

中文社会科学引文索引(CSSCI)来源集刊

NANKAI LINGUISTICS

南开语言学刊

南开大学文学院 汉语言文化学院/编

2013年

第1期

(总第21期)



商务印书馆

南开语言学刊

Nankai Linguistics

2013 年第 1 期

(总第 21 期)

南开大学

文学院 汉语言文化学院 编



2013 年 · 北京

图书在版编目(CIP)数据

南开语言学刊. 2013 年. 第 1 期: 总第 21 期/南开
大学文学院, 汉语言文化学院编. —北京: 商务印书
馆, 2013

ISBN 978-7-100-10007-6

I. ①南… II. ①南… ②汉… III. ①语言学—丛刊
IV. ①H0-55

中国版本图书馆 CIP 数据核字(2013)第 121331 号

所有权利保留。
未经许可, 不得以任何方式使用。

NÁNKĀI YǔYÁN XUÉKĀN

南 开 语 言 学 刊

2013 年第 1 期(总第 21 期)

南开大学文学院 汉语言文化学院 编

商 务 印 书 馆 出 版

(北京王府井大街36号 邮政编码 100710)

商 务 印 书 馆 发 行

北京瑞古冠中印刷厂印刷

ISBN 978-7-100-10007-6

2013 年 10 月第 1 版

开本 787×1092 1/16

2013 年 10 月北京第 1 次印刷

印张 10 ½

定价: 27.00 元

《南开语言学刊》编辑委员会

顾问 王士元 刘叔新 黄正德
主编 马庆株 石 锋

编审委员会(按姓氏笔画顺序排列)

王洪君	冯胜利	邢向东	朱庆之	朱晓农	刘丹青
江蓝生	孙茂松	麦 耘	远藤光晓	李行德	李宇明
吴福祥	沈国威	沈家煊	张洪明	张 敏	陆丙甫
岩田礼	罗仁地	罗 端	袁毓林	徐大明	唐钰明
黄 行	曹志耘	储泽祥	游汝杰	蔡维天	潘悟云
魏培泉					

编务委员会(按姓氏笔画顺序排列)

马庆株	王红旗	石 锋	卢福波	李 兵	杨 琳
张文忠	张洪明	周 荐	施向东	郭继懋	曾晓渝
意西微萨·阿错					

执行编辑

王 萍 冉启斌

本刊实行双向匿名审稿制度

目 录

• 特稿 •

王士元:划时代的语言学家

- 《大江东去——王士元先生 80 华诞庆祝文集》前言 … 石 锋、彭 刚 1
- 语言学正面临战略转移的重要时刻 …… 冯志伟 7
- 试谈语言观和汉语观的更新——语言研究的当下与可能 …… 史有为 20

• 语法 •

- 汉语受事主语句范围的限定与中动句的特性 …… 王国栓 29
- 澳门立法语言中情态动词使用的若干问题 …… 苏小妹、马庆株 36
- 广东廉江粤语的补语标记与补语分类 …… 林华勇、马 喆 44

• 语音 •

- 论音节模式、结构类型对三音节单位词感的影响 …… 李晋霞、王忠玲 53
- 穗、港、澳三地粤语的语调格局——以陈述句的起伏度为例 …… 向 柠 65
- 超声波成像技术在生理语音学中的应用 …… 陈 彧 72

• 音韵 •

- 上古汉语突厥语对应词——以“合、盍、甲、夹”谐声系列为例 …… 王艳红 81
- 敦煌汉藏对音中部分溪母字读见母音的现象 …… 史淑琴 89
- 古音字表勘误三种(下) …… 郑张尚芳 95

• 应用语言学 •

不同年级学生词意识发展的研究	张兰兰、闫国利、白学军	112
对外汉语情态副词教学中的一些问题	关 键	121
中级汉语水平韩国留学生词义猜测的实证研究	刘善涛	129

• 译介 •

《鼻化元音在元音感知空间中的表现》述介	冉启斌、段文君	139
语言接触与汉语	马诗帆著;严艳群译	151
《南开语言学刊》稿约		19

Contents

Feature

A Preface to <i>Gone with the Yangzi River</i>	Shi, Feng; Peng, Gang 1
Linguistics: In the Important Time of Strategic Transit	Feng, Zhiwei 7
The New Progress of the Philosophical Views of Chinese and Language: The Present and Possibilities of Language Research	Shi, Youwei 20

