

以“计算”的眼光看汉语语法研究的“本位”问题

詹卫东

zwd@pku.edu.cn 北京大学中文系, 北京, 100871

摘要 对于汉语语法研究中的“本位”问题, 本文的观点是: 并非只有作为“实体”的“语法单位”才能充当“本位”。“关系”是贯穿全部语言研究的, 任何语法体系都不可能回避的“基础概念”; 作为“实体”的各“语法单位”之间很难分出谁比谁更适合当“本位”; 不同“本位”的“语法体系”之间的差异可能并不像人们“想像”的那么大; 一个“本位”也并非只对应一个“语法体系”(反之亦然); 对满足“计算”的需求来说, 汉语语法研究真正需要解决的问题是如何从功能角度对各种语法成分进行分类, 以及如何构建语法成分之间的组合规则系统, 而不是选择哪一个“语法单位”作为“本位”。

关键词 汉语语法 语法单位 语法体系 本位 计算 规则

无论是自觉还是不自觉, 无论是直接还是间接, 从事汉语语法研究工作的人, 似乎永远离不开国外语法理论或者语法体系的影响。关于这种影响最常见的一个表述, 就是所谓“印欧语的眼光”。

似乎从《马氏文通》诞生开始, 汉语语法的研究者就一直在做着要摆脱这种“印欧语的眼光”, 换一种“眼光”来看汉语的努力。这些努力, 一方面表现为前辈和时贤所做的大量的扎实的汉语语法研究的**具体成果**, 另一方面也表现为各种具有“旗帜”意味的**汉语语法体系的名称**。这些名称包括¹: (1) 汉语“词本位”语法体系; (2) 汉语“句(成分)本位”语法体系; (3) 汉语“词组本位”语法体系; (4) 汉语“字本位”语法体系; (5) 汉语“小句中枢”语法体系; (6) 汉语“双本位”语法体系(具体说是“以词和词组为本位”的汉语语法体系)。

因为这些名称的存在, 从事汉语语法研究工作的人似乎就很难回避一个问题: 应该选择其中哪一个“本位”呢?

从逻辑上讲, 其实有多种可能性, 比如:

一个也不选, 即“无本位”的语法体系²(如果觉得“本位”并非是一个语法体系所必需的东西的话);

同时选其中的两个或三个或更多³(如果这些“本位”之间并非相互排斥的话);

另起炉灶, 看看是否还有创建一个新的“某某本位”语法体系的余地(如果觉得现有的这些“本位”都不尽如人意的话)。

本文尝试来回答上面这个问题⁴。下面展开具体的分析。先说“本位”, 再说“计算”, 最后合起来说从“计算”的角度对汉语语法研究提出了什么要求。

¹ 这里提及的汉语语法体系, 仅仅是跟“本位”这个说法关系比较密切的。汉语语法研究的成果中还有不少没有自称“某某本位”的语法体系, 这里没有列出。此外, 像徐杰(2001)提出的“原则本位”的语法学说, 因为跟中国大陆学者所说的“本位”相去甚远, 也没有列在本文评论范围之内。

² 邵敬敏(2003), 袁毓林(2003)都曾提到过“无本位”。金立鑫(2003)也持类似的看法。

³ 除马庆株(1998)外, 史有为(1991, 1998, 2000)提到过所谓的“移动本位”说, 也即“多本位”。

⁴ 我在做思考时, 试图从所谓“计算”的眼光来看这个“本位”问题, 但我不能确定这个“计算”的眼光是否摆脱了“印欧语的眼光”之后的一种纯粹的眼光, 还是其中多多少少仍然是杂揉着“印欧语的眼光”的眼光——一如本文开头就已经谈到的, 我们似乎永远也离不开“印欧语的眼光”。不过值得指出的是, 我的“印欧语”水平目前只是在能够不太困难地阅读专业文献的程度, 因此即便有“印欧语的眼光”的成分, 大概也是不多的。此外, 我本人的大部分研究工作是有关汉语句法的计算分析的, 算是积累了一些“计算”的实践经验, 因而基本可以保证“计算”的眼光应该是我审视上面这个问题时的主要眼光。

一 “本位”、“语法单位”、“语法体系”

这一节讨论的内容并不直接关乎“计算”，而是有关“本位”的两个很基本的方面（这两个方面在以往讨论“本位”的文献中尽管都有所提及，但实际上尚未得到充分的讨论），厘清这两个方面的问题，是进一步讨论的基础。

1.1 “本位”与“语法单位”

撇开概念的细节不论，单从名称上，就很容易看到，“本位”是某种“语法单位”。

邢福义（1996）在《汉语语法学》一书的“导言”中指出了汉语的七种语法实体：1. 语素；2. 词；3. 短语；4. 小句；5. 复句；6. 句群；7. 句子语气。其中前六种是“音节实体”，就是所谓的“语法单位”，最后一种是“非音节实体”，不属于“语法单位”。

除上面这些单位外，汉语研究者提到过的汉语语法单位还有“语素组”（郭锐，1996），“字”（徐通锵，1994a, 1994b, 2001），“辞”、“块”、“读”、“句”（徐通锵，1997），以及“语位”（鲁川，2002），“字位”（鲁川，2003），等等。

对于这些大大小小的“语法单位”，不同的研究者对它们的地位和重要性的认识就可能存在很大的差异，这种差异最后就落在了“本位”这个概念上。如果一个语法体系认为“字”这个单位特别重要，就是“字本位”语法体系，如果一个语法体系认为“词组”这个单位特别重要，就是“词组本位”语法体系。如此这般。

然而，这里的问题是：（1）是什么决定了一个“语法单位”比其他的“语法单位”**更重要**（或者说**更基本**）？（2）对于一个语法体系来说，“语法单位”真的是**最重要**（**最基本**）的吗？

其实上面两个问题的顺序应该是先问第二个问题，再问第一个问题，因为如果对第二个问题的回答是否定的话，第一个问题似乎就不必再问了。但是由于以往的讨论大多集中在第一个问题上，所以这里也还是从讨论这个问题开始。

