Combinatory Categorial Grammars for Robust Natural Language Processing

Mark Steedman
with Jason Baldridge, Çem Boszahin, Ruken Çakıcı, Stephen Clark,
James Curran, Julia Hockenmaier, Tom Kwiatkowski, Mike Lewis, Jong Park
Emily Thomforde, Prachya Boonkwan, Greg Copolla, Sharon Goldwater,
Luke Zettlemoyer *and others*

June 2012

School of informatics

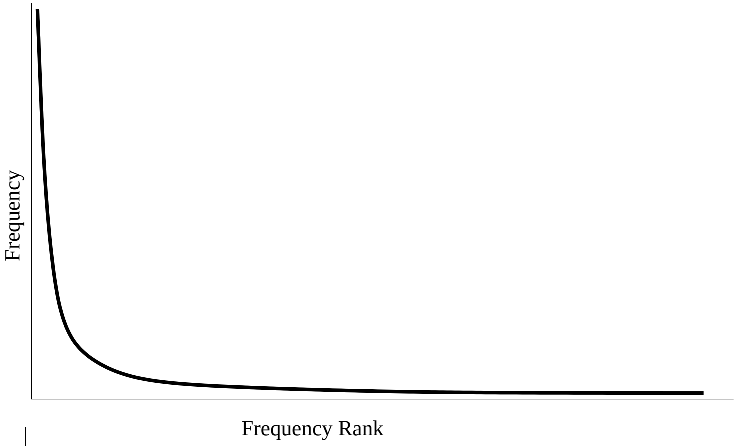
Prospectus

- Prologue: Why use CCG for NLP?
- I: Combinatory Categorial Grammar
- II: Wide Coverage Parsing with CCG
- III: The *Statistical* Problem of Language Acquisition
- IV: Towards a Robust Semantics
- V: The Surface Compositional Semantics of Intonation and Information Structure

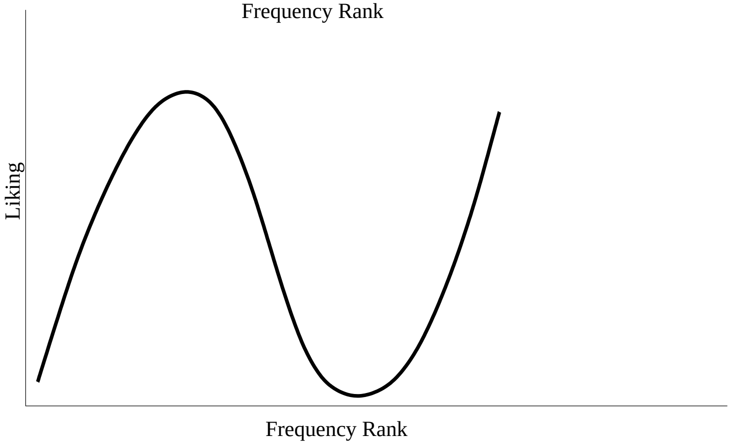
Prologue: Why Use CCG for NLP?

The Long Tail and the Uncanny Valley

- Zipf's Law:



- The Uncanny Valley:



☡ Ignoring the long tail can engender the uncanny:

In the Uncanny Valley

- TREC 2005:

Q77.6 Name opponents who Foreman defeated.
Q77.7 Name opponents who defeated Foreman.

- A QA Program (Kor 2005):

Opponents who Foreman defeated:
George Foreman
Joe Frazier
Ken Norton
Sonny
Archie Moore

Opponents who defeated Foreman:
George Foreman
Joe Frazier
Ken Norton
Sonny
Archie Moore

The Problem

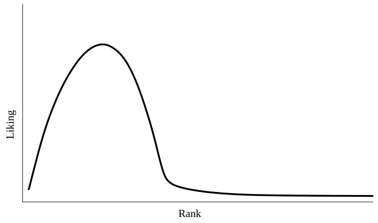
- The contribution of certain constructions to determining system acceptability is disproportionate to their low frequency.

- This is bad news.

- ☡ Machine learning is very bad at acquiring systems for which crucial information is in rare events.

The Darkling Plain

- If the distribution of event types really is a power law curve, then there is no other side to the Uncanny Valley accessible to brute-force machine learning.



- We shall see that, for certain categories of parser error, up to half the error rate is due to unseen grammatical event types (such as lexical entries), and up to half is due to unseen model tokens for seen types (such as head word dependencies).

- So the long tail is already hurting us badly.

What To Do

- The distribution of grammatical event types isn't a true power law, because there is a finite number of them, defined generatively, ultimately by a universal semantics.
- In principle, we can enumerate the types.
- ☡ But there are more constructions than you can shake a stick at (Goldberg 1995)
- Induce them from labeled data. (Or get linguists to enumerate them).
- If we knew what that semantics was, we might be able to solve the model problem as well.
- ☡ But none of the existing logicist semantic formalisms will do (MacCartney and Manning 2007).

How To Do It

- We need a readily extensible, construction-based grammar.
- It must be robustly and efficiently parsable with wide coverage
- It must be transparent to a “natural” semantics, supporting cheap inference.

I: Combinatory Categorial Grammar

Categorial Grammar

- Categorial Grammar replaces PS rules by lexical categories and general combinatory rules (**Radical Lexicalization**):

(1) ~~$S \rightarrow NP \quad VP$~~
~~$VP \rightarrow TV \quad NP$~~
~~$TV \rightarrow \{\textit{proved}, \textit{finds}, \ldots\}$~~

- Categories:

(2) proved := $(S\backslash NP)/NP$

(3) think := $(S\backslash NP)/_{\diamond}S$

Categorial Grammar

- **Categorial Grammar** replaces PS rules by lexical categories and general combinatory rules (**Lexicalization**):

(1) $$\xcancel{\begin{array}{lcl} S & \rightarrow & NP \quad VP \\ VP & \rightarrow & TV \quad NP \\ TV & \rightarrow & \{\mathit{proved},\ \mathit{finds},\ \ldots\} \end{array}}$$

- **Categories** with semantic intepretations:

(2) proved := $(S\backslash NP)/NP : \mathit{prove}'$

(3) think := $(S\backslash NP)/_{\diamond}S : \mathit{think}'$

Applicative CG Derivation = CFG

- (4)

$$\begin{array}{cccc}
\text{Marcel} & \text{proved} & \text{completeness} & \\
\cline{1-1}\cline{2-2}\cline{3-3}
NP & (S\backslash NP)/NP & NP & \\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP} & > \\
\cline{1-3}
\multicolumn{3}{c}{S} & < \\
\end{array}$$

(5)

$$\begin{array}{cccccc}
\text{I} & \text{think} & \text{Marcel} & \text{proved} & \text{completeness} & \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}
NP & (S\backslash NP)/_{\diamond}S & NP & (S\backslash NP)/NP & NP & \\
\cline{4-5}
 & & & \multicolumn{2}{c}{S\backslash NP} & > \\
\cline{3-5}
 & & \multicolumn{3}{c}{S} & < \\
\cline{2-5}
 & \multicolumn{4}{c}{S\backslash NP} & > \\
\cline{1-5}
\multicolumn{5}{c}{S} & < \\
\end{array}$$

Applicative CG Derivation = CFG

- (4) $$\begin{array}{cccl} \text{Marcel} & \text{proved} & \text{completeness} & \\ \cline{1-1}\cline{2-2}\cline{3-3} NP : marcel' & (S\backslash NP)/NP : prove' & NP : completeness' & \\ \cline{2-3} & \multicolumn{2}{c}{S\backslash NP : \lambda y.prove'completeness'y} & > \\ \cline{1-3} \multicolumn{3}{c}{S : prove'completeness'marcel'} & < \end{array}$$

(5) $$\begin{array}{ccccccl} \text{I} & \text{think} & \text{Marcel} & \text{proved} & \text{completeness} & \\ \cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5} NP : i' & (S\backslash NP)/_{\diamond}S : think' & NP : marcel' & (S\backslash NP)/NP : prove' & NP : completeness' & \\ \cline{4-5} & & & \multicolumn{2}{c}{S\backslash NP : \lambda y.prove'completeness'y} & > \\ \cline{3-5} & & \multicolumn{3}{c}{S : prove'completeness'marcel'} & < \\ \cline{2-5} & \multicolumn{4}{c}{S\backslash NP : think'(prove'completeness'marcel')} & > \\ \cline{1-5} \multicolumn{5}{l}{S : think'(prove'completeness'marcel')i'} & < \end{array}$$

Combinatory Categorial Grammar (CCG)

- Combinatory Rules:

Application :

$$\frac{X/_{\star}Y \qquad Y}{X}> \qquad \frac{Y \qquad X\backslash_{\star}Y}{X}<$$

Composition :

$$\frac{X/_{\diamond}Y \qquad Y/_{\diamond}Z}{X/_{\diamond}Z}>\mathbf{B} \qquad \frac{Y\backslash_{\diamond}Z \qquad X\backslash_{\diamond}Y}{X\backslash_{\diamond}Z}<\mathbf{B}$$

Crossed Composition :

$$\frac{X/_{\times}Y \qquad Y\backslash_{\times}Z}{X\backslash_{\times}Z}>\mathbf{B}_{\times} \qquad \frac{Y/_{\times}Z \qquad X\backslash_{\times}Y}{X/_{\times}Z}<\mathbf{B}_{\times}$$

- All arguments are type-raised in the (morpho) lexicon:

Type Raising :

$$\frac{X}{\mathrm{T}/(\mathrm{T}\backslash X)}>\mathbf{T} \qquad \frac{X}{\mathrm{T}\backslash(\mathrm{T}/X)}<\mathbf{T}$$

- We omit a further family of rules based on the combinator **S**

Combinatory Categorial Grammar (CCG)

- Combinatory Rules:

Application :

$$\frac{X/_{\star}Y\colon f \quad Y\colon g}{X\colon f(g)}> \qquad \frac{Y\colon g \quad X\backslash_{\star}Y\colon f}{X\colon f(g)}<$$

Composition :

$$\frac{X/_{\diamond}Y\colon f \quad Y/_{\diamond}Z\colon g}{X/_{\diamond}Z\colon \lambda z.f(g(z))}>\mathbf{B} \qquad \frac{Y\backslash_{\diamond}Z\colon g \quad X\backslash_{\diamond}Y\colon f}{X\backslash_{\diamond}Z\colon \lambda z.f(g(z))}<\mathbf{B}$$

Crossed Composition :

$$\frac{X/_{\times}Y\colon f \quad Y\backslash_{\times}Z\colon g}{X\backslash_{\times}Z\colon \lambda z.f(g(z))}>\mathbf{B}_{\times} \qquad \frac{Y/_{\times}Z\colon g \quad X\backslash_{\times}Y\colon f}{X/_{\times}Z\colon \lambda z.f(g(z))}<\mathbf{B}_{\times}$$

- All arguments are type-raised in the (morpho) lexicon:

Type Raising :

$$\frac{X\colon x}{\mathrm{T}/(\mathrm{T}\backslash X)\colon \lambda f.f(x)}>\mathbf{T} \qquad \frac{X\colon x}{\mathrm{T}\backslash(\mathrm{T}/X)\colon \lambda f.f(x)}<\mathbf{T}$$

- We omit a further family of rules based on the combinator **S**

Slash Typing

- The features $\star, \diamond, \times$ were introduced by Baldridge 2002 following Hepple (1987)
- They form a lattice

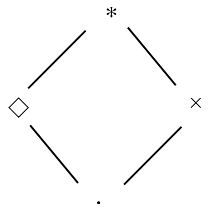
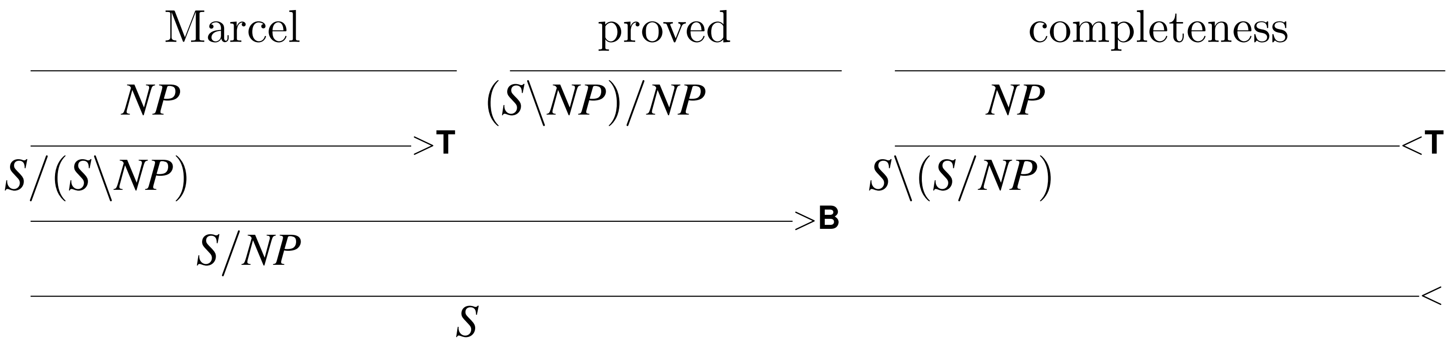


Figure 1: CCG type hierarchy for slash modalities (Baldridge and Kruijff 2003)

- $\cdot$ type written as bare slash e.g. α/β means any rule can apply
- $\diamond$ type e.g. $\alpha/_{\diamond}\beta$ means any rule except $\times$ can apply.
- $\times$ type e.g. $\alpha/_{\times}\beta$ means any rule except $\diamond$ can apply.
- $\star$ type e.g. $\alpha/_{\star}\beta$ means no rule except $\star$ can apply.

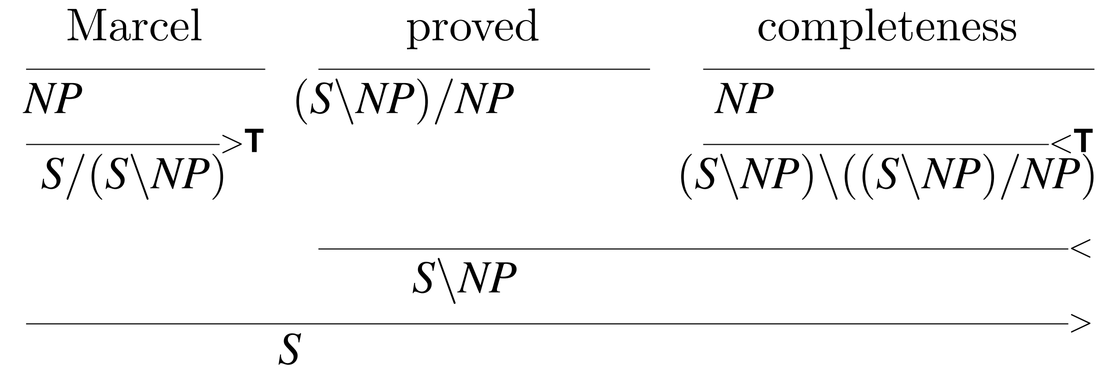
Combinatory Derivation

(6)



$$\begin{array}{lll} \text{Marcel} & \text{proved} & \text{completeness} \\ \hline NP & (S\backslash NP)/NP & NP \\ \hline S/(S\backslash NP) \;\; >\mathbf{T} & & S\backslash(S/NP) \;\; <\mathbf{T} \\ \hline \multicolumn{2}{l}{S/NP \;\; >\mathbf{B}} & \\ \hline \multicolumn{3}{l}{S \;\; <} \end{array}$$

(7)



$$\begin{array}{lll} \text{Marcel} & \text{proved} & \text{completeness} \\ \hline NP & (S\backslash NP)/NP & NP \\ \hline S/(S\backslash NP) \;\; >\mathbf{T} & & (S\backslash NP)\backslash((S\backslash NP)/NP) \;\; <\mathbf{T} \\ \hline & \multicolumn{2}{l}{S\backslash NP \;\; <} \\ \hline \multicolumn{3}{l}{S \;\; >} \end{array}$$

Combinatory Derivation

(6)

$$\begin{array}{ccc}
\text{Marcel} & \text{proved} & \text{completeness}\\
\hline
NP : \mathit{marcel'} & (S\backslash NP)/NP : \mathit{prove'} & NP : \mathit{completeness'}\\
\hline
\quad\quad\quad\quad >\mathbf{T} & & \quad\quad\quad\quad <\mathbf{T}\\
S/(S\backslash NP) : \lambda f.f\ \mathit{marcel'} & & S\backslash(S/NP) : \lambda p.p\ \mathit{completeness'}\\
\hline
\multicolumn{2}{r}{>\mathbf{B}} & \\
\multicolumn{2}{c}{S/NP : \lambda x.\mathit{prove'x\ marcel'}} & \\
\hline
\multicolumn{3}{r}{<}\\
\multicolumn{3}{c}{S : \mathit{prove'completeness'marcel'}}
\end{array}$$

(7)

$$\begin{array}{ccc}
\text{Marcel} & \text{proved} & \text{completeness}\\
\hline
NP : \mathit{marcel'} & (S\backslash NP)/NP : \mathit{prove'} & NP : \mathit{completeness'}\\
\hline
>\mathbf{T} & & <\mathbf{T}\\
S/(S\backslash NP) & & (S\backslash NP)\backslash((S\backslash NP)/NP)\\
: \lambda f.f\ \mathit{marcel'} & & : \lambda p.p\ \mathit{completeness'}\\
\hline
 & \multicolumn{2}{r}{<}\\
 & \multicolumn{2}{c}{S\backslash NP : \lambda y.\mathit{prove'completeness'y}}\\
\hline
\multicolumn{3}{r}{>}\\
\multicolumn{3}{c}{S : \mathit{prove'completeness'marcel'}}
\end{array}$$

Case as Type-Raising

- The type-raising combinator **T** is directly related to case systems, as in Latin:

(8) a. Balb*us* ambulat. b. Livia Balb*um* amat. c. Livia Balb*o* murum dabit.
“Balbus walks.” “Livia loves Balbus.” “Livia gave Balbus a wall.”

- This involves the following fragment of Latin lexicon:

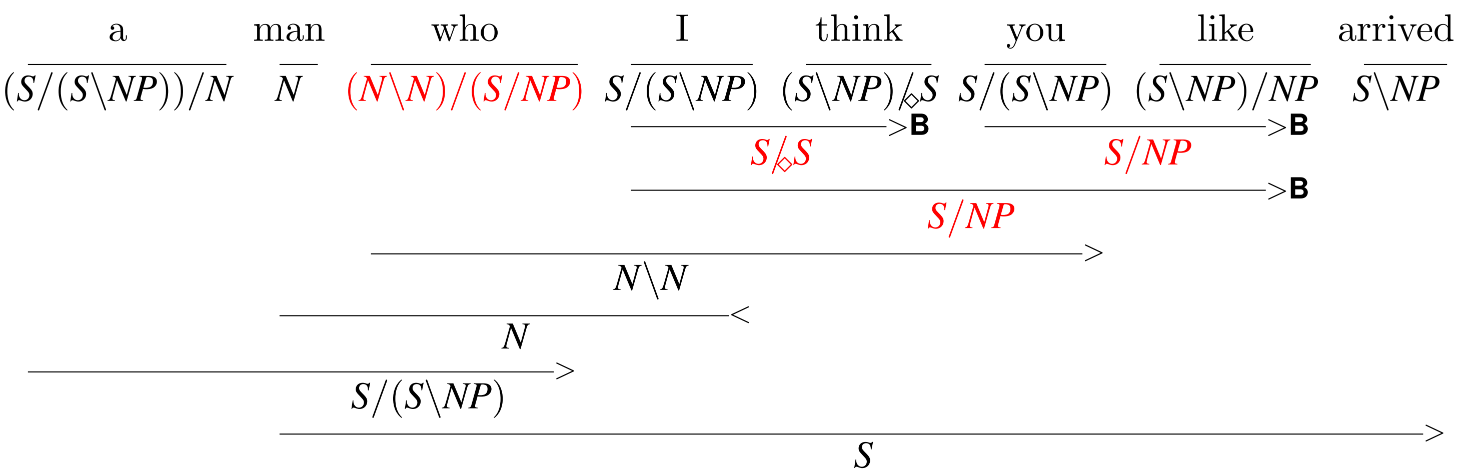
$$(9)\ \left\{\begin{array}{lll} \text{Balb+us} & := & S/(S\backslash NP) : \lambda p_{(e,t)}.p\ balbus' \\ \text{Balb+um} & := & S/(S/NP)\lambda p_{(e,(e,t))}\lambda y.p\ balbus'y \\ & := & (S\backslash NP)/((S\backslash NP)/NP) : \lambda p_{(e,(e,t))}\lambda y.p\ balbus'y \\ \text{Balb+o} & := & ((S\backslash NP)/NP)/(((S\backslash NP)/NP)/NP) : \lambda p_{(e,(e,(e,t)))}\lambda y\lambda z.p\ balbus'yz \\ & & \&c. \end{array}\right\}$$

- Even English possesses a case system in this sense. It’s just very ambiguous.
- We will sometimes schematize cased/raised forms as $NP^{\uparrow}$

Linguistic Predictions: Unbounded “Movement”

- The combination of type-raising and composition allows derivation to project lexical function-argument relations onto “unbounded” constructions such as relative clauses and coordinate structures, without transformational rules:

(10)



a man who I think you like arrived

$(S/(S\backslash NP))/N$ N $(N\backslash N)/(S/NP)$ $S/(S\backslash NP)$ $(S\backslash NP)/_{\diamond}S$ $S/(S\backslash NP)$ $(S\backslash NP)/NP$ $S\backslash NP$

$>\mathbf{B}$: $S/_{\diamond}S$

$>\mathbf{B}$: S/NP

$>\mathbf{B}$: S/NP

$>$: $N\backslash N$

$<$: N

$>$: $S/(S\backslash NP)$

$>$: S

Linguistic Predictions: Constraints on “Movement”

- We also predict the following assymetry without stipulation, since we can allow a without allowing b, but can’t allow b without also getting c:

(11) a. a man who(m) [I think that]$_{S/S}$ [Keats likes]$_{S/NP}$
b. *a man who(m) [I think that]$_{S/S}$ [likes Keats]$_{S\backslash NP}$
c. *[I think]$_{S/S}$ Chapman [that]$_{S/S}$ [likes Keats]$_{S\backslash NP}$.

- In transformational (MP) terms, CCG reduces MOVE to MERGE.

Predictions: English Intonation

- A minimal pair of contexts and contours:

(12) Q: I know who proved soundness. But who proved COMPLETENESS?
A: (MarCEL)(proved COMPLETENESS).
H*L L+H* LH%

(13) Q: I know which result Marcel PREDICTED. But which result did Marcel PROVE?
A: (MARcel PROVED)(COMPLETENESS).
L+H* LH% H* LL%

- Crossing contexts and responses yields complete incoherence.
- In transformational (MP) terms, CCG reduces intonational phrases to surface constituents.

Predictions: Argument-Cluster Coordination

- The following construction is predicted on arguments of symmetry.