Grammar

The Range of Subject Construction and the Quality of the Middle Construction	Wang, Guoshuan 29
Some Issues on the Modal Verbs in the Chinese Legislative Language of Macao	Su, Xiaomei, Ma, Qingzhu 36
The Markers of Verb-Complements and Categories of Complement in Lianjiang Yue	Lin, Huayong; Ma, Zhe 44

Phonetics

On the Restrictions of Syllable Pattern and Structure Type to the Word Sense of Three-Syllable Language Units	Li, Jinxia; Wang, Zhongling 53
Intonation Pattern of Yue Dialect in Guangzhou, Hong Kong and Macao: Take the Undulating Scale of the Intonation of Declarative Sentences for Example	Xiang, Ning 65
An Review of the Application of Ultrasonic Imaging Technology in Articulatory Phonetics	Chen, Yu 72

Phonology

- Ancient Chinese and Turkic Equivalents: “合(he)、盍(he)、
甲(jia)、夹(jia)” Harmonic Series Wang, Yanhong 81
- A Phenomenon of Some Initials Xi (溪母) Being Read as Initials Jian (见母)
in Sino-Tibetan Transliteration Sources from Dun-huang Shi, Shuqin 89
- Three Types of Collation of Ancient Chinese Characters’
Pronunciation Table(B) Zhengzhang, Shangfang 95

Applied Linguistics

- Research on the Development of Students’
Word Awareness Zhang, Lanlan; Yan, Guoli; Bai, Xuejun 112
- Teaching Nowadays of Chinese Modal Adverb in Teaching Chinese as
Second Language Guan, Jian 121
- Empirical Research of the Understanding of Words Meaning for
Intermediate Korea International Students Liu, Shantao 129

Review

- The Behavior of Nasalized Vowels in the Perceptual*
Vowel Space Ran, Qibin; Duan, Wenjun 139
- Language Contact and Chinese Yan, Yanqun 151

王士元:划时代的语言学家

——《大江东去——王士元先生 80 华诞庆祝文集》前言

石 锋 彭 刚

翻开语言学悠久的历史长卷,闪烁着一个个熟悉的名字。众多的语言学家把自己的人生和智慧投入到语言学研究之中,在探索语言奥秘的道路上,前赴后继,奋力前行。王士元先生是 20 世纪以来最具活力的语言学家之一。

《王士元先生 80 华诞庆祝文集》的编纂得到中国大陆、中国香港、中国台湾及欧美等众多语言学家的热情支持,汇集了 60 余篇当今语言学各领域的精美文章,分为中文、英文两辑,在王先生 80 寿诞之日正式发行。这是中国语言学界的一件大事,也是国际语言学界的一件大事。

诗人苏轼的“大江东去”四字,既简洁质朴又气象宏大,抒发了诗人对英雄豪杰的怀念,富含千古兴亡的哲理,自古以来,无人超越。文集名为《大江东去》,有两层含义:一是比喻王士元先生的语言学研究如滚滚长江奔涌东流,一江春水汇入大海;一是形容当代国际语言学研究从西向东的潮流气势磅礴,充满创新的活力。

我们庆祝王士元先生 80 华诞,不仅是因为王先生在语言学研究中的卓越贡献,更是因为王先生所代表的当代语言学的目标和方向。从国际语言学的现状和全局来观察,从现代语言学的历史和发展来思考,王士元先生 50 余年来的语言学研究工作给我们以珍贵的启示,是半个多世纪以来语言学行进足迹的投影。

一 从雅柯布森说起

索绪尔(Ferdinand de Saussure)和雅柯布森(Roman Jakobson)是 20 世纪语言学的两座高峰。然而,他们在中国的境遇却迥然不同。雅柯布森似乎还鲜为人知。这是在非正常的历史和社会条件下形成的亟待改变的学术盲区现象。

