# Subcategorial considerations in statistical categorial parsing

—

Aditya Bhargava

# Syntactic structure in NLP

# Meaning is what we want

When was the dog-wolf genetic split?

between 27,000 and 40,000 years ago

---

“Text Bob, ‘How to recognize speech?’”

“Texting Bob, ‘How to wreck a nice beach’”

---

I would like eight pizzas, please

मुझे दो पिज्जा चाहिए

## Meaning is what we want

When was the dog-wolf genetic split? — between 27,000 and 40,000 years ago

---

“Text Bob, ‘How to recognize speech?’” — “Texting Bob, ‘How to wreck a nice beach’”

---

I would like eight pizzas, please — मुझे दो पिज्जा चाहिए

# How can syntax help?

# Syntax: a vehicle for semantics

Book a train from Toronto to Montréal

vs.

Book a train from Montréal to Toronto

# Categorial grammars (CGs)

- CG categories can “be thought of as a shorthand for the semantics” (Ades and Steedman, 1982)
- Rules have semantic correspondences as well
  - E.g., CCG composition: $\mathbf{B}{=}\lambda abc.a(bc)$
- CG derivations carry semantics

# CG lexical categories vs. parts of speech

N
N/N
$NP_{nb}$/N
(NP\NP)/NP
((S\NP)\(S\NP))/NP
Conj
NP
PP/NP
(S\NP)\(S\NP)
((S\NP)\(S\NP))/((S\NP)\(S\NP))
(((S\NP)\(S\NP))\((S\NP)\(S\NP)))/NP
...

CC
CD
DT
EX
FW
IN
JJ
JJR
JJS
LS
MD
NN
NNS
NNP
NNPS
PDT
POS
PRP

PRP$
RB
RBR
RBS
RP
SYM
TO
UH
VB
VBD
VBG
VBN
VBP
VBZ
WDT
WP
WP$
WRB

# CG lexical categories vs. parts of speech

N
N/N
$NP_{nb}/N$
(NP\NP)/NP
((S\NP)\(S\NP))/NP
Conj
NP
PP/NP
(S\NP)\(S\NP)
((S\NP)\(S\NP))/((S\NP)\(S\NP))
(((S\NP)\(S\NP))\((S\NP)\(S\NP)))/NP
…

CC
CD
DT
EX
FW
IN
JJ
JJR
JJS
LS
MD
NN
NNS
NNP
NNPS
PDT
POS
PRP
PRP$
RB
RBR
RBS
RP
SYM
TO
UH
VB
VBD
VBG
VBN
VBP
VBZ
WDT
WP
WP$
WRB

# Statistical LCG & CCG parsing

- LCG parsing: non-statistical only
  - No corpus until LCGbank (Fowler, 2016)
  - Relevance of category structure in proof nets (Roorda, 1991; Penn, 2004; Fowler, 2010)
- CCG parsing: lacking in consideration of category structure
  - Early work in the context of statistical methods for other formalisms (e.g., Hockenmaier, 2001; Clark, 2002)
  - Still-influential statistical CCG parser (Clark and Curran, 2007)

# Thesis

Subcategorial information in CG lexical categories is useful for statistical parsing

# Thesis

Subcategorial information in CG lexical categories is useful for statistical parsing

1. Supertagging with CCG primitives

# Thesis

Subcategorial information in CG lexical categories is useful for statistical parsing

1. Supertagging with CCG primitives
2. Proof net structure for neural LCG parsing

# Thesis

Subcategorial information in CG lexical categories is useful for statistical parsing

1. Supertagging with CCG primitives
2. Proof net structure for neural LCG parsing
3. Decomposed scoring of CCG dependencies

# Contributions

- Subcategorial supertagging improves:
  - Supertagging accuracy
  - Parser $F_1$
  - Parser coverage
- Subcategories allow for more effective use of prediction history
- Enables OOV category prediction

1. Supertagging with CCG primitives
2. Proof net structure for neural LCG parsing
3. Decomposed scoring of CCG dependencies

