

面向自然语言处理的现代汉语词组本位语法体系*

A NLP-Oriented Phrase-Based Grammar of Contemporary Chinese

詹卫东

提要 本文从计算机处理自然语言的角度审视现代汉语语法的理论建设问题,较为全面地讨论了现代汉语词组本位语法体系的思想原则和分析方法,及其在自然语言处理应用方面的表现。词组本位体系以它已有的丰富研究成果为依托,以信息时代计算机技术的飞速发展和广泛应用为契机,一方面在自然语言处理研究中进一步检验其科学性和形式化的可操作性;另一方面也在面对实践中新的挑战时不断发展完善。

§ 1 全景扫描:词组本位语法体系面面观

有关词组本位语法体系的明确提法,是朱德熙先生在他那本篇幅不长,却影响颇大的小册子《语法答问》¹中首先加以阐发的。不过这一体系在词类划分的原则、汉语语法各单位间的关系以及如何析句等语法理论基本问题上所持看法的形成却由来已久。这不仅显著地体现在它跟不同语法体系各个时期论战的激烈交锋上,也饱蕴在这一体系阵营里几代学者对汉语事实及其规律的努力探索中。本文首先就立足于有关词组本位语法体系的经典理论描述,面向自然语言处理,概括介绍词组本位语法体系的基础理论和分析方法,以及对歧义这一困扰着人,更困扰着计算机的问题是如何看待并处理的。下面的讨论不准备过分纠缠一些概念的浅层辨析,而是力求触及一种思想或者分析方法的精神实质。

1.1 一以贯之的功能分类思想

词组本位语法体系坚持以功能(也即分布 Distribution)标准对汉语的词进行分类。之所以采取这种分类标准,可从分类目的着眼得到合理的解释。

对汉语的词进行分类,一个很直接的目的就是构建汉语的语法描述系统,用来描写汉语语言事实和规律,解释各种语言现象及其相互间的内在联系。基于这一目的,对词的分类就必须满足一个条件,即分类结果应该能够有效地说明各类词之间的位置关系(也包括搭配关系)。譬如:“红”跟“红色”两个词,按功能分类的标准,前者可划入形容词类,后者则属名词类。在句法形式上表现为“红”可以在“不、很”的后面出现形成合法的结构。而“红色”不能;“红”可以后附“着、了、过”等助词,而“红色”不能。等等。不难看出,功能分类的结果实际上正是对词跟词之间相互位置关系以及能否形成结构的直接描述。

* 本文研究工作得到国家 863 项目(编号 863-306-03-06-2)和国家自然科学基金项目(编号 69483003)支持。北大计算语言所俞士汶教授在本文思路形成和写作过程中都给予很多指导;文中不少想法得益于跟中科院计算所二室刘群副研究员、北大计算机系常宝宝博士的讨论,在此一并表示感谢。

对词进行功能分类的思想很容易扩展到词组的分类上²。词组功能分类的结果直接概括了一个词组向外组合的性质特征。而通常对词组采用的结构分类方式则类似描述词的内部构成，虽也可在一定程度上说明一个词组对外的结合能力，但跟功能分类方式相比，还是隔了一层，是间接的。譬如说，“打球”，按功能分类，属动词性词组（VP）；按结构分类，属述宾词组。对VP而言，我们可以像描述动词那样，方便地概括它跟其他功能类的词组之间的搭配关系，如可前加介词词组（“跟他打球”）形成状中关系；对述宾结构而言，要概括它跟其他结构类的词组之间的搭配关系，就麻烦得多。因为我们并不能根据一个词组的内部结构类型而直接地知道它的外部组合情况。此外，词组还可跟词直接组合，前者是结构类，后者是功能类，描述它们之间的搭配关系显然也是不协调的。词组本位语法体系在词和词组两级语法单位上贯彻功能分类标准，为构建一个简明有效的语法描写系统打下了扎实基础。

