

Large-Scale Language Learning



Research
at Google

Slav Petrov

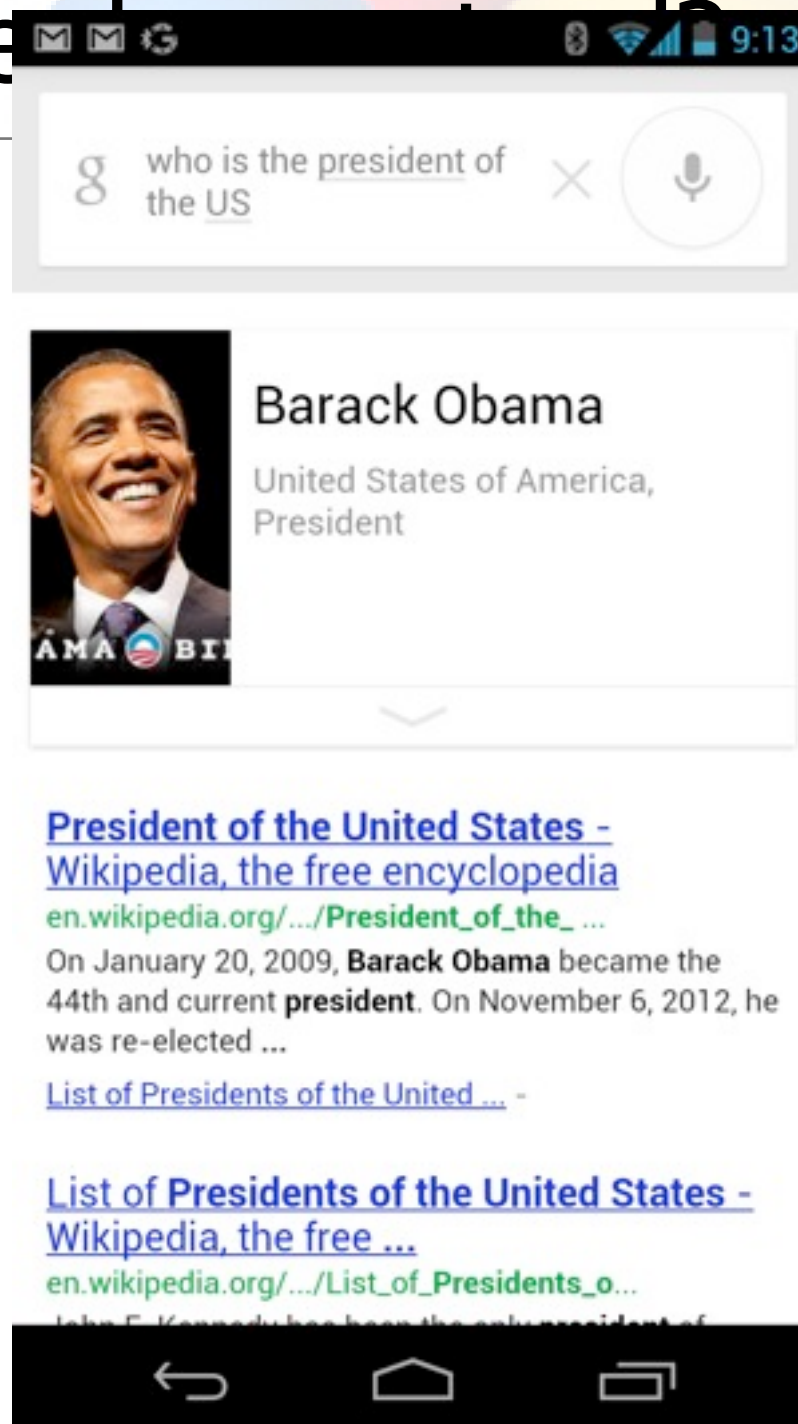
Thanks to

Ryan McDonald, Keith Hall, Alexander Rush, Dipanjan Das,
Hao Zhang, Michael Ringgaard, Terry Koo and Kuzman Ganchev
(a.k.a. the Natural Language Parsing Team at Google)

NLP: Where do we stand?



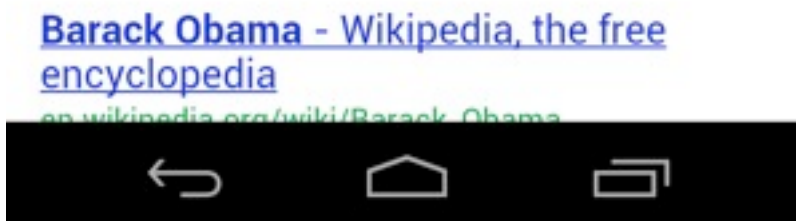
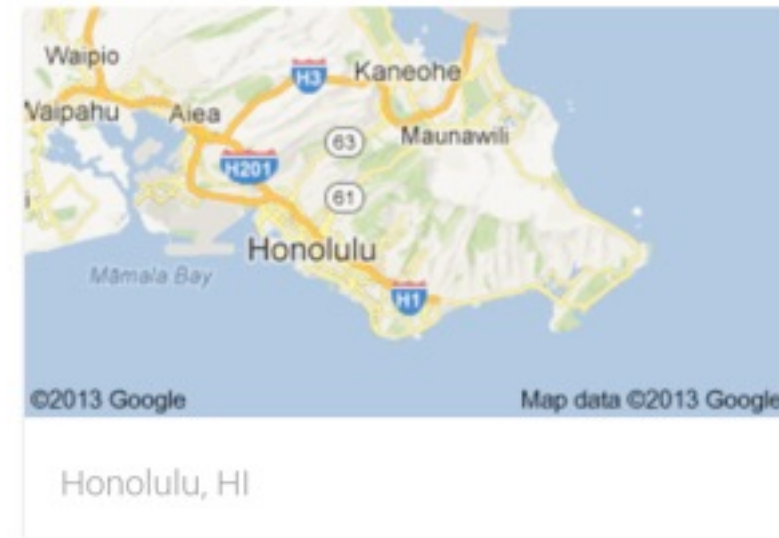
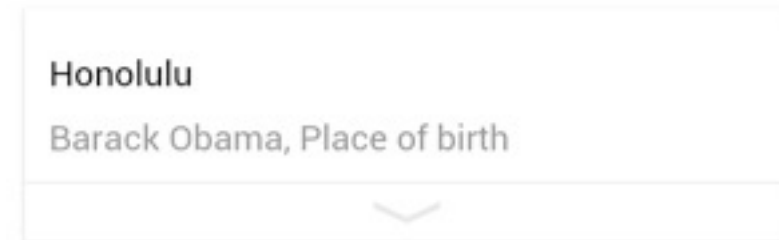
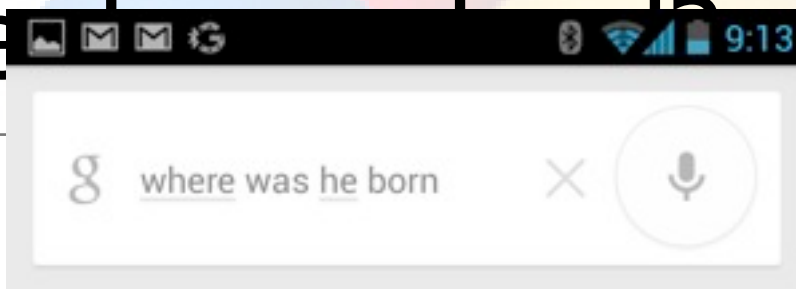
NLP: Where



NLP: Where do we stand?



NLP: Where



NLP: Where do we stand?



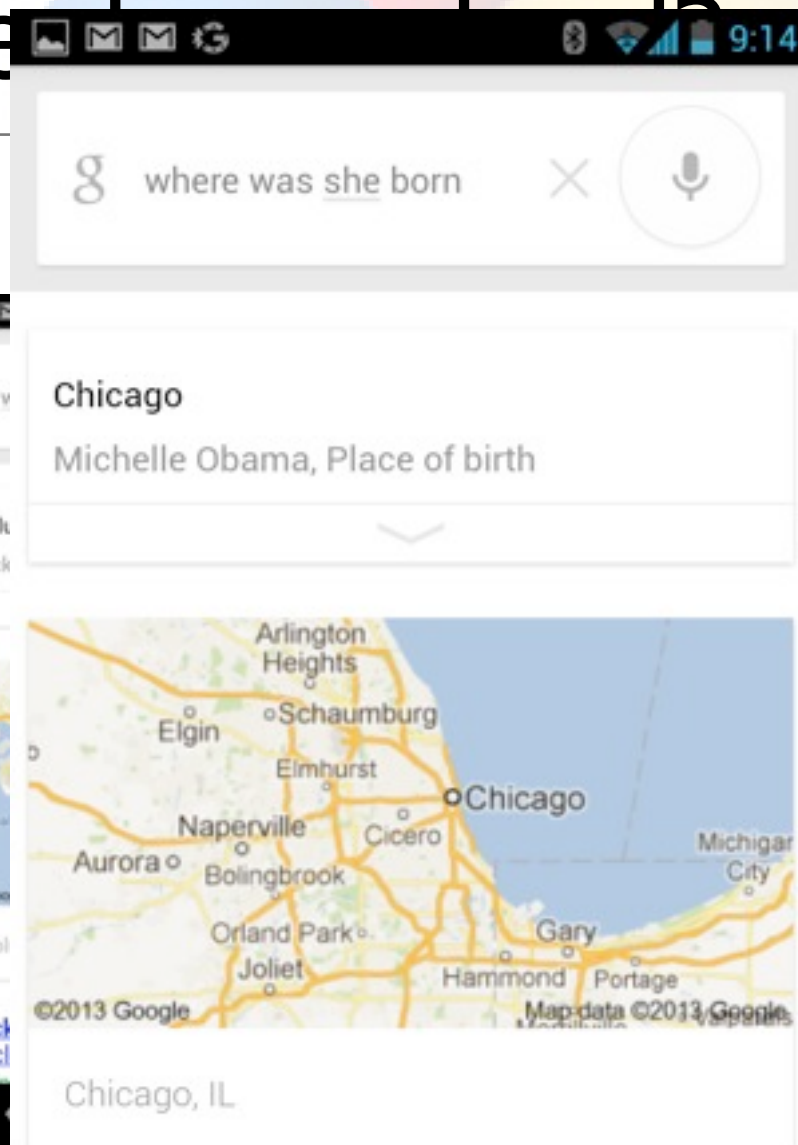
NLP: Where



NLP: Where do we stand?



NLP: Where



[Michelle Obama - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/wiki/Michelle_Obama](#)

NLP: Where do we stand?



who is the president of the US



Barack Obama
United States of America, President

[President of the United States - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/.../President_of_the_...
On January 20, 2009, **Barack Obama** became the 44th and current **president**. On November 6, 2012, he was re-elected ...
[List of Presidents of the United States - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/.../List_of_Presidents_o...

where was he born

Honolulu
Barack Obama, Place of birth



Honolulu, HI

[Barack Obama - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Barack_Obama

who is his wife



Michelle Obama (m. 1992)
Barack Obama, Spouse

[Michelle Obama - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Michelle_Obama
Michelle LaVaughn Robinson **Obama** (born January 17, 1964) is the **wife** of the 44th and current President of the United ...
[Craig Robinson - Valerie Jarrett - Magnet school](#)

[Family of Barack Obama - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/.../Family_of_Barack_...

where was she born

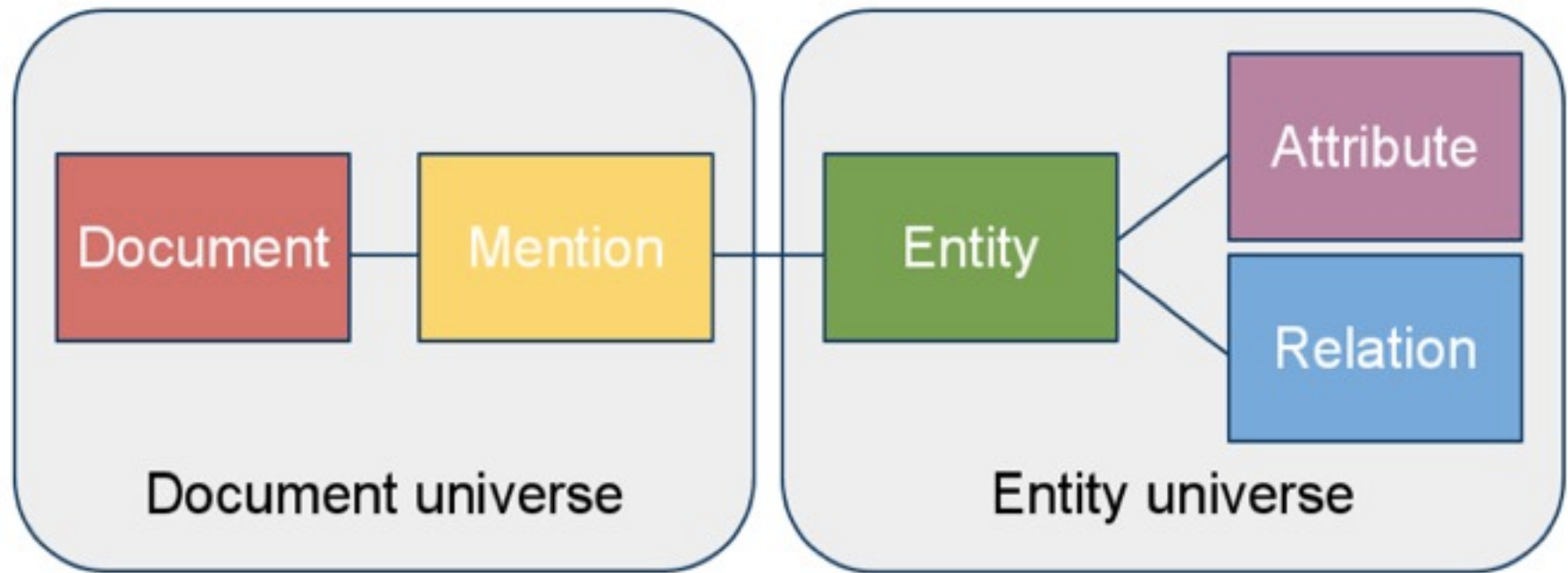
Chicago
Michelle Obama, Place of birth



Chicago, IL

[Michelle Obama - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Michelle_Obama

Building a Repository



- Find *all Mentions* of Entities in **Documents**
- Resolve Mentions to the *Identities* of the **Entities**
 - Extract typed *Relations* between Entities
 - Extract typed *Attribute* values for Entities

Semantic Annotation of Documents

President Barack Obama has been re-elected to a second term, defeating Republican challenger Mitt Romney. America's first black president secured more than the 270 votes in the electoral college needed to win. In his victory speech before supporters in Chicago, Mr Obama said he would talk to Mr Romney about "where we can work together to move this country forward". Mr Obama prevailed despite lingering dissatisfaction with the economy and a hard-fought challenge by Mr Romney.

His Democrats also retained their majority in the Senate, which they have held since 2007. The Republicans kept control of the House of Representatives, which analysts say will likely result in more of the gridlock that characterized Mr Obama's first term, with the House and the president at loggerheads on most legislation.

In his address, the president challenged his opponents, asking them to work with him. With only Florida's 29 electoral votes still undecided, Mr Obama won 303 electoral votes to Mr Romney's 206. The popular vote, which is symbolically and politically important but not decisive in the race, remains very close.

Mr Obama congratulated Mr Romney and Republican vice-presidential candidate Paul Ryan on their hard-fought campaign. "We have picked ourselves up, we have fought our way back and we know in our hearts that for the United States of America the best is yet to come," he said. Mr Obama said he was returning to the White House "more determined, and more inspired than ever about the work there is to do, and the future that lies ahead". He pledged to work with Republican leaders in Congress to reduce the government's budget deficit, fix the tax code and reform the immigration system.

Semantic Annotation of Documents

E₁ Barack Obama		PER
Mentions	Barack Obama, Obama (6), Mr (6), President, He, His, he (3), his	
Gender	Male ♂	
Profile	Barack Obama	
Freebase MID	/m/02mjmr	
Wikipedia	Barack Obama	
E₂ term		NON
Mentions	term	
E₃ Mitt Romney		PER
Mentions	Mitt Romney, Romney (4), Mr (4), challenger	
Gender	Male ♂	
Profile	Mitt Romney	
Freebase MID	/m/0271_s	
Wikipedia	Mitt Romney	
E₄ Republican		ORG
Mentions	Republican (3)	
Profile	Republican Party (United States)	
Freebase MID	/m/07wbk	
Wikipedia	Republican Party (United States)	
E₅ president		PER
Mentions	president	
E₆ United States of America		LOC
Mentions	America, United States of America	
Profile	United States	
Freebase MID	/m/09c7w0	
Wikipedia	United States	
E₇ votes		NON

s <President>₁ [Barack Obama]₁ has been re-elected to a second <term>₂, defeating [Republican]₄ <challenger>₃ [Mitt Romney]₃. s [America]₆'s first black <president>₅ secured more than the <270>₄ <votes>₇ in the <electoral college>₈ needed to win. s In [his]₁₁ <victory speech>₉ before <supporters>₁₀ in [Chicago]₁₁, <Mr>₁ [Obama]₁ said <he>₁ would talk to <Mr>₃ [Romney]₃ about "<where>₁₁ {we}<10 can work together to move this <country>₁₂ forward". s <Mr>₁ [Obama]₁ prevailed despite lingering <dissatisfaction>₁₃ with the <economy>₁₄ and a hard-fought <challenge>₁₅ by <Mr>₃ [Romney]₃.

s [His]₁ [Democrats]₁₆ also retained <their>₁₆ <majority>₁₇ in the [Senate]₁₈, <which>₁₈ {they}<16 have held since <2007>₄. s The [Republicans]₁₉ kept <control>₂₀ of the [House of Representatives]₂₁, <which>₂₁ <analysts>₂₂ say will likely result in more of the <gridlock>₂₃ <that>₂₃ characterized <Mr>₁ [Obama]₁'s first <term>₂₄, with the [House]₂₁ and the <president>₂₅ at <loggerheads>₂₆ on most <legislation>₂₇.

s In [his]₂₈ <address>₂₉, the <president>₂₈ challenged <his>₂₈ <opponents>₃₀, asking {them}<30 to work with {him}<28. s With only [Florida]₃₂'s <29>₄ <electoral votes>₃₁ still undecided, <Mr>₁ [Obama]₁ won <303>₄ <electoral votes>₃₃ to <Mr>₃ [Romney]₃'s <206>₄. s The popular <vote>₃₄, <which>₃₄ is symbolically and politically important but not decisive in the <race>₃₅, remains very close.

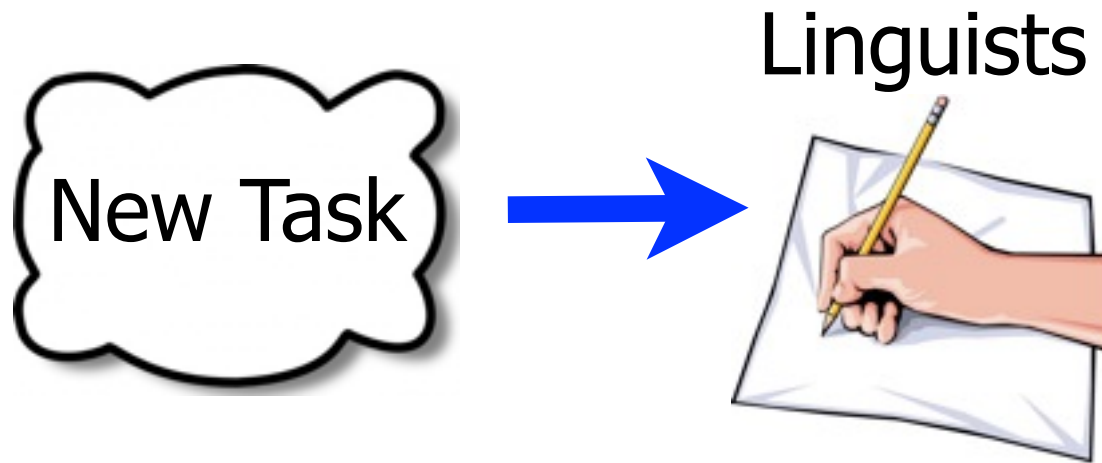
s <Mr>₁ [Obama]₁ congratulated <Mr>₃ [Romney]₃ and [Republican]₄ <vice-presidential candidate>₃₆ [Paul Ryan]₃₆ on {their}<33 hard-fought <campaign>₃₇. s "[We]<33 have picked {ourselves}<33 up, {we}<33 have fought {our}<33 <way>₃₈ back and {we}<33 know in {our}<33 <hearts>₃₉ that for the [United States of America]₆ the best is yet to come," <he>₁ said. s <Mr>₁ [Obama]₁ said <he>₁ was returning to the [White House]₄₀ "more determined, and more inspired than ever about the <work>₄₂ {there}<43 is to do, and the <future>₄₁ <that>₄₁ lies ahead". s {He}<1 pledged to work with [Republican]₄ <leaders>₄₄ in [Congress]₄₅ to reduce the

Traditional Supervised NLP

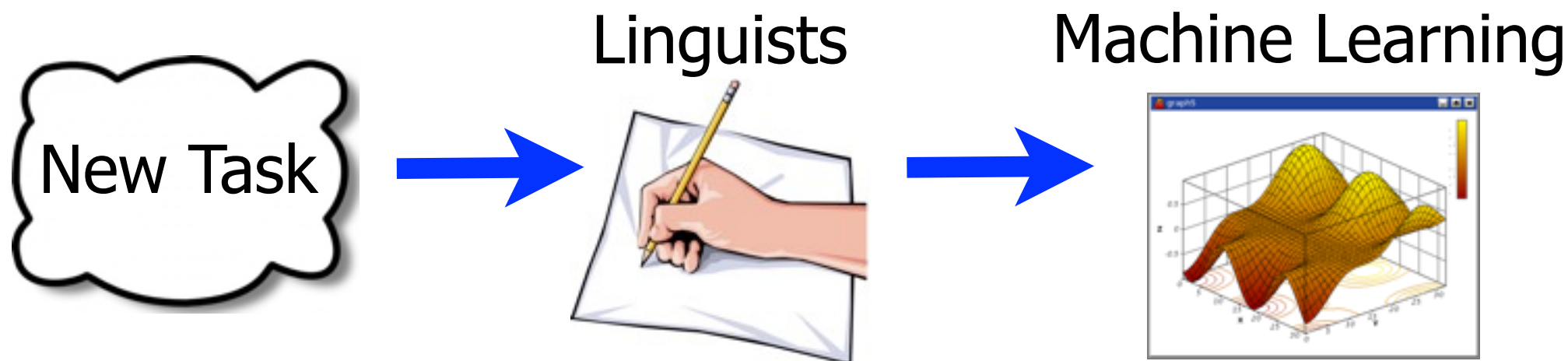


New Task

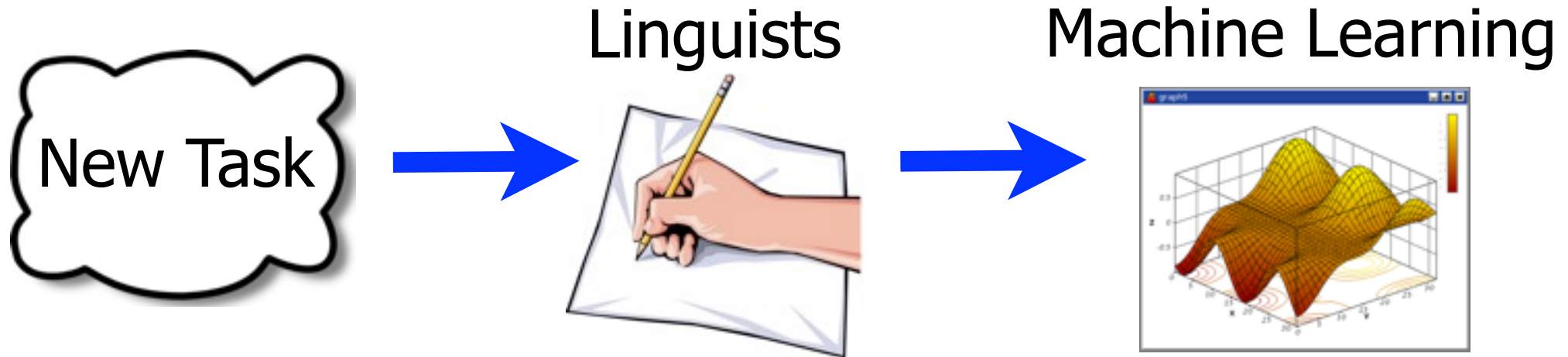
Traditional Supervised NLP



Traditional Supervised NLP

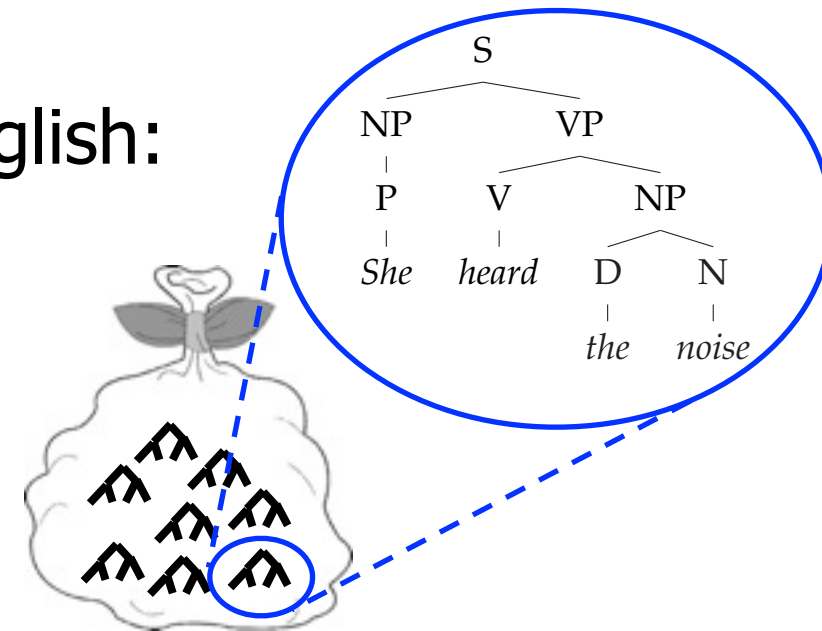


Traditional Supervised NLP



- Example: Syntactic analysis for English:

- Not solved, but accuracies are high
- 97% **(or 90%?)** for parts-of-speech
- 93% **(or 83%?)** for parse trees



Treebank:
2 million words

A Note on Parallel Computing



A Note on Parallel Computing



- Most NLP models these days operate at sentence level
 - Trivially parallelizable
 - Focus on efficient single-core models

A Note on Parallel Computing



- Most NLP models these days operate at sentence level
 - Trivially parallelizable
 - Focus on efficient single-core models
- Examples:
 - Web-Scale Information Extraction
 - Machine Translation

A Note on Parallel Computing

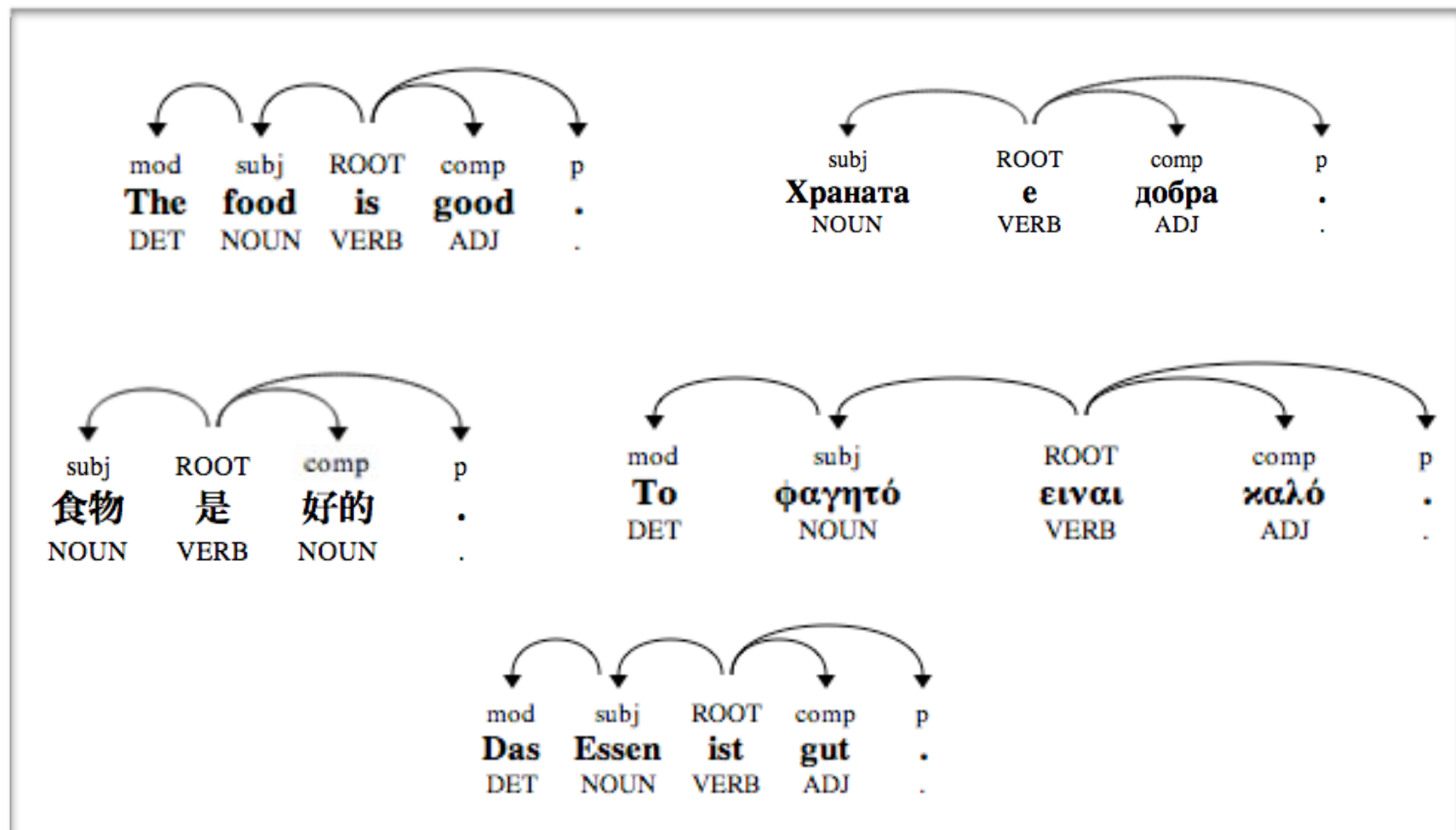


- Most NLP models these days operate at sentence level
 - Trivially parallelizable
 - Focus on efficient single-core models
- Examples:
 - Web-Scale Information Extraction
 - Machine Translation
- Distributed systems might be needed for storage:
 - Distributed Language Models, Phrase Tables, ...
in Machine Translation, Speech Recognition, ...

