CCG Supertagging with a Recurrent Neural Network

Wenduan Xu [1] Michael Auli [2] Stephen Clark [1]

[1]Cambridge University

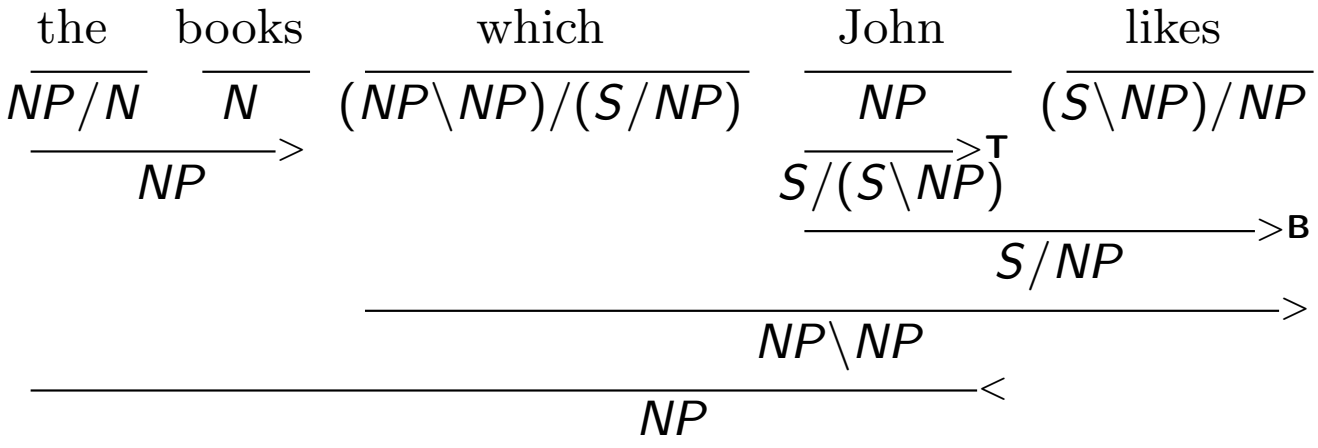
[2]Facebook AI Research

July, 2015

Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]

Combinatory Categorial Grammar (CCG)



Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]

Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]
- 425 lexical categories in the standard models [Clark & Curran]

Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]
- 425 lexical categories in the standard models [Clark & Curran]
  - compared to about 50 POS tags in PTB

Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]
- 425 lexical categories in the standard models [Clark & Curran]
  - compared to about 50 POS tags in PTB
- Only a dozen combinatory rules

Combinatory Categorial Grammar (CCG)

- A lexicalized formalism [Steedman, 2000]
- 425 lexical categories in the standard models [Clark & Curran]
  - compared to about 50 POS tags in PTB
- Only a dozen combinatory rules
  - compared to over 500K for a PTB parser [Petrov and Klein, 2007]

CCG Parsing

- Method 1: leave all lexical ambiguity to the parser
  - was done in the first CCG parser [Hockenmaier, 2003]

CCG Parsing

- Method 1: leave all lexical ambiguity to the parser
  - was done in the first CCG parser [Hockenmaier, 2003]
- Method 2: use only 1-best supertags
  - accuracy not good enough (C&C supertagger accuracy is about 92%)

CCG Parsing

- Method 1: leave all lexical ambiguity to the parser
  - was done in the first CCG parser [Hockenmaier, 2003]
- Method 2: use only 1-best supertags
  - accuracy not good enough (C&C supertagger accuracy is about 92%)
- Current optimal solution, two stage process [Clark & Curran, 2007]:
  - multi-supertagging + parsing

CCG Parsing

- Method 1: leave all lexical ambiguity to the parser
  - was done in the first CCG parser [Hockenmaier, 2003]
- Method 2: use only 1-best supertags
  - accuracy not good enough (C&C supertagger accuracy is about 92%)
- Current optimal solution, two stage process [Clark & Curran, 2007]:
  - multi-supertagging + parsing
- In [Auli and Lopez, 2011]
  - called integrated supertagging and parsing
  - still, the same supertagging model and two-stage process as in C&C

Adaptive Supertagging [Clark & Curran, 2007]