(14)
$$\begin{array}{cccccc}
\text{give} & \text{a teacher} & \text{an apple} & \text{and} & \text{a policeman} & \text{a flower} \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}\cline{6-6}
DTV & TV\backslash DTV \;{<}\mathbf{T} & VP\backslash TV \;{<}\mathbf{T} & (X\backslash_{\star}X)/_{\star}X & TV\backslash DTV \;{<}\mathbf{T} & VP\backslash TV \;{<}\mathbf{T} \\
\cline{2-3}\cline{5-6}
 & \multicolumn{2}{c}{VP\backslash DTV \;{<}\mathbf{B}} & & \multicolumn{2}{c}{VP\backslash DTV \;{<}\mathbf{B}} \\
\cline{4-6}
 & & & \multicolumn{3}{c}{(VP\backslash DTV)\backslash_{\star}(VP\backslash DTV) \;{>}} \\
\cline{2-6}
 & \multicolumn{5}{c}{VP\backslash DTV \;{<}} \\
\hline
\multicolumn{6}{c}{VP \;{<}}
\end{array}$$

—**where** $VP = S\backslash NP$; $TV = (S\backslash NP)/NP$; $DTV = ((S\backslash NP)/NP)/NP$.

- A variant like the following cannot occur in an SVO language like English:

(15) *A policeman a flower and give a teacher an apple.

Syntax = Type-Raising and Composition

- CCGs combination of type-raising and composition yields a permuting and rebracketing string calculus closely tuned to the needs of natural grammar.
- The argument cluster coordination construction (14) is an example of a universal tendency for "deletion under coordination" to respect basic word order: in all languages, if arguments are on the left of the verb then argument clusters coordinate on the left, if arguments are to the right of the verb then argument clusters coordinate to the right of the verb (Ross 1970):

(16) SVO: *SO and SVO SVO and SO
VSO: *SO and VSO VSO and SO
SOV: SO and SOV *SOV and SO

- In transformational (MP) terms, CCG reduces COPY/DELETE to MERGE.

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)))))
```

- It could cost taxpayers 15 million and __ BPC residents 1 million

These Things are Out There (contd.)

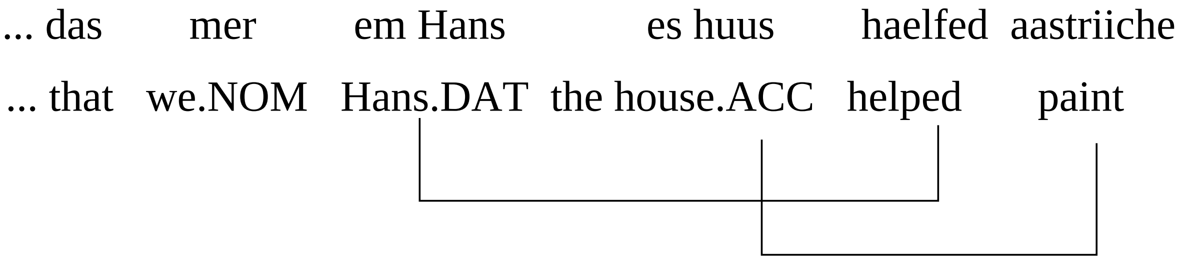
- Parasitic Gaps (at least 6 in WSJ treebank):

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

- HK's relation with C will constrain _, though not inhibit _, long-term growth.

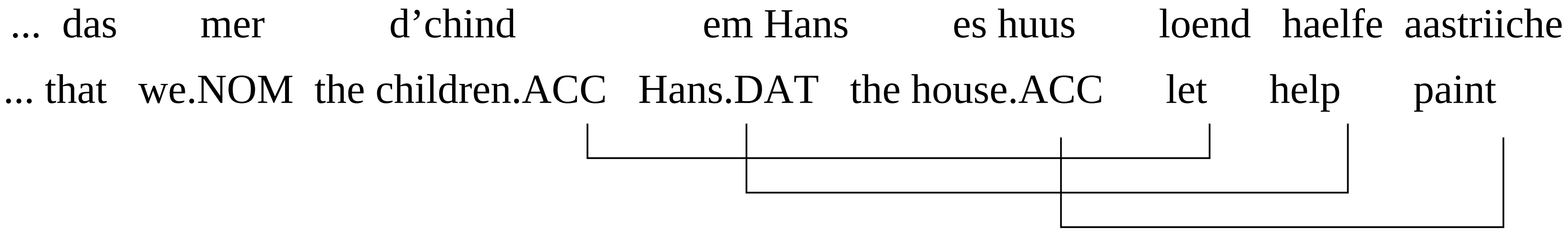
A Trans-Context Free Natural Language

- CCG can capture unboundedly crossed dependencies in Dutch and Zurich German (examples from Shieber 1985):



... das	mer	em Hans	es huus	haelfed	aastriiche
... that	we.NOM	Hans.DAT	the house.ACC	helped	paint

'... that we helped Hans paint the house.'

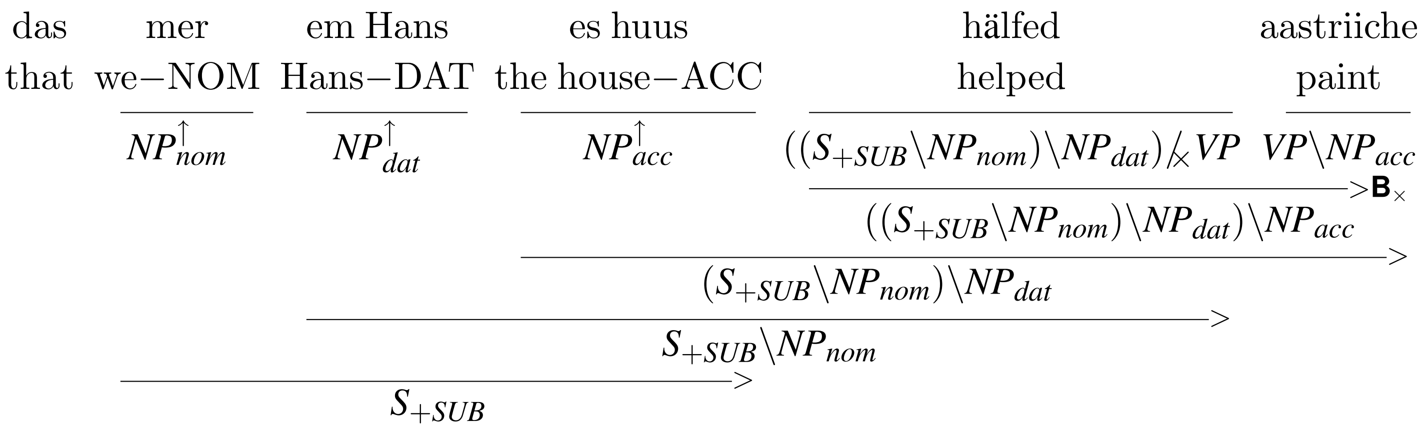


... das	mer	d'chind	em Hans	es huus	loend	haelfe	aastriiche
... that	we.NOM	the children.ACC	Hans.DAT	the house.ACC	let	help	paint

'... that we let the children help Hans paint the house.'

A Trans Context-Free CCG Analysis

(17)



das mer em Hans es huus hälfed aastriiche

that we−NOM Hans−DAT the house−ACC helped paint

$NP^{\uparrow}_{nom}$ $NP^{\uparrow}_{dat}$ $NP^{\uparrow}_{acc}$ $((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})/_{\times}VP$ $VP\backslash NP_{acc}$

$$((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc} \quad >\mathbf{B}_{\times}$$

$$(S_{+SUB}\backslash NP_{nom})\backslash NP_{dat} \quad >$$

$$S_{+SUB}\backslash NP_{nom} \quad >$$

$$S_{+SUB} \quad >$$

“that we helped Hans paint the house”

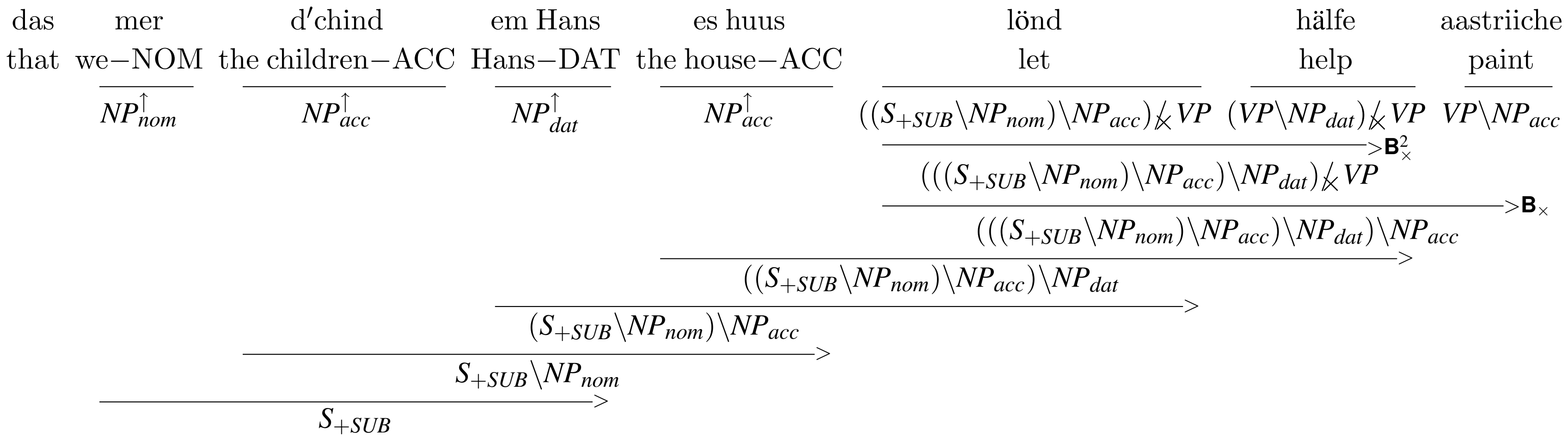
- The following alternative word order is correctly also allowed:

(18) Das mer em Hans hälfed es huus aastriiche.

◇ The corresponding word order is *disallowed* in the related Dutch construction.

A Trans Context-Free CCG Analysis

(18)



das / that

mer / we−NOM: $NP^{\uparrow}_{nom}$

d′chind / the children−ACC: $NP^{\uparrow}_{acc}$

em Hans / Hans−DAT: $NP^{\uparrow}_{dat}$

es huus / the house−ACC: $NP^{\uparrow}_{acc}$

lönd / let: $((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})/_{\times}VP$

hälfe / help: $(VP\backslash NP_{dat})/_{\times}VP$

aastriiche / paint: $VP\backslash NP_{acc}$

$$\xrightarrow{>\mathbf{B}^2_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/_{\times}VP$$

$$\xrightarrow{>\mathbf{B}_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}$$

$$\xrightarrow{>} ((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat}$$

$$\xrightarrow{>} (S_{+SUB}\backslash NP_{nom})\backslash NP_{acc}$$

$$\xrightarrow{>} S_{+SUB}\backslash NP_{nom}$$

$$\xrightarrow{>} S_{+SUB}$$

“that we let the children help Hans paint the house”

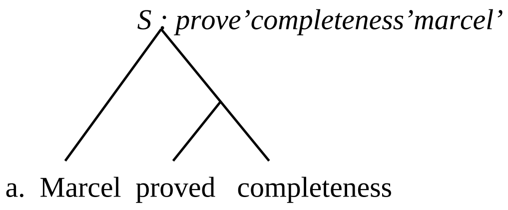
- Again, other word orders are correctly allowed.

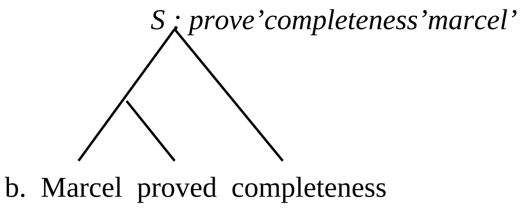
On So-called "Spurious" Ambiguity

- Examples like (78), and (14), embody the claim that fragments like "Marcel proved", and "a policeman a flower", are *constituents*, comparable to "proved completeness".

- If "Marcel proved" can be constituent in right node raising, then it can be a constituent of a canonical transitive sentence.

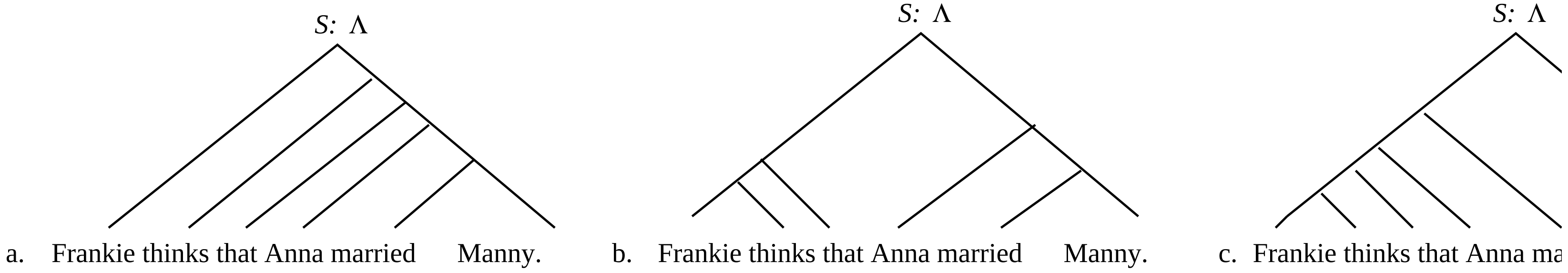
- Even such simple sentences are *derivationally ambiguous*:





On So-called “Spurious” Ambiguity (Contd.)

- More complex sentences are multiply ambiguous:



- This has been referred to (misleadingly) as “Spurious” ambiguity, since all the derivations have the same interpretation Λ.

- Interestingly, so called “spurious” constituents include many **left prefixes**.

- Left prefixes are relevant to language modeling for purposes like ASR.

Parsing in the Face of “Spurious Ambiguity”

- **All** grammars exhibit derivational ambiguity—even CFG.

- **Any** grammar that captures coordination at all will have the **same** derivational ambiguity as CCG.

- Use standard table-driven parsing methods such as CKY, with packed charts, either:
  - checking identity of **underlying** representation of table entries (Steedman 2000b), rather than identity of derivation, or:
  - parsing normal-form derivations (Eisner 1996)

CCG is “Nearly Context-Free”

- CCG and TAG are provably weakly equivalent to Linear Indexed Grammar (LIG) Vijay-Shanker and Weir (1994).
- Hence they are not merely “Mildly Context Sensitive” (Joshi 1988), but rather “Nearly Context Free,” or “Type $1.\dot{9}$” in the Extended Chomsky Hierarchy.

Language Type	**Automaton**	**Rule-types**	**Exemplar**
Type 0: RE	Universal Turing Machine	$\alpha \rightarrow \beta$	
Type 1: CS	Linear Bound Automaton (LBA)	$\phi A \psi \rightarrow \phi \alpha \psi$	$\mathscr{P}(a^n b^n c^n)$ (?)
I	Nested Stack Automaton(NSA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi C_{[(i),...]} \xi$	a^{2^n}
LCFRS (MCS)	ith-order EPDA	$A_{[[(i),...]...]} \rightarrow \phi B_{[[(i),...]...]} \psi$	$a^n b^n c^n \ldots m^n$
LI/CCG/TAG	Embedded PDA (EPDA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi$	$a^n b^n c^n$
Type 2: CF	Push-Down Automaton (PDA)	$A \rightarrow \alpha$	$a^n b^n$
Type 3: FS	Finite-state Automaton (FSA)	$A \rightarrow \begin{cases} \text{a B} \\ \text{a} \end{cases}$	a^n

CCG is Nearly Context-Free (contd.)

- It has polynomial parsing complexity (Vijay-Shanker and Weir 1990)

- Hence it has nice "Divide and Conquer" algorithms, like CKY, and Dynamic Programming.

- For real-life sized examples like parsing the newspaper, such algorithms must be statistically optimized.

Processing

- CCG was widely expected to be completely useless for processing, because of so-called "spurious" derivational ambiguity
- However, any theory that covers the range of grammatical phenomena covered by CCG has the same ambiguity.
- Moreover, everyone has to use a probabilistic parsing model to limit search arising from standard ambiguity, so who cares about a little more?
- Near context free power guarantees polynomial parsability hence applicability of standard algorithms.
- CCG supports one of the fastest and most accurate wide coverage parsers (Clark and Curran 2004; Auli and Lopez 2011) (with the bonus of capturing unbounded dependencies.)

CKY

(19)
```
1. for $j := 1$ to $n$ do
   begin
   $t(j, j) := \{A | A$ is a lexical category for $a_j\}$
2. for $i := j - 1$ down to $0$ do
   begin
3. for $k := i$ down to $0$ do
   begin
   $t(k, j) := pack\{A |$ for all $B \in t(k, i), C \in t(i + 1, j)$ end
                      such that $B\ C \Rightarrow A$ for some
                      combinatory rule in $R$
                      and $admissible(B\ C \Rightarrow A)\}$
   end
   end
```

CKY

- The procedure *pack* packs all categories A in the chart entry $t(k, j)$ with the same syntactic type Σ_A but different logical forms Λ_A into a single disjunctive structure-sharing entry

- The boolean function *admissible* stands for one of a number of possible conditions on the inclusion of A in the chart entry $t(k, j)$ that are necessary to keep the algorithm polynomial.

II: Wide-Coverage Parsing with CCG

Human and Computational NLP

- No handwritten grammar ever has the coverage that is needed to read the daily newspaper.

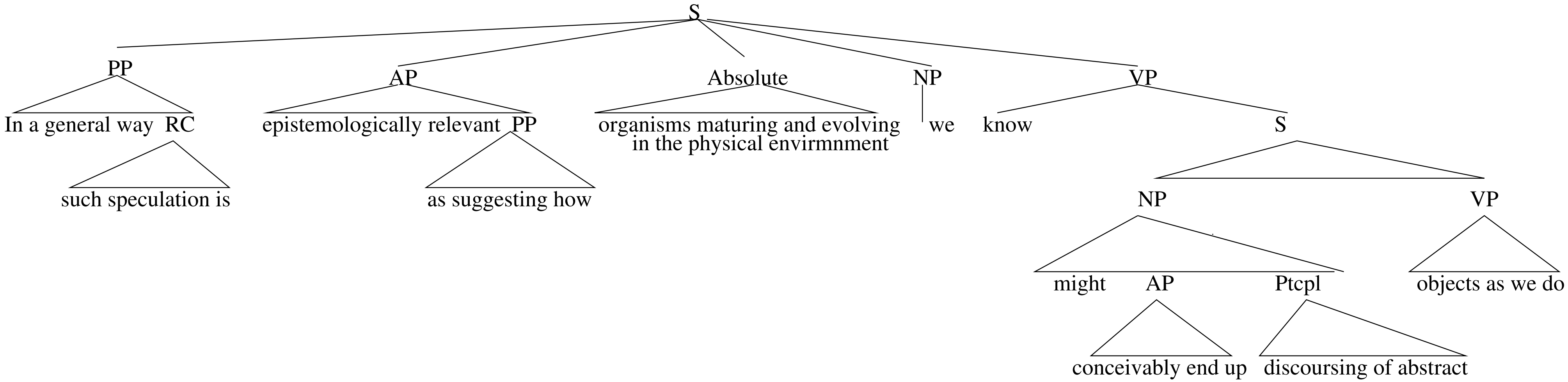
- Language is syntactically highly ambiguous and it is hard to pick the best parse. Quite ordinary sentences of the kind you read every day routinely turn out to have hundreds and on occasion thousands of parses, albeit mostly semantically wildly implausible ones.

- High ambiguity and long sentences break exhaustive parsers.

For Example:

- “In a general way such speculation is epistemologically relevant, as suggesting how organisms maturing and evolving in the physical environment we know might conceivably end up discoursing of abstract objects as we do.” (Quine 1960:123).

- —yields the following (from Abney 1996), among many other horrors:



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, Nested Push-Down, etc.);
  - A search Algorithm characterized as left-to-right (etc.), bottom-up (etc.), and the associated working memories (etc.);
  - A Model, to resolve ambiguity.

- The model 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 mostly use it the former way.

Competence and Performance

- Linguists (Chomsky 1957, *passim*), have always insisted on the methodological independence of “Competence” (the grammar that linguists study) and “Performance” (the mechanisms of language use).

- This makes sense: there are many more parsers than there are grammars.

- Nevertheless, Competence and Performance must have evolved as a single package, for what evolutionary edge does a parser without a grammar have, or a grammar without a parser?

- Any theory that does not allow a one-to-one relation between the grammatical and derivational constituency has some explaining to do.

Human Sentence Processing

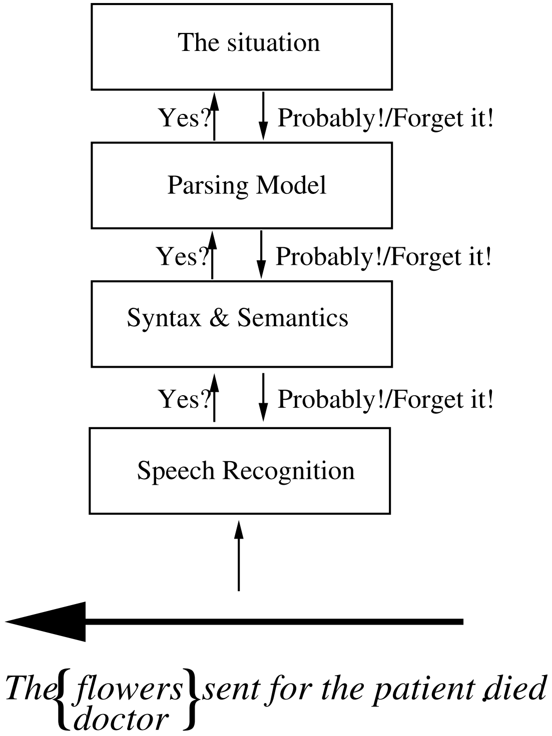
- “Garden path” sentences are sentences which are grammatical, but which naive subjects fail to parse.
- Example (20a) is a garden path sentence, because the ambiguous word “sent” is analysed as a tensed verb:

(20) a. # The doctor sent for the patient died.
b. The flowers sent for the patient died.

- However (20b) is not a garden path.
- So garden path effects are sensitive to world knowledge (Bever 1970).
- They are even sensitive to referential context: (Altmann and Steedman 1988) showed that (simplifying somewhat) if a context is established with two doctors, one of whom was sent for a patient, then the garden path effect is reversed.

The Architecture of the Human Sentence Processor

- This requires a “cascade” architecture:



The $\left\{\begin{matrix}\textit{flowers}\\ \textit{doctor}\end{matrix}\right\}$ *sent for the patient died*

Grammar and Incrementality

- Most left prefix substrings of sentences are typable constituents in CCG, for which alternative analyses can be compared using the parsing model

- The fact that (21a,b) involve the nonstandard constituent [The doctor sent for]$_{S/NP}$, means that constituent is also available for (21c,d)

(21) a. The patient that [the doctor sent for]$_{S/NP}$ died.
b. [The doctor sent for]$_{S/NP}$ and [The nurse attended]$_{S/NP}$ the patient who had complained of a pain.
c. #[The doctor sent for]$\left\{\begin{array}{lll} S/NP & & \\ (S/(S\backslash NP))/N & N & (N\backslash N)/NP \end{array}\right\}$[the patient]$_{NP}$ died$_{S\backslash NP}$.
d. [The flowers sent for]$\left\{\begin{array}{lll} \#S/NP & & \\ (S/(S\backslash NP))/N & N & (N\backslash N)/NP \end{array}\right\}$[the patient]$_{NP}$ died$_{S\backslash NP}$.

- (22) a. #[The doctor sent for the patient]$_{S}$ died$_{S\backslash NP}$.
b. [The flowers sent for the patient]$_{S\backslash NP}$ died$_{S\backslash NP}$.

The Strict Competence Hypothesis

- Since the spurious constitutent [#The flowers sent for]$_{S/NP}$ is available in the chart, so that its low probability in comparison with the probabilities of the unreduced components can be detected (according to some "figure of merit" (Charniak *et al.* 1998) discounting the future), the garden path in (20b) is avoided, even under the following very strong assumption about the parser:
  - *The Strict Competence Hypothesis*: the parser only builds structures that are licensed by the Competence Grammar as typable *constituents*.
- This is an attractive hypothesis, because it allows the Competence Grammar and the Performance Parser/Generator to evolve as a package deal, with parsing completely transparent to grammar, as in standard bottom-up algorithms.
- But is such a simple parser possible? We need to look at some real-life parsing programs.

Wide Coverage Parsing: the State of the Art

- Early attempts to model parse probability by attaching probabilities to rules of CFG performed poorly.
- Great progress as measured by the ParsEval measure has been made by combining statistical models of headword dependencies with CF grammar-based parsing (Hindle and Rooth 1993; Collins 1997; Charniak 2000)
- However, the ParsEval measure is very forgiving. Such parsers have until now been based on highly overgenerating context-free covering grammars. Analyses depart in important respects from interpretable structures.
- In particular, they typically fail to represent the long-range “deep” semantic dependencies that are involved in relative and coordinate constructions.

Head-dependencies as Model

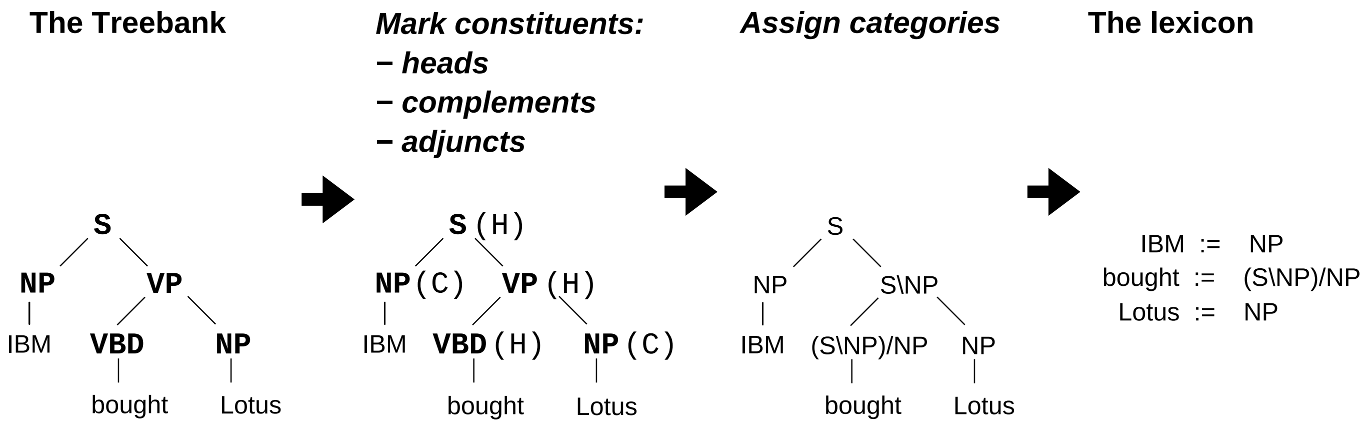
- Head-dependency-Based Statistical Parser Optimization works because it approximates an oracle using real-world knowledge.
- In fact, the knowledge- and context- based psychological oracle may be much more like a probabilistic relational model augmented with associative epistemological tools such as typologies and thesauri and associated with a dynamic context model than like traditional logicist semantics and inferential systems.
- Many context-free processing techniques generalize to the “mildly context sensitive” grammars.
- The “nearly context free” grammars such as LTAG and CCG—the least expressive generalization of CFG known—have been treated by Xia (1999), Hockenmaier and Steedman (2002a), and Clark and Curran (2004).

Nearly Context-Free Grammar

- Such Grammars capture the deep dependencies associated with coordination and long range dependency.
- Both phenomena are frequent in corpora, and are explicitly annotated in the Penn WSJ corpus.
- Standard 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. (around 3% of sentences in WSJ include long-range object dependencies, but those dependencies 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!

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).



The Treebank

Mark constituents:
– heads
– complements
– adjuncts

Assign categories

The lexicon

IBM := NP
bought := (S\NP)/NP
Lotus := NP

- 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);
```