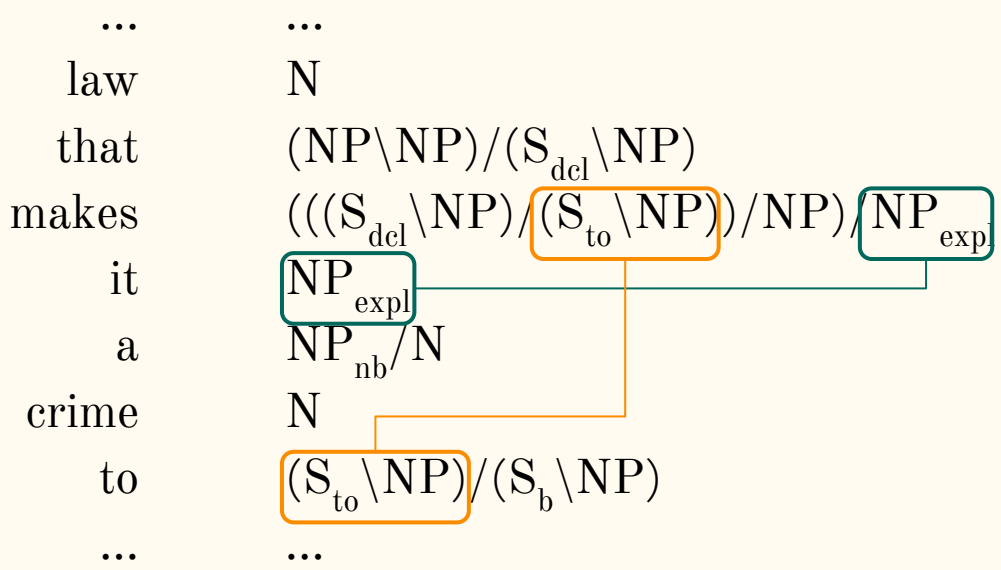
# Category internals are useful for supertagging

| | |
|---|---|
| ... | ... |
| law | N |
| that | $(NP\backslash NP)/(S_{dcl}\backslash NP)$ |
| makes | $(((S_{dcl}\backslash NP)/(S_{to}\backslash NP))/NP)/NP_{expl}$ |
| it | $NP_{expl}$ |
| a | $NP_{nb}/N$ |
| crime | N |
| to | $(S_{to}\backslash NP)/(S_{b}\backslash NP)$ |
| ... | ... |

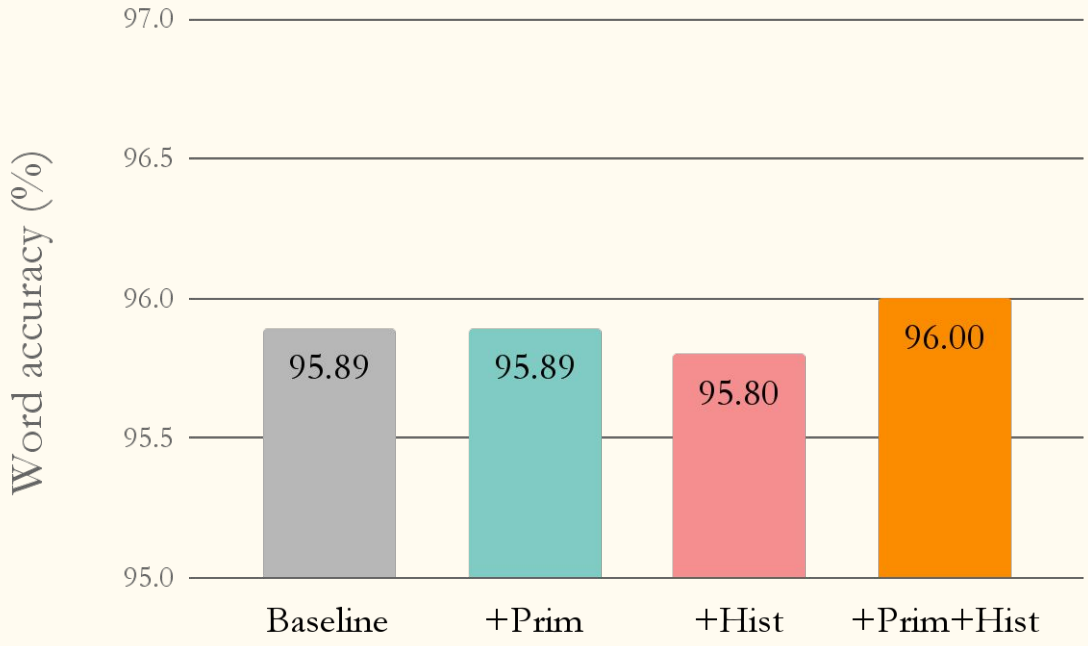
# Category internals are useful for supertagging