Start with an initial prob. cutoff β

$$\frac{\text{He}}{NP} \quad \frac{\text{reads}}{(S[pss]\backslash NP)/NP} \quad \frac{\text{the}}{NP/N} \quad \frac{\text{book}}{N}$$

Adaptive Supertagging [Clark & Curran, 2007]

Prune a category, if its probability is below β times the prob. of the best category

$$\frac{\text{He}}{NP} \quad \frac{\text{reads}}{(S[pss]\backslash NP)/NP} \quad \frac{\text{the}}{NP/N} \quad \frac{\text{book}}{N}$$

Adaptive Supertagging [Clark & Curran, 2007]

Decrease β if no spanning analysis

$$\frac{\text{He}}{\begin{array}{c}NP\\ N\\ N/N\end{array}} \quad \frac{\text{reads}}{\begin{array}{c}(S[pss]\backslash NP)/NP\\ (S\backslash NP)/NP\\ S\backslash NP\end{array}} \quad \frac{\text{the}}{\begin{array}{c}NP/N\\ NP/NP\end{array}} \quad \frac{\text{book}}{\begin{array}{c}N\\ (S\backslash NP)/NP\end{array}}$$

Adaptive Supertagging [Clark & Curran, 2007]

Decrease β if no spanning analysis

He	reads	the	book
NP	$(S[pss]\backslash NP)/NP$	NP/N	N
N	$(S\backslash NP)/NP$	NP/NP	$(S\backslash NP)/NP$
N/N	$S\backslash NP$		
NP/NP	$(S[pt]\backslash NP)/NP$		
	$(S[dcl]\backslash NP)/NP$		

