



# 计算语言学概论

Computational Linguistics: An Introduction

---

詹卫东

北京大学中文系

<http://ccl.pku.edu.cn/doubtfire/>

zwd@pku.edu.cn



# 课程安排

---

- |   |            |        |
|---|------------|--------|
| 1 | 计算语言学概述    | (2 学时) |
| 2 | 语言知识的形式化表示 | (4 学时) |
| 3 | 汉语的形式句法体系  | (4 学时) |
| 4 | 语义知识的形式表示  | (4 学时) |
| 5 | 语篇知识的形式表示  | (2 学时) |
| 6 | 语料库与语言知识工程 | (4 学时) |
| 7 | 词法分析       | (4 学时) |
| 8 | 句法分析       | (4 学时) |
| 9 | 机器翻译       | (2 学时) |
| a | 课堂讨论交流     | (2 学时) |



# 授课对象与教学目标

---

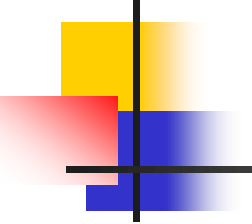
- 授课对象：

中文系语言学专业研究生和高年级本科生。

- 教学目标：

具备以计算机为背景思考语言学问题，对语言知识进行形式化描述的能力；

了解计算语言学的基本理论和自然语言处理中的常用模型及算法，为将来从事以应用为导向的语言研究打好基础。



# 参考资料：按时间顺序分组

---

1. 语言和计算机编辑组编，1985-1986 《语言和计算机》（3册） 中国社会科学出版社。
2. 钱锋（1990）《计算语言学引论》，学林出版社1990年版。
3. 陆致极（1990）《计算语言学导论》，上海教育出版社 1990年版。
4. 刘开瑛、郭炳炎（1991）《自然语言处理》，科学出版社 1991年版。
5. 冯志伟（1991）《数学与语言》，湖南教育出版社1991年版。
  
6. 白硕（1995）《语言学知识的计算机辅助发现》，科学出版社1995年版。
7. 冯志伟（1995）《自然语言机器翻译新论》，语文出版社1995年版。
8. 姚天顺等（1995）《自然语言理解 —— 一种让机器懂得人类语言的研究》，清华大学出版社、广西科学技术出版社1995年版。
9. 冯志伟（1997）《自然语言的计算机处理》，上海外语教育出版社1997年版。
10. 吴立德 等（1997）《大规模中文文本处理》，复旦大学出版社1997年版。
  
11. 翁富良、王野翊（1998）《计算语言学导论》，中国社会科学出版社1998年版。
12. 俞士汶 等（1998/2003）《现代汉语语法信息词典详解》（第二版），清华大学出版社、广西科学技术出版社1998年版。
13. 陈小荷（2000）《现代汉语自动分析》，北京语言文化大学出版社2000年版。
14. 赵铁军（2000）《机器翻译原理》，哈尔滨工业大学出版社2000年版。
15. 詹卫东（2000）《面向中文信息处理的现代汉语短语结构规则研究》，清华大学出版社、广西科学技术出版社2000年版。



# 参考资料（续）

16. James Allen, 1995, *Natural Language Understanding*, 2<sup>nd</sup> Edition, The Benjamin/Cummings Publishing Company, Inc. (译本: 刘群 等译, 2005, 《自然语言理解》, 电子工业出版社)
17. Christopher D. Manning & Hinrich Schutze, 1999, *Foundations of Statistical Natural Language Processing*, The MIT Press. (译本: 苑春法 等译, 2005, 《统计自然语言处理基础》, 电子工业出版社)
18. Daniel Jurafsky & James H. Martin, 2008, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2<sup>nd</sup> Edition, Prentice-Hall Inc.. (译本: 冯志伟、孙乐译, 2005, 《自然语言处理综论》, 电子工业出版社 (原著D.Jurafsky & J.H.Martin, 1<sup>st</sup> Edition, 2000))
19. 刘颖, 2002, 《计算语言学》, 清华大学出版社。
20. 俞士汶、常宝宝、詹卫东, 2003, 《计算语言学概论》, 商务印书馆。
21. 冯志伟, 2004, 《机器翻译研究》, 中国对外翻译出版公司。
22. 冯志伟, 2010, 《自然语言处理的形式模型》, 中国科学技术大学出版社。
23. 冯志伟, 2012, 《自然语言处理简明教程》, 上海外语教育出版社。
24. 刘群, 2008, 《汉英机器翻译若干关键技术研究》, 清华大学出版社。
25. 宗成庆, 2008/2013, 《统计自然语言处理》(第二版), 清华大学出版社。
26. Koehn, Philipp, 2009, *Statistical Machine Translation*, Cambridge University Press (译本: 宗成庆, 张霄军 译, 2012, 《统计机器翻译》, 电子工业出版社)
27. Steven Bird, Ewan Klein, Edward Loper, 2009, *Natural Language Processing with Python*, O'Reilly Media. (译本: 陈涛 等译, 2014 《Python自然语言处理》 人民邮电出版社)
28. James Pustejovsky, Amber Stubbs, 2012, *Natural Language Annotation for Machine Learning*, O'Reilly Media. (东南大学出版社, 2013年影印版)
29. Emily M. Bender, 2013, *Linguistic Fundamentals for Natural Language Processing*, Morgan & Claypool Publishers.
30. Yoav Goldberg, 2017, *Neural Network Methods for Natural Language Processing*, Morgan & Claypool Publishers. (车万翔 等译, 2018 《基于深度学习的自然语言处理》, 机械工业出版社)



# 参考资料

---

<http://framenet.icsi.berkeley.edu/> (FrameNet)

<http://wordnet.princeton.edu/> (WordNet)

<https://openhownet.thunlp.org/> (HowNet)

<https://universaldependencies.org/> (Universal Dependency)

<https://aclanthology.org/>

<http://jcip.cipsc.org.cn/CN/home> (中文信息学报)

<https://www.cipsc.org.cn/> (中国中文信息学会)



# 考核

---

- 两次作业（ $2 \times 20\%$ ）
- 期末报告（ $60\%$ ）



# 致谢

---

- 在撰写本课程讲义和教学过程中，得到许多专家学者的帮助：

俞士汶、常宝宝、于江生、孙斌、张化瑞、胡俊峰（北京大学计算语言学研究所）

刘群、王斌、白硕（中科院计算所）

周强、刘颖（清华大学）

在此深表谢意