1.2 层次和关系并重、多种方法递进运用的分析策略

词组本位语法体系用以进行句法分析的基本手段是所谓层次分析法。这种方法背后的语法思想则是语言的结构观念。从结构观念出发对一个线性的语符串进行分析，就是将表面的线性表达式分解还原成深层的按一定结构层级组合的图式。一个线性语符串变成了若干层次上的大小不等的语言片段的组合。反映了各个语言成分之间结合紧密程度以及先后顺序及关系的不同。这是语言系统客观存在的一种性质。而语法分析必须揭示这种层次组合的性质，才可能对语言现象做出清晰的描写跟合理的解释。在自然语言处理领域广泛使用的乔姆斯基范式（Chomsky Normal Form），正因为包含了语言系统内在的这种层次特点，才成为自然语言处理应用系统语法形式表达的基础。八十年代国内有的学者提出对句子进行组块分析，本质上也是体现了语言系统的这种结构层次性质的³。

以层次分析为基础，词组本位语法体系进而逐步发展出变换分析法、语义特征分析法⁴、语义指向分析法⁵等多种分析手段，对语言成分间的句法关系和语义关系，以及各个成分本身的语法特征和语义特征及其相互间的制约关系，都能有效地分层加以描写刻画。

就关注对象而言，层次分析法针对的是单个的句法结构。一方面它只能分析同一层次上相邻直接成分（immediate constitute）间的关系但无法顾及不同层次上非直接成分间的关系，另一方面它只是指出同一层次上直接成分间的句法关系却没有涉及语义关系。变换分析法把关注对象从集中于单个结构转向不同结构间的变换关系上。基于相同的语义关系对应相同的变换关系这个前提，通过考察不同结构间形式上的变换关系，可以对直接成分间的语义关系做出判断，离析出表面模式相同而深层语义关系有别的句法结构来。从这点上说，变换分析法把单个句法结构范围内解决不了的问题转化为相关句法结构间的变换来求解，比层次分析法扩大了考察视野。这类似于几何学中添加辅助线的办法。至于结构模式相同而语义关系不同的原因，则可通过细致比较语言成分间语义特征的差异来解释。跟变换分析、语义特征分析关注直接成分间的语义关系相比，语义指向分析关注的则是非直接成分间的语义关系。虽也是着眼于单个句法结构，但语义指向分析不同于层次分析，是跨层次进行分析的。由此看来，上述几种分析方法可以说是覆盖了一个语符串内部结构从句法到语义的所有方面，既有对同一层次上直接成分间句法语义关系进行分析的，又有对不同层次上非直接成分间远距离语义关系进行分析的，形成了一套逐级递进的全方位分析机

制。

不难发现，前一小节对词和词组的功能分类的讨论实质上也就是如何确定语言成分的聚合性质（Paradigmatic）。这一小节对几种分析手段的比较归纳目的则是探求语言成分的组合格质（Syntagmatic）。这两部分内容有机结合，正是在语法层面对结构主义语言学所谓的语言系统聚合组合两大基本关系做出的整体说明。通过这样的分析处理模式得到的语言学研究成果，可以紧扣自然语言处理的要求，为实际应用系统提供源源不断的语言知识养分。

1.3 词组本位语法体系的歧义处理模式

基于上面的语法处理模式，词组本位语法体系形成了层次分明行之有效的汉语歧义分析和消解技术。在对自然语言中广泛存在的歧义问题的认识上，词组本位语法体系坚持一贯的层次观念，区别在各个层次上造成歧义的不同原因，包括同形多义词⁶，结构层次组合歧义，句法结构关系歧义，语义关系歧义，语义指向歧义等等。后两种歧义直接牵扯到语义问题，消歧难度比结构歧义要大得多。特别是在自然语言处理的现有研究水平上，一时还很难找到有效的消歧策略。因此把近期目标定位在结构歧义问题上，应该是明智之举。

