The Chomsky Hierarchy of Languages

- Regular Languages
  - Context-Free Languages
  - Context-Sensitive Languages
  - Deterministic Context-Free (DCFL)
  - Non-Deterministic Context-Free (NDCFL)

There exists a correspondence between the class of languages and the structure of the grammar.

In particular, the class of languages in one class properly includes the languages in lower classes.

The hierarchy is ordered by their complexity. The higher the language is in the hierarchy, the more complex it is.

A hierarchy of classes of languages, viewed as sets of strings:

- Regular languages are the lower in the hierarchy. The languages they generate are the simplest.
- Context-Free languages are more restricted than regular languages, and the format of phrase-structure rules necessary for generating all its languages.
- Context-Sensitive languages are more restricted than context-free languages, and the format of phrase-structure rules necessary for generating all its languages.

Languages with grammars of this type are called context-sensitive languages.

Languages that require some auxiliary storage are called context-dependent languages.

Languages that require no auxiliary storage are called context-free languages.

A hierarchy of natural languages:

- Natural language: the expressive power of (and complexity of) natural languages
- Auxiliary storage: the resources needed in order to process class of languages
- Computational complexity: what makes individual constructions linguistic complexity: what makes individual constructions more difficult to understand

Languages like killed, is killed, are killed, ate, and that killed, that Car that the dog, that the car, that rat, that the rat, that the cat, that the cat, that the dog are more difficult to understand.

This is the rat that the cat that the dog built.

This is the rat that the cat that the dog killed.
The Chomsky Hierarchy of Languages

### Type 0 (Recursively Enumerable)

- Computational device: General Rewriting Systems
- Languages: Recursively enumerable (type-0)

### Type 1 (Context-Sensitive)

- Computational device: Linear-bounded automata (Turing machine with a finite tape, linearly bounded by the length of the input string)
- Languages: Context-sensitive (type-1)

### Type 2 (Context-Free)

- Computational device: Pushdown automata
- Languages: Context-free (type-2)

### Type 3 (Regular)

- Computational device: Finite state automata
- Languages: Regular (type-3)

Grammar:
- Context-free grammars
- Context-sensitive grammars
- General Rewriting Systems

Rule Form:
- Type 0: $A \rightarrow a$ where $a \in \Sigma$ and $A \in \{U \cup Z\}$
- Type 1: $A \rightarrow a$ where $a \in \Sigma$ and $A \in \{U \cup Z\}$
- Type 2: $A \rightarrow a$ where $a \in \Sigma$ and $A \in \{U \cup Z\}$
- Type 3: $A \rightarrow a$ where $a \in \{U \cup Z\}$
introduce a new kind of rule that will permit us to do so. The significant aspects of language structure. We must phrase structure. Grammars cannot possibly represent all one language, namely English, we can be sure that to represent an obviously significant generalization about since there seem to be no way of using such P rules

(Akmajian and Heny, 1976)

An introduction to the principles of transformational syntax

How not to do it

English is a regular language. I do not know whether or not

Chomsky (1957)

Where are natural languages located?

wered knille a the

This is the mail that the rat that the cat that the dog

built.

rat that ate the mail, that lay in the house that Jack

This is the dog, that wored the cat, that killed the

Performance vs. Competence

Language is context-free? How about Hebrew?

When viewed as a set of strings, is English a regular

What exactly is the question?

Where are natural languages located?

It can shed light on questions of human processing of

complexity of natural languages

The hierarchy represents some informal notion of the

Why is it interesting?

Where are natural languages located?

English is a regular language. I do not know whether or not
English is not a context-free language. Agreement is insufficient to guarantee the accuracy of the grammatical phenomenon of Subject-Predicate.

**How not to do it**

English grammar...clearly the inadequacy of context-free phrase-structure.

*the girl kiss the boy
*the girl saw the boy

Transformational Grammar (Chomsky & Elgin, 1973)

**How not to do it**

Hence, the phrase structure analysis will not be sufficient there will always be an infinite number of such sentences. A finite number of rules like (2.5), (2.52) and (2.54) is not general for any phrase structure grammar containing

Syntax (Peter Culicover, 1976):
Close under homomorphism

Close under intersection

The pumping lemma for regular languages

Proof techniques:

How not to do it right

The notion of "context-freeness"

What is the source for this confusion?

How not to do it

A realistic transformational grammar (Bresnan, 1978):

Syntactic theory (Bach 1974):
English is not a regular language

Since \( L_L \) is a regular language, \( L_L \) is also a regular language.

Since \( L_L \) is a regular language, \( L_L \) is also a regular language.

The following is a sequence of grammatical English sentences:

**English is not a regular language**

\( \{ \{ q + a \mid x \} \} \)
Linguistic complexity

We have wanted to let the kids help Hans paint the house.

The cat the dog the rat bit chased likes tuna fish

someone must have more sentences more difficult to understand than

Natural languages are not context-free

\{uPw^2q_wv \}

\{ \ast \{q+n \} \in \ast \mid xx \}

Some languages that are not context-free:

Languages under homomorphisms and under interaction with regular

Free languages is not context-free is the pumping lemma for context-

The common proof technique for showing that a language

Which is trans-context-free.

Jan sent das meer (d.chin) (em Hans) es huns hand wele

the above regular expression with Swiss-German yields

similarity for accusatives. Intersecting the language defined by

must equal the number of dative NPs (em Hans) and

However, the number of verbs requiring dative objects (hastig)

Jan sent das meer (d.chin) (em Hans) es huns hand wele

verbs must precede accusative-taking verbs:

Dative NPs must precede accusative NPs’s and dative-taking

Is English context-free?
Weak and strong generative capacity.

Two grammars are strongly equivalent if they are weakly equivalent and, in addition, assign the same structure to each sentence.

When grammars for natural languages are involved, we say that grammars \( C_1 \) and \( C_2 \) are weakly equivalent if their string languages are identical:

\[ L(C_1) = L(C_2) \, \quad (C_1 \equiv C_2) \]

In formal language theory, the natural equivalence relation on the string languages is known as the string equivalence relation.