| | |
|---|---|
| ... | ... |
| law | N |
| that | $(NP\backslash NP)/(S_{dcl}\backslash NP)$ |
| makes | $(((S_{dcl}\backslash NP)/(S_{to}\backslash NP))/NP)/NP_{expl}$ |
| it | $NP_{expl}$ |
| a | $NP_{nb}/N$ |
| crime | N |
| to | $(S_{to}\backslash NP)/(S_{b}\backslash NP)$ |
| ... | ... |

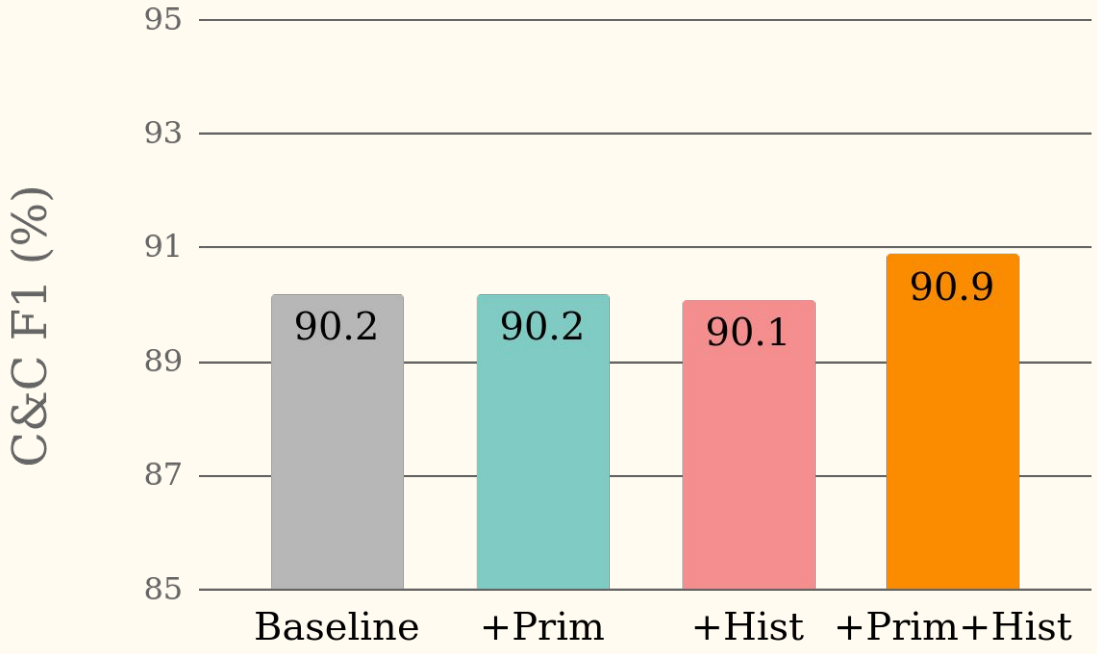
# Category internals are useful for supertagging



