Research presentation

Matthijs Westera, Universitat Pompeu Fabra

Two research strands

Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

2) Understanding discourse structure (goals, topics)
   - implicit questions
   - referent predictability

Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

2) Understanding discourse structure (goals, topics)
   - implicit questions
   - referent predictability

Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

2) Understanding discourse structure (goals, topics)
   - implicit questions
   - referent predictability

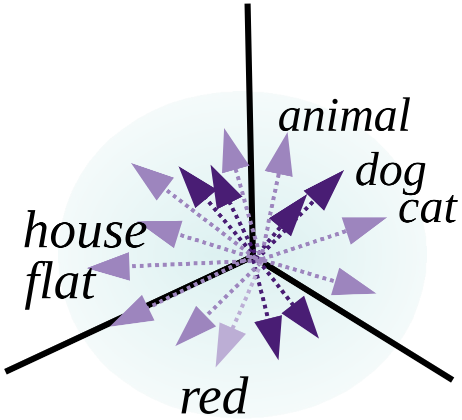
1. Distributional semantics

Work with Abhijeet Gupta, Sebastian Padó & Gemma Boleda.

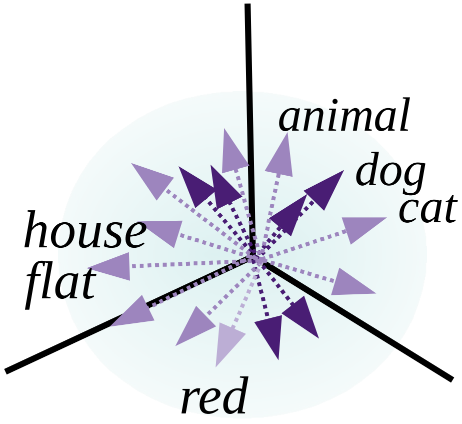


Westera & Boleda (2019, IWCS):

What does Distributional Semantics model?



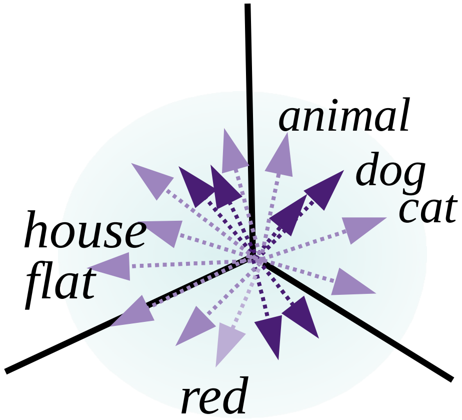
Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?



- The vectors of DS are *abstractions over occurrences.*

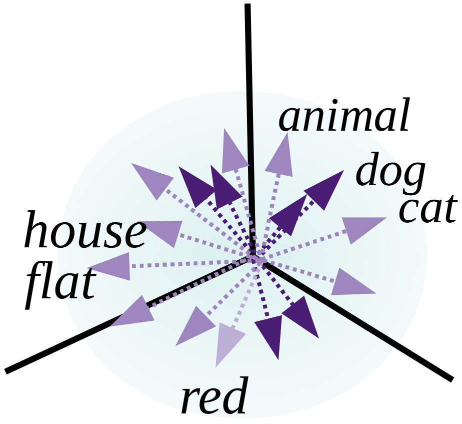
Westera & Boleda (2019, IWCS):

What does Distributional Semantics model?



- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

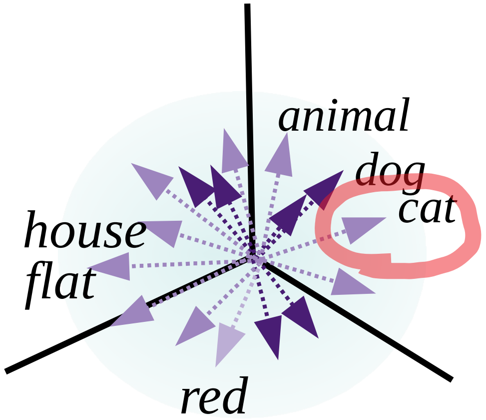
Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?



- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

But what sort of concepts does DS model?

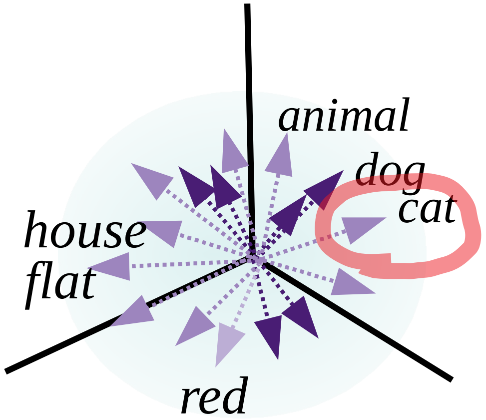
Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?



- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

But what sort of concepts does DS model?

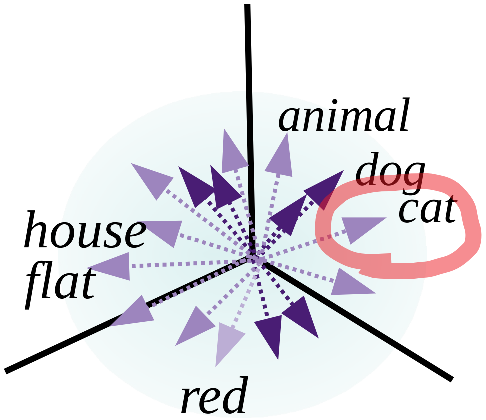
Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?



- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

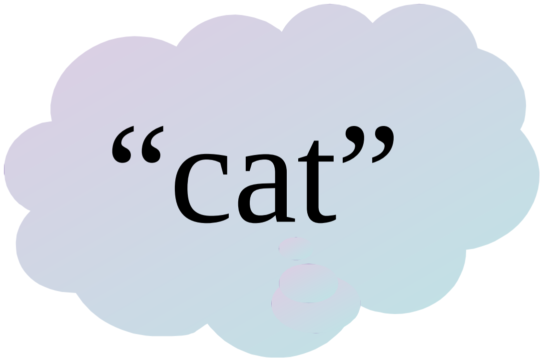
But what sort of concepts does DS model?

Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?

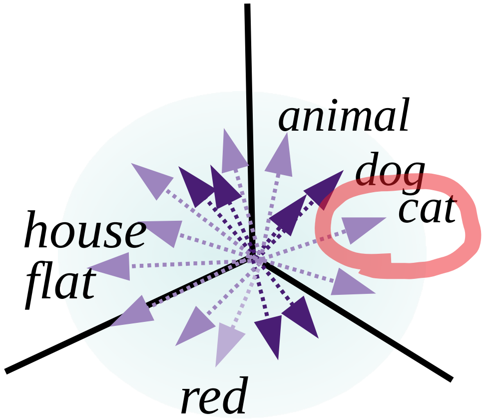


- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

But what sort of concepts does DS model?



Westera & Boleda (2019, IWCS):
What does Distributional Semantics model?

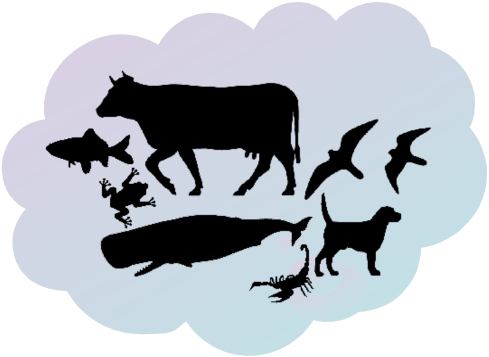


- The vectors of DS are *abstractions over occurrences.*
- And so are *concepts* (e.g., Piaget).

But what sort of concepts does DS model?

Westera & Boleda (2019, IWCS):
Should Distributional Semantics account for *entailment*?

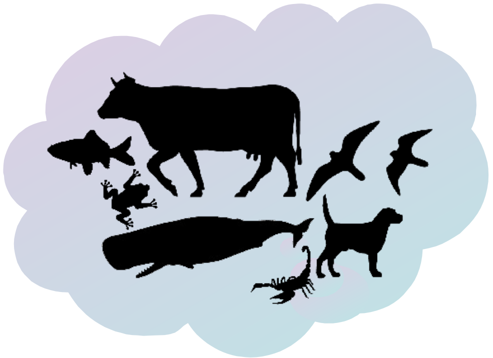
Westera & Boleda (2019, IWCS):
Should Distributional Semantics account for *entailment*?



Westera & Boleda (2019, IWCS):
Should Distributional Semantics account for *entailment*?

Westera & Boleda (2019, IWCS):
Should Distributional Semantics account for *entailment*?

“cat”

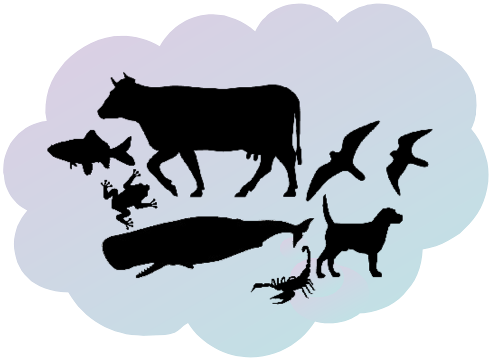


“animal”

Westera & Boleda (2019, IWCS):
Should Distributional Semantics account for *entailment*?

