# Grammar Acquisition by Child and Machine:
## *The Combinatory Manifesto*

Mark Steedman

*with*

Jason Baldridge, Cem Bozsahin, Stephen Clark, James Curran, & Julia Hockenmaier

Informatics, University of Edinburgh

CoNNL *and* ACL PsychoCompLA Workshop, Ann Arbor MI, June 2005

`www.informatics.ed.ac.uk/~steedman/papers.html`

# 1. Introduction for PsychoComp: Statistical Parsing

- The only way that anyone has so far been able to induce reasonably sound, wide coverage, adult-sized grammars for realistic corpora, by machine, is via Supervised Learning, based on Human-annotated data, such as that in the Penn Wall Street Journal corpus (Collins 1999; Charniak 2000).

- The only way that anyone has been able to write programs that parse accurately using grammars of that size and consequent degree of ambiguity is by using statistical models based on the same labelled data, such as Head-Dependency models(Collins 1999) .

- Accuracy is still not that good. The grammars and models that you can extract from 1M words of labeled data are still small by human standards, BUT

- Unsupervised learning of grammars and models to within any specified margin of error, though possible in principle, in practice is hard to achieve for realistic cases.

# Intro. for CoNNL: Human Language Acquisition

- The only plausible source for the positive evidence that the child brings to bear on the of induction of grammar from strings is access to meaning representations.
- The only plausible source for negative evidence, *inasfar as it is needed at all, rather than showing that the theory of grammar is wrong*, is statistical properties of the corpus the child is exposed to
- There is clear evidence that human sentence processing of adult grammar relies on a model of semantic and pragmatic coherence for ambiguity resolution.

Such knowledge is “AI complete” and hard to model.

# Thesis

- This paper argues that these two fields have essentially the same problem, and can learn from each other.
- Computational Linguists should apply their learning and modeling techniques to linguistically expressive grammars that support semantic interpretation.
- Developmental Linguists should think about realistic—i.e. very big, very ambiguous—grammars, and machine learning.

# Antithesis

- In fact, these communities seem hardly to be talking to each other
- Linguists are quite sure that you cannot capture human grammar. with systems less than or equal in expressive power to context-free grammar.
  - In fact they have mostly stopped worrying about restricting the expressive power of their theories.
  - They think computational linguists have forgotten about long-range dependencies.
- Most computational linguists currently use grammars that are less than or equal in expressive power to context free grammars (because they want to learn them from data, and the only systems you can effectively learn by machine are finite-state.)
  - They think that linguists have forgotten about Zipf's law.
- Time for a **Synthesis**.

# 2. The Anatomy of a Parser

- Every parser can be identified by three elements:
  - A Grammar (Regular, Context Free, Linear Indexed, etc.) and an associated automaton (Finite state, Push-Down, Embedded Push-Down, etc.);
  - A search Algorithm characterized as left-to-right (etc.), bottom-up (etc.), and the associated working memories (etc.);
  - An Oracle, to resolve ambiguity.
- The oracle can be used in two ways, either to actively limit the search space, or in the case of an "all paths" parser, to rank the results.
- In wide coverage parsing, we have to use it in the former way.

# Head-dependencies as Oracle

- Head-dependency-Based Statistical Parser Optimization works because it approximates an oracle using semantics and real world inference.
- Its probably as close as we will get to the real thing for the foreseeable future.
- In fact, if suitable enhanced by associative Knowledge Representations and Contextual Model management. backing-off word dependencies to ontology-based hypernyms, etc., it may *be* the real thing.
- There is a strong case for generalizing the method to more expressive grammars.
- Many context-free processing techniques generalize to the mildly context sensitive class.
- The "nearly context free" grammars such as LTAG and CCG—the least expressive generalization of CFG known—are interesting candidates.

