

语料库与社会语言学研究方法

苏金智 肖航

(教育部语言文字应用研究所, 北京 100010)

[摘要] 语料库和语料库方法的出现对社会语言学书面语研究起到了重要的促进作用, 用语料库进行社会语言学研究还有很多改进的空间。利用国家语委语料库检验 134 个汉语音译词的变化情况得到的结果说明, 这些汉语音译词被汉语固有词代替的结论基本上符合现代汉语的语料实际, 这表明语料库方法对社会语言学的语言变化趋势分析具有重要作用, 而 119 个音译词查不到频次说明语料库存在不足之处。建设社会语言学需要的语料库, 就要全面了解语料库的作用以及对语料库的需求。社会语言学研究所需要的语料库除了具有一般语料库所具有的特点外, 还应该包含丰富完整的社会语言学信息, 其中部分信息应该通过标注手段获得。有了好的语料库, 利用语料库和语料库方法研究社会语言学, 将会进一步推动社会语言学研究定量研究与定性研究的完美结合。

[关键词] 语料库; 社会语言学; 方法论; 标注

Corpus and Sociolinguistic Methodology

Su Jinzhi Xiao Hang

(Institute of Applied Linguistics, Ministry of Education, PRC, Beijing 100010, China)

Abstract: Methodology of sociolinguistics is different from traditional structural linguistics, in particular, from the Chomsk's transformational-generative grammar. The distinctive methodological feature of Sociolinguistics is to analyze language elements in the speech communities with qualitative and quantitative approaches. In the earlier studies, sociolinguists paid much more attention on spoken forms because these forms could be analyzed in the speech communities and be easily explained from the relationship between language variables and sociolinguistic variables such as ethnic group, age, social class and gender. On the contrary, the study of written forms got inattention because they could not be researched as fully as spoken forms. Collecting data from the literatures, which is the traditional method used in exploring written forms in sociolinguistic, had restricted it form further development. Corpus and corpus-based approach has turned out to be useful in the study of written language since they emerged. With the use of corpus approach, a large number of achievements have been made such as in the studies of gender language, Chinese speech communities words in mainland China, Taiwan, Hong Kong and Macao, Chinese register,

[收稿日期] 2011-09-15

[本刊网址·在线杂志] <http://www.journals.zju.edu.cn/soc>

[在线优先出版日期] 2012-04-27

[作者简介] 1. 苏金智, 男, 教育部语言文字应用研究所研究员, 中国社会科学院研究生院教授, 博士生导师, 哲学博士, 主要从事社会语言学研究; 2. 肖航, 男, 教育部语言文字应用研究所副研究员, 主要从事计算语言学和语料库语言学研究。

and discourse analysis. These studies are pioneering ones and have accumulated experiences for the future study, but they are still limited and need to be improved in some sense. The result of loanwords studies by Su (2010) shows that 134 Chinese phonetic matching loanwords, one percent of the total loanwords in *A Dictionary of Loan Words and Hybrid Words in Chinese* (1984), have already been replaced by Chinese indigenous words, but this finding was not testified by the contemporary Chinese corpus. The result of the reexamination of the changing histories of 134 Chinese phonetic matching loanwords by using the Contemporary Chinese Corpus established by State Language Committee of PRC supports Su's finding and reinforces the function of corpus approach in analyzing the trends of language change. Another fact that 119 Chinese phonetic matching loanwords have zero frequency when they were searched in this corpus suggests that there are some limitations in corpus, such as it is unbalanced with time spans or no Chinese phonetic matching loanwords is tagging in this corpus. Before using corpus and corpus approach in sociolinguistic studies, one should understand the comprehensive functions of corpus and what and how corpus can help the research in sociolinguistics. In addition to representativeness, balance and good distribution as a general corpus has, the corpus sociolinguistics needs should have: 1) Universality, namely including various styles or registers, various language variants from different domains and different regions; 2) much more sociolinguistic information about language producers such as ethnic group, age, social class and gender, part of which should be able to be attained from sociolinguistically annotation; 3) sociolinguistic annotation. If there is a better corpus, sociolinguistic studies with the help of corpus and corpus approaches will enhance a perfect combination of qualitative analysis and quantitative analysis.