对结构歧义，朱德熙先生很早就提出了“歧义格式”的概念⁷。冯志伟教授进一步将这一概念深化成为“歧义结构的潜在性”⁸。区分了格式固有的隐性歧义和具体实例的显性歧义。我们认为，对所有的结构歧义而言，潜在性是语言系统赋予的自然属性。从具有潜在歧义的系统备用格式向表层语言符号串发展，既可能造成歧义实例，也可能造成没有歧义的实例。语言系统本身提供了丰富的手段或者说是调节机制来完成这一过程。面向自然语言处理，还应该根据从隐性的格式歧义向显性的实例歧义映射的不同情况，把歧义格式进一步分化为不同的类型⁹。举例说来，像“V+A+N”这样的格式¹⁰，既可形成无歧义的实例如“踢新球、踢碎热水瓶”等，也可形成歧义实例“踢破球”，可以称之为真歧义格式；像“D+V+N”这样的格式，虽从结构上可以有[D [V N]]和[[D V] N]两种构成方式，但它投射出来的实例都没有歧解，例如“非常喜欢电影”，不管按哪种构成方式都不影响意义理解，可称之为伪歧义格式；像“P<被>+V+V”这样的格式，结构上类似“D+V+N”，也有两种构成方式，投射出来的每个实例本身都没有歧解，但只能按两种构成方式中的一种来理解意义。如“被政府邀请参加了大会”和“被警察抓住罚款”，前者“参加”不受“被”修饰，跟“邀请”不在同一结构层次上，后者“罚款”受“被”修饰，跟“抓住”在同一结构层次上，两者的内部构成方式有别，并且各自都只能唯一地分析为一个意思。这种情况可称之为准歧义格式。以上三种歧义格式类型都反映了汉语句法结构的潜在歧义性。同时又有程度差异。从这个角度考察汉语的结构歧义，就能针对歧义的不同特点，更有效地解决歧义分析问题。

§ 2 特点透视：词组本位语法体系的自适应性和可扩充性

在对一个语法体系有了全局性的总体把握之后，对其特色之处作进一步剖析无疑会有助于更深入的了解。而在审视一个语法体系的特色时，体系内各部分间的相互协调能力也即系统自适应性，对新的语法思想的兼容能力以及对语法规律的挖掘能力也即系统的可扩

充性，是两个重要窗口。本节我们就来讨论词组本位语法体系在这两方面的表现。

2.1 词组本位语法体系的自适应性

一个语法体系的自我协调性能如何，取决于它是否充分反映了所描述语言对象的内在特质。就汉语而言，以下三个特征当在考虑之列。（1）词、词组、句子各级语法单位一定程度上同构。（2）词、词组及句子三者并无天然的形式界限（这一点词和词组在书面上的表现更为显著）。（3）词组构造可以在相当宽松的条件下套叠。

联系上一章对词组本位语法体系的介绍不难看出，以词组为本位即立足词组，统摄词和句子展开汉语的全面句法分析，正与上面这三个特征吻合。如果剔除言语运用中的附加因素（如语气、独立语、插入语等等），仅在抽象的语言系统层面看汉语词、词组、句子各级语法单位的构造，很容易发现这三级单位的同构关系。单纯词谈不上内部构造可以撇开不论，合成词的内部构成方式有述宾式、偏正式、主谓式、并列式等等，词组的构造同样是这样几种类型，而正如朱德熙先生早就指出过的那样，汉语句子可以看作是由词组实现（realization）得到的，大多情况下构造跟词组也基本一致¹¹。再在具体的语言交际使用中看三者的关系。汉语语法学家都不否认有所谓独词句的存在，如“干杯、水”等，前者用于祝酒辞，后者用于请求给予实物，就功能而言，当属句子无疑。也就是说，在一定场合下，词也可直接作为句子使用。而象“来信、打手”这样的发生在词和词组之间的跨级歧义现象¹²，也昭示着汉语词跟词组之间形式界限的模糊。值得一提的是，这种现象正是造成计算机对汉语进行分词处理时碰到所谓组合型歧义的直接原因。显然，汉语三级语法单位形式上存在连续性，没有天然分隔界限；在构造上基本采用相同模式。这就决定了位居中间的词组的特殊地位。尽管浑然一体的单纯词以及不规整的流水句和复杂句等都不能纳入词组模式之内，但同构部分仍是主流。词是功能单位，可以看作是内部封装起来的一个实体或者是原子。语法系统只关心它对外结合其他成分的性质。句子则是结构单位。它不再跟其他任何成分发生结合关系。语法系统只关心它的内部状态。只有词组兼具结构单位和功能单位双重角色。进一步看，词不能由词构成，句子也不能再构成句子，只有词组可以由简单而复杂套叠生成更大的词组¹³，语言结构的递归（recursive）性质在词组一级上体现得最为鲜明。汉语各种类型的词组构造可以自相套叠且不需任何形态变化（例如：“这个人的确心眼儿很好”中“心眼儿很好”本身是个主谓词组，被包孕在更大的主谓词组也就是整句之中）。综合这些因素，选择词组作为组织语法描述系统的切入点，对词的属性判断可以直接在构成词组时得到检验，对句子的分析理解也可转化为对词组的层层剖析。同时，以词组为本位，并不意味着研究重心向词组倾斜而置词和句子的个性于不顾。词组本位不是研究对象意义上的本位，而是方法论意义上的以词组为本。也正是在这个意义上，词组本位语法体系以词组构造规律的研究辐射对词和句子性质的探求，理顺了三级语法单位间的关系。同时以层次分析为基础，运用变换分析，语义指向分析，语义特征分析等多种方法，针对不同的语言现象，各种方法各有侧重，在句法语义不同层次上对语言成分间的结构关系和意义关系展开深入的解析，显著地体现了一个语法体系内部各部分间以及不同分析方法运用之间的自适应性。

