Ogden's Lemma for ET0L Languages

Max Rabkin

March 2012

Overview

What are ET0Ls?

An ET0L system $(V, \Sigma, \mathcal{T}, S)$ generates a language by

- a **L**indenmayer system

$$A_1 A_2 \ldots A_n \Rightarrow u_1 u_2 \ldots u_n$$

- which uses **zero** context

$$A \rightarrow u$$

- is **t**able-driven

$$\Big\{\{Q \rightarrow QS, S \rightarrow Q, a \rightarrow \lambda\}, \{Q \rightarrow a, S \rightarrow a, a \rightarrow a\}\Big\}$$

- and **e**xtended

$$L = \{w : S \Rightarrow^* w\} \cap \Sigma^*$$

Why ET0Ls?

L-systems:

- natural parallelism
- biological inspiration

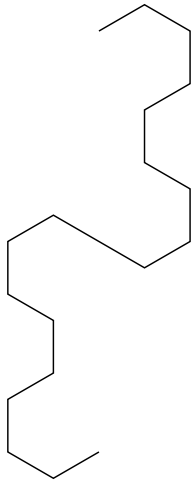
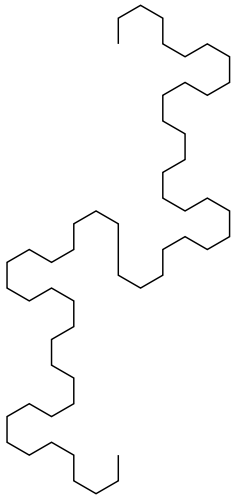
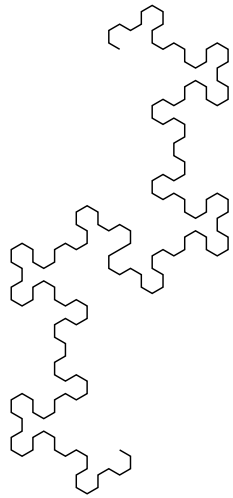
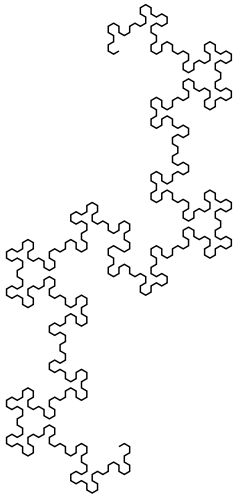
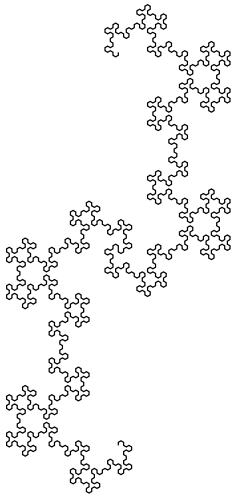
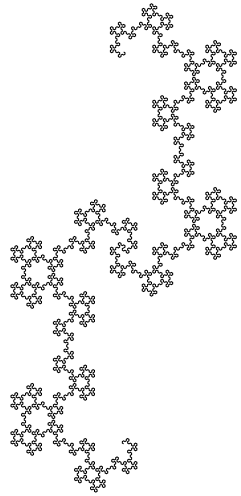
Why ET0Ls?

L-systems:

- natural parallelism
- biological inspiration

ET0Ls:

- generate all the context-free languages
- have nice closure properties (AFL)

Ogden's Lemma for CFLs

Recall

Theorem

If L is a context-free language, then there exists an integer l such that for any $u \in L$ with at least l positions marked, u can be written as $u = vwxyz$ such that

1. *x and at least one of w or y both contain a marked position;*
2. *wxy contains at most l marked positions; and,*
3. *$vw^m xy^m z \in L$ for all $m \in \mathbb{N}$.*

Direction of generalisation

Features to lose:

- arithmetic progressions
- no moving

Direction of generalisation

Features to lose:

- arithmetic progressions
- no moving

and to keep:

- infinite sequence of new words
- duplication of marked symbols

Ogden's Lemma for ET0L Languages

Theorem

If L is an ET0L language, then there exists an $l \in \mathbb{N}$ such that for any word $w \in L$ with at least l marked positions,

