

# 汉语形式语法中的空位和非常序

罗建林

(中国计算机软件与技术服务总公司)

语言工程事业部

## Empty slot and abnormal order in chinese formal grammar

*Luo Jianlin*

CS&S 100081

[Abstract] Chinese formal grammar is based on the theory of general grammar, but in the parsing of spoken chinese, there are two common kinds of problems that cannot be dealt with effectively in the framework of general grammar — components omitting and order reversing. We in this paper call these two phenomena in spoken chinese empty slot and abnormal order, respectively, we discuss their basic types and distributions in chinese, and we provide methods for tackling them. We also point out that these two problems are very important, but extremely difficult. A deep study on them will on the one hand, supplement chinese kernel grammar research, and on the otherhand, improve the parsing quality of chinese.

[Keywords] formal grammar, general grammar, spoken language, written language, empty slot, abnormal order.

【内容提要】汉语形式语法是建立于普遍语法的理论基础上的,但在分析汉语口语时常常会遇到许多普遍语法无法解释的现象,归纳起来有两大类型:成分缺省和语序倒错,本文称之为“空位”和“非常序”。鉴于这两类现象在汉语中的普遍性,汉语形式语法不可能回避或视而不见。本文探讨了汉语中空位和非常序的基本类型以及在汉语口语中的分布,指出了与传统语言学所谓“省略”“空语类”和“倒装”的异同,提供了形式语法的处理策略。同时,本文也指出,“空位”和“非常序”是一个极其复杂艰难而又极有意义的问题,对这个问题的解决,一方面是对核心语法的有益补充,可以促进汉语语法研究,另一方面也有助于提高汉语分析的质量。

【关键词】形式语法 普遍语法 口语 书面语 空位 非常序

### 1. 引言

从表面上看,汉语的确是一种很独特的语言,比如说没有形态变化,没有词间隔等等,另外,汉语中还存在大量的语法成分不完整的现象。在自然语言理解和机器翻译中,这些先天的不足使得汉语处理困难重重。然而这些现象又是不可避免的,长期的习惯和使用者的随意造成了语言的不规范,“名无固宜,约定俗成谓之宜”,这些不规范的语言一旦被人们接受,便成为一股不可抗拒的力量,变得非人力所能改变。因此,怎样表示和处理这些为汉语所独有的问题,就

成为形式语法面临的任务。

本文主要探讨汉语口语(定义详后)中语法成分的缺省和语言形式的不正常序列,属于核心语法的外围层次,是核心语法有益的补充。

为什么要把口语作为一个特殊对象研究?作者曾经提出建立汉语形式语法的设想[1],并指出将以形式语法中的核心语法为首要任务。核心语法即规范书面语,是语法中最根本和最稳定的一部分。对口语的探讨一方面有助于确定规范书面语的范围,有利于建立汉语的核心语法,另一方面有利于用形式化方法分析口语。

## 2. 名词术语界定

### 2.1 口语和书面语

语言因其表现形式可以分为两大类:口语和书面语。通常认为:

口语是:“同书面语相对。指语言的口头存在形式。”“由于口语常常有语境的补充作用,所以多用简化、省略形式。口语和书面语有交叉。例如广播、报告的语言是书面语的口头形式;谈话的记录则是口头语言的书面形式。”书面语是:“同口语相对。指某一交际集体在书面语上使用的语言。”“一种语言的标准语一般是在书面语中定型下来的,书面语是一种语言进行提炼加工和规范的基础。”[2]

从这种角度对语言进行分类也许不无意义,但用在形式语法中却有许多不足之处。首先是二者的边界不清,正如上述“有交叉”,你中有我,我中有你;其次是无法确定各自在语法上明显的确定的特点,难以进行形式化。为此,作者在一篇文章中[1]把口语和书面语重新定义为:

- (1) 语法形式与语法意义一致,在语法上无遗漏成分,且由此可获得其语言意义的句子,称作规范书面语。简言之不用语义即可分析的句子。
- (2) 语法形式与语法意义虽不相一致,在语法上有缺陷,但可用规则弥补的句子称作非规范书面语,简言之靠语义能分析的句子。
- (3) 语法形式与语法意义相距甚远,语法成分有遗漏,且无法用规则补上,不能由此获得其语言意义的句子称作口语(非书面语)。简言之靠语义也难分析的句子。

