

语言工程

陈力为 袁 琦 主编

Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering
Language Engineering

清华大学出版社

语 言 工 程

LANGUAGE ENGINEERING

陈力为 袁琦 主编

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是全国第四届计算语言学联合学术会议(1997年8月15—17日,北京)的论文集。书中选录的62篇论文全文和12篇论文详摘是从大会征集到的113篇论文中精选出来的。内容包括下列七类:(1)汉语的句法和语义;(2)语料库建设与语料加工技术;(3)基于语料库的语言分析方法;(4)文本分析与生成;(5)机器翻译;(6)文本检索、自动文摘、文本校对;(7)智能型汉字输入方法和人机接口。

本书展现了近两年来我们计算语言学研究与应用的最新进展,对今后中文信息处理的基础研究和产品开发具有重要的参考价值。本书可供计算机、语言学等专业的科研人员、工程技术人员、大学教师和研究生选读。

书 名: 语言工程

作 者: 陈力为等

出 版: 清华大学出版社(北京清华大学校内,邮编 100084)

印 刷: 清华大学印刷厂

发 行: 新华书店总店北京科技发行所

开 本: 787×1092 1/16 印张: 26 字数: 666 千字

版 次: 1997年8月第1版 1997年8月第1次印刷

书 号: ISBN 7-302-02618-1/TP·1342

印 数: 001~600

定 价: 60.00 元

全国第四届计算语言学联合学术会议组织情况

日期：1997年8月15日～17日

地点：北京清华大学

发起单位：中国中文信息学会
中国计算机学会
中国人工智能学会
北京市语言学会

赞助单位：电子工业部计算机与微电子发展研究中心
美国IBM公司中国研究中心
微软（中国）有限公司
台湾英业达集团电子技术有限公司
中国科学院计算机语言信息工程研究中心
清华大学计算机科学与技术系

组织单位：北京清华大学

大会名誉主席：陈力为
大会主席：姚天顺
大会副主席：刘开瑛 张 普

程序委员会主席：黄昌宁
副 主 席：俞士汶 董振东
委 员：（以姓氏拼音为序）
蔡莲红 曹右琦 陈群秀 傅爱平 嘎日迪 黄河燕 姬东鸿
林杏光 刘开瑛 鲁 川 茅于杭 孙茂松 许寿椿 姚天顺
易绵竹 苑春法 张 普 周 强

组织委员会主席：陈群秀
副 主 席：曹右琦 苑春法
委 员：张凤中 王建章 来春红 赵 军 李涓子 张 磊

语言工程

前言

由中国中文信息学会、中国计算机学会、中国人工智能学会和北京市语言学会共同发起的“全国第四届计算语言学联合学术会议”(JSCL-97)即将在北京清华大学举行。我们仅代表上述四个发起单位和全体与会代表向主办本届大会的东道主——北京清华大学的同仁表示由衷的感谢，并预祝大会取得圆满成功。

当今世界信息化浪潮席卷全球。Internet 的发展非常迅速，目前全世界 Internet 的用户数已达到 6 千万左右，并且每年仍在以极高的速度增加。中国的 Internet 用户已在两、三年中迅速发展到 20 万户。在面向 21 世纪的网络通信时代，以文本信息的智能化处理为主要对象的语言工程，已成为国际上关注的热点。1996 年由联合国大学高等研究所 (UNU/IAS) 发起的通用网络语言 (UNL) 工程，将为实现网络时代的自然语言处理应用开辟广阔的领域。从另一方面看，当前计算机国际互联网是一个以英语为主导语言的网路，因此极大的限制了我国在网络上进行信息交换与实现资源共享的能力。对于我国计算语言学工作者来说，以汉语的句法语义分析为基础的应用研究是一项任重道远的任务。目前，根据市场需求牵引的指导原则，我们认为具有网络环境功能、基于自然语言处理技术的中文文本自动分类、信息检索、信息提取与生成，以及机器翻译等智能化技术将是我国“九五”期间语言工程的重点研究开发任务。

本届大会在全国各地和海外一共征集到 113 篇论文。大会的程序委员会本着严肃认真的态度对全部应征论文进行了评审，最后精选出 62 篇论文在本届大会上宣读交流。评选出的论文充分展示了国内计算语言学研究与应用的最新进展。注重实用性的研究与开发依然是它们的一个重要特点。本书中选录的论文大体上可归纳为如下 7 类：

- (1) 汉语的句法和语义
- (2) 语料库建设与语料加工技术
- (3) 基于语料库的语言分析方法
- (4) 文本分析与生成
- (5) 机器翻译
- (6) 文本检索、自动文摘、文本校对
- (7) 智能型汉字输入方法与人机接口

