Project no.: 027657

Project full title: Perception, Action & Cognition through learning of Object-Action Complexes

Project Acronym: PACO-PLUS

Deliverable no.: D5.2

Title of the deliverable: Grounded Grammar Induction for a Robot

Contractual Date of Delivery to the CEC:	31 January 2007
Actual Date of Delivery to the CEC:	30 January 2007
Organisation name of lead contractor for this deliverable:	UEDIN
Author(s): Chris Geib, Ron Petrick, and Mark Steedman	
Participant(s): UEDIN, UL, AAU, BCCN	
Work package contributing to the deliverable:	WP4,WP5
Nature:	R
Version:	Final
Total number of pages:	21
Start date of project:	1^{st} Feb. 2006 **Duration:** 48 month

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

The core focus of WP5 is the generalization of the action representation developed in WP2, WP3, and WP4 to cover communicative acts, and the formalization of syntax and semantics for communication and interaction in natural language with situated purposeful agents, together with mechanisms for the acquisition of grammar from sentence-meaning pairs. The deliverable and the attached paper are exclusively concerned with the nature of the problem of language acquisition on the basis of paired presentations of sentences of any human language and contextually supported meanings for those sentences. The paper shows that a very simple statistical model can simulate the general course of acquisition, including certain patterns of overgeneralization, without adherence to any subset principle, and without the use of parametric triggers and attendant ordering principles that have been postulated in the recent literature.

Keyword list: Combinatory Categorial Grammar (CCG); Language Acquisition; Grammatical Bootstrapping; Generative Statistical Models of Grammars and Parsers

Table of Contents

1. Executive Summary

The core focus of WP5 is the generalization of the action representation developed in WP2, WP3 and WP4 to cover communicative acts, and the formalization of syntax and semantics for communication and interaction in natural language with situated purposeful agents, together with mechanisms for the acquisition of grammar from sentence-meaning pairs. The deliverable and the attached paper are exclusively concerned with the nature of the problem of language acquisition on the basis of paired presentations of sentences of any human language and contextually supported meanings for those sentences. The paper shows that a very simple statistical model can simulate the general course of acquisition, including certain patterns of overgeneralization, without adherence to any subset principle, and without the use of parametric triggers and attendant ordering principles that have been postulated in the recent literature. The associated deliverable D5.1 shows how the LDEC action representation and the associated PKS planner developed under WP4 and described in D4.3.1 can both be induced from lower-level representations of states and state transitions, and provide a basis for natural language semantics at the higher level of Combinatory Categorial Grammar, providing the input to the system for either the child or the PACOPLUS agent. Both of these papers are theoretical and look ahead to the next phase of the project, as was anticipated in the plan of work in the annex, and the account of KRA 4 in the Annex (Section 6), since at this stage the low-level modules are not delivering object-concepts at a level appropriate to the formulation of semantics. In particular linguistic semantics grounded in robot sensory-motor schemata that will provide the basis for learning is yet to be developed.

Combinatory Categorial Grammar (CCG, Steedman 2000) is a theory of grammar according to which all language-specific grammatical information resides in the lexicon. A small universal set of strictly type-driven, non-structure dependent, syntactic rules (based on Curry's combinators **B**, **S**, and **T**) then "projects" lexical items into sentence-meaning pairs and defines the mapping from one to the other.

Steedman (2002b,a) showed how the same set of combinatory operations were involved in human and animal non-linguistic planning, and defined a Linear Dynamic version of the Event Calculus (LDEC) as a notation for such a planner. Work by UEDIN under PACOPLUS support reported under deliverable D4.3.1, implements LDEC as a high-level symbolic planner under the PKS framework of Petrick and Bacchus (2002, 2004).

The present report analyzes the problem of connecting this planner to a mechanism for inducing a language-specific CCG grammar from presentations of sentences and (probably ambiguous, possibly noisy) contextually-supported meanings. CCG is being used as a basis for interaction with semantically grounded robots in a number of other European and American projects, notably under EU FP6 IST IP CoSy (Kruijff and Brenner 2006) and in Leslie Kaelbling's group at MIT (Zettlemoyer, Pasula and Kaelbling 2005). The present paper offers a basis for a completely general and strikingly simple account of language acquisition in human and artificial systems for any semantics, including semantics defined on the basis of the kind of dialog actions considered in deliverable D5.1., Annex B. It is potentially applicable to all of these systems.

The document consists of a single paper describing this work, included in the present paper as Annex A.

A: The Computational Problem of Language Acquisition (to be submitted: presented at the Institute of Research in Cognitive Science (IRCS) Colloquium, University of Pennsylvania, January 2007). This paper outlines a complete model of language acquisition. It uses the framework of CCG but is applicable to any lexicalized grammar formalism, such as Tree adjoining Grammar (TAG, Joshi and Schabes 1992), Lexical-Functional Grammar (LFG, Bresnan 1982), Head driven Phrase-Structure Grammar (HPSG, Pollard and Sag 1994), and Type-Logical Grammar (TLG, Morrill (1994)). To the extent that Construction Grammar (ConstG, Goldberg (1995)) can be lexicalized (which appears to be completely) it also applies to that.

The paper is an extension of work by Zettlemoyer and Collins 2005, who also use CCG as a framework. The present paper differs in doing language learning in the full space of universal grammar as captured in CCG, and in using a generative statistical model, rather than the discriminative Maximum Entropy model used by them. The advantage of a generative model of lexical acquisition is that, because it learns probabilities $P(Syntax, Semantics|Word)$ of adult utterance, rather than discriminative weights, the model can be inverted to yield predictions about the probabilities of errorful utterance by the child $P(Word, Syntax|Semantics)$ These probabilities can be used to make quantitative predictions about the type of error that will be made by the child under conditions of forced elicitation of the kind investigated by Crain and Thornton (1998), and about the learning curve of the target construction.

2. Role of Language Acquisition in PACOPLUS

The relation of prelinguistic semantics, grounded in sensory motor experience, to high level cognition including language is a central concern of PACOPLUS. The solution presented here to the problem of language acquisition is a very general one. The research has as much to gain from involvement with grounded agents learning action representations as the agents have in terms of provision of spoken interfaces. That is why we are not limiting the language interface to a fixed set of slot-and-filler sentence templates, hand tailored to the PACOPLUS domain, an exercise that would be entirely without scientific interest.

3. Relation to Demonstrator 8.1

The capabilities of Demonstrator 8.1 are decidedly sensory-motor. It is likely that the scope for language learning will be limited, though it will be explored as far as possible. The impact of this research is planned according to the PACOPLUS Annex 1 (see section 6 KRA4) for a later phase, at which point a substantial conceptual base of robot object-action complex (OAC) concepts will have been built up to act as a substrate for a grounded linguistic semantics.

4. Principal Scientific Results

The paper in Annex A shows that the simplest possible generative model predicts the general shape of the childs progress from an initial unstable state in which almost any alternative allowed by universal grammar may be elicited, via a process of exponential reinforcement and extinction which may give the appearance of parametric "switch-setting", to stable adherence to a single form. This result resembles the somewhat different statistical model of Yang (2002), but eschews the use of parameters entirely. It provides a good model for language learning in robots, where problems of error in interpreting the situation and (if standard speech-recognition technology is used) in identifying the string correctly demand a probabilistic approach. There are interesting implications of these results for the purely syntactic, parameter-based approaches of Wexler and Fodor, and for the notion of "syntactic bootstrapping" advanced by Gleitman.

