Matt's grand plan for natural language understanding (with CCG)

Matthew Honnibal

September 6, 2010

NLU is currently divided into several tasks

Example

The company was persuaded to buy Power Set.

- **Syntax:** ((The company) (was (persuaded (to buy (Power Set)))))
- **Semantics:** persuade(x, company, buy(company, Power Set))
- **Sense Disambiguation:** *company* as in *army unit*?
- **Named Entity Recognition and Disambiguation:** Power Set → http://en.wikipedia.org/PowerSet
- **Coreference? Sentiment? Discourse? More?**

Pros and cons of dividing up the task

Pros

- **Reductionism:** It may be easier to make progress on the tasks in isolation.
- **Modularity:** Don't like one parser? Just plug in another!

Cons

- **Accuracy:** Information can only flow in one direction.
- **Efficiency:** The same work is repeated many times.
- **Plausibility:** Is a pipeline a realistic model of natural language understanding? Should we be trying to find one?

Intuition behind joint modelling

- $H(W_s)$: information to disambiguate the words in s
- $H(R_s)$: information to assign semantic role labels to s
- If word senses are good features for SRL, then $H(R_s|W_s) < H(R_s)$
- But if $H(R_s|W_s) < H(R_s)$, then $H(W_s|R_s) < H(W_s)$
- **If WSD helps SRL, then SRL must be able to help WSD.**
- **So: model $P(R_s, W_s)$ instead of $P(W_s)$ and $P(R_s|W_s)$**
- The grand plan: jointly model all the sentence understanding tasks by bringing all the information into a CCG parse.

Categorial Grammar: few rules, complex categories

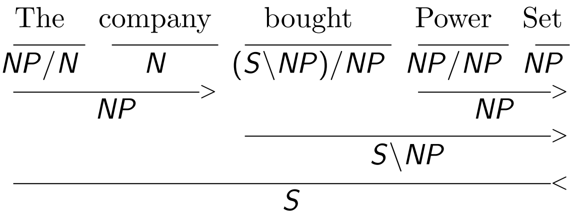
Table: Categorial Grammar has only 2 rule schemas, and 3 atomic types.

Rules				Types
X	$\rightarrow$	X/Y	Y	N
X	$\rightarrow$	Y	$Y\backslash X$	PP
				S

Table: Production rules get ‘translated’ into complex categories.

PSG				CG
NP	$\rightarrow$	DT	N'	NP/N
PP	$\rightarrow$	IN	NP	PP/NP
S	$\rightarrow$	NP	VP	$S\backslash NP$
VP	$\rightarrow$	V	NP	$(S\backslash NP)/NP$
VP	$\rightarrow$	VP	$ADVP$	$(S\backslash NP)\backslash(S\backslash NP)$

Example Categorial Grammar Derivation



The company bought Power Set

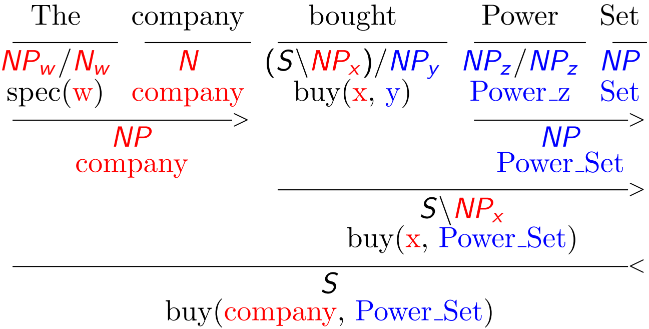
NP/N N $(S\backslash NP)/NP$ NP/NP NP

NP > NP >

$S\backslash NP$ >

S <

The pay-off: semantic transparency



The company bought Power Set

NP_w/N_w spec(w)

N company

$(S\backslash NP_x)/NP_y$ buy(x, y)

NP_z/NP_z Power_z

NP Set

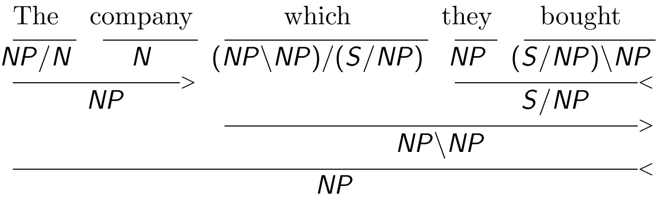
$>$ NP company

$>$ NP Power_Set

$>$ $S\backslash NP_x$ buy(x, Power_Set)

$<$ S buy(company, Power_Set)