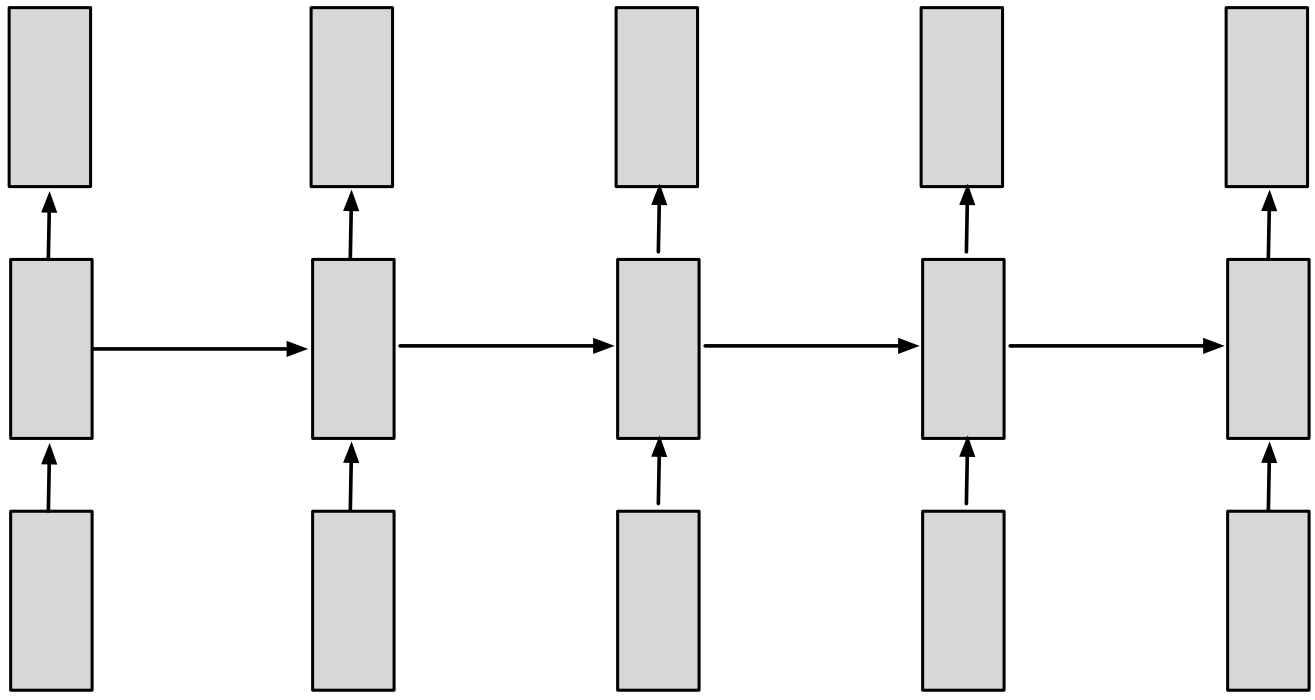
CCG Supertagging

- Affects search for training and testing [Clark & Curran, 2007]
- Also affects parsing efficiency [Curran et all, 2006]
- The standard MaxEnt model
  - relies heavily on POS tags
  - uses only sparse indicator features
  - considers only local contexts
- The recent feed-forward neural supertagger [Lewis & Steedmanm, 2014]
  - no POS tag features
  - all dense features
  - still considers only local contexts

