

# 现代汉语多级标注语料库的建设及其应用

段慧明 朱学锋 俞士汶\* 吴云芳

北京大学<sup>†</sup>

## 提要

本文作者与北京大学计算语言学研究所 (ICL/PKU) 的同仁一道, 自 1986 年起积 25 年之努力建成“综合型语言知识库” (简称 CLKB)。现代汉语多级标注语料库是 CLKB 中的一项重要的语言知识库。本文在介绍 CLKB 的概要之后,<sup>1</sup> 论述 ICL/PKU 研制多级标注语料库的理念、已经取得的成果及其应用情况。

## 关键词

计算语言学 中文信息处理 综合型语言知识库 现代汉语多级标注语料库 现代汉语语法信息词典

THE CONSTRUCTION AND APPLICATION OF MANDARIN  
CHINESE MULTI-LEVEL ANNOTATED CORPUS

**Huiming Duan Xuefeng Zhu Shiwen Yu Yunfang Wu**

*Peking University*

ABSTRACT

The Comprehensive Language Knowledge Base (CLKB) which has been under construction by the authors and the colleagues of Institute of Computational Linguistics at Peking University since 1986. Mandarin Chinese multi-level annotated corpus is one of the important language knowledge bases of CLKB. After a brief introduction of CLKB, this paper describes the leading ideas, the achievement and application of our multi-level annotated corpus.

KEYWORDS

Computational linguistics Chinese information processing

Comprehensive Language Knowledge Base

Mandarin Chinese Multi-level Annotated Corpus

Grammatical Knowledge-base of Contemporary Chinese