Annotating Anaphoric Shell Nouns with their Antecedents

Varada Kolhatkar and Graeme Hirst
University of Toronto

Heike Zinsmeister
Universität Stuttgart

Anaphoric Shell Nouns (ASNs)

(Schmid, 2000)

Our paper: Annotating antecedents of ASNs such as *this issue*, *this fact*, *this possibility*

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

Anaphoric Shell Nouns (ASNs)

(Schmid, 2000)

Our paper: Annotating antecedents of ASNs such as *this issue*, *this fact*, *this possibility*

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers v[antecedent]meras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

anaphoric shell noun (ASN)

Why ASN annotation?

Ubiquity of Shell Nouns

- Occur frequently in all kinds of texts
- *fact*, *idea*, *problem*: among 100 most frequently occurring nouns (Schmid 2000)
- ~25 million occurrences in the NYT corpus (~1.3 billion tokens)

Organizing Discourse

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- *Characterize* and *label* chunks of information
- Cohesive devices
- Topic boundary markers

Organizing Discourse

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- *Characterize* and *label* chunks of information
- Cohesive devices
- Topic boundary markers

Organizing Discourse

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- *Characterize* and *label* chunks of information
- Cohesive devices
- Topic boundary markers

Organizing Discourse

idea

mixed views

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- *Characterize* and *label* chunks of information
- Cohesive devices
- Topic boundary markers

Current Research: Gap

Corpus	ASN instances
Poesio and Artstein, 2008 ARRAU	455 abstract anaphor instances, very few ASNs
Kolhatkar and Hirst, 2012	188 *this issue* instances from Medline abstracts
Botley, 2006	462 ASN instances (not available)

Current Research: Gap

Corpus	ASN instances
Poesio and Artstein, 2008 ARRAU	455 abstract anaphor instances, very few ASNs
Kolhatkar and Hirst, 2012	188 *this issue* instances from Medline abstracts
Botley, 2006	462 ASN instances (not available)

ASNs are largely ignored in CL

Need a large-scale ASN antecedent corpus

Annotation challenges

What to Annotate?

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- Antecedents: complex and abstract entities
- Heterogeneous set of markables
  (e.g., NPs, VPs, sentences, clauses, ...)

What to Annotate?

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

- Antecedents: complex and abstract entities
- Heterogeneous set of markables
  (e.g., NPs, VPs, sentences, clauses, ...)

Leads to large search space

What's the "right" answer?

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

What's the "right" answer?

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

What's the "right" answer?

whether to

rk is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

Annotation data

Shell Nouns: Categorization

(Schmid, 2000)

Category	Examples
Factual	*fact, problem, reason*
Linguistic	*news, proposal, question*
Mental	*idea, belief, decision, issue*
Modal	*possibility, need, trend*
Eventive	*act, reaction, attempt*
Circumstantial	*situation, approach*

Shell Nouns: Selection

	Category	Examples
✔	Factual	*fact, problem, reason*
✔	Linguistic	*news, proposal, question*
✔	Mental	*idea, belief, decision, issue*
✔	Modal	*possibility, need, trend*
	~~Eventive~~	~~*act, reaction, attempt*~~
	~~Circumstantial~~	~~*situation, approach*~~

Shell Nouns: Selection

	Category	Examples
✔	Factual	*fact*, *problem*, *reason*
✔	Linguistic	*news*, *proposal*, *question*
✔	Mental	*idea*, *belief*, *decision*, *issue*
✔	Modal	*possibility*, *need*, *trend*
	~~Eventive~~	~~*act*, *reaction*, *attempt*~~
	~~Circumstantial~~	~~*situation*, *approach*~~

High-frequency nouns from each category

The ASN Corpus

- Base corpus: The New York Times corpus (Sandhaus, 2008)
- ~475 instances per 6 selected shell nouns
  fact, reason, issue, decision, question, possibility
- Total: 2,822 ASN instances

Annotation methodology

Annotation Platform

- Crowdsourcing: CrowdFlower
  - Quality control
    - Gold questions
    - Training phase
- Detailed results
  - Aggregated and full results with annotators' demographic information

CrowdFlower Confidence

Annotator	Trust	Answer
A	0.75	"a_1"
B	0.75	"a_2"
C	1.0	"a_2"

CrowdFlower Confidence

CrowdFlower

Annotator	Trust	Answer
A	0.75	"a_1"
B	0.75	"a_2"
C	1.0	"a_2"

score for "a_2" = 1.75

CrowdFlower Confidence

Annotator	Trust	Answer
A	0.75	"a_1"
B	0.75	"a_2"
C	1.0	"a_2"

score for "a_1" = 0.75

score for "a_2" = 1.75

Crowd's answer:
"a_2" with confidence 0.7=1.75/(1.75+0.75)

Annotation Tasks

ASN instances from the NYT

CrowdFlower Expt. 1

Identify the sentence containing antecedent

CrowdFlower Expt. 2

Identify the precise antecedent

Annotated ASN Corpus

Simple tasks do best with crowdsourcing
(Madnani et al. 2010, Wang et al. 2012)

CrowdFlower Expt. 1

Identifying the sentence containing the antecedent

(a3) New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. (a2) Some lawmakers worry that cameras might compromise the rights of the litigants. (a1) But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. (b) New York's backwardness on this issue hurts public confidence in the judiciary...

CrowdFlower Expt. 1

Identifying the sentence containing the antecedent

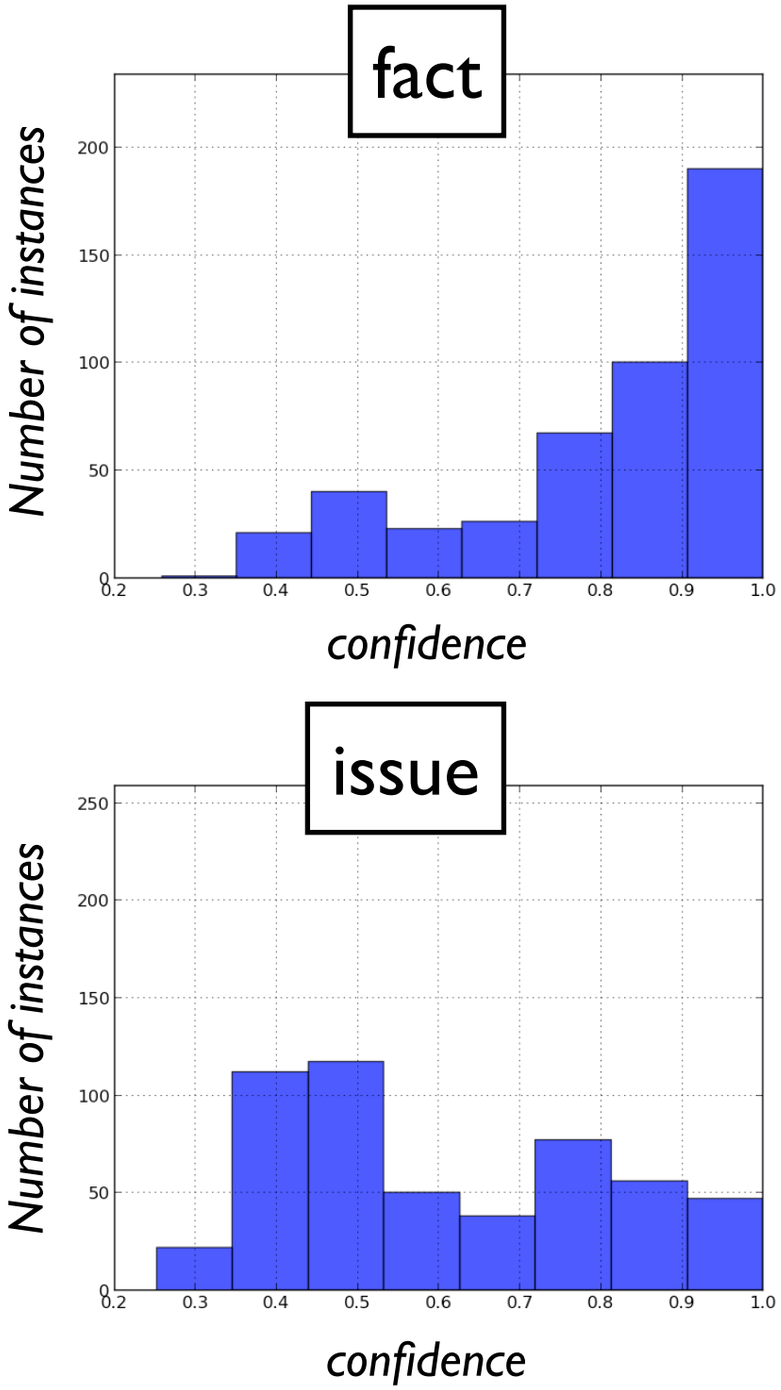
(a3) New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. (a2) Some lawmakers worry that cameras might compromise the rights of the litigants. (a1) But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. (b) New York's backwardness on this issue hurts public confidence in the judiciary...