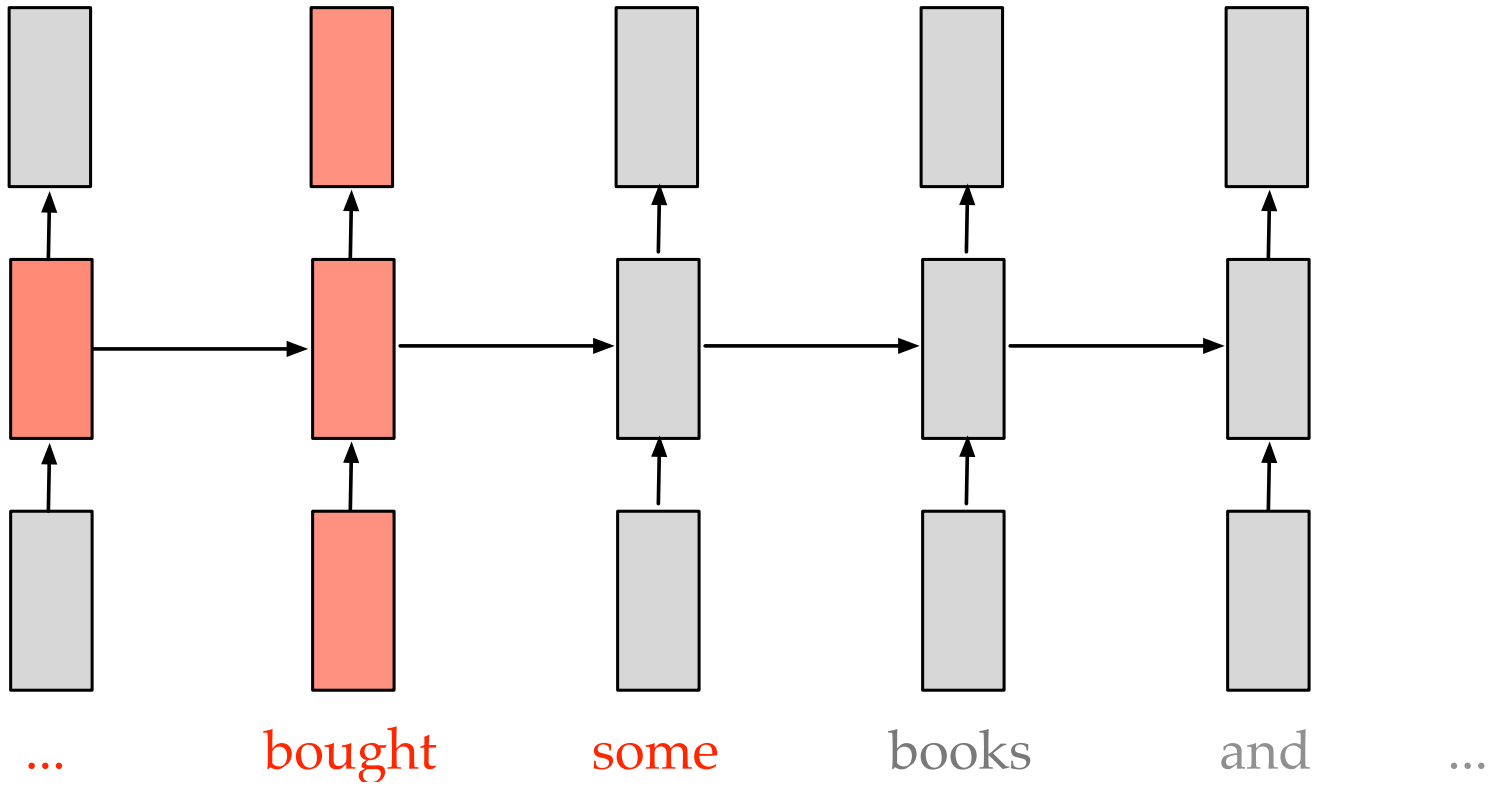
Supertagging with a RNN

- Using only dense features
  - word embedding
  - suffix embedding
  - capitalization
- The input layer is a concatenation of all embeddings of all words in a context window

