计算语言学
Computational Linguistics

教师：孙茂松
Tel:62781286

Email:sms@tsinghua.edu.cn

TA:李文浩

Email:wh-li20@mails.tsinghua.edu.cn

郑重声明

- 此课件仅供选修清华大学计算机系研究生课《计算语言学》(70240052)的学生个人学习使用，所以只允许学生将其下载、存贮在自己的电脑中。未经孙茂松本人同意，任何人不得以任何方式扩散之（包括不得放到任何服务器上）。否则，由此可能引起的一切涉及知识产权的法律责任，概由该人负责。
- 此课件仅限孙茂松本人讲课使用。除孙茂松本人外，凡授课过程中，PTT文件显示此《郑重声明》之情形，即为侵权使用。

第二章
自然语言的特点及其计算复杂性（Part 1）

2.1. 自然语言的特点

离散无限性 (Discrete Infinity)

The discrete infinity of language means **unlimited productivity** from the finite means as a major design feature of language (Irvine, 2014).

Discreteness means that the boundary between linguistic symbols is clear. Since the linguistic symbols are discrete, the chain of linguistic symbols can be segmented part by part until the smallest linguistic symbols are assigned.

https://blogs.commons.georgetown.edu/cctp-748-spring2018/2018/02/07/discrete-infinity-of-language/

2.1. 自然语言的特点

"For example, the most elementary property of the language faculty is the property of discrete infinity; you have six-word sentences, seven-word sentences but you don't have six-and-a-half-word sentences. Furthermore, there is no limit; you can have ten-word sentences, twenty-word sentences and so on indefinitely. That is the property of discrete infinity. This property is virtually unknown in the biological world. There are plenty of continuous systems, plenty of finite systems but try to find a system of discrete infinity! The only other one that anybody knows is the arithmetical capacity, which could well be some offshoot of the language faculty. The more you go on the more it seems true."
Noam Chomsky, The Architecture of Language

"Language makes infinite use of finite means."
Wilhelm von Humboldt

2.1. 自然语言的特点

+ 语言符号的任意性 (Arbitrariness)

The meaning of linguistic signs is not predictable from its word form, nor is the word form dictated by its meaning/function.

The human language is completely arbitrary with very few exceptions of onomatopoeia and sound symbolism.

It frees up the communication system. Because it is not a must for word form to have a correlation to the meanings, this makes it much convenient for words to refer to abstract concepts.

+ 符号的线性序列

Journalists say that when a dog bites a man that is not news, but when a man bites a dog that is news (Pinker, 1994).