根据程序委员会的决定，本书除选录上述 62 篇论文的全文之外，还择优录用了另外 12 篇论文的详细摘要。

最后,衷心的感谢中国中文信息学会计算语言学专委会主任、本届大会程序委员会主席黄昌宁教授与陈群秀副教授在大会论文的组织 and 评审中所做出的贡献。

编者 1997年6月18日 于
电子部计算机与微电子发展研究中心(CCID)

语 言 工 程

目 录

论 文 全 文

I. 汉语的句法和语义

“请”的句法语义特征和兼语句的理解策略	刘永耕..... 1
“甚”的语法功能的聚类分析与词性演变模型的建立	王秋月..... 7
隐含型语句函子的求解方法	陆汝占 靳光瑾..... 13
现代汉语槽关系研究	林杏光 张庆旭..... 19
名词空位的控制性同指、照应性同指与词汇性同指	沈 阳..... 25
自然语言理解的人本中心论	易绵竹 汤庆国..... 31
对汉语中包含结果补语的句子分析策略	胡正微 胡继红..... 37
汉语句子的信息安排及其句法后果 —— 以“周遍句”为例	温锁林..... 43
现代汉语自动词句模研究	苗传江..... 49
现代汉语插入语研究	邢红兵..... 55
语言的多媒体性与多媒体语言知识的作用	张 普..... 61

II. 语料库建设与语料加工技术

实现 500 万词级《现代蒙古语文数据库》的主要措施	华沙宝..... 69
《现代蒙古语词频统计系统》的设计与实现	
吉日木图 嘎日迪 赛音 白晓玲 达·巴特尔 阿拉腾敖道..... 76	
几种常用汉语词典收词的统计分析	张永奎 齐新战..... 82
大规模语料库中词语接续对的统计与分析	邱超捷 宋 柔 欧阳龙根..... 88
并行语料库中的双语知识获取	周莉娜 廖雅国 俞士汶..... 95
汉语句子的主题 —— 主语标注	陈小荷 石定栩..... 102
基于概率模型的语言的聚类方法：根据多语种语料库进行语言系统树的再构造	
北研二..... 109	
万维网上的双语文本的获取及对齐	许玉祥..... 115
基于英汉双语语料库的词汇自动对齐实验系统	李 竹..... 121

III. 基于语料库的语言分析方法

局部统计在汉语未登录词辨识中应用和实现方法	沈达阳 孙茂松 黄昌宁	127
基于语料库统计方法在汉字文本识别应用中的若干问题		
	夏 莹 马少平 朱小燕 金奕江 姜 哲 常新功	133
基于特征关联度的汉语文本自动分类系统的设计与实现	张玥杰 姚天顺	139
汉语语义关联网的研究	苑春法 许 伟 黄昌宁	145
现代汉语紧缩结构定量分析	梁蕴华	151
从标注语料库中归纳语法规则: “V + N” 序列实验分析	孙宏林	157
基于组合实例的双向优化聚类	李涓子 姬东鸿 黄昌宁	164

IV. 文本分析与生成

基于词典的汉语名词语义信息的自动分析与获取

	翟高寿 张永奎 杨尔弘 查建中	170
汉语天气预报文本内容规划器的设计与实现	王 纤 姚天昉	176
歧义字段的最大概率切分算法	刘 挺 王开铸 姜兴海 邵艳秋	182
汉语句子分块研究	王厚峰 戴大为	188
汉语匹配分析算法的实现	周 强	194
中文文本歧义切分技术研究	郑家恒 刘开瑛	201
A Lexicalized Tree Adjoining Grammar for the Generation of Chinese Weather Forecasts	Yu-Fang Wang	207
基于维语中心语驱动句法分析器的结构及其实现		
	玉素甫·艾白都拉 吾守尔·斯拉木	214

V. 机器翻译

面向受限领域的汉英口语翻译	王海峰 高 文 李 生	219
汉语方言机器翻译	张小衡 石定栩	225
实用型日语自动分词系统的算法及初步实现	陈群秀 孙勇 陈利人 莫林峰	232
汉语句子规范化处理策略的研究	宗成庆 章 森 陈肇雄 黄河燕	239
基于译后启发式信息的知识辅助获取系统: HIKAS	汪 冰 黄河燕 陈肇雄	245
朝语情态问题研究	毕玉德	251
基于依存关系分析的日语句法分析器	简幼良 唱红涛 王秀坤 黄德根	258
日汉机器翻译系统中的断段分析法研究与实现	曹睿妮 陈群秀	266
用概率方法处理汉英机器翻译系统中的歧义	刘 颖 常宝宝 刘 群 王 斌	274
基于软件构件的机器翻译研究方法	刘 群 张 祥	280
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		
	詹卫东 刘 群	286
机器翻译基于从属关系的语义分析	傅爱平	292
带有对译信息的惯用表示的收集	田中康仁	298

