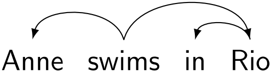
Configurable Dependency Tree Extraction from CCG Derivations

Kilian Evang

Heinrich Heine University Düsseldorf

2020-12-13
Universal Dependencies Workshop

Anne	swims	in	Rio	
NP	S[dcl]\ NP	((S \ NP)\(S \ NP))/ NP	NP	
		(S \ NP)\(S \ NP)		$>^0$
	S[dcl]\ NP			$<^0$
S[dcl]				$<^0$

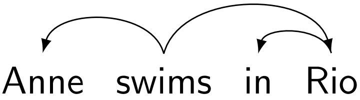

Anne swims in Rio

Combinatory Categorial Grammar (CCG)

Anne	swims	in	Rio	
NP	S[dcl]\ NP	((S \ NP)\(S \ NP))/ NP	NP	
		(S \ NP)\(S \ NP)		$>^0$
	S[dcl]\ NP			$<^0$
S[dcl]				$<^0$

- grammar formalism developed by Mark Steedman
- transparent syntax-semantics interface, elegant handling of coordination
- popular in computational semantics; Parallel Meaning Bank (PMB)

Why Convert CCG Derivations to Dependency Trees?



- feature extraction
- cross-framework parser comparison
- convert word-based to span-based annotation of semantic roles
- ...

Example 1

$$
\begin{array}{ccc}
\text{John} & \text{sees} & \text{Mary} \\
\hline
\mathsf{NP} & (\mathsf{S}\backslash\mathsf{NP})/\mathsf{NP} & \mathsf{NP}
\end{array}
$$

$$
\frac{(\mathsf{S}\backslash\mathsf{NP})/\mathsf{NP} \quad \mathsf{NP}}{\mathsf{S}\backslash\mathsf{NP}} >^{0}
$$

$$
\frac{\mathsf{NP} \quad \mathsf{S}\backslash\mathsf{NP}}{\mathsf{S}} <^{0}
$$

Example 1

$$\dfrac{\dfrac{\text{John}}{NP_1} \quad \dfrac{\dfrac{\text{sees}}{(S_2 \backslash NP_1) / NP_3} \quad \dfrac{\text{Mary}}{NP_3}}{S_2 \backslash NP_1} >^0}{S_2} <^0$$

Example 1

John	sees	Mary
NP_1	$(S_2 \backslash NP_1) / NP_3$	NP_3

Example 1

John	sees	Mary
NP_1	$(S_2 \backslash NP_1)/NP_3$	NP_3

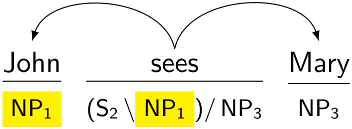
Example 1

John	sees	Mary
NP_1	$(S_2 \backslash NP_1) / NP_3$	NP_3

Example 1

John	sees	Mary
NP_1	$(S_2 \backslash NP_1) / NP_3$	NP_3

Example 1



Example 2

$$\text{Anna} \quad \text{swims} \quad \text{in} \quad \text{Rio}$$

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

$$(S \backslash NP)\backslash(S \backslash NP) \quad >^{0}$$

$$S \backslash NP \quad <^{0}$$

$$S \quad <^{0}$$

Example 2

$$
\begin{array}{cccc}
\underline{\text{Anna}} & \underline{\text{swims}} & \underline{\text{in}} & \underline{\text{Rio}} \\
\mathrm{NP}_1 & \mathrm{S}_2 \backslash \mathrm{NP} & ((\mathrm{S}_3 \backslash \mathrm{NP}_1)\backslash(\mathrm{S}_2 \backslash \mathrm{NP}))/\,\mathrm{NP}_4 & \mathrm{NP}_4
\end{array}
$$

$$
\frac{((\mathrm{S}_3 \backslash \mathrm{NP}_1)\backslash(\mathrm{S}_2 \backslash \mathrm{NP}))/\,\mathrm{NP}_4 \quad \mathrm{NP}_4}{(\mathrm{S}_3 \backslash \mathrm{NP}_1)\backslash(\mathrm{S}_2 \backslash \mathrm{NP})} >^0
$$

$$
\frac{\mathrm{S}_2 \backslash \mathrm{NP} \quad (\mathrm{S}_3 \backslash \mathrm{NP}_1)\backslash(\mathrm{S}_2 \backslash \mathrm{NP})}{\mathrm{S}_3 \backslash \mathrm{NP}_1} <^0
$$

$$
\frac{\mathrm{NP}_1 \quad \mathrm{S}_3 \backslash \mathrm{NP}_1}{\mathrm{S}_3} <^0
$$

