Atitit 语言分类 **形式语言**

目录

[1. 形式语言（英语：Formal language）是用精确的数学或机器可处理的公式定义的语言。 1](#_Toc29630)

[1.1. 形式语言理论是用数学方法研究自然语言(如英语)和人工语言(如程序设计语言) 1](#_Toc24130)

[1.2. 形式语言学 2](#_Toc27631)

[2. 形式语言的特点 4](#_Toc29930)

[2.1. 1,高度的抽象化(采用形式化的手段-专用符号,数学公式-来描述语言的结构关系,这种结构关系是抽象的) 4](#_Toc14056)

[2.2. 2,是一套演绎系统(形式语言本身的目的就是要用有限的规则来推导语言中无限的句子,提出形式语言的哲学基础也是想用演绎的方法来研究自然语言) 4](#_Toc16719)

[2.3. 3,具有算法的特点.(比如说句法分析中采用不同的算法来构造句子的句法推导树) 4](#_Toc15209)

[3. 乔姆斯基把文法分成4种类型，即0型，1型，2型，和3型 4](#_Toc8665)

[3.1. 。0型文法也称短语文法，0型文法的能力相当于图灵机（Turing）， 4](#_Toc20029)

[3.2. 2型文法也称上下文无关法，其能力相当于非确定的下推自动机。 5](#_Toc20517)

[3.3. 3型文法也称右线性文法，由于这种文法等价于正规式，所以也称正规文法。 5](#_Toc671)

[3.4. 从文法描述语言的能力来说，0型文法最强，3型文法最弱。 5](#_Toc6732)

[4. 形式语言与自然语言有两个重要的区别。 5](#_Toc32489)

[4.1. 形式语言的界限是明确的，而自然语言的界限往往不明确。 5](#_Toc25430)

[4.2. 其次，自然语言不管如何庞大，它总是有限的。形式语言则以无限的语言为主要研究对象。 5](#_Toc32689)

[5. 常见形式语言 6](#_Toc3183)

[5.1. 例如数学家用的数字和运算符号 6](#_Toc17588)

[5.2. 、化学家用的分子式等。 6](#_Toc13463)

[5.3. 编程语言也是一种形式语言，是专门设计用来表达计算过程的形式语言。 [2] 6](#_Toc15943)

[5.4. 6](#_Toc27355)

# **形式语言**（[英语](https://baike.baidu.com/item/%E8%8B%B1%E8%AF%AD/109997" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)：Formal language）是用精确的数学或机器可处理的公式定义的语言。

## 形式语言理论是用数学方法研究自然语言(如英语)和人工语言(如程序设计语言)

形式语言（Formal Language）是为了特定应用而人为设计的语言。例如数学家用的数字和运算符号、化学家用的[分子式](https://baike.baidu.com/item/%E5%88%86%E5%AD%90%E5%BC%8F" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)等。编程语言也是一种形式语言，是专门设计用来表达计算过程的形式语言。 [2]

如[语言学](https://baike.baidu.com/item/%E8%AF%AD%E8%A8%80%E5%AD%A6/3632" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)中语言一样，形式语言一般有两个方面: [语法](https://baike.baidu.com/item/%E8%AF%AD%E6%B3%95/2447258" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)和[语义](https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89/9716033" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)。专门研究语言的语法的数学和计算机科学分支叫做**形式语言理论**，它只研究语言的语法而不致力于它的语义。在形式语言理论中，**形式语言**是一个[字母表](https://baike.baidu.com/item/%E5%AD%97%E6%AF%8D%E8%A1%A8/1314769" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)上的某些有限长[字符串](https://baike.baidu.com/item/%E5%AD%97%E7%AC%A6%E4%B8%B2/1017763" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)的[集合](https://baike.baidu.com/item/%E9%9B%86%E5%90%88/73081" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)。一个形式语言可以包含无限多个字符串。