VI. 文本检索、自动文摘、文本校对

基于选择生成文摘法的自动文摘系统研究与实现	杨晓兰 宋 帆 钟义信..... 313
汉语文本校对字词级查错处理的研究	孙 才 罗振声..... 319
中文校对系统中的修改建议提供算法	郭志立 裘照明..... 325
汉语文摘系统中文本结构的自动分析	薛翠芳 李晓黎 郭炳炎..... 332
中文校对软件标准评测系统的构造	郑 佩 杨力平 裘照明..... 338

VII. 智能型汉字输入方法与人机接口

语句拼音 — 汉字转换的智能处理	章 森 宗成庆 陈肇雄 黄河燕..... 345
以杂讯通道 — 讯息重建模型自动评测语言处理系统	张照煌..... 351
语音在人机接口中的作用和特点	方志刚 吴晓波..... 357
汉语查询可移植接口的设计与分析	邵晓英 童 颖 徐洁磐..... 363
容错型拼音汉字转换系统 — 一些调查及基本方案	
(日) 叶 斌 应江黔 川上裕忠 松本忠博 後藤宗弘..... 369	
计算机汉字语句输入的最短码长	王 轩 王晓龙 377
基于扩充 LR 分析的数据库汉语接口的句法分析	王力红 383
基于短语分析和动态语义搭配的拼音汉字智能转换	张小衡 陆发兴 389

论 文 详 摘

基于类比的计算语义处理机制	刘海涛..... 396
有限词汇、非特定人关键词识别系统	温艳红 李 维 张子范 张吉庆 王 粤..... 397
会议公告信息抽取系统	付宝宇 李侯运 吕克林..... 398
无词典德语自动分析	孙介铭..... 399
专用机器翻译系统 SCDOS 的技术研究与实现	张琳琳 王能忠..... 400
意义段划分问题研究	刘 挺 吴 岩 王开铸 邵艳秋..... 401
一种应用句法分析技术的汉语自动分词模型	单玉秋..... 402
基本词典与专业词典的关系	刘志杰 刘 倬..... 403
一种新的适合中文计算机处理的文法理论—核心推导语法	
	宋继平 陈肇雄 黄河燕 王能忠..... 404
汉英机器翻译中的简单语境处理	刘海军 陈肇雄 黄河燕..... 405
维语词法分析器的探讨	
	玉素甫·艾白都拉 吾守尔·斯拉木 赛依提·阿不都拉..... 406
实用型汉英机器翻译系统的研究与实现	孙广范..... 407

“请”的句法语义特征和兼语句的理解策略

刘永耕

(福建师范大学中文系现代汉语教研室)

摘要：“请”的使令义是语法意义。“请 N_2V_2 ”中的“请”具有[+手段]，[+信息]义素，使 V_2 具有了信息性质，成为非现实的行为，于是“请”成为带一个体词宾语和一个谓词宾语的三价动词，而且“请”是先与 N_2 组合成动词短语再带宾语 V_2 的。然而：(1)当句中有“来”并且“请”先与“来”结合然后再与 V_2 结合，就会使“请”失去[+信息]义素，从而使 V_2 变成目的补语。(2)当 V_2 带动态助词“了”成为实现的行为时，它就成为结果补语。

The Syntactic and Semantic Features of “Qing (invite)” and the Tactics of Understanding the Doublesentence

Liu Yong-geng

Modern Chinese Teaching & Researching Section Chinese Department
Fujian Teachers University

Abstract: The causative meaning of “Qing (invite)” is a grammatical meaning “Qing (invite)” in the structure “qing N_2V_2 ” possesses the sememe of (+ means) (+ information), which makes it possible to invest N_2 with the informational qualities and substantiate N_2 as a nonrealistic action. This results in the fact that “qing invite” becomes a frivalent verb with a nominal object and a predicate object, and that “qing (invite)” constitutes a verbal phrase with N_2 and then is followed by object v_2 .

However, ①“qing” loses its sememe of (+information), consequently transforming v_2 into a complement of purpose when, in a sentence, “qing (invite)” combines with “lai (come)” subsequently followed by V_2 . ②“Qing (invite)” turns into a complement of result when V_2 followed dynamic auxiliary “le” becomes a realized action.