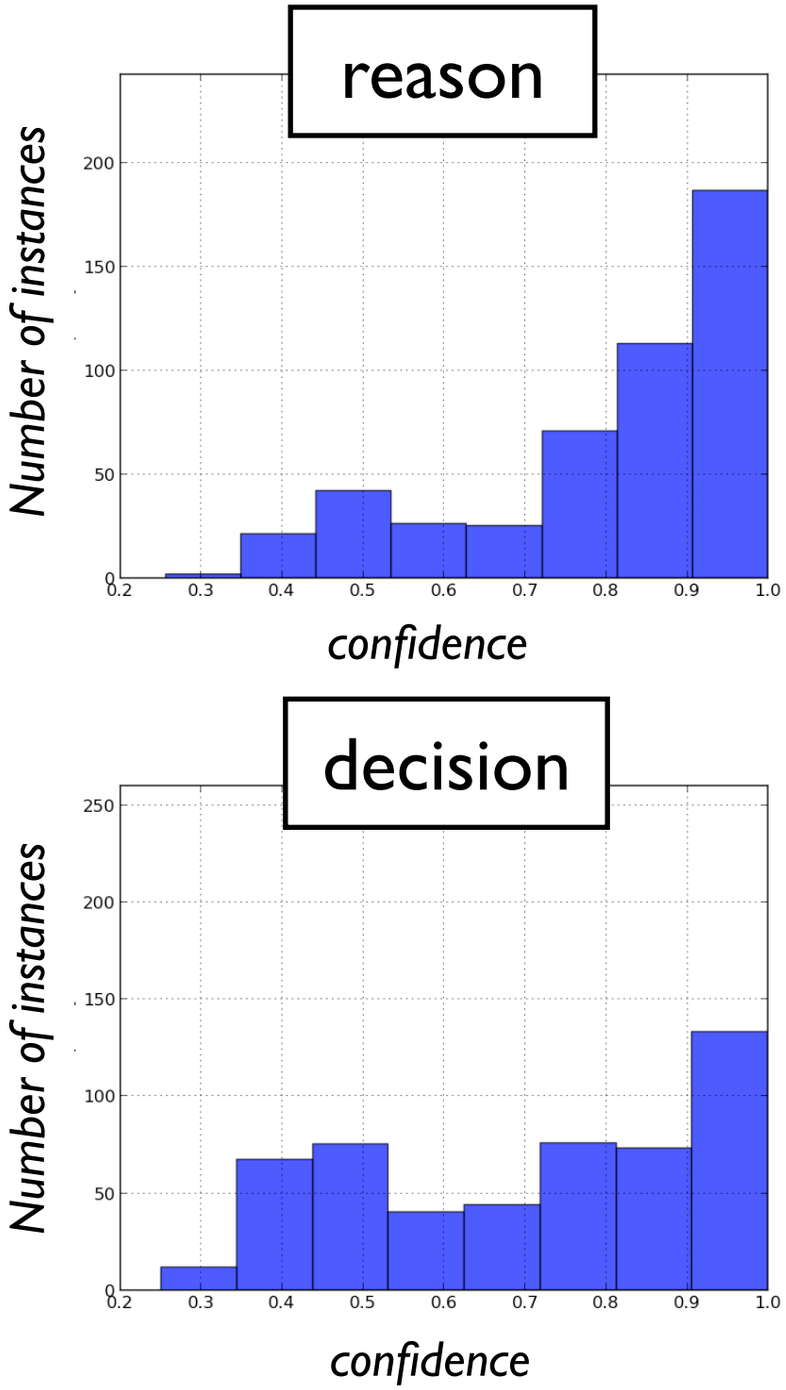
Settings

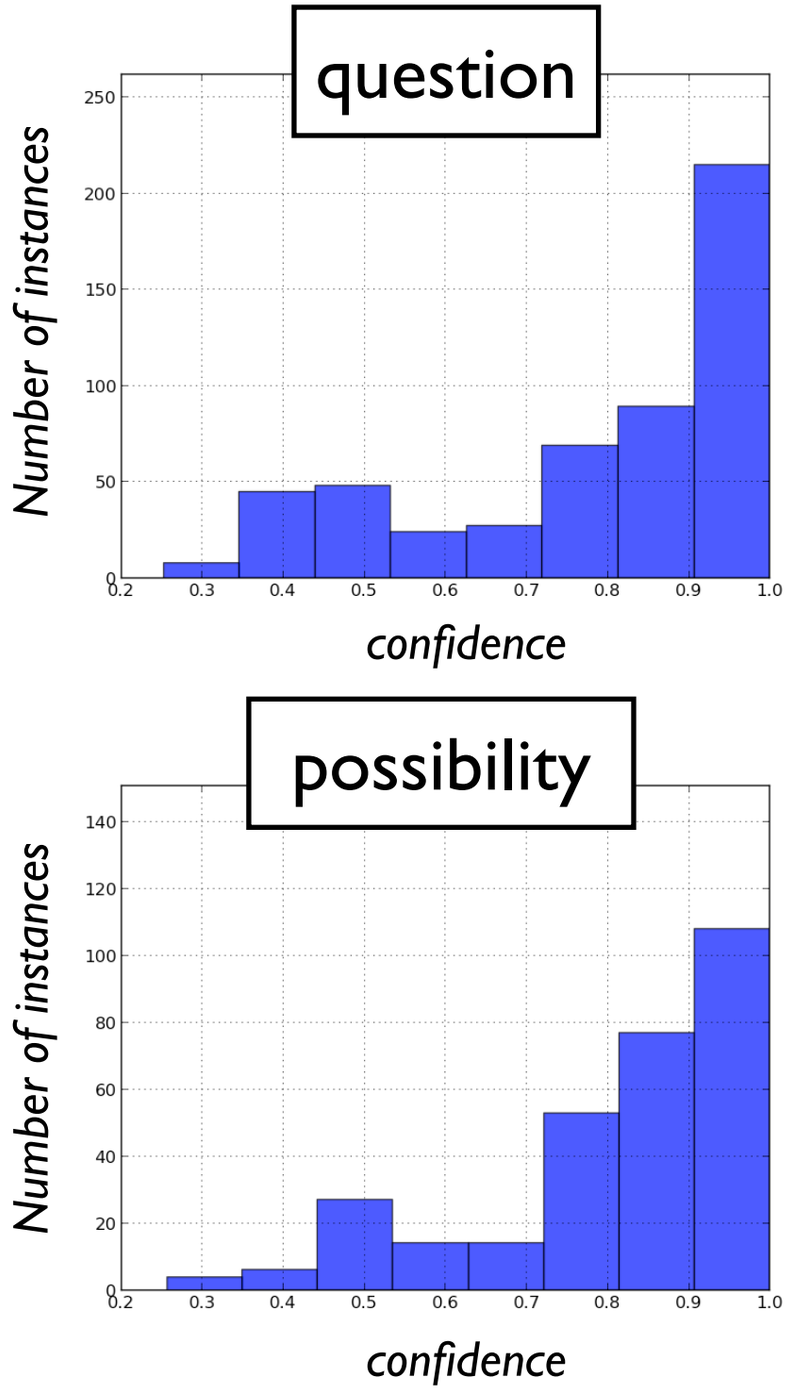
- 2,822 instances
- 8 judgements per instance
- 8 cents per annotation
- Completion time: 3 days