Key words: corpus; sociolinguistic; methodology; annotation

一、引言

社会语言学之所以成为一门独立的学科,不仅在语言观方面与传统结构语言学,尤其是转换生成语法学派不同,在研究方法上也有较大的差别。关于语言观上的不同,本文不加以讨论,本文重点讨论方法上的差别,尤其是语料库方法对社会语言学研究的促进作用。《语言学方法论》一书把语言学的研究方法分为上中下三篇,上篇为“理论方法篇”,中篇为“描写方法篇”,下篇为“实验方法篇”。社会语言学研究方法部分放在中篇“描写方法篇”,主要涉及社会语言学的抽样方法、数据的收集、描写与分析^[1]。社会语言学的研究方法主要是描写的还是解释的,我们这里也暂且不加以讨论。我们认为从社会语言学角度对数据进行分析描写是有必要的,但仅仅是分析描写远远不够。《社会语言学研究方法的理论与实践》一书把社会语言学研究方法的特征概括为三个:定性与定量研究结合、解释性、实证性。该书的这一概括还是比较全面的,但在解释“语料或数据”时,认为定性研究所收集的数据主要是词语而不是数字,定量研究所收集的数据主要是数字^[2]。这种解释似乎不够全面,利用语料库进行社会语言学研究,收集的数据既需要词语,也需要数字。该书还没有把语料库方法作为社会语言学的一种重要方法,没有看到语料库对社会语言学研究的重要作用。《社会语言学教程》第六章设专章讨论语料库与社会语言学研究,认为语料库语言学为社会语言学带来崭新的研究工具^{[3]131}。我们认为,随着现代信息技术的不断发展,利用语料库方法研究社会语言学将会是一种定量研究与定性研究完美结合的社会语言学研究的重要方法。

二、基于语料库方法的社会语言学研究现状

社会语言学研究在方法上的最大特点是在对言语社区语言使用进行定量和定性分析的基础上,揭示语言变量与社会变量之间的关系。早期的社会语言学研究更多的是注重口语的研究,因为口语的研究比较容易同言语社区结合起来,从而探讨语言变异与民族、年龄、社会阶层、性别等社会变量之间的关系。而这类研究一般要进行社会调查,调查一般采用的是抽样的方法,主要是判断抽样的方法。早期书面语的研究一般采用文献调查的方法,因为缺乏先进的手段,在语言变异研究方面很难有较大的突破并取得重大的成果。语料库语言学的出现不但给社会语言学研究书面语带来了生机,同时也大大推动了这一领域的定性定量有机结合的分析研究。

McEnery 和 Wilson 的《语料库语言学》(*Corpus Linguistics*)一书第4章第7小节“语料库在语言研究中的应用”专门介绍了语料库对社会语言学研究方法的影响。作者在探讨一般意义上的社会语言研究所收集的语料与语料库的语料之间的区别时指出,社会语言研究所收集的是专门性的语料,而不是一般的来自自然语言的语料,这些语料不一定是用来作定量分析用的,因此不一定经过严格的抽样;而语料库却可以提供来自自然语言的具有代表性的样本,且可以用来作定量分析。当然该书作者也指出,利用语料库方法进行社会语言学研究还只停留在相对简单的词汇层面上的语言与性别方面的研究^[4]。

到目前为止,利用语料库方法进行社会语言学研究取得重要成果并值得关注的主要有四个方面:一是语言与性别的研究,二是汉语社区词的研究,三是语域研究,四是话语分析方面的研究。