兼语句是具有汉语特色的一种句式，一般认为这种句式不易作直接成分分析，很难进行结构层次的切分，因此成为计算机理解汉语的一块绊脚石。本文意在以“请”字兼语句为代表，探究使令类动词的语法特征，进而寻求兼语句结构分析和意义理解的合理策略。因为现代汉语使令类动词最典型的的就是“请”、“派”、“命令”这一类，它们所构成的 $N_1V_1N_2V_2$ 式，如“我们请李教授作报告”，向来被公认为兼语式的代表。

一、“请”的句法语义特征

在计算语言学中，词的语义解释是直接针对句法模式的，使令类动词“请”的语义特征从总体上说可以记为：[+施动/事者]，[+配动者]，[+目的]。具体分析起来，“请”的语义特征主要体现于下：

(一)“请”的意义包括两个方面：(1)施事者通过口头或书面等手段支配受事者；(2)施事者希望受事者发出某动作。这里，(1)与词汇义紧密相连，其语义特征可以记为[+手段][+信息][+动作]；(2)是语法意义，记为[+使令]。

关于(2)，首先应该注意，所谓“希望”就是“想要促使”，而不是“促使”，也就是说，想要受事者发出的行为是非现实的，从逻辑上说，“请”是现实世界的行为，而“作报告”只是可能世界的行为，试比较“我们请李教授作报告”的如下两种变换式：

A、我们请李教授，(要他)作报告。 B、我们请李教授，他作报告。

哪一种变换式更合符原句的语用价值，显然是A式，因为原句并未涉及“李教授作报告”这件事。

其次，“希望”、“要促使”都只是附加于词汇义之上的潜在意义，这种潜在意义只有在“请 N_2V_2 ”结构中才能得到显示。试看如下两句：

A、请问李教授在家吗？ B、李教授被我们请来作报告。

显然，A和B中的“请”都没有使令义。这可以证明“请”的[+使令]确是语法意义。

(二)从语义上看，“请”的施事者所希望受事者发出的行为应该说只能是“请”的目的。例如“请李教授作报告”中的“作报告”只表示“请”的目的，而没有表达其他更多的意思。

(三)“请 N_2 ”意义上不能自足，如果只说“我们请李教授”，听者必定要问：“请他干什么？”这就是说， V_2 可以省略，并且可以是不确定的，“我们请李教授”既可以是“我们请李教授作报告”的省略式，也可以是“我们请李教授讲课（或开会……）”的省略式。

(四)“请 V_2 ”则有两种情况：(1)当听者无法确定 V_2 的施事 N_2 时，“请 V_2 ”是不能说的。“请作报告”单说，固然不知所云，即便是在回答“请李教授干什么”这一问话时，也不能回答“请作报告”，而只能回答“请他作报告”或者“作报告”。(2)只有当 N_2 是“你”时，才可以说“请 V_2 ”。例如，我们只有在报告即将开始时，才可以对报告人说：“请作报告。”也就是说，“请作报告”只有作为祈使句“请你作报告”的省略式才有语用价值。

(五)综合(二)与(三)可以进一步看出，“请 N_2V_2 ”中 V_2 可以自由省略，而 N_2 则不能自由省略，这就是说， V_2 可以是变项，而 N_2 ，相对于 V_2 来说，只能是常项，这证明 V_2 与 N_2 在句中不是处于同一个层面。再从主谓关系的特点看，一般情况下句中的主语比谓语更容易省略，而 N_2V_2 的情况却正相反，这可以旁证 N_2 和 V_2 不是主谓关系，因而不在于同一个层面。

(六)当我们把观察深入到词以下的结构层面时，发现情况正与前面所说相反：当“请”和 N_2 、 V_2 都作为语素时，“请 N_2 ”和“请 V_2 ”都能够自足，而且，“请 V_2 ”还比“请 N_2 ”具有更大的能产性。词中“请 N_2 ”结构只有“请客”一个，而“请 V_2 ”则有“请示、请教、请援、请便、请罪、请功、请酒”等一大批。在这种“请 V_2 ”中， N_2 不但是不确定的，而且还是不能出现的，要出现只能出现于词以上的结构层面。如我们只能说“请示上级”、“请教老