2.2 词组本位语法体系的可扩充性

可扩充性是着眼于横向拓展和纵向深化两方面来考察一个语法体系的伸缩性能。横向主要看该体系对新语法思想的反应，纵向则看它对语言现象和规律的认识能否逐步深入。

就横向拓展能力而言，在词组本位语法体系的框架下开展的有关汉语格语法（Case Grammar）、配价语法（Valent Grammar）的研究¹⁴，已经生动地说明了这一体系对语法学新思想良好的消化吸收能力。究其原因，还是在于词组本位语法体系以词组为本的观念。以词组为本很容易把格语法和配价语法关注语言成分间搭配性质的思想从词扩展到词组。举例来说，动词“带”后面只能跟具体事物类名词宾语，不能跟抽象事物类名词宾语。如能说“带书包”，不能说“带马克思列宁主义”。但当“带”跟“来”构成述补式动词词组后，“带来”就可以跟抽象事物类名词发生语义联系了，如既可以说“带来了书包”，也可以说“带来了马克思列宁主义”。再譬如“吃”跟“吃饱”，前者不仅可以带一般的受事宾语（如“吃饭”），还可带处所宾语（如“吃食堂”），工具宾语（如“吃大碗”），跟名词的搭配面比较宽；后者则不能带处所和工具宾语，搭配面较窄。不过除了能带受事宾语（如“吃饱了饭”）外，“吃饱”比“吃”另外又增加了一项搭配选择，可以带“肚子”作宾语（能说“吃饱了肚子”）。而同样是述补式动词结构，“吃完”跟“吃饱”又不相同，只能带受事宾语（如“吃完饭”），搭配面比“吃饱”更窄。像这样的情况在汉语动词的配价研究中不是孤立存在而是普遍的成系统的现象。这就使得在汉语的格或配价研究中不局限于词一级而是充分重视词组的配价问题更显意义重大。词组本位语法体系在这方面进行了有益的扩充。

至于纵向深化方面，不妨以词组本位语法体系对汉语歧义现象的不断挖掘为例来略做说明。如果从歧义对环境（context）的影响程度不同着眼，句法结构层面的歧义可分为自围型歧义和他围型歧义两种类型¹⁵。前者的歧义实例取不同的可能意义时，连带着会有整个结构体组合功能上的明显差异。如“发现了敌人的哨兵、出租汽车”等；后者的歧义实例取不同的可能意义时，整个结构体对外的组合功能没有显著不同。如“放大了一点儿、想起来”等。具体说来，“发现了敌人的哨兵”整个结构体既可理解为动词性词组，也可理解为名词性词组。内部构造差异会造成对外组合能力的不同。“放大了一点儿”不管内部构造如何，整个结构体功能上都属动词性词组，并不明显地对外表现出组合差异。“出租汽车”和“想起来”也是如此。前者内部结构关系有定中和述宾两种可能性，分别对应着不同的整体功能类名词性词组和动词性词组，整体对外组合情况会有较大不同；后者内部结构关系虽然也有述补和述宾两种可能性，但两种情况下整体功能类都是动词性词组，按不同关系理解意义，对外组合情况并无显著差别。基于词组本位语法体系的这些歧义研究，深入发掘了汉语句法结构歧义客观存在的不同类型，细化了人们对歧义现象的认识。

