

Lecture 1: Fundamentals

Learnability and Language Acquisition

Alexander Clark

Department of Computer Science
Royal Holloway, University of London

January 2012
LOT 2012

Topic

What is the course about?

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

- Understand the problem.
- Sketch a possible solution.

Course outline

Roughly:

- Monday: Fundamentals
- Tuesday: Expressivity
- Wednesday: Learnability (Shalom Lappin)
- Thursday: Distributional learning
- Friday: Empirical learnability

Outline

Introduction

Linguistics

The key debate: Linguistic Nativism

The Argument from the Poverty of the Stimulus

Empirical APS

Formal versions of the APS

Conclusion

Outline

Introduction

Linguistics

The key debate: Linguistic Nativism

The Argument from the Poverty of the Stimulus

Empirical APS

Formal versions of the APS

Conclusion

What is linguistics?

- The scientific study of language.
- What happens in linguistics departments.

Mainstream Chomskyan view

Linguistics is a branch of psychology

Ultimately of biology

Object of study

Primary object of study

Adults have the ability to generate and understand sentences in their native language(s).

- We will call this “knowledge of language”
- Not interested in questions like “Is it propositionally represented?”
- Think of it as some “machinery”

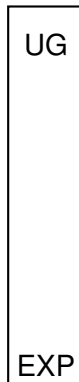
Chomskyan view

Chomsky's questions (1986)

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

Where does language come from?

Two factors: UG and experience



Where does language come from?

Principles and Parameters



Where does language come from?

MP/Biolinguistics



Where does language come from?

Distributional learning



Three factors

Chomsky (2005)

- “1. Genetic endowment, apparently nearly uniform for the species, which interprets part of the environment as linguistic experience, a nontrivial task that the infant carries out reflexively, and which determines the general course of the development of the language faculty. . . .
2. Experience, which leads to variation, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally.
3. Principles not specific to the faculty of language.”

Third factor principles

Chomsky, 2005/2006

“The third factor falls into several subtypes: (a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints that enter into canalization, organic form, and action over a wide range, including principles of efficient computation, which would be expected to be of particular significance for computational systems such as language. It is the second of these subcategories that should be of particular significance in determining the nature of attainable languages.”

“It also might turn out that general cognitive principles that enter into language acquisition pose conditions on FL design.”

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

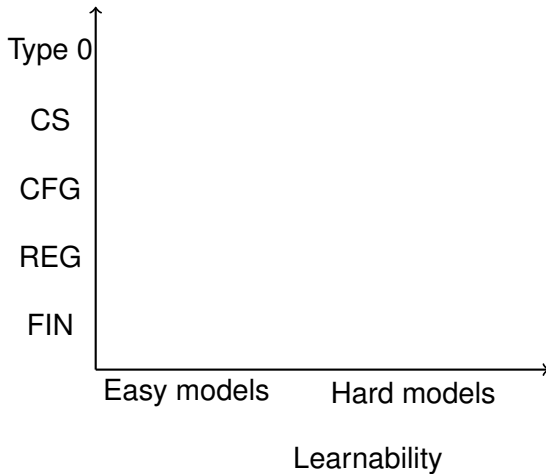
Learnability in nonlinguistic domains

Computational learning theory/grammatical inference studies
learnability in general:

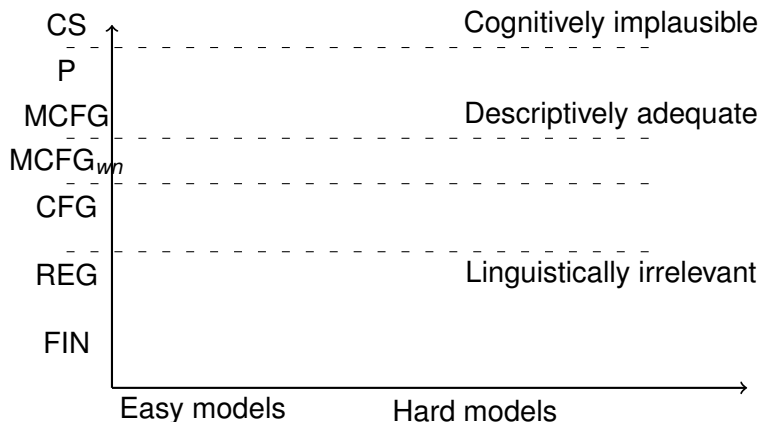
- Simple Markov models/locally testable languages
- Finite state models under various assumptions
- Pattern languages
- Many other classes of languages that are not linguistically relevant are learnable