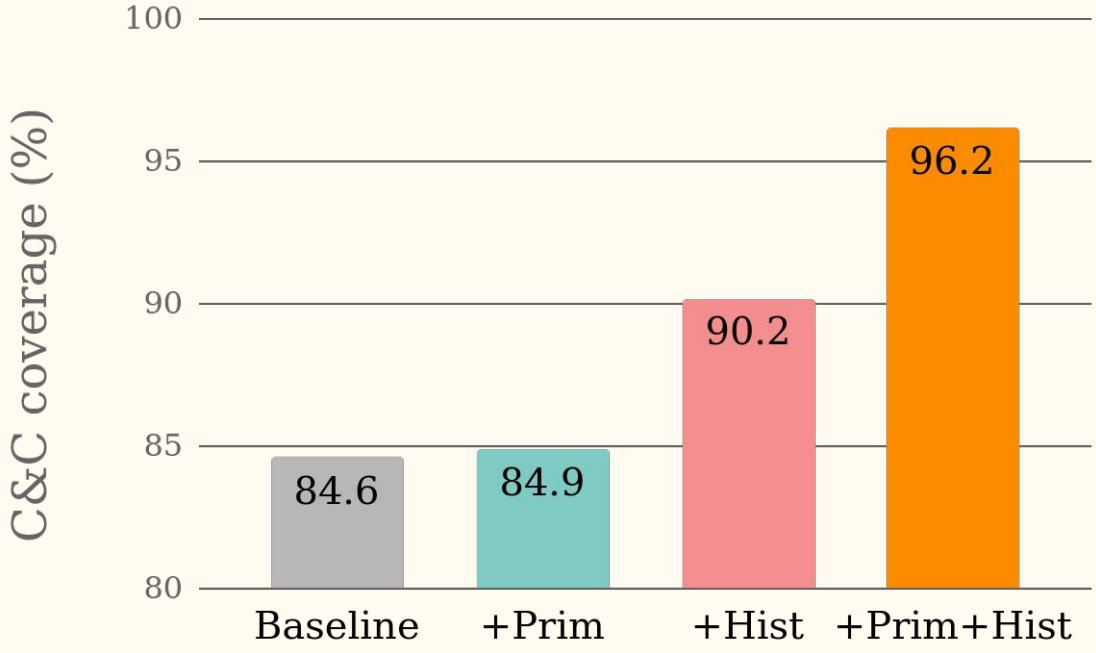
# Supertagging word accuracy



# C&C parser $F_1$


C&C F1 (%)
95
93
91
89
87
85
90.2
90.2
90.1
90.9
Baseline
+Prim
+Hist
+Prim+Hist

# C&C parser coverage

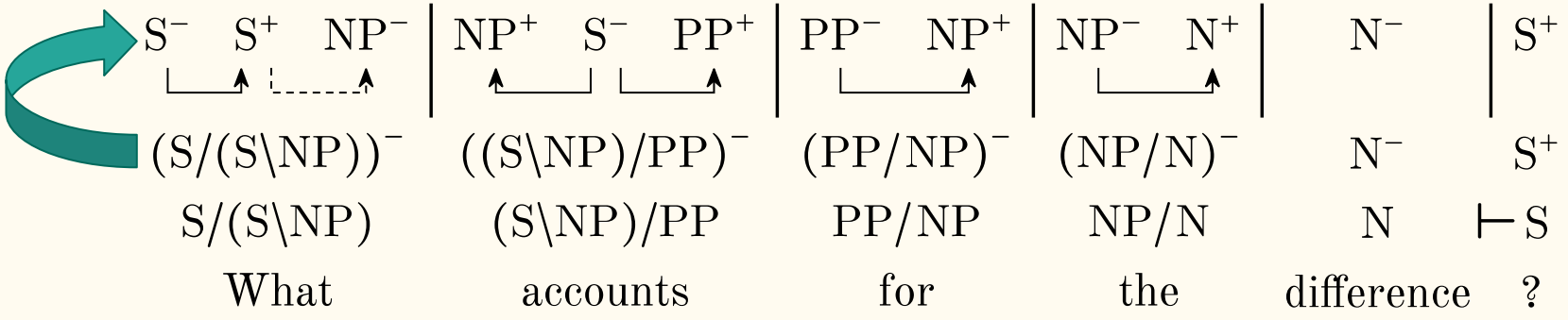


# Contributions

- First statistical LCG parser
- Proof net parsing with lexical decomposition
- Novel loss functions and structural constraints
  - Increase parser performance
  - Enable training without ground truth

