



Practical CCG

CCGBank
Tgrep2
CCG Parse

CCGBank

Julia Hockenmaier and Mark Steedman

CCGBank: translation of Penn Tree Bank to CCG

Pairs syntactic derivations with word-word dependencies

CCGBank covers 99.44% of Penn Tree Bank
(and fixes some errors/inconsistencies)

Supports Tgrep access

Example Lexical Entries

Entry

Word (mail)

Category (N)

Word probability ($P(\text{mail} | \text{N})$)

Category probability ($P(\text{N} | \text{mail})$)

Word-category frequency in DB

mail	N	0.00013495	0.508772	29
mail	N/N	0.000168503	0.438596	25
mail	S[b]\NP	0.000367647	0.0175439	1
mail	((S[b]\NP)/NP)/NP	0.00359712	0.0175439	1
mail	((S[dcI]\NP)/PP)/NP	0.000619963	0.0175439	1

CCGBank access

Within CMU network

`/afs/cs.cmu.edu/academic/class/11722-s08/DATA/CCG`

Manual

142 page

Corpus

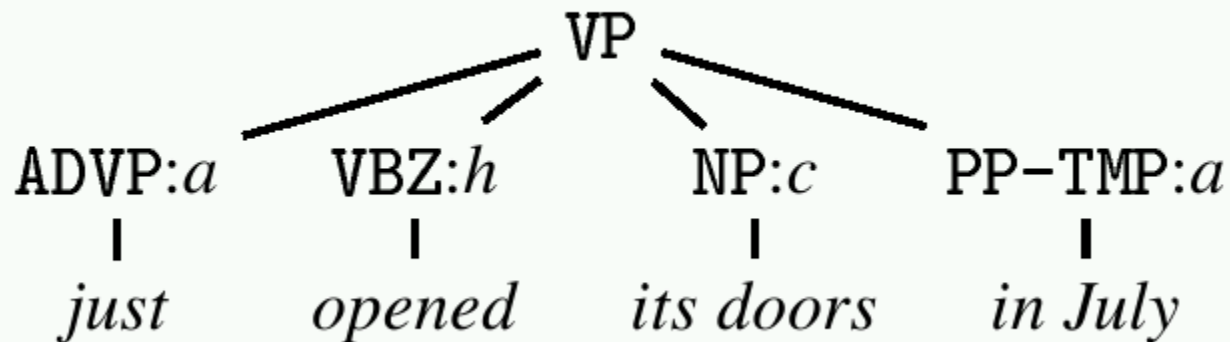