Universal Tagging/Parsing



- Goal: high accuracy parsing in all languages with a single universal representation of syntax



Syntactic Processing for the Web



Syntactic Processing for the Web



Syntactic Processing for the Web



- Fast and accurate supervised parsing
 - Many labeled resources are English only
 - Nonetheless: Use them! [Rush & Petrov '12]

Syntactic Processing for the Web



- Fast and accurate supervised parsing
 - Many labeled resources are English only
 - Nonetheless: Use them! [Rush & Petrov '12]
- Weakly supervised domain adaptation
 - Training data is not representative
 - Learn from weak signals [Hall et al. '11, Ganchev et al. '12]

Syntactic Processing for the Web



- Fast and accurate supervised parsing
 - Many labeled resources are English only
 - Nonetheless: Use them! [Rush & Petrov '12]
- Weakly supervised domain adaptation
 - Training data is not representative
 - Learn from weak signals [Hall et al. '11, Ganchev et al. '12]
- Multilingual tagger and parser projection
 - Will never have labeled resources for all languages
 - Use parallel data to project information [Das & Petrov '11, McDonald et al. '11]

Syntactic Processing for the Web



- Fast and accurate supervised parsing
 - Many labeled resources are English only
 - Nonetheless: Use them! [Rush & Petrov '12]
- Weakly supervised domain adaptation
 - Training data is not representative
 - Learn from weak signals [Hall et al. '11, Ganchev et al. '12]
- Multilingual tagger and parser projection
 - Will never have labeled resources for all languages
 - Use parallel data to project information [Das & Petrov '11, McDonald et al. '11]

Dependency Parsing

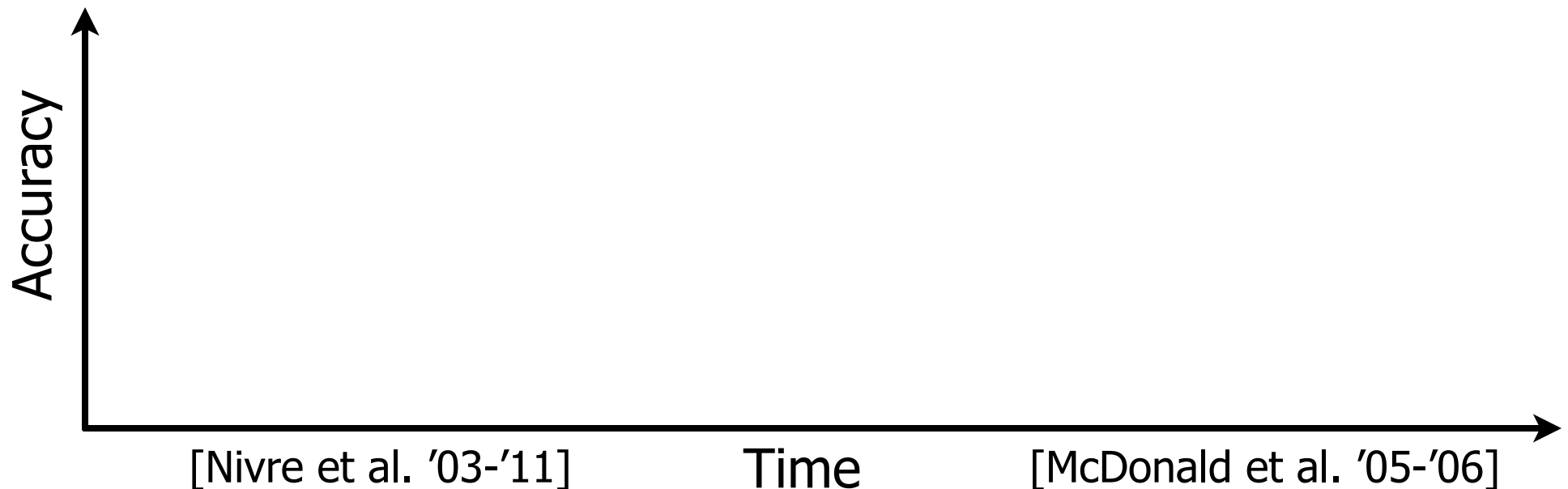


- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations

Dependency Parsing



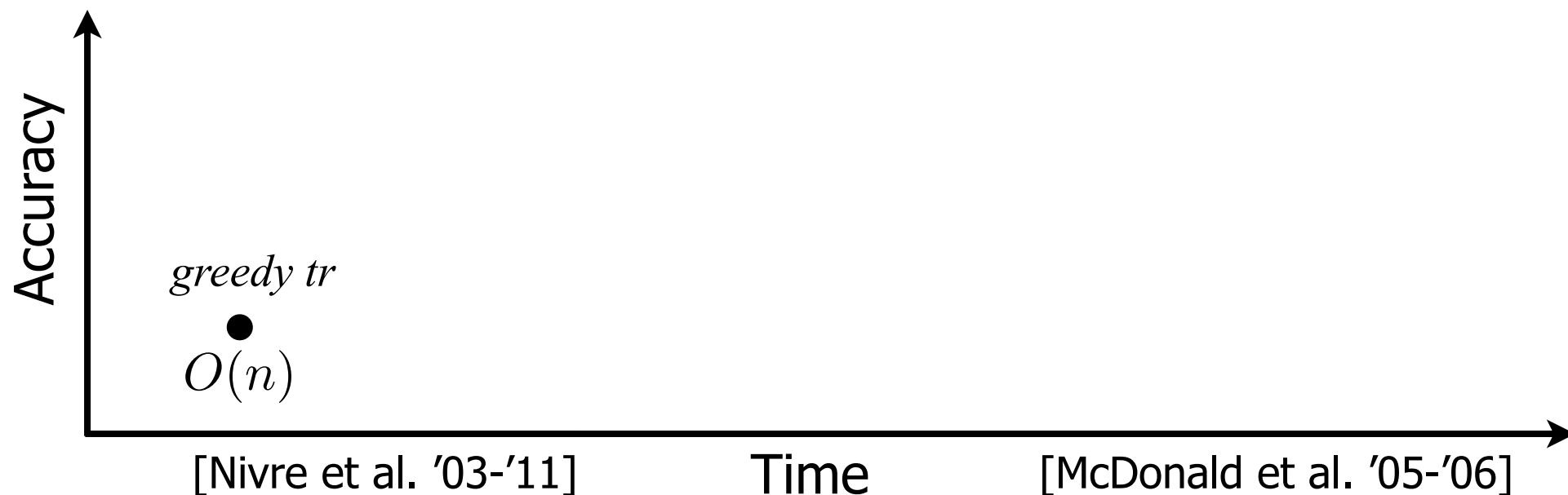
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



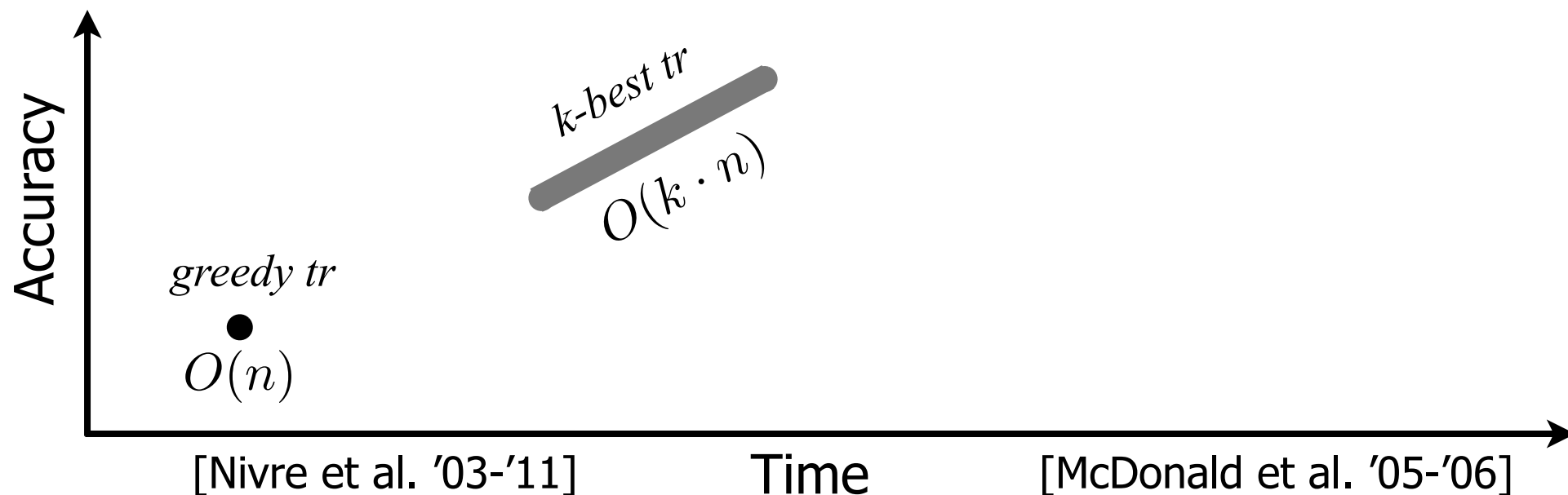
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



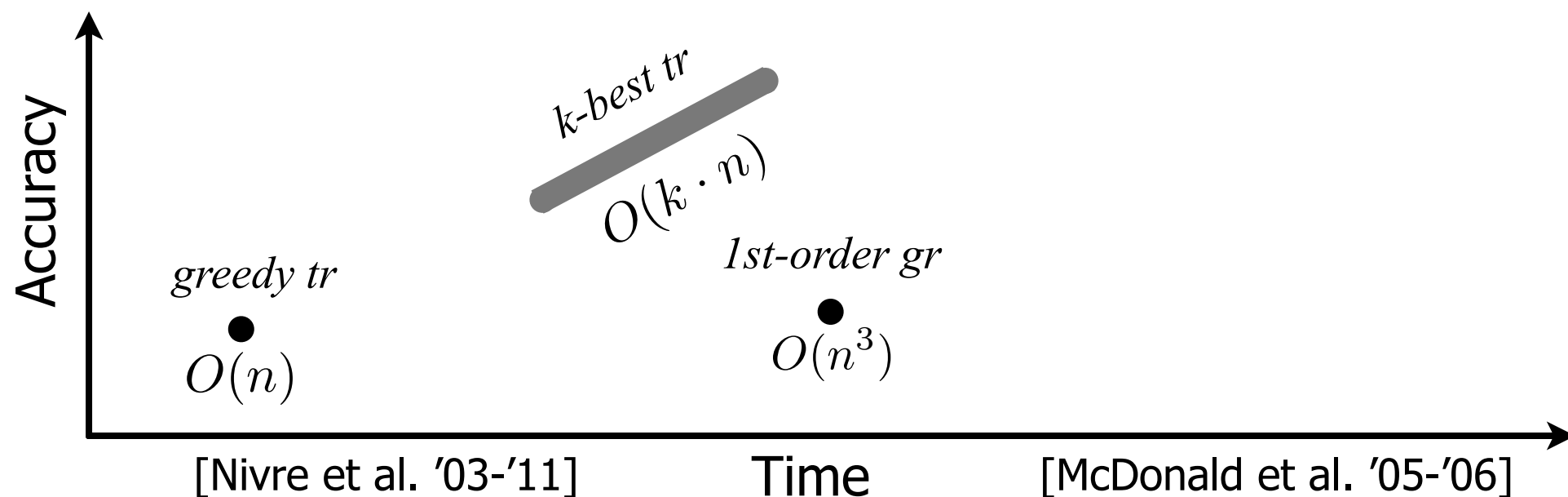
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



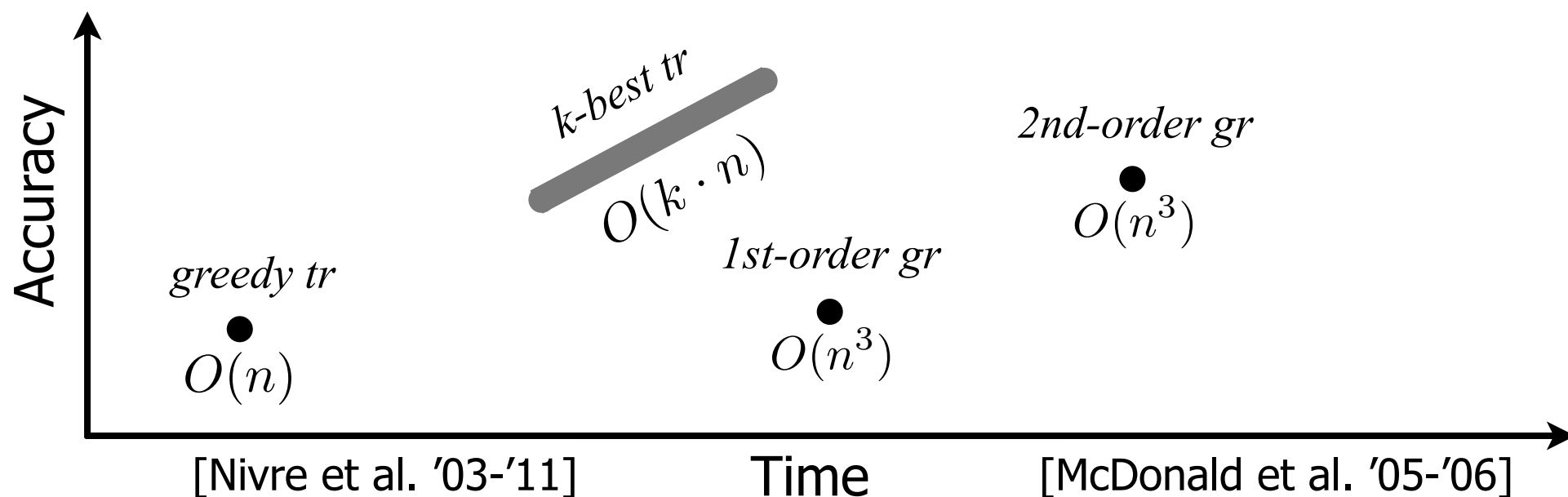
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



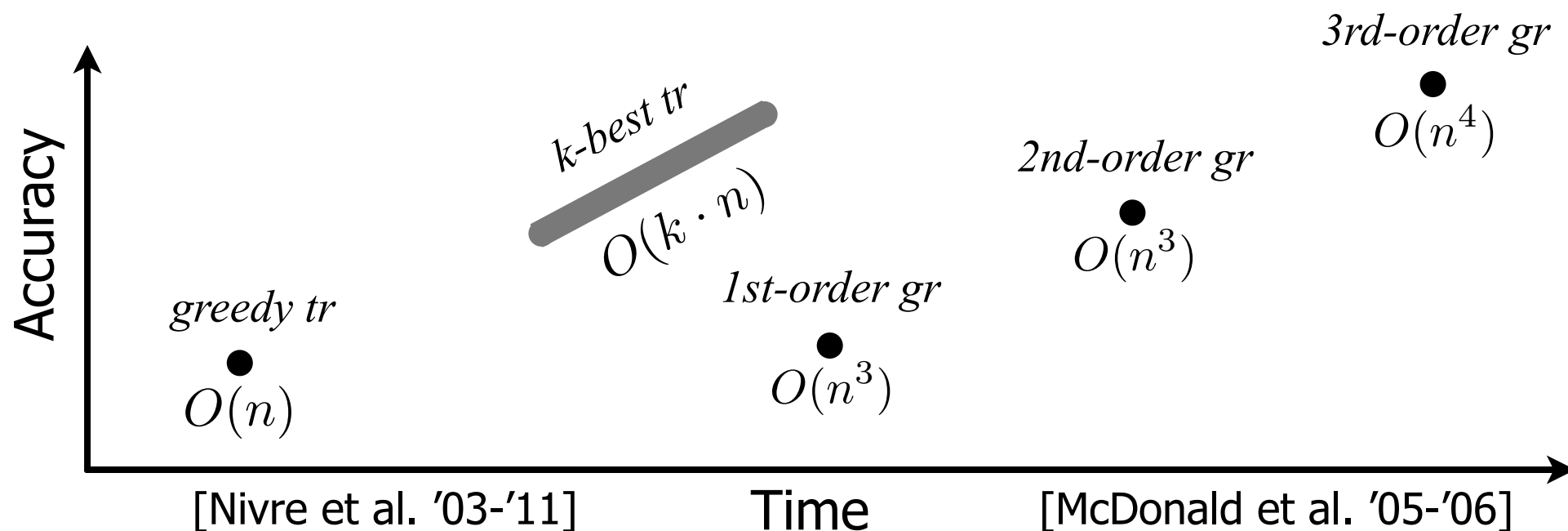
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



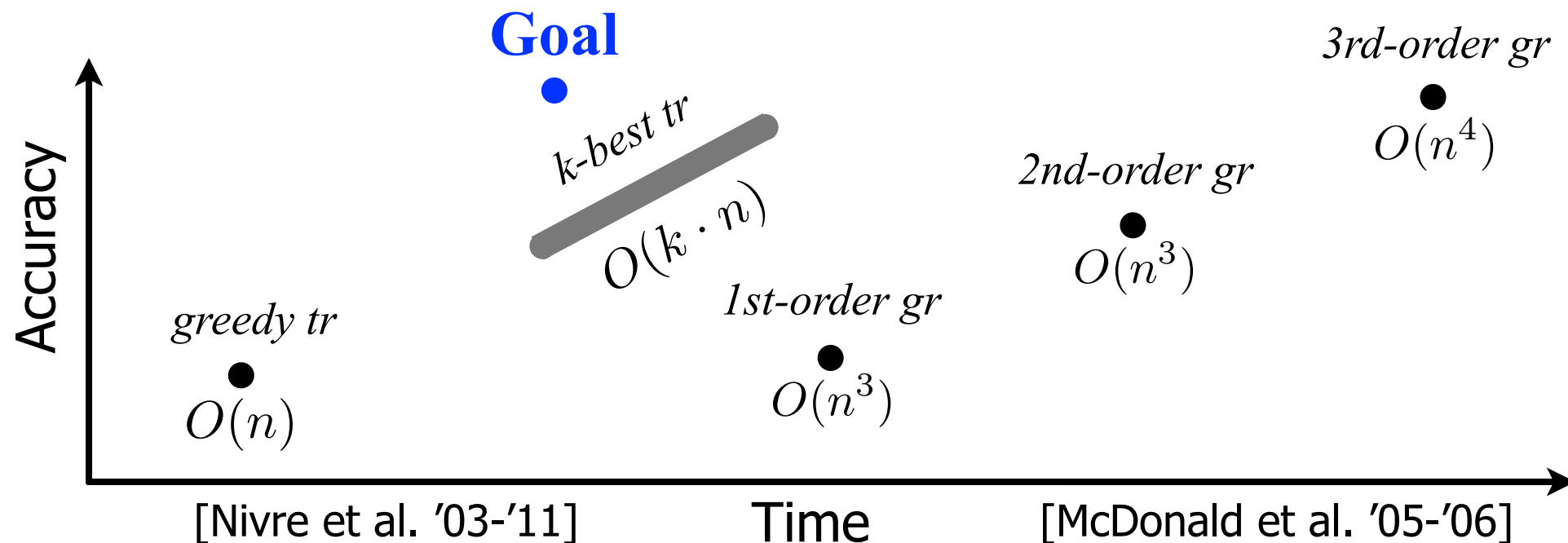
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



Dependency Parsing



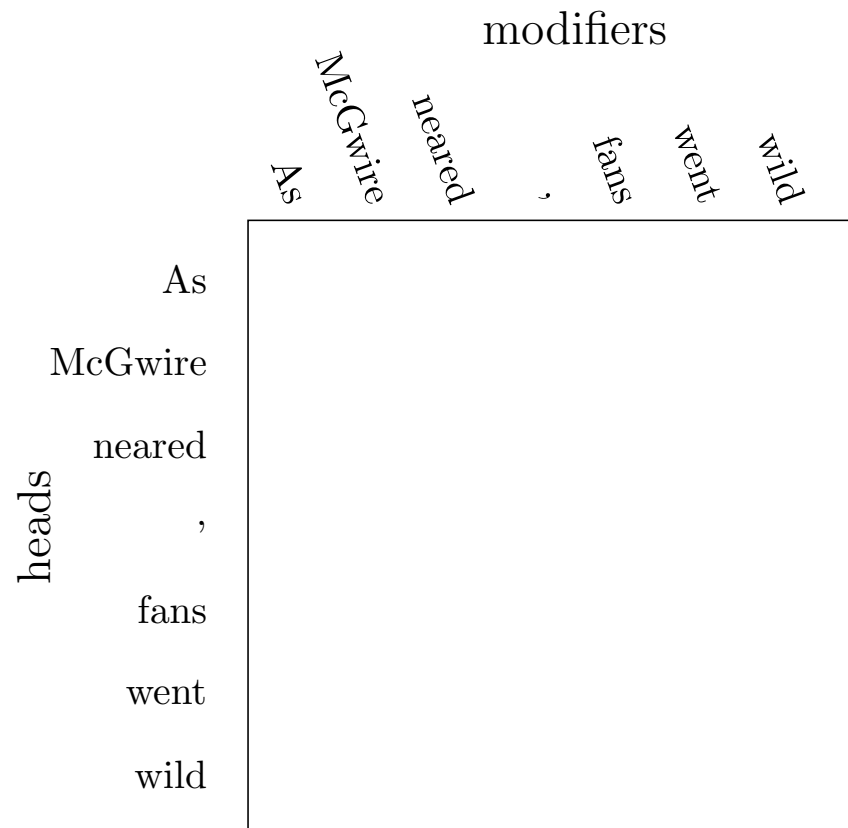
- Transition-Based (tr)
 - Fast, greedy, linear time inference algorithms
 - Trained for greedy search
 - Beam search
- Graph-Based (gr)
 - Slower, exhaustive, dynamic programming inference algorithms
 - Higher-order factorizations



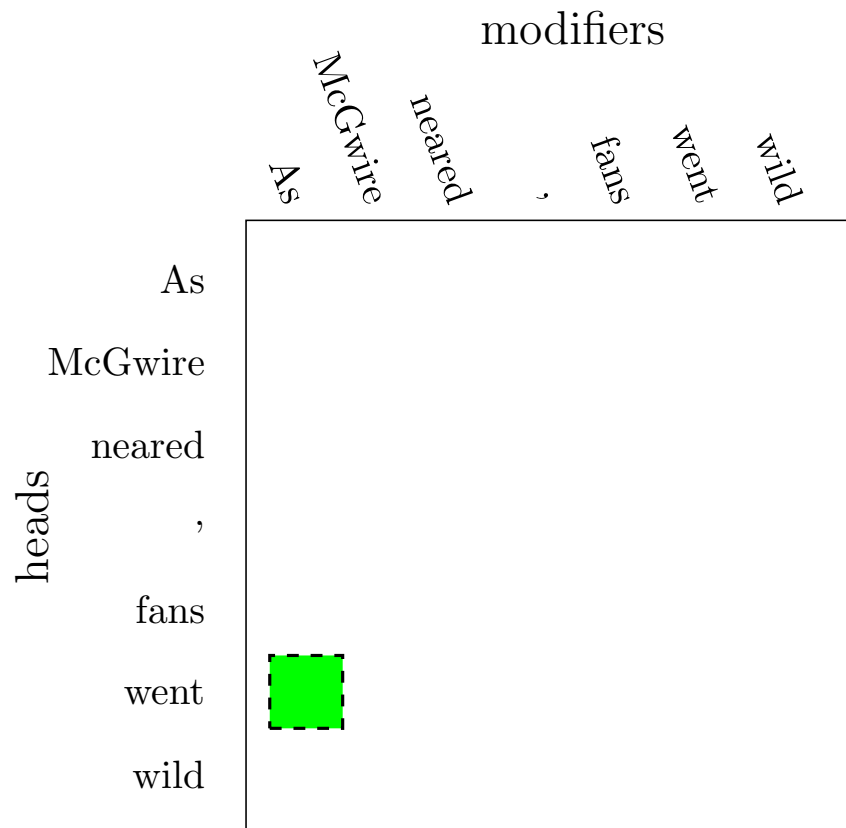
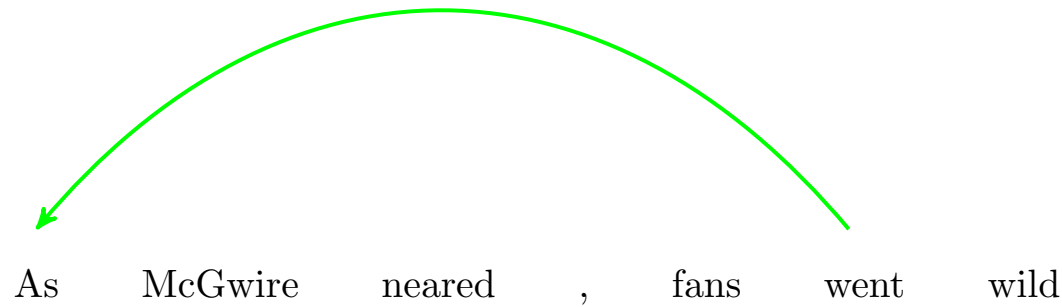
Dependency Representation



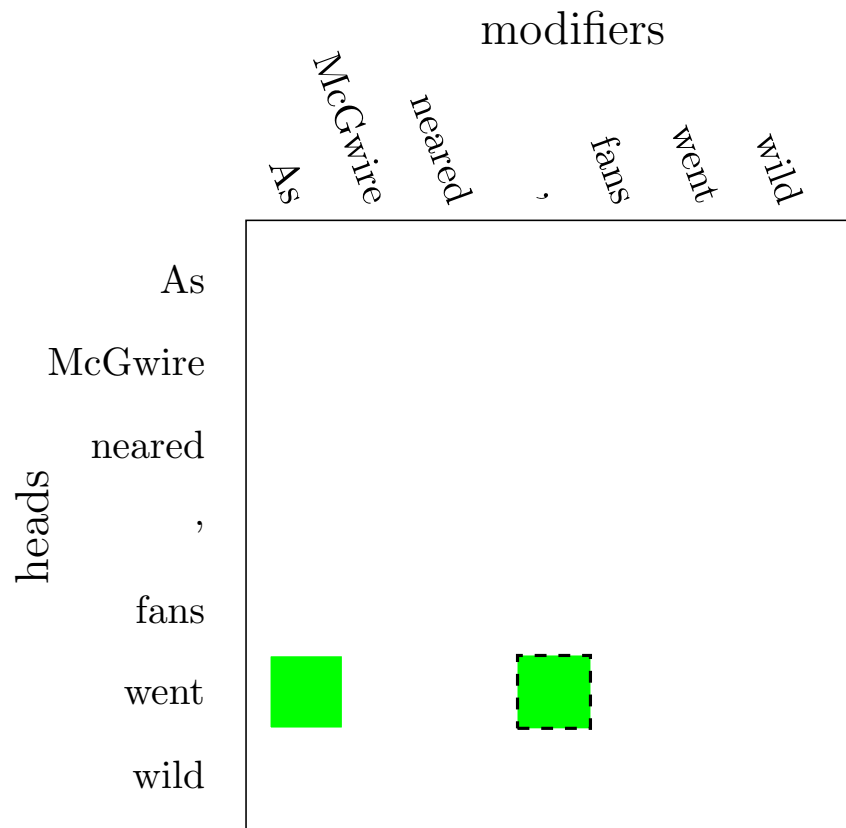
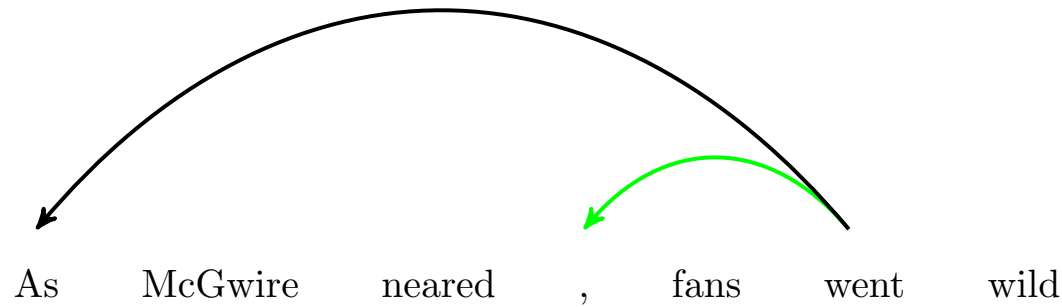
As McGwire neared , fans went wild



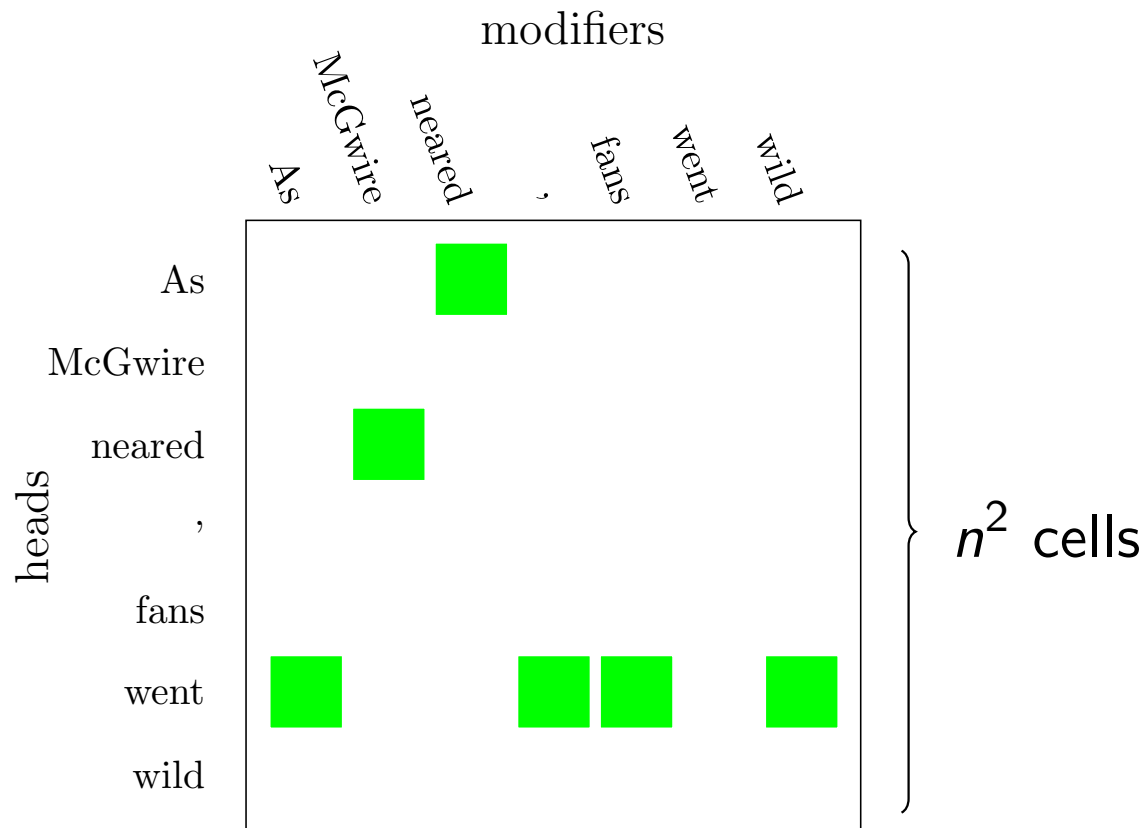
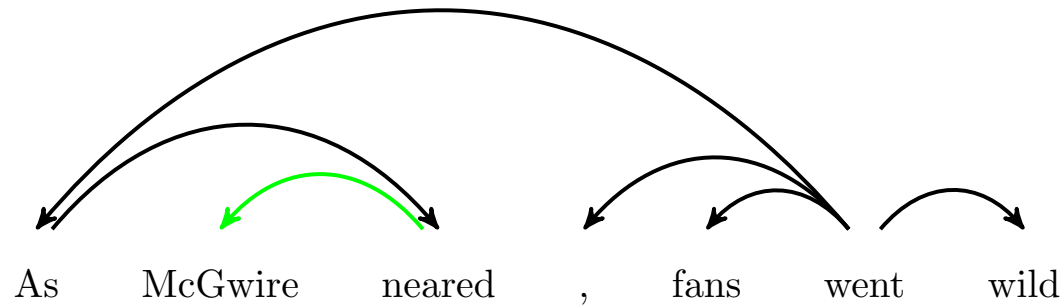
Dependency Representation



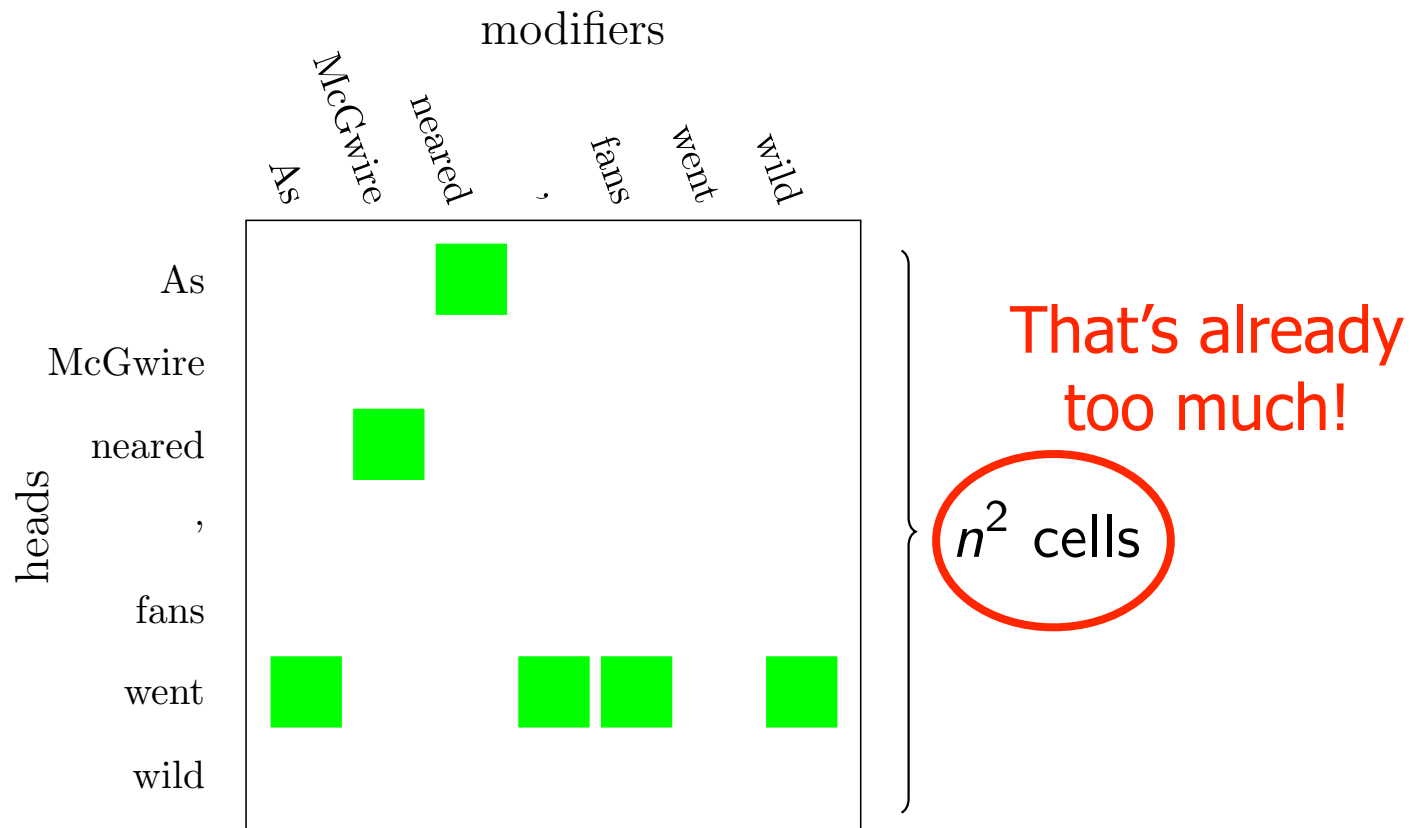
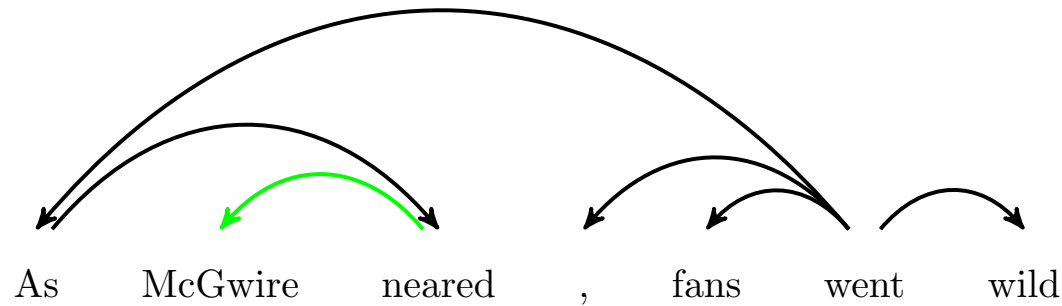
Dependency Representation



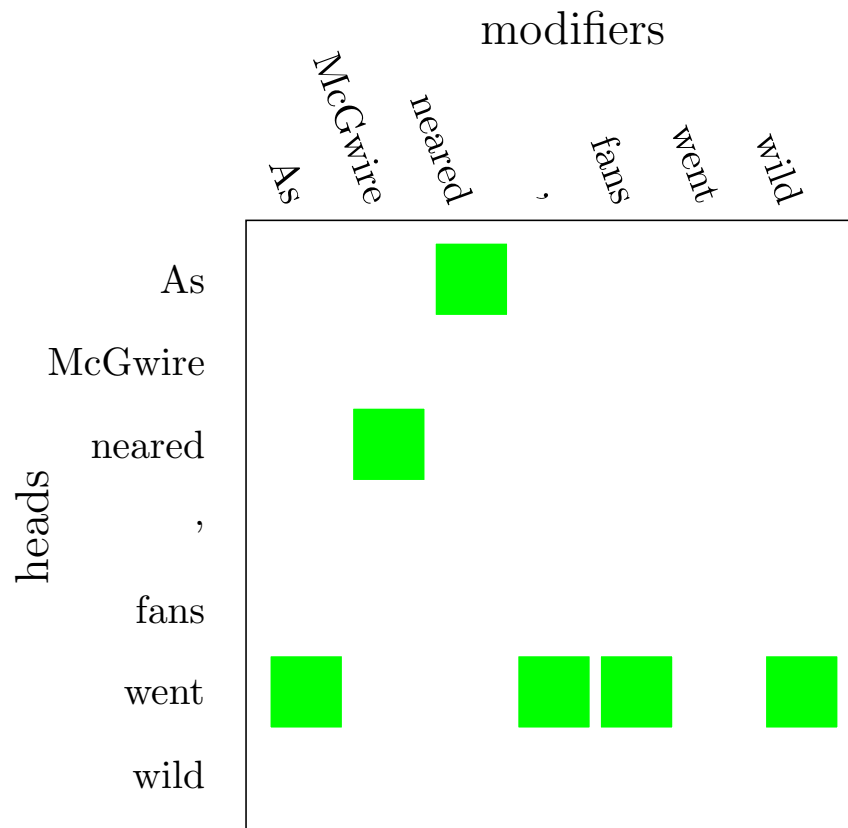
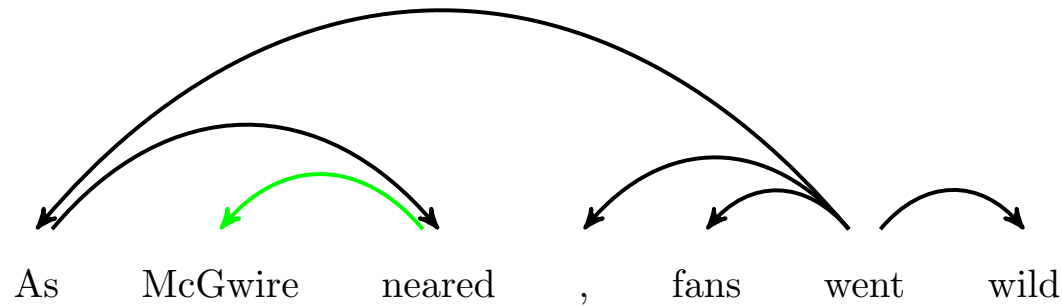
Dependency Representation



Dependency Representation



Dependency Representation



That's already too much!