Kjellmer^[5]和 Holmes^[6]利用语料库方法进行语言与性别研究取得了一些重要成果。Kjellmer 利用 Brown 语料库和 LOB 语料库检验美国英语和英国英语中的男性偏爱。他专门检索了男性代词与女性代词(he 和 she)的出现情况以及词项 man/men 与 woman/women 的出现情况,结果发现在两个语料库中女性词项出现频率均大大低于男性词项的出现频率,但英国英语的女性词项出现频率高于美国英语。他还发现这种差异的比率是由体裁决定的,总的来说,女性更富于想象,爱情小说中女性出现的频率最高。Kjellmer 还发现,女性相对缺少主动性,但经常出现客观动词而不是主观动词的假定是没有根据的,实际上男女都有类似的主客观比率。Holmes 对比了 Ms 与 Miss/Mrs 的出现频率,分析了带有男权主义色彩的后缀和作为普通类别词语 man 的用法。Holmes 在方法上有两个重要看法值得注意:一是要注意语境和是否有合适的替代成分,如 policeman 或 policewomen 可以用 police officer 代替,但在 Duchess of York(约克的公爵夫人)中后缀-ess 没有合适的替代成分,因此这类后缀在计算男权主义的后缀时不应该算进去。二是指出语义形式的分类很难。例如 man 什么时候指男性,什么时候作为普通的类(不分性别),判别起来并不容易。

邹嘉彦等利用香港城市大学语言科学资讯研究中心在20世纪90年代建立起来的共时语料库(Linguistic Variety in Chinese Communities, LIVAC)的基础上,进行了中文五地(香港、澳门、上海、台湾和新加坡)社区词的研究,取得了一批令人瞩目的成果。该语料库不仅可以从共时平面进行语言变异的研究,也可以从历时的角度进行语言变化的研究。《社会语言学教程》第六章“基于语料库的社会语言学研究”专门讨论如何利用语料库进行社会语言学研究。该章利用这一共时语料库从共时角度比较了各地中文词汇使用的不同情况,分析了新词的发展变化,并从港澳报刊地名词的演变观察社会的变化,开拓了我国社会语言学变异研究的新路子^[3]。

黄昌宁等认为,语料库方法是解决不同语域的语言特性描写问题的最好方法,并且介绍了 Biber 进行口语和书面语两种不同语域变体的研究^[7]。语域(register)是社会语言学的一个重要概念,陈瑞端把这个英文词翻译成“语体”,并对大陆、香港和台湾三地的科学语体、事务语体和报道语

体的许多语言成分进行了定量分析^①。这些研究是社会语言学利用语料库方法在语域研究方面的开拓性研究,为我们积累了经验。

桂诗春等介绍了话语分析研究中的语料库方法^[1]。Hope 把语料库方法与社会语言学的研究结合起来研究一些莎士比亚有争议的剧作者的归属问题。Hope 通过第二人称代词和助动词的使用,观察作家的出生年份和使用规则化句子的关系来判定作品的归属^[8]。Sotillo 和 Wang-Gempp 的《运用语料库语言学的方法探讨网络政治话题讨论中的阶级、意识形态、说话跑题》,对一个有 46 300 个词的语料库的称呼语、夸张用法、词汇选择、词语搭配、认知动词和人称代词的使用情况进行了分析,其语料来自美国北新泽西五个城镇居民有关政治话题的在线讨论。该研究发现,词汇选择(lexical choices)和负面的语义韵律(negative semantic prosody)可用于破坏候选人的形象,第一人称和第二人称单数代词比第一人称和第三人称复数代词出现的频率高。参与者用第二人称代词完成以下四个语用功能:劝说有可能投票的人;告诫某一候选人不要出现;排除不同政见;在更广泛的意义上与网民交流^[9]。我国学者也重视利用语料库的方法进行话语分析研究,马博森在《当代中国话语语料库的建构问题》一文中认为运用语料库语料及语料库分析方法研究不同类型的话语,可以帮助人们进一步认识语言与社会之间的关系。他还提出了建构当代话语语料库的基本设想及总体框架。其总体框架是分别建书面语语料库和口语语料库。书面语又根据语域和体裁建子语料库,口语分社会话语和家庭话语两大类建库^[10]。