要回答第一个问题，就得定出决定一个“语法单位”重要与否的标准来。但我们觉得这样的标准实在是很难定。这里暂且借用李宇明（1997）的标准来加以讨论。

李宇明（1997）认为：语法的‘本位’或者‘中枢’，最主要的是要给语法研究寻找到一个稳固的立足点和视野广阔、对语法现象看得真切细致的观察站。而能够做“语法研究的立足点”的单位应该具备两个条件：（a）容易识别；（b）通过对这种单位的研究，可以波及到一切语法单位及语法规则。能够做“语法研究的观察站”的单位应该具备四个条件：（c）语法因素齐全；（d）是各种语法单位的联络中心；（e）具有控他性；（f）能发现在其他单位上不易发现或不能发现的问题。

按照上述标准，依次检验“字”“词”“词组”“小句”“句子”等等单位，就可以排出一个重要程度由大到小的顺序来。李宇明（1997）的综合排序结果是“小句”最重要，其次是“词组”，最不重要的是“字”⁵。

我们的认识是：为什么选这些标准来排序，本身似乎缺乏论证，同时，用这些标准去检查一个“语法单位”的重要程度也缺乏可操作性，因而排出的序可能是带有较多主观性的。拿条件（a）来说，一个语法单位是否“容易识别”可能涉及到很多复杂的因素，现在还不知道如何实验才能证明“小句”比“词组”更容易识别，也不知道如何证明“字”比“词组”更容易识别（“字本位”里的“字”可不是简单的“汉字”，即便可以证明“汉字”很容易“识

⁵ 有意思的是，在李宇明（1997）的三次分项排序中，“小句”总是与“词组”紧邻，并“刚好”优于“词组”（对“小句”和“词组”的关系，本文1.2小节还有更进一步的发挥）。在最后的综合排序中，顺序当然也是如此。

别”，也证明不了“字本位”的“字”很容易“识别”)，如果不做实验，纯粹在理论上来“识别”的话，可能得不出任何结论。以“词组”的识别为例，在“词组”本位语法体系的极端支持者眼里，一篇文章也不过是一个“词组”而已，难道不是吗？哪篇文章不是“一个词一个词组合起来”的？因此“词组”的识别根本不会有任何障碍。也许有人说这是抬杠，但从理论上说，当“句子不过是独立的词组而已”成为一个理论信条之后，就无法否认“篇章也不过是更大的独立的词组而已”（至少在纯粹的抽象的结构层面是如此）。再说条件（b），这个条件同样缺乏可操作性的判别标准，实际上任何一个语法单位都可以波及到一切语法单位及语法规则，或者直接波及，或者间接波及。而且，如果研究者甲构建的语法体系是间接波及的话，比研究者甲更高明的研究者乙很可能会想出办法来让它们直接波及。说到底，一个语法单位如何跟其他语法单位发生关系，这是语法体系构建者的理论能力决定的，而不是“所谓的客观的语言事实”决定的。

其余的四个条件这里不一一讨论了。下面仅举一个实例来说明，在李宇明（1997）的综合排序中最不重要的“字”有时候也折射出有趣的语法现象，似乎在“控制”着人们对语言的理解。

	A	——	B
例 1	“你怎么来了”		“你怎么来的”
例 2	“那本书你借谁了”		“那本书你借谁的”

例 1A 和例 1B，例 2A 和例 2B 可见的差别是表层语言形式上 A 组以“了”字结尾，B 组以“的”字结尾。例 1A 中的“怎么”应该理解为“为什么”，例 1B 中的“怎么”应该理解为“如何”；例 2A 中的“借”应该理解为“借出”，例 2B 中的“借”应该理解为“借入”。是“了”和“的”造成了这种差异吗？它们又是如何造成这种词义差异的呢？我们目前还无法看出，回答这些问题，跟从哪个“观察站”来观察有什么直接的联系。

综上所述，我们对第一个问题的看法是：很难确定哪个“语法单位”最重要。我们看不出“字”比“词组”更重要，也看不出“词组”比“字”更重要，其他单位亦然。

下面讨论第二个问题：对于构建汉语语法体系来说，“语法单位”是最重要或者说是最基本的吗？

对这个问题，我们现在也还没有明确的经过严格论证的答案。不过在学习和研究工作中形成了一点看法，写在这里就教于方家。那就是“语法单位”之间的关系可能比“语法单位”本身（**实体**）更重要。决定一个语法体系整体面貌的是什么？是语法单位的类（category），是建立在各种语法单位类之上的规则。尽管名称不同，但“语法单位”无非都是大大小小的语言成分而已（甲称之为“词组”的，乙可能称之为“字组”），语法体系的差别主要表现在对这些单位的分类上，比如甲语法体系把“词”分成 15 类，乙语法体系可能分成 20 类，丙语法体系可能没有“词”这个概念，而是“字”或“字组”的概念，那么同样也需要把“字”或“字组”分成若干类。类的多少，如何分类，是体系间最基本的差异。那么，决定分类的又是什么呢？是“关系”，或者更具体的说，是“结构关系”⁶。以朱德熙（1982）所代表的“词组”本位语法体系为例，体系的根基实际上是“主谓结构”“定中结构”“状中结构”“述宾结构”“述补结构”“联合结构”，……等等基本结构关系，词类是在这些结构关系基础上定义的结果。离开结构关系，整个语法体系无从谈起。作为一个“词组”本位语法体系的研究者，可以用这套结构关系的眼光去看“词”的内部构造，也可以用这个眼光去看“词组”，看“句子”。考察的对象——语法单位（或语言实体）——可以变大变小，不变的是“关系”。同样的，当一个人主张“字”本位语法体系，或者“小句”本位语法体系时，最基础

⁶ 我们的理解是：“结构”是人们对各种“关系”加以系统化之后所形成的认识。Piaget（1968）在《结构主义》一书的结尾中指出，“……结构首先是，并且主要是一束转换关系……”。

的工作仍然是假设一套基本的“结构关系”，然后用这个“关系”去观察大大小小的“实体”。

事实上，作为研究对象，“关系”的地位高于“实体”这种认识在数学界也普遍存在着。20世纪的数学大师，德国数学家希尔伯特（David Hilbert）在《几何基础》（*Foundations of Geometry*, 1899）中指出，“虽然用的概念是点、线、面等，但如果它们遵从所涉及的公理的话，可以是啤酒杯、椅子或任何物体”⁷。而公理实际上就是一组抽象“关系”的说明。法国数学家彭加勒（Henri Poincaré）在论及数学的抽象性时也说“数学是一门为不同事物起相同名字的艺术”。我们理解，这实际上也就是对事物（实体）间的各种“关系”的抽象。