§3 多维显影：词组本位语法体系的语言工程实践

上面的讨论侧重在理论层面，这一节我们则从实践角度来概观词组本位语法体系在自然语言处理的实际应用中表现如何。限于篇幅，不能面面俱到。即使下面重点介绍的几项基于词组本位语法体系开展的研究工作，也只是点到为止，没有展开论述。但我们期望这些工作一方面可以丰富并完善对词组本位语法体系的认识；另一方面也对理论本身有个检验，达到在实践中不断提高的目的。

3.1 现代汉语语法信息词典的开发

现代汉语语法信息词典¹⁶的最初设想，源于北大计算语言学研究所俞士汶教授提出的“现代汉语词语语法信息库”的开发计划。这个计划作为国家七五攻关项目“自然语言理解与人机接口”中的一个子专题，有意把汉语信息处理的研究重点首先放在对汉语词语语法属性的描述上。时逢北大中文系朱德熙先生正承担“现代汉语词类研究”这一社科攻关项目。两家在对汉语词语语法属性的全面研究有着同样的紧迫认识并对汉语词类问题有着广泛认同的基础上，开始了长期紧密合作。1990年，“现代汉语词语语法信息库”取得阶段性成果，1995年，“现代汉语语法信息词典”通过电子工业部技术鉴定。如今，这部凝聚开发人员十余年心血的词典已达到5万多词的规模，总信息量约占16兆字节存贮空间。

联系国际自然语言处理研究的大环境来看，随着八十年代中期词汇功能语法（LFG）、功能合一语法（FUG）等一批以复杂特征描述为主要表达手段的新语法学派的崛起，大词库，小规则集成为语言研究的潮流。而作为生成语法学派鼻祖的乔姆斯基，在不断修正发展其理论体系的过程中也是顺应潮流而动。从GB到最简方案，一直在加强词库建设。一时间，词库成为人类语言知识最合适不过的栖身之地。在这种理论和实践背景下，中国语法学界和自然语言处理研究领域的专家们站在了时代和科研发展的前沿，敏锐地把握住了全面展开汉语词语语法属性研究的良机。

词组本位语法体系在这一世纪工程中以导演的身份安排了一切。首先是选取一些具体的功能标准确定了汉语的词语分类系统，并对照一个词语的句法功能表现将它归入某个词类；然后是以功能理念指导词语语法属性项目的设置，并根据一个词语实际使用的情况标记它的属性值。词典中属性项目可设置得相当多。例如作为研究重点的动词在词典中共设立了100多项属性，来标记一个动词能否重叠、能否直接受名词修饰、能否作“有”的宾语、是带体词宾语还是带谓词宾语等等。项目数量的多少取决于对一个词的功能描述的广度和深度要求。从学习的角度来说，这可以看成是一部面向计算机的关于汉语词语的用法详解词典。对句法分析而言，以这样一部词典为基础，逐步提高计算机的排歧能力将不再是纸上谈兵。

在词组本位语法体系的框架下，词语语法属性的组织和描写都可以相当灵活。层次分析法、变换分析法、语义特征分析法、语义指向分析法都能大显身手，对成千上万的汉语词语各自丰富的语法个性以及同一类词的语法共性，既能尽量细致地刻划，又可避免词典信息量冗余度过大。不光描述语法信息，还可进一步增加语义信息；不光是在词一级上描述，还可拓展到词组一级进行描述。这样最终就能建立起面向自然语言处理的语言知识库通用平台。

3.2 汉英机器翻译系统的研制

作为自然语言处理最有活力的应用领域之一，机器翻译从开始简单机械的词对词硬译发展到今天基于句法规则方法（rule-based）、基于语料库实例方法（example-based）、以及基于统计方法（statistic-based）等等多种策略并存的局面，不断在探索自然语言之间实现同义转换的最佳途径。实践经验告诉人们，没有对自然语言本身透彻的理解，要让机器完成这样的转换是不可能的。脱离自然语言的具体研究奢谈语言处理的工程实践无异于建造