回顾过去的研究,其重点主要集中在词汇层面上,语法方面和话语方面虽然也出现了一些成果,但从总体上看还不能与词汇研究相比,这与语料库的设计和建设有直接关系,因此,利用语料库方法进行社会语言学研究还有许多开拓空间。

三、利用语料库检验汉语外来词变化的例子

苏金智对刘正焱等编的《汉语外来词词典》^[11]中汉语外来词的消长进行了研究,判断其中约占词典 1% 的 134 个音译词基本上被汉语固有词语所取代^[12]。这篇文章立论时只是将是否收入《现代汉语词典》作为标准,没有语料库的支持。现在我们通过国家语委现代汉语语料库的查询系统^②对 134 个音译词逐个进行查询,看看这些词语的出现情况,以证明这一判断是否正确。国家语委现代汉语语料库是一个大型的国家的平衡语料库,包含 1 亿字生语料库、5 500 万字标注语料库、100 万字句法树库、语料库建设加工系列软件和语料库查询检索系统^[13]。语料库选材时间跨度为 1919 至 2003 年,1997 年以前约 7 000 万字,1997 年以后约 3 000 万字。

通过该检索系统对这 134 个汉语音译词的出现情况逐一进行搜索,其中只有 15 个词能统计到词频,如表 1 所示,其余 119 个词未统计到词频。

表 1 统计到词频的音译词

序号	音译词	外文原形词	语料库频次	现通用词
1	马克	march	54	边境,边界
2	摩托	motor	18	发动机

① 陈瑞端《三种汉语书面语体研究》,香港理工大学 2000 年博士学位论文。[Chan Shui Duen, "Linguistic Characterization of Three Written Chinese Registers," Ph.D Dissertation, the Hong Kong Polytechnic University, 2000.]

② 语料库在线网站: www.ncorpus.org。

续表 1

序号	音译词	外文原形词	语料库频次	现通用词
3	喀斯特	caste	17	种姓制度
4	胶姆糖	chewing gum	11	口香糖
5	费厄泼赖	fair play	6	公平比赛
6	安那其(琪)	anarchy	4	无政府主义
7	安琪儿	angel	3	天使
8	白脱(油)	butter	3	奶油,黄油,牛油
9	德律(利)风	telephone	3	电话
10	德谟克拉西	democracy	2	民主
11	回丝	waste	2	废料,废品
12	密斯	miss	2	小姐
13	赛因斯	science	2	科学
14	莱塞	laser	1	激光
15	水门汀	cement	1	水泥

经原文核对,上面 15 个音译词中只有 12 个完全对应于外语原形词在语料库中出现,它们是:“安那其(琪)”、“安琪儿”、“白脱(油)”、“德律(利)风”、“德谟克拉西”、“费厄泼赖”、“回丝”、“胶姆糖”、“莱塞”、“密斯”、“赛因斯”、“水门汀”。其他 3 个词虽然出现,但语料库中的例子显示与《汉语外来词词典》所对应的原形词无关,而是对应于其他外语原形词。如“喀斯特”出现 17 次,全部是地质用语(这个用法已经收入中国社会科学院语言研究所编的《现代汉语词典》),即 Karst 的音译,与《汉语外来词词典》所列的义项“种姓制度”无关;“摩托”一词出现 18 次,是摩托车的意思,与发动机无关;“马克”出现 54 次,44 次作为货币名称出现,10 次作为人名出现,与“边界”和“边境”无关。