也许人们的直觉是先有“实体”，然后才有“关系”。但很可能这只是个错觉。如果“先”和“后”不仅仅是在绝对的时间意义上说的话，我们“构造”的语言学理论或体系，也就完全可以先有“关系”，然后才有“实体”。或者说，“实体”只能在“关系”的基础上被认识，离开了“关系”，“实体”也就无从谈起。

如果以上认识可以被接受的话，那么一个新的“本位”就应运而生了，即“结构关系本位”的语法体系，或者简单说，就是“结构本位”的语法体系。问题是，提出这样一个新的“本位”，又有多大的意义呢？

从积极的方面说，研究者选择不同的“语法单位”作为“本位”，可以从多个角度来审视汉语的语法现象，应该是有助于加深我们对语法现象的认识的。徐杰（2001，p1）、袁毓林（2003，p56）等学者都表达过类似的看法；但从消极的方面说，强调某一种语法单位的“特殊”地位，容易厚此薄彼，甚至将语法研究带入认识上的“盲区”或者“误区”。陆丙甫（1992）、邵敬敏（2003）就表示过这样的担忧。如果大家陷入争论这些语法单位的地位谁重要的泥潭中，大概就会像马季的群口相声《五官争功》⁸所表现的那样，其实谁离了谁都不好过。从这点上说，强调观察语法现象的立足点的差异而淡化“本位”的争论也许更好一些。选择站在什么地方看“语言学风景”，就好象一个人选择到什么地方旅游一样，应该只是一个个人偏好的问题。

下面这个小节的讨论试图说明，即便是“本位”不同，也不一定就意味着“语法体系”不同。

1.2 “本位”与“语法体系”

以往人们在谈到“本位”的时候，似乎有一个潜台词，那就是“本位”与“语法体系”之间是不可分割的联系在一起的。某个“本位”就决定了某个“语法体系”。但果真如此吗？

我们认为，这里也至少有两个问题值得反思：（1）不同的“本位”是否意味着本质上不同的“语法体系”？（2）一个“本位”是否意味着只有一个“语法体系”（反之亦然）？

对于这两个比较“宏大”的问题，要全面地回答，不是一篇文章可以胜任的，因此下面的讨论多少有些“点到为止”的意味。我们不妨把这两个大问题转成两个具体一点的“小”问题来问，并简单阐述我们的看法。

（1）“词组本位”的语法体系跟“小句中核”的语法体系有本质（很大）的区别吗⁹？

我们的看法是没有多少本质的区别。当然，“说有易，说无难”，要说没有本质的区别实际上是很难的。而且，到底以什么标准来判别两个语法体系是否存在本质区别，也是一个极难的问题。大多数表面名称（提法、表述）的差异也许会误导人们的感受。同时，相同的

⁷ 转引自 Morris Kline（1980）著，李宏魁（1997）译，p189。

⁸ 该相声是在 1987 年春节联欢晚会上跟观众见面的，后来发表于《人民文学》1987 年第 3 期。

⁹ 这个问题同样适合去问其他“本位”的语法体系。朱德熙先生在《语法答问》第六章讨论“汉语语法体系”时指出过，“语法体系在很大程度上是指语法事实和语法规律的表述系统。……并不是企图揭示语法构造的本质和规律的理论。”我们赞同朱先生的看法，并认为这个看法在讨论各种“本位”的“语法体系”之间的区别时有很强的指导意义。

名称在不同的体系中也也许又有完全不同的理解，从而造成纠缠不清的种种争端，让人觉得差异是如此的巨大。

这里仅以“词组本位”语法体系的代表作《语法讲义》与“小句中枢”语法体系的代表作《汉语语法学》来做简单对比。在我们看来，二者只有“量”上的区别，而没有“质”的区别。《语法讲义》把大量的篇幅给了“词组”，只给“复句”留了一章的“地盘”（但我们不知道朱先生如果可以继续给我们讲汉语语法，是否会考虑把“复句”的内容扩展到可以跟“词组”相当的程度¹⁰）；在《汉语语法学》中，有关“词组”的篇幅大约是有关“复句”的篇幅的一倍。显然，跟《语法讲义》最大的不同是，有关“复句”的讨论大大加强了。然而这种不同，只是“量”的不同。是“多”和“少”的差别，而不是“有”和“无”的差别。

此外，“小句”去掉“语气”就是“词组”，“词组”加上“语气”就是“小句”。如果不论“语气”，“小句”跟“词组”就没有什么太大的差别了。换言之，如果只关心“结构”问题，“词组本位”和“小句中枢”几乎是一样的。即便要加上“语气”，二者也还是没有多大差别，因为无论是“词组本位”，还是“小句中枢”，都没有反对在“表达”层面进行汉语的研究，也就是说，“语气”同样也兼容于两家语法体系。在下面这些例子中，划线的部分是“复句”，但被包含在“词组”中了，因而也是“词组”。这些“词组”在说出来的时候，当然也可以有各种各样的“语气”（比如“嘲讽”“愤怒”“焦虑”等等——如果这些算“语气”的话）。

例3 你可以藏在一个你看得见他，可是他看不见你的地方。（复句占据“定语”位置）

例4 我就不相信如果你跟他离了婚，他还敢来问你要钱。（复句占据“宾语”位置）

例5 你有时候冷，有时候热，就是非典的症状之一。（复句占据“主语”位置）

从这些例子不难看出，“小句”也好，“复句”也好，在“结构”层面，都是“词组”。反过来，不管什么“词组”，只要带上了“语气”，在“表达”层面，就可以将它确定为“小句”（极端一点，像“从北京”，“被张三”这样的所谓“粘着性词组”，在实际对话中就未必不能带上“语气”而成为“独立”的“小句”）。从这点上说，“小句”和“词组”之间几乎可以划等号了（当然这里可以划等号的“小句”和“词组”不是日常语言中的“小句”和“词组”，而是一个理论体系中被“科学”定义了的“术语”）。大概也正因为如此，在李宇明（1997）的排序中，“小句”和“词组”的地位是如此的接近，以至于几乎可以合在一个位置上了。