师”、“向朋友请援”，而不能说“请上级示”、“请朋友援”等。

(七) 将(一)至(五)对比(六)，更可以看出：

(1) N_2 的常项性质只是语用层面造成的，而并非结构上的必然要求。

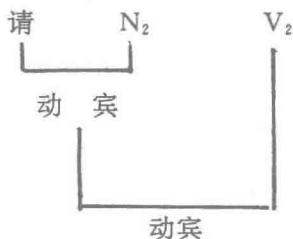
(2) N_2 与 V_2 必须同时出现，也只是语用的要求而并非结构的要求，这意味着 N_2 与 V_2 并不发生直接的结构关系。

(3) “请示上级”、“请教老师”等都是“请 $V_2 N_2$ ”式，既然这里 V_2 与 N_2 不处于同一结构层面—— V_2 处于语素层面， N_2 处于词的层面，那么同理“请李教授作报告”中， N_2 “李教授”与 V_2 “作报告”也不处于同一层面。

(八) 在“请示”、“请教”这一类“请 V_2 ”式的词中， V_2 显然是“请”的宾语，同理，祈使句“请作报告”中，“作报告”也是“请”的宾语。

二、兼语句的理解策略

“请”的语法特征的揭示，为我们制定“请”字兼语句的理解策略打下了坚实基础。首先它使我们看清了兼语式“请 $N_2 V_2$ ”中 N_2 与 V_2 虽然在语义上有构成施事者与行为的可能，但是在句法平面上二者无实际的直接关系，相反， N_2 和 V_2 都是 V_1 “请”的宾语。其次也使我们认识到“请”的这两个宾语在结构上不处于同一层面，这实质是：二者是分为两次与“请”进行组合的。这样，“请 $N_2 V_2$ ”的结构层次应该这样切分：



这从我们对它的提问方式也可看得清楚，以“我们请李教授作报告”为例：

问：你们请谁？答：我们请李教授。 问：请李教授干什么？答：作报告。

可见，“作报告”是“请李教授”这个动宾短语的宾语。这样，它的直接成分就清楚了。

“作报告”虽然是谓词性短语，但是它只是“请”所涉及的内容，也就是说，“请”这个行为的目的，是以它的宾语的形式表达出来的。这也是由“请”的词汇义所决定的：“请”的词义所表达的“请”这一行为的手段，是发出信息，不论是一番谦恭热情的话语还是一张大红的请柬，都只是表达了信息。所以，“请”的目的只能包含在这个信息之中，故只能是内容。从这个意义上，“请 $N_2 V_2$ ”式可以称为“请”的内容式。

下面仍以“请李教授作报告”作“请 $N_2 V_2$ ”式的代表来进一步讨论。这句话作为基本式在实际交际中还可以有如下语用变体：

(A) 请了李教授作报告

说 A 时，尽管已发出了邀请，但“请”仍然只是信息行为，所以“作报告”仍然只是“请”的内容，A 式与基本式在变换方式和提问方式上都没有什么不同。试比较：

a、基本式变换：请李教授，(要他)作报告。 b、A 式变换：请了李教授，(要他)作

报告。

c、基本式提问：请李教授干什么？作报告。 d、A 式提问：请李教授干什么？作报告。

(B) 请李教授来作报告 其变换式：请李教授，(要他)来作报告。

提问式：请李教授干什么？来作报告。 或：请李教授来干什么？作报告。

可见 B 式也同基本式一样。

(C) 请了李教授来作报告 其变换式：请了李教授来，(为了让他)作报告。

提问式：请李教授来干什么？来作报告。

“请……来干什么”、“来作报告”一问一答中的“干什么”、“作报告”，十分明显是“来”的目的，因此，“作报告”是“请了李教授来”的补语。

那么，能否把 (C) 的变换式改为这样：请了李教授，(要他)来作报告。

我们说，不可以。首先，邀请既已发出了，后面又紧接李教授的“来”，那么，这个“来”自然应该属前成为“请了”的结果，而不应该与“作报告”先行结合。试将 (C) 的句中停顿放在“来”之前来朗读，就会感觉不合理。其次，按照这第二种变换，“来作报告”只是“请”的内容，既然作为要求的内容，不是现实世界的行为，自然“来作报告”就与“作报告”没有本质区别了，这反过来又证明第二种变换式实质上正是 (A) 的变换式，而不是 (C) 的变换式。

(C) 的提问式也不可能是“请了李教授干什么”，因为这样问的意思是指“你们请了李教授之后还要准备干什么”，这就不是针对 (C) 的发问了，这也说明“来”既不能属后，也不能省去，否则就不是 (C) 了。

