

ACS Advanced Syntax and Semantics of Natural Language

Lecture 5: CCG Applied to Real Text



UNIVERSITY OF
CAMBRIDGE

Stephen Clark

Natural Language and Information Processing (NLIP) Group

sc609@cam.ac.uk

Pierre Vinken, 61 years old, will join the board as a non-executive director Nov. 29.

Activation of the CD28 surface receptor provides a major costimulatory signal for T cell activation resulting in enhanced production of interleukin-2 (IL-2) and cell proliferation.

The Trust's symbol, a sprig of oak leaves and acorns, is thought to have been inspired by a carving in the cornice of the Alfriston Clergy House.

- Can we really move from simple “linguistic” examples to sentences like these found in the real world?

Pierre|N/N Vinken|N ,|, 61|N/N years|N old|(S[adj]\NP)\NP
,|, will|(S[dcl]\NP)/(S[b]\NP) join|((S[b]\NP)/PP)/NP
the|NP/N board|N as|PP/NP a|NP/N nonexecutive|N/N
director|N Nov.|((S\NP)\(S\NP))/N 29|N .|. .

- Needs an $N \rightarrow NP$ rule
- $S[adj]\backslash NP$ is for predicative adjectives, e.g. *the man is old*
- We need a *unary type-changing rule*: $S[adj]\backslash NP \rightarrow NP\backslash NP$
- We need special rules in the parser to deal with punctuation
- Only need application in this example (no composition or type-raising)

-
- S category often has a grammatical feature which indicates the kind of sentence or verb phrase
 - $S[dcl]$ declarative sentence
 - $S[q]$ yes/no questions
 - $S[b]$ bare infinitives
 - $S[to]$ to infinitives
 - $S[pss]$ past participles in passive mode
 - $S[pt]$ past participles in active mode
 - $S[ng]$ present participles
 - ...
 - See p.47 of Julia's thesis for full list
 - S in adverbial modifiers, e.g. $(S \setminus NP) / (S \setminus NP)$, effectively has a variable feature: $(S[X] \setminus NP) / (S[X] \setminus NP)$, which unifies with the feature on the argument and transfers to the result

Activation|N of|(NP\NP)/NP the|NP/N CD28|N/N surface|N/N
receptor|N provides|(S[dc1]\NP)/NP a|NP/N major|N/N
costimulatory|N/N signal|N for|(NP\NP)/NP T|(N/N)/(N/N)
cell|N/N activation|N resulting|(S[ng]\NP)/PP in|PP/NP
enhanced|N/N production|N of|(NP\NP)/NP interleukin-2|N
(|(IL-2|N) |) and|conj cell|N/N proliferation|N .|. .

- Needs a unary type-changing rule: $S[ng]\backslash NP \rightarrow (S\backslash NP)\backslash(S\backslash NP)$
- Need special rules to deal with brackets
- Still only needs application

The|NP/N Trust|N 's|(NP/N)\NP symbol|N ,|, a|NP/N sprig|N
of|(NP\NP)/NP oak|N/N leaves|N and|conj acorns|N ,|,
is|(S[decl]\NP)/(S[pss]\NP) thought|(S[pss]\NP)/(S[to]\NP)
to|(S[to]\NP)/(S[b]\NP) have|(S[b]\NP)/(S[pt]\NP)
been|(S[pt]\NP)/(S[pss]\NP) inspired|S[pss]\NP
by|((S\NP)\(S\NP))/NP a|NP/N carving|N in|(NP\NP)/NP
the|NP/N cornice|N of|(NP\NP)/NP the|NP/N
Alfriston|(N/N)/(N/N) Clergy|N/N House|N .|.

- Still only need application
- No unary type-changing rules in this example

- Without type-changing rules (notice that the category for *used* is non-standard and the category for *once* changes also):

<i>A form of asbestos</i>	<i>once</i>	<i>used</i>	<i>to make Kent cigarettes</i>
NP	$(NP \backslash NP) / (NP \backslash NP)$	$(NP \backslash NP) / (S[to] \backslash NP)$	$S[to] \backslash NP$

- With type-changing rules (uses standard categories for *used* and *once*):

<i>A form of asbestos</i>	<i>once</i>	<i>used</i>	<i>to make Kent cigarettes</i>
NP	$(S \backslash NP) / (S \backslash NP)$	$(S[pss] \backslash NP) / (S[to] \backslash NP)$	$S[to] \backslash NP$
	$S[pss] \backslash NP$		
	$NP \backslash NP$		

- Type-changing rules increase the compactness of the lexicon (capturing generalisations) and reduce the number of categories assigned to modifiers such as *once*

- Object extraction from a relative clause, using type-raising and forward composition:

$$\begin{array}{ccccccc} \textit{That} & \textit{finished} & \textit{the job} & \textit{that} & \textit{Captain Chandler} & \textit{had} & \textit{begun} \\ \hline NP & (S[dcl]\backslash NP)/NP & NP & (NP\backslash NP)/(S[dcl]/NP) & NP & (S[dcl]\backslash NP)/(S[pt]\backslash NP) & (S[pt]\backslash NP)/NP \end{array}$$

- Question with an object extraction:

$$\begin{array}{ccccccc} \textit{What} & \textit{books} & \textit{did} & \textit{he} & \textit{author} & \textit{?} & \\ \hline (S[wq]/(S[q]/NP))/N & N & (S[q]/(S[b]\backslash NP))/NP & NP & (S[b]\backslash NP)/NP & - & \end{array}$$

Real Coordination Example

children *were* *more* *apt* *to* *avoid*
 $\frac{NP}{(S[dcl]\backslash NP)/(S[adj]\backslash NP)} \quad \frac{(S\backslash NP)/(S\backslash NP)}{(S[adj]\backslash NP)/(S[to]\backslash NP)} \quad \frac{(S[to]\backslash NP)/(S[b]\backslash NP)}{(S[b]\backslash NP)/NP}$

or *be* *shocked* *by* *nudity*
 $\frac{conj}{(S[b]\backslash NP)/(S[pss]\backslash NP)} \quad \frac{S[pss]\backslash NP}{((S\backslash NP)\backslash (S\backslash NP))/NP} \quad \frac{NP}{NP}$

-
- CCGbank: A Corpus of CCG Derivations and Dependency Structures Extracted from the Penn Treebank. Julia Hockenamier and Mark Steedman. Computational Linguistics. 2007
 - Data and models for statistical parsing with Combinatory Categorical Grammar, Julia Hockenmaier, PhD thesis, Edinburgh, 2003