Exploring (Multimodal) Model Abilities in Grounding *Vague* Expressions

BlackBoxNL - May 6, 2019

Sandro Pezzelle

ILLC - Institute for Logic, Language & Computation, University of Amsterdam
s.pezzelle@uva.nl
https://sandropezzelle.github.io/

Overview

- **Vague expressions** (most, small) ubiquitous in language

Overview

- **Vague expressions** (most, small) ubiquitous in language
- Single but non-specified meaning: **Context dependence**

Overview

- **Vague expressions** (most, small) ubiquitous in language
- Single but non-specified meaning: **Context dependence**
- Real-world **contexts are multimodal**, e.g. Language + Vision

Overview

- **Vague expressions** (most, small) ubiquitous in language
- Single but non-specified meaning: **Context dependence**
- Real-world **contexts are multimodal**, e.g. Language + Vision
- Can they be learned from **use in *grounded* (visual) contexts**?

Quantifiers

Quantity expressions

Quantity expressions

Numbers

Three dogs

Comparatives

More dogs than cats

Proportions

60% of pets are dogs

Quantifiers

Some pets are dogs

Quantity expressions: Precise vs fuzzy

Numbers

Three/~~two~~ dogs

Comparatives

More/~~less~~ dogs than cats

Proportions

60%/~~50%~~ of pets are dogs

Quantifiers

Some/~~all~~ pets are dogs

Quantity expressions: Precise vs fuzzy

Numbers

Three/~~two~~ dogs

Comparatives

More/~~less~~ dogs than cats

Proportions

60%/~~50%~~ of pets are dogs

Quantifiers

Some/~~all~~ pets are dogs

some ∼ **a few** ∼ **several** ∼ **most**

pets are dogs

Computational study

Questions

Can the meaning of **quantifiers** be learned from vision?

Can a single model **jointly** learn quantifiers, comparatives, and proportions?

Pezzelle, Sorodoc, Bernardi. *NAACL-HLT*, 2018

Motivation

Comparatives, quantifiers, proportions:

- describe (increasingly-complex) **relations** between sets

Motivation

Comparatives, quantifiers, proportions:

- describe (increasingly-complex) **relations** between sets
- subtend different operation compared to **numbers**

  [Fabbri et al., 2012]

Dataset

Comparatives

more

Quantifiers

most 0.43, **some** 0.29, etc.

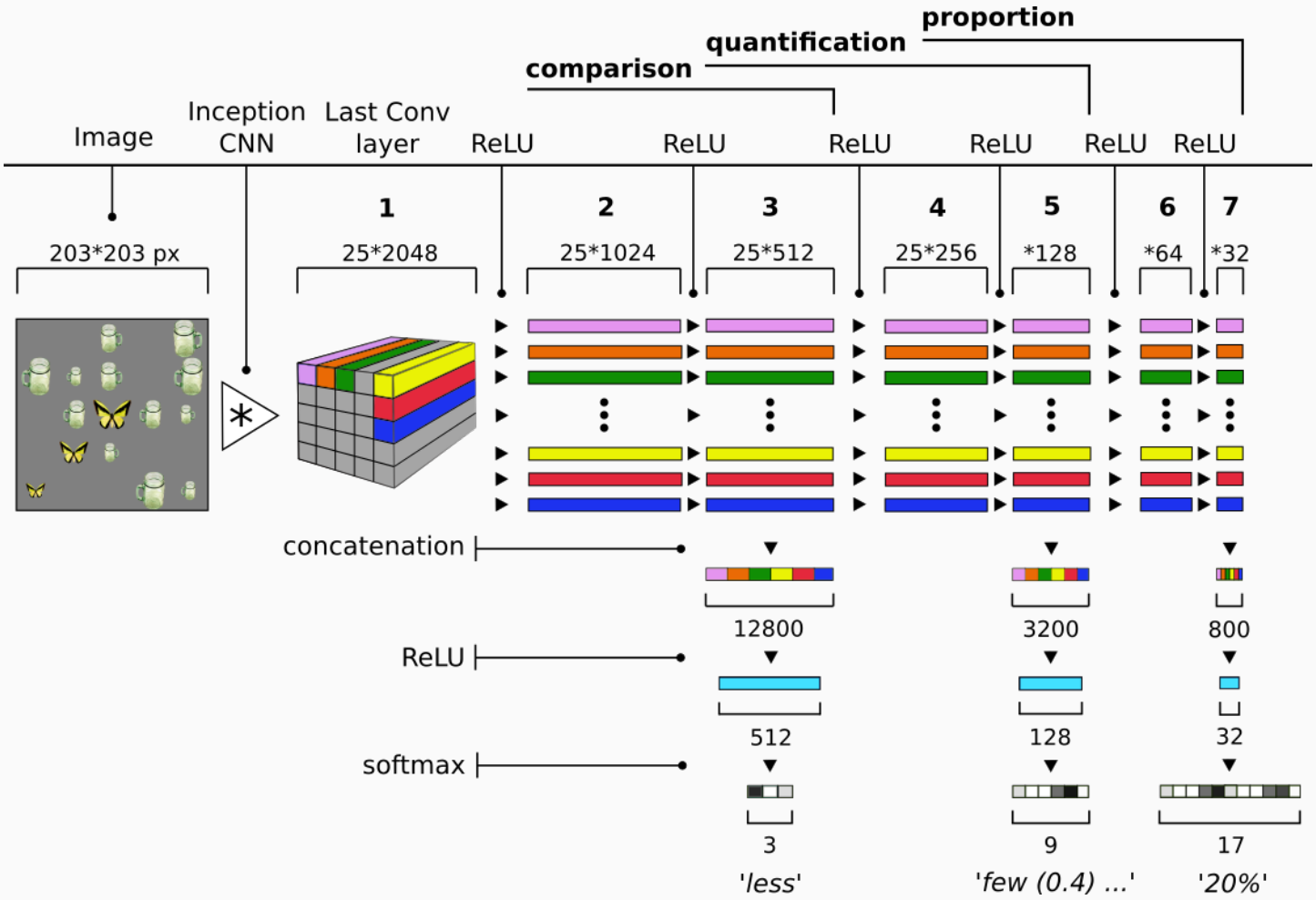
(Pezzelle, Bernardi, Piazza. *Cognition*, 2018)

Proportions