总之，(C) 中的“作报告”只能是“请了李教授来”的目的，这不仅是语义上的目的，而且实实在在是句法结构上的目的补语。

(D) 请来了李教授作报告 其变换式：请来了李教授，(让他)作报告。

提问式：请来李教授干什么？作报告。 或：把李教授请来干什么？请来作报告。

既然李教授已经请来，那么“作报告”就自然不会是一种信息内容，只能是目的补语。

(E) 请来李教授作报告 其变换式：请来李教授，(让他)作报告。

提问式：请来李教授干什么？作报告。 或：把李教授请来干什么？作报告。

值得注意的是：提问式完全与 (D) 相同。所谓“请来”，

就是用请的方法使之来，这样，“作报告”自然也不会是一种信息内容。因此，(E) 与 (C)、(D) 一样，“作报告”都是目的补语。

(F) 请李教授作了报告 其变换式：请李教授，(要他作报告，结果他)作了报告。

既然“报告”已经“作了”，那它就不再可能是信息内容或目的，自然只能是结果，因此，“作了报告”是“请李教授”的结果补语。

(G) 请李教授来作了报告

其变换式：请李教授来(作报告，结果他)作了报告。

(H) 请来李教授作了报告

其变换式：请来要教授(让他作报告，结果他)作了报告。

显然，(G)、(H) 都同 (F) 一样，“作了报告”是结果补语。

对于以上理解策略，也许有人会想不通：“作报告”或“作了报告”怎么能是“请”或“请来”的补语？其实，80 年代以来关于补语的语义指向研究已经回答了这个问题：补语既可

指向施事主语,也可指向受事宾语。例如“我吃饱了饭”中补语“饱”是指向主语“我”,而“我们吃光了一锅饭”中补语“光”则指向宾语“一锅饭”,是“饭光了”。汉语动词具有特殊的灵活性,仅在谓语中它就可以自由充当谓语核心、状语、补语和宾语,它的胸前并不须要佩戴工号、代表证或工作牌等等,站到什么岗位就是什么身份。兼语句之所以具有汉语特色,就在于它充分地开发利用了汉语动词的这种潜能。

下面再从语用角度对以上理解策略作一检验。

(一)对于基本式“请李教授作报告”来说,最可能出现的语境莫过于以下场合:报告会主持人在简单的开场白之后宣布:a“请李教授作报告。”如果李教授是应邀在倪萍主持的某次晚会上作报告,那么,主持人倪萍在宣布完a后还会热情地大声邀请:b₁“有请李教授!”或者,按照倪萍的习惯,她更可能说b₂“有请!”并且等李教授走近话筒后把右手一摊躬身说:c₁“请作报告!”或者更可能说:c₂“请!”然后飘然离去。

就倪萍宣布的a句来说,基本意义仍然是:(1)请李教授,(要他)作报告。

但这时基本义实际已具体化为言内意义:(2)现在我代表晚会组织者、全体观众以及我自己盛情邀请李教授,请求的内容只有一个——作报告,(b₁b₂是这意思的浓缩)。

此外,还有一个言外意义:(3)下面的节目是:李教授的报告(“报告”是动词)。

这样说来,a句就是一语双关:从“请”的内容关照出节目的内容。

就倪萍对李教授说的C句(正如b₂是b₁的省略式,c₂是c₁的省略式)来说,其语用意义显然是重申邀请的内容——作报告。就是说,在这个场合,不论是完整地说a,还是单说b或单说c,“作报告”作为内容是始终一贯的。

(二)对(A)句来说,最适宜的语境也许是这样的场合:李教授的学生甲在这次晚会之后向朋友介绍李教授时。甲可能会这样说:“我们李教授是顶风趣顶幽默的人,不久前中央电视台的晚会就请了李教授作报告。”尽管此时报告早已作过,但甲这时不会用“请李教授作了报告”的说法,因为甲的主旨在于说明中央电视台看中了(具体表现为“请了”)李教授的幽默风趣,所以请了他,提的要求正是“作报告”。

(三)对于(B)句来说,最适宜的语境应该是晚会策划者甲、乙两人当初正在策划的时候。甲提议说:“李教授很风趣很幽默,请李教授来作报告,一定效果很好。”焦点在“请”,“来作报告”是“请”的具体要求。当然,这时还可说“把李教授请来作报告”,这样说时“作报告”明显是目的而不是信息内容:“请来干什么?”而不是“请求什么?”

