

CaCL  
Spring 2012  
Tuesdays, 3.30-5.00  
Oxley 024

### Revised Schedule (3/27/2012)

#### Bayesian learning and grammar

4/3 (William)

Pearl, L. & Goldwater, S. (forthcoming - updated 12/1/11). Statistical Learning, Inductive Bias, and Bayesian Inference in Language Acquisition, In J. Lidz, W. Snyder, & C. Pater (eds), The Oxford Handbook of Developmental Linguistics. (Language acquisition is a problem of induction: the child learner is faced with a set of specific linguistic examples and must infer some abstract linguistic knowledge that allows the child to generalize beyond the observed data, i.e., to both understand and generate new examples. Many different generalizations are logically possible given any particular set of input data, yet different children within a linguistic community end up with the same adult grammars. This fact suggests that children are biased towards making certain kinds of generalizations rather than others. The nature and extent of children's inductive bias for language is highly controversial, with some researchers assuming that it is detailed and domain-specific (e.g., Chomsky 1973, Baker 1978, Chomsky 1981, Huang 1982, Fodor 1983, Bickerton 1984, Lasnik & Saito 1984, Gleitman & Newport 1995) and others claiming that domain-general constraints on memory and processing are sufficient to explain the consistent acquisition of language (e.g., Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett 1996, Sampson 2005). In this chapter, we discuss the contribution of an emerging theoretical framework called Bayesian learning that can be used to investigate the inductive bias needed for language acquisition.)

#### Anaphoric *one*

4/10 (Peter)

Pearl, L. & Lidz, J. (2009) When domain general learning fails and when it succeeds: Identifying the contribution of domain specificity, *Language Learning and Development*, 5(4), 235-265. (We identify three components of any learning theory: the representations, the learner's data intake, and the learning algorithm. With these in mind, we model the acquisition of the English anaphoric pronoun **one** in order to identify necessary constraints for successful acquisition, and the nature of those constraints. Whereas previous modeling efforts have succeeded by using a domain-general learning algorithm that implicitly restricts the data intake to be a subset of the input, we show that the same kind of domain-general learning algorithm fails when it does not restrict the data intake. We argue that the necessary data intake restrictions are domain-specific in nature. Thus, while a domain-general algorithm can be quite powerful, a successful learner must also rely on domain-specific learning mechanisms when learning anaphoric *one*.)

Pearl, L., & Mis, B. (submitted - updated 1/13/12). What Indirect Evidence Can Tell Us About Universal Grammar: Anaphoric One Revisited. (read with Foraker, et al. 2009, Lidz et al 2003) [updated version of Pearl, L., & Mis, B. (2011). How Far Can Indirect Evidence Take Us? Anaphoric One Revisited, In L. Carlson, C. Hölscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society, 879-884. Austin, TX: Cognitive Science Society. (A controversial claim in linguistics is that children learning their native language face an induction problem: the data in their input are insufficient to identify the correct language knowledge as rapidly as children do. If this is true, children must bring some helpful learning biases to the problem, and the nature of these biases is often debated. In particular, induction problems are often used to motivate innate, domain-specific biases which are part of Universal Grammar. We examine the case study of English anaphoric *one*, an induction problem receiving recent attention in the computational modeling literature, and consider whether indirect evidence leveraged by an online probabilistic learner from a broader input set could be effective. We find our learner can reproduce child learning behavior, given child-directed speech. We discuss what learning biases are required for acquisition success, and how this impacts the larger debate about Universal Grammar.)