空中楼阁。不管是理性主义的规则方法还是经验主义的语料库方法，实质上都需要人自身先对语言知识有科学的描写和解释能力。而选择恰当的符合一种自然语言实际情况的语法体系，显然会令关于该语言的知识表述无论是呈现在人还是计算机面前，都更为清晰，易于操作富于效率。以词组本位语法体系为理论框架建立的“现代汉语语法信息词典”，成功地将汉语词语丰富的语法功能信息纳入“属性项：属性值”对这种简单有效的二维描述模式，为汉外机器翻译提供了坚实的词库基础。在此基础上，北大计算语言所又与中科院计算所合作研制了一套描述汉语句法结构规律的形式语法系统，语法系统跟词库相结合，共同构成了联合开发的汉英机器翻译模型系统语言知识库的主干¹⁷。

构造形式语法系统的工作，也就是把词组本位语法体系框架下的有关汉语句法结构规律的研究成果转写成机器可读的形式。上下文无关文法的产生式和功能合一文法的合一等式相结合，是完成这一工作的理想途径。产生式描述汉语句法系统允许的可能组合类型；合一等式给出语言成分相互间发生组合关系的条件。前者对决定一个形式系统覆盖真实语料的能力起主要作用；后者则对形式系统的句法分析能力排歧能力有显著影响。目前，我们用于汉英机器翻译模型系统的规则已有 300 余，绝大多数规则都给出了约束条件限制的说明。对汉语句法结构的覆盖面和分析深度也达到一定水准。系统对 3000 余句封闭语料进行测试，达到 90% 的正确率。

3.3 受限汉语研究

顾名思义，受限汉语¹⁸ (Restricted Chinese) 是对汉语加以限制造成的一种带有人为特色的自然语言。研究这种受限自然语言的初衷也如同它的名字一样明白易解。在计算机面对自然语言的充分复杂性而又应变乏术的状况下，作为研究者至少可采取这样两种策略。一种是打计算机的主意，多想办法从技术上提高计算机处理自然语言的能力；另一种就是动自然语言的脑筋，考虑如何降低处理对象的复杂度。受限汉语就是后一种想法的直接结果。

在理论层面上看，上面提到的两种策略对人的要求没有本质差别。不管怎样做，都是需要人对自然语言系统内在的规律性先有明确精细的认识。否则两种策略都无法真正具体的落实。不过解决问题的路线不同，研究重点还是会有很大差异。就受限汉语的研究而言，在语言系统的精确度和表达意义的自由度之间找到一个合适的平衡点，是主要努力方向。而这是必须在科学的语法理论指导下才有实现可能的。分别审视字、词、词组、句子、篇章等从小到大各级语言单位的受限问题，字、词显然相对简单容易操作；句子和篇章则因素复杂难以下手。而处在中间枢纽环节的词组，本身结构相对稳定，不象句子语序那样多变。同时大量的歧义问题又集中在这一层级上，从词组入手开展系统的汉语受限研究在理论上和可操作性上都具明显的优势，以词组本位语法体系为理论指导具体落实这一工作也就顺理成章了。结合上文对汉语句法结构歧义的讨论，由于认识到结构类型歧义跟实例歧义间存在不同的对应关系，歧义结构对上下文环境还有不同影响程度等现象，在对汉语词组结构进行受限研究时就有所依据，可以选择无歧义或歧义程度低的结构作为汉语句法结构系统最核心的部分。在此基础上考虑表达需要做适当有条件限制的增补，就能得到合理的受限汉语词组结构系统。此外，运用变换分析法，可以揭示语言中表达能力相近的句法格式之间的同异，使得对汉语句式的选择更有科学依据。这样，以相对稳定的词组结

