

汉语网络媒体监测语料库及博客语言应用调查

何婷婷* 涂新辉

华中师范大学[†], 武汉

提要

汉语网络媒体监测语料库由国家语言资源监测与研究中心网络媒体语言分中心¹2005年开始建设,包含网络新闻、博客、论坛的语料。基于该语料库2009年、2010年的博客语料的统计数据,对比分析了一般博客用户和名博在博客发帖量、用字用语上的特点;基于不同性别的作者的博客文本,对比分析了男、女性作者在用字用词上的特点。

关键词

语料库 博客 语言调查 性别语言 频率比值

CHINESE NETWORK MEDIA MONITORING CORPUS AND
EMPIRICAL INVESTIGATION OF LANGUAGE USE ON BLOG

Tingting He Xinhui Tu

Central China Normal University, Wuhan

ABSTRACT

Chinese network media monitoring corpus, which includes network news, blog and forum texts, has been maintained by network media branch of national language resources monitoring and research center since 2005. Based on the corpus's blog texts in 2009 and 2010, we compare some characteristics of language use for famous blog users and general blog users. And we also conduct a comparative analysis of characteristics of language use on female users and male users.

KEYWORDS

Corpus Blog Language use survey Gender linguistics
Frequency ratio