CCGbank: Hockenmaier and Steedman 2007

- The trees in CCG-bank are (Rightward-Branching Normalized) CCG derivations, and in cases like Argument Cluster Coordination and Relativisation they depart radically from Penn Treebank structures.

- 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.

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.

- Both have been applied to CCG parsing

Generative Models (Hockenmaier)

- A problem: standard generative models for the local dependencies characteristic of CFGs do not immediately 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 it can *recover* long range dependencies).
- It uses "Normal-form modeling", where the derivations 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 mildly context sensitive grammars as it is for CFG.
- Log Linear models offer another solution (Clark and Curran 2003, 2004, and see below)

Hockenmaier 2002/2003: Overall Dependency Recovery

- Hockenmaier and Steedman (2002b)

		Parseval				Surface dependencies	
Model	LexCat	LP	LR	BP	BR	$\langle PHS \rangle$	$\langle\rangle$
Baseline	87.7	72.8	72.4	78.3	77.9	81.1	84.3
HWDep	92.0	81.6	81.9	85.5	85.9	84.0	90.1

- Collins (1999) reports 90.9% for unlabeled $\langle\rangle$ “surface” dependencies.

- CCG benefits greatly from word-word dependencies.
  (in contrast to Gildea (2001)’s observations for Collins’ Model 1)

- This parser is available on the project webpage.

Long Range Dependencies (Hockenmaier 2003)

- Extraction:
  - Dependencies involving subject relative pronoun
    $(\mathbf{NP}\backslash\mathbf{NP})/(\mathbf{S}[\mathbf{dcl}]\backslash\mathbf{NP})$: 98.5%LP, 95.4%LR (99.6%UP, 98.2%UR)
  - Lexical cat. for embedded subject extraction (Steedman 1996b)
    $((\mathbf{S}[\mathbf{dcl}]\backslash\mathbf{NP})/\mathbf{NP})/(\mathbf{S}[\mathbf{dcl}]\backslash\mathbf{NP})$: 100.0%P, 83.3%R
  - Dependencies involving object relative pronoun (including ES)
    $(\mathbf{NP}\backslash\mathbf{NP})/(\mathbf{S}[\mathbf{dcl}]/\mathbf{NP})$: 66.7%LP, 58.3%LR (76.2%UP, 58.3%UR)
- Coordination:
  - VP coordination (coordination of $\mathbf{S}[.]\backslash\mathbf{NP}$): 67.3%P, 67.0%R
  - Right-node-raising (coordination of $(\mathbf{S}[.]\backslash\mathbf{NP})/\mathbf{NP}$): 73.1%P, 79.2%R

Log-Linear Conditional CCG Parsing Models

- Features f_i encode evidence indicating good/bad parses
- (23) $p(d|S) = \frac{1}{Z(S)} e^{\sum_i \lambda_i f_i(d,S)}$
- Use standard Maximum Entropy techniques to train a FSM "supertagger" Clark (2002) to assign CCG categories, **multitagging ($n \approx 3$) at over 98% accuracy** (Clark and Curran 2003, 2004).
- Clark and Curran use a conditional log-linear model such as Maximum Entropy of **either**:
  - The derived structure or parse yield;
  - All derivations;
  - All derivations with Eisner Normal Form constraints.

Conditional CCG Parsing Models (Contd.)

- 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
Clark and Curran 2004	**86.6**	**86.3**	**92.5**	**92.1**	**93.6**
Hockenmaier (POS)	83.1	83.5	91.1	91.5	91.5
C&C (pos)	**84.8**	**84.5**	**91.4**	**91.0**	**92.5**

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

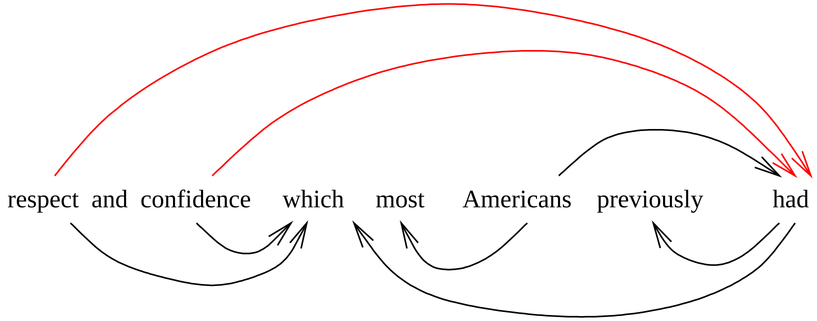
- To maintain comparability to Collins, Hockenmaier (2003) did not use a Supertagger, and was forced to use beam-search. With a Supertagger front-end, the Generative model might well do as well as the Log-Linear model. We have yet to try this experiment.

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 an aggressive "Best-First increasing" mode (Clark and Curran 2004), and behaves as an "almost parser" (Bangalore and Joshi 1999)
- Clark and Curran 2006 show that CCG all-paths almost-parsing with supertagger-assigned categories loses only 1.3% dependency-recovery F-score against parsing with a full dependency model
- C&C has been ported to the TREC QA task (Clark *et al.* 2004) using a hand-supertagged question corpus, and applied to the entailment QA task (Bos *et al.* 2004), using automatically built logical forms.

Recovering Deep or Semantic Dependencies

Clark *et al.* (2004)



respect and confidence which most Americans previously had

lexical_item	category	slot	head_of_arg
which	$(\mathsf{NP}_X\backslash\mathsf{NP}_{X,1})/(\mathsf{S}dcl_2/\mathsf{NP}_X)$	2	*had*
which	$(\mathsf{NP}_X\backslash\mathsf{NP}_{X,1})/(\mathsf{S}dcl_2/\mathsf{NP}_X)$	1	*confidence*
which	$(\mathsf{NP}_X\backslash\mathsf{NP}_{X,1})/(\mathsf{S}dcl_2/\mathsf{NP}_X)$	1	*respect*
had	$(\mathsf{S}dcl_{had}\backslash\mathsf{NP}_1)/\mathsf{NP}_2)$	2	*confidence*
had	$(\mathsf{S}dcl_{had}\backslash\mathsf{NP}_1)/\mathsf{NP}_2)$	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.

✓ 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.

✓ 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.

✓ 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.

✓ 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.

✓ 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.

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

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

✓ 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 associated with the object relative pronoun category $(NP_X\backslash NP_X)/(S[dcl]/NP_X)$;
- The extracted object, relative pronoun and verb are in italics; sentences marked with a ✓ are cases where the parser correctly recovers all object dependencies

Clark *et al.* (2004): Full Object Relatives

- 24 cases of extracted object in Section 00:
- 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

Clark *et al.* (2004): Free Relatives

- 14/17 (82%) recall 14/15 (93%) precision for the single dependency.

- Better performance on long-range dependencies can be expected with more features such as regular expressions for Max Ent to work on.

- Other varieties of deep dependency (Control, subject relatives, reduced relatives) discussed in Hockenmaier (2003); Clark *et al.* (2002, 2004).

- It looks as though about half the errors arise because the lexicon is too small, and about half because the head-dependency model is too weak.

- 1M words of treebank is nothing like enough data

Experiments with Porting the Parser

- As with all treebank grammars, almost any practical application involves porting the parser to a different grammar and model.
- For example, in ongoing experiments with open domain question answering, we would like to use the parser for parsing the questions.
- However, all treebank grammars including this one do appallingly badly on the TREC question database, because WSJ contains almost no direct questions, and none at all of some common patterns.
- Hand-labelling data for retraining is usually not possible.
- However, semi-automatically hand-supertagging a few thousand sentences and retraining the supertagger with those included is quite practical.
- Clark *et al.* 2004 did the 1,171 *What* questions from TREC in a week.

Porting to Questions: Results

- 171 *What*-question development set. 1000 for training (and testing using tenfold cross-validation), average length 8.6 words.
- Since the gold standard question data is only labelled to the level of lexical category we can only evaluate to that level.
- However, supertagger accuracy and sentence accuracy correlate very highly with dependency and category recall by the parser, and we know we need around 97% per word and 60% per sentence for the original WSJ performance
-

MODEL	1 CAT ACC	SENT ACC	1.5 cats /word	SENT ACC
CCGbank	72.0	1.8	84.8	11.1
Qs	92.3	66.7	96.6	80.7
Qs+CCGbank	93.1	61.4	98.1	86.5

Table 2: Accuracy of Supertagger on Development set Question Data

Porting to Questions: Results

Supertagging/ parsing method	CAT Acc	Sent Acc	What Acc
Increasing av. cats	94.6	81.8	91.2
Decreasing av. cats	89.7	65.3	80.0
Increasing cats (rand)	93.4	79.4	88.2
Decreasing cats (rand)	64.0	9.4	21.2
Baseline	68.5	0.0	60.6

Table 3: Category accuracy of parser on dev question data

- For the *What* object questions, per word/sentence accuracies were 90%/71%, suggesting that they are harder than the average question.
- Object dependency recall by the parser for these questions was 78%.

Applications: CCG Parsers as Language Models

- Standard technique/baseline is Trigram modeling, *strikingly akin to Elman's Simply Recurrent Networks*.
- Strict left-to-right parsing interpolated with trigram model does better: Chelba and Jelinek (1998); Roark (2001).
- Immediate-Head parser modeling alone does even better, even with a non-left-to-right algorithm: Charniak (2001).

CCG Parsers as Language Models

- CCG type-raising treats head and complement as **dual**: In some sense, it makes **all** constructions head first.
- Hence many left prefixes are constituents, even in Dutch/German/Japanese.
- While any grammar can in principle be mapped onto a prefix grammar with a generative model (Jelinek and Lafferty 1991), CCG already *is* (nearly) a prefix grammar and probabilities for prefix dependency structures can be derived from the standard dependency model.
- CCG similarly offers a direct way to use prosodic information (Steedman 2000a). (see (77) and (78), above, and cf. Charniak 2001).

CCG Parsers as Language Models

- For example, in Dutch the prefix *dat Cecilia een hond een knok . . .* ("that Cecilia a dog a bone . . . ") has (via type-raising and composition) a category $S/(((S\backslash NP)\backslash NP)\backslash NP)$.
- The type of this constituents tells you how to *invert* the dependency model to obtain a left-to-right prediction.
- It predicts a ditransitive verbgroup and tells you all you need to know to estimate its Arg Max from verbs of that class. (For example, the "give" stem is going to come out ahead of the "sell" stem.)
- *dat een hond een knok Cecilia . . .* is going to make a quite different predictions.
- So are some of the alternative derivations of *dat Cecilia een hond een knok . . .*

Aside: Can You Do this with Lambek Proof Nets?

- Yes—its just like making a DCG into a PDCG with a head-dependency model. You can even use our lexicon.
- But its not clear that its the same enterprise.
- For reasons of theoretical and computational economy, it seems *very odd* to relegate word order to post-hoc Linear Precedence (LP) rules, as in some Linear Logic and proof-net -based generalizations of the Lambek calculus.

Aside: Can You Do this with Proof Nets?

- First, categories *already* express linear order.
- Second, multiple occurrences of words so treated induce factorially many spurious derivations:
  - Police police police police police.
  - This is the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built.
- So string position must be part of the *resource*, not extra logical.

Aside: Can You Do this with Proof Nets?

- But string positions are what makes DCGs, Lambek, and all linguistic grammars and parsers linear (in both senses of the term) in the first place—so is *that* what we are axiomatizing?

- This seems to be what is implied by talk of “compiling” Linear Logic grammars into more standard formalisms like TAG.

- But to anyone who cares about actually parsing, that seems to be an odd way of thinking about it. When we’re computing arithmetic, we don’t use Peano’s axioms, or even think of our operations as “compiled” Peano Arithmetic.

Where do we Go from Here?

- 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 smoothing can help with
- But a worse problem is words that *have* been seen, but *not with the necessary category*.
- The only answer to this problem is to generalize the grammar and the model, using
  - Active learning over unreliable parser output from unlabeled data, or
  - High precision low recall methods over web-scale amounts of data.

Moral

- You can have the linguistic expressivity that is needed to build interpretable structure *and* parse efficiently with wide coverage—with an automatically induced CCG lexicon and a statistical head-dependency model

III: The Statistical Problem of Language Acquisition

- Combinatory Categorial Grammar (CCG) and Surface Compositional Semantics (Steedman 2012b)
- CCG-based Induction of “Semantic Parsers” for GeoQueries, ATIS, etc. datasets (Kwiatkowski *et al.* 2010, 2011).
- Semantic Parsing as a model of Child Language Acquisition.
- Results from using the CHILDES “Eve” dataset for learning (Kwiatkowski *et al.* 2012).
- Incorporating Information Structure.
- Comparisons with Fodor, Yang, etc. Parameter Setting.

Inducing Semantic Parsers with CCG

- Thompson and Mooney (2003); Zettlemoyer and Collins (2005, 2007); Wong and Mooney (2007); Lu *et al.* (2008); Börschinger *et al.* (2011); Liang *et al.* (2011) generalize the problem of inducing parsers from language-specific treebanks like WSJ to that of inducing parsers from paired sentences and unaligned language-independent logical forms.
  - The sentences might be in any language.
  - The logical forms might be database queries, λ-terms, robot action primitives, etc.

- This is a harder problem: in the worst case, we would have to consider all possible pairings of all possible substrings of the sentence with all possible subtrees of the logical form.

Inducing Semantic Parsers: GeoQuery

- Most of these programs invoke (English) language-specific assumptions.
- Kwiatkowski *et al.* (2010, 2011) have applied a more language-general approach to the problem of inducing multilingual grammars from the GeoQueries database of sentence meaning pairs (Thompson and Mooney, 2003):
  - Which states border states through which the mississippi traverses?
  - $\lambda x.\exists y[state(x) \land state(y) \land loc(mississippi\_river, y) \land next\_to(x, y)]$
- GeoQuery is all about *wh*-dependency
- Model is discriminative (log-linear), learned by batch mode inside-outside EM using stochastic gradient descent, iterated, evaluated by 10-fold cross-validation.
- Learning is accelerated by inititialization with Giza++ alignment between strings and logical forms.

Inducing CCG Semantics Parsers: GeoQuery 250

- % of unseen test sentences parsed correctly by induced grammars:

	UBL-s	λ-WASP	Lu08
English	**81.8**	75.6	72.8
Spanish	**81.4**	80.0	79.2
Japanese	**83.0**	81.2	76.0
Turkish	**71.8**	68.8	66.8

- This is done without the language-specific engineering of the other approaches. Constraints on splits are universal (e.g. ATB, A-over-A, semantic-syntactic types mapping).
- See Kwiatkowski *et al.* (2011) for effect of factored lexical generalization, and competitive results on the much harder ATIS travel bookings dataset.

II: Child and Computer Language Development

- The child's problem is similar to the problem of inducing a semantic parser (Siskind 1992; Villavicencio 2002, 2011; Buttery 2006).
    - Children too have unordered logical forms in a universal language of thought, not language-specific ordered WSJ trees.
    - So they too have to work out which words (substrings) 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).

- A word may correspond to any substructure of the logical form

Child and Computer Language Development

- Children do not seem to have to deal with a greater amount of illformedness than that in the Penn WSJ treebank (MacWhinney 2005).
  - But they need to learn big grammars.
  - They are faced with contexts which support irrelevant logical forms.
  - They need to be able to recover from temporary wrong lexical assignments.
  - And they need to be able to handle serious amounts of lexical ambiguity.

The Statistical Problem of Language Acquisition

- 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 a universal Combinatory Projection Principle, and a mapping from semantic types to the set of all universally available lexical syntactic types.

- Once the lexicon is learned, CCG will handle unbounded projection for free.

- In CCG, all dependencies are projective—even so-called “non-projective” ones.

Indefinitely Expandable Model

- At the start of the learning process the child does not have access to the scope of the final grammar or lexicon.

- We need to model an indefinitely expandable set of *grammar rules* and *lexical items*.

- This is done using Dirichlet priors.

- All unseen rules and lexical items are drawn from geometric base distributions.

The Algorithm

- Variational Bayes (cf. Sato 2001; Hoffman *et al.* 2010)
- Incremental Two-stage Expectation/Maximisation Algorithm
- The intuition:
  - Compute the probabilities of all analyses of the new sentence on the basis of the previous model.
  - Update the model on the basis of weighted counts of events in the new sentence.

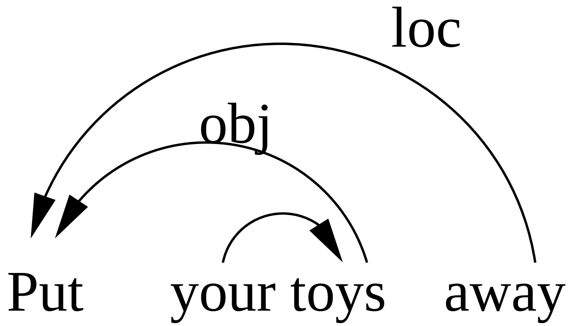
Incremental Variational Bayes EM

- For n string-interpretation pairs $\{(S_i, I_i); i = 1 \ldots n\}$:

1. Find all derivations **D** that map string S to I
2. For each derivation $D \in \mathbf{D}$, calculate its probability:
   $P(D|S,I;\theta) = \frac{1}{Z}\prod_{r \in D} P(r|\theta)$
3. Calculate expectation of each rule r being used in all derivations D:
   $E(r|S,\mathbf{D};\theta) = \sum_{D \in \mathbf{D}} count(r \in D) \times P(D|S,I;\theta)$
4. Update model parameters with rule expectations:
   $\theta^{t+1} := \theta^t + E(r|\mathbf{D};\theta)$

Experiment: Learning from CHILDES Data (Kwiatkowski *et al.* 2012 this conference)

- Part of the CHILDES corpus (“Eve”) is annotated with dependency graphs.



- These are English-specific.
- We can ignore linear order and treat them as impoverished logical forms.
- In fact we automatically map them into equivalent λ-terms.

Limitations of CHILDES Data

- The resulting pseudo-logical forms are still partly lexically English-specific.

- We will learn constructions in other languages that are *more synthetic* than English as multi-word items.

- Milou traverse la rue à la course! *(Milou runs across the road!)*

(24)

$$\begin{array}{cc}
\text{Milou} & \text{traverselaruеàla}\textsc{course} \quad ! \\
\hline
S/(S\backslash NP) : \lambda p.p\ milou' & S\backslash NP : \lambda y.run'(across'road')y \\
\hline
\multicolumn{2}{c}{S : run'(across'road')\ milou'} \quad >
\end{array}$$

- CHILDES isn't annotated in the Language of Thought accessed by the child.

Using CHILDES Data

- Nevertheless, we can learn any construction in any language that is *less synthetic* than English.

- Ranges tes jouets! (*Put away your toys!*)

(25)
$$\begin{array}{lcc}
\textsc{Ranges} & \textsc{tes jouets} & ! \\
\cline{1-1}\cline{2-2}
S/NP : \lambda x.put'away'x\ you' & NP : toys' & \\
\cline{1-2}
\multicolumn{2}{c}{S : put'away'toys'you'} & >
\end{array}$$

Results

- Following Alishahi and Stevenson (2008), we train on chronological sections $1-n$ and test on $n+1$. We see steady learning for both this program and Kwiatkowski *et al.* (2010) (the latter with the Giza++ alignment initialization turned off and run for 10 iterations over $1-n$.)
- The present program learns around a steady 8% better than the latter State-of-the-Art semantic parser inducer.
- Even with Giza alignment it is around 1% better. Full results in Kwiatkowski *et al.* (2012).
- Absolute accuracy of all systems is low because we can only learn from 33% of Eve, excluding stuff like “MMMMMM” and “DOGGIE DOGGIE DOGGIE!”
- The Eve corpus is also a tiny proportion of what the child has to work with, so test-on-$n+1$ is very brutal.

