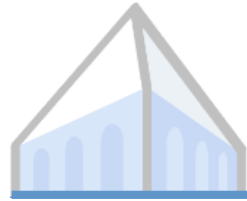


Error-Driven Analysis of Challenges in Coreference Resolution

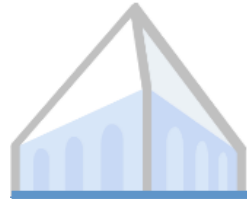


Jonathan K. Kummerfeld and Dan Klein
University of California, Berkeley



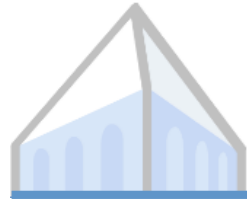
Coreference

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 ...



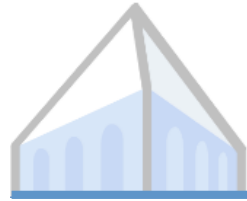
Coreference

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 ...



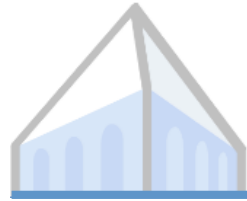
Coreference

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 ...



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 ...



Coreference

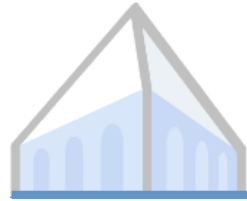
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Coreference

Span Error

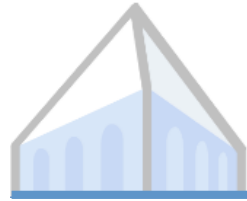
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Coreference

Span Error

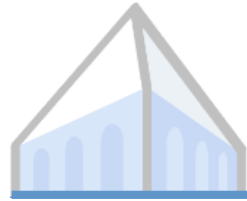
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Coreference

Span Error

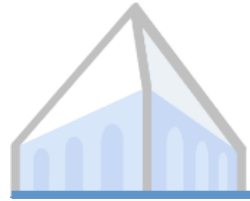
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Coreference

Span Error
Conflated Entities

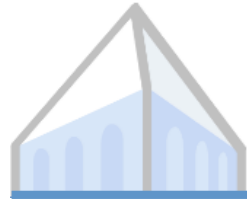
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Coreference

Span Error
Conflated Entities

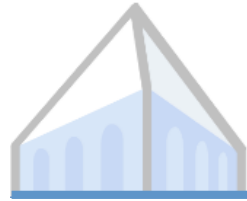
Or is it checkbook diplomacy ?
... pursuing checkbook diplomacy .



Coreference

Span Error
Conflated Entities

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



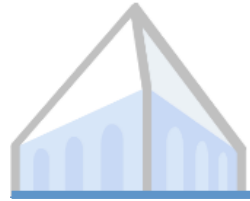
Coreference

Span Error
Conflated Entities
Extra Entity

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

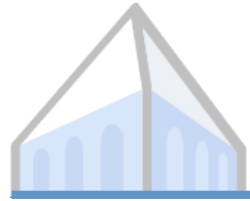


Previous Work



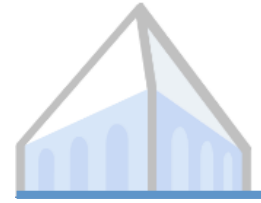
Previous Work

- ▶ Reconcile system analysis
Stoyanov et al. (2009)



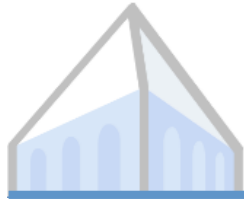
Previous Work

- ▶ Reconcile system analysis
Stoyanov et al. (2009)



Our Approach

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 ...



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

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 ...



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 ...

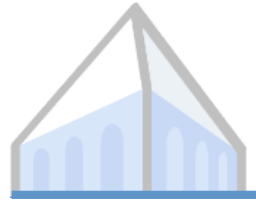


Transform

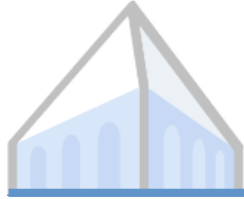
Today **President Bush** unveiled a seven point one billion dollar ...
... drawing Al Gore and **George W. Bush** to the Sunshine State ...
... nonpartisan analysis confirms **George Bush 's** plan sets aside ...
... **Jeb Bush** was once expected to deliver **his** brother this state ...



Errors

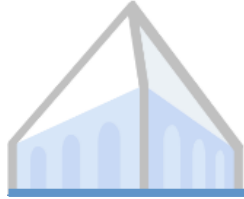


Transformations



Transformations

Alter Span – Split – Remove – Introduce – Merge



Transformations

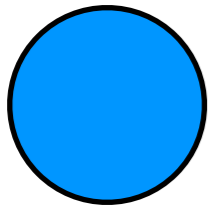
Alter Span – Split – Remove – Introduce – Merge

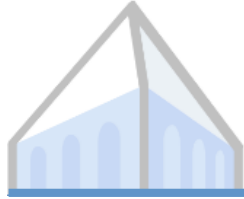
George
W. Bush



Transformations

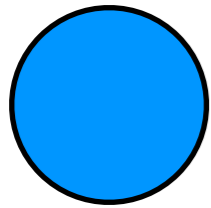
Alter Span – Split – Remove – Introduce – Merge





Transformations

Alter Span – Split – Remove – Introduce – Merge

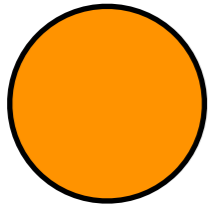
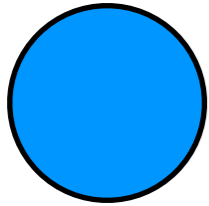


Jeb
Bush



Transformations

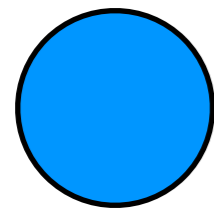
Alter Span – Split – Remove – Introduce – Merge



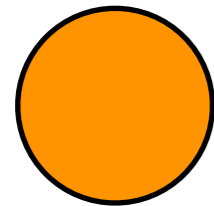


Transformations

Alter Span – Split – Remove – Introduce – Merge



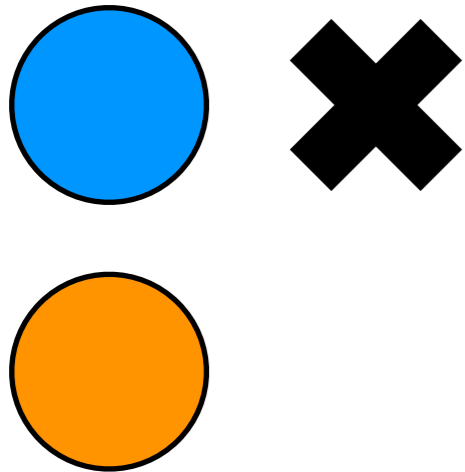
Today
President
Bush

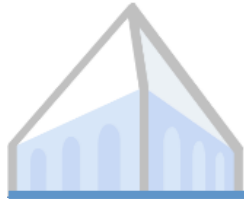




Transformations

Alter Span – Split – Remove – Introduce – Merge

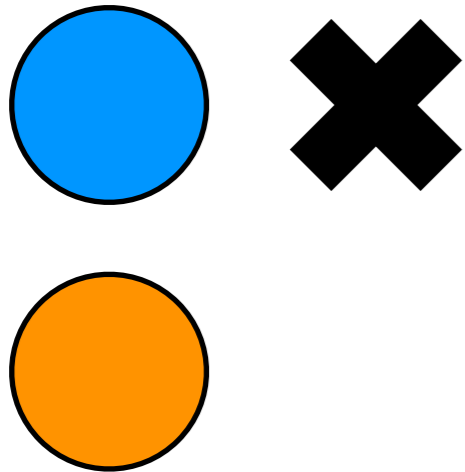




Transformations

Alter Span – Split – Remove – Introduce – Merge

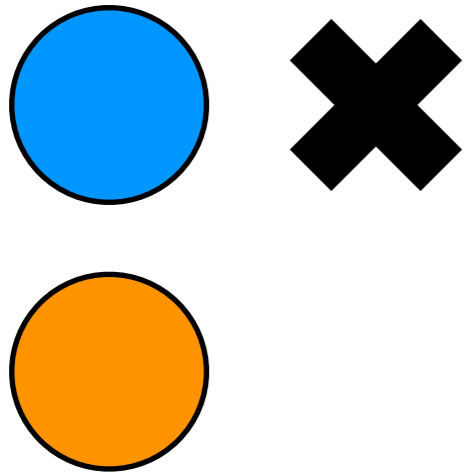
President
Bush

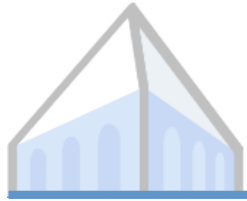




Transformations

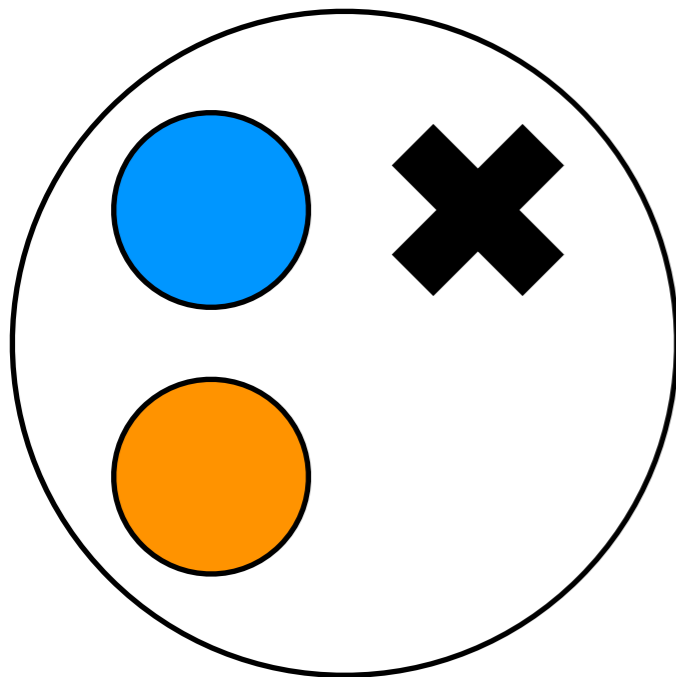
Alter Span – Split – Remove – Introduce – Merge

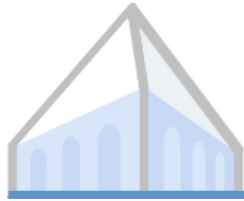




Transformations

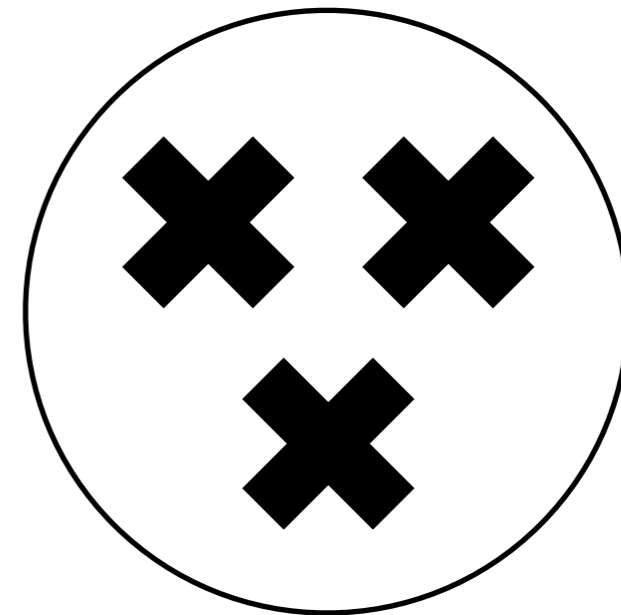
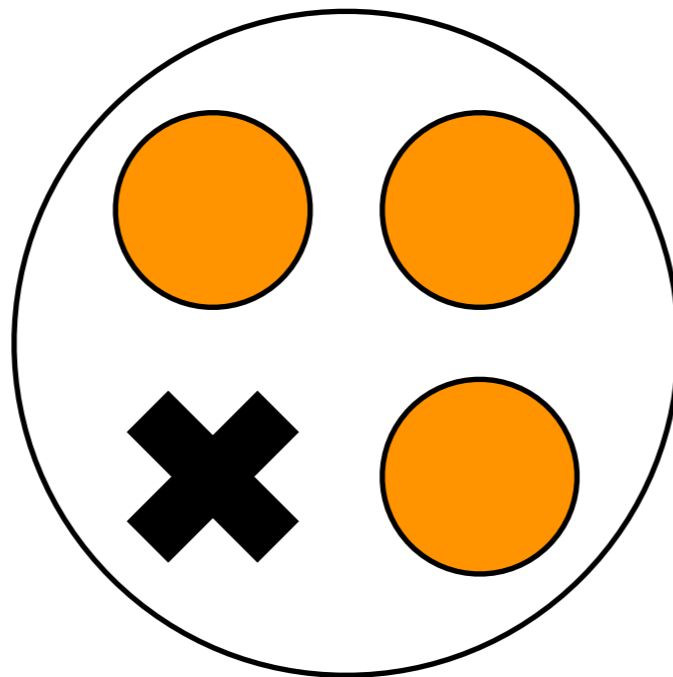
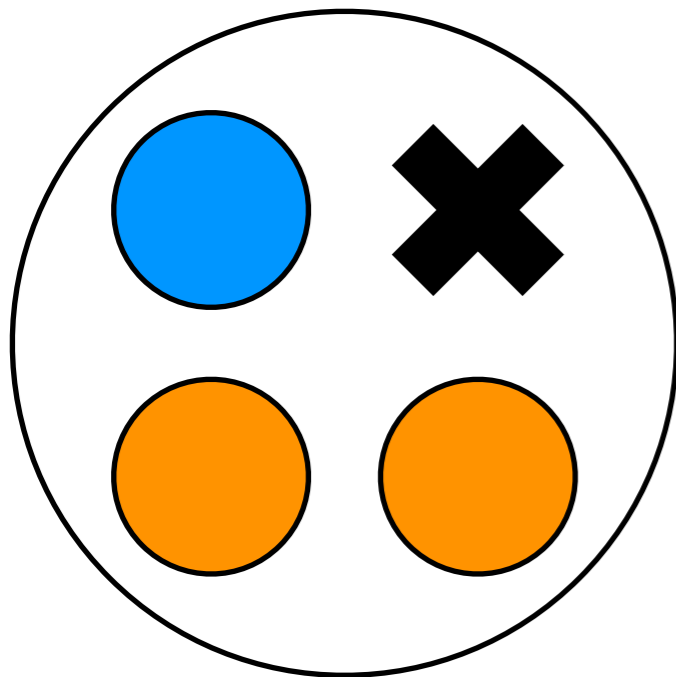
Alter Span – Split – Remove – Introduce – Merge

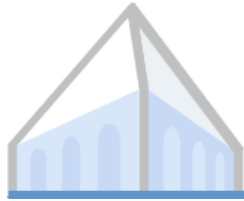




Transformations

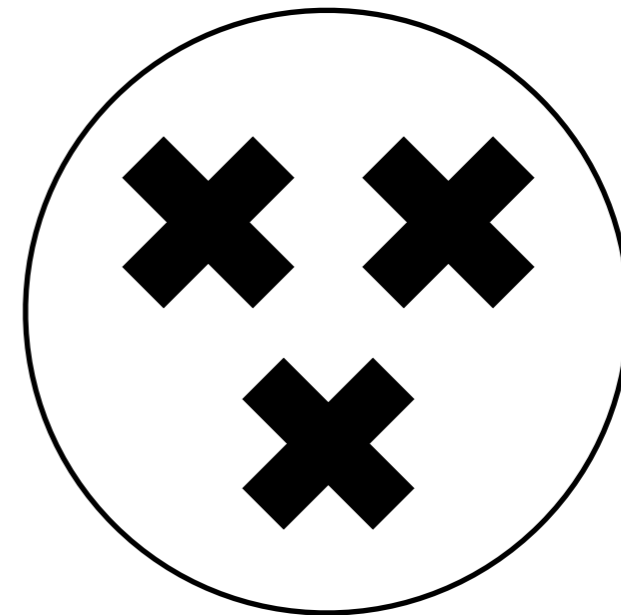
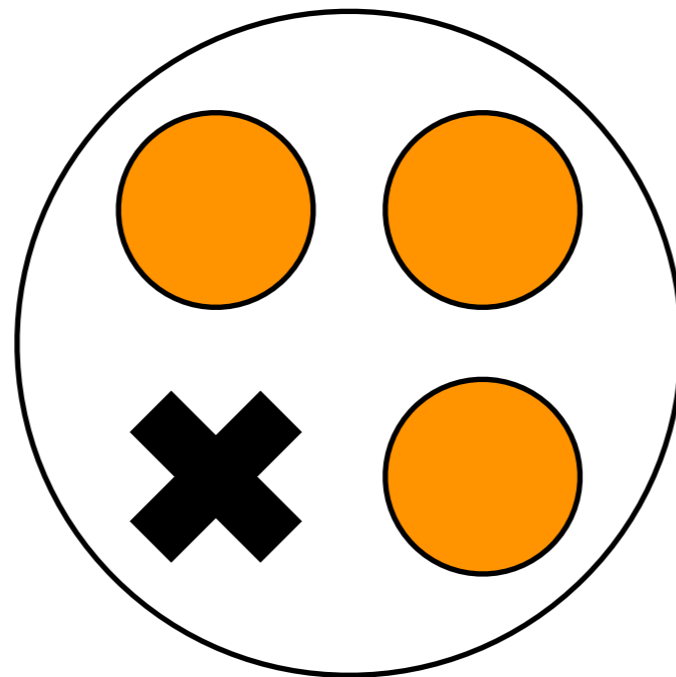
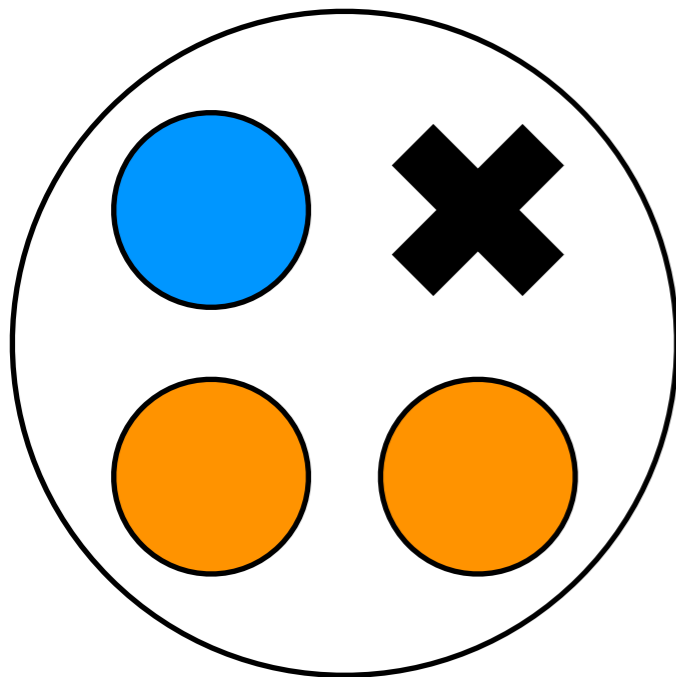
Alter Span – Split – Remove – Introduce – Merge