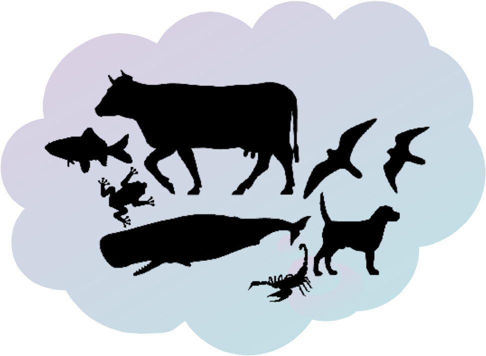
No.

“cat”



“animal”

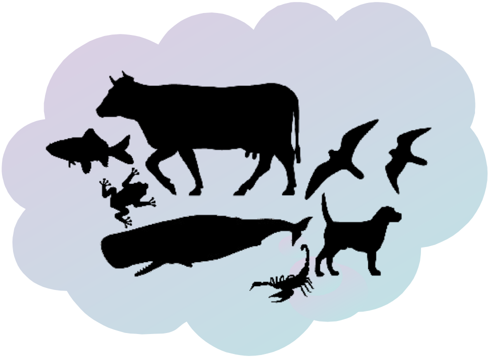
“cat”



“animal”

Language and the world align only *roughly*

“cat”



“animal”

Language and the world align only *roughly*

Language and the world align only *roughly*

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

- Are some parts of language closer to the world than other parts?

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

- Are some parts of language closer to the world than other parts?
- Does this show in Distributional Semantics?

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

- Are some parts of language closer to the world than other parts?
- Does this show in Distributional Semantics?
- Can we exploit this?

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

- Are some parts of language closer to the world than other parts?
- Does this show in Distributional Semantics?
- Can we exploit this?

Some expressions are used more *rigidly* than others...

Language and the world align only *roughly*

This is not (just) a technical challenge, but *interesting*.

- Are some parts of language closer to the world than other parts?
- Does this show in Distributional Semantics?
- Can we exploit this?

Some expressions are used more rigidly than others... (Kripke, '80)

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.

    E.g., for Scientist, the word vector of "scientist"

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.
    E.g., for Scientist, the word vector of "scientist"
  - **Name-based:**
    Average of the word vectors of names of instances of the category.

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.
    E.g., for Scientist, the word vector of "scientist"
  - **Name-based:**
    Average of the word vectors of names of instances of the category.
    E.g., the average of vectors for "Albert Einstein", "Emmy Noether", ...

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.
    E.g., for Scientist, the word vector of "scientist"
  - **Name-based:**
    Average of the word vectors of names of instances of the category.
    E.g., the average of vectors for "Albert Einstein", "Emmy Noether", …
- Evaluation against human judgments of category relatedness.

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:**
    Word vector of a common noun that is used to denote the category.
    E.g., for Scientist, the word vector of "scientist"
  - **Name-based:** ← more 'rigid'
    Average of the word vectors of names of instances of the category.
    E.g., the average of vectors for "Albert Einstein", "Emmy Noether", …
- Evaluation against human judgments of category relatedness.

Westera, Gupta, Padó & Boleda (subm. Cognitive Science)

- We compare two kinds of representations of category concepts:
  - **Noun-based:** ← less 'rigid'
    Word vector of a common noun that is used to denote the category.
    E.g., for Scientist, the word vector of "scientist"
  - **Name-based:** ← more 'rigid'
    Average of the word vectors of names of instances of the category.
    E.g., the average of vectors for "Albert Einstein", "Emmy Noether", …
- Evaluation against human judgments of category relatedness.

Main results

Table 3: Main results of Experiment 1 (Spearman correlation coefficients).

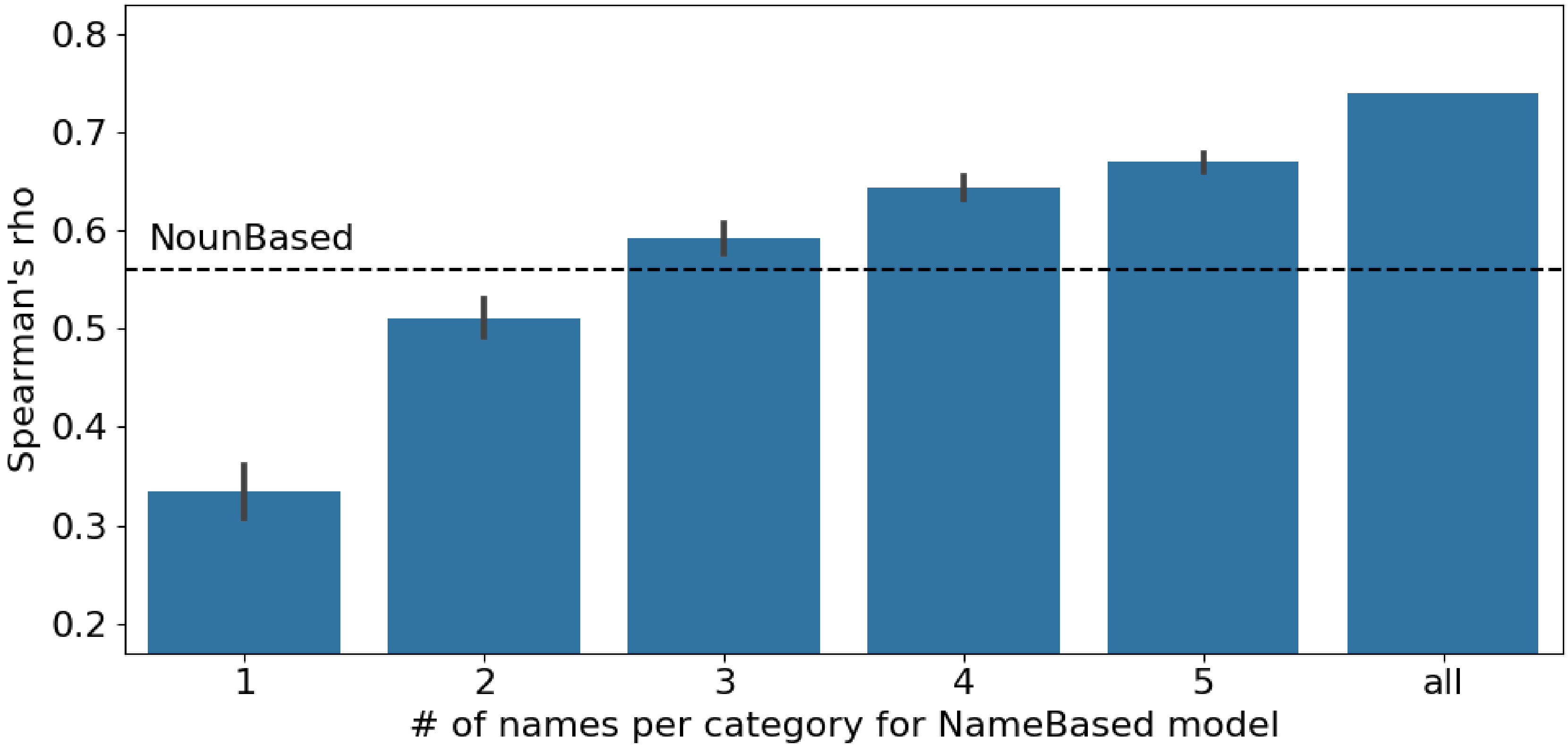
	all	match	unclear	within-domain	between-domain
number of pairs	981	626	355	474	507
NounBased	0.56	0.62	0.45	0.57	0.64
NameBased	**0.74**	**0.73**	**0.75**	**0.68**	0.69

Main results

Table 3: Main results of Experiment 1 (Spearman correlation coefficients).

