

# 翻译研究技术： 应用于翻译研究的 电子工具

---

萧世昌

2022年6月11日

# 背景



人工智能发展



广泛应用领域



翻译研究

# 大纲



人工智能和翻译



「翻译研究技术」框架



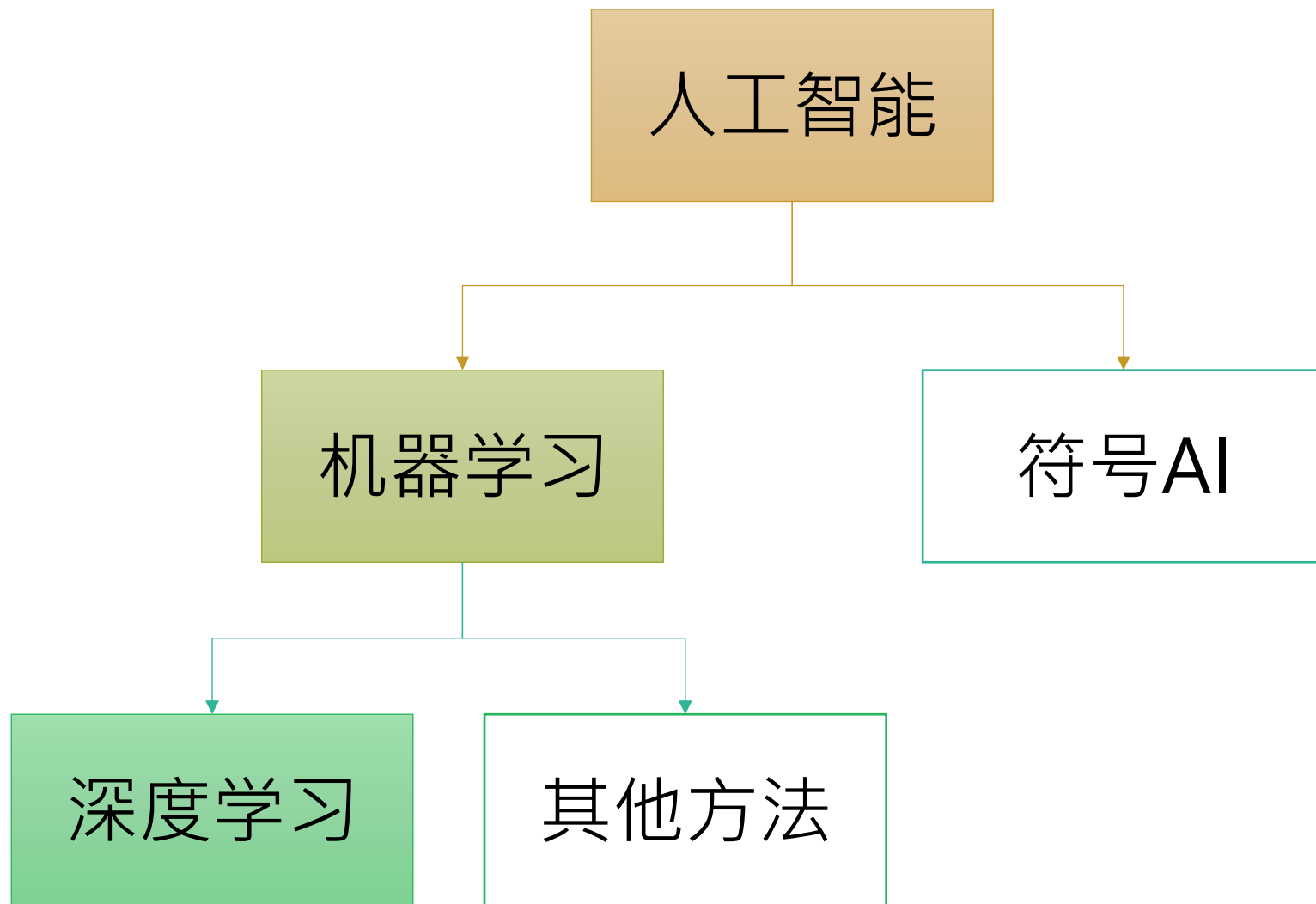
具体方向



结语

# 人工智能和翻译







# 翻译技术

- 神经机器翻译
- 语音翻译

# 机器翻译



基于规则的机  
器翻译



基于实例的机  
器翻译