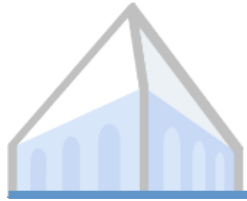




Transformations

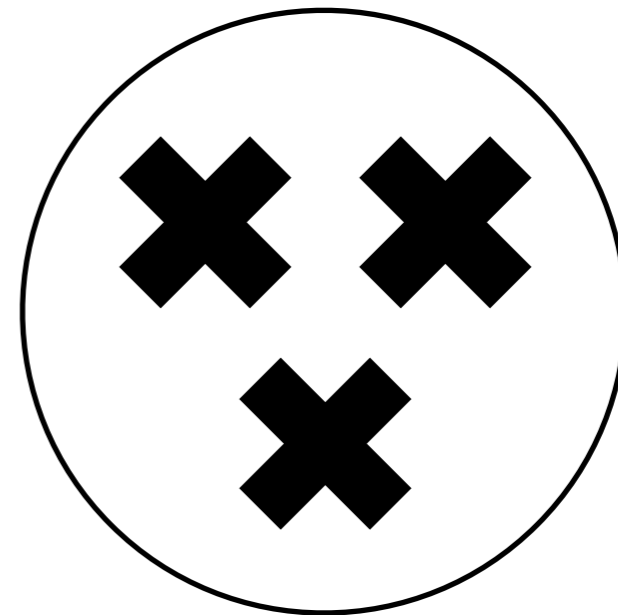
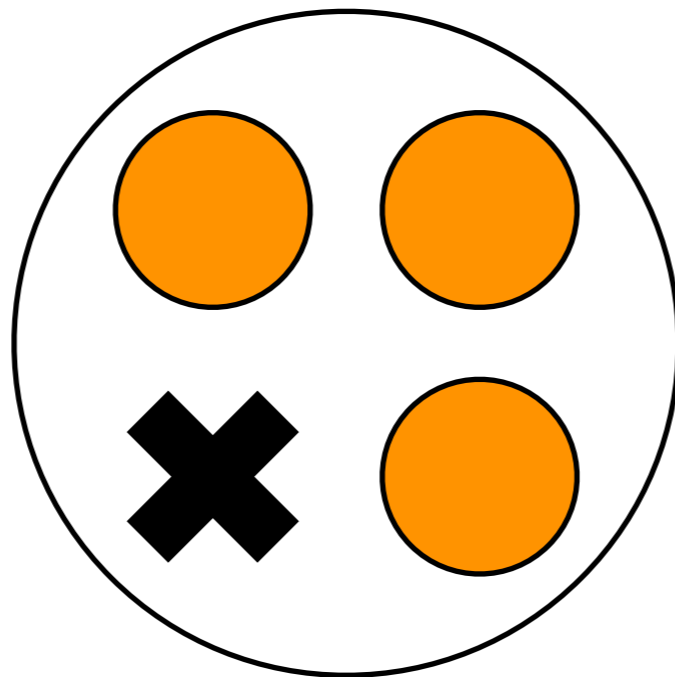
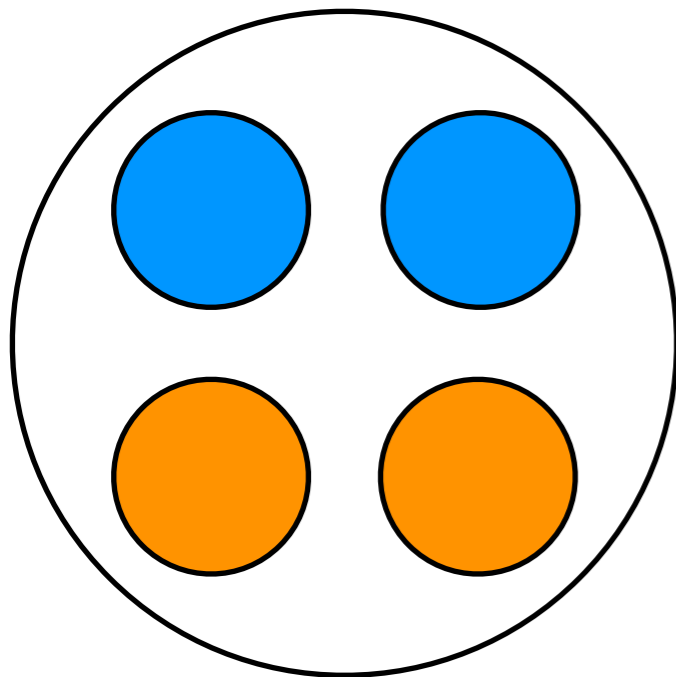
Alter Span – Split – Remove – Introduce – Merge

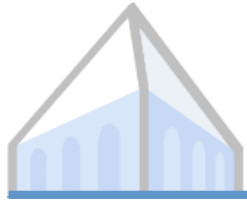




Transformations

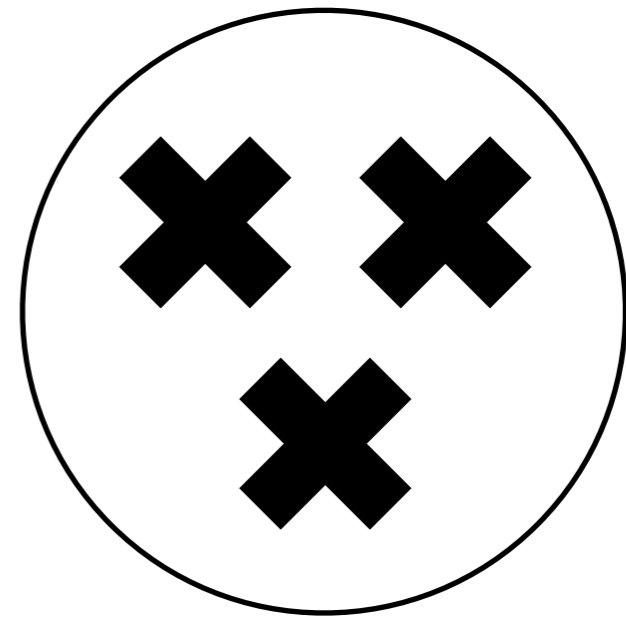
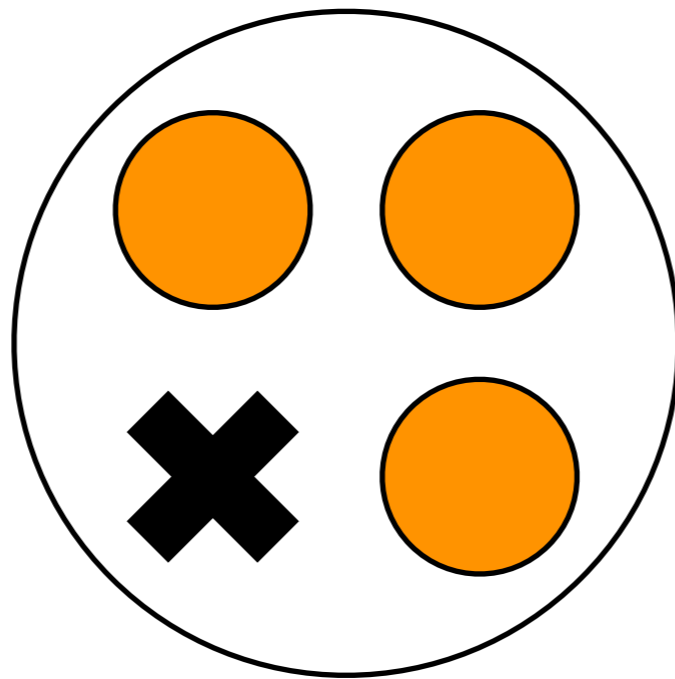
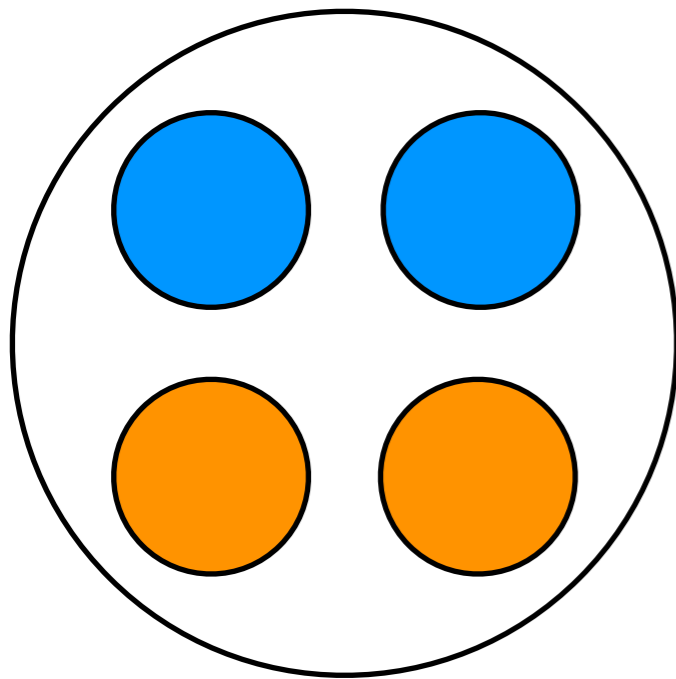
Alter Span – Split – Remove – Introduce – Merge

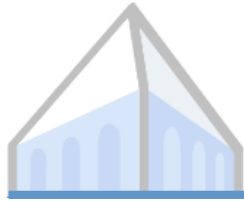




Transformations

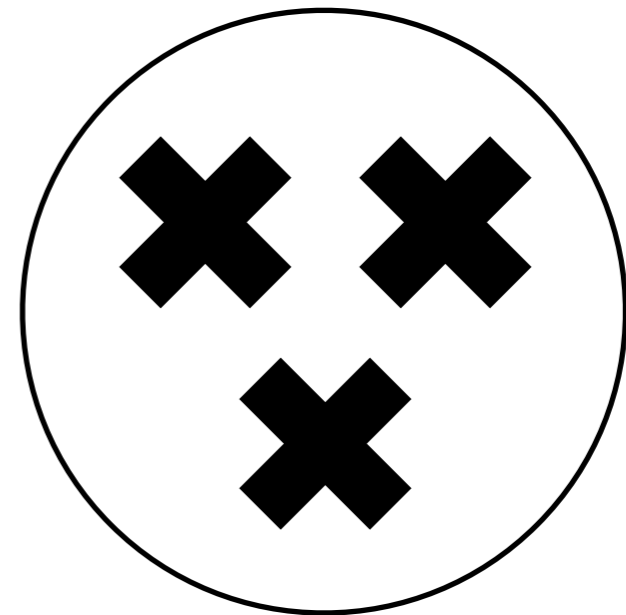
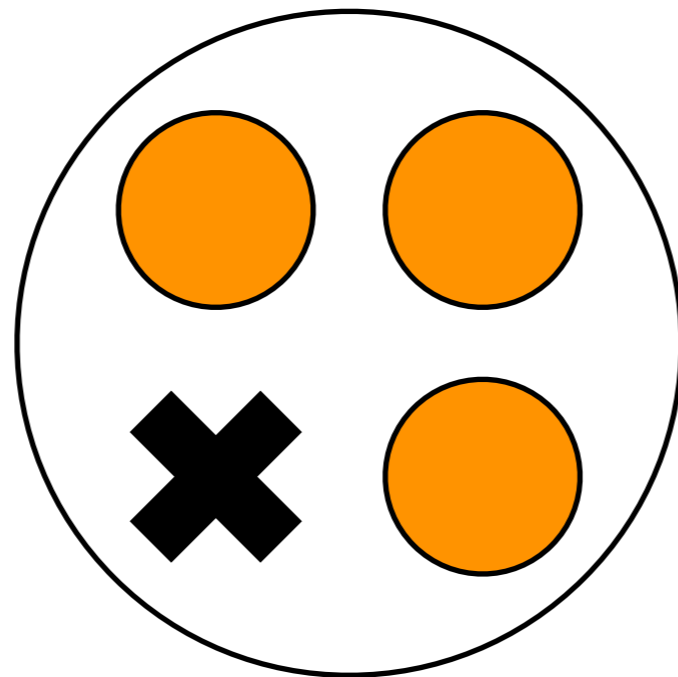
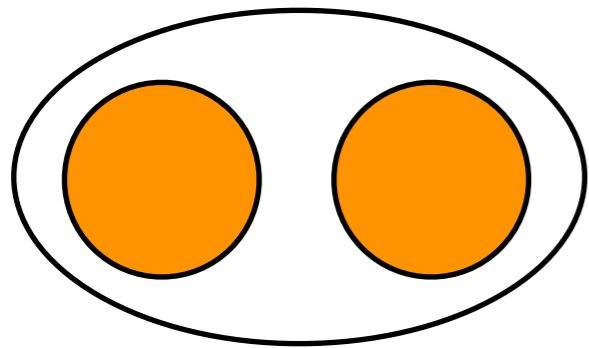
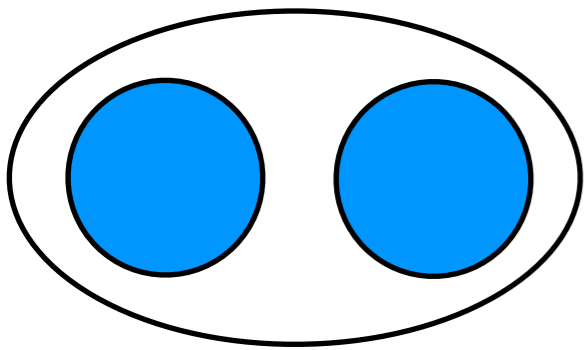
Alter Span – **Split** – Remove – Introduce – Merge

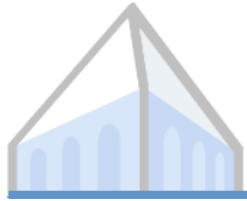




Transformations

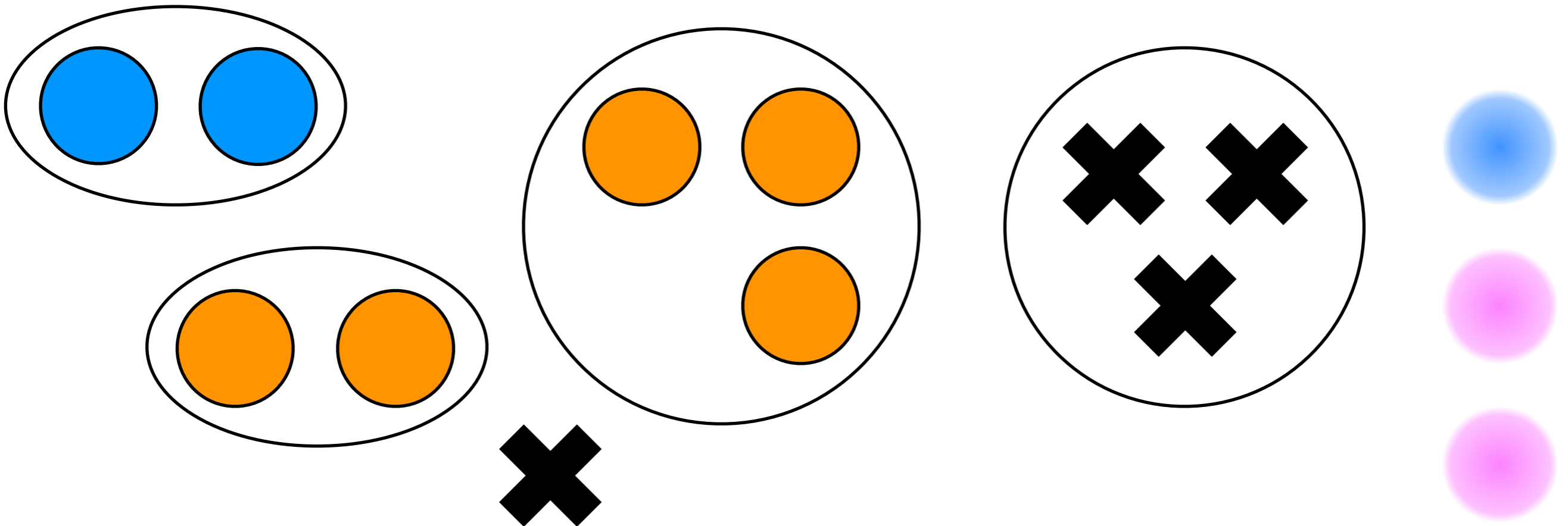
Alter Span – **Split** – Remove – Introduce – Merge





Transformations

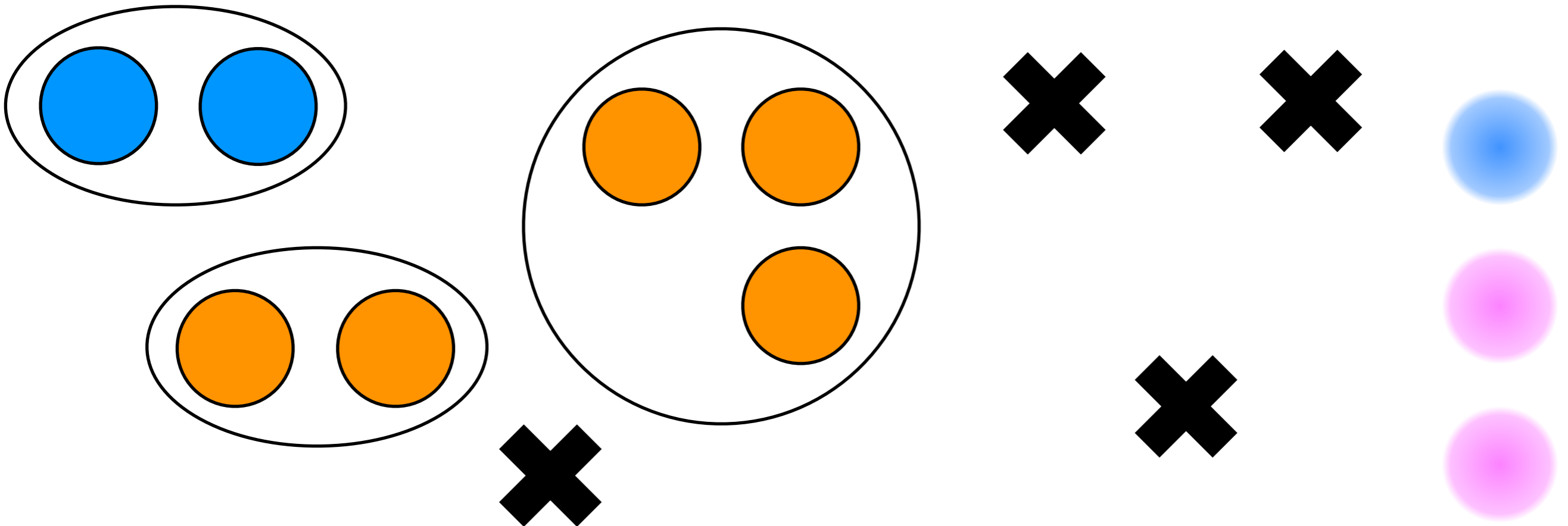
Alter Span – **Split** – Remove – Introduce – Merge

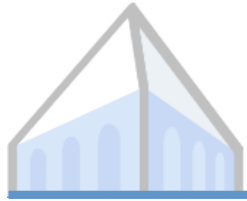




Transformations

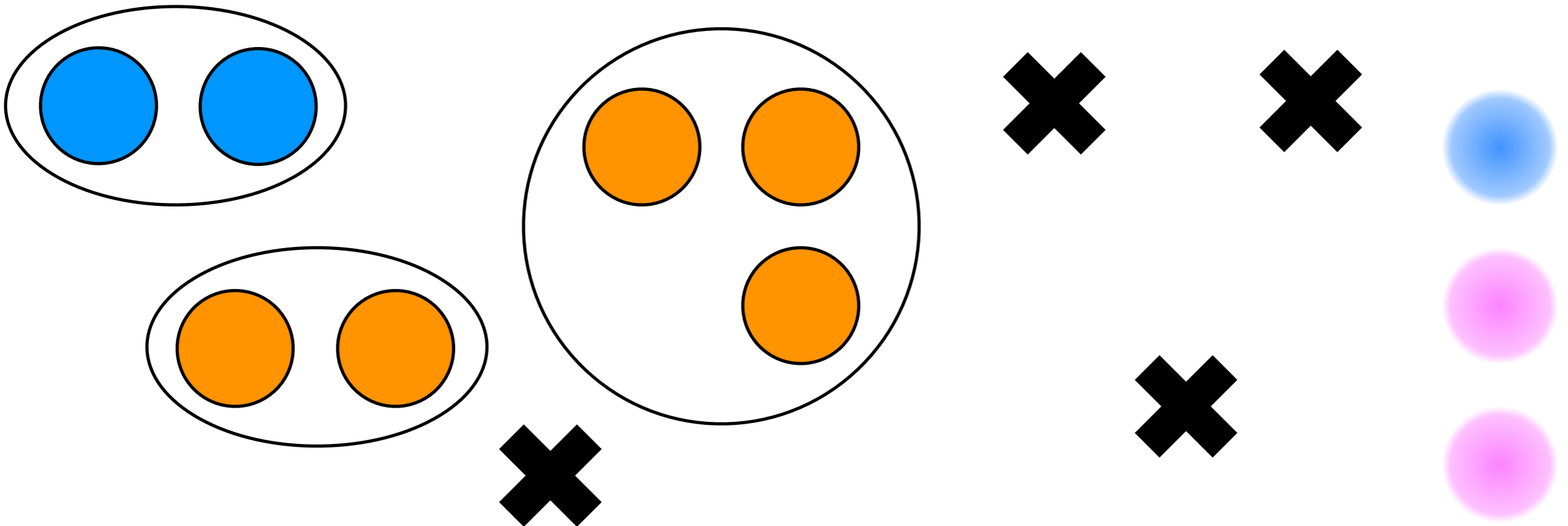
Alter Span – **Split** – Remove – Introduce – Merge