Example 2

$$\frac{\text{Anna}}{NP_1} \quad \frac{\text{swims}}{S_2 \backslash NP} \quad \frac{\text{in}}{((S_3 \backslash NP_1)\backslash(S_2 \backslash NP))/ NP_4} \quad \frac{\text{Rio}}{NP_4}$$

Example 2

Anna	swims	in	Rio
NP_1	$S_2 \backslash NP$	$((S_3 \backslash NP_1)\backslash(S_2 \backslash NP))/ NP_4$	NP_4

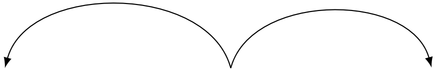
Example 2

Anna	swims	in	Rio
NP_1	$S_2 \backslash NP$	$((S_3 \backslash NP_1)\backslash(S_2 \backslash NP))/NP_4$	NP_4

Example 2

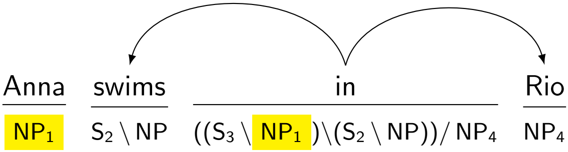
Anna	swims	in	Rio
NP_1	$S_2 \backslash NP$	$((S_3 \backslash NP_1)\backslash(S_2 \backslash NP))/NP_4$	NP_4

Example 2

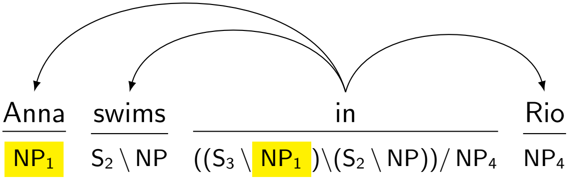


Anna	swims	in	Rio
NP_1	$S_2 \backslash NP$	$((S_3 \backslash NP_1) \backslash (S_2 \backslash NP)) / NP_4$	NP_4

Example 2


Anna
swims
in
Rio
NP_1
$S_2 \backslash NP$
$((S_3 \backslash NP_1) \backslash (S_2 \backslash NP)) / NP_4$
NP_4

Example 2



Example 2, Attempt 2

Anna	swims	in	Rio
NP_1	$S_2 \backslash NP$	$((S_3 \backslash NP_1)\backslash(S_2 \backslash NP))/NP_4$	NP_4

Example 2, Attempt 2

Anna	swims	in	Rio
$NP_{1,F}$	$S_2 \backslash NP$	$((S_3 \backslash NP_{1,F})\backslash(S_2 \backslash NP))/ NP_4$	NP_4

Example 2, Attempt 2

Anna	swims	in	Rio
$NP_{1,F}$	$S_2 \backslash NP$	$((S_3 \backslash NP_{1,F}) \backslash (S_2 \backslash NP)) / NP_{4,A}$	$NP_{4,A}$

Example 2, Attempt 2

$$\frac{\text{Anna}}{NP_{1,F}} \quad \frac{\text{swims}}{S_{2,A} \backslash NP} \quad \frac{\text{in}}{((S_3 \backslash NP_{1,F}) \backslash (S_{2,A} \backslash NP)) / NP_{4,A}} \quad \frac{\text{Rio}}{NP_{4,A}}$$

Example 2, Attempt 2

Anna	swims	in	Rio
$NP_{1,F}$	$S_{2,A} \backslash NP$	$((S_3 \backslash NP_{1,F}) \backslash (S_{2,A} \backslash NP)) / NP_{4,A}$	$NP_{4,A}$

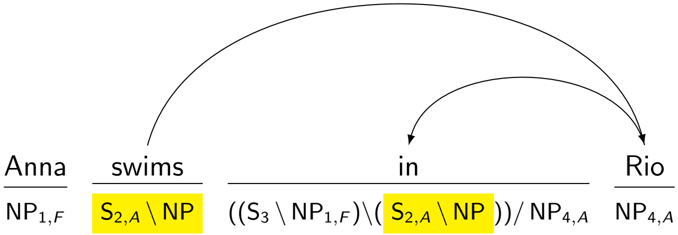
Example 2, Attempt 2

Anna	swims	in	Rio
$NP_{1,F}$	$S_{2,A} \backslash NP$	$((S_3 \backslash NP_{1,F}) \backslash (S_{2,A} \backslash NP)) / NP_{4,A}$	$NP_{4,A}$