Inter-annotator agreement for expt. 1

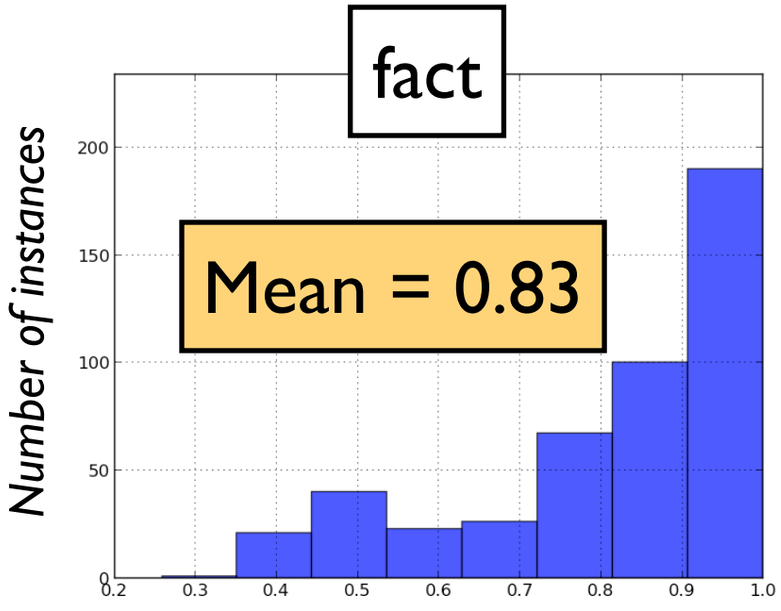
Confidence Distribution

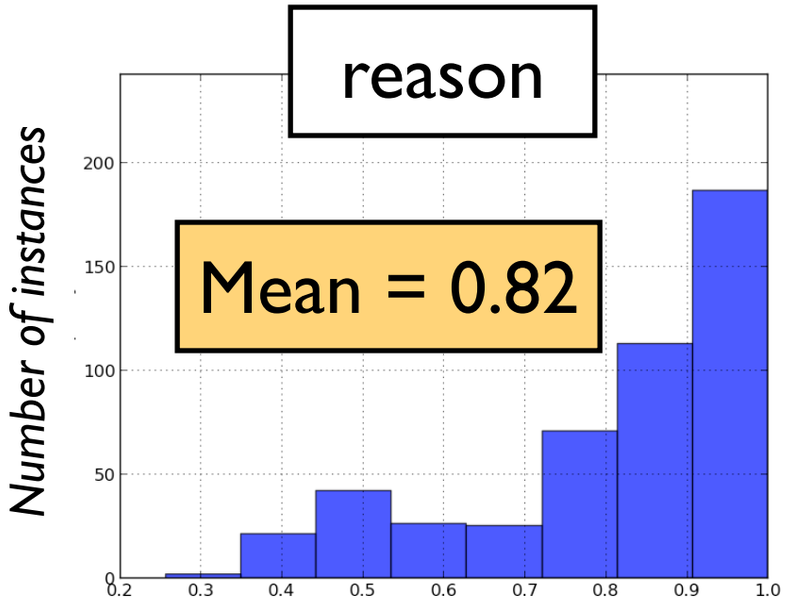


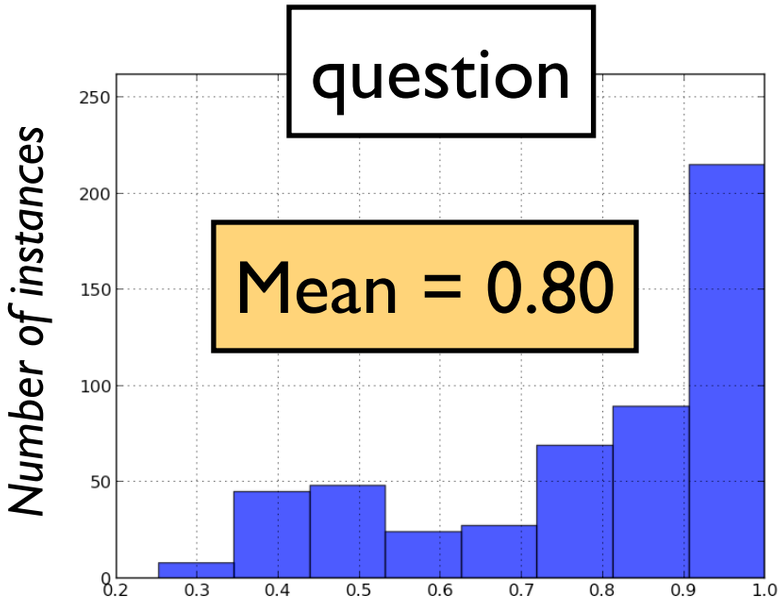


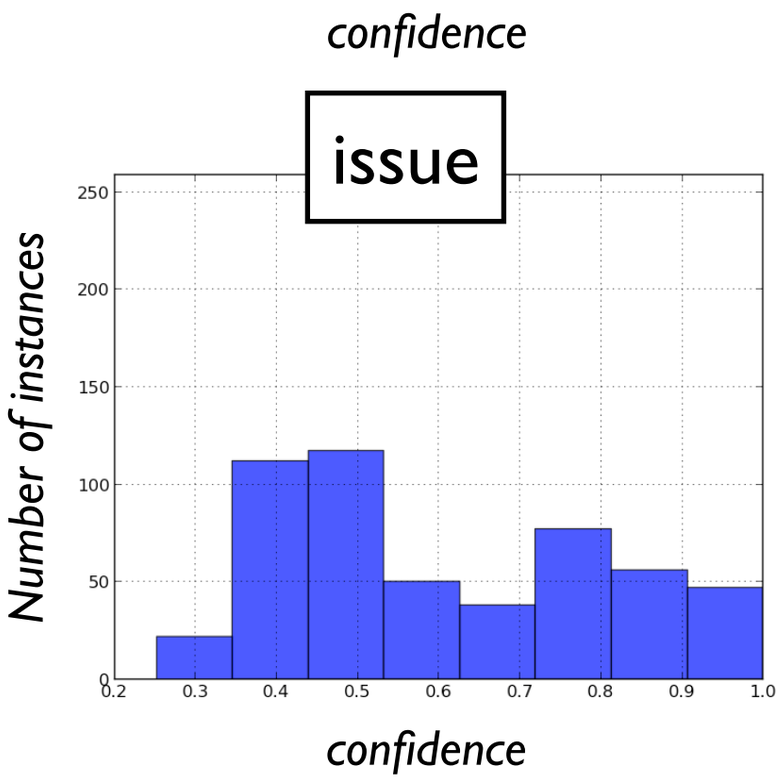


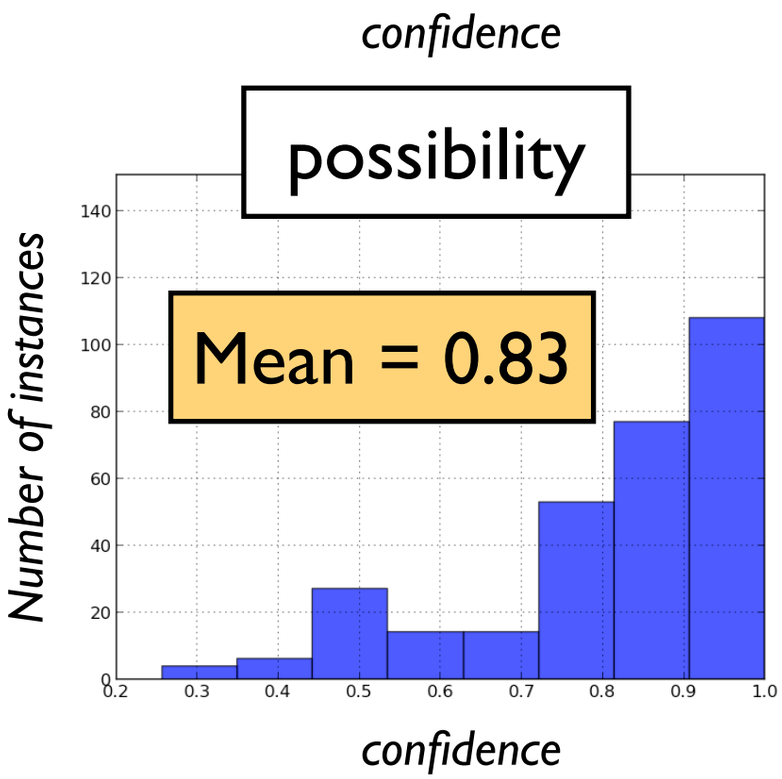
Confidence Distribution



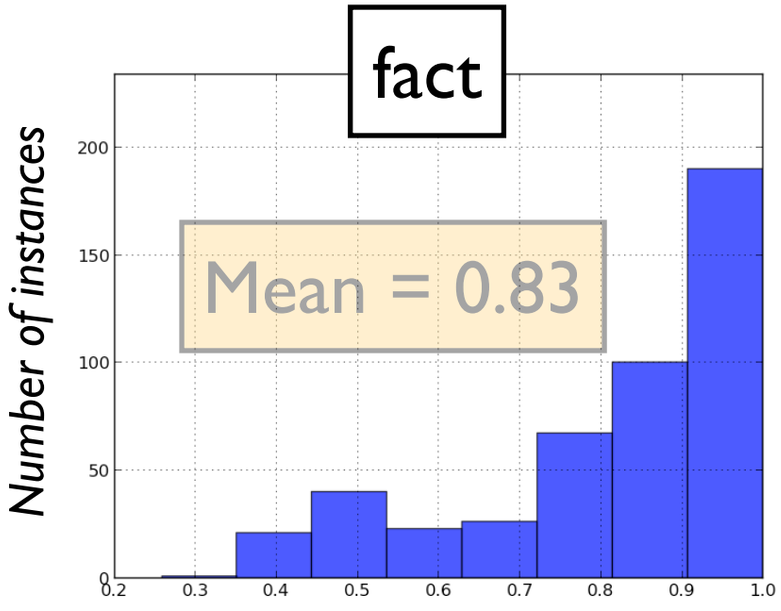


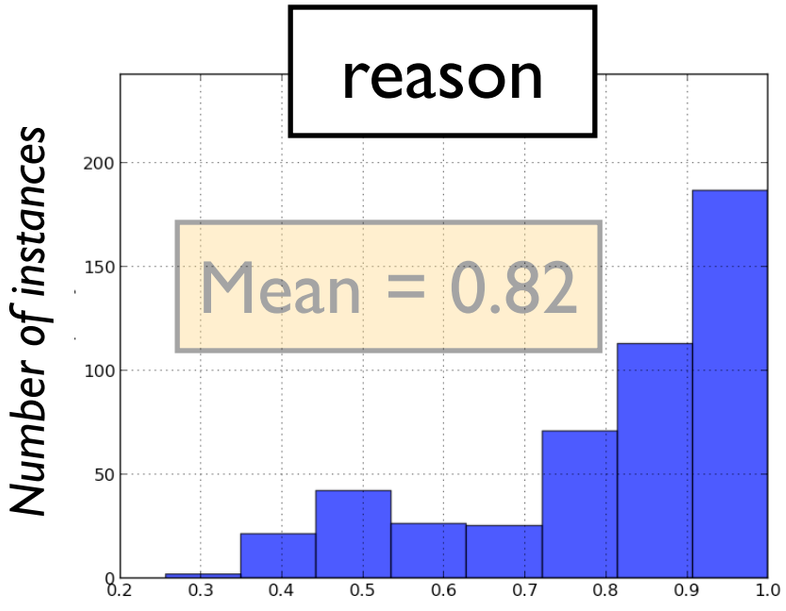


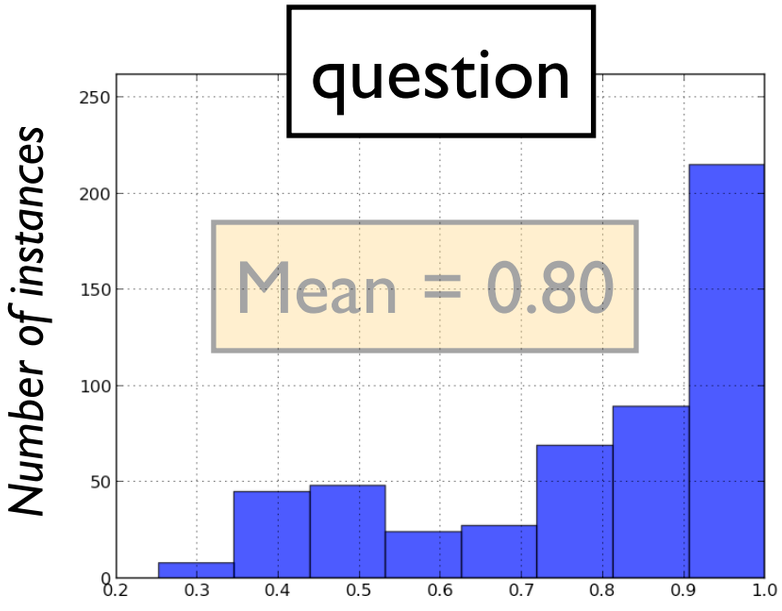


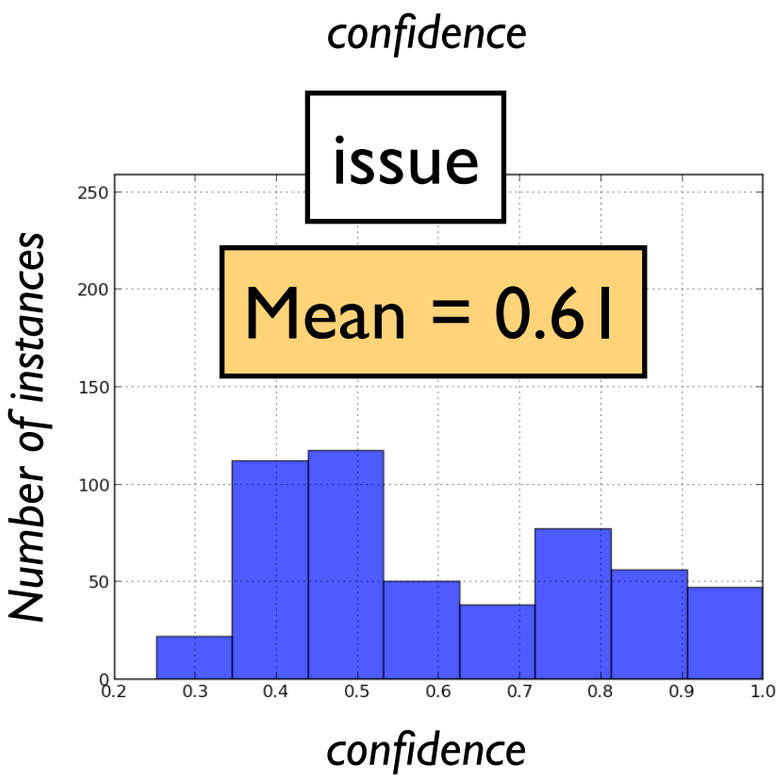


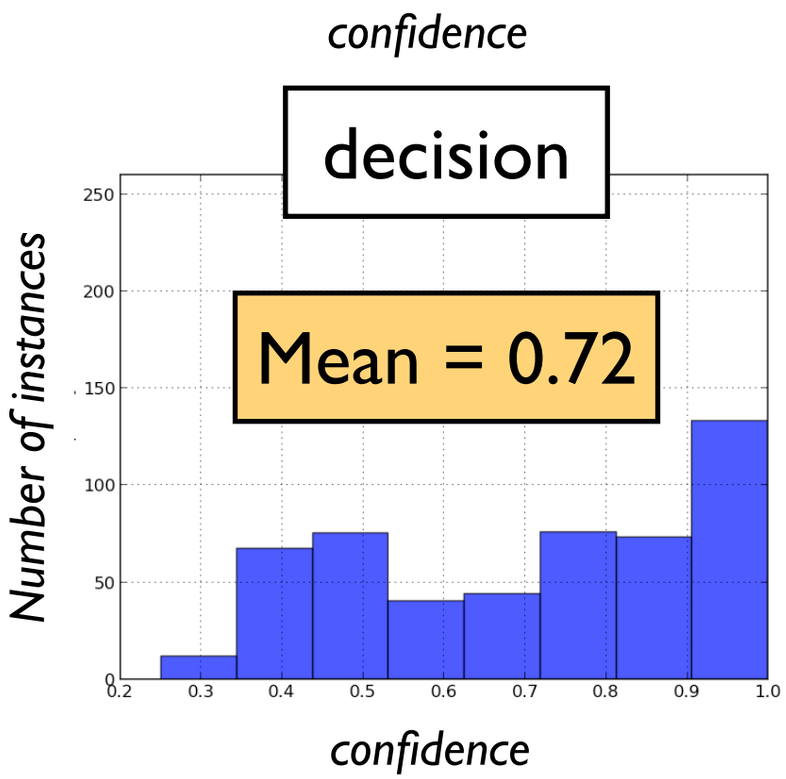
Confidence Distribution

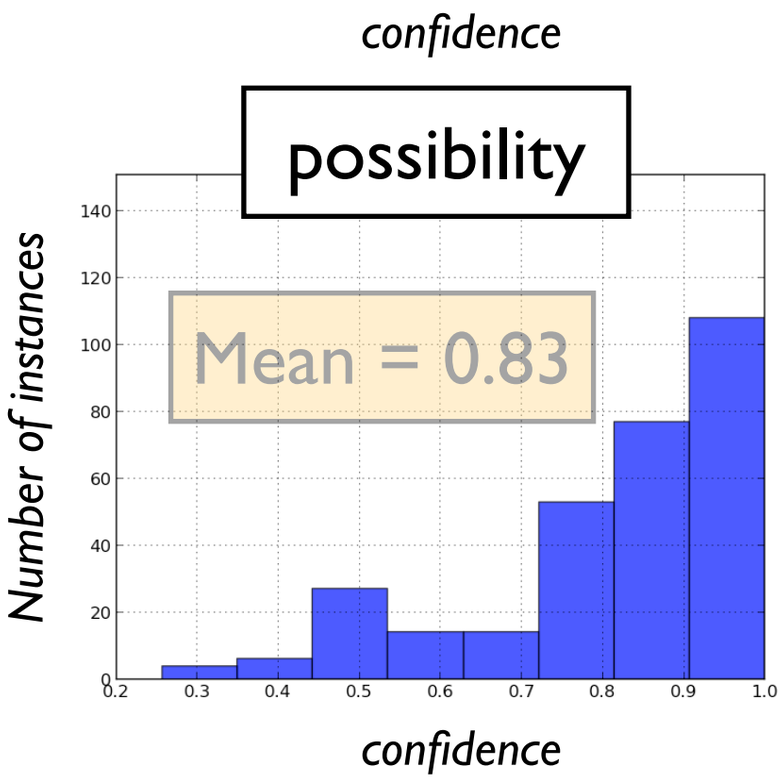












CrowdFlower Expt. 2

Identifying the precise antecedent

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...

CrowdFlower Expt. 2

Identifying the precise antecedent

New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings.

Select one of the options (required)

- None
- one of only three states
- some form of audio-visual coverage
- some form of audio-visual coverage of court proceedings
- that do not allow some form of audio-visual coverage of court proceedings
- audio-visual coverage of court proceedings
- some form
- New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings.
- only three states that do not allow some form of audio-visual coverage of court proceedings
- one of only three states that do not allow some form of audio-visual coverage of court proceedings
- allow some form of audio-visual coverage of court proceedings

Settings

- 2,323 high-confidence instances from CrowdFlower expt. 1
- 8 judgements per instance
- 6 cents per annotation
- Completion time: 7 days

Inter-annotator agreement for expt. 2

Challenge

It is believed that between 20 percent and 30 percent of Italy's economic output is submerged. The decision to finally take this fact into account...

Answer 1: that between 20 percent and 30 percent of Italy's economic output is submerged

Answer 2: between 20 percent and 30 percent of Italy's economic output is submerged