Transformations

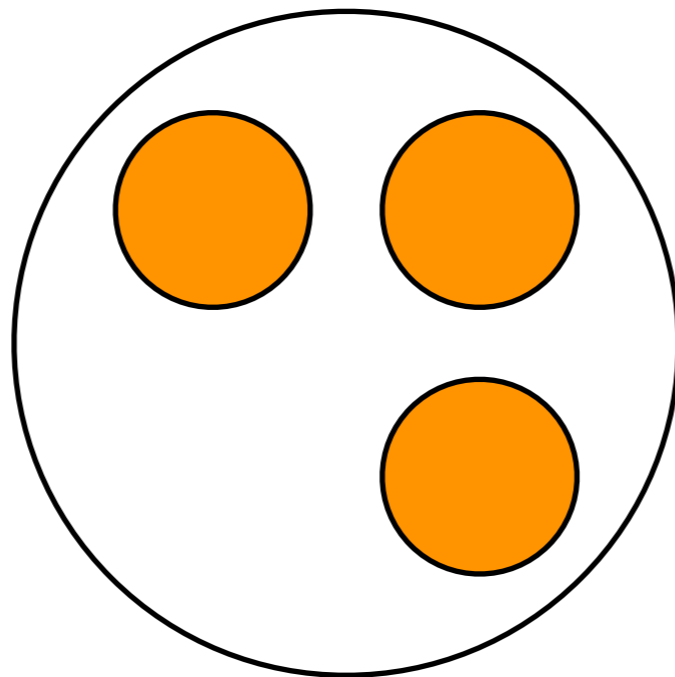
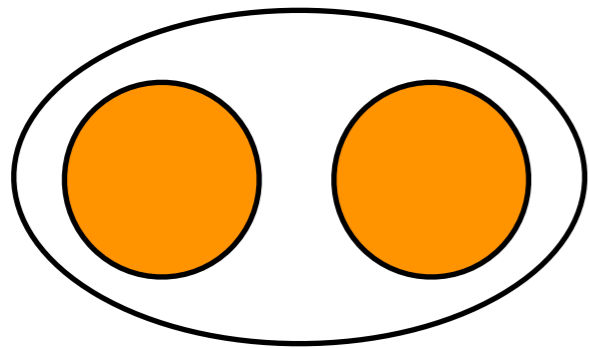
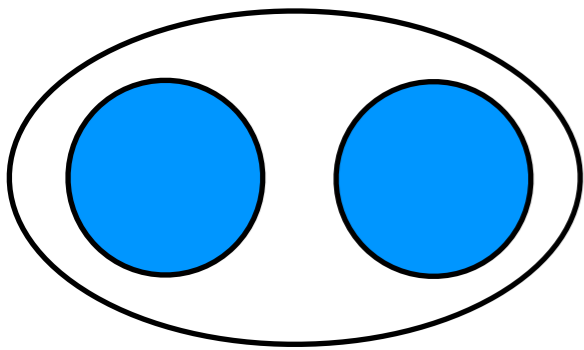
Alter Span – Split – **Remove** – Introduce – Merge

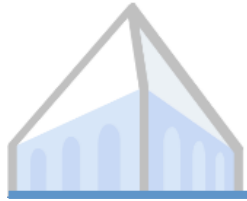




Transformations

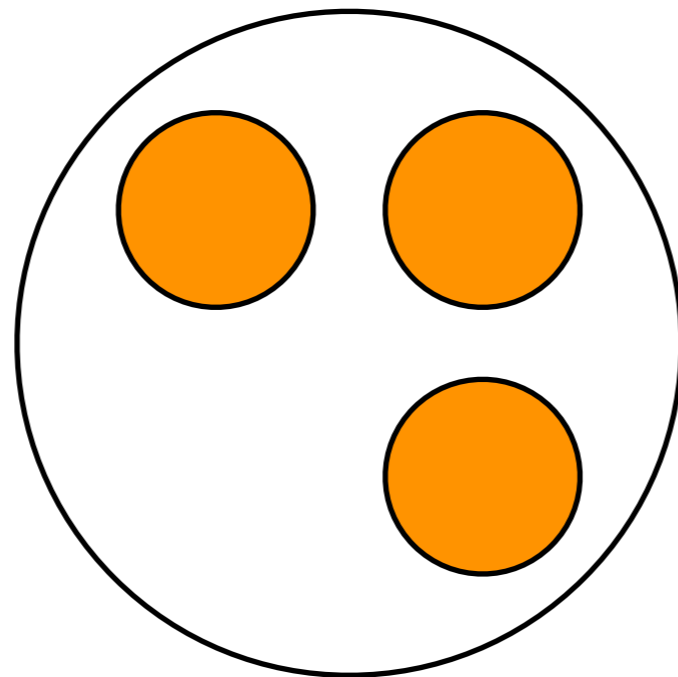
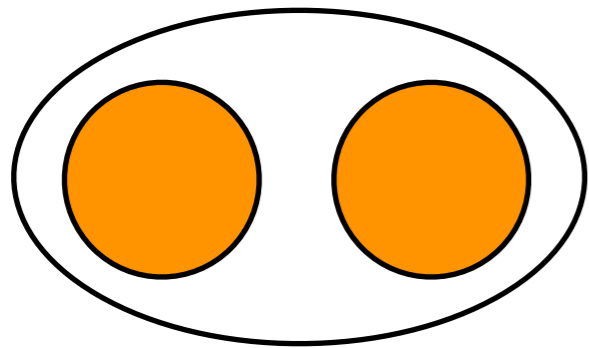
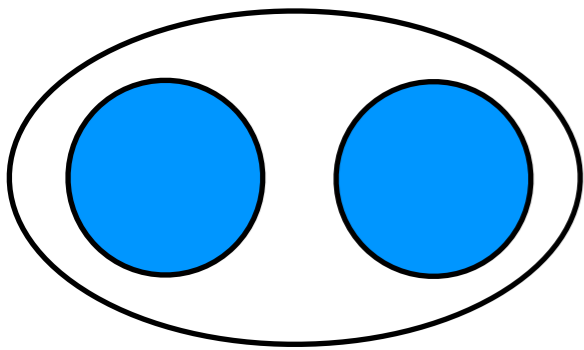
Alter Span – Split – **Remove** – Introduce – Merge

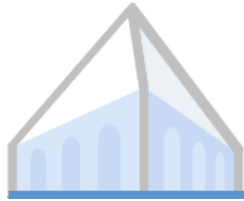




Transformations

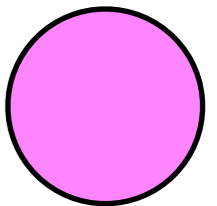
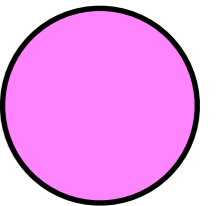
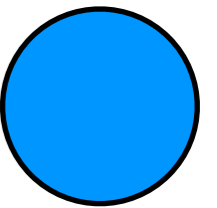
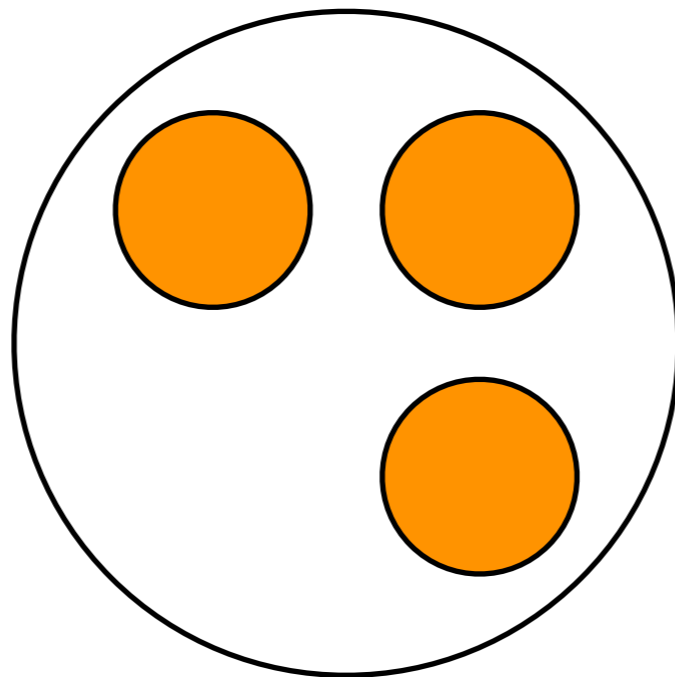
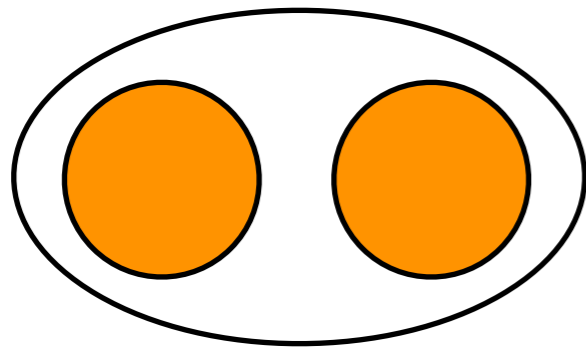
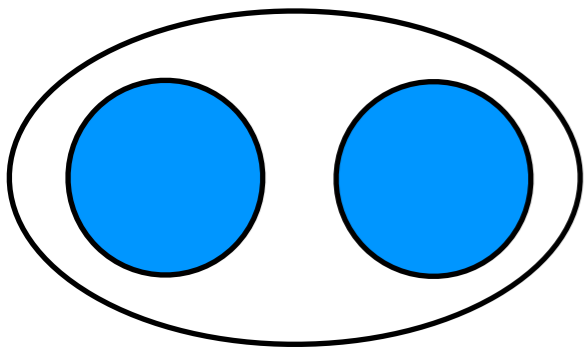
Alter Span – Split – Remove – **Introduce** – Merge

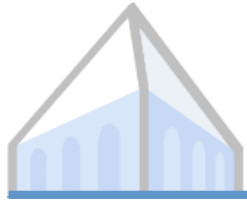




Transformations

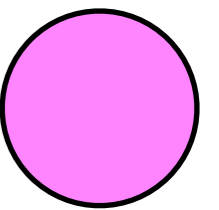
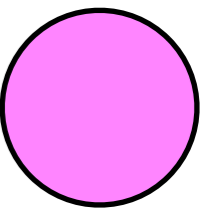
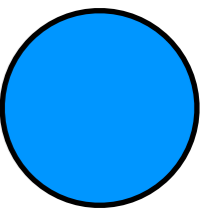
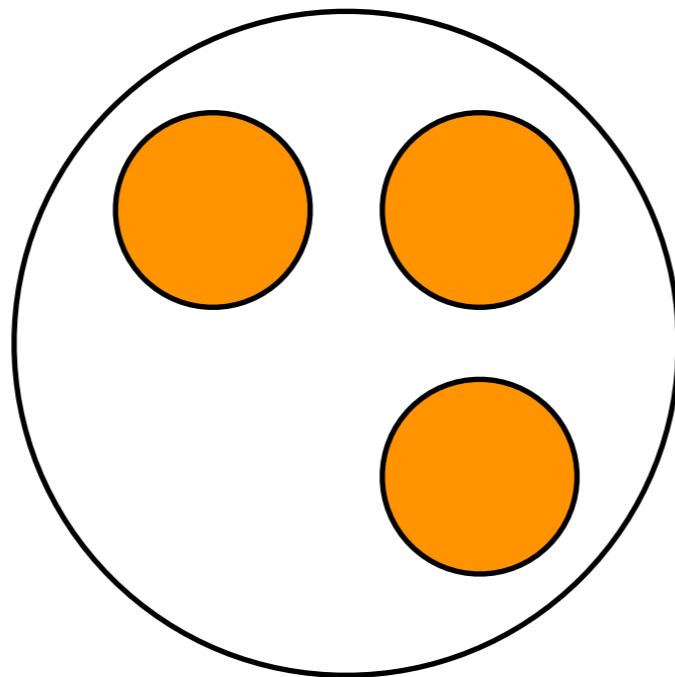
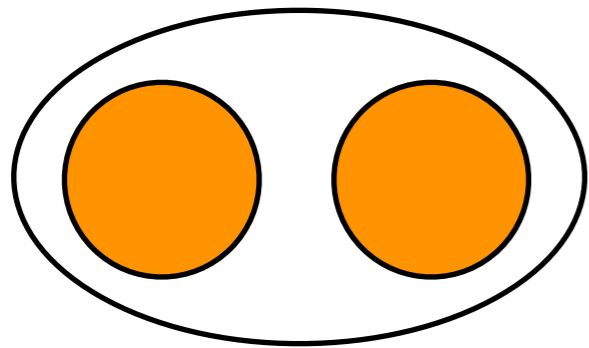
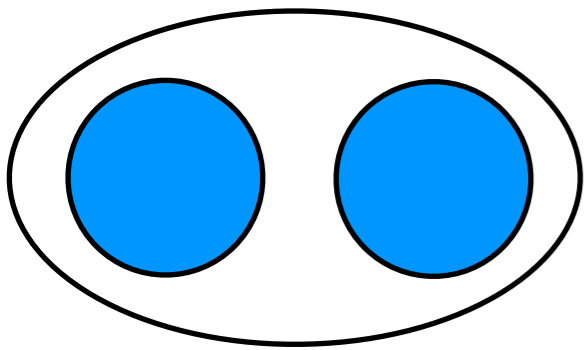
Alter Span – Split – Remove – **Introduce** – Merge

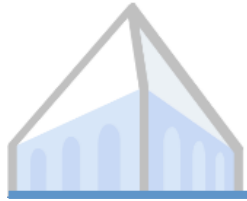




Transformations

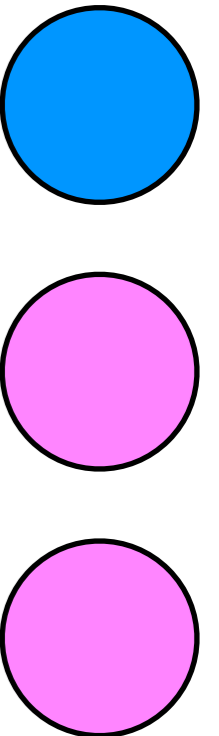
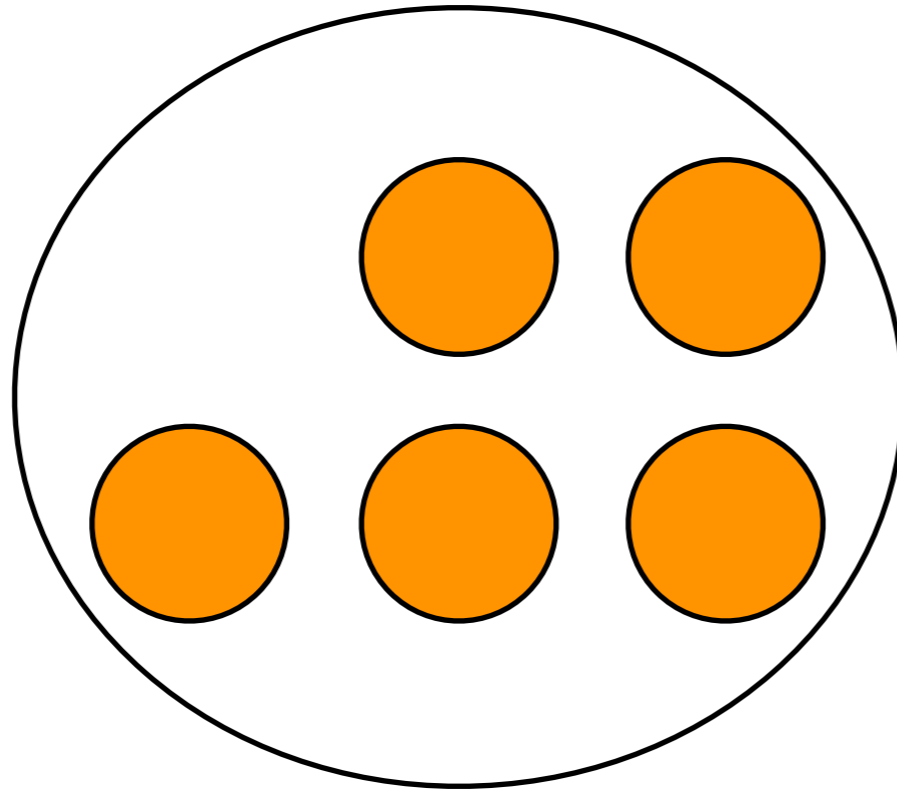
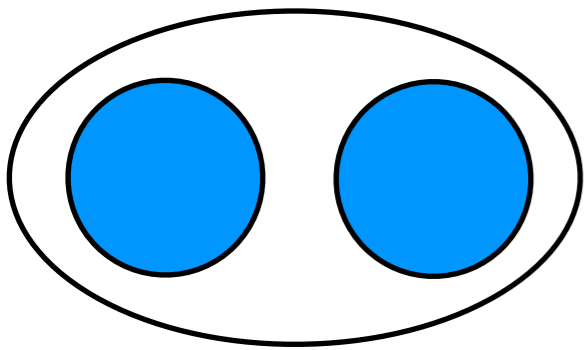
Alter Span – Split – Remove – Introduce – **Merge**

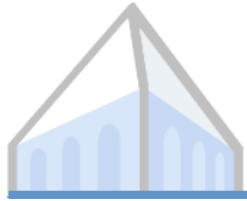




Transformations

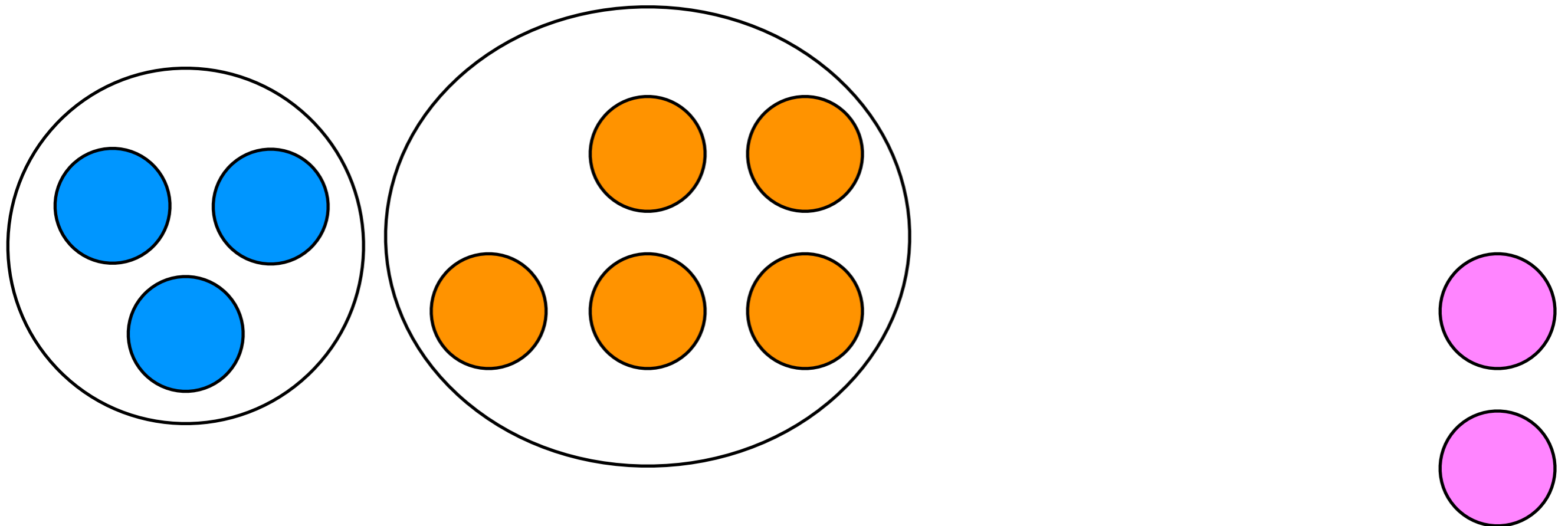
Alter Span – Split – Remove – Introduce – Merge





Transformations

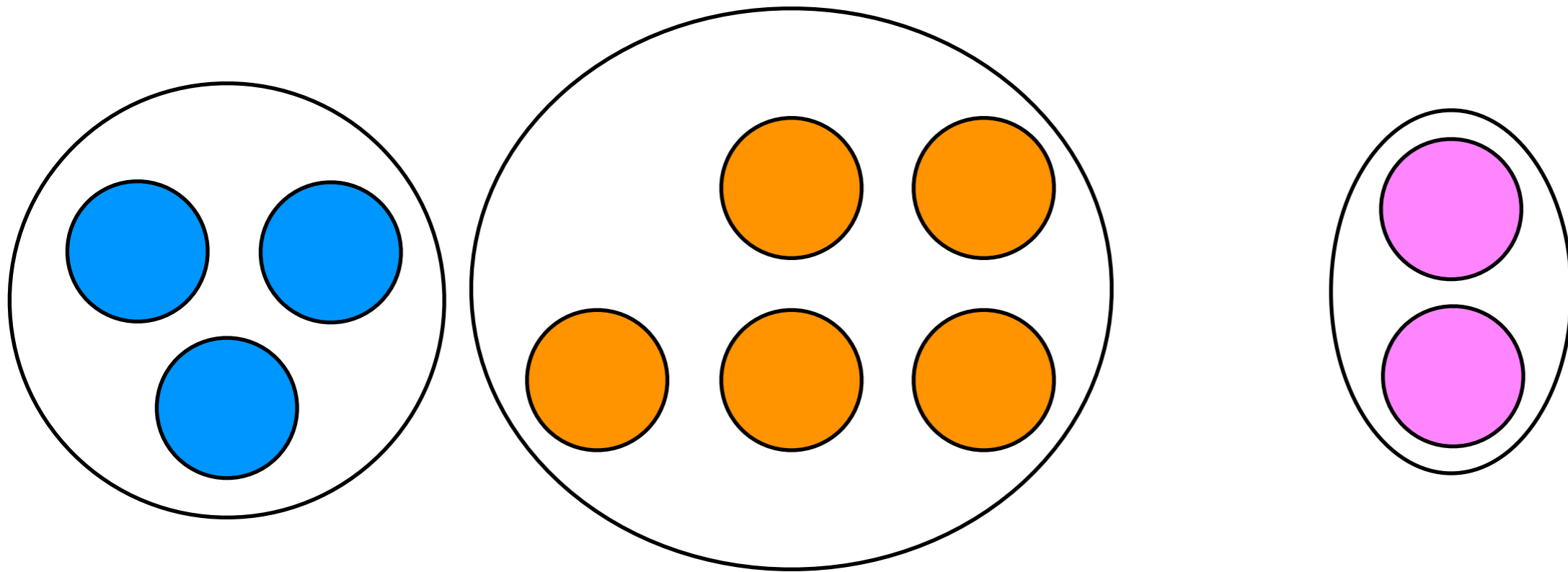
Alter Span – Split – Remove – Introduce – **Merge**

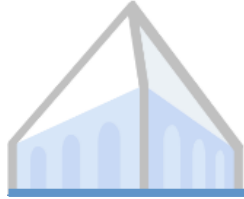




Transformations

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 ...