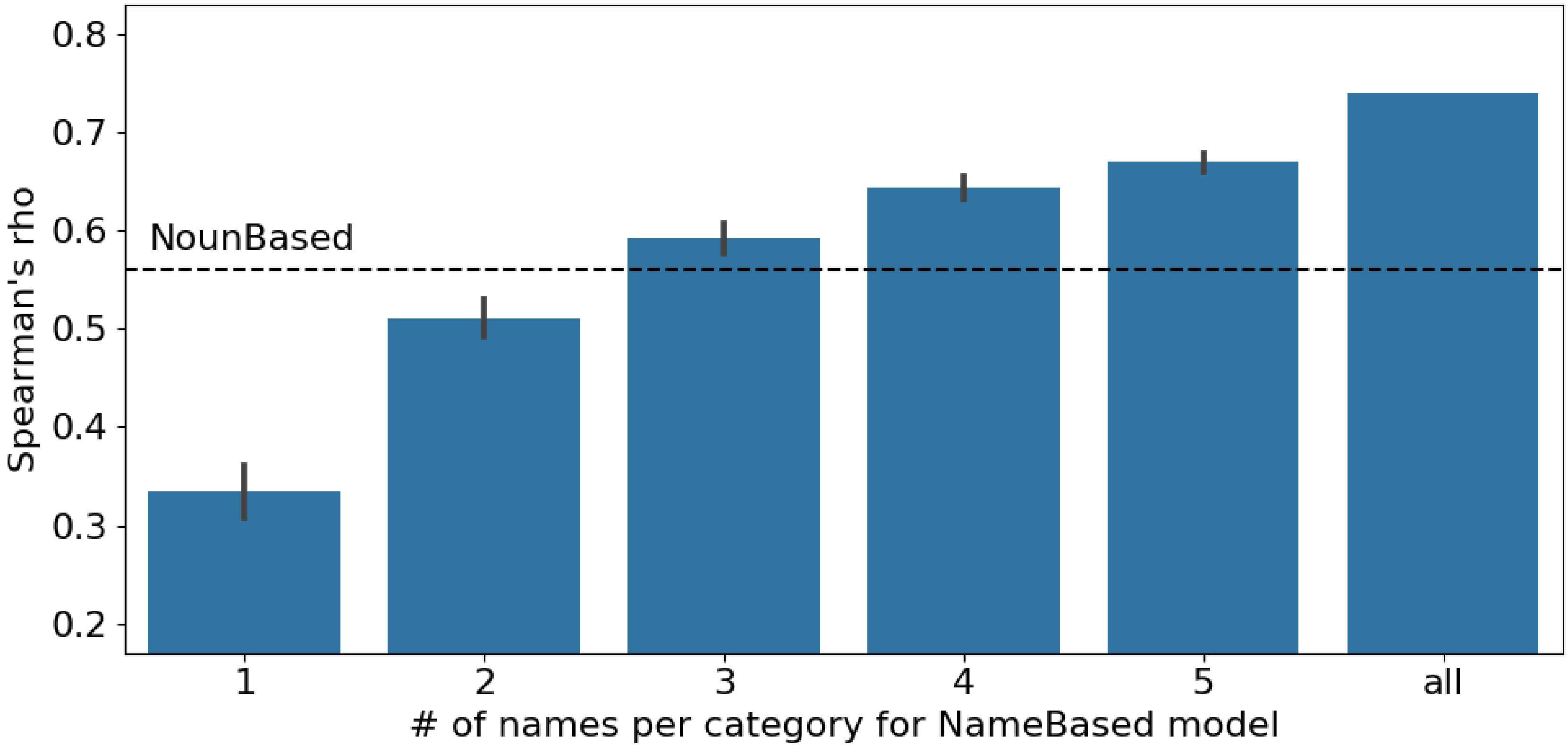
	all	match	unclear	within-domain	between-domain
number of pairs	981	626	355	474	507
NounBased	0.56	0.62	0.45	0.57	0.64
NameBased	**0.74**	**0.73**	**0.75**	**0.68**	0.69

How many names do we need?



How many names do we need?

Surprisingly few!



Conclusions

Conclusions

- Name-based representations of category concepts are better.

Conclusions

- Name-based representations of category concepts are better.
- We should try to understand what our computational models do/don't represent.

Conclusions

- Name-based representations of category concepts are better.
- We should try to understand what our computational models do/don’t represent.
- Crucial distinction: linguistic vs. extralinguistic concepts.

Conclusions

- Name-based representations of category concepts are better.
- We should try to understand what our computational models do/don't represent.
- Crucial distinction: linguistic vs. extralinguistic concepts.

(cf. semantics/pragmatics)

2. Discourse expectations

Main motivation

Main motivation

“I saw Sue at the protest.”

Main motivation

Who were at the protest?

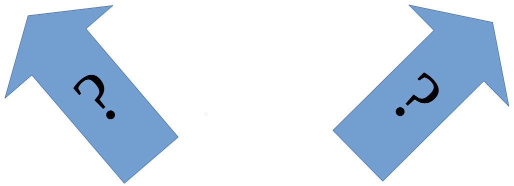
?

"I saw Sue at the protest."

Main motivation

Who were at the protest?

Where was Sue?



"I saw Sue at the protest."

Main motivation

Who were at the protest? *Where was Sue?*

My PhD:

? ?

"I saw Sue at the protest."

Main motivation

"I went to Sue and Bob's place but they weren't home."

Who were at the protest?

Where was Sue?

My PhD:

"I saw Sue at the protest."

Main motivation

"I went to Sue and Bob's place but they weren't home."

Who were at the protest?

Where was Sue?

My PhD:

"I saw Sue at the protest."

Main motivation

“I went to Sue and Bob’s place but they weren’t home.”

Who were at the protest?

Where was Sue?

My PhD:

?

?

“I saw Sue at the protest.”

2.1. Evoked questions

Published & ongoing work with Hannah Rohde, Laia Mayol and Jacopo Amidei.



Today was the worst day of my life.

Today was the worst day of my life.

▶ **Please enter a question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

▶ **In the text, highlight the main word or short phrase that evokes this question.**

Today was the worst day of my life.

▶ **Please enter a question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

What happened?

▶ **In the text, highlight the main word or short phrase that evokes this question.**

Today was the worst day of my life.

▶ **Please enter a question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

What happened?

▶ **In the text, highlight the main word or short phrase that evokes this question.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

You entered the following question:

What happened?

▶ **Was that question answered in the new piece of text?**

Not answered at all. 1 2 3 4 5 *Completely answered.*

▶ **Enter the (complete/partial) answer in your own words:**

▶ **In the new piece of text, highlight the main word or short phrase suggesting this answer.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

You entered the following question:

What happened?

▶ **Was that question answered in the new piece of text?**

Not answered at all. 1 2 3 4 5 *Completely answered.*

▶ **Enter the (complete/partial) answer in your own words:**

▶ **In the new piece of text, highlight the main word or short phrase suggesting this answer.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

You entered the following question:

What happened?

▶ **Was that question answered in the new piece of text?**

Not answered at all. 1 2 3 4 5 *Completely answered.*

▶ **Enter the (complete/partial) answer in your own words:**

The person got fired due to oversleeping.

▶ **In the new piece of text, highlight the main word or short phrase suggesting this answer.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

You entered the following question:

What happened?

▶ **Was that question answered in the new piece of text?**

Not answered at all. 1 2 3 4 5 *Completely answered.*

▶ **Enter the (complete/partial) answer in your own words:**

The person got fired due to oversleeping.

▶ **In the new piece of text, highlight the main word or short phrase suggesting this answer.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

▶ **Please enter another question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

▶ **In the text, highlight the main word or short phrase that evokes this question.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

▶ **Please enter another question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

What else made it the worst day?

▶ **In the text, highlight the main word or short phrase that evokes this question.**

Today was the worst day of my life. First of all, my alarm didn't go off, so I arrived late at work again. The boss decided to fire me this time.

▶ **Please enter another question the text evokes for you at this point.**
(The text so far must *not* yet contain an answer to the question!)

What else made it the worst day?

▶ **In the text, highlight the main word or short phrase that evokes this question.**

The TED-Q dataset (Westera, Mayol & Rohde, 2020 LREC)

Elicitation phase:		Comparison phase:	
texts:	6	question pairs:	4516
words:	6975	participants/pair:	6
probe points:	460	participants:	163
participants/probe:	5+	judgments:	30412
participants:	111	RELATED mean:	1.21
questions:	2412	RELATED std:	0.79
answers:	1107	Agreement (AC_2):	.46
ANSWERED mean:	2.50		
ANSWERED std:	1.51		

The TED-Q dataset (Westera, Mayol & Rohde, 2020 LREC)

Elicitation phase:		**Comparison phase:**	
texts:	6	question pairs:	4516
words:	6975	participants/pair:	6
probe points:	460	participants:	163
participants/probe:	5+	judgments:	30412
participants:	111	RELATED mean:	1.21
questions:	2412	RELATED std:	0.79
answers:	1107	Agreement (AC_2):	.46
ANSWERED mean:	2.50		
ANSWERED std:	1.51		

Experiment: Predictability ~ Implicitness

Experiment: Predictability ~ Implicitness

Background:

- Uniform Information Density (Frank and Jaeger, 2008);

Experiment: Predictability ~ Implicitness

Background:

- Uniform Information Density (Frank and Jaeger, 2008);
- Lexical predictability leads to acoustic reduction (Jurafsky et al., 2001; Arnold 1999, 2001)

Experiment: Predictability ~ Implicitness

Background:

- Uniform Information Density (Frank and Jaeger, 2008);
- Lexical predictability leads to acoustic reduction (Jurafsky et al., 2001; Arnold 1999, 2001)
- Applied to discourse structure by Asr and Demberg (2012).

Experiment: Predictability ~ Implicitness

Background:

- Uniform Information Density (Frank and Jaeger, 2008);
- Lexical predictability leads to acoustic reduction (Jurafsky et al., 2001; Arnold 1999, 2001)
- Applied to discourse structure by Asr and Demberg (2012).

Our approach:

- Our source texts came from TED-MDB (Zeyrek et al. 2018), annotated for *explicit* and *implicit* discourse connectives.

