Error-Driven Analysis of Challenges in Coreference Resolution

Jonathan K. Kummerfeld and Dan Klein
University of California, Berkeley
Today President Bush unveiled a seven point one billion dollar . . . 
. . . drawing Al Gore and George W. Bush to the Sunshine State . . . 
. . . nonpartisan analysis confirms George Bush 's plan sets aside . . . 
. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush 's plan sets aside . . .
. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar ... ... drawing Al Gore and George W. Bush to the Sunshine State ... ... nonpartisan analysis confirms George Bush's plan sets aside ... ... Jeb Bush was once expected to deliver his brother this state ...
Today President Bush unveiled a seven point one billion dollar . . .

. . . drawing Al Gore and George W. Bush to the Sunshine State . . .

. . . nonpartisan analysis confirms George Bush 's plan sets aside . . .

. . . Jeb Bush was once expected to deliver his brother this state . . .

Coreference
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush 's plan sets aside . . .
. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush 's plan sets aside . . .
. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar ... ... drawing Al Gore and George W. Bush to the Sunshine State ... ... nonpartisan analysis confirms George Bush 's plan sets aside ... ... Jeb Bush was once expected to deliver his brother this state ...
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush 's plan sets aside . . .
. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar ... drawing Al Gore and George W. Bush to the Sunshine State ... nonpartisan analysis confirms George Bush's plan sets aside ... Jeb Bush was once expected to deliver his brother this state ...
Coreference

Or is it checkbook diplomacy?  
...pursuing checkbook diplomacy.
Or is it **checkbook diplomacy**? ...
...pursuing **checkbook diplomacy**.
Coreference

Span Error
Conflated Entities
Extra Entity

Or is it checkbook diplomacy?
...pursuing checkbook diplomacy.
Previous Work
Previous Work

- Reconcile system analysis
  Stoyanov et al. (2009)
Previous Work

- Reconcile system analysis
  Stoyanov et al. (2009)
Our Approach
Today President Bush unveiled a seven point one billion dollar . . .

. . . drawing Al Gore and George W. Bush to the Sunshine State . . .

. . . nonpartisan analysis confirms George Bush’s plan sets aside . . .

. . . Jeb Bush was once expected to deliver his brother this state . . .
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush’s plan sets aside . . .
Jeb Bush was once expected to deliver his brother this state . . .
Our Approach

Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush’s plan sets aside . . .
Jeb Bush was once expected to deliver his brother this state . . .

Transform

Errors
Our Approach

Today President Bush unveiled a seven point one billion dollar... 
... drawing Al Gore and George W. Bush to the Sunshine State... 
... nonpartisan analysis confirms George Bush's plan sets aside... 
Jeb Bush was once expected to deliver his brother this state...
Transformations
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge

George W. Bush
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge

Jeb
Bush
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge

Today
President Bush
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge

President Bush
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span  –  Split  –  Remove  –  Introduce  –  Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

Alter Span  –  Split  –  Remove  –  Introduce  –  Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Transformations

- Alter Span
- Split
- Remove
- Introduce
- Merge
Transformations

Alter Span – Split – Remove – Introduce – Merge
Our Approach

Today President Bush unveiled a seven point one billion dollar... drawing Al Gore and George W. Bush to the Sunshine State... nonpartisan analysis confirms George Bush's plan sets aside... Jeb Bush was once expected to deliver his brother this state...
Today President Bush unveiled a seven point one billion dollar . . .
. . . drawing Al Gore and George W. Bush to the Sunshine State . . .
. . . nonpartisan analysis confirms George Bush's plan sets aside . . .
Jeb Bush was once expected to deliver his brother this state . . .
Classify Operations
Today President Bush unveiled a...
Today President Bush unveiled a...
The most deadly form... The virus could mutate... it’s just a matter of time.
The most deadly form of the virus could mutate... it's just a matter of time.

The most deadly form of the virus could mutate... it's just a matter of time.
The most deadly form... the virus could mutate...
...it's just a matter of time.
Classify Operations

Extra Entity

Or is it checkbook diplomacy?

...pursuing checkbook diplomacy.

Or is it checkbook diplomacy?

...pursuing checkbook diplomacy.
Classify Operations

Extra Entity

Split

Remove

Or is it checkbook diplomacy?
...pursuing checkbook diplomacy.

Or is it checkbook diplomacy?
...pursuing checkbook diplomacy.
Classify Operations

Missing Mention

...usually a search takes ... 
...calling off that search

Why call it off?
Classify Operations

Missing Mention

...usually a search takes...
...calling off that search
Why call it off?

...usually a search takes...
...calling off that search
Why call it off?
Classify Operations

Missing Mention

Introduce

Merge

... usually a search takes ...
... calling off that search
Why call it off ?

... usually a search takes ...
... calling off that search
Why call it off ?
We took a taxi . . .

. . . the car was stopped . . .
We took a taxi . . .

. . . the car was stopped . . .

