Course Overview

CSC2540S Machine Learning and Universal Grammar
Department of Computer Science, University of Toronto

Shalom Lappin*
Department of Philosophy
King's College, London

*Joint work with Alex Clark
Department of Computer Science
Royal Holloway College, London

January 6, 2009

Outline

Topics and Schedule

Week 1: January 6, 2009

- Overview of the problem
- The argument from the poverty of stimulus (APS) and Universal Grammar

Week 2: January 13

- Nativism and cognitive architecture
- The evolution of language

Topics and Schedule

Week 1: January 6, 2009

- Overview of the problem
- The argument from the poverty of stimulus (APS) and Universal Grammar

Week 2: January 13

- Nativism and cognitive architecture
- The evolution of language

Topics and Schedule

Week 3: January 20

- Clarifying the APS
- Some linguistic arguments for the APS

Week 4: January 27

- The nature of primary linguistic evidence (PLD)
- The role of negative evidence in language acquisition

Topics and Schedule

Week 3: January 20

- Clarifying the APS
- Some linguistic arguments for the APS

Week 4: January 27

- The nature of primary linguistic evidence (PLD)
- The role of negative evidence in language acquisition

Topics and Schedule

Week 5: February 3

- Identification in the Limit: The Gold paradigm of grammar induction

Week 6: February 10

- An alternative computational learning model: PAC learning

Topics and Schedule

Week 5: February 3

- Identification in the Limit: The Gold paradigm of grammar induction

Week 6: February 10

- An alternative computational learning model: PAC learning

Topics and Schedule

Week 7: February 17

- Reading Week

Week 8: February 24

- Modifying PAC: Using probability distributions on learning samples

Topics and Schedule

Week 7: February 17

- Reading Week

Week 8: February 24

- Modifying PAC: Using probability distributions on learning samples

Topics and Schedule

Week 9: March 3

- Unsupervised grammar induction
- Recent work on acquiring grammars through the application of machine learning methods to large natural language corpora

Week 10: March 10

- Parameters in linguistic theory and in probabilistic language models
- A bootstrapping approach to parameter estimation for language models

Topics and Schedule

Week 9: March 3

- Unsupervised grammar induction
- Recent work on acquiring grammars through the application of machine learning methods to large natural language corpora

Week 10: March 10

- Parameters in linguistic theory and in probabilistic language models
- A bootstrapping approach to parameter estimation for language models

Topics and Schedule

Week 11: March 17

- The FOXP2 Gene and specific language impairment (SLI)
- Psychological evidence for Bayesian grammar induction

Weeks 12-14: March 24, March 31, and April 7

- Project presentations

Topics and Schedule

Week 11: March 17

- The FOXP2 Gene and specific language impairment (SLI)
- Psychological evidence for Bayesian grammar induction

Weeks 12-14: March 24, March 31, and April 7

- Project presentations

The Course Grade

- Course assessment depends upon
  - (a) a project, which will provide 80% of the final grade, and
  - (b) class participation, which counts for 20% of the grade.
- For the project
  - (a) you can write a term paper, or
  - (b) you can build an implemented system.
- Class participation is evaluated on the basis of
  - (a) a project presentation.

The Course Grade

- Course assessment depends upon
  - (a) a project, which will provide 80% of the final grade, and
  - (b) class participation, which counts for 20% of the grade.
- For the project
  - (a) you can write a term paper, or
  - (b) you can build an implemented system.
- Class participation is evaluated on the basis of
  - (a) a project presentation.

The Course Grade

- Course assessment depends upon
  (a) a project, which will provide 80% of the final grade, and
  (b) class participation, which counts for 20% of the grade.
- For the project
  (a) you can write a term paper, or
  (b) you can build an implemented system.
- Class participation is evaluated on the basis of
  (a) a project presentation.

Term Paper

- A term paper should be 5000-8000 words in length.
- It must provide an in depth critical study of one or more of the issues taken up in the course.
- The paper should discuss salient parts of the literature relevant to the issues it addresses.
- It need not offer a new proposal.
- However, a paper defending an original hypothesis with interesting arguments is, of course, welcome.

Term Paper

- A term paper should be 5000-8000 words in length.
- It must provide an in depth critical study of one or more of the issues taken up in the course.
- The paper should discuss salient parts of the literature relevant to the issues it addresses.
- It need not offer a new proposal.
- However, a paper defending an original hypothesis with interesting arguments is, of course, welcome.

Term Paper

- A term paper should be 5000-8000 words in length.
- It must provide an in depth critical study of one or more of the issues taken up in the course.
- The paper should discuss salient parts of the literature relevant to the issues it addresses.
- It need not offer a new proposal.
- However, a paper defending an original hypothesis with interesting arguments is, of course, welcome.

Term Paper

- A term paper should be 5000-8000 words in length.
- It must provide an in depth critical study of one or more of the issues taken up in the course.
- The paper should discuss salient parts of the literature relevant to the issues it addresses.
- It need not offer a new proposal.
- However, a paper defending an original hypothesis with interesting arguments is, of course, welcome.