Answer 3: between 20 percent and 30 percent

Need coefficients that incorporate the notion of the distance between strings

Challenge

It is believed that between 20 percent and 30 percent of Italy's economic output is submerged. The decision to finally take this fact into account...

Answer 1: that between 20 percent and 30 percent of Italy's economic output is submerged

Answer 2: between 20 percent and 30 percent of Italy's economic output is submerged

Answer 3: between 20 percent and 30 percent

Need coefficients that incorporate the notion of the distance between strings

Challenge

It is believed that between 20 percent and 30 percent of Italy's economic output is submerged. The decision to finally take this fact into account...

Answer 1: that between 20 percent and 30 percent of Italy's economic output is submerged

less distant

Answer 2: between 20 percent and 30 percent of Italy's economic output is submerged

more distant

Answer 3: between 20 percent and 30 percent

Need coefficients that incorporate the notion of the distance between strings

Krippendorff's α

$$\alpha = 1 - \frac{D_o}{D_e}$$

observed disagreement

expected disagreement

α = 1, perfect reliability

α = 0, absence of reliability

α < 0, systematic disagreement or small sample size

α with Distance Metrics

	Jaccard			Dice		
	D_o	D_e	α	D_o	D_e	α
A&P, 2006	.53	.95	.45	.43	.94	.55
Our results	.47	.96	.51	.36	.92	.61

$$\text{Jaccard}(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad \text{Dice}(A,B) = \frac{2\,|A \cap B|}{|A| + |B|}$$

Artstein and Poesio, 2006: 20 annotators, 16 instances
Our work: 8 annotators, 2,323 instances

α with Distance Metrics

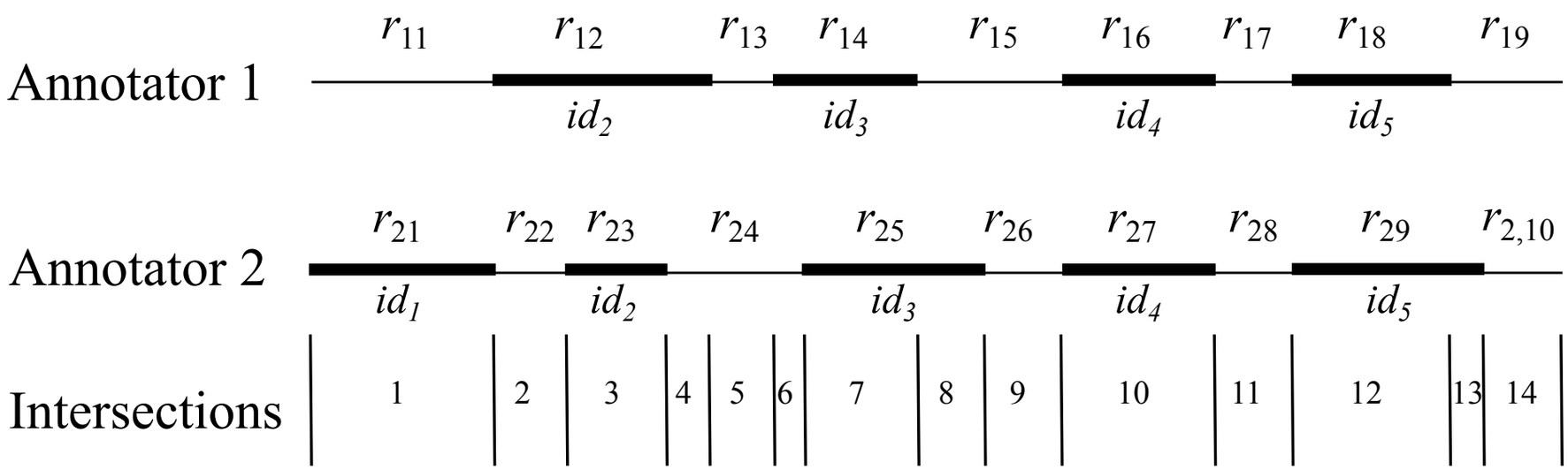
	Jaccard			Dice		
	D_o	D_e	α	D_o	D_e	α
A&P, 2006	.53	.95	.45	.43	.94	.55
Our results	.47	.96	.51	.36	.92	.61

$$\text{Jaccard}(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad \text{Dice}(A,B) = \frac{2\,|A \cap B|}{|A| + |B|}$$

