

语料库与社会语言学研究

苏金智*

教育部语言文字应用研究所, 北京

提要

社会语言学的研究方法不同于传统的结构语言学和乔姆斯基的转换生成语法, 其重要特点是把语言放在言语社区中进行定量与定性的分析。早期的社会语言学更多注重口语的研究, 把口语的研究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; 而书面语的研究一般采用文献调查的方法, 缺乏先进的手段和方法, 在语言变异研究方面很难取得重大的成果, 因此没有得到足够的重视。语料库和语料库语言学出现后, 社会语言学利用语料库研究书面语取得了一系列成果。这些成果主要有语言与性别的研究、利用 LIVAC 语料库进行两岸四地汉语社区词的研究、汉语语体 (register) 研究和话语分析方面的研究。这些研究是开拓性的研究, 积累了很好的经验, 但还有许多可以改进之处。苏金智 (2010) 的一项有关汉语外来词变化的研究结果显示, 《汉语外来词词典》(刘正琰等 1984) 中所收的外来词约占词典 1% 的 134 个音译词已经基本上被汉语固有词语所取代, 遗憾的是这个结论没有用语料库的大量语料进行检验。利用国家语委现代汉语语料库检验 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为汉语固有词代替的结论基本上符合现代汉语的语料实际, 这进一步说明了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。文章[†]最后对如何利用语料库进行社会语言学研究提出了建议。

* Corresponding author: sjzhi@sina.com.cn

[†] This current paper is a revised paper from the Roundtable Conference on Linguistic Corpus and Corpus Linguistics in the Chinese Context, held in May 2011 at the Hong Kong Institute of Education. A major part of this article was published after the conference as “Yuliaoku yu shehui yuyanxue yanjiu fangfa 语料库与社会语言学研究方法 (Corpus and sociolinguistic methodology)” by Su Jinzhi 苏金智 and Xiao Hang 肖航 in *Zhejiang daxue xuebao* (renwen shehui kexue ban) 浙江大学学报(人文社会科学版) 42(4):87-95.

关键词

语料库 社会语言学 音译词 语言变化 建议

LINGUISTIC CORPUS AND SOCIOLINGUISTIC STUDY

Jinzhi Su

Institute of Applied Linguistics, Ministry of Education of the PRC, Beijing

ABSTRACT

Methodology of sociolinguistics is different from traditional structural linguistics, in particular, from the Chomskyan transformational generative grammar. Sociolinguistic distinctive methodological feature is to analyze language elements in the speech communities in qualitative and quantitative approaches. In the early studies, sociolinguists paid much more attention on spoken forms because these forms could be analyzed in the speech communities and be explained from the relationship between language variables and sociolinguistic variables such as ethnic group, age, social class and gender easily. On the contrary, the study of written forms got inattention because they could not be done as well as the spoken forms due to the use of traditional method to collect data from the literatures. Corpus and corpus approach has had a great help to the study of written language since it emerged. Sociolinguistic studies reap the benefits from corpus approach such as in the study of gender language, the study of words from Chinese speech communities using the LIVAC, the study of Chinese register, and discourse analysis. These studies are

pilot ones and this line of study has accumulated experiences for the future study but is still limited and there is clearly need for further investigation. The result of loanwords studies by Su (2010) shows that 134 Chinese phonetic matching loanwords, one percent among the total loanwords in *The Dictionary of Chinese Loanwords* (1984), have been replaced by the 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 for analyzing the trends of language change. Finally some suggestions are made for the improvement of sociolinguistic study by using Chinese corpus.

KEYWORDS

Corpus Sociolinguistics Phonetic matching loanwords
Language change Suggestions