Introduction and Methodology

Theoretical Linguistics and Formal Learning Theory

Alexander Clark

Manchester

6 November 2013

Outline

1. Methodology
    1.1 Learnability in Linguistic Theory
    1.2 Weak, Strong, Semantic Learning
2. Learning theory in general
    2.1 Nontechnical introduction
    2.2 Gold (1967)
    2.3 Probabilistic learning
3. Distributional Learning of Context-Free Grammars
    3.1 Weak learning of substitutable
    3.2 Strong learning
    3.3 Lattice based theories
4. Beyond Context-free grammars; and Conclusion.

Topic

What is the course about?

The central problem in linguistics is the tension between learnability and expressive power.

- Understand the problems of learnability.
- Sketch a possible solution.

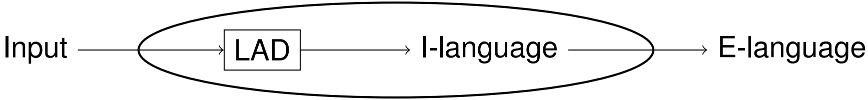
Outline

Chomskyan view

Chomsky's questions (1986)

1. What constitutes knowledge of a language?
2. How is this knowledge acquired by its speakers?

Standard view



Input → LAD → I-language → E-language

Levels of adequacy

Chomsky, 1986

> *To achieve descriptive adequacy it often seems necessary to enrich the system of available devices, whereas to solve our case of Plato's problem we must restrict the system of available devices so that only a few languages or just one are determined by the given data. It is the tension between these two tasks that makes the field an interesting one, in my view.*

Chomsky, Hauser and Fitch (2005)

Two basic conditions that UG must satisfy are that it (1) accommodate the attainable I-languages, and (2) account for their acquisition.

Tension between two goals

Find a class of languages:

Expressive power

large and **rich** enough to account for what we observe in natural languages.

Learnability

small enough to be learnable

These goals are in conflict!.

Expressive power

One goal at a time

Two components:

- Rich enough – where in the Chomsky hierarchy are NLs?
- Large enough – how varied are they? Can any context-free language be a natural language?

Mathematical linguistics

Try to define the right level of complexity.

Descriptive linguistics

Interesting in describing natural languages
Most syntax!

Learnability

A range of models

Learnability is not that clear cut

The real situation of the learner is very complex

- The child hears certain utterances
- Can extract prosodic and phonological cues
- Observes situational context
- Can infer the speakers intention
- Can generate her own partial utterances
- Interacts with the speakers and other participants

There are many different models, some harder than others, that idealise in different ways.

Learnability

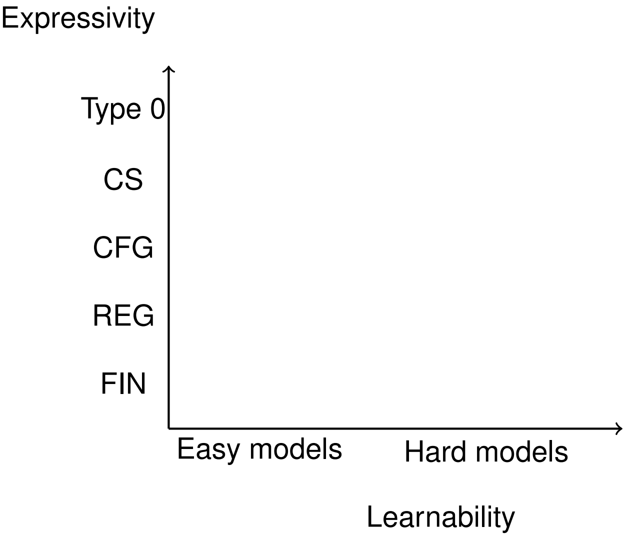
One goal at a time

Learnability in nonlinguistic domains

Computational learning theory/grammatical inference studies learnability in general:

- Simple Markov models/locally testable languages
- Finite state models under various assumptions
- Pattern languages
- Many other classes of languages that are not linguistically relevant are learnable

Two criteria



Outline

The fundamental problem of ornithology

Problem

Birds are heavier than air and yet they fly

Answer

Dogs, humans etc cannot fly:
Birds must have an innate biologically determined ability to fly:
A “faculty of flight”

Aristotle

The parts of animals/the movement of animals

- For it is of the essence of a bird that it shall be able to fly; and it is by the extension of wings that this is made possible.
- So too flying and swimming things progress, the one straightening and bending their wings to fly, the other their fins to swim.
- And so if the wings be cut off a bird can neither stand still nor go forwards.