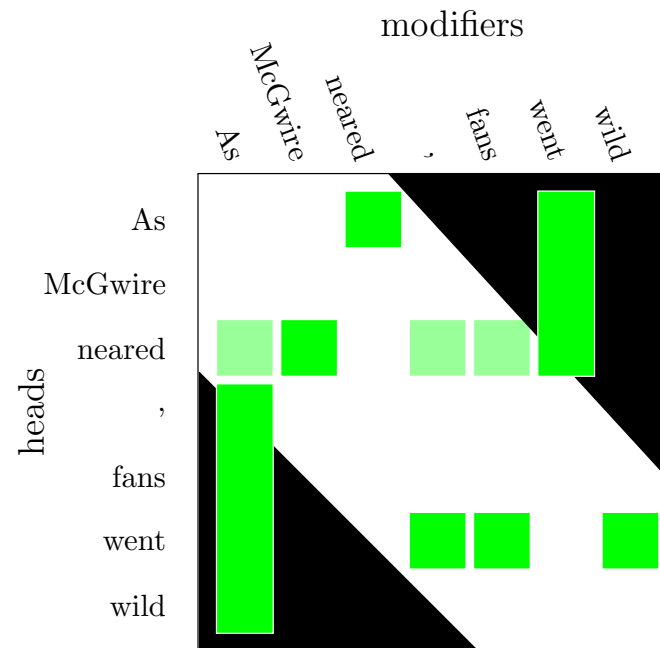
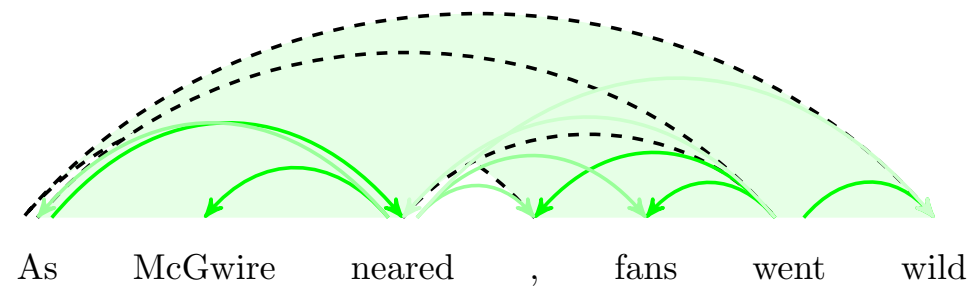
n^2 cells

Exploit problem structure!

Vine Pruning



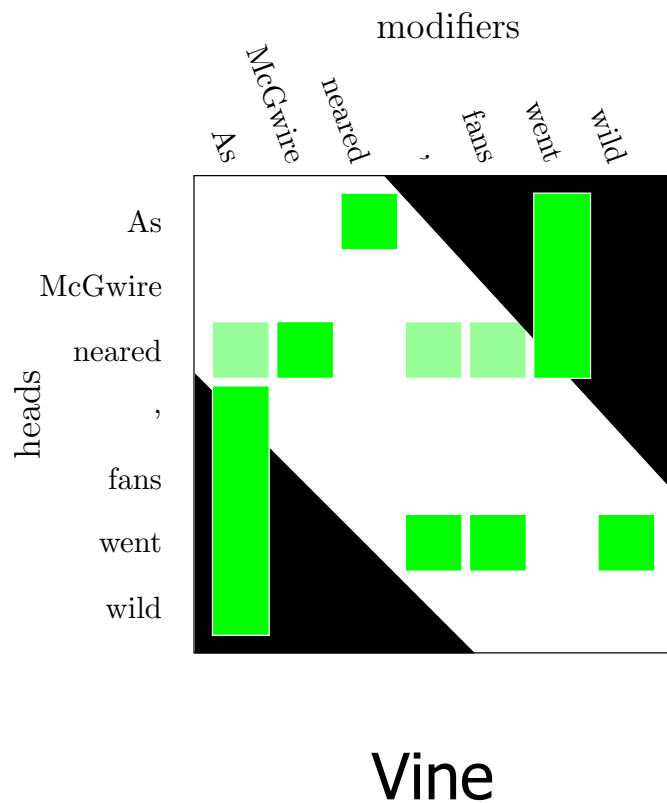
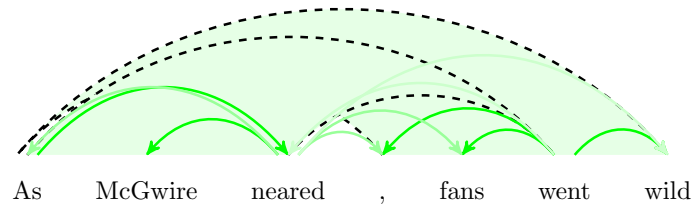
[Rush & Petrov '12]



Coarse-to-Fine Cascades



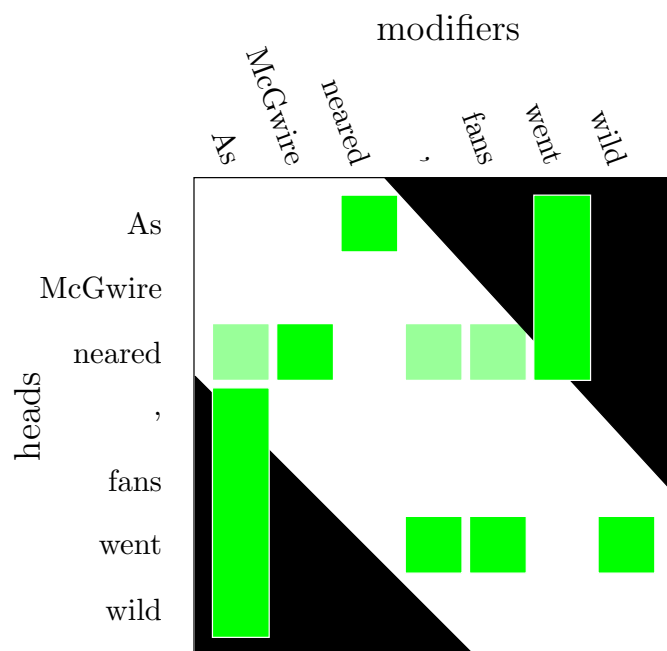
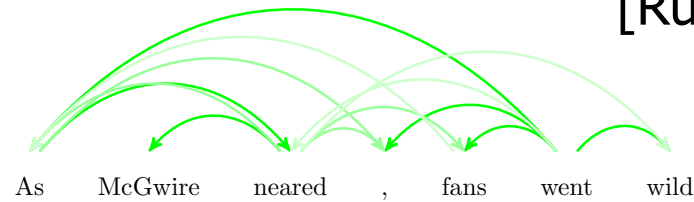
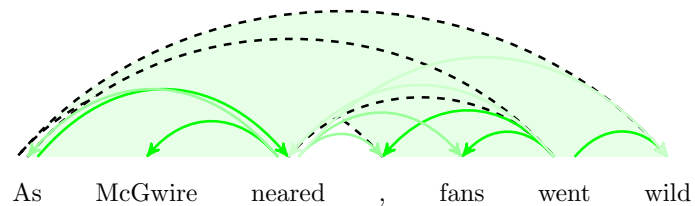
[Rush & Petrov '12]



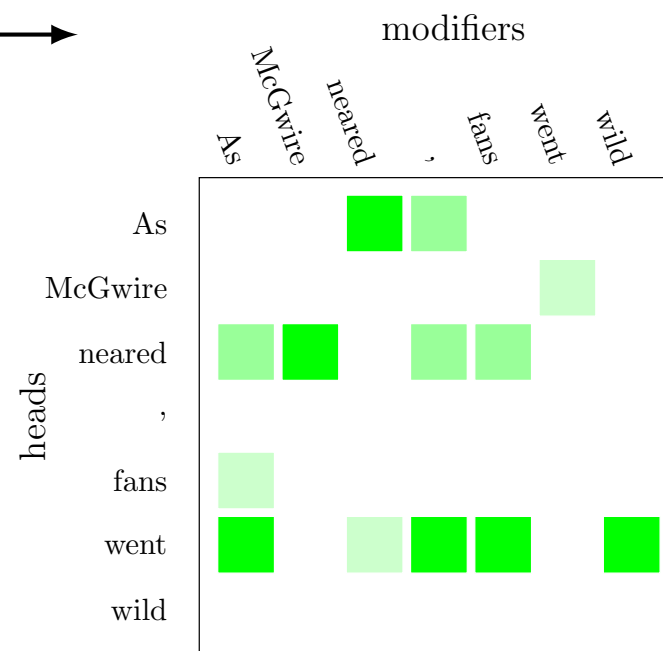
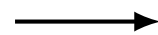
Coarse-to-Fine Cascades



[Rush & Petrov '12]



Vine

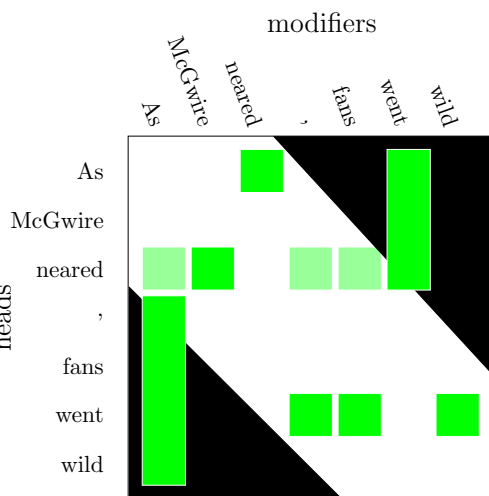
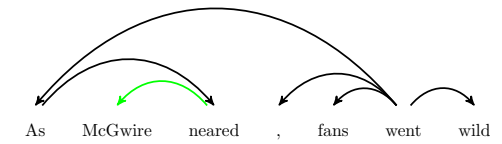
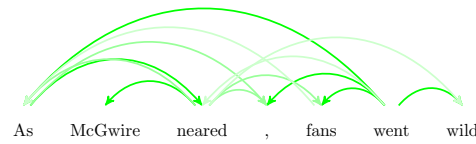
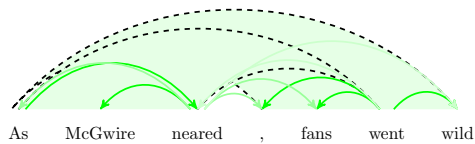


First-Order

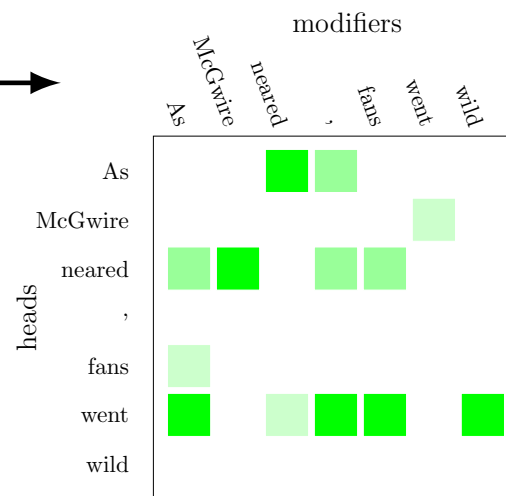
Coarse-to-Fine Cascades



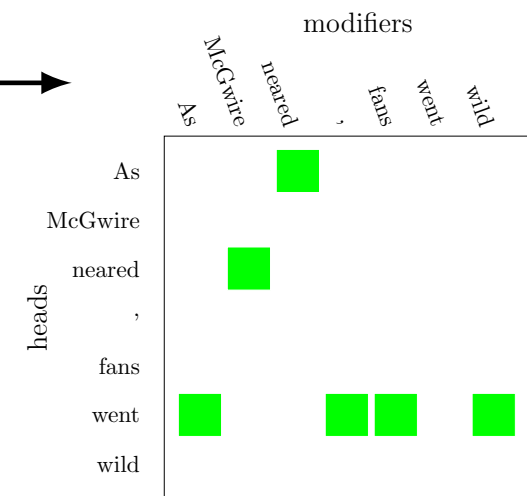
[Rush & Petrov '12]



Vine



First-Order

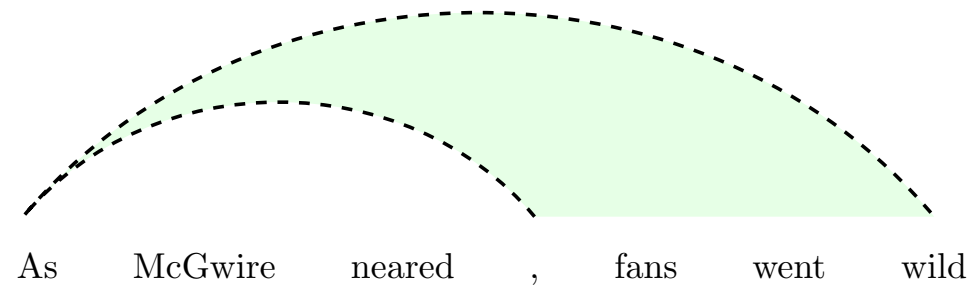


Second-Order

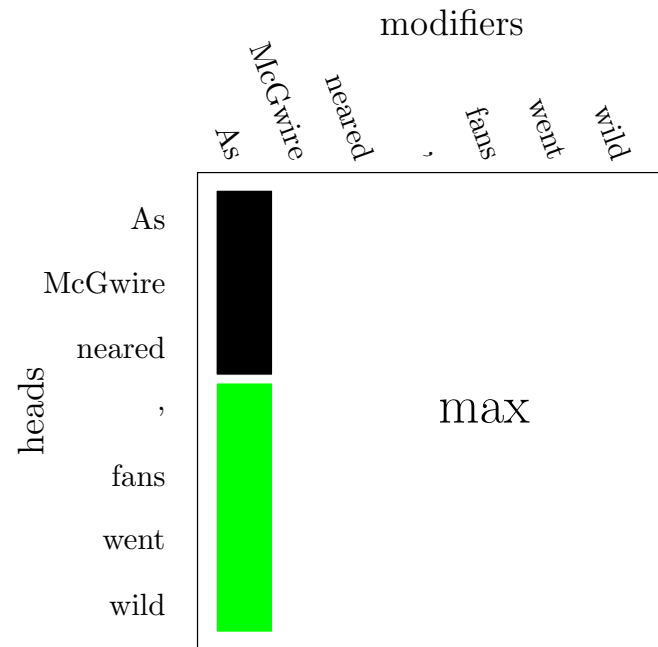
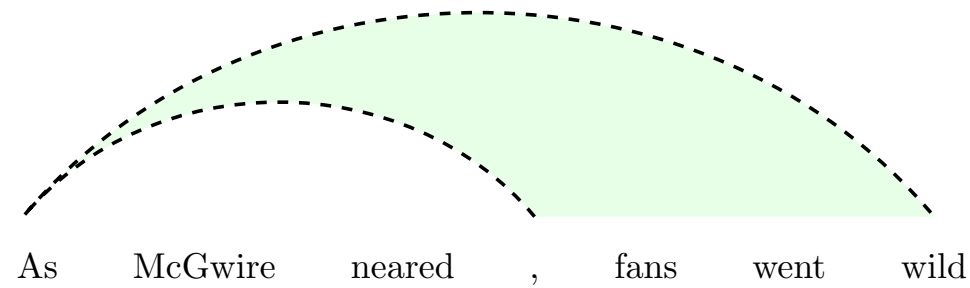

$$y^* = \arg \max_{y \in \mathcal{Y}} y \cdot w$$
$$m(a) = \arg \max_{y \in \mathcal{Y}: a \in A} y \cdot w$$

	As	McGwire	nearer	,	fans	went	wild
As							
McGwire							
nearer							
,							
fans							
went							
wild							

Max-Marginals



Max-Marginals





Structured Prediction Cascades Training

[Weiss & Taskar '10]

Structured Prediction Cascades Training

[Weiss & Taskar '10]

- Train to minimize pruning error (rather than 1-best)

Structured Prediction Cascades Training

[Weiss & Taskar '10]

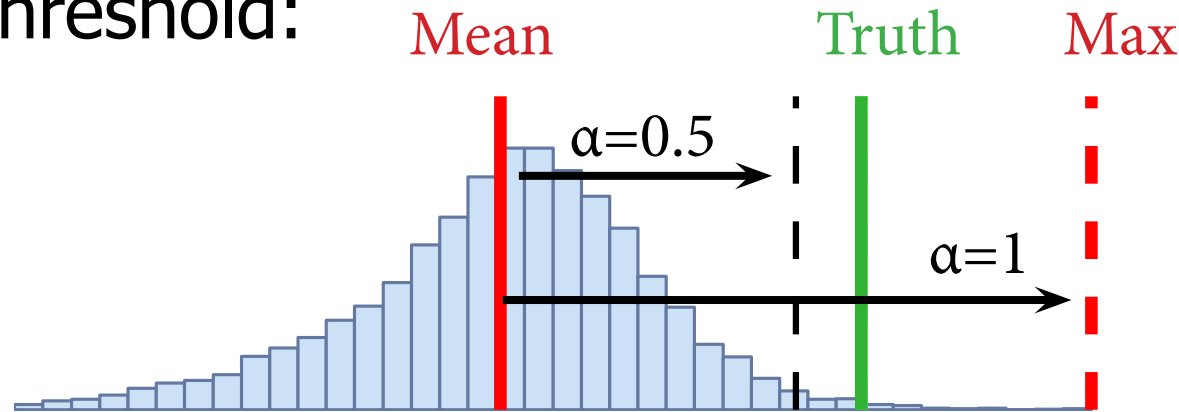
- Train to minimize pruning error (rather than 1-best)
- Pruning threshold:

$$t_{\alpha}(w) = \alpha y^* \cdot w + (1 - \alpha) \frac{1}{|A|} \sum_{a \in A} m(a) \cdot w$$

Structured Prediction Cascades Training

[Weiss & Taskar '10]

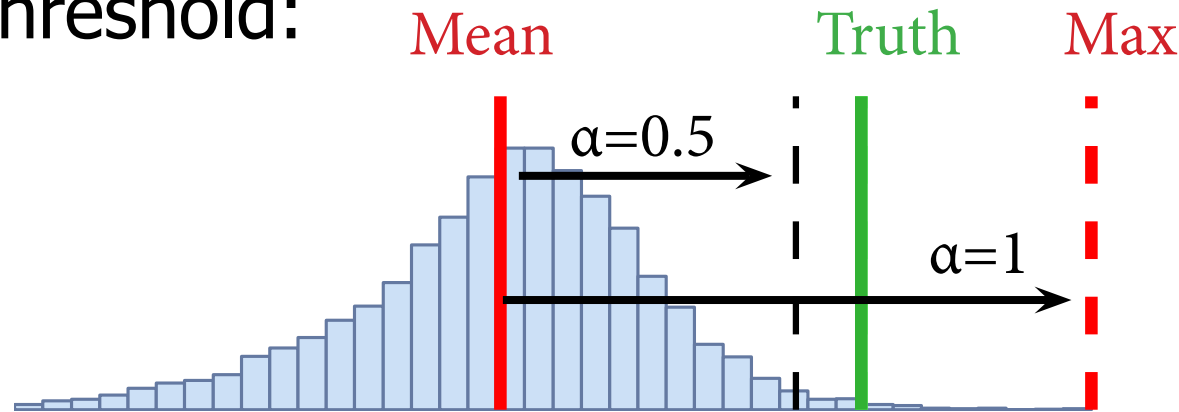
- Train to minimize pruning error (rather than 1-best)
- Pruning threshold:



Structured Prediction Cascades Training

[Weiss & Taskar '10]

- Train to minimize pruning error (rather than 1-best)
- Pruning threshold:



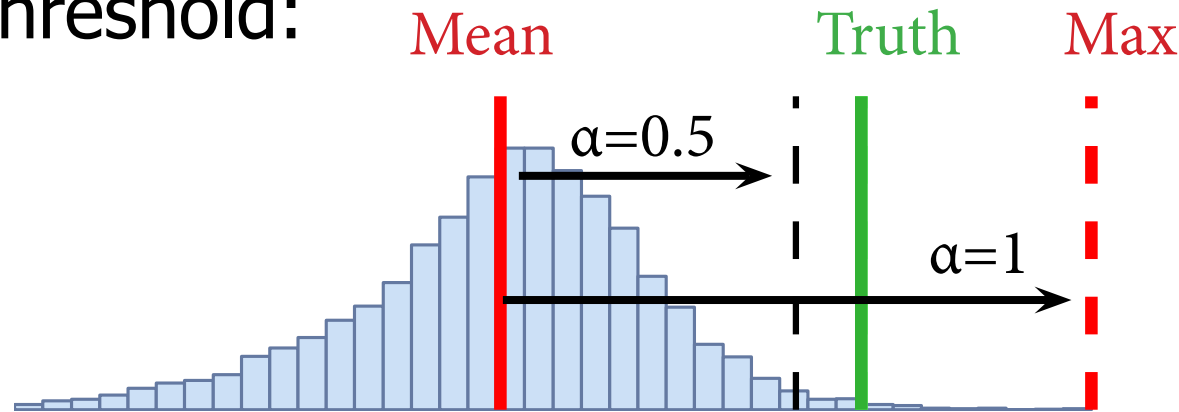
- Training objective:

$$\min_w \lambda \|w\|^2 + \frac{1}{M} \sum_{m=1}^M \max\{0, 1 + t_\alpha(w) - y^m \cdot w\}$$

Structured Prediction Cascades Training

[Weiss & Taskar '10]

- Train to minimize pruning error (rather than 1-best)
- Pruning threshold:

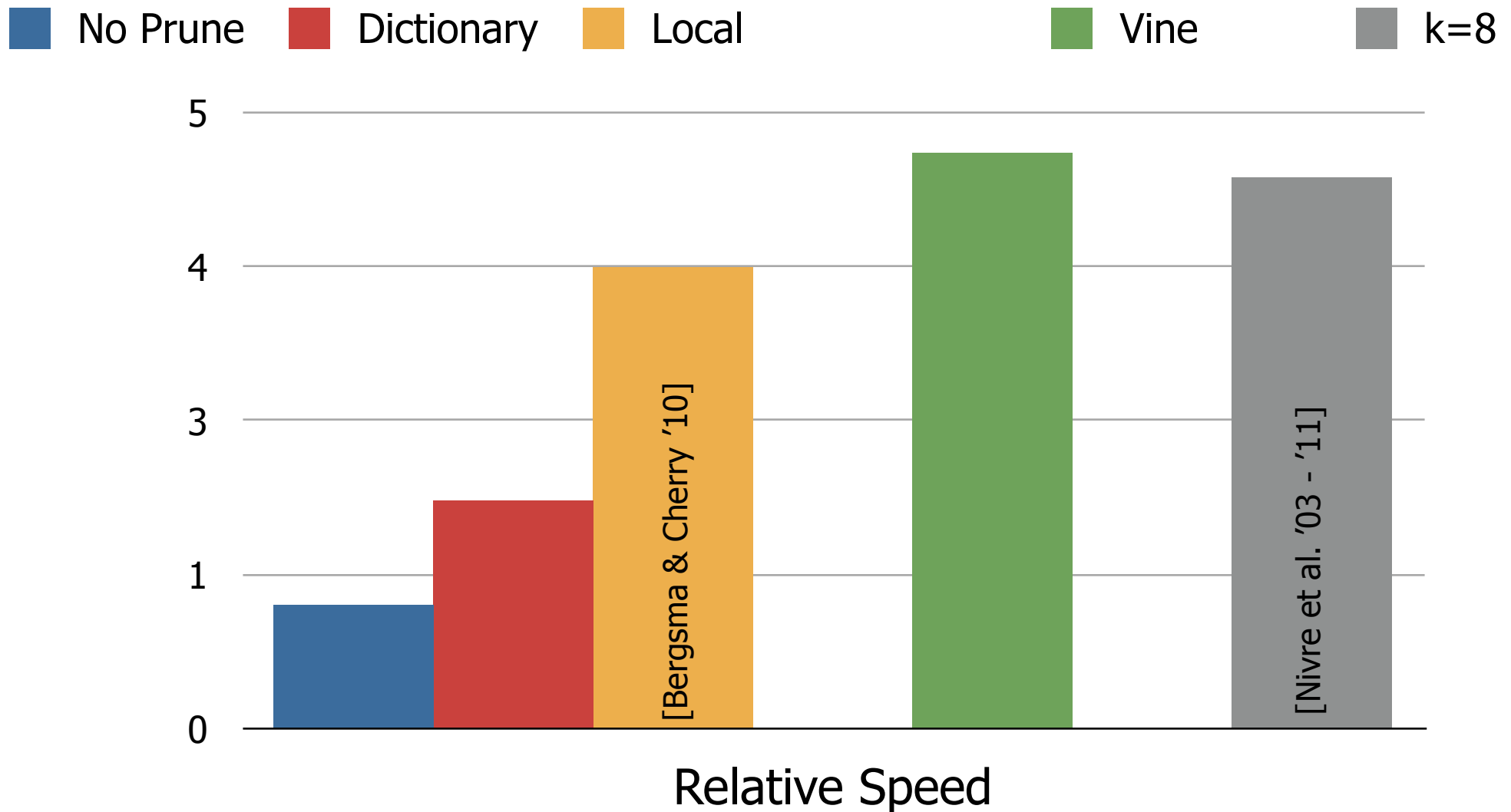


- Training objective:
 - Filter as many arcs as possible
 - While preserving gold arc
 - Optimize with stochastic gradient decent (not so different from perceptron updates)

First-Order Parsing



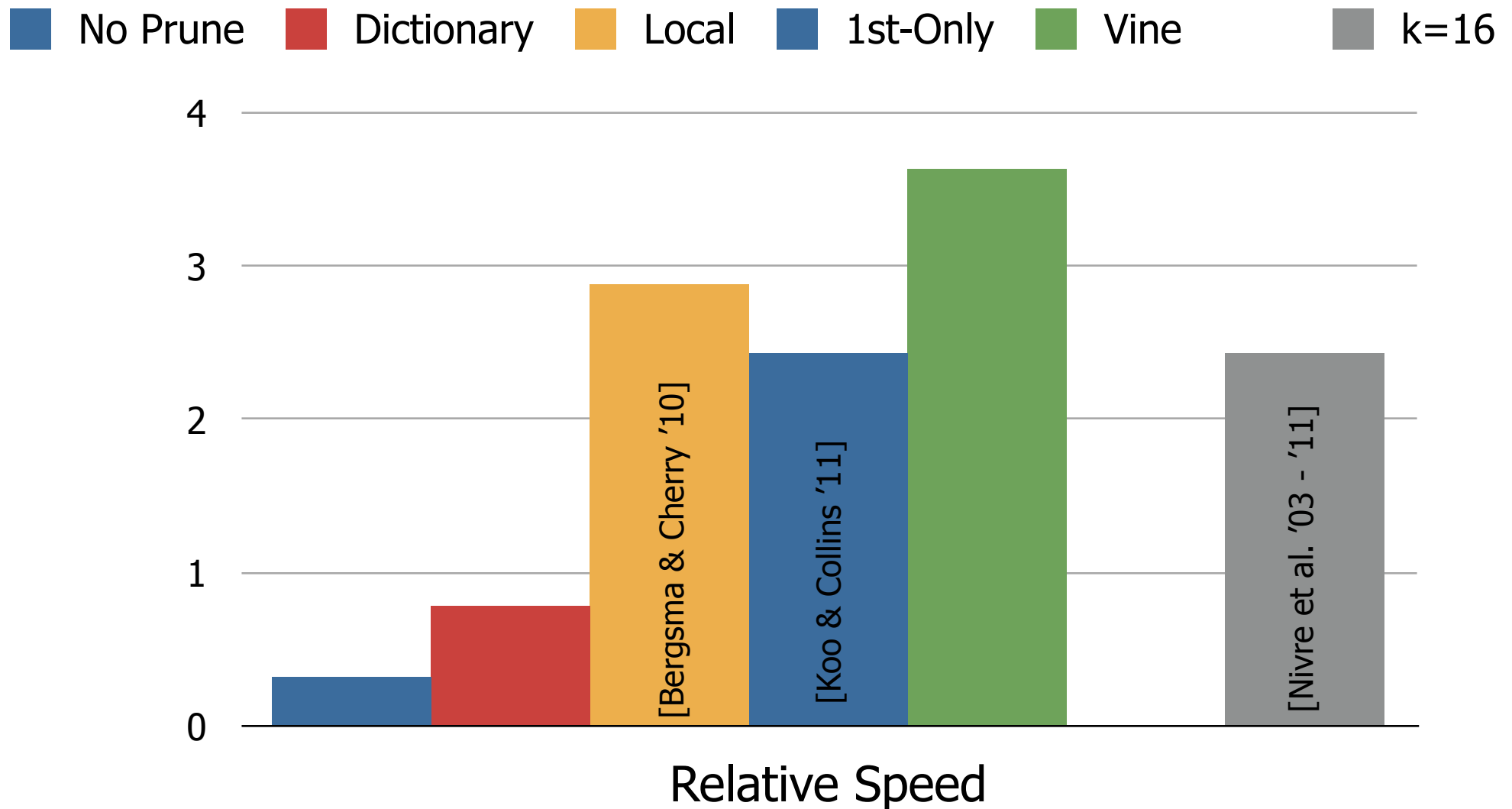
UAS 91.0, Set pruning thresholds for no loss in accuracy