（2）以“词组”为“本位”是否意味着只有一个“词组本位语法体系”？

虽然一个“语法体系”通常有一些具体的“语法著作”作为基础，但“语法体系”这个概念的所指本身还是非常抽象的。人们在谈到“词组本位语法体系”的时候，大都以《语法讲义》作为代表。但我们认为，“词组本位语法体系”应该是多种可能性的，并非只此一家，至少理论上如此。以对“词类”的处理为例，温锁林（2001）全面反对“词组”本位语法体系，认为“汉语词的兼类现象普遍又自由”，反对朱德熙先生提出的“汉语词类的多功能性”。这就比较容易给人造成一种印象，“词组本位语法体系”认同词的“多功能说”，反对词的“兼类说”。这里我们不展开评述这两种词类观。只是觉得有必要指出，以“词组”为本位建立语法体系，既可以像朱德熙先生那样，反对词的“兼类”，也可以像温锁林先生那样，认同词的“兼类”。换言之，以“词组”为本位构建汉语语法体系，实际上并不就一定意味着要如何来处理“词类”问题。这二者之间没有必然的联系（陈小荷，1998；胡明扬，1996）。推广开去，对其他一些语法问题的认识和处理，也是同样的道理。理论上应该完全具有这样的可能性：甲和乙同举“词组”本位的大旗，但各自构建的具体的“语法体系”的

¹⁰ 注意：在“词组”本位语法体系的眼里，“复句”不过是一个长一点（大一点）的“词组”。这里把“复句”跟“词组”对举，是按照一般对“词组”和“复句”的认识来说的，实际上“复句”只能跟其他“非复句词组”对举。从这里也不难体会到，当“语法体系”不同时，人们对同一个名称（比如“复句”）的理解，差异何止千里。

面貌却不一样。

二 对语言进行“计算”

什么是对语言进行“计算”？这可以从很多角度进行理解，有兴趣的读者可以参看俞士汶等（2003）以及詹卫东（2000）的介绍。这里仅举一个例子略做说明，读者需要关心的不是“计算”的“技术细节”，而是对语言进行“计算”的“基本模式”。

下面这三个例子都是以所谓层次分析法来分析汉语结构的实例。

<p>例 6</p> <p>老公 丢了 钱包</p> <p> 主 谓 </p> <p> 述 宾 </p>	<p>例 7</p> <p>老公 丢了 钱包 也丢了</p> <p> 联 合 </p> <p> 主 谓 主 谓 </p>	<p>例 8</p> <p>老公 丢了 钱包 也 丢了 人</p> <p> 主 谓 </p> <p> 联 合 </p> <p> 述 宾 状 中 </p>
--	--	--

无论采用哪种“本位”的语法体系，相信都不会否认以下基本的“语言事实”：这三个例子中包含一个共同的符号串¹¹“老公丢了钱包”，并且这三个例子都是汉语中可以说的符号串（一般人们称之为“句子”）。听到例 6 时，我们知道它的“意思”是“某人的老公的钱包丢了”；听到例 7 时，我们知道它的“意思”是“某人的老公丢了，某人的钱包丢了”；听到例 8 时，我们知道它的“意思”是“某人的老公的钱包丢了，并且某人的老公丢人了”。

这里的问题是：我们是怎么知道的？

这是对人类的智慧构成极大挑战的一个问题（因为这是对“智慧”本身的发问）。要回答这个问题目前还非常困难。以下的讨论应该看作是我们试图接近答案所做的努力。为便于说明，这里把“知道”的含义简化为：如果计算机也能像人一样进行层次分析，就算它也“知道”了这些例子所要表达的“意思”。于是问题变成：这些层次是怎么划出来的？

下面给出一种划层次（也就是对语言进行“计算”）的办法。限于篇幅，仅以例 6 为例来说明（当然，这套办法也适合对例 7，例 8 以及其他任意有限长度字符串的层次分析）。

表 1 是一套“划层次”的规则（以下简称 G1），其中 S, dj, …u 等字母是“非终结符”（non-terminal）；“张三”“钱包”…“了”等是“终结符”；“->”可以解释为“变成”或“由……组成”。“非终结符”可以看作是对“终结符”的一种分类。“dj”“vp”等等代表什么并不重要，可以看作是纯粹的“记号”（如果你愿意，可以把这些记号换成“⊙”或“♥”什么的）。重要的是它们之间的关系，比如规则(5)所显示的关系是：“vp”可以被换成“vp np”。正是这种抽象的关系，决定了语言的结构是“这样”，而不是“那样”。决定了一个字符串（终结符串）是“合法的”（grammatical），或者是“不合法的”（ungrammatical）¹²。

(1) S -> dj	(2) S -> dj dj	(3) dj -> np vp	(4) np -> n
(5) vp -> vp np	(6) vp -> vp vp	(7) vp -> v u	(8) n -> 老公
(9) n -> 钱包	(10) n -> 人	(11) v -> 丢	(12) u -> 了