Analysis: Learning Curves

- The following slides show learning curves for

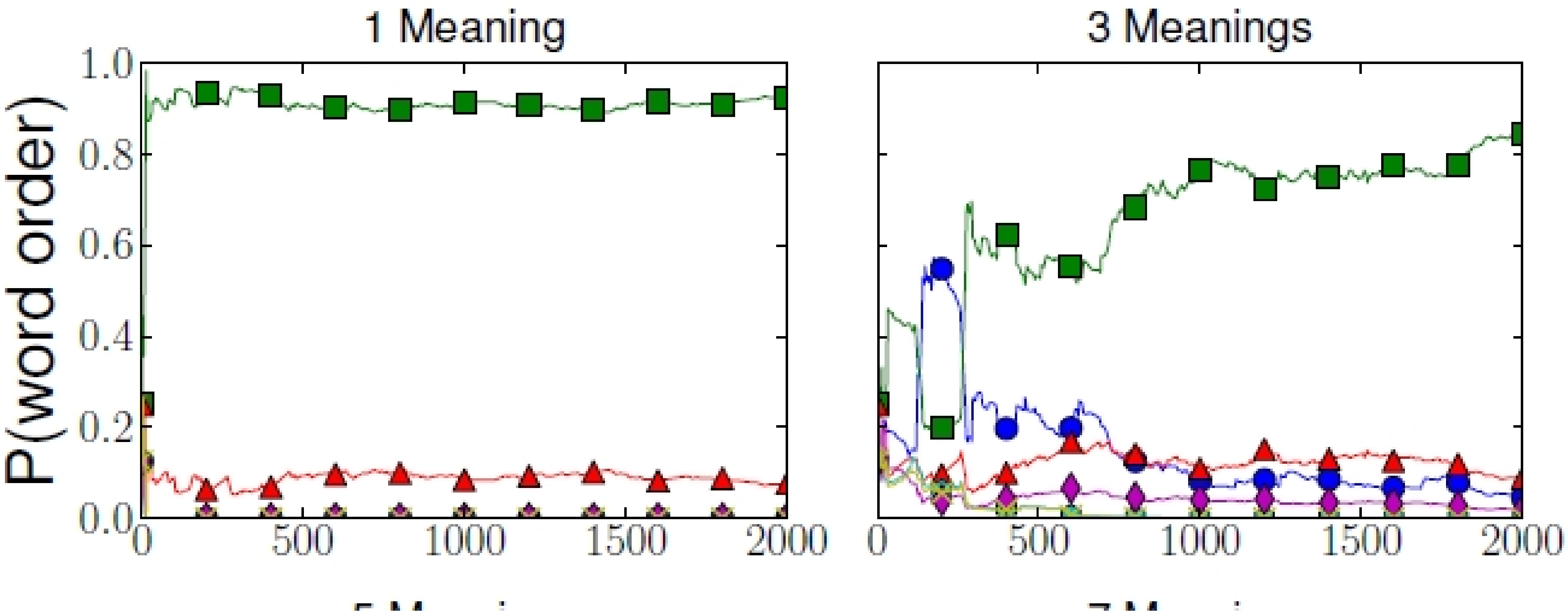
  1. Learning that verbs are SVO rather than SOV, VSO, VOS, OSV, or OVS
  2. Learning that determiners are Det-N rather than N-Det;

  In each case, curves are shown for learning with

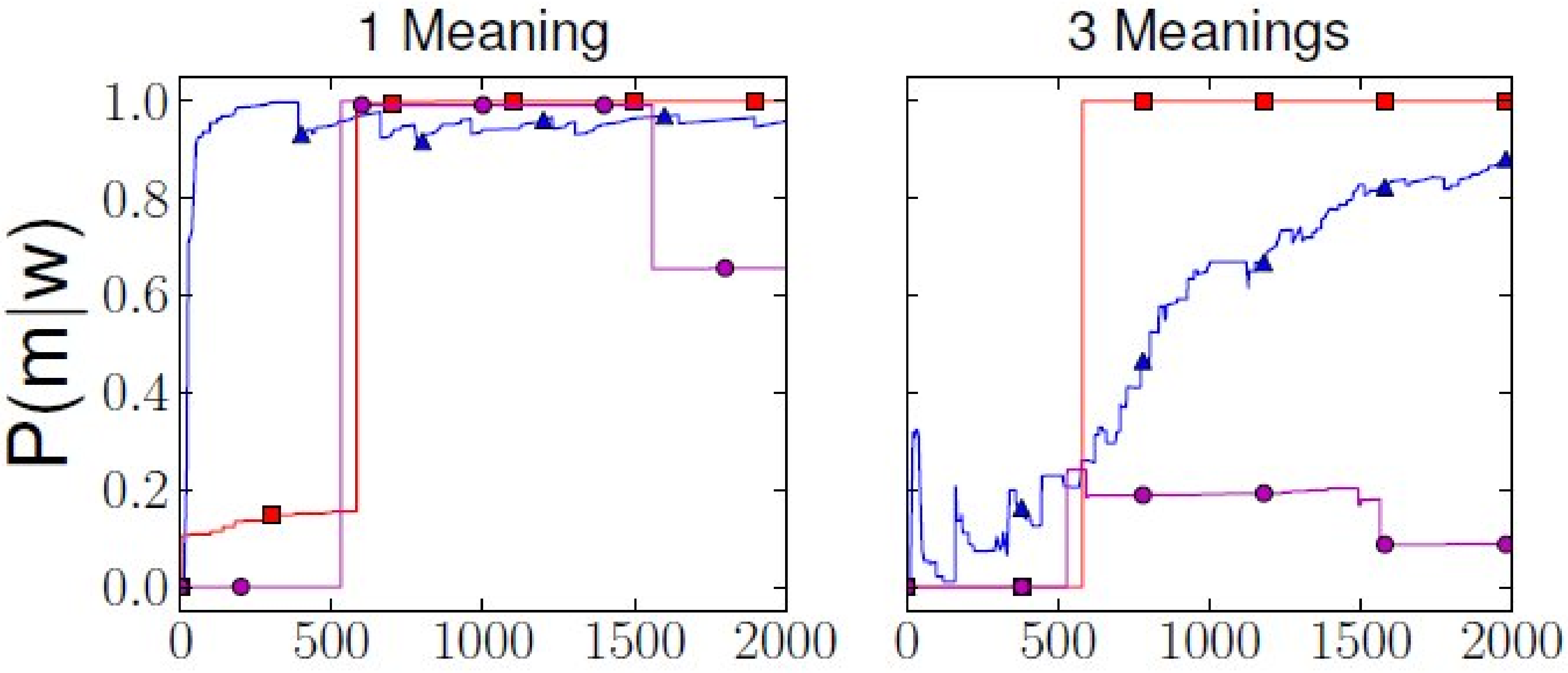
  1. The correct logical form alone;
  2. The correct logical form plus the logical forms from the preceding and succeeding turn, as irrelevant distractors.
  3. Cases with even more distractors are discussed in Kwiatkowski *et al.* (2012).

- In the later case learning is slower, but still converges.

Learning SVO Word Order



Learning Determiners *a, another, any*



Fast Mapping

- Frequent determiners like "a" $(f = 168)$ are learned slowly and continuously with high stability,

- By the time low frequency determiners like "another" $(f = 10)$ and "any" $(f = 2)$ are actually encountered, the prior on the category NP/N has grown to the point where learning may only need a few trials.

- Such "fast mapping" is characteristic of later child language acquisition (Carey and Bartlett 1978).

Later Development

- The latter effect is all that is needed to explain the phenomenon of "syntactic bootstrapping" (Gleitman 1990), where at a later stage of development, the child can learn lexical entries for words for which the corresponding concept is not salient, or is even entirely lacking to the child.

- Transitive verbs could in principle be assigned either of the two syntactic categories in (27), both of which support a derivation of a different logical form supported by the same contexts:

(26) Grover flees Big Bird! := $S : \mathit{flee'bigbird'grover'}$

(27) a. flee := $(S\backslash NP)/NP : \lambda x\lambda y.flee'xy$
b. flee := $*(S/NP)\backslash NP : \lambda x\lambda y.chase'xy$

Bootstrapping “Flee” against Competing “Chase”

- Pairs of verbs which a single situation necessarily supports are relatively rare, and one member is usually much rarer
- There is exactly one occurrence of any form of “flee” in the entire CHILDES corpus, in comparison to 162 occurences of inflected forms of the verb “chase”.
- We are therefore justified in assuming that situations unambiguously supporting the correct transitive category will predominate.
- Providing everything else in the sentence is known, this should be enough to ensure that the priors for the derivation that supports the correct category (27a) with the nonsalient or unavailable meaning will be more probable than that with the nonstandard category (27b) with a salient meaning.

Bootstrapping Known and Unknown Meanings

- 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 high probability to the correct lexical entry, and can productively proceed to investigate its meaning further (Thomforde and Steedman 2011).

(28) "Why can't you cut ice with A. Smaill?"

Bootstrapping

- Gleitman 1990 has described the process by which the child resolves this 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 syntactic bootstrapping is emergent from the statistical model resulting from primary semantic bootstrapping.

Bootstrapping

- Like the related proposals of Siskind; Villavicencio; Zettlemoyer and Collins and the somewhat different probabilistic approach of Yang 2002, this proposal considerably simplifies the logical problem of language acquisition:
  - No "subset principle."
  - No "triggers" other than reasonably short reasonably interpretable sentences in context, drawn from a reasonably representative sample.
  - Hence no "trigger ordering problem."
  - No "parameters"

☡ We need more datasets! (Commentaries? Call centers?)

Conclusion

- The theory presented here somewhat resembles the proposals of Fodor 1998 as developed in Sakas and Fodor (2001) and Niyogi (2006), and Yang (2002) in treating the acquisition of grammar as in some sense parsing with a universal "supergrammar".

- However, rather than learning over the space of *all possible grammars* corresponding to all possible parameter settings, the present theory adjusts probabilities in a model of all elements of the grammar for which there is positive evidence from all processable utterances.

- "Parameters" like V2 vs. free order are simply statements about probability distributions over lexical types and rule types.

- Nevertheless, learning is typically step-like, like parameter-setting.

Moral: Against “Parameter Setting”

- If parameters are implicit in the rules or categories themselves, and you can learn the rules or categories directly, why should the child (or a truly Minimal theory) bother with parameters at all?

- For the child, all-or-none parameter-setting is counterproductive, as many languages include inconsistent items.

- Consider English expressions like *Doggies galore!*

◇ “Galore” is one of a tiny group of phrase-final determiner in E. (It came from Irish. the others are “a-plenty” (Norman French) and “a gogo” (also Fr))

IV: Towards a Robust Semantics

Building Interpretations

- The combinatory rules guarantee “surface compositionality” with any compositional semantic representation.
- Thus the process of interpretation building can be built into the categories and combinatory rules, and can be done in parallel to derivation, as in (4)
- To make such a semantics wide-coverage involves specifying a semantics or a morphological stem-based semantic schema for the 400-500 most frequent category types (Hockenmaier *et al.* 2004; Bos et al. 2004)
- Generalize categories for open-class content words.
- Use 1st order logics such as DRT, using λ-calculus as “glue language”.
- Example (Bos *et al.* (2004): *From 1953 to 1955 , 9.8 billion Kent cigarettes with the filters were sold , the company said .*

```
  -------------   ------------------------------------------------------------------
 | x1          | | x2 x3                                                            |
 |-------------| |------------------------------------------------------------------|
(| company(x1) |A| say(x2)                                                          |)
 | single(x1)  | | agent(x2,x1)                                                     |
 |_____________| | theme(x2,x3)                                                     |
                 | proposition(x3)                                                  |
                 |        ------------------    ------------   -----------------    |
                 |       | x4               |  | x5         | | x6 x7 x8        |   |
                 |   x3: |------------------|  |------------| |-----------------|   |
                 |      (| card(x4)=billion |;(| filter(x5) |A| with(x4,x5)     |)) |
                 |       | 9.8(x4)          |  | plural(x5) | | sell(x6)        |   |
                 |       | kent(x4)         |  |____________| | patient(x6,x4)  |   |
                 |       | cigarette(x4)    |                 | 1953(x7)        |   |
                 |       | plural(x4)       |                 | single(x7)      |   |
                 |       |__________________|                 | 1955(x8)        |   |
                 |                                            | single(x8)      |   |
                 |                                            | to(x7,x8)       |   |
                 |                                            | from(x6,x7)     |   |
                 |                                            | event(x6)       |   |
                 |                                            |_________________|   |
                 | event(x2)                                                        |
                 |__________________________________________________________________|
```

The Poverty of Logicism

- Parsing with C&C 2004, and feeding such logical forms to a battery of FOL theorem provers, Bos and Markert (2005) attained quite high precision of 76% on the 2nd PASCAL RTE Challenge Problems.

- However, recall was only 4%, due to the overwhelming search costs of full FOL theorem proving.

- MacCartney and Manning (2007) argue that entailment must be computed more directly, from the surface form of sentences, using edit distance, and fast inference such as Modus Ponens implicit in polarity marking and resources such as WordNet.

- It is the latter that does the real work.

Polarity

- It is well-known that explicit and implicit *negation* systematically switches the “upward” or “downward direction of entailment of sentences with respect to ontology-based inference:

(29) Egon walks $\vdash$ Egon moves
$\nvdash$ Egon walks quickly
Egon doesn’t walk $\vdash$ Egon doesn’t walk quickly
$\nvdash$ Egon doesn’t move

- Sánchez Valencia (1991) and Dowty (1994) point out that polarity can be computed surface-compositionally using CG.

Taking Scope (Steedman 2012b, hereafter *TS*)

- (30) Everybody loves somebody.

- (31) a. $\forall x[person'x \rightarrow \exists y[person'y \wedge loves'yx]])$
  b. $\exists y[person'y \wedge \forall x[person'x \rightarrow loves'yx]$

- (32) An effective silencer must be fitted to every vehicle.

- Appears not to allow computation of LF from the simple combinatorics of grammatical derivation.

☡ Has motivated "quantifying in," "covert quantifier movement," morpholexically unmotivated type-changing operations, and the dreaded "underspecification."

The Problem with Underspecification/ Movement/&c.

- The following two sentences from the Rondane treebank of MRS-based underspecified logical forms respectively generate 3,960 readings all falling into one equivalence class, and 480 readings falling into two semantically distinct equivalence classes (Koller and Thater 2006):

(33) a. For travelers going to Finnmark there is a bus service from Oslo to Alta through Sweden.
b. We quickly put up the tents in the lee of a small hillside and cook for the first time in the open.

- We should stick to surface-compositionality, using nothing but the derivational combinatorics of surface grammar to deliver all and only the attested readings.

Scope Alternation: The Universals

- The universal quantifiers *every* and *each* can invert scope in the strong sense of binding (unboundedly) c- or lf-commanding indefinites, subject to certrain island conditions.
- Such quantifier “movement” appears to be subject to the same “Across-the-Board” condition as *wh*-movement, as in examples like the following (Geach 1972):

(34) Every boy admires, and every girl detests, some saxophonist.

Two readings, not four. (Another problem for covert movement, underspecification, et.,)

Scope (Non)Alternation: The Existentials

- Existential quantifiers like *some, a*, and *at least/at most/exactly three* appear able to take wide scope over unboundedly c- or lf-commanding universals, and are *not* sensitive to island boundaries.

- However, existentials in general cannot invert scope in the strong sense of distributing over a structurally-commanding indefinite:

- Maybe existentials don't really move at all.

Deriving Scope from Grammatical Combinatorics

- Existentially quantified NPs are replaced by a generalization of standard *Skolem terms*.

- Skolem terms are obtained by replacing all occurrences of a given existentially quantified variable by a term applying a unique functor to all variables bound by universal quantifiers in whose scope the existential quantifier falls.

- Such Skolem terms denote dependent "narrow-scope" indefinite individuals.

- If there are no such universal quantifiers, then the Skolem term is a constant.

- Since constants behave as if they "have scope everywhere", such terms denote nondependent "wide-scope" specific-indefinites.

Generalized Skolem Terms

- We generalize the notion of Skolem terms by analogy to generalized quantifiers by packaging the restriction p (and any associated cardinality property c) inside the functor over arguments $\mathscr{A}$, together with a number n identifying the originating NP,in a term of the form $sk_{n:\ p;c}^{(\mathscr{A})}$,
- We can usually ignore n and c
- The ambiguity of (32) can be expressed by the following two logical forms, which differ only in the generalized skolem terms $sk_{person'}^{(x)}$ (denoting a dependent or "narrow-scope" beloved) and $sk_{person'}$, a Skolem constant.

(35) a. $\forall x[person'x \rightarrow loves' sk_{person'}^{(x)} x)]$
b. $\forall x[person'x \rightarrow loves' sk_{person'} x)]$

The Model Theory

- We need an explicit model theory because Generalized Skolem Terms are first class citizens of the logic, rather than being derived from existentials via prenex normal form. They need to carry information about their scope with them, to avoid problems arising from their interaction with negation.

(36) a. Some farmer owns no donkey.
b. $\neg_i owns' -_i sk_{donkey'} + sk_{farmer'}$

- Because of the involvement of Skolem terms and their restrictors, which are λ-terms in L, we need to identify a notion of **level** for terms of L. Object symbols, variables, and the related pro-terms are terms of level 0.
- The model theory also treats implication as $\neg P \vee (P \wedge Q)$, rather than material implication, because of duplication of Skolem terms in donkey sentences.

Syntax

1. If $a_1,\ldots,a_n$ are terms whose maximum level is i, then $R_n(a_1,\ldots,a_n)$ is a wff of level i.
2. If X is a wff of level i then $[\neg X]$ is a wff of level i.
3. If X and Y are wff for which i is the higher of their respective levels, then $[X \wedge Y]$, $[X \vee Y]$, and $[X \rightarrow Y]$ are all wff of level i.
4. If X is a wff of level i then $[\forall x[X]]$ is a wff of level i
5. If X is a wff of level i then $sk^{\mathscr{A}}_{\lambda x.X}$ is a term of level $i+1$ where $\mathscr{A}$ is the set of arguments of the Skolem functor $sk_{\lambda x.X}$ and $\mathscr{A}$ is a superset of the free variables of X other than x.

A complete formula or *sentence* of L is then a wff all of whose variables are bound.

Semantics: Preliminaries

- We refer to a generalized Skolem term $sk_{p;c}^{\mathscr{A}}$ with no free variables among its arguments $\mathscr{A}$ (and hence none in its λ-term p) as *saturated*.
- There is a basic correspondence $\mathscr{C}_0$ from model objects and relations to L object symbols, relation symbols, and pro-terms.
- If a correspondence $\mathscr{C}$ includes $\mathscr{C}_0$, but does not map any object of $\mathfrak{M}$ to a particular saturated generalized Skolem term t, then we will speak of a correspondence $\mathscr{C}'$ obtained by adding to $\mathscr{C}$ a pair $\langle \mathfrak{a}, t\rangle$ (together with all the related pronoun pairs $\langle \mathfrak{a}, pro't\rangle$, $\langle \mathfrak{a}, pro'(pro't)\rangle$, . . .) for some object $\mathfrak{a} \in \mathfrak{M}$ as an "extension of $\mathscr{C}$ to t" and of $\mathfrak{a}$ as the "value" named by t in $\mathscr{C}'$. We will refer to the set of correspondences obtained by extending $\mathscr{C}$ to some set of saturated generalized Skolem terms in L (including the null set) as the "extensions" of $\mathscr{C}$. (That is to say that the extensions of $\mathscr{C}$ include $\mathscr{C}$.)
- The function $\mathscr{C}^{-1}$ on the range of a correspondence $\mathscr{C}$ is the *inverse* of $\mathscr{C}$.

Semantics

1. $\mathscr{C}$ satisfies an atomic formula $R(a_1,\ldots,a_n)$ in L if and only if there is an there is an extension $\mathscr{C}'$ of $\mathscr{C}$ for which the terms $a_1,\ldots,a_n$ are all in the range of $\mathscr{C}'$ and:
   (a) The n-tuple $\langle \mathscr{C}'^{-1}(a_1),\ldots,\mathscr{C}'^{-1}(a_n)\rangle$ is in the relation $\mathscr{C}'^{-1}(R)$ in $\mathfrak{M}$;
   (b) For all a_i that are Skolem terms of the form $sk^{\mathscr{A}}_{p;c}$, $\mathscr{C}'$ also satisfies $p(sk^{\mathscr{A}}_{p;c})$ and $c(sk^{\mathscr{A}}_{p;c})$;
   (c) For all such Skolem terms of the form $sk^{\mathscr{A}}_{p;c}$ whose value under $\mathscr{C}'$ is a set object $\mathfrak{a}'$, there is no correspondence $\mathscr{C}''$ differing from $\mathscr{C}'$ only in the value $\mathfrak{a}''$ named by $sk^{\mathscr{A}}_{p;c}$ that satisfies the atomic formula and $p(sk^{\mathscr{A}}_{p;c})$ in which $\mathfrak{a}''$ is a proper superset of $\mathfrak{a}'$;
2. Given two sentences Y and Z in L:
   (a) $\mathscr{C}$ satisfies a sentence $\neg Y$ if and only if $\mathscr{C}$ does not satisfy Y;

(b) $\mathscr{C}$ satisfies a sentence $Y \vee Z$ if and only if $\mathscr{C}$ satisfies at least one of Y or Z;

(c) $\mathscr{C}$ satisfies a sentence $Y \wedge Z$ if and only if there is an extension $\mathscr{C}'$ of $\mathscr{C}$ to all and only the saturated generalized Skolem terms common to Y and Z that are not in the range of $\mathscr{C}$ such that $\mathscr{C}'$ satisfies both Y and Z;

(d) $\mathscr{C}$ satisfies a sentence $Y \rightarrow Z$ if and only if every extension $\mathscr{C}'$ of $\mathscr{C}$ to all and only the saturated generalized Skolem terms common to Y and Z that are not in the range of $\mathscr{C}$ that satisfies Y also satisfies Z;

3. Given a well-formed formula $Y(x)$ in L, in which x and no other variable is free:

(a) $\mathscr{C}$ satisfies a sentence $\forall x[Y(x)]$ if and only if there is an extension $\mathscr{C}'$ of $\mathscr{C}$ to all saturated generalized Skolem terms in $Y(x)$ such that for all object symbols a, in L $\mathscr{C}'$ satisfies $Y(a)$.

We then define truth of a sentence Y in a model $\mathfrak{M}$ as follows: Y is true in $\mathfrak{M}$ relative to a correspondence $\mathscr{C}$ if and only if $\mathscr{C}$ satisfies Y.

Example

- Consider a model containing six individuals: Giles, George, Elizabeth, Pedro, Modeſtine, and Maxwelton. The unary relation farmer holds for Giles, George, and Elizabeth. The unary relation donkey holds for Pedro, Modeſtine, and Maxwelton. The binary relation own holds for the pairs {Giles, Pedro}, {Giles, Modeſtine}, and {Elizabeth, Maxwelton}. The binary relation feed holds for the pairs {Giles, Pedro}, {Giles, Modeſtine}, and {Elizabeth, Maxwelton}.