1. Supertagging with CCG primitives
2. Proof net structure for neural LCG parsing
3. Decomposed scoring of CCG dependencies

# LCG proof nets

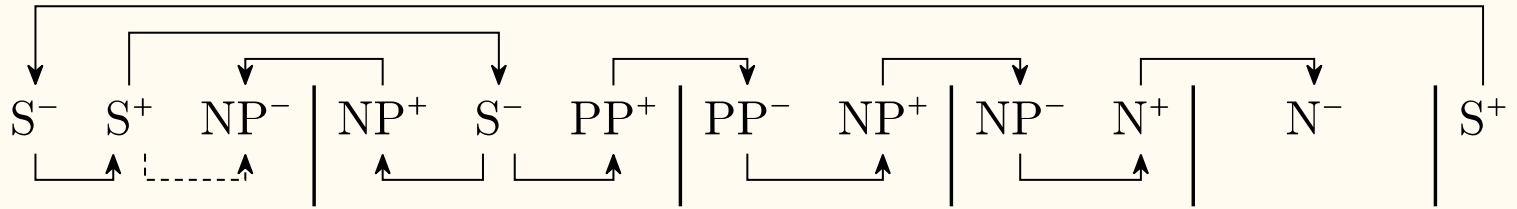


# LCG parsing with proof nets: validity conditions

T1. Linkage must be half-planar
- No crossing edges in half-plane above vertices

T2. No regular cycles (solid edges)

T3. Each Lambek edge must have regular path between its vertices (dashed edges)

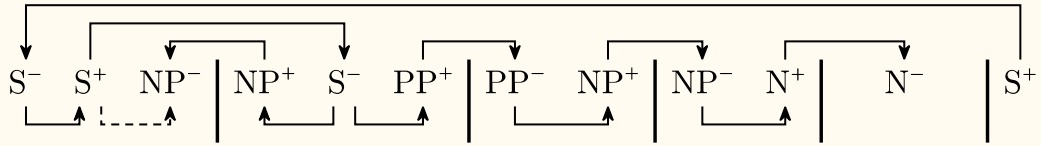


## LCG parsing with proof nets: validity conditions

T1. Linkage must be half-planar
- No crossing edges in half-plane above vertices

T2. No regular cycles (solid edges)

T3. Each Lambek edge must have regular path between its vertices (dashed edges)




Each condition can be expressed as a differentiable function of the *graph*—no ground truth linkages needed

# Parsing performance

| Condition | Link Acc | Sent Acc | Coverage |
|---|---|---|---|
| Base | 97.7 | 86.2 | 97.3 |
| Enhanced model | 97.9 | 87.4 | 98.4 |
| Enhanced model + losses | 97.9 | 87.2 | 98.7 |

# Training without ground truth

| Condition | Link Acc |
|---|---|
| Enhanced model + losses | 91.2 |
| —T1 loss | 84.5 |
| —T2 loss | 72.9 |
| —T3 loss | 70.6 |
| —Link filter | 73.9 |
| —planar attention — T1 loss | 19.2 |

# Training without ground truth

| Condition | Link Acc |
|---|---|
| Enhanced model + losses | 91.2 |
| —T1 loss | 84.5 |
| —T2 loss | 72.9 |
| —T3 loss | 70.6 |
| —Link filter | 73.9 |
| —planar attention — T1 loss | 19.2 |