Transform

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 ...



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 ...



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



Classify Operations



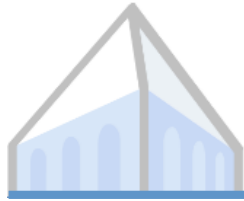
Classify Operations

Span Error

Today **President Bush** unveiled a ...

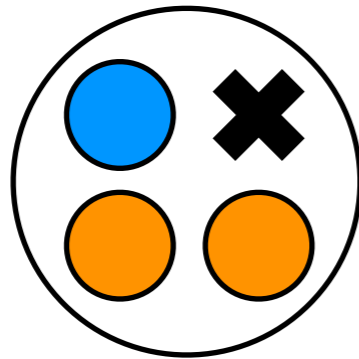


Today **President Bush** unveiled a ...

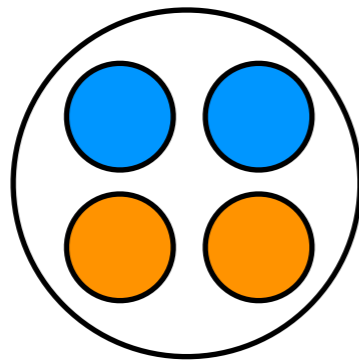
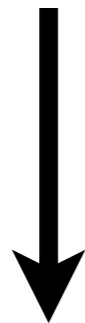


Classify Operations

Span Error



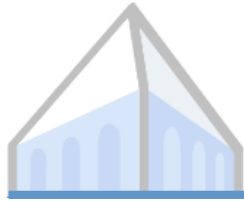
Alter Span



Today **President Bush** unveiled a ...



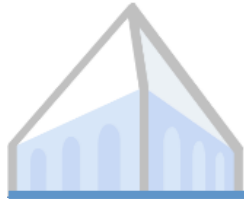
Today **President Bush** unveiled a ...



Classify Operations

Extra Mention

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



Classify Operations

Extra Mention

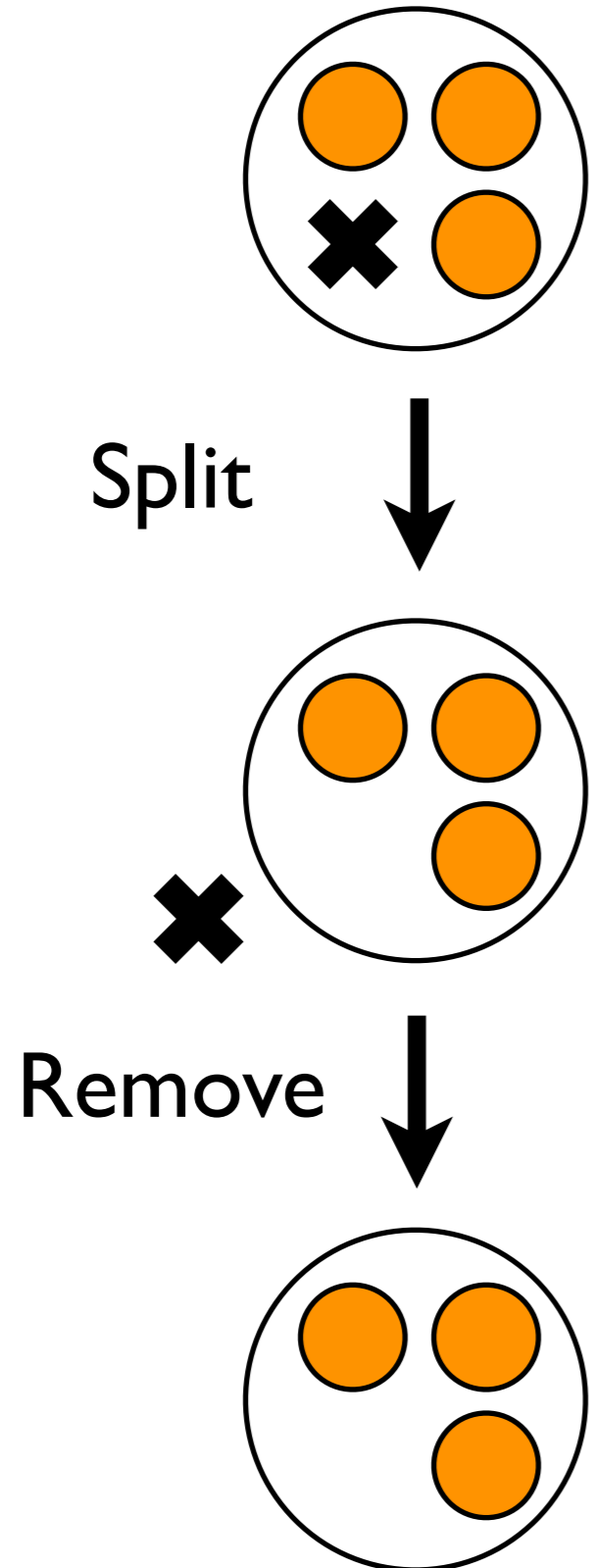
The most deadly form ...
... the virus could mutate ...
... it 's just a matter of time .



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



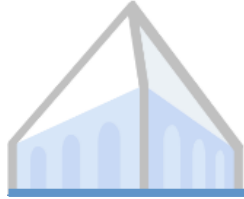
Classify Operations



Extra Mention

The most deadly form ...
... 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 .



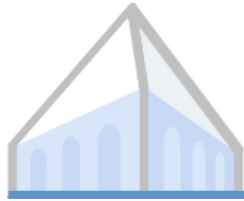
Classify Operations

Extra Entity

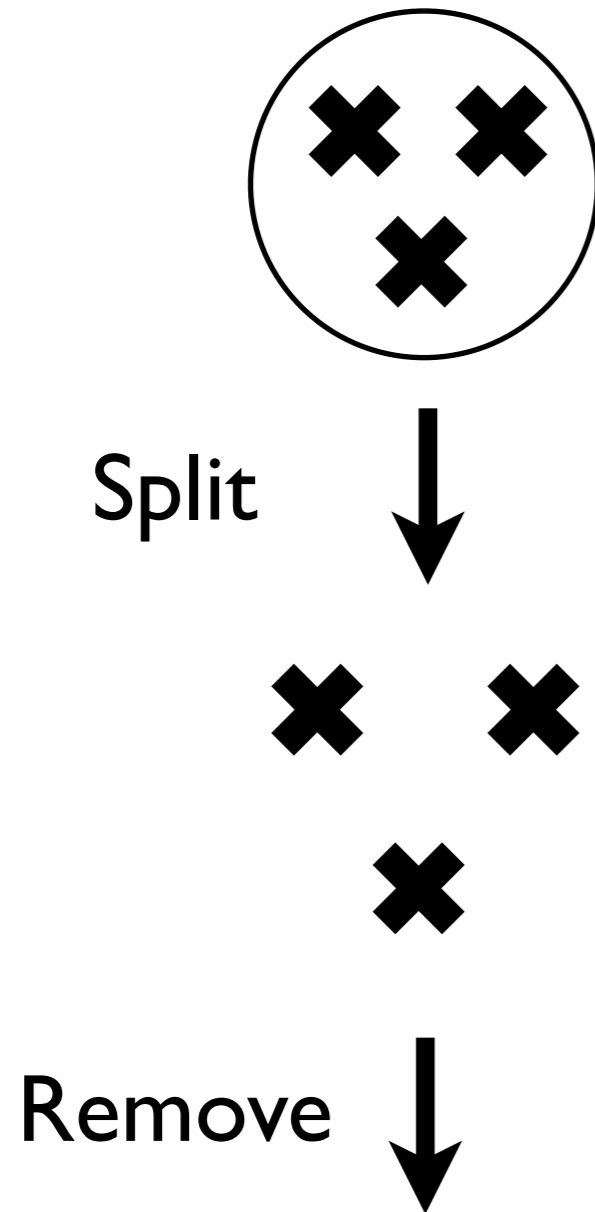
Or is it checkbook diplomacy ?
... pursuing checkbook diplomacy



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



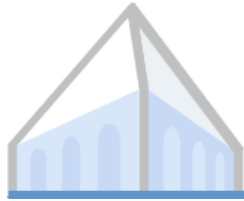
Classify Operations



Extra Entity

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

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



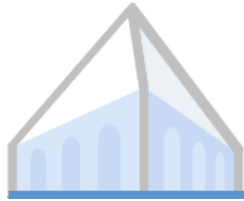
Classify Operations

Missing Mention

... usually **a search** takes ...

... calling off **that search**

Why call **it** off ?



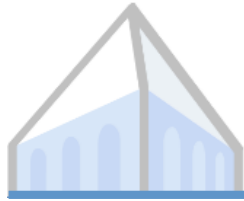
Classify Operations

Missing Mention

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

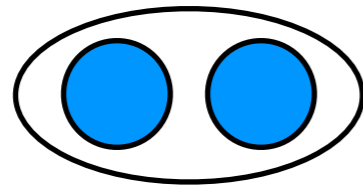


... usually **a search** takes ...
... calling off **that search** .
Why call **it** off ?

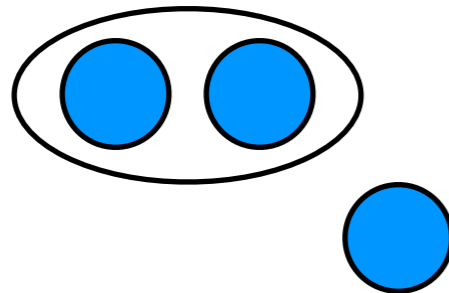


Classify Operations

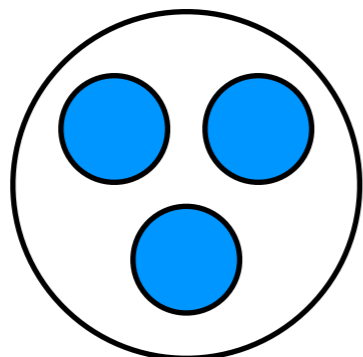
Missing Mention



Introduce



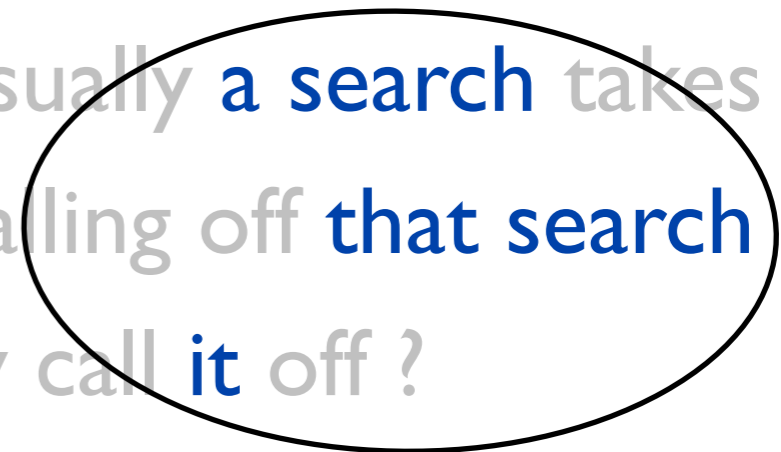
Merge



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



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



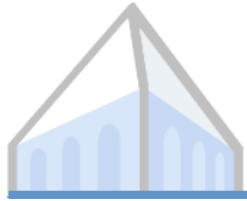


Classify Operations

Missing Entity

We took a taxi ...

... the car was stopped ...



Classify Operations

Missing Entity

We took a taxi ...

... the car was stopped ...



We took a taxi ...

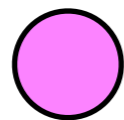
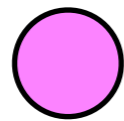
... the car was stopped ...



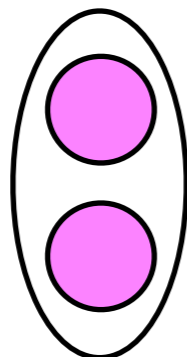
Classify Operations

Missing Entity

Introduce



Merge



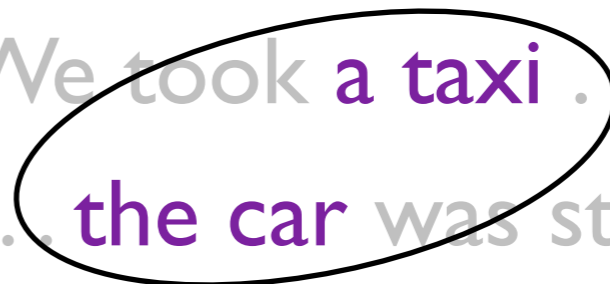
We took a taxi ...

... the car was stopped ...



We took a taxi ...

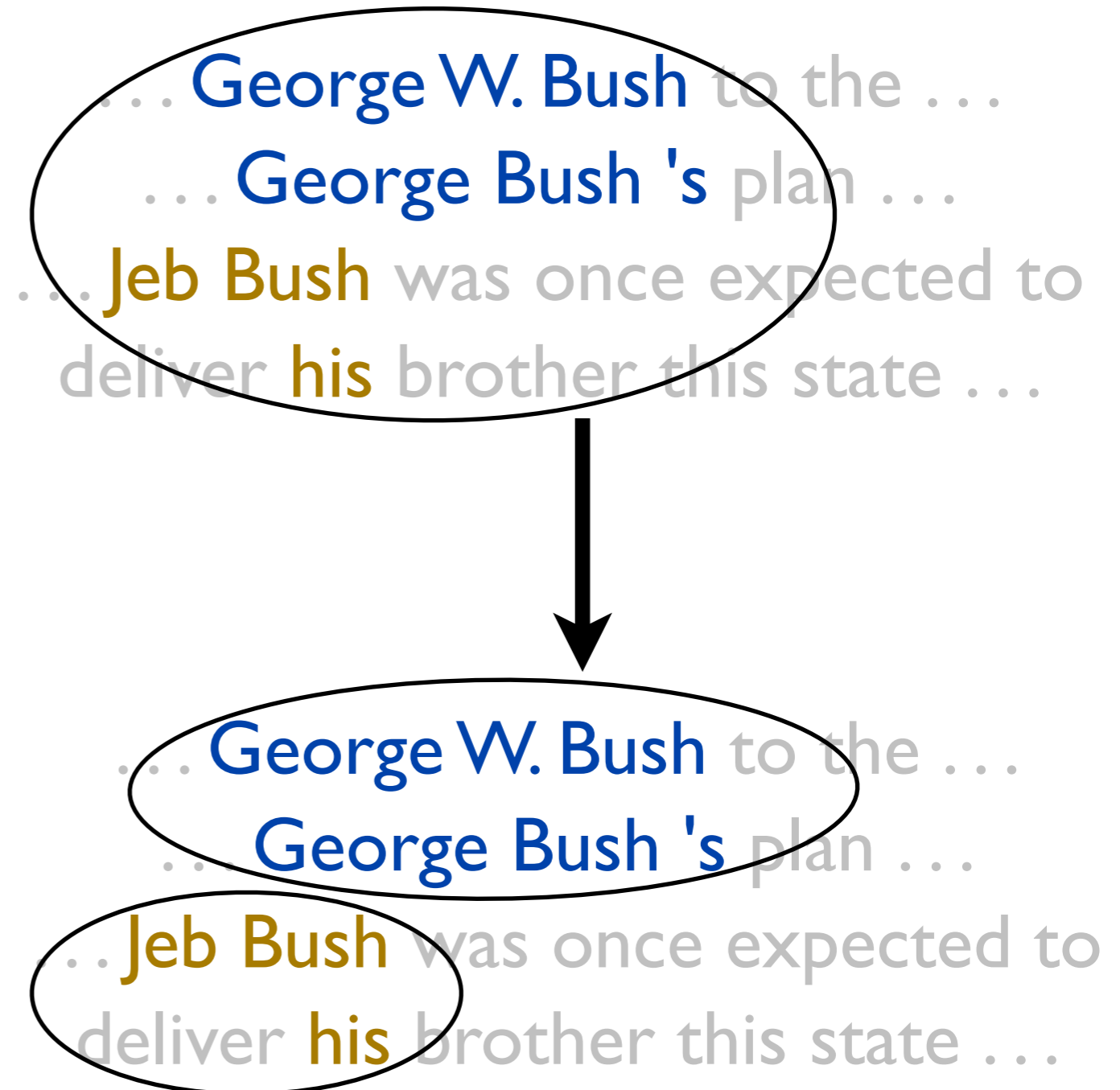
... the car was stopped ...