为了方便讨论,我们把这 12 个音译词的出现情况在表 2 中列出。

表 2 音译词在语料库中出现的情况

音译词	出现次数	出现时间	语言环境
安那其(琪)	4	1930	散文
安琪儿	3	1982	散文
白脱(油)	3	1989	小说
德律(利)风	2	1987,1996	研究性论文,作为历史词提及
德谟克拉西	2	1954,1980	研究性文章,作为历史词提及
费厄泼赖	7	1976(5),1983(1),1998(1)	6 例与鲁迅作品有关,1 例出现在文学作品中
回丝	2	1985	小说
胶姆糖	11	1991	科普读物
莱塞	1	1987	研究性文章,作为历史词提及
密斯	2	1980	散文
水门汀	1	1937	散文
赛因斯	2	1980,1981	研究性文章,作为历史词提及

从表 2 可以看出,“安那其(琪)”、“水门汀”两词都只出现在 20 世纪 30 年代;“德律(利)风”、“莱塞”、“德漠克拉西”、“赛因斯”等 4 个词作为历史词提及;“费厄泼赖”出现 7 例,6 例实际上也是作为历史词提及,只有 1 例是在实际的语言环境中使用。另外,1980 年以后实际语言环境中使用的外来词是“安琪儿”、“白脱(油)”、“回丝”、“胶姆糖”和“密斯”等 5 个词。从下面表 3 中 5 个词相对应的汉语固有词与音译词出现的频率的对比中可以发现,固有词出现的频率均大大多于音译词。这里需要说明的是,胶姆糖本来是英文 chewing gum 的半音译词,汉语固有词是口香糖,但实际使用中这个词也包括了泡泡糖,所以两者都计算在内。

表 3 音译词与固有词的出现次数比较

音译词	出现次数	固有词	出现次数
安琪儿	3	天使	49
白脱(油)	3	奶油、黄油、牛油	30 (26、0、4)
回丝	2	废品	46
胶姆糖	11	泡泡糖、口香糖	28 (22、6)
密斯	2	小姐	523

在 134 个音译词中,除了上面分析的 15 词外,还有 119 个没有统计到词频,没有统计到词频的词占绝大多数。由此可以证明,《语言接触与语言借用——汉语借词消长研究》一文中所分析的 134 个音译词被汉语固有词代替的结论基本上符合现代汉语的语料实际。

通过利用语料库检验汉语外来词变化的例子可以看到,语料库能够在社会语言学研究起到重要作用,但同时我们也看到了上述语料库所存在的局限。这个局限主要体现在两个方面,一是语料库历时分布的均衡性不足,二是音译词没有得到有效的标注。下文将进行比较详细的讨论。

四、社会语言学研究需要什么样的语料库?

利用语料库和语料库方法进行社会语言学研究需要注意两方面的问题:一是语料库的作用是什么,如何建设一个好的语料库;二是社会语言学研究需要什么样的语料库。

语料库的通常定义是语言研究收集的、用电子形式保存的语言材料,由自然出现的书面语或口语的样本汇集而成,用来代表特定的语言或语言变体。经过科学选材和标注、具有适当规模的语料库能够反映和记录语言的实际使用情况。人们可以通过语料库观察和把握语言事实,分析和研究语言系统的规律。语料库语言学通常被认为不属于语言本身某个侧面的研究,而是一种以现实生活中人们运用语言的实例为基础进行语言研究的方法,以语料库为语言描述的起点,用语料来验证有关语言假说。它为语言研究提供了一种方法论基础,同时又给语言学研究提供了新的思路,所以它是介于理论和方法论之间的一种研究^[14]。语料库语言学是语言研究与相应的计算机技术相伴相生的产物,是语言研究中定量和定性方法相结合的典范。相对传统语言学,语料库语言学不仅研究哪些词语、结构和使用是可能出现的,而且还要统计其出现的概率和方式。

语料库设计和开发最主要的出发点是使其基础上开展的语言调查是合理、可靠的。Kennedy 指出,语料库设计所面临的最基本的问题是语料库所采集的语言数据是否能真正代表某种期望的语言或语体^[15]。McEnery 和 Wilson 认为,语料库并不是语篇的简单堆砌,它应具有样本广泛、规模有限、机器可读等基本特征^[4]。