(四)对于(C)、(D)两句,最适宜的语境是相同的:在晚会现场,李教授的学生甲、乙两人都是观众,乙突然看见李教授在后台,惊奇地说:“李教授怎么在后台?”知道原委的甲回答:“策划人看中了李教授幽默风趣的演讲艺术,请了李教授来作报告(或:请来了李教授作报告)。”这时“作报告”的目的性质非常明显,而且恰好回答了乙的疑点“怎么”。这时(C)、(D)在语用上完全同义,并且又都有另外两个语用变体:a、把李教授请了来作报告,b、把李教授请来作报告了。注意:b句的“了”是附于全句的,表示整个邀请工作的完成而并不表“作报告”的实现,可见这四句的“作报告”都是可能世界的行为。这四句同B句及其变体的差别在于多一个“了”,这个问题在讨论(C)句的变换式时已谈过。这四句以及(B)句及其变体同(A)句之间的差别在于多一个“来”。“请”和“请来”的差别是:仅有“请”,其使令义是隐含的,由此决定“作报告”的目的性质也是隐含的,因而在句法上表现为表内容的宾语。而在“请来”中,使令义表达出来了,这就是“来”有了使动义,正是这个使动义

造成了“作报告”失去作为信息内容的性质，成了目的补语。因为“请来”的核心是“来”，是“用请的方法使他来”，“来”可以要求目的而不能要求内容。

(五) 对于(E)句，最适宜的语境是客观叙述，并且着重在分析晚会策划人的意图。例如策划人对朋友说：“当时，我们想到李教授非常幽默，为了增添晚会幽默的气氛，特意请来李教授作报告，果然效果很好。”李教授的幽默适于表现为“作报告”，这样，“作报告”的目的性质恰好突出了策划人的意图。

(六) 最后三句——(F)、(G)、(H)在语用上也是同义的，因此适宜的语境也相同，最基本的是：时间在李教授作过报告之后，另外“来”也决定了说话人是晚会的主人。我们假定是晚会的策划人在事后说：“李教授的幽默风趣早已闻名全国，这次晚会请李教授作了报告，果然效果很好。”此话主旨是用事实来证明李教授的名不虚传，相应的，事实“作了报告”是这里(F)(G)(H)三句的焦点，“请李教授”只不过是背景。另外，报告既已作过，说不说“来”也无关紧要。因此这三句的深层结构大体是一样的。

以上从语用角度验证了我们对“请”字兼语句理解策略的合理性。当然，基本式的语用变体还远不止上述8种，例如把“来”换为“去”，又会出现一批新的变体。但是我们的理解策略是立足于“请”的句法语义特征和“请 N_2V_2 ”的语法性质的，这个原则不论对哪一种变体，也不论对哪一种兼语句都是适合的。这一理解策略可以表述如下：由于 V_1 “请”的手段具有信息性质，所以，当 V_2 “作报告”在句中被表述为信息内容时，则 V_2 是 V_1 的宾语（称为“内容宾语”），这时“请 N_2V_2 ”是内容式，如基本式及(A)(B)式；当 V_2 不被表述为信息内容时，它就不是宾语而是补语。有两个因素的介入可以使 V_2 失去其信息性质：其一是改造 V_1 “请”使其信息手段不再显示作用（例如“请”后加“来”），从而使 V_2 的目的性质能显示到句法层面上来，使 V_2 成为目的补语，这时“请 N_2V_2 ”是目的式，如(C)(D)(E)式；其二是使 V_2 本身成为现实的、完成了的行为，从而成为 V_1 的结果补语，这时“请 N_2V_2 ”是结果式，如(F)、(G)、(H)式。上述理解策略使得兼语式完全可以归入人类大多数语言普遍具有的基本结构——动宾和动补结构之内。

以上理解策略的制定原则是：以语义理解为基础，以句法分析为核心，辅以语用的解释和检验。这个原则，实际上就是理解一切句子的共同原则，只有这样，才能合语言本身之理，才能处处有据，从而成为人和计算机都可理解、可操作的实实在在的策略。

参 考 文 献

- [1]、吕叔湘：《中国文法要略》，商务印书馆，1982年。
- [2]、王力：《中国现代语法》，商务印书馆，1985年。
- [3]、朱德熙：《语法答问》，商务印书馆，1985年。
- [4]、吕叔湘：《汉语语法分析问题》，商务印书馆，1979年。
- [5]、胡附、文炼：《现代汉语语法探索》，商务印书馆，1989年。
- [6]、石安石：《语义论》，商务印书馆，1993年。
- [7]、徐烈炯：《语义学》，语文出版社，1990年。
- [8]、(美)诺姆·乔姆斯基：《句法理论的若干问题》，黄长著等译，中国社会科学出版社，1986年。
- [9]、(波兰)沙夫：《语义学引论》，罗兰、周易译，商务印书馆，1979年。