Progress

Aerodynamics

Progress depended on developing a theory of aerodynamics, a branch of physics:

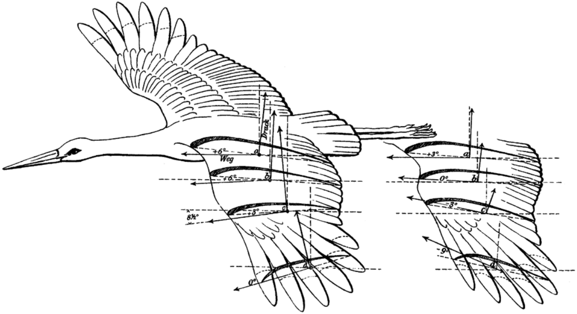
- Newton (1687)
- Navier (1802) and Stokes (1842-1845)

Application to bird flight

- Lord Rayleigh (1883)
- Otto Lilienthal (1889)

Bird Flight as the Basis of Aviation

Lilienthal, 1889



Flügel eines Storches *beim Niederschlag* *beim Aufschlag*

Poverty of lift problem

Dickinson et al., Science (1999)

While such complex aerial feats involve many physiological and anatomical specializations that are poorly understood, perhaps the greatest puzzle is how flapping wings can generate enough force to keep an insect in the air.

A recurrent problem in this field:

- Rayleigh (1883), Rayner(1979) , . . .

Poverty of lift problem

Dickinson et al., Science (1999)

When insect wings are placed in a wind tunnel and tested over the range of air velocities that they encounter when flapped by the animal, the measured forces are substantially smaller than those required for active flight

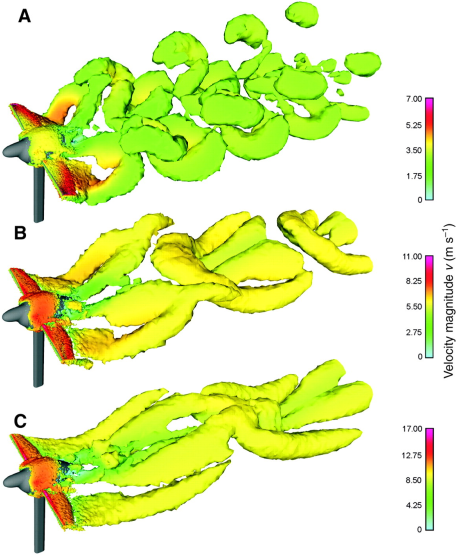
Poverty of lift problem

Dickinson et al., Science (1999)

When insect wings are placed in a wind tunnel and tested over the range of air velocities that they encounter when flapped by the animal, the measured forces are substantially smaller than those required for active flight

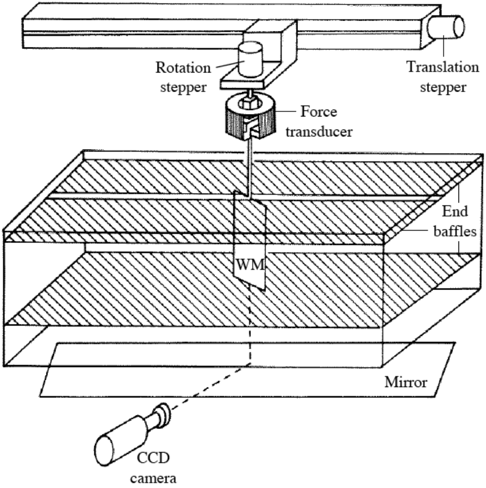
Answer: clap and fling, LEV, delayed stall, rotational circulation, and wake capture

Combine observation and simulations



Modeling mosquito and fruit fly

Dickinson et al. (1993)



Underlying problem

- Animal flight is above all a problem of physics: You cannot explain flight without aerodynamics.
- A necessary preliminary to constructing a theory of animal flight is to construct a theory of aerodynamics, that covers the appropriate conditions.

Birds	
	• in air at about 1 bar • speeds from 10-200mph • Reynolds number 10^3 to 10^5

Connecting theory to the empirical problem

Idealised mathematical model of aerodynamics

- Rigid 2d-airfoil in incompressible fluid in a steady state.

Connecting theory to the empirical problem

Idealised mathematical model of aerodynamics

- Rigid 2d-airfoil in incompressible fluid in a steady state.

Connecting aerodynamics to bird flight

- We observe that the shape of the wing is like an airfoil.
    1. This gives a (partial) explanation of bird flight.

Connecting theory to the empirical problem

Idealised mathematical model of aerodynamics

- Rigid 2d-airfoil in incompressible fluid in a steady state.