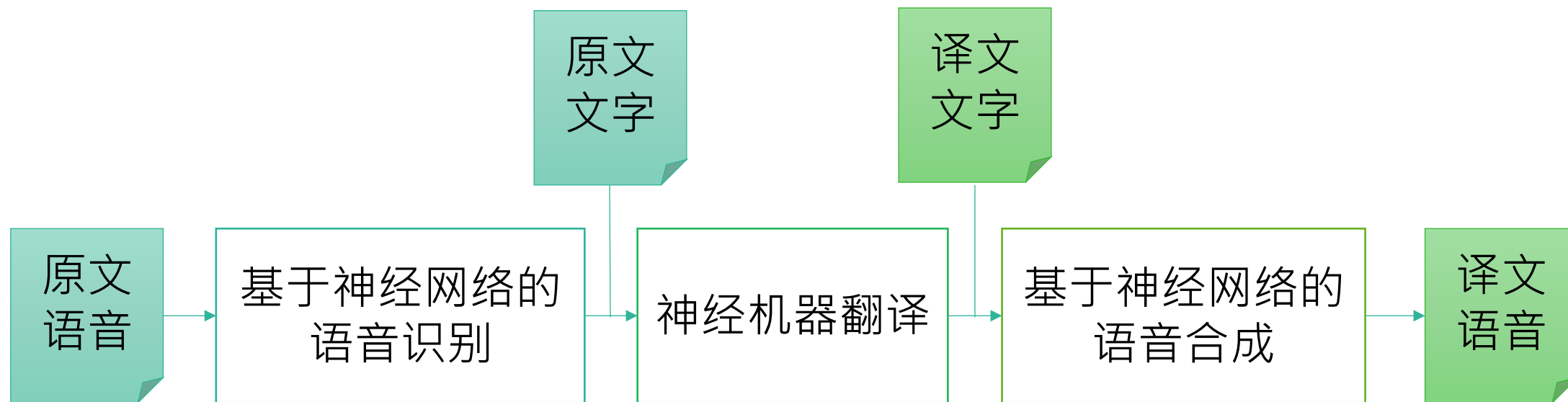


统计机器翻译



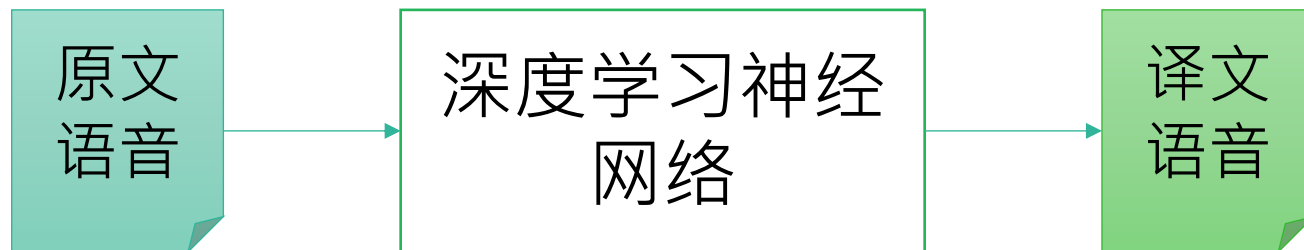
神经机器翻译

# 语音翻译

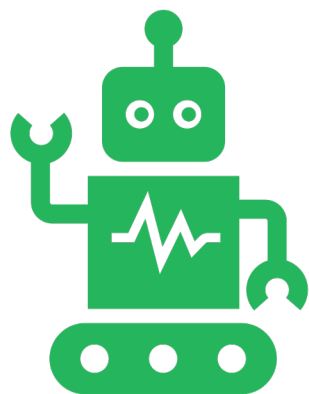




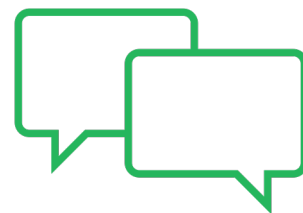
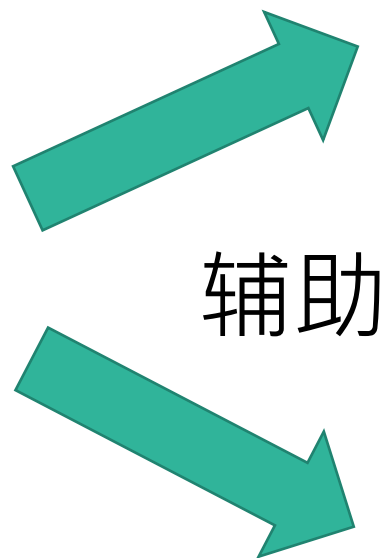
# 端到端语音翻译