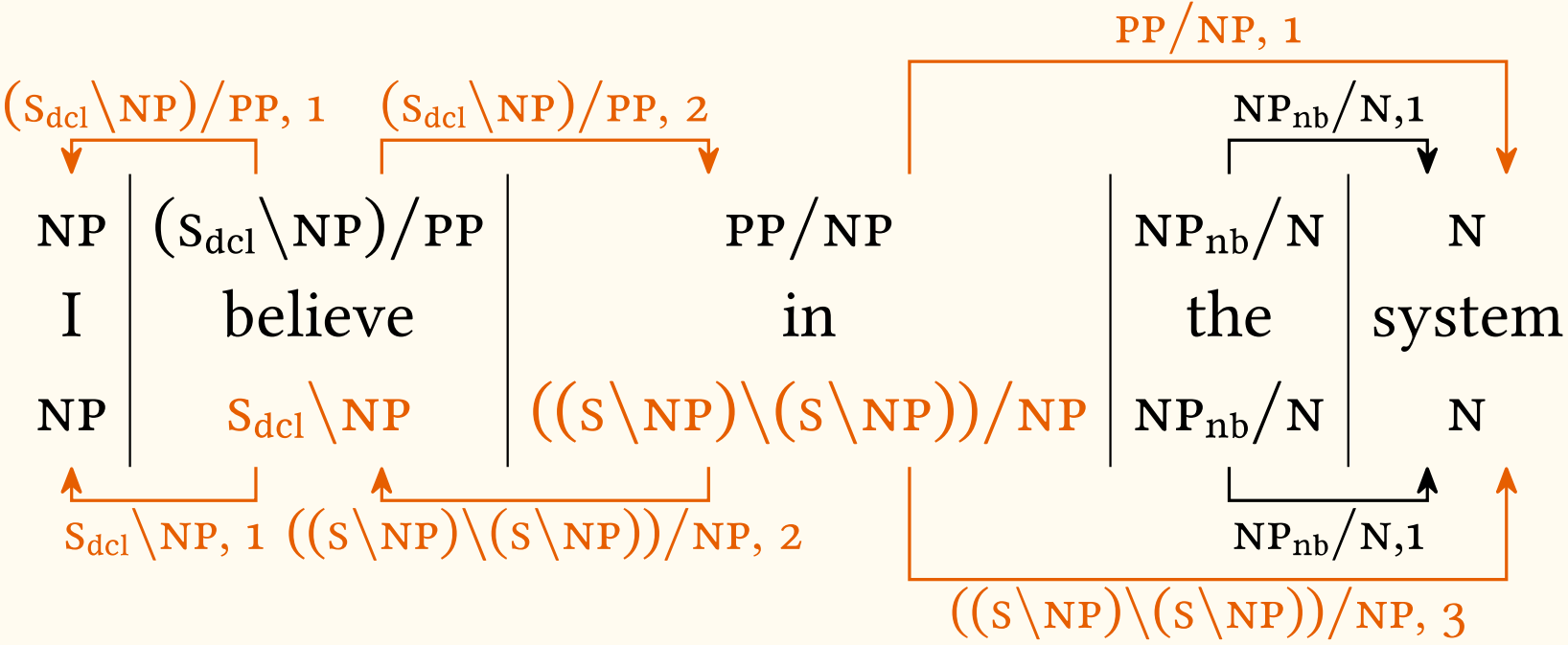
All planarity information removed

# Contributions

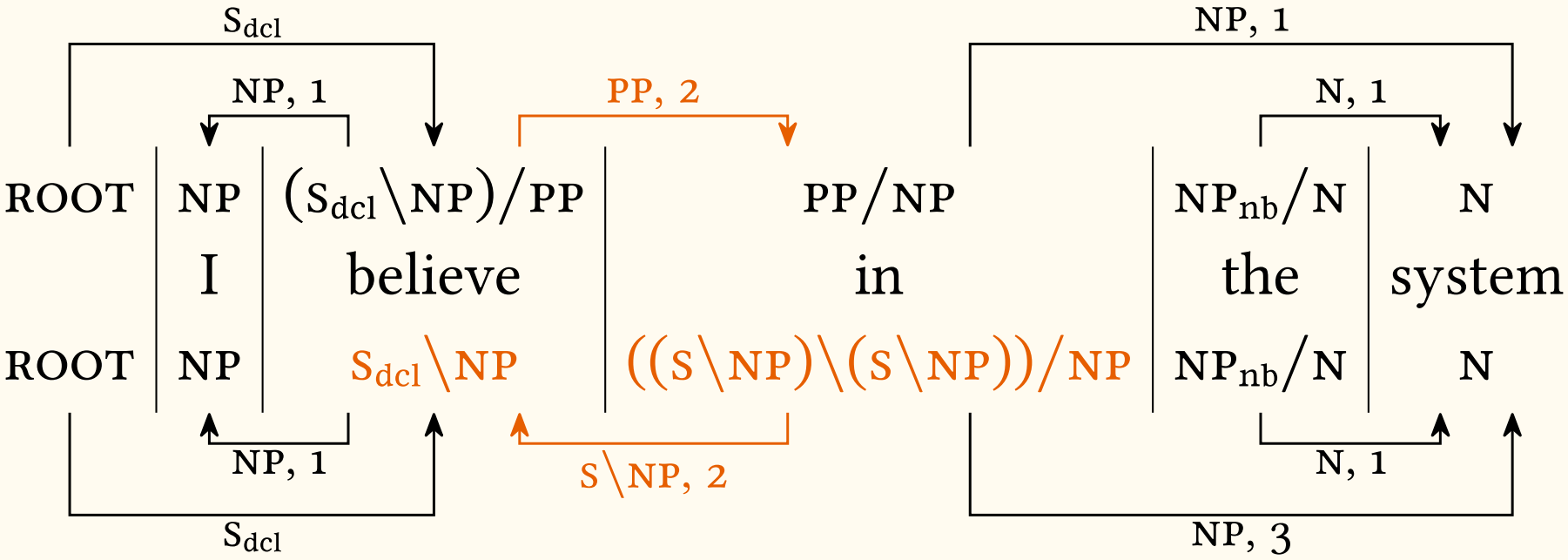
- Decomposed scoring fixes cases of error amplification
- Systematic human validation of statistical parsing scoring method
  - Expert judges find in favour of decomposed scoring (overall)
  - Raises questions about the validity of aggregated parser scores

1. Supertagging with CCG primitives
2. Proof net structure for neural LCG parsing
3. Decomposed scoring of CCG dependencies

# CCG dependency evaluation



# Decomposed scoring



# Expert judgements

- Intrinsic task: direct comparison of scoring methods
  - Strong agreement among judges in favour of decomposed scoring
- Extrinsic task: pairwise rank inversions
  - Overall agreement among judges in favour of decomposed scoring, but...
  - High disagreement: 50% ties in pilot study; 24% ties in main study

# Future directions

# End-to-end categorial parsing

- End-to-end systems pass training signals from output back to original input
  - Would be helpful for parsing errors to propagate back to supertagging decisions
- Recent “end-to-end” parsers aren’t truly end-to-end (Kasai et al., 2018; Kogkalidis et al., 2020)
  - Better described as “joint” or “multi-task”
- Tighter coupling between subcategorial supertagger and parser

# Mildly context-sensitive proof nets

- LCG is weakly CF-equivalent
- Phenomena in some languages are known to be trans-CF (Bresnan et al., 1982; Huybregts, 1984; Shieber, 1985)
- Development of proof nets for CCG (Buch, 2009)
- Encoding of proof net conditions for MCS extension of LCG (Komatsu, 2021)

# Low-data/unsupervised categorial parsing

- Recent unsupervised parsers have relatively *ad hoc* structure (Drozdov et al., 2019)
  - Advantages of formalism-driven structure unexplored
- Novel loss functions suggest a path to structured learning from unlabelled data
  - Structural constraints may help lower data requirements

# Subcategorial considerations in statistical categorial parsing

—

Aditya Bhargava