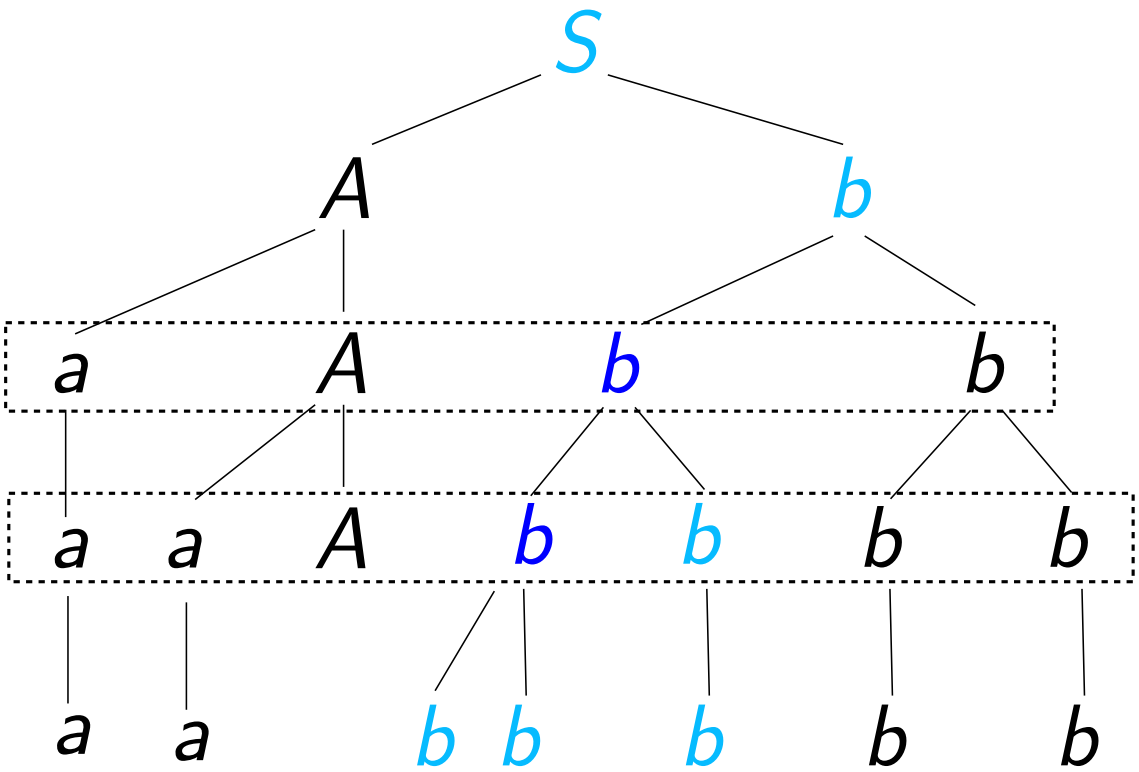
- *w can be written as $w = u_1 u_2 \cdots u_n$ and each u_i can be written $u_i = v_{(i,1)} v_{(i,2)} \cdots v_{(i,n_i)}$ (we will denote the set of subscripts of v, i.e. $\{(i,j) : i \in [n], j \in [n_i]\}$, by I);*
- *there is a map $\phi : I \to [n]$ such that if each $v_{(i,j)}$ is replaced with $u_{\phi(i,j)}$, then the resulting word is still in L, and this process can be applied iteratively to always yield a word in L;*
- *if $v_{(i,j)}$ contains a marked position then so does $u_{\phi(i,j)}$;*
- *there is an $(\mathbf{i},\mathbf{j}) \in I$ such that $\phi(\mathbf{i},\mathbf{j}) = \mathbf{i}$, and there are at least two marked positions in $v_{(\mathbf{i},\mathbf{j})}$ and at least one in $u_{\mathbf{i}}$ but outside of $v_{(\mathbf{i},\mathbf{j})}$.*