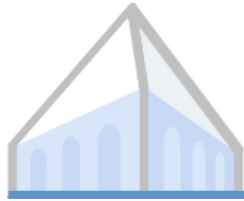




Classify Operations

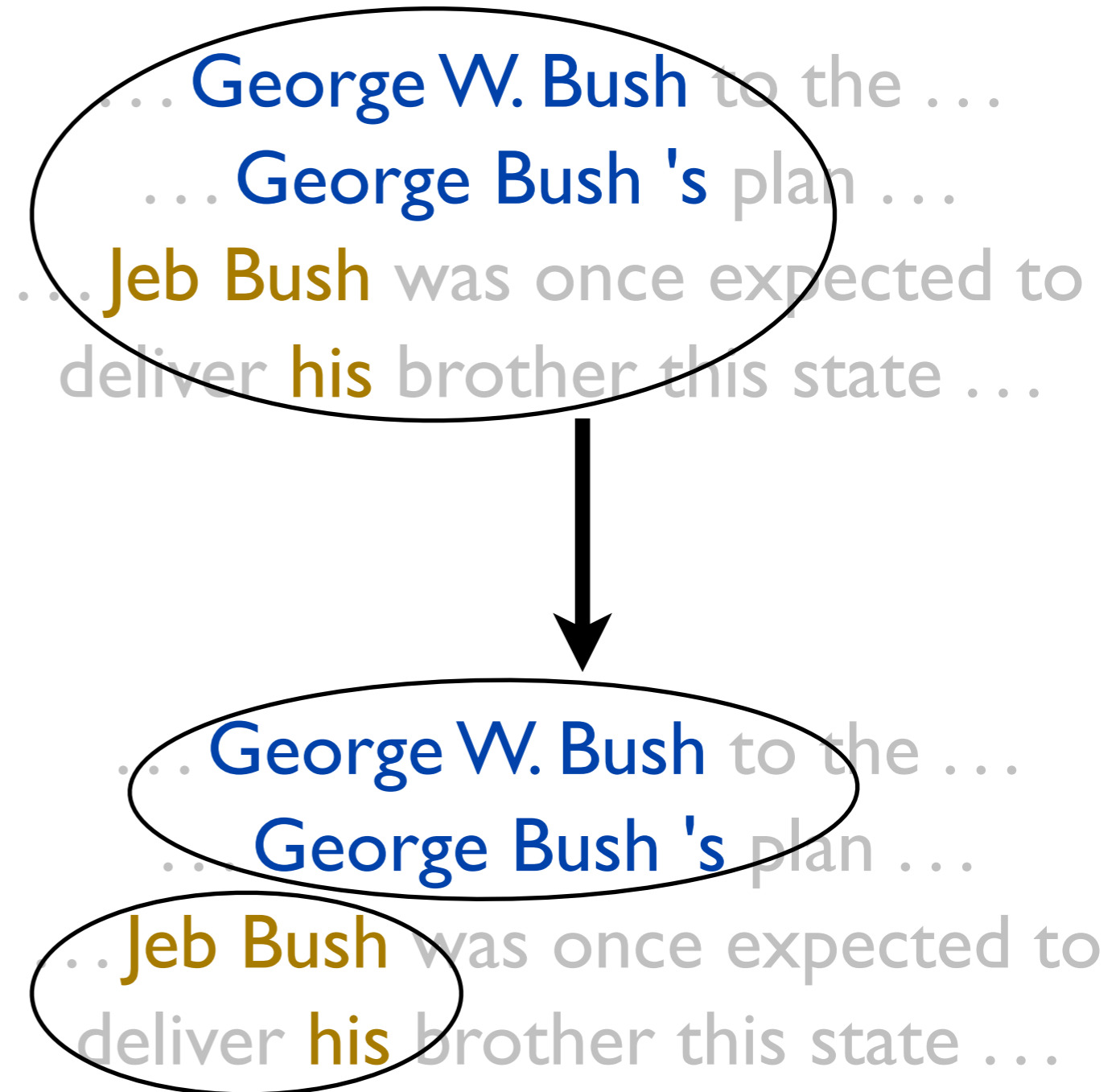
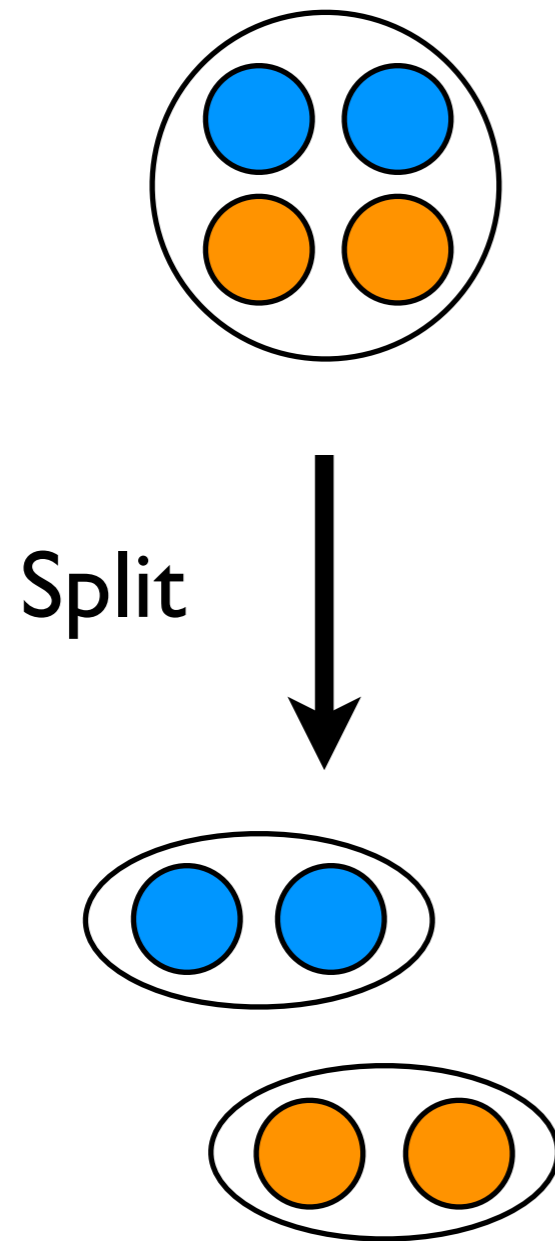
Conflated Entities

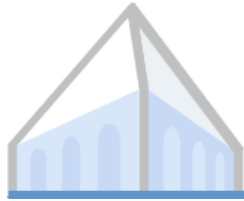




Classify Operations

Conflated Entities





Classify Operations

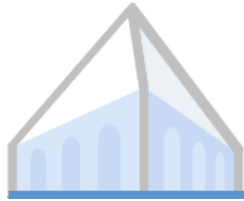
Divided Entity

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 ...

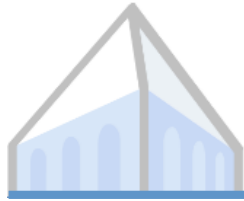


President Clinton is questioning ...

... and Mr. Clinton praised ...

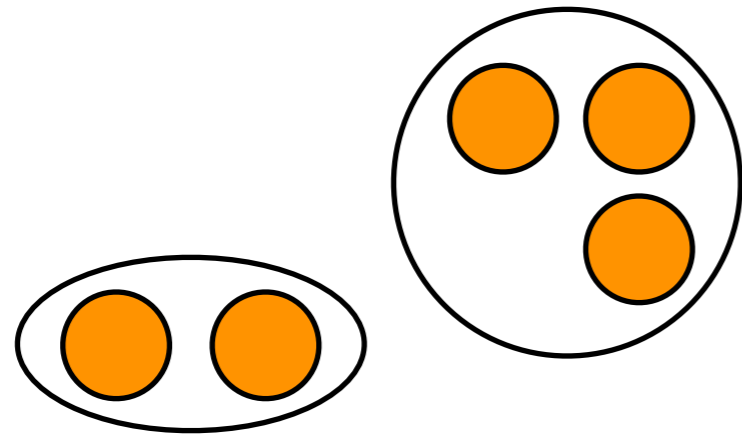
“ I want to thank Bill Daley ...

I think he did a brilliant job ...

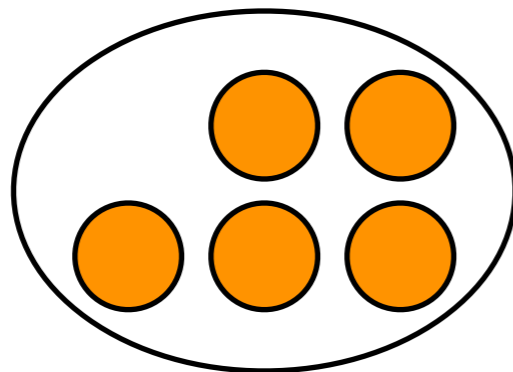


Classify Operations

Divided Entity



Merge



President Clinton is questioning ...
... and Mr. Clinton praised ...
“ I want to thank Bill Daley ...
I think he did a brilliant job ...

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

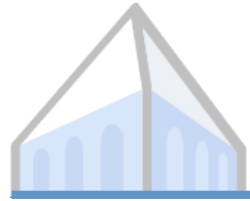


Results



Results

- ▶ System comparison



Results

- ▶ System comparison

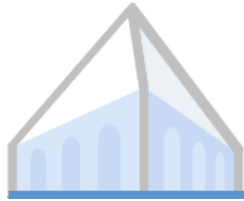


Results

- ▶ **System comparison**
- ▶ Investigation of common errors



System Comparison



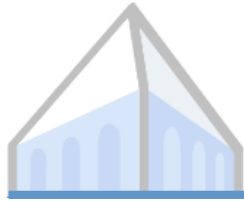
System Comparison

System	Metric F ₁		Span Error	Conflated Entities	Extra Mention	Extra Entity	Divided Entity	Missing Mention	Missing Entity
	MUC	B ³							
Publicly Available Systems									
Berkeley	66.43	66.17							
IMS	64.71	64.73							
Stanford	61.40	63.06							
Reconcile	49.40	54.90							
Bart	48.44	54.34							
UIUC	45.21	52.88							
CherryPicker	40.71	51.39							
CoNLL, Predicted Mentions									
Lee-open	61.03	62.96							
Lee	59.56	61.88							
Sapena	59.54	61.28							
Song	59.95	60.08							
Chang	57.13	61.75							
Cai-open	57.86	60.89							
Nugues	58.61	59.75							
Uryupina-open	57.63	58.74							
Santos	56.65	59.48							
Stoyanov	58.43	57.35							
Hao	54.46	55.82							
Yang	52.31	55.85							
Charton	52.49	55.61							
Klenner-open	49.86	55.62							
Sobha	50.48	54.85							
Zhou	48.96	53.42							
Kobdani	48.62	53.00							
Zhang	47.88	52.76							
Xinxin	46.62	51.50							
Kummerfeld	42.70	50.05							
Irwin-open	27.21	44.29							
Zhekova	24.08	41.42							
Irwin	19.98	42.73							
CoNLL, Gold NP Spans									
Lee-open	65.39	65.88							
Lee	63.90	64.70							
Nugues	62.12	61.67							
Chang	59.77	62.97							
Santos	59.52	61.35							
Stoyanov	61.53	59.07							
Song	55.48	58.04							
Kobdani	53.94	55.82							
Zhang	51.64	54.77							
Zhekova	35.22	45.80							
CoNLL, Gold Mentions									
Lee-open	81.56	75.95							
Chang	82.52	73.68							
Most Errors			11.6	18.6	13.3	25.6	23.1	9.8	15.6

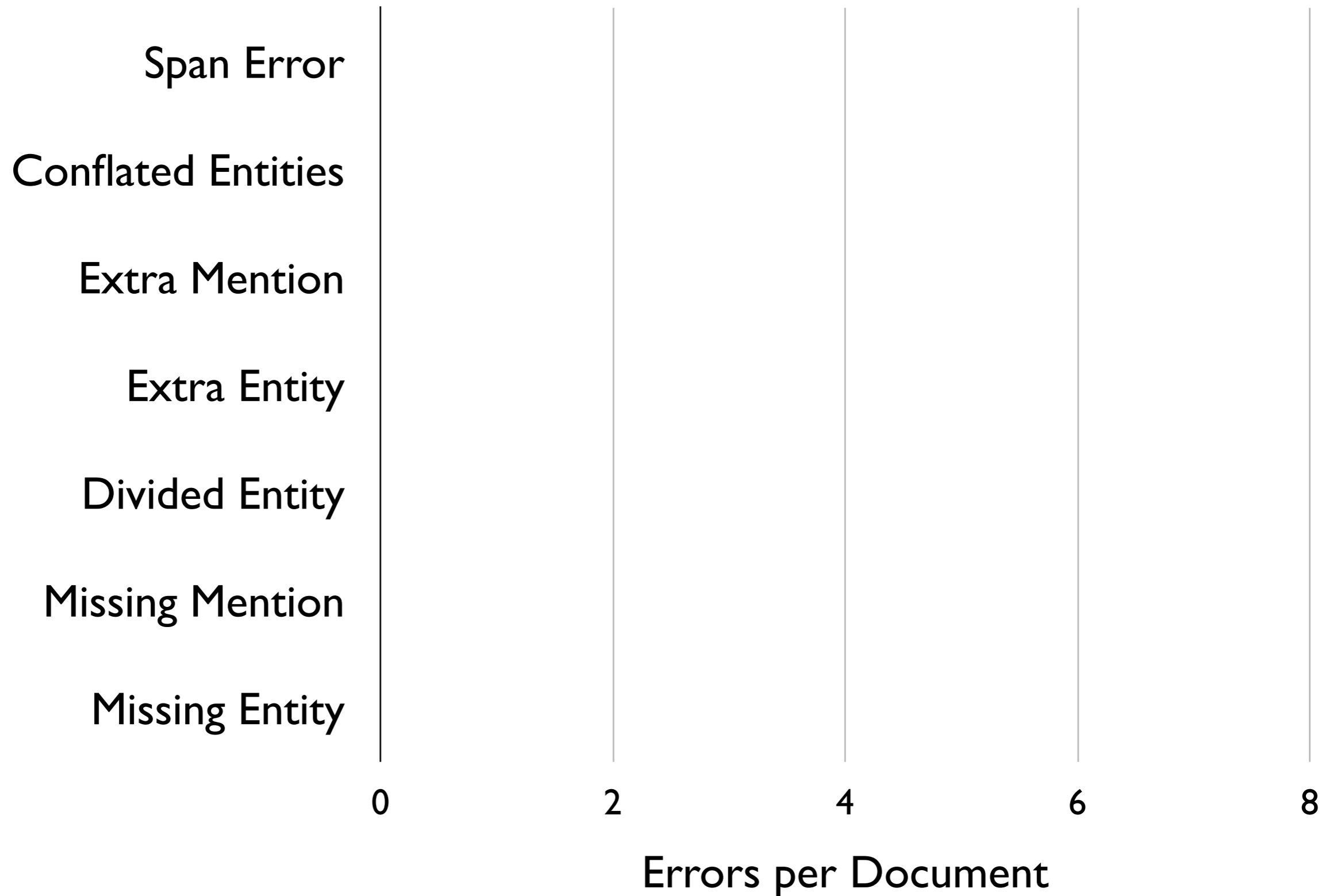


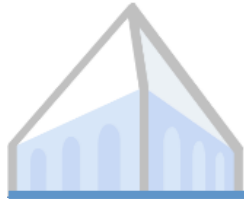
System Comparison

System	Metric F_1		Span Error	Conflated Entities	Extra Mention	Extra Entity	Divided Entity	Missing Mention	Missing Entity
	MUC	B^3							
Publicly Available Systems									
Berkeley	66.43	66.17							
IMS	64.71	64.73							
Stanford	61.40	63.06							

