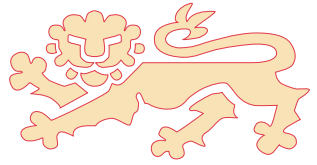
Adaptive Supertagging for Faster Parsing

Jonathan K. Kummerfeld
Supervisor – Dr. James Curran

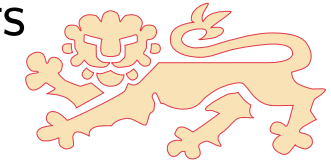


Motivation – Natural Language Processing

Build systems that use text intelligently

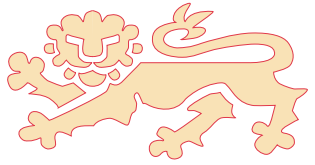
- Question Answering
  "When was the PM of Australia not the leader of the majority party in the House of Representatives?"
- Machine Translation
  Translate this website from French into English
- Anaphora Resolution
  "Steve is the leader of Apple, and without him it would be far less successful."

All of these rely on syntactic information, produced by parsers



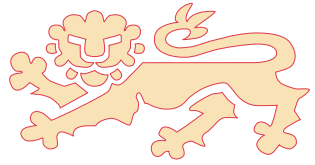
Tagging and Parsing

$$\text{One} \quad \text{claims} \quad \text{he} \quad \text{is} \quad \text{pro} - \text{choice}$$



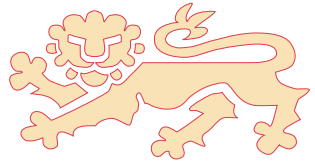
Part of Speech Tagging

One	claims	he	is	pro – choice
NN	*VBZ*	*PRP*	*VBZ*	*JJ*



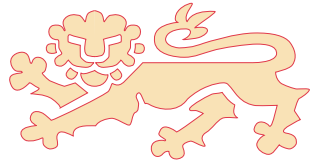
Combinatory Categorial Grammar (CCG) – Supertagging

$$\frac{\text{One}}{N} \quad \frac{\text{claims}}{(S\backslash NP)/S} \quad \frac{\text{he}}{NP} \quad \frac{\text{is}}{(S\backslash NP)/(S\backslash NP)} \quad \frac{\text{pro}-\text{choice}}{S\backslash NP}$$



Combinatory Categorial Grammar (CCG) – Parsing

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{One}}{N}}{NP} \quad \dfrac{\text{claims}}{(S\backslash NP)/S} \quad \dfrac{\text{he}}{NP} \quad \dfrac{\dfrac{\text{is}}{(S\backslash NP)/(S\backslash NP)} \quad \dfrac{\text{pro}-\text{choice}}{S\backslash NP}\,>}{S\backslash NP}\,<}{S}\,>}{S\backslash NP}\,<}{S}$$



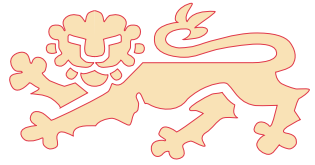
Motivation – Parsing

Parsers are slow:

- State-of-the-art, usually <1 sentence / sec
- Fastest state-of-the-art, <50 sentences / sec

Especially for our data sets:

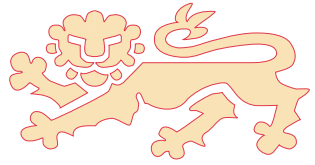
- Estimated 10,000,000,000,000 words online
- More coming



Motivation – Parsing

The key idea behind the speed of the fastest parsers today is to shift work from parsing to tagging:
For n words, each with k tags

- Tagging – $O(nk)$
- Parsing – $O(n^3k^2)$



Contributions

Core Novel Ideas

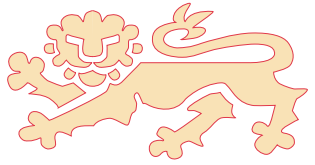
- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- Perceptron Algorithms
- Parallelisation

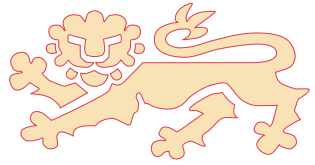
Demonstration

- Improved performance
- Domain adaptation
- Explored Supertagger – Parser interaction



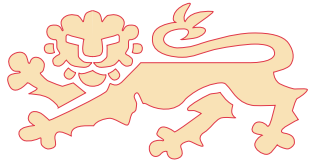
Ideal World

$$\frac{\text{One}}{N} \quad \frac{\text{claims}}{(S\backslash NP)/S} \quad \frac{\text{he}}{NP} \quad \frac{\text{is}}{(S\backslash NP)/(S\backslash NP)} \quad \frac{\text{pro} - \text{choice}}{S\backslash NP}$$