Second-Order Parsing



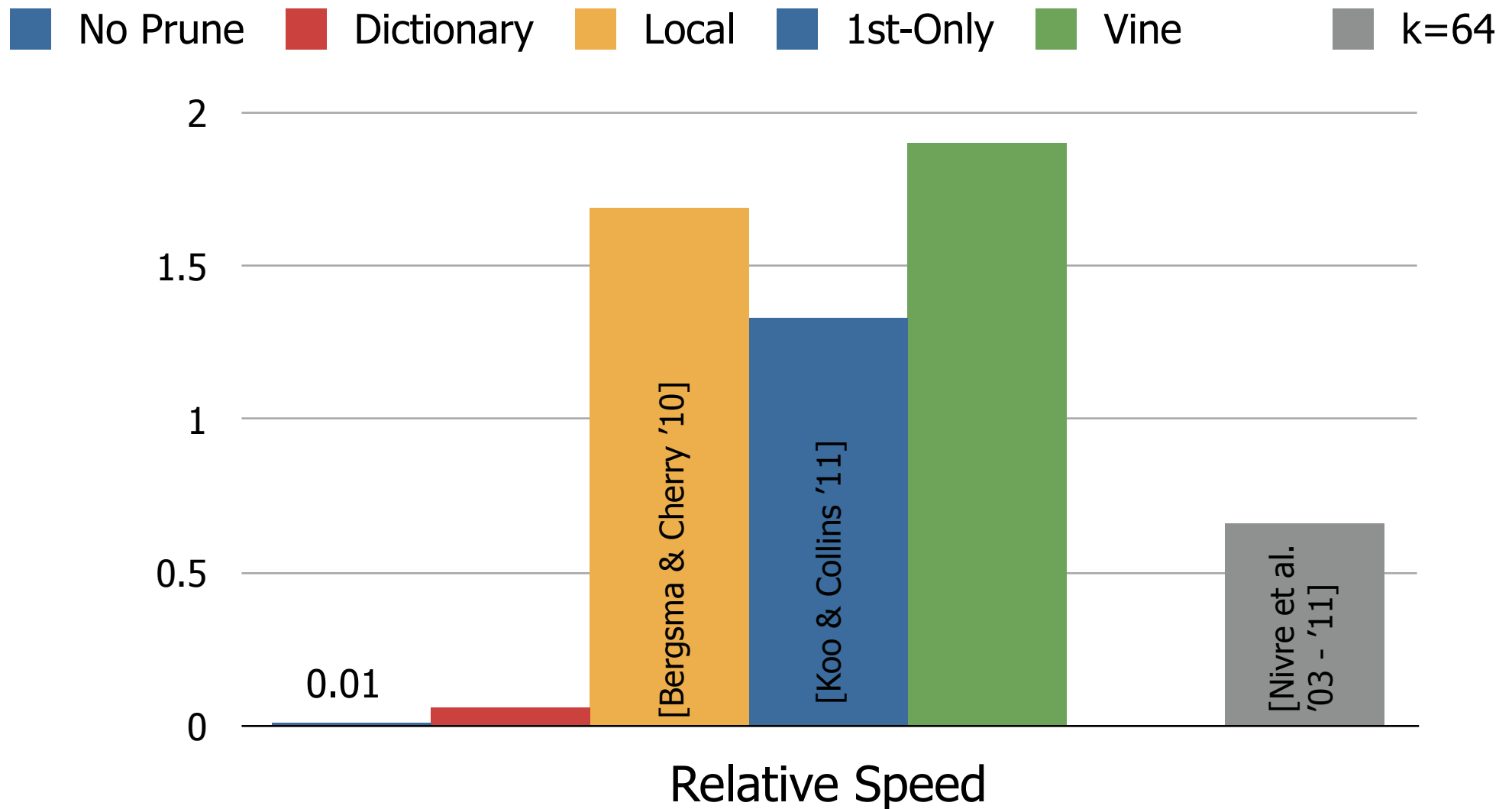
UAS 92.1, Set pruning thresholds for no loss in accuracy



Third-Order Parsing



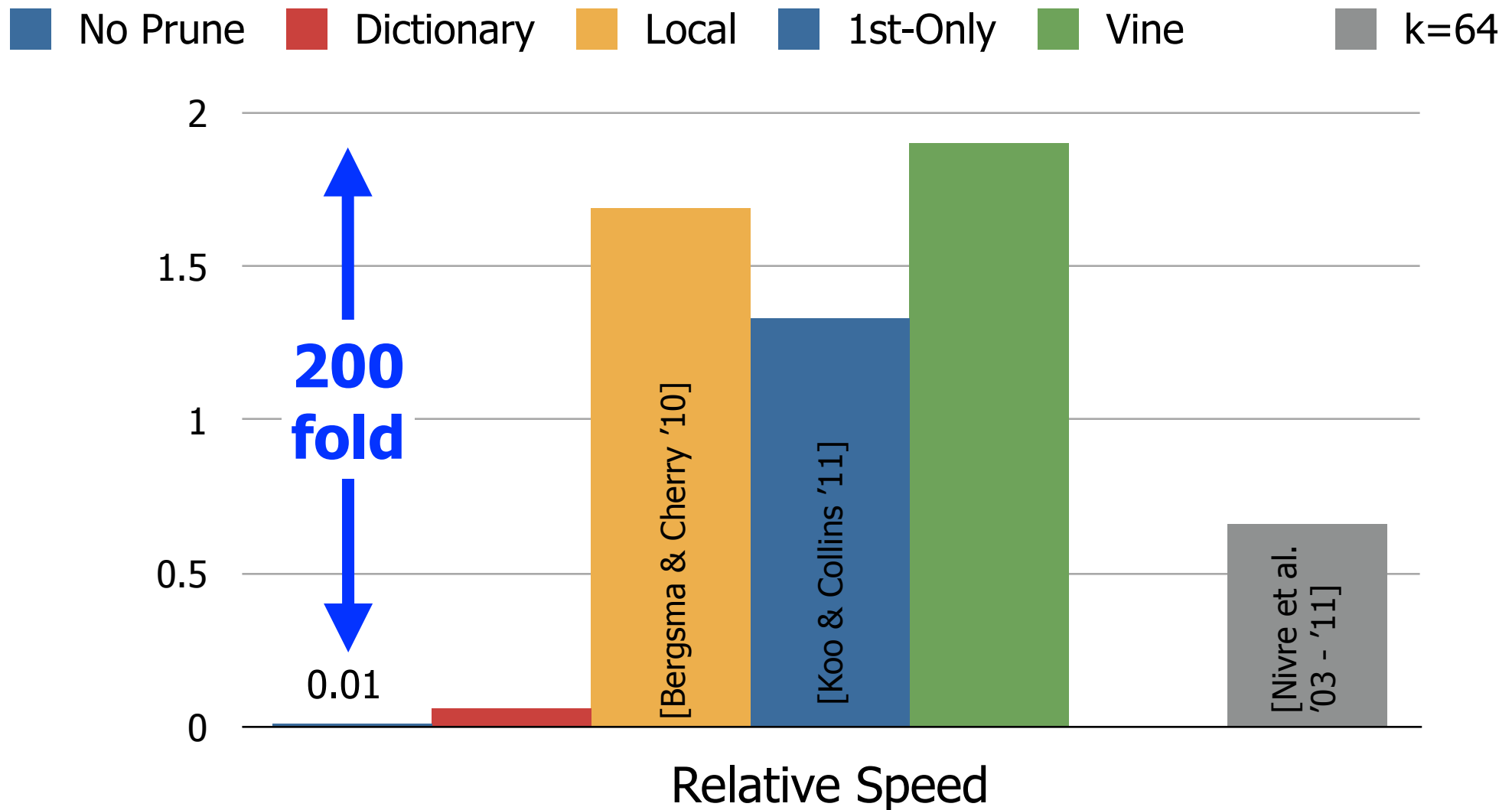
UAS 92.9, Set pruning thresholds for no loss in accuracy



Third-Order Parsing



UAS 92.9, Set pruning thresholds for no loss in accuracy



Google Web Treebanks



[Petrov & McDonald '12]

- Google Web Treebank
 - Funded by Google, annotated and released by LDC
 - 5 domains: Blogs, Newsgroups, Reviews, Emails, Q&A
 - ~2,000 manually annotated sentences (PTB-style)
 - >100,000 unlabeled sentences
- Shared Task at NAACL '12 Workshop
 - Constituency Trees or Stanford Dependencies
 - Train on WSJ + unlabeled data
 - 2 domains released for development
 - Test on remaining 3 domains
- New: Google Multilingual Treebank (6 languages)

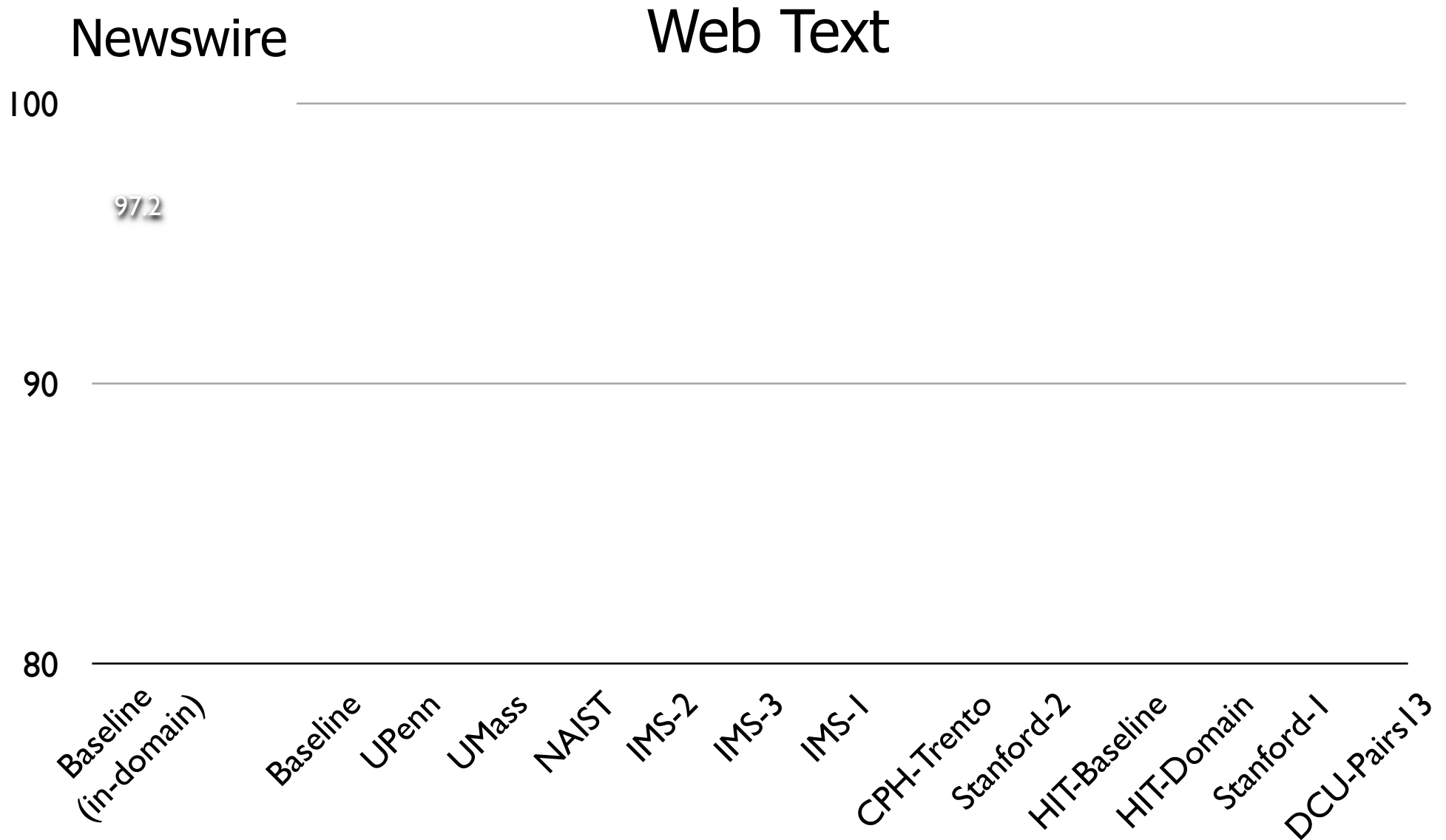
POS Accuracy (SANCL Shared Task)

News wire

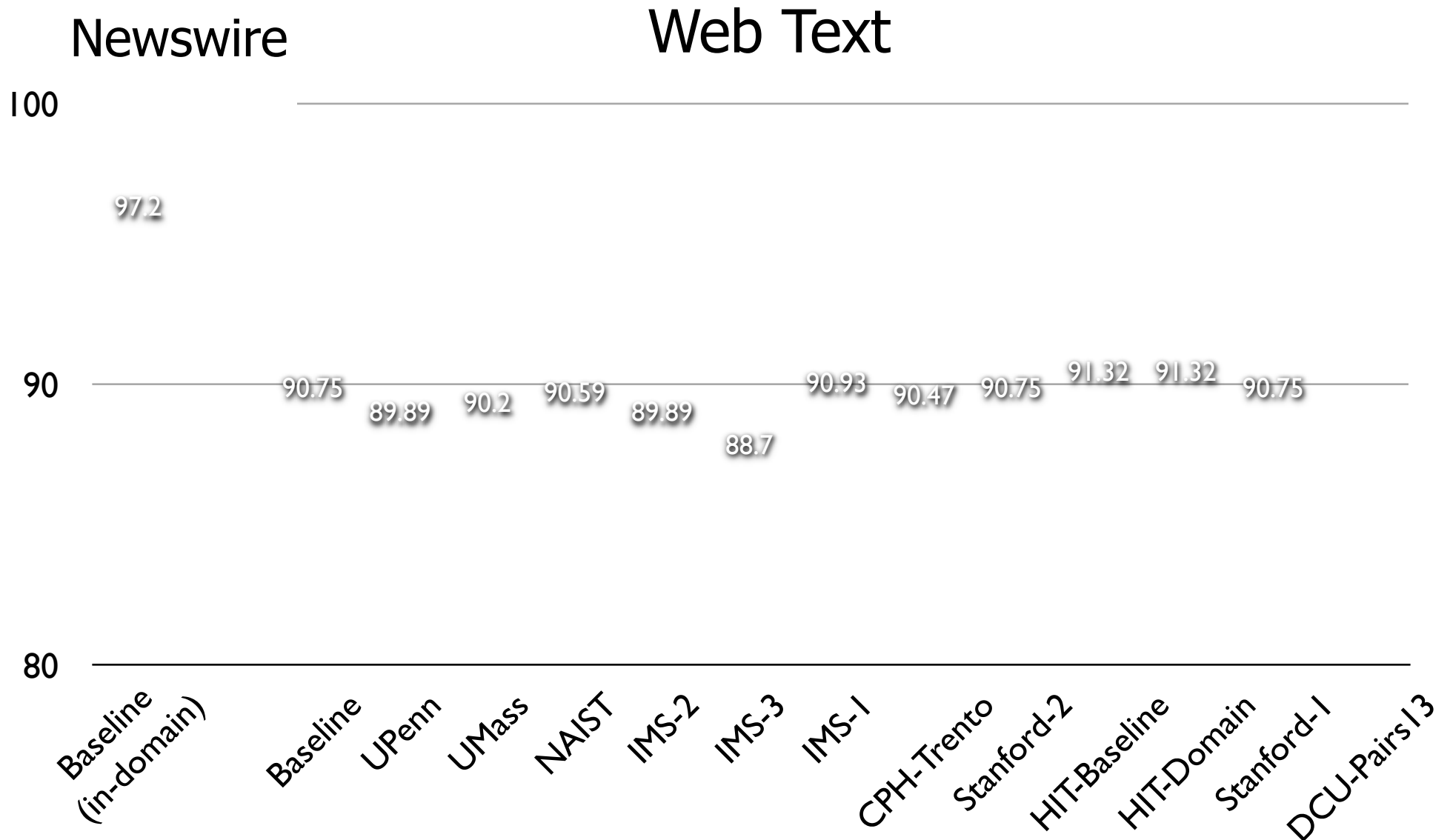
Web Text

Baseline

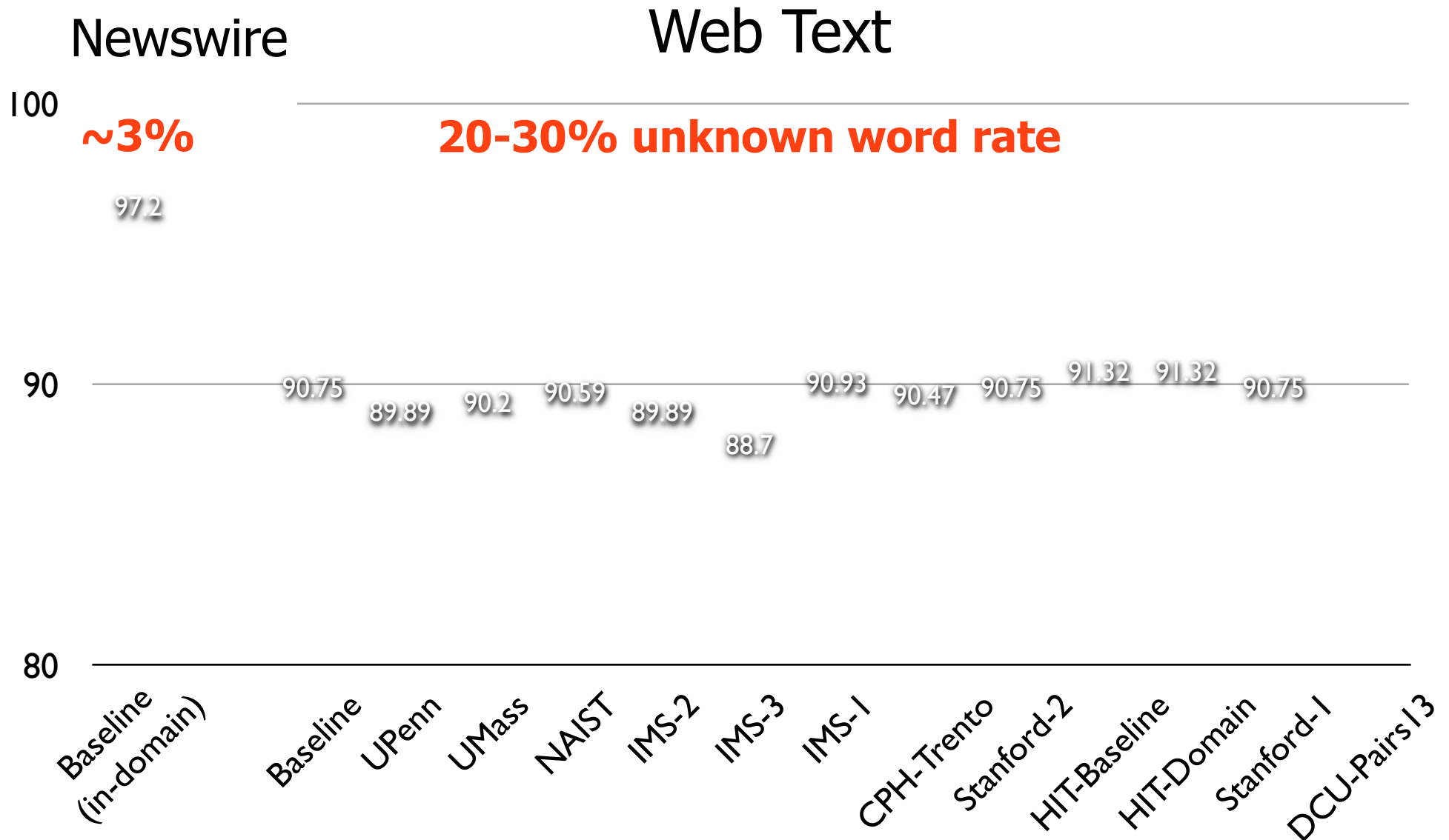
POS Accuracy (SANCL Shared Task)



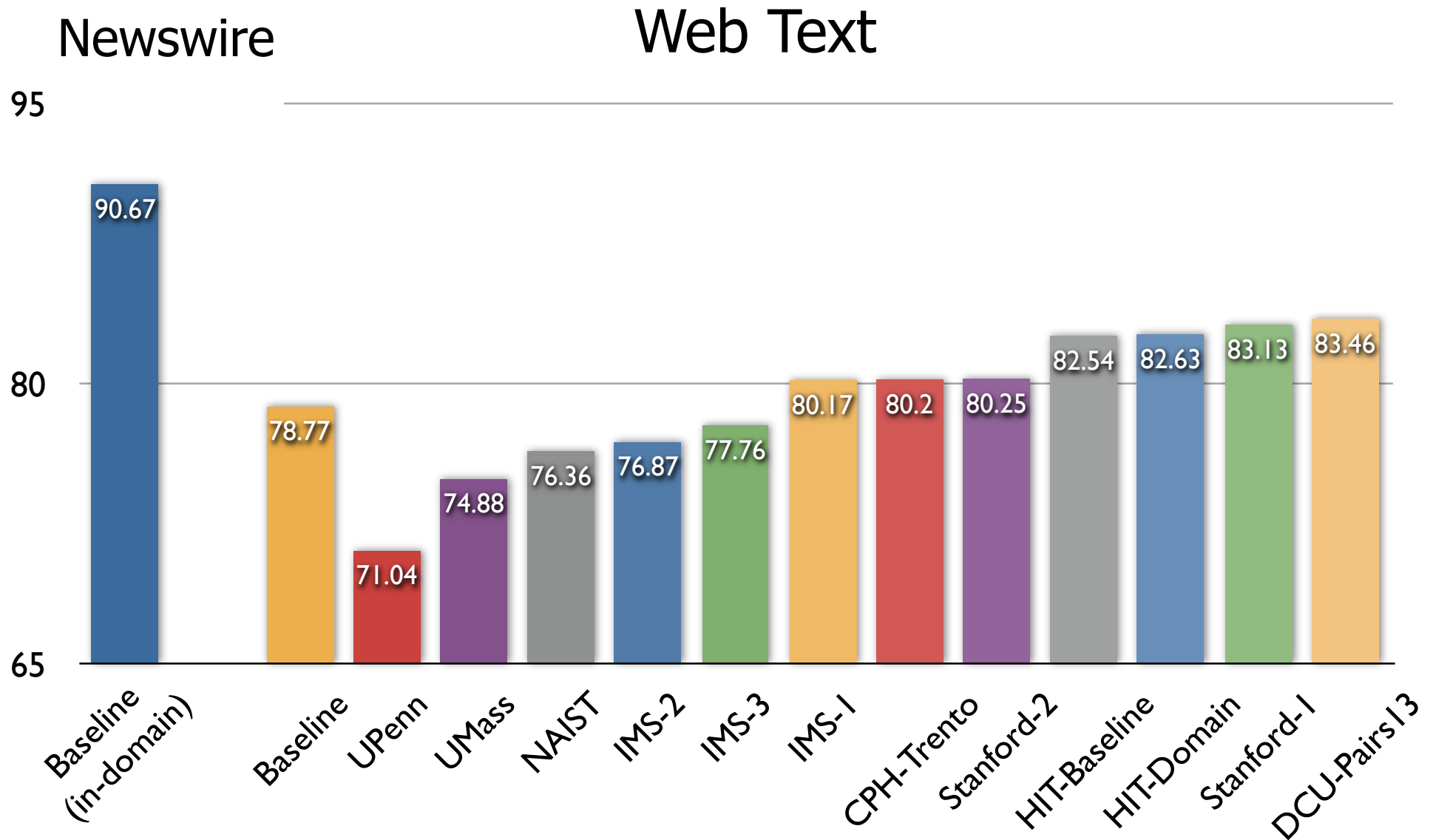
POS Accuracy (SANCL Shared Task)



POS Accuracy (SANCL Shared Task)

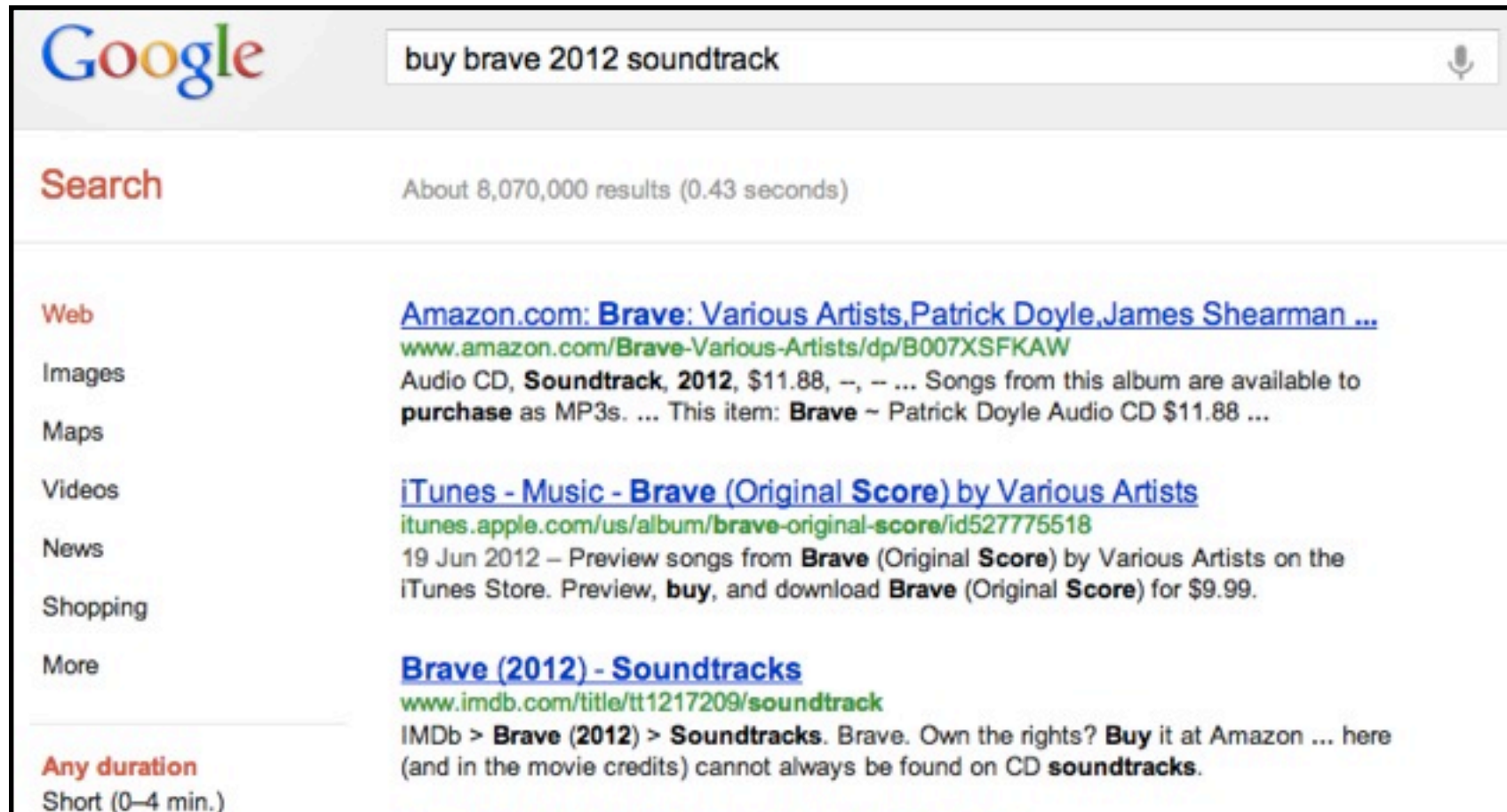


Parsing Accuracy (SANCL Shared Task)



Tagging Search Queries

[Ganchev et al. '12]



The image shows a screenshot of a Google search results page. At the top, the Google logo is on the left, and the search bar contains the text "buy brave 2012 soundtrack" with a microphone icon on the right. Below the search bar, the word "Search" is in red, followed by "About 8,070,000 results (0.43 seconds)". On the left side, there is a vertical menu with links: "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below this menu, there are filters: "Any duration" and "Short (0-4 min.)". The main content area displays three search results. The first result is from Amazon.com, titled "Brave: Various Artists, Patrick Doyle, James Shearman ...", with a green URL "www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW". The description says "Audio CD, Soundtrack, 2012, \$11.88, -, - ... Songs from this album are available to purchase as MP3s. ... This item: Brave ~ Patrick Doyle Audio CD \$11.88 ...". The second result is from iTunes, titled "iTunes - Music - Brave (Original Score) by Various Artists", with a green URL "itunes.apple.com/us/album/brave-original-score/id527775518". The description says "19 Jun 2012 - Preview songs from Brave (Original Score) by Various Artists on the iTunes Store. Preview, buy, and download Brave (Original Score) for \$9.99.". The third result is from IMDb, titled "Brave (2012) - Soundtracks", with a green URL "www.imdb.com/title/tt1217209/soundtrack". The description says "IMDb > Brave (2012) > Soundtracks. Brave. Own the rights? Buy it at Amazon ... here (and in the movie credits) cannot always be found on CD soundtracks."

Google

buy brave 2012 soundtrack

Search

About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Any duration

Short (0-4 min.)

[Amazon.com: Brave: Various Artists, Patrick Doyle, James Shearman ...](http://www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW)
www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack**, 2012, \$11.88, -, - ... Songs from this album are available to purchase as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

[iTunes - Music - Brave \(Original Score\) by Various Artists](http://itunes.apple.com/us/album/brave-original-score/id527775518)
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 - Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave (Original Score)** for \$9.99.

[Brave \(2012\) - Soundtracks](http://www.imdb.com/title/tt1217209/soundtrack)
www.imdb.com/title/tt1217209/soundtrack
IMDb > **Brave (2012)** > **Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

Search About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Any duration

Short (0–4 min.)

[Amazon.com: **Brave**: Various Artists, Patrick Doyle, James Shearman ...](#)
www.amazon.com/Brave-Various-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack**, 2012, \$11.88, –, – ... Songs from this album are available to **purchase** as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

[iTunes - Music - **Brave \(Original Score\)** by Various Artists](#)
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 – Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave (Original Score)** for \$9.99.

[Brave \(2012\) - Soundtracks](#)
www.imdb.com/title/tt1217209/soundtrack
IMDb > **Brave (2012)** > **Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

ADJ or NOUN?

Search About 8,070,000 results (0.43 seconds)

Web

[Amazon.com: **Brave**: Various Artists, Patrick Doyle, James Shearman ...](http://www.amazon.com/Brave-Various-Artists/dp/B007XSFKAW)
www.amazon.com/Brave-Various-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack**, **2012**, \$11.88, —, — ... Songs from this album are available to **purchase** as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

Images

Maps

Videos

News

Shopping

More

Any duration
Short (0–4 min.)

[iTunes - Music - **Brave** \(Original Score\) by Various Artists](http://itunes.apple.com/us/album/brave-original-score/id527775518)
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 – Preview songs from **Brave** (Original Score) by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave** (Original Score) for \$9.99.

[Brave \(2012\) - Soundtracks](http://www.imdb.com/title/tt1217209/soundtrack)
www.imdb.com/title/tt1217209/soundtrack
IMDb > **Brave** (2012) > **Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

ADJ or NOUN?

Search About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Amazon.com: Brave: Various Artists, Patrick Doyle, James Shearman ...
www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack**, **2012**, \$11.88, --, -- ... Songs from this album are available to **purchase** as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

iTunes - Music - Brave (Original Score) by Various Artists
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 – Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave (Original Score)** for \$9.99.

Brave (2012) - Soundtracks
www.imdb.com/title/tt1217209/soundtrack
IMDb > **Brave (2012)** > **Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

Any duration
Short (0–4 min.)

buy	brave	2012	soundtrack
VB	JJ	CD	NN
VERB	ADJ	NUM	NOUN

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

ADJ or NOUN?

Search About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Any duration Short (0–4 min.)

Amazon.com: **Brave**: Various Artists, Patrick Doyle, James Shearman ...
www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack**, 2012, \$11.88 ... – ... Songs from this album are available to purchase as MP3s. ... This item: **Brave** - Patrick Doyle Audio CD \$11.88 ...

iTunes - Music - **Brave (Original Score)** by Various Artists
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 – Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, buy, and download **Brave (Original Score)** for \$9.99.

