Computational Psycholinguistics

Lecture 9: Computational Syntax Acquisition

Afra Alishahi

January 12, 2009

Human Language Acquisition

- Representation of the linguistic knowledge
  - What is innate, what is learnable?
  - How is the knowledge organized in mind and brain?
    - Are there separate areas/levels for representing lexical/syntactic/semantic knowledge?
- Acquisition of the linguistic knowledge
  - What are the processes involved in language learning?
  - Are different types of knowledge acquired in order?

Learnability in Acquisition

- Representation of the linguistic knowledge
  - What is innate, what is learnable?
  - How is the knowledge organized in mind and brain?
    - Are there separate collections/levels for representing lexical/syntactic/semantic knowledge?
- Acquisition of the linguistic knowledge
  - What are the processes involved in language learning?
  - Are different types of knowledge acquired in order?

Modularity in Acquisition

- Representation of the linguistic knowledge
  - What is innate, what is learnable?
  - How is the knowledge organized in mind and brain?
    - Are there separate collections/levels for representing lexical/syntactic/semantic knowledge?
- Acquisition of the linguistic knowledge
  - What are the processes involved in language learning?
  - Are different types of knowledge acquired in order?

Acquisition vs. Processing

- How is acquisition related to processing?
- Competence: what it means to "know" a language
  - syntactic and semantic rules and representations provided by a linguistic theory
- Performance: how is such knowledge used on-line to recover the meaning for a given sentence
  - a psychologically plausible parsing algorithm

Computational Lang. Acquisition

- Computational modeling of human language acquisition:
  - Providing cognitively plausible formalisms for representing linguistic knowledge
  - Developing algorithms that can acquire knowledge of language from exposure to linguistic data

Cognitive Modeling: Evaluation

- Cognitive models cannot be solely evaluated based on their accuracy in performing a task
    - The behavior of the model must be compared against observed human behavior
    - The errors made by humans must be replicated and explained
- Evaluation of cognitive models dependents highly on experimental studies of language

Experimental Studies

- Collected data on child language development
    - CHILDES database (MacWhinney, 1995)
- Experimental methods
    - Neuroscientific methods
    - Preferential looking studies

Nativism

- The Innateness Hypothesis (IH):
  - Humans have innately specified, domain specific knowledge in several areas, in particular language
  - The hypothesis must ultimately settled by neurological evidence, but for now, we have to use indirect evidence from psycholinguistics
- Localization:
  - Our ability to process language is localized to specific regions of the brain (Bates, 1994)
- Innateness is not the same as localization

Domain-specificity of Language

- The key claim of innateness is that the humans' innate abilities of language are domain-specific
  - Specific to language
  - Include highly detailed linguistic knowledge
- Many species have domain-specific, innately specified abilities or behaviours
  - E.g., spiders weaving complex webs on their first attempts

Innateness of Language

- Newborns exhibit few complex behaviours immediately after birth
- But it is claimed that children acquire language (esp. syntax) without being exposed to sufficient stimulus
- Therefore, there must be a pre-existing domain-specific innate structure that partially specifies the structure of their knowledge of language (Chomsky, 1986; Pinker, 1994)

Argument from the Poverty of the Stimulus

- APS: main argument for Innateness Hypothesis
- Argument from the Poverty of the Stimulus (Chomsky, 1965): linguistic experience of a child is not sufficiently rich for learning the grammar of the language
- Children learn the language, thus they must have access to some innate source of information to constrain the search for the correct grammar

Premise of APS

- Knowing a language involves knowing a grammar
    - A domain-specific form of knowledge representation that permits the creation of an infinite set of well-formed utterances
- There are no general learning algorithms that can learn grammars from the linguistic evidence that children are exposed to.

Universal Grammar

- Universal Grammar: a limited set of rules which organize language in the human brain (Chomsky)
  - Underlying assumption: all languages have a common structural basis
- Most of the UG rules have the form *"if a language has a feature X, it will also have the feature Y."*
  - Example: *"If a language is head-initial, it will have prepositional phrases. If it is head-final, it will have post-positional phrases."*

Principles and Parameters

- Principles and Parameters is a framework for representing Universal Grammar.
- P&P: a humans' syntactic knowledge can be modeled with two formal mechanisms:
  - A finite set of fundamental principles that are common to all languages (e.g., a sentence must have a subject)
  - A finite set of parameters that determine syntactic variability amongst languages (e.g., a binary parameter that determines whether or not the subject of a sentence must be overtly pronounced)

