

## 詹卫东：记忆在语言知识表达中的重要性

非常感谢邀请我来参加这个交流，对我来说是一个非常好的学习机会，刚才听了刘挺老师的报告，学到了很多。另外李院士讲到记忆认知非常重要，跟我要讲的内容也比较契合。在语言学研究里面，我有这样一些认识，就是强调记忆在语言知识表达中的重要性。下面要谈的构式语法可能是在座的计算机专家比较陌生的名词，但我理解它的实质是更加强调记忆方面的。

在实际对话系统中有很多综合的语言因素，不过语言研究跟前沿的技术发展相比有点滞后，还是关注比较基础的东西，比如句法语义层次的东西，最基本的范畴与词类还没有完全搞清楚，但是在实用层面、语言工程层面都已经要标很多语料，实际上还是有各种各样的问题。在语言本体方面，词类的研究已经有几十年了，到现在还在研究，借这个机会，希望能让计算机行业的专家也可以了解一下这个情况。

下面我主要说三点：一是构式语法跟传统语法观念上的区别，二是把我收集整理学术界已有的关于汉语词类体系的认识概要地介绍一下，三是从构式的视角看词类会有一些什么新的看法。相对于构式语法来说，主流的语法理论强调语言的规则，就是提取语言中规则的知识系统。这方面主要刻划了语言单位在组合时候的递归性，就是大的语言单位的意义由小的语法单位组合

得到。构式语法强调另外一方面，它认为这不是语言的本质性质，而是更加强调约定俗成的观念，也就是说语言单位是形式和意义一种约定俗成的配对，没有那么多道理可讲。一个形式配一个意义，构式有独立的意义，不能由其组成部分的意义组合得到，它强调整体的部分，这是最核心的观念上的区别。

词类体系是处理其他东西的基础，首先得把词分类，在这个基础上才能做后面很多理解的事情。基本上，它是依据组合上的分布差异，有差异分成不同类，有共性聚成一个类，这是一个非常朴素观念。实际上，关键在于组合的框怎么定，语言学家从内省出发来定组合框架，觉得好像很清楚，就定下来了，实际上碰到真实语料问题就非常多。现在我们用构式的观念来看，觉得问题非常多。

举一些例子，对比一下大家就很清楚了，以前我们习惯的短语结构是一个树来表达的，可以递归，可以无限生成，比如最普通的、简单的主谓结构。而构式就是像下面这样的形式，比如“他这眼睛是看书看的”，“人多了去了”，“你罚你的款，他违他的章”，大家一看到这样的句子，马上就会觉得它有字面之外的意义，很自然的就有这样的感觉。再比如“不是办法的办法”，“天才中的天才”，就是这些意思不是在这个词层面上的，还有“一天比一天严重”，这些例子很普通，我们会经常碰到，另外也有些新颖的说法，比如“被套路”。还有像“这一去就是三十

年”，“他是儿子，我是女儿”，这是俩人在讨论各自的家庭情况的时候说的话。再比如“有一种优秀叫别人家的孩子”，带有修辞性色彩的用法。这些都是广义的构式，共同特点非常清楚，一定有超越字面的意思，就是那个意思不仅仅是辞典里面描写的那个词的意思，构式语法认为有一个构式的意义在里面，它特别强调这点，从这个出发，后来引发一整套理论体系。

现在简单回顾一下现有的词类体系，先看汉语学界的词类体系，20 世纪的 11 家词类体系，再到 21 世纪的 8 家词类体系，还有就是信息处理用的词类标记集，有 7 家单位 10 个标记集。

从语法学本体研究方面来看发展趋势很明显，词类分得越来越细，对于汉语的一些词类范畴特点的认识越来越深入了，最早《马氏文通》比较接近英语的 9 类，到 20 世纪 90 年代的时候大概 15 类。

到 21 世纪之后，主要的教材、语法书里面的词类发展到 20 类，比如北大的语法教科书里面词类体系现在定的是 20 类，是目前最多的。商务印书馆的《现代汉语词典》应该是影响最大的语文辞书，它从第五版之后开始标词类，大类是 12 类，但是实际上里面有些类别分了小类，实质类别应该是 17 类，总的来说各家的词类还是共性更多。

下面再简单地分析一下，词类、词性对于后续的句法语义分析所起的作用。词类信息包含这个词的组合方向以及对组合对象

的约束，在信息处理往后做分析的时候功能是非常明确的，不管做依存的也好，或者做短语结构树的分析也好，都是这样。比如朱德熙先生提出汉语中的区别词，这个词类的效果就非常好，因为组合方向以及对组合对象的约束都明确。但是汉语里面的名动形三大类实词，它们的组合方向不清楚，组合的约束对象也不清楚，信息量比较小，对句法分析能够起到的帮助作用就不是很大，这是我们对现有词类体系的基本看法。这样一个词类体系用到信息处理中还有很大的改进空间。

像代词、拟声词等等这些词类，它的功能都不是太清楚，还有汉语里面的助词这一类。我有个调侃的说法，助词的唯一共性恐怕是每个助词都很有个性，因为它们有这个“不是共性的共性”，所以就归到助词一类里面了，这样的范畴，信息量是很低的，对后面的分析能提供什么帮助呢？实际上提供不了太多帮助。

下面我们从构式视角来看词的分布，刚才那些名词、动词、形容词从语言学理论角度来看是基于一个给定的句法关系的框架来看它能分布还是不能分布在某个框架中，得出的一个结果，实际上那个句法框架覆盖不了这里提到的所谓的构式，比如“我胖我的”中的“胖”，一般定性是形容词，但是处理为形容词有什么意义呢？还有“啊来啊去，啊个没完”，构式里面有很多这样的东西。再比如汉语里面有一个非常特殊的构式是数字缩略构

式，比如“三不一没有”等等，传统的词类信息在这样的结构中就抓瞎了，说明语言符号是约定俗成的。

再比如朱德熙先生提出的“这本书的出版”的例子，语法学界争论这个“出版”到底是名词还是动词，从构式来看既不是名词也不是动词，而是在特定构式中的特定的一类词。另外，大家可以体会一下下面的例子，比如“你说的好，他说的不好”，“你说的好，做起来完全不是那么回事”，“你说的好，还不是真的好”，“你说的好，声调有些不太准确”。这里面的“好”，如果只是判断它是一个形容词，也解决不了这些句子的语义分析问题。从构式语法的角度来看这些超越了主谓宾等等结构关系。组合规则那一套怎么组合也组合不出一个好的分析结果，可能需要更多强调记忆的内容。

这里都是举一些例子，实际上我们也没有什么所谓的新观点和新方法，只是把一些现象跟大家做一些交流，传统的词类划分依赖句法结构关系，句法结构关系以往在语言学者的心目中认为是普适性的，甚至是跨语言普适性的。词类是怎么出来的呢？词类就是看这个词能不能做谓语，那个词能不能做宾语，你给了一个结构框架后，再看那些词能不能进入结构框架的某一个位置。词类也是普适性的，是针对所有词而言的一个词类划分。构式语法认为可能没有这种普适性的东西，形式和意义的配对是个别性的。从构式语法的视角看，一类词可能只是针对某一个具体的构

式来说的。这也有点类似语义学方面认识的发展，比如以往论元角色的分类，是针对所有动词而言的，现在以框架语义学为代表的语义表示，认为每个词在不同的框架中确定其概念义，在框架中确定其搭配的框架角色，所以它把概括性（普适性）的东西下放到具体的词例知识（个别性）。语言学上这些认识的变化，某种意义上看，跟 NLP 技术路线上的转变好像有一定的契合之处。以上是我的看法，谢谢大家！