Scalar Anaphora: Annotating Degrees of Coreference in Text

Bingyang Ye, Jingxuan Tu, Kyeongmin Rim, James Pustejovsky

December 6, 2023

CRAC 2023

Brandeis

Concept of Anaphora Resolution

- Anaphora resolution is the task of identifying when two or more narrative entities should be resolved as anaphorically bound.
  - Coreference: entities are semantically identical
  - Anaphora: entities are semantically related

Binary doesn’t work

1. On homecoming night ***Postville*** feels like Hometown, USA, but a look around this town of 2,000 shows ***it***’s become a miniature Ellis Island . . . For those who prefer the old Postville, Mayor John Hyman has a simple answer. (ACE)
2. On homecoming night ***Postville*** feels like Hometown, USA, but a look around this town of 2,000 shows ***it***’s become a miniature Ellis Island . . . For those who prefer ***the old Postville***, Mayor John Hyman has a simple answer. (OntoNotes)

There’s also a gap

- ***The United States*** has officially restored diplomatic relations with Yugoslavia . . . ***The White House*** said the United States will provide 45 million dollars in food aid to Yugoslavia.

We need a middle ground

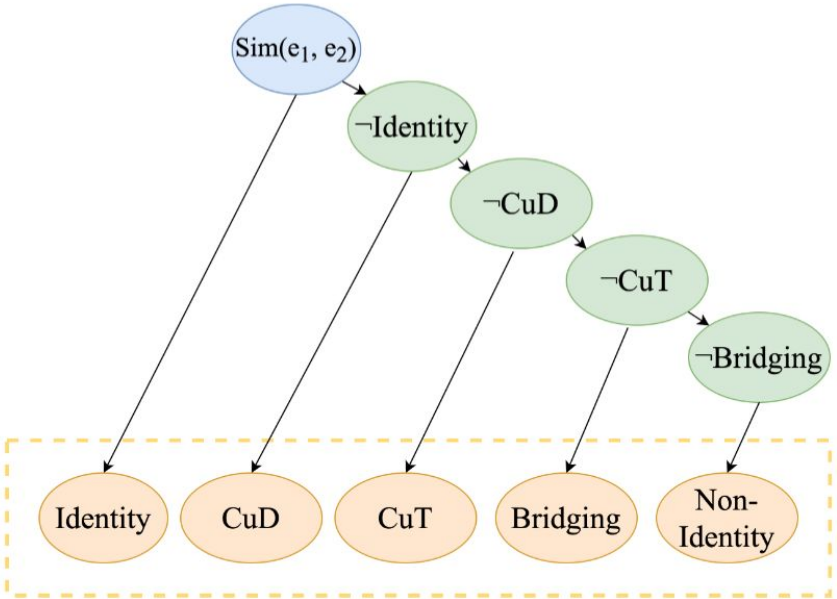
- Near Identity (NIDENT)
  - Recasens et al. 2011
  - A concept of partial identity relations
  - A typology of fifteen types
  - A corpus of weak/strong relations

We need a middle ground

- Scalar Anaphora
  - Determine the degree of similarity between mention pairs
  - Simpler typology then NIDENT
  - Fill the gap between anaphoric relations (e.g., bridging) and coreference

Scalar Anaphora

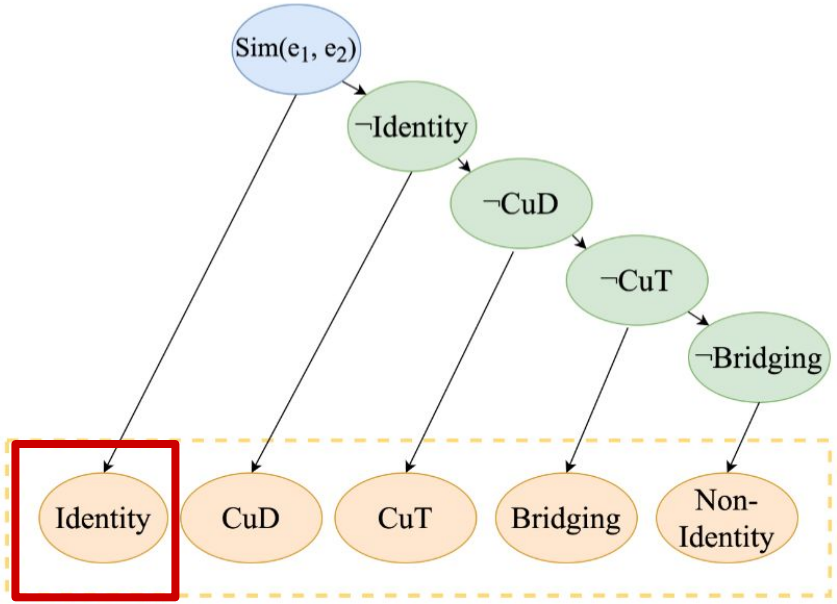
- A coreference scale from strict identity to strict non-identity
- Ordered by referential similarity