- Consider the correspondence $\mathscr{C}_0$ consisting of the following pairs:

(37) {Giles, $giles'$}
{George, $george'$}
{Elizabeth, $elizabeth'$}
{Pedro, $pedro'$}
{Modeſtine, $modestine'$}
{Maxwelton, $maxwelton'$}
{farmer, $farmer'$}
{donkey, $donkey'$}
{own, own'}
{feed, $feed'$}

- Consider the donkey sentence, $\forall x[\mathit{farmer}'x \wedge \mathit{own}'\mathit{sk}^{(x)}_{\mathit{donkey}'}x \rightarrow \mathit{feed}'(\mathit{pro}'\mathit{sk}^{(x)}_{\mathit{donkey}'})x]$
- By 3a, $\mathscr{C}_0$ satisfies this sentence if and only if for all object symbols a in L there is an extension of $\mathscr{C}_0$ to the saturated generalized skolem term $\mathit{sk}^{(a)}_{\mathit{donkey}'}$ that satisfies $\mathit{farmer}'a \wedge \mathit{own}'\mathit{sk}^{(a)}_{\mathit{donkey}'}a \rightarrow \mathit{feed}'(\mathit{pro}'\mathit{sk}^{(a)}_{\mathit{donkey}'})a$.
- By 2d, the interesting cases are $a = \mathit{giles}'$ and $a = \mathit{elizabeth}'$, and the respective extensions by the pairs $\{\mathfrak{Pedro},\ \mathit{sk}^{(\mathit{giles}')}_{\mathit{donkey}'}\}$, $\{\mathfrak{Modestine},\ \mathit{sk}^{(\mathit{giles}')}_{\mathit{donkey}'}\}$, and $\{\mathfrak{Maxwelton},\ \mathit{sk}^{(\mathit{elizabeth}')}_{\mathit{donkey}'}\}$, since for all object symbols there is either no extension that satisfies the antecedent or all of these extensions satisfy both antecedent and consequent, once the skolem terms are unpacked via rule 1b.
- This is the “strong” reading of the donkey sentence, because of 2d.

Related Approaches

- DRT:
  - Like DRT with generalized Skolem terms as discourse referents, except:
  - Quantifies over farmers rather than farmer donkey pairs, hence no proportion problem.
  - Gets the strong reading for donkey sentences, unlike duplex.
- E-type pronouns:
  - No covert definites masquerading as pronouns, hence:
  - No uniqueness problem, and hence,
  - No call for minimal situations, nor ensuing ramifications concerning bishops meeting other bishops, split antecedents etc.

Related Approaches (contd.)

- Other referential accounts of indefinites:
  - Unlike Fodor 1982, Reinhart 1987, Park 1995, indefinites have *only* the non-quantificational reading.
  - Unlike storage and underspecification accounts, possibilities for scope-taking are closely tied to derivation and syntactic combinatorics.

Universals ARE Generalized Quantifiers in CCG

- The universals *every* and *each* are Good Old-Fashioned generalized quantifier determiners:

(38) every, each := $NP^{\uparrow}_{3SG}/_{\diamond}N_{3SG} : \lambda p\lambda q\lambda\ldots\forall x[px \rightarrow qx\ldots]$

- $NP^{\uparrow}$ schematizes over all NP types raised over functions of the form $T|NP$. $\lambda\ldots$ schematizes over the corresponding arguments.

☡ This is analogous to lexicalizing covert quantifier movement to SPEC-of-IP/CP. but once again there is no movement or equivalent syntactic type-lifting, only MERGE, a.k.a. unification of variables

Existentials NOT Generalized Quantifiers in CCG

- All other “quantifiers” are referential (cf. Woods 1975; VanLehn 1978; Webber 1978; Fodor and Sag 1982; Park 1996).

  (39) a, an, some $:= NP^{\uparrow}_{agr}/_{\diamond}N_{agr} : \lambda p\lambda q.q(skolem'p)$

- In the present theory, existentials entirely lack quantificational senses.

Indefinites as Generalized Skolem Terms

- We do this by making the meaning of NPs underspecified Skolem terms of the form $skolem'_n : (p;c)p$, (Again, p is a predicate such as $donkey'$, corresponding to the restrictor of a generalized quantifier, c is a cardinality condition which may be null, and n is a number unique to the originating NP which we usually suppress.)

- We then define a notion of an environment for Skolem terms:

(40) *The environment* $\mathscr{E}$ of an unspecified skolem term $\mathscr{T}$ is a tuple comprising all variables bound by a universal quantifier or other operator in whose structural scope $\mathscr{T}$ has been brought at the time of specification, by the derivation so far.

Indefinites as Generalized Skolem Terms

- Skolem term Specification (simplified) can then be defined as follows:

  (41) *Skolem specification* of a term t of the form $skolem'_{n}p;c$ in an environment $\mathscr{E}$ yields a generalized Skolem term $sk^{\mathscr{E}}_{n,p;c}$, which applies a generalized Skolem functor $sk_{n,p}$ to the tuple $\mathscr{E}$, defined as the environment of t at the time of specification, which constitutes the *arguments* of the generalized Skolem term.

  We will suppress the number n from now on, since it usually does no work.

- There is more to say about negation and polarity here—see *TS*.

Narrow-scope Saxophonist Reading

$$
\begin{array}{lccc}
(42) & \text{Every boy} & \text{admires} & \text{some saxophonist}\\
\cline{2-2}\cline{3-3}\cline{4-4}
 & S/(S\backslash NP_{3SG}) & (S\backslash NP_{3SG})/NP & (S\backslash NP)\backslash((S\backslash NP)/NP)\\
 & :\lambda p.\forall y[boy'y \rightarrow py] & admire' & :\lambda q.q(skolem'sax')\\
\cline{3-4}
 & & \multicolumn{2}{c}{S\backslash NP \quad <}\\
 & & \multicolumn{2}{c}{:admires'(skolem'sax')}\\
\cline{2-4}
 & \multicolumn{3}{c}{S:\forall y[boy'y \rightarrow admires'(skolem'sax')y] \quad >}\\
 & \multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
 & \multicolumn{3}{c}{S:\forall y[boy'y \rightarrow admires'sk^{(y)}_{sax'}y]}
\end{array}
$$

- Unlike FCG/TLG, **the left-branching derivation allows the same logical form.**
- That has to be the case, because of the Geach sentence.

Wide-scope Saxophonist Reading

$$
\begin{array}{lcccl}
(43) & \text{Every boy} & \text{admires} & \text{some saxophonist} & \\
\cline{2-2}\cline{3-3}\cline{4-4}
 & S/(S\backslash NP_{3SG}) & (S\backslash NP_{3SG})/NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & \\
 & :\lambda p.\forall y[boy'y \rightarrow py] & admires' & :\lambda q.q(skolem'sax') & \\
 & & & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
 & & & :\lambda q.q\ sk_{sax'} & \\
\cline{3-4}
 & & \multicolumn{2}{c}{S\backslash NP : admires' sk_{sax'}} & < \\
\cline{2-4}
 & \multicolumn{3}{c}{S : \forall y[boy'y \rightarrow admires' sk_{sax'} y]} & >
\end{array}
$$

- Unlike FCG/TLG, **the left-branching derivation allows the same logical form**.
- That has to be the case, because of the Geach sentence.

How Universals Invert Scope

- (44)

$$\begin{array}{ccc}
\text{Some boy} & \text{admires} & \text{every saxophonist}\\
\hline
S/(S\backslash NP_{3SG}) & (S\backslash NP_{3SG})/NP & (S\backslash NP)\backslash((S\backslash NP)/NP)\\
:\lambda p.p(skolem'boy') & :\lambda x\lambda y.admires'xy & :\lambda q.\forall x[sax'x\rightarrow qx]\\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3SG}:\lambda y.\forall x[sax'x\rightarrow admires'xy]} \; < \\
\hline
\multicolumn{3}{c}{S:\forall x[sax'x\rightarrow admires'x(skolem'boy')]} \; > \\
\multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
\multicolumn{3}{c}{S:\forall x[sax'x\rightarrow admires'x\; sk_{boy'}^{(x)}]}
\end{array}$$

- The SVO grammar of English means that embedded subjects in English are correctly predicted neither to extract nor to allow universals to take scope over their matrix subject in examples like the following (Cooper 1983, Farkas 2001):

Non-Inversion of Embedded Subject Universals

- (45) a. *a boy who(m) [I know that]$_{S/_{\diamond}S}$ [admires some saxophonist]$_{S\backslash NP}$

  b. [Somebody knows (that)]$_{S/_{\diamond}S}$ [every boy]$_{S/(S\backslash NP)}$ [admires]$_{(S\backslash NP)/NP}$ some saxophonist.

  $\neq \forall x[boy'x \rightarrow know'(admire'sk_{saxophonist'}x)sk^{(x)}_{person'}]$

  $\neq \forall x[boy'x \rightarrow know'(admire'sk^{(x)}_{saxophonist'}x)sk^{(x)}_{person'}]$

- This sort of thing is very common in German (Kayne 1998; Bayer 1990, 1996; *SP*)

☡ To allow bare complement subjects to extract a quite different “antecedent governed” category $(VP/NP_{-LEX,agr})/(S\backslash NP_{agr})$ must be added to the English lexicon for *know*. However, *Every boy* cannot combine with that.

How Universals Invert Scope Out of NP Modifiers

- (46) a. Some apple in every barrel was rotten.
  b. Someone from every city despises it/#the dump

- Cf. #A City that every person from admires sincerity.

- But also cf. A city that every person from despises

(47)

$$
\begin{array}{ccc}
\text{Some apple in} & \text{every barrel} & \text{was rotten} \\
\hline
(S/(S\backslash NP))/NP : \lambda x\lambda p.p(skolem'\lambda y.apple'y \wedge in'x\ y) & NP^{\uparrow} : \lambda p.\forall x[barrel'x \rightarrow px] & S\backslash NP : rotten' \\
\hline
\multicolumn{2}{c}{S/(S\backslash NP) : \lambda p.\forall x[barrel'x \rightarrow p(skolem'\lambda y.apple'y \wedge in'x\ y)]} < & \\
\hline
\multicolumn{3}{c}{S : \forall x[barrel'x \rightarrow rotten'(skolem'\lambda y.apple'y \wedge in'x\ y)]} > \\
\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\
\multicolumn{3}{c}{S : \forall x[barrel'x \rightarrow rotten' sk^{(x)}_{\lambda y.apple'y \wedge in'x\ y}]}
\end{array}
$$

Inverse Scope Limits Readings

- This process only supports four distinct readings for the following:

(48) a. Some representative of every company saw every sample.
b. Every representative of some company saw every sample.

- We will return to this example below.

Why Non Universals Don't Invert Scope

- Non-universals cannot invert scope because they are not quantificational:

(49) a. Some linguist can program in at most two programming languages.
b. Most linguists speak at least three/many/exactly five/no/most languages.

☡ Chierchia (1995) points out that apparent exceptions like "a Canadian flag was hanging in front of at least five windows," crucially involve unaccusatives, passives, etc.

☡ Hirschbüller pointed out that, exceptionally, they support inversion out of VP ellipsis. Something else is going on.

Binding and Distributivity are Lexicalized

- (50) a. eat a pizza:= $S\backslash NP_{PL} : \lambda y.eat'(skolem'pizza')y$
  b. eat a pizza := $S\backslash NP_{PL} : \lambda y.\forall w[w \in y \rightarrow eat'(skolem'pizza')w]$
- Binding and Distributivity are lexicalized via the verb (cf. Link 1983, *passim*):

(51)
$$\begin{array}{cccc}
\text{Three} & \text{boys} & \text{ate} & \text{a pizza}\\
\hline
NP^{\uparrow}_{PL}/N_{PL} & N_{PL} & (S\backslash NP_{PL})/NP & NP^{\uparrow}\\
:\lambda n\lambda p.p(skolem'n\ ;\ \lambda s.|s|=3)) & :boy' & :\lambda x\lambda y.\forall z[z\in y\rightarrow eat'xz] & :\lambda p.p(skolem'pizza'
\end{array}$$

$$\overline{NP^{\uparrow}_{PL} : \lambda p.p(skolem'boy'\ ;\ \lambda s.|s|=3))}^{>} \qquad \overline{S\backslash NP : \lambda y : \forall z[z \in y \rightarrow eat'(skolem'pizza')z]}^{<}$$

$$NP^{\uparrow}_{PL} : \lambda p.p(sk_{boy'\ ;\ \lambda s.|s|=3}))$$

$$\overline{S : \forall z[z \in sk_{boy'\ ;\ \lambda s.|s|=3} \rightarrow eat'(skolem'pizza')z]}^{>}$$

$$S : \forall z[z \in sk_{boy'\ ;\ \lambda s.|s|=3} \rightarrow eat'sk^{(z)}_{pizza'}z]$$

Crucially, the term on the left of the implication is *not* a Generalized Skolem term, but a bound variable z.

Binding and Distributivity are Lexicalized

- Compare Greenlandic Eskimo and Chinese, in which distributivity is explicitly morphologically marked on the verb, and English "collective-only" verbs, as in *Three boys/#Every boy gathered in the library*.
- Localizing distributivity on the verb predicts mixed readings for the following:

(52) a. Three boys ate a pizza and lifted a piano.
b. Three boys gathered in the bar and ate a pizza.
c. Three boys met each other and ate a pizza.

- As Robaldo points out, This account means that *Exactly two boys read exactly one book* is false in a model in which there are two boys who read the same book, and one of them read some other book.
- I claim that result is correct for natural quantifiers.

The Donkey

(53)

$$\begin{array}{cccccc}
\text{Every} & \text{farmer} & \text{who} & \text{owns} & \text{a donkey} & \text{feeds it} \\
\hline
(S/(S\backslash NP_{3SG})/_{\diamond}N_{3SG} & N_{3SG} & (N_{agr}\backslash_{\diamond}N_{agr})/_{\diamond}(S\backslash NP_{agr}) & (S\backslash NP_{3SG})/NP & (S\backslash NP)\backslash(S\backslash NP)/NP & S\backslash NP_{3SG} \\
:\lambda n\lambda p.\forall x[nx\rightarrow px] & :farmer' & :\lambda q\lambda n\lambda y.ny\wedge qy & :\lambda x\lambda y.own'xy & :\lambda p.p(skolem'donkey') & :\lambda y.feed'it'y
\end{array}$$

$$\overline{}^{<}$$

$$S\backslash NP_{3SG}$$
$$:\lambda y.ny\wedge own'(skolem'donkey')y$$

$$\overline{}^{>}$$

$$N_{3SG}\backslash N_{3SG}:\lambda n\lambda y.ny\wedge own'(skolem'donkey')y$$

$$\overline{}^{<}$$

$$N_{3SG}:\lambda y.farmer'y\wedge own'(skolem'donkey')y$$

$$\overline{}^{>}$$

$$S/(S\backslash NP_{3SG}):\lambda p.\forall x[farmer'x\wedge own'(skolem'donkey')x\rightarrow px]$$

$$\overline{}^{>}$$

$$S:\forall x[farmer'x\wedge own'(skolem'donkey')x\rightarrow feed'it'x]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S:\forall x[farmer'x\wedge own'sk_{donkey'}^{(x)}x\rightarrow feed'it'x]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S:\forall x[farmer'x\wedge own'sk_{donkey'}^{(x)}x\rightarrow feed'(pro'sk_{donkey'}^{(x)})x]$$

Strength *sans* Proportion/Uniqueness Problems

- (54)

$$\begin{array}{ccc}
\text{Most} & \text{farmers who own a donkey} & \text{feed it}\\
\hline
NP^{\uparrow}_{PL}/_{\diamond}N_{PL} & N_{PL} & S\backslash NP_{PL}\\
:\lambda n\lambda p.p(skolem'n\ ;\ most') & :\lambda x.farmer'x\wedge own'(skolem'donkey')x & :\lambda y.\forall z[z\in y\rightarrow feeds'(pronoun'it')z]\\
\multicolumn{2}{l}{\overline{\hspace{30em}}>} & \\
\multicolumn{2}{l}{NP^{\uparrow}_{PL}:\lambda p.p(skolem'(\lambda x.farmer'x\wedge own'(skolem'donkey')x\ ;\ most'))} & \\
\multicolumn{2}{l}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} & \\
\multicolumn{2}{c}{NP^{\uparrow}_{PL}:\lambda p.p(sk_{\lambda x.farmer'x\wedge own'(skolem'donkey')x\ ;\ most'})} & \\
\multicolumn{3}{l}{\overline{\hspace{50em}}>}\\
\multicolumn{3}{c}{S:\forall z[z\in sk_{\lambda x.farmer'x\wedge own'(skolem'donkey')x\ ;\ most'}\rightarrow feeds'(pronoun'it')z]}\\
\multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
\multicolumn{3}{c}{S:\forall z[z\in sk_{\lambda x.farmer'x\wedge own'sk^{(z)}_{donkey'}x\ ;\ most'}\rightarrow feeds'(pronoun'it')z]}\\
\multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
\multicolumn{3}{c}{S:\forall z[z\in sk_{\lambda x.farmer'x\wedge own'sk^{(z)}_{donkey'}x\ ;\ most'}\rightarrow feeds'sk^{(z)}_{donkey'}z]}
\end{array}$$

- $most' = \lambda s.|s| > 0.5 * |all'(\lambda x.farmer'x \wedge own(skolem'donkey'))|.$
- Quantifies over farmers z not farmer-donkey pairs, avoiding proportion problem. It is the distributive universal quantifier that ensures strong reading without recourse to devices like “minimal situations,” avoiding uniqueness problem.

Coordination Constraints on Scope Alternation

- *SP* showed that, by contrast with distributivity, localizing quantification and Skolem terms on the NP disallows mixed readings:

- Narrow-scope saxophonist reading of (34):

(55)

$$\begin{array}{cc}
\text{Every boy admires and every girl detests} & \text{some saxophonist}\\
\hline
S/NP & S\backslash(S/NP)\\
:\lambda x.\forall y[boy'y \rightarrow admires'xy] \wedge \forall z[girl'z \rightarrow detests'xz] & :\lambda q.q(skolem'sax')\\
\hline
\multicolumn{2}{r}{<}\\
\multicolumn{2}{c}{S:\forall y[boy'y \rightarrow admires'(skolem'sax')y] \wedge \forall z[girl'z \rightarrow detests'(skolem'sax')z]}\\
\multicolumn{2}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
\multicolumn{2}{c}{S:\forall y[boy'y \rightarrow admires'sk^{(y)}_{sax'}y] \wedge \forall z[girl'z \rightarrow detests'sk^{(z)}_{sax'}z]}
\end{array}$$

Coordination Constraints on Scope Alternation

- The same categories also yield the wide-scope saxophonist reading of (34):

(56)

$$
\begin{array}{cc}
\text{Every boy admires and every girl detests} & \text{some saxophonist} \\
\hline
S/NP & S\backslash(S/NP) \\
:\lambda x.\forall y[boy'y \rightarrow admires'xy] \wedge \forall z[girl'z \rightarrow detests'xz] & :\lambda q.q(skolem'sax') \\
 & \cdots\cdots\cdots\cdots\cdots \\
 & :\lambda q.q(sk_{sax}) \\
\hline
\multicolumn{2}{c}{S : \forall y[boy'y \rightarrow admires'sk_{sax}y] \wedge \forall z[girl'z \rightarrow detests'sk_{sax}z]}
\end{array}
<
$$

There are no mixed readings.

Polarity and Directional Entailment

- (57) doesn't$^{\circ}$ $:= (S^{\circ}\backslash NP)/(S^{\bullet}_{inf}\backslash NP) : \lambda p.\bullet p$

- $\circ$ stands for the polarity of the syntactic/semantic environment, and $\bullet$ stands for $-\circ$, its inverse.

- Crucially, this category inverts the polarity of the predicate alone.

Negation, Polarity, and Directional Entailment

- Sánchez Valencia (1991) and Dowty (1994) point to the natural compatibility of CG and polarity.

- (58)

$$\begin{array}{ccc}
\text{Enoch} & \text{doesn}'\text{t} & \text{walk} \\ \cline{1-1}\cline{2-2}\cline{3-3}
\text{Enoch}^{+} := & \text{doesn}'\text{t}^{\circ} := & \text{walk}^{\circ} := \\
S^{\circ}/(S^{\circ}\backslash NP^{+}) & (S^{\circ}\backslash NP)/(S^{\bullet}_{inf}\backslash NP) & S^{\circ}_{inf}\backslash NP \\
: \lambda p.p\ {+enoch'} & : \lambda p\lambda x.{\bullet p}\ {\circ x} & : {\circ walk'} \\ \cline{2-3}
 & \multicolumn{2}{c}{\text{doesn}'\text{t}^{\circ}\text{walk}^{\bullet} := S^{\circ}\backslash NP : {\bullet walk'} \quad >} \\ \cline{1-3}
\multicolumn{3}{c}{\text{Enoch}^{+}\text{doesn}'\text{t}^{+}\text{walk}^{-} := S^{+} : {-walk'}{+enoch'} \quad >}
\end{array}$$

- This engenders a number of complications to the model theory of *TS*, notably that Skolem terms must carry polarity, and in the case it is negative, a binding to a particular negative operator

NPIs &c.

(59) any $:= (S^{\bullet}/(S^{\bullet}\backslash NP))/N^{-}) : \lambda p \lambda q.q - (skolem' p)$

$(S^{\bullet}/(S^{\bullet}/NP))/N^{-}) : \lambda p \lambda q.q - (skolem' p)$

$((S^{\bullet}\backslash NP)\backslash((S^{\bullet}\backslash NP)/NP))/N^{-}) : \lambda p \lambda q.q - (skolem' p)$

&c.

Split Scope

- Błaszczak and Gärtner (2005) point out that CCG predicts split scope readings under such an analysis:

(60)

$$\begin{array}{cccl}
\text{They asked us} & \text{to review} & \text{no book} & \\
\cline{1-1}\cline{2-2}\cline{3-3}
S/VP_{to-inf} & VP^{\circ}_{to-inf}/NP & VP^{\circ}_{to-inf}\backslash(VP^{\bullet}_{to-inf}/NP) & \\
:\lambda p.ask'(p\ us')\circ us'\circ they' & :review' & :\lambda p\lambda y.\neg p\ {-sk_{book'}}y & \\
\cline{2-3}
 & \multicolumn{2}{c}{VP^{\circ}_{to-inf}:\lambda y.\neg review'{-sk_{book'}}y} & < \\
\cline{1-3}
\multicolumn{3}{c}{S^{+}:ask'(\neg review'{-sk_{book'}}{+us'}){+us'}{+they'}} & >
\end{array}$$

Split Scope (contd.)

$$
\begin{array}{ccccc}
(61) & \text{They} & \text{asked us to review} & \text{no book} & \\
\cline{2-2}\cline{3-3}\cline{4-4}
 & S/VP & VP^{\circ}/NP & VP^{\circ}\backslash(VP^{\bullet}/NP) & \\
 & \lambda p.p\ \mathit{they}' & :\lambda x\lambda y.\mathit{ask}'(\mathit{review}'x\circ\mathit{us}')\circ\mathit{us}'y & :\lambda p\lambda y.\neg p\ {-}\mathit{sk}_{\mathit{book}'}y & \\
\cline{3-4}
 & & \multicolumn{2}{c}{VP^{\circ}:\lambda y.\neg\mathit{ask}'(\mathit{review}'{-}\mathit{sk}_{\mathit{book}'}\mathit{us}')\circ\mathit{us}'y} & < \\
\cline{2-4}
 & \multicolumn{3}{c}{S^{+}:\neg\mathit{ask}'(\mathit{review}'{-}\mathit{sk}_{\mathit{book}'}{+}\mathit{us}'){+}\mathit{us}'{+}\mathit{they}'} & >
\end{array}
$$

- See *TS* for MIT publication sentences, cross linguistic differences, double negation, negtive concord, &c.