# 目前状况



技术



专业译者

机器翻译、翻译记忆库、  
术语库、项目管理等等



翻译研究



# 重点



提出框架



探索工具

# 「翻译研究技术」框架

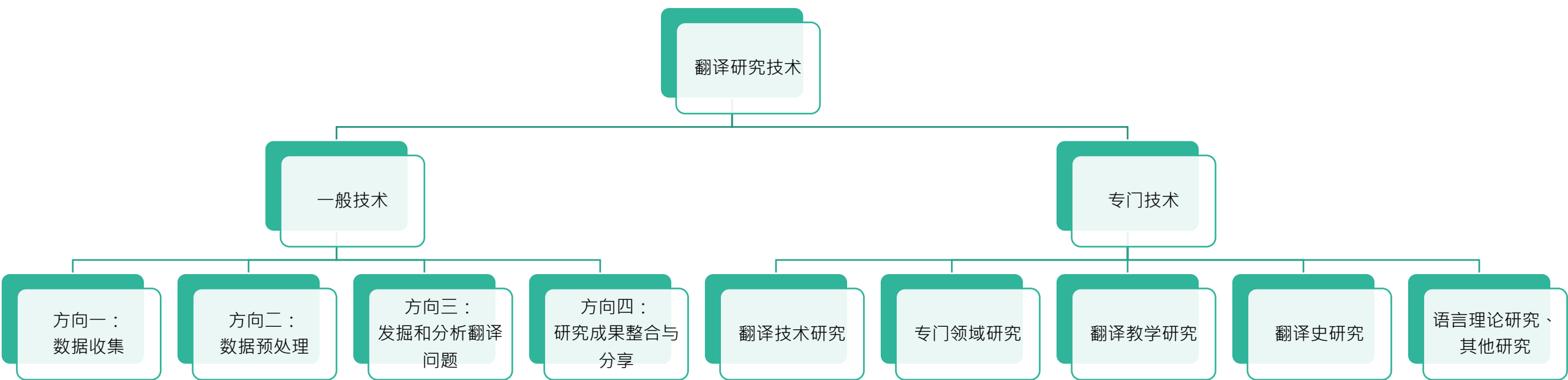


# 电子资源的三大类别

- 单机系统
- 在线系统
- 代码库



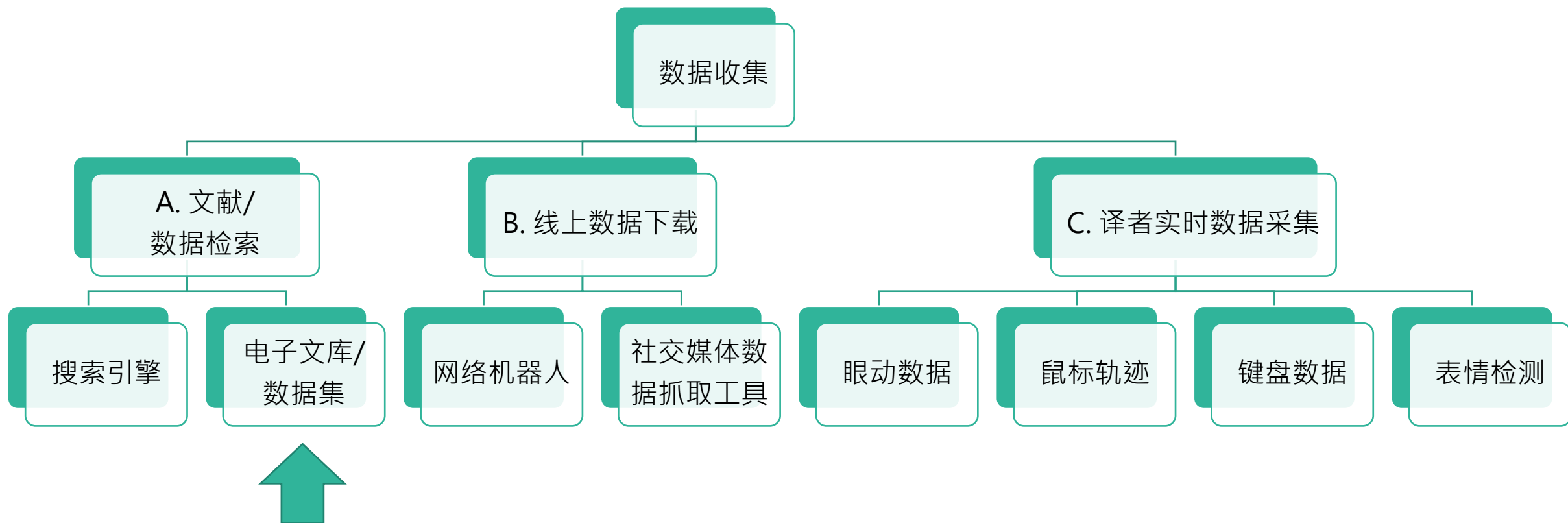
# 关键方向



# 具体方向



# 方向一：数据收集

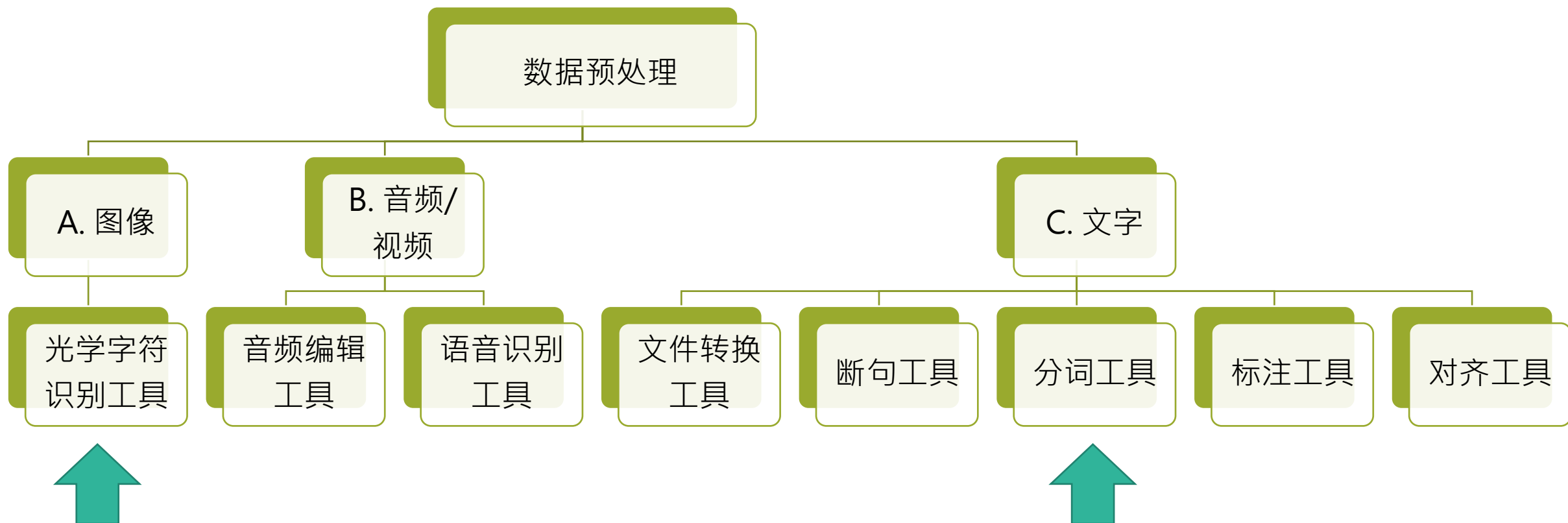