因此,并非“说出来的话就是口语”,也并非所有能形诸眼睛的语言表现方式如写出来、印刷出来的文字都是书面语。规范的书面语应该具备以下两方面的性质:

- i. 语法成分完整无缺;
- ii. 语法和语义的分析结果一致。

与此相应,口语是不具备上述条件的语言表现形式。作者摒弃了传统以感官感知语言的方式对口语和书面语的分类,希望以全新的思路使口语和书面语的研究导向科学化。

### 2.2 语法空位

何谓“语法空位”?所谓空位是语义上当有而语法上没有的语言形式,简言之就是遗漏了功能形式词(定义参见[1],简称功能词)。本文涉及的功能词主要有介词、系词等。从概念上来说,空位所表示的语义是理解时必不可少的,虽然有时不易辨别出来,如:

鲁迅浙江绍兴人

人们是怎样理解这个句子的呢?从语法形式来看,是“N + N + N + N”的组合,从语法意义上来看,这些名词的序列表达的是一个判断:“鲁迅是浙江绍兴人”,系词“是”在实际的语流中省略了,即出现了系词空位。不这样理解,就无法弄清它们之间的关系,这句话应当理解成:

鲁迅(是)浙江绍兴人

这是可以通过规则补出语法成分的非规范书面语。在语法分析时,语法空位是可以而且必须通过规则来弥补的。

### 2.3 常序和非常序

常序就是在语法链和语义链上按正常次序展开的语言符号。常序是核心语法的基础,其语法形式和语法意义是一致的。无论从形式还是意义出发,都能导致正确的分析结果。具体来说,常序的最基本的形式是

S + V + O

非常序即非 S-V-O 序,在语义上表现为 AGENT 置于动词后或者 OBJECT 置于动词前的语言形式。作者认为,S-V-O 的 S 和 O 之间必定有语义上的施受关系,否则就不是 S-V-O 形式,而是有语法空位或者非常序。从理论上来说,非常序应当可以通过规则调整为常序。

非常序与语感:非常序不是语感上的不通顺,实际上,以汉语为母语的人,对非常序是不易察觉的。在各种语言运用的场合,因为交际的需要,移位的手段被广泛使用,人们也习以为常,语感上就不认为有任何不妥了。

### 2.4 空位和非常序与口语的关系

空位和非常序是口语的两大特征,他们和口语是相辅相依的关系。解决了空位和非常序,口语的问题也就迎刃而解了。

当然,也有个别句子成分冗余,在深层结构中必须删除,如:

(有)人来了                      (是)他干的好事

但这种情况与空位和非常序相比,就显得不太重要了。

### 2.5 普通话口语的分布(阶层、环境)

语言的运用是丰富多彩的。所谓语法规则只是人们分析、归纳、总结出来的,而不是按照事先规定好的。正因为语言产生的自发性,所以,语言也是纷乱复杂的,各个不同的社会阶层、各种不同的文体、不同的语言使用环境,都会影响到语言的纯洁性,在语言运用上显示出不同的特色。语言的使用具有以下的特点:

- i. 使用者文化程度越高,语言越规范,口语越少;反之则口语多。
- ii. 正式的场合、严肃的环境,语言较规范;随便的场合,轻松的环境,语言较灵活。
- iii. 公文语体、政论语体、科技语体的语言规范,文艺语体、口语体的语言较活泼。

所谓“灵活、活泼、不规范”等,即语法空位和非常序出现得多,更多借助语境和常识来表情达意。本文针对的就是这一部分。不过以上的分类是就大体而言的,具体到每一场合或每一文体,都会有规范和不规范的成分,只不过是程度不同而已。在机器翻译和自然语言理解中,目前主要对象是科技文章,其中的口语成分较少,但即便如此,也足以使人感到无从入手了。

## 3. 大致类型

从非规范书面语到口语,其中的难易程度是不一样的,据 2.1 的定义,非规范语是可通过规则弥补的句子,因而也是本文探讨的主要对象,而口语的文体过于复杂,必须借助其它手段才能处理,本文不作重点讨论。

本文符号定义:

TC=处所名词 TD=时间词 N1=OBJECT N2=SUBJECT

, =搭配关系 — =结合关系

### 3.1 空位

普通话口语中的空位,主要有:

#### 3.1.1 介词空位

TC + V + N2: ()台上坐着主席团 ()桌上放着一本书 ()大地开遍红花

N1 + N2 + V: ()这件事我不知道——>对这件事我不知道

#### 3.1.2 系词空位

N + N:鲁迅()浙江绍兴人 今天()星期三

N + AJ:鸡()很肥 汽车的速度()特别快

#### 3.1.3 其他空位

你去收拾一下(),我们马上走 商店有袋装康师傅,还不赶快去买()

鸡不吃()了

### 3.2 非常序

#### 3.2.1 主语非常序

TC + V + N2: 台上坐着主席团

TD + V + N2: 前面来了一个人

#### 3.2.2 宾语非常序

N1 + N2 + V: 饭我吃了 这件事我不知道——> 我不知道这件事 N1 + V: 信寄走了 窗户破了 鸡不吃了(歧义)

#### 3.2.3 状语非常序:这件事我不知道——> 我对这件事不知道

### 3.3 语用(口语)空位和非常序

这种类型的空位和非常序既多且杂,下面仅举几例:

“怎么啦,你?”

“喊什么喊?”

“谁?”“我!”

“天气太冷了(快把窗户关上)”

“大会是谁的主席?”

“王冕七岁上死了父亲”“李大可四十九岁生了孩子”

“王府井四毛”

“说句笑话解解乏你还不让吗?”

“这稿子请让我看一次校样”(可参看 [9])

对口语中对话的省略本文不涉及,如上所列,完全是由语境造成的,要很好地解决这些问题决非易事。

空位和非常序有时表现在一个句子上,从不同的角度可分别看成空位和非常序。在处理时可分别不同情况归入。看成空位和看成非常序在语义上有差异,但这种差异常常是语用上的,在语法上可忽略。

本文旨在说明形式语法中一种语言现象的处理,所以不刻意于列举各种空位和非常序的种种类型和表现,只举大致类型。

## 4. 语言学的态度

空位和非常序的问题并不是作者第一次发现和提出,语言学对此早已有研究,只是在名

称、范围、处理方法上不太一样。“空位”被称为“省略”“缺省”“隐含”等，“非常序”被称为“宾提动前”“宾踞句首”“反宾为主”“副冠句首”“主退谓后”[6]、“主语后置”、“移位”等。

关于省略有很多种说法：

(1) 指句子因交际的需要省去其中的某个成分，这省去的成分是可以根据上下文明确添补出来的。包括：对话时的省略，自述时的省略，承前的省略。（黎锦熙：新著）

(2) 狭义的省略，与“隐含”相对。指句中省去的是可以补出来并且只有一种补法的词语。这种省略有两个条件：①如果一句话离开上下文或者说话的环境意思就不清楚；②经过添补的话是实际上可以有的，并且添补的词语只有一种可能，这样才能说是省略了这个词语。……广义的省略，包括意思里有而话语里不出现的所谓“隐含”。（吕叔湘：中国语文，1986.1）

(3) 指的是在一定的语法条件下没有出现，结构上必不可少的成分。……省略的说法不宜滥用，特别是不能因为一个句子意义不足就主观地说它省略了什么成分。如：一张动物园！（公共汽车上乘客对售票员说）是完整的句子，没有省略什么。…（朱德熙：讲义）

(4) 指句子在意念上或结构上或交际上省去某一成分。省略属于语言的三个不同平面：①意念上的省略，属语义平面。②结构上的省略，属语法平面，必须是在语言的表层结构中有形可寻，并且是语言中的一种特定格式。③交际上的省略，即语境省略，一般都可以通过上下文明确添补出来的。（王惟贤：中国语文，1985.6）

(5) 王力分为：“承前的省略”和“习惯的省略”。（以上所引均见于[2]P284~286）

概而言之，省略指原本有而在对话或上下文中因语言的经济原则而没有出现，或习惯上不说。对于省略部分的补出，有人认为能补出且只有一种补法的就是省略，补出后不成话（语感上不认）就是“蕴涵”（隐含），与我们的空位说不完全一致，主要区别在一部分空位现象（如“这件事我不知道”）语言学界不认为有省略。

语言学对“省略”类型的限定请参见[4]。

比较 2.3 可以看出，“省略”所辖范围与空位不一样，主要区别在：

- i. 认为形容词和名词可以直接做谓语，故无系词空位；
- ii. 认为句首的名词性成分必为主语，故无介词空位和主宾语非常序；
- iii. 认为语用中的省略是正常现象。

乔姆斯基也研究了语言中的缺省现象。在他 EST(扩充标准理论)的管辖与约束(GB)理论中，定义了一种重要成分“空语类”(empty category)，他用来指称“具有一定范畴意义而没有指派实在的语音内容的范畴”[2] P207。空语类主要有移位后产生的语迹(trace)和无声、无形的空代词 PRO [3]，如：

(1) Which book have you read t ?