Brandeis

Scalar Anaphora

- Identity
  - e_1 and e_2 are substitutible under both transparent and opaque contexts

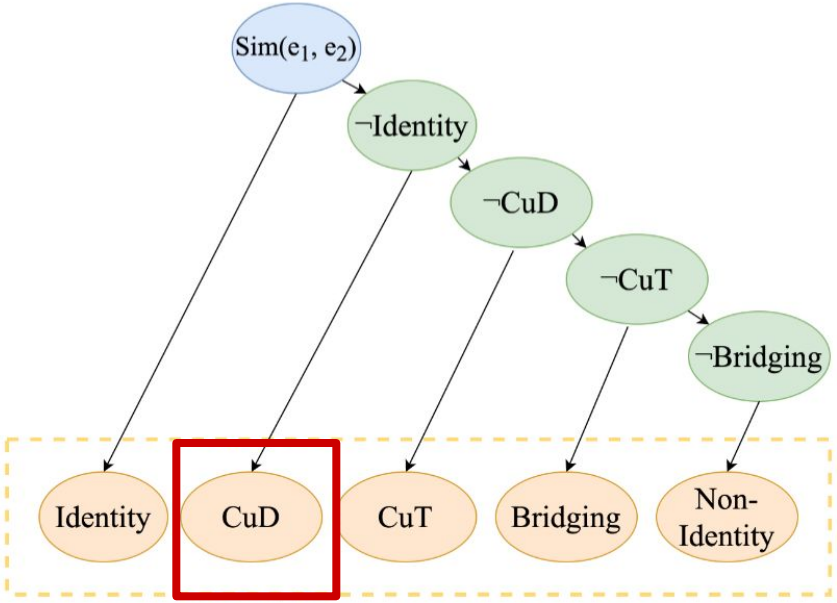


Brandeis

Scalar Anaphora

- Coreference under Description
  - e_1 and e_2 are **not** substitutible under both transparent and opaque contexts
  - Arises with occupational and functional descriptions

(1) **Clinton**[ANTECEDENT], **the Senator**[ANAPHOR] from New York, voiced her concerns about the proposed bill during the congressional hearing.



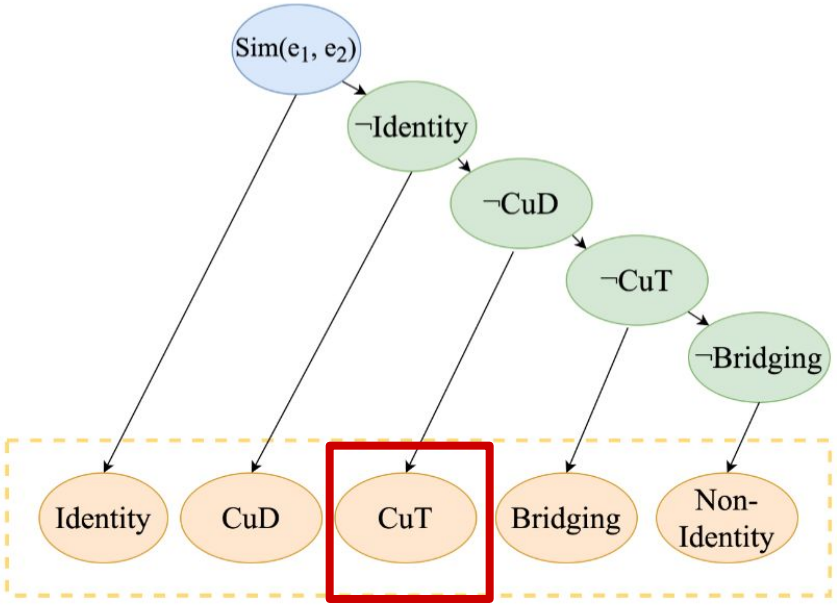
Scalar Anaphora

- Coreference under Transformation (Rim et al. (2023))
  - the formal difference between