表 1 句法结构规则示例

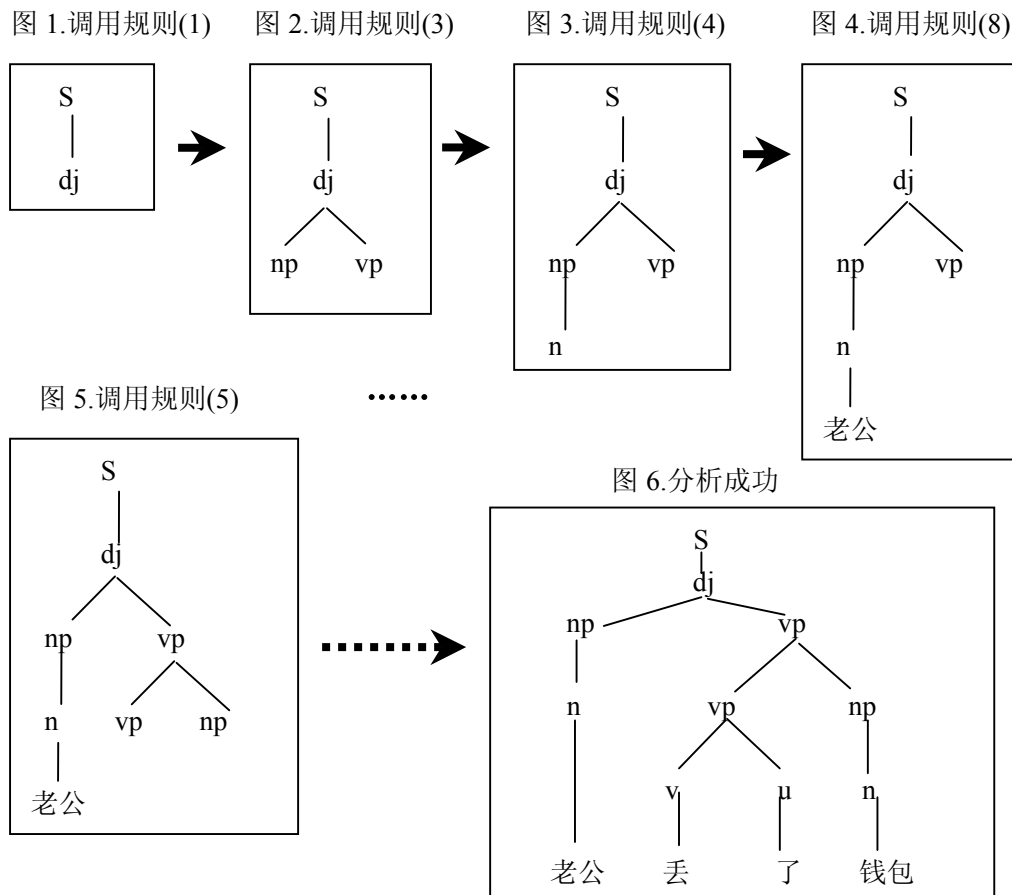
在 G1 的基础上，我们还需要定义一套使用规则的具体办法。这就是所谓的分析算法

¹¹ 我们不说这是“句子”，以免引起不必要的争论。“符号串”就是一串“符号”。“符号”在语言学中，就像几何学中的“点”一样，无法定义。

¹² “合法”的句子可以定义为“根据规则能划出层次”的句子；“不合法”的句子就是“根据规则划不出层次”的句子。

(parsing algorithm)。在计算机分析语言的实践中，人们已经发展出多种针对 G1 这样的规则的分析算法 (James Allen, 1995)。这里介绍自顶向下 (top-down) 分析算法，该算法的基本思路是：从 S 开始 (S 可以理解为“START/开始符”)，不断调用规则，将短的符号串转换成更长的符号串，直到这个符号串与当前待分析的字符串完全匹配为止。如果完全匹配，则分析成功，得到当前待分析字符串的“层次结构”；如果所有规则都调用后也无法产生与待分析字符串完全匹配的符号串，则分析失败，可判定当前字符串“不合法”。

限于篇幅，这里回避了算法的诸多技术细节问题，比如如何决定当前应该调用哪条规则，如何“不断”调用规则，如何“记住”已经调用过的规则等等。下面图 1 到图 6 显示了以“自顶向下”分析算法分析例 6 的过程，不难看出，原则是每次只调用一条规则进行“变换”操作 (图 1 到图 5 即是如此。只是为节省篇幅从图 5 到图 6 一下子“跳过”了多个操作步骤)。



计算语言学的假设¹³是：人在“理解”例 6 时经过了上述步骤 (这些步骤是在一瞬间完成的)，这里以类似“照相”的方式把这些步骤一帧一帧地记录了下来。这些步骤是完全“机械”的，用褒义的说法，就是完全“客观”的，无论是张三还是李四，在同一套规则基础上，按照这种方法来分析一个“有限长度的字符串”，得到的结果总是一样的 (最终结果跟张三和李四的个人喜好没有关系)，有兴趣的读者可以按照类似的方法来分析例 7 和例 8 (不管由人来分析还是由计算机来分析，结果也都是完全一样的)，都可以“计算”出它们的层次来。也可以试试用 G1 分析“老公丢了钱包人”“老公钱包丢了”等等“字符串”，结果一定都是“失败”，即分析不出层次，因而这些“字符串”将被 G1 “认为”是“不合法”的。

从“计算”的眼光看，“语法”(或“语法体系”)不是别的，就是类似 G1 的这样一套规则。这样一套规则决定一个“字符串”是否能被“识别”(“理解”)，或者说是否能划出层

¹³ 当然这是很笼统的说法，人们完全可以把对语言的“计算”假设成另外的模型，参见俞士汶等 (2003)。

次来。这样一种“面向计算”的“语法观”有两个值得强调的要点：

(1)“语法”的作用，以及它如何起作用，都非常清楚。“语法”可以用来“识别”一个字符串是不是一个“合法”的单位，并且在“识别”的同时，“语法”还能揭示该字符串的“层次”。任何人（或计算机）来“操作”一套“语法”，都要按照一定的步骤来“机械”地进行，操作结果都必然是一样的。以 G1 为例，它无法告诉我们“丢钱包”是“不如意的事情”。反过来，人们也不应该这样来要求它；它可以告诉我们“老公丢了钱包”是汉语中“合法”的句子，以及这个句子的结构层次是什么。无论谁来用 G1 对“老公丢了钱包”这个字符串进行判别，都是同样的操作过程，结果也都相同。

(2)“语法”完全可以，而且应该看作是“想像”的产物¹⁴。像“S, dj, vp, …, n, v, …”等“非终结符”是完全人造的概念¹⁵，这些范畴是否“客观存在”并不重要，重要的是我们通过“定义”它们之间的关系，来模拟人为什么“知道”一个符号串的“层次”（也就是一种“意义”）。换言之，为了达到目的，可以“各择手段”¹⁶。张三的规则和李四的规则可以完全不同，各玩各的花样¹⁷，当然真理仍然只有一个，那就是不管如何假设规则，最后都得来分析无限多的“字符串”，分析结果跟人的“语感”越接近，就说明假设的规则可能越好；此外，尽管具体的规则不同，但规则的组织方式和原则从根本上来说是一样的。

三 面向“计算”的汉语语法系统 —— 需要研究者解决哪些问题？

如果把“语法”比喻成一个对“字符串”进行 X 光透视（照相）的装置（device），那么 G1 无疑只是一个“模型”（prototype），虽然它可以展示这种装置的基本工作原理，但确实太简陋了，很多活儿干不了。比如上面提到的“老公钱包丢了”这样的字符串，在人看来是“合法”的，有“层次”的，但 G1 却无法“识别”它（因而认为它是“不合法”的）。研究者的主要任务就是把像 G1 这样的“语法”完善和丰富起来。为了能够识别更多的“字符串”，研究者需要构造出更好的语法系统 G₂, G₃, …, G_n 来。