Term Paper

- A term paper should be 5000-8000 words in length.
- It must provide an in depth critical study of one or more of the issues taken up in the course.
- The paper should discuss salient parts of the literature relevant to the issues it addresses.
- It need not offer a new proposal.
- However, a paper defending an original hypothesis with interesting arguments is, of course, welcome.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Progamming Project

- An implemented system must be relevant to the course.
- It should illustrate some of the computational learning questions that we discuss.
- The program can be a small scale prototype learning algorithm rather than a wide coverage corpus system.
- It must be an original piece of work developed for this course.
- A system demo will be given in class.
- The code must be properly commented, and it will be accompanied by a brief project report that clearly explains the system.

Project Dates

- February 6: Project proposals are due (by email).
- March 13: Brief progress reports are due (by email)
- March 24, March 31, and April 7: These classes are devoted to presentations of your projects.
- April 20: Final version of the project is due.

Project Dates

- February 6: Project proposals are due (by email).
- March 13: Brief progress reports are due (by email)
- March 24, March 31, and April 7: These classes are devoted to presentations of your projects.
- April 20: Final version of the project is due.

Project Dates

- February 6: Project proposals are due (by email).
- March 13: Brief progress reports are due (by email)
- March 24, March 31, and April 7: These classes are devoted to presentations of your projects.
- April 20: Final version of the project is due.

Project Dates

- February 6: Project proposals are due (by email).
- March 13: Brief progress reports are due (by email)
- March 24, March 31, and April 7: These classes are devoted to presentations of your projects.
- April 20: Final version of the project is due.

Domain General and Domain Specific Learning Mechanisms

- The APS has been invoked as the primary motivation for a strong version of linguistic nativism.
- It posits a distinct language faculty with innate learning biases for language acquisition.
- This type of cognitive nativism is distinct from the claim that human cognitive activity depends on a rich set of innate learning priors.
- The former posits strong domain specific learning biases, while the latter allows for the possibility that most (all?) types of learning are achieved through domain general procedures, initialized for specific functions.

Domain General and Domain Specific Learning Mechanisms

- The APS has been invoked as the primary motivation for a strong version of linguistic nativism.
- It posits a distinct language faculty with innate learning biases for language acquisition.
- This type of cognitive nativism is distinct from the claim that human cognitive activity depends on a rich set of innate learning priors.
- The former posits strong domain specific learning biases, while the latter allows for the possibility that most (all?) types of learning are achieved through domain general procedures, initialized for specific functions.

Domain General and Domain Specific Learning Mechanisms

- The APS has been invoked as the primary motivation for a strong version of linguistic nativism.
- It posits a distinct language faculty with innate learning biases for language acquisition.
- This type of cognitive nativism is distinct from the claim that human cognitive activity depends on a rich set of innate learning priors.
- The former posits strong domain specific learning biases, while the latter allows for the possibility that most (all?) types of learning are achieved through domain general procedures, initialized for specific functions.

Domain General and Domain Specific Learning Mechanisms

- The APS has been invoked as the primary motivation for a strong version of linguistic nativism.
- It posits a distinct language faculty with innate learning biases for language acquisition.
- This type of cognitive nativism is distinct from the claim that human cognitive activity depends on a rich set of innate learning priors.
- The former posits strong domain specific learning biases, while the latter allows for the possibility that most (all?) types of learning are achieved through domain general procedures, initialized for specific functions.

Chomsky on Statistical Modeling of Grammar

- Chomsky (1957) rejects the use of statistical methods to represent the distinction between grammatical and ungrammatical strings.
  1. Colourless green ideas sleep furiously.
  2. Furiously sleep ideas green colourless.
- (1) and (2) both have a probability approaching nil (in 1957) of appearing in a corpus or actual speech.
- (1) is syntactically well formed, even if semantically anomalous, while (2) is not.

Chomsky on Statistical Modeling of Grammar

- Chomsky (1957) rejects the use of statistical methods to represent the distinction between grammatical and ungrammatical strings.
  1. Colourless green ideas sleep furiously.
  2. Furiously sleep ideas green colourless.
- (1) and (2) both have a probability approaching nil (in 1957) of appearing in a corpus or actual speech.
- (1) is syntactically well formed, even if semantically anomalous, while (2) is not.

Chomsky on Statistical Modeling of Grammar

- Chomsky (1957) rejects the use of statistical methods to represent the distinction between grammatical and ungrammatical strings.
  1. Colourless green ideas sleep furiously.
  2. Furiously sleep ideas green colourless.
- (1) and (2) both have a probability approaching nil (in 1957) of appearing in a corpus or actual speech.
- (1) is syntactically well formed, even if semantically anomalous, while (2) is not.

Chomsky on Statistical Modeling of Grammar

- Chomsky (1957) rejects the use of statistical methods to represent the distinction between grammatical and ungrammatical strings.
  1. Colourless green ideas sleep furiously.
  2. Furiously sleep ideas green colourless.
- (1) and (2) both have a probability approaching nil (in 1957) of appearing in a corpus or actual speech.
- (1) is syntactically well formed, even if semantically anomalous, while (2) is not.