Two criteria

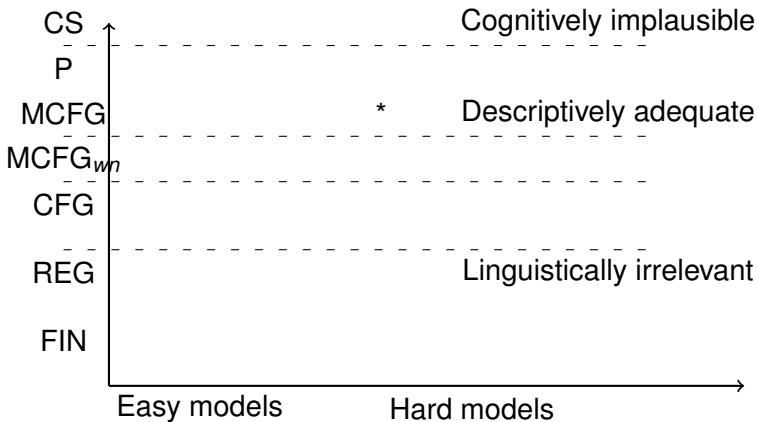
Expressivity



Two criteria

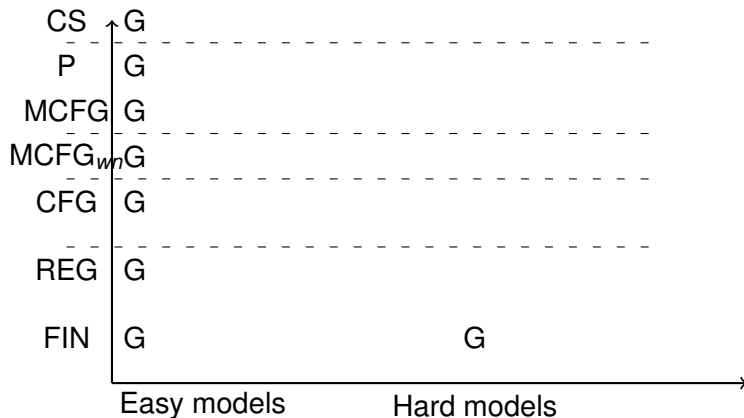


Goal



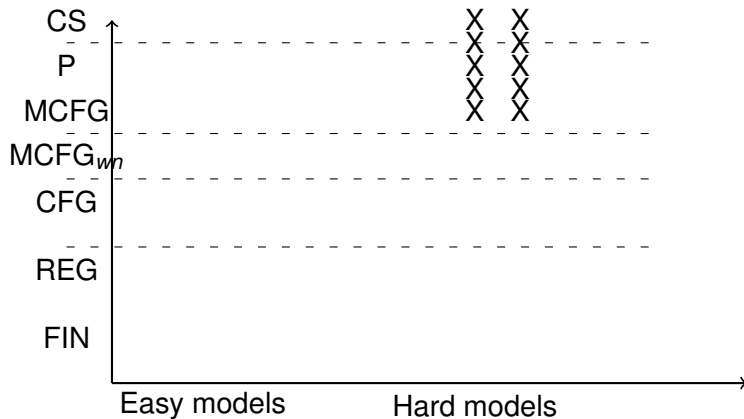
Two criteria

Gold, 1967

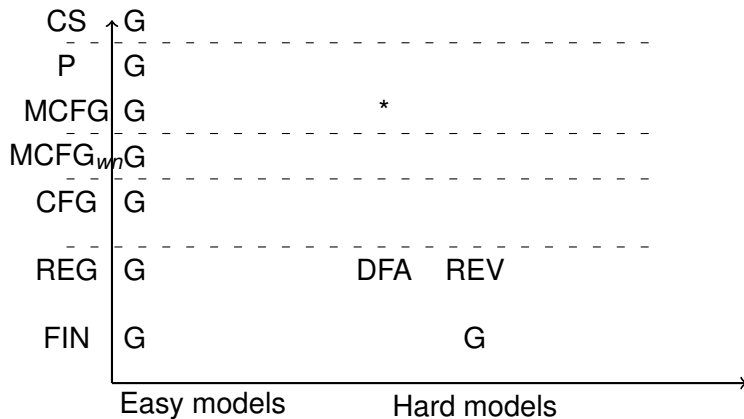


Two criteria