Learning as Parameter Setting

- In the P&P framework, learning a language involves setting the parameters of UG to the appropriate values for the current language
  - Fixing the values of a finite set of parameters to select a single fully-specified grammar
- Formal models of parameter setting are developed for a small set of grammars (Clark 1992, Gibson & Wexler 1994, Niyogi & Berwick 1996, Briscoe 2000, Buttery 2006)

Computational Simulation of P&P

- Gibson & Wexler (1994):
  - Each trigger (i.e., sentence) signals the value of some parameter and can guide to the target grammar
  - Learner must update a parameter so that the trigger can be parsed appropriately
- Briscoe (2000)
  - Parameter setting in a Generalized Categorial Grammar (GCG)
  - Learning is based on a partial ordering on the updating of parameter settings

Limitations of Computational P&P

- Parameter setting framework predicts a huge space of possible grammars
  - 20 binary parameters lead to > 1 million grammars
- The search spaces for a grammar contain local maxima, which may cause a learner to converge to an incorrect grammar
- Most of the models are psychologically implausible because they predict that a child may repeatedly revisit the same hypothesis and/or jump randomly around the hypothesis space.

P&P and UG: Criticism

- Formalizing a UG that covers all of the existing languages has been a challenge
- Learning in P&P relies on well-formed, complete sentences as input, but conversation analysis shows that speakers often use incomplete data
- P&P ignores the role of linguistic experience in learning, and cannot explain frequency effects
    - Infants are shown to be sensitive to transitional probabilities in artificial languages (Saffran et al, 1996)

Learning Grammar from Corpora

- A number of computational models are proposed to show that learning a grammar from corpus data is possible (mostly CFG)
- A variety of machine learning techniques are used, mainly to induce a grammar that fits the corpus data best
- Most of these models are not incremental, and focus on syntax acquisition without taking semantics into account

Example: MOSAIC

- MOSAIC (Model Of Syntax Acquisition In Children; Jones et al, 2000)
  - Learns from raw text, and produces utterances similar to what children produce
  - Uses a discrimination network, where nodes represent single words and links present a generative link
  - Learning involves expanding the network based on the input data, and production involves traversing the network and outputting the contents of the links

Example: Clark (2001)

- A model of syntax acquisition (Clark, 2001):
  - Unsupervised induction of stochastic context-free grammars from tagged text
  - Sets of tag sequences are clustered together based on their context
  - A grammar is iteratively built by forming clusters and defining rules that best describe data
  - No lexical information is learned by the algorithm

Distributed Representation as an Alternative to Grammar

- Claim: knowing a language is not equated with knowing a grammar.
  - Knowledge of language is developed in the course of learning to perform primary communicative tasks of comprehension and production.
  - Neural networks: different levels of linguistic representation are emergent structures that a network develops in the course of learning
  - E.g., Elman (1990, 1991), Allen (1997), Allen & Seidenberg (1999)

Usage-based Accounts of Language Acquisition

- Claim: children learn language regularities from the input alone, without guidance in the form of the innate principles
- Motivation: experimental studies on language comprehension and generation in children
  - Children build their linguistic knowledge around individual items, rather than adjusting some general grammar rules they already possess

Verb Island Hypothesis

- Verb Island Hypothesis (Tomasello, 1992):
  - Young children initially learn verbs and their arguments as lexical constructions, and on an item-by-item basis
  - Each verb forms its own 'island' consisting of verb-specific constructions with open nominal slots
  - More general constructions emerge over time as children generalize the patterns that they have learned for one verb to another.

Non-domain-specific Mechanisms

- Claim: children use cognitive processes to gradually categorize the syntactic structure of their item-based constructions
  - Imitation: reproducing the language adults produce for the same communicative function
  - Analogy: detecting similarities between individual items' behaviour
  - Structure mapping: detecting both structural and functional similarities in utterances independent of the specific words involved

Syntax vs. Semantics

- Structure mapping: detecting both structural and functional similarities in utterancesnt
- How is the surface structure (i.e., syntax) linked to the underlying meaning (i.e., semantics)?
    - Nativist account: syntax is learned independently of semantics
    - Usage-based account: syntax and semantics are learned at the same time

Marr's Levels of Modeling

- Theories often provide a relatively high-level characterization of a process
- Marr (1982) identifies three levels of describing cognitive processes:
    - Computational level: defines *what* is computed
    - Algorithmic level: specifies *how* computation takes place
    - Implementation level: states how the algorithms are actually *realized* in brain