2713 paragraphs

48,934 sentences

83,540 lexical entries

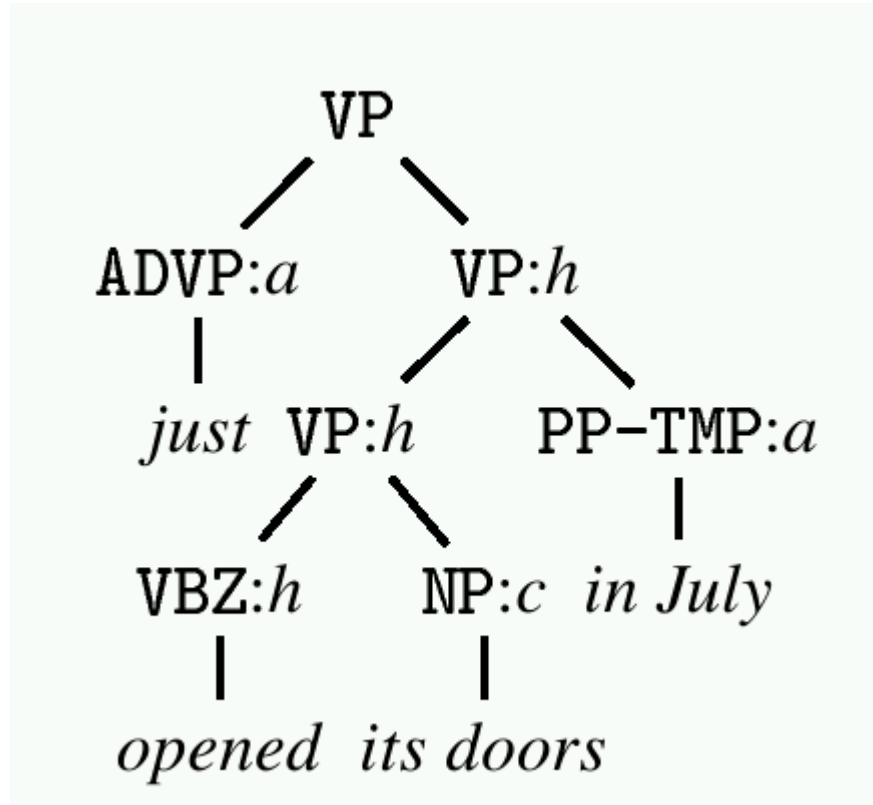
Tree Conversion

Basic UPenn tree



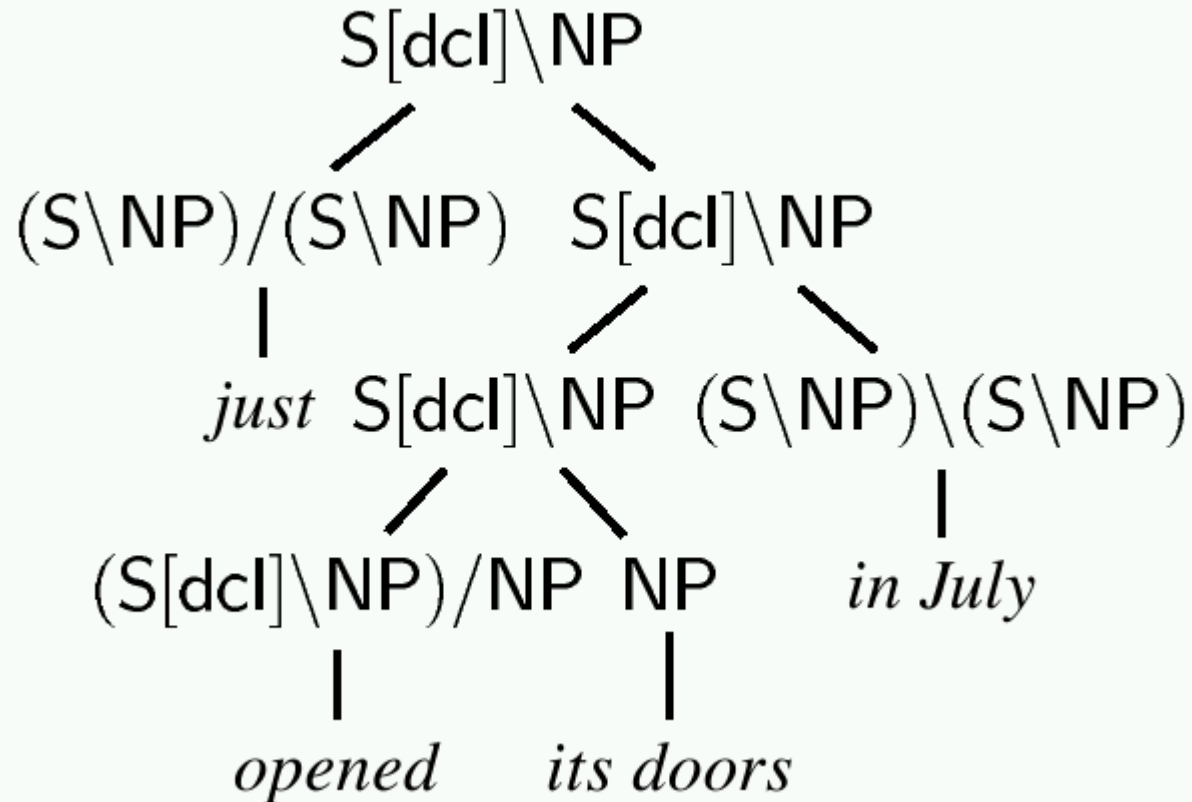
Binary Branching Tree

Converted to Binary Branching Tree



CCG Tree

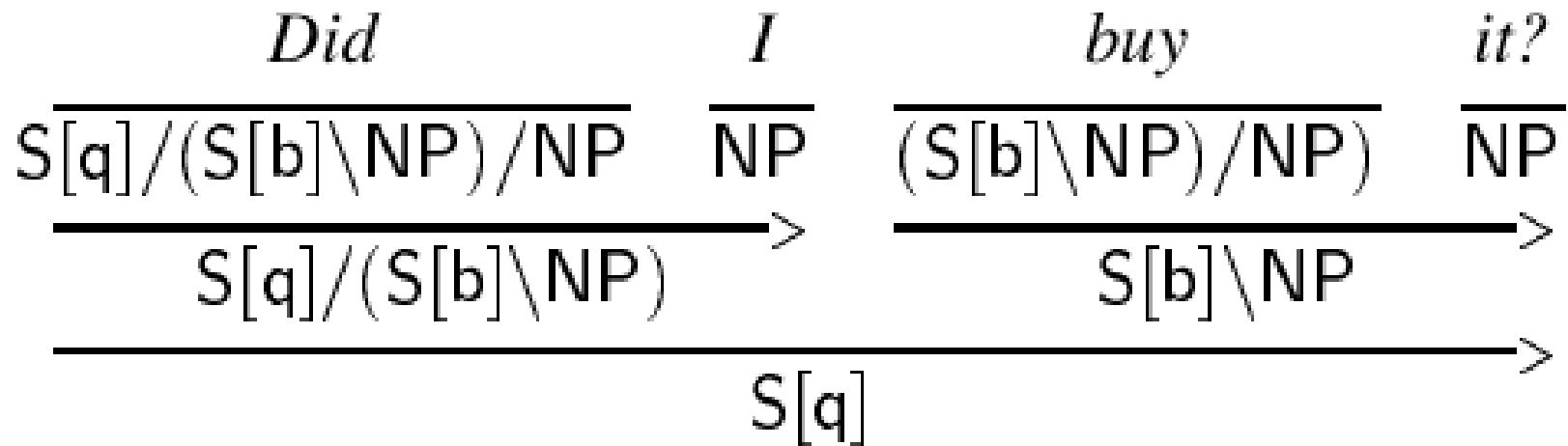
Re-labelled CCG tree



Additional Categories

- Sentences identified by type:
 - S[decl], S[pss], S[q], S[b]
 - S[pss]\NP (passive VP)