Chomsky on Statistical Modeling of Grammar

Chomsky (1957) (p. 17)

If we rank the sequences of a given length in order of statistical approximation to English, we will find both grammatical and ungrammatical sequences scattered throughout the list; there appears to be no particular relation between order of approximation and grammaticalness. Despite the undeniable interest and importance of semantic and statistical studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances. I believe that we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure.

A Smoothed Bigram Model

- Chomsky moves from the claim that information theoretic methods cannot identify the set of grammatical sentences in the PLD to the conclusion that they are irrelevant to characterizing syntactic structure.
- This argument is not sound.
- Chomsky assumes a bigram model in which probability of a word in a string depends on the word that immediately precedes it.
- Pereira (2000) constructs a smoothed bigram model in which the probability of a word depends on the class of the prior word.

A Smoothed Bigram Model

- Chomsky moves from the claim that information theoretic methods cannot identify the set of grammatical sentences in the PLD to the conclusion that they are irrelevant to characterizing syntactic structure.
- This argument is not sound.
- Chomsky assumes a bigram model in which probability of a word in a string depends on the word that immediately precedes it.
- Pereira (2000) constructs a smoothed bigram model in which the probability of a word depends on the class of the prior word.

A Smoothed Bigram Model

- Chomsky moves from the claim that information theoretic methods cannot identify the set of grammatical sentences in the PLD to the conclusion that they are irrelevant to characterizing syntactic structure.
- This argument is not sound.
- Chomsky assumes a bigram model in which probability of a word in a string depends on the word that immediately precedes it.
- Pereira (2000) constructs a smoothed bigram model in which the probability of a word depends on the class of the prior word.

A Smoothed Bigram Model

- Chomsky moves from the claim that information theoretic methods cannot identify the set of grammatical sentences in the PLD to the conclusion that they are irrelevant to characterizing syntactic structure.
- This argument is not sound.
- Chomsky assumes a bigram model in which probability of a word in a string depends on the word that immediately precedes it.
- Pereira (2000) constructs a smoothed bigram model in which the probability of a word depends on the class of the prior word.

A Smoothed Bigram Model

- Pereira's model computes the conditional probability of a word w_i in a string with the formula

$$P(w_i \mid w_{i-1}) \approx \sum_c P(w_i \mid c)P(c \mid w_{i-1})$$

  where c is the class of w_{i-1}.
- We can use distributional patterns of words in a corpus to learn their classes from training data.
- Other procedures allow us to compute the values of the parameters $P(w_i \mid c)$ and $P(c \mid w_{i-1})$ from this data.
- When applied to (1) and (2), this model yields a five order of magnitude difference between their probability values for a corpus of newspaper text.

A Smoothed Bigram Model

- Pereira's model computes the conditional probability of a word w_i in a string with the formula

$$P(w_i \mid w_{i-1}) \approx \sum_{c} P(w_i \mid c) P(c \mid w_{i-1})$$

  where c is the class of w_{i-1}.
- We can use distributional patterns of words in a corpus to learn their classes from training data.
- Other procedures allow us to compute the values of the parameters $P(w_i \mid c)$ and $P(c \mid w_{i-1})$ from this data.
- When applied to (1) and (2), this model yields a five order of magnitude difference between their probability values for a corpus of newspaper text.

A Smoothed Bigram Model

- Pereira's model computes the conditional probability of a word w_i in a string with the formula

$$P(w_i \mid w_{i-1}) \approx \sum_{c} P(w_i \mid c) P(c \mid w_{i-1})$$

where c is the class of w_{i-1}.
- We can use distributional patterns of words in a corpus to learn their classes from training data.
- Other procedures allow us to compute the values of the parameters $P(w_i \mid c)$ and $P(c \mid w_{i-1})$ from this data.
- When applied to (1) and (2), this model yields a five order of magnitude difference between their probability values for a corpus of newspaper text.

A Smoothed Bigram Model

- Pereira's model computes the conditional probability of a word w_i in a string with the formula

$$P(w_i \mid w_{i-1}) \approx \sum_{c} P(w_i \mid c) P(c \mid w_{i-1})$$

where c is the class of w_{i-1}.
- We can use distributional patterns of words in a corpus to learn their classes from training data.
- Other procedures allow us to compute the values of the parameters $P(w_i \mid c)$ and $P(c \mid w_{i-1})$ from this data.
- When applied to (1) and (2), this model yields a five order of magnitude difference between their probability values for a corpus of newspaper text.

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)

UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)

UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)

UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)

UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)

UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)

UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)

UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)

UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Elements of UG: Chomsky 1965

- UG(1) An enumeration of the class $s_1, s_2, \ldots$ of possible sentences (**set of strings that each grammar generates**)
- UG(2) An enumeration of the class $SD_1, SD_2, \ldots$ of possible structural descriptions (**the set of syntactic representations that these grammars assign to the strings that they produce**)
- UG(3) An enumeration of the class $G_1, G_2, \ldots$ of possible generative grammars (**the hypothesis space of possible grammars for natural languages**)
- UG(4) Specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j, for arbitrary i, j (**the function that maps a grammar to the set of representations for a string**)
- UG(5) Specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number) (**an evaluation measure that ranks the possible grammars**)

Descriptive and Explanatory Adequacy

- Chomsky distinguishes between the descriptive and explanatory adequacy of a theory of grammar.
- A descriptively adequate theory generates and assigns syntactic representations to the sentences of a language in a way that captures their observed structural properties.
- An explanatorily adequate theory encodes the function that children apply to select a single grammar from among several incompatible grammars, all of which are descriptively adequate for the PLD.
- This notion of explanatory adequacy requires UG to account for central aspects of language acquisition.