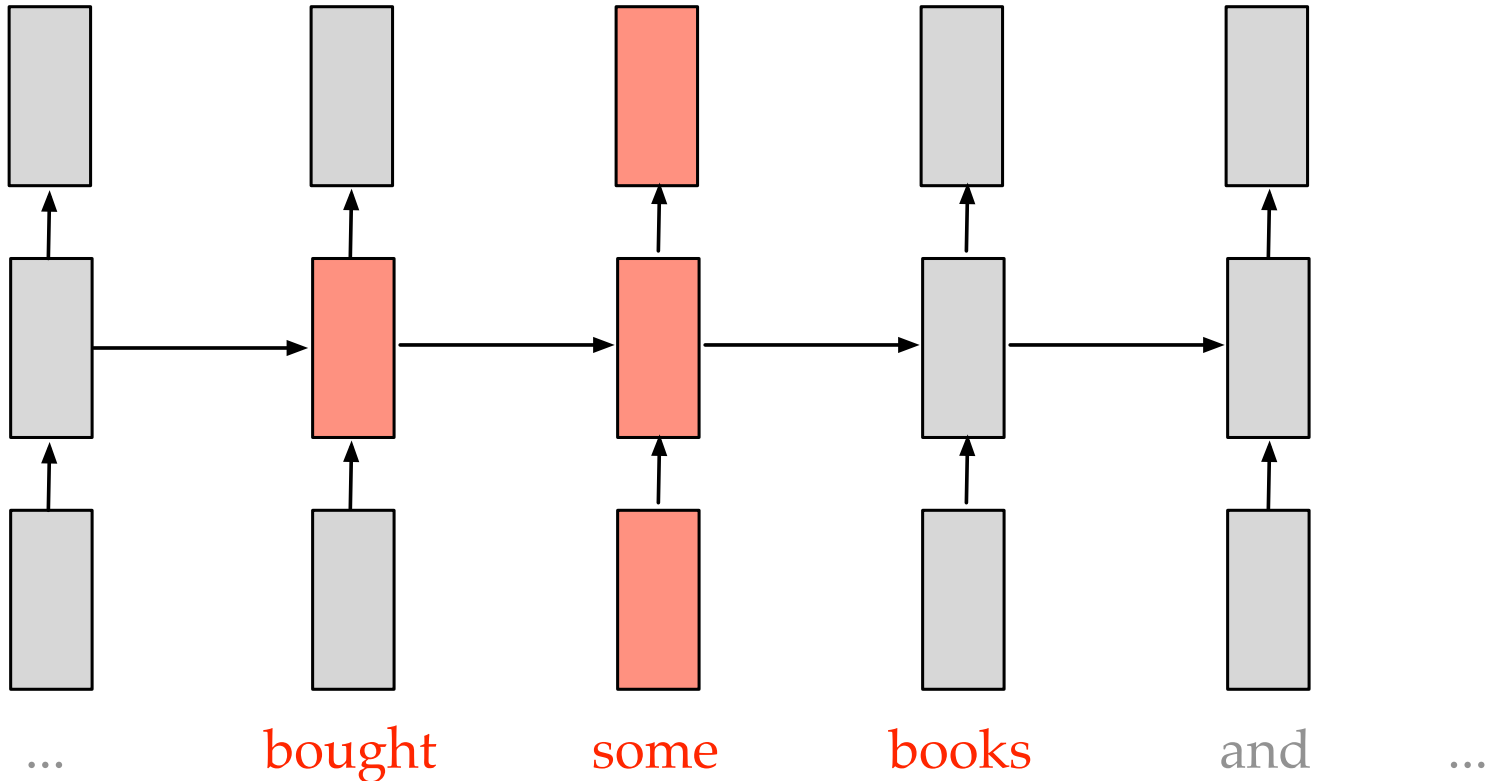
Supertagging with a RNN



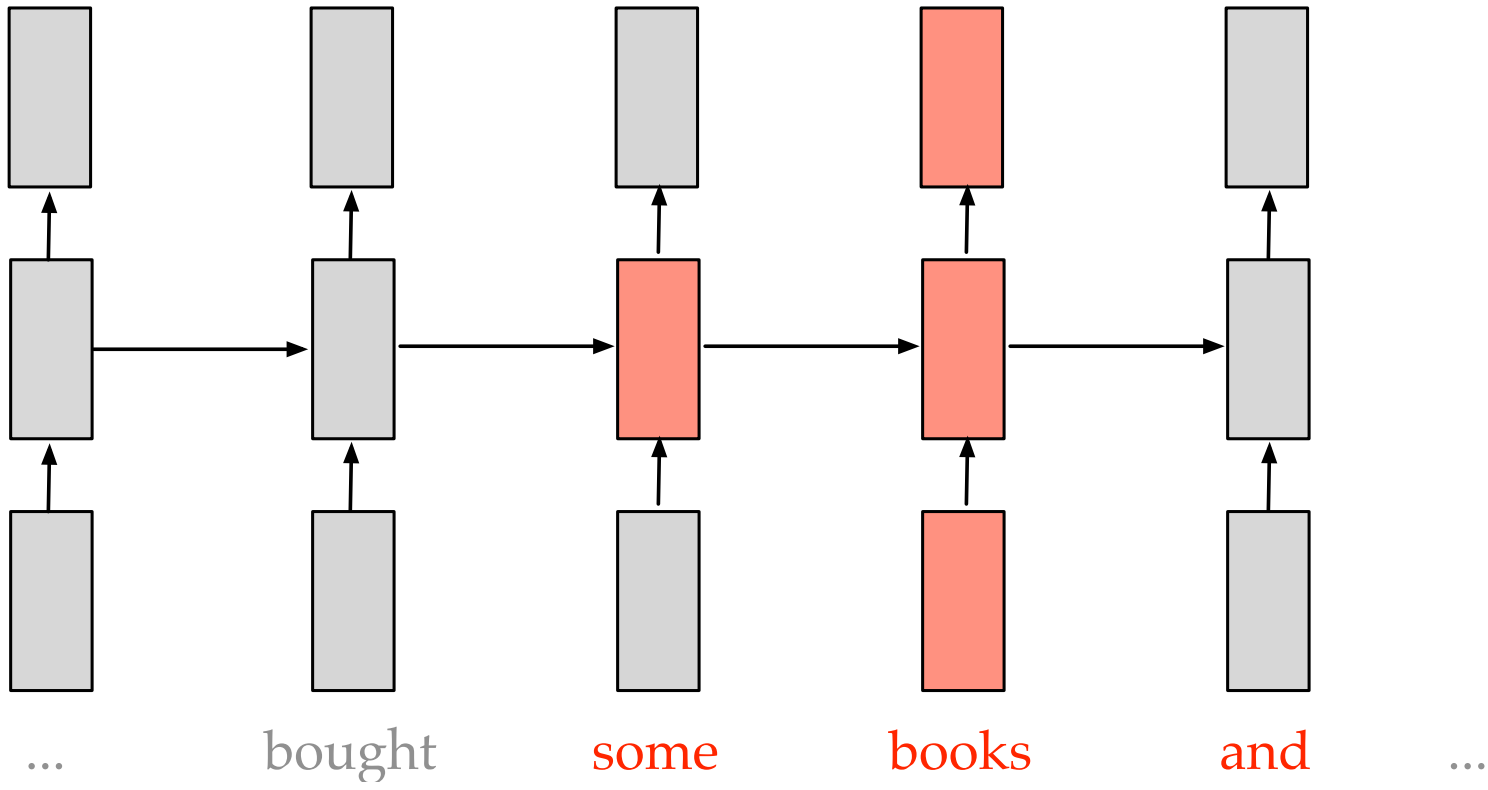
Supertagging with a RNN



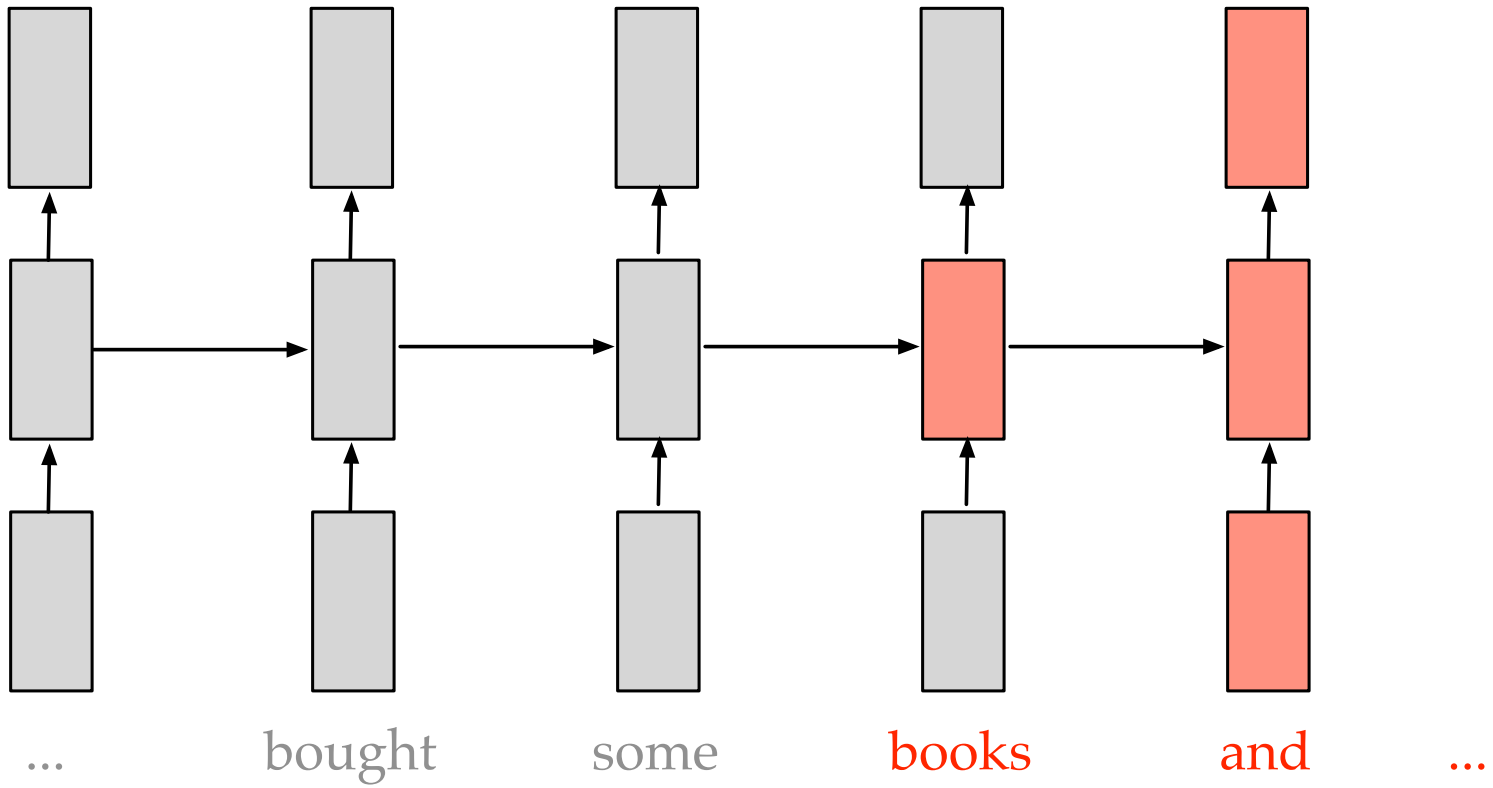
Supertagging with a RNN



Supertagging with a RNN



Supertagging with a RNN



Training & Experiments

- Mini-batched BPTT [Rumelhart et al., 1988; Mikolov, 2012]
- A context window-size of 7, a BPTT step size of 9
- 50-dim scaled Turian embeddings [Turian et al., 2010]
- Other two look-up tables randomly initialized
- Embedding fine-tuning during training
- Dropout regularization
- Parsing experiments: use the same supertagger prob. cutoff values as C&C