# CCG as Grammar

- CCG was invented to capture human language phenomena like coordination and long range dependency.
- Both phenomena are frequent in corpora, and are explicitly annotatedin the Penn WSJ corpus.
- Most treebank grammars ignore this information and fail to capture these phenomena entirely.
- Zipf's law says using it won't give us much better overall numbers. ( 5% of sentences in WSJ include long-range object dependencies, but LRoDs are only a small proportion of the dependencies in those sentences.)
- But there is a big difference between getting a perfect eval-b score on a sentence including an object relative clause and interpreting it!
- Making the generalization will give us: less under- and over-generalizing parsers, better models, and semantically interpretable output.

# Categorial Grammar

- (1) proves := $(S\backslash NP_{3s})/NP : \lambda x.\lambda y.prove'xy$
- (2) *Functional application*

  a. $X/_{\star}Y : f \quad Y : a \quad \Rightarrow \quad X : fa \qquad (>)$

  b. $Y : a \quad X\backslash_{\star}Y : f \quad \Rightarrow \quad X : fa \qquad (<)$

- For present purposes we can ignore modalities like $\star$ on slashes (Baldridge 2002), which limit the applicability of particular rules to particular categories.

(3)

$$\dfrac{\dfrac{\text{Marcel}}{NP_{3sf} : marcel'} \quad \dfrac{\dfrac{\text{proves}}{(S\backslash NP_{3s})/NP : \lambda x.\lambda y.prove'xy} \quad \dfrac{\text{completeness}}{NP : completeness'}}{S\backslash NP_{3s} : \lambda y.prove'completeness'y}>}{S : prove'completeness'marcel'}<$$

# Combinatory Categorial Grammar

(4) *Forward composition* ($>\mathbf{B}$)

$X/_{\diamond}Y : f \quad Y/_{\diamond}Z : g \quad \Rightarrow_{\mathbf{B}} \quad X/_{\diamond}Z : \lambda x.f(gx)$

(5) *Forward type-raising* ($>\mathbf{T}$)

$NP : a \quad \Rightarrow_{\mathbf{T}} \quad \mathrm{T}/_{\tau}(\mathrm{T}\backslash_{\tau}NP) : \lambda f.fa$

(6)

$$
\begin{array}{cccccc}
\text{Marcel} & \text{proved} & \text{and} & \text{I} & \text{disproved} & \text{completeness} \\
\hline
NP & (S\backslash NP)/NP & (X\backslash_{\star}X)/_{\star}X & NP & (S\backslash NP)/NP & NP \\
\hline
S/(S\backslash NP) \; {>}\mathbf{T} & & & S/(S\backslash NP) \; {>}\mathbf{T} & & \\
\hline
\multicolumn{2}{c}{S/NP \; {>}\mathbf{B}} & & \multicolumn{2}{c}{S/NP \; {>}\mathbf{B}} & \\
\hline
\multicolumn{5}{c}{S/NP \; {<}{>}} & \\
\hline
\multicolumn{6}{c}{S \; {>}}
\end{array}
$$

# Coordination and Relativization

- We correctly predict the fact that right-node raising is unbounded, as in a, below, and also provide the basis for an analyis of the similarly unbounded character of leftward extraction, as in b (see Steedman (1996, 2000) for details):

(7) a. [I conjectured]$_{S/NP}$ and [you think Marcel proved]$_{S/NP}$ completeness.

b. The result [which]$_{(N\backslash N)/(S/NP)}$ [you think Marcel proved]$_{S/NP}$.

- Other predictions include non-constituent coordination and parasitic gaps.

# These Things are Out There in the Treebank!

- Full Object Relatives ( 570 in WSJ treebank)
- Reduced Object Relatives ( 1070 in WSJ treebank)
- Argument Cluster Coordination ( 230 in WSJ treebank):

```
(S (NP-SBJ It)
   (VP (MD could)
       (VP (VP (VB cost)
               (NP-1 taxpayers)
               (NP-2 $ 15 million))
           (CC and)
           (VP (NP=1 BPC residents)
               (NP=2 $ 1 million)))))
```