Brave (2012) - Soundtracks
www.imdb.com/title/tt1217209/soundtrack
IMDb: **Brave (2012)** > **Soundtracks**. Brave. Own the rights? Buy it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

buy brave 2012 soundtrack

VB
VERB

JJ
ADJ

CD
NUM

NN
NOUN

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

ADJ or NOUN?

Search About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

[Amazon.com: Brave: Various Artists, Patrick Doyle, James Shearman ...](#)
[www.amazon.com/Brave/Variou-Artists/dp/B007XSFKAW](#)
Audio CD, **Soundtrack**, 2012, \$11.88, --, -- ... Songs from this album are available to purchase as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

[iTunes - Music - Brave \(Original Score\) by Various Artists](#)
[itunes.apple.com/us/album/brave-original-score/id527775518](#)
19 Jun 2012 – Preview songs from **Brave** (Original Score) by Various Artists on the iTunes Store. Preview, buy, and download **Brave** (Original Score) for \$9.99.

Preview	songs	from	Brave	(	Original	Score	)	by	Various	Artists	.
NN	NNS	IN	NNP	(	JJ	NN	)	IN	JJ	NNS	.
NOUN	NOUN	ADP	NOUN	.	ADJ	NOUN	.	ADP	ADJ	NOUN	.

buy	brave	2012	soundtrack
VB	JJ	CD	NN
VERB	ADJ	NUM	NOUN

Tagging Search Queries

[Ganchev et al. '12]

The screenshot shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "buy brave 2012 soundtrack". Below the search bar, the text "VERB NOUN NUM NOUN" is displayed in purple. Below this, the word "Search" is in red, followed by "About 8,070,000 results (0.43 seconds)". On the left side, there is a vertical menu with links: "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". The main content area displays three search results. The first result is from Amazon.com, titled "Brave: Various Artists, Patrick Doyle, James Shearman ...", with a green URL "www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW". The description mentions "Audio CD, Soundtrack, 2012, \$11.88, -, - ... Songs from this album are available to purchase as MP3s. ... This item: Brave ~ Patrick Doyle Audio CD \$11.88 ...". The second result is from iTunes, titled "iTunes - Music - Brave (Original Score) by Various Artists", with a green URL "itunes.apple.com/us/album/brave-original-score/id527775518". The description mentions "19 Jun 2012 - Preview songs from Brave (Original Score) by Various Artists on the iTunes Store. Preview, buy, and download Brave (Original Score) for \$9.99.". The third result is from IMDb, titled "Brave (2012) - Soundtracks", with a green URL "www.imdb.com/title/tt1217209/soundtrack". The description mentions "IMDb > Brave (2012) > Soundtracks. Brave. Own the rights? Buy it at Amazon ... here (and in the movie credits) cannot always be found on CD soundtracks.".

Google

buy brave 2012 soundtrack

VERB NOUN NUM NOUN

Search

About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Any duration

Short (0-4 min.)

Amazon.com: **Brave: Various Artists, Patrick Doyle, James Shearman ...**
www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW
Audio CD, **Soundtrack, 2012**, \$11.88, -, - ... Songs from this album are available to **purchase** as MP3s. ... This item: **Brave ~ Patrick Doyle** Audio CD \$11.88 ...

iTunes - Music - **Brave (Original Score) by Various Artists**
itunes.apple.com/us/album/brave-original-score/id527775518
19 Jun 2012 - Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave (Original Score)** for \$9.99.

Brave (2012) - Soundtracks
www.imdb.com/title/tt1217209/soundtrack
IMDb > **Brave (2012) > Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

buy brave 2012 soundtrack

VB
VERB

JJ
ADJ

CD
NUM

NN
NOUN

Tagging Search Queries

[Ganchev et al. '12]

Google

buy brave 2012 soundtrack

VERB NOUN NUM NOUN

Search About 8,070,000 results (0.43 seconds)

Web

Images

Maps

Videos

News

Shopping

More

Any duration Short (0–4 min.)

[Amazon.com: **Brave**: Various Artists, Patrick Doyle, James Shearman ...](#)
[www.amazon.com/Brave-Variou-Artists/dp/B007XSFKAW](#)
Audio CD, **Soundtrack**, 2012, \$11.88, –, – ... Songs from this album are available to **purchase** as MP3s. ... This item: **Brave** ~ Patrick Doyle Audio CD \$11.88 ...

[iTunes - Music - **Brave \(Original Score\)** by Various Artists](#)
[itunes.apple.com/us/album/brave-original-score/id527775518](#)
19 Jun 2012 – Preview songs from **Brave (Original Score)** by Various Artists on the iTunes Store. Preview, **buy**, and download **Brave (Original Score)** for \$9.99.

[Brave \(2012\) - Soundtracks](#)
[www.imdb.com/title/tt1217209/soundtrack](#)
IMDb > **Brave (2012)** > **Soundtracks**. Brave. Own the rights? **Buy** it at Amazon ... here (and in the movie credits) cannot always be found on CD **soundtracks**.

Baseline

buy	brave	2012	soundtrack
VB	JJ	CD	NN
VERB	ADJ	NUM	NOUN

Retrained

buy	brave	2012	soundtrack
VB	NNP	CD	NN
VERB	NOUN	NUM	NOUN



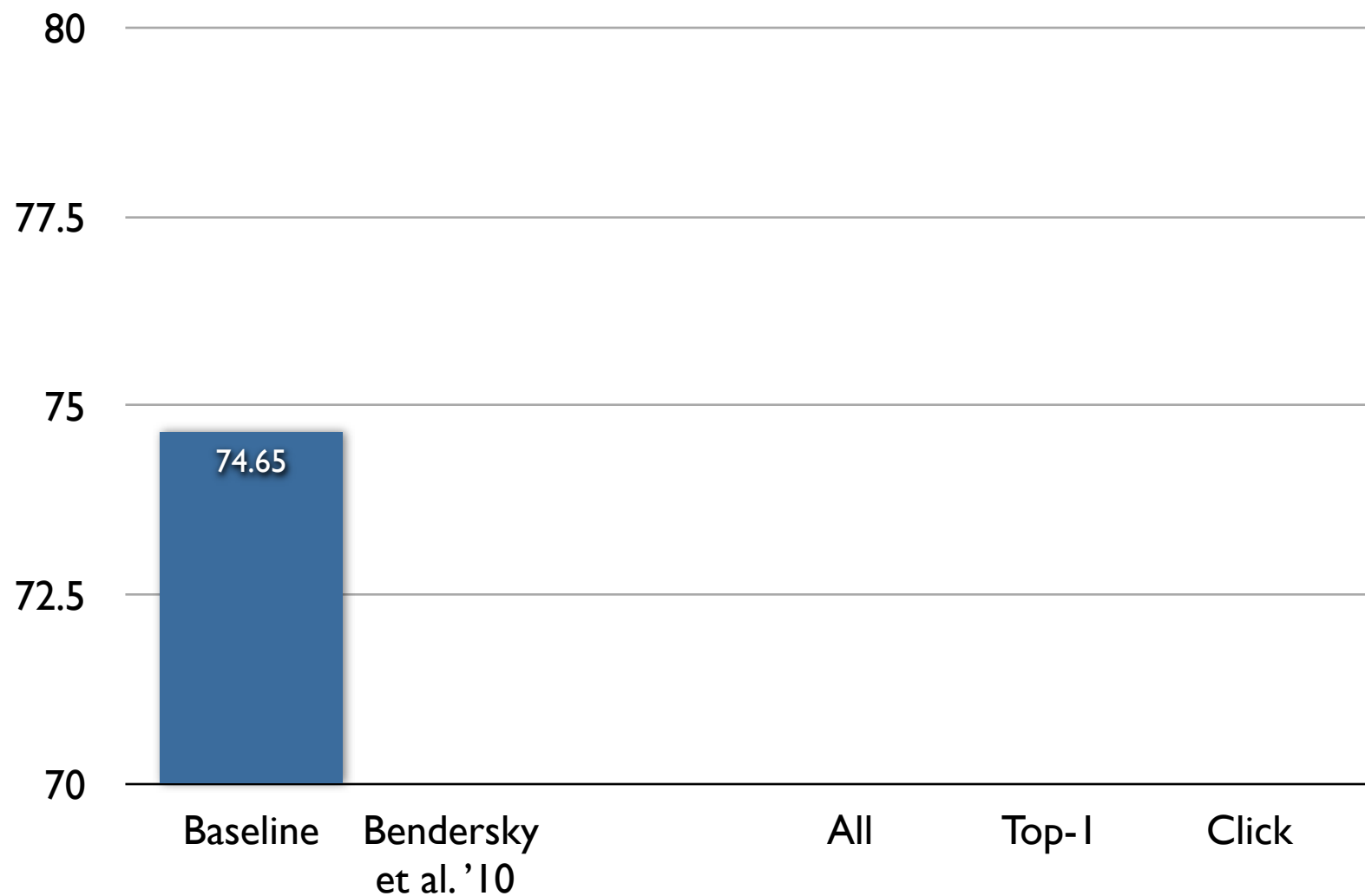
Query Tagging Results

[Ganchev et al. '12]

et al. '10

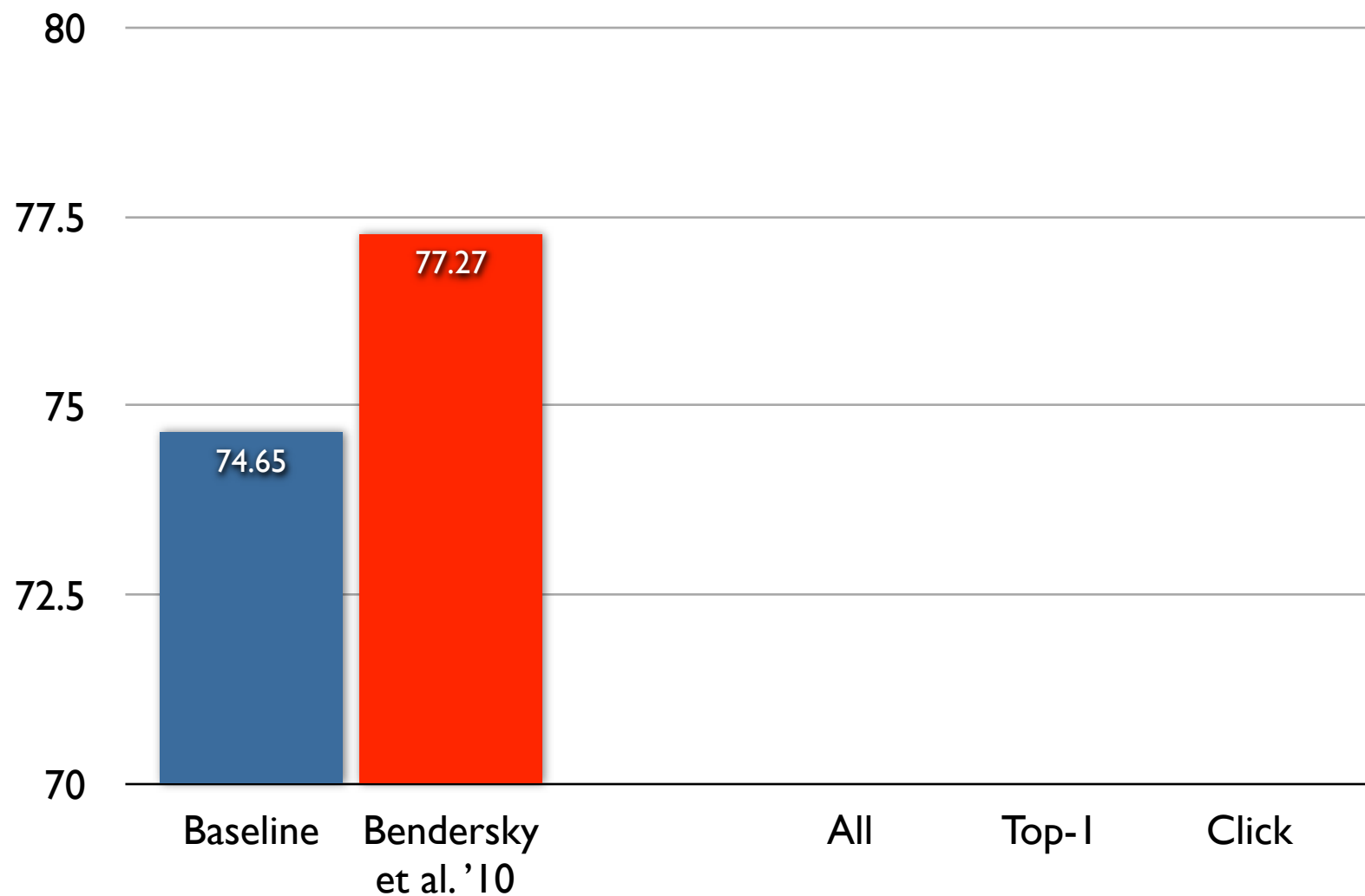
Query Tagging Results

[Ganchev et al. '12]



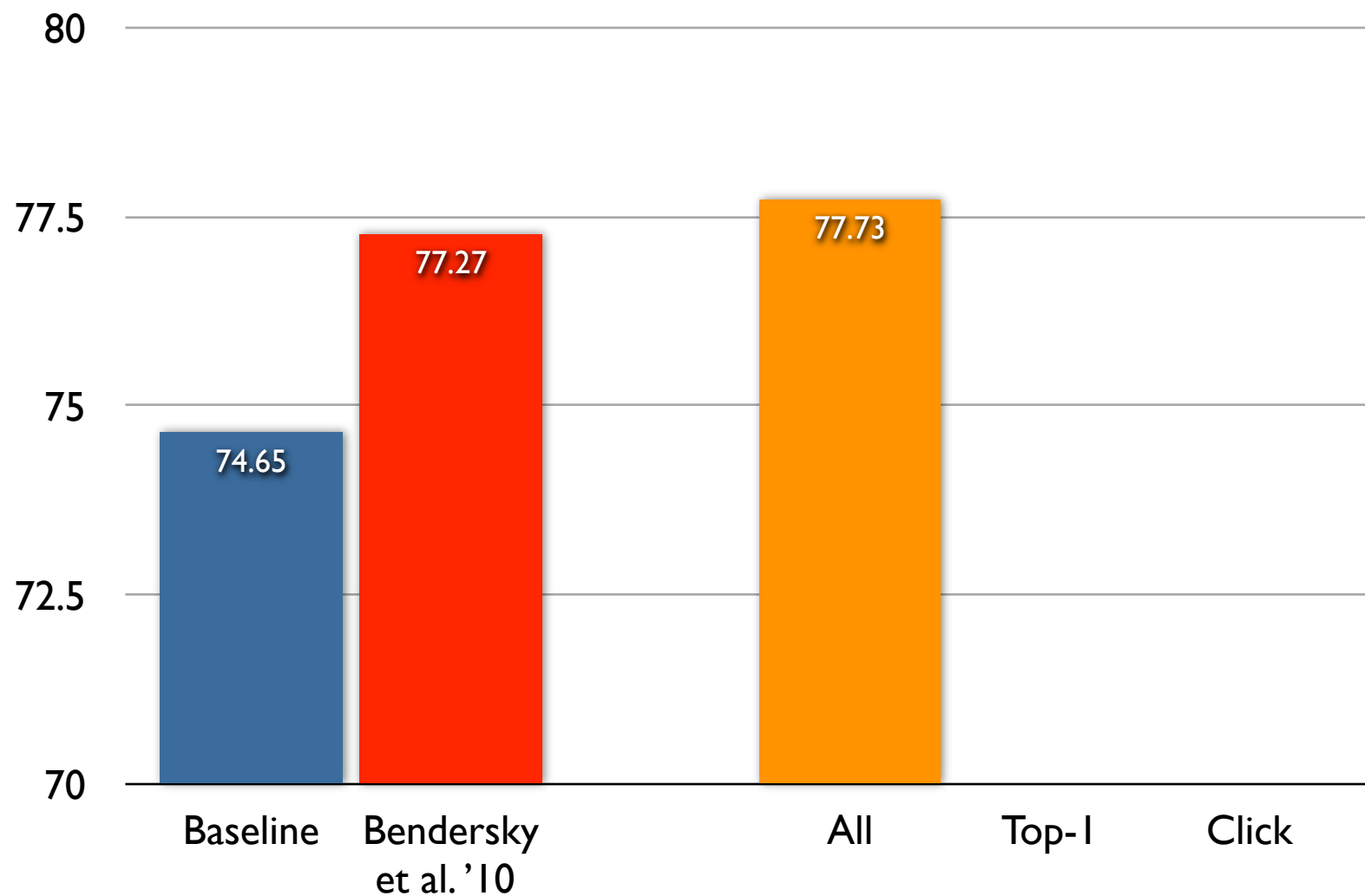
Query Tagging Results

[Ganchev et al. '12]



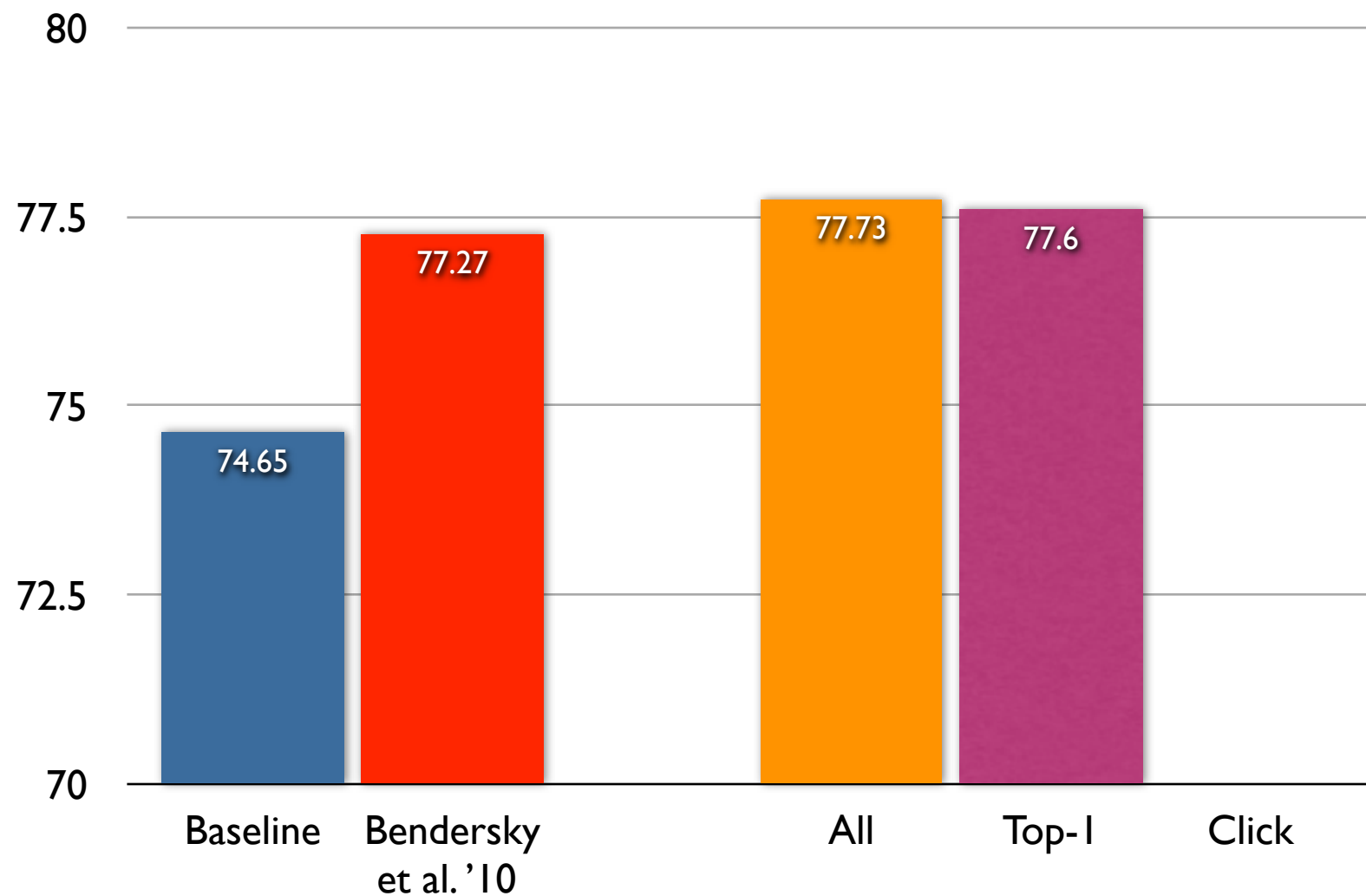
Query Tagging Results

[Ganchev et al. '12]



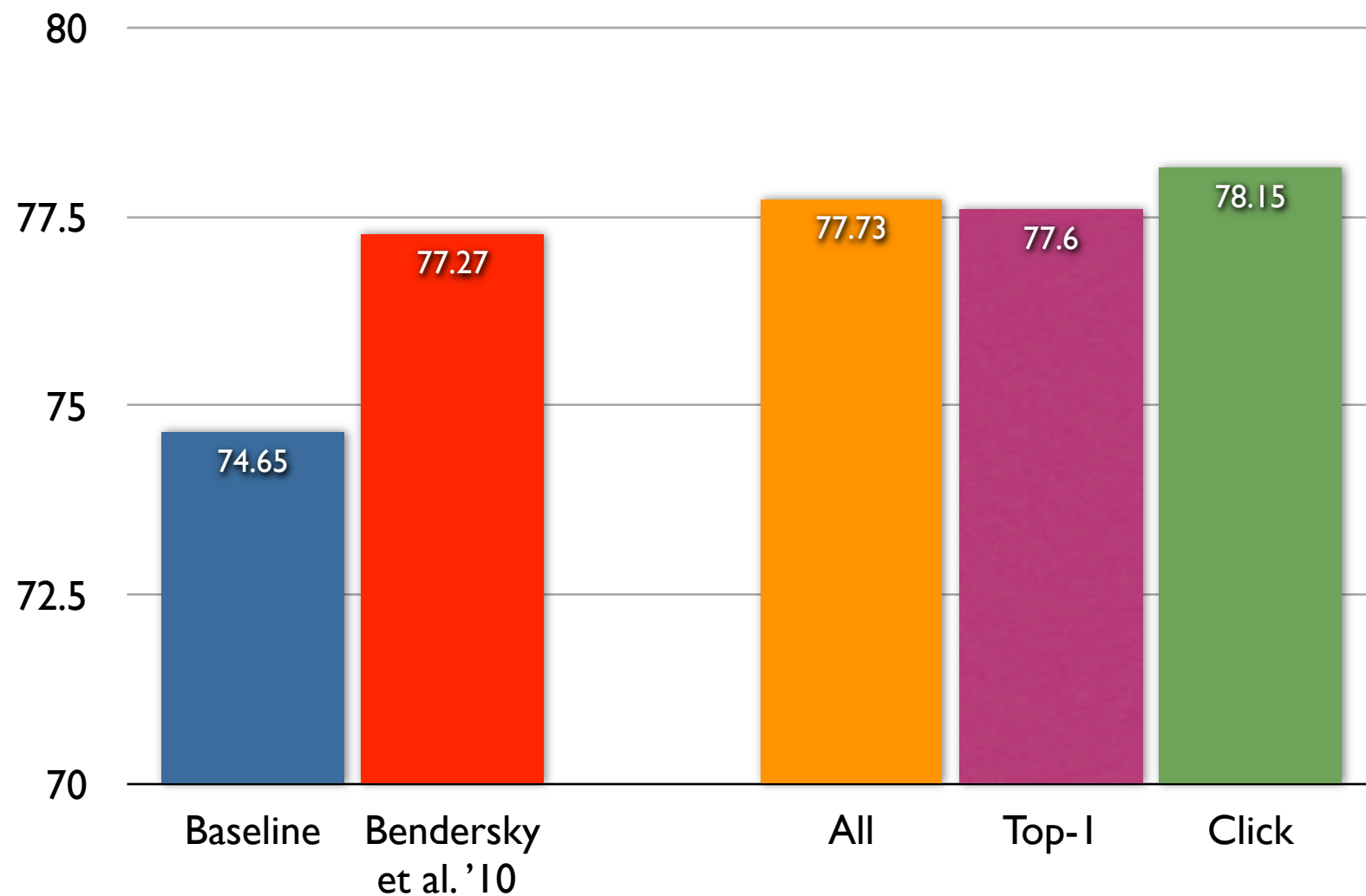
Query Tagging Results

[Ganchev et al. '12]



Query Tagging Results

[Ganchev et al. '12]



The breaking bad case...



who stars in breaking bad

Google

who stars in breaking bad

Web Images Maps Shopping News More Search tools

Breaking Bad cast

Bryan Cranston
Walter White

Aaron Paul
Jesse Pinkman

RJ Mitte
Walter White...

Anna Gunn
Skyler White

Giancarlo Esposito
Gustavo Fring

Dean Norris
Hank Schrader

Jonathan Banks
Mike Ehrmantraut

Betsy Brandt
Marie Schrader

Bob Odenkirk
Saul Goodman

Laura Fraser
Lydia Rodarte-Blanco

Feedback / More info

[Breaking Bad Cast - TV.com](#)
[www.tv.com/shows/breaking-bad/cast/](#)
70+ items -- **Breaking Bad Cast:** Bryan Cranston, Aaron Paul, Anna Gunn, ...
Bob Odenkirk Saul Goodman (Season 3)
Giancarlo Esposito Gus Fring (Season 3-4)

[News for who stars in breaking bad](#)

[Breaking Bad costumes donated to homeless in New Mexico](#)
[Star Magazine UK](#) - 13 hours ago
STAR Magazine :: Breaking News :: **Breaking Bad** costumes ... The items, worn by Bryan Cranston and his co-stars in the hit drama, were ...
Starpulse.com

[Breaking Bad Stars Seize the Silver Screen in 2013 ... - AMC Blogs](#)
[blogs.amctv.com/breaking-bad/2013/.../cast-in-movies-2013.p...](#)
Jan 2, 2013 -- Bryan Cranston, Aaron Paul, Dean Norris, Giancarlo Esposito and more are ... RJ Mitte He may seem straight-faced on **Breaking Bad**, but in the ...

[Breaking Bad - Cast - AMC](#)
[www.amctv.com/shows/breaking-bad/cast](#)
Read about the **Breaking Bad** cast and the characters they play. ... **Breaking Bad** Gus Fring Season 4. Gus Fring is the owner of a fast food chicken restaurant ...
Jane Margolis - Walter White, Jr. - Gus - Walter White

[Breaking Bad - A Look Ahead: Inside Breaking Bad - AMC](#)

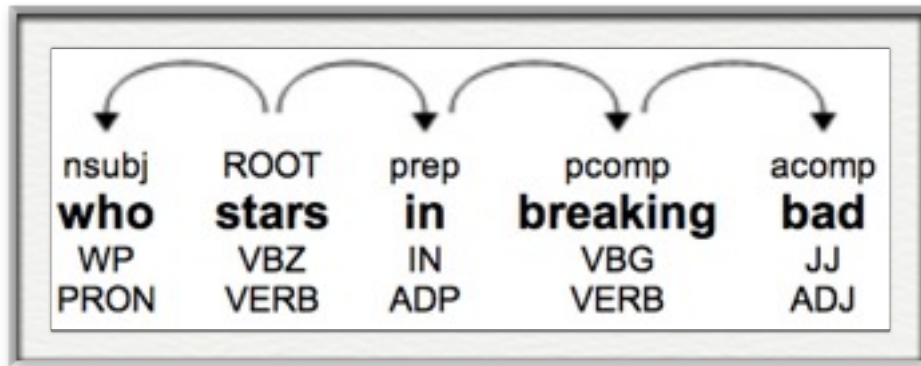
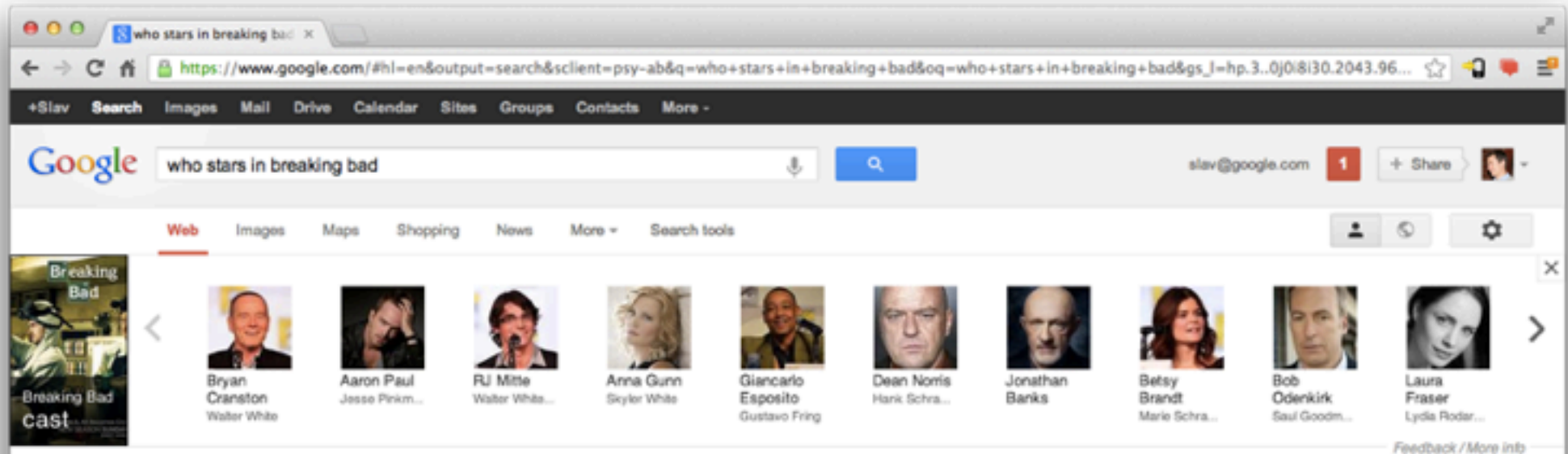
Breaking Bad

Breaking Bad is an American television drama series created and produced by Vince Gilligan. Set and produced in Albuquerque, New Mexico, **Breaking Bad** is the story of Walter White, a struggling high ... Wikipedia

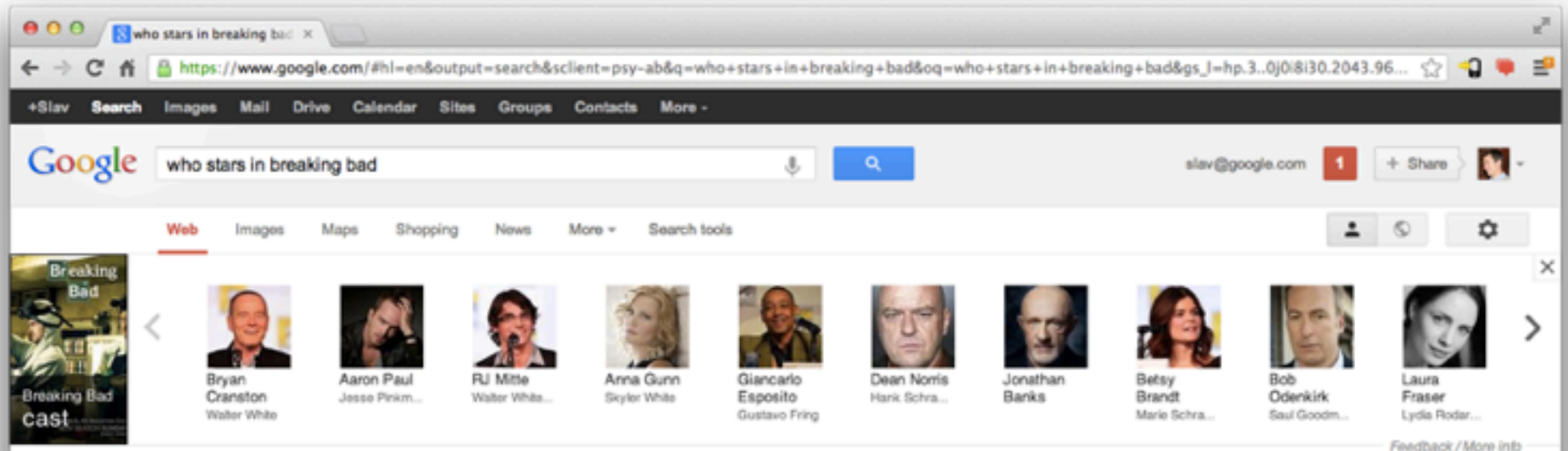
First episode: January 20, 2008
Theme song: **Breaking Bad** Theme
Network: AMC
Program creator: Vince Gilligan
Writers: Vince Gilligan, John Shiban, George Mastras, Sam Catlin, Patty Lin, Thomas Schnauz

People also search for

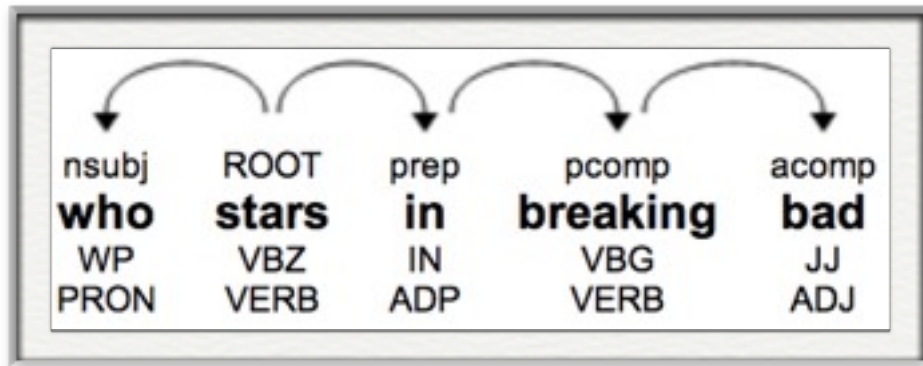
The breaking bad case...