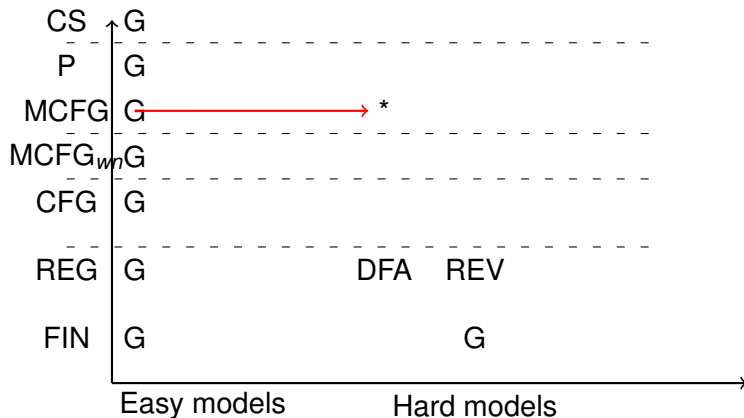
Principles and Parameters



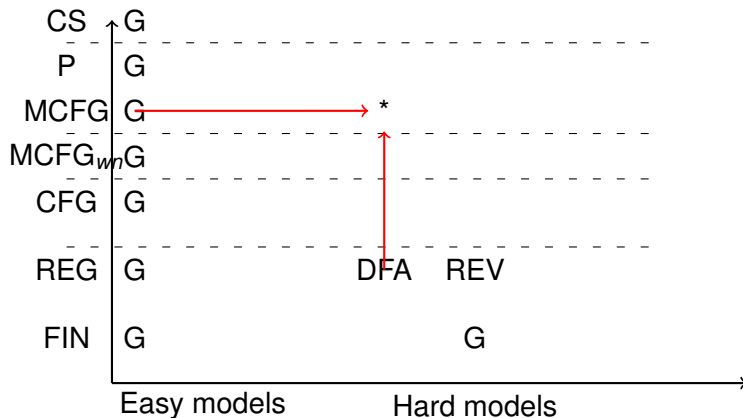
Two criteria



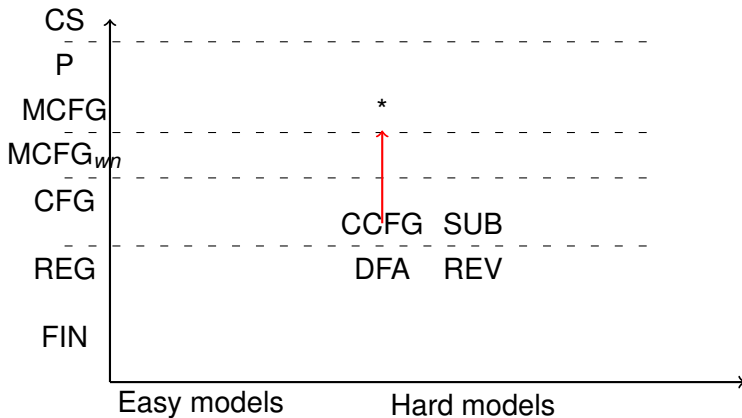
Two criteria



Two criteria



Two criteria



What is linguistics a theory of?