Experiment: Predictability ~ Implicitness

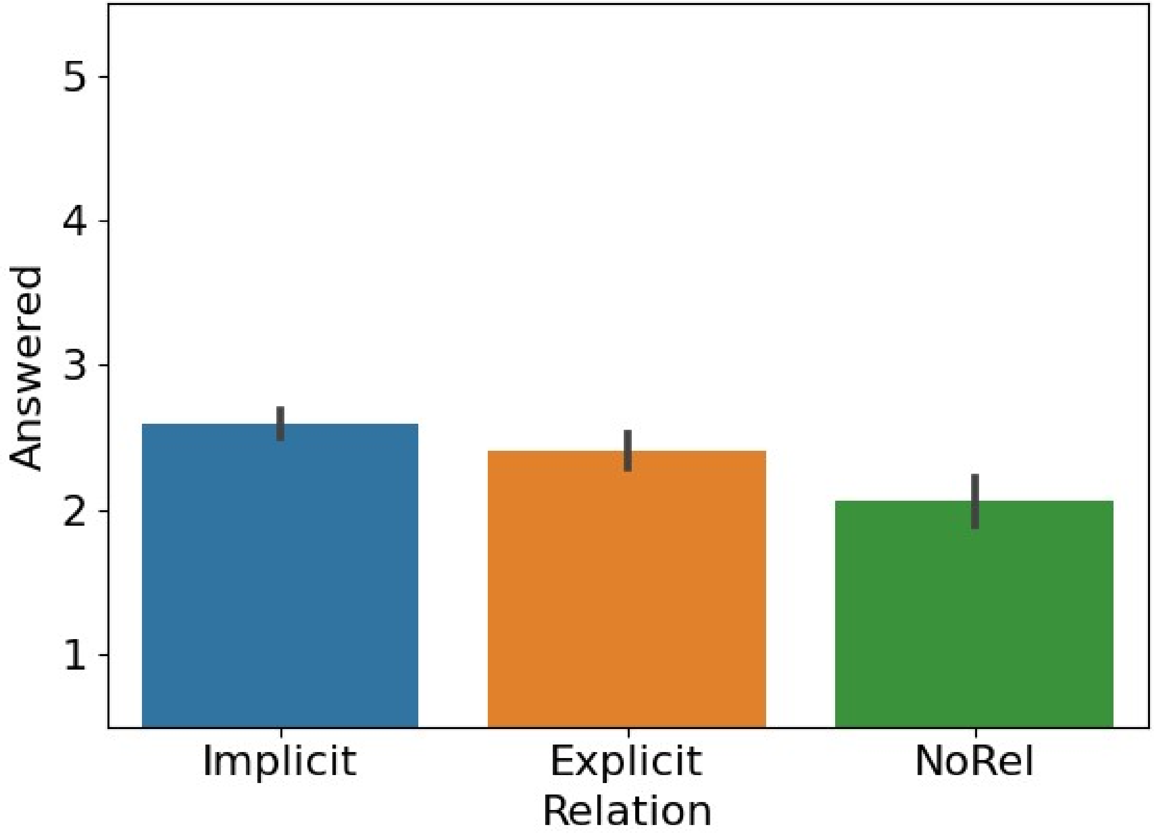
Background:

- Uniform Information Density (Frank and Jaeger, 2008);
- Lexical predictability leads to acoustic reduction (Jurafsky et al., 2001; Arnold 1999, 2001)
- Applied to discourse structure by Asr and Demberg (2012).

Our approach:

- Our source texts came from TED-MDB (Zeyrek et al. 2018), annotated for *explicit* and *implicit* discourse connectives.
- TED-Q's ANSWERED scores ~ discourse structure predictability.

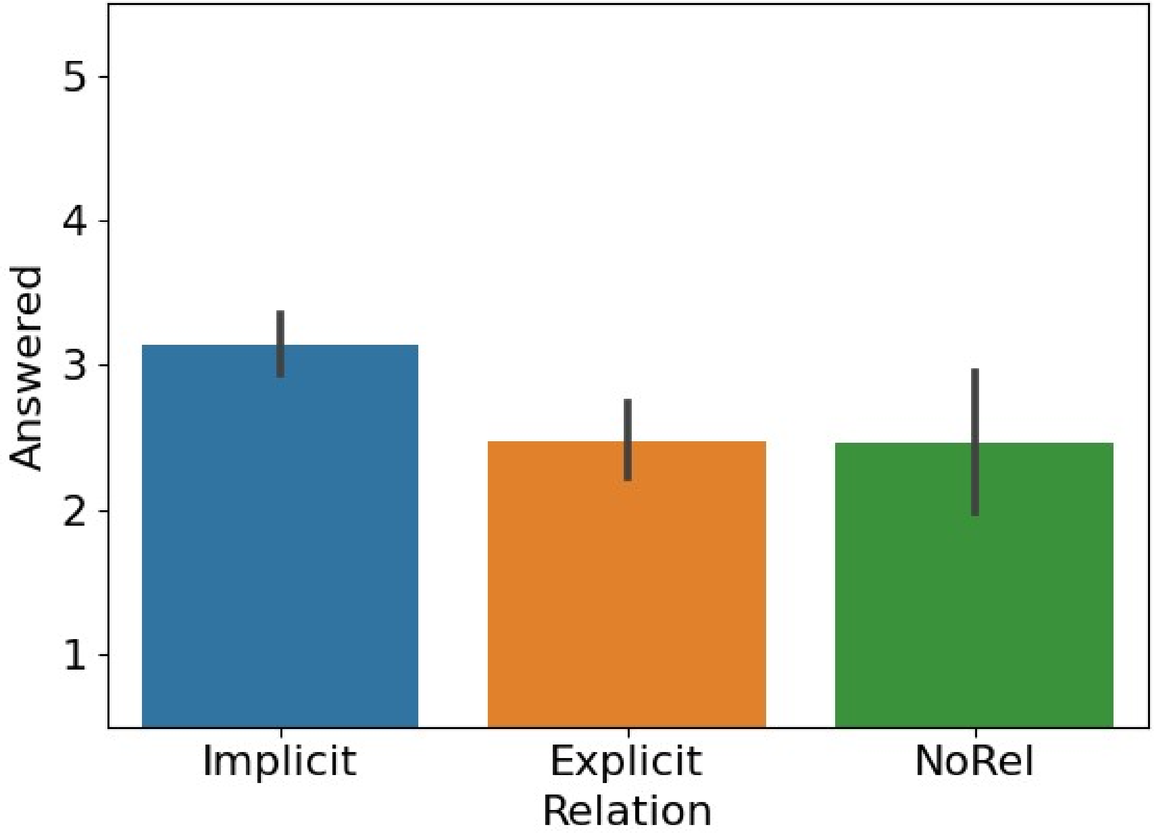
Main finding



Kruskal-Wallis H-test: $p=6.8e$-7

Dunn's post-hoc test (Bonferroni):
- Implicit, Explicit: 0.044
- Implicit, NoRel: $4.3e$-07
- (Explicit, NoRel: 0.003)

Main finding

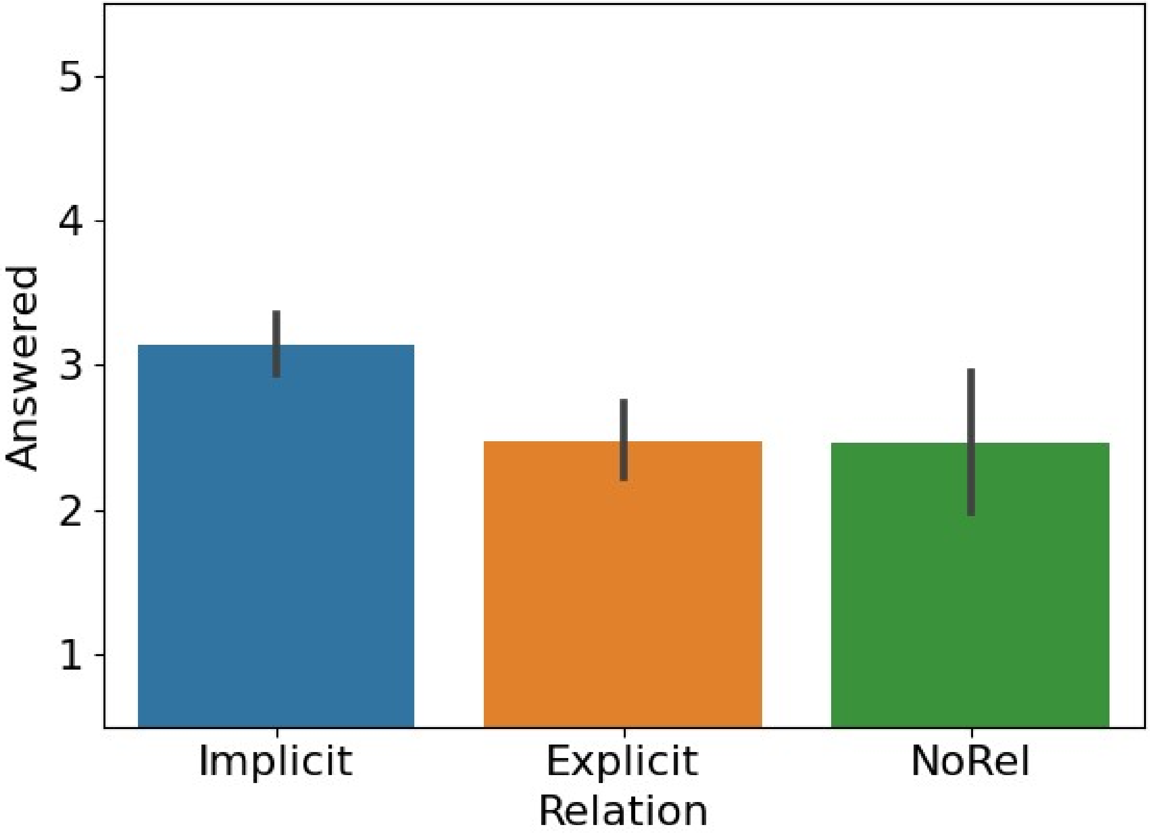


Kruskal-Wallis H-test: $p = 0.0001$

Dunn’s post-hoc test (Bonferroni):
- Implicit, Explicit: 0.00023
- Implicit, NoRel: 0.025
- (Explicit, NoRel: 1.000)

Main finding

Restricted to the top 25% places with highest agreement about the evoked questions.