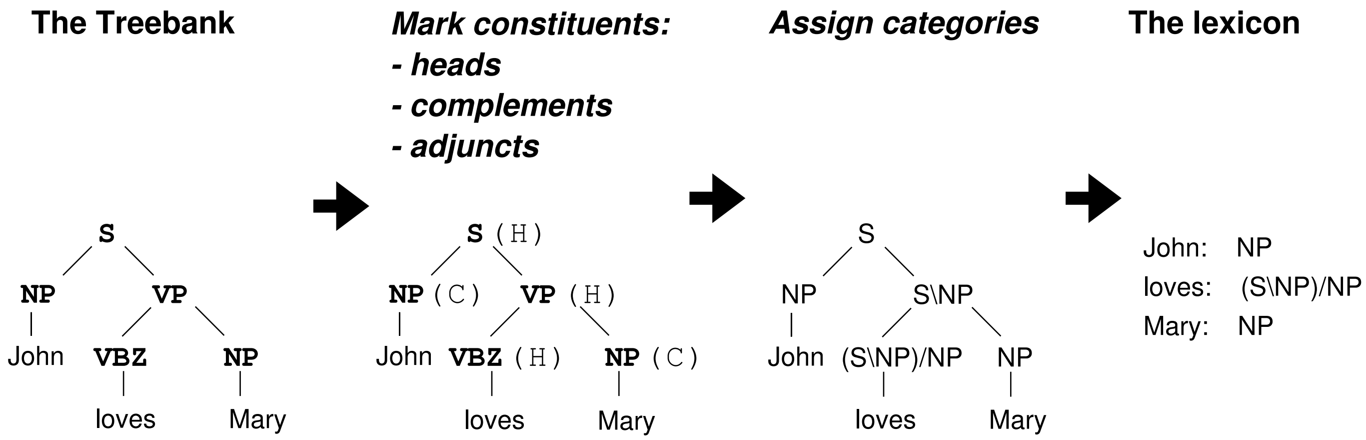
# These Things are Out There (contd.)

- Parasitic Gaps (at least 6 in WSJ treebank):

```
(S (NP-SBJ Hong Kong's uneasy relationship with China)
   (VP (MD will)
       (VP (VP (VB constrain)
               (NP (-NONE- *RNR*-1)))
           (PRN (: --)
                (IN though)
                (VP (RB not)
                    (VB inhibit)
                    (NP (-NONE- *RNR*-1)))
                (: --))
           (NP-1 long-term economic growth))))
```

# 3. Supervised CCG Induction by Machine

- Extract a CCG lexicon from the Penn Treebank: Hockenmaier and Steedman (2002a), Hockenmaier (2003) (cf. Buszkowski and Penn 1990; Xia 1999).




- This trades lexical types (500 against 48) for rules (around 3000 instantiated binary combinatory rule types against around 12000 PS rule types) with standard Treebank grammars.

# Supervised CCG Induction: Full Algorithm

- `foreach tree T:`

  ```
  preprocessTree(T);
  preprocessArgumentCluster(T);
  determineConstituentType(T);
  makeBinary(T);
  percolateTraces(T);
  assignCategories(T);
  treatArgumentClusters(T);
  cutTracesAndUnaryRules(T);
  ```

- The resulting treebank is somewhat cleaner and more consistent, and is offered for use in inducing grammars in other expressive formalisms. It was released in June 2005 by the Linguistic Data Consortium with documentation and can be searched using t-grep.

# 4. How the Child Induces a CCG Lexicon

- The child's problem is similar but a little harder.
  - They have unordered logical forms, not language-specific ordered derivation trees.
  - So 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).
- They do not seem to have to deal with a greater amount of error than the Penn WSJ treebank has (cf. McWhinnie, this conference).
  - But they may need to deal with situations which are ambiguous as to a number of logical forms.
  - And they need to be able to recover from temporary wrong lexical assignments.
  - And they need to be able to handle lexical ambiguity.

# Example

- The child has encountered a dog. Then she encounters more dogs.