显然，在构造语法系统的过程中，有两个核心问题需要研究者解决：（1）对一个语言来说（比如汉语），有哪些“非终结符”？（2）这些“非终结符”如何组织成一个规则系统？要回答这两个问题，人们还需要做很多具体的研究工作。这里我们简要讨论应该如何认识非终结符的性质。这是直接关系到如何对“语法单位”进行分类的问题。

仍以对“老公丢了钱包”这个字符串的层次分析为例，例 6 中给出的是一般现代汉语语法教科书中常见的分析结果，图 6 则是计算语言学中习惯的表示方式，为清晰起见，我们把两个层次分析的结果重新表示如下（例 6 对应到图 7；图 6 则对应到图 8）

¹⁴ 比“凭空想像”好听些的说法是所谓“公理化方法”——规则就是公理。参见冯志伟（1991），第 4 章。

¹⁵ 不造这些非终结符同样可以“玩语法游戏”，比如“链语法”（Link Grammar）就完全基于终结符“词”（word-based）来构造语法系统，有兴趣的读者可以访问 <http://www.link.cs.cmu.edu/link/> 了解详情。

¹⁶ 这里有必要提到胡明扬（2002）在评价《新著国语法》时的说法。胡先生称“《新著国语法》……是唯一一部经过了语言教学和自然语言计算机处理的检验和考验的著作”。我们不知道“经过检验和考验”的具体标准是什么，但只要看一下吴蔚天、罗建林（1994）第 4 章第 5 节和第 5 章第三节（3.2）就不难发现：（1）吴、罗的系统综合了各家的语法观念（也就是，凡是我认为有用的就拿来，这是非常典型的“工程师原则”）；（2）他们采用的语法规则系统从形式上跟本文 G1 所示的这种规则系统（Chomsky 所谓的 2 型文法）是完全一样的，差别只是在选用非终结符时有所不同（本文第三节将对此展开讨论）。也就是说，如果吴、罗的系统真的经过了检验和考验，也不能说是《新著国语法》经过了自然语言计算机处理的检验和检验，只能说汉语跟其他语言一样，也可以用 Chomsky 的短语结构文法来描述。我们没有丝毫贬低《新著国语法》的地位的意思，只是想说明，设计自然语言分析系统的工程师在干活的时候，从来都是遵循“为达目的，不择手段（或各种手段都可以用）”的原则。就计算语言学目前的发展水平来说，要通过所谓实践上的成功来证明某种语法体系的成功，还是很困难的一件事情（潜台词：计算机的水平还没有高到可以当裁判的程度）。

¹⁷ 从某种意义上说，“语法”也“平民化”了，它不是高高在上的“法”，只是一套游戏规则而已。

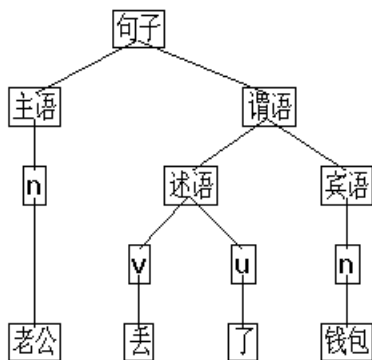


图 7

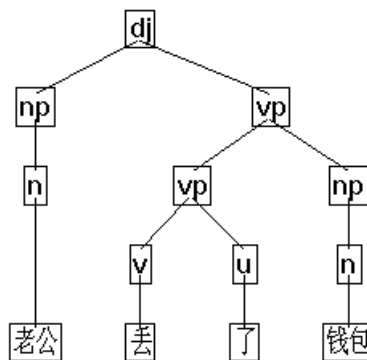


图 8

很显然，图 7 跟图 8 的结构层次完全相同，差别是“树结构”上有些结点 (node) 的标签 (label)，即“非终结符”不同。比如图 7 的“谓语”对应为图 8 的“vp”。前文已经说过，“非终结符”可以看作是对“终结符”的分类。那么，图 7 和图 8 实际上就反映了人们对语言单位的分类的差异。如果采用图 7 这种分类，规则就是“句子 → 主语 谓语”，“宾语 → n”这样的形式 (吴蔚天、罗建林 1994 采用的正是这样的规则形式，当然细节仍有不少差别)。不妨把这种规则系统记作 G2。值得注意的是，当代形式语法学著作中的“非终结符”系统均是图 8 所对应的 G1，而很难见到采用 G2 这种系统的 (Mark Baltin & Chris Collins, 2000)。为什么呢？

G1 和 G2 的差别不是“名称”上的 (前文已经说过，如果愿意，“主语”或“np”这些名称也可以被换作“西红柿”或者“啤酒杯”什么的)。差别在于不同的分类所反映出的对“关系”认识的差异。“vp”等“非终结符”是所谓功能分类 (这里的“功能”仅指分布位置) 的结果，这种分类刻画了一个“实体”向外跟其他实体发生关系的性质，反映的是“外部”关系；“谓语”等“非终结符”则是结构分类，这是向内观察一个“实体”的构成所得到的分类结果，反映的是“内部”关系。很显然，G2 混用了两种分类所得到的范畴 (“谓语”是结构分类，“v”是功能分类)，因此 G2 的非终结符系统是“非匀质的” (unhomogeneous)；而 G1 的非终结符系统是“匀质的” (homogeneous)。

如果进一步考虑 G2 为何是“非匀质的”，不难发现，G2 中的“结构类”基本是对“词”这个层次之上 (比“词”更大) 的单位的刻画，而对“词”的刻画 (分类) 只能采用“功能类”的办法。换句话说，“功能类”可以不依赖“结构类”独立构造规则系统，而“结构类”必然要依赖“功能类”才能构造出完整的规则系统。再比如下面这些例子中加下划线的“词”，如果从“内部构成” (“结构类”) 来看，分属不同的类型，但从“外部功能” (“功能类”) 来看，属于同一类 (即一般说的“名词”)。

- 一 种 思想 (“思想”的结构类型：并列)
- 一 本 专著 (“专著”的结构类型：偏正)
- 一 位 司机 (“司机”的结构类型：支配)
- 一 个 房间 (“房间”的结构类型：主从)