CCG adds more rules to reduce category ambiguity



The company which they bought

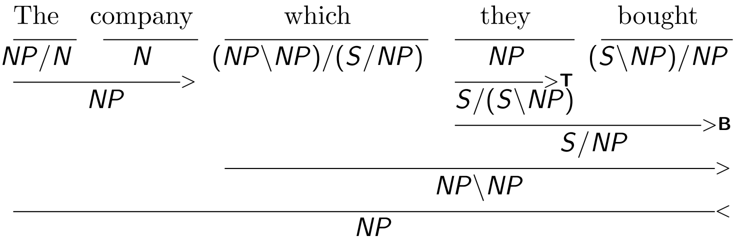
NP/N N $(NP\backslash NP)/(S/NP)$ NP $(S/NP)\backslash NP$

$\xrightarrow{>}$ NP

$\xrightarrow{<}$ S/NP

$\xrightarrow{>}$ $NP\backslash NP$

$\xrightarrow{<}$ NP



The company which they bought

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

$\xrightarrow{>}$ NP

$\xrightarrow{>\mathbf{T}}$ $S/(S\backslash NP)$

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

$\xrightarrow{>}$ $NP\backslash NP$

$\xrightarrow{<}$ NP

PropBank and NomBank: Penn Treebank SRL layers

- A **predicate** heads a proposition (but might not assert it)
- **Arguments** can be **core** or **peripheral**.

Example predicate-argument structures

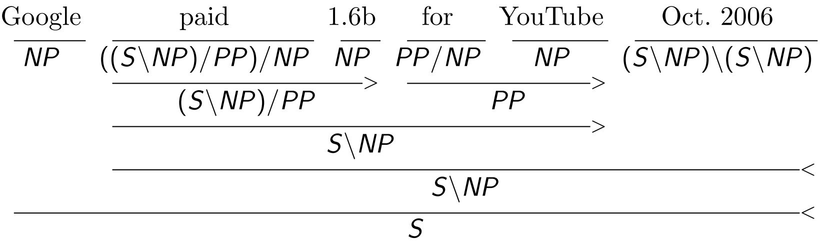
(1)	*Google*	*bought*	*YouTube*	*October 2006*	*for 1.6bn*
	Arg-0	Predicate	Arg-1	Arg-TMP	Arg-3
(2)	*Google*	*paid*	*1.6bn*	*for YouTube*	*October 2006*
	Arg-0	Predicate	Arg-1	Arg-3	Arg-TMP
(3)	*Google's*	*1.6bn*	*acquisition*	*of YouTube*	*October 2006*
	Arg-0	Arg-1	Predicate	Arg-3	Arg-TMP

- **PropBank:** Propositions headed by **verbs** in the PTB.
- **NomBank:** Propositions headed by **nouns** in the PTB.

Integrating PropBank annotation into CCG

- Target: CCG derivations that map unambiguously to PropBank analyses.
- **Predicates** will be identified by the **semantic category** assigned to them.
- **Core arguments** will be **syntactic complements**. Argument labels will be assigned by the syntax-semantics mapping.
- **Peripheral arguments** will be **syntactic adjuncts**. Their type will be specified in their semantics.

Distinguishing core and peripheral arguments in CCG



Google paid 1.6b for YouTube Oct. 2006

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

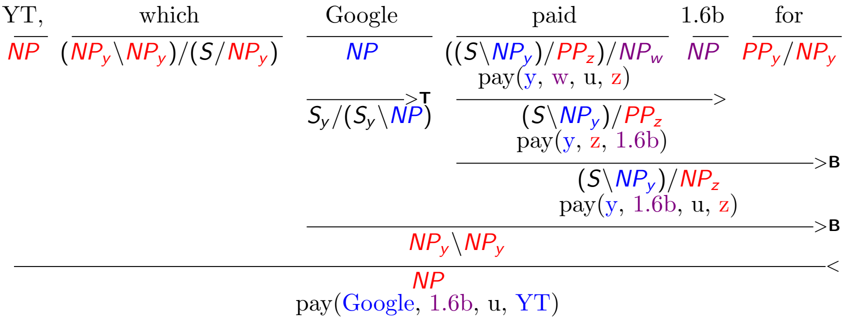
$>$ $(S\backslash NP)/PP$ $>$ PP

$>$ $S\backslash NP$

$<$ $S\backslash NP$

$<$ S

Compositional semantics for SRL with CCG



YT, which Google paid 1.6b for

NP $(NP_y\backslash NP_y)/(S/NP_y)$ NP $((S\backslash NP_y)/PP_z)/NP_w$ NP PP_y/NP_y

pay(y, w, u, z)