Artstein and Poesio, 2006: 20 annotators, 16 instances
Our work: 8 annotators, 2,323 instances

Unitizing α

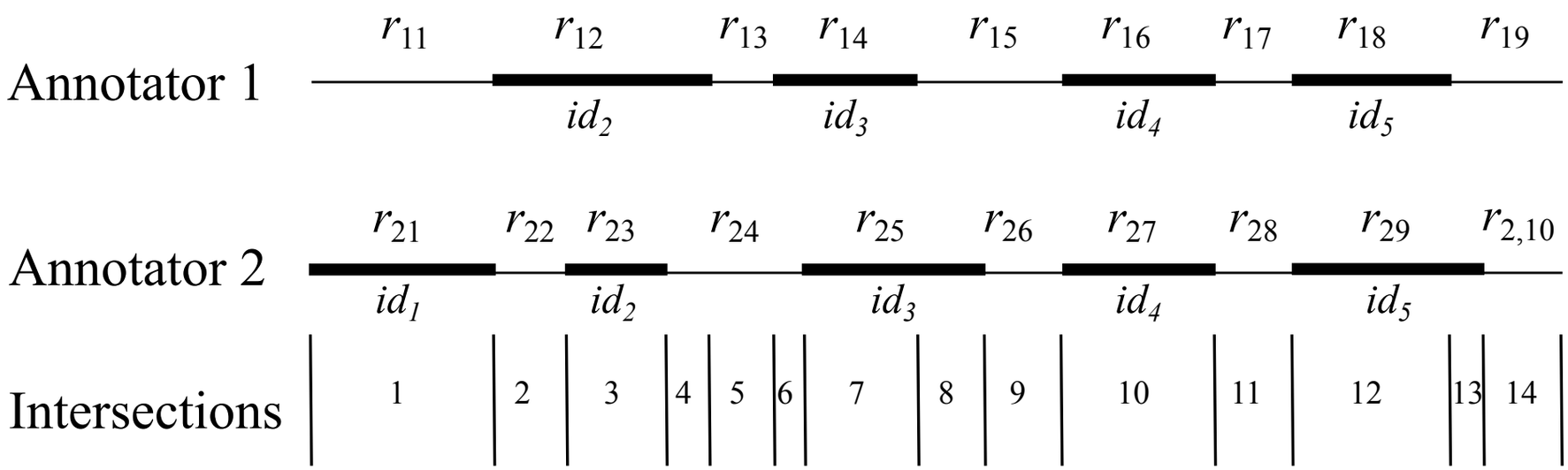
(Krippendorff, 2013)



Distance function:
square-distance between non-overlapping tokens

Unitizing α

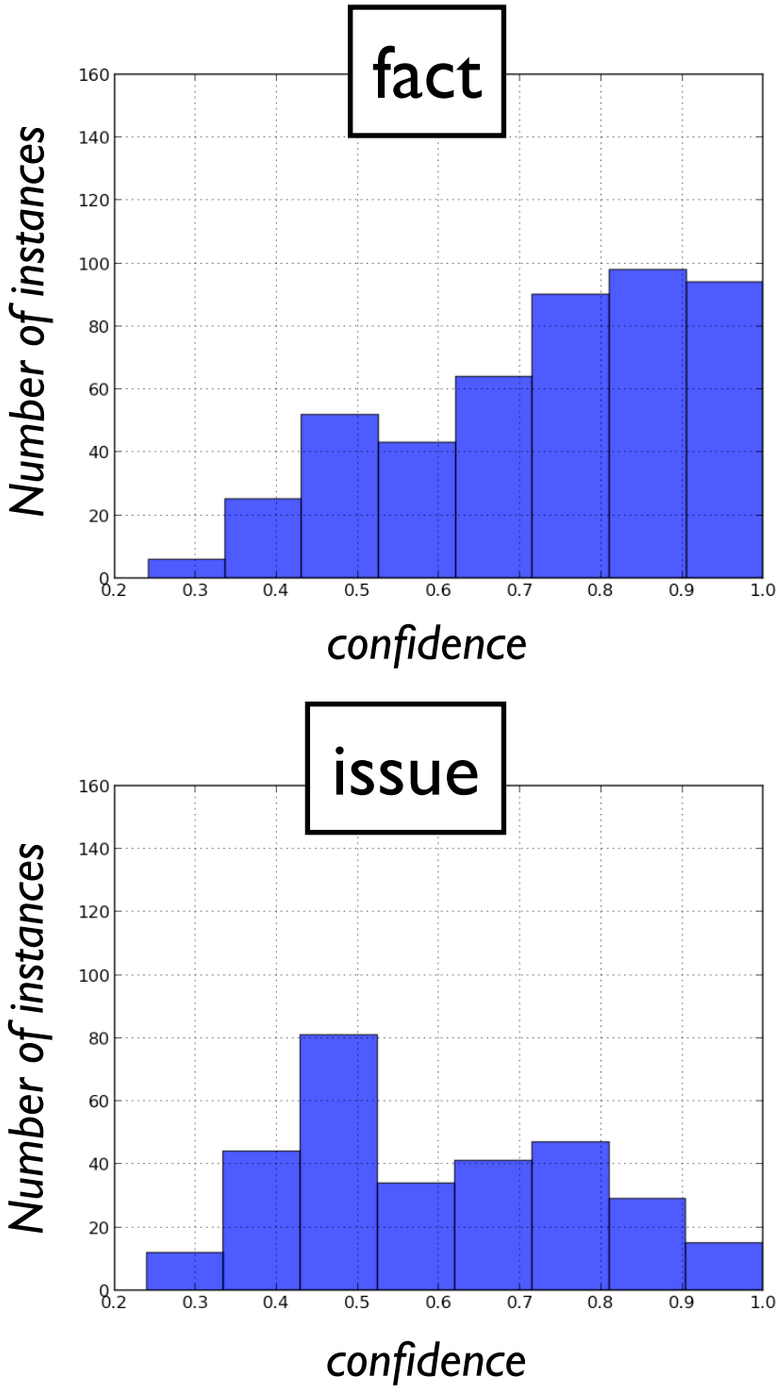
(Krippendorff, 2013)