5. Future Work

A number of questions remain open at the time of this report and constitute further work.

1. Children show a number of biases which may work to make this process easier. For example, verbs are acquired later than comparably frequent nouns. It is not clear whether this is an intrinsic cognitive

intellectual development, or whether it is an artefact of the way the data is presented to the child, and is predicted by the model. Answering this question requires closer attention to corpora like CHILDES than we have so far been able to afford.

2. The actual sensory-motor derived semantics that real children bring to bear on this task is almost entirely opaque. One of the objectives of PACOPLUS is to say what such a semantics might look like. The major effort in the remaining period for this work package is to define such a semantics for the robot agents in its own right, in the hope of shedding light on the nature of the child's own, via exploration of language learning on the basis of such an artificial semantics grounded in sensory motor interaction with the world.

6. Publications Associated with D5.1

1. M. Steedman and J. Hockenmaier, 2006: "The Computational Problem of Language Acquisition" (to be submitted: presented at the Institute of Research in Cognitive Science (IRCS) Colloquium, University of Pennsylvania, January 2007)

References

Bresnan, Joan, ed. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Crain, Stephen, and Rosalind Thornton. 1998. *Investigations in Universal Grammar*. Cambridge MA: MIT Press.

Goldberg, Adèle. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: Chicago University Press.

Joshi, Aravind, and Yves Schabes. 1992. "Tree-Adjoining Grammars and Lexicalized Grammars." In Maurice Nivat and Andreas Podelski, eds., *Definability and Recognizability of Sets of Trees*. Princeton, NJ: Elsevier.

Kruijff, Geert-Jan, and Michael Brenner. 2006. "A Cross-modal Approach to Spatio-Temporal Comprehension in Situated Dialog for Human-Robot Interaction." Deliverable D9.3, EU FP6-004250 CoSy.

Morrill, Glyn. 1994. *Type-Logical Grammar*. Dordrecht: Kluwer.

Petrick, Ronald P. A., and Fahiem Bacchus. 2002. "A Knowledge-Based Approach to Planning with Incomplete Information and Sensing." In Malik Ghallab, Joachim Hertzberg, and Paolo Traverso, eds., *Proceedings of the Sixth International Conference on Artificial Intelligence Planning and Scheduling (AIPS-2002)*, 212–221. Menlo Park, CA: AAAI Press.

Petrick, Ronald P. A., and Fahiem Bacchus. 2004. "Extending the Knowledge-Based Approach to Planning with Incomplete Information and Sensing." In Shlomo Zilberstein, Jana Koehler, and Sven Koenig, eds., *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS-04)*, 2–11. Menlo Park, CA: AAAI Press.

Pollard, Carl, and Ivan Sag. 1994. *Head Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark. 2002a. “Formalizing Affordance.” In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society, Fairfax VA, August*, 834–839. Mahwah NJ: Lawrence Erlbaum.

Steedman, Mark. 2002b. “Plans, Affordances, and Combinatory Grammar.” *Linguistics and Philosophy*, 25, 723–753.

Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Zettlemoyer, Luke, and Michael Collins. 2005. “Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars.” In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. Held in conjunction with IJCAI 2005, Edinburgh.

Zettlemoyer, Luke S., Hanna M. Pasula, and Leslie Pack Kaelbling. 2005. “Learning Planning Rules in Noisy Stochastic Worlds.” In *National Conference on Artificial Intelligence (AAAI),*. AAAI.

7. Annexes

A. The Computational Problem of Language Acquisition

Mark Steedman and Julia Hockenmaier

CCG is a theory of grammar in which all language-specific grammatical information resides in the lexicon. A small universal set of strictly type-driven, non-structure dependent, syntactic rules (based on Curry’s combinators B, S, and T) then “projects” lexical items into sentence-meaning pairs. The task that faces the child in the earliest stages of language acquisition can therefore be seen as learning a lexicon on the basis of exposure to (probably ambiguous, possibly somewhat noisy) sentence-meaning pairs, given this universal combinatory “projection principle”, and a mapping from semantic types to the set of all universally available lexical syntactic types.

The paper argues that a very simple statistical model allows children to arrive at a target lexicon without navigation of subset principles, or attention to any attendant notion of trigger other than the notion “reasonably short sentence in a reasonably understandable situation drawn from a reasonably representative sample”. The model explains the pattern of errors that have been found in elicitation experiments. The linguistic notion of “parameter” appears to be redundant to this process.

The Computational Problem of Natural Language Acquisition

Annex A

Mark Steedman
School of Informatics
University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW
steedman@inf.ed.ac.uk

Julia Hockenmaier
Computer and Information Science
University of Pennsylvania
Philadelphia PA 19104
juliahr@cis.upenn.edu

Abstract

CCG is a theory of grammar in which all language-specific grammatical information resides in the lexicon. A small universal set of strictly type-driven, non-structure dependent, syntactic rules (based on Curry's combinators B, S, and T) then "projects" lexical items into sentence-meaning pairs. The task that faces the child in the earliest stages of language acquisition can therefore be seen as learning a lexicon on the basis of exposure to (probably ambiguous, possibly somewhat noisy) sentence-meaning pairs, given this universal combinatory "projection principle", and a mapping from semantic types to the set of all universally available lexical syntactic types.

The paper argues that a very simple statistical model allows children to arrive at a target lexicon without navigation of subset principles, or attention to any attendant notion of trigger other than the notion "reasonably short sentence in a reasonably understandable situation drawn from a reasonably representative sample". The model explains the pattern of errors that have been found in elicitation experiments. The linguistic notion of "parameter" appears to be redundant to this process.

1 Introduction

It seems highly likely that the child's acquisition of a first language is, in machine learning terms, an example of *supervised* learning. That is not to say that they are explicitly instructed by adults, but in coming to know which words of the language are the verbs and which the nouns, and in what linear spatio-temporal order(s) the two may occur, children must have access to something more than the mere strings of words constituting a subset of the legal sentences of the languages.

This agreement is based in part on observation of the extreme rapidity with which language acquisition proceeds, and the absence of negative data While it is theoretically possible, using probabilistic models and unsupervised machine learning, to approximate grammars of any class to any desired degree of accuracy, the computational costs of such learning for realistic grammars are prohibitive. The consensus also rests on the observation that no-one has actually managed to make these techniques work very well computationally for natural language.

The "something more" that the child brings to language acquisition is sometimes referred to as "Universal Grammar", and as such is sometimes talked about in exclusively syntactic terms, as in the "parameter-setting" account of acquisition of Hyams (1986) and much subsequent work, according to which a homunculus "flips switches" corresponding to syntactic parameters such as head-finality and *pro*-drop until the "universal grammar engine" uniquely specifies the language *modulo* its lexicon, in a process that has been likened to a game of Twenty-Questions (Yang 2006:Ch.7).

Such accounts seem to raise as many questions as they answer about the mechanism by which such learning could proceed. In particular, the specific inventory of parameters that this universal machine embodies, the way in which the very large search spaces engendered by even quite small sets of binary independent parameter can be effectively explored (Clark & Roberts 1993), and the aspects of the data that "trigger" their setting (Gibson & Wexler 1995) remain rather unclear. One is uneasily reminded of the warnings of Newell (1973) in a different context, concerning the likely outcome of playing Twenty-Questions with nature.