$>\mathbf{T}$ $S_y/(S_y\backslash NP)$

$>$ $(S\backslash NP_y)/PP_z$

pay(y, z, 1.6b)

$>\mathbf{B}$ $(S\backslash NP_y)/NP_z$

pay(y, 1.6b, u, z)

$>\mathbf{B}$ $NP_y\backslash NP_y$

$<$ NP

pay(Google, 1.6b, u, YT)

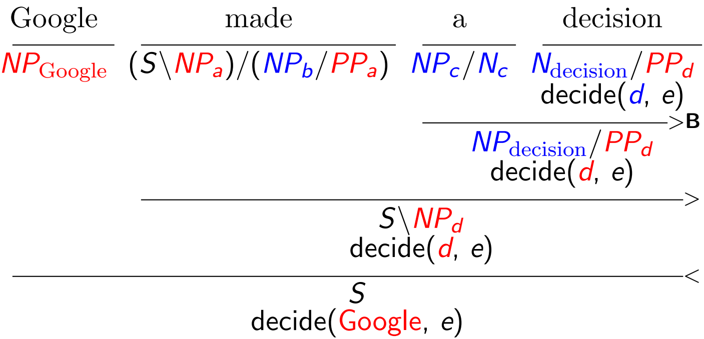
Compositional semantics for nominal predicates in CCG

$$\begin{array}{cccc}
\text{acquisition} & \text{of} & \text{YouTube} & \\ \hline
N_{acquisition}/PP_b & PP_c/NP_c & NP_{\text{YouTube}} & \\
\text{acquire}(a,\ b) & & & \\ \cline{2-3}
 & \multicolumn{2}{c}{PP_{\text{YouTube}}} & > \\ \cline{1-3}
\multicolumn{3}{c}{N_{\text{acquisition}}} & > \\
\multicolumn{3}{c}{\text{acquire}(a,\ \text{YouTube})} &
\end{array}$$

Nominal predicates require some creative analyses

$$\begin{array}{ccc}
\text{YouTube} & \text{'s} & \text{acquisition} \\
\hline
NP_{\text{YouTube}} & (NP_a/(N_a/PP_b))\backslash NP_b & N_{\text{acquisition}}/PP_d \\
 & & \mathsf{acquire}(c,\ d) \\
\cline{1-2}
\multicolumn{2}{c}{NP_a/(N_a/PP_{\text{YouTube}})} & < \\
\hline
\multicolumn{3}{c}{NP_{\text{acquisition}}} \\
\multicolumn{3}{c}{\mathsf{acquire}(c,\ \mathsf{YouTube})}
\end{array} >$$

Nominal predicates with support verbs



Google made a decision

NP_{Google} $(S\backslash NP_a)/(NP_b/PP_a)$ NP_c/N_c N_{decision}/PP_d

decide(d, e)

$>\mathbf{B}$

$NP_{\text{decision}}/PP_d$

decide(d, e)

$>$

$S\backslash NP_d$

decide(d, e)

$<$

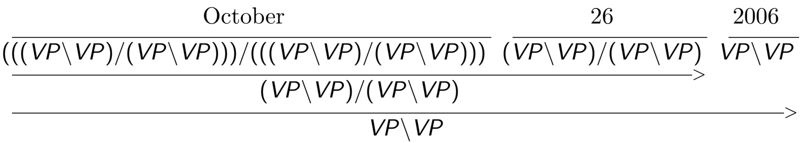
S

decide(Google, e)

Joint Named Entity Recognition and PTB parsing

- Named entities recognition is usually modelled as a **sequence tagging** task, e.g.
  - **Power|ORG Set|ORG**
- This makes it difficult to account for **nested named entities**, e.g.
  - **New York Stock Exchange**
  - **Sydney, Australia**
  - **David and Melissa Smith**
- Finkel and Manning (2009) joint NER and parsing:
  - Up to 1.36% F-measure parsing improvement;
  - Up to 9% F-measure NER improvement.