Descriptive and Explanatory Adequacy

- Chomsky distinguishes between the descriptive and explanatory adequacy of a theory of grammar.
- A descriptively adequate theory generates and assigns syntactic representations to the sentences of a language in a way that captures their observed structural properties.
- An explanatorily adequate theory encodes the function that children apply to select a single grammar from among several incompatible grammars, all of which are descriptively adequate for the PLD.
- This notion of explanatory adequacy requires UG to account for central aspects of language acquisition.

Descriptive and Explanatory Adequacy

- Chomsky distinguishes between the descriptive and explanatory adequacy of a theory of grammar.
- A descriptively adequate theory generates and assigns syntactic representations to the sentences of a language in a way that captures their observed structural properties.
- An explanatorily adequate theory encodes the function that children apply to select a single grammar from among several incompatible grammars, all of which are descriptively adequate for the PLD.
- This notion of explanatory adequacy requires UG to account for central aspects of language acquisition.

Descriptive and Explanatory Adequacy

- Chomsky distinguishes between the descriptive and explanatory adequacy of a theory of grammar.
- A descriptively adequate theory generates and assigns syntactic representations to the sentences of a language in a way that captures their observed structural properties.
- An explanatorily adequate theory encodes the function that children apply to select a single grammar from among several incompatible grammars, all of which are descriptively adequate for the PLD.
- This notion of explanatory adequacy requires UG to account for central aspects of language acquisition.

Problems with the Evaluation Measure

- The evaluation measure is required in order to resolve conflicts among alternative grammars which are compatible with the PLD.
- However, it is not clear how it can be formally specified, or what sort of evidence should be invoked to motivate an account of its design.
- By assumption, it ranks grammars that enjoy the same degree of descriptive adequacy, and so it is not accessible to PLD.

Problems with the Evaluation Measure

- The evaluation measure is required in order to resolve conflicts among alternative grammars which are compatible with the PLD.
- However, it is not clear how it can be formally specified, or what sort of evidence should be invoked to motivate an account of its design.
- By assumption, it ranks grammars that enjoy the same degree of descriptive adequacy, and so it is not accessible to PLD.

Problems with the Evaluation Measure

- The evaluation measure is required in order to resolve conflicts among alternative grammars which are compatible with the PLD.
- However, it is not clear how it can be formally specified, or what sort of evidence should be invoked to motivate an account of its design.
- By assumption, it ranks grammars that enjoy the same degree of descriptive adequacy, and so it is not accessible to PLD.

Problems with the Evaluation Measure

- Notions of formal simplicity of the sort used to choose among rival scientific theories do not offer an appropriate grammar ranking procedure because
  - (i) they are notoriously difficult to formulate as global metrics that are both precise and consistent, and
  - (ii) if one could define a workable simplicity measure of this kind, then it would not be part of a domain specific UG.
- Chomsky's (1965) evaluation measure is a domain specific simplicity measure internal to UG.
- But it is a mechanism whose description is inaccessible to the empirical assessment that Chomsky insists is the only basis for understanding its design.

Problems with the Evaluation Measure

- Notions of formal simplicity of the sort used to choose among rival scientific theories do not offer an appropriate grammar ranking procedure because
  - (i) they are notoriously difficult to formulate as global metrics that are both precise and consistent, and
  - (ii) if one could define a workable simplicity measure of this kind, then it would not be part of a domain specific UG.
- Chomsky's (1965) evaluation measure is a domain specific simplicity measure internal to UG.
- But it is a mechanism whose description is inaccessible to the empirical assessment that Chomsky insists is the only basis for understanding its design.

Problems with the Evaluation Measure

- Notions of formal simplicity of the sort used to choose among rival scientific theories do not offer an appropriate grammar ranking procedure because
  - (i) they are notoriously difficult to formulate as global metrics that are both precise and consistent, and
  - (ii) if one could define a workable simplicity measure of this kind, then it would not be part of a domain specific UG.
- Chomsky's (1965) evaluation measure is a domain specific simplicity measure internal to UG.
- But it is a mechanism whose description is inaccessible to the empirical assessment that Chomsky insists is the only basis for understanding its design.

Problems with the Evaluation Measure

- Notions of formal simplicity of the sort used to choose among rival scientific theories do not offer an appropriate grammar ranking procedure because
  - (i) they are notoriously difficult to formulate as global metrics that are both precise and consistent, and
  - (ii) if one could define a workable simplicity measure of this kind, then it would not be part of a domain specific UG.
- Chomsky's (1965) evaluation measure is a domain specific simplicity measure internal to UG.
- But it is a mechanism whose description is inaccessible to the empirical assessment that Chomsky insists is the only basis for understanding its design.