构去控制语序灵活的句子，以句式间同义变换为客观基础科学地精简表达，最终就可发展出一套受限汉语的句型系统。

北大计算语言学研究所词组本位语法体系的框架下，已经有步骤地进行了多个词组结构的调查，如“被”字、“把”字结构、连谓结构等等，并综合这些单项研究形成了一个初步的受限汉语句型系统¹⁹。以此为依托，结合机器翻译、辅助写作系统等等应用开发项目的研究，就能不断调整汉语受限的尺度，向形式精确度与表达自由度的最佳平衡逐步逼近。

结 语

尽管面向自然语言处理的汉语语法研究已有不少理论上和实践上的探索，但这一领域总的来说还是处在起步阶段。必要的经验积累和及时的总结可以更好的把握航向。本文的写作，最初也就是源于这样的动机。我们期望中国的理论语言学界和计算语言学界都多一些理论深层的思考。虽然这种思考与扎实的具体而微的研究须臾不可分，但仍有单独提出来强调一下的必要。在借鉴国外先进语法理论的大环境下，基于对汉语语言事实的独立认知发展中国的现代语法学，及时把握住自然语言处理带给语言学发展的新机遇，不能不说是带有些当务之急的意味。这些见解恰当与否，文中对词组本位语法体系理论和应用上的把握是不是准确，都敬请专家读者批评指正。

附 注：

¹ 朱德熙（1985）《语法答问》，商务印书馆 1985 年版。

² 可参见周强 俞士汶（1996）《汉语短语标注标记集的确立》，载《中文信息学报》1996 年第 4 期。詹卫东等（1997a）《现代汉语短语本位语法体系在机器翻译中应用及其问题》，载《智能计算机接口与应用进展（97）》，电子工业出版社 1997 年版。

³ 参见陆丙甫（1986）《语句理解的同步组块过程及其数量描述》，载《中国语文》1986 年第 2 期。

⁴ 陆俭明（1993）《八十年代中国语法研究》，商务印书馆 1993 年版。

⁵ 陆俭明（1996）《关于语义指向分析》，载《当代中国语言学》总第一期，1996 年。

⁶ 可参见詹卫东（1997b）刘群《词的语义分类在汉英机器翻译中所起的作用以及难以处理的问题》，载《语言工程》，清华大学出版社 1997 年版。

⁷ 朱德熙（1980）《汉语句法中的歧义现象》，载《现代汉语语法研究》，商务印书馆 1980 年版。

⁸ 冯志伟（1995）《论歧义结构的潜在性》，载《中文信息学报》，1995 年第 4 期。

⁹ 参见詹卫东（1996）《现代汉语 VP 的结构层次和结构关系判定》，北京大学硕士学位论文。

¹⁰ V 表示动词；A 表示形容词；N 表示名词；D 表示副词。

¹¹ 同注 1。

¹² 参见陆俭明（1988）《名词性“来信”是词还是词组》，载《中国语文》1988 年第 5 期。

¹³ 参见陆俭明（1990）《汉语句法成分特有的套叠现象》，载《中国语文》1990 年第 2 期。

¹⁴ 参见林杏光 鲁川（1994）《论深化现代汉语格关系的研究》，载《动词大词典·序言》，中国物资出版

社 1994 年版；沈阳 郑定欧 主编《现代汉语配价语法研究》，北京大学出版社 1995 年版。

¹⁵ 同注 9；还可参见孙茂松 黄昌宁（1989）《汉语中的兼类词、同形词类组及其处理策略》，载《中文信息学报》1989 年第 4 期。

¹⁶ 有关现代汉语语法信息词典的详细情况，可参阅朱学锋等《〈现代汉语语法信息词典〉的开发与应用》，载新加坡《中文与东方语言信息处理学会通讯》1995 年第 2 期；俞士汶等《〈现代汉语语法信息词典〉规格说明书》，载《中文信息学报》1996 年第 2 期。

¹⁷ 参见詹卫东（1997a）；刘群 詹卫东等《一个汉英机器翻译系统的计算模型与语言模型》，载《智能计算机接口与应用进展（97）》，电子工业出版社 1997 年版。

¹⁸ 参见俞士汶（1995）《关于受限的规则汉语的设想》，载王均主编《语文现代化》，山东教育出版社 1995 年版。

¹⁹ 参见张伟（1997）《受限汉语的句型》，北大计算语言学研究所内部报告。