显然，只有从“功能类”的角度来刻画“词”这个语言单位，才“方便”我们去说明“词”之间的组合 (层次)。对“复合词”是如此，对“单纯词”就更是如此，因为“单纯词”已经没有“内部构成”可言，只能论“功能”了。对于“词”的这种认识，完全可以，而且必需拓展到其他语言单位 (比如“词组”和“小句”)。如果不这样，就出现像 G2 那样的系统，在“词”层次上采用“功能类”，在“词”之上的层次采用“结构类”，形成一个“非匀质的”的“非终结符”系统。尽管这样的系统在一定范围内也可以在计算机上处理自然语言，但理解起来非常别扭，特别是表达语言结构的递归性非常困难。因此对各级语言单位进行全面的“功能分类”研究，是非常有必要的。陆俭明 (1993) 指出“词的语法功能”是任何语言划分

词类的本质依据”，实际上，根据以上简要的分析不难看出，这个说法还可以进一步扩展。从“计算”的眼光来看，语言成分的分类，无论词、词组还是小句，都必需是“功能类”。如果以“字”或“字组”来构造语法系统的话，同样也得按照“功能”来分类，舍此没有别的可能。无论给“语法单位”起什么样的“名称”，所有的“语法单位”实际上可以分为三级：**(1) 最小的语法单位：这种语法单位没有结构类，只有功能类（最小的语法单位可以定义为没有内部结构的单位）。****(2) 最大的语法单位：这种语法单位没有功能类，只有结构类。因为对最大的单位来说，压根就没有“功能”（“分布位置”）这个概念，而只有“结构”的概念。****(3) 处在最小和最大之间的语法单位：既有功能类，也有结构类。但如果从构造规则的角度看，只有功能分类，才是建立规则系统的合理选择¹⁸。**

综上所述，面向“计算”的汉语语言研究，首先就要求人们从功能的角度出发来对语法单位进行分类。目前有关汉语词语功能分类的理论和实践都有了较多的研究成果，而对汉语词组（短语）、小句的功能分类研究却还远远不能满足“计算”的需要¹⁹。比如《汉语语法学》对“复句”这种“词组”进行了大量富于创见的“内部结构”研究，但却没有从“向外组合”的“功能”角度对“复句”进行分类和研究。而计算机要分析像上文例 3-5 那样的包含了“复句词组”的“字符串”，首先就需要掌握有关“什么复句能出现在什么句法位置”的知识，即“复句”的功能分类知识。

四 余论

本文以“印欧语的眼光”开始我们的话题，最后还是以有关这种“眼光”的讨论来做结。到底是用“印欧语的眼光”看汉语，还是用“汉语自己的眼光”看汉语，似乎是语法学界从未停止过争论的一个问题。有趣的是，甲认为自己是用“汉语的眼光”看“汉语的特点”，但往往会被乙批评只不过是“在用‘另一种印欧语的眼光’看汉语（徐通锵，1994b）”。曾经被认为是“汉语特点”的“现象”，在更广阔的理论视野下，却发现其实是“语言的共性”（程工，1999，p.126）。李宇明（1997）说得好，“近百年的语法史一直在不断争论，是强调汉语语法的特点，还是强调人类语法的共性。这种争论看来针锋相对，但采用的却是一个共同的思维模式，即把特点的研究与共性的研究对立起来。而事实很可能是另外的一种情况：只有对特点的深入研究，才能发现最具语法共性的东西；只有用共性的眼光来审视特点，才能把特点看得更为真切。”

从“计算”的眼光来看汉语语法研究的问题，我们看到更多的是汉语与其他语言的共性，而非差异，我们常常觉得看问题的角度不够宽广，“眼光”不够丰富，而非已经具备能用“印欧语的眼光”看问题的能力。在“计算”汉语的实践中，我们感觉到有很多具体的问题需要解决，而且往往涉及到所有的“语法单位”，可以从任何“本位”出发进行研究。研究能否取得成果，重点不在于选择什么“本位”，而在于我们对语言成分本身认识到什么程度（比如前文例 1 和例 2 到底显示出“了”和“的”的差异是什么）。

从满足“计算”的需要来说，不管是“印欧语的眼光”，还是“汉语的眼光”，能解决具体问题的“眼光”就是“好眼光”。此外，无论是用哪种眼光来看汉语，为了“计算”的目的构造汉语语法系统，必需强调“可操作性”，即语法体系中的概念务必是严格而清晰的。

¹⁸ 必需指出的是，“功能”和“结构”绝对不是对立的。我们说“功能分类”很重要，并不意味着“结构”就不重要了。事实上，功能分类必然要建立在“结构关系”的基础上（参见詹卫东，2000）。也正因为如此，我们说“结构关系”是贯穿所有“语法实体”，任何语法体系都无法回避的“基础概念”。

¹⁹ 实际上也不能满足汉语语法本体研究的需要。国外形式学派语法理论的主要表述工具就是树结构和短语的功能范畴。现代形式语法理论的主要研究成果也往往表现为提出新的短语功能范畴，比如用 CP（补足语短语）和 IP（屈折特征短语）取代原来的 S（句子），由一般的 NP（名词性短语）发展出 DP（限定性短语），等等。每次提出新的短语功能范畴，都标志着对句法结构的认识的深化（Mark Baltin & Chris Collins, 2000）。

“柔性”作为一种研究的技巧或许值得称赏，但若作为一种研究的结果却毫无“计算”的价值。

参考文献

Charles C.Fries (1952) *The Structure of English : An Introduction to The Construction of English Sentences*, Harcourt, Brace And Company, INC.

James Allen, 1995, *Natural Language Understanding*, Second Edition, The Benjamin/Cummings Publishing Company, Inc..

Jean Piaget (1968) 著, *Le Structuralisme*, 倪连生、王琳译《结构主义》, 商务印书馆 1996 年版。

Mark Baltin & Chris Collins, 2000, *The Handbook of Contemporary Syntactic Theory*, Blackwell Publishers Ltd, 2000, 外语教学与研究出版社 2001 年原版引进。

Morris Kline (1980) *Mathematics: The Loss of Certainty*, 中译本: 李宏魁《数学: 确定性的丧失》, 湖南科学技术出版社 1997 年版, 第一推动丛书。

Noam Chomsky (1956) *Three Models for The Description of Language*, IRE Transactions on Information Theory, IT-2: pp.113--124, *Proceedings of the symposium on information theory*.

