Atitit 形式语言 与自然语言

作者：L晓霓  
链接：https://www.zhihu.com/question/24278291/answer/28400475  
来源：知乎  
著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。

在“语言”上的分类，自然语言是处于最千变万化，表达能力最强大，同时也最难以摸索出统一的规律的一种语言。在这个基础上，语言学家们才提出一个概念叫“形式语言”，即formal language，这是一种用精确的数学公式就能把一个语言中的语法表达清楚的语言。

形式语言的概念到了[诺姆·乔姆斯基](https://link.zhihu.com/?target=http://zh.wikipedia.org/wiki/%E8%AF%BA%E5%A7%86%C2%B7%E4%B9%94%E5%A7%86%E6%96%AF%E5%9F%BA" \t "_blank)，就形成了一个新的理论（乔姆斯基谱系）。这个谱系把形式语言分成四种：  
文法 语言 自动机 产生式规则  
0-型 递归可枚举语言 图灵机 无限制  
1-型 上下文相关语言 线性有界非确定图灵机 α*A*β -> αγβ  
2-型 上下文无关语言 非确定下推自动机 *A* -> γ  
3-型 正规语言 有限状态自动机 *A* -> *aB A* -> *a*

这四种语言中，从上到下的语言是变得越来越容易用公式表达，但是表达能力也越来越弱。到了最后一种语言，已经变成最普通的正规语言了。

计算机编程语言是属于2-型，即上下文无关语言(context-free grammar)。只有这种语言，才可以用我们现在研究出来的编译器进行编译，也只有用这种语言写出来的代码，才可以经过编译器的解释之后被电脑理解。

回到楼主的问题，如果你想用自然语言来编程，就像楼上说的，你的逻辑能力已经足够强大，你已经可以用你强大的逻辑能力来写出复杂的自然语言程序，但问题是，那个编译器怎么办？这个编译器必须理解自然语言中那种千变万化的语法，而这个是我们现行的数学理论中还无法建立起来的体系。

楼上有人说，这是自然语言处理。实际上，自然语言处理现在所处的阶段仍然是在用各种数学上的统计模型去理解自然语言的。这些统计模型到目前为止，已经相当复杂，可以比较准确地推算出自然语言中的很多任务，比如翻译，归类，等等。但是还不能保证100%正确。

所以我认为，用自然语言写程序，你首先要用那个不是完全正确的自然语言处理机去把你的程序变成一个上下文无关语言（势必要造成很多错误，还有损失很多细节），然后再用编译器去编译。

所以，我认为这不是逻辑的问题，而是以现行的数学理论，根本不可能做到。