Kruskal-Wallis H-test: $p = 0.0001$

Dunn's post-hoc test (Bonferroni):
- Implicit, Explicit: 0.00023
- Implicit, NoRel: 0.025
- (Explicit, NoRel: 1.000)

Next steps

- Computational modeling (Westera, Amidei & Mayol, submitted to CoLing)
- Look into the ‘highlighting’ data – relation to information structure.
- *Maybe* getting more/better/controlled data.

2.2. Referent predictability

Ongoing work with Xixian Liao, Laura Aina, Laia Mayol and Gemma Boleda.




Referent predictability and referring expression choice

Referent predictability and referring expression choice

Background:

- Again: Uniform Information Density (Frank and Jaeger, 2008).

Referent predictability and referring expression choice

Background:

- Again: Uniform Information Density (Frank and Jaeger, 2008).
- Applied to referring expressions (e.g., Stevenson et al., 1994; Tily & Piantadosi 2009; Kehler & Rohde 2013); no consensus.

Referent predictability and referring expression choice

Background:

- Again: Uniform Information Density (Frank and Jaeger, 2008).
- Applied to referring expressions (e.g., Stevenson et al., 1994; Tily & Piantadosi 2009; Kehler & Rohde 2013); no consensus.

Idea:

- Use *coreference resolution model* to compute a proxy for referent predictability.

Coreference resolution model

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

[new entity]

Sue was at the protest with Bill and his daughter. She called me later.

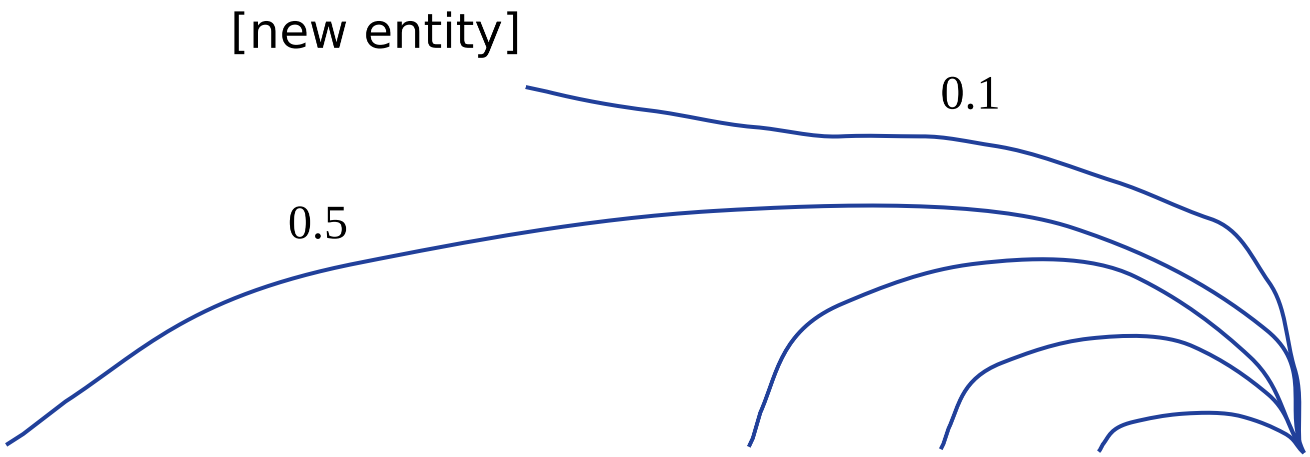
Coreference resolution model

[new entity]

0.1

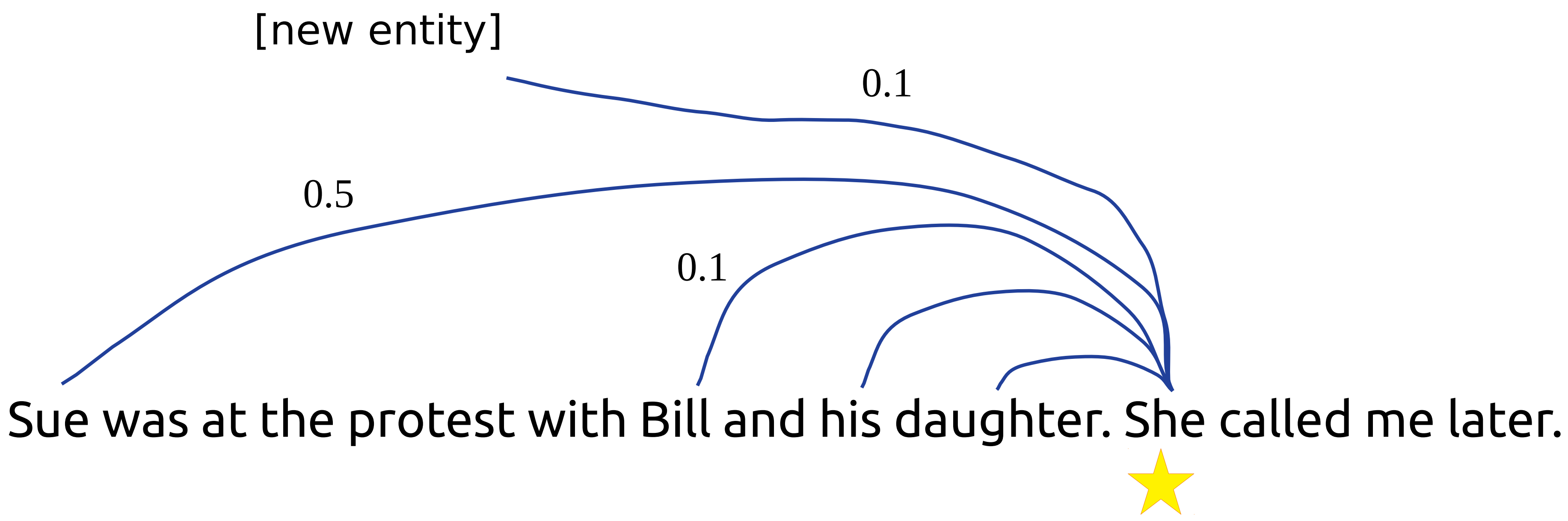
Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model

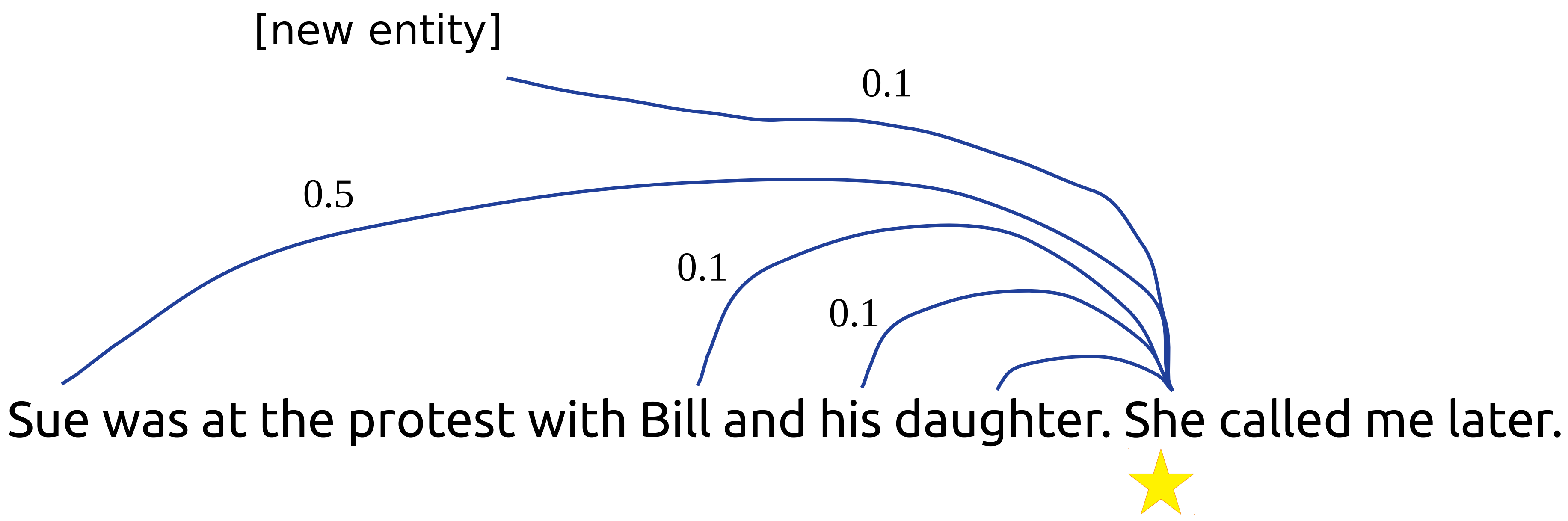


Sue was at the protest with Bill and his daughter. She called me later.