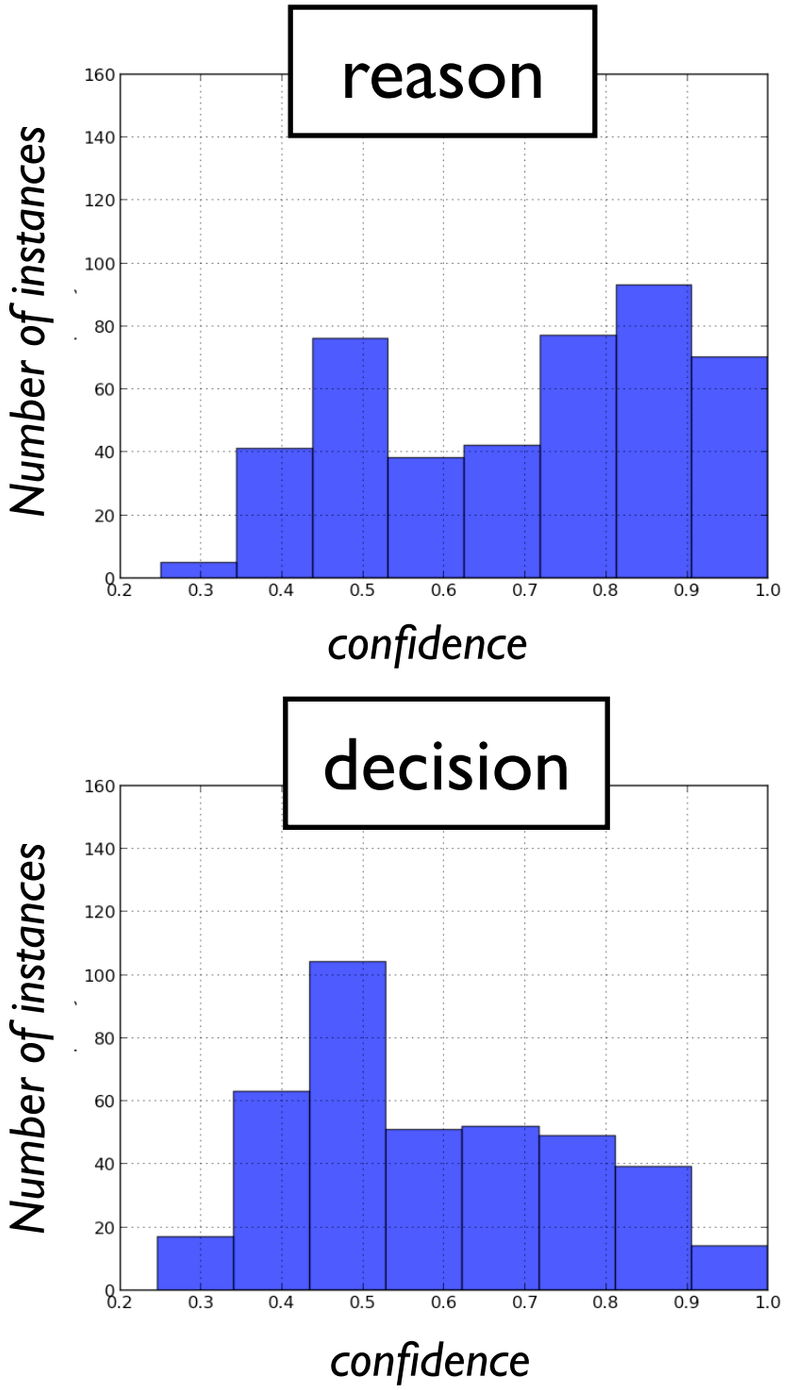


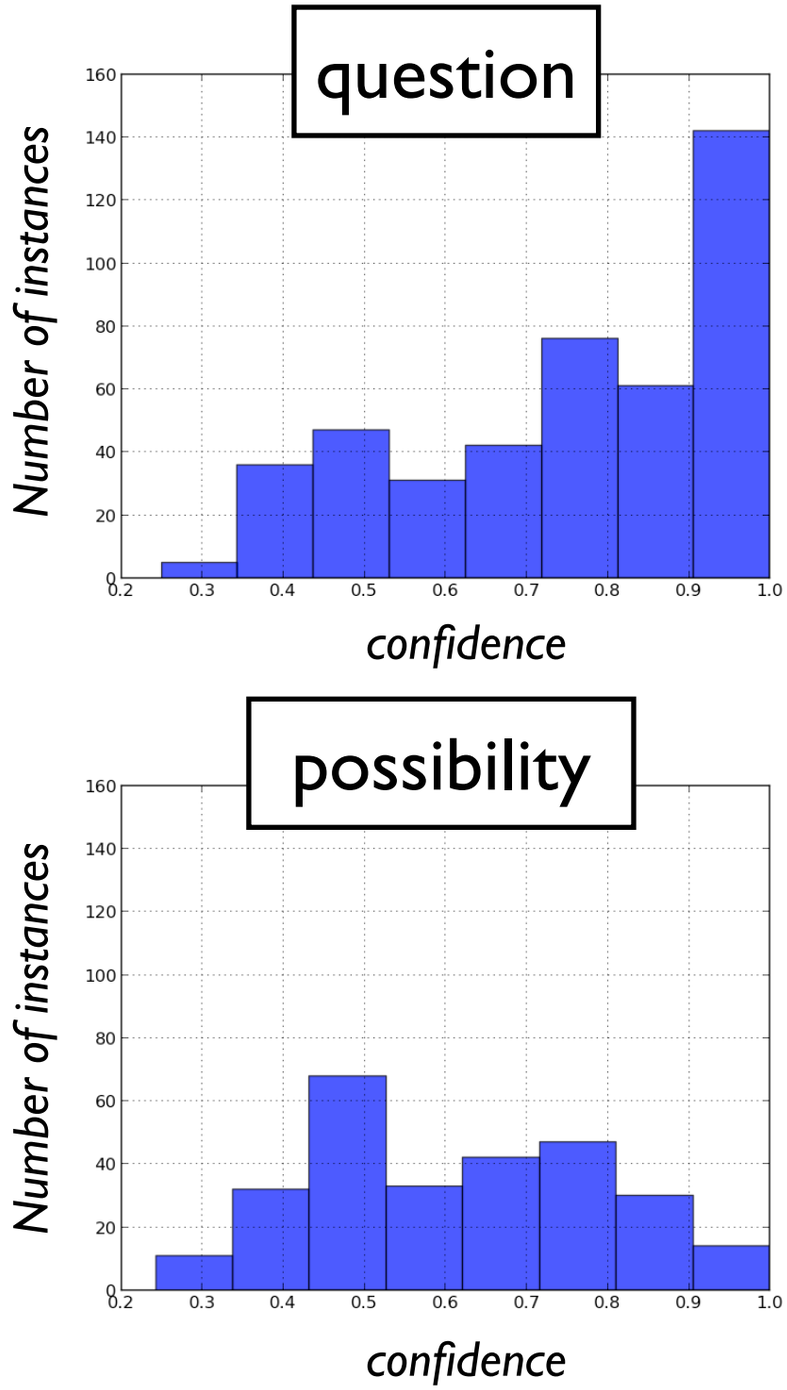
Distance function:
square-distance between non-overlapping tokens