Government Binding (GB) and the Principles and Parameters (P&P) Model of UG

- Chomsky (1981): UG consists of schematic constraints on the representations in a syntactic derivation of a sentence, and on the movement operation which maps between adjacent levels in the derivation.
- These constraints include parameters that allow for a finite number of possible values (ideally binary).
- Assigning values to all the parameters of UG yields a particular grammar.

Government Binding (GB) and the Principles and Parameters (P&P) Model of UG

- Chomsky (1981): UG consists of schematic constraints on the representations in a syntactic derivation of a sentence, and on the movement operation which maps between adjacent levels in the derivation.
- These constraints include parameters that allow for a finite number of possible values (ideally binary).
- Assigning values to all the parameters of UG yields a particular grammar.

Government Binding (GB) and the Principles and Parameters (P&P) Model of UG

- Chomsky (1981): UG consists of schematic constraints on the representations in a syntactic derivation of a sentence, and on the movement operation which maps between adjacent levels in the derivation.
- These constraints include parameters that allow for a finite number of possible values (ideally binary).
- Assigning values to all the parameters of UG yields a particular grammar.

Language Acquisition as Parameter Setting

- Language acquisition consists in setting parameter values through exposure to a small amount of data from a language.
- UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values.
- Therefore it defines a finite set of possible (core) grammars for natural language.
- Identifying values for the parameterized constraints of UG yields a unique grammar for the string set of a language, on the basis of the PLD.
- The grammar evaluation measure of the *Aspects* model is no longer needed.

Language Acquisition as Parameter Setting

- Language acquisition consists in setting parameter values through exposure to a small amount of data from a language.
- UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values.
- Therefore it defines a finite set of possible (core) grammars for natural language.
- Identifying values for the parameterized constraints of UG yields a unique grammar for the string set of a language, on the basis of the PLD.
- The grammar evaluation measure of the *Aspects* model is no longer needed.

Language Acquisition as Parameter Setting

- Language acquisition consists in setting parameter values through exposure to a small amount of data from a language.
- UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values.
- Therefore it defines a finite set of possible (core) grammars for natural language.
- Identifying values for the parameterized constraints of UG yields a unique grammar for the string set of a language, on the basis of the PLD.
- The grammar evaluation measure of the *Aspects* model is no longer needed.

Language Acquisition as Parameter Setting

- Language acquisition consists in setting parameter values through exposure to a small amount of data from a language.
- UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values.
- Therefore it defines a finite set of possible (core) grammars for natural language.
- Identifying values for the parameterized constraints of UG yields a unique grammar for the string set of a language, on the basis of the PLD.
- The grammar evaluation measure of the *Aspects* model is no longer needed.

Language Acquisition as Parameter Setting

- Language acquisition consists in setting parameter values through exposure to a small amount of data from a language.
- UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values.
- Therefore it defines a finite set of possible (core) grammars for natural language.
- Identifying values for the parameterized constraints of UG yields a unique grammar for the string set of a language, on the basis of the PLD.
- The grammar evaluation measure of the *Aspects* model is no longer needed.

Finite Hypothesis Spaces and the Tractability of Learning

- Chomsky (1981): the finiteness property of the P&P model of UG "trivializes" important aspects of grammar induction.
- In fact, a finite hypothesis space is neither a necessary nor a sufficient condition for efficient learning.
- Grammar induction in a finite hypothesis space can be intractable.
- Efficient learning in certain infinite spaces is possible.
- The complexity of learning (both data samples and time needed for acquisition) is crucial to the learnability of a language class, even when the string sets are decidable.

Finite Hypothesis Spaces and the Tractability of Learning

- Chomsky (1981): the finiteness property of the P&P model of UG "trivializes" important aspects of grammar induction.
- In fact, a finite hypothesis space is neither a necessary nor a sufficient condition for efficient learning.
- Grammar induction in a finite hypothesis space can be intractable.
- Efficient learning in certain infinite spaces is possible.
- The complexity of learning (both data samples and time needed for acquisition) is crucial to the learnability of a language class, even when the string sets are decidable.

Finite Hypothesis Spaces and the Tractability of Learning

- Chomsky (1981): the finiteness property of the P&P model of UG "trivializes" important aspects of grammar induction.
- In fact, a finite hypothesis space is neither a necessary nor a sufficient condition for efficient learning.
- Grammar induction in a finite hypothesis space can be intractable.
- Efficient learning in certain infinite spaces is possible.
- The complexity of learning (both data samples and time needed for acquisition) is crucial to the learnability of a language class, even when the string sets are decidable.

Finite Hypothesis Spaces and the Tractability of Learning

- Chomsky (1981): the finiteness property of the P&P model of UG "trivializes" important aspects of grammar induction.
- In fact, a finite hypothesis space is neither a necessary nor a sufficient condition for efficient learning.
- Grammar induction in a finite hypothesis space can be intractable.
- Efficient learning in certain infinite spaces is possible.
- The complexity of learning (both data samples and time needed for acquisition) is crucial to the learnability of a language class, even when the string sets are decidable.