(8) a. Child: (thinks:) $more' dog'$

b. Adult: "More doggies!"

c. Child's lexical candidates:

| | | |
|---|---|---|
| more | $:= NP/NP : \lambda x.x$ | doggies $:= NP/NP : \lambda x.x$ |
| more | $:= NP\backslash NP : \lambda x.x$ | doggies $:= NP\backslash NP : \lambda x.x$ |
| more | $:= NP/N : more'$ | doggies $:= NP/N : more'$ |
| more | $:= NP\backslash N : more'$ | doggies $:= NP\backslash N : more'$ |
| more | $:= N : dog'$ | doggies $:= N : dog'$ |
| more | $:= NP : more' dog'$ | doggies $:= NP : more' dog'$ |
| more doggies | $:= more' dog'$ | |

- Sometimes she gets it wrong, starting to use "doggies" to mean "more". But she soon corrects in the light of further evidence.

- Where $more' dog'$ came from is a different question—see Quine (1960).

# Computational Accounts: Siskind and Villavicencio

- Siskind (1995, 1996) and following him, Villavicencio (2002) offer computational models of this process, the latter explicitly using CCG.
- Both theories make strong assumptions about the association of words with elements of logical form.
- Both make strong assumptions about universally available parametrically specified rule- or category- types, the latter in the form of a type hierarchy
- Both deal with noise and homonymy probabilistically.
- Both do the learning in two stages:
  - Association of logical forms with words.
  - Induction of phrase structure rules (Siskind) or directional CCG categories (Villavicencio).
  - The first is reminiscent of alignment in MT data. The second is reminiscent of learning recursive phrasal MT rules by Chiang (2005).

# Computational Accounts: Zettlemoyer and Collins

- In CCG, there is no reason to separate the two processes of associating meaning and syntactic type. Zettlemoyer and Collins (UAI 2005) combine the two in a single pass in an algorithm which seems both simpler and more general.

- 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 (8) 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.

# Zettlemoyer and Collins (Contd.)

- 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.

# Zettlemoyer and Collins (Contd.)

- Because it allows multiword elements (MWE) to be lexical entries, it avoids the problem that two words which consistently collocate, like *want* and *to* fails to reveal which of them means $want'$ and which means $to'$. They can be learned as a single item *want to*
- So can idioms/MWEs 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 "triggers" distinct from reasonably short string-meaning pairs 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.

# 5. Statistical Models for Wide-Coverage Parsers

- There are two kinds of statistical models:
  - Generative models directly represent the probabilities of the rules of the grammar, such as the probability of the word *eat* being transitive, or of it taking a nounphrase headed by the word *integer* as object.
  - Discriminative models compute probability for whole parses as a function of the product of a number of weighted features, like a Perceptron. These features typically include those of generative models, but can be anything.

# Generative Models (Hockenmaier)

- A problem: standard generative models for the local dependencies characteristic of CFGs do not easily generalize to the reentrant dependencies generated by these more expressive grammars (Abney 1997).
- The generative model of Hockenmaier and Steedman 2002b only models probability for Collins-style local dependencies (although the parser *recovers* long range dependencies).
- It uses "Normal-form modeling", where the derivations in the CCG treebank that are modeled are those in which type-raising and composition are only used when there is no alternative.
- Hockenmaier (2003) showed that a sound full generative model is as possible for CCG grammars as for CFG.

# Log-Linear Discriminative CCG Parsing Models

- Features $f_i$ encode evidence indicating good/bad parses
- (9) $p(d|S) = \frac{1}{Z(S)} e^{\sum_i \lambda_i f_i(d,S)}$
- Use standard Maximum Entropy techniques to train a FSM "supertagger" Clark (2002) to assign CCG categories, multitagging ($n \approx 3$) at over 98% accuracy
- Then use a conditional log-linear model such as Maximum Entropy of **either**:
  - The derived structure or parse yield;
  - All derivations;
  - All derivations with Eisner (1996) Normal Form constraints.

# Clark and Curran (2003, 2004)

- Discriminative estimation via the limited-memory BFGS algorithm is used to set feature weights
- Estimation is computationally expensive, particularly for “all derivations”:
  - Beowulf cluster allows complete Penn Treebank to be used for estimation.
  - The fact that the supertagger is very accurate makes this possible.