    e_1 and e_2 is a result of

    transformative action

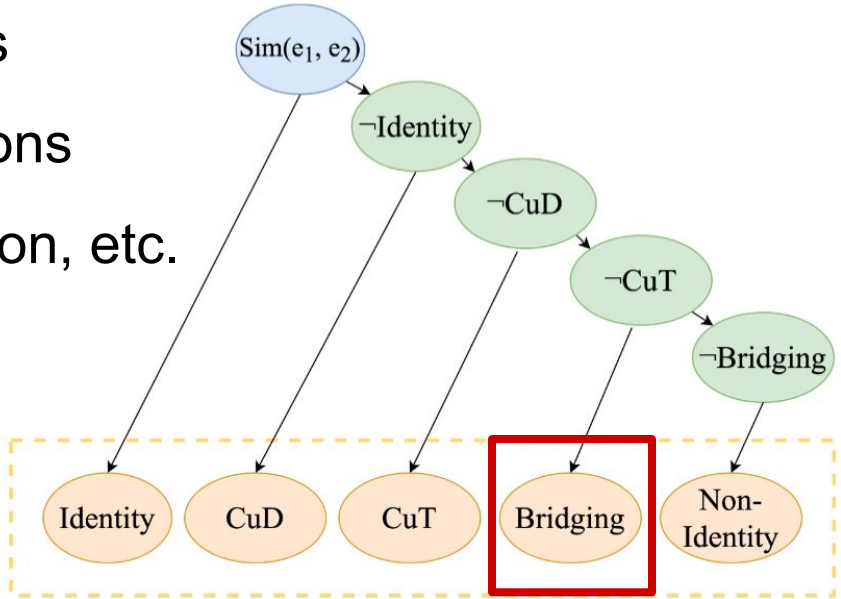
  - e_1 and e_2 are substance identical
  - ***an onion*** and ***chopped onion***



Brandeis

Scalar Anaphora

- Bridging
  - e_1 and e_2 are of different types
  - e_1 and e_2 holds tangible relations
  - E.g. part-of, member-of, location, etc.



Brandeis

Scalar Anaphora Annotation

- 35 Wikipedia docs from Phrase Detective 3.0 gold
- Use the annotators' disagreement to automatically extract SA relations
- Disagreement Score = number of disagreements / total number of judgements
- For each doc, we randomly sampled three pairs for each Disagreement Score bin ([0, 0.4), [0.4, 0.7), [0.7, 1.0])
- 308 pairs split into 3 batches after excluding edge cases

Scalar Anaphora Annotation

- IAA

	$F1$	Cohen's κ
Round 1	51.43	0.31
Round 2	66.67	0.51
Round 3	76.19	0.64

Table 1: IAA of each annotation round.

Brandeis

Scalar Anaphora Annotation

- Stats

	Count	Ratio (%)	IAA (F1)
IDENTITY	114	37.0	75.98
CUD	31	10.1	40.68
CUT	18	5.8	37.04
BRIDGING	129	41.9	70.87
NON-IDENTITY	16	5.2	36.36
OVERALL	308	100	65.31

Table 2: Statistics of annotation in terms of SA relation.

Brandeis

Scalar Anaphora Resolution

- Formalize the task as identifying the SA relation between each mention pair given the sentence context of the entity
- Data split in terms of relation type
- Add “negative” examples

Experiment 1: SA Resolution with T5

- Set the input sequence as question answering format

input text:
queston: What is the relation between [mainly wealthier nations] and [these countries]?
context: VHEMT spreads its message ... reaching [mainly wealthier nations] .
A few of [these countries] already have fertility rates below ...
output text:
Bridging

Figure 3: Example of T5 model input and output for SA resolution task.

Experiment 1: SA Resolution with T5

- Results

	P	R	F1
IDENTITY	56.25	78.26	65.45
CUD	100	16.67	28.57
CUT	0	0	0
BRIDGING	60.00	57.69	58.82
NON-IDENTITY	0	0	0
NEGATIVE	100	28.57	44.44
OVERALL	52.71	30.20	32.88

Table 5: Pairwise relation classification results on the test set with T5.

Experiment 2: SA Resolution with GPT-4

- Prompt tuning
  - Human Instruction
    - Flat
    - Hierarchical
  - Exemplar
    - 0 shot
    - Few shot
      - Random
      - In-domain
      - CoT

Experiment 2: SA Resolution with GPT-4

- Results on different prompt settings

		P	R	F1
Flat	0-shot	32.41	**40.00**	33.63
	5-shot-random	36.94	33.33	26.79
	5-shot-domain	20.50	30.00	21.30
	5-shot-CoT	40.00	**40.00**	36.41
Hierarchy	0-shot	44.44	30.00	34.78
	5-shot-random	46.30	30.00	27.78
	5-shot-domain	41.32	30.00	34.76
	5-shot-CoT	**50.11**	36.67	**37.90**

Table 7: Pairwise relation classification results on 25 random examples with different prompt settings.

Brandeis

Experiment 2: SA Resolution with GPT-4

- Results

	P	R	F1
IDENTITY	46.81	95.65	62.86
CUD	10.00	16.67	12.50
CUT	40.00	50.00	44.44
BRIDGING	66.67	15.38	25.00
NON-IDENTITY	0	0	0
NEGATIVE	100	14.29	25.00
OVERALL	43.91	32.00	28.30

Table 8: Pairwise relation classification results on the test set with GPT-4.

Conclusion

- Reintroduce the complexity of conference, non-coreference and anaphora.
- A unified typology of different degrees of semantic similarity to address the issue of near identity and fill the gap between coreference and bridging.
- A corpus of the new Scalar Anaphora typology and some preliminary results using LLMs.
- A potential integration into the Universal Anaphora schema.