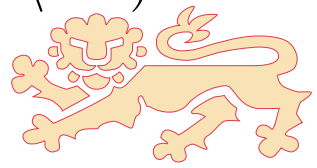
Current World – Problem

$$\frac{\text{One}}{N} \quad \frac{\text{claims}}{(S\backslash NP)/NP} \quad \frac{\text{he}}{NP} \quad \frac{\text{is}}{(S\backslash NP)/(S\backslash NP)} \quad \frac{\text{pro} - \text{choice}}{S\backslash NP}$$



Current World – Solution

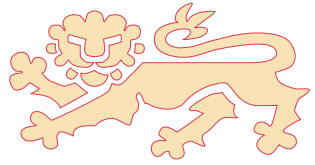
$$\dfrac{\dfrac{\text{One}}{\dfrac{N/N}{N}}_{(S/S)/(S/S)} \quad \dfrac{\text{claims}}{\dfrac{(S\backslash NP)/NP}{N}} \quad \dfrac{\text{he}}{NP} \quad \dfrac{\text{is}}{\dfrac{(S\backslash NP)/(S\backslash NP)}{(S\backslash NP)/NP}_{(S\backslash NP)/(S\backslash NP)}} \quad \dfrac{\text{pro}-\text{choice}}{\dfrac{S\backslash NP}{\dfrac{(S\backslash NP)\backslash(S\backslash NP)}{\dfrac{(S\backslash NP)/S}{\dfrac{N}{\dfrac{(S\backslash NP)/PP}{\dfrac{(S\backslash NP)/NP}{\dfrac{N/N}{(S\backslash NP)/(S\backslash NP)}}}}}}}}$$



Adaptive Supertagging

$$\begin{array}{cccccc}
\underline{\text{One}} & \underline{\text{claims}} & \underline{\text{he}} & \underline{\text{is}} & \underline{\text{pro}-\text{choice}} \\
N/N & N & NP & (S\backslash NP)/(S\backslash NP) & S\backslash NP \\
 & & & (S\backslash NP)/NP & (S\backslash NP)/PP \\
 & & & (S\backslash NP)/(S\backslash NP) & (S\backslash NP)/NP \\
 & & & & N/N
\end{array}$$

How do we teach the supertagger to produce these tags?
Use the parser!



Contributions

Core Novel Ideas

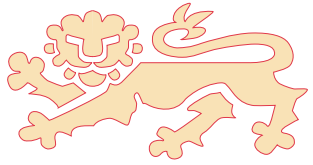
- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- **Perceptron Algorithms**
- **Parallelisation**

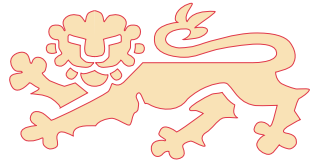
Demonstration

- Improved performance
- Domain adaptation
- Explored Supertagger – Parser interaction

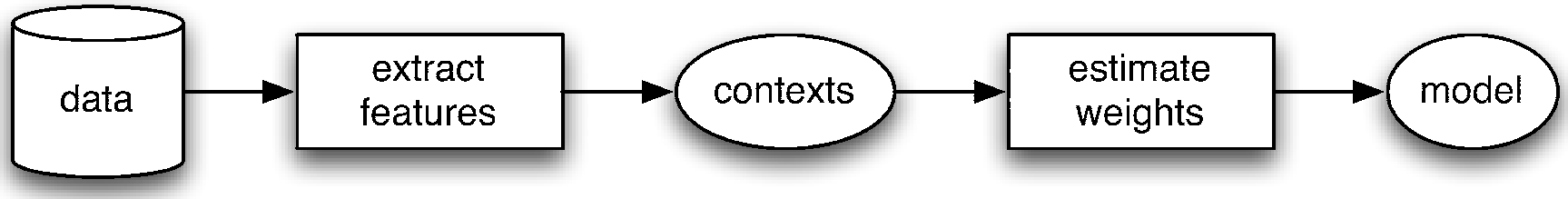


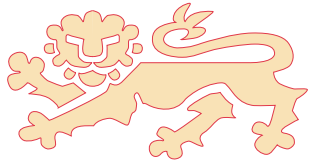
Implementation

Component	Initial System	Additions
Statistical Feature Extraction	3 Types	+9 Types
	Single thread	Parallel
Parameter Estimation	BFGS, GIS	AP, MIRA
	Single thread	Parallel