We took a taxi . . .

. . . the car was stopped . . .
Introduce

Merge

Missing Entity

We took a taxi . . .

. . . the car was stopped . . .

We took a taxi . . .

. . . the car was stopped . . .
Classify Operations

Conflated Entities

George W. Bush
George Bush's plan...
Jeb Bush was once expected to deliver his brother this state...

George W. Bush
George Bush's plan...
Jeb Bush was once expected to deliver his brother this state...
Classify Operations

Conflated Entities

George W. Bush to the ...
George Bush's plan ...
Jeb Bush was once expected to deliver his brother this state ...

George W. Bush to the ...
George Bush's plan ...
Jeb Bush was once expected to deliver his brother this state ...
President Clinton is questioning . . .

. . . and Mr. Clinton praised . . .

“

I want to thank Bill Daley . . .

I think he did a brilliant job . . .
Classify Operations

Divided Entity

President Clinton is questioning . . .

. . . and Mr. Clinton praised . . .

“ I want to thank Bill Daley . . .

I think he did a brilliant job . . .
Classify Operations

Divided Entity

President Clinton is questioning . . .

. . . and Mr. Clinton praised . . .

“ I want to thank Bill Daley . . .

I think he did a brilliant job . . .

Merge

President Clinton is questioning . . .

. . . and Mr. Clinton praised . . .

“ I want to thank Bill Daley . . .

I think he did a brilliant job . . .
Error Types

- Span Error
- Extra Mention
- Extra Entity
- Conflated Entities
- Missing Mention
- Missing Entity
- Divided Entity
Results
Results

- System comparison
Results

- System comparison
Results

- System comparison
- Investigation of common errors
System Comparison
## System Comparison

<table>
<thead>
<tr>
<th>System</th>
<th>Metric $F_1$</th>
<th>$\text{MUC}$</th>
<th>$\beta^1$</th>
<th>Span Conflated</th>
<th>Extra</th>
<th>Extra Divided</th>
<th>Missing</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publicly Available Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berkeley</td>
<td>66.43</td>
<td>66.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>64.71</td>
<td>64.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford</td>
<td>61.40</td>
<td>63.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reconcile</td>
<td>49.40</td>
<td>54.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bart</td>
<td>48.44</td>
<td>54.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UIUC</td>
<td>45.21</td>
<td>52.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CherryPicker</td>
<td>40.71</td>
<td>51.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoNLL, Predicted Mentions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee-open</td>
<td>61.03</td>
<td>62.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee</td>
<td>59.56</td>
<td>61.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sapena</td>
<td>59.54</td>
<td>61.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song</td>
<td>59.95</td>
<td>60.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang</td>
<td>57.13</td>
<td>61.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cai-open</td>
<td>57.86</td>
<td>60.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nugues</td>
<td>58.61</td>
<td>59.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uryupina-open</td>
<td>57.63</td>
<td>58.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santos</td>
<td>56.65</td>
<td>59.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoyanov</td>
<td>58.43</td>
<td>57.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hao</td>
<td>54.46</td>
<td>55.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yang</td>
<td>52.31</td>
<td>55.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charton</td>
<td>52.49</td>
<td>55.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klenner-open</td>
<td>49.86</td>
<td>55.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sobha</td>
<td>50.48</td>
<td>54.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou</td>
<td>48.96</td>
<td>53.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kobdani</td>
<td>48.62</td>
<td>53.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>47.88</td>
<td>52.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xinxin</td>
<td>46.62</td>
<td>51.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kummerfeld</td>
<td>42.70</td>
<td>50.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irwin-open</td>
<td>27.21</td>
<td>44.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhekova</td>
<td>24.08</td>
<td>41.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irwin</td>
<td>19.98</td>
<td>42.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoNLL, Gold NP Spans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee-open</td>
<td>65.39</td>
<td>65.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee</td>
<td>63.90</td>
<td>64.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nugues</td>
<td>62.12</td>
<td>61.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang</td>
<td>59.77</td>
<td>62.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santos</td>
<td>59.52</td>
<td>61.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoyanov</td>
<td>61.53</td>
<td>59.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song</td>
<td>55.48</td>
<td>58.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kobdani</td>
<td>53.94</td>
<td>55.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>51.64</td>
<td>54.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhekova</td>
<td>35.22</td>
<td>45.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoNLL, Gold Mentions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee-open</td>
<td>81.56</td>
<td>75.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang</td>
<td>82.52</td>
<td>73.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most Errors</td>
<td>11.6</td>
<td>18.6</td>
<td>13.3</td>
<td>25.6</td>
<td>23.1</td>
<td>9.8</td>
<td>15.6</td>
<td></td>
</tr>
</tbody>
</table>
## System Comparison