# Overall Dependency Recovery

| | LP | LR | UP | UR | cat |
|---|---|---|---|---|---|
| Clark et al. 2002 | 81.9 | 81.8 | 90.1 | 89.9 | 90.3 |
| Hockenmaier 2003 | 84.3 | 84.6 | 91.8 | 92.2 | 92.2 |
| **Log-linear** | **86.6** | **86.3** | **92.5** | **92.1** | **93.6** |
| Hockenmaier (POS) | 83.1 | 83.5 | 91.1 | 91.5 | 91.5 |
| **Log-linear (POS)** | **84.8** | **84.5** | **91.4** | **91.0** | **92.5** |

Table 1: Dependency evaluation on Section 00 of the Penn Treebank

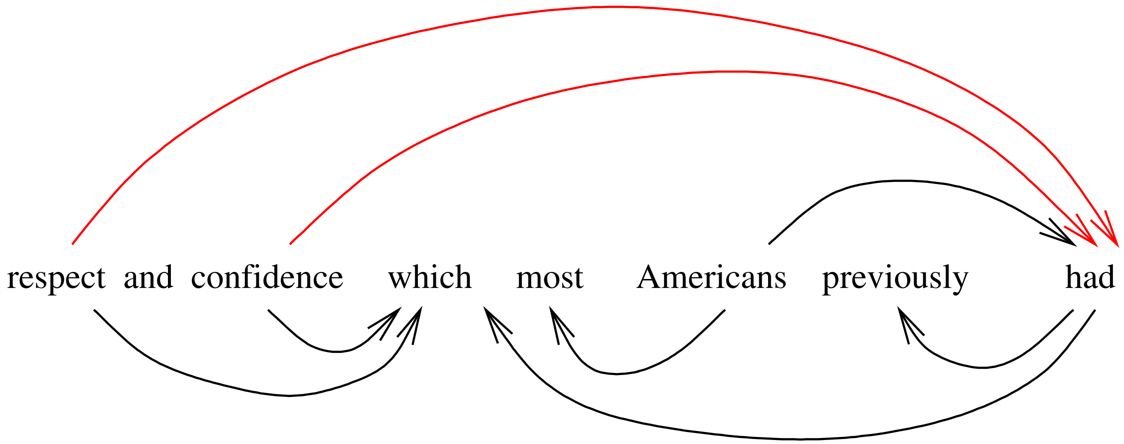
- To maintain comparability to Collins, Hockenmaier (2003) did not use a Supertagger, and was forced to use beam-search. With a Supertagger front-end, the Generative model might well do as well as the Log-Linear model. We have yet to try this experiment.

# Log-Linear Overall Dependency Recovery

- The C&C parser has state-of-the-art dependency recovery.
- The C&C parser is very fast ($\approx$ 30 sentences per second)
- The speed comes from highly accurate supertagging which is used in a "Best-First increasing" mode (Clark and Curran 2004), and behaves as an "almost parser" (Bangalore and Joshi 1999).
- It has been ported to the TREC QA task (Clark *et al.* 2004), and applied to the entailment QA task (Bos *et al.* 2004), using automatically built logical forms.

# Recovering Deep or Semantic Dependencies

Clark *et al.* (2002)




| lexical_item | category | slot | head_of_arg |
|---|---|---|---|
| *which* | $(NP_X \backslash NP_{X,1})/(S[dcl]_2/NP_X)$ | 2 | *had* |
| *which* | $(NP_X \backslash NP_{X,1})/(S[dcl]_2/NP_X)$ | 1 | *confidence* |
| *which* | $(NP_X \backslash NP_{X,1})/(S[dcl]_2/NP_X)$ | 1 | *respect* |
| *had* | $(S[dcl]_{had} \backslash NP_1)/NP_2)$ | 2 | *confidence* |
| *had* | $(S[dcl]_{had} \backslash NP_1)/NP_2)$ | 2 | *respect* |

# Full Object Relatives in Section 00

- 431 sentences in WSJ 2-21, 20 sentences (24 object dependencies) in Section 00.

