从计算视角对语言学相关课程体系的初步考察与分析

清华大学计算机科学与技术系

孙茂松

第二届语言资源与智能国际学术研讨会

2018年12月16日，北京语言大学，北京
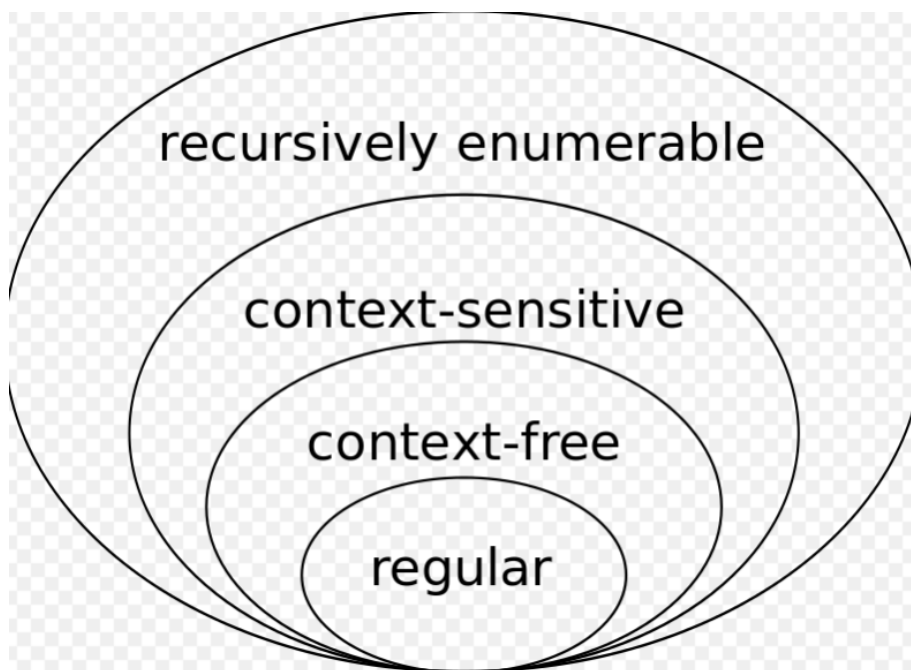
从**Noam Chomsky**谈起

- ## Chomsky hierarchy(1957)

THREE MODELS FOR THE DESCRIPTION OF LANGUAGE[*]

Noam Chomsky

Department of Modern Languages and Research Laboratory of Electronics
Massachusetts Institute of Technology
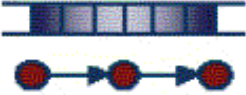Cambridge, Massachusetts

recursively enumerable

context-sensitive

context-free

regular

Chomsky and automata theory(1959)

On Certain Formal Properties of Grammars*

NOAM CHOMSKY

Massachusetts Institute of Technology, Cambridge, Massachusetts and The Institute for Advanced Study, Princeton, New Jersey

Language	Automaton	Grammar	Recognition	Dependency
Recursively Enumerable Languages	Turing Machine	Unrestricted Baa → A	Undecidable	Arbitrary
Context-Sensitive Languages	Linear-Bounded	Context-Sensitive At → aA	NP-Complete	Crossing
Context-Free Languages	Pushdown (stack)	Context-Free S → gSc	Polynomial	Nested
Regular Languages	Finite-State Machine	Regular A → cA	Linear	Strictly Local

THEOREM 2. Every recursively enumerable set of strings is a type 0 language (and conversely).[7]

That is, a grammar of type 0 is a device with the generative power of a Turing machine. The theory of type 0 grammars and type 0 languages is thus part of a rapidly developing branch of mathematics (recursive function theory). Conceptually, at least, the theory of grammar can be viewed as a study of special classes of recursive functions.