### **Architectures**

4/17 (Abby)  
Cutler & Clifton (1999)

### **Processing representations**

4/24 (Kodi)  
Norris, Cutler & McQueen (2006)

### **Syntactic principles**

5/1 (Marten, Peter)  
Perfors, A., Tenenbaum, J.B., Gibson, E., and Regier, T. (2010) How recursive is language? A Bayesian exploration. In H. van der Hulst (ed.) *Recursion and human language*. Mouton: DeGruyter: 159-175. [\[pdf\]](#) Materials can be found [here](#). (Recursion involves an inherent tradeoff between simplicity and goodnessofit: a grammar with recursive rules might be simpler than one without, but will predict the sentences in any finite corpus less exactly. As a result, one cannot conclude that any particular grammar or grammatical rule is recursive, given a corpus, without some way to quantify and calculate this tradeoff in a principled way. We present a Bayesian framework for performing rational inference that enables us to quantitatively evaluate grammars with and without recursive rules and normatively determine which best describe the sentences in a corpus of childdirected spoken English. Our results suggest three main points. First, they suggest that rational principles would favor a grammar with a specific type of recursive rule, even if there are relatively few instances of particular recursivelygenerated sentences in the input. Second, they suggest that the optimal grammar may occupy a representational middle ground between fully recursive

and nonrecursive. Finally, our results suggest that the optimal grammar may represent subject NPs distinctly from object NPs. We suggest that our method and insights can be usefully applied to address other questions in linguistics and the study of recursion.)

Perfors, A., Tenenbaum, J.B., and Regier, T. (2011) The learnability of abstract syntactic principles. *Cognition* 118(3): 306-338. [\[pdf\]](#). Materials can be found [here](#). (Children acquiring language infer the correct form of syntactic constructions for which they appear to have little or no direct evidence, avoiding simple but incorrect generalizations that would be consistent with the data they receive. These generalizations must be guided by some inductive bias – some abstract knowledge – that leads them to prefer the correct hypotheses even in the absence of directly supporting evidence. What form do these inductive constraints take? It is often argued or assumed that they reflect innately specified knowledge of language. A classic example of such an argument moves from the phenomenon of auxiliary fronting in English interrogatives to the conclusion that children must innately know that syntactic rules are defined over hierarchical phrase structures rather than linear sequences of words (e.g., Chomsky, 1965, 1971, 1980; Crain & Nakayama, 1987). Here we use a Bayesian framework for grammar induction to address a version of this argument and show that, given typical child-directed speech and certain innate domain-general capacities, an ideal learner could recognize the hierarchical phrase structure of language without having this knowledge innately specified as part of the language faculty. We discuss the implications of this analysis for accounts of human language acquisition.)

5/8 (Mike W.)

Kwiatkowski, Tom, Luke Zettlemoyer, Sharon Goldwater, Mark Steedman. (2011). Lexical generalization in CCG grammar induction for semantic parsing. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. (We consider the problem of learning factored probabilistic CCG grammars for semantic parsing from data containing sentences paired with logical-form meaning representations. Traditional CCG lexicons list lexical items that pair words and phrases with syntactic and semantic content. Such lexicons can be inefficient when words appear repeatedly with closely related lexical content. In this paper, we introduce factored lexicons, which include both lexemes to model word meaning and templates to model systematic variation in word usage. We also present an algorithm for learning factored CCG lexicons, along with a probabilistic parse-selection model. Evaluations on benchmark datasets demonstrate that the approach learns highly accurate parsers, whose generalization performance benefits greatly from the lexical factoring.)

### **Metrical phonology**

5/15 (Mike Ph.)

Pearl, L. (2011). When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition*, 18(2), 87-120. (Parametric systems have been proposed as models of how humans represent

knowledge about language, motivated in part as a way to explain children's rapid acquisition of linguistic knowledge. Given this, it seems reasonable to examine if children with knowledge of parameters could in fact acquire the adult system from the data available to them. That is, we explore an argument from acquisition for this knowledge representation. We use the English metrical phonology system as a nontrivial case study and test several computational models of unbiased probabilistic learners. Special attention is given to the modeled learners' input and the psychological plausibility of the model components in order to consider the learning problem from the perspective of children acquiring their native language. We find that such cognitively inspired unbiased probabilistic learners uniformly fail to acquire the English grammar proposed in recent metrical studies from English child-directed speech, suggesting that probabilistic learning alone is insufficient to acquire the correct grammar when using this parametric knowledge representation. Several potential sources of this failure are discussed, along with their implications for the parametric knowledge representation and the trajectory of acquisition for English metrical phonology.)