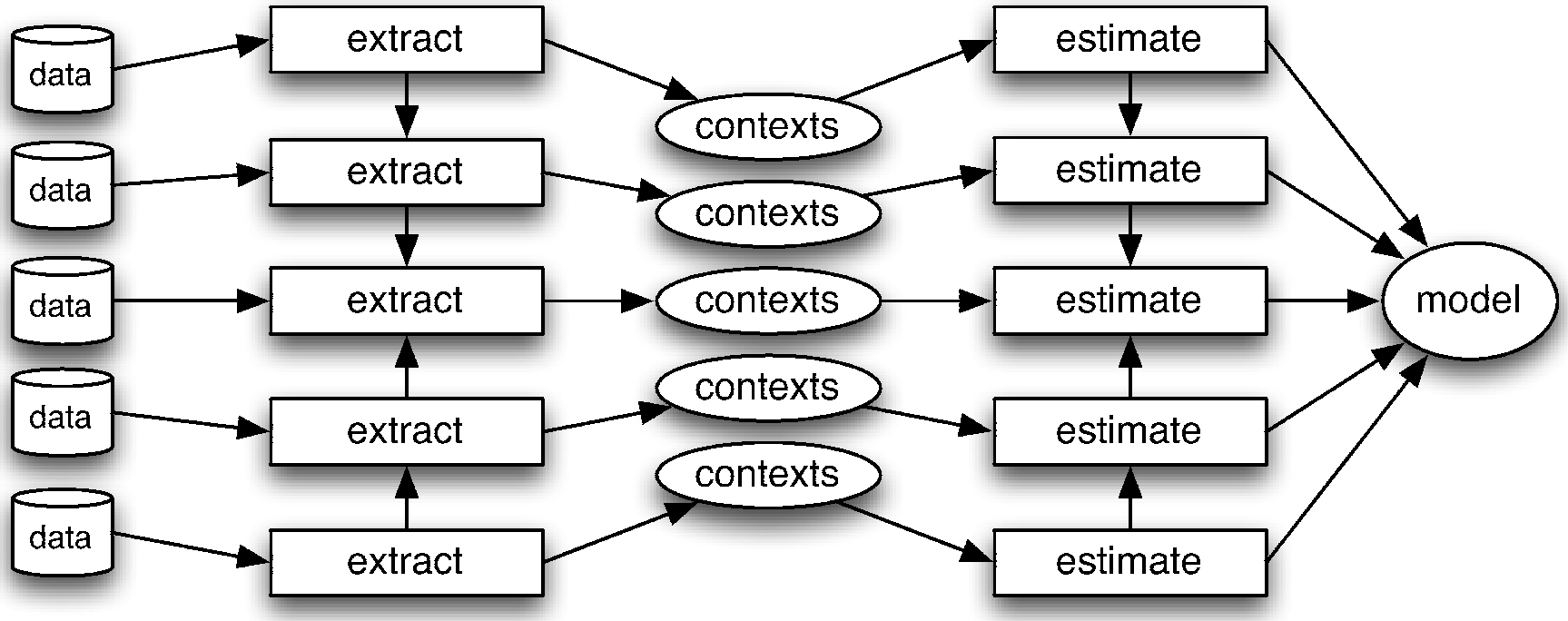


Implementation – Initial System

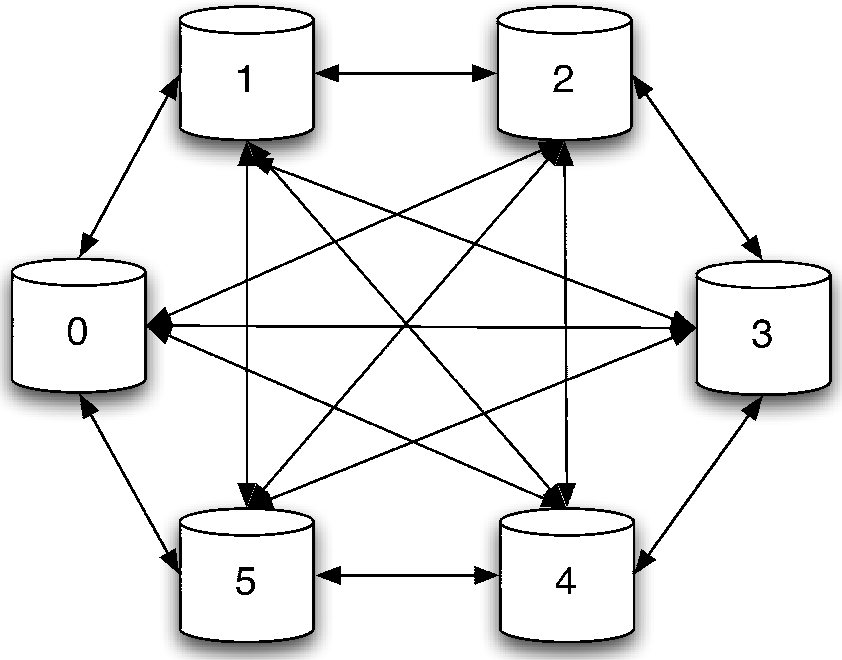


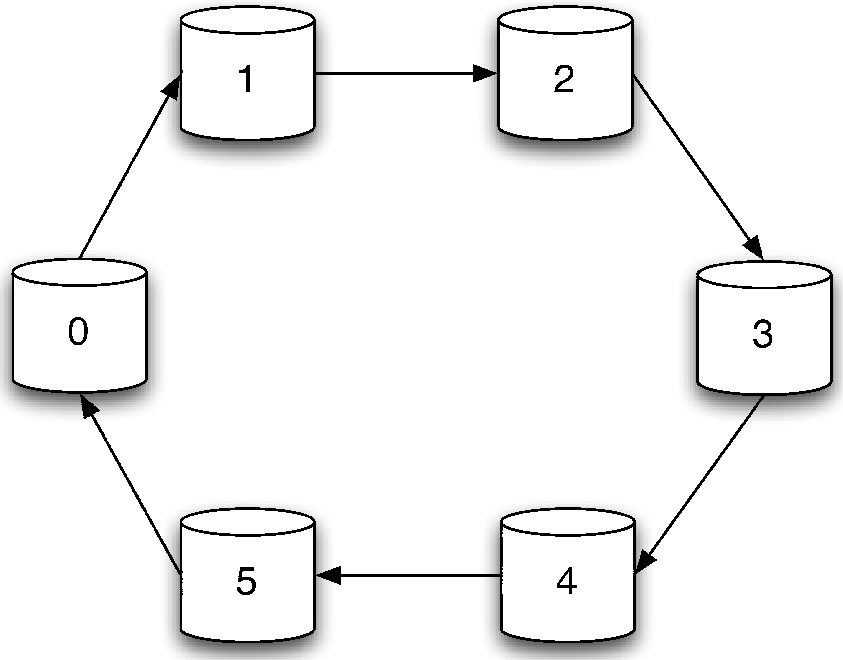


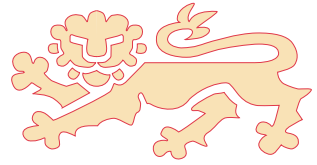
Implementation – Parallelised



Implementation – Parallelised Weight Estimation

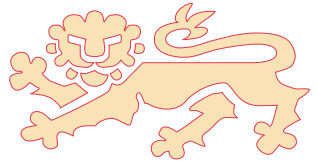






Implementation – AP and MIRA

Algorithm	Training Time (sec)		
	40k	80k	440k
GIS	7,200	14,000	*
BFGS	6,300	13,000	*
AP	76	160	950
MIRA	96	200	1,200



Contributions

Core Novel Ideas

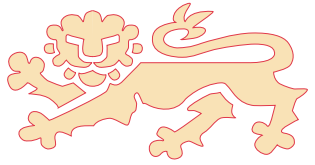
- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- Perceptron Algorithms
- Parallelisation

Demonstration

- **Improved performance**
- **Domain adaptation**
- **Exploration of Supertagger – Parser interaction**



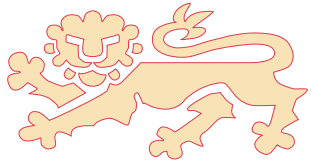
Tests

> 300 Models trained

- Algorithm – GIS, BFGS, AP, MIRA
- Training data source – Wall Street Journal, Wikipedia
- Various amounts of training data – 40k, 80k . . . 26,000k
- Number of cores – 1, 2 . . . 64
- Statistical features – Subtractive analysis of 9 new feature types

$> 12,000$ Tests

- Tested with tuned and default settings
- Evaluated on Wall Street Journal and Wikipedia
- Systematic analysis of 1,913 sentences parsed with 2,200 variations of parameters