按一定规律构成的句子或符号串的有限或无限的集合

形式语言理论是从语言学衍生而来，作为一种理解自然语言的句法规律。在计算机科学中，形式语言通常作为定义编程语言和语法的基础，是正式版本的自然语言的子集。在计算复杂性理论中，决策问题通常定义为形式语言，复杂类被定义为形式语言的集合，它能被具有有限计算能力的机器所解析。在逻辑和数学基础中，形式语言是用来表示公理系统的语法。 [1]

形式语言有严格的[语法](https://baike.baidu.com/item/%E8%AF%AD%E6%B3%95" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)（Syntax）规则，例如，3+3=6是一个语法正确的数学等式，而3=+6$则不是，H2O是一个正确的分子式，而2Zz则不是。语法规则是由关于符号（Token）和结构（Structure）的规则所组成的。Token的概念相当于自然语言中的单词和标点、数学式中的数和运算符、化学分子式中的元素名和数字。关于Token的规则称为词法（Lexical）规则，而关于语句结构的规则称为语法（Grammar）规则。 [2]

## 形式语言学

也称[代数语言学](https://baike.baidu.com/item/%E4%BB%A3%E6%95%B0%E8%AF%AD%E8%A8%80%E5%AD%A6" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank),它研究一般的[抽象符号](https://baike.baidu.com/item/%E6%8A%BD%E8%B1%A1%E7%AC%A6%E5%8F%B7" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)系统,运用形式模型对语言(包括[人工语言](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E8%AF%AD%E8%A8%80" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)和自然语言)进行理论上的分析和描写.

[形式文法](https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E6%96%87%E6%B3%95" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank):是一种格式,用来说明什么句子在该语言中是合法的,并指明把词组合成短语和句子的规则.

描述语言有三种途径:1,穷举2,文法 3,自动机 其中文法是指的产生过程,而自动机是指的识别过程.一种语言,如果存在对它的识别过程,就一定存在对它的产生过程,反之亦然.

 穷举法— 只适合句子数目有限的语言。语法描述— 生成语言中合格的句子自动机— 对输入的句子进行检验，区别哪些是语言中的句子，哪些不是语言中的句子

某种意义上，形式语言和自然语言之间的不同正像诗歌和说明文的区别：

诗歌

词语的发音和意思一样重要，全诗作为一个整体创造出一种效果或者表达一种感情。歧义和非字面意思不仅是常见的而且是刻意使用的。

说明文

词语的字面意思显得更重要，并且结构能传达更多的信息。诗歌只能看一个整体，而说明文更适合逐字逐句分析，但仍然充满歧义。

程序

计算机程序是毫无歧义的，字面和本意高度一致，能够完全通过对Token和结构的分析加以理解。

---------------------

首先请记住形式语言远比自然语言紧凑，所以要多花点时间来读；其次，结构很重要，从上到下从左到右读往往不是一个好办法，而应该学会在大脑里解析－－识别Token，分解结构；最后，请记住细节的影响，诸如拼写错误和标点错误这些在自然语言中可以忽略的小毛病会把形式语言搞得面目全非

式语言理论是用数学方法研究[自然语言](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80/4146019" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)(如英语)和[人工语言](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E8%AF%AD%E8%A8%80/7595857" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)(如[程序设计语言](https://baike.baidu.com/item/%E7%A8%8B%E5%BA%8F%E8%AE%BE%E8%AE%A1%E8%AF%AD%E8%A8%80/2317999" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank))的语法的理论。它只研究语言的组成规则，不研究语言的含义。形式语言理论在自然语言的理解和翻译、[计算机语言](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%AF%AD%E8%A8%80/4456504" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)的描述和编译、社会和自然现象的模拟、语法制导的[模式识别](https://baike.baidu.com/item/%E6%A8%A1%E5%BC%8F%E8%AF%86%E5%88%AB/295301" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)等方面有广泛的应用。形式语言的研究始于20世纪初，50年代中期将形式语言用于描述自然语言。 [1] 式语言理论是用数学方法研究[自然语言](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80/4146019" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)(如英语)和[人工语言](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E8%AF%AD%E8%A8%80/7595857" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)(如[程序设计语言](https://baike.baidu.com/item/%E7%A8%8B%E5%BA%8F%E8%AE%BE%E8%AE%A1%E8%AF%AD%E8%A8%80/2317999" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank))的语法的理论。它只研究语言的组成规则，不研究语言的含义。形式语言理论在自然语言的理解和翻译、[计算机语言](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%AF%AD%E8%A8%80/4456504" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)的描述和编译、社会和自然现象的模拟、语法制导的[模式识别](https://baike.baidu.com/item/%E6%A8%A1%E5%BC%8F%E8%AF%86%E5%88%AB/295301" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)等方面有广泛的应用。形式语言的研究始于20世纪初，50年代中期将形式语言用于描述自然语言。 [1]