Connecting aerodynamics to bird flight

- We observe that the shape of the wing is like an airfoil.
    1. This gives a (partial) explanation of bird flight.
    2. This goes ’beyond explanatory adequacy’ to give an explanation of the shape of bird wings.

Interpreting failure

(Until recently) aerodynamics could not explain insect flight.

Poverty of lift/stimulus

- The failure of steady state aerodynamics to explain all aspects of flight is not a reason to reject aerodynamics, and hypothesize other forces and other explanations.

Interpreting failure

(Until recently) aerodynamics could not explain insect flight.

Poverty of lift/stimulus

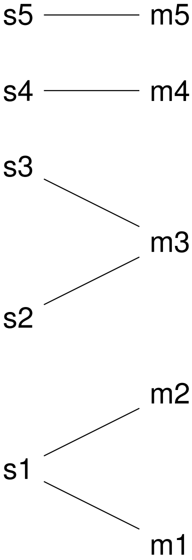
- The failure of steady state aerodynamics to explain all aspects of flight is not a reason to reject aerodynamics, and hypothesize other forces and other explanations.
- It is a reason take up aerodynamics!

Outline

What are the linguistic data?

- (E-)Languages are collections of sound/meaning pairs
- More precisely, we observe sounds/gestures.
- We don't observe the meanings, but we are aware of entailment relations between propositions. (or something similar)
- The sound/meaning relation is not one-to-one, but there is a lot of structure. (knowing what was said tells you a lot about the meaning!)

Sound/meaning pairs



s5 m5

s4 m4

s3

m3

s2

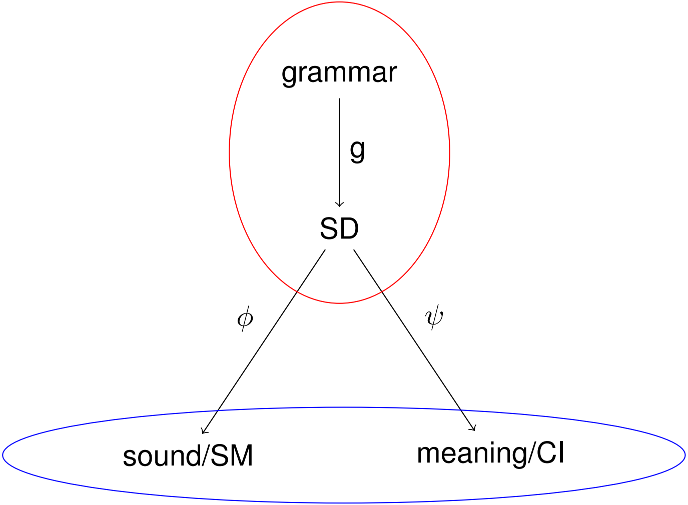
m2

s1

m1

Structural descriptions

Traditional view



grammar

g

SD

ϕ

ψ

sound/SM

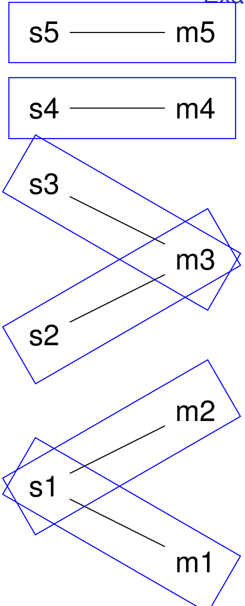
meaning/CI

Example

- Grammar might be a context-free grammar, augmented with some compositional semantics.
- Structure might be a labeled derivation (parse) tree.
- ϕ is just the yield of the tree (a function).
- ψ just computes the semantics (another function).

Standard model

Exactly one SD per string/meaning pair



Issues

Description

If you are just describing the data, then you can stipulate this as a requirement.
If you are *learning* then you may not be able to learn grammars that satisfy this stipulation.
So you need to decide whether this stipulation is necessary.