THEOREM 3. Each type 1 language is a decidable set of strings.[7a]

That is, given a type 1 grammar G, there is an effective procedure for determining whether an arbitrary string x is in the language enumerated by G. This follows from the fact that if φ_i, φ_{i+1} are successive lines of a derivation produced by a type 1 grammar, then φ_{i+1} cannot contain fewer symbols than φ_i, since φ_{i+1} is formed from φ_i by replacing a single symbol A of φ_i by a non-null string ω. Clearly any string x which has a

Chomsky normal form

CYK algorithm (the Cocke–Younger–Kasami algorithm)

CYK table

S						
	VP					
S						
	VP			PP		
S		NP			NP	
NP	V, VP	Det.	N	P	Det	N
she	eats	a	fish	with	a	fork

Brief Observation on the Department of Linguistics and Philosophy, MIT

- Undergraduate program

   Program 1: Philosophy

   Program 2: Linguistics

- Curriculum

Required subject common to both tracks↵

One of the following two subjects:↵

- 24.900 Introduction to Linguistics (CI-H)↵
- 24.9000 How Language Works↵

Required Subjects for Linguistics Track

24.901 Language and Its Structure I: Phonology
24.902 Language and Its Structure II: Syntax
24.903 Language and Its Structure III: Semantics and Pragmatics
24.918 Workshop in Linguistic Research (CI-M)

One of the following three Linguistic Analysis subjects:

24.909 Field Methods in Linguistics (LAB, CI-M)
24.910 Advanced Topics in Linguistic Analysis (CI-M)
24.914 Language Variation and Change (CI-M)

One of the following four Philosophy subjects:

24.08J Philosophical Issues in Brain Science (CI-H)
24.09 Minds and Machines (CI-H)
24.241 Logic I
24.251 Introduction to Philosophy of Language

One of the following four Experimental Results subjects:

24.904 Language Acquisition
24.905J Psycholinguistics
24.906J The Linguistic Study of Bilingualism (CI-H)
24.915 Linguistic Phonetics

Required Subjects for Philosophy track

24.251 Introduction to the Philosophy of Language (CI-M)
24.260 Topics in Philosophy (CI-M)

One Logic Subject

24.118 Paradox and Infinity
24.241 Logic I
24.242 Logic II
24.243 Classical Set Theory
24.244 Modal Logic
24.245 Theory of Models

or a logic subject from another department (e.g., Mathematics) with the approval of the major advisor.

Two of the following Knowledge and Reality subjects:

24.05 Philosophy of Religion (CI-H)
24.08J Philosophical Issues in Brain Science (CI-H)
24.09 Minds and Machines (CI-H)
24.111 Philosophy of Quantum Mechanics
24.112 Space, Time, and Relativity
24.115 Philosophy and Time
24.211 Theory of Knowledge
24.215 Topics in the Philosophy of Science
24.221 Metaphysics (CI-M)
24.253 Philosophy of Mathematics
24.280 Foundations of Probability

One of the following four subjects:

9.65 Cognitive Processes
24.903 Language and Its Structure III: Semantics and Pragmatics
24.904 Language Acquisition
24.905J Psycholinguistics

Graduate Program in Linguistics at MIT

- Founded in 1961, and produced its first PhDs in 1965.

- In the 1990's links with the Department of Brain and Cognitive Science were established to expand the range of research tools and methodologies available for the study of human language. In the current decade the program has integrated research in experimental phonetics and computational modeling of language learning.

Specialization in Experimental Linguistics

- **Quantitative Methods requirement:** A two semester graduate level course sequence in statistics and experimental design such as Harvard Psy_1950 + Psy_1952.

- **Additional subjects:** Three graduate level courses in diverse areas of experimental or computational research on language and its interfaces. The list of possible courses is vast, and relevant course offerings vary across years. A sample of the possibilities includes: Systems Neuroscience (9.011J); Cognitive Science (9.012);); Natural Language and the Computer representation of Knowledge (9.611J); Computational Cognitive Science (9.660); Functional MRI Investigations of the Human Brain (9.71); Special Subject in Brain and Cognitive Sciences (9.S911); Speech Communication (6.541J); Machine Learning (6.867); Topics in Experimental Phonology (24.967). One of these courses may be an independent study or independent laboratory research course.'

一位典型的跨界学者

Christopher Manning

Thomas M. Siebel Professor in Machine Learning, Professor of Linguistics and of Computer Science
Director, Stanford Artificial Intelligence Laboratory (SAIL)

Brief CV

- I'm Australian ("I come from a land of wide open spaces ...")
- BA (Hons) Australian National University 1989 (majors in mathematics, computer science and linguistics)
- PhD Stanford Linguistics 1994
- Asst Professor Carnegie Mellon University Computational Linguistics Program 1994-96
- Lecturer University of Sydney Dept of Linguistics 1996-99
- Asst Professor Stanford University Depts of Computer Science and Linguistics 1999-2006
- Assoc Professor Stanford University Depts of Linguistics and Computer Science 2006-2012
- Professor Stanford University Depts of Linguistics and Computer Science 2012-
- President of the Association for Computational Linguistics 2015