# 双语数据：以香港翻译为例

- 电子版香港法例 ( <https://www.elegislation.gov.hk/> )
- 香港特区政府新闻公报  
( [https://www.info.gov.hk/gia/ISD\\_public\\_Calendar\\_tc.html](https://www.info.gov.hk/gia/ISD_public_Calendar_tc.html) )
- 香港立法会议事录资料库  
( <https://app.legco.gov.hk/hansarddb/sc/search.aspx> )

## 方向二：数据预处理



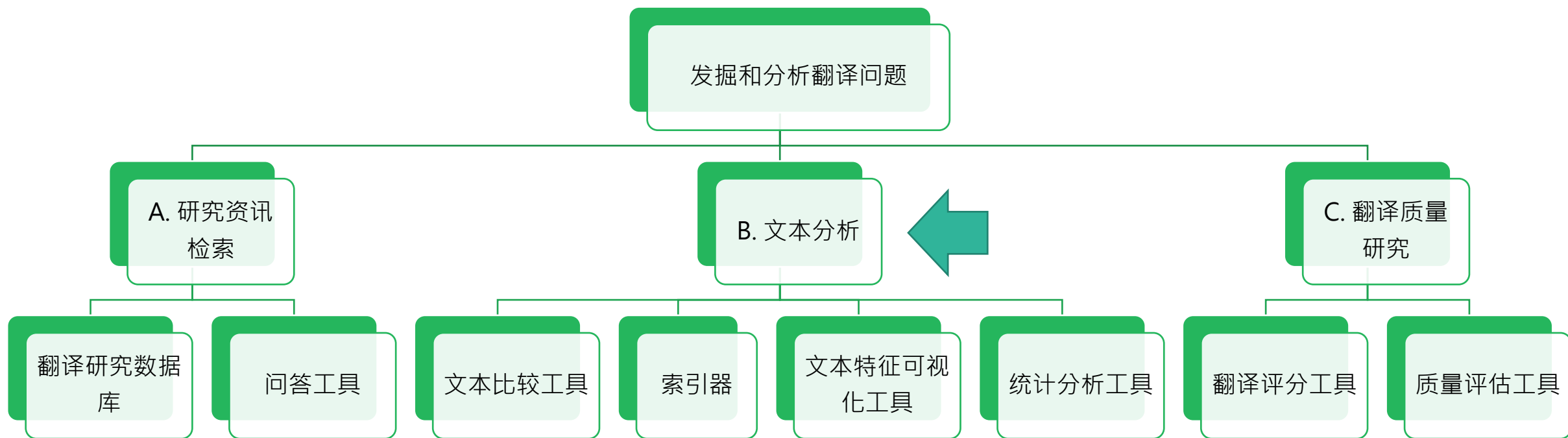
# 光学字符识别工具

- Youdao OCR (<https://ai.youdao.com/product-ocr-print.s>)
- Transkribus (<https://readcoop.eu/transkribus>)
- 中文古籍 OCR (<https://ocr.gj.cool/>)