1. Commonwealth Edison now faces an additional court-ordered *refund* on its summerwinter rate differential collections *that* the Illinois Appellate Court has *estimated* at DOLLARS.

2. Mrs. Hills said many of the 25 *countries that* she *placed* under varying degrees of scrutiny have made genuine progress on this touchy issue.

√ 3. It's the petulant complaint of an impudent *American whom* Sony *hosted* for a year while he was on a Luce Fellowship in Tokyo – to the regret of both parties.

√ 4. It said the *man*, *whom* it did not *name*, had been found to have the disease after hospital tests.

5. Democratic Lt. Gov. Douglas Wilder opened his gubernatorial battle with Republican Marshall Coleman with an abortion *commercial* produced by Frank Greer *that* analysts of every political persuasion *agree* was a tour de force.

6. Against a shot of Monticello superimposed on an American flag, an announcer talks about the strong *tradition* of freedom and individual liberty *that* Virginians have *nurtured* for generations.

√ 7. Interviews with analysts and business people in the U.S. suggest that Japanese capital may produce the economic *cooperation that* Southeast Asian politicians have *pursued* in fits and starts for decades.

8. Another was Nancy Yeargin, who came to Greenville in 1985, full of the *energy* and *ambitions that* reformers wanted to *reward*.

9. Mostly, she says, she wanted to prevent the *damage* to self-esteem *that* her low-ability students would *suffer* from doing badly on the test.

√ 10. Mrs. Ward says that when the cheating was discovered, she wanted to avoid the morale-damaging public *disclosure that* a trial would *bring*.

$\surd$ 11. In CAT sections where students' knowledge of two-letter consonant sounds is tested, the authors noted that Scoring High concentrated on the same *sounds that* the test *does* – to the exclusion of other *sounds that* fifth graders should *know*.

$\surd$ 12. Interpublic Group said its television programming *operations* – *which* it *expanded* earlier this year – agreed to supply more than 4,000 hours of original programming across Europe in 1990.

13. Interpublic is providing the programming in return for advertising *time*, *which* it *said* will be valued at more than DOLLARS in 1990 and DOLLARS in 1991.

$\surd$ 14. Mr. Sherwood speculated that the *leeway that* Sea Containers *has* means that Temple would have to substantially increase their bid if they're going to top us.

$\surd$ 15. The Japanese companies bankroll many small U.S. companies with promising products or ideas, frequently putting their money behind *projects that* commercial banks won't *touch*.

$\surd$ 16. In investing on the basis of future transactions, a role often performed by merchant banks, trading companies can cut through the *logjam that* small-company owners often *face* with their local commercial banks.

17. A high-balance *customer that* banks *pine for*, she didn't give much thought to the rates she was receiving, nor to the fees she was paying.

$\surd$ 18. The events of April through June damaged the *respect* and *confidence which* most Americans previously *had* for the leaders of China.

$\surd$ 19. He described the situation as an escrow *problem*, a timing *issue*, *which* he *said* was rapidly rectified, with no losses to customers.

$\surd$ 20. But Rep. Marge Roukema (R., N.J.) instead praised the House's acceptance of a new youth training wage, a *subminimum that* GOP administrations have *sought* for many years.

Cases of object extraction from a relative clause in 00; the extracted object, relative pronoun and verb are in italics; sentences marked with a $\surd$ are cases where the parser correctly recovers all object dependencies

# Clark *et al.* (2004): Object Relatives Error Analysis

- 24 cases of extracted object in Section 00 associated with object relative pronoun category $(NP_x \backslash NP_x)/(S[dcl]/NP_x)$
- 15/24 (62.5%) recovered with all dependencies correct (15/20 (75%) precision)
  - That is, with both noun attachment and rel_pronoun-verb dependency correct—comparable to 58.3%/67% labelled recall/precision by Hockenmaier 2003 and significantly better than Clark *et al.* (2002) 42% recall
  - 1 sentence (1) failed to parse at all (necessary category for seen verb *estimated* unseen in 2-21).
  - 5 were incorrect because wrong category assigned to relative pronoun, of which: in two (5, 9) this was only because again the necessary category for a seen verb was unseen in 2-21, and one (17) was incorrect because the POS tagger used for back-off labeled the entirely unseen verb incorrectly
  - 3 incorrect only because relative clause attached to the wrong noun