<table>
<thead>
<tr>
<th>System</th>
<th>Publicly Available Systems</th>
<th>Metric $F_1$</th>
<th>Span</th>
<th>Conflated</th>
<th>Extra</th>
<th>Extra</th>
<th>Divided</th>
<th>Missing</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MUC</td>
<td></td>
<td>B$^3$</td>
<td>Error</td>
<td>Mention</td>
<td>Entity</td>
<td>Mention</td>
<td>Entity</td>
</tr>
<tr>
<td>Berkeley</td>
<td></td>
<td>66.43</td>
<td>66.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td></td>
<td>64.71</td>
<td>64.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford</td>
<td></td>
<td>61.40</td>
<td>63.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Top Public Systems

Errors per Document

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity
Top Public Systems

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Errors per Document

- Berkeley
Top Public Systems

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Errors per Document

- Berkeley

0 2 4 6 8
Top Public Systems

Errors per Document

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Berkeley
IMS
Top Public Systems

Errors per Document

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Berkeley
IMS
Stanford

0 2 4 6 8
Top Public Systems

Errors per Document

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Berkeley
IMS
Stanford
Results

- System comparison
- Investigation of common errors
Results

- System comparison
- Investigation of common errors
1 – Span Errors due to Parsing
1 – Span Errors due to Parsing

... Today President Bush ...
1 – Span Errors due to Parsing

... Today President Bush ...
1 – Span Errors due to Parsing

... Today President Bush ...
## 1 – Span Errors due to Parsing

<table>
<thead>
<tr>
<th>Label</th>
<th>Missing or Extra (per Document)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>0.34</td>
</tr>
</tbody>
</table>
## 1 - Span Errors due to Parsing

<table>
<thead>
<tr>
<th>Label</th>
<th>Missing or Extra (per Document)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>0.34</td>
</tr>
<tr>
<td>POS</td>
<td>0.34</td>
</tr>
<tr>
<td>,</td>
<td>0.29</td>
</tr>
<tr>
<td>SBAR</td>
<td>0.18</td>
</tr>
<tr>
<td>PP</td>
<td>0.17</td>
</tr>
</tbody>
</table>
2 – Misleading String Features
Or is it checkbook diplomacy?

...pursuing checkbook diplomacy
2 – Misleading String Features

<table>
<thead>
<tr>
<th>% of Extra Entities</th>
<th>0</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
</table>

**Note:** The table shows the percentage of misleading string features across different categories.
2 – Misleading String Features

% of Extra Entities

Head Match

* * *

86
2 – Misleading String Features

% of Extra Entities

- Head Match: 86%
- Exact Match: 53%
2 – Misleading String Features
We took a taxi . . .

. . . the car was stopped . . .
2 – Misleading String Features

% of Missing Entities

0
20
40
60
80
100
2 – Misleading String Features

% of Missing Entities

Head Match

0 20 40 60 80 100

48
2 – Misleading String Features

% of Missing Entities

Head Match

Exact Match

48

18
You can enjoy the most charming . . .
You can ascend Victoria Peak . . .
3 – Pronoun Cluster Errors

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Errors (per Document)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Entity</td>
<td>0.63</td>
</tr>
<tr>
<td>Missing Entity</td>
<td></td>
</tr>
</tbody>
</table>
3 – Pronoun Cluster Errors

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Errors (per Document)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Entity</td>
<td>0.63</td>
</tr>
<tr>
<td>Missing Entity</td>
<td>0.1</td>
</tr>
</tbody>
</table>
4 – Impact on Metrics
4 – Impact on Metrics

- Apply transformation to correct all errors of one type
4 – Impact on Metrics

- Apply transformation to correct all errors of one type
4 – Impact on Metrics

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

F-score impact
4 – Impact on Metrics

![Bar chart showing F-score impact for various entity errors.]

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

MUC

F-score impact
4 – Impact on Metrics

- Span Error
- Conflated Entities
- Extra Mention
- Extra Entity
- Divided Entity
- Missing Mention
- Missing Entity

Impact on Metrics (F-score)

- Span Error: Impact
- Conflated Entities: Impact
- Extra Mention: Impact
- Extra Entity: Impact
- Divided Entity: Impact (highest)
- Missing Mention: Impact
- Missing Entity: Impact

MUC Scale (0-12)
4 – Impact on Metrics

Impact on Metrics for different entity-related errors:

- **Span Error**: F-score impact is 3.
- **Conflated Entities**: F-score impact is 6.
- **Extra Mention**: F-score impact is 3.
- **Extra Entity**: F-score impact is 4.
- **Divided Entity**: F-score impact is 9.
- **Missing Mention**: F-score impact is 9.
- **Missing Entity**: F-score impact is 12.

The chart compares the impact on metrics for different entity-related errors between MUC and B3 models.
Conclusion
Meaningful error analysis does not have to be time consuming
Meaningful error analysis does not have to be time consuming

Download and run our tool on your output: nlp.cs.berkeley.edu/software
Conclusion

- Meaningful error analysis does not have to be time consuming

- Download and run our tool on your output: nlp.cs.berkeley.edu/software

Thank you!