# 形式语言的特点

## 1,高度的抽象化(采用形式化的手段-专用符号,[数学公式](https://baike.baidu.com/item/%E6%95%B0%E5%AD%A6%E5%85%AC%E5%BC%8F" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)-来描述语言的结构关系,这种结构关系是抽象的)

## 2,是一套演绎系统(形式语言本身的目的就是要用有限的规则来推导语言中无限的句子,提出形式语言的[哲学基础](https://baike.baidu.com/item/%E5%93%B2%E5%AD%A6%E5%9F%BA%E7%A1%80" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)也是想用演绎的方法来研究自然语言)

## 3,具有[算法](https://baike.baidu.com/item/%E7%AE%97%E6%B3%95" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)的特点.(比如说句法分析中采用不同的算法来构造句子的句法推导树)

# 乔姆斯基把文法分成4种类型，即0型，1型，2型，和3型

## 。0型文法也称短语文法，0型文法的能力相当于[图灵机](https://baike.baidu.com/item/%E5%9B%BE%E7%81%B5%E6%9C%BA" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)（Turing），

或者说任何0型语言都是递归可枚举的。1型文法也称上下文有关法，其能力相当于线性界限自动机。

## 2型文法也称上下文无关法，其能力相当于非确定的[下推自动机](https://baike.baidu.com/item/%E4%B8%8B%E6%8E%A8%E8%87%AA%E5%8A%A8%E6%9C%BA" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)。

## 3型文法也称右线性文法，由于这种文法等价于[正规式](https://baike.baidu.com/item/%E6%AD%A3%E8%A7%84%E5%BC%8F" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)，所以也称[正规文法](https://baike.baidu.com/item/%E6%AD%A3%E8%A7%84%E6%96%87%E6%B3%95" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)。

## 从文法描述语言的能力来说，0型文法最强，3型文法最弱。

# [形式语言](https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/2277363" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)与[自然语言](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80/4146019" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80%E7%90%86%E8%AE%BA/_blank)有两个重要的区别。

## 形式语言的界限是明确的，而自然语言的界限往往不明确。

因为自然语言有许多方言和习惯用法，而且处于不断发展之中。

## 其次，自然语言不管如何庞大，它总是有限的。形式语言则以无限的语言为主要研究对象。

例如，所有由*n*个*ɑ*构成的字（*n*≥1）组成一个语言*L*ɑ={*ɑ*，*ɑɑ*，*ɑɑɑ*，…},它就是无限的。因此，研究形式语言遇到的第一问题就是描述问题。描述的手段必须是严格的，而且必须能以有限的手段描述无限的语言。

# 常见形式语言

## 例如数学家用的数字和运算符号

## 、化学家用的[分子式](https://baike.baidu.com/item/%E5%88%86%E5%AD%90%E5%BC%8F" \t "https://baike.baidu.com/item/%E5%BD%A2%E5%BC%8F%E8%AF%AD%E8%A8%80/_blank)等。

## 编程语言也是一种形式语言，是专门设计用来表达计算过程的形式语言。 [2]

## 