# HanLP.com

- 主页 ( <https://www.hanlp.com/index.html> )
- 多语言分词 ( <https://www.hanlp.com/product-token.html> )
- 词性标注 ( <https://www.hanlp.com/product-pos.html> )
- 命名实体识别 ( <https://www.hanlp.com/product-ChineseName.html> )

# 方向三：发掘和分析翻译问题



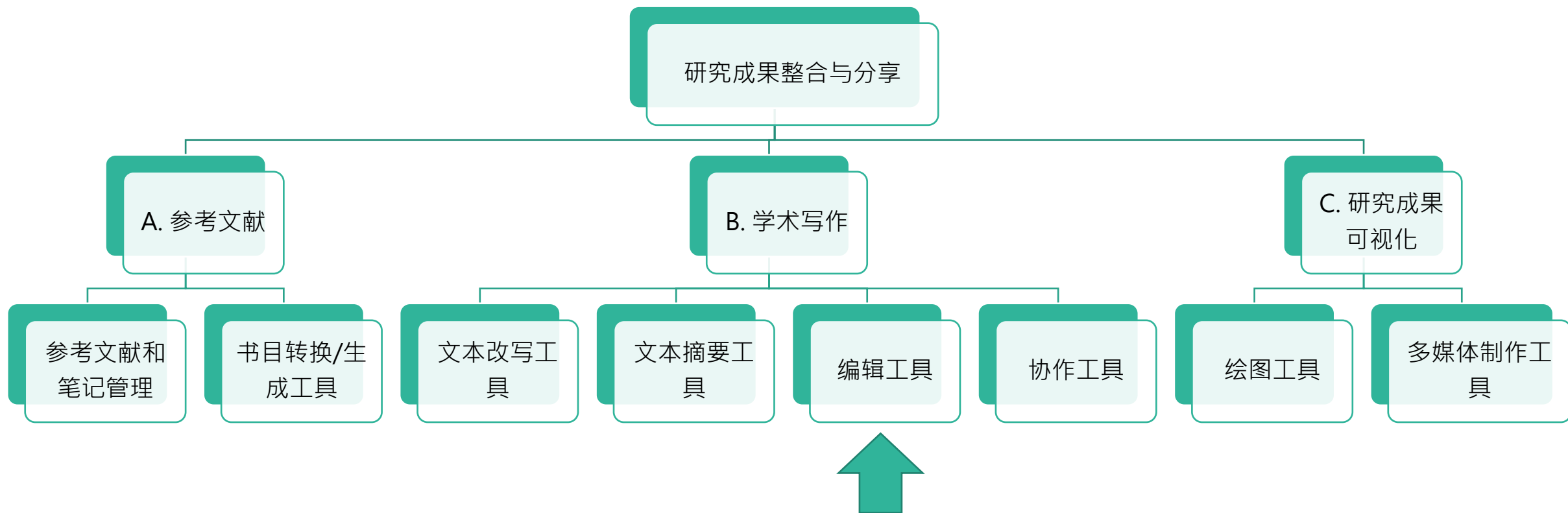
# 文本特征可视化和统计分析

- Voyant Tools ( <https://voyant-tools.org/> )
- Rstudio ( <https://www.rstudio.com/> )

# Memsource

- 主页 ( <https://www.memsource.com/> )
- 自动对齐
- 预料检索
- 文本比较
- 质量评估工具

# 方向四：研究成果整合与分享





# 编辑工具

- Wordtune ( <https://www.wordtune.com/> )
- Wordvice AI ( <https://wordvice.ai/> )

# 讯飞开放平台

- 主页 ( <https://www.xfyun.cn/> )
- 例子 : [https://www.xfyun.cn/services/text\\_rewrite](https://www.xfyun.cn/services/text_rewrite)

# 结语



工具目录

实例分析

自然语言处理、  
人工智能、  
编程教学

# 谢谢！



[scsiu@hsu.edu.hk](mailto:scsiu@hsu.edu.hk)