索绪尔因克服 19 世纪语言学的极端历史主义倾向而被誉为“现代语言学之父”。然而,一种倾向掩盖另一种倾向,索绪尔强调共时研究又走向另一极端。他“就语言而研究语言”,排除“所有属于‘外部语言学’的东西”(1980:20),提倡封闭的小语言学。语言研究越走越窄,走进象牙之塔。一个学科矫枉过正了。

雅柯布森是 20 世纪最有才华和最富想象力的语言学家与文学家之一。他受到索绪尔影

响,最早提出结构-功能的分析方法,奠定了布拉格学派的理论基础,被誉为“结构主义之父”。他把结构作为分析方法,而不是研究目标。他把人类学引入语言学,使诗学和语言学结合,到医院里考察失语症(1941),与声学家合作提出语音区别特征(1951)。雅柯布森提倡开放的大语言学。他不是简单地重复索绪尔的词句,而是真正继承发展了索绪尔创新的精髓。他是一位跨学科、跨领域、跨学派的开放式语言学研究的先驱。

王士元先生的语言学理念跟雅柯布森一脉相承,从语言演化的高度,把考古学、遗传学跟语言学作为三个窗口(1998),从广度和深度上开拓了语言学研究的领域。王先生开阔的跨学科视野集中表现在两个方面:一是王先生最早把语言学和计算机结合起来,使得语言学插上了现代科技的翅膀;二是王先生把人类在脑科学上的进展应用于语言演化研究,一直探索在语言研究的最前沿。

早在20世纪60年代,王先生与郑锦全等同事率先做成第一部汉语方言计算机词典,即最早的语言资料数据库。词汇扩散理论的创立就是当时以这个数据库为基础而引发的,是最早的计算机和语言学结合得到的成果。他曾经跟世界著名的数学家弗里曼(D. A. Freeman)合作利用概率计算的方法解决人类语言的单起源还是多起源的问题(1996)。近年来他利用计算机建模仿真方法模拟语言起源的约定俗成过程(2001),还成功地把计算机建模用于研究语言演化中横向的和纵向的传播(2005)。他跟“遗传算法之父”贺兰(John H. Holland)密切合作,推进了对语言这个复杂适应系统的建模仿真研究。

王先生从20世纪60年代开始进行汉语声调的听感实验(1967),还与著名心理语言学家曾志朗等同事探究语言与大脑的联系(1977),开启了汉语和汉字的心理语言学与神经语言学研究。他们70年代在加州大学伯克利校区合开了一门课“语言的生物基础”,多年后曾先生还怀念那段彼此切磋的精彩时光带来的“淋漓尽致的畅快”。王先生一直没有停止前进的脚步,他的《语言、演化与大脑》(2011)、《演化语言学论集》(2013)以及一系列论文反映了研究的最新进展。

王先生以他丰硕的独创性的研究成果,走出了一条多学科结合的语言研究的新路,改变了当代语言研究的面貌,使语言学跟人类科学同步发展。语言学理论体系以及研究方式由此引发的深刻变革将随时间的推移而愈加显现。

二 跟赵元任同事

赵元任先生是现代中国语言学的开拓者。赵先生的研究具有多面性。赵先生跟雅柯布森很相近,雅柯布森是语言学家兼诗学家,赵先生是语言学家兼音乐家。他用田野调查和实验方法研究方言,录制国语留声片,设计国语罗马字,创制五度值,提出音位分析的多能性(1934),出版《中国话的文法》,被誉为“汉语语言学之父”。

王士元先生早年应赵先生之约,从俄亥俄转到伯克利,成为赵先生的同事和忘年交。赵先生与王先生都来自中国,一个来自常州,一个来自上海;他们大学都是理工科,一个是数学,一个是电子;最重要的是他们都有一颗中国心,一份赤子情,以振兴中国语言学为己任。

