A Probabilistic Model of First Language Acquisition
Tom Kwiatkowski, Sharon Goldwater, Mark Steedman

Introduction
We present a universal model of syntactic acquisition that is initialised with a Universal CCG Grammar and then learns language specific syntax from sentences paired with contextually-allowed meaning-representations.

Meaning Representations
Each training sentence is paired with a single, correct, predicate-argument representation of meaning relations.

Sentence: Can you get the blocks out?
Meaning Representation: can(get(out(the(blocks))),you)

These Meaning Representations can be split into sub-parts using the lambda-calculus.

Universal Grammar
The Universal Grammar which the model is initialised is defined in terms of a mapping from semantic types to syntactic categories. Given a meaning-representation and sentence, all parses that project the meaning-representation onto the sentence are defined by this mapping along with the combiners of CCG.

CCG
CCG lexical items are < word : syntactic category : meaning > triples. For example, the English lexicon contains:

- the := NP/N : Ax . the( x )
- blocks := N : blocks

Lexical items are combined through the use of combinators that build the syntactic derivation and semantic interpretation of the sentence synchronously. The slash directions in CCG categories are used to control the semantic type to syntactic category. Given a sentence, the model builds a parse tree by combining lexical items using combinators of CCG.

From Sentence Meaning pairs to lexical items
CCG assumes a functional mapping from semantic type to possible syntactic categories. The meaning-representation in the pair below refers to an entity and is therefore given syntactic category NP. The combinators of CCG can be reversed to generate the full set of lexical items that could be used in building any given phrase-meaning pair.

Universal CCG lexicon.

Phrase: the blocks
Meaning: the ( blocks )

The logical form the is a place-holder for the presumed universal semantics of definites. The lexical items created here support determiner first languages (such as English), determiner final languages (such as Lakota) and languages that do not separately lexicalise these definites.

Probabilistic Parsing Model
The Universal CCG grammar supports many parses for each sentence meaning-representation pair in the training corpus. We use a generative model to assign probabilities to each of the parses.

We test the model by using it to predict the most probable meaning-representations for a set of unseen test sentences and then comparing these meaning-representation to a gold standard annotation.

Parameter estimation
Parameter estimation is done using online Variational Bayesian Expectation Maximisation. Updates are strictly online in that:

1. Each training example is seen sequentially.
2. Each training example is seen once and forgotten before the next training example is observed.

The Eve corpus
We use the Eve corpus of the CHILDEs database. This corpus is made up of child directed utterances and it has been tagged with dependency graphs (Sagae et al. 2007) of the sort given below.

Experiments
The model is trained on 3599 child-directed sentence meaning-representation pairs collected between the ages of 1:6 and 2:1 (years/months) from the first 13 files of the Eve corpus. Sentences of up to length 6 were used giving 10^10 word candidates for which the universal grammar licenses 2x10^9 possible lexical items. The test set is taken from files 14 and 15 of the Eve corpus, collected at age 2:1.

At test time the model is used to predict the most probable meaning-representation for each test sentence. This meaning representation is compared to a gold standard annotation in two ways. Exact match accuracy scores the prediction as correct if it exactly matches the gold standard. Partial match accuracy scores each of the directed, labelled dependency relations within the meaning-representations.

We compare to a baseline model which memoizes each of the sentence meaning-representation pairs seen in the training set.

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>f-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact-match</td>
<td>baseline</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>model</td>
<td>51</td>
<td>28</td>
</tr>
<tr>
<td>partial-match</td>
<td>baseline</td>
<td>100</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>model</td>
<td>67.5</td>
<td>61.9</td>
</tr>
</tbody>
</table>

This graph shows the probabilities of each of the 6 possible transitive-verb orders over the course of 2500 training examples. The correct SVO order is learned with a step-like transition in rule probability after a brief period in which OVS is favoured due to insufficient training data.