

东北大学



语言信息技术的研究

姚天顺

东北大学计算机科学研究所

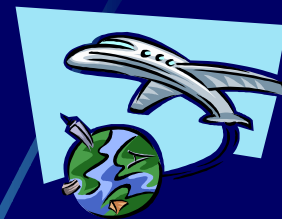
简单的历史

- 1980年，研究所的领导人作为访问学者在香港中文大学《孔安道机器翻译研究中心》工作半年。
- 1984年，成立NLP研究组，从事机器翻译研究
- 1987年，参加日本CICC（7年）

CICC(国际情报化协力组织, Center of the International Cooperation for Computerization)
负责汉语分析组组长工作。

开发了自己的机器翻译系统—CETRAN，尤其是
汉语分析器





● 先后得到的国家重大项目：

《国家重点基础理论规划计划（973）》	1 项
《国家高技术研究发展计划（863）》	3 项
《国家自然科学基金项目》	9 项
《国家博士点基金资助项目》	6 项
《国家科技进步项目》	8 项

27 项



先后与国际友好单位合作项目

- 日本CICC《亚洲多国语机器翻译系统项目》
- 香港城市大学《计算机自动摘要系统》合作
- 美国Berliz公司汽车手册的机译人助的翻译工作
- 韩国浦项科技大学（POSTECH）
《多国语机器翻译系统》
- 美国Motorola中国研究中心关于汉语分析器
（语音合成前理）合作研究
- 美国普林斯顿松下信息与网络技术研究所
（Panasonic Infor. & Networking Tech. Lab.），
网络复印机OCR语言学后处理

组织与 人员情况

- 姚天顺
所长
- 朱靖波
副所长
- 许桂清
副所长
- 刘丽晨
秘书

博士生导师
国际中文电脑学会理事
博士， 副教授

副教授



● 中文信息处理研究组

朱靖波

博士，副教授

陈文亮

硕士

吴学岩

硕士

顾 莉

硕士

● 语言工程研究组

姚天顺

教授

杨 莹

博士生，副教授

张 俐

博士生，副教授

刘明强

硕士

张 乐

硕士

王大禹

硕士

● 信息检索研究组

王家华

教授

战学刚

博士，副教授

童瑜锟

硕士

沈嫣娜

硕士

● 主要从事称为：

人类语言技术 (Human Language Technology)

**Ronald A. Cole: Survey of the State of the Art in Human
Language Technology)**



1. 语言资源



- 1) 歧义切分词段, 47,000 个
- 2) 可以作为姓的 300 字库
- 3) 可以作为名的 1000 字库
- 4) 基本词库 300,000 词
- 5) 25类专业词库 1,000,000 词
- 6) 汉语语法树库, 10,000 棵
- 7) 动态特征库, 240,000 词
- 8) 双语语料库, 50,000 字
- 9) 汉语带标语料库, 1,000,000字
- 10) 半结构汉语语料库, 1,000,000 词
- 11) 中文WordNet (动词部分)

中文WordNet（动词部分）

演示

C:\¥WN16¥BIN¥WISHWN.EXE



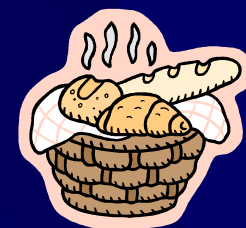
2. 语言分析与生成

1. 1987年，作为**CICC (Center of the International Cooperation for Computerization)** 的汉语分析的负责单位，主要从事汉语分析器的研究。
2. 与美国的**Berlitz**公司，利用我们的分析器，进行机译人助，翻译克莱斯勒用户手册。
3. 与韩国浦项科技大学合作，利用我们的汉语分析器，进行多国语机器翻译系统。
4. 与美国**Motorola**中国研究中心合作：汉语分析器（用于语音合成前处理）



汉语分析器演示

..¥..¥..¥korean¥AnaChi.exe



3. 文本信息处理

- 交互式信息检索系统 (**Interactive Track**)
- 问-答系统 (**Question-Answering Track**)
- 交叉-语言检索系统 (**CE Cross-Language Track**)
- 语音文档检索系统 (**Spoken Document Retrieval**)
- 文本过滤系统 (**Filtering Track**)
- **Web**检索系统 (**Web Track**)
- 图像检索系统 (**Image Retrieve**)



文本过滤演示

..¥..¥..¥Filter¥Filtering.exe