Chomskyan view:

- Grammars are natural objects in the brain
- We don't want a theory of individual grammars
- We want theories of language acquisition

Alternate view:

- A grammar is a theory of a language
- We want simple grammars because we want simple theories

Where do grammars come from?

Traditional view

(Linguists write grammars)

- Efficient for parsing (perhaps)
- Intuitive and easy for linguists to use
- Expressive enough to capture any generalisation we want to capture
- (We want them to be small)

Psychological view

(Children construct grammars)

- Efficient for processing
- Learnable from Primary Linguistic Data
- Expressive enough that we can have a compact representation
- (They are the size they are)

Size of grammars

Discussion

Two competing analogies:

- Grammars are scientific theories
- Grammars are like the genome of an organism

Comments:

- It's not a criticism of genetics that it claims that we have 2.9 billion BPs.
- Maybe there are different notions of simplicity? Type/token confusion? Methodological/Empirical/Metaphysical?

Outline

Introduction

Linguistics

The key debate: Linguistic Nativism

The Argument from the Poverty of the Stimulus

Empirical APS

Formal versions of the APS

Conclusion

The Central Debate

How does language acquisition proceed?

Empiricist hypothesis

It proceeds largely through domain-general processes of induction, generalisation and so on.

Linguistic Nativism

It proceeds largely through processes that are specific to the domain of language.

General nativism

Uncontroversial

- Lobsters can't learn language
- Humans can
- Therefore: there is some part of our innate, genetically determined endowment that allows us to acquire and use language.

The “rocks and kittens” argument (Barbara Scholz)

Common points

All parties (except a few philosophers) accept that there are innate mechanisms.

The behaviorist is knowingly and cheerfully up to his neck in innate mechanisms of learning readiness.

W. Quine

Linguistic nativism

Working definition

(First) Language acquisition proceeds primarily through domain-specific mechanisms rather than through general-purpose learning mechanisms.

Linguistic nativism

Working definition

(First) Language acquisition proceeds primarily through domain-specific mechanisms rather than through general-purpose learning mechanisms.

- Distinction between domain-specific innate *knowledge* and domain-specific *processes* seems hard to sustain.
- Empirical debate not philosophical

Linguistic nativism

Working definition

(First) Language acquisition proceeds primarily through domain-specific mechanisms rather than through general-purpose learning mechanisms.

- Distinction between domain-specific innate *knowledge* and domain-specific *processes* seems hard to sustain.
- Empirical debate not philosophical
- Chomsky does not accept this definition: there are merely specific theories, rather than a general thesis.

Innateness

Philosophical analysis tends to muddy the waters. Originally a philosophical term predates discussion of genetics, but we will use it in a naive sense.

- Not learned from experience
- Genetically determined

Wexler

Chomsky's hypothesis is that many aspects of the formal structure of languages are encoded in the genome

Domain specific

- Spectrum from truly general purpose to domain specific.
- Parts of speech: clearly domain specific
- Sequence models: maths, music and other domains that might be parasitic on language.
- Hierarchical “tree” structures: occurs in nature. H. Simon’s observations on *The architecture of complexity*.
- David Papineau: innate ability to recognize animals.

Species specificity

Hauser

- Only humans know language.

Species specificity

Hauser

- Only humans know language.
- Therefore, if a cognitive ability is language specific, then it must be human specific.

Species specificity

Hauser

- Only humans know language.
- Therefore, if a cognitive ability is language specific, then it must be human specific.
- If we observe this ability in non human animals, then it is not language specific.

Universal Grammar

- Debate is about how much/what parts of language is innate.
- Important to avoid fallacy of equivocation in the term *UG*
 - UG can refer to the empirical claim that some non trivial parts of language are innate
 - UG can refer to a class of languages (that is learnable by the language faculty)
 - UG can refer to the initial state of the language faculty

Opposing views

- Chomskyan family of views:
 - Principles and Parameters
 - Minimalist grammars
- Non-Chomskyan views
- Connectionist models
 - Rumelhart and McClelland
 - Elman et al. *Rethinking Innateness*
- Usage based/Emergentist/Constructionist approaches:
 - Tomasello (2000)
 - Goldberg (2003)

Minimalist Program

slightly confused situation

Appear to reject strong nativism and parameter setting

Hauser, Chomsky and Fitch (2002)

Boeckx (2009/11)

Remain committed to a parametric view

Roeper and Yang (2011)

Confusion in the MP (Longa and Lorenzo, 2008)

Many arguments for LN

- Argument from Poverty of the stimulus
- Neurological localisation
- Double dissociation: Williams syndrome and SLI
- Universals of language/typological arguments
- Developmental evidence
- Critical period hypothesis
- Genetic evidence – SLI and FOXP2 gene.

