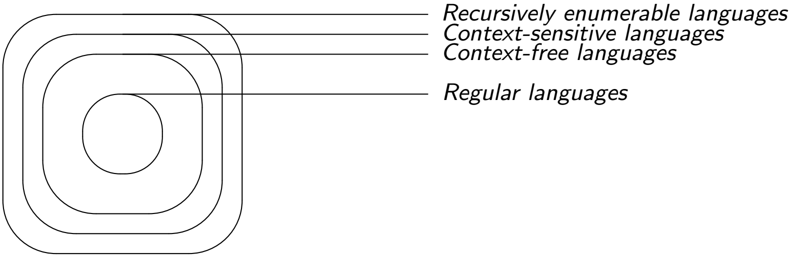
The Language of Primitive Words

Géza Horváth

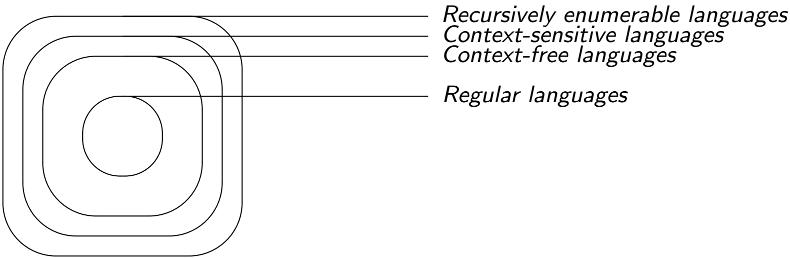
Brno, November 2, 2011.

Chomsky hierarchy



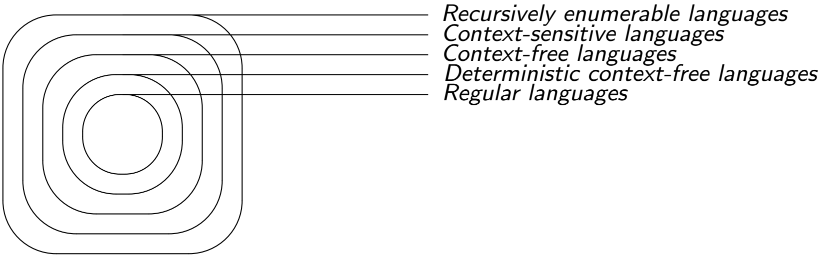
The Generative (formal) Grammar is an universal tool for creating languages.
The Chomsky Hierarchy is a containment hierarchy of classes of formal grammars.

Language of primitive words



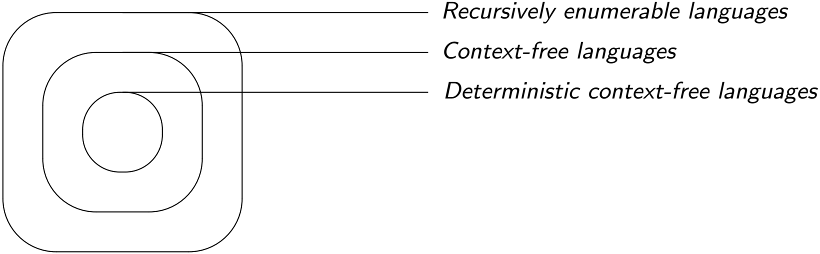
A nonempty word is said to be primitive if it is not a proper power of another word.
Examples over the $\{a, b\}$ alphabet:
Primitive: bbaa, aaaab, ababa
Non-primitive: $ababab = (ab)^3$, $aaaaa = a^5$, $babbbabb = (babb)^2$

Deterministic context-free languages



A context-free language is called deterministic context-free if it can be accepted by deterministic pushdown automaton.
The set of deterministic context-free languages are closed under complementation.

Known facts

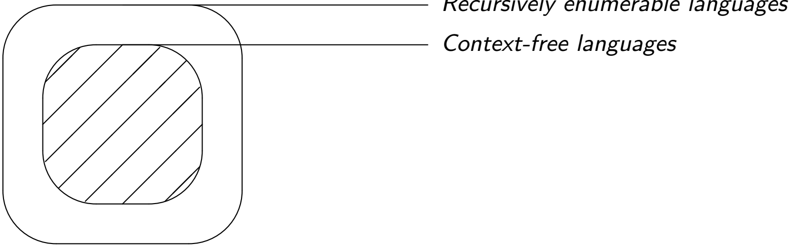


The language of non-primitive words is not context-free. (Bar-Hillel lemma.)

$$\Rightarrow$$

The language of primitive words is not deterministic context-free.