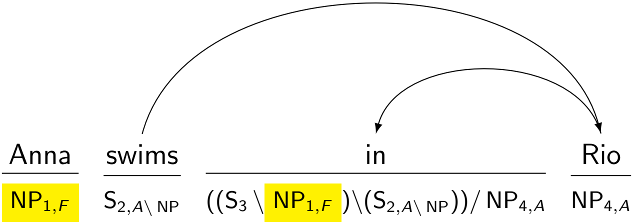
Example 2, Attempt 2

Anna	swims	in	Rio
$NP_{1,F}$	$S_{2,A} \backslash NP$	$((S_3 \backslash NP_{1,F}) \backslash (S_{2,A} \backslash NP)) / NP_{4,A}$	$NP_{4,A}$

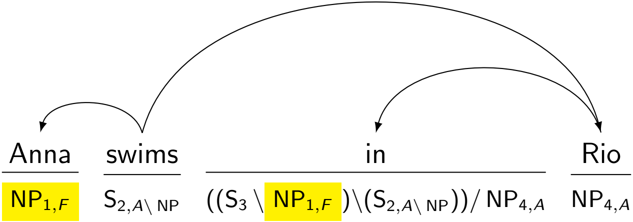
Example 2, Attempt 2



Example 2, Attempt 2



Example 2, Attempt 2



Categories with Argument-head Marking

Module	Example
Coordinating conjunctions	$(X\backslash X)/X$
Modifiers	$X\|X$
Copulas*	$(\mathrm{S}[X]\backslash\mathrm{NP})\|\mathrm{NP}$
Auxiliaries and modals*	$(\mathrm{S}[X]\backslash\mathrm{NP})\|(\mathrm{S}[Y]\backslash\mathrm{NP})$
Adpositions	$\mathrm{PP}\|\mathrm{NP}$
	$(X\|X)\|\mathrm{NP}$
Determiners	$\mathrm{NP}\|\mathrm{N}$
Possessive suffix	$(\mathrm{NP}/(\mathrm{N}/\mathrm{PP}))\backslash\mathrm{NP}$
Subordinating conjunctions	$(\mathrm{S}\|\mathrm{S})\|\mathrm{S}[X]$
Complementizers	$\mathrm{S[em]}/\mathrm{S[dcl]}$
Relativizers	$(\mathrm{N}\backslash\mathrm{N})/(\mathrm{S[dcl]}\|\mathrm{NP})$
Fronted *wh*-words	$\mathrm{S[wq]}/(\mathrm{S[q]}/(\mathrm{S[adj]}/\mathrm{NP}))$

* need semantic tags/symbols to identify

Comparison of ccg2dep Algorithms

Clark et al. (2002), Koller and Kuhlmann (2009), Bisk and Hockenmaier (2013)

	CHS02	KK09	BH13	This work
Compact specification	-	+	+	+
Nonlocal dependencies	+	-	-	-
Labeled dependencies	+	-	-	-
Configurable dependency directions	+	-	+	+
Supports all cats, rules in PMB	+	-	-	+

Evaluation (PMB gold, parts 00 and 01)

	dev				test			
	en	de	it	nl	en	de	it	nl
Statistics								
Sentences	383	100	73	37	208	35	27	13
Mean length	6.0	5.53	5.1	5.7	6.4	5.4	5.6	5.1
Scores								
UAS	.982	.993	.936	.921	.977	.937	.976	1
Error analysis								
PMB annotation error	1	1						
Attachment ambiguity	3				2	1		
Punctuation	4				1			
Tag question	2							
Embedded question					1			
Date					1			
Subordinating conjunction					1			
Possessive suffix					1	1		
Copula/auxiliary/modal				1	1		1	
Expletive nominal			2	1				
Question verb		3	1					
Inverted copula	1							
Pseudo-copula		1	1	2				
Non-local dependencies	4							

Conclusions

- simple: only list function word categories
- configurable: turn on/off as needed
- limited: no labels, no non-local dependencies
- imperfect: treebank has long tail of constructions

Thank you! Questions?

Bibliography I

Bisk, Y. and Hockenmaier, J. (2013). An HDP model for inducing combinatory categorial grammars. *Transactions of the Association for Computational Linguistics*, 1:75–88.

Clark, S., Hockenmaier, J., and Steedman, M. (2002). Building deep dependency structures using a wide-coverage CCG parser. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 327–334, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Koller, A. and Kuhlmann, M. (2009). Dependency trees and the strong generative capacity of CCG. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 460–468, Athens, Greece. Association for Computational Linguistics.