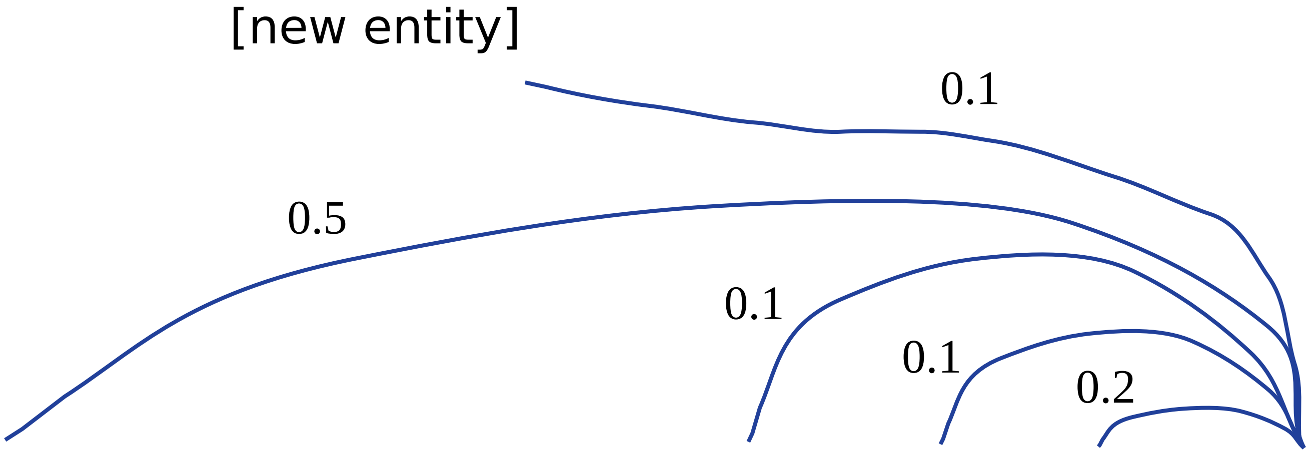
Coreference resolution model



Coreference resolution model

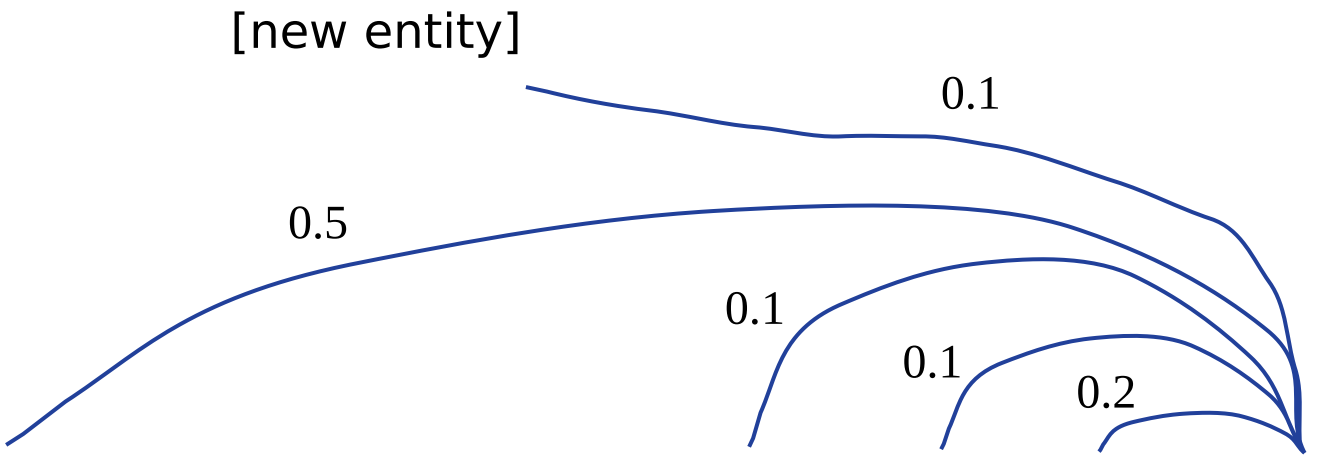


Coreference resolution model



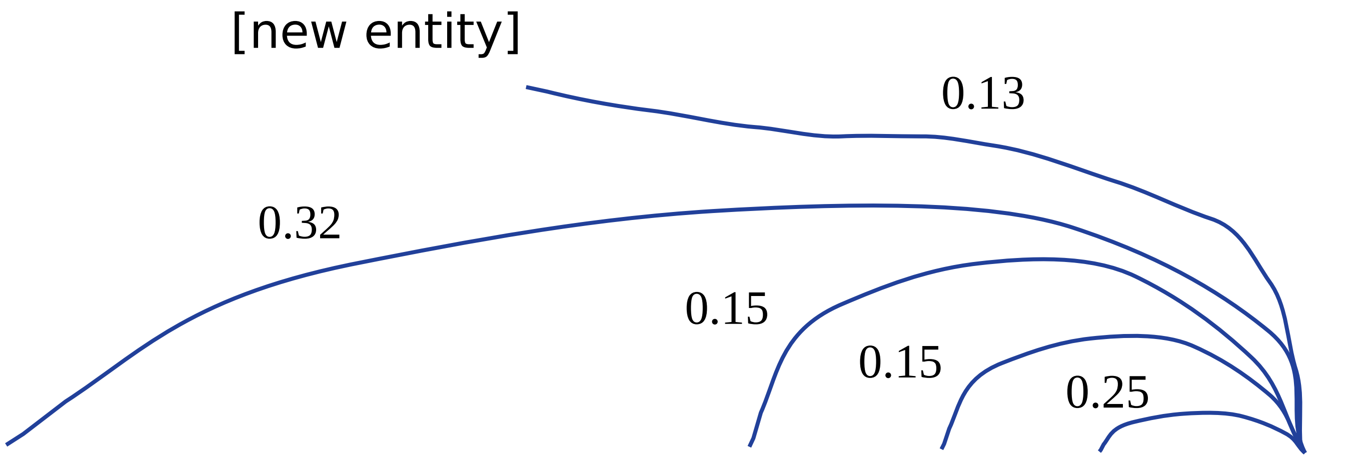
Sue was at the protest with Bill and his daughter. She called me later.

Coreference resolution model



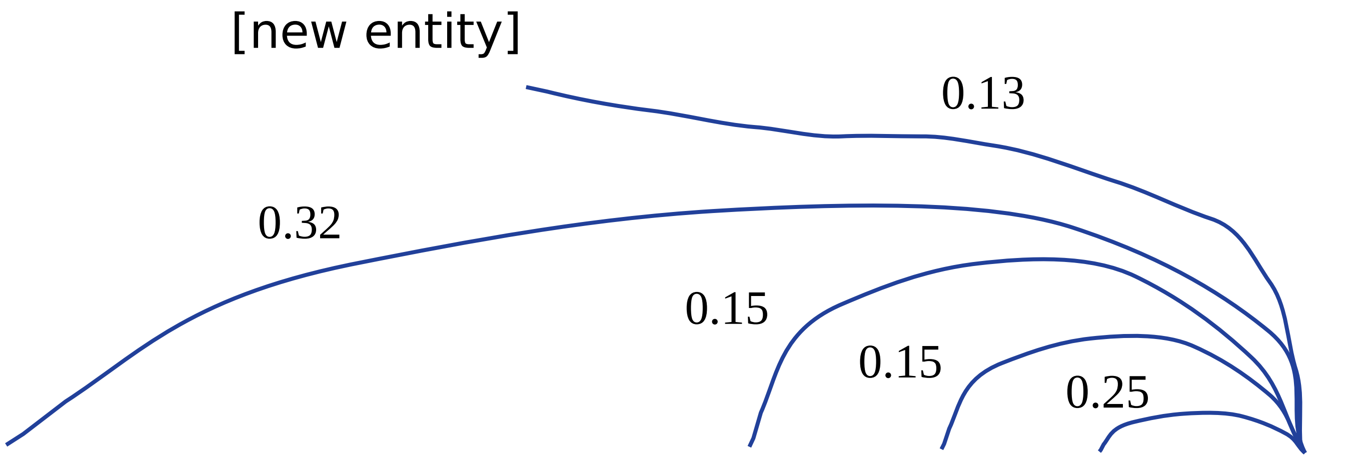
Sue was at the protest with Bill and his daughter. [redacted] called me later.

Coreference resolution model



Sue was at the protest with Bill and his daughter. [redacted] called me later.