Nevertheless, there is something deeply right in the idea that the process of language learning proceeds by entertaining all possible grammars, and eliminating all alternatives but one, because that is exactly what the child's developmental behavior looks like, once you know how to look at it. In particular, Crain & Thornton (1998) and their students have shown (using ingeniously forced elicitations) that learning is characterized by great initial variation in productions for any given construction, apparently covering alternatives characteristic of many other languages, followed by abrupt transitions to stable adherence to the correct form for the target language. Yang (2002) offers a probabilistic account of this process in terms of classical Mathematical Learning Theory. While Thornton & Tesan (2006) argue that changes they observe are too abrupt and switch-like to support that particular model, probabilistic models in general are capable of approximating catastrophic, switch-like behavior, so they should not be ruled out.

The present paper uses a computational model derived from work by Siskind (1996), Villavicencio (2002), and Zettlemoyer & Collins (2005) to argue that the notion of parameter setting is metatheoretical, and entirely redundant to the specification of language learning of this kind. The only notion of trigger that it requires is the notion "reasonably short sentence with an independently accessible meaning". The only notion of language specific grammar it needs is the lexicon for the language. The only notion of universal grammar that it needs is a universal mapping from each semantic type to the possible lexical types, together with a universal machine for merging or projecting lexical types and their meaning representations onto grammatical derivations.

2 Semantically Grounded Grammar Acquisition

The only remotely plausible source that has ever been proposed for universal grammar is a universal *semantics*, in the form of structured meanings or logical forms to which the child already has access as language acquisition begins, to which syntactic forms are rather directly attached, and which drastically limit the search space.

To say this much is not very helpful in psychological or linguistic terms, since (as Chomsky never tires of pointing out) linguists don't know that much about how to articulate the semantics. However, the child doesn't *need* to articulate it. They just need to label it, so our theories need to represent it somehow. As a temporary stopgap we'll use terms of the lambda calculus, and defer the problem of what the semantics actually looks like till section 5.

This approach makes the child's problem resemble that of treebank grammar induction for wide coverage parsing (Collins 1997; Charniak 2000; Hockenmaier & Steedman 2002), where sentences hand-annotated with syntactic trees are used to derive a grammar and a statistical parser-model. However, the child's task is a little harder. First, they have to induce the grammar from strings paired with *unordered logical forms*, rather than language-specific ordered derivation trees. That is, they have to work out *which word(s) go with which element(s) of logical form*, as well as the directionality of the syntactic categories (which are otherwise universally determined by the semantic types of the latter). Second, while they do not seem to have to deal with a greater amount of error than is found in the Penn WSJ treebank (McWhinnie 2005), they may need to deal with *situations which support a number of logical forms*. Third, they need to be able to recover from temporary *wrong lexical assignments*. Fourth, they need to tolerate *lexical ambiguity*.

3 Previous Work

Siskind (1995, 1996), Villavicencio (2002), and Zettlemoyer & Collins (2005) offer computational models of this process, the latter two explicitly using CCG.

Siskind and Villavicencio make strong assumptions about the association of words with elements of logical form. Both make similarly strong assumptions about universally available parametrically specified rule- or category- types, the latter assuming a type hierarchy. Both deal with noise and homonymy probabilistically.

Both do the learning in two stages, first associating logical forms with words, then inducing phrase structure rules (Siskind) or directional CCG categories (Villavicencio).

However, there is no necessity to separate the two processes of associating meaning and syntactic type. Zettlemoyer and Collins (UAI 2005) combine the two in a single pass CCG induction algorithm. Crucially, their algorithm allows *any contiguous substring* of the sentence to be a lexical item, so that for the given logical form, the learner has to search the cross-product of the substring powerset of the string with the set of pairs of legal categories with substructure powerset of the logical form, as in the example (9) below, for categories that yield combinatory derivations that yield the correct logical form. Learning is via a log-linear model using lexical entries as features and gradient descent on their weights, iterating over successive sentences of a corpus of sentence-logical form pairs.

The algorithm as presented in 2005 learns only a very small rather unambiguous fragment of English, hand-labeled with uniquely identified database queries as logical forms, and an English specific inventory of possible syntactic category types in lieu of Universal Grammar. However, Siskind's and Villavicencio's results already tell us that the algorithm should work with multiple candidate logical forms. Similarly, their results show that a universal set of category types can be used without overwhelming the learner.

All of these models depend on availability to the learner of short sentences paired with logical forms, since complexity is determined by a cross-product of powersets both of which are exponential in sentence length. A number of techniques are available to make search efficient including *association of incrementally adjusted Bayesian priors with category-types*.

Because it allows multiword elements (MWE) to be lexical entries, Zettlemeyer and Collins' program avoids the problem that two words which consistently collocate, like *want* and *to* fail to reveal which of them means *want*′ and which means *to*′. They can be learned as a single item *want to*. So can idioms and multi-word expressions like "buy the farm," and "take advantage of"

As with Siskind's version, lexical items can have complex meanings—corresponding for example to causatives, whose availability may differ (*swim across* vs. *traverser à la nâge*) across languages. No notion of trigger distinct from that of "reasonably simple string-meaning pair" is necessary.

It is possible to use the statistics of the lexicon itself to implicitly represent "parameters" such as verb-finality, via incrementally adjusted prior probabilities on the members of the set of universally available category types.

4 The Proposal

We will assume as a theory of grammar a version of Combinatory Categorial Grammar (CCG, Steedman 2000b; Steedman & Baldridge 2006) in which all language-specific information resides in the lexicon, and a universal set of combinatory rules including functional composition and lexicalized type-raising as well as function application, projects strings of lexical items onto meanings, and vice versa.

The task that faces the child is to learn the categorial lexicon on the basis of exposure to (probably ambiguous, possibly somewhat noisy) sentence-meaning pairs, given this universal combinatory projection principle, and a mapping from semantic types to the set of all universally available lexical syntactic types.

For a corpus of sentences S_i, each with a number of interpretations I_j, each of which has an number of derivations D_k, the relative frequency f of a lexical entry ϕ, σ, μ for a word with phonology ϕ, syntactic category σ and meaning μ is given by:

(1)
$$f(\langle\phi,\sigma,\mu\rangle) = \sum_i \sum_j P(I_j|S_i) \sum_k P(D_k|I_j,S_i) \cdot n_{D_k}(\langle\phi,\sigma,\mu\rangle)$$

By Bayes' Rule,

(2)
$$\begin{aligned} P(D|I,S) &= \frac{P(D,I,S)}{P(I,S)} \\ &\propto P(D,I,S) \end{aligned}$$

We will assume that $P(D,I,S)$ is a generative model for an (exhaustive) parser, rather than the discriminative model of Zettlemoyer & Collins. One advantage of generative models besides their closeness to competence grammar is that we can invert the parser model to define the probability of an utterance given a meaning.