王士元先生倾力推动中国语言学发展。王先生在俄亥俄州立大学同时创设了语言学与东亚语言文化系。他在加州大学伯克利分校创建的音系学实验室是美国最早的语音实验室之一。他创办的《中国语言学报》(*Journal of Chinese Linguistics*, 简称 JCL) 是中国语言学第一个国际性的刊物, 创刊号就是献给赵元任先生的, 同时刊载了“非汉语语言学之父”李方桂先生关于汉藏语系属的经典文章。王先生曾在伯克利主持语言分析研究所 (Project on Linguistic Analysis, 简称 POLA) 三十余年, 使 POLA 成为中国语言学的标志。很多前辈学者至今还在心中珍藏着“永远的 POLA”(2005)。他担任国际中国语言学学会首届会长, 在美国语言学会所办的暑期讲习班以及圣菲研究所 (Santa Fe Institute) 的暑期课程都曾多次授课, 他还全力帮助在中国举办首次中国语言学暑期讲习班并亲临南开大学讲学。

王先生曾亲赴吉尔吉斯斯坦考察东干语跟汉语西北方言的联系, 曾亲赴泰国北部、中国南方调查民族语言, 完成汉藏语言的计算机集群分析。前面已经讲到, 他探索汉语声调和汉字的心理认知和大脑加工机制, 以汉语词汇句法的资料进行语言演化的计算机仿真建模研究, 获得多种成果, 使汉语研究立于国际语言学之林。

王士元先生还为中国语言学培养了众多人才, 桥本万太郎、余蔼芹、陈渊泉、谢信一、柴谷方良、廖秋忠、邹嘉彦、连金发、沈钟伟、小仓美惠子、孔江平、彭刚、汪锋、龚涛等等, 都出自王先生门下。王士元先生待人谦和真诚、胸怀宽广, 很多人都是在他的直接或间接的启迪和影响下, 走上了语言学的道路。

赵先生在 20 世纪 30 年代从汉语方言的语音资料引发出音位分析多能性理论; 王士元先生在 60 年代据汉语方言计算机词典创立词汇扩散理论, 完成了中国语言学影响国际语言学的又一次重要的创新。词汇扩散理论的主旨是, 语音的演变不只是音类的转移, 而是与词汇有互动关系, 某一音变其实就是在这一音类所辖的语词中随时间扩散的过程。词汇扩散理论对语言演变的微观或宏观的研究提供了开创性的研究方法和理论模式。

斯坦福大学教授、遗传学家卡瓦里-斯佛扎 (Cavalli-Sforza, 1994) 认为: “词汇扩散理论已经公认为是语言演化研究中最重要创新理论。”中国语言学成果得到国际学术界杰出科学家的认同, 已经进入各种理论语言学论著和百科全书。

哈佛大学教授、人类学家张光直 (1994) 曾这样评论 20 世纪中国人文社会学科对国际学术界的贡献和影响: “只有在语言学上, 中国语言学者如赵元任、李方桂、王士元等先生的著作在普通语言学的书刊里给人引用。”在这一方面, 王士元先生与张光直先生有着共同的理念: “争取主流地位, 甚至可以说是中国人文社会科学家对于世界的责任。”

如果说 19 世纪的语言学瞩目欧洲, 20 世纪的语言学瞩目美国, 那么, 21 世纪的语言学将开始瞩目中国。这其中离不开王士元先生的努力, 他为中国语言学引入了新观念和新方法, 为多学科发展注入了巨大的活力。王士元先生是当代中国语言学的开拓者之一, 是划时代的国际语言学家。

三 与拉波夫对话

2012 年 5 月在香港中文大学有一场别开生面的世纪学术盛事——拉波夫和王士元两位

语言学巨匠以“音变:过去、现在和未来”为主题的深度对话。他们在几十年的学术生涯中,各自从不同的起点出发,朝着相同的方向,构建了变异的演化的语言观,如今相聚在一起。

拉波夫(William Labov)在社会中研究语言的变异,主张语言研究不应仅限于语言内部,更要考察各种社会因素如何影响语言的演变,拨转了索绪尔以来的语言学方向。拉波夫调查语言不仅到田野中,而且到街区中、商店里,获取成百上千的样品,做实验、做统计。他借助庞大的资料库阐明语言是异质有序的大系统。拉波夫为语言变化研究开拓了一个新方向。