Example

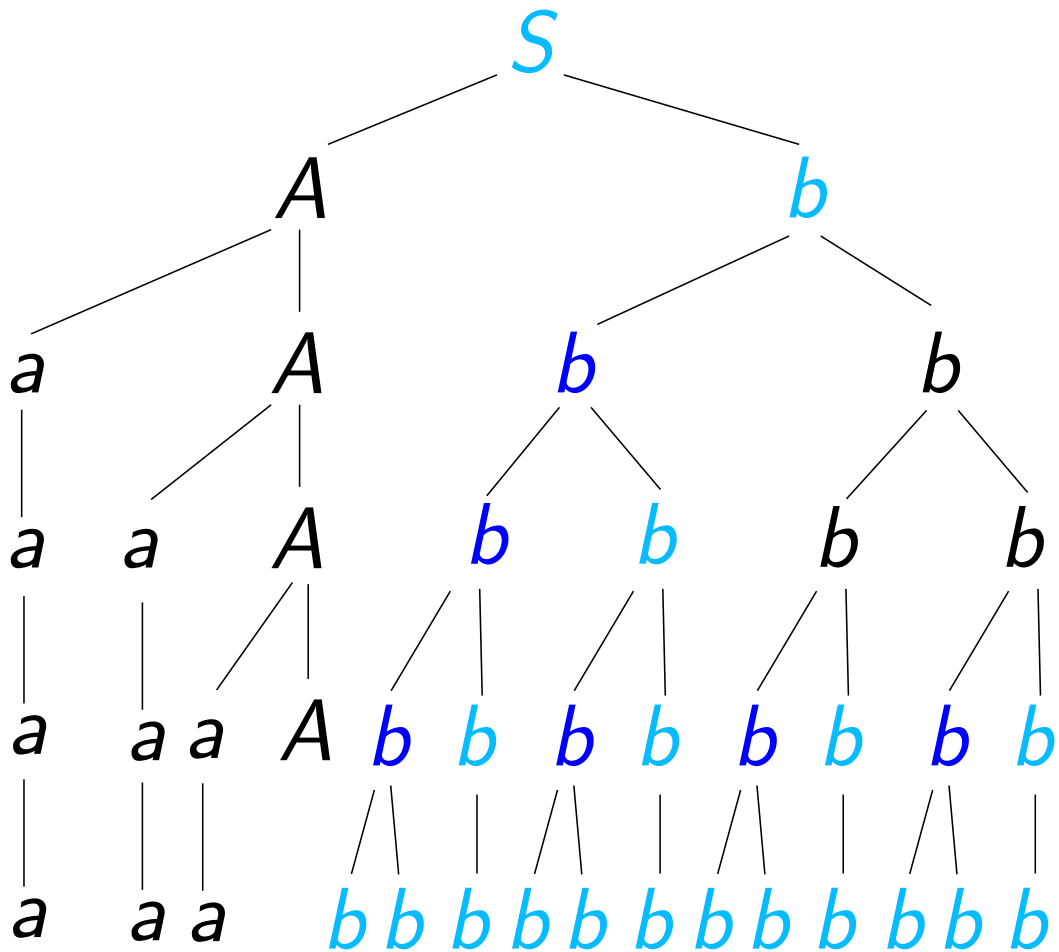
$$L = \{a^n b^m : n \in \mathbb{N}, m \geq 2^n\}$$

$$\Big\{ \{S \to Ab, A \to aA, b \to bb, a \to a\},$$
$$\{S \to S, A \to \lambda, b \to b, b \to bb, a \to a\} \Big\}$$

Proof idea



Proof idea



Lemma is not sufficient

$$L(K) = \left\{ a^{2^n(2k+1)} : n \in \mathbb{N}, k \in K \right\}$$

- K uncomputable implies $L(K)$ uncomputable
- uncountably many examples

Proving a language is not ET0L

$$L = \left\{ a^n b^m : n \in \mathbb{N}, m \geq 2^{2^n} \right\}$$

Need facts about pumping:

- unbounded growth
- no super-exponential growth

Rare and nonfrequent symbols

A symbol is called *nonfrequent* in a language if there is an upper bound on how many appear in words

A symbol is called *rare* if instances of it appear far apart in long enough words

Ehrenfeucht and Rozenberg's rare-nonfrequent theorem

rare $\Rightarrow$ nonfrequent

Can prove frequent $\Rightarrow$ non-rare using our theorem:

- Mark all the symbol
- It's frequent, so above the threshold
- Pumping makes long words with a fixed subword
- So it's not rare

Need fact about pumping: subword with two marked symbols is preserved

What about a shrinking version?
And the shrinking lemma for rFcls?