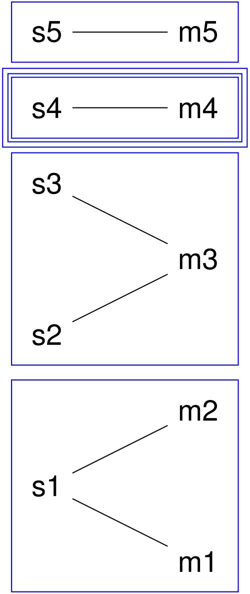
Issues

Description

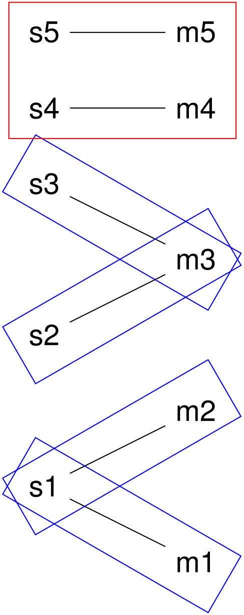
If you are just describing the data, then you can stipulate this as a requirement.
If you are *learning* then you may not be able to learn grammars that satisfy this stipulation.
So you need to decide whether this stipulation is necessary.

- Everybody here speaks two languages.
- Either way is fine.
- I gave him a cake and her a biscuit/I gave him a cake.
- John likes but Mary hates Sibelius/John likes Sibelius

Alternative model



Invalid structures



Weak and Strong

Weak generative capacity

Set of strings that a grammar can generate
Mathematically tractable

Strong generative capacity

Set of structures that a grammar can generate
Linguistically interesting but unobservable

From a learnability point of view, this is not so crucial: the data that the child sees consists of semantically well formed utterances.

Outline

Introduction

Birds and bees

Weak and Strong

Example

Running example

Propositional logic

Alphabet

propositional symbols	$A_1, A_2, \ldots$
binary connectives	$\wedge, \vee, \rightarrow, \leftrightarrow$
negation symbol	$\neg$
brackets	$(,)$

- A_1
- $(A_1 \rightarrow A_3)$
- $(A_1 \rightarrow (\neg A_4))$

Running example

Propositional logic

Alphabet

propositional symbols	rain, snow, hot, cold, danger
binary connectives	and, or, implies, iff
negation symbol	not
brackets	open, close

- rain
- open snow implies cold close
- open snow implies open not hot close close

Classic model

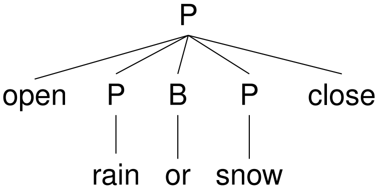
open rain or snow close

Classic model

open rain or snow close

Q	R	S	
T	T	T	T
T	T	F	T
T	F	T	T
. . .	. . .	. . .	
F	F	F	F

Classic model



open rain or snow close

Q	R	S	
T	T	T	T
T	T	F	T
T	F	T	T
. . .	. . .	. . .	
F	F	F	F

Simple example

“You must study the simplest system you think has the properties you are interested in”[Platt(1964)]

- semantically interpreted language
- Infinite non-regular language
- Hierarchically structured expressions
- Missing: ambiguity, vagueness, reference, illocutionary force…

Models of learning

Three classes of objects

strings, meanings, and trees.

Accessibility for the child:

- Strings – complete
- Meanings – partial
- Trees – no information at all

Weak learning

Learning model

	Inputs	Outputs
Weak	strings	strings

Learner must acquire a grammar that defines an infinite set of strings on the basis of a finite set of strings.

Grammatical Strings only

Learning models

open rain and snow close

Grammatical Strings only

Learning models

open rain and snow close
cold

Grammatical Strings only

Learning models

open rain and snow close
cold
open snow implies open not hot close close

A second model

Two mathematically reasonable models

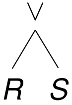
	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

What meanings?

open rain or snow close

What meanings?

open rain or snow close



What meanings?

open rain or snow close

Q	R	S	
T	T	T	T
T	T	F	T
T	F	T	T
...	...	...	
F	F	F	F

Standard view: Semantic Bootstrapping

[Pinker(1995)]

> *Many models of language acquisition assume that the input to the child consists of a sentence and a representation of the meaning of that sentence, inferred from context and from the child's knowledge of the meanings of the words (e.g. Anderson, 1977; Berwick, 1986; Pinker, 1982, 1984; Wexler & Culicover, 1980). Of course, this can't literally be true – children don't hear every word of every sentence, and surely don't, to begin with, perceive the entire meaning of a sentence from context.*

Language meanings

Trivial

John likes cookies

$\textsc{likes}(\textsc{john},\textsc{cookies})$

Non-trivial [Kanazawa(2001)]

John is looking for a cookie

$\mathbf{try}(\lambda z.\exists(\lambda y. \wedge (\mathbf{cookie}\, y(\mathbf{find}\, y\, z)))\mathbf{john}$

Arguments against this view

1. Impossible for children actually to do this.
2. Assumes precisely the ability which needs to be explained.
3. Language would be unnecessary.
4. Language dependence of the semantic representation.
5. Language acquisition starts very early.
6. (Blind children acquire language with only minor delay.)