王士元先生提出词汇扩散的理论(1969),与文莱奇(Weinreich)和拉波夫等人发表语言变异文章(1968)几乎是在同时。拉波夫在1971年就任美国语言学会主席的任职演说中最早肯定了词汇扩散理论的意义。他在后来的三卷巨著《语言变化原理》开头致谢中写道:“本书的许多研究、调查、论述都是对于王士元在竞争性语音演变、词汇扩散、方言混合方面极具远见卓识的开创性研究工作的响应。……虽然我自己的团队在这些问题上的研究跟他们不都一样,但是他们理论观点的冲击力和与之密切相关的语言资料对我影响至深。”(1994:xvi)

确实,拉波夫的《语言变化原理》采用词群(word class)分析方法,根据共有同一音位的词群成员的变化来考察语音的演变。虽然在具体的方法上不完全一样,但是在宏观的思维方式上跟词汇扩散理论具有本质的一致性。王先生讲过,词汇扩散理论预测所有活的语言都会呈现大量的“有序异质性”(orderly heterogeneity),既肯定了拉波夫的说法非常有用,又阐明了词汇扩散与语言变异之间的密切联系。王士元和拉波夫的思想是一种全新的语言观和语言演化观。在这个意义上,王士元和拉波夫代表着语言学的新变革、新方向。

王士元先生的词汇扩散理论实际上是运用生物进化论的变异-选择的观念解释语音演变,王先生由此走上演化语言学之路。他认为:语言演化和生物演化有许多重要的共同点,很可能物理界和社会体系的演化之间,也存在诸多的共同点。语言就是一种复杂适应系统(complex adaptive systems)。在这个理念下,王先生的研究团队对于语言起源、词汇涌现、声调发展、词汇和句法共变、语言消亡等语言学重要问题进行计算机建模及多个体仿真模拟(multi-agent simulation)分析,取得了新进展,拓展了新天地。

从最远古的语言产生到最现代的大脑机制,都汇集为演化语言学。在这个方向上,语言的系统性和社会性完美整合,语言的共时研究和历时研究高度统一。语言学者的天职是解开语言之谜,揭开语言演化之谜。语言之谜也就是人类之谜。在这个意义上,研究语言也就是研究人类自己。难怪王先生感慨:不识语言真面目,只缘身在言语中。

四 结语

量子论创始人普朗克指出:“科学是内在的整体,它被分解为单独的部门不是取决于事物的本质,而是基于人类认识能力的局限性。实际上存在着从物理学到化学,通过生物学和人类学到社会学的链条,这是任何一处都不能被打断的链条。”语言研究的不同领域从整体上来看,就是这个总的链条上从人类学到社会学之间的一环。

王士元先生讲过:“不同学科之间的边界犹如画在沙滩上的线条,随着每一次先进知识的

波涛到来,这边界就会发生变化,甚至完全消失。人类的知识,特别是研究语言的知识,应该是彼此相连的,并且最终是相互贯通的。”跨学科、泛学科的开放式研究是当代科学的重要特征。

赢得人类学、基因学、数学、心理学等其他学科中最杰出学者的肯定和认同,是王士元先生研究工作的一个重要特征。王先生多年来致力于在多学科视野下研究语言。这就是他的巨大学术魅力的来源:总是充满活力地站在研究的最前沿。

