

运用现代科学技术研究中国传统文化

——国家社科基金项目“汉语俗语语料的计算机处理及相关语言学问题”结项

发布日期：2006-10-03 作者：侯晓斌

【打印文章】

本报讯（通讯员侯晓斌）由山西省社科院研究员温端政主持的国家社科基金项目“汉语俗语语料的计算机处理及相关语言学问题”日前结项。专家指出，把现代化的研究手段运用于传统文化资源研究，是汉语俗语研究方法的新突破，将推动汉语俗语研究的繁荣和发展。

该成果指出，俗语可以根据叙述的内容和方式分为以下三种类型：一是表述语，即谚语，属于表述性俗语，表达某种推理和判断，传授某种知识（含经验）。如“远水不救近火”、“留得青山在，不怕没柴烧”等。二是描述语，即惯用语，属于描述性俗语，描述事物的形象、性质或状态。如“走后门”、“喝西北风”、“挂羊头，卖狗肉”等。三是引述语，即歇后语，属于引述性俗语，由“引子”引出“注释”，表示某种形象、性质或状态。如“木匠的斧子——一面砍”、“泥菩萨过河——自身难保”等。

专家指出，这种分类标准与语类定性，在编俗语辞书时具有可操作性，对于创建俗语语料数据库，具有学术上的指导意义。该成果与同类成果相比，不仅规模大，而且体例、内容和方法上皆有所创新。在体例上，每种大全都分上下编，上编收通行于口语的，下编辑录语料；上编，对比较难以理解的作提示性注释；下编，用“语出”表示语目的出处，例证见于古代（清代以前，含清代），尽量收录；现当代重点收录名家名作里的例证；在该课题语言学问题的研究成果《俗语研究与探索》（论文集）部分，着重探讨了俗语及其属类谚语、歇后语、惯用语的性质、范围和分类，为大全的编写和语料数据库的创建提供学术支撑；建立了汉语俗语语料数据库（光盘），该数据库由古代经史子集中的俗语，近代戏剧、小说中的俗语，现代名作中的俗语，口语中的谚语，口语中的歇后语和口语中的惯用语6个部分组成。并设立了包括专书检索、语目音序检索、语目分类检索、语目任意字，以及语境任意字等多种检索手段。其中语目任意字和语境任意字检索尤为方便，在课题计划中原来没有这部分内容，是在课题研究过程中开发并增加的部分。

汉语俗语是历代群众创造并在口语中广泛流传的语言单位。它题材广泛，内容丰富，几乎涉及社会生活的各个方面。许多俗语富有哲理性，涉及如何处理人与大自然的关系、人与人的关系以及个人的修养等方面的问题，是亿万人民世代集体经验和智慧的结晶，是中华民族先进文化的组成部分。因此，汉语俗语语料数据库既是集汉语古今俗语之大成的语言宝库，也是中华民族文化的宝库。

该课题最终成果的应用转化，将是创建我国汉语俗语语料数据库，该库具有语料翔实丰富、收条规范、分类合理、检索方便等特点，具有多方面的实用价值：一是为汉语俗语的专题研究和俗语类辞书的编纂提供资料。二是历代俗语为研究方言、历史语言、民俗文化者提供了丰富的素材，为汉语特点的研究提供参考资料；其他文史方面的学者，也可以从中采用所需语料。三是可以用来进行语频统计。如对《左传》、《国语》、《史记》、《汉书》及《金瓶梅》、《红楼梦》、《儿女英雄传》等专书使用俗语频率的统计，可为汉语史研究提供资料；对现代汉语里通用俗语使用频率的统计，可为编写常用俗语词典提供依据。四是可以发现俗语在文本中形式上的标志，为计算机的自动识别提供多种方法，如标注识别法（俗语前面多有“语曰”、“谚曰”、“鄙语曰”、“俗话说”、“常言道”等标注）；标点符号识别法（如歇后语前后两部分之间多有破折号）；谐音识别法（如运用谐音的歇后语或惯用语，本字之后多注明谐音字）；结构识别法（如由两个部分组成的谚语和惯用语，前后多对称）等。这些方法，不仅有助于俗语语料数据库的长远建设，而且从一个侧面提高了中文信息处理的技术水平。

该课题阶段性成果“汉语俗语语料汇编”以《中国谚语大全》（上下）、《中国惯用语大全》和《中国歇后语大全》为书名，于2004年5月由上海辞书出版社出版，计700余万字。最终成果之一《汉语俗语语料数据库》（光盘），于2005年7月完成，1465万字，其中数据库字数762万，附录三本“大全”字数703万。最终成果之二《汉语俗语研究》（论文集），于2005年5月由上海辞书出版社出版。

专题[口头传统](#)的相关文章

- 《人类口头和非物质遗产代表作条例》
- 家乡民俗学
- 台静农：歌谣乡土研究的遗产
- 民俗学家乡研究的理论反思
- “ 非遗 ” 有望纳入国民教育体系 弘扬传统文

中国民族文学网



ᠴᠤᠷᠭ᠎ᠠ ᠮᠢᠨᠵᠢᠴᠤᠷ ᠪᠠᠭᠠᠨ ᠮᠤᠭᠦᠬᠦ

جوڭگو مىنلله نلەر ئەدەبىياتى تورى

Curggoz Minzcuz Vwnzyoz Muengx