Finite Hypothesis Spaces and the Tractability of Learning

- Chomsky (1981): the finiteness property of the P&P model of UG "trivializes" important aspects of grammar induction.
- In fact, a finite hypothesis space is neither a necessary nor a sufficient condition for efficient learning.
- Grammar induction in a finite hypothesis space can be intractable.
- Efficient learning in certain infinite spaces is possible.
- The complexity of learning (both data samples and time needed for acquisition) is crucial to the learnability of a language class, even when the string sets are decidable.

UG in the Minimalist Program

- The Minimalist Program (MP) (Chomsky 1995, 2001, 2005, 2007) eliminates intermediate derivational levels of representation (D- and S-Structure in the GB model).
- Only the two interface levels of LF (Logical Form, the Conceptual-Intentional interface) and PF (Phonetic Form, the Sensory Motor interface) remain, as the outputs of a syntactic derivation from a selection of lexical items (a numeration).
- A single operation, Merge, combines lexically projected functional heads with their complements and their adjuncts to produce a hierarchical tree structure.

UG in the Minimalist Program

- The Minimalist Program (MP) (Chomsky 1995, 2001, 2005, 2007) eliminates intermediate derivational levels of representation (D- and S-Structure in the GB model).
- Only the two interface levels of LF (Logical Form, the Conceptual-Intentional interface) and PF (Phonetic Form, the Sensory Motor interface) remain, as the outputs of a syntactic derivation from a selection of lexical items (a numeration).
- A single operation, Merge, combines lexically projected functional heads with their complements and their adjuncts to produce a hierarchical tree structure.

UG in the Minimalist Program

- The Minimalist Program (MP) (Chomsky 1995, 2001, 2005, 2007) eliminates intermediate derivational levels of representation (D- and S-Structure in the GB model).
- Only the two interface levels of LF (Logical Form, the Conceptual-Intentional interface) and PF (Phonetic Form, the Sensory Motor interface) remain, as the outputs of a syntactic derivation from a selection of lexical items (a numeration).
- A single operation, Merge, combines lexically projected functional heads with their complements and their adjuncts to produce a hierarchical tree structure.

Syntactic Derivations and Interface Representations

- Syntactic movement is characterized as "internal merge", a procedure that copies sub-constituents of a tree in a specified position at the left or right edge of the containing phrase structure.
- Movement is triggered by the need to check features in the lexical head of a constituent against those in the target site.
- Uninterpretable features must be eliminated through checking prior to the construction of the interface levels, as they are "illegible" to the cognitive device that an interface feeds.
- Syntactic derivations are subject to locality constraints, which are stated as economy conditions.

Syntactic Derivations and Interface Representations

- Syntactic movement is characterized as "internal merge", a procedure that copies sub-constituents of a tree in a specified position at the left or right edge of the containing phrase structure.
- Movement is triggered by the need to check features in the lexical head of a constituent against those in the target site.
- Uninterpretable features must be eliminated through checking prior to the construction of the interface levels, as they are "illegible" to the cognitive device that an interface feeds.
- Syntactic derivations are subject to locality constraints, which are stated as economy conditions.

Syntactic Derivations and Interface Representations

- Syntactic movement is characterized as "internal merge", a procedure that copies sub-constituents of a tree in a specified position at the left or right edge of the containing phrase structure.
- Movement is triggered by the need to check features in the lexical head of a constituent against those in the target site.
- Uninterpretable features must be eliminated through checking prior to the construction of the interface levels, as they are "illegible" to the cognitive device that an interface feeds.
- Syntactic derivations are subject to locality constraints, which are stated as economy conditions.

Syntactic Derivations and Interface Representations

- Syntactic movement is characterized as "internal merge", a procedure that copies sub-constituents of a tree in a specified position at the left or right edge of the containing phrase structure.
- Movement is triggered by the need to check features in the lexical head of a constituent against those in the target site.
- Uninterpretable features must be eliminated through checking prior to the construction of the interface levels, as they are "illegible" to the cognitive device that an interface feeds.
- Syntactic derivations are subject to locality constraints, which are stated as economy conditions.

UG as a Perfect Computational System

- The guiding principle behind the MP is that UG is a "perfect" computational system.
- It provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF.
- Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.
- As a result, it is not clear what predictions they make concerning the formal properties of UG.
- It is not obvious how to test the comparative "perfection" of alternative theories of UG within the MP.

UG as a Perfect Computational System

- The guiding principle behind the MP is that UG is a "perfect" computational system.
- It provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF.
- Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.
- As a result, it is not clear what predictions they make concerning the formal properties of UG.
- It is not obvious how to test the comparative "perfection" of alternative theories of UG within the MP.

UG as a Perfect Computational System

- The guiding principle behind the MP is that UG is a "perfect" computational system.
- It provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF.
- Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.
- As a result, it is not clear what predictions they make concerning the formal properties of UG.
- It is not obvious how to test the comparative "perfection" of alternative theories of UG within the MP.