语料库的代表性和平衡性是语料库建设的基本要求。作为对语言的抽样,语料的代表性是语料

库数据合理性和可靠性的保证。Leech认为,一个语料库具有代表性是指从该语料库获得的分析结果可以概括成为这种语言整体或其指定部分的特性^[16]。所有语料库建设者都力图使其创建的语料库足以代表或反映其所要研究的目标语域或整个语言的语言事实,因此在创建初期都会对语料库的设计、取样进行科学的分析^[17]。但有一点我们必须认识到,无论如何也无法穷尽“某种语言的全体使用者说出来(或写下来)的和尚未说出来(或写下来)的所有话语”,因为它是一个开放集,理论上无法真正捕获到这些语言全貌的。因而为了尽可能地(至少在统计上)反映语言的实际情况,对语言进行抽样的方法在一定程度上可以满足研究的需要,与“抽样”紧密关联的一个重要概念就是“代表性”问题,也即所收集的语料是否可以在统计上代表各种类型的真实话语。语料平衡性不能简单理解为文本的不同来源。实际上,不同来源语料的使用、发行和读者数都不一样。Sinclair提出,至少应该区分小说/非小说、图书/期刊/报纸、正式/非正式出版物、作者属性等^[18]。然而,由于语言自身的高度复杂性,关于语料的代表性和平衡性还没有找到能为大家普遍接受的答案。

语料库规模也是建设语料库需要考虑的一个重要问题。多大的语料库才够用是伴随语料库作为一种研究方法产生的,也是最难回答的问题。一般就语言研究而言,例如研究词汇和词典编撰,或者比较各种文本的语言结构,语料库越大越好。但鉴于语句的无限性及语言的不断变化,语料是不能穷尽的,代表性也是相对而言的。因此,从量化研究的需要出发,语料的规模应是有限的。

语料库的标注加工是语料库能否发挥作用最重要的一环。一个语料库的功能主要与三个因素有关:一是语料库的规模,二是语料的分布,三是语料的加工程度。规模的大小关系到统计数据是否可靠,语料的分布涉及统计结果的适用范围,语料加工的深度则决定这个语料库能为使用者提供什么样的语言学信息。为了从语料库中抽取信息,必须首先从多个层面对语料库进行分析,并且将分析结果标注到语料库上去,这个过程就是语料库标注。语料库标注是指给口语或书面语语料库增添解释的(interpretative)和语言的(linguistic)信息的实践^[16]。“标注”也可以视为这个过程的最终产品,即附加、关联或分散在语料中的语言标记。语料库标注给语料库带来巨大的附加价值,是语料库建设中最为关键的一环。Leech提出了英语语料库已经达到或可以达到的七个语言学标注层级,分别为拼写(orthographic)、语音(phonetic)、韵律(prosodic)、语法(grammatical)、句法(syntactic)、语义(semantic)和语用(pragmatics/discourse)^[16]。

大型语料库的建立为各类语言研究提供了大量的真实语料,也为各种语言成分的定量分析提供了条件。但由于语料库的规模、分布和标注等方面存在局限,数据稀疏的问题还普遍存在。比如上文我们在讨论音译词时,提到134个音译词中有119个词查不到频次,即查询结果为0。这很可能主要是语料时间分布上不平衡出现的问题。

社会语言学研究对语料的代表性和平衡性有严格的要求,具有良好代表性、类别多样、时间分布平衡的语料库对于社会语言学的语言变异的研究意义更大。目前语料库的标注还远远不能适应社会语言学研究的要求,上面所说的七个语言学标注层级中,语用标注对社会语言学研究的帮助最大,但已标注语用信息的语料库一般都是用于专题研究性质的小型语料库,目前国内外的大型语料库大都缺少语用方面的标注。社会语言研究需要的语料库大致可从以下几个方面考虑:

(1) 语料样本具有广泛性、代表性和平衡性;样本分布既有共时意义也有历时意义。所谓广泛性,指语料库收集的语料应该涉及各种文体,涵盖不同行业 and 不同地区的语言使用者。代表性、平衡性则是一般语料库共有的要求。样本分布具有共时意义可以用于共时的语言变异研究,具有历时意义则可以用于历时的语言变化研究。

(2) 语料样本信息丰富完整,包含作者的民族、性别、年龄、出生地、母语等更多背景信息,其中部分信息需要通过标注获得。这些信息是社会语言学研究中的重要社会因素,是研究语言与社会共变的重要资料。

(3) 语料库具有多层次、多方位的标注。例如,在词语标注方面,不仅有词性标注,有些具有地域特点的词语还应该标注地区来源的标注。汉语里至少有四类词可以标注来源:第一类是方言词,可以标注出来源自何地何种方言;第二类是社区词,可以标注出来源自大陆、台湾、香港、澳门或其他华语社区;第三类是少数民族语言的借词,可以标注出来源自何地何种少数民族语言;第四类是外来词,可以标注出来源自哪个国家哪种语言,最好还能够标注出外来词的次类,如原形词、字母词、音译词、半音译半意译词、意译词等等,外来词如果有这样的标注,上面提到的没有频次的 119 个音译词的数量一定会大大减少。

五、期待与希望

近年来,语料库研究不断取得新的进展,相信在计算语言学界和社会语言学界的共同努力下,语料库的建设会进一步适应社会语言学的需求,推动社会语言学向定性研究和定量研究完美结合的方向发展。

社会语言学研究方法不仅是描写性的,更重要的是解释性的,解释语言在社会中发生的变化。语料库方法能够为语言变化的研究提供现代化的手段,可使这一领域的研究发生质的变化。我们期待计算语言学家更多关注语言变化研究,关注社会语言学研究,为社会语言学研究提供更多的理论方法和技术支持。从汉语语料库的建设和中国社会语言学研究的角度出发,我们认为以下三个问题值得重视:

一是现有的古代汉语语料库和现代汉语语料库应该进一步整合,建设成为可以进行连续检索、分析汉语历时变化的重要语料库,使语料库资源得到更便捷、更充分的利用。

二是 LIVAC 语料库在跟踪研究现代汉语不同社区词的共时变化上已经发挥了重要作用,但应该进一步丰富检索功能,让更多的研究人员利用这一平台。

三是语料库建设时应该进一步研究如何满足社会语言学的需求,进行社会语言学研究所需信息的标注,以便深入开发利用。

[参 考 文 献]

- [1] 桂诗春、宁春岩:《语言学方法论》,北京:外语教学与研究出版社,1997年。[Gui Shichun & Ning Chunyan, *Linguistic Methodology*, Beijing: Foreign Language Teaching and Research Press, 1997.]
- [2] 张廷国、郝树壮:《社会语言学研究方法的理论与实践》,北京:北京大学出版社,2008年。[Zhang Tingguo & Hao Shuzhuang, *Sociolinguistic Research Methodology: Theory and Practice*, Beijing: Peking University Press, 2008.]
- [3] 游汝杰、邹嘉彦:《社会语言学教程》,上海:复旦大学出版社,2004年。[You Rujie & B.K.T'sou, *A Course in Sociolinguistics*, Shanghai: Fudan University Press, 2004.]
- [4] T. McEnery & A. Wilson, *Corpus Linguistics*, Edinburgh: Edinburgh University Press, 1996.
- [5] G. Kjellmer, "The Lesser Man: Observations on the Role of Women in Modern English Writings," in J. Aarts & W. Meijs (eds.), *Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora*, Amsterdam: Rodopi, 1986, pp. 163-76.
- [6] J. Holmes, "Inferring Language Change from Computer Corpora: Some Methodological Problem," *ICAME Journal*, No. 18 (1994), pp. 27-40.
- [7] 黄昌宁、李涓子:《语料库语言学》,北京:商务印书馆,2007年。[Huang Changning & Li Juanzi, *Corpus Linguistics*, Beijing: The Commercial Press, 2007.]