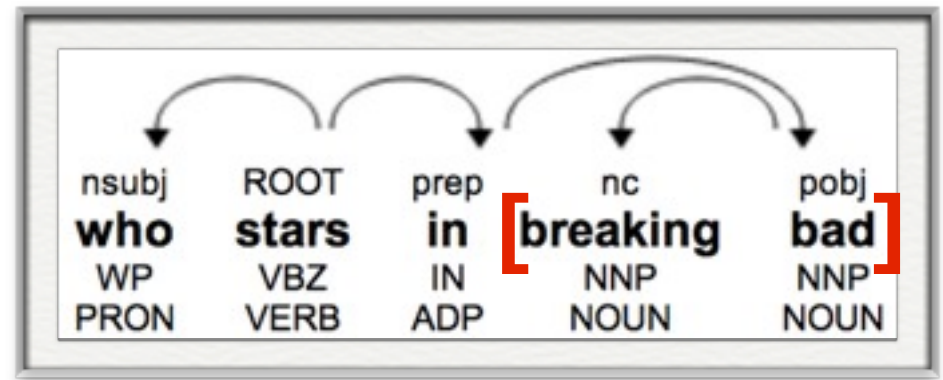
The breaking bad case...



Before

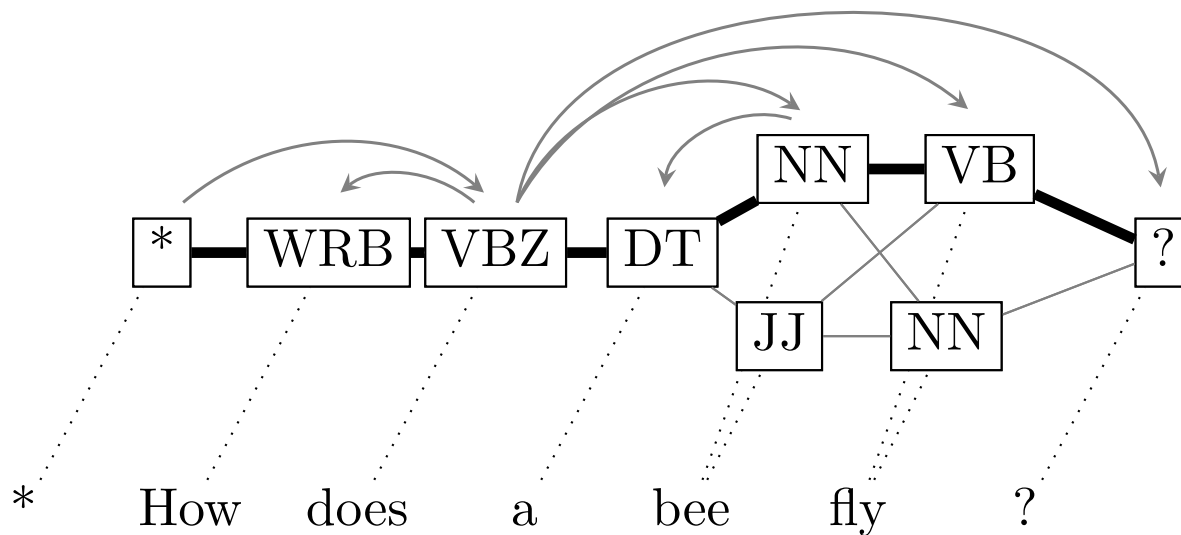


After



Use freebase entries as features

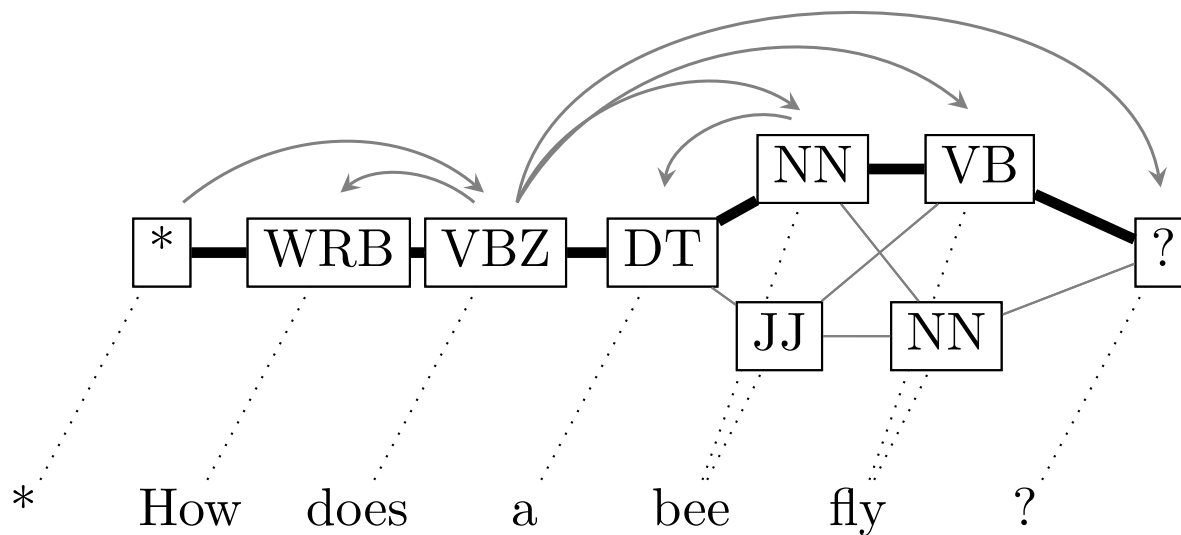
Integrated Tagging & Parsing



Integrated Tagging & Parsing



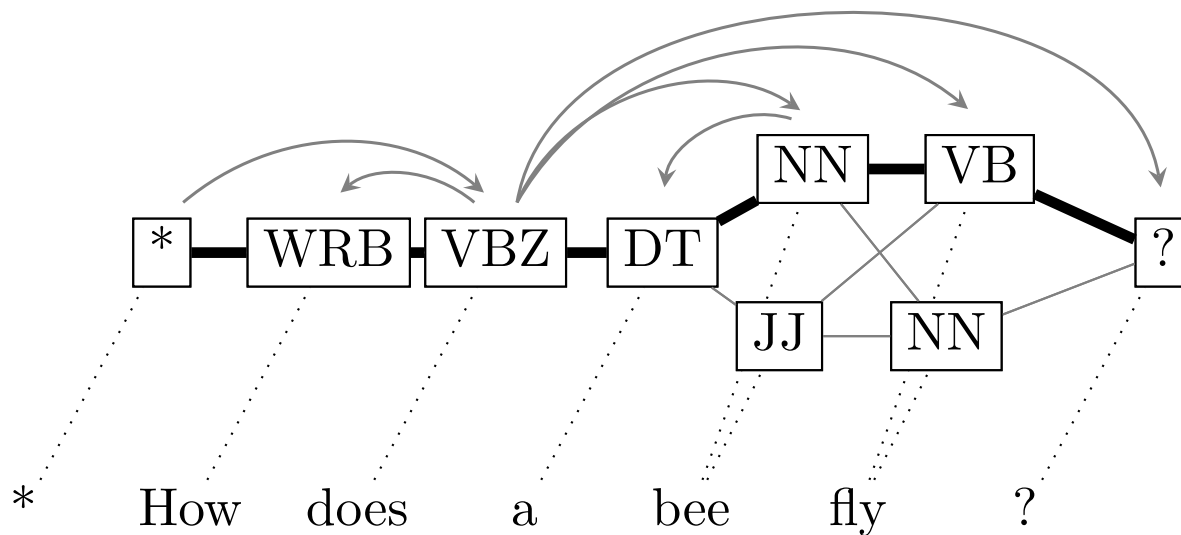
- Sequential Markov Models have limitations:



Integrated Tagging & Parsing



- Sequential Markov Models have limitations:

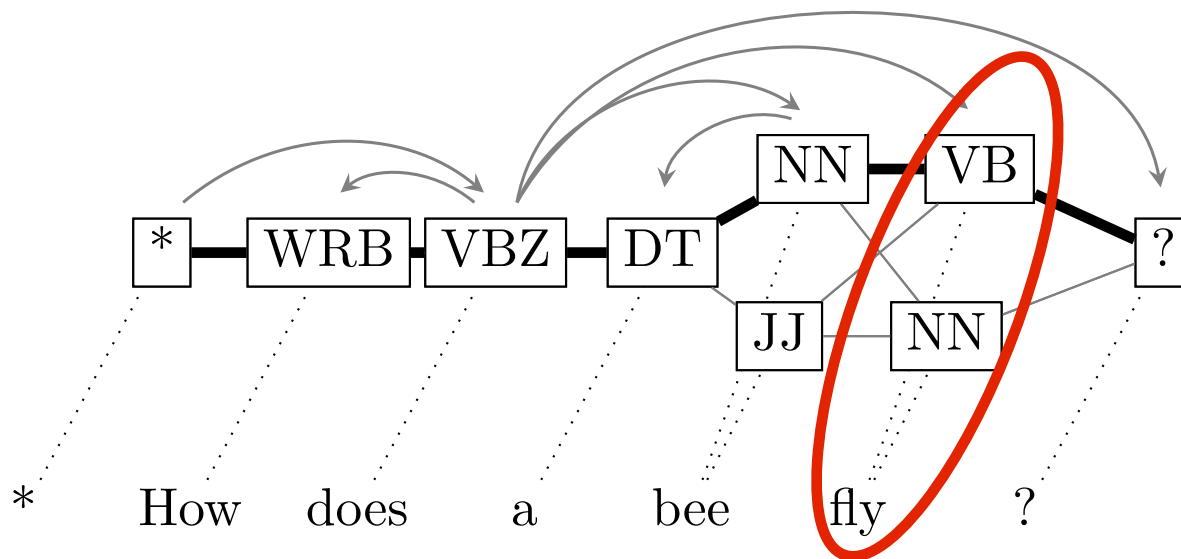


... 'to coach kim'

Integrated Tagging & Parsing



- Sequential Markov Models have limitations:

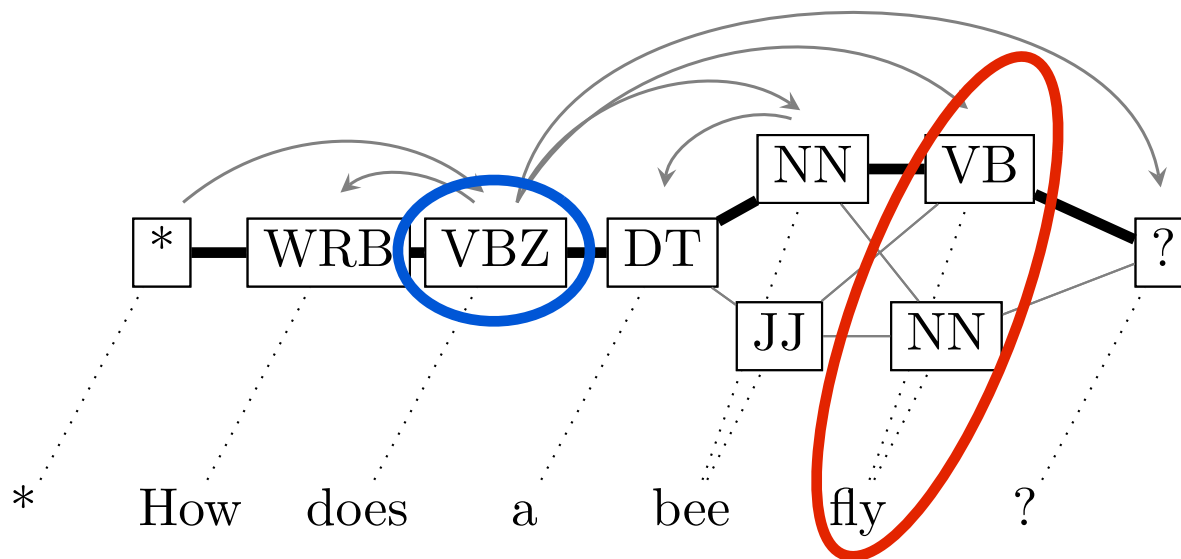


... 'to coach kim'

Integrated Tagging & Parsing



- Sequential Markov Models have limitations:

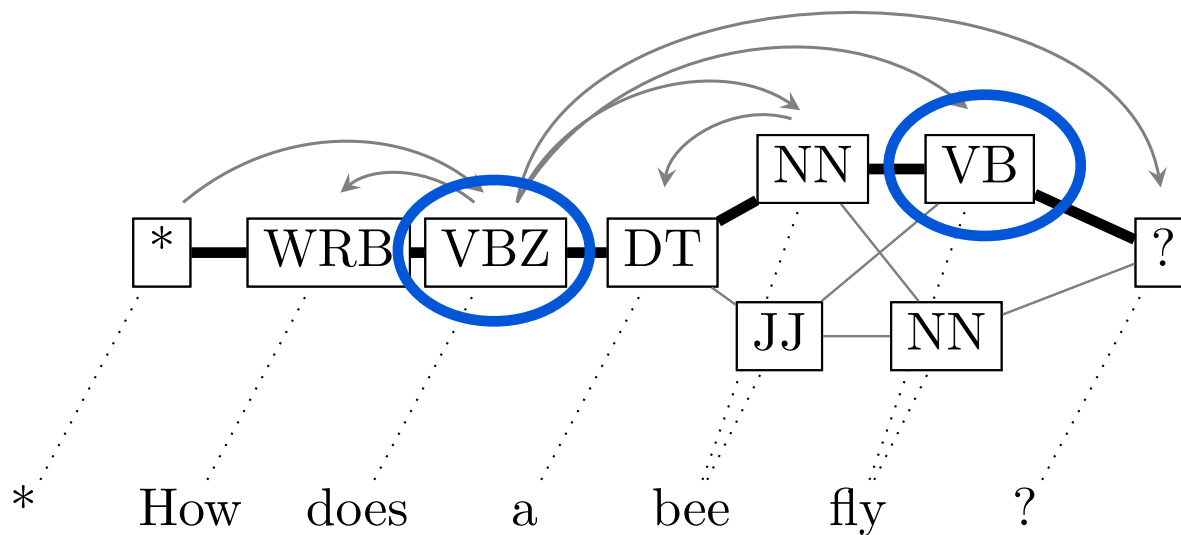


... 'to coach kim'

Integrated Tagging & Parsing



- Sequential Markov Models have limitations:

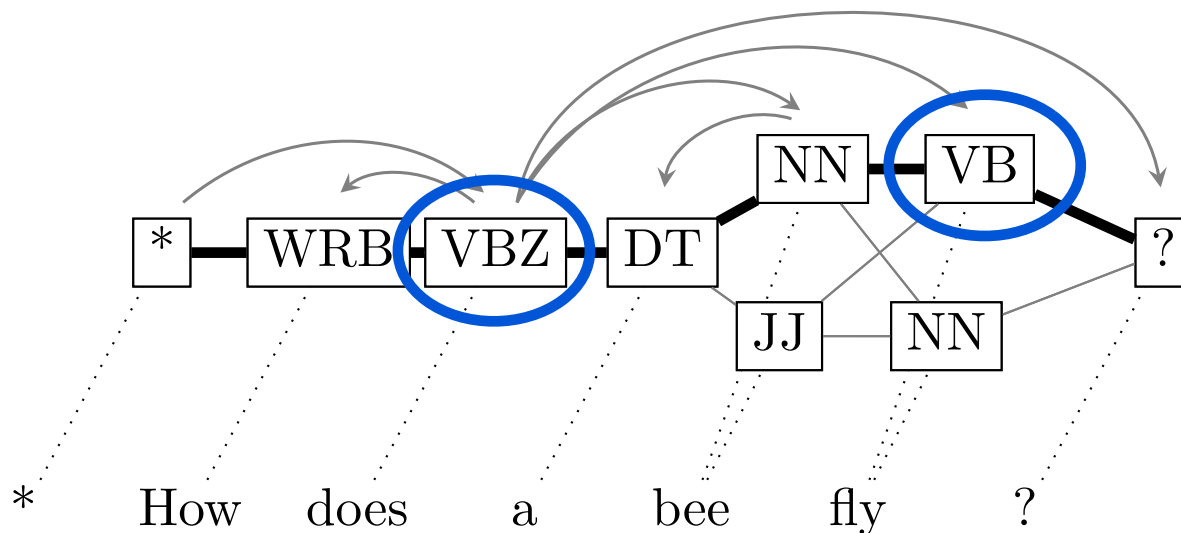


... 'to coach kim'

Integrated Tagging & Parsing



- Sequential Markov Models have limitations:



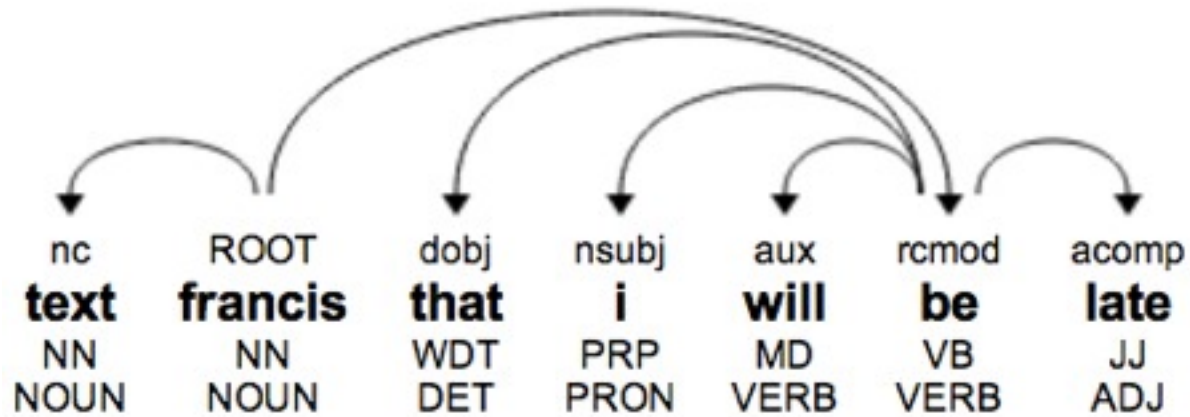
... 'to coach kim'

- Ongoing work:
Search jointly over POS tags and parse trees.

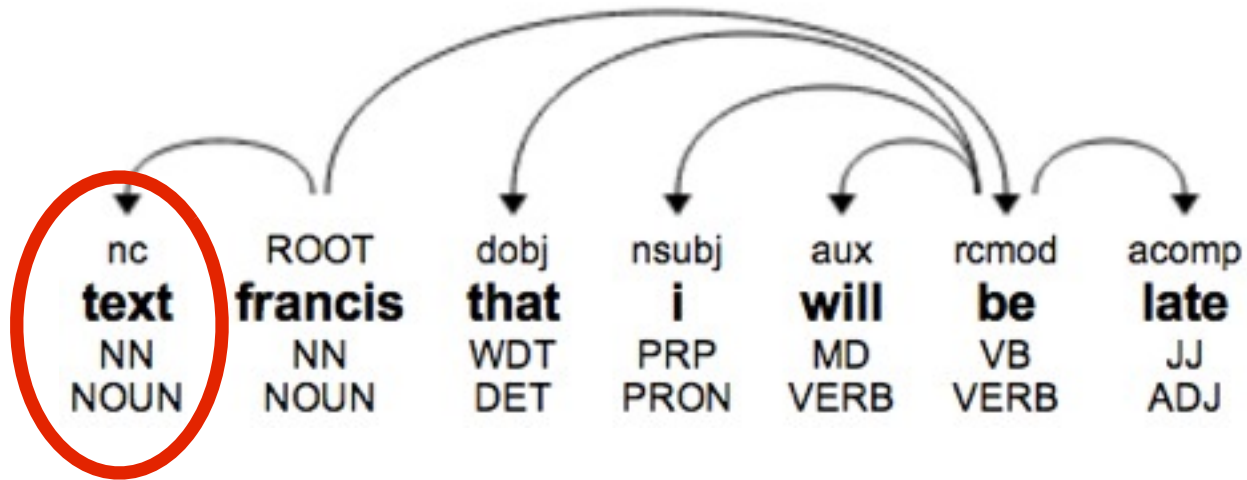
Manual Intervention



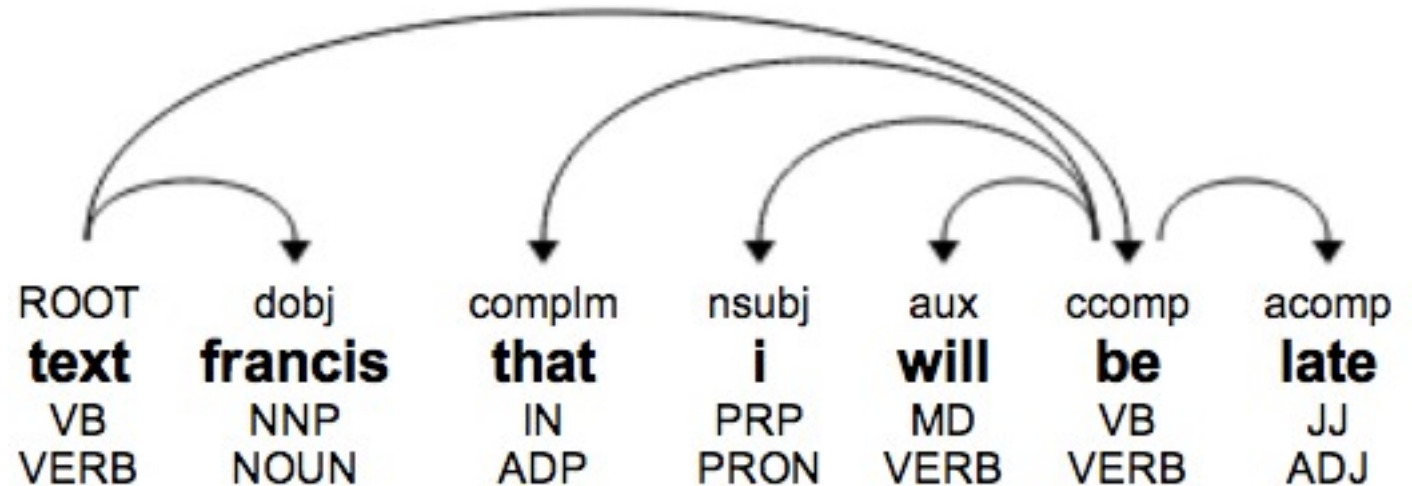
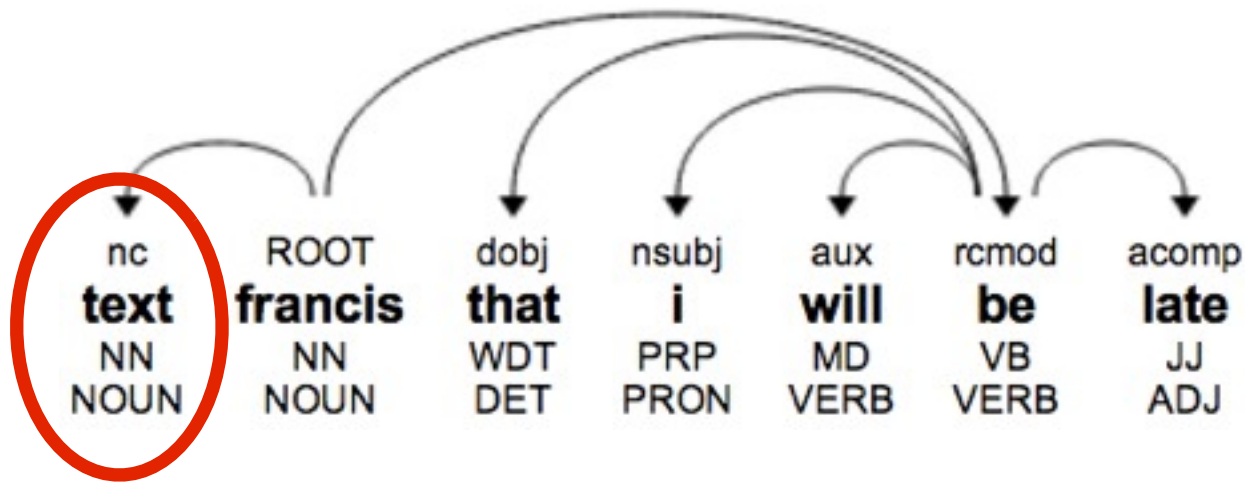
Manual Intervention



Manual Intervention



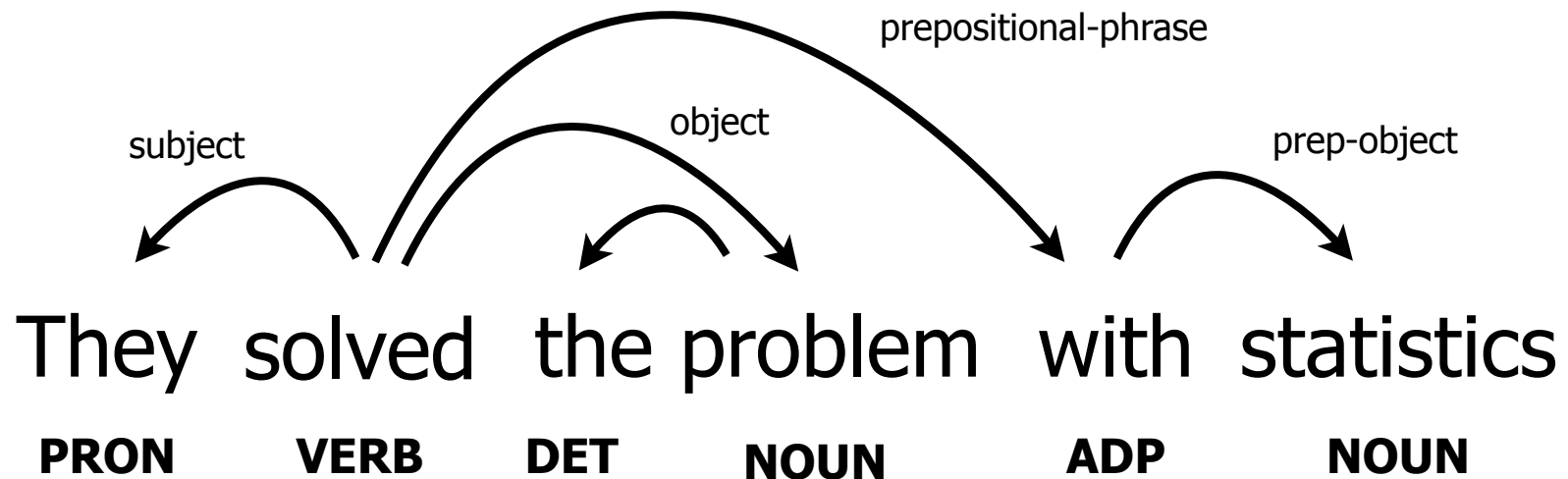
Manual Intervention



Machine Translation Reordering



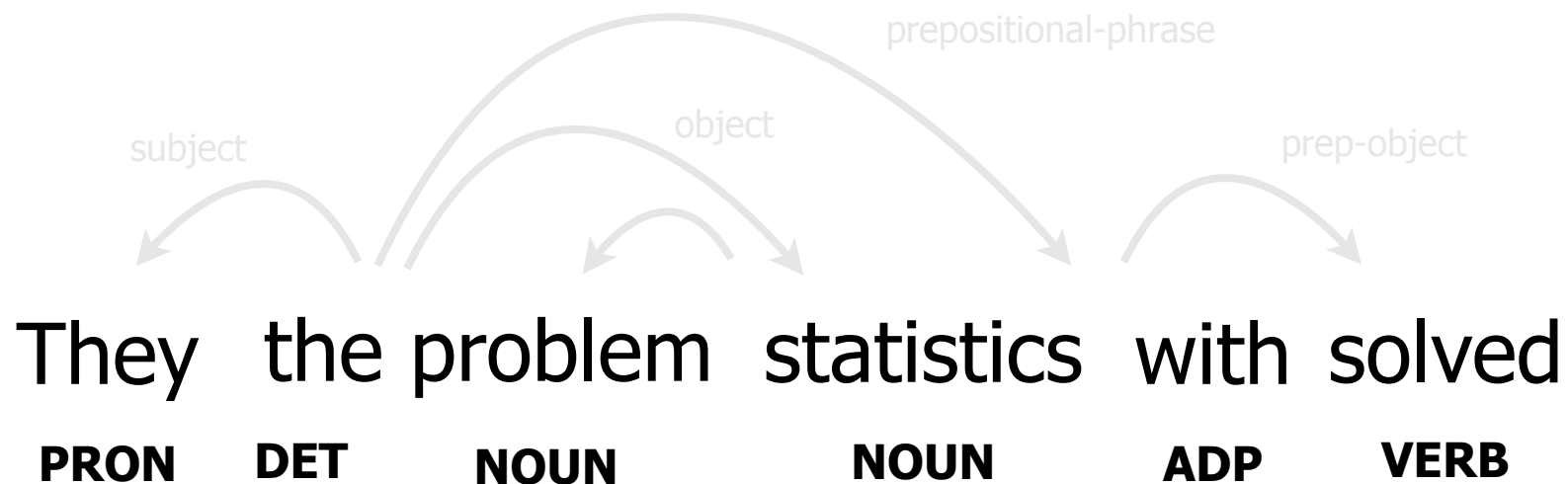
- Source-side reordering for Japanese-English MT
[Collins et al. '05]
- Dependency-based reordering for English-Japanese
[Xu et al. '09]



Machine Translation Reordering



- Source-side reordering for Japanese-English MT
[Collins et al. '05]
- Dependency-based reordering for English-Japanese
[Xu et al. '09]



Reordering Score



[Talbot et al. '11]

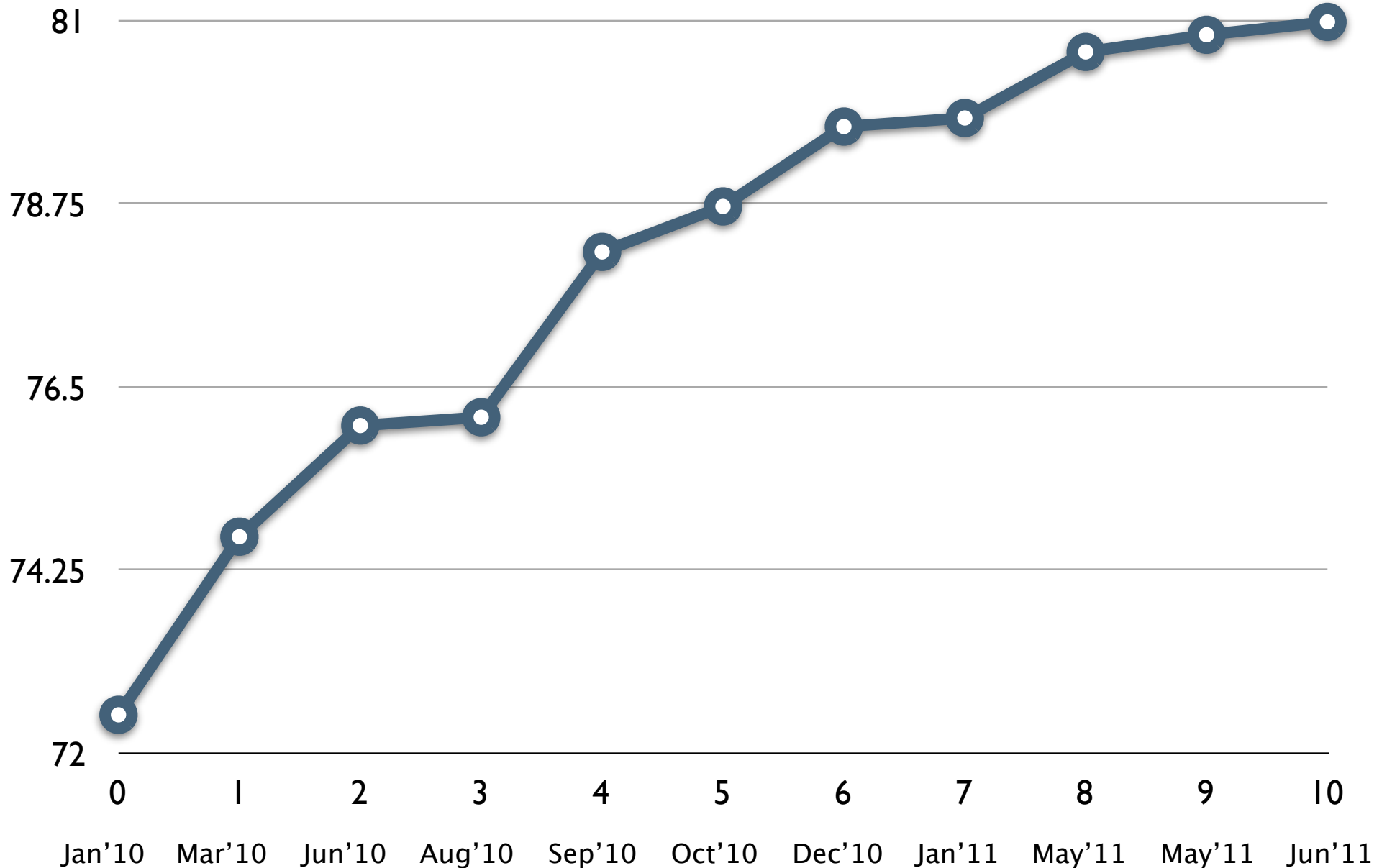
- Source-side reordering for Japanese-English MT
[Collins et al. '05]
- Hand generated reordering data (English + Jenglish)
 - ~ 10k sentences for Augmented Loss training
 - ~ 6k evaluation sentences
- Score based on reordering penalty of METEOR

$$\text{reorder-score} = \frac{\# \text{ chunks} - 1}{\# \text{ unigrams matched} - 1}$$

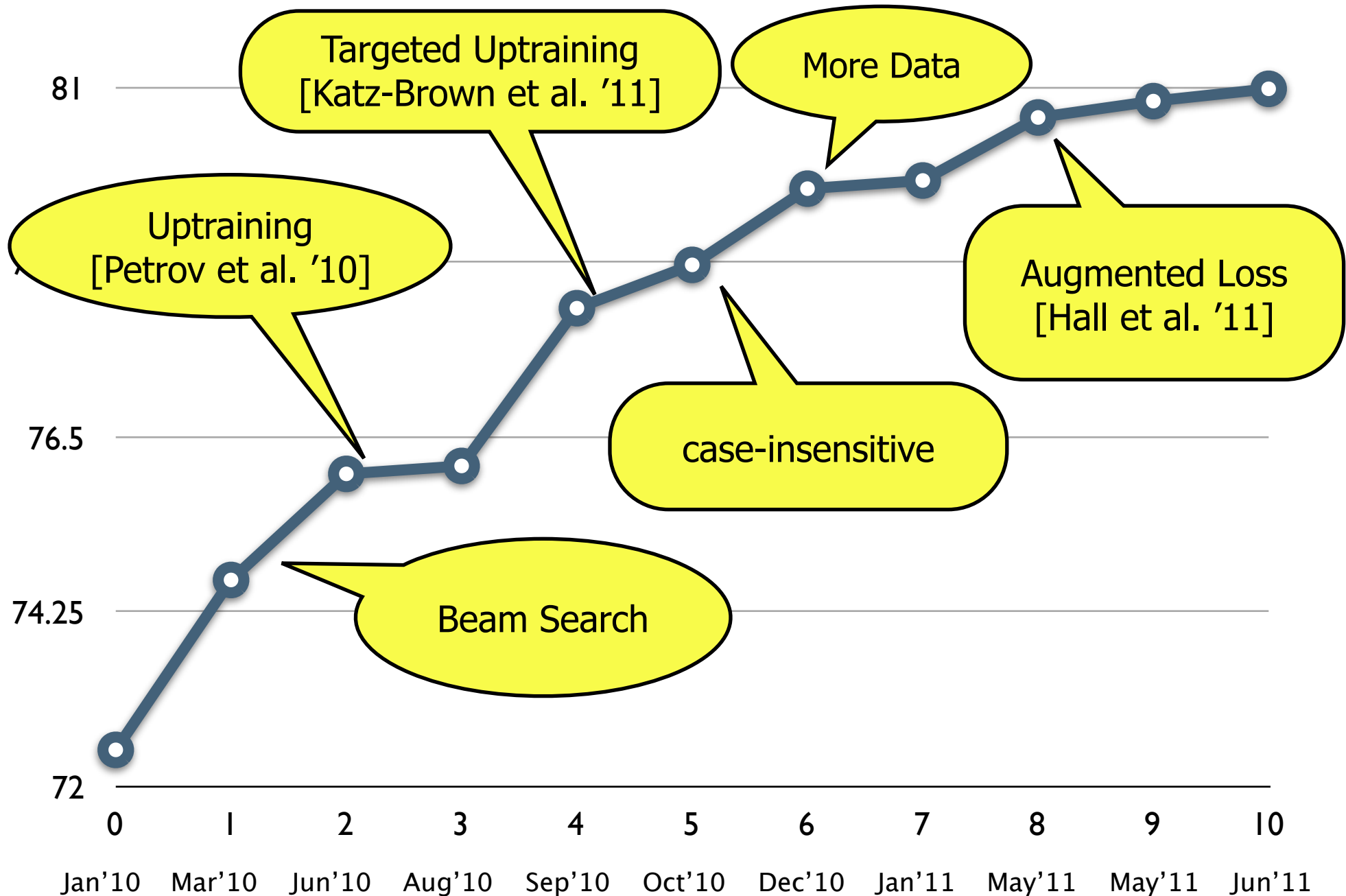
$$\text{reorder-loss} = 1 - \text{reorder-score}$$

- Very well correlated with human eval scores.