(2) It is easy PRO to be friendly.

他的“空语类”中的语迹是通过句型转换比较辨别出来的，不同于本文的空位，它们并不缺少任何语义成分，另外，乔姆斯基把空范畴局限在名词性成分，本文则涉及语法的各个方面。语迹问题从另一角度看就是非常序，只有非常序才有语迹问题。空代词是一种空位现象，但本文不限于此。

《中国语言学大辞典》(P367)定义“空位”：“为一个语言项目规定的上下文空格，如“一个()走了”，这一空位是为名词性词语准备的，即空位中只能填进名词性词语。”[7]这句话很令人费解，何谓“语言项目”，何谓“空格”，不得而知。另外查阅了《实用语法学词典》(职工教育出版社)[8]、《语法学词典》均无“空位”条目，与本文所述有关的内容只见于“省略”词条。

对“非常序”，争论很大，争论的背景是传统语法和结构主义语法两大体系的矛盾，也就是意义标准和形式标准的矛盾。传统语法是建立在逻辑思维的基础之上的，一切从概念出发，思维的普遍性，决定了语言的普遍性，因此，汉语研究者比附外语语法，把汉语中不合于印欧语言之处概以“倒装”“提前”“后置”来解释，以黎锦熙为代表。后来情况有所变化，结构主义语法学理论引进后，早期的摹仿之作遭到一致的批评，认为抹杀了汉语的特色，因而以后的汉语研究中倾向于从形式上分析语言现象，把非常序当作是汉语区别于其它语言的不同之处，是汉语的常序（除“要点什么，你？”等以外），所以在后期的语法书中大量出现“名词谓语句”、“形容词谓语句”、“主谓谓语句”。详情可参看 [6,9]。

## 5. 形式语法的理解

所谓形式语法的处理就是使之变成“条件—动作”的偶对 [5]，以便于计算机的分析和转换。

### 5.1 空位和非常序的辨认

自然语言处理可分为辨认和处理两个步骤，空位和非常序的辨认或理解是处理的基础。因为所谓空位和非常序是在语法分析中为解决语义成分的缺省而设的，在语言形式上并没有明显的标识，所以在理解时有一定的困难。

从第三节所列举的类型来看，属于语用层次的类型很复杂，对语境和常识的依赖过大，基本上无法辨认。

#### 5.1.1 空位的辨认

i. 若句首为处所词，其后为 V + N 形式，且 V 和 N 之间有动宾关系，则为介词空位。可表示为：

判据： $\text{cond}(\text{TC} + \text{V} + \text{N}_2)$  and  $\text{cond}(\text{N}_2, \text{V})$

若真则： $(\text{P}) + \text{TC} + \text{V} + \text{N}_2$

举例：台上坐着主席团 学校有一口井

ii. 若语言形式为 N1 + N2，且 N1 和 N2 之间有①整体与部分②事物和属性③人物和籍贯④时间和时间等等关系，则为系词空位。可表示为：

判据： $\text{cond}(\text{N} + \text{N})$  and  $\text{cond}(\text{语义为整体与部分等等})$

若真则： $\text{N} + \text{N} \text{ --- } \text{N} + (\text{是}) + \text{N}$

举例：老张黄胡子 鲁迅浙江绍兴人 今天星期三

#### 5.1.2 非常序的辨认：

i. 若语言形式为 N1 + N2 + V，N1 和 N2 之间不能形成定中关系，且 N2 和 V 之间有主谓关系，N1 和 V 之间有动宾关系，则可确定该语言形式为非常序。可表示为：

判据： $\text{cond}(\text{N}_2, \text{V})$  and  $\text{cond}(\text{V}, \text{N}_1)$  and  $\text{Ncond}(\text{N}_1 - \text{N}_2)$

若真则： $\text{N}_1 + \text{N}_2 + \text{V} \text{ --- } \text{N}_2 + \text{V} + \text{N}_1$