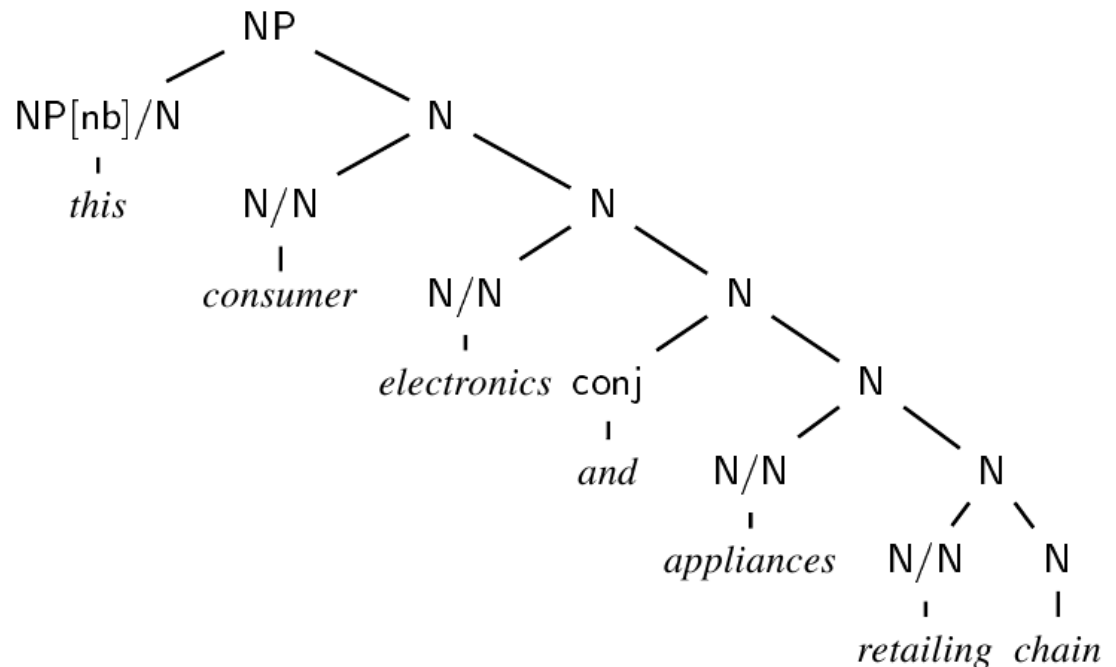
Example parse



Compound Nouns

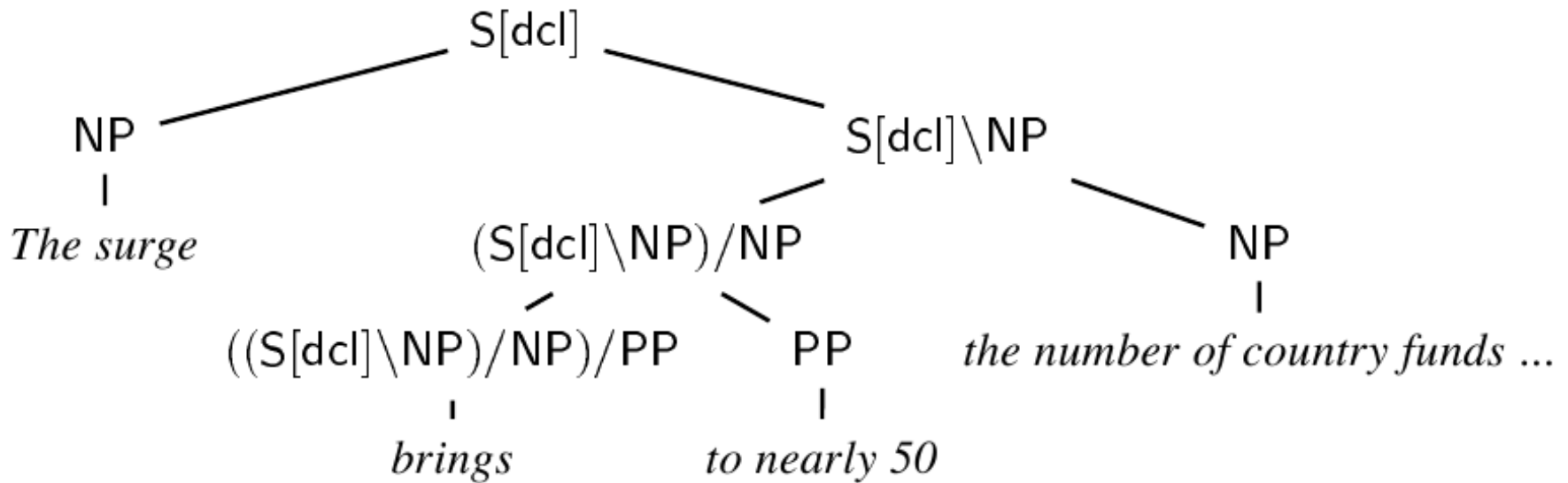
- Penn TreeBank, compounds are flat
Conversions must build binary branching

(NP (DT this)
(NN consumer)
(NNS electronics)
(CC and)
(NNS appliances)
(NN retailing)
(NN chain))



Actual Heavy NP Treatment

- Shifts aren't identified in TreeBank



Tgrep2 searching

- ◆ *./tgrep2 -c data/ccgbank.00-24.t2c **EXPR***
 - *Where **EXPR** may be*
 - *'S\[pss\]</NP//S\[pss\]/'*
 - *"/.*</^[Tt]hat/'*
- ◆ *Regex notes:*
 - *Backslash the backslash*
 - *Backslash [] literals*
 - *Practice a lot ...*

CCG Parser



Some specific examples

Page 29 of manual

Some problems

Due to original syntactic derivations
compound noun structure left as flat

Co-indexing issues

“I was early yesterday and late today”

early/late modify was, should be separate
events

Fails to capture control dependencies in
argument cluster coordination

“I want him to go and you to stay.”

Translation Algorithm

Translate each parse tree to word-word dependencies

For each tree:

determineConstituentType();

makeBinary();

assignCategories();

CCGBank Uses

Statistical Categorical Grammar Parsing

Julia's own work

Identifying semantic roles

Various wide-coverage CCG parsers