# 6. Can Wide-Coverage NLP Learn from Children?

- This performance is still bad by human standards.
- The main obstacle is that 1M words of annotated training data is not nearly enough,
- There are lots of words that never occur at all in the TreeBank at all.
  - This is a problem that the supertagger can help with. (In fact the front-end supertagger is already crucial to performance.)
- But a worse problem is words that *have* been seen, but *not with the necessary category*.
  - The answer to this problem is lexical smoothing: either
  - supervised, using morphology (!), ontologies, thesauri, machine-readable dictionaries, etc., or
  - unsupervised, using clustering, principal components analysis, and the like over unannotated text.

# Large-Volume Text as a Substitute for the World

- These techniques haven't worked very well. Can we understand how to do better from developmental linguists?
- Children learn by assocating words with semantic conceptual representations.
- This process is hard to simulate.
- But if language and the world are that closely coupled, maybe we can invert the process, to infer world-like structure from large volumes of text?
- The problem is that words are ambiguous, unlike the world, and they are ambiguous in different ways.
- This is one reason why simple clustering gives less than satisfactory results (especially for verbs).

# Unsupervised Lexical Smoothing

- One technique that is closer to the "text as universe" view, and seems very promising for CCG has been proposed by Brent (1993) and Deane (2003).
- The proposal relies on the fact that certain function-word contexts, such as _ *him a* and _ *through the* are highly diagnostic of certain classes of verb, like the *give* and *pass* classes.
- Because they are made up of function words, these n-grams are relatively frequent on the billion word scale, and can be used to detect unseen word-class pairings.
- Because they include pronouns and specifiers, they also offer a basis for building a simple head dependency model.
- In languages with rich inflection, and even in English, morphological analysis can perform a similar role.

# A Preliminary Experiment

- We know that all CCG category types have already been encountered in sections 2-21.
- Of the nine object relative dependencies in section 00 missed by the C&C parser discussed earlier, four arise because the necessary category for the head word is missing. Of these four, three have been seen in other inflected forms in sections 2-21, while one (*pine (for)*) is entirely unseen.
- All four were found in a lexicon extracted at Edinburgh from the English Gigaword corpus using function word regular expressions generated semiautomatically from the treebank lexical category types and morphological smoothing.
- A study of overall impact on dependency recovery is under way.
- The difficult part is estimating a *model* for the new lexical items.

# 7. Do Children Have Statistical Language Models?

- None of the computational models of language acquisition discussed here use parser models. This only works because the grammars are small and unambiguous enough for lexical lookup to almost fully determine parse. (Cf. Supertagger as "almost parser".)
- Realistic parsers need not only a grammar and an algorithm, but also an oracle. Is the child's oracle a head dependency model?
- Animals can compute statistical models.
- For example, rats in an environment with several food sources with different probabilities of paying off will visit low-payoff sites with frequency proportional to their probability of paying off (Empathy exercise: Why?)
- An *n*-gram model is the baseline hypothesis.

# 8. Synthesis

- Computational Linguists can afford the linguistic expressivity that is needed to build interpretable structure and still parse efficiently with wide coverage, using automatically-induced CCG lexicalized grammars and statistical head-dependency lexicalized model.
- Psychologists can model the child learning CCG lexicalized grammar in the face of moderate amounts of error, noise and ambiguity, probabilistically inducing a categorial lexicon from paired strings and contextually available meanings, without any other triggers, and without explicit parameter-setting.
- Linguists have nothing to lose but their Chains.

## References

Abney, Steven, 1997. "Stochastic Attribute-Value Grammars." *Computational Linguistics* 23:597–618.