举例：这件事我不知道 饭我吃了

ii. 若句首为副词（包括时间词），其后为 V + N 形式，且 N 和 V 之间有施动关系，则为非常序。可表示为：

判据： $\text{cond}(\text{TD} + \text{V} + \text{N}_2)$  and  $\text{cond}(\text{N}_2, \text{V})$

若真则： $\text{TD} + \text{V} + \text{N}_2 \text{ --- } \text{N} + \text{TD} + \text{V}$

举例：前面来了一个人 高高地跳下来一个人

## 5.2 空位和非常序的处理方法

空位与非常序是语言的异质部分,对这种情况通常有两种处理办法:一、承认它是一种特殊现象,尽量通过添加、移位等方法使之变为常序;二、认为它是汉语中一种正常现象,没有空位,也不是非常序,可单独制定规则来处理。

两种方法都有优点和局限。采用第一种方法,与外语寻求语法上的一致,尽量减少语言转换规则,比较适于机器翻译系统;第二种方法充分利用语言的本来面目,不用增加额外的转换,比较适于自然语言理解系统。

但二者也都有缺陷。语言学几十年的研究已经证明第二种方法对汉语研究利少弊多,第一种方法也远非完善。对于空位,有的能够通过添加、移位等手段来使之完整,有的则不行;非常序也有一些根本无法处理。这在本文的 3.3 节已有限定,但目前似乎还没有比这更有效的办法,因而本文采取前者来解决汉语的空位和非常序问题。

## 5.3 具体处理策略

### 5.3.1 对空位的处理:

主要通过添加手段来处理。如上述几例可转换为:

介词空位:台上坐着主席团 =====> (在)台上坐着主席团

学校有一口井 =====> (在)学校有一口井

系词空位:老张黄胡子 =====> 老张(是)黄胡子

今天星期三 =====> 今天(是)星期三

鲁迅浙江绍兴人 =====> 鲁迅(是)浙江绍兴人

### 5.3.2 对非常序的处理:

主要通过移位手段来处理。如上述几例可变换为:

宾语非常序:这件事我不知道 =====> 我不知道这件事

饭我吃了 =====> 我吃了饭

主语非常序:前面来了一个人 =====> 一个人前面来了

高高地跳下一个人 =====> 一个人高高地跳下

因为变换的结果不是实际的语言,只是用来进行分析的中间结果,所以不考虑变换后的句子的可读性和是否流畅。

### 5.3.3 对空位和非常序交叉情况的处理

有的句子既有空位问题,又有非常序的问题,有时调整语序后可附带消除空位,有时互不相关,必须视不同情况分别处理,各个击破。

## 6. 结语

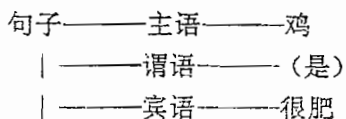
分别语法空位和非常序的理论意义和实用意义何在?

作者认为,分别空位和非常序的理论意义在于:空位问题的提出是迈向普遍语法的十分重要的一步;可以推动汉语研究的进一步深入发展。

实用意义在于:可以解决核心语法所鞭长莫及的问题,有利于提高汉语以及机器翻译的译文质量。

例如:可以正确生成汉语完全语法树 [1],否则会出现主动词空缺和分析错误,

“鸡很肥”的语法树是：



### 致 谢

本文得到中软总公司吴蔚天老师的指导,在此深表谢意!

### 参考文献:

1. 吴蔚天、罗建林:“关于建立汉语形式语法体系的探讨”,《机器翻译研究进展》,135~144,电子工业出版社,1992.
2. 王惟贤主编:《语法学词典》,P208,299~300,284~286,浙江教育出版社,1992.
3. 徐烈炯:《生成语法理论》,P58、298,上海教育出版社,1988.
4. 杨成凯:“语言学中的省略概念”,《语言学和汉语教学》,189~205,语言学院出版社,1990
5. 冯志伟:“计算语言学对理论语言学的挑战”,《语言文字应用》,1992. 1.
6. 《中国语文》杂志社编:《汉语的主语宾语问题》,60~75,中华书局,1956.
7. 陈海洋主编:《中国语言学大辞典》,江西教育出版社,1991.
8. 陈高春主编:《实用汉语语法大词典》,职工教育出版社,1989.
9. 吕叔湘:“汉语句法的灵活性”,《中国语文》,1986. 1