Processing Scope in CCG

- One might expect Skolem Specification to induce further Spurious Ambiguity

(62) A representative of a company saw a sample.

- The parser will have to keep track of eight distinct underspecified logical forms, representing all possible combinations of specification versus nonspecification of three generalized Skolem terms.

- This ambiguity too is real, and must be allowed for in any framework. For example, if there is a dependency-inducing universal, as follows, then all eight interpretations are semantically distinct.

(63) Every exhibitor knows that a representative of a company saw a sample.

Son of Spurious Ambiguity

- Since Skolem specification can happen at any point in a derivation, it might therefore appear that there is a danger of an even greater proliferation of semantically spurious ambiguity.

Sharing Structure

- Unlike related sets of traditional logical forms using traditional quantifiers, all eight partially specified logical forms are structurally homomorphic.

- Rather than maintaining a single underspecified logical form as in UDRT, the multiple specified readings can instead be efficiently stored as a single packed shared structure, which we might visualize as follows:

(64) $saw'(\left\{\begin{matrix} skolem' \\ sk \end{matrix}\right\} sample')(\left\{\begin{matrix} skolem' \\ sk \end{matrix}\right\} \lambda x.(representative' x \wedge of' \left\{\begin{matrix} skolem' \\ sk \end{matrix}\right\} company'\ x))$

- Since unspecified terms can be eliminated at the end of derivation, this ends up as:

(65) $saw'(\{sk\}\, sample')(\{sk\}\, \lambda x.(representative' x \wedge of'\, \{sk\}\, company'\ x))$

Sharing Structure

- The related ambiguous example (66) delivers a shared structure (67), off which four distinct specified readings can be read directly.

(66) A representative of every company saw a sample.

(67) $\forall y \left[company'y \rightarrow saw'(\left\{ \begin{array}{c} sk^{(y)} \\ sk \end{array} \right\} sample')(\left\{ \begin{array}{c} sk^{(y)} \\ sk \end{array} \right\} \lambda x.representative'x \wedge of'yx) \right]$

- (Cf. example (48).)

Sharing Structure

- The four readings are as follows:

(68) a. $\forall y[\mathit{company}'y \rightarrow \mathit{saw}'\mathit{sk}^{(y)}_{\mathit{sample}'}\mathit{sk}^{(y)}_{\lambda x.\mathit{representative}'x \wedge \mathit{of}'yx}]$

b. $\forall y[\mathit{company}'y \rightarrow \mathit{saw}'\mathit{sk}^{(y)}_{\mathit{sample}'}\mathit{sk}_{\lambda x.\mathit{representative}'x \wedge \mathit{of}'yx}]$

c. $\forall y[\mathit{company}'y \rightarrow \mathit{saw}'\mathit{sk}_{\mathit{sample}'}\mathit{sk}^{(y)}_{\lambda x.\mathit{representative}'x \wedge \mathit{of}'yx}]$

d. $\forall y[\mathit{company}'y \rightarrow \mathit{saw}'\mathit{sk}_{\mathit{sample}'}\mathit{sk}_{\lambda x.\mathit{representative}'x \wedge \mathit{of}'yx}]$

Controlling Specification

- In order to avoid duplicating specified generalized Skolem terms in the logical form, we need only to add a test for nonredundant specification to the admissibility condition *admissible* of the algorithm for adding a new entry A to the chart.

- Such an admissibility condition can be incorporated by comparing the environment associated with each daughter category B, C with that of A to determine whether Skolem specification could possibly affect the structure-sharing logical form Λ_A by specifying all instances of a given unspecified Skolem term, say $skolem'_{39}sample'$.

- In which case, the specification operation is applied to its instances, and the result is structure shared, and we iterate.

Example

- (69) Every man who read a book loves every woman.

- The result consists of two packed logical forms corresponding to the two possible scopes of the two universals with respect to each other:

(70) a. $S : \forall x[man'x \wedge read'(\left\{\begin{array}{c} sk^{(x)} \\ sk \end{array}\right\} book')x \rightarrow \forall y[woman'y \rightarrow loves'yx]]$

b. $S : \forall y[woman'y \rightarrow \forall x[man'x \wedge read'(\left\{\begin{array}{c} sk^{(x,y)} \\ sk^{(x)} \\ sk \end{array}\right\} book')x \rightarrow loves'yx]]$

- Each of these two packed logical forms subsumes two interpretations, one with a wide-scope Skolem constant book, and another in which books are dependent on men. The latter generates a further reading in which books are dependent on both men and women.

Representatives

- (71) Every representative of a company saw most samples.

  The result reveals only four readings, not the five claimed by Hobbs and Shieber, and by Keller, and predicted by their higher-order unification-based mechanism.

- These four readings are represented by a single packed structure, repeated here, since there is only one true quantifier. It is therefore immediately apparent that they are semantically distinct.

(72) $S : \forall x[rep'x \wedge of'(\left\{\begin{array}{c} sk^{(x)} \\ sk \end{array}\right\} comp')x \rightarrow saw'(\left\{\begin{array}{c} sk^{(x)} \\ sk \end{array}\right\}(samples';most'))x]$

Geach Sentence and Packing

- In unpacking logical forms like the following for the Geach sentence, which has more than one occurrence of the same generalized Skolem term $skolem'_{39}saxophonist'$, we must ensure that *all* instances are interpreted as the first, or as the second, etc. specified form.

$$(73)\quad \forall y[boy'y \rightarrow admires'(\left\{\begin{array}{c} sk_{39}^{(y)} \\ sk_{39} \end{array}\right\} sax')y] \wedge \forall z[girl'z \rightarrow detests'(\left\{\begin{array}{c} sk_{39}^{(z\ )} \\ sk_{39} \end{array}\right\} sax')z]$$

- This move does not compromise the competence-theoretic account of why there are only two readings for the Geach sentence. It is simply a consequence of the use of packing in the performance representation.

Geach Sentence

- Thus, the Geach sentence ends up with just two interpretations:

(74) a. $\forall y[boy'y \rightarrow admires'(\left\{sk_{39}^{(y)}\right\} sax')y] \wedge \forall z[girl'z \rightarrow detests'(\left\{sk_{39}^{(z)}\right\} sax')z]$

b. $\forall y[boy'y \rightarrow admires'(\{sk_{39}\} sax')y] \wedge \forall z[girl'z \rightarrow detests'(\{sk_{39}\} sax')z]$

Remarks

- Most so-called quantifiers aren't generalized quantifiers. (Many languages appear to entirely lack true generalized quantifiers—Baker 1995; Bittner 1994; Aoun and Li 1993).

- The account combines the advantages of both DRT and E-type theories with a movement-free syntax and semantics.

- It escapes the Scylla of the proportion problem and the Charybdis of the uniqueness problem, without the involvement of category ambiguity for existentials or minimal situations.

What about the Open-Class Words?

- We need to get beyond the semantics of “walks” being *walks′*”
- Hand built resources like Wordnet, FrameNet, etc. are useful but incomplete.
- We need unsupervised “Machine Reading” (Etzioni *et al.* 2007; Mitchell *et al.* 2009) over web-scale corpora.
- Work by Harrington and Clark (2009) using C&C and spreading activation semantic networks is interesting.
- This work has only just begun. Ideas of ”distributed semantics” have hardly been road tested.

Conclusion

- Scope relations are defined lexically at the level of logical form, and projected onto the sentence by combinatory derivation The pure syntactic combinatorics of CCG is the source of all and only the grammatically available readings.

- All logical-form level constraints on scope-orderings can be dispensed with—a result related to, but more powerful than, that of Pereira 1990, as extended in Dalrymple et al. 1991, Shieber et al. 1996 and Dalrymple et al. 1997.

- Some but not all of these results transfer to other non-TG frameworks, such as LTAG, LFG, HPSG, and recent MP and DRT.

- However, the interactions of scope and coordinate structure discussed here seem to demand the specific syntactic combinatorics of CCG.

V: The Surface Compositional Semantics of English Intonation

Outline

1. The Four Dimensions of Information-Structural Meaning:
   (a) Contrast/Background
   (b) Theme/Rheme
   (c) Presence in Common Ground
   (d) Speaker/Hearer Agency

2. The Surface Compositionality of Intonational Semantics

3. Conclusion: Intonation Structure = Information Structure = Derivation Structure

Dimensions of Information-Structural Meaning

Dimension 1: Contrast/Background

- Accents are properties of *words*.

- In English (and perhaps in every language), primary accents mark words as contributing via their interpretation to *contrast* between the speakers actual utterance, and various *other* utterances that they might have made, as in the Alternatives Semantics of Karttunen (1976) and Rooth (1985).

(75) Q: Who was that lady I saw you with last night?
A: That was my WIFE.
H* LL%

- *Contrast* in this sense is a property of *all* (primary) accents–cf. Vallduví and Vilkuna (1998) "kontrast".

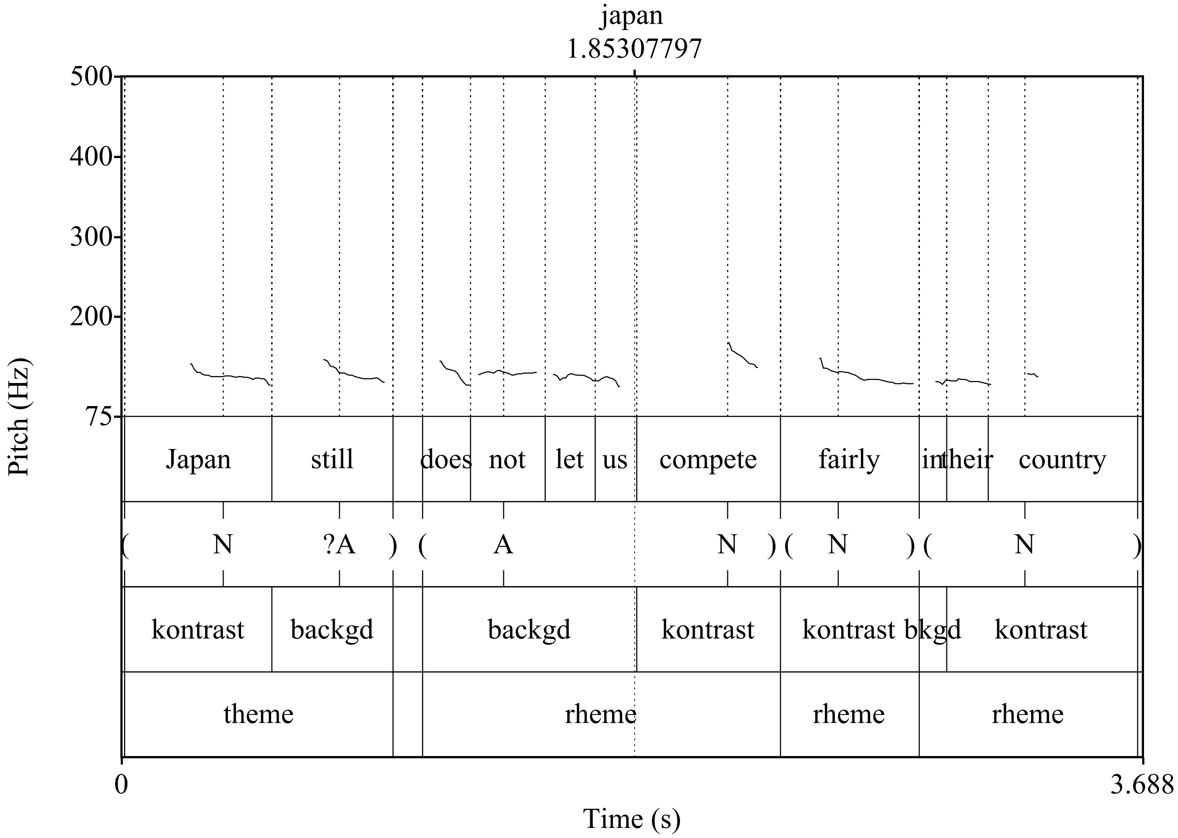
Accents are Not Necessarily Pitch-Accents

- While many English speakers (including the present one) mark the various species of accent by pitch contour, and we accordingly use the notation of Pierrehumbert 1980 to distinguish those species, such labels do not presuppose the use of pitch alon as a marker.

- They are abstract phonological categories reflecting a trade-off between a number of artulatory dimensions, including length, syllabic alignment and relative height, as well as pitch (Calhoun 2006, 2010).

- Some speakers, including those without a larynx (such as Miles Davis), certain non native speakers (sucxh as Finns), and certain native speakers mark the same distinctions without any pitch excursion.

Accents are Not Necessarily Pitch-Accents



Semantics of Contrast

- We follow Rooth (1992) in assuming that all logical forms of all linguistic elements come in pairs (Λ^o, Λ^a), and Steedman 2012b (hereafter *TS*) in assuming that non-universals translate as generalized Skolem terms, rather than as existential generalized quantifiers.
- Λ^a is an "alternative" logical form , in which the constants c in the "ordinary" logical form Λ^o corresponding to words bearing an accent have been replaced by unique free variables of the same type τ_c as c, defining an "alternatives set" $\{\Lambda^a\}$.
- For example, the alternative semantic content of the all-rheme example (75), *That was my wife* might be written as follows:

$$(76)\quad \left\{ \begin{array}{l} \mathit{was}\ \mathit{sk}_{\lambda x.\mathit{wife}\ x \wedge \mathit{mine}}\ \mathit{that} \\ \mathit{was}\ \mathit{sk}_{\lambda x.v_{\tau_{\mathit{wife}}}\ x \wedge \mathit{mine}\ x}\ \mathit{that} \end{array} \right\}$$

An Extension to the Earlier Model

- A model $\mathfrak{M}$ for the logical language L of *TS* includes a *correspondence* $\mathscr{C}$ from the *objects* $\{\mathfrak{anna},\ \mathfrak{manny},\ \ldots\}$ and *relations* $\{\mathfrak{man},\ \mathfrak{marry},\ \mathfrak{introduce},\ \ldots\}$ in $\mathfrak{M}$ into a set of *object symbols* $\{anna, manny, \ldots\}$ (not including any generalized Skolem terms or free variables), and a set of *relation symbols* $\{man, marry, introduce, \ldots,\}$ in L. The function $\mathscr{C}^{-1}$ on the range of the correspondence $\mathscr{C}$ is defined as the *inverse* of $\mathscr{C}$. Then:

1. The correspondence $\mathscr{C}$ *satisfies* a formula $Ra_1 \ldots a_n$ in which R is a relation symbol in L and all a_i are object symbols in L in the standard way, if and only if the n-tuple $\langle \mathscr{C}^{-1}(a_1), \ldots, \mathscr{C}^{-1}(a_n) \rangle$ is in the relation $\mathscr{C}^{-1}(R)$ in $\mathfrak{M}$.

2. The correspondence $\mathscr{C}$ satisfies a formula $Ra_1 \ldots a_n$ in which in which R is a relation symbol in L and some a_i are generalized Skolem terms sk_{p_i} if and only if there is an interpretation for each Skolem term sk_{p_i} as an object symbol a'_i in L such that a'_i satifies the restrictor condition p of the skolem term sk_{p_i}, and when the Skolem terms sk_{p_i} are replaced by the object symbols a'_i, $\mathscr{C}$ satisfies $Ra_1 \ldots a_n$.
3. The correspondence $\mathscr{C}$ satisfies a formula $Ra_1 \ldots a_n$ in which in which R and/or some a_i are free variables v_{τ_R} and/or $v_{\tau_{p_i}}$ if and only if there is an interpretation for each free variable as a relation symbol R' or an object symbol a'_i in L such that, when the free variables are replaced by the relation and/or object symbols a'_i, $\mathscr{C}$ satisfies $Ra_1 \ldots a_n$.

- There is much more to say (not least about semantics of negation), but for present purposes, we can assume that the rest of the model theory for *TS* is much like a standard model of first-order predicate logic, with rules for each connective and for the sole (Universal) quantifier.

Dimension 2: Theme/Rheme

- Accents also define information-structural role, which syntax projects onto constituents delimited by boundary tones:

(77) Q: I know EMMA will marry ARNIM. But who will marry MANNY?
A: (ANNA)(will marry MANNY).
H* L+H* LH%

(78) Q: I know EMMA will marry ARNIM. But who will ANNA marry?
A: (ANNA will marry)(MANNY).
L+H* LH% H* LL%

- The claim: L+H* (and L*+H) mark **theme** (roughly, the topic or "Question Under Discussion") in English. H* (and H*+L, L*, and H+L*) mark **rheme** (roughly, the comment or part that advances that discussion).

The Speaker Defines the Theme

- That is not to say that information structure is uniquely determined by contexts such as wh-questions.
- Speakers **establish** information structure *by their utterance*, as in the following variant of (78):

(79) Q: I know EMMA will marry ARNIM. But who will ANNA marry?
A: (ANNA)(will marry MANNY).
L+H* LH% H* LL%

- The hearer **accomodates** the speaker's consistent presupposition that Anna (as opposed to somebody else) is the theme, and marrying Manny (as opposed to someone else) is the rheme. This obviates criticism by Joshi (1990) and Pulman (1997).

Common Ground

- We follow Stalnaker and Thomason in assuming that common ground consists in *the set of propositions that a conversational participant supposes to be mutually agreed to for the purposes of the conversation*.

☡ $\neq$ the set of propositions that all participants actually believe.

Theme and Rheme

- In these terms, we can informally define theme and rheme as follows:
  - A theme is a part of the meaning of an utterance that some participant in the conversation supposes (or fails to suppose) **already to be** common ground;
  - A rheme is a part of the meaning of an utterance which some participant **makes** (or fails to make) common ground.
- Cf. Gussenhoven 1983 SELECTION/ADDITION
- Cf. Brazil 1997 REFERRING/PROCLAIMING.
- Cf. Roberts 1996; Ginzburg 1996 QUD

Unmarked Themes

- In cases where there is only one theme, known to all participants, the theme lacks any contrast and any accent.

- E.g., the following responses to the questions in (77) and (78) are possible:

(80) Q: I know Emma will marry ARNIM. But who will marry MANNY?
A: (ANNA)(will marry Manny).
H* LL%

(81) Q: I know Anna DATED ARNIM. But who will she MARRY?
A: (Anna will marry)(MANNY).
H* LL%

- Such non-contrastive themes are referred to as “unmarked” themes.

Unmarked Themes vs. All-Rheme utterances

- Sentences with unmarked themes are *ambiguous*, and a number of information structures (including the “all-rheme” utterance (82c) that is appropriate to the “out-of-the-blue” context) are acoustically indistinguishable.

(82) a. Q: What will Anna do?
A: $(\text{Anna will})_\theta(\text{ marry MANNY})_\rho$.
H* LL%
b. Q: What about Anna?
A: $(\text{Anna})_\theta(\text{will marry MANNY})_\rho$.
H* LL%
c. Q: What’s new?
A: $(\text{Anna will marry MANNY})_\rho$.
H* LL%

All-Rheme Utterances with Non-Final Accent

- In English, the following all-rheme example succeeds as an out-of-the-blue rheme just in case phoning is a background activity of the absent mother:

(83) Q: What's new?
A: (Your MOTHER called)$_\rho$.
H* LL%

☡ However, the possibility of such a subject-accented all-rheme utterance does not appear in general to extend to transitive examples like the following:

(84) Q: What's new?
A: #(ANNA married Manny)$_\rho$.
H* LL%

Dimension 3: Presence in Common Ground

- English accents are distinguished along an orthogonal dimension of whether their information unit is (or becomes) present in common ground (H*, L+H*, H*+L) or not (L*, L*+H, H+L*):

(85) a. You put my TROUSERS in the MICROWAVE!
H* H* LL%
b. You put my TROUSERS in the MICROWAVE?
L* L* LH%

- In (85a), the speaker marks the proposition as being made common ground.

- In (85b), the speaker marks the proposition as *NOT* being made common ground, thereby possibly implicating disbelief.

Dimension 4: Speaker/Hearer Agency

- A further dimension of intonational meaning is carried by the *boundaries*, rather than the accents.

- The level boundaries L, LL% and HL% mark information units as being (or not being) supposed to be (or made to be) common ground by the *speaker*.

- The rising boundaries H, LH% and HH% mark information units as being (or not being) supposed to be (or made to be) common ground by the *hearer*.

Generativity of Intonational Tunes in English

- The system relating these four dimensions of information structural meaning can be set out as follows, where θ signifies theme, and ρ signifies rheme, while $\top$ and $\neg$ signify success/failure of supposition/update over the Common Ground by the speaker/hearer agents S and H.

Generativity of Intonational Tunes in English

	$\top$	$\bot$
θ	L+H*	L*+H
ρ	H*, H*+L	L*, H+L*

Table 4: Meaning Elements Contributed by the Accents

$\mathbb{S}$	L, LL%, HL%
$\mathbb{H}$	H, HH%, LH%

Table 5: Meaning Element Contributed by the Boundaries

Semantics of Theme/Rheme, Common Ground, and Agency

- We define the common ground as a (sub)model $\mathfrak{C}$, and the property of a proposition holding in $\mathfrak{C}$ as a logical modality [C]. The thematic function of *being already supposed present in common ground* can then be represented as θ, and the rhematic function of *being made present in common ground* as ρ, defined as follows:[1]

(86) $\theta =_{def} \lambda p \lambda x.suppose([\mathrm{C}]theme\ p^{o} \wedge \forall a \in \{p^{a}\}\,[theme\ a \rightarrow a = p^{o}])x$

(87) $\rho =_{def} \lambda p \lambda x.[\mathrm{C}]update\ \mathfrak{C}\ p^{o}\ x \vee \exists t[theme\ t \wedge update\ \mathfrak{C}\ (p^{o}\ t)\ x$

—where:

---

[1] The latter definition is simplified here by omitting any mention of the alternative semantic value p^{a}.

1. p is a polymorphic variable ranging over pairs (p^o, p^a) where p^o is a function of any valency (including propositions of zero valency), and p^a is a function of the same valency that includes at least one free variable;
2. $\{p^a\}$ is the alternative set characterized by p^a;
3. $suppose$ can be thought of as a modal version of Beaver's (2001) fallible presupposition operator ∂—roughly, verify or update with respect to the common ground $\mathfrak{C}$;
4. the predicate *theme* is assumed to be directly interpreted in the common ground model $\mathfrak{C}$ as a (polymorphic) property $\mathfrak{theme}$.
5. $update$ can be thought of as a fallible update predicate which fails if its argument is not a proposition, and which either extends the common ground model $\mathfrak{C}$ by the denotation of a proposition p, or finds a theme t and extends it by the denotation of the result of applying p to t.
6. x is the agent S or H.

Alternative Semantics for CCG: Accents

- The proper name *Anna* bearing an H* accent has the following nominative category, among other case-like type-raised categories:

(88) $\text{A\small{NNA}} := S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho}) : \left\{ \begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array} \right\}$
H*

- A subject bearing no accent has the following category:

(89) $\text{Anna} := S_{\pi,\eta}/(S_{\pi,\eta}\backslash NP_{\pi,\eta}) : \left\{ \begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ anna \end{array} \right\}$

(Where Λ^o and Λ^a are identical as here we will write them as one e.g. $\lambda p.p\ anna$.)