Historically and conceptually the APS is the most fundamental.

The APS is the existence proof for the possibility of cognitive science. Fodor (1981)

Outline

Introduction

Linguistics

The key debate: Linguistic Nativism

The Argument from the Poverty of the Stimulus

Empirical APS

Formal versions of the APS

Conclusion

Many different names

- Plato's problem
- the logical problem of language acquisition
- The argument from the poverty of the stimulus
- the projection problem

The Argument from the Poverty of the Stimulus

APS

The general form of the APS is an argument from the premise that

1. The primary linguistic data (PLD) to which children have access is not sufficient to support the acquisition of adult linguistic competence through data driven learning procedures.

The Argument from the Poverty of the Stimulus

APS

The general form of the APS is an argument from the premise that

1. The primary linguistic data (PLD) to which children have access is not sufficient to support the acquisition of adult linguistic competence through data driven learning procedures.

to the conclusion that

2. Language acquisition requires a rich set of innate language specific learning constraints encoded in a Universal Grammar (UG).

The APS and Linguistic Nativism

- Using the APS to support nativist claims for innate domain specific knowledge goes back at least to Plato's argument in the *Meno* for *a priori* human knowledge of geometry.
- Linguistic nativism has been dominant in linguistics and cognitive science for the past fifty years.
- Chomsky (1965) gives an early statement of this view.

"A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of language can be learned by an organism initially uninformed as to its general character." (p. 57-58)

Empirical versions of the APS

General argument

Children learn facts for which there is no evidence in the input.

Specific arguments

Auxiliary fronting

One anaphora

English auxiliaries etc.

Auxiliary Inversion: A Case Study

- Chomsky (1971, 1975), Crain and Nakayama (1987), and Crain (1991) (among others) take auxiliary inversion in polar questions to be an instance of children acquiring a structure-dependent rule without having access to the relevant evidence in the PLD.

Auxiliary Inversion: A Case Study

- Chomsky (1971, 1975), Crain and Nakayama (1987), and Crain (1991) (among others) take auxiliary inversion in polar questions to be an instance of children acquiring a structure-dependent rule without having access to the relevant evidence in the PLD.
- Crain (1991) describes this as the "parade case of an innate constraint".

Auxiliary Inversion: A Case Study

- Chomsky (1971, 1975), Crain and Nakayama (1987), and Crain (1991) (among others) take auxiliary inversion in polar questions to be an instance of children acquiring a structure-dependent rule without having access to the relevant evidence in the PLD.
- Crain (1991) describes this as the "parade case of an innate constraint".
- It is possible to specify an operation for forming polar interrogatives in English by means of two distinct rules.

Two Possible Inversion Rules

- (1a) The student in the garden is hungry.
- (b) Is the student in the garden hungry?

Two Possible Inversion Rules

(1a) The student in the garden is hungry.

(b) Is the student in the garden hungry?

(2a) Front the first auxiliary in the string to the beginning of the main clause.

Two Possible Inversion Rules

- (1a) The student in the garden is hungry.
 - (b) Is the student in the garden hungry?
-
- (2a) Front the first auxiliary in the string to the beginning of the main clause.
 - (b) Front the auxiliary of the main VP to the beginning of the main clause.

Linear Order vs. Constituent Structure

(2a), a linear counting rule, is the simpler rule, but (2b), a constituent structure dependent operation, is the correct one.

Linear Order vs. Constituent Structure

(2a), a linear counting rule, is the simpler rule, but (2b), a constituent structure dependent operation, is the correct one.