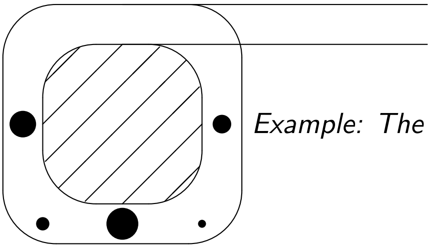
The Bar-Hillel lemma



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

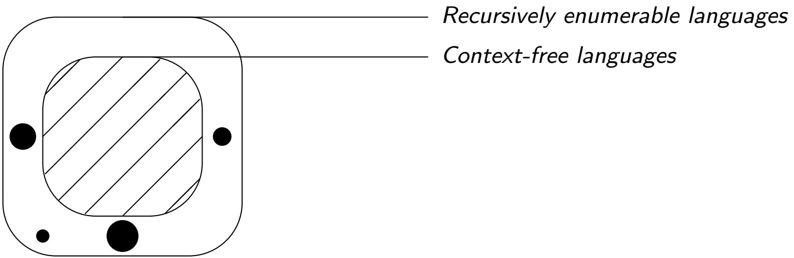


Example: The $L = (a^i b a^i b a^i b \mid i \geq 0)$ *language is not context-free.*

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* $i \geq 0$.

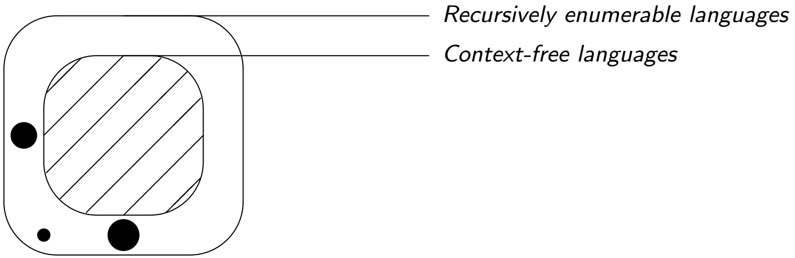
The Ogden Lemma



Lemma

All context-free language satisfies the Ogden restriction.

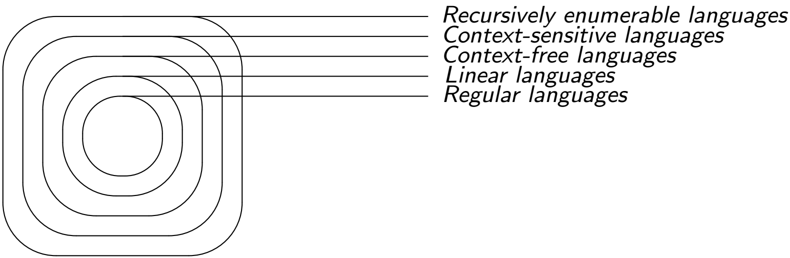
The strong Bader-Moura lemma



Lemma

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

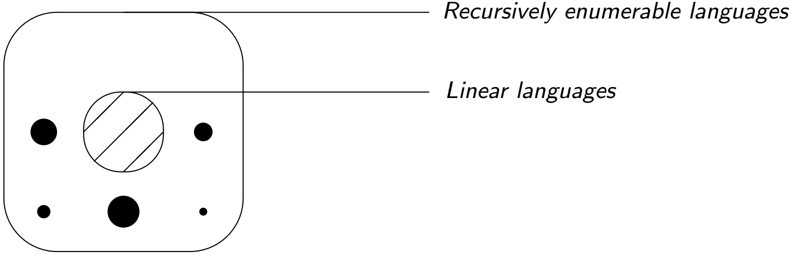
Linear languages



Definition

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

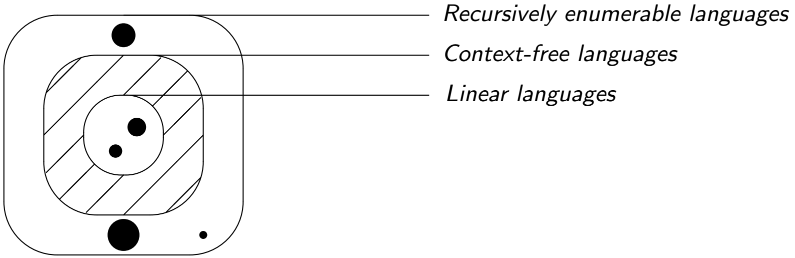
Pumping lemma for linear languages



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* $i \geq 0$.