参考文献

- 索绪尔 1980 《普通语言学教程》,高名凯译,商务印书馆。
- 王士元 2010 《王士元语音学论文集》,世界图书出版公司。
- 2011 《语言、演化与大脑》,商务印书馆。
- 2013 《演化语言学论集》,商务印书馆。
- 王士元、柯津云 2001 《语言的起源及建模仿真初探》,《中国语文》第3期;另载王士元《语言学论文集》,2002。
- 曾志朗 2004 《人人都是科学人》,台北远流出版公司。
- 张光直 1994 《中国人文社会科学该跻身世界主流》,《亚洲周刊》7月10日。
- Cavalli-Sforza 1994 An evolutionary view in linguistics. In Matthew Chen and Ovid Tzeng (eds.) *Interdisciplinary Studies on Language and Language Change*. 17–28. Taipei: Pyramid Press.
- Chao, Yuen Ren 1934 The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 4:363–397.
- Freedman, D. A. & W. S-Y. Wang 1996 Language polygenesis: A probabilistic model. *Anthropological Science* 104.2:131–138.
- Ho, D.-A. & Tzeng, O. J. L. (eds.) 2005 *POLA Forever: Festschrift in Honor of Professor William S-Y. Wang on His 70th Birthday*. Taipei: Institute of Linguistics, Academia Sinica.
- Jakobson, Roman 1941 *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell, Allan, R. Keiler, trans. 1968. *Child Language, Aphasia, and Phonological Universals*. The Hague.
- Jakobson, Roman, Gunnar M. Fant & Morris Halle 1951 *Preliminaries to Speech Analysis*. MIT Press.
- 王力译,《语音分析初探:区别性特征及其相互关系》,《国外语言学》,(1981)。
- Labov, William 1994 *Principles of Language Change*. Vol. 1: Internal Factors, Blackwell Publishing Ltd.
- Tzeng, Ovid J. L., D. L. Hung & Wang, William S-Y. 1977 Speech recording in Chinese characters. *Journal of Experimental Psychology: Human Learning and Language* 3.6: 621–630. Reprinted in Wang, W. S-Y. 1991. *Explorations in Language*. 249–262.
- Wang, W. S-Y. 1969 Competing changes as a cause of residue. *Language* 45.1:9–25. Reprinted with postscript in P. Baldi and R. N. Werth (eds.) *Readings in Historical Phonology: Chapters in the Theory of Sound Change*. 236–257. Pennsylvania University Press. 1978. Reprinted in Wang, W. S-Y. 1991. *Explorations in Language*. 3–19. Also in Wang 2010. 132–153.
- 1991 *Explorations in Language*. Taipei: Pyramid Press.
- 1998 Three windows on the past. In Mair, V. (ed.) *The Bronze Age and Early Iron Age Peoples*

and Eastern Central Asia. University of Pennsylvania. 508—534.

- Wang, W. S-Y. & J. W. Minett 2005 Vertical and horizontal transmission in language evolution. *Transactions of the Philological Society* 103. 2:121—146. 张静芬译,《语言演变中的横向传递与纵向传递》,见王士元(2013)。
- Wang, W. S-Y. & Kung-Pu Li 1967 Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10. 3. 焦立为译,《北京话的第三调》,见《语言的探索——王士元语言学论文选译》,北京语言大学出版社。
- Weinreich, Uriel, W. Labov & M. Herzog 1968 Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel (eds.) *Directions for Historical Linguistics*. 97—195. University of Texas Press.

(石 锋 南开大学语言所 300071; 彭 刚 香港中文大学现代语言和语言学系)

语言学正面临战略转移的重要时刻

冯 志 伟

提 要 当前计算语言学中已经实现了战略转移。本文首先分析了计算语言学中所实现的这一场战略转移,说明了在传统语言学中也有必要进行这样的战略转移,并指出了在这场战略转移中,应当把理性主义方法和经验主义方法结合起来。

关键词 计算语言学 传统语言学 规则推导方法 语料库方法 理性主义 经验主义

一 计算语言学中的战略转移

20 世纪 90 年代以前,从事计算语言学研究的绝大多数学者,都把自己的目标局限于某个十分狭窄的专业领域之中,他们采用的主流技术是基于规则的句法-语义分析,尽管这些应用系统在某些受限的“子语言”(sub-language)中也曾经获得一定程度的成功,但是,要想进一步扩大这些系统的覆盖面,用它们来处理大规模的真实文本,仍然有很大的困难。因为从自然语言处理系统所需要装备的语言知识来看,其数量之浩大和颗粒度之精细,都是以往的任何系统所远远不及的。而且,随着系统拥有的知识在数量上和程度上发生的巨大变化,系统在如何获取、表示和管理知识等基本问题上,不得不另辟蹊径。

这样,就提出了大规模真实文本的自动处理问题。1990 年 8 月在芬兰赫尔辛基举行的第十三届国际计算语言学会会议(即 COLING'90)为会前讲座确定的主题是:“处理大规模真实文本的理论、方法和工具”。这说明,实现大规模真实文本的处理将是计算语言学在今后一个相当长的时期内的战略目标。为了实现战略目标的转移,需要在理论、方法和工具等方面实行重大的革新。

