

# Some of Them Can be Guessed!

## Exploring the Effect of Linguistic Context in Predicting Quantifiers

Sandro Pezzelle<sup>1</sup>, Shane Steinert-Threlkeld<sup>2</sup>, Raffaella Bernardi<sup>1,3</sup>, Jakub Szymanik<sup>2</sup>

<sup>1,3</sup>CiMeC - Center for Mind/Brain Sciences, <sup>3</sup>DISI, University of Trento  
<sup>2</sup>ILLC - Institute for Logic, Language and Computation, University of Amsterdam  
[sandro.pezzelle@unitn.it](mailto:sandro.pezzelle@unitn.it) | [quantit-clic.github.io](https://github.com/quantit-clic)

### Motivation

**Quantifiers** ('few', 'some', 'all') are interesting because:

- They are typically considered as function words (as opposed to nouns, verbs, etc.), but they have a rich semantics
- They are of central importance in linguistic semantics and its interface with cognitive science [1,2,3]
- Their choice depends on both local and global context [4]
- Larger contexts are claimed to be detrimental for the prediction of function words in cloze test [5]

### Hypotheses

- Human performance **boosted** by more context (proportional Qs)
- Models (very) effective with local context, **hurt** by broader context

### Task & Datasets

#### Cloze test

<qnt> the island's breeding birds are endemic.

The island is one of the world's most biologically diverse areas, with many endemic species.

<qnt> the island's breeding birds are endemic.

Other endemic species include the red-bellied lemur, the indri, and the aye-aye.

#### Quantifiers

a few of  
 all of  
 almost all of  
 few of  
 many of  
**more than half of**  
 most of  
 none of  
 some of

#### Datasets

**1-Sent** 10350 target sentences (quantifier+ of at beginning): <s<sub>t</sub>>

**3-Sent** 10350 preceding + s<sub>t</sub> + following: <S<sub>p</sub>,S<sub>t</sub>,S<sub>f</sub>>

### Human Evaluation

#### Crowdsourcing

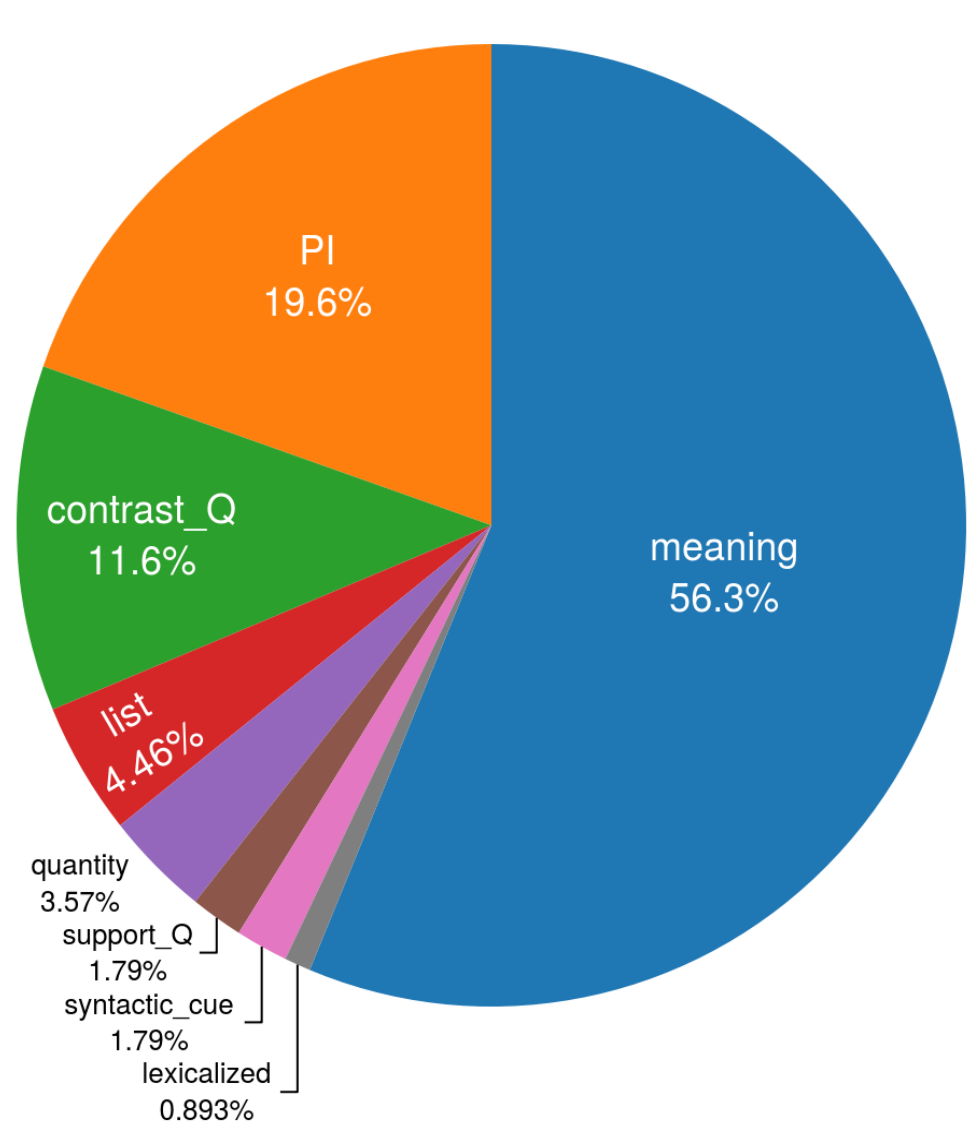
- Two experiments, one per condition (**1-Sent**, **3-Sent**)
- 506 examples from validation set (same in two conditions)
- 3 judgments/datapoint; correctly-guessed w/ agreement > 0.66
- Higher accuracy in **3-Sent** (0.258) compared to **1-Sent** (0.221)