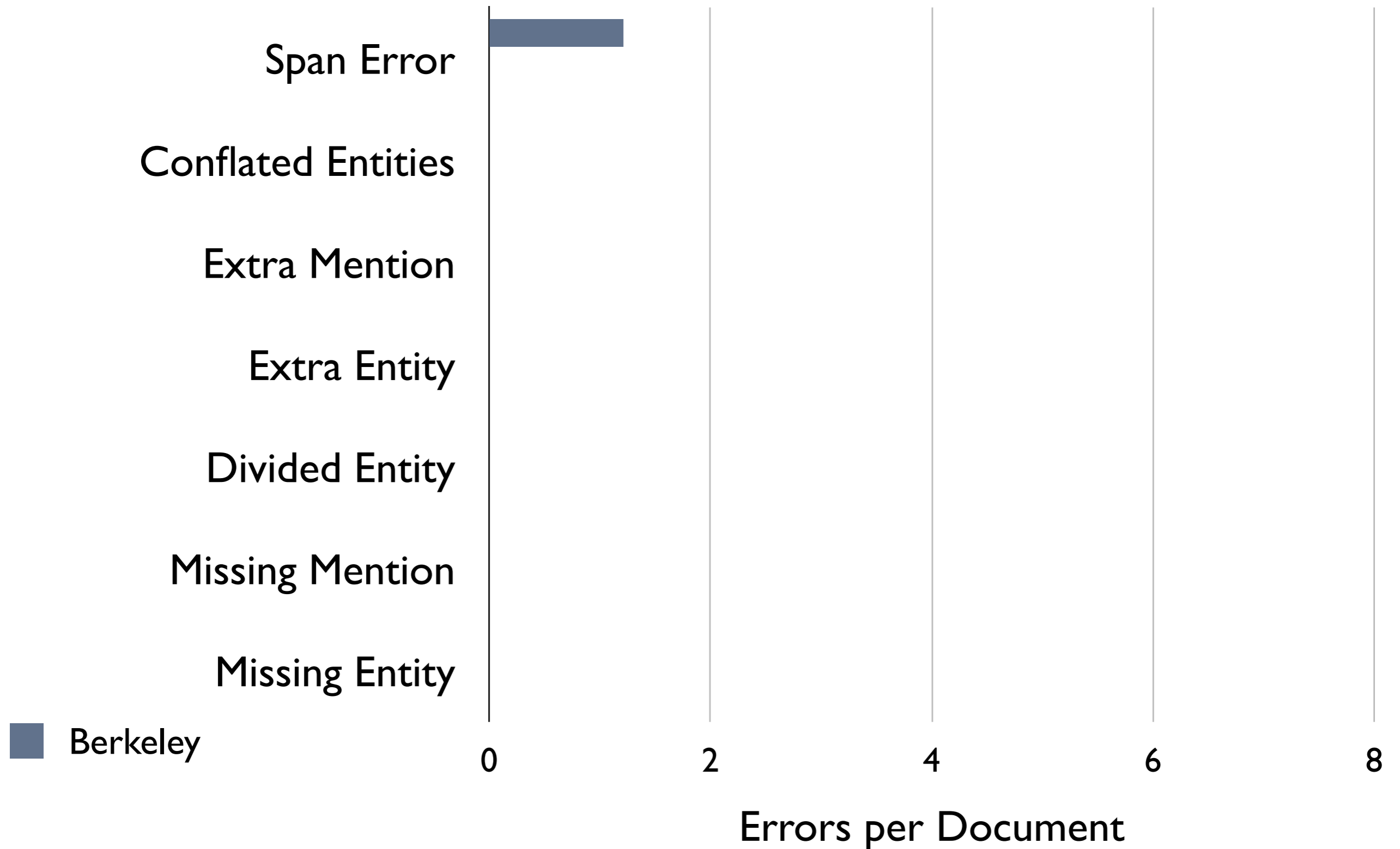


Top Public Systems



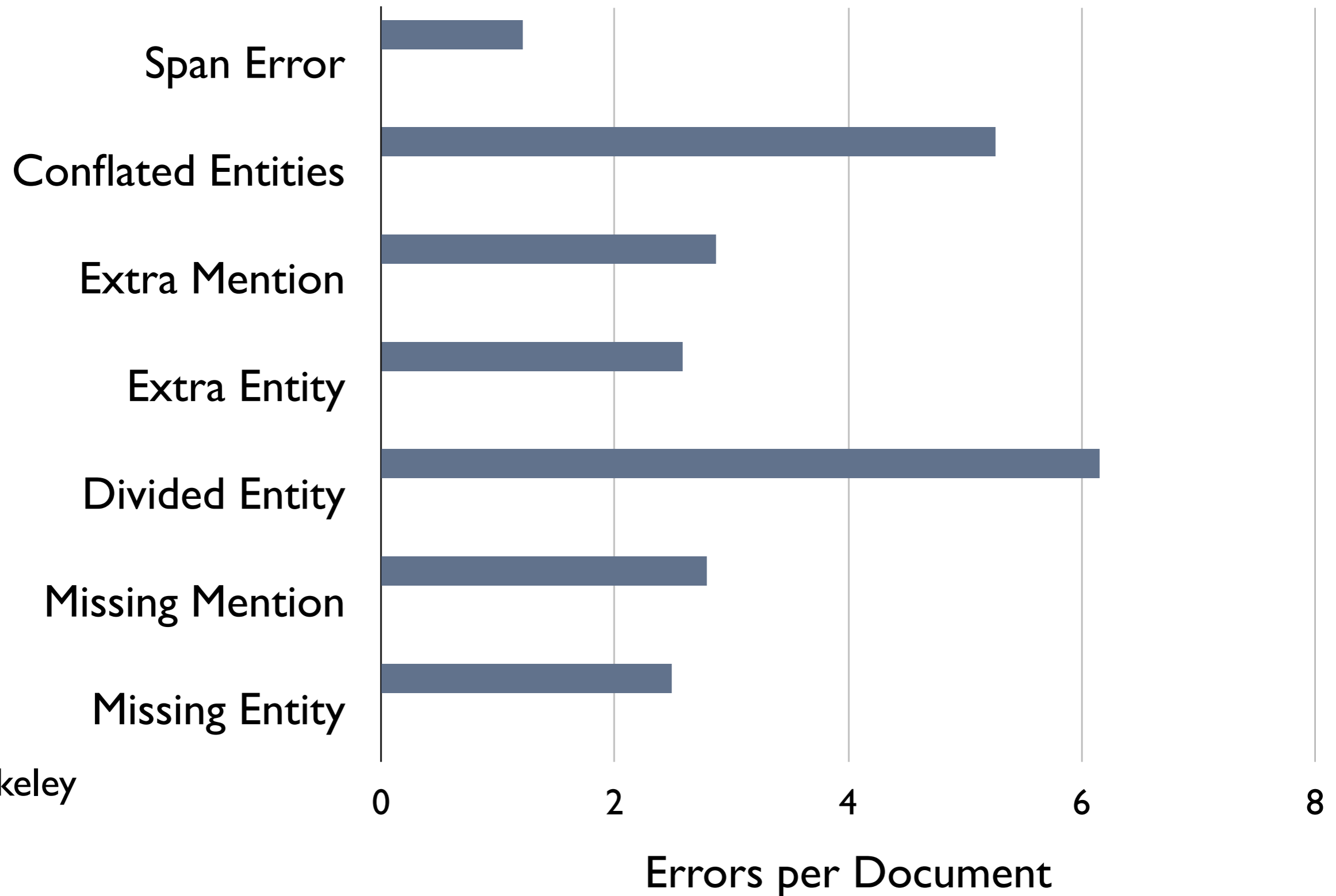


Top Public Systems



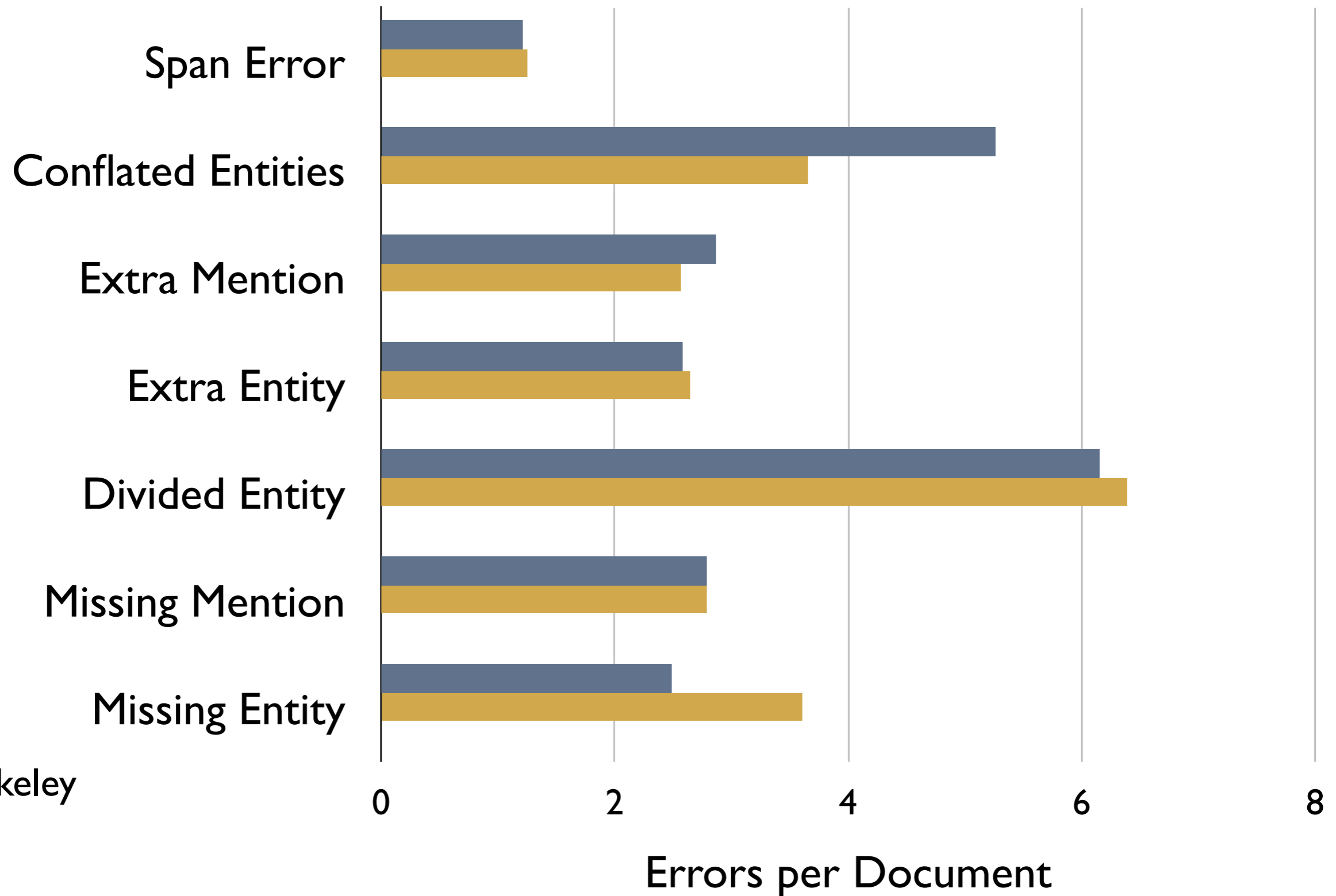


Top Public Systems

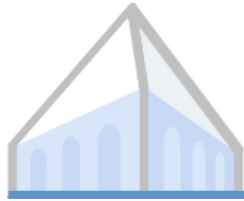




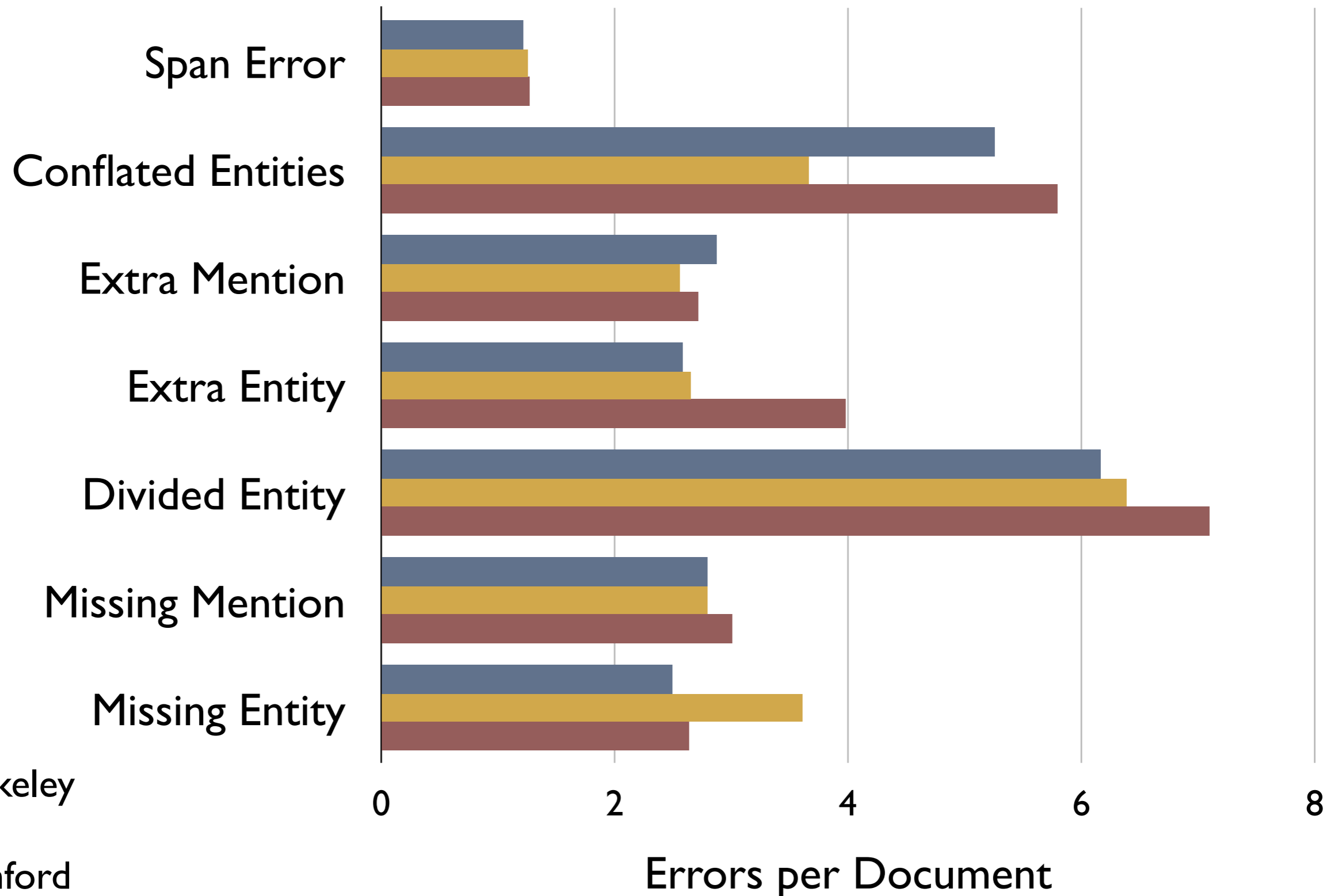
Top Public Systems

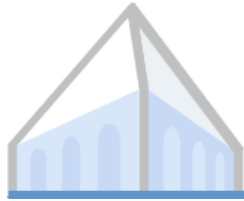


■ Berkeley
■ IMS

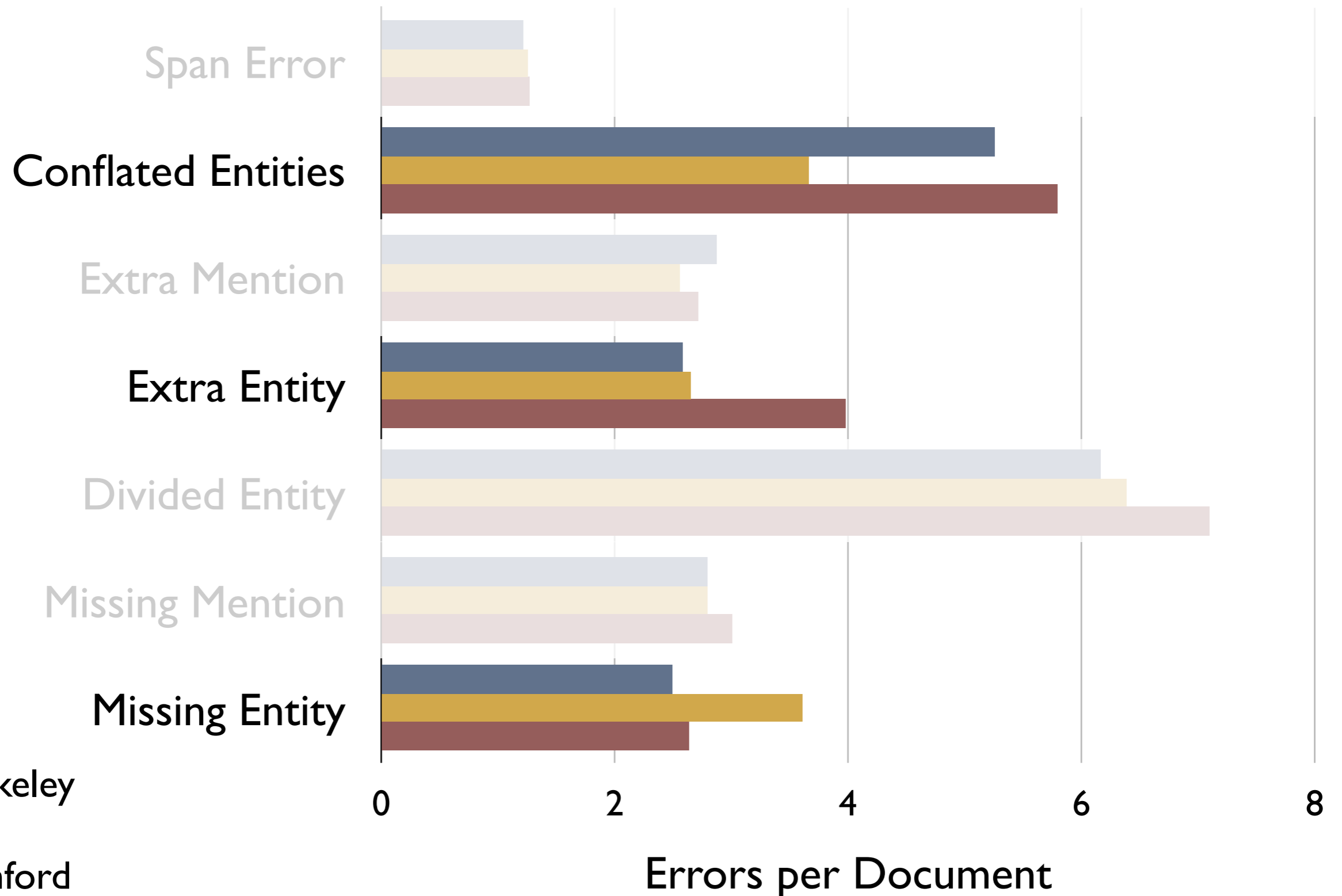


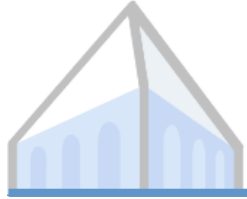
Top Public Systems





Top Public Systems





Results

- ▶ **System comparison**
- ▶ Investigation of common errors

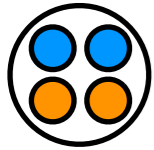
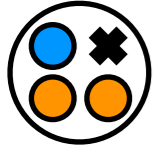


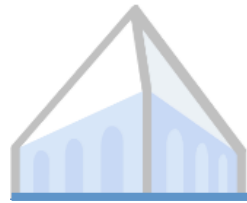
Results

- ▶ System comparison
- ▶ Investigation of common errors

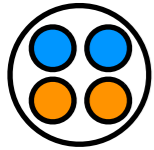
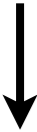


1 – Span Errors due to Parsing

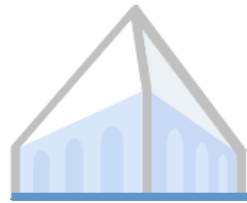




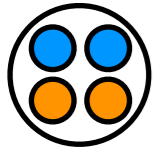
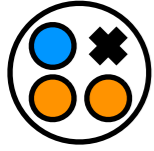
1 – Span Errors due to Parsing



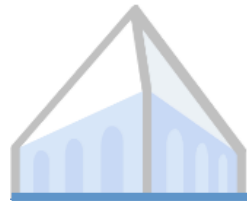
... Today **President Bush** ...



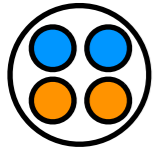
1 – Span Errors due to Parsing



... Today President Bush ...

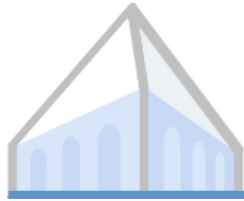


1 – Span Errors due to Parsing

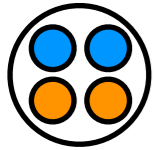
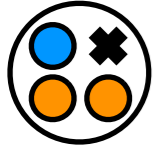


... Today President Bush ...

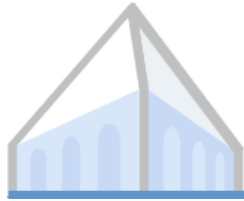
NP



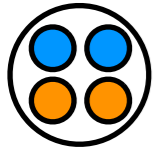
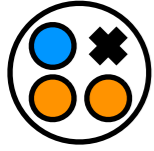
1 – Span Errors due to Parsing



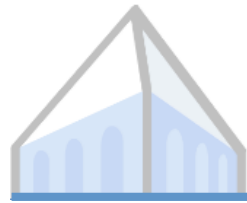
Label	Missing or Extra (per Document)
NP	0.34



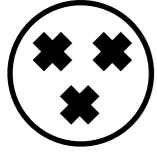
1 – Span Errors due to Parsing

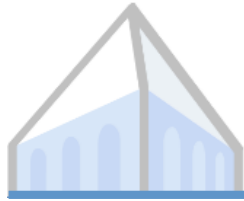


Label	Missing or Extra (per Document)
NP	0.34
POS	0.34
,	0.29
SBAR	0.18
PP	0.17

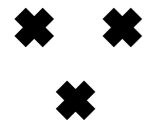
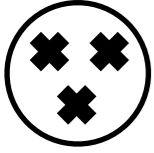


2 – Misleading String Features

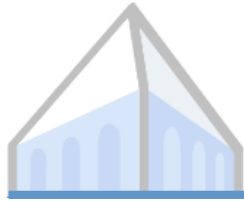




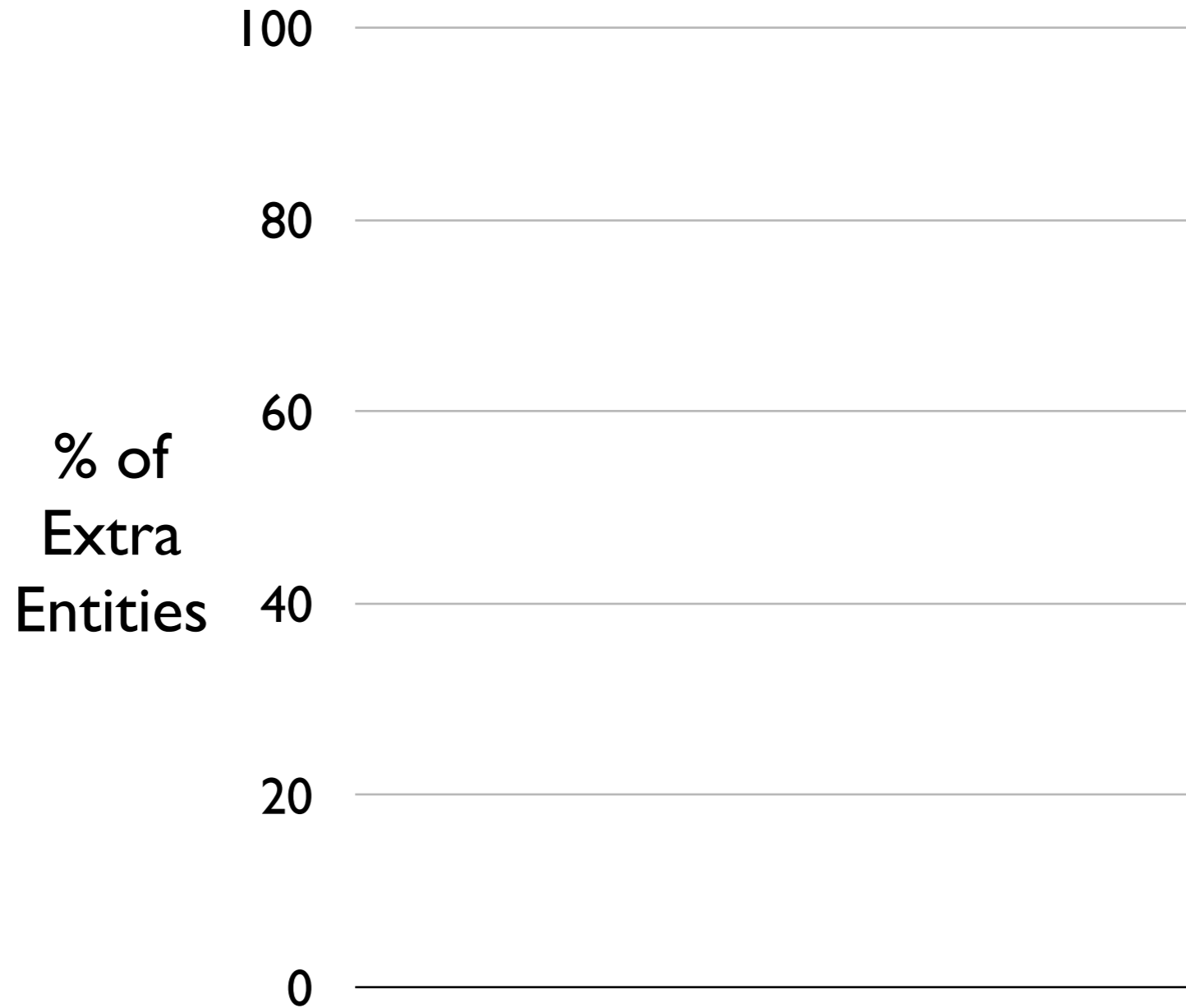
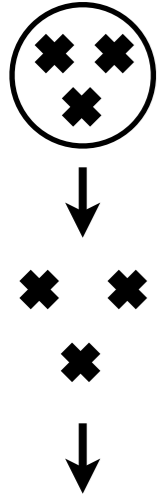
2 – Misleading String Features