当前语料库的建设和语料库语言学的崛起,正是计算语言学战略目标转移的一个重要标志。随着人们对大规模真实文本处理的日益关注,越来越多的学者认识到,基于语料库的分析方法(即经验主义的方法)至少是对基于规则的分析方法(即理性主义的方法)的一个重要补充。因为从“大规模”和“真实”这两个因素来考察,语料库才是最理想的语言知识资源。但是,要想使语料库名副其实地成为自然语言的知识库,就有必要首先对语料库中的语料进行自动加工,以便于人们从中提取丰富的语言知识。

计算语言学现在面临着一场战略转移(strategic transit)。这场战略转移的关键是知识的

获取方式和方法:从依靠“自省”(introspection)方式转向依靠“语料”的方式,从基于“规则”的方法转向基于“语料库”的方法(也就是基于“统计”的方法)。

目前,计算语言学中的基于统计的方法已经获得了很大的成功。2003年7月,在美国马里兰州巴尔的摩(Baltimore, Maryland),由美国商业部国家标准与技术研究所 NIST/TIDES (National Institute of Standards and Technology) 主持的评比中,德国 Aachen 大学年轻的博士生 Franz Josef Och(奥赫)获最好成绩,他使用统计方法,在很短的时间之内就构造了阿拉伯语和汉语到英语的若干个机器翻译系统。

伟大的希腊科学家 Archimedes(阿基米德)说过:“只要给我一个支点,我就可以移动地球。”(“Give me a place to stand on, and I will move the world.”)而现在 Och 也模仿着阿基米德说:“只要给我充分的并行语言数据,那么,对于任何的两种语言,我就可以在几小时之内给你构造出一个机器翻译系统。”(“Give me enough parallel data, and you can have translation system for any two languages in a matter of hours.”)这反映了新一代的机器翻译研究者朝气蓬勃的探索精神和继往开来的豪情壮志。看来,Och 似乎已经找到了机器翻译的有效方法,至少按照他的路子走下去,也许有可能开创出机器翻译研究的一片新天地,使我们在探索真理的曲折道路上看到耀眼的曙光。

早在 1947 年,美国洛克菲勒基金会的副总裁 Weaver(韦弗)在他的以《翻译》为题的备忘录中,就提出了使用解读密码的方法来进行机器翻译,他认为翻译类似于解读密码的过程。他说:“我可以这样说,一本用中文写的书实际上是用英语写的,只不过它是用中文的代码符号编了码而已,这样的说法是很有诱惑力的。”(Weaver,1949)他的这段话非常重要,后来以不同的形式广为流传。

下面。我们举例说明 Weaver 的思想。

如果我们有一篇看不懂的文件,上面写满了奇奇怪怪的密码(图 1)。

	e	e	e	e
<u>ingcmpngsnwf</u>	<u>cv</u>	<u>fpn</u>	<u>owoktvcv</u>	
<u>hu ihgzsnwfv</u>	<u>rqcffnw</u>	<u>cw</u>	<u>owgcnwf</u>	
<u>kowazoanv</u>	...			

图 1

	e	e	e	e
<u>ingcmpngsnwf</u>	<u>cv</u>	<u>fpn</u>	<u>owoktvcv</u>	
	e	e	e	
<u>hu ihgzsnwfv</u>	<u>rqcffnw</u>	<u>cw</u>	<u>owgcnwf</u>	
	e			
<u>kowazoanv</u>	...			

图 2

我们可以假定这段密码原来是用英语写的,只不过在传输过程中被一种特殊的符号转写了,因此变成了我们看不懂的密码。为了解读这段密码,我们首先假定,密码中的字符 n 是由英语中的 e 转换来的,于是,我们得到图 2:

接着,我们又可以假定密码中的字符串 fp 是由英语中的 th 转换来的,这样,密码中的字符串 fpn 就可能是从英语的 the 转换来的,我们得到图 3。

	e	e	e	the
<u>ingcmpngsnwf</u>	<u>cv</u>	<u>fpn</u>	<u>owoktvcv</u>	
	e	e	e	
<u>hu ihgzsnwfv</u>	<u>rqcffnw</u>	<u>cw</u>	<u>owgcnwf</u>	
	e			
<u>kowazoanv</u>	...			