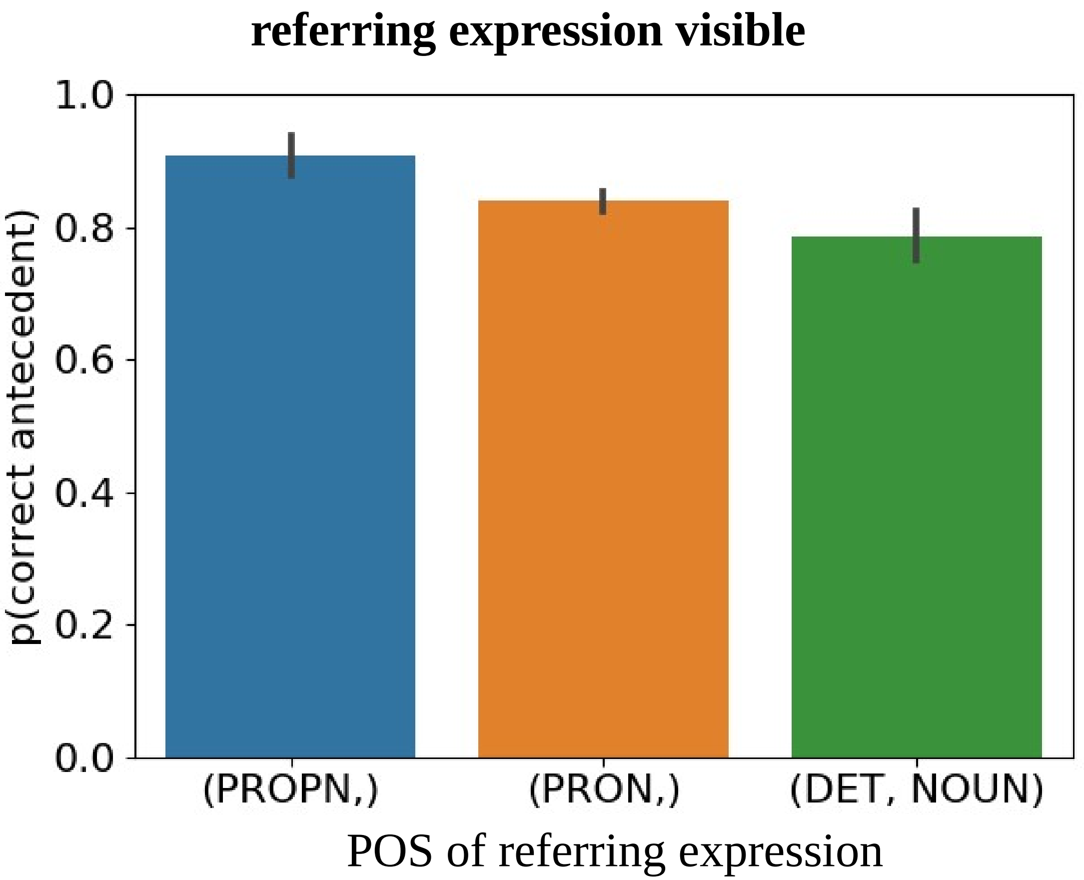
Coreference resolution model



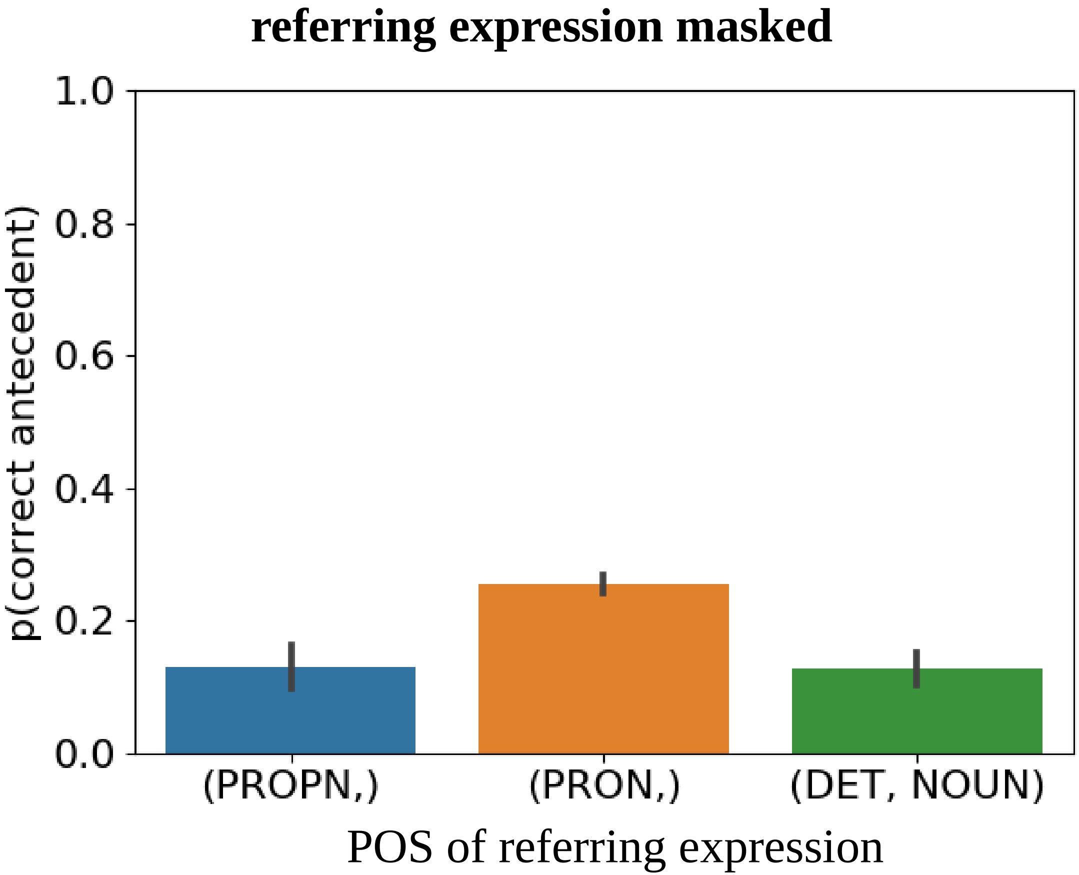
Sue was at the protest with Bill and his daughter. [redacted] called me later.

- I used the SpanBERT model Joshi et al. (2019).

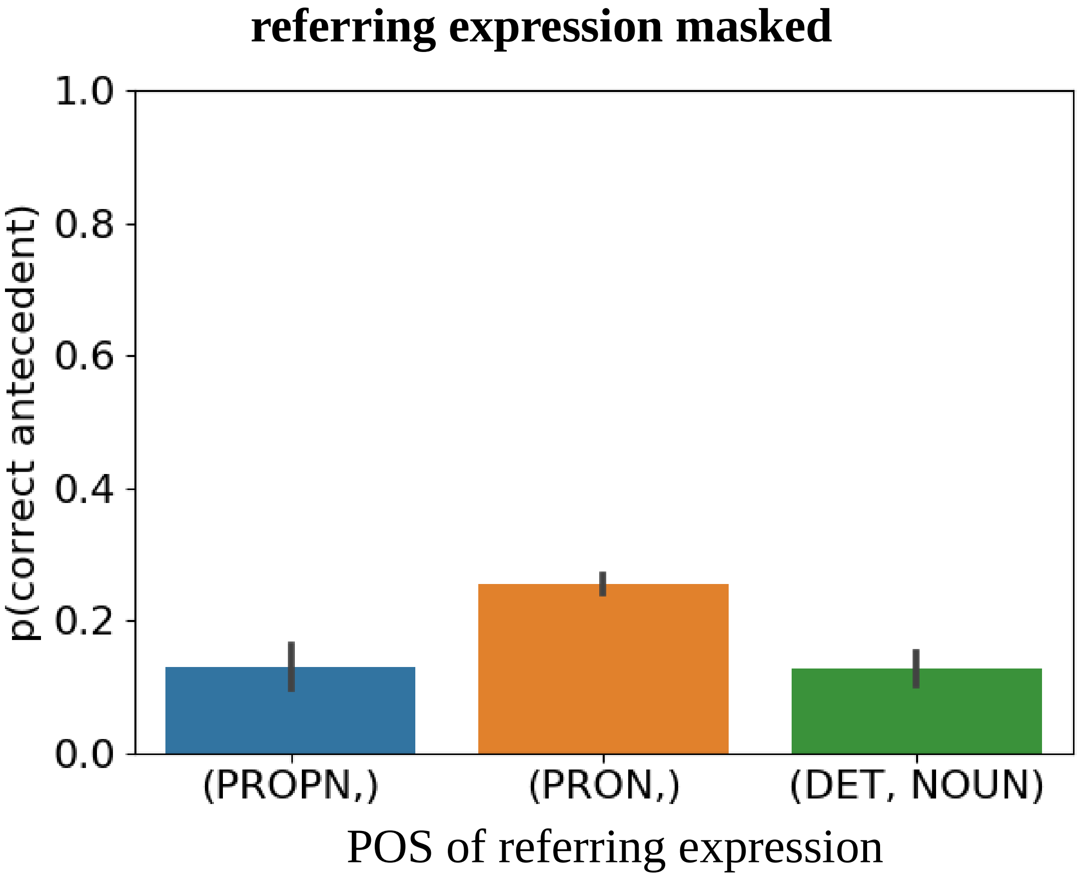
Results (pilot)



Results (pilot)



Results (pilot)



ANOVA: p=1e-19

Tukey's HSD:

- (PRON), (DET,NOUN): 0.001
- (PRON), (PROPN): 0.001
- (DET,NOUN), (PROPN): 0.9

Conclusion

- Pronouns have more expected antecedents (*according to this model)*.

Conclusion

- Pronouns have more expected antecedents (*according to this model)*.
- Plausibly use machine learning models as a proxy for human processing.