### Models & Results

#### Models

8 models tested: 3 BoW baselines, 1 CNN, 4 LSTMs  
 2 conditions: **1-Sent**, **3-Sent**  
 Data: 80% train, 10% val, 10% test

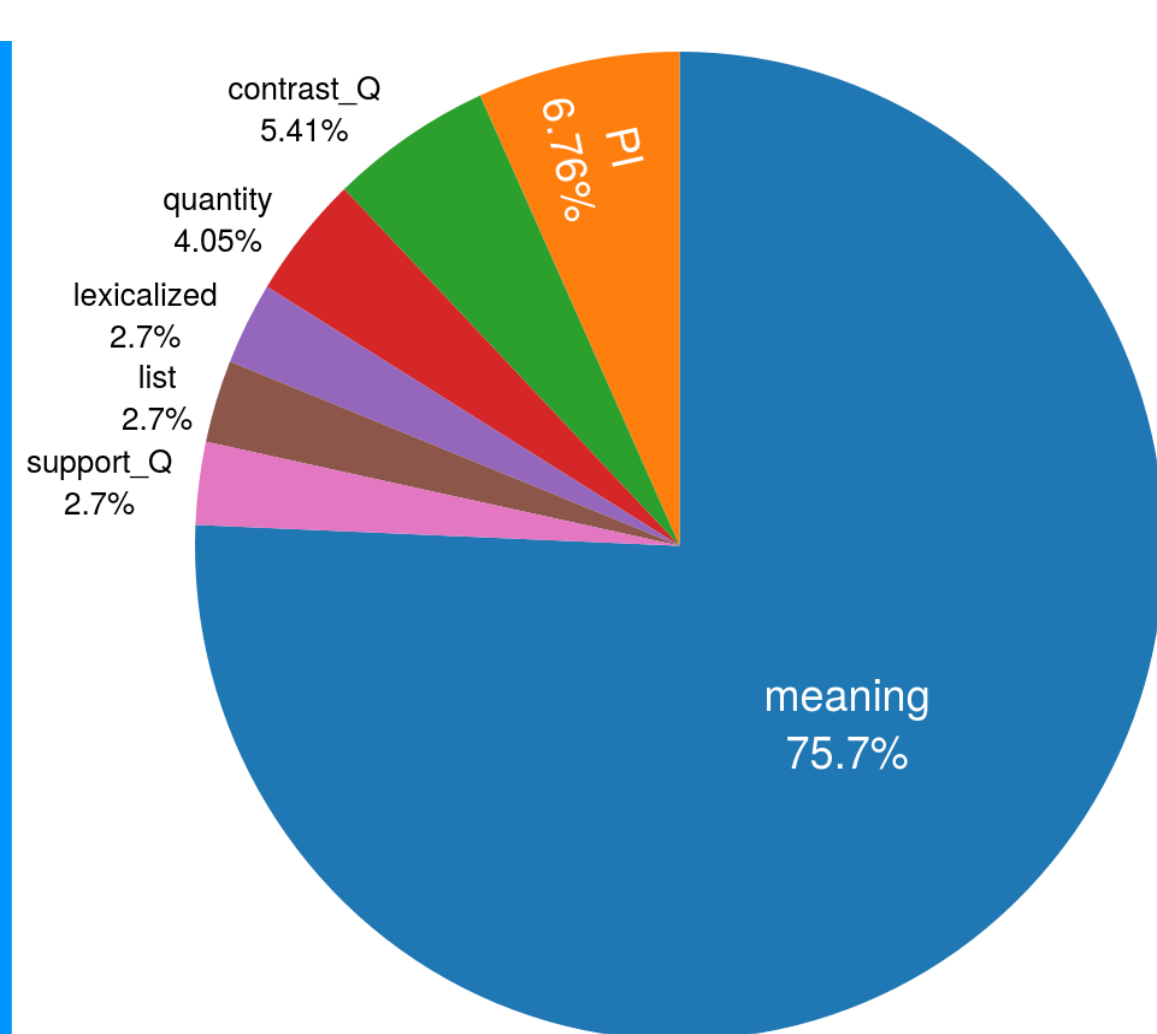
#### 1-Sent



#### Linguistic cues

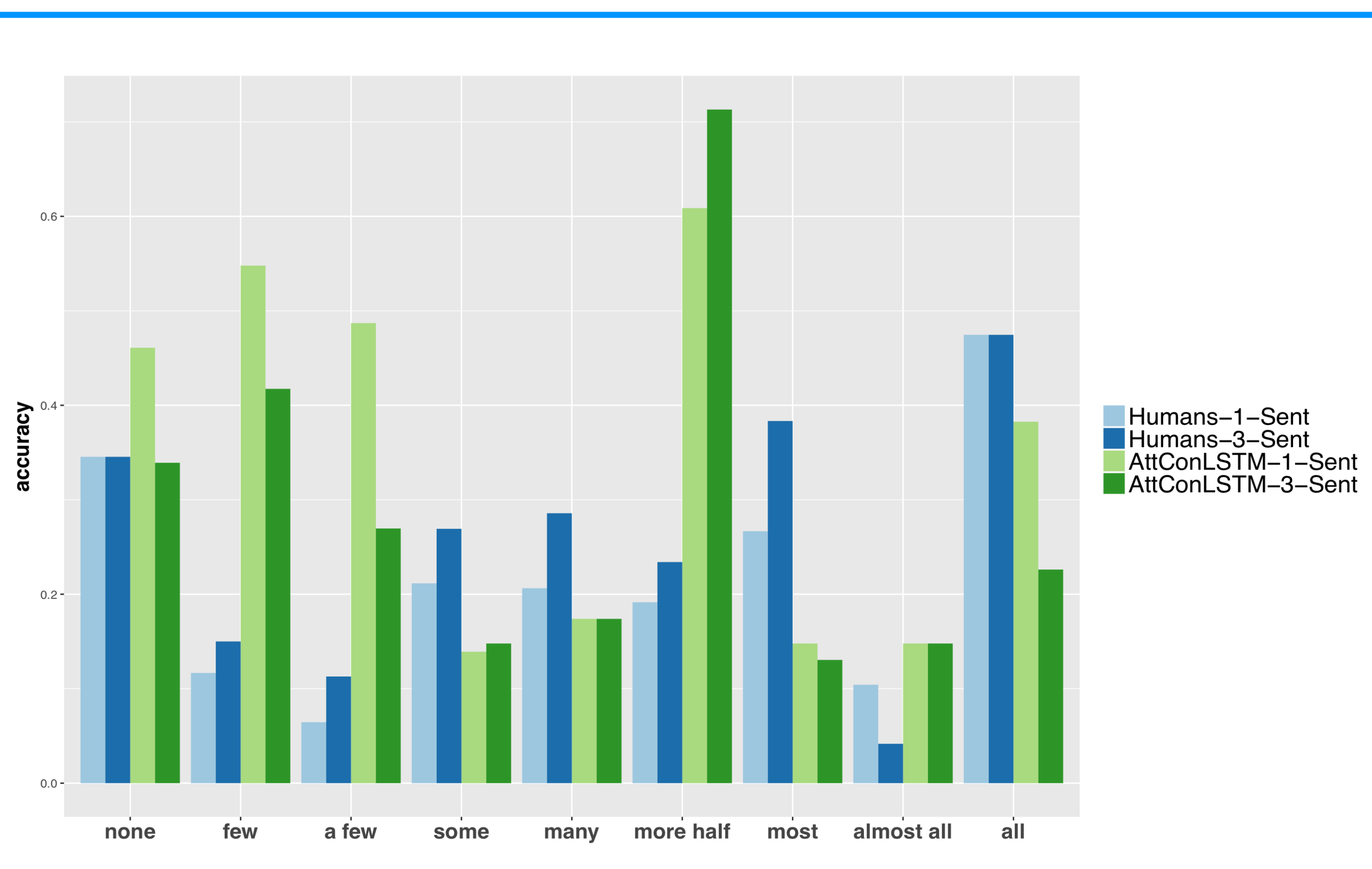
Meaning  
 Polarity Item  
 Contrast Q  
 Support Q  
 Quantity  
 List  
 Lexicalization  
 Syntax

#### 3-Sent



	1-Sent		3-Sent	
	val	test	val	test
chance	0.111	0.111	0.111	0.111
BoW-conc	0.270	0.238	0.224	0.207
BoW-sum	0.308	0.290	0.267	0.245
fastText	0.305	0.271	0.297	0.245
CNN	0.310	0.304	<b>0.298</b>	0.257