As the acquisition process begins, this generative model corresponds to the P_{UG}, the probability model of Universal Grammar, which we will assume for present purposes assigns uniform probabilities to everything. This model can be regarded as a log-linear model in which all weights λ_j are unknown and all counts f_j are zero. As the child is exposed to more language, it updates the counts in a language specific model P_G and adjusts a weight λ ($0 \leq \lambda \leq 1$) representing their confidence in G:

(3)
$$\tilde{P}(D,I,S) = \lambda \hat{P}_G(D,I,S) + (1-\lambda) \cdot P_{UG}(D,I,S)$$

The probability of a lexical entry can be defined in terms of (1) as:

(4) $$P_{lex}(\langle \phi, \sigma, \mu \rangle) = \frac{f(\langle \phi, \sigma, \mu \rangle)}{\sum_i f(\langle \phi, \sigma, \mu \rangle_i)}$$

The probability of a semantic intepretation μ_τ of type τ and a syntactic category σ given a word ϕ is given by

(5)
$$P(\sigma, \mu_\tau | \phi) = P(\mu_\tau | \phi) \cdot P(\sigma | \mu_\tau, \phi) \approx P(\mu_\tau | \phi) \cdot P(\sigma | \tau)$$

where

(6) $$P(\mu_\tau | \phi) = \sum_i P_{lex}(\langle \phi, \sigma_i, \mu \rangle)$$

and

(7) $$\begin{aligned} P(\sigma|\tau) &= \frac{P(\sigma,\tau)}{P(\tau)} \\ &= \frac{P(\sigma,\tau)}{\sum_i P(\mu_{\tau_i})} \\ &= \frac{P(\sigma,\tau)}{\sum_i \sum_j P_{lex}(\langle \phi_j, \sigma_j, \mu_{\tau_i} \rangle)} \\ &= \frac{\sum_{i,j} P_{lex}(\langle \phi_i, \sigma, \mu_{\tau_j} \rangle)}{\sum_i \sum_j P_{lex}(\langle \phi_j, \sigma_j, \mu_{\tau_i} \rangle)} \end{aligned}$$

Hence, crucially, we can obtain from the above definitions the probability of uttering a word ϕ, such as "more" or "doggies", given a logical form μ_τ, such as $more'_{((e,t),e)}$, by Bayes' rule:

(8) $$P(\phi|\mu) = \frac{P(\mu|\phi) \cdot P(\phi_i)}{\sum_i P(\mu|\phi_i) \cdot P(\phi_i)} = \frac{P(\mu|\phi)}{\sum_i P(\mu|\phi_i)}$$

The course of language acquisition can then be accounted for as follows.

4.1 The First Few Words

Consider an adult-accompanied child at Piagetian Stage VI who has yet to learn her first word of such a grammar. She encounters a dog, and shows an interest, but fails to grasp the word "doggie". Later, she encounters some *more dogs*. The adult observes the child's evident delight, and says "MORE DOGGIES!."

We can assume that the child has already learned some phonological regularities of the language, and in particular is in a position to consider the possibility that the utterance consists of more than one word (Mattys et al. 1999; Mattys & Juszyk 2001).

What the child must do is consider the cross-product of every non-empty substring ϕ of the utterance "More doggies!" with every connected typed subterm μ_τ of type τ the logical form $more'doggies'$, together with all syntactic categories σ_i that universal grammar allows for the semantic type τ of each such subterm.

We might as a first oversimplification think of the situation as follows:[1]

(9) a. The child thinks: $(more'dogs')_e$
b. The adult says: "MORE DOGGIES!"
c. All possible lexical candidates:

more:=	$\mathbf{NP/N : more'_{((e,t),e)}}$
	$NP\backslash N : more'_{((e,t),e)}$
	$N : dogs'_{(e,t)}$
doggies:=	$NP/N : more'_{((e,t),e)}$
	$NP\backslash N : more'_{((e,t),e)}$
	$\mathbf{N : dogs'_{(e,t)}}$
more doggies:=	$NP : (more'dogs')_e$

All of these candidates are permitted by the universal lexical principles of UG. However, not all of them are consistent with this utterance in this language. For some of them, such as doggies:=$NP/N : more'_{((e,t),e)}$, the universal syntactic projection principle of UG fails to offer any derivation yielding $NP : (more'dogs')_e$. Such candidates may be supported by other utterances, but the present utterance does not give any information on them. They are therefore dropped from further consideration in this cycle, leaving the following reduced set of candi-

[1]The assumption that the child immediately considers the hypothesis that more is a determiner is particularly far-fetched, and will be reviewed later.

dates:

(10) The child's lexical candidates:

more:=	$\mathbf{NP/N : more'_{((e,t),e)}}$
	$N : dogs'_{(e,t)}$
doggies:=	$NP\backslash N : more'_{((e,t),e)}$
	$\mathbf{N : dogs'_{(e,t)}}$
more doggies:=	$NP : (more'dogs')_e$

For each of these candidates, if there is not already a corresponding entry in the lexicon, such an entry is added, with a zero count. Then for each candidate, its count is incremented by 1. Since we are assuming this is the first utterance the child has processed, the lexicon now contains two entries for each of the words "more" and "doggies", each with one count, and one entry for the holophrastic or multi-word entity "more doggies", all with one count. If we assume that the various hypotheses afforded by UG are equiprobable, then by (5) (or by inspection) the conditional probabilities $P(\sigma,\mu|\phi)$ for the former categories are all $\frac{1}{2}$, while that for the latter is 1.

Since for the example so far, $P(\sigma|\mu)$ is always 1, we have the following probabilistic lexicon:

(11) *The Child's First Lexicon*:

ϕ	σ,μ	f	$P(\sigma,\mu\|\phi)$	$P(\phi\|\mu)$
more:=	$\mathbf{NP/N : more'_{((e,t),e)}}$	1	0.5	0.5
	$N : dogs'_{(e,t)}$	1	0.5	0.5
doggies:=	$NP\backslash N : more'_{((e,t),e)}$	1	0.5	0.5
	$\mathbf{N : dogs'_{(e,t)}}$	1	0.5	0.5
more doggies:=	$NP : (more'dogs')_e$	1	1.0	1.0

Since the word counts and conditional probabilities for "more" and "doggies" with them meaning $more'_{((e,t),e)}$ are all equal at this stage, the child may well make errors of overgeneration, using some approximation to "doggies" to mean "more".[2]

However, even on the basis of this very underspecified lexicon, the child will not overgenerate "*doggies more". Moreover, further observations involving utterances like "Bad doggies!" "More cookies!", and "Bad cookies!", with further updates to frequency counts, will rapidly lower the estimated conditional probability of the spurious hypotheses concerning categories and substrings in comparison to the correct ones, indicated in bold type, as follows:

(12) *The Corpus:*

a. More doggies!
b. Bad doggies!
c. More cookies!
d. Bad cookies!

$P(\sigma|\mu)$ is still always 1, so the lexicon is now

(13) *The Child's Lexicon*:

ϕ	σ,μ	f	$P(\sigma,\mu\|\phi)$	$P(\phi\|\mu)$
more:=	$\mathbf{NP/N : more'_{((e,t),e)}}$	2	0.50	0.50
	$N : dogs'_{(e,t)}$	1	0.25	0.25
	$N : cookies'_{(e,t)}$	1	0.25	0.25
bad:=	$\mathbf{NP/N : bad'_{((e,t),e)}}$	2	0.50	0.50
	$N : dogs'_{(e,t)}$	1	0.25	0.25
	$N : cookies'_{(e,t)}$	1	0.25	0.25
doggies:=	$NP\backslash N : more'_{((e,t),e)}$	1	0.25	0.25
	$NP\backslash N : bad'_{((e,t),e)}$	1	0.25	0.25
	$\mathbf{N : dogs'_{(e,t)}}$	2	0.50	0.50
cookies:=	$NP\backslash N : more'_{((e,t),e)}$	1	0.25	0.25
	$NP\backslash N : bad'_{((e,t),e)}$	1	0.25	0.25
	$\mathbf{N : cookies'_{(e,t)}}$	2	0.50	0.25
more doggies:=	$NP : (more'dogs')_e$	1	1.0	1.0
bad doggies:=	$NP : (bad'dogs')'_e$	1	1.0	1.0
more cookies:=	$NP : (more'cookies')_e$	1	1.0	1.0
bad cookies:=	$NP : (bad'cookies')_e$	1	1.0	1.0