1-best Supertagging Results: dev

Model	Accuracy	Time
C&C (gold POS)	92.60	-
C&C (auto POS)	91.50	0.57
NN	91.10	21.00
RNN	92.63	-
RNN+dropout	93.07	2.02

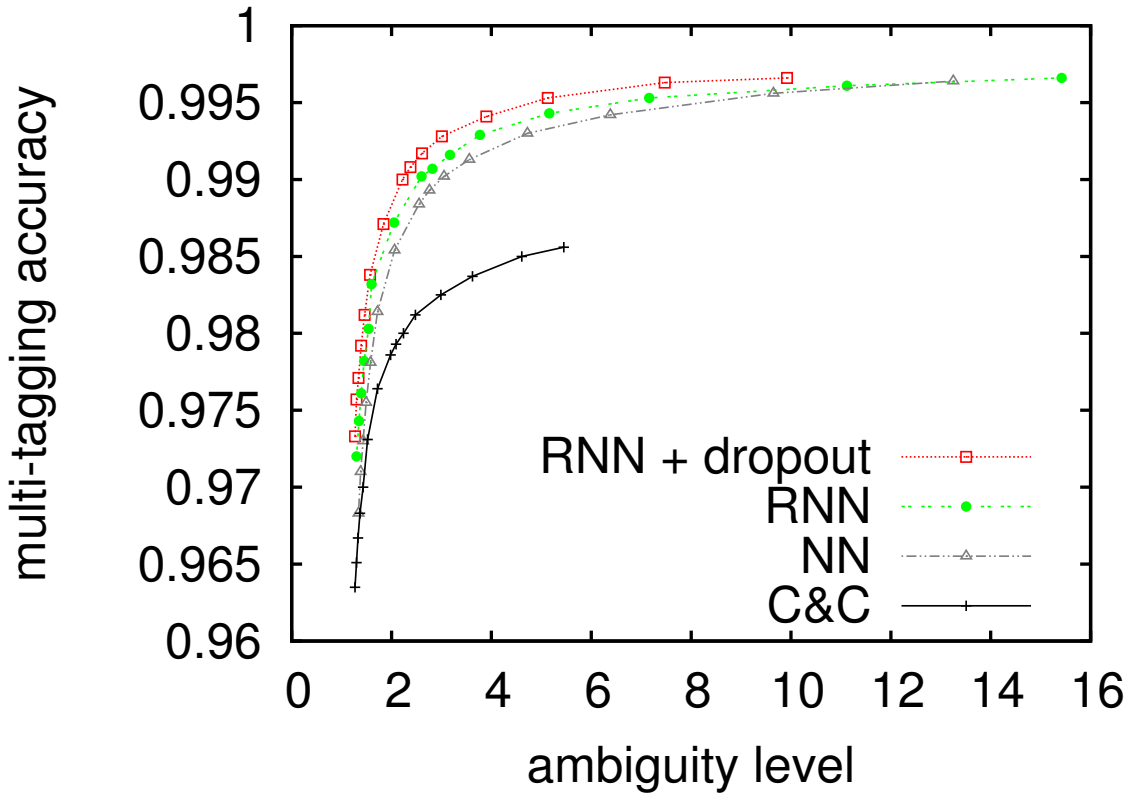
Table 1 : 1-best tagging accuracy and speed comparison on CCGBank Section 00 with a single CPU core (1,913 sentences), tagging time in secs.