- [8] J. Hope, *The Authorship of Shakespeare's Plays: A Socio-linguistics Study*, Cambridge: Cambridge University Press, 1994.
- [9] S.M. Sotillo & J. Wang-Gempp, "Using Corpus Linguistics to Investigate Class, Ideology, and Discursive Practices in Online Political Discussions," in U. Connor & T.A. Upton (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam: Rodopi, 2004, pp. 91-122.
- [10] 马博森:《当代中国话语语料库的建构问题》,见施旭主编:《当代中国话语研究》第1卷第1辑,杭州:浙江大学出版社,2008年,第44-53页。[Ma Bosen, "Issues in the Construction of Corpus of Contemporary Chinese Discourse," in Shi Xu (ed.), *Journal of Contemporary Chinese Discourse Studies: Vol. 1*, Hangzhou: Zhejiang University Press, 2008, pp. 44-53.]
- [11] 刘正琰等编:《汉语外来词词典》,上海:上海辞书出版社,1984年。[Liu Zhengyan et al. (ed.), *A Dictionary of Loan Words and Hybrid Words in Chinese*, Shanghai: Shanghai Lexicographical Publishing House, 1984.]
- [12] 苏金智:《语言接触与语言借用——汉语借词消长研究》,《中国语言学报》2010年第14期,第71-83页。[Su Jinzhi, "Language Contact and Linguistic Borrowing: A Case Study on the Growth and Decline of Loanwords in Chinese," *Journal of Chinese Linguistics*, No. 14 (2010), pp. 71-83.]
- [13] 肖航:《现代汉语通用平衡语料库建设与应用》,《华文世界》2010年第106期,第24-29页。[Xiao Hang, "On the Construction and Application of Contemporary Chinese Corpus," *Journal of Chinese World*, No. 106 (2010), pp. 24-29.]
- [14] T. Bonelli, *Corpus Linguistics at Work*, Amsterdam: John Benjamins Publishing Company, 2001.
- [15] G. Kennedy, *An Introduction to Corpus Linguistics*, London: Longman, 1998.
- [16] G. Leech, "Corpus Annotation Schemes," *Literary and Linguistic Computing*, Vol. 8, No. 4 (1993), pp. 275-281.
- [17] D. Biber, "Representativeness in Corpus Design," *Literary and Linguistics Computing*, Vol. 8, No. 4 (1993), pp. 243-257.
- [18] J. Sinclair, *Corpus, Concordance, Collocation*, Oxford: Oxford University Press, 1991.

书 名: 走出书斋的史学

作 者: 李埏、李伯重、李伯杰

定 价: 48.00 元

· 新 书 架 ·

浙江大学出版社出版



内容简介: 让史学走出史学家的书斋,把正确的历史知识介绍给社会大众,是当今的一项重要工作,也是本书的主要目的。本书把著名历史学者李埏、李伯重、李伯杰教授的史学随笔和读书札记进行了整理,分为政治史、经济史、社会史、文化史、世界史几个栏目。已故的李埏教授写过不少关于中国历史的短文,发表在不同的报纸杂志上。这些文章文笔优美,深受社会各界的好评。李伯重和李伯杰教授分别是中国经济史和德国文化史方面的专家,他们在过去的十多年中也写了一些中国历史和德国历史方面的随笔和读书札记。这些文章是面向社会大众而非专家,因此所谈内容也不像专业学术论文那样学科的界限泾渭分明。

发行地址: 杭州市天目山路148号浙江大学出版社

邮 编: 310028

电子邮箱: faxingbu2004@zju.edu.cn

电 话: 0571-88925591、88273163