Why is this view held?

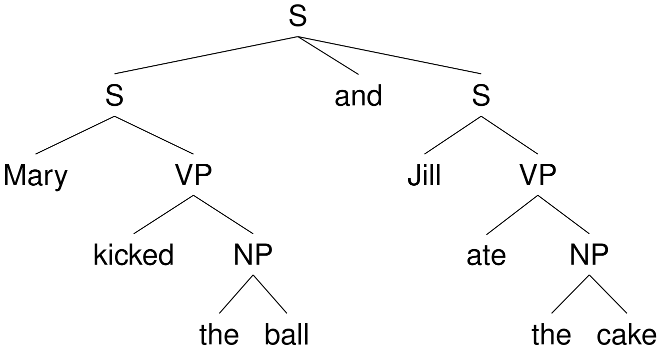
[Steedman(1996)]

> *As soon as it is recognized that the very earliest stages in acquiring syntax require some language-independent source of information about grammatical categories and grammatical relations, the only plausible source that has ever been identified is the semantic interpretation that underlies the utterance. . . . sooner or later, the child needs access to semantic interpretations in order to acquire syntactic competence.*

English example

Which structure gives you the right dependencies?

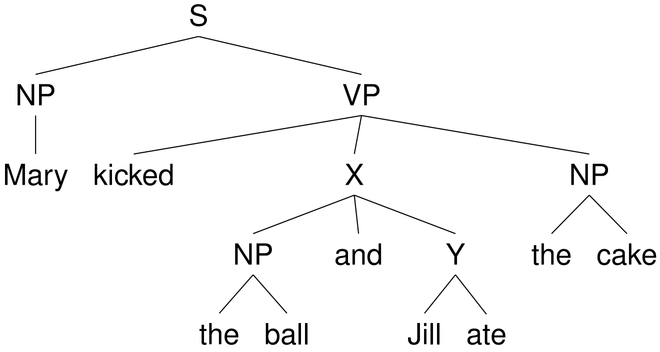
- Mary kicked the ball and Jill ate the cake



English example

Which structure gives you the right dependencies?

- Mary kicked the ball and Jill ate the cake



Semantically well-formed

- Mary kicked the ball and Jill ate the cake
- # Mary fed the ball and Jill ate the cake
- # Mary kicked the cake and Jill ate the cake
- # Mary fed the ball and Jill ate the cat
- Mary fed the cat and Jill ate the cake

Crucial to consider semantic wellformedness, not just syntactic wellformedness, especially when we come to think about learnability.

Strong learning

Two mathematically reasonable models

	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

Strong learning

Two mathematically reasonable models

	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

A mathematically unreasonable model

[Wexler and Culicover(1980)]

	Inputs	Outputs
Strong learning	strings	strings + trees

Individuating languages

Suppose we have two speakers who agree on all sound/meaning pairings.
Do we expect them to have:

- Grammars that are identical
- Grammars that generate the same sets of structural descriptions?

[Berwick et al.(2011)Berwick, Pietroski, Yankama, and Chomsky]

> *Put another way, language acquisition is not merely a matter of acquiring a capacity to associate word strings with interpretations. Much less is it a mere process of acquiring a (weak generative) capacity to produce just the valid word strings of a language. Idealizing, one can say that each child acquires a procedure that generates boundlessly many meaningful expressions, and that a single string of words can correspond to more than one expression.*

Bibliography I

- R.C. Berwick, P. Pietroski, B. Yankama, and N. Chomsky.
  Poverty of the stimulus revisited.
  Cognitive Science, 35:1207–1242, 2011.
- Makoto Kanazawa.
  Learning word-to-meaning mappings in logical semantics.
  In *Proceedings of the Thirteenth Amsterdam Colloquium*, pages 126–131, 2001.
- Steven Pinker.
  Language acquisition.
  In Daniel Osherson, Lila R. Gleitman, and Mark Liberman, editors, *An invitation to cognitive science: Volume 1 Language*, chapter 6, pages 135–182. MIT Press, 2nd edition, 1995.

Bibliography II

John R Platt.
Strong inference.
Science, 146(3642):347–353, 1964.

M. Steedman.
Phrasal intonation and the acquisition of syntax.
In J.L. Morgan and K. Demuth, editors, *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pages 331–342. Lawrence Erlbaum Associates, 1996.

K. Wexler and P. W. Culicover.
Formal Principles of Language Acquisition.
MIT Press, Cambridge, MA, 1980.