At this point, the child is exponentially less likely to generate "doggie" when she means "more". By contemplating the definition (8), the reader should be able to satisfy themselves that this effect will be even stronger for more realistic corpora in which the frequency distribution of words is highly skewed, with open class words like "doggie" being exponentially rarer (hence with lower values for $P(\phi)$) than closed class words like "more". Experimental sampling by elicitation of child utterances during such exponential extinction may well give the appearance of all-or-none setting of parameters like NEG-placement and *pro*-drop claimed by Thornton & Tesan (2006).[3]

This lexicon includes non-standard holophrastic lexical items such as "more doggies". Such spurious lexical entries can later be pruned if necessary on grounds of low relative frequency in the corpus as a whole, along with the spurious entries. Nevertheless, holophrastic lexical items such as "All gone," may be sufficiently common as to be useful in their own right, and persist in the developing lexicon in

[2]The example is constructed, but was inspired by being told of a case in real life when this particular error appears to have occurred (C. Urwin, p.c.).

[3]This effect is related to the "winner-take-all" effect observed in Steels' 2004 game-based account of the otherwise rather different process of establishing a shared vocabulary among agents who have no preexisting language.

parallel with their components.

It is of course possible that the adult will on occasion mistake the proposition that the child has in mind, or that the child will choose such a proposition wrongly, leading to false lexical associations. However, provided the two get it right most of the time, the same process of Bayesian re-estimation of conditional probabilities of these lexical hypotheses for each word will allow the latter to arrive at a correct lexicon.

4.2 Transitives

Up till now, we have been able to ignore the influence of the childs estimate of the prior conditional probability $P(\sigma|\tau)$ of a syntactic category given a semantic type in calculating $P(\sigma_i, \mu|\phi)$ in computing (5), the probability of a syntactic and semantic category given a word, because the examples have only admitted one syntactic category per semantic type per word.

However, unlike intransitive predicates and the determiner category considered in section 4.1, transitive verbs as presented in examples like the following could in principle be assigned either of the two syntactic categories in (15), both of which support a derivation of the logical form:[4]

(14) I see you! := $S : see'you'i'$

(15) a. see := $(S\backslash NP)/NP : \lambda x \lambda y.see'xy$
b. see := $*(S/NP)\backslash NP : \lambda y \lambda x.see'xy$

No SVO language/construction has ever been seriously argued to have a surface syntax corresponding to the second category. We can therefore safely assume either that it is not included in the universal set of possible syntactic categories for interpretations of type $(e,(e,t))$ at all, or that it has an extremely low prior probability.

Specifically, we will assume that the universally permitted set of transitive categories is the following, corresponding to the six basic constituent orders, here listed in order of decreasing frequency of attestation of the order in question.[5]

(16) a. SOV := $(S\backslash NP)\backslash NP : \lambda x \lambda y.see'xy$
b. SVO := $(S\backslash NP)/NP : \lambda x \lambda y.see'xy$
c. VSO := $(S/NP)/NP : \lambda y \lambda x.see'xy$
d. VOS := $(S/NP)/NP : \lambda x \lambda y.see'xy$
e. OVS := $(S/NP)\backslash NP : \lambda x \lambda y.see'xy$
f. OSV := $(S\backslash NP)\backslash NP : \lambda y \lambda x.see'xy$

The decreasing frequency of these orders appears to reflect two independent defeasible constraints. One favors linearization of subject before object. The other favors keeping the syntactic command relations between subject and object as reflected in order of combination the same as those in the logical form.[6]

Since (15b) violates the second of these constraints, we are justified in assuming it has a lower prior. Thus the child faced with the pair (14) effectively has only one candidate category for the transitive verb. However, this does not exhaust the problem of learning transitive verbs, because a context may support more than one category.

4.3 Contextual Ambiguity

Many languages, perhaps all, allow a number of lexical alternations of transitives, as in the case of English "chase/flee" where the same physical situation seems to support more than one logical form. How do children faced with examples like the following avoid the error of making an OVS lexical entry for "flee" with the meaning *chase′*?

(17) Pussies flee doggies!

It is important that examples of the verb class of which "flee" is the most common representative are rare. In particular, in comparison to 162 occurences of inflected forms of the verb "chase," there is exactly one occurrence of any form of "flee" in the entire CHILDES corpus. We are therefore justified in assuming that the child will have encountered plenty of unambiguous transitive verbs in utterances like (14) before encountering examples like (17).

This means that the probability of the category type $(S\backslash NP)/NP : \mu_{((e,(e,t))}$ will be substantial at the time they eventually do encounter (17)—

[4]We continue to assume for the sake of simple exposition that there is only one logical form supported by the context. In particular, we assume that the corresponding passive is not salient, or that if it is it has a distinct logical form from the active. We will abandon these restrictions later.

[5]We assume, following Baldridge (2002), that free word-order languages simply have more than one of these categories.

[6]Two of these categories, VSO and OSV, "wrap" their most oblique argument O(object) around their least oblique argument S(ubject). (These categories are forced under the account of argument cluster coordination and the restriction to the combinators **BTS** in CCG—Steedman 2000b).

for the sake of illustration let's conservatively assume they have seen 1000 tokens—and adds one count each for these two categories. In that case, by (5), since $P(\mu_\tau|\phi)$ is the same for both, and $P((S\backslash NP)/NP|\text{"flee"})$ is $\frac{.25\cdot 1000}{1001} = .25$, while $P((S/NP)\backslash NP|\text{"flee"})$ is $\frac{.25\cdot 1}{1001} = .00025$, the lexical probability for the two entries stand in a ratio of 1000:1.

Thus, *provided the adult's intended meaning is available*, even if with low prior probability, then the child is in a position to assign the correct hypothesis a high probability. (Even if it is not available, the child will assign a low probability to the spurious lexical entry for *chase′*.)

Gleitman 1990 has described the process by which the child resolves contextual ambiguity as "syntactic bootstrapping," meaning that it is the childs knowledge of the language-specific grammar, as opposed to the semantics, that guides lexical acquisition. However, in present terms such an influence on learning is simply emergent from the statistical model used in semantic bootstrapping. We will return to this point in the Discussion.

Like the related proposals of Siskind; Villavicencio; Zettlemoyer & Collins and the somewhat different probabilistic approach of Yang 2002, this proposal considerably simplifies the logical problem of language acquisition. In particular, it allows us to eliminate the Subset Principle of Berwick (1985), and attendant requirements for ordered presentation of unambiguous parametric triggers, both of which appear to present serious problems for the language learner (Angluin 1980; Becker 2005; Fodor & Sakas 2005). Nor does this move contradict widely-held assumptions concerning the "poverty of the stimulus", and in particular the unavailability to the child of negative evidence. The child's progression from the universal superset grammar to the language-specific target grammar is entirely determined by positive evidence raising the probability of correct hypotheses at the expense of incorrect ones. The incorrect hypotheses that are eliminated in this way include any that are introduced by error and noise. The only evidence that the child needs in order to learn their language is a reasonable proportion of utterances involving sentences which are sufficiently short for them to deal with.