Alternative semantics for CCG: Boundaries

- Boundaries are not properties of words or phrases, but independent string elements in their own right.

- They bear a category which “freezes” $\pm,\theta/\pm,\rho$-marked constituents as complete information-/intonation- structural units, making them unable to combine further with anything except similarly complete prosodic units.

- For example, the speaker-supposition- signaling LL% boundary bears the following category:

  (90) $\text{LL\%} := S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} : \lambda f.\pi(\eta f \mathrm{S})$

A Derivation

(91)

$$
\begin{array}{ccccc}
\textsc{Anna} & \text{married} & & \textsc{Manny} & . \\
\text{L}{*}{+}\text{H} & & \text{LH\%} & \text{H}{*} & \text{LL\%} \\
\overline{\hphantom{xxxxxxxxxxxxxxxx}}{}^{>\mathbf{T}} & \overline{\hphantom{xxxxxxxxxxxxxx}} & \overline{\hphantom{xxxxxxxxx}} & \overline{\hphantom{xxxxxxxxxxxxxxxx}}{}^{<\mathbf{T}} & \overline{\hphantom{xxxxxxxxxxxx}} \\
S_{\perp,\theta}/(S_{\perp,\theta}\backslash NP_{\perp,\theta}) & (S\backslash NP)/NP & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} \\
:\left\{\begin{array}{l}\lambda f.f\ \mathit{anna}\\ \lambda p.p\ v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.\mathit{married}\ xy & :\lambda f.\pi(\eta\ f\ \mathrm{H}) & :\left\{\begin{array}{l}\lambda p.p\ \mathit{manny}\\ \lambda p.p\ v_{\tau_{manny}}\end{array}\right\} & :\lambda g.\pi(\eta\ g\ \mathrm{S}) \\
\multicolumn{2}{c}{\overline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{}^{>\mathbf{B}}} & & & \\
\multicolumn{2}{c}{S_{\perp,\theta}/NP_{\perp,\theta}:\left\{\begin{array}{l}\lambda x.\mathit{married}\ x\ \mathit{anna}\\ \lambda x.\mathit{married}\ x\ v_{\tau_{anna}}\end{array}\right\}} & & & \\
\multicolumn{3}{c}{\overline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{}^{<}} & \multicolumn{2}{c}{\overline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{}^{<}} \\
\multicolumn{3}{c}{S_{\phi}/NP_{\phi}:\perp(\theta\left\{\begin{array}{l}\lambda x.\mathit{married}\ x\ \mathit{anna}\\ \lambda x.\mathit{married}\ x\ v_{\tau_{anna}}\end{array}\right\}\mathrm{H})} & \multicolumn{2}{c}{S_{\phi}\backslash(S_{\phi}/NP_{\phi}):\top(\rho\left\{\begin{array}{l}\lambda p.p\ \mathit{manny}\\ \lambda p.p\ v_{\tau_{manny}}\end{array}\right\}\mathrm{S})} \\
\multicolumn{5}{c}{\overline{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}{}^{<}} \\
\multicolumn{5}{c}{S_{\phi}:\top(\rho\left\{\begin{array}{l}\lambda p.p\ \mathit{manny}\\ \lambda p.p\ v_{\tau_{manny}}\end{array}\right\}\mathrm{S})(\perp(\theta\left\{\begin{array}{l}\lambda x.\mathit{married}\ x\ \mathit{anna}\\ \lambda x.\mathit{married}\ x\ v_{\tau_{anna}}\end{array}\right\}\mathrm{H}))} \\
\multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{5}{c}{S:\left\{\begin{array}{l}\mathit{married}\ \mathit{manny}\ \mathit{anna}\\ \mathit{married}\ v_{\tau_{manny}}\ v_{\tau_{anna}}\end{array}\right\}}
\end{array}
$$

“You do not suppose the question of who Anna (as opposed to anyone else) married to be common ground, I make it common ground that she married Manny (as opposed to anyone else)”

Remarks

- Theme/Rheme marking is projected onto phrasal constituents by syntactic derivation alone.
- It is bounded by combination of the phrase with a boundary tone.
- No independent extrasyntactic mechanism of “Focus Projection” is needed to achieve the semantics of “broad focus”

Unmarked Theme

- (92) *Prosodic phrase promotion rule* (%)
  $$S\$_{\pi,\eta} : f \quad \Rightarrow_{\%} \quad S\$_{\phi} : \pi(\eta f \mathrm{S})$$

(93)

$$\begin{array}{cccc}
\text{Anna} & \text{married} & \text{M{\small ANNY}} & \\
 & & \text{H}* & \text{LL\%} \\
\overline{\quad\quad}^{>\mathbf{T}} & \overline{\quad\quad\quad\quad} & \overline{\quad\quad\quad\quad}^{<\mathbf{T}} & \overline{\quad\quad} \\
S/(S\backslash NP) & (S\backslash NP)/NP : & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} : \\
: \lambda f.f\ anna & \lambda x.\lambda y.married\ xy & : \left\{ \begin{array}{l} \lambda p.p\ manny) \\ \lambda p.p\ v_{\tau_{manny}}) \end{array} \right\} & : \lambda g.\pi(\eta\ g\ \mathrm{S})
\end{array}$$

$$\overline{\quad S/NP : \lambda x.married\ x\ anna \quad}^{>\mathbf{B}}$$

$$\overline{\quad S_{\phi}/NP_{\phi} : \pi(\eta\,\{\lambda x.married\ x\ anna\}\,\mathrm{S}) \quad}^{\%} \qquad \overline{\quad S_{\phi}\backslash(S_{\phi}/NP_{\phi}) : \top(\rho \left\{ \begin{array}{l} \lambda p.p\ manny \\ \lambda p.p\ v_{\tau_{manny}} \end{array} \right\} \mathrm{S}) \quad}^{<}$$

$$\overline{\quad S_{\phi} : \top(\rho \left\{ \begin{array}{l} \lambda p.p\ manny \\ \lambda p.p\ v_{\tau_{manny}} \end{array} \right\} \mathrm{S})(\pi(\eta(\lambda x.married\ x\ anna)\mathrm{S})) \quad}^{<}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S : \left\{ \begin{array}{l} married\ manny\ anna \\ married\ v_{\tau_{manny}}\ anna \end{array} \right\}$$

Accent-Final All-Rheme Utterance

- The above contour also allows an alternative analysis as a all-rheme utterance:

(94)

$$\begin{array}{cccc}
\text{Anna} & \text{married} & \begin{array}{c}\text{M{\scriptsize ANNY}}\\ \text{H}*\end{array} & \text{LL\%}\\
\overline{}^{\,>\mathbf{T}} & \overline{} & \overline{\phantom{S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho})}}^{\,<\mathbf{T}} & \overline{\phantom{S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta}}}\\
S/(S\backslash NP) & (S\backslash NP)/NP & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta}\\
:\lambda f.f\ anna & :\lambda x.\lambda y.married\ xy & :\left\{\begin{array}{c}\lambda p.p\ manny\\ \lambda p.p\ v_{\tau_{manny}}\end{array}\right\} & :\lambda g.\pi(\eta\ g\ \mathrm{S})
\end{array}$$

$$\overline{S/NP:\lambda x.married\ x\ anna}^{\,>\mathbf{B}}$$

$$\overline{S_{\top,\rho}:\left\{\begin{array}{l}married\ manny\ anna\\ married\ v_{\tau_{manny}}\ anna\end{array}\right\}}^{\,<}$$

$$\overline{S_{\phi}:\top(\rho\left\{\begin{array}{l}married\ manny\ anna\\ married\ v_{\tau_{manny}}\ anna\end{array}\right\}\mathrm{S})}^{\,<}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S:\left\{\begin{array}{l}married\ manny\ anna\\ married\ v_{\tau_{manny}}\ anna\end{array}\right\}$$

"I make it common ground that Anna married Manny"

Non-Final Accent All-Rheme Utterance

- Since English unaccented verbs are also unmarked as to the theme/rheme dimension, there is also a all-rheme analysis for intransitives like the following:

$$
\begin{array}{llll}
(95)\ \text{Your} & \underset{\text{H}*}{\text{MOTHER}} & \text{called} & \text{LL\%} \\
\multicolumn{2}{l}{\overline{}{>}\mathbf{T}} & \overline{} & \overline{} \\
\multicolumn{2}{c}{S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho})} & S\backslash NP & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} \\
\multicolumn{2}{c}{:\left\{\begin{array}{l}\lambda f.f\,(your\ mother)\\ \lambda f.f\,(your\ v_{\tau_{mother}})\end{array}\right\}} & :\lambda x.called\ x & :\lambda g.\pi(\eta\ g\ \mathrm{S}) \\
\multicolumn{3}{l}{\overline{}{>}} & \\
\multicolumn{3}{c}{S_{\top,\rho}:\left\{\begin{array}{l}called\ (your\ mother)\\ called\ (your\ v_{\tau_{mother}})\end{array}\right\}} & \\
\multicolumn{4}{r}{\overline{}{<}} \\
\multicolumn{4}{c}{S_{\phi}:\top(\rho\left\{\begin{array}{l}called\ (your\ mother)\\ called\ (your\ v_{\tau_{mother}})\end{array}\right\}\mathrm{S})} \\
\multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
\multicolumn{4}{c}{S:\left\{\begin{array}{l}called\ (your\ mother)\\ called\ (your\ v_{\tau_{mother}})\end{array}\right\}}
\end{array}
$$

No English Non-final Rheme-Accented Transitives

- We are free to make unaccented accusatives theme-marked:

(96) Manny $:= (S_{\top,\theta}\backslash NP_{\top,\theta})\backslash((S_{\top,\theta}\backslash NP_{\top,\theta})/NP_{\top,\theta}) : \lambda p.p\ manny'$
$S_{\top,\theta}\backslash(S_{\top,\theta}/NP_{\top,\theta}) : \lambda p.p\ manny'$

No English Non-final Rheme-Accented Transitives

- This category allows a derivation for (80) with an unmarked theme:

(97)

$$\begin{array}{cccc}
\begin{array}{c}\textsc{Anna}\\ \mathrm{H*}\end{array} & \text{married} & \text{Manny} & \mathrm{L\dot{L}\%}\\
\hline
\xrightarrow{\quad >\mathbf{T}\quad} & & \xleftarrow{\quad <\mathbf{T}\quad} & \\
S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho}) & (S\backslash NP)/NP & (S_{\top,\theta}\backslash NP_{\top,\theta})\backslash((S_{\top,\theta}\backslash NP_{\top,\theta})/NP_{\top,\theta}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta}\\
:\left\{\begin{array}{l}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.married\ xy & :\lambda p.p\ manny & :\lambda g.\pi(\eta\ g\ \mathrm{S})
\end{array}$$

$$\xrightarrow{\qquad\qquad\qquad\%}$$

$$S_{\phi}/(S_{\phi}\backslash NP_{\phi}):\top(\rho\left\{\begin{array}{l}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}})\end{array}\right\}\mathrm{S}$$

$$\xleftarrow{\qquad\qquad\qquad<}$$

$$S_{\top,\theta}\backslash NP_{\top,\theta}:\lambda x.married\ manny\ x$$

$$\xleftarrow{\qquad\qquad\qquad<}$$

$$S_{\phi}\backslash NP_{\phi}:\top(\theta\{\lambda x.married\ manny\ x\}\mathrm{S})$$

$$\xrightarrow{\qquad\qquad\qquad>}$$

$$S_{\phi}:\top(\rho\left\{\begin{array}{l}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}}\end{array}\right\}\mathrm{S})(\top(\theta\,\{\lambda x.married\ manny\ x\}\,\mathrm{S}))$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S:\left\{\begin{array}{l}married\ manny\ anna\\ married\ manny\ v_{\tau_{anna}}\end{array}\right\}$$

"I make Anna (as opposed to anyone else) common ground, I suppose the question of who married Manny to be common ground"

Non-Final Accent Rheme Disallowed

- However, the theme-marked object category prevents an all-rheme analysis

(98)

$$\begin{array}{cccc}
\textsc{Anna} & \text{married} & \text{Manny} & \\
\text{H}* & & & \mathrm{L}\dot{\mathrm{L}}\% \\
\overline{}{>}\mathbf{T} & \overline{} & \overline{}{<}\mathbf{T} & \overline{} \\
S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho}) & (S\backslash NP)/NP & (S_{\top,\theta}\backslash NP_{\top,\theta})\backslash((S_{\top,\theta}\backslash NP_{\top,\theta})/NP_{\top,\theta}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} \\
:\left\{\begin{array}{c}\lambda f.f\;anna\\ \lambda f.f\;v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.married\;xy & :\lambda p.p\;manny & :\lambda g.\pi(\eta\;g\;\mathrm{S}) \\
 & \multicolumn{2}{c}{\overline{}{<}} & \\
 & \multicolumn{2}{c}{S_{\top,\theta}\backslash NP_{\top,\theta}:\lambda x.married\;manny\;x} & \\
\multicolumn{3}{c}{\overline{}{*}} &
\end{array}$$

"#I make it common ground that Anna (as opposed to anyone else) married Manny"

- Hence the anomaly of the out-of-the-blue utterance (84) with the same contour. (The all-theme version is OK in context.)

Another Impossible Non-Final Accent Rheme

- The same observation applies to examples like the following, consistent with Ladd's 1996 analysis of related examples (cf. Steedman 2000a:119, (62)):

(99) Q: Has Anna read *Ulysses*?
A: (Anna doesn't READ)$_\rho$ (books)$_\theta$.
H* LL%

- (99) cannot be uttered out of the blue.

"Anchoring" Objects and Superman Sentences

- Examples like the following can be uttered out-of-the-blue:

(100) a. I have to SEE a guy.
b. You need to TALK to someone.
c. Your MOTHER called you.
d. He was reading SUPERMAN to some kid. (Neeleman and Szendrői 2004)

- Such non-theme-marked objects are reminiscent of Prince's 1981 "anchoring" given modifiers:

(101) Anna married a GUY I know.

(The alternatives are *people*, not *people I know*.)

Focusing Particles such as *only*

(102) $\text{only} := NP^{\uparrow}/NP^{\uparrow} : \lambda np \lambda p \lambda \ldots . np^{o}\; p\; \ldots \wedge \forall a \in \{np^{a}\}[a\; p\; \ldots \rightarrow (a = np^{o})]$

(103)

$$\begin{array}{ccccc}
\text{Anna} & \text{married} & \text{only} & \begin{array}{c}\text{M{\small ANNY}}\\ \text{H}*\end{array} & \text{L}\dot{\text{L}}\% \\
\overline{}{}^{>\mathbf{T}} & \overline{} & \overline{\phantom{NP^{\uparrow}/NP^{\uparrow}NP^{\uparrow}/NP^{\uparrow}}} & \overline{\phantom{S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho})}}{}^{<\mathbf{T}} & \overline{\phantom{S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta}}} \\
S/(S\backslash NP) & (S\backslash NP)/NP & NP^{\uparrow}/NP^{\uparrow} & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) & S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta} \\
: \lambda f.f\; anna & : \lambda x.\lambda y.married\; xy & : \lambda np\lambda p.np^{o}\; p \wedge \forall a \in \{p^{a}\}[a\; p \rightarrow (a = np^{o})] & : \left\{\begin{array}{c}\lambda p.p\; manny\\ \lambda p.p\; v_{\tau_{manny}}\end{array}\right\} & : \lambda g.\pi(\eta\; g\; \mathrm{S})
\end{array}$$

$$\begin{array}{cc}
\overline{}{}^{>\mathbf{B}} & \overline{\phantom{S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) : \lambda p.p\; manny \wedge \forall a}}{}^{>} \\
S/NP & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) \\
: \lambda x.married\; x\; anna & : \lambda p.p\; manny \wedge \forall a \in \left\{\lambda p.p\; v_{\tau_{manny}}\right\}[a\; p \rightarrow (a = \lambda p.p\; manny)] \\
\overline{}{}^{\%} & \overline{\phantom{S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho}) : \lambda p.p\; manny \wedge \forall a\; S\$_{\phi}}}{}^{<} \\
S_{\phi}/NP_{\phi} & S_{\phi}\backslash(S_{\phi}/NP_{\phi}) \\
: \pi(\eta\,\{\lambda x.married\; x\; anna\}\mathrm{S}) & : \top(\rho\,\{\lambda p.p\; manny \wedge \forall a \in \left\{\lambda p.p\; v_{\tau_{manny}}\right\}[a\; p \rightarrow (a = \lambda p.p\; manny)]\}\mathrm{S})
\end{array}$$

$$\begin{array}{c}
\overline{S_{\phi} : \top(\rho\,\{\lambda p.p\; manny \wedge \forall a \in \left\{\lambda p.p\; v_{\tau_{manny}}\right\}[a\; p \rightarrow (a = \lambda p.p\; manny)]\}\mathrm{S})(\pi(\eta(\lambda x.married\; x\; anna)\mathrm{S}))}{}^{<} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
S : married\; manny\; anna \wedge \forall a \in \left\{\lambda p.p\; v_{\tau_{manny}}\right\}[a(\lambda x.married\; x\; anna) \rightarrow (a = \lambda p.p\; manny)]
\end{array}$$

"I suppose the question of who Anna married to be common ground, I make it common ground she married Manny and none of the alternatives."

Only and “Second Occurence Focus”

- Who ate only TOFU?

(104)

$$\begin{array}{ccccc}
\text{A}\textsc{nna} & \text{ate} & \text{only} & \text{tofu} & \\
\text{H}* & & & & \text{L}\dot{\text{L}}\% \\
\overline{S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho})}^{>\mathbf{T}} & \overline{(S\backslash NP)/NP} & \overline{NP^{\uparrow}/NP^{\uparrow}} & \overline{(S\backslash NP)\backslash((S\backslash NP)/NP)}^{<\mathbf{T}} & \overline{S\$_{\phi}\backslash_{\star}S\$_{\pi,\eta}} \\
:\left\{\begin{array}{c}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.ate\ xy & :\lambda np\lambda p\lambda y.np^{o}\ p\ y\wedge\forall a\in p^{a}[a\ p\ y\rightarrow(a=np^{o})] & :\lambda p\lambda y.p\ tofu\ y & :\lambda g.\pi(\eta\ g\ \mathrm{S})
\end{array}$$

$$\overline{\qquad}^{\%}\quad S_{\phi}/(S_{\phi}\backslash NP_{\phi}) : \top(\rho\left\{\begin{array}{c}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}}\end{array}\right\}\mathrm{S})$$

$$\overline{\qquad}^{>}\quad (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.\lambda y.p\ tofu\ y\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ p\ y\rightarrow a=\lambda p\lambda y.p\ tofu\ y]$$

$$\overline{\qquad}^{<}\quad S\backslash NP : \lambda y.ate\ tofu\ y\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ ate\ y\rightarrow a=\lambda p\lambda y.p\ tofu\ y]$$

$$\overline{\qquad}^{<}\quad S_{\phi}\backslash NP_{\phi} : \pi(\eta\,\{\lambda y.ate\ tofu\ y\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ ate\ y\rightarrow a=\lambda p\lambda y.p\ tofu\ y]\}\,\mathrm{S})$$

$$\overline{\qquad}^{>}\quad S_{\phi} : \top(\rho\left\{\begin{array}{c}\lambda f.f\ anna\\ \lambda f.f\ v_{\tau_{anna}}\end{array}\right\}\mathrm{S})(\pi(\eta\,\{\lambda y.ate\ tofu\ y\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ ate\ y\rightarrow a=\lambda p\lambda y.p\ tofu\ y]\}\,\mathrm{S}))$$

$$\cdots\cdots\cdots\quad S:\left\{\begin{array}{l}ate\ tofu\ anna\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ ate\ anna\rightarrow(a=\lambda p\lambda y.p\ tofu\ y)]\\ ate\ tofu\ v_{\tau_{anna}}\wedge\forall a\in\{\lambda p\lambda y.p\ tofu\ y\}\,[a\ ate\ v_{\tau_{anna}}\rightarrow(a=\lambda p\lambda y.p\ tofu\ y)]\end{array}\right\}$$

Second occurrence isn't contrastive

- Category (102) does not require Λ^o, Λ^a values to be distinct.
- (104) does not evoke any alternatives to tofu, because tofu is unaccented.
- But that is because we have already disposed of the alternatives to tofu in a previous utterance such as *Who ate only* TOFU*?*
- (104) is only admissible at all in such contexts, in which by definition such alternatives are not only given, but also explicitly denied by the previous utterance. So why should they be evoked?
- Cf. Rooth 1992 *People who* GROW *rice generally only* EAT *rice*.

◇ The semantics of *Only* is independent of focus, at least in the sense of contrast with alternatives.

“Nested” Focus

- Adverbial *only* and *also*

(105) only := $(S\backslash NP)/\$/(S\backslash NP)/\$: \lambda p \lambda x \ldots .p^o x \ldots \wedge \forall a \in \{p^a\}\,[a\, x \ldots \rightarrow (a = p^o)]$

(106) also := $(S\backslash NP)/\$/(S\backslash NP)/\$: \lambda p \lambda x \ldots .p^o x \ldots \wedge \exists a \in \{p^a\}\,[a\, x \ldots \wedge a \neq p^o)]$

- Wold (1996) notes that Rooth makes the wrong prediction for “nested focus” examples like the following elaborated answer to the question “Who did John introduce to Bill?”:

(107) a. Anna only introduced SUE to Bill.
b. Anna also only introduced Sue to TOM

Wold's Problem

- The available reading supported by the context is one in which Anna introduced Sue and no one else to both Bill and Tom.

- However, if both the second mention focus and the novel focus in the second sentence are captured by *only*, (b) means (counterfactually) that Anna introduced only Sue to only Tom.