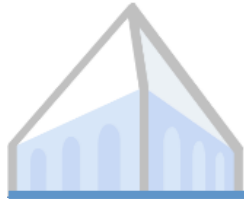


Or is it checkbook diplomacy ?
... pursuing checkbook diplomacy

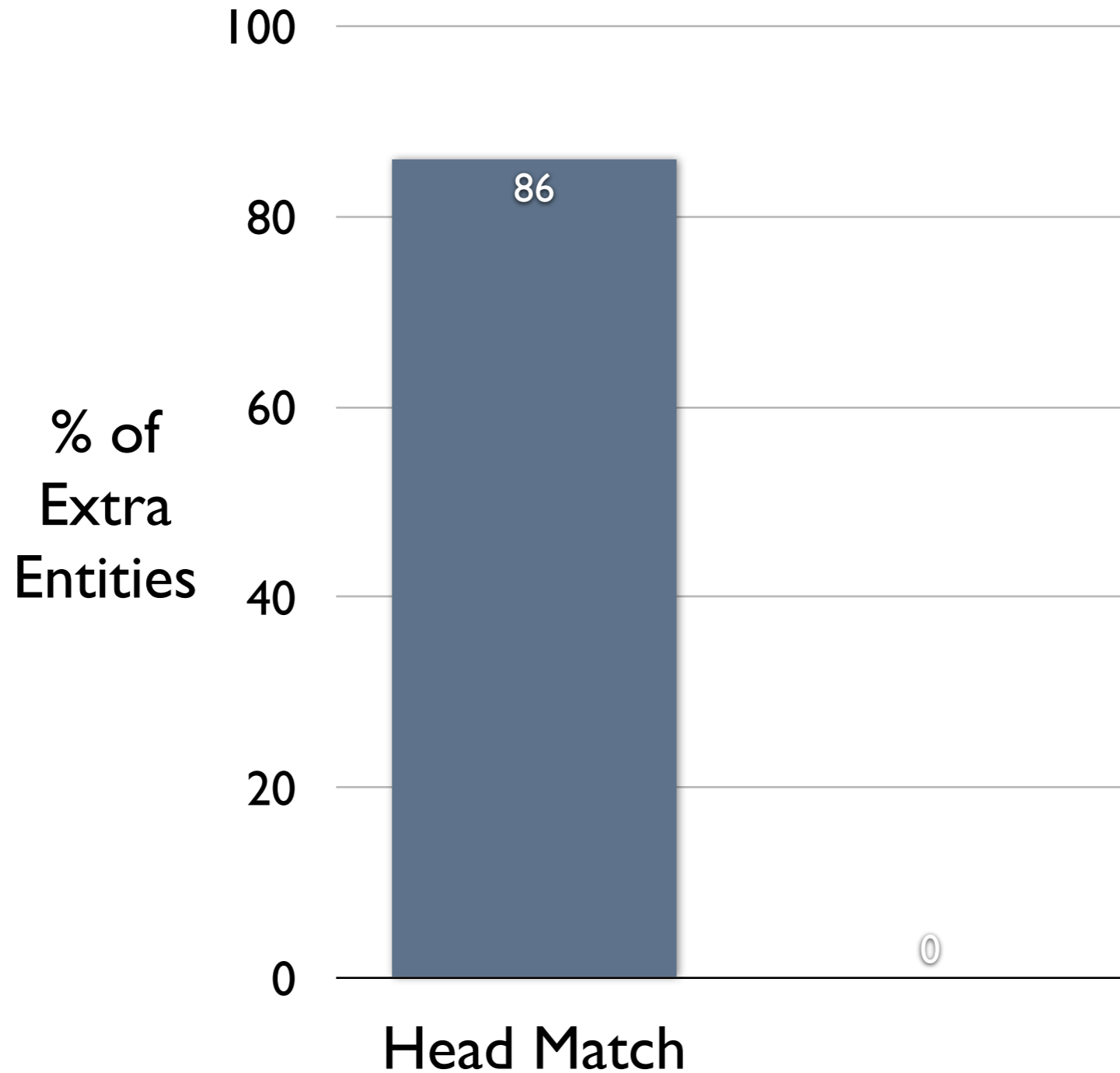
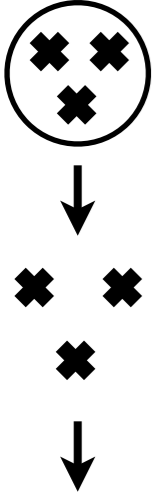


2 – Misleading String Features



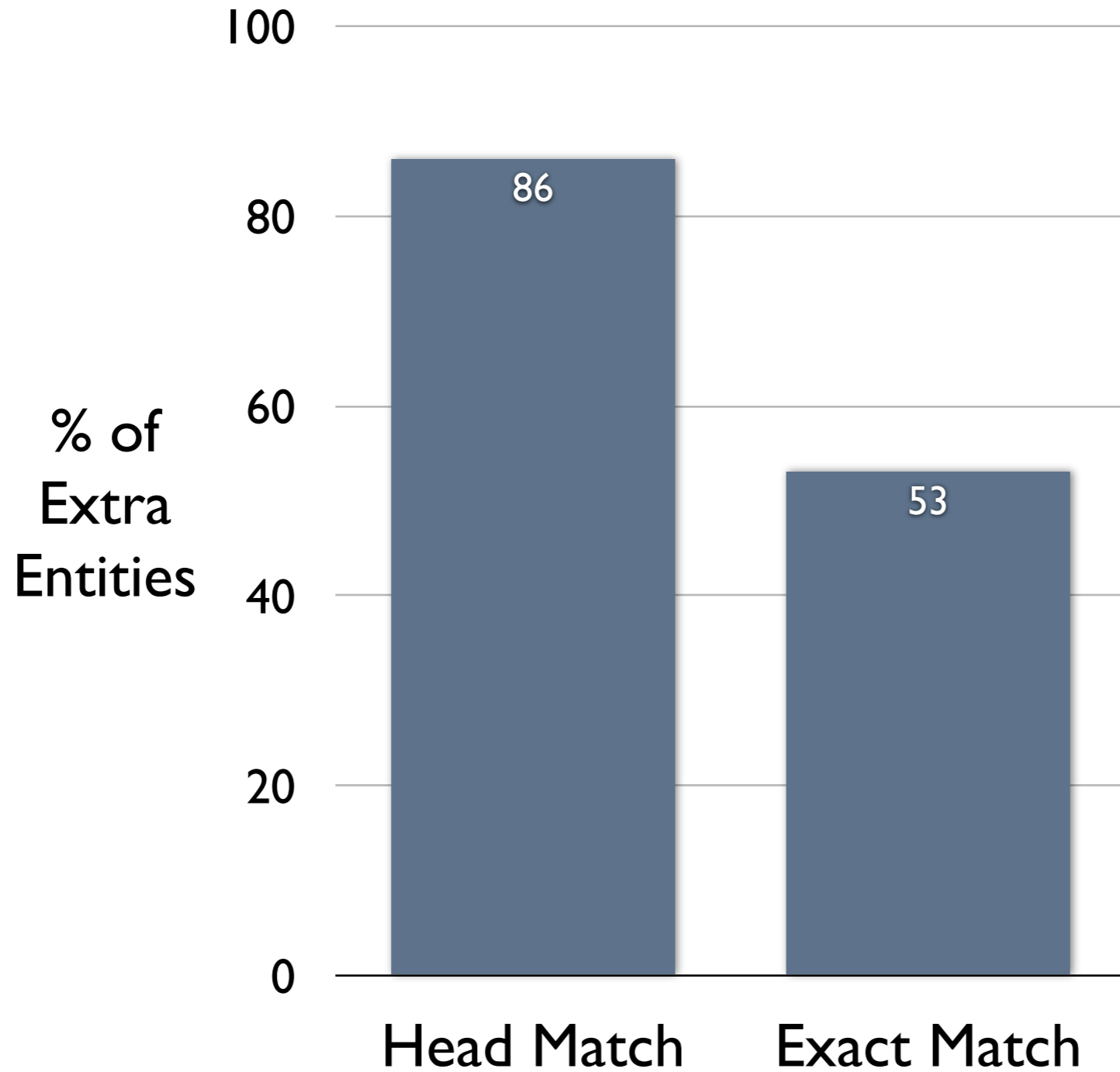
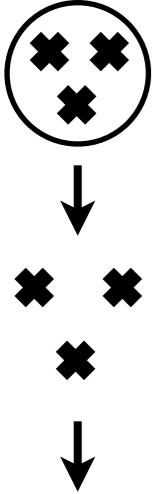


2 – Misleading String Features



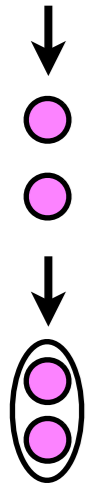


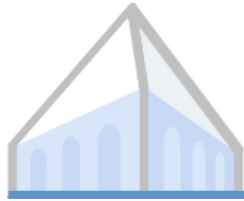
2 – Misleading String Features



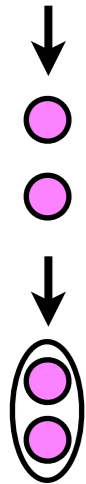


2 – Misleading String Features



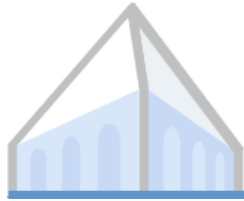


2 – Misleading String Features

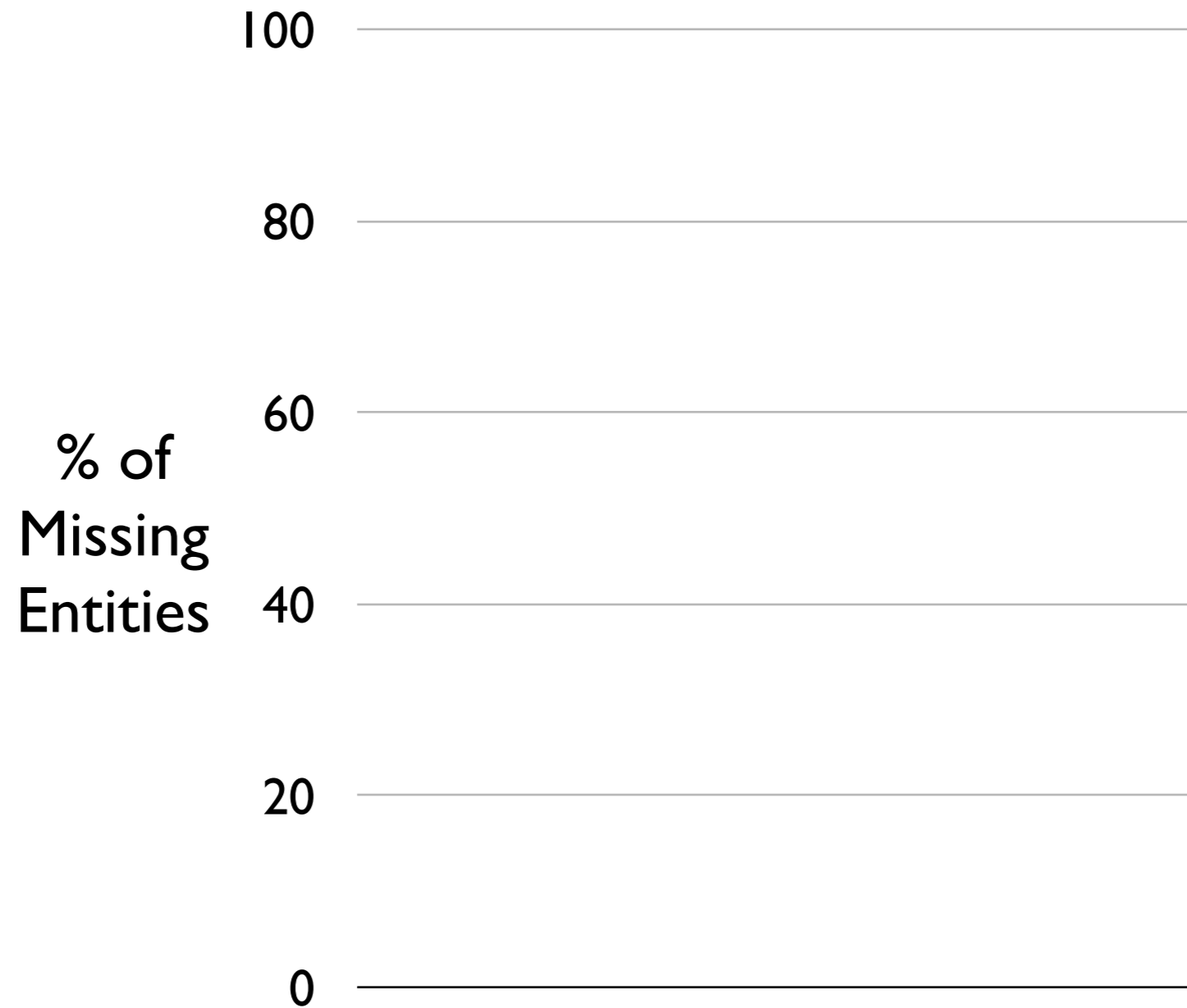
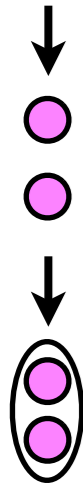


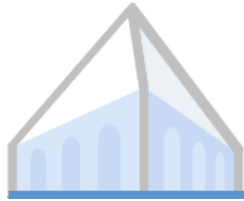
We took a **taxi** ...

... **the car** was stopped ...

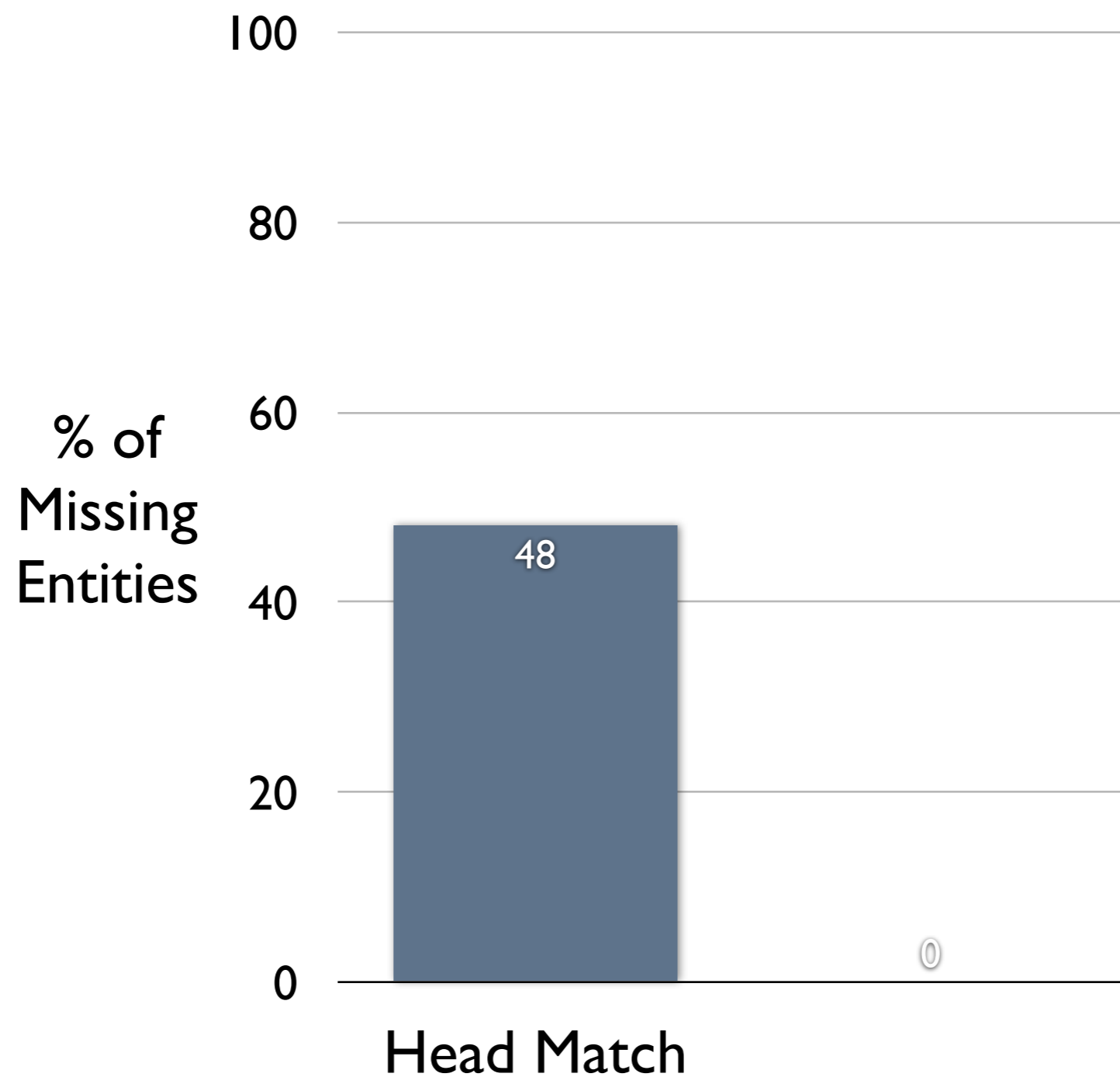
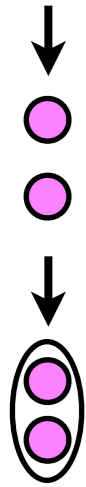


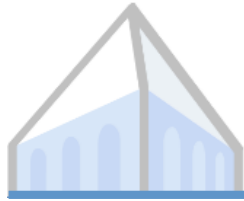
2 – Misleading String Features



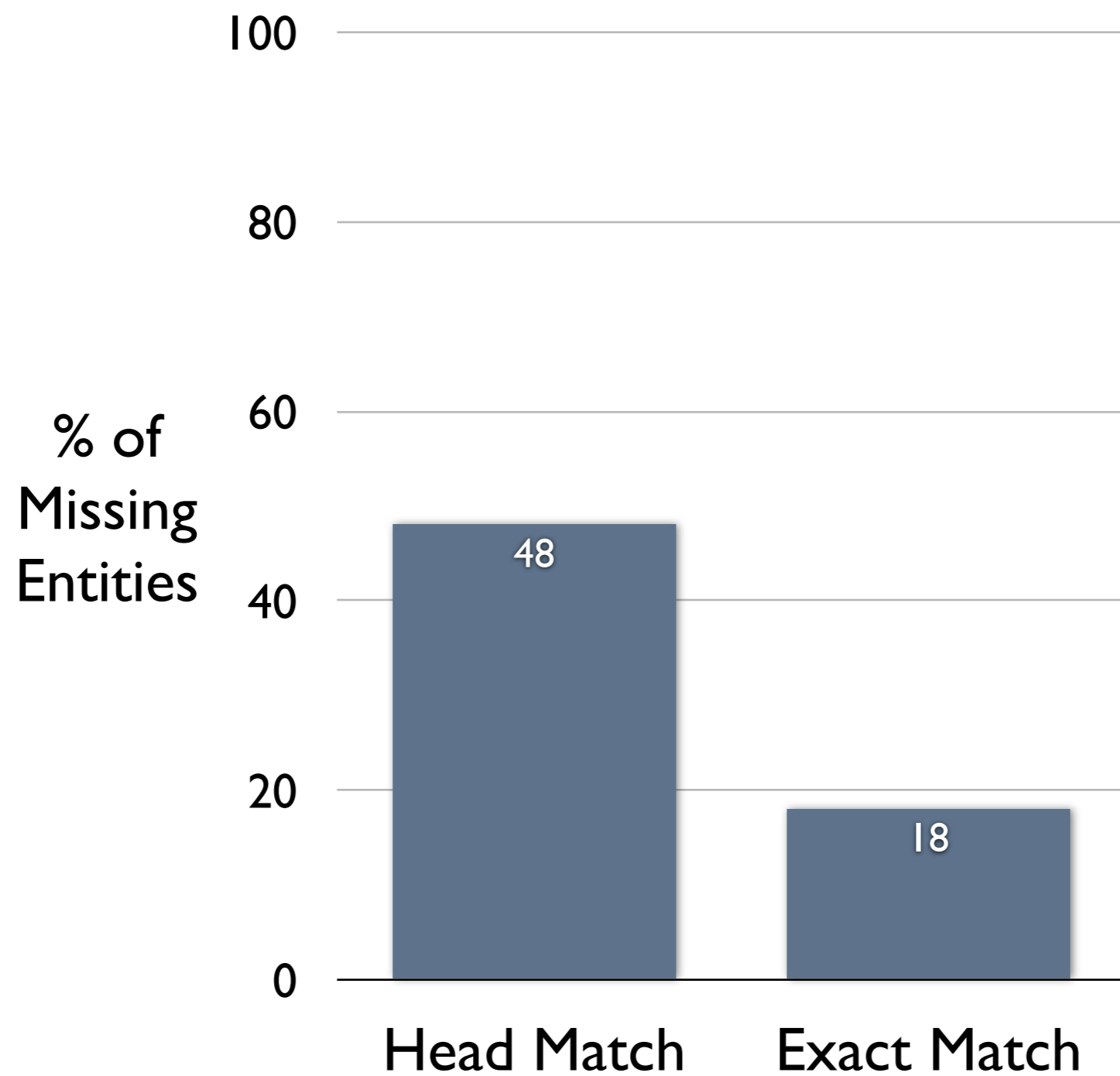
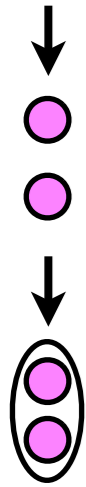


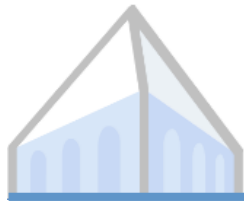
2 – Misleading String Features



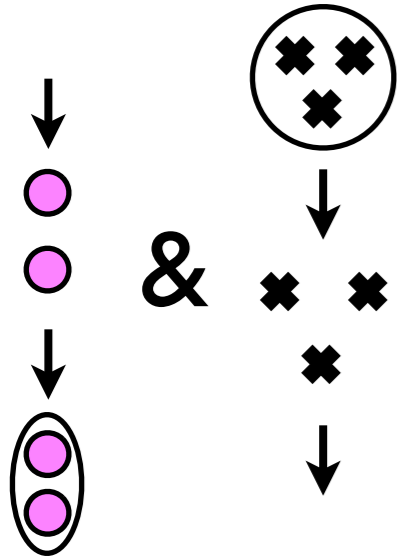


2 – Misleading String Features



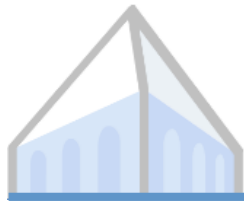


3 – Pronoun Cluster Errors

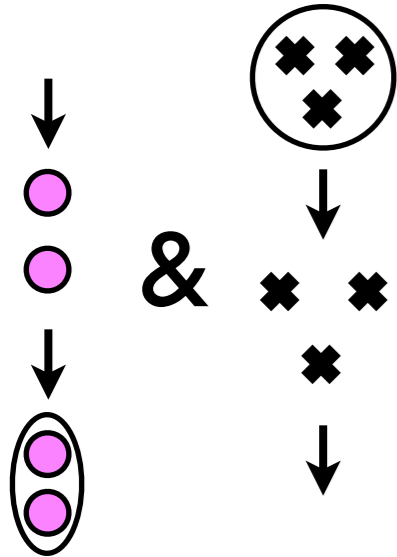


You can enjoy the most charming ...