Conclusion

- Pronouns have more expected antecedents (*according to this model)*.
- Plausibly use machine learning models as a proxy for human processing.

Next:

- Explore some different implementations of this idea.

Conclusion

- Pronouns have more expected antecedents (*according to this model)*.
- Plausibly use machine learning models as a proxy for human processing.

Next:

- Explore some different implementations of this idea.
- Compare more POS, different genres, fine-grained distinctions (e.g., definite/indefinite; subject/object), different languages (e.g., Pro-drop).

Summary

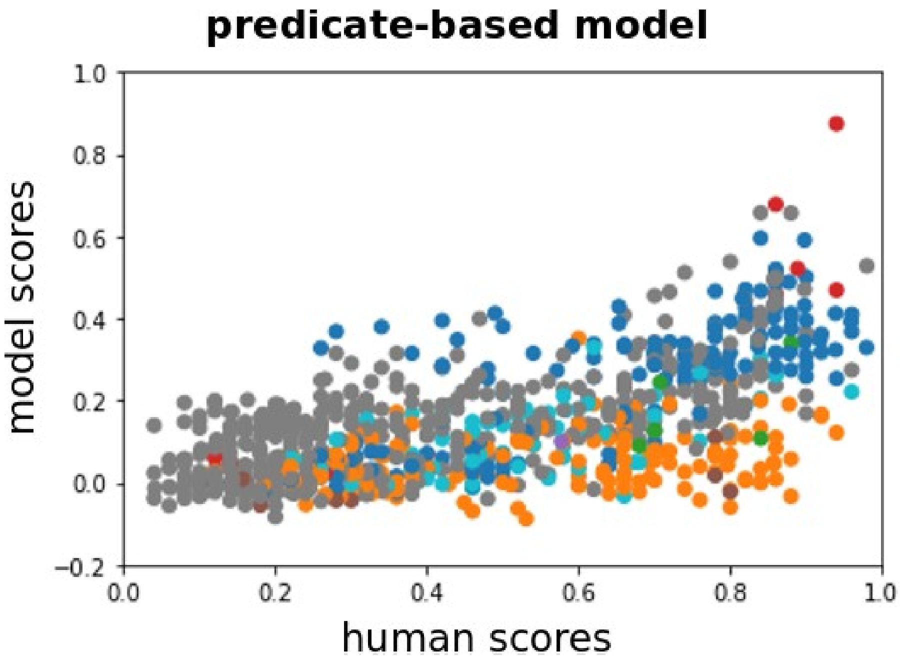
Two research strands

1) What is meaning?
   - semantics vs. pragmatics
   - distributional vs. formal semantics
   - neural networks vs. linguistic theory.

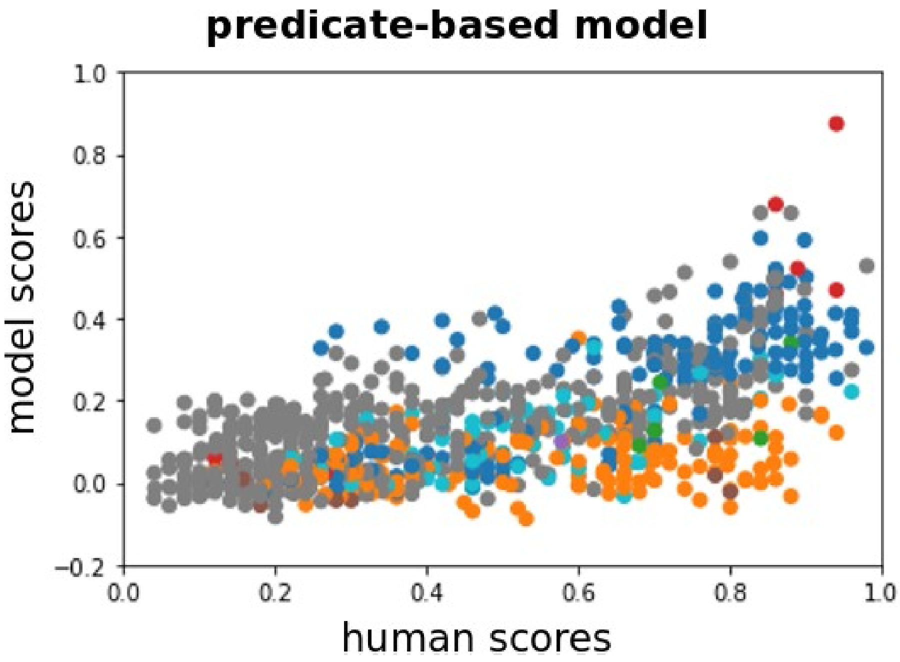
2) Understanding discourse structure (goals, topics)
   - implicit questions
   - referent predictability

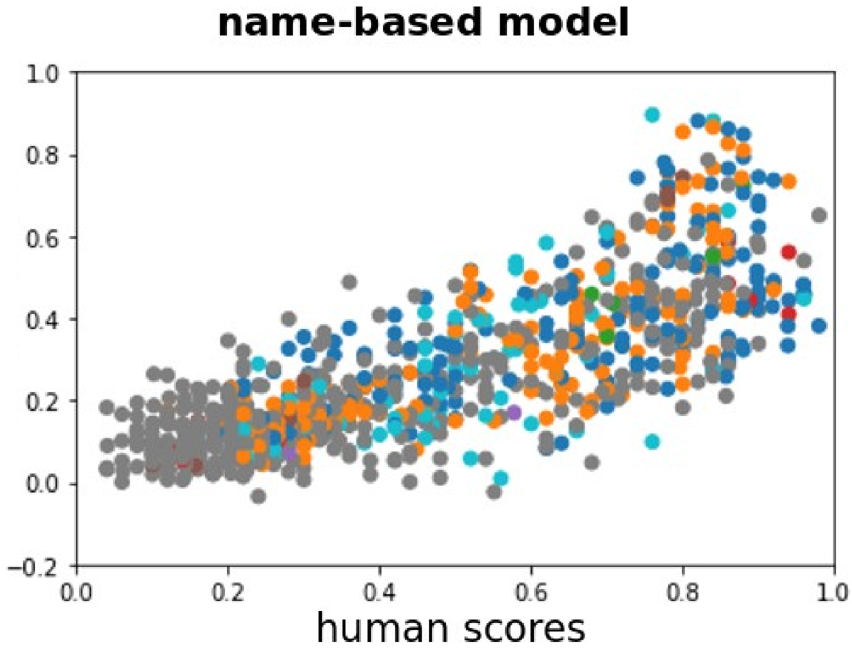
Appendix

Confetti plots



Confetti plots





Entities need to be *representative*

Entities need to be *representative*

- E.g., the Name-based model overestimates *surgeon ~ siege*…

Entities need to be *representative*

- E.g., the Name-based model overestimates *surgeon ~ siege*...
- Instances of surgeon in the Instantiation dataset:
  - William Cowper
  - James Parkinson
  - Alexis Carrel
  - Walter Reed
  - William Beaumont
  - Joseph Lister

Entities need to be *representative*

- E.g., the Name-based model overestimates *surgeon* ~ *siege*...
- Instances of surgeon in the Instantiation dataset:
  - William Cowper
  - James Parkinson
  - Alexis Carrel
  - Walter Reed ← Involved in WW1
  - William Beaumont
  - Joseph Lister

Entities need to be *representative*

- E.g., the Name-based model overestimates *surgeon* ~ *siege*…
- Instances of surgeon in the Instantiation dataset:
  - William Cowper
  - James Parkinson
  - Alexis Carrel
  - Walter Reed ← Involved in WW1
  - William Beaumont
  - Joseph Lister

Members of US military corps

Entities need to be *representative*

- E.g., the Name-based model overestimates *surgeon ~ siege*...
- Instances of surgeon in the Instantiation dataset:
  - William Cowper
  - James Parkinson — Wrote "the siege of chester" (?)
  - Alexis Carrel
  - Walter Reed — Involved in WW1
  - William Beaumont
  - Joseph Lister

Members of US military corps

Correlation with polysemy

Table 4: Spearman correlations between model error (absolute rank difference) and number of synsets, with p-values in parenthesis.

	all	match	unclear	within-domain	between-domain
number of pairs	981	626	355	484	497
NounBased	**0.13** (2e−5)	**0.11** (0.007)	**0.16** (0.002)	**0.21** (3e−6)	0.043
NameBased	0.023	0.078	-0.053	0.024	0.030

Which questions are ‘the same’?

- http://mwestera.humanities.uva.nl/crowdsource/evoque.html

▸ **Please read the snippet:**

> [...] Who here knows that in many cities across the United States it is now illegal to sit on the sidewalk, to wrap oneself in a blanket, to sleep in your own car, to offer food to a stranger?

▸ **Next, compare the questions it evoked:**

	Questions:	How related are target (T) and comparison (C) question?				
Target (T):	**Which are some cities who have these laws?**					
Comparison (C):	**Why is it illegal to offer food to someone?**	T=C	C T	C T	C T	?
Comparison (C):	**What cities are affected?**	T=C	C T	C T	C T	?
Comparison (C):	**Why would it be illegal to offer food to a stranger?**	T=C	C T	C T	C T	?
Comparison (C):	**Why would it be against the law to be kind?**	T=C	C T	C T	C T	?

ANSWERED & RELATED