Improved Performance

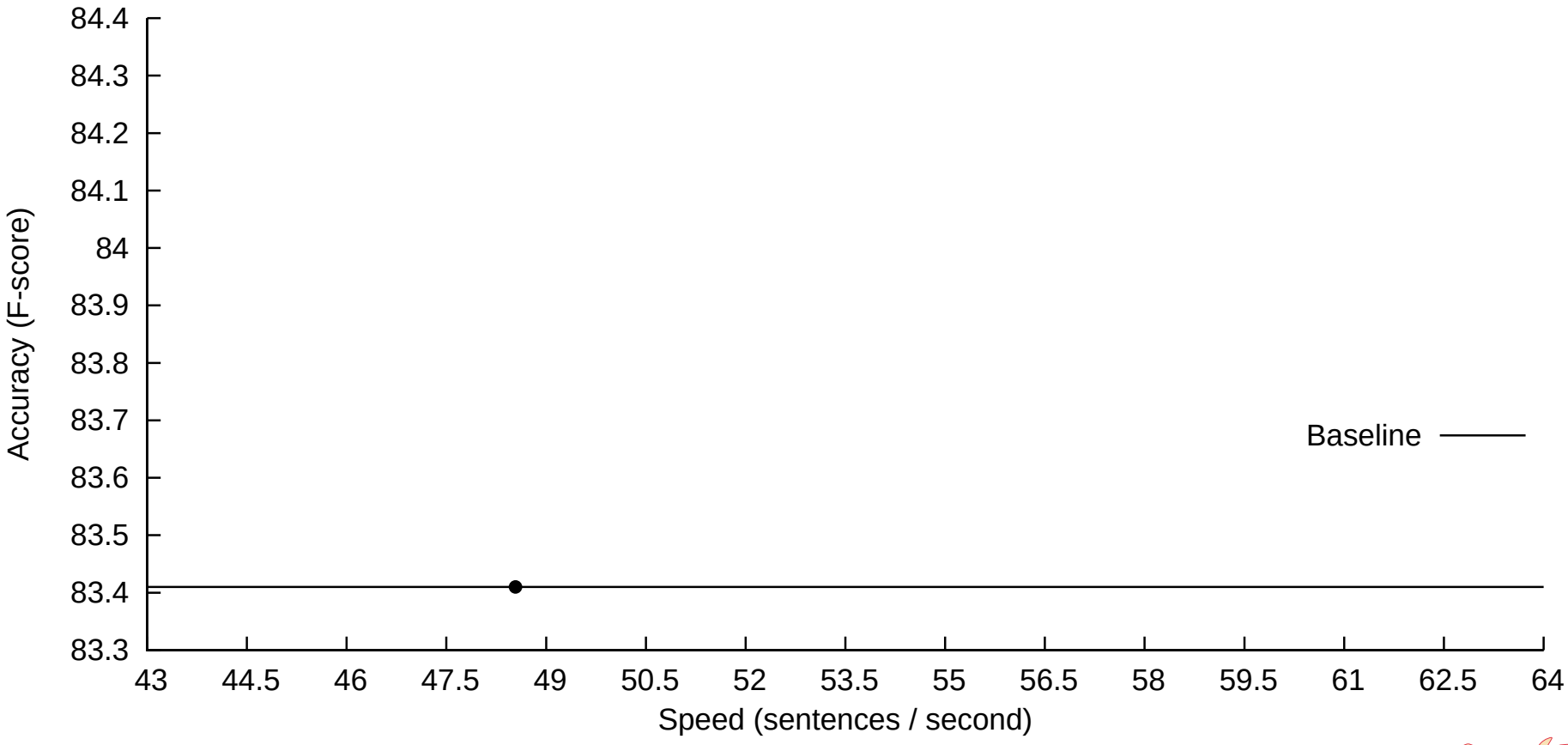


Figure: Evaluation on the Wall Street Journal

Improved Performance

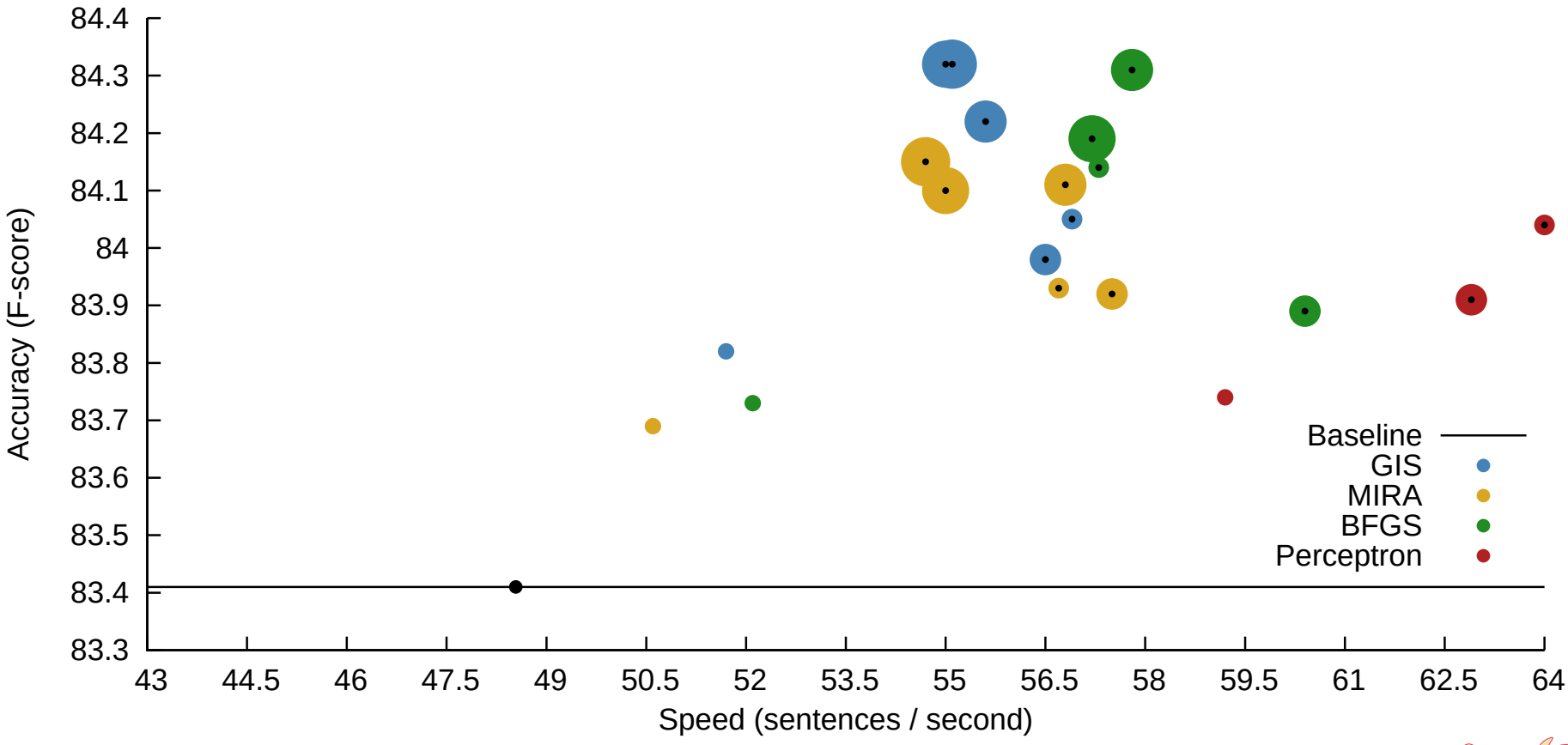


Figure: Evaluation on the Wall Street Journal

Domain Adaptation

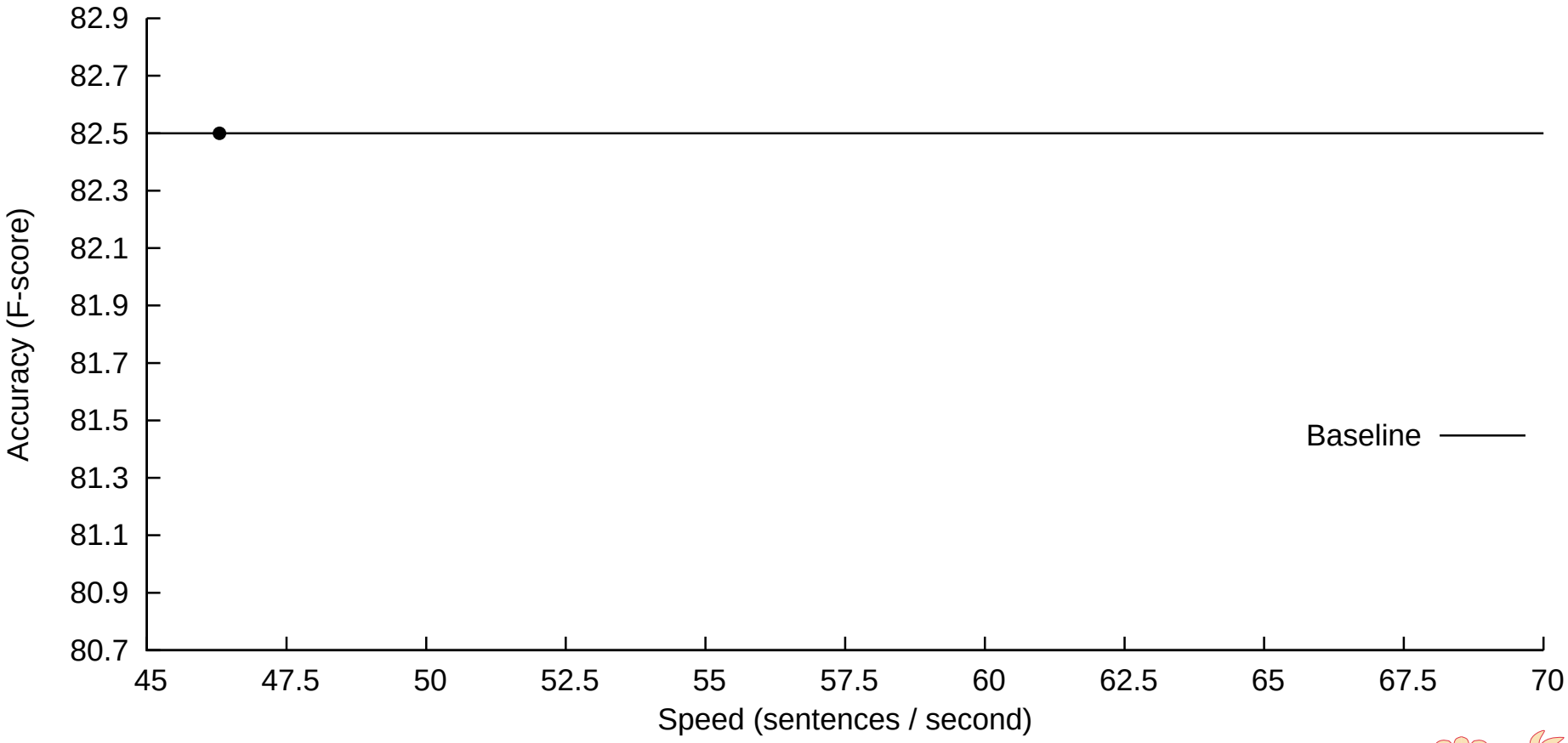