图 3

	e	he	e	of the	fof
<u>ingcmpngsnwf</u>	<u>cv</u>	<u>fpn</u>	<u>owoktvcv</u>		
	e	f	o	e	o
<u>hu ihgzsnwfv</u>	<u>rqcffnw</u>	<u>cw</u>	<u>owgcnwf</u>		oe t
	e				
<u>kowazoanv</u>	...				

图 4

这时,密码中的 f 对应于英语的 t,p 对应于英语的 h,我们把这个转换推广到整篇的密码。我们进一步假定密码中的字符串 cv 对应于英语的 of,并把密码中的 cv 对应于英语 of 的假定推广到整篇密码,得到图 4。

这样的推广似乎有些问题,vcv 转换成了英语的 fof,而在英语中很少出现 fof 这样的字符串,于是我们推翻 cv 对应于英语 of 的假定,假定密码中的 cv 对应于英语的字符串 is,并且把这样的假定推广到整篇密码,最后我们可以得到这篇密码对应的英文,如图 5 所示:

```
decipherment is the analysis
ingcmpngsnwf cv fpn owoktvcv
of documents written in ancient
hu ihgzsnwfv rqcffnw cw owgcnwf
languages ...
kowazoanv ...
```

图 5

通过解读密码,我们得到的与这篇密码相对应的英语:

Decipherment is the analysis of document written in ancient language ...
(解读密码就是分析用古代语言写的文献……)

解读密码(decipherment)是古典文献研究的一个重要内容,历代学者们曾经依靠自己的聪明才智出色地解读了不少古代的铭文,或者通过铭文中已知的部分来解读铭文中未知的文字。罗塞塔石碑(Rosetta Stone)上古代埃及文字的解读,能够给我们使用解读密码技术来进行机器翻译不少的启发。

罗塞塔石碑由上至下共刻有同一段诏书的三种语言版本,是用埃及象形文字(Egyptian hieroglyphs,又称为圣书体,代表献给神明的文字)、埃及通俗文字(Egyptian demotic,又称草书体,是古代埃及平民使用的文字)和古希腊文(Greek,代表统治者的语言,这是因为当时的埃及已臣服于希腊的亚历山大帝国之下,来自希腊的统治者要求统治领地内所有的此类文书都需要添加希腊文的译版)三种不同的文字写成的,刻于公元前 196 年,现藏大英博物馆。

罗塞塔石碑如图 6 所示:

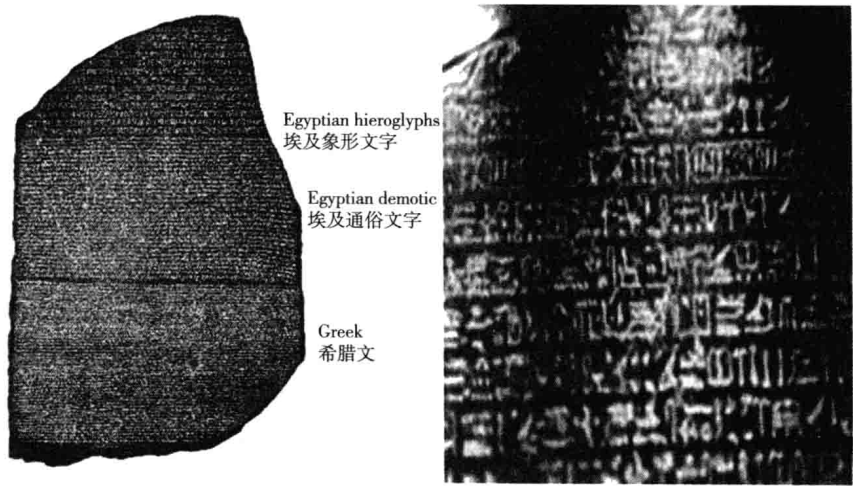


图 6 罗塞塔石碑(右侧为埃及象形文字之局部)