4.4 A More Realistic Lexicon

If children's exposure to language were merely confined to recitations of propositions they already had in mind, it would be a dull affair. It is not even clear why they would bother to learn language at all, as Clark (2004) points out in defence of a PAC learning model.

However, the worked example above is deliberately simplified in respect of the child's syntax and semantics. We know from Fernald et al. (1989) and Fernald (1993) that infants are sensitive to interpersonal meanings of intonation from a very early age. In English, intonation contour is used to convey a complex system of information-structural elements, including topic/comment markers and given/newness markers (Bolinger 1965; Halliday 1967; Ladd 1996), and is exuberantly used in speech by and to infants. It is this part of the meaning that constitutes the whole point of the exercise for the child, providing the motivation that Clark questions.

For example, it is likely that the child's representation of the utterance "MORE DOGGIES! is more like (18), which uses the notation of Steedman 2000a, 2006b, in which [S] represents speaker modality (contributed by the LL% boundary tone), ρ indicates a rheme or comment (contributed by the H* pitch-accents), * marks emphasis or kontrast (also contributed by the pitch-accents), and the category NP is "type-raised", indicated by the annotation $NP^{\uparrow}$:[7]

(18)

$$\dfrac{\dfrac{\underset{\text{H}*\quad\text{H}*}{\text{MORE DOGGIES}}}{NP^{\uparrow}_{+,\rho} : \lambda p.p(*more'*dogs')} \quad \dfrac{\underset{\text{LL\%}}{!}}{X_{\phi}\backslash_{\star}X_{\pi,\eta} : \lambda g.\pi[\text{S}]\eta\, g}}{NP^{\uparrow}_{\phi} : [S]\rho\lambda p.p(*more'*dogs')}<$$

"Mummy makes the property afforded by more dogs common ground."

The set of type-raised NP categories licenced by UG that is schematized in (18) as $NP^{\uparrow}$ denotes the set of all order-preserving functions over functions-over-NP onto the results of applying those functions to the original NP. It includes categories of the following two forms, where T is a variable over all cat-

[7]The term kontrast, due to Vallduví & Engdahl (1996) means much the same as Halliday's "new", and is so spelled to distinguish it from other notions of contrast, in particular any distinct notion of "topic contrast".

egory types::

(19) $\mathrm{T}/(\mathrm{T}\backslash \mathrm{NP}) : \lambda p \lambda x.px$
$\mathrm{T}\backslash(\mathrm{T}/NP) : \lambda p \lambda x.px$

We also need the following related non-order-preserving "extracting" categories, in which S_x indicates a distinct type of clause:

(20) $S_x\backslash(S\backslash NP) : \lambda p \lambda x.px$
$S_x/(S/NP) : \lambda p \lambda x.px$

While, up until now, we have only seen one syntactic type per semantic type in the child's lexicon for English, in general a single semantic type may be realized by many syntactic types in a single language, and this is the case for English NPs. Such ambiguity is perfectly compatible with the learning procedure defined earlier and exemplified at (13): it just means that there will be several categories with substantial conditional probability mass $P(\sigma|\phi)$

It may seem surprising that a language should allow so much ambiguity in such a basic linguistic category type as NP. However, this is simply the same proliferation of syntactic types that would be disambiguated in a language with overt morphological *case*. English just happens to be a language which has so-called structural case, implicit in linear order. We shall see that the child will not find this a problem. But first we need to consider the role of intonation in the child's grammar.

While intonation has been shown to be even more markedly discrepant from traditional syntactic structure in child-directed and child-originated speech (Fisher & Tokura 1996; Gerken et al. 1994; Gerken 1996) than in adult dialog, in CCG intonation structure is united with a freer notion of derivational structure. Consider the child in a similar situation faced with the following utterance, from Fisher & Tokura (1996) as discussed in Steedman 1996:

(21)

$$
\begin{array}{cccc}
\text{You} & \text{LIKE} & & \text{the doggies!} \\
 & \text{H}* & \text{L} & \text{LL\%} \\
\hline
S/(S\backslash NP) & (S\backslash NP)/NP & X_\phi\backslash_\star X_{\pi,\eta} & S_\phi\backslash(S_\phi/NP_\phi) \\
: \lambda p.p\ you' & *like' & : \lambda g.\pi[\mathrm{S}]\eta\ g & : [S]\eta\lambda q.q\ dogs' \\
\hline
\multicolumn{2}{l}{S/NP : \lambda x.*like\ x\ you' \quad >\mathbf{B}} & & \\
\hline
\multicolumn{3}{c}{S_\phi/NP_\phi : [S]\rho\lambda x.*like' x\ you' \quad <} & \\
\hline
\multicolumn{4}{c}{S_\phi : ([S]\theta\lambda p.p\ dogs')([S]\rho\lambda x.*like' x\ you') \quad <} \\
\cdots & \cdots & \cdots & \cdots \\
\multicolumn{4}{c}{S : like' dogs' you'}
\end{array}
$$

"Mummy supposes what property the dogs afford to be common ground, Mummy makes it common ground it's me liking them."

Fisher points out that the L intermediate phrase boundary that she observed after the verb makes the intonation structure inconsistent with standard assumptions about surface constituency. However, this intonation structure is isomorphic to the CCG derivation above, which delivers the corresponding theme/rheme information partition directly.

Thus, here too, the availability of the full semantic interpretation, including information-structural information, directly reveals the target grammar. In this case, since the derivation requires the use of the forward composition rule, indexed $>\mathbf{B}$, the child gets information not only about the probability of the verb, the nominative, and the accusative categories of English, but also about the probability of applying the composition rule to the first two categories, the probability that the subject of "like" will be headed by "you", and its object be headed by "doggies". Thus, the child can build the parser model in parallel with learning the grammar. (Indeed, the grammar and parser model are essentially one.)

4.5 Smoothing and Generalization

A standard assumption in wide-coverage parsing using treebank grammars is that the grammar must be generalized and the statistical model must be smoothed with respect to unseen words and word-category pairs. Since all language-specific information in CCG resides in the lexicon, this amounts to predicting unseen word-category pairs and head-dependencies.

Generalizing grammars is a tricky business: Fodor & Sakas offer as an example the observation that the child should assume on the basis of seen topicalizations in English that all NPs can undergo topicalization. However, they should not assume on the basis of observations of negative placement with repect to auxilliaries that the same process can apply to all verbs.

This problem looks rather different from the present perspective. Since we are learning a probabilistic instance of universal grammar, the grammar is already generalized, and predicts all possible word-category pairs. Since topicalization is a lexically-specified contruction in CCG, when the child hears the following as its first example of the construction, it still has available all possible categories for “doggies”, including the preposing topicalized one that supports this derivation:

(22) DOGGIES you LIKE !

$$\begin{array}{ll}
\text{L+H}* \quad \text{LH\%} & \text{H}* \quad \text{LL\%} \\
\overline{\qquad\qquad\qquad}{<} & \overline{\qquad}{>}\mathbf{B} \\
Stop_\phi/(S_\phi/NP_\phi) & \overline{\qquad\qquad\qquad\qquad}{<} \\
[H]\theta\lambda p.p\ {*dogs'} & S_\phi/NP_\phi \\
 & :[S]\rho\lambda x.{*like'}x\ you' \\
\multicolumn{2}{l}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{>}} \\
\multicolumn{2}{l}{S_\phi : ([H]\theta\lambda p.p\ dogs')([S]\rho\lambda x.{*like'}x\ you')} \\
\multicolumn{2}{l}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{2}{c}{S : like'dogs'you'}
\end{array}$$

“I suppose what property dogs (as opposed to something else) afford to be common ground, Mummy makes it common ground it’s me liking them.”