1-best Supertagging Results: test

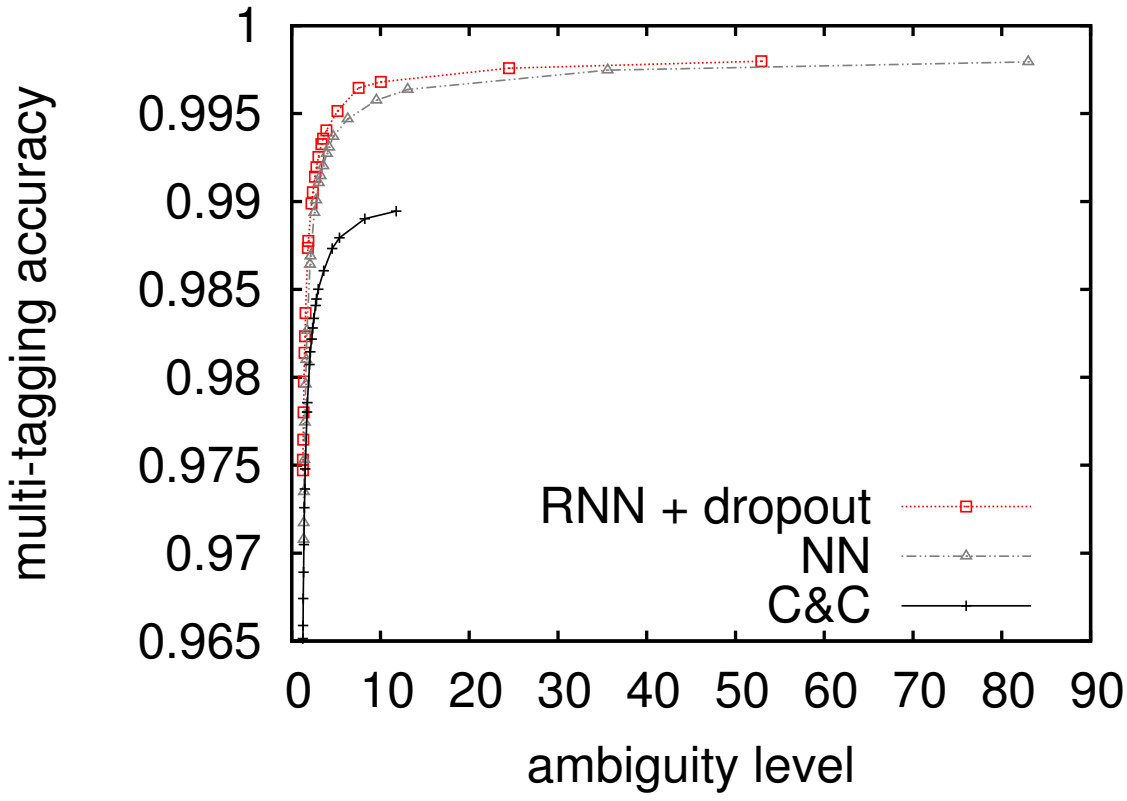
Model	Section 23	Wiki	Bio
C&C (gold POS)	93.32	88.80	91.85
C&C (auto POS)	92.02	88.80	89.08
NN	91.57	89.00	88.16
RNN	93.00	90.00	88.27

Table 2 : 1-best tagging accuracy comparison on CCGBank Section 23 (2,407 sentences), Wikipedia (200 sentences) and Bio-GENIA (1,000 sentences).

Multi-tagging Results: dev



Multi-tagging Results: test



Final Parsing Results

	CCGBank Section 23				Wikipedia			
	LP	LR	LF	cov.	LP	LR	LF	
C&C	86.24	84.85	85.54	99.42	81.58	80.08	80.83	99.50
(NN)	86.71	85.56	86.13	99.92	82.65	81.36	82.00	**100**
(RNN)	**87.68**	**86.47**	**87.07**	**99.96**	**83.22**	**81.78**	**82.49**	**100**
C&C	86.24	84.17	85.19	100	81.58	79.48	80.52	100
(NN)	86.71	85.40	86.05	100	-	-	-	-
(RNN)	**87.68**	**86.41**	**87.04**	100	-	-	-	-

Table 3 : Parsing test results (auto POS). We evaluate on all sentences (100% coverage) as well as on only those sentences that returned spanning analyses (% cov.). RNN and NN both have 100% coverage on the Wikipedia data.

The End

Thank You!