## **Islands**

5/22 (Peter)

- Pearl, L. & Sprouse, J. (forthcoming - updated 2/1/12) Computational Models of Acquisition for Islands, In J. Sprouse & N. Hornstein (eds), *Experimental Syntax and Islands Effects*. Cambridge University Press. (In this chapter, we examine child-directed speech input in order to formalize the apparent induction problem that has been claimed by linguists. We then explore a statistical learning model of island constraints that is based upon the frequency of certain abstract structures in the input. The model is tested on input derived from child-directed speech (from CHILDES: MacWhinney (2000)) as well as input derived from adult-directed speech (Switchboard section of Treebank-3: Marcus et al. 1999) and adult-directed text (Brown section of Treebank-3: Marcus et al. 1999). We use this statistical model to investigate the types of learning biases that are necessary to learn these constraints from the input, with the goal of determining whether any innate domain-specific biases (i.e., UG) are necessary. Our results suggest that a learner only requires the following biases to learn syntactic island constraints from child-directed input, none of which are considered specific to the nativist/UG approach to language acquisition: (i) perceive the input with a phrase-structure-based representation of sentences (i.e., a parser) (ii) track the frequency of sequences of three phrase structure nodes (trigrams of phrase structure nodes), and their associated probability of occurring (iii) construct a longer dependency by combining trigrams of phrase structure nodes, and assess that dependency's grammaticality based on that combination.
- Pearl, L., & Sprouse, J. (submitted). Syntactic Islands without Universal Grammar: A computational model of the acquisition of constraints on long-distance dependencies. (The induction problems facing language learners have played a

central role in debates about the types of learning biases that exist in the human brain. Many linguists have argued that the necessary learning biases to solve these language induction problems must be both innate and language-specific (i.e., the Universal Grammar (UG) hypothesis). Though there have been several recent high-profile investigations of the necessary types of learning biases, the UG hypothesis is still the dominant assumption for a large segment of linguists due to the lack of studies addressing central phenomena in generative linguistics. To address this, we focus on how to learn constraints on long-distance dependencies, sometimes called syntactic islands. We use formal acceptability judgment data to identify the target state of learning for syntactic island constraints, and conduct a corpus analysis of child-directed data to affirm that there does appear to be an induction problem when learning these constraints. We then create a computational model that successfully learns the pattern of acceptability judgments observed in formal experiments, based on realistic input data. Crucially, while this modeled learner does require several types of learning biases to work in concert, it does not require any (clearly) innate, domain-specific biases. This suggests that syntactic islands constraints can in principle be learned without relying on UG. We discuss the consequences of this learner for the learning bias debates, as well as questions raised by the nature of the linguistic knowledge that is required by this learner.)

### **Word order generalizations**

5/29 (Andrea)

Maurits, L., Perfors, A., and Navarro, D.J. (2010). Why are some word orders more common than others? A uniform information density account. In *Advances in Neural Information Processing Systems 23*: 1585-1593. [\[pdf\]](#). Materials can be found [here](#). (Languages vary widely in many ways, including their canonical word order. A basic aspect of the observed variation is the fact that some word orders are much more common than others. Although this regularity has been recognized for some time, it has not been well-explained. In this paper we offer an informationtheoretic explanation for the observed word-order distribution across languages, based on the concept of Uniform Information Density (UID). We suggest that object-first languages are particularly disfavored because they are highly nonoptimal if the goal is to distribute information content approximately evenly throughout a sentence, and that the rest of the observed word-order distribution is at least partially explainable in terms of UID. We support our theoretical analysis with data from child-directed speech and experimental work.)