LSTM	0.315	0.310	0.277	0.253
bi-LSTM	0.341	<b>0.337</b>	0.279	0.265
Att-LSTM	0.319	0.324	0.287	<b>0.291</b>
AttCon-LSTM	<b>0.343</b>	0.319	0.274	0.288
Humans	0.221*	—	0.258*	—

### Humans vs Models



### Discussion & References

#### Discussion

Humans **do better** w/ broader contexts especially on proportional Qs; models **suffer** due to their inability to handle longer sequences

Models capitalize more on **lexical** cues compared to humans: 41% cases in **3-Sent** (hum. 24%) and 50% cases in **1-Sent** (hum. 44%)

#### References

- [1] Jon Barwise and Robin Cooper. 1981. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4(2):159-219.
- [2] Stanley Peters and Dag Westerståhl. 2006. *Quantifiers in Language and Logic*. Clarendon Press, Oxford.
- [3] Jakub Szymanik. 2016. *Quantifiers and Cognition. Logical and Computational Perspectives. Studies in Linguistics and Philosophy*. Springer.
- [4] Kevin B. Paterson, Ruth Filik, and Linda M. Moxey. 2009. *Quantifiers and Discourse Processing*. *Language and Linguistics Compass*.
- [5] Frank Smith. 1971. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. Holt, Rinehart & Winston.