EnJa Fuzzy Reordering Evolution



EnJa Fuzzy Reordering Evolution





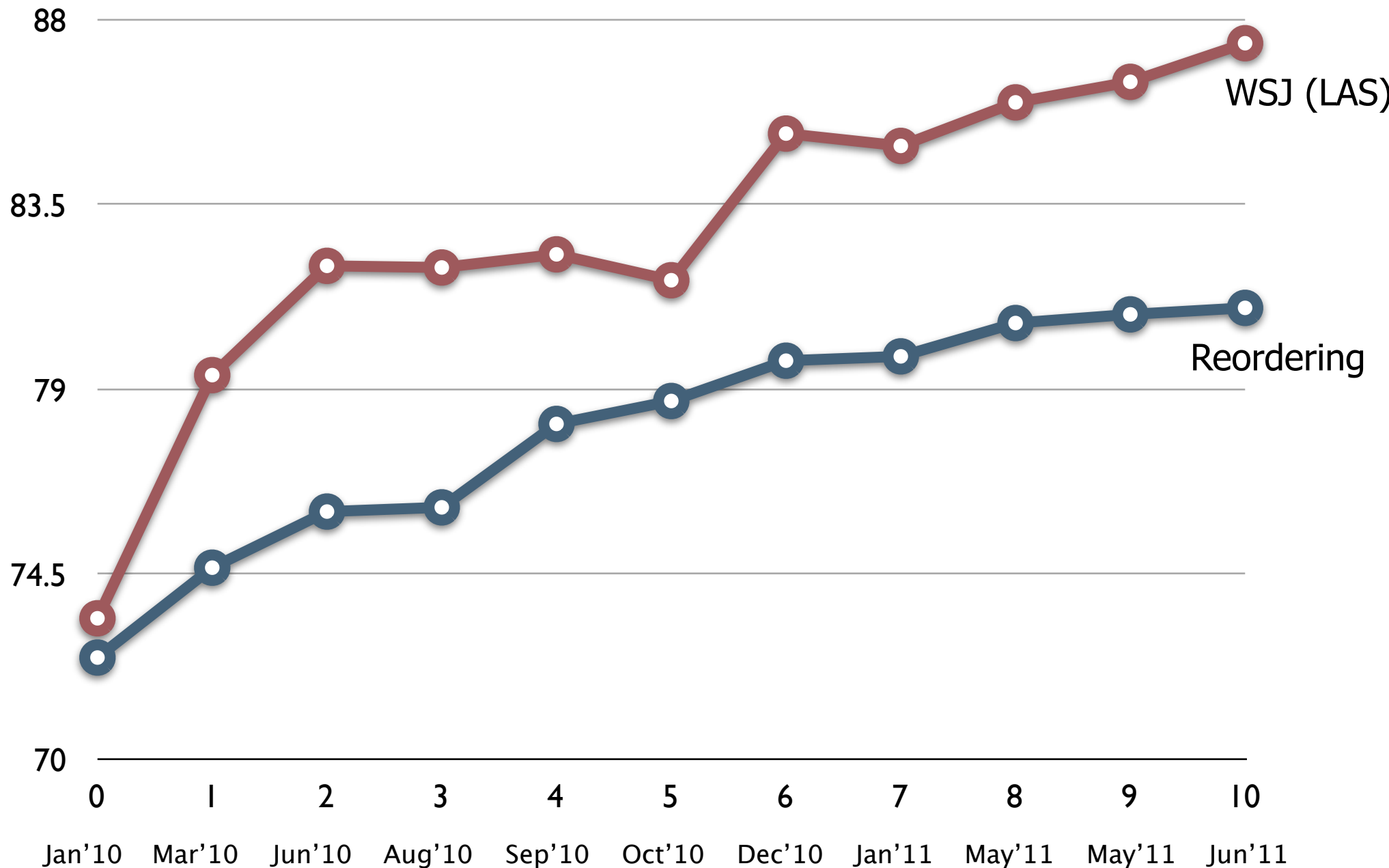
EnJa Fuzzy Reordering Evolution

WSJ (LAS)

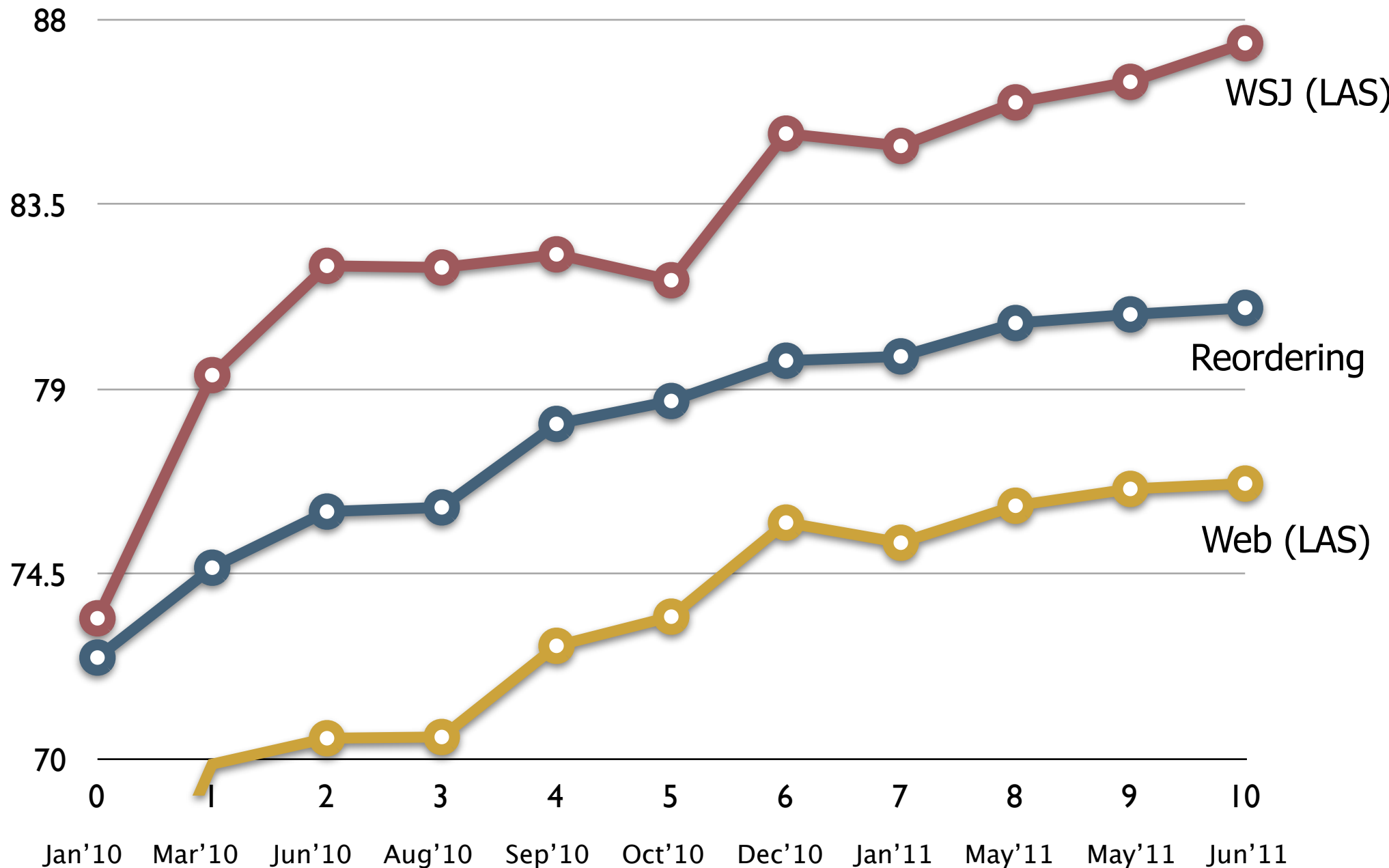
Reordering

Jan'10 Mar'10 Jun'10 Aug'10 Sep'10 Oct'10 Dec'10 Jan'11 May'11 May'11 Jun'11

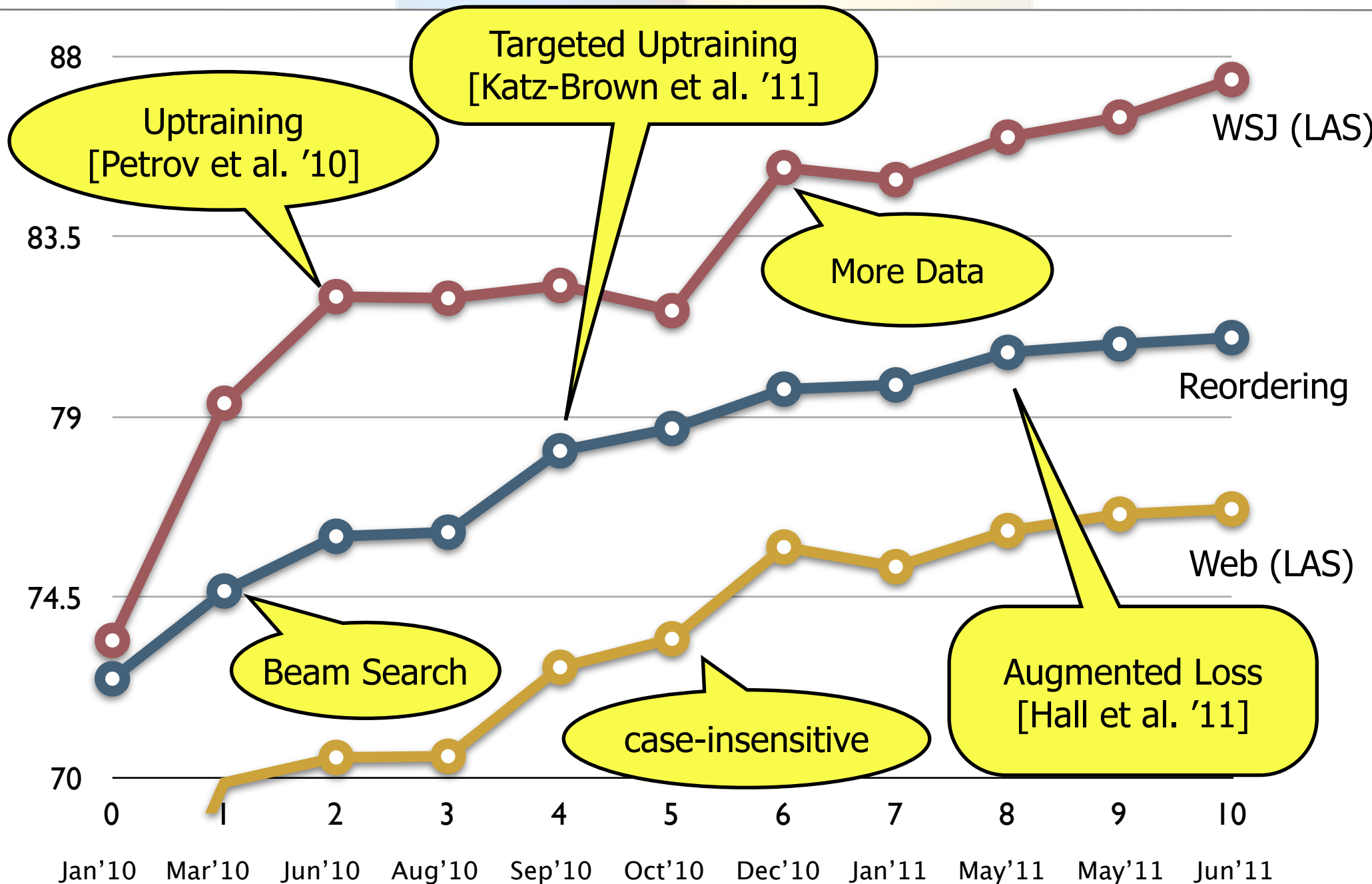
EnJa Fuzzy Reordering Evolution



EnJa Fuzzy Reordering Evolution



EnJa Fuzzy Reordering Evolution



Retrospective Analysis



final parser

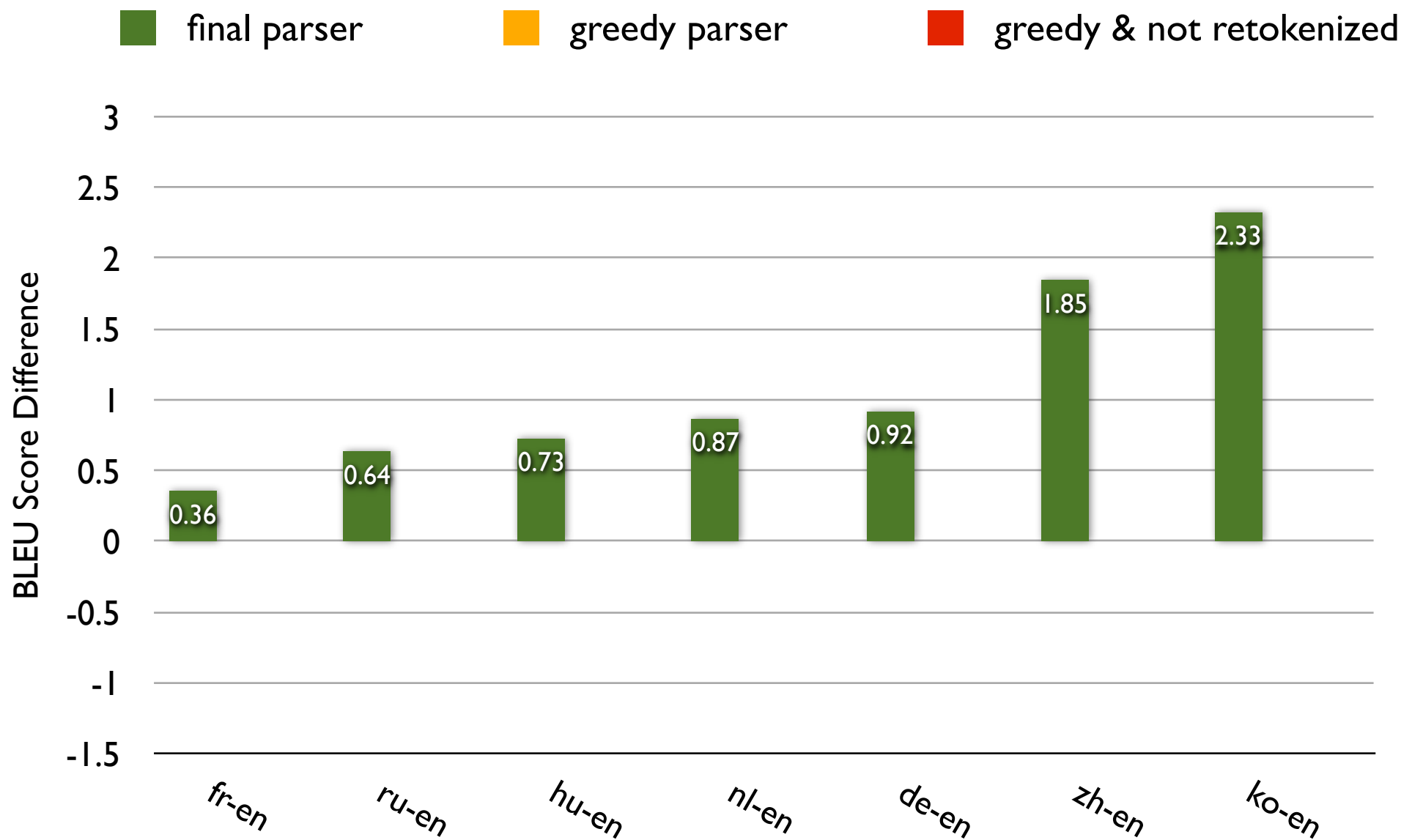


greedy parser

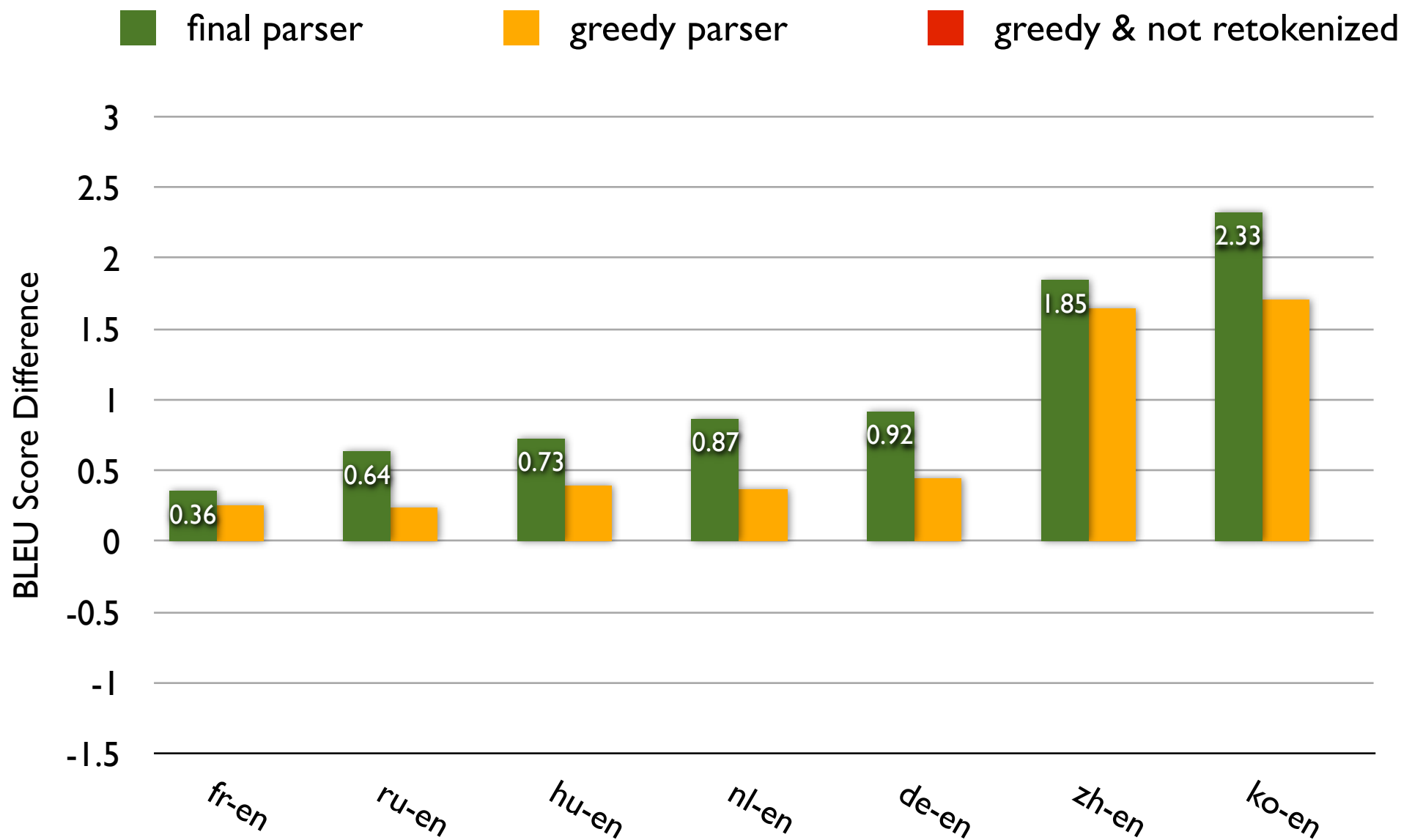


greedy & not retokenized

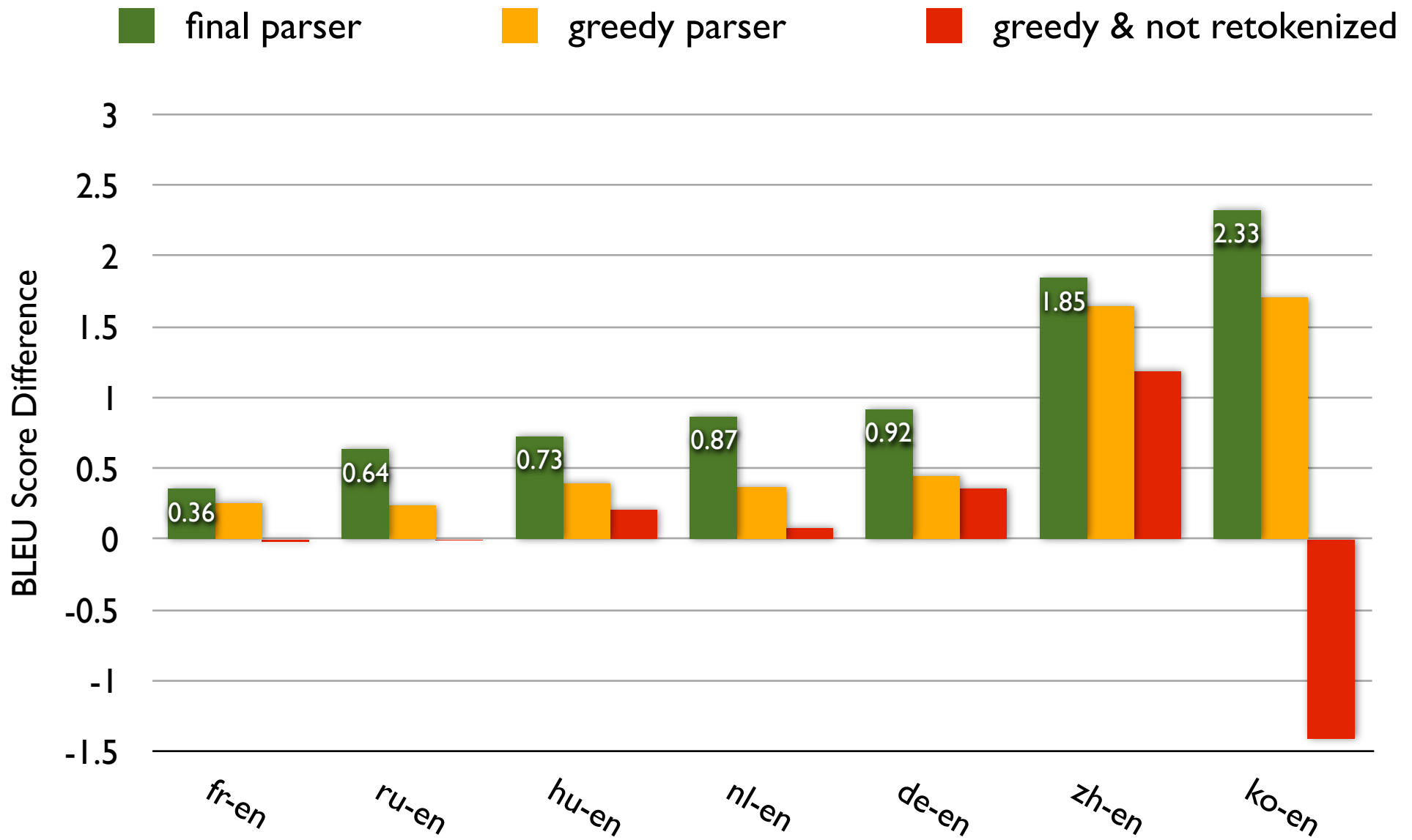
Retrospective Analysis



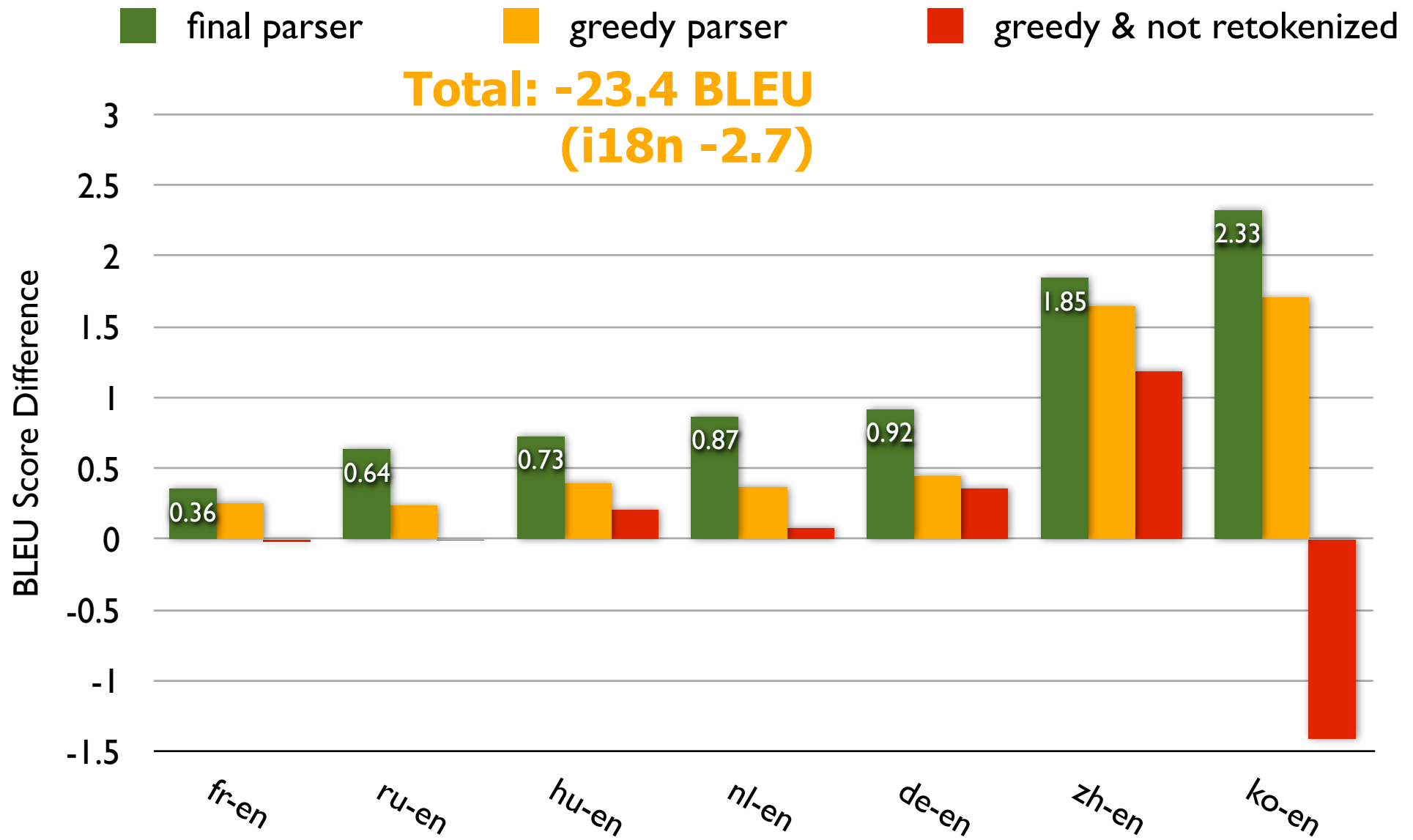
Retrospective Analysis



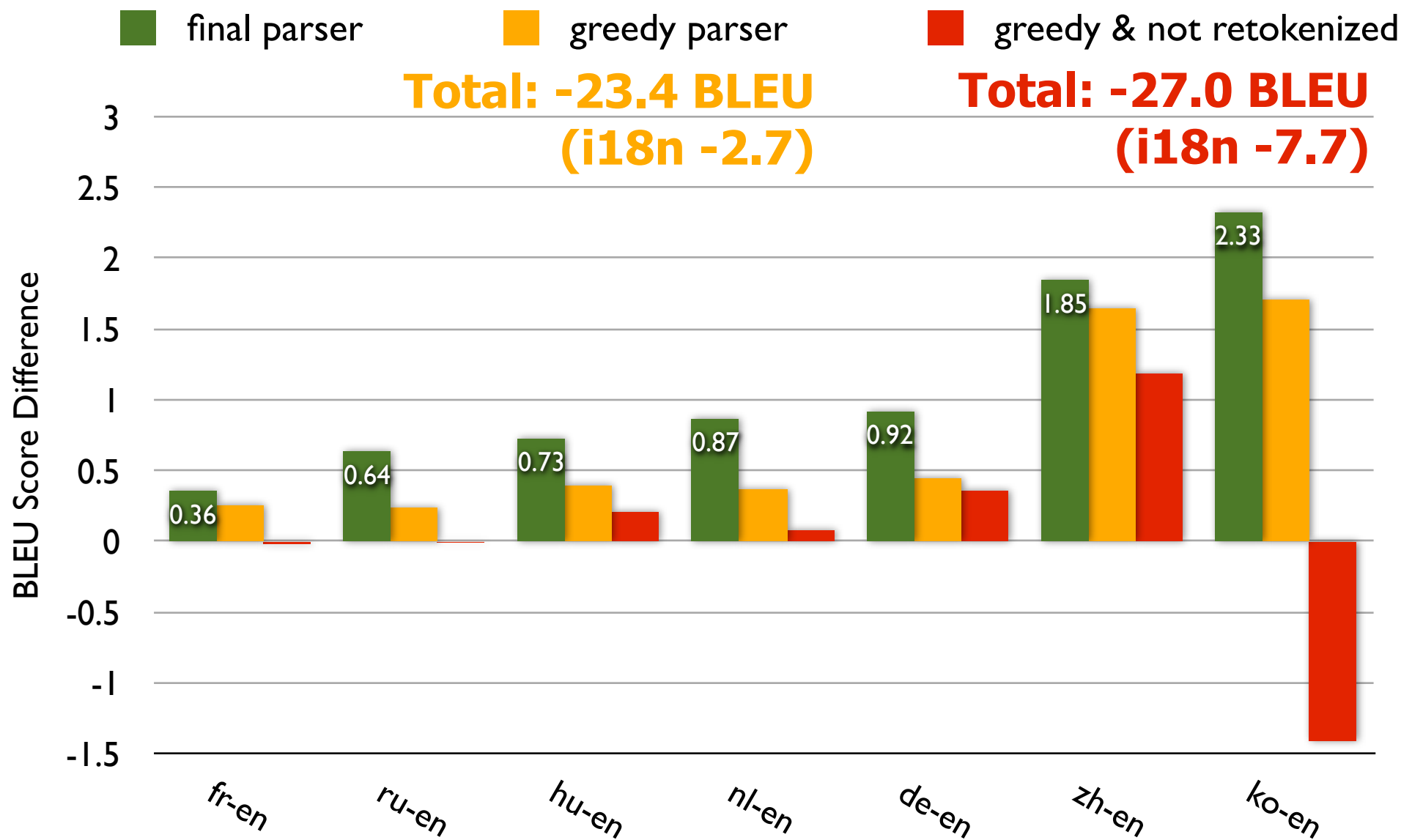
Retrospective Analysis



Retrospective Analysis



Retrospective Analysis



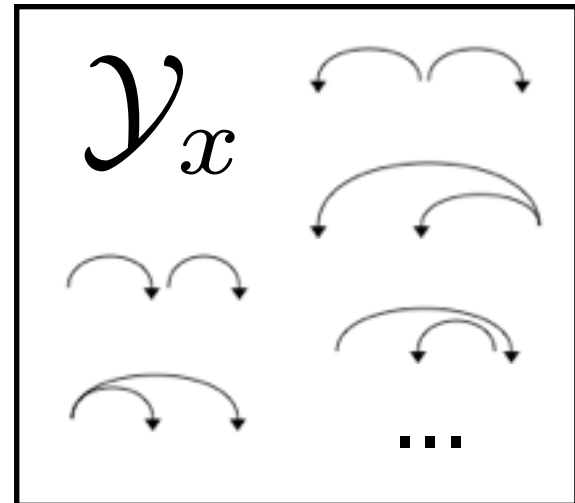
Standard Perceptron Training

Standard Perceptron

x = John likes Mary

y^* =

nsubj	ROOT	dobj
John	likes	Mary
NNP	VBZ	NNP
NOUN	VERB	NOUN



$$y = \arg \max_{y \in \mathcal{Y}_x} w \cdot \phi(y)$$

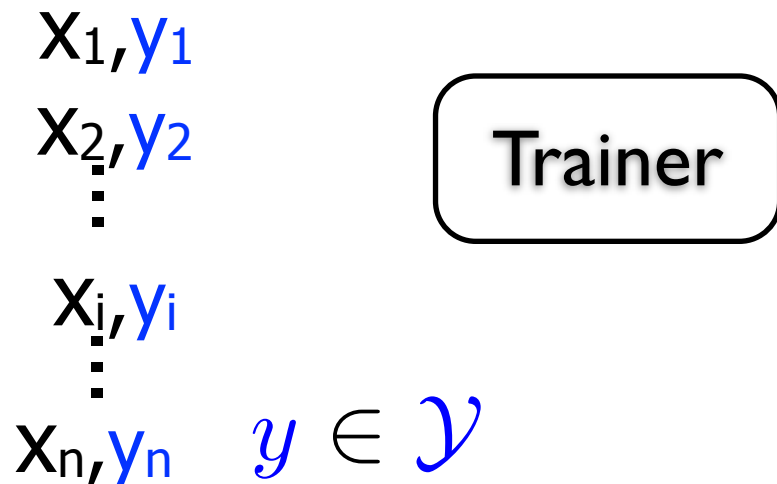
$$\theta = \theta + \phi(y^*) - \phi(y)$$