Figure: Evaluation on Wikipedia

Domain Adaptation

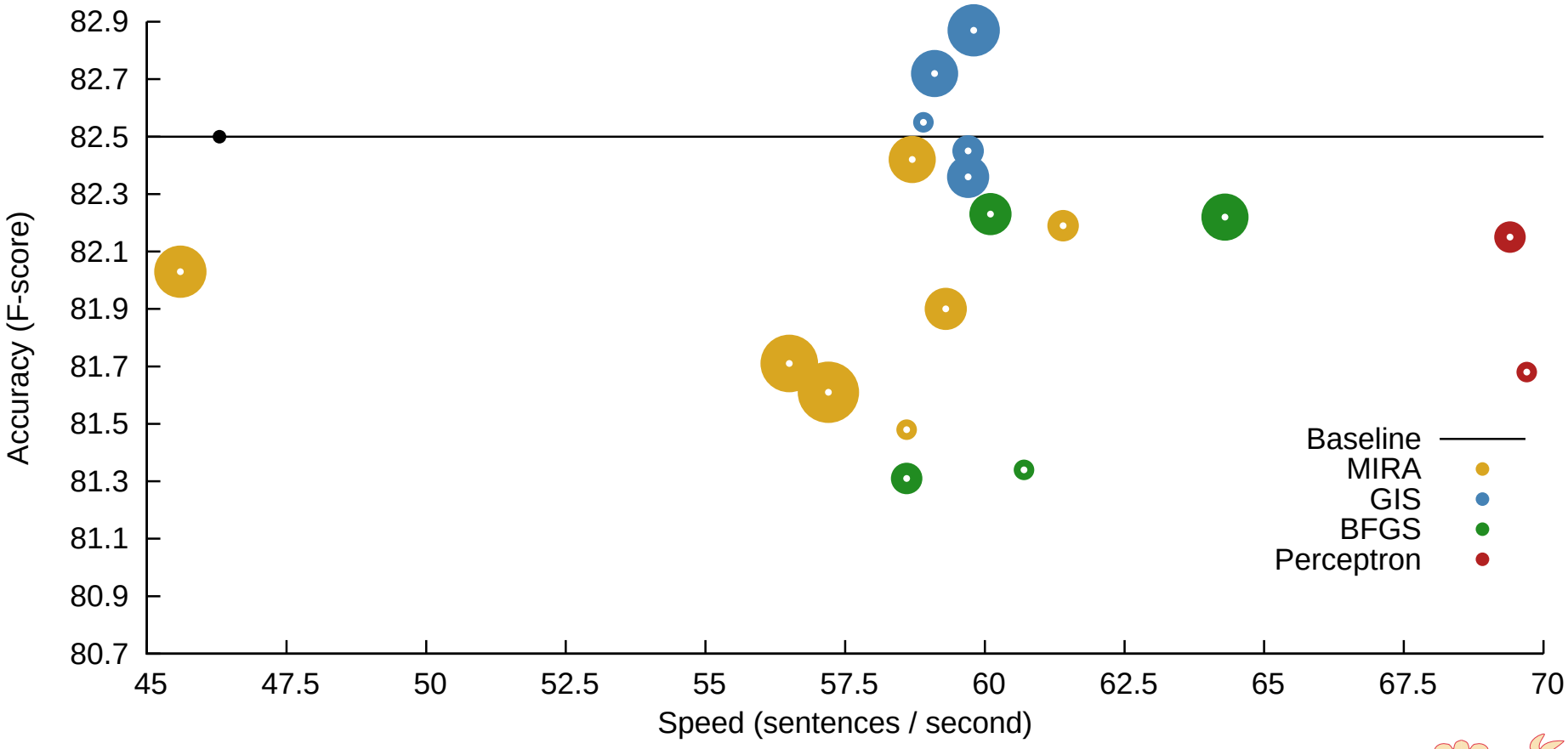


Figure: Evaluation on Wikipedia

Domain Adaptation

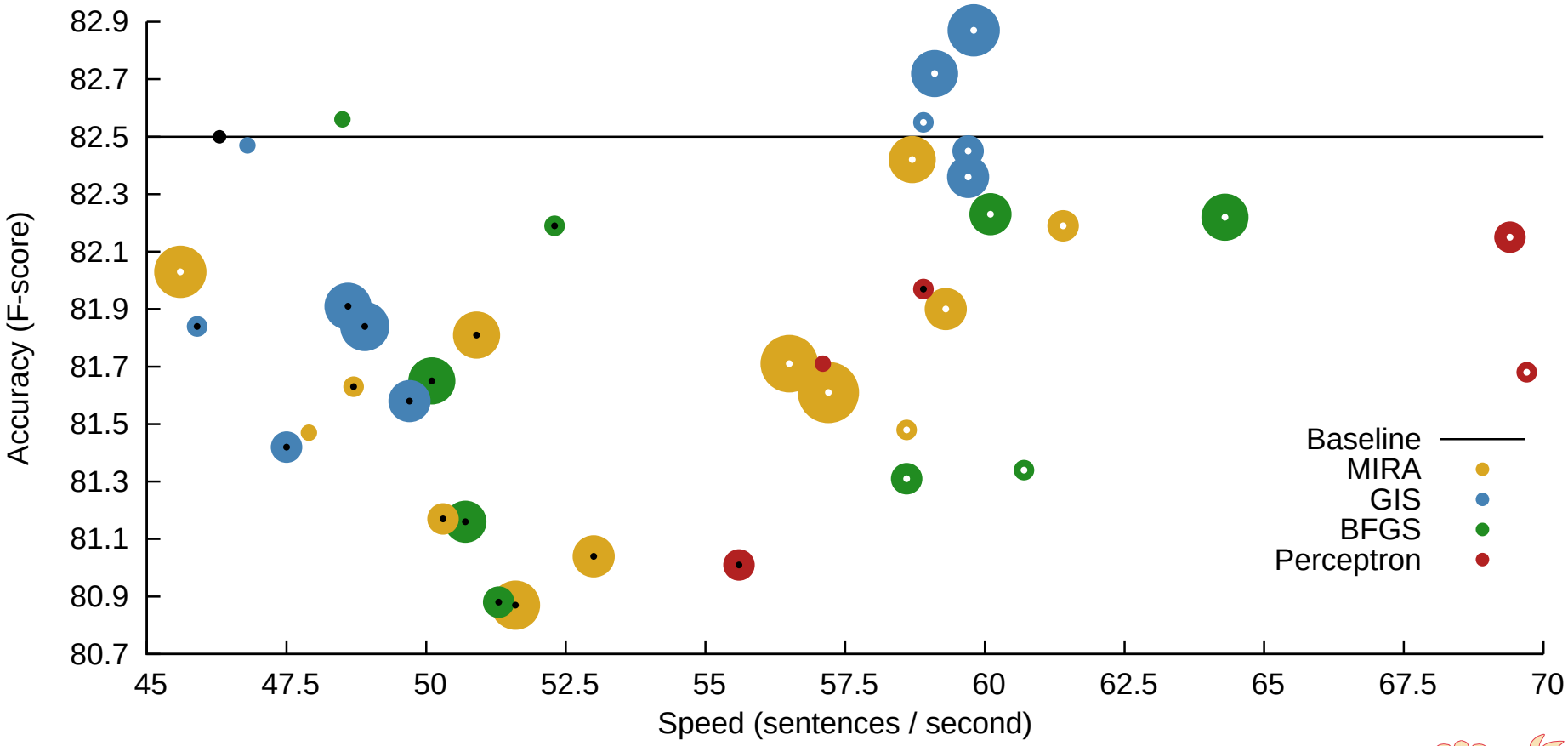
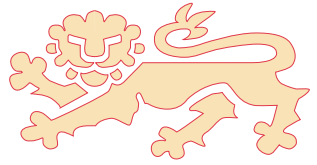


Figure: Evaluation on Wikipedia

Contributions

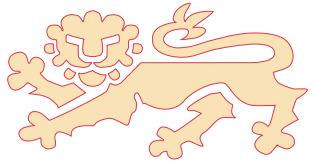
Metric	Baseline	Final Models
Wall Street Journal		
Accuracy (%)	83.41	83.99
Speed (sents / sec)	48.5	90.2
Wikipedia		
Accuracy (%)	82.5	83.3
Speed (sents / sec)	46.3	60.5

This will lead directly to benefits in NLP systems



Acknowledgements

- Johns Hopkins University, CLSP Summer Workshop
  in particular, Jessika Rosener, collaborator on parallelisation
- Early results to appear as:
  Jonathan K. Kummerfeld and Dr James R. Curran
  Faster Parsing and Supertagging Model Estimation, ALTA 2009



Second Core Idea

Figure: Parsing behaviour for the 1791^{st} sentence in section 00.