UG as a Perfect Computational System

- The guiding principle behind the MP is that UG is a "perfect" computational system.
- It provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF.
- Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.
- As a result, it is not clear what predictions they make concerning the formal properties of UG.
- It is not obvious how to test the comparative "perfection" of alternative theories of UG within the MP.

UG as a Perfect Computational System

- The guiding principle behind the MP is that UG is a "perfect" computational system.
- It provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF.
- Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.
- As a result, it is not clear what predictions they make concerning the formal properties of UG.
- It is not obvious how to test the comparative "perfection" of alternative theories of UG within the MP.

Parameters in the MP

- In the GB model parameters are located in general constraints on levels of representation and on movement.
- In the MP they have been moved to the functional heads of the lexicon.
- In fact Boeckx (2008) proposes eliminating parameters from "narrow syntax" entirely.
- He situates them at the PF interface as part of the "Spell Out" of representations.
- LF representations are uniform across languages, and so their realization is not subject to parameteric variation.

Parameters in the MP

- In the GB model parameters are located in general constraints on levels of representation and on movement.
- In the MP they have been moved to the functional heads of the lexicon.
- In fact Boeckx (2008) proposes eliminating parameters from "narrow syntax" entirely.
- He situates them at the PF interface as part of the "Spell Out" of representations.
- LF representations are uniform across languages, and so their realization is not subject to parameteric variation.

Parameters in the MP

- In the GB model parameters are located in general constraints on levels of representation and on movement.
- In the MP they have been moved to the functional heads of the lexicon.
- In fact Boeckx (2008) proposes eliminating parameters from "narrow syntax" entirely.
- He situates them at the PF interface as part of the "Spell Out" of representations.
- LF representations are uniform across languages, and so their realization is not subject to parameteric variation.

Parameters in the MP

- In the GB model parameters are located in general constraints on levels of representation and on movement.
- In the MP they have been moved to the functional heads of the lexicon.
- In fact Boeckx (2008) proposes eliminating parameters from "narrow syntax" entirely.
- He situates them at the PF interface as part of the "Spell Out" of representations.
- LF representations are uniform across languages, and so their realization is not subject to parameteric variation.

Parameters in the MP

- In the GB model parameters are located in general constraints on levels of representation and on movement.
- In the MP they have been moved to the functional heads of the lexicon.
- In fact Boeckx (2008) proposes eliminating parameters from "narrow syntax" entirely.
- He situates them at the PF interface as part of the "Spell Out" of representations.
- LF representations are uniform across languages, and so their realization is not subject to parameteric variation.

UG as Recursion

- Hauser et. al (2002) reduce the "narrow" language faculty (FLN) to recursion, which allows for operations that generate unbounded sets of expressions and hierarchical syntactic structures.
- Recursion may have first emerged in other cognitive domains (such as computation with numbers), and then been adapted to language.
- Pinker and Jackendoff (2005) observe that recursion is pervasive across various primate and human cognitive domains (navigation, recognition of kinship structures, and perception of geometric mereology).
- It is not clear in what sense recursion is constitutive of the language faculty.

UG as Recursion

- Hauser et. al (2002) reduce the "narrow" language faculty (FLN) to recursion, which allows for operations that generate unbounded sets of expressions and hierarchical syntactic structures.
- Recursion may have first emerged in other cognitive domains (such as computation with numbers), and then been adapted to language.
- Pinker and Jackendoff (2005) observe that recursion is pervasive across various primate and human cognitive domains (navigation, recognition of kinship structures, and perception of geometric mereology).
- It is not clear in what sense recursion is constitutive of the language faculty.

UG as Recursion

- Hauser et. al (2002) reduce the "narrow" language faculty (FLN) to recursion, which allows for operations that generate unbounded sets of expressions and hierarchical syntactic structures.
- Recursion may have first emerged in other cognitive domains (such as computation with numbers), and then been adapted to language.
- Pinker and Jackendoff (2005) observe that recursion is pervasive across various primate and human cognitive domains (navigation, recognition of kinship structures, and perception of geometric mereology).
- It is not clear in what sense recursion is constitutive of the language faculty.

UG as Recursion

- Hauser et. al (2002) reduce the "narrow" language faculty (FLN) to recursion, which allows for operations that generate unbounded sets of expressions and hierarchical syntactic structures.
- Recursion may have first emerged in other cognitive domains (such as computation with numbers), and then been adapted to language.
- Pinker and Jackendoff (2005) observe that recursion is pervasive across various primate and human cognitive domains (navigation, recognition of kinship structures, and perception of geometric mereology).
- It is not clear in what sense recursion is constitutive of the language faculty.

Language Acquisition in a Reduced UG

- The MP is a drastic retreat from the rich domain specific mechanisms of previous theories.
- Chomsky argued that these elaborate devices were required because domain general procedures could not overcome the poverty of PLD for language acquisition.
- The current theory fades into an application of procedures shared with other cognitive capacities.
- It is not clear how such an impoverished theory can solve the learning problems for which the earlier models were designed.
- In fact explaining language acquisition appears to have been demoted to a peripheral interest in the current theory.