- (3a) The student who is in the garden is hungry.
- (b) Is the student who is in the garden hungry?
- (c) *Is the student who in the garden is hungry?

Structure Dependence and the APS

Advocates of this instance of the APS make the following claims.

Structure Dependence and the APS

Advocates of this instance of the APS make the following claims.

- Considered as a hypothesis about the available data, (2b) is a less natural rule than (2a)

Structure Dependence and the APS

Advocates of this instance of the APS make the following claims.

- Considered as a hypothesis about the available data, (2b) is a less natural rule than (2a)
- Children do not generally make errors like (3c).

Structure Dependence and the APS

Advocates of this instance of the APS make the following claims.

- Considered as a hypothesis about the available data, (2b) is a less natural rule than (2a)
- Children do not generally make errors like (3c).
- Sentences like *Is the student who is in the garden hungry?* (3b) do not occur (or do not occur frequently) in the linguistic data available for language acquisition.

Empirical debate

About frequency of the key examples:

- Chomsky
- Pullum and Scholz (2002)
- Legate and Yang (2002)

CHILDES

1. Where's the little blue crib that was in the house before?
2. Where's the other dolly that was in here?
3. Where's the other doll that goes in there?

But what are the key sentences exactly?

Problem

What is the relevant class of constructions?

- Impossible to decide how large the class should be without a learning theory.
- Target sentence: *Is the student who is in the garden hungry?*

Problem

What is the relevant class of constructions?

- Impossible to decide how large the class should be without a learning theory.
- Target sentence: *Is the student who is in the garden hungry?*
- Possible data:
 - *Is the student who is in the garden hungry?*

Problem

What is the relevant class of constructions?

- Impossible to decide how large the class should be without a learning theory.
- Target sentence: *Is the student who is in the garden hungry?*
- Possible data:
 - *Is the student who is in the garden hungry?*
 - *Is the professor who is in the house hungry?*

Problem

What is the relevant class of constructions?

- Impossible to decide how large the class should be without a learning theory.
- Target sentence: *Is the student who is in the garden hungry?*
- Possible data:
 - *Is the student who is in the garden hungry?*
 - *Is the professor who is in the house hungry?*
 - *Did the man over there give me a biscuit?*
 - Any sentences that support hierarchical structure.

A general problem with this class of argument.

A priori APS

Other versions of the APS rely on formal general arguments that derive from the formal theory of learnability

The strongest most central arguments for innateness thus continue to be the arguments from APS and learnability theory. . . . The basic results of the field include the formal, mathematical demonstration that without serious constraints on the nature of human grammar, no possible learning mechanism can in fact learn the class of human grammars. Wexler, MIT Encyclopedia of the Cognitive Sciences.

Taking these arguments seriously

- First, try to formulate them precisely.

Taking these arguments seriously

- First, try to formulate them precisely.
- There are serious learnability problems as we shall see
- The arguments tend to have a number of technical flaws
- Even if we fix them, the arguments do not support linguistic nativism over general nativism.
- We still need to take the challenge seriously: how can we learn in an “empiricist” way?
- The APS is in a sense the right question.

Outline

Introduction

Linguistics

The key debate: Linguistic Nativism

The Argument from the Poverty of the Stimulus

Empirical APS

Formal versions of the APS

Conclusion

Tension

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.

Tension

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.

Principles and Parameters

No tension.

Unsupervised Learning

Fundamental problem of linguistics

Chomsky's questions

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

Pinker (1990)

To understand how X is learned, you first have to understand what X is.

Crain and Pietroski (2001)

First, one tries to find principles that characterize human grammars; *then* one tries to determine which aspects of these grammars could plausibly be learned from experience, and which are more likely to be innately specified.

The subject matter of linguistics



Standard methodology

- Step 1: Construct a descriptively adequate representation
- Step 2: Try to design learning algorithms for those representations

Step 1

Construct a descriptively adequate grammar

This failed

- No-one ever managed to make a descriptively adequate grammar for any natural language, not even English.
- In order to account for new facts (e.g. Swiss German) representations were made more powerful and expressive.
- Statistical parsers do not separate grammatical from ungrammatical sentences (Okanohara and Tsujii, 2007; Berwick and Fong, 2008)
- Generative grammarians have largely abandoned the task of constructing large scale grammars.