So the conditional probability of this category given this type $P(Stop_\phi/(S_\phi/NP_\phi)|((e,t),t)$ will grow and become available to other words, supporting generalization.

We must correspondingly assume that the non-generalization of the negative category is based on a semantically distinct type of verb.

5 Gavagai!

One might ask at this point how the child or machine comes to have access to the logical form *more′dogs′* (or whatever), and why she does not entertain other candidates, such as *more′tails′*. As Quine (1960) pointed out, this is a different kind of question, whose answer lies in the nature of the child’s sensory-motor interactions with the world, and depends as much on mammalian evolution as on learning in the individual child.

Nevertheless, this observation carries a warning that the semantics that emerges from that interaction and those evolutionary processes may be very unlike the semantics that naive logicist assumptions suggest. For example, the logical form that the child brings to (21) is likely be something more like *give′pleasure′you′dogs′*, so that the lexical entry for “like” of type $(e,(e,t))$ is the following, exhibiting the same “quirky” relation between (structural) nominative case and an underlying dative role that Icelandic exhibits morphologically for the corresponding verb:

(23) like := $(S\backslash NP)/NP : \lambda x\lambda y.give'pleasure'\ y\ x'$

Similarly, it is quite possible that the childs initial representation of the meaning of “more” is as a predicate $S/NP : more$, and that it is the resulting prior on the conditional probability $P(S/NP|e \rightarrow t)$ that is generalized to “allgone”, leading to transient non-standard orders like “Allgone milk”. Or “all gone” may be misanalysed as a proto-determiner like “no more.” These questions are much harder to investigate. While one can annotate corpora such as CHILDES with logical forms, as Villavicencio did, one has very little idea of what relation such logical forms bear to a psychologically real adult semantics, let alone a child’s. This fact makes quantitative testing of the present theory difficult.

One we way around this is to do linguistics, meditating on the huge collection of phenomena to do with binding, case, classification, tense and aspect, and so on, that seem to dimly reveal an underlying system of meanings, in the hope of discerning the real semantics. This is a very hard problem, and progress seems slow.

Another alternative is to investigate the question qualitatively, using simulated language learners. Since everyone believes that the semantics is determined by the child’s sensory-motor experience of acting in the physical world, this makes the use of physically grounded robots particularly interesting. Projects of this kind are under investigation by a number of groups, including those led by Luc Steels, Deb Roy, and Geert-Jan Kruijff. These groups are looking at emergence of agreed vocabulary among prelinguistic agents (Steels & Baillie 2003; Steels 2004), plans and plan-recognition as a basis for situated language understanding (Roy 2005; Gorniak & Roy 2007), and context-dependent spatial models for natural language semantics (Kelleher et al. 2006). However, these projects so far rely on forms of semantics that are designed top-down, using the robot tasks as a forcing function, rather than on a semantics developed bottom-up from action representations themselves. Delivering semantic representations that are grounded in the same sense that mechanisms developed over hundreds of millions of years of evolution is much harder. Steedman (2002b,a) argues that the combinators **B** and **T** that do most of the projective syntactic work in CCG are directly related to operations of seriation and affordance in the planner. This suggests that mechanisms for state-based reactive planning of the kind investigated by Petrick & Bacchus (2002, 2004) may offer a way towards a more distinctively action-based semantics for natural language (cf. Steedman 2006a, Geib & Steedman 2007).

6 Conclusion

This paper has argued that syntax is learned on the basis of preexisting semantic interpretations afforded by the situation of adult utterance, using a statistical model over a universal set of grammatical possibilities. The existence of the model itself helps the child to rapidly acquire a correct grammar even in the face of competing ambiguous semantics.

The fact that the onset of syntactically productive language at the end of the Piagetian sensory-motor develomental phase is accompanied by an explosion of advances in qualitatively different "operational" cognitive abilities suggests that the availability of language has a feedback effect, facilitating access to concepts that the child would not otherwise have access. Early work by Oléron (1953) and Furth (1961) on specific cognitive deficits concerning non-perceptually evident concepts arising in deaf children who had been linguistically deprived by being denied access to sign supports this view.

This means that Gleitman's (1990) influential suggestion that it is the availability of syntax that enables the child to "syntactically bootstrap" lexical entries for verbs (such as "think") that are not situationally evident is essentially correct. However, we have seen from the case of learning the verb "flee" in the face of competition from the meaning $chase'$ that it is the availability to the child of *a model of the relation between language-specific syntax and universal semantics* that makes this possible. It follows that the effects observed by Oléron and Furth, and Gleitman herself must have the character of *directing the child's attention* to alternatives that are available to them, but which they would otherwise overlook, by sheer force of Bayesian priors on the conditional probability $P(\sigma|\tau)$ of a syntactic category given a semantic type. In that sense, we should probably refer to this effect as "grammatical" boostrapping, since it is an effect that is both syntactic and semantic.

The theory presented here resembles the proposal of Fodor 1998 as developed in Sakas & Fodor (2001) and Niyogi (2006) in that it treats the acquisition of grammar as in some sense parsing with a universal "supergrammar". As in that proposal, both parameters and triggers are simply properties of the language-specific grammar itself—in their case, rules over independently learned parts of speech, in present terms, lexical categories.

It differs in assuming that the unordered logical form for the utterance is mostly available, with tolerable degrees of error and ambiguity. This means that the problem of syntactically ambiguous sentences to which STL is heir does not arise.

It also differs in the algorithm by which it converges on the target grammar. Rather than learning rules in an all or none fashion on the basis of unambiguous sentences that admit of only one analysis, it adjusts probabilities in a model of all elements of the grammar for which there is positive evidence for *all* processable utterances. In this respect, it more closely resembles the proposal of Yang (2002). However it differs from both in eschewing the view that grammar learning is parameter setting.

In equating language-specific grammar with a statistical model for parsing with universal grammar, the proposal bears an intriguing relation to the Maximum Spanning Tree (MST) parser (McDonald et al. 2005; McDonald & Pereira 2006b,a). This parser searches for the maximum-valued spanning tree-forming subgraph of a totally connected graph over the words of the string, using a perceptron-like maximum-margin discriminative model trained using pairs of strings and dependency trees. It has been applied to parsing "non-projective" or long-range dependencies, including crossing dependencies. It works best when the features over which the model is trained are grammar-like features such as position with respect to the verb, or morphological features. In particular, Çakıcı (2007) has shown that using CCG categories as features in a dependency-model of Turkish improves performance over the baseline in McDonald & Pereira (2006b). MST could therefore be seen as offering an alternative, discriminative, version of the present approach, according to which it could be used to learn weights for a language-specific set of features or categories drawn from a larger universal set.

If the parameters are implicit in the rules or categories themselves, and you can learn the rules or categories directly, why should the child or the theory bother with parameters at all? For the child, all-or-none parameter-setting is counterproductive, as it will make it hard to learn the many languages which have inconsistent settings of parameters across lexical types and exceptional lexical items, as in German and Dutch head finality, and English expressions like the following:

(24) Doggies galore!