- Because the CCG account of the projection of rheme focus (that is, accent) is strictly via the derivation, the preferred consistent reading is correctly derived here (brace yourself):

Anna — also — only — introduced — Sue — to
H∗

Anna:

$$\dfrac{}{NP^{\uparrow}}{>}\mathbf{T}$$

$$: \lambda f.f\ anna$$

$$\dfrac{}{NP^{\uparrow}_{\phi}}\%$$

$$\pi(\eta\{\lambda f.f\ anna\})$$

also:

$$(S\backslash NP)/(S\backslash NP)$$

$$: \lambda p\lambda y.p^{o}y \land \exists a \in \{p^{a}\}\,[a\ y \land (a \neq p^{o})]$$

only:

$$((S\backslash NP)/PP)/((S\backslash NP)/PP)$$

$$\lambda p\lambda z\lambda y.p^{o}zywedge\forall a \in \{p^{a}\}\,[a\ zy \rightarrow (a = p^{o})]$$

introduced:

$$((S\backslash NP)/PP)/NP$$

$$: \lambda x\lambda z\lambda y.introduced\ zxy$$

Sue:

$$NP^{\uparrow}$$

$$: \lambda g.g\ sue$$

to H∗:

$$: \left\{ \begin{array}{l} \lambda h. \\ \lambda h. \end{array} \right.$$

introduced Sue $<$:

$$(S\backslash NP)/PP$$

$$\lambda z\lambda y.introduce\ z\ sue\ y$$

only introduced Sue $>$:

$$(S\backslash NP)/PP$$

$$: \lambda z\lambda y.introduce\ z\ sue\ y$$

$$\land \forall a \in \{\lambda z\lambda y.introduce\ z\ sue\ y\}\,[a\ y \rightarrow (a = (\lambda z\lambda y.introduce\ z\ sue\ y))]$$

$$S_{\top,\rho}\backslash NP_{\top,\rho}$$

$$: \left\{ \begin{array}{l} \lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ tom\ s \\ \lambda y.introduce\ v_{\tau_{tom}}\ sue\ y \land \forall a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ v_{\tau_{tom}} \end{array} \right.$$

$$S_{\top,\rho}\backslash NP_{\top,\rho}$$

$$\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ tom\ sue\ y)]$$

$$\land \exists a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y \land \forall a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\}$$

$$[a\ y \rightarrow (a = \lambda y.introduce\ v_{\tau_{tom}}\ sue\ y)]\}[a\ y \land (a \neq (\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ to$$

$$S_{\phi}\backslash NP_{\phi}$$

$$\top(\rho\{\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}[a\ y \rightarrow (a = (\lambda y.introduce\ tom\ sue\ y))]$$

$$\land \exists a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y \land \forall a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\}[a\ y \rightarrow (a = \lambda y.introduce\ v_{\tau_{tom}}\ sue\ y)]\}$$

$$[a\ y \land (a \neq (\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ tom\ sue\ y)]))]\}\mathrm{S}$$

$$S_{\phi}$$

$$\pi(\eta(\lambda f.f\ anna))(\top(\rho\{\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}[a\ y \rightarrow (a = (\lambda y.introduce\ tom\ sue\ y))]$$

$$\land \exists a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y \land \forall a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\}[a\ y \rightarrow (a = \lambda y.introduce\ v_{\tau_{tom}}\ sue\ y)]\}$$

$$[a\ y \land (a \neq (\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ tom\ sue\ y)]))]\}\mathrm{S}))$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$introduce\ tom\ sue\ anna \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}[a\ anna \rightarrow (a = (\lambda y.introduce\ tom\ sue\ y))]$$

$$\land \exists a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\} \land \forall a \in \{\lambda y.introduce\ v_{\tau_{tom}}\ sue\ y\}[a\ y \rightarrow (a = \lambda y.introduce\ v_{\tau_{tom}}\ sue\ y)]\}$$

$$[a\ anna \land (a \neq (\lambda y.introduce\ tom\ sue\ y \land \forall a \in \{\lambda y.introduce\ tom\ sue\ y\}\,[a\ y \rightarrow (a = \lambda y.introduce\ tom\ sue\ y)]))]\}\mathrm{S}))$$

Coda: Intonational Phrases are Constituents

- The present theory makes intonation structure as defined by intonational boundaries isomorphic with the top-level constituency of surface syntactic derivational structure.

- Surface derivational structure is also, as we have seen, isomorphic to coordinate structure and the domain of relativization.

- It follows that this theory predicts the strongest possible relation between intonation structure, information structure, coordination, and movement, as follows (cf. Steedman 1991):
  - All and only those substrings that can either undergo coordination or be extracted over can be intonational phrases and elements of information structure, and *vice versa*.

Syntactic Islands are Intonational Islands

- (108) a. *(Three MATHEMATICIANS) (in ten derive a LEMMA).
  L+H* LH% H* LL%
  b. *(Seymour prefers the NUTS) (and bolts APPROACH).
  L+H* LH% H* LL%
  c. *(They only asked whether I knew the woman who CHAIRS) (the ZONING board).
  L+H* LH% H* LL%

- (109) a. *Three mathematicians in ten derive a lemma and in a hundred can cook a boiled egg.
  b. *The nuts which Seymour prefers and bolts approach
  c. *Which boards did they ask whether you knew the woman who chairs?

Coda: The Real Problem of Language Acquisition

- If all there was to language was an encoding of propositions that the child already has in mind, as in part III, it is not clear why they should bother to learn language at all, as Clark (2004) points out, in defence of a PAC learning model (!).

- We know from Fernald (1993) that infants are sensitive to interpersonal aspects 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.

Towards a More Realistic Syntax and Semantics

- For example, it is likely that the child’s representation of the utterance “More doggies! is more like (110):

(110)

$$\begin{array}{ccl}
\text{More doggies} & ! & \\
\text{H}* \quad \text{H}* & \text{LL\%} & \\
\cline{1-1}\cline{2-2}
NP^{\uparrow}_{+,\rho} & X_{\phi}\backslash_{\star}X_{\pi,\eta} & \\
:\left\{\begin{array}{l}\lambda p.p\,(more'dogs')\\ \lambda p.p\,(v_{more'}v_{dogs'})\end{array}\right\} & :\lambda g.\pi[\mathrm{S}]\eta\; g & \\
\cline{1-2}
\multicolumn{2}{c}{NP^{\uparrow}_{\phi} : [S]\rho\left\{\begin{array}{l}\lambda p.p\,(more'dogs')\\ \lambda p.p\,(v_{more'}v_{dogs'})\end{array}\right\}} & <
\end{array}$$

“Mummy makes the property afforded by more dogs (as opposed to the alternatives) common ground.”

- Consider the child then faced with the following, from Fisher and Tokura (1996), as the next utterance (cf. Steedman 1996a):

(111)

$$\begin{array}{ccccc}
\text{You} & \overset{\text{LIKE}}{\text{H}*} & \text{L} & \text{the doggies!} & \text{LL\%} \\
\hline
S/(S\backslash NP) & (S\backslash NP)/NP & X_{\phi}\backslash_{\star}X_{\pi,\eta} & S_{\phi}\backslash(S_{\phi}/NP_{\phi}) & \\
:\lambda p.p\ you' & :\left\{\begin{array}{l}\lambda x\lambda y.like'xy\\ \lambda x\lambda y.v_{like'}xy\end{array}\right\} & :\lambda g.\pi[\mathrm{S}]\eta\ g & :[S]\eta\lambda q.q\ dogs' &
\end{array}$$

$$\xrightarrow{\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad}{>}\mathbf{B}$$

$$S/NP:\left\{\begin{array}{l}\lambda x.like'x\ you'\\ \lambda x.v_{like'}x\ you'\end{array}\right\}$$

$$\xleftarrow{\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad}{<}$$

$$S_{\phi}/NP_{\phi}:[S]\rho\left\{\begin{array}{l}\lambda x.like'x\ you'\\ \lambda x.v_{like'}x\ you'\end{array}\right\}$$

$$\xleftarrow{\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad}{<}$$

$$S_{\phi}:([S]\theta\lambda p.p\ dogs')([S]\rho\left\{\begin{array}{l}\lambda x.like'x\ you'\\ \lambda x.v_{like'}x\ you'\end{array}\right\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S:\left\{\begin{array}{l}like'dogs'you'\\ v_{like'dogs'you'}\end{array}\right\}$$

“Mummy supposes what property the dogs afford to be common ground, and makes it common ground it’s me liking (as opposed to anything else) them.”

CCG as “Motherese”

- Fisher points out that the L boundary after the verb makes the intonation structure inconsistent with standard assumptions about surface constituency.
- However, this intonation structure is homomorphic to the CCG derivation, 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 $>$**B**, the child gets information about the probability of applying the composition rule to the first two categories to yield $S\backslash NP$.
- Thus, the child can build the entire parsing model in parallel with learning the grammar,
- Long range dependencies come for free.

Coda: Information Structure and Gapping

- I conjecture that the Alternative Logical Form defined in this section is the locus of the Gap information in the English gapped construction.

(112)

$$\begin{array}{ccc}
\textsc{Anna}\ \mathit{married}\ \textsc{Manny} & \text{and} & \textsc{Tom}\ \textsc{Sue}\\
\overline{S:\left\{\begin{array}{l} married'manny'anna'\\ married'v_{\tau_{manny'}}v_{\tau_{anna'}}\end{array}\right\}} & \overline{(X\backslash_{\star}X)/_{\star}X} & \overline{S\backslash((S/NP)/NP_{SG}):\left\{\begin{array}{l}\lambda tv.tv\ sue'tom'\\ \lambda tv.tv\ v_{\tau_{sue'}}v_{\tau_{tom'}}\end{array}\right\}}
\end{array}$$

- This would fill a hole in the account of gapping as constituent coordination in Steedman (1990)

Conclusion

- Intonation Structure is just Surface Syntactic Derivation, a.k.a. PF.
- Information Structure is just Logical Form, a.k.a. LF.
- PF and LF are the only "interface levels"
- LF is the only structural representational level.

◇ covert/overt MOVE (a.k.a. COPY/DELETE ETC.) = MERGE = LF surface composition

References

Abney, Steven, 1996. “Statistical Methods and Linguistics.” In Judith Klavans and Philip Resnik (eds.), *The Balancing Act*, Cambridge, MA: MIT Press. 1–26.

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

Alishahi, Afra and Stevenson, Suzanne, 2008. “A Computational Model of Early Argument Structure Acquisition.” *Cognitive Science* 32:789–834.

Altmann, Gerry and Steedman, Mark, 1988. "Interaction with Context During Human Sentence Processing." *Cognition* 30:191–238.

Auli, Michael and Lopez, Adam, 2011. "A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, OR: ACL, 470–480.

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

Baldridge, Jason and Kruijff, Geert-Jan, 2003. "Multi-Modal Combinatory Categorial Grammar." In *Proceedings of 11th Annual Meeting of the European Association for Computational Linguistics*. Budapest, 211–218.

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

Beaver, David, 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford, CA: CSLI Publications.

Bever, Thomas, 1970. “The Cognitive Basis for Linguistic Structures.” In John Hayes (ed.), *Cognition and the Development of Language*, New York: Wiley. 279–362.

Błaszczak, Joanna and Gärtner, Hans-Martin, 2005. “Intonational Phrasing, Discontinuity, and the Scope of Negation.” *Syntax* 8:1–22.

Bolinger, Dwight, 1965. *Forms of English*. Cambridge, MA: Harvard University Press.

Börschinger, Benjamin, Jones, Bevan K., and Johnson, Mark, 2011. "Reducing Grounded Learning Tasks To Grammatical Inference." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1416–1425.

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, Geneva*. ACL, 1240–1246.

Bos, Johan and Markert, Katja, 2005. "Combining Shallow and Deep NLP Methods for Recognizing Textual Entailment." In *Proceedings of the First PASCAL Challenge Workshop on Recognizing Textual Entailment*. `http://www.pascal-network.org/Challenges/RTE/`: Pascal, 65–68.

Brazil, David, 1997. *The Communicative Value of Intonation in English*. Cambridge University Press, second edition.

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

Buttery, Paula, 2006. *Computational Models for First Language Acquisition*. Ph.D. thesis, University of Cambridge.

Calhoun, Sasha, 2006. *Intonation and Information Structure in English*. Ph.D. thesis, University of Edinburgh.

Calhoun, Sasha, 2010. “The Centrality of Metrical Structure in Signaling Information Structure: A Probabilistic Perspective.” *Language* 86:1–42.

Carey, Sue and Bartlett, Elsa, 1978. "Acquiring a Single New Word." In *Proceedings of the 15th Stanford Child Language Conference*. Stanford, Papers and Reports on Child Language Development, 15, 17–29.

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.

Charniak, Eugene, 2001. "Immediate-Head Parsing for Language Models." In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. Toulouse: ACL, 116–123.

Charniak, Eugene, Goldwater, Sharon, and Johnson, Mark, 1998. "Edge-Based Best-First Chart Parsing." In *Proceedings of the 6th Workshop on Very Large Corpora, Montreal, August*. 127–133.

Chelba, Ciprian and Jelinek, Frederick, 1998. “Exploiting Syntactic Structure for Language Modeling.” In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*. Montreal: Coling/ACL, 225–231.

Chomsky, Noam, 1957. *Syntactic Structures*. The Hague: Mouton.

Clark, Alexander, 2004. “Grammatical Inference and First Language Acquisition.” In *CoLing Workshop on Psycho-computational Models of Human Language Acquisition*.

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 Annual Meeting of the Association for Computational Linguistics*. Barcelona, Spain: ACL, 104–111.

Clark, Stephen and Curran, James R., 2006. "Partial Training for a Lexicalized Grammar Parser." In *Proceedings of the Human Language Technology Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL '06)*. New York: ACL.

Clark, Stephen, Hockenmaier, Julia, and Steedman, Mark, 2002. "Building Deep Dependency Structures with a Wide-Coverage CCG Parser." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, 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: ACL, 111–118.

Collins, Michael, 1997. “Three Generative Lexicalized Models for Statistical Parsing.” In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*. Madrid: ACL, 16–23.

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

Dowty, David, 1994. “The Role of Negative Polarity and Concord Marking in Natural Language Reasoning.” In *Proceedings of the 4th Conference on Semantics and Theoretical Linguistics*. Rochester, NY: CLC Publications, Cornell University.

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: Morgan Kaufmann, 79–86.

Etzioni, Oren, Banko, Michele, and Cafarella, Michael, 2007. “Machine Reading.” In *Proceedings of AAAI Spring Symposium on Machine Reading*. Menlo Park, CA: AAAI Press.

Fernald, Anne, 1993. “Approval and Disapproval: Infant Responsiveness to Vocal Affect in Familiar and Unfamiliar Languages.” *Child Development* 64:657–667.

Fisher, Cynthia and Tokura, Hisayo, 1996. “Prosody in Speech to Infants: Direct and Indirect Acoustic Cues to Syntactic Structure.” In James Morgan and Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, Erlbaum. 343–363.

Fodor, Janet Dean, 1998. “Unambiguous Triggers.” *Linguistic Inquiry* 29:1–36.

Gildea, Dan, 2001. “Corpus Variation and Parser Performance.” In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*. Pittsburgh, PA, 167–202.

Ginzburg, Jonathan, 1996. “Interrogatives: Questions, Facts, and Dialogue.” In Shalom Lappin (ed.), *Handbook of Contemporary Semantic Theory*, Oxford: Blackwell. 385–422.

Gleitman, Lila, 1990. “The Structural Sources of Verb Meanings.” *Language Acquisition* 1:1–55.

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

Gussenhoven, Carlos, 1983. *On the Grammar and Semantics of Sentence Accent*. Dordrecht: Foris.

Halliday, Michael, 1967. *Intonation and Grammar in British English*. The Hague: Mouton.

Harrington, Brian and Clark, Stephen, 2009. “ASKNet: Creating and Evaluating Large Scale Integrated Semantic Networks.” *International Journal of Semantic Computing* 2:343–364.

Hepple, Mark, 1987. *Methods for Parsing Combinatory Grammars and the Spurious Ambiguity Problem*. Master’s thesis, University of Edinburgh.

Hindle, Donald and Rooth, Mats, 1993. “Structural Ambiguity and Lexical Relations.” *Computational Linguistics* 19:103–120.

Hockenmaier, Julia, 2003. *Data and Models for Statistical Parsing with CCG*. Ph.D. thesis, School of Informatics, University of Edinburgh.

Hockenmaier, Julia, Bierner, Gann, and Baldridge, Jason, 2004. "Extending the Coverage of a CCG System." *Journal of Logic and Computation* 2:165–208.

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 Association for Computational Linguistics*. Philadelphia, 335–342.

Hockenmaier, Julia and Steedman, Mark, 2007. "CCGbank: a Corpus of CCG Derivations and Dependency Structures Extracted from the Penn Treebank." *Computational Linguistics* 33:355–396.

Hoffman, Matthew., Blei, David, and Bach, Francis, 2010. "Online Learning for Latent Dirichlet Allocation." *Advances in Neural Information Processing Systems* 23:856–864.

Jelinek, Fred and Lafferty, John, 1991. "Computation of the Probability of Initial Substring Generation by Stochastic Context-Free Grammars." *Computational Linguistics* 17:315–323.

Joshi, Aravind, 1988. "Tree-Adjoining Grammars." In David Dowty, Lauri Karttunen, and Arnold Zwicky (eds.), *Natural Language Parsing*, Cambridge: Cambridge University Press. 206–250.

Joshi, Aravind, 1990. “Phrase Structure and Intonation: Comments on the Papers by Marcus and Steedman.” In Gerry Altmann (ed.), *Cognitive Models of Speech Processing*, Cambridge, MA: MIT Press. 457–482.

Karttunen, Lauri, 1976. “Discourse Referents.” In James McCawley (ed.), *Syntax and Semantics*, Academic Press, volume 7. 363–385.

Koller, Alexander and Thater, Stefan, 2006. “An Improved Redundancy Elimination Algorithm for Underspecified Descriptions.” In *Proceedings of the International Conference on Computational Linguistics/Association for Computational Linguistics*. Sydney: Coling/ACL, 409–416.

Kor, Kian Wei, 2005. *Improving Answer Precision and Recall of List Questions*. Ph.D. thesis, Edinburgh.

Kwiatkowski, Tom, Steedman, Mark, Zettlemoyer, Luke, and Goldwater, Sharon,

2012. “A Probabilistic Model of Language Acquisition from Utterances and Meanings.” In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*. Avignon: ACL, 234–244.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2010. “Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: ACL, 1223–1233.

Kwiatkowski, Tom, Zettlemoyer, Luke, Goldwater, Sharon, and Steedman, Mark, 2011. “Lexical Generalization in CCG Grammar Induction for Semantic Parsing.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1512–1523.

Ladd, D. Robert, 1996. *Intonational Phonology*. Cambridge: Cambridge University Press. 2nd edition revised 2008.

Liang, Percy, Jordan, Michael, and Klein, Dan, 2011. “Learning Dependency-Based Compositional Semantics.” In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, OR: ACL, 590–599.

Lu, Wei, Ng, Hwee Tou, Lee, Wee Sun, and Zettlemoyer, Luke S., 2008. “A Generative Model for Parsing Natural Language to Meaning Representations.” In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: ACL, 783–792.

MacCartney, Bill and Manning, Christopher D., 2007. “Natural Logic for Textual Inference.” In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Prague: ACL, 193–200.

MacWhinney, Brian, 2005. “Item Based Constructions and the Logical Problem.”

In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition. CoNNL-9*. ACL, 53–68.

Mitchell, Tom, Betteridge, Justin, Carlson, Andrew, Hruschka, Estevam, and Wang, Richard, 2009. “Populating the Semantic Web by Macro-Reading Internet Text.” In *Proceedings of the 8th International Semantic Web Conference (ISWC 2009)*. Karlsuhe: Semantic Web Science Association.

Neeleman, Ad. and Szendrői, Kriszta, 2004. “Superman Sentences.” *Linguistic Inquiry* 35:149–159.

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

Pierrehumbert, Janet, 1980. *The Phonology and Phonetics of English Intonation*.

Ph.D. thesis, MIT. Distributed by Indiana University Linguistics Club, Bloomington.

Prince, Ellen, 1981. “Towards a Taxonomy of Given-New Information.” In Peter Cole (ed.), *Radical Pragmatics*, New York: Academic Press. 223–256.

Pulman, Stephen, 1997. “Higher Order Unification and the Interpretation of Focus.” *Linguistics and Philosophy* 20:73–115.

Quine, Willard Van Orman, 1960. *Word and Object*. Cambridge, MA: MIT Press.

Roark, Brian, 2001. “Probabilistic top-down parsing and language modeling.” *Computational Linguistics* 27:249–276.

Robaldo, Livio, 2008. “Skolem Theory and Generalized Quantifiers.” In *Workshop* *on Logic, Language, Information, and Computation (WoLLIC-08*. Berlin: Springer, Lecture Notes in Artificial Intelligence 5110, 286–297.

Roberts, Craige, 1996. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *OSU Working Papers in Linguistics* 49:91–136.

Rooth, Mats, 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats, 1992. "A Theory of Focus Interpretation." *Natural Language Semantics* 1:75–116.

Ross, John Robert, 1970. "Gapping and the Order of Constituents." In Manfred Bierwisch and Karl Heidolph (eds.), *Progress in Linguistics*, The Hague: Mouton. 249–259.

Sakas, William and Fodor, Janet Dean, 2001. “The Structural Triggers Learner.” In S. Bertolo (ed.), *Language Acquisition and Learnability*, Cambridge: Cambridge University Press. 172–233.

Sánchez Valencia, Víctor, 1991. *Studies on Natural Logic and Categorial Grammar*. Ph.D. thesis, Universiteit van Amsterdam.

Sato, Maso-Aki, 2001. “Online Model Selection based on the Variational Bayes.” *Neural Computation* 13(7):1649–1681.

Siskind, Jeffrey, 1992. *Naive Physics, Event perception, Lexical Semantics, and Language Acquisition*. Ph.D. thesis, MIT.

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

Steedman, Mark, 1990. “Gapping as Constituent Coordination.” *Linguistics and Philosophy* 13:207–263.

Steedman, Mark, 1991. “Structure and Intonation.” *Language* 67:262–296.

Steedman, Mark, 1996a. “The Role of Prosody and Semantics in the Acquisition of Syntax.” In James Morgan and Katherine Demuth (eds.), *Signal to Syntax*, Hillsdale, NJ: Erlbaum. 331–342.

Steedman, Mark, 1996b. *Surface Structure and Interpretation*. Linguistic Inquiry Monograph 30. Cambridge, MA: MIT Press.

Steedman, Mark, 2000a. “Information Structure and the Syntax-Phonology Interface.” *Linguistic Inquiry* 34:649–689.

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

Steedman, Mark, 2007. "Information-Structural Semantics for English Intonation." In Chungmin Lee, Matthew Gordon, and Daniel Büring (eds.), *Topic and Focus:Cross-linguistic Perspectives on Meaning and Intonation*, Dordrecht: Kluwer, number 82 in Studies in Linguistics and Philosophy. 245–264. Proceedings of the LSA Workshop on Topic and Focus, Santa Barbara CA July, 2001.

Steedman, Mark, 2012a. "The Surface Compositional Semantics of English Intonation." *Submitted* .

Steedman, Mark, 2012b. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Thomforde, Emily and Steedman, Mark, 2011. "Semi-Supervised CCG Lexicon Extension." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 1246–1256.

Thompson, Cynthia and Mooney, Raymond, 2003. "Acquiring Word-Meaning Mappings for Natural Language Interfaces." *Journal of Artificial Intelligence Research* 18:1–44.

Vallduví, Enric and Vilkuna, Maria, 1998. "On Rheme and Kontrast." In Peter Culicover and Louise McNally (eds.), *Syntax and Semantics, Vol. 29: The Limits of Syntax*, San Diego: Academic Press. 79–108.

Vijay-Shanker, K. and Weir, David, 1990. "Polynomial Time Parsing of Combinatory Categorial Grammars." In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*. Pittsburgh: ACL, 1–8.

Vijay-Shanker, K. and Weir, David, 1994. "The Equivalence of Four Extensions of Context-Free Grammar." *Mathematical Systems Theory* 27:511–546.

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

Villavicencio, Aline, 2011. “Language Acquisition with Feature-Based Grammars.” In Robert Boyer and Kirsti Börjars (eds.), *Non-Transformational Syntax: A Guide to Current Models*, Blackwell. 404–442.

Wold, Dag, 1996. “Long Distance Selective Binding: The Case of Focus.” In *Proceedings of SALT*. volume 6, 311–328.

Wong, Yuk Wah and Mooney, Raymond, 2007. “Learning Synchronous Grammars for Semantic Parsing with Lambda Calculus.” In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. ACL, 960–967.

Xia, Fei, 1999. “Extracting Tree-Adjoining Grammars from Bracketed Corpora.”

In *Proceedings of the 5th Natural Language Processing Pacific Rim Symposium (NLPRS-99)*.

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

Zettlemoyer, Luke and Collins, Michael, 2005. "Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars." In *Proceedings of the 21st Conference on Uncertainty in AI (UAI)*. Edinburgh: AAAI, 658–666.

Zettlemoyer, Luke and Collins, Michael, 2007. "Online Learning of Relaxed CCG Grammars for Parsing to Logical Form." In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL)*. Prague: ACL, 678–687.