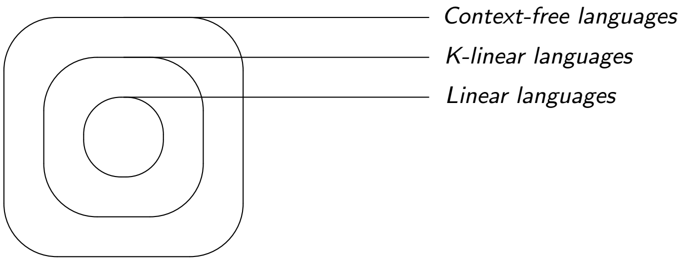
Pumping lemma for non-linear context-free languages



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 $i, j \geq 0$.

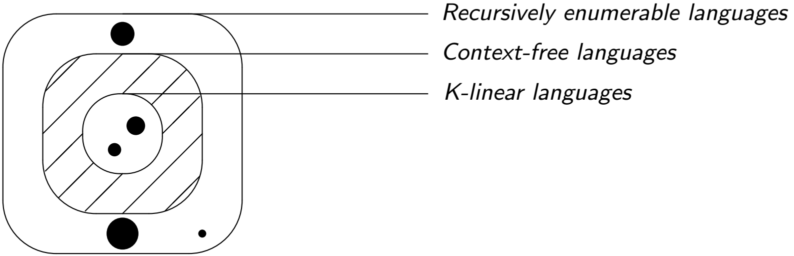
K-linear languages



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.

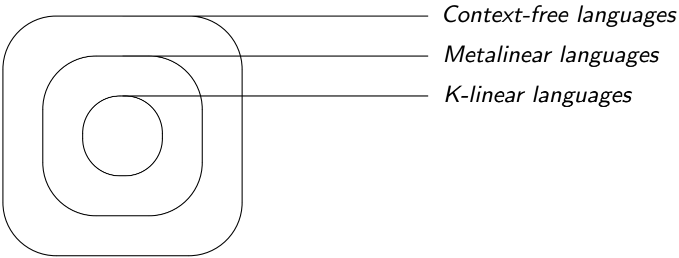
Pumping lemma for not k-linear context-free languages



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 $0 \leq j \leq k$.

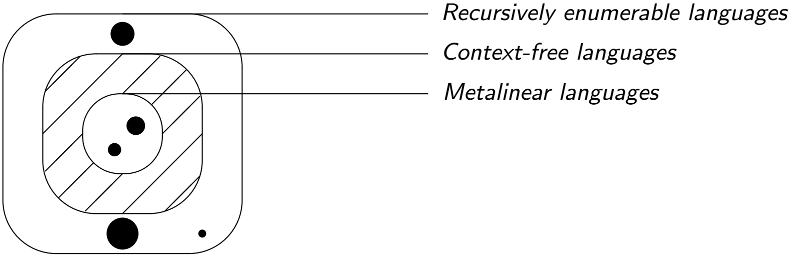
Metalinear languages



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 $0 \leq j \leq k$.

Conjectures

Definition

Denote by L_p the language of primitive words over the $\{a, b\}$ alphabet. Let $L_1 = L_p \cup a^+ \cup b^+ \backslash \{a\} \backslash \{b\}$. If L_p is context-free, then L_1 will be context-free language as well. L_1 generates all the words for length i if and only if i is prime number.

Conjecture

The language L_1 is not context-free.

Conjecture

Denote by L_2 any language which generates all the words for length i if and only if i is prime number. The language L_2 is not context-free.

Conjectures

Definition

We call a language all-word-periodic, if there exists integers n, m, such that the language generates all the words for length $n + i * m$, $i \geq 0$.

Conjecture

If a context-free grammar generates all the word for infinite many length, then the language will be all-word-periodic.

Bibliography

- 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).
- Dömösi, P., Horváth, S., Ito, M.: Formal languages and primitive words, (1991).
- Horváth, G.: New Pumping Lemma for Non-Linear Context-Free Languages, (2006).
- Horváth, G., Nagy, B.: Pumping Lemmas for Linear and Nonlinear Context-free Languages, (2010).