Robust Conversion of CCG Derivations to Phrase Structure Trees

Jonathan K. Kummerfeld†, Dan Klein† and James R. Curran‡
UC Berkeley† and U Sydney‡
Motivation
Motivation

- Various parsing formalisms: HPSG, CCG, LTAG, etc.
Motivation

- Various parsing formalisms: HPSG, CCG, LTAG, etc.

- Scientific Motivation: Investigate the relationship between formalisms
Motivation

- Various parsing formalisms: HPSG, CCG, LTAG, etc.

- Scientific Motivation: Investigate the relationship between formalisms

- Engineering Motivation: Create a useful conversion tool
labeled his death a suicide
labeled his death a suicide
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide.

(labeled (his death) a suicide)

(((S[dcl] \ NP)/NP)/NP NP/N N NP/N N NP/N N NP/N N NP/N N NP/N N NP/N N)

((S[dcl] \ NP)/NP) → ((S[dcl] \ NP)/NP)
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide.
Naive Solution
Naive Solution

Categories → Labels
Naive Solution

Categories → Labels

Combinator Applications → Brackets
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide.
Italian magistrates labeled his death a suicide

Clark and Curran (2009)
Italian magistrates labeled his death a suicide.
Clark and Curran (2009)

more than doubled …

\[
\frac{(S\backslash NP)}{(S\backslash NP)} \quad S[dcl]\backslash NP
\]

\[
\frac{S[dcl]\backslash NP}{S[dcl]\backslash NP}
\]
Clark and Curran (2009) also worried ... 

\[
\text{more than doubled ... } \begin{array}{c}
(S\backslash NP)/(S\backslash NP) \\
S[\text{dcl}]\backslash NP
\end{array} \rightarrow \begin{array}{c}
S[\text{dcl}]\backslash NP
\end{array}
\]

\[
\text{also worried ... } \begin{array}{c}
(S\backslash NP)/(S\backslash NP) \\
S[\text{dcl}]\backslash NP
\end{array} \rightarrow \begin{array}{c}
S[\text{dcl}]\backslash NP
\end{array}
\]
Clark and Curran (2009)

\[
\begin{align*}
\text{VP} & \quad \text{ADVP} \\
& \quad \text{VBD} \\
\text{more than} \quad \text{doubled} \quad \ldots \\
(S\backslash NP)/(S\backslash NP) & \quad S[\text{dcl}]\backslash NP
\end{align*}
\]
Clark and Curran (2009)
Clark and Curran (2009)

more than doubled ...
(S\NP)/(S\NP)  S[dcl]\NP

also worried ...
(S\NP)/(S\NP)  S[dcl]\NP

VP

ADVP
more than
(S\NP)/(S\NP)  S[dcl]\NP

VBD

doubled ...
S[dcl]\NP

ADVP

...  RB

VP

VBN

more than doubled ...
(S\NP)/(S\NP)  S[dcl]\NP

NP

S
Clark and Curran (2009)

more than 
(S\NP)/(S\NP) 
S[dcl]\NP

doubled ...  
S[dcl]\NP

also
(S\NP)/(S\NP)
S[dcl]\NP

worried ...
(S\NP)/(S\NP)
S[dcl]\NP

(S\NP)/(S\NP)
S[dcl]\NP

(VP)

ADVP

VBD

VP

S

...
Resolving the Issue
Resolving the Issue

also

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

than

\( ((S \backslash NP)/(S \backslash NP)) \backslash (S[adj] \backslash NP) \)
Conversion from Gold

Clark and Curran (2009) vs This Work

F-score:
- Clark and Curran (2009): 94.6
- This Work: 96.3
Conversion from Parsed

F-score

Clark and Curran (2009)  This Work

F-score

84.6  86.2
Conversion from Parsed

F-score

100
96
92
88
84
80

All Sentences
Clean Sentences

86.2
Conversion from Parsed

<table>
<thead>
<tr>
<th></th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Sentences</td>
<td>86.2</td>
</tr>
<tr>
<td>Clean Sentences</td>
<td>91.7</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

- A more robust and accurate conversion method
Conclusion

- A more robust and accurate conversion method
- Insight into treebank differences
Conclusion

- A more robust and accurate conversion method
- Insight into treebank differences
- Code is available at: nlp.cs.berkeley.edu/software
Conclusion

- A more robust and accurate conversion method
- Insight into treebank differences
- Code is available at: nlp.cs.berkeley.edu/software

Thank you!
Comparison with PTB Parsers

<table>
<thead>
<tr>
<th></th>
<th>All Sentences</th>
<th>Clean Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCG - Auli and Lopez (2011)</td>
<td>86.2</td>
<td>90.1</td>
</tr>
<tr>
<td>PTB - Klein and Manning (2003)</td>
<td>85.8</td>
<td>89.8</td>
</tr>
<tr>
<td>PTB - Petrov and Klein (2007)</td>
<td></td>
<td>93.6</td>
</tr>
</tbody>
</table>
Parsing vs. Converting

![Scatter plot showing the distribution of converted C&C and Parseval values.](scatter_plot.png)
Parsing vs. Converting

![Graph showing the comparison between Native C&C, Ideps and Converted Gold, EVALB. The graph is a scatter plot with the x-axis labeled Converted Gold, EVALB and the y-axis labeled Native C&C, Ideps. The plot indicates a trend where higher values of Converted Gold correspond to higher values of Native C&C, suggesting a positive correlation.](image-url)