Baldridge, Jason, 2002. *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Ph.D. thesis, University of Edinburgh.

Bangalore, Srinivas and Joshi, Aravind, 1999. "Supertagging: An Approach to Almost Parsing." *Computational Linguistics* 25:237–265.

Bos, Johan, Clark, Stephen, Steedman, Mark, Curran, James R., and Hockenmaier, Julia, 2004. "Wide-Coverage Semantic Representations from a CCG Parser." In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04), Geneva*. ACL, 1240–1246.

Brent, Michael, 1993. "From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax." *Computational Linguistics* 19:243–262.

Buszkowski, Wojciech and Penn, Gerald, 1990. "Categorial Grammars Determined from Linguistic Data by Unification." *Studia Logica* 49:431–454.

Charniak, Eugene, 2000. “A Maximum-Entropy-Inspired Parser.” In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*. Seattle, WA, 132–139.

Chiang, David, 2005. “A Hierarchical Phrase-Based Model for Statistical Machine Translation.” In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor MI*. ACL, 263–270.

Clark, Stephen, 2002. “A Supertagger for Combinatory Categorial Grammar.” In *Proceedings of the TAG+ Workshop*. Venice, 19–24.

Clark, Stephen and Curran, James R., 2003. “Log-Linear Models for Wide-Coverage CCG Parsing.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Sapporo, Japan, 97–104.

Clark, Stephen and Curran, James R., 2004. “Parsing the WSJ using CCG and Log-Linear Models.” In *Proceedings of the 42nd Meeting of the ACL*. Barcelona, Spain, 104–111.

Clark, Stephen, Hockenmaier, Julia, and Steedman, Mark, 2002. “Building Deep Dependency Structures with a Wide-Coverage CCG Parser.” In *Proceedings of the 40th Meeting of the ACL*. Philadelphia, PA, 327–334.

Clark, Stephen, Steedman, Mark, and Curran, James R., 2004. “Object-Extraction and Question-Parsing Using CCG.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain, 111–118.

Collins, Michael, 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

Deane, Paul, 2003. “Cooccurrence and Constructions.” In Luuk Lagerwerf, Wilbert Spooren, and Liesbeth Degand (eds.), *Determination of Information and Tenor in Texts: Multidisciplinary Approaches to Discourse*. Stichting Neerlandistiek VU Amsterdam, Münster: Nodus Publikationen, 277–304.

Eisner, Jason, 1996. “Efficient Normal-Form Parsing for Combinatory Categorial Grammar.” In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, Santa Cruz, CA*. San Francisco, CA: Morgan Kaufmann.

Hockenmaier, Julia, 2003. *Data and models for statistical parsing with CCG*. Ph.D. thesis, School of Informatics, University of Edinburgh.

Hockenmaier, Julia and Steedman, Mark, 2002a. “Acquiring Compact Lexicalized

Grammars from a Cleaner Treebank." In *Proceedings of the Third International Conference on Language Resources and Evaluation*. Las Palmas, Spain, 1974–1981.

Hockenmaier, Julia and Steedman, Mark, 2002b. "Generative Models for Statistical Parsing with Combinatory Categorial Grammar." In *Proceedings of the 40th Meeting of the ACL*. Philadelphia, PA, 335–342.

Quine, Willard van Ormond, 1960. *Word and Object*. Cambridge MA: MIT Press.

Siskind, Jeffrey, 1995. "Grounding Language in Perception." *Artificial Intelligence Review* 8:371–391.

Siskind, Jeffrey, 1996. "A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings." *Cognition* 61:39–91.

Steedman, Mark, 1996. *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.

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

Villavicencio, Aline, 2002. *The Acquisition of a Unification-Based Generalised Categorial Grammar*. Ph.D. thesis, University of Cambridge.

Xia, Fei, 1999. “Extracting Tree Adjoining Grammars from Bracketed Corpora.” In *Proceedings of the 5th Natural Language Processing Pacific Rim Symposium(NLPRS-99)*.

Zettlemoyer, Luke and Collins, Michael, 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.