陈小荷 (1998) 《从自动句法分析角度看汉语词类问题》, 发表于' 98 现代汉语语法学国际学术会议, 北京大学. 1998 年 8 月, 载陆俭明主编《面临新世纪挑战的现代汉语语法研究》, 山东教育出版社 2000 年版。

程工 (1999) 《语言共性论》, 上海外语教育出版社。

冯志伟 (1991) 《数学与语言》, 湖南教育出版社 1991 年版。

郭锐 (1996) 《汉语语法单位及其相互关系》, 《汉语学习》1996 年第 1 期。

胡明杨 (1996) 主编《词类问题考察》, 北京语言文化大学出版社。

胡明扬 (2002) 《现代汉语语法的开创性著作——〈新著国语文法〉的再认识和再评价》, 《语言科学》2002 年第 1 卷第 1 期。

胡作玄 (1985) 《第三次数学危机》, 四川人民出版社 1985 年版, 走向未来丛书。

金立鑫、白水振 (2003) 《现代汉语语法特点和汉语语法研究的本位观》, 《汉语学习》2003 年第 5 期。

黎锦熙 (1955) 《新著国语文法》, 商务印书馆 1955 年版。

李宇明 (1997) 《汉语语法“本位”论评 —— 兼评邢福义“小句中枢说”》, 《世界汉语教学》1997 年第 1 期。

刘群、俞士汶 (1998) 《汉英机器翻译的难点分析》, 载黄昌宁主编《1998 中文信息处理国际会议论文集》, pp507-514, 清华大学出版社。

鲁川 (2002) 《汉语的“语位”》, 《语言教学与研究》2002 年第 4 期。

鲁川 (2003) 《立足汉语实际的信息语言学 —— 确立汉语基元单位和构建角色序列模式》, 《语言科学》2003 年第 4 期。

陆丙甫 (1992) 《从“短语本位”看“词”的地位和判别 —— 谈必需给“词”在句法分析中留有一席之地》, 《汉语学习》1992 年第 2 期。

陆俭明 (1993) 《八十年代中国语法研究》, 商务印书馆 1993 年版。

陆俭明、郭锐 (1998) 《现代汉语语法研究所面临的挑战》, 发表于' 98 现代汉语语法学国际学术会议, 北京大学. 1998 年 8 月, 载陆俭明主编《面临新世纪挑战的现代汉语语法研究》, 山东教育出版社 2000 年版。

马庆株 (1997) 《词组的研究》, 《语言教学与研究》1997 年第 4 期。

- 马庆株 (1998) 《结构、语义、表达研究琐议》，《中国语文》1998 年第 3 期。
- 邵敬敏 (2003) 《关于新世纪汉语语法研究的几点思考》，《语言科学》2003 年第 4 期。
- 史有为 (1991) 《多元 柔性 主体 —— 80 - 90 年代语法研究大势之我见》，《世界汉语教学》1991 年第 4 期。
- 史有为 (1998) 《汉语语法研究百年回顾 —— 汉语语法学史的再认识》(上、下)，《语文建设通讯》58 期，59 期。
- 史有为 (2000) 《迎接新世纪：语法研究的百年反思》，《语言教学与研究》2000 年第 1 期。
- 王洪君 (2000) 《汉语语法的基本单位与研究策略》，《语言教学与研究》2000 年第 2 期。
- 温锁林 (2001) 《现代汉语语用平面研究》，北京图书馆出版社 2001 年版。
- 吴蔚天、罗建林 (1994) 《汉语计算语言学——汉语形式语法和形式分析》，电子工业出版社 1994 年版。
- 邢福义 (1996) 《汉语语法学》，东北师范大学出版社 1996 年版。
- 徐杰 (2001) 《普遍语法原则与汉语语法现象》，北京大学出版社 2001 年版。
- 徐通锵 (1994a) 《“字”和汉语的句法结构》，《世界汉语教学》1994 年第 2 期。
- 徐通锵 (1994b) 《“字”和汉语研究的方法论 —— 兼评汉语研究中的“印欧语的眼光”》，《世界汉语教学》1994 年第 3 期。
- 徐通锵 (1997) 《语言论 —— 语义型语言的结构原理和研究方法》，东北师范大学出版社 1997 年版。
- 徐通锵 (2001) 《字和汉语语义句法的基本结构原理》，《语言文字应用》2001 年第 1 期。
- 杨成凯 (1991) 《动词作主宾语是汉语的语法特点吗？ —— 汉语语法特点散论之一》，《汉语学习》1991 年第 6 期。
- 杨成凯 (1993) 《关于短语和句子构造原则的反思 —— 汉语语法特点散论之二》，《汉语学习》1993 年第 2 期。
- 杨成凯 (1994) 《关于汉语语法单位的反思 —— 汉语语法特点散论之三》，《汉语学习》1994 年第 6 期。
- 俞士汶、朱学锋 (2002) 《关于汉语信息处理的认识及其研究方略》，《语言文字应用》2002 年第 2 期。
- 俞士汶、常宝宝、詹卫东 (2003) 编著《计算语言学概论》，商务印书馆 2003 年版。
- 袁毓林 (2003) 《走向多层面互动的汉语研究》，《语言科学》2003 年第 2 卷第 6 期。
- 詹卫东 (1997) 《PP<被>+VP1+VP2 格式歧义的自动消解》，《中国语文》1997 年第 6 期。
- 詹卫东、常宝宝、俞士汶 (1999) 《汉语短语结构定界歧义类型分析及分布统计》，《中文信息学报》1999 年第 3 期。
- 詹卫东 (2000) 《面向中文信息处理的现代汉语短语结构规则研究》，清华大学出版社 2000 年版。
- 詹卫东、常宝宝、俞士汶 (2002) 《机器翻译与语言研究》，《语言科学》2002 年第 1 卷第 1 期。
- 朱德熙 (1982) 《语法分析和语法体系》，《中国语文》1982 年第 1 期。
- 朱德熙 (1982) 《语法讲义》，商务印书馆 1982 年版。
- 朱德熙 (1985) 《语法答问》，商务印书馆 1985 年版。