“甚”的语法功能的聚类分析与词性演变模型的建立

王秋月

(清华大学)

摘要: 本文运用多元统计分析方法,对先秦两汉时期至魏晋南北朝时期的十二部古典文献中的“甚”字的语法功能进行了聚类分析。分析表明,在古典文献中,“甚”的语法功能主要有两大类:一是在先秦两汉时期,“甚”的主要语法功能是作谓语;二是在魏晋南北朝时期,“甚”的语法功能主要是作状语。从“甚”的语法功能演变可以推论出“甚”的词性演变,即在先秦两汉时期“甚”是形容词,而到了魏晋南北朝时期,“甚”的词性主要是副词。聚类分析方法不仅适用于汉语词性演变的建模研究,而且对古籍鉴定也有重要的参考价值。

关键词: 聚类分析方法, 汉语词性演变模型, 古籍断代鉴定

Study on modeling of SHEN's character development by analysis of its grammatical function using hierarchical clustering method

Wang Qiuyue

(Tsinghua University)

Abstract: Statistical analysis on the grammatical function of SHEN used in 12 ancient books from Qin and Han Dynasty to Wei and Jin Dynasty is carried out by hierarchical clustering method. The results show that SHEN has two grammatical functions: (1) predicate during Qin and Han Dynasty; (2) adverbial modifier in the Wei and Jin Dynasty. The model of SHEN's character development is inferred from the evolution of its grammatical function. SHEN is an adjective in Qin and Han dynasty, while it is an adverb during Wei and Jin dynasty. The hierarchical clustering method is useful not only in modeling the Chinese character development, but also in determining the publishing period of ancient book.

Key words: Hierarchical clustering method, Chinese character development model, division of ancient book into periods

一. 引言

“物以类聚，鸟以群分”。分类学是人类认识世界的一门基础学科。分类学发展至今经历了从多半凭经验和本学科的专业知识来进行分类的定性分类阶段到利用数学方法进行更科学的分类的数值分类学两个阶段。近几十年来，多元统计分析技术的引进，从数值分类学中逐渐分离出了聚类分析这个新分支。聚类分析已广泛应用于天气与气候预报、服装标准的制订、考古以及传统的地学研究[1][2][3]。

聚类分析应用于语言学研究则是随着计算语言学的诞生而得到迅速发展。美国伊利诺斯大学的郑锦全教授开创了计算机在汉语音韵和方言学上的应用研究[5]。特别是大型汉语语料库的建立，基于语料库的统计和规则相结合的汉语计算语言学研究方法，为汉语的自然语言理解与处理研究提供了强有力的工具，并取得了丰硕的成果[6]。

本文试图采用聚类分析方法，对先秦两汉至魏晋南北朝时期的十二部典籍中的“甚”字的语法功能进行统计分析，由“甚”的语法功能演变推论出其词性演变规律，从而以“甚”为例，探讨建立汉语词性演变模型的方法，为汉语的计算语言学研究与古文献鉴定等奠定基础。

二. “甚”的语法功能的统计与分析

研究汉语词性演变规律可以从考察词的语法功能的演变入手。以“甚”字为例。为了建立“甚”的词性演变模型，对先秦两汉至魏晋南北朝时期的十二部典籍中的“甚”字的全部用例，按照语法功能进行统计，结果如表1所示。其中，1-8号文献是先秦时期的作品；9号文献是东汉时期的作品；10-12号文献是魏晋南北朝时期的作品。

表1的统计结果表明，在上古汉语中，“甚”的语法功能大体上可分为两大类：一是在先秦两汉时期，“甚”的语法功能主要是作谓语和状语，作谓语的占用例的50%以上，例如：天之爱民甚矣（《左传·襄公十四年》），作状语的占用例的30%以上；例如：动刀甚微（《庄子·养生主》），二是在魏晋南北朝时期，“甚”的语法功能发生了重大变化，其作状语的用例高达90%以上，例如：佛法甚微（《法显传》）；众生甚苦（《百喻经》）。值得注意的是，《谷梁传》在先秦两汉作品中有特殊性的一面，其中“甚”的语法功能主要是作谓语。

在古代汉语中，形容词既能作谓语，又能作状语。因此，“甚”在先秦两汉时期是形容词。而到了魏晋南北朝时期，“甚”已演变成了副词[4]。

上述关于“甚”字的语法功能分类与词性演变分析主要是基于定性分类方法的，能否应用多元统计分析中的聚类分析方法得出同样的结论呢？