Therefore, the fact that languages show violable tendencies to consistency for values of parameters like headedness across categories for related semantic types such as verbs and prepositions probably stems from considerations of overall encoding efficiency for the grammar as a whole, of the kind captured in notions like Minimal Description Length (MDL). Such considerations may be relevant to comparing entire grammars for the purpose of explaining language change, as in the work of Briscoe (2000). Their presence will under the present theory make the task of learning easier, by raising prior probabilities in the model for rules and categories that actually do recur. But it is less clear that representing them explicitly, rather than leaving them implicit in th model, will help the individual child learning a specific grammar, word-by-word.

References

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Information and Control*, *45*, 117–135.

Baldridge, J. (2002). Lexically specified derivational control in combinatory categorial grammar. Ph.D. thesis, University of Edinburgh.

Becker, M. (2005). Raising, control, and the subset principle. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, pp. 52–60, Somerville MA. Cascadilla Proceedings Project.

Berwick, R. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.

Bolinger, D. (1965). *Forms of English*. Cambridge, MA: Harvard University Press.

Briscoe, E. (2000). Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, *76*.

Çakıcı, R. (2007). Parser models for a highly inflected language. Ph.D. thesis, University of Edinburgh. in preparation.

Charniak, E. (2000). A maximum-entropy-inspired parser. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 132–139, Seattle, WA.

Clark, A. (2004). Grammatical inference and first language acquisition. In *CoLing Workshop on Psycho-computational Models of Human Language Acquisition*.

Clark, R. & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry*, *24*, 299–345.

Collins, M. (1997). Three generative lexicalized models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid*, pp. 16–23, San Francisco, CA. Morgan Kaufmann.

Crain, S. & Thornton, R. (1998). *Investigations in Universal Grammar*. Cambridge MA: MIT Press.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*, 657–667.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to infants. *Journal of Child Language*, *16*, 477–501.

Fisher, C. & Tokura, H. (1996). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, pp. 343–363. Erlbaum.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry*, *29*, 1–36.

Fodor, J. D. & Sakas, W. (2005). The subset principle in syntax: Costs of compliance. *Journal of Linguistics*, *41*, 513–569.

Furth, H. (1961). The influence of language ofn the development of concept formation in deaf children. *Journal of Abnormal and Social Psychology*, *63*, 386–389.

Geib, C. & Steedman, M. (2007). On natural language processing and plan recognition. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1612–1617.

Gerken, L. (1996). Prosodic structure in young children's language production. *Langauge*, *72*, 683–712.

Gerken, L., Jusczyk, P., & Mandel, D. (1994). When prosody fails to cue syntactic structure. *Cognition*, *51*, 237–265.

Gibson, E. & Wexler, K. (1995). Triggers. *Linguistic Inquiry*, *25*, 355–407.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 1–55.

Gorniak, P. & Roy, D. (2007). Situated language understanding as filtering perceived affordances. *Cognitive Science*, *31*. to appear.

Halliday, M. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Hockenmaier, J. & Steedman, M. (2002). Generative models for statistical parsing with Combinatory Categorial Grammar. In *Proceedings of the 40th Meeting of the ACL*, pp. 335–342, Philadelphia, PA.

Hyams, N. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Kelleher, J. D., Kruijff, G.-J. M., & Costello, F. J. (2006). Proximity in context: An empirically grounded computational model of proximity for processing topological spatial expressions. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 745–752, Sydney, Australia. Association for Computational Linguistics.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Mattys, S. & Juszyk, P. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91–121.

Mattys, S., Juszyk, P., Luce, P., & Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.

McDonald, R. & Pereira, F. (2006a). Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the 10th Conference on Natural Language Learning*, New Brunswick. ACL.

McDonald, R. & Pereira, F. (2006b). Online learning of approximate dependency parsing algorithms. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics,*, pp. 81–88, New Brunswick. ACL.

McDonald, R., Pereira, F., Ribarov, K., & Hajic, J. (2005). Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the Joint Conference on Human Laguage Technologies and Empirical Methods in Natural Language Processing HLT/EMNLP,*, pp. 523–530, New Brunswick. ACL.

McWhinnie, B. (2005). Item based constructions and the logical problem. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition. CoNNL-9*, pp. 53–68, New Brunswick. ACL.

Newell, A. (1973). You can't play twenty questions with nature and win. In W. Chase (ed.), *Visual Information Processing*, pp. 283–308. New York NY: Academic Press.

Niyogi, P. (2006). *Computational Nature of Language Learning and Evolution*. Cambridge MA: MIT Press.

Oléron, P. (1953). Conceptual thinking of the deaf. *American Annals of the Deaf*, *98*, 304–310.

Petrick, R. P. A. & Bacchus, F. (2002). A knowledge-based approach to planning with incomplete information and sensing. In M. Ghallab, J. Hertzberg, & P. Traverso (eds.), *Proceedings of the Sixth International Conference on Artificial Intelligence Planning and Scheduling (AIPS-2002)*, pp. 212–221, Menlo Park, CA. AAAI Press.

Petrick, R. P. A. & Bacchus, F. (2004). Extending the knowledge-based approach to planning with incomplete information and sensing. In S. Zilberstein, J. Koehler, & S. Koenig (eds.), *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS-04)*, pp. 2–11, Menlo Park, CA. AAAI Press.

Quine, W. v. O. (1960). *Word and Object*. Cambridge MA: MIT Press.

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, *167*, 170–205.

Sakas, W. & Fodor, J. D. (2001). The structural triggers learner. In S. Bertolo (ed.), *Language Acquisition and Learnability*, pp. 172–233. Cambridge: Cambridge University Press.

Siskind, J. (1995). Grounding language in perception. *Artificial Intelligence Review*, *8*, 371–391.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Steedman, M. (1996). The role of prosody and semantics in the acquisition of syntax. In J. Morgan & K. Demuth (eds.), *Signal to Syntax*, pp. 331–342. Hillsdale, NJ: Erlbaum.

Steedman, M. (2000a). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *34*, 649–689.

Steedman, M. (2000b). *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, M. (2002a). Formalizing affordance. In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society, Fairfax VA, August*, pp. 834–839, Mahwah NJ. Lawrence Erlbaum.

Steedman, M. (2002b). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, *25*, 723–753.

Steedman, M. (2006a). Origins of universal grammar in planned action. In M. Christiansen, C. Collins, & S. Edelman (eds.), *Language Universals*, pp. 143–152. Oxford: Oxford University Press. to appear.

Steedman, M. (2006b). Surface compositional semantics of intonation. *Submitted*.

Steedman, M. & Baldridge, J. (2006). Combinatory categorial grammar. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, vol. 2, pp. 610–622. Oxford: Elsevier, 2nd edn.

Steels, L. (2004). Constructivist development of grounded construction grammars. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 9–14, Barcelona.

Steels, L. & Baillie, J. C. (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, *43*, 163–173.

Thornton, R. & Tesan, G. (2006). Categorical acquisition: Parameter setting in universal grammar. *Submitted*.

Vallduví, E. & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, *34*, 459–519.

Villavicencio, A. (2002). The acquisition of a unification-based generalised categorial grammar. Ph.D. thesis, University of Cambridge.

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Yang, C. (2006). *The Infinite Gift*. New York NY: Scribner.

Zettlemoyer, L. & Collins, M. (2005). Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. Held in conjunction with IJCAI 2005, Edinburgh.