Language Acquisition in a Reduced UG

- The MP is a drastic retreat from the rich domain specific mechanisms of previous theories.
- Chomsky argued that these elaborate devices were required because domain general procedures could not overcome the poverty of PLD for language acquisition.
- The current theory fades into an application of procedures shared with other cognitive capacities.
- It is not clear how such an impoverished theory can solve the learning problems for which the earlier models were designed.
- In fact explaining language acquisition appears to have been demoted to a peripheral interest in the current theory.

Language Acquisition in a Reduced UG

- The MP is a drastic retreat from the rich domain specific mechanisms of previous theories.
- Chomsky argued that these elaborate devices were required because domain general procedures could not overcome the poverty of PLD for language acquisition.
- The current theory fades into an application of procedures shared with other cognitive capacities.
- It is not clear how such an impoverished theory can solve the learning problems for which the earlier models were designed.
- In fact explaining language acquisition appears to have been demoted to a peripheral interest in the current theory.

Language Acquisition in a Reduced UG

- The MP is a drastic retreat from the rich domain specific mechanisms of previous theories.
- Chomsky argued that these elaborate devices were required because domain general procedures could not overcome the poverty of PLD for language acquisition.
- The current theory fades into an application of procedures shared with other cognitive capacities.
- It is not clear how such an impoverished theory can solve the learning problems for which the earlier models were designed.
- In fact explaining language acquisition appears to have been demoted to a peripheral interest in the current theory.

Language Acquisition in a Reduced UG

- The MP is a drastic retreat from the rich domain specific mechanisms of previous theories.
- Chomsky argued that these elaborate devices were required because domain general procedures could not overcome the poverty of PLD for language acquisition.
- The current theory fades into an application of procedures shared with other cognitive capacities.
- It is not clear how such an impoverished theory can solve the learning problems for which the earlier models were designed.
- In fact explaining language acquisition appears to have been demoted to a peripheral interest in the current theory.

Conclusions and the Focus of the Course

- The APS claims that domain general learning procedures are not adequate for solving the learning task of grammar induction from PLD.
- A strong domain specific learning bias, encoded as a rich theory of UG, is posited as the solution to this problem.
- It is not clear how the drastically truncated model of UG proposed in recent linguistic theory can account for language acquisition.

Conclusions and the Focus of the Course

- The APS claims that domain general learning procedures are not adequate for solving the learning task of grammar induction from PLD.
- A strong domain specific learning bias, encoded as a rich theory of UG, is posited as the solution to this problem.
- It is not clear how the drastically truncated model of UG proposed in recent linguistic theory can account for language acquisition.

Conclusions and the Focus of the Course

- The APS claims that domain general learning procedures are not adequate for solving the learning task of grammar induction from PLD.
- A strong domain specific learning bias, encoded as a rich theory of UG, is posited as the solution to this problem.
- It is not clear how the drastically truncated model of UG proposed in recent linguistic theory can account for language acquisition.

Conclusions and the Focus of the Course

- In this course we will not generally be concerned with the historical development of Chomsky's theory of UG.
- Our main interest is to formulate the APS as a problem in computational learning theory.
- We will explore the extent to which domain general learning procedures of the kind used in machine learning can efficiently achieve grammar induction.
- To the extent that this is possible, it undermines the force of the APS for a strong bias model of language acquisition.
- We will consider a variety of computational and cognitive issues that bear on this question.

Conclusions and the Focus of the Course

- In this course we will not generally be concerned with the historical development of Chomsky's theory of UG.
- Our main interest is to formulate the APS as a problem in computational learning theory.
- We will explore the extent to which domain general learning procedures of the kind used in machine learning can efficiently achieve grammar induction.
- To the extent that this is possible, it undermines the force of the APS for a strong bias model of language acquisition.
- We will consider a variety of computational and cognitive issues that bear on this question.

Conclusions and the Focus of the Course

- In this course we will not generally be concerned with the historical development of Chomsky's theory of UG.
- Our main interest is to formulate the APS as a problem in computational learning theory.
- We will explore the extent to which domain general learning procedures of the kind used in machine learning can efficiently achieve grammar induction.
- To the extent that this is possible, it undermines the force of the APS for a strong bias model of language acquisition.
- We will consider a variety of computational and cognitive issues that bear on this question.

Conclusions and the Focus of the Course

- In this course we will not generally be concerned with the historical development of Chomsky's theory of UG.
- Our main interest is to formulate the APS as a problem in computational learning theory.
- We will explore the extent to which domain general learning procedures of the kind used in machine learning can efficiently achieve grammar induction.
- To the extent that this is possible, it undermines the force of the APS for a strong bias model of language acquisition.
- We will consider a variety of computational and cognitive issues that bear on this question.

Conclusions and the Focus of the Course

- In this course we will not generally be concerned with the historical development of Chomsky's theory of UG.
- Our main interest is to formulate the APS as a problem in computational learning theory.
- We will explore the extent to which domain general learning procedures of the kind used in machine learning can efficiently achieve grammar induction.
- To the extent that this is possible, it undermines the force of the APS for a strong bias model of language acquisition.
- We will consider a variety of computational and cognitive issues that bear on this question.