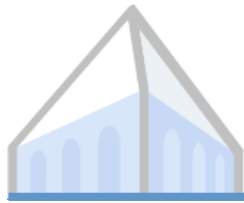
You can ascend Victoria Peak ...



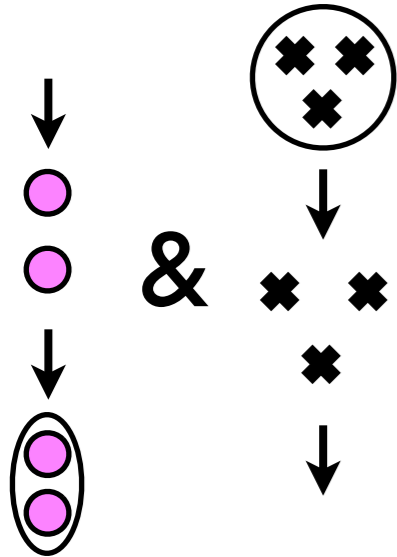
3 – Pronoun Cluster Errors



Error Type	Errors (per Document)
Extra Entity	0.63
Missing Entity	



3 – Pronoun Cluster Errors



Error Type	Errors (per Document)
Extra Entity	0.63
Missing Entity	0.1

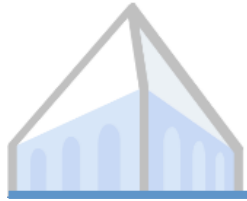


4 – Impact on Metrics



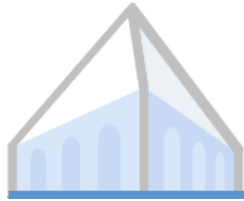
4 – Impact on Metrics

- ▶ Apply transformation to correct all errors of one type

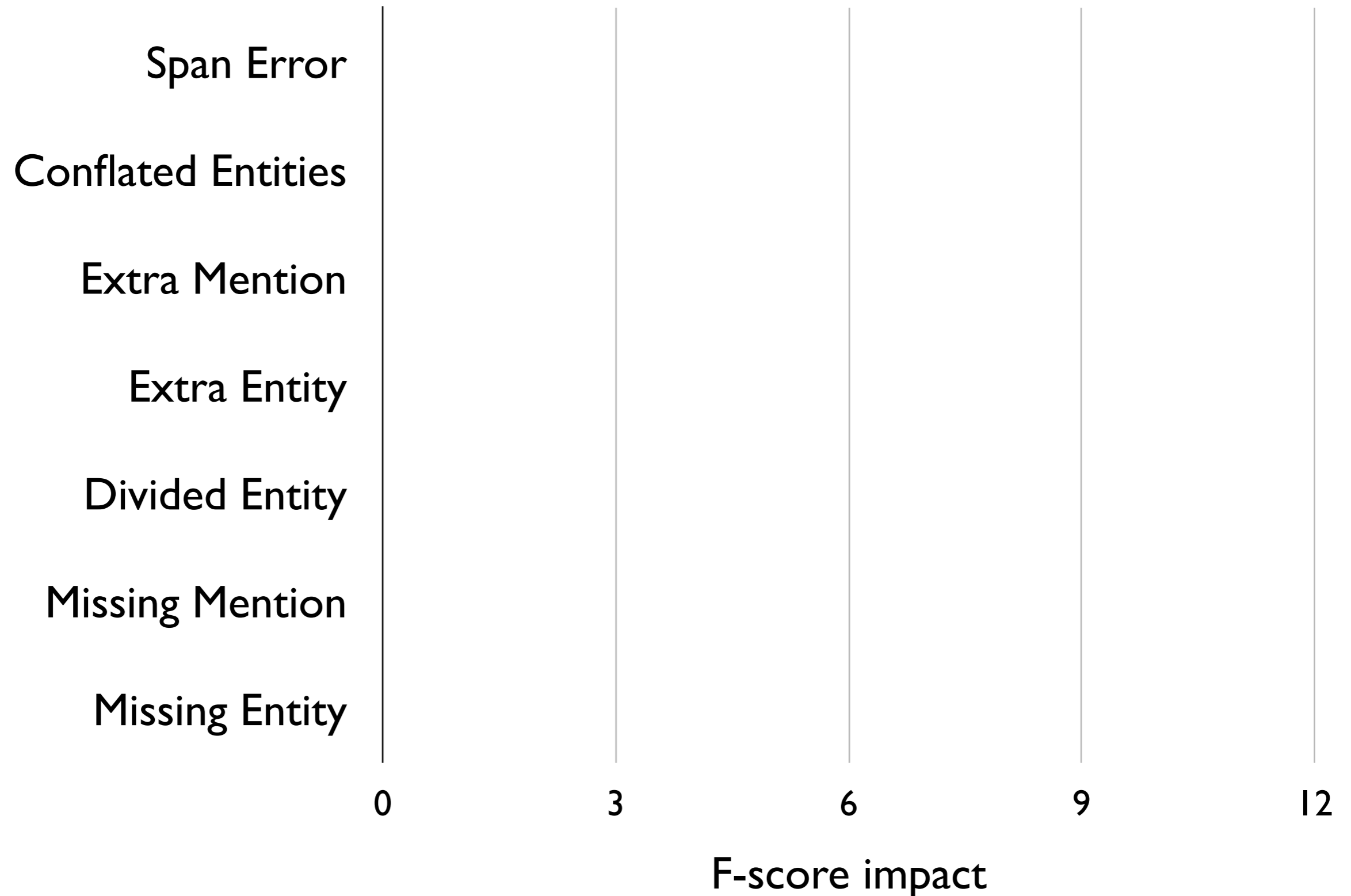


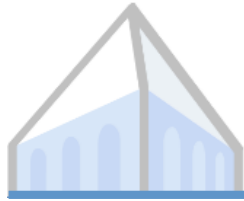
4 – Impact on Metrics

- ▶ Apply transformation to correct all errors of one type

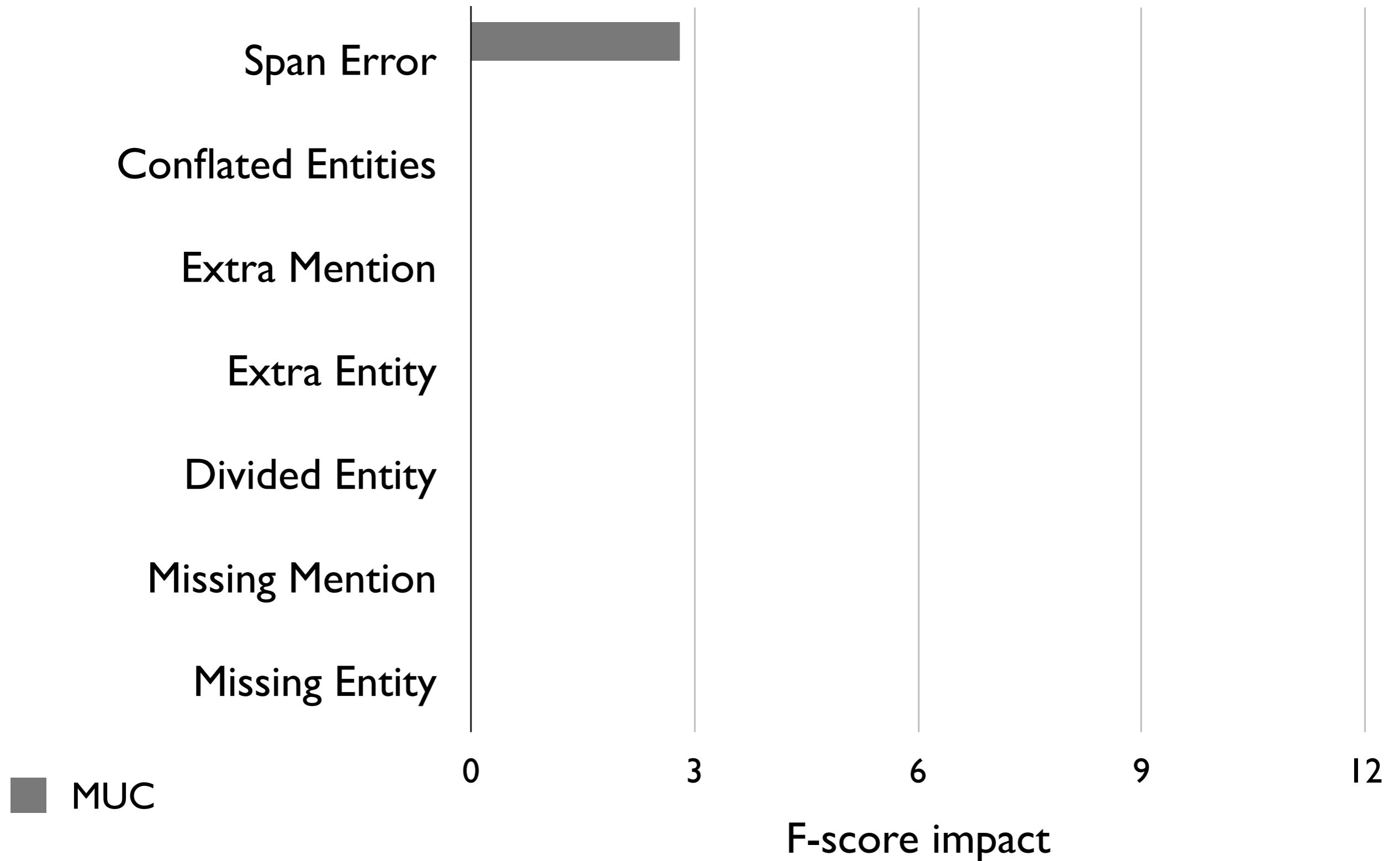


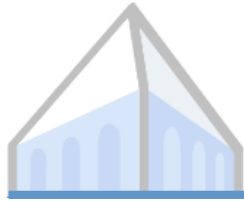
4 – Impact on Metrics



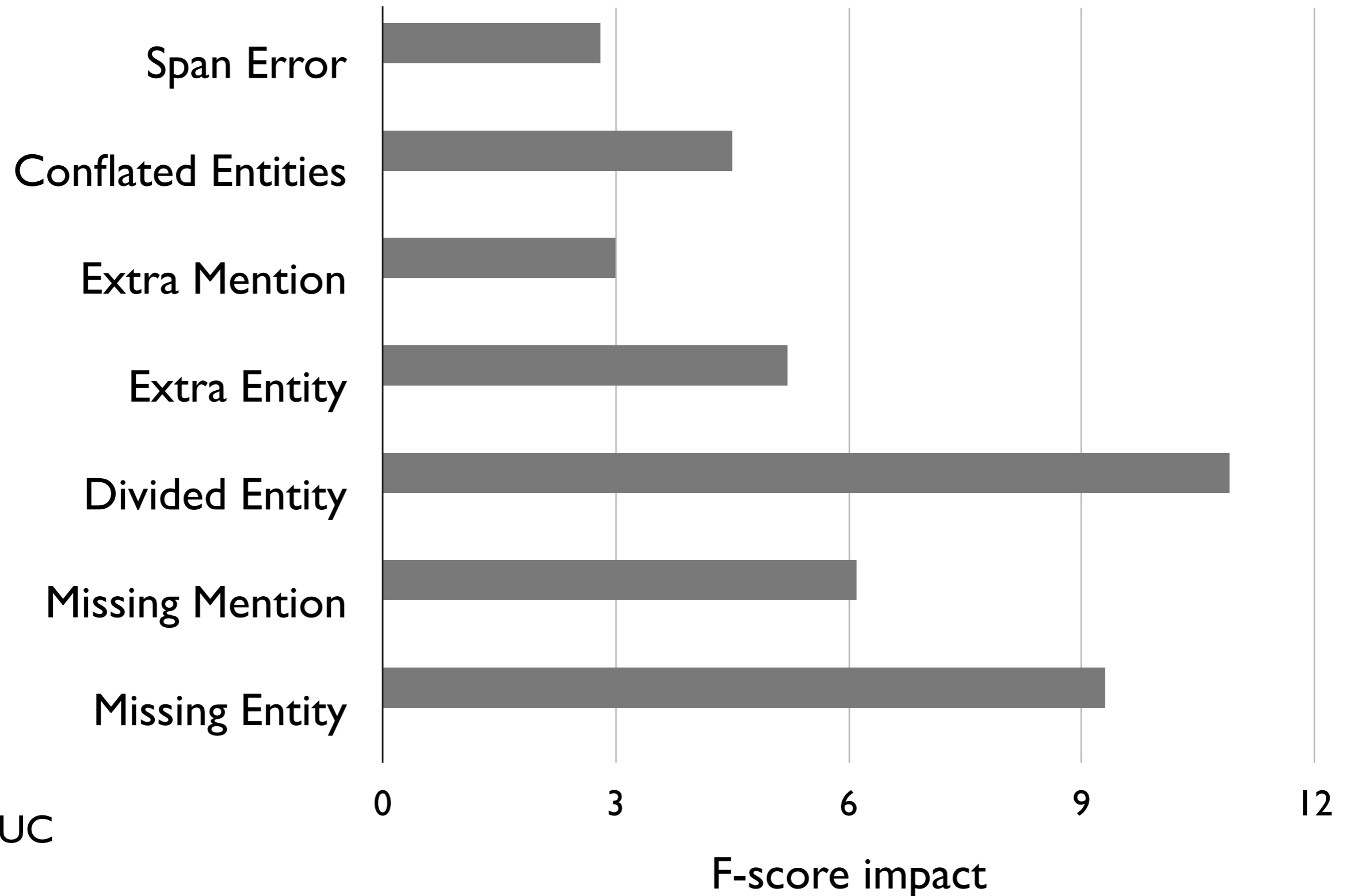


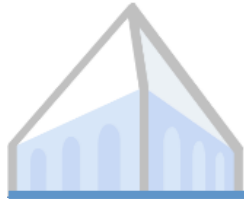
4 – Impact on Metrics



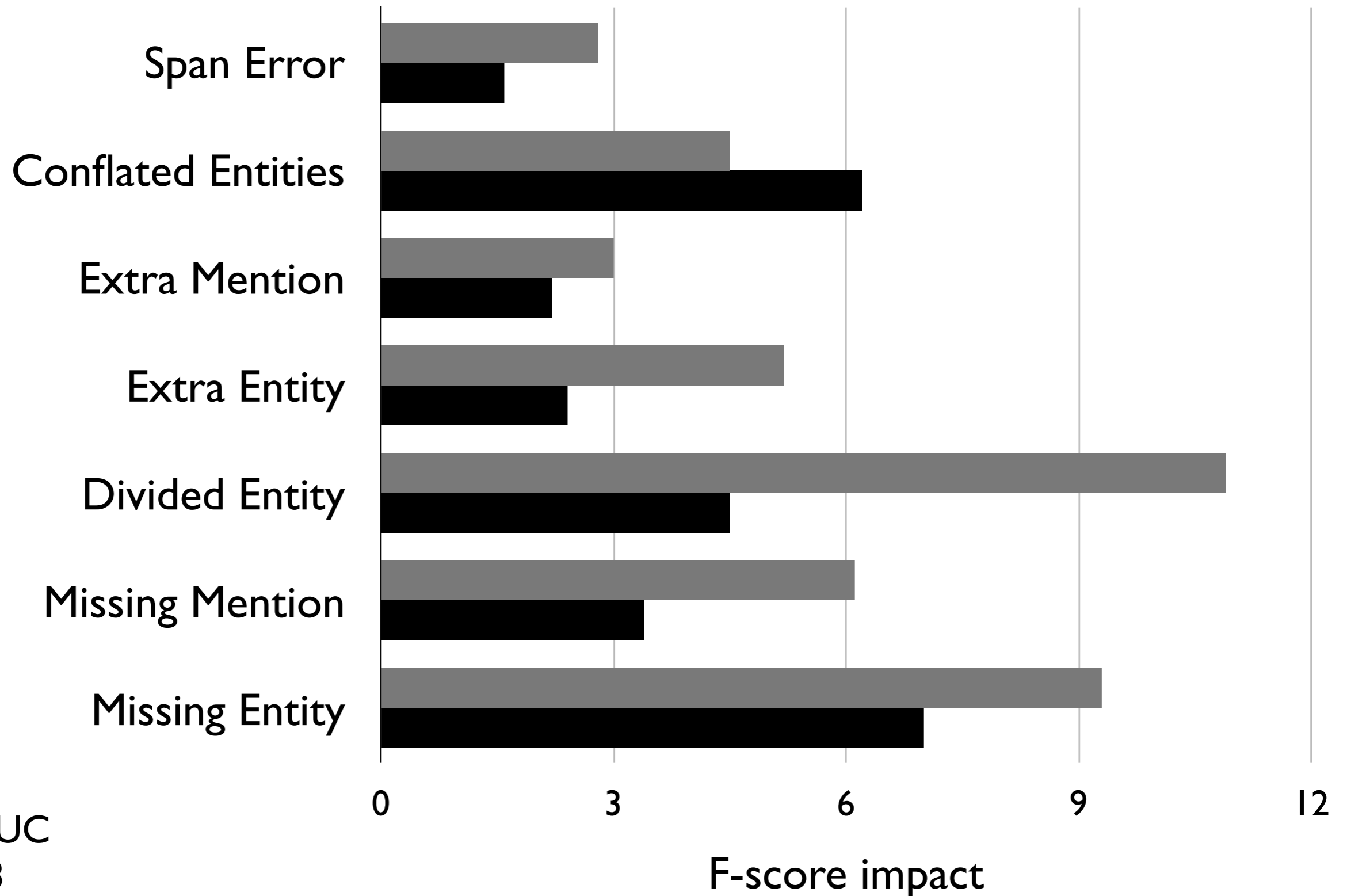


4 – Impact on Metrics



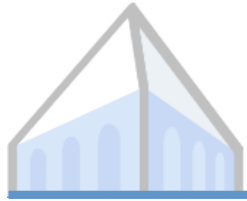


4 – Impact on Metrics



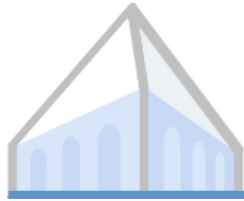


Conclusion



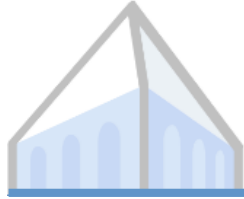
Conclusion

- ▶ Meaningful error analysis does not have to be time consuming



Conclusion

- ▶ 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!