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data



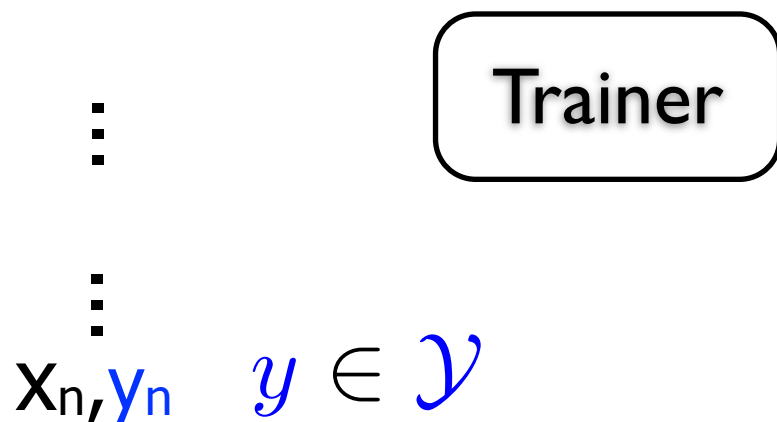
Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data



Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data

x_1, y_1

x_2, y_2

$\vdots$

x_i, y_i

$\vdots$

x_n, y_n

$y \in \mathcal{Y}$

Trainer

Extrinsic Data

x'_1, y'_1

x'_2, y'_2

$\vdots$

x'_i, y'_i

$\vdots$

x'_n, y'_n

$y' \in \mathcal{Y}'$

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data

x_1, y_1

x_2, y_2

$\vdots$

x_i, y_i

$\vdots$

x_n, y_n

$y \in \mathcal{Y}$

Trainer

Extrinsic Data

x'_1, y'_1

x'_2, y'_2

$\vdots$

x'_i, y'_i

$\vdots$

x'_n, y'_n

$y' \in \mathcal{Y}'$

$l(\hat{y}_i, y'_i) \quad \hat{y} \in \mathcal{Y}$

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data

x_1, y_1

x_2, y_2

$\vdots$

x_i, y_i

$\vdots$

x_n, y_n

$y \in \mathcal{Y}$

Trainer

Extrinsic Data

x'_1, y'_1

x'_2, y'_2

$\vdots$

x'_i, y'_i

$\vdots$

x'_n, y'_n

$y' \in \mathcal{Y}'$

$l(\hat{y}_i, y'_i) \quad \hat{y} \in \mathcal{Y}$

$\mathcal{Y} \neq \mathcal{Y}'$

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data

$\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$y \in \mathcal{Y}$

Trainer

Extrinsic Data

x'_1, y'_1

x'_2, y'_2

$\vdots$
 x'_i, y'_i
 $\vdots$

x'_n, y'_n

$l(\hat{y}_i, y'_i) \quad \hat{y} \in \mathcal{Y}$

$y' \in \mathcal{Y}'$

$\mathcal{Y} \neq \mathcal{Y}'$

Similar to [Chang et al. '08, McAllester et al. '10]

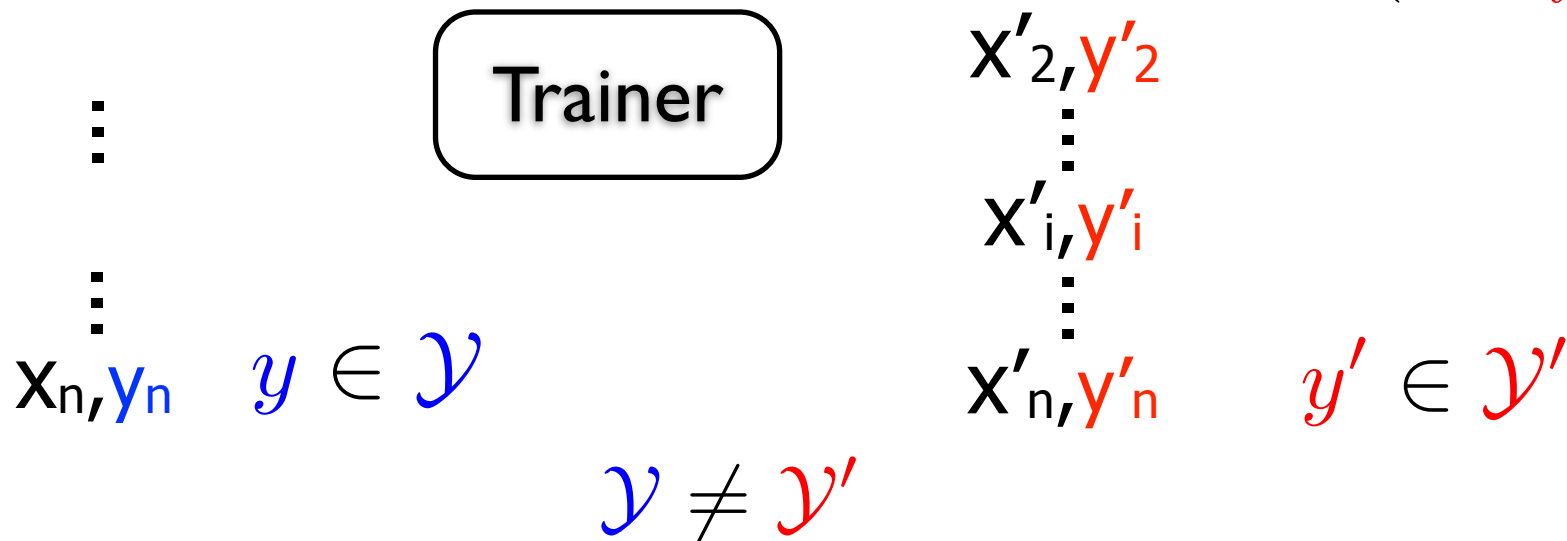
Augmented Loss (Online) Training



[Hall et al. '11]

Intrinsic Data

Extrinsic Data



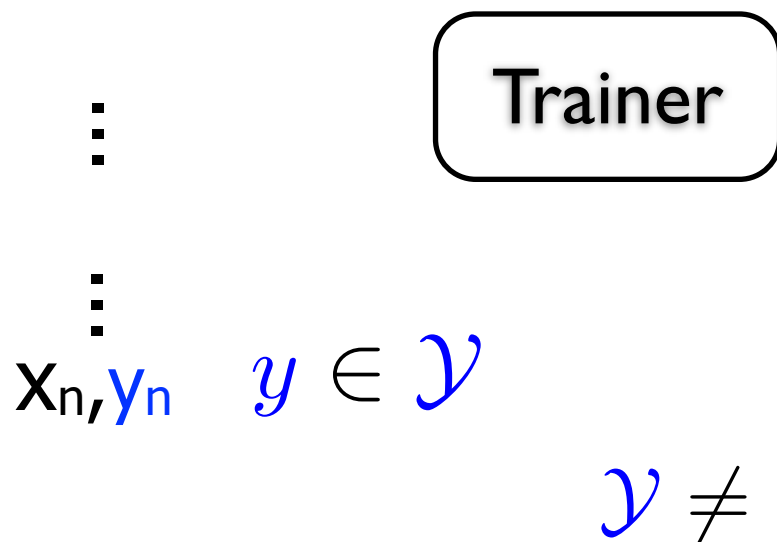
Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss (Online) Training



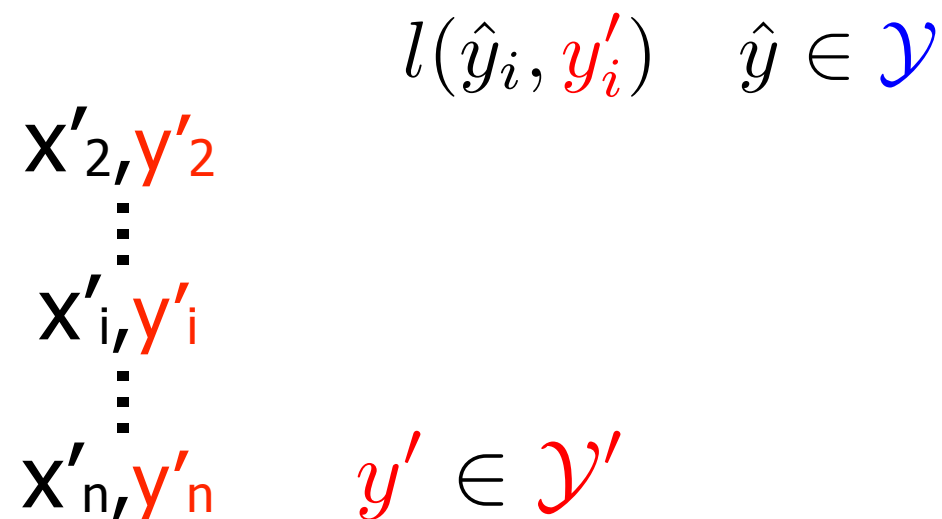
[Hall et al. '11]

Intrinsic Data



Intrinsic: supervised training data/objective

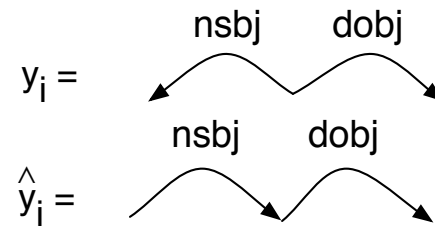
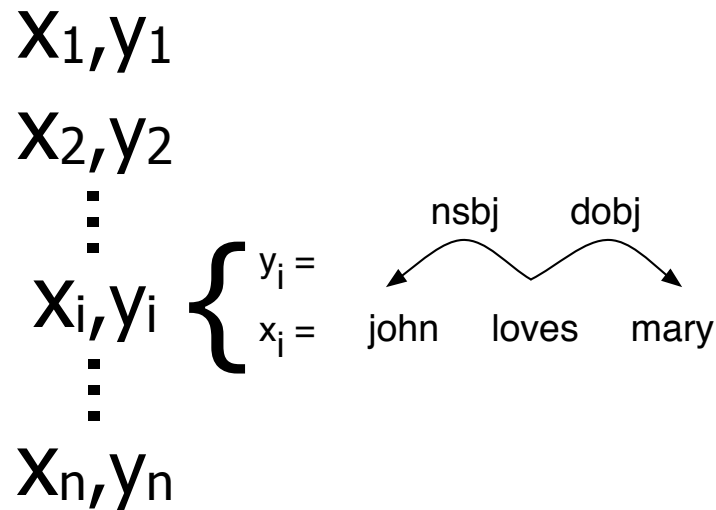
Extrinsic Data



Extrinsic: downstream task which "uses" outputs of model

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Extrinsic
Data

x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 x'_i, y'_i
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



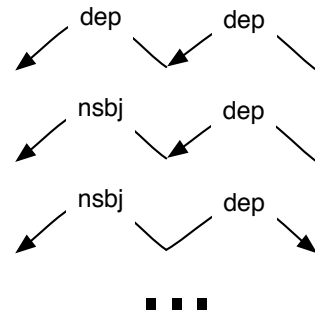
x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Parse(x'_i) {



Extrinsic
Data

x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



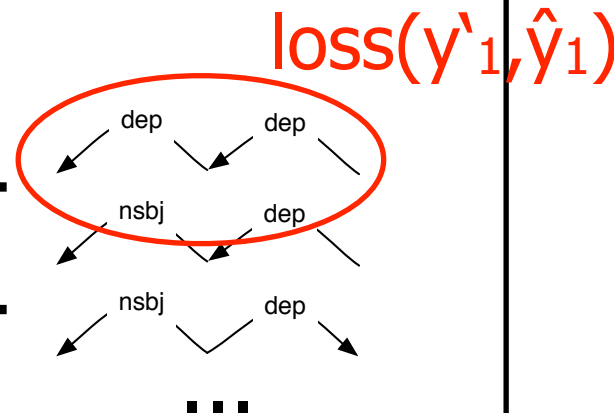
x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Parse(x'_i) {



Extrinsic
Data

x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



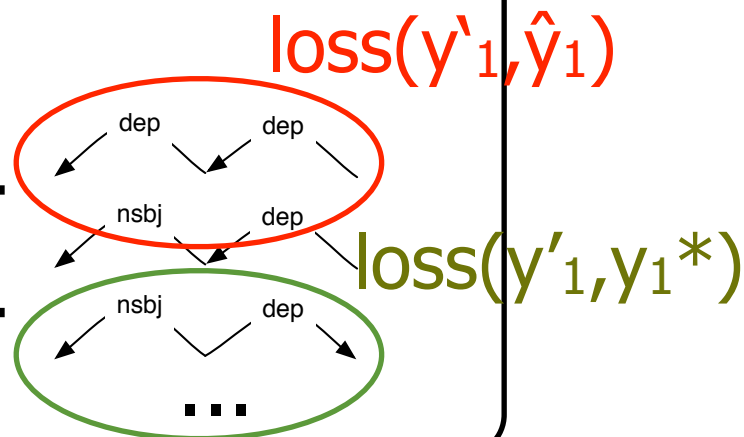
x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

Parse(x'_i) {



Extrinsic
Data

x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_{x'_i}} \theta \cdot \phi(x'_i, y)$$

$$y_i^* = \arg \min_{y \in \text{k-best-parse}(x'_i)} \text{loss}(y, y'_i)$$

if $\text{loss}(y_i^*, y'_i) < \text{loss}(\hat{y}_i, y'_i)$ then

$$\theta^t = \theta^{t-1} + \phi(y_i^*) - \phi(\hat{y}_i)$$

Extrinsic
Data

x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Augmented Loss Perceptron



x_1, y_1
 x_2, y_2
 $\vdots$
 x_i, y_i
 $\vdots$
 x_n, y_n

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_x} \theta \cdot \phi(x_i, y)$$

if $y_i \neq \hat{y}_i$ then

$$\theta^t = \theta^{t-1} + \phi(y_i) - \phi(\hat{y}_i)$$

$$\hat{y}_i = \arg \max_{y \in \mathcal{Y}_{x'_i}} \theta \cdot \phi(x'_i, y)$$

$$y_i^* = \arg \min_{y \in \text{k-best-parse}(x'_i)} \text{loss}(y, y'_i)$$

if $\text{loss}(y_i^*, y'_i) < \text{loss}(\hat{y}_i, y'_i)$ then

$$\theta^t = \theta^{t-1} + \phi(y_i^*) - \phi(\hat{y}_i)$$

Extrinsic
Data

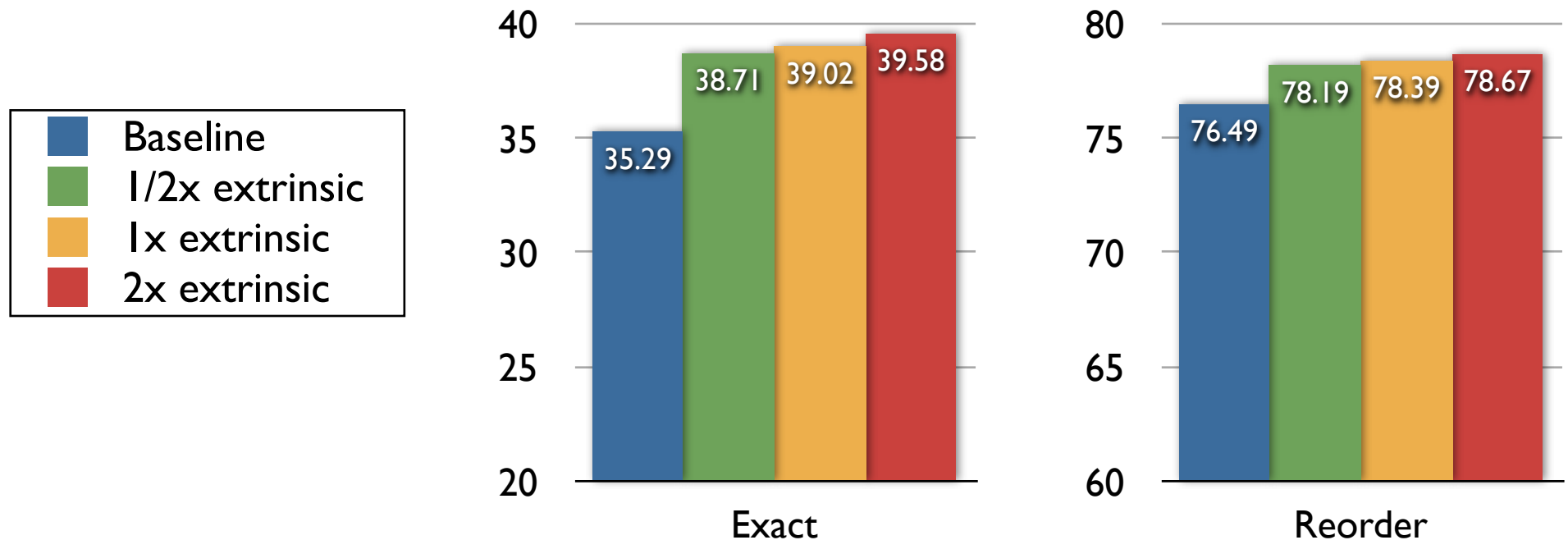
x'_1, y'_1
 x'_2, y'_2
 $\vdots$
 $\vdots$
 x'_n, y'_n

Similar to [Chang et al. '08, McAllester et al. '10]

Example MT-Reordering



“Clean” experiment



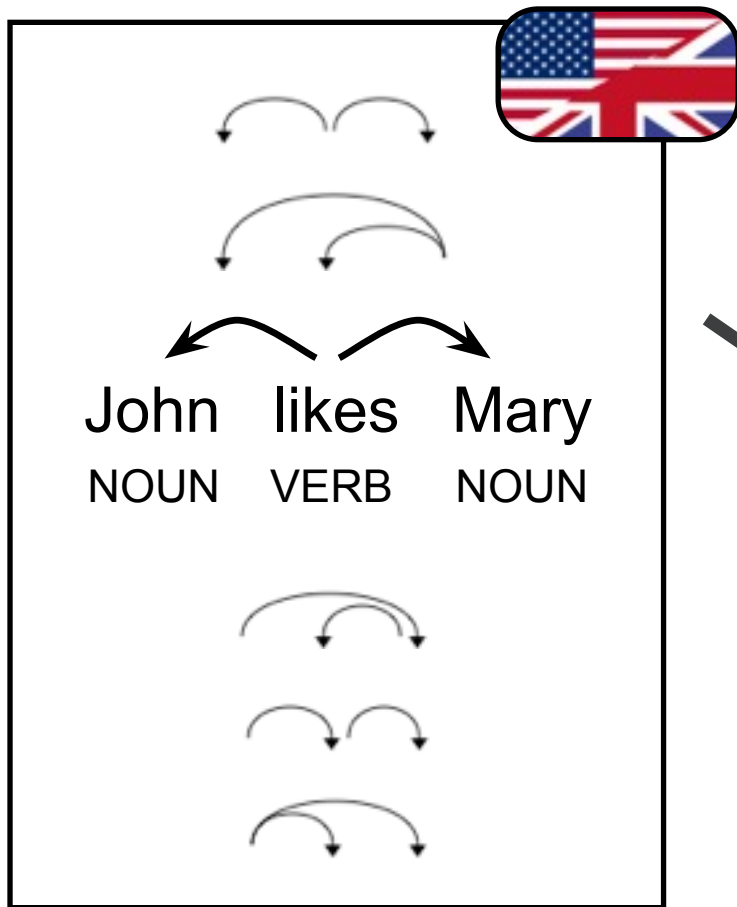
Production: 0.786 -> 0.792 enja fuzzy score

On top of targeted up/self-training [Katz-Brown et al. '11]

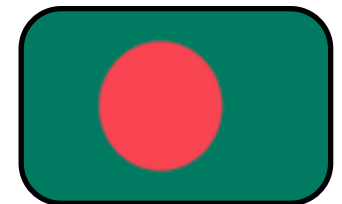
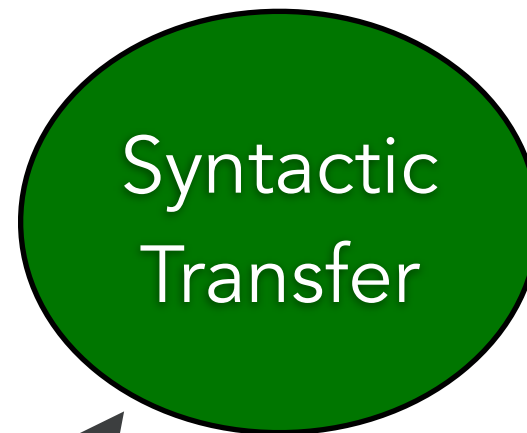
Syntactic Transfer



English Treebank



- Learn parsers for resource-poor languages from resource-rich languages
- Hwa et al. 2005 and earlier



...



Dictionaries:



Confidence Estimation



how tall is the eiffel tower

Google

how tall is the eiffel tower

Web Images Maps Shopping Videos More Search tools

About 2,430,000 results (0.57 seconds)

1,063' (324 m)

Eiffel Tower, Height

[Eiffel Tower - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/wiki/Eiffel_Tower](#) 
 Further information: List of tallest towers in the world and Observation deck ...
[Gustave Eiffel - List of Eiffel Tower replicas](#) - Category:Eiffel Tower

[How Tall Is The Eiffel Tower? | Things To Learn](#)
[www.thingstolearn.net/2012/06/how-tall-is-the-eiffel-tower/](#) 
 The Answer is: 1,063 feet tall, or 324 meters (taller than the Washington Monument).
 The Eiffel Tower is the tallest Structure in Paris, France. This structure has ...

[How tall is the Eiffel Tower - WikiAnswers](#)
[wiki.answers.com](#) > ... > France > Paris > Eiffel Tower 
 This answer is closed to changes. This is done in rare cases when questions are being
 vandalized or answers have become debates. E-mail reopen ...

[Eiffel Tower](#)
[www.tour-eiffel.fr/](#)  Translate this page
 9h30 à 23h00 - 9h00 à 00h00 l'été. Calc Préparer sa visite. Calc Explorer la tour Eiffel
 Calc A savoir sur la tour Eiffel Calc L'actu. Calc Professionnels
 Score: **26 / 30** - **2,283 Google reviews** - [Write a review](#) - \$208 

 Champ de Mars 5 Avenue Anatole France, 75007 Paris, France
 +33 892 70 12 39

[How tall is the Eiffel Tower | Paris Attractions](#)




Eiffel Tower

Observation tower in Paris, France

[Directions](#)

The Eiffel Tower is an iron lattice tower located on the Champ de Mars in Paris, named after the engineer Gustave Eiffel, whose company designed and built the tower. [Wikipedia](#)

Height: 1,063' (324 m)

Construction started: January 28, 1887

Floors: 3

Address: Champ de Mars, 5 Avenue Anatole France, 75007 Paris, France

Architect: [Stephen Sauvestre](#)

Architecture firms: [Eiffel & Cie](#), [Barbier, Benard and Turenne](#)

People also search for







Confidence Estimation



- Sometimes it is possible to say: “I don’t know”
- Ten blue links are better than triggering an incorrect answer
- The (English) web is redundant:
 - Skip examples with low confidence predictions
 - Discount low-confidence contradicting predictions

In Summary



- Efficiency:
 - Exploit problem structure and domain knowledge
 - Train models specifically for pruning
- Adaptation:
 - Use unsupervised data
 - Use indirect signals
 - “Manual intervention”
- Confidence:
 - ?

Learning from Indirect Signals



Learning from Indirect Signals



Labeled Data



Learning from Indirect Signals



Labeled Data



Unlabeled Data

Learning from Indirect Signals



Structured Data



Labeled Data



Unlabeled Data



Learning from Indirect Signals



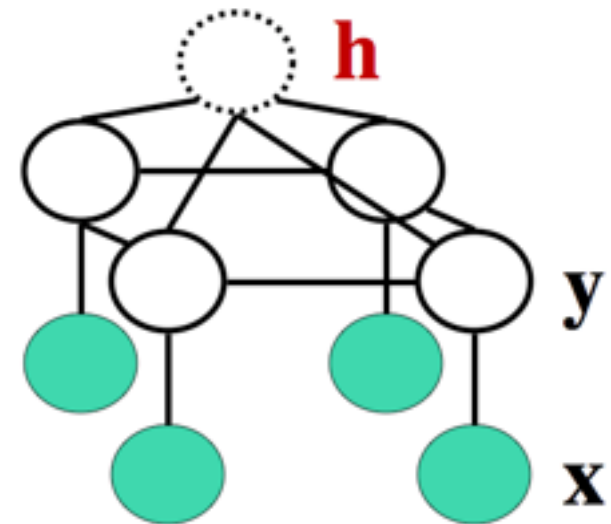
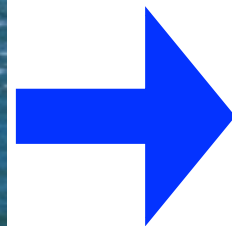
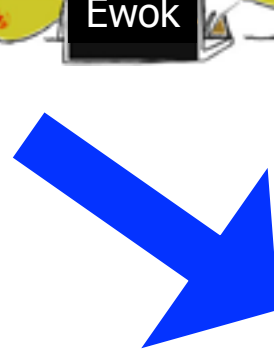
Structured Data



Labeled Data



Unlabeled Data

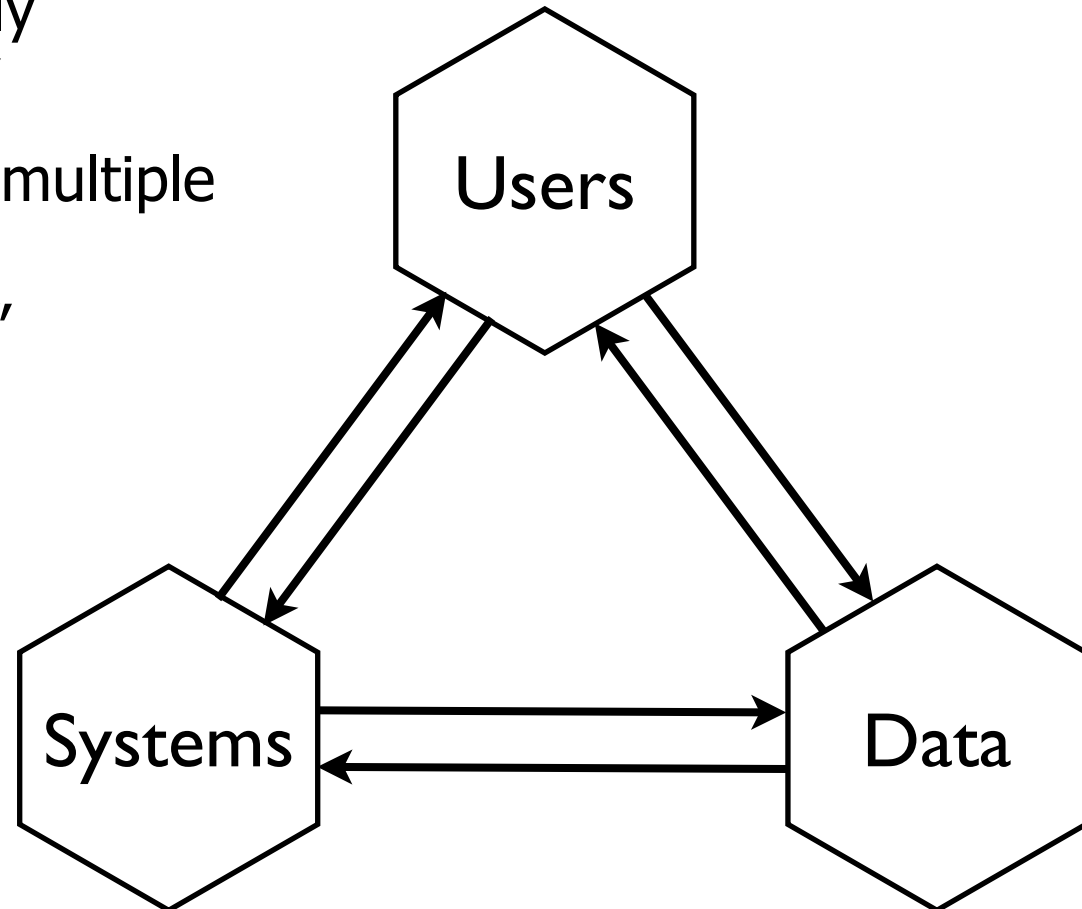


Learning & Modeling

Hypothesis



- Understanding arises from machine learning of relationships implicit in web content and use
 - Some expert annotation may be needed to start
 - Most evidence is not explicitly annotated: text “in the wild”
 - Aggregate information from multiple unstructured sources into a structured “knowledge base”
 - Exploit user interactions and implicit user feedback



[Based on slide from Fernando Pereira]

Thank you!



Research
at Google

slav@google.com