Core Courses

Students must take at least one course each from two of the following three areas below:

- Phonetics and Phonology: 105/205A, 110
- Morphology and Syntax: 120, 121A, 121B, 222A
- Semantics and Pragmatics: 130A/230A, 130B, 230B, 232A

Breadth Courses

Students must take at least one course each from two of the following four areas:

- Historical Linguistics and Language Change: 160, 260A, 260B
- Sociolinguistics: 150, 156, 157/257, 250
- Psycholinguistics: 35, 140, 145
- Computational Linguistics: 180/280, 188/288, 278, 284

Depth Courses

- At least two 200-level **Linguistics courses**.
- If the courses are cross-listed at the 100 and 200 levels, please enroll at the 200 level, if you wish the course to count for this requirement.

卡内基梅隆大学计算机科学学院 "人工智能专业"（**2018年招生**）

* Carnegie Mellon has led the world in artificial intelligence education and innovation since the field was created.

* B.S. in AI (BSAI) : the nation's first bachelor's degree in artificial intelligence.

* The BSAI program includes faculty members from the school's Computer Science Department, Human-Computer Interaction Institute, Institute for Software Research, Language Technologies Institute, Machine Learning Department and Robotics Institute.

ARTIFICIAL INTELLIGENCE CORE
4 Classes

Concepts in Aritificial Intelligence

Introduction to AI Representation and Problem Solving

Introduction to Machine Learning

Introduction to Natural Language Processing *OR* Introduction to Computer Vision

COMPUTER SCIENCE CORE
6 Classes

Freshman Immigration Course

Principles of Imperative Computation

Principles of Functional Programming

Parallel and Sequential Data Structures and Algorithms

Introduction to Computer Systems

Great Theoretical Ideas in Computer Science

MATH AND STATISTICS CORE
6 Classes

Math Foundations of Computer Science

Differential and Integral Calculus

Integration and Approximation

Matrices and Linear Transformations

Probability Theory for Computer Scientists

Modern Regression

AI CLUSTER ELECTIVES
4 Classes

Take one course from each of the following areas:

Decision Making and Robotics Cluster

Machine Learning Cluster

Perception and Language Cluster

Human-AI Interaction Cluster

ETHICS ELECTIVE
1 Class

Choose from one of the following:

Freshman Seminar: Artificial Intelligence and Humanity

Ethics and Policy Issues in Computing

AI, Society and Humanity

HUMANITIES AND ARTS
7 Classes

BSAI students take seven courses in the humanities and arts as part of the SCS General Education requirements. Of the seven Humanities and Arts courses in the curriculum, one must be in cognitive science or cognitive psychology.

SCIENCE AND ENGINEERING
4 Classes

BSAI students take four courses in science and engineering as part of the SCS General Education requirements.

两个极端

- *But it must be recognized that the notion "probability of a sentence" is an entirely useless one, under any known interpretation of this term.* Noam Chomsky(1960).

  poverty of the stimulus
  语言的关键部分是天生的

- *Anytime a linguist leaves the group the recognition rate goes up.* Fred Jelinek (IBM speech group, 1988)

18 Nov.1932 - 14 Sept.2010

Frederick Jelinek, Who Gave Machines the Key to Human Speech, Dies at 77

By STEVE LOHR
Published: September 24, 2010

Frederick Jelinek, who survived the Nazi occupation of Czechoslovakia to become a pioneer in computer research in America, helping to make it possible for computers to decipher and translate human speech, died on Sept. 14 in Baltimore. He was 77.

Enlarge This Image

The cause was a heart attack, his son, William, said. Mr. Jelinek was stricken while he was in his office at Johns Hopkins University, where he was a professor. He lived in Baltimore and New York.

Today, computerized speech recognition is becoming a mainstream technology. A few words spoken into a smartphone can summon an Internet search; doctors use voice-transcription software for patient records; drivers talk to speech-recognition systems in cars that reply with driving directions; and customer questions to call centers are increasingly being answered by automated speech systems.

The Johns Hopkins University
Frederick Jelinek

Add to Portfolio

建议

■ 定位：纯粹的理论语言学？

■ 语言及应用语言学
   + 加强统计学
   + 加强计算思维，会编程
   + 加强形式逻辑学
   + 人工智能微专业
   + 邀请计算机的教授作为语言学系的双聘教授

谢谢！