Unitizing α = 0.54

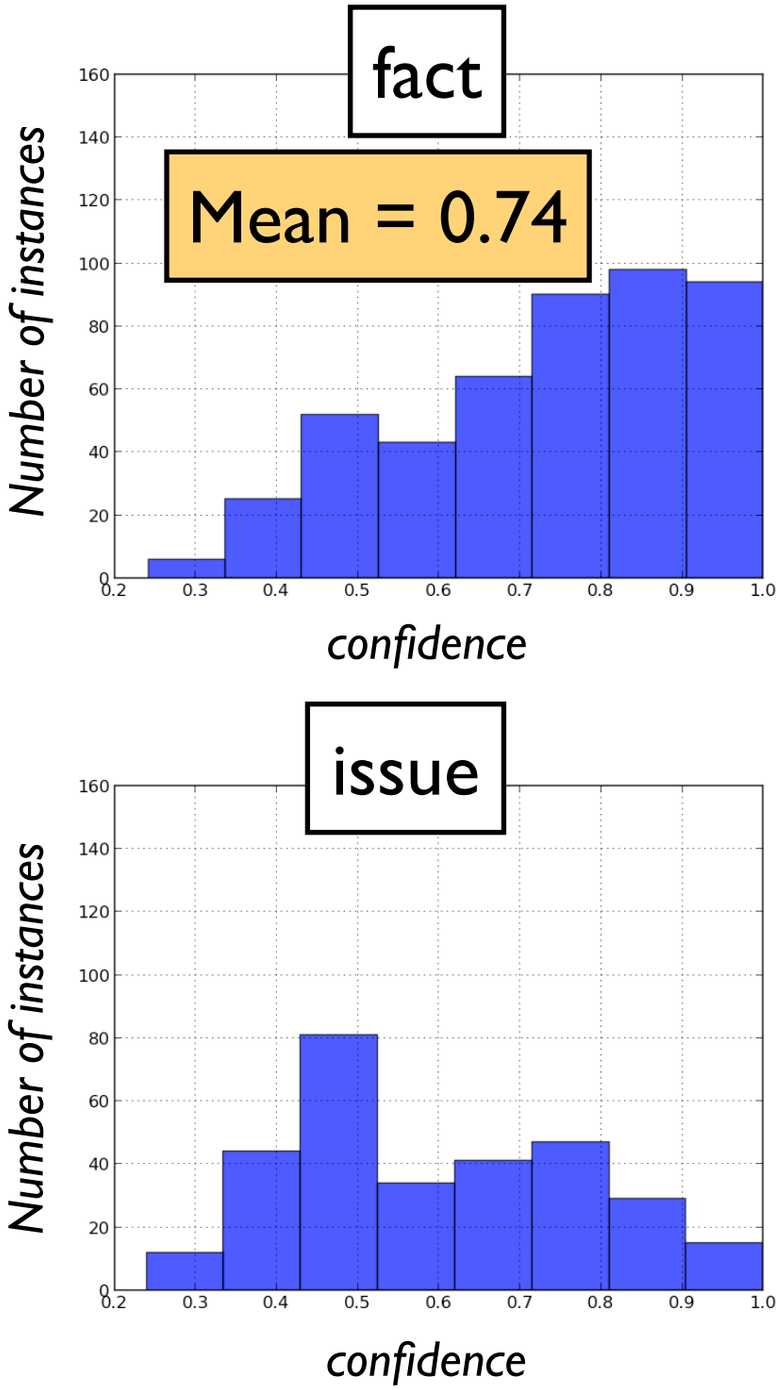
CrowdFlower Confidence

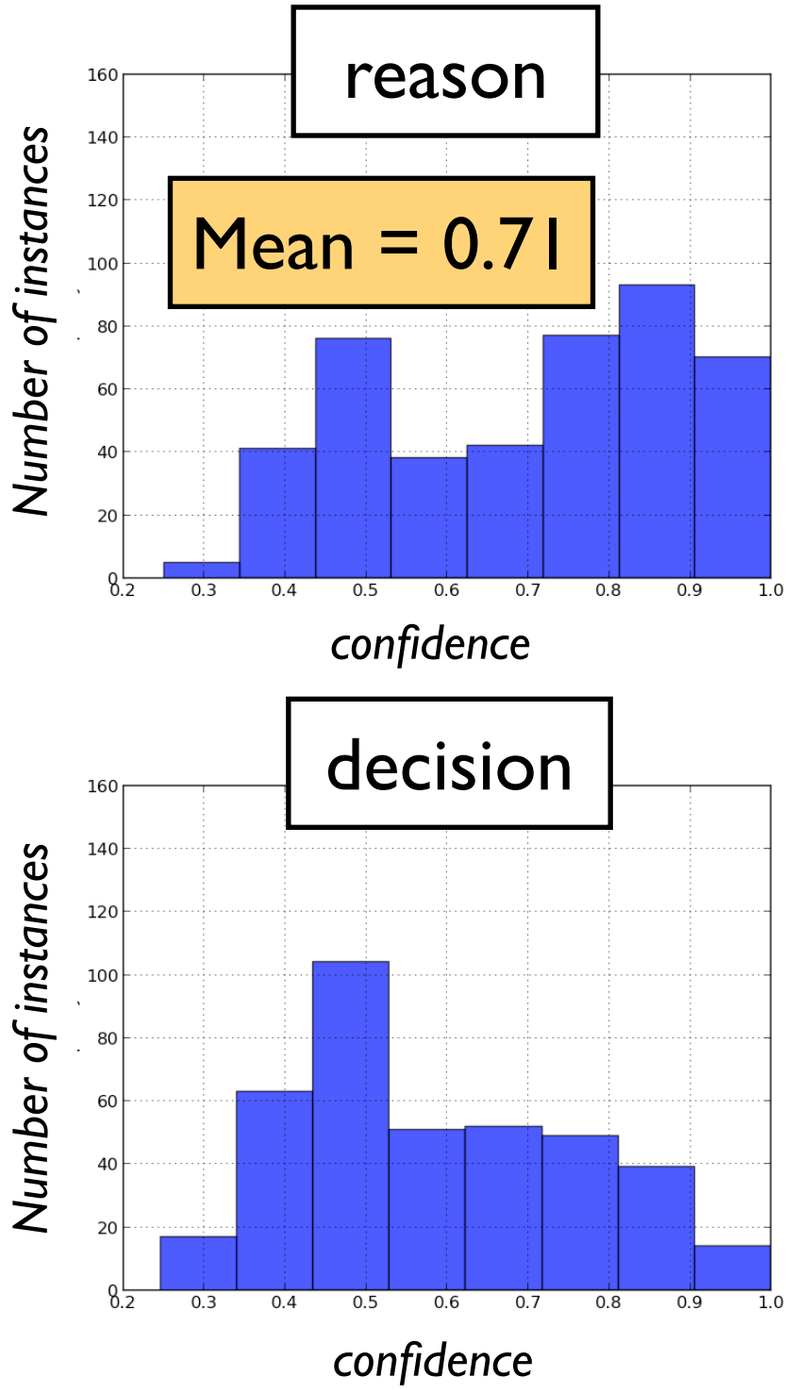


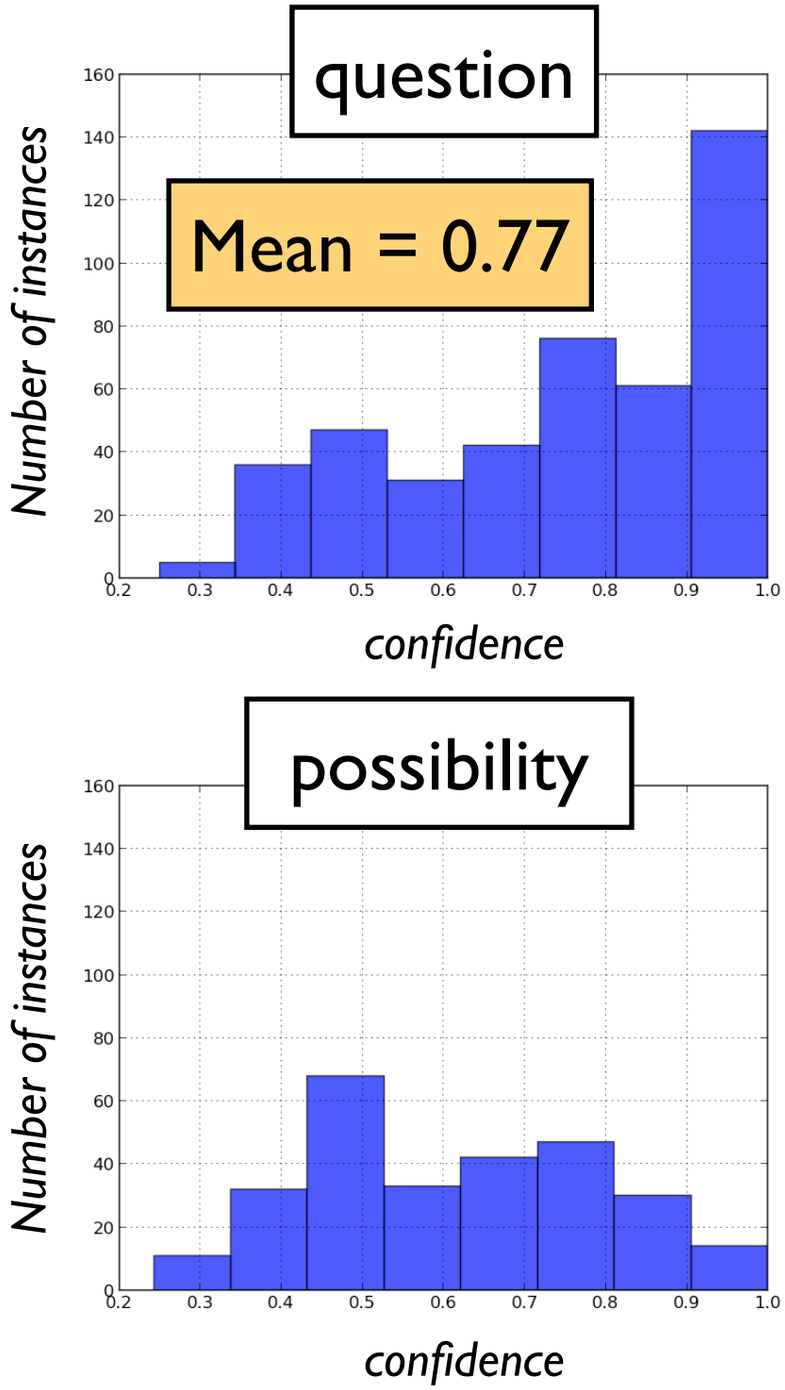




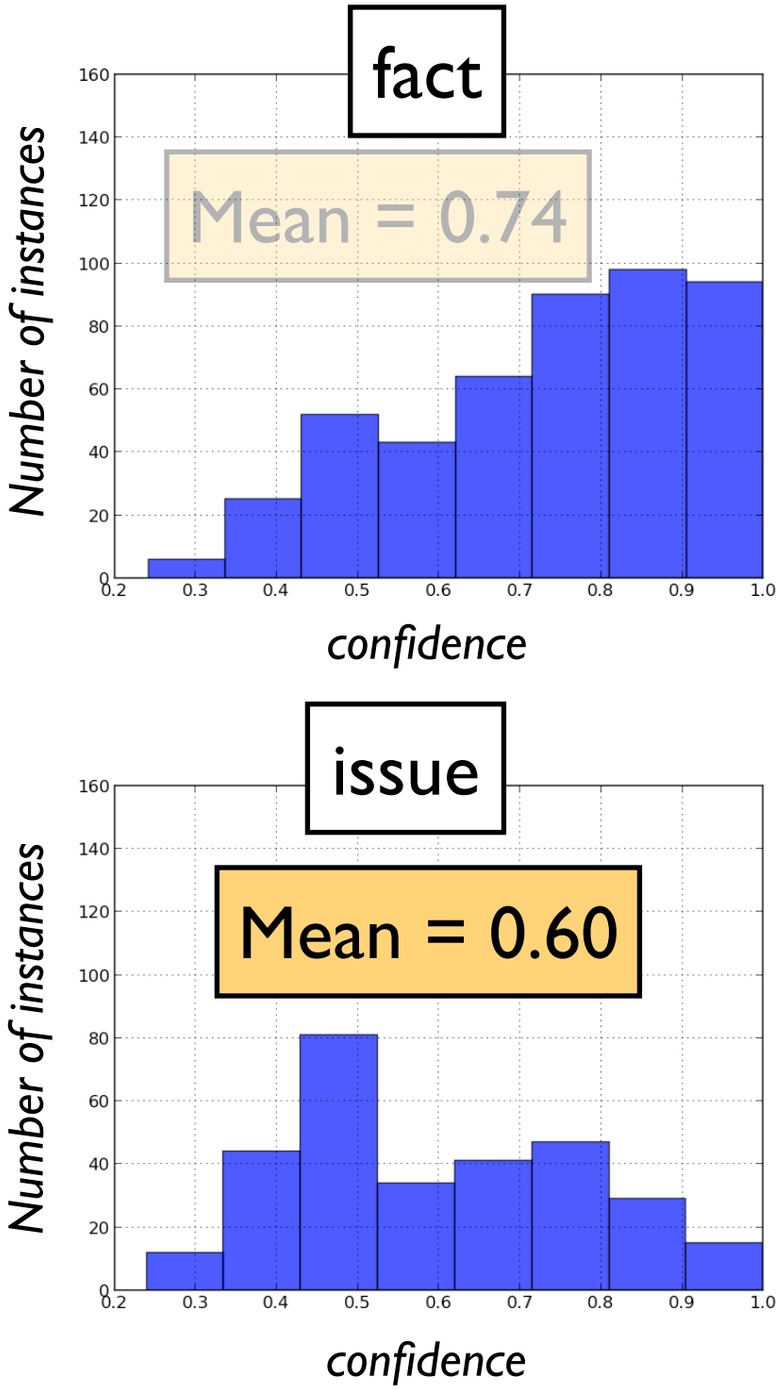
CrowdFlower Confidence

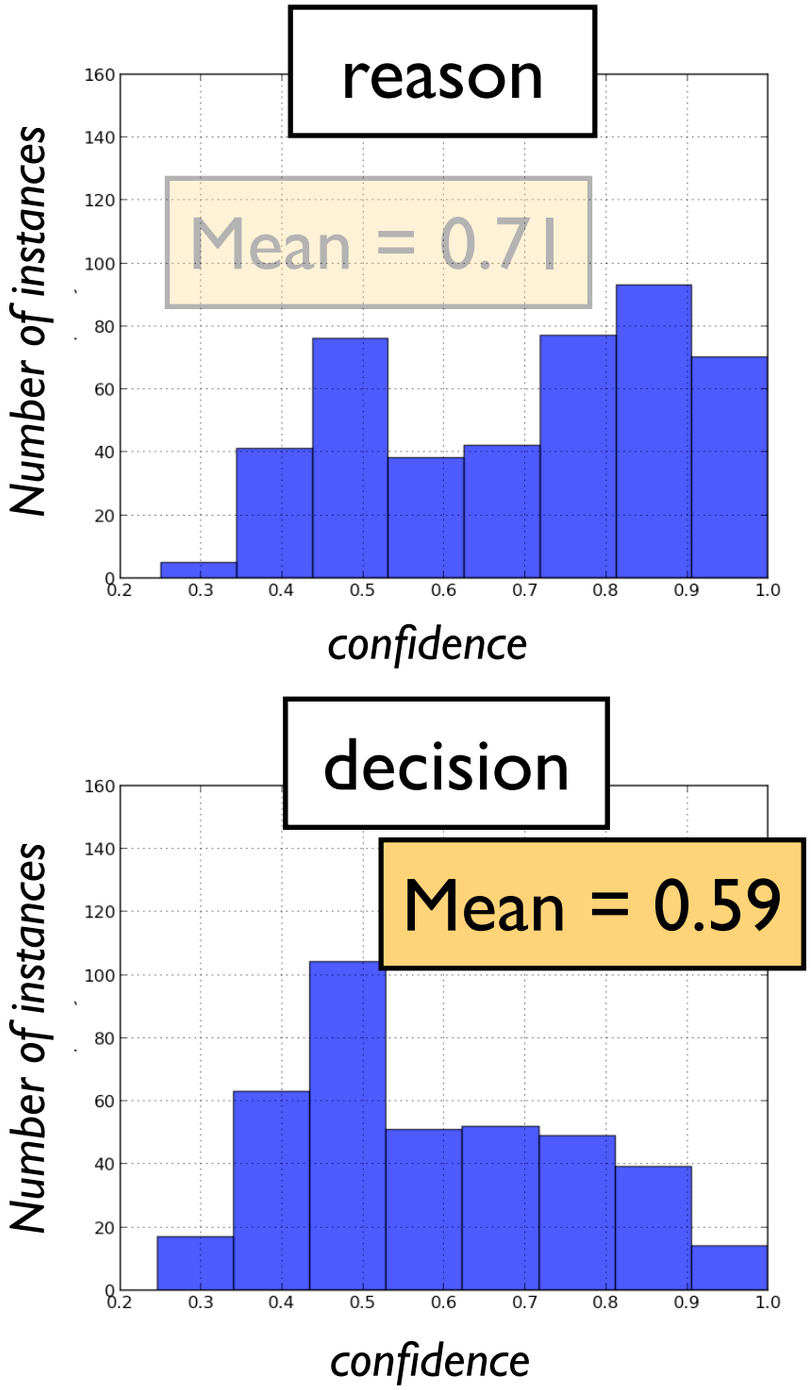


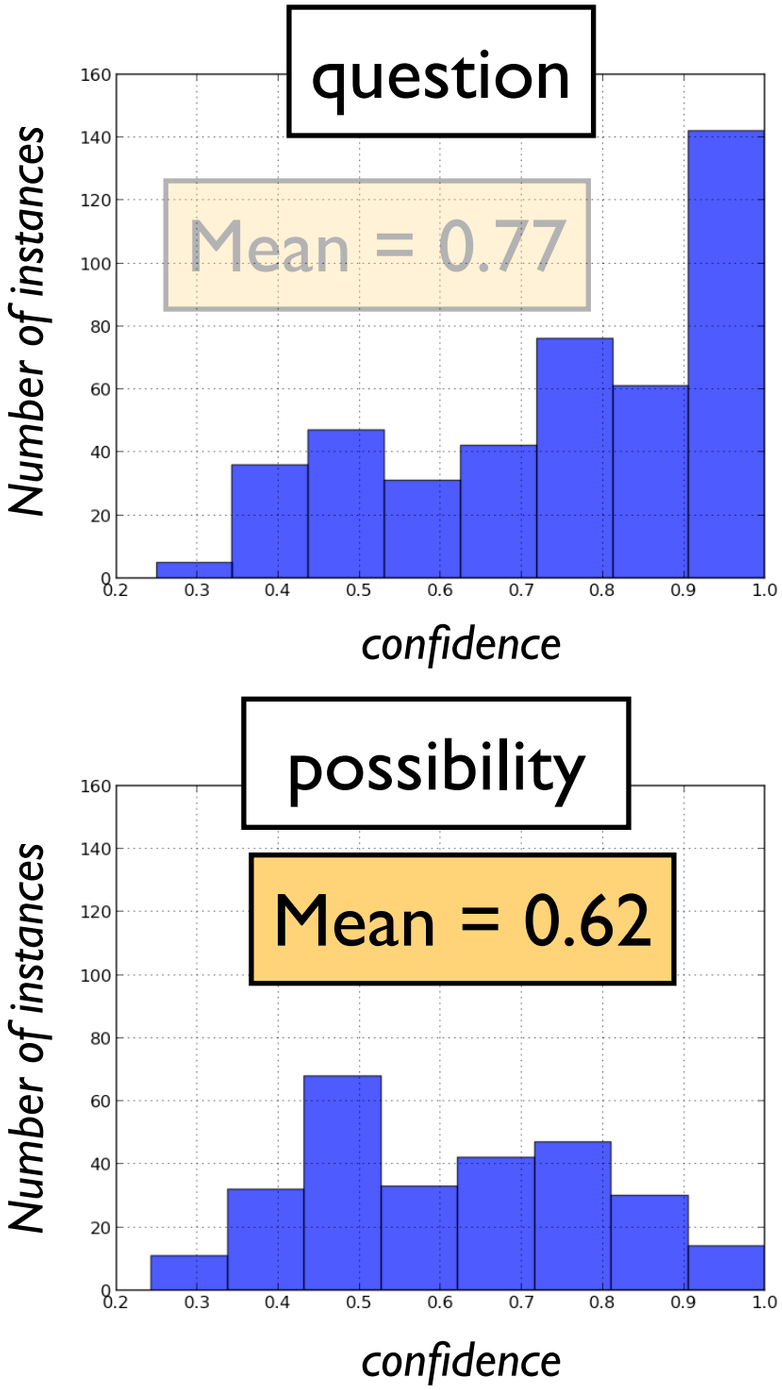




CrowdFlower Confidence







ASN Antecedent Corpus

- 1,810 high confidence (confidence ≥ 0.5) instances from the CrowdFlower expt. 2

Evaluation by Experts

- Goal
  Examine acceptability of crowd's answers
- Judges
  - Two highly-qualified academic editors
- Evaluation options
  - *Perfectly, Reasonably, Implicitly, Not at all*

Evaluation by Experts

Perfectly (P), Reasonably (R), Implicitly (I), Not at all (N)

		Judge B				
		P	*R*	*I*	*N*	Total
Judge A	*P*	**171**	44	11	7	233
	R	12	**27**	7	4	50
	I	2	4	**6**	1	13
	N	1	2	0	**1**	4
	Total	186	77	24	13	300

Evaluation by Experts

Perfectly (P), Reasonably (R), Implicitly (I), Not at all (N)

		Judge B				
		P	*R*	*I*	*N*	Total
Judge A	*P*	**171**	44	11	7	233
	R	12	**27**	7	4	50
	I	2	4	**6**	1	13
	N	1	2	0	**1**	4
	Total	186	77	24	13	300

According to the experts, ~95% instances had acceptable annotations

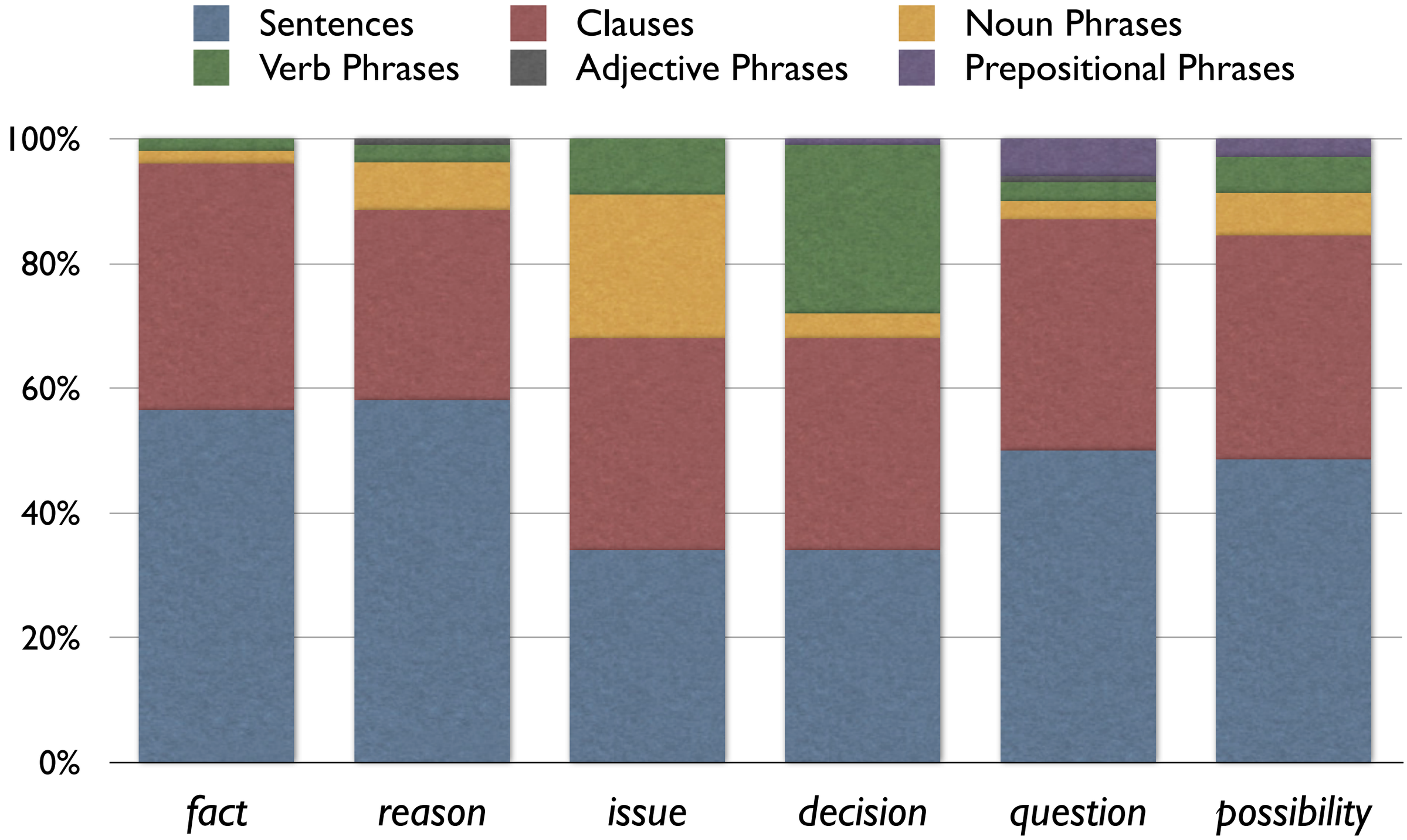
Conclusion

- Examined feasibility of annotating antecedents of ASNs (e.g., *this issue*, *this fact*) using crowdsourcing
  - Most of the time, at least half of the annotators converged on one answer
- Evaluated crowd-annotations using experts
- Resulted in an ASN antecedent corpus containing 1,810 instances with high-confidence annotation

Error Analysis

- Problem with agreeing on *None*
- Multiple possible answers
- Hard instances
- Different strings representing similar concepts

Syntactic Type Distribution



Hard Examples

Several Vatican officials said, however, that any such talk has little meaning because the church does not take sides in elections. But the statements by several American bishops that Catholics who vote for Mr. Kerry would have to go to confession have raised the question in many corners about whether this is an official church position.

The church has not addressed this question publicly and, in fact, seems reluctant to be dragged into the fight...”

Hard Examples

Any biography of Thomas More has to answer one fundamental question. Why? Why, out of all the many ambitious politicians of early Tudor England, did only one refuse to acquiesce to a simple piece of religious and politica opportunism? What was it about More that set him apart and doomed him to a spectacularly avoidable execution?

The innovation of Peter Ackroyd's new biography of More is that he places the answer to this question outside of More himself.