Named entities screw up CCG parses if handled naively

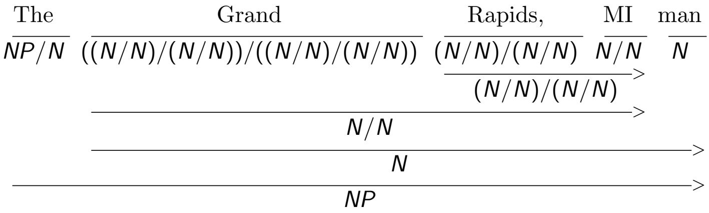


October 26 2006

$$(((VP\backslash VP)/(VP\backslash VP)))/(((VP\backslash VP)/(VP\backslash VP))) \quad (VP\backslash VP)/(VP\backslash VP) \quad VP\backslash VP$$

$$\xrightarrow{>} (VP\backslash VP)/(VP\backslash VP)$$

$$\xrightarrow{>} VP\backslash VP$$



The Grand Rapids, MI man

$$NP/N \quad ((N/N)/(N/N))/((N/N)/(N/N)) \quad (N/N)/(N/N) \quad N/N \quad N$$

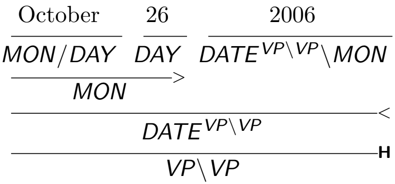
$$\xrightarrow{>} (N/N)/(N/N)$$

$$\xrightarrow{>} N/N$$

$$\xrightarrow{>} N$$

$$\xrightarrow{>} NP$$

Integrating NER into CCG with Hat Categories



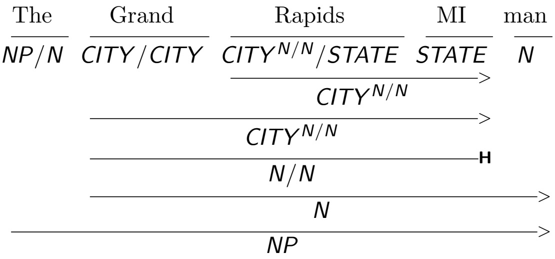
October 26 2006

MON/DAY DAY $DATE^{VP\backslash VP}\backslash MON$

$>$ MON

$<$ $DATE^{VP\backslash VP}$

H $VP\backslash VP$



The Grand Rapids MI man

NP/N $CITY/CITY$ $CITY^{N/N}/STATE$ $STATE$ N

$>$ $CITY^{N/N}$

$>$ $CITY^{N/N}$

H N/N

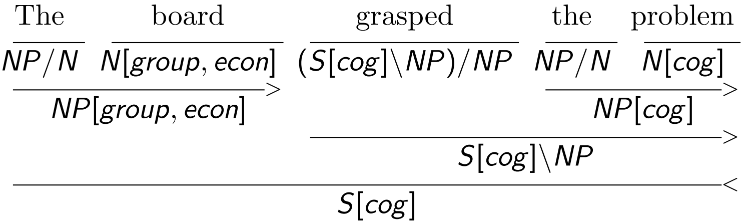
$>$ N

$>$ NP

Tentative thoughts on Word Sense Disambiguation

- Full word sense disambiguation involves many fine-grained labels
- Integrating these labels into CCG category sets may cause sparse data problems
- What if I just use super senses and WordNet Domains?
-
    - 41 supersenses e.g. noun.food, noun.group, verb.cognition.
    - 46 domains, e.g. economy, sport, fashion, sexuality

Adding SuperSenses and domains as category features



WordNet senses for ‘board’ and ‘problem’

Sense	Super sense	Definition
1	noun.group	A committee having supervisory powers *the board has seven members*
2	noun.substance	A stout length of sawn timber; made in a wide variety of sizes and used for many purposes
4	noun.food	Food or meals in general *room and board*
9	noun.artifact	A flat portable surface (usually rectangular) designed for board games. *he got out the board and set up the pieces*
1	noun.state	A state of difficulty that needs to be resolved: *she and her husband are having problems*
2	noun.communication	A question raised for consideration or solution: *our homework consisted of ten problems to solve*
3	noun.cognition	A source of difficulty *one trouble after another delayed the job*

Progress so far

- PropBank/CCGbank integration complete
- Most difficult NomBank/CCGbank integration complete
- Preliminary parsing experiments on modified corpora
- Nicky has BBN and CCG aligned and is working on the integration
- Mike White's group have done something with CCG and discourse parsing

Current priorities

- Get oracle figures for CCGbank-to-SRL
- Error analysis over oracle errors. Further improvements? Problems with CCG?
- Parse with SRL-CCGbank to get joint model performance.
- Tinker with WSD/CCG ideas at some point.

Conclusion

- I am focussing on a representation problem, rather than the learning problem. But do these tasks all fit in one hypothesis space? Will the task be tractable?
- It's currently very difficult to deploy a system that makes use of all the NLU modules.
- If my approach works, it will produce a very efficient all-singing-all-dancing NLU solution.
- The project also raises a lot of questions about our current theories of compositional semantics.