Step 1

Construct a descriptively adequate grammar

This failed

- No-one ever managed to make a descriptively adequate grammar for any natural language, not even English.
- In order to account for new facts (e.g. Swiss German) representations were made more powerful and expressive.
- Statistical parsers do not separate grammatical from ungrammatical sentences (Okanohara and Tsujii, 2007; Berwick and Fong, 2008)
- Generative grammarians have largely abandoned the task of constructing large scale grammars.

Why is it so hard?

Step 2

Come up with a learning algorithm

This also failed.

- Learning even regular grammars is computationally hard: Angluin and Kharitonov (1995)
- We have some heuristic algorithms that can induce crude constituent structure (Klein and Manning, 2004)
- The classes of representations we need have even richer, deeper and more abstract hidden structure: (LTAG, $ACG_{2,4}$, ...)
- It is out of the question to construct learning algorithms for these classes.

PSGs were meant to be learnable

Chomsky (1968/2006)

“The concept of “phrase structure grammar” was explicitly designed to express the richest system that could reasonable be expected to result from the application of Harris-type procedures to a corpus.”

Linguists don't know what the representations are

A Cambridge quote

“At the most fundamental level, it is not clear that there is any meaningful empirical motivation for the representational assumptions of any current formal model of syntax.”

(Blevins, J., 2009)

Linguists cannot agree whether the head of “the cat” is “the” or “cat”. Nor can they produce any empirical evidence to decide between the two.

(Matthews, P.; 2007)

Linguists don't know what the representations are

A Cambridge quote

“At the most fundamental level, it is not clear that there is any meaningful empirical motivation for the representational assumptions of any current formal model of syntax.”

(Blevins, J., 2009)

Linguists cannot agree whether the head of “the cat” is “the” or “cat”. Nor can they produce any empirical evidence to decide between the two.

(Matthews, P.; 2007)

- We don't know what the representations are but we do know that they are learnable!

Reasonable Research Strategy

Slogan

Put learnability first!

- If you construct a super-powerful class of languages with no thought of learnability, you won't be able to learn them.
- Rather, design representations from the ground up to be learnable.

Reasonable Research Strategy

Slogan

Put learnability first!

- If you construct a super-powerful class of languages with no thought of learnability, you won't be able to learn them.
- Rather, design representations from the ground up to be learnable.

Strategy

- Step 1: build simple learnable representations
- Step 2: gradually try to increase their expressive power, while maintaining learnability

Reverse direction

Normal direction

Function from representation to language

Context free grammar $G \rightarrow$ context free language $L(G)$

Non-terminal \rightarrow set of strings derived from non-terminal

Reverse direction

Normal direction

Function from representation to language

Context free grammar $G \rightarrow$ context free language $L(G)$

Non-terminal \rightarrow set of strings derived from non-terminal

Opposite Direction

Function from language to representation

$L \rightarrow R(L)$

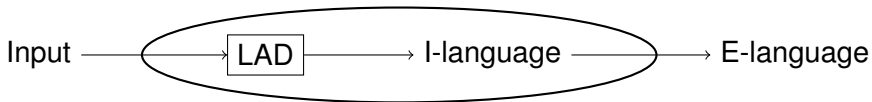
From set of strings \rightarrow representational primitive of formalism

Ideally $L(R(L)) = L$.

The subject matter of linguistics



The subject matter of linguistics



Empiricism and Linguistic Empiricism

Empiricism

The philosophical doctrine that knowledge has its origin in experience.

Linguistic Empiricism

The empiricam claim that linguistic knowledge has its origin in linguistic experience.

Empiricist models

Slogan

The structure of the representation should be based on the structure of the language, not something arbitrarily imposed on it from outside.

- Identify some structure in the language
- Show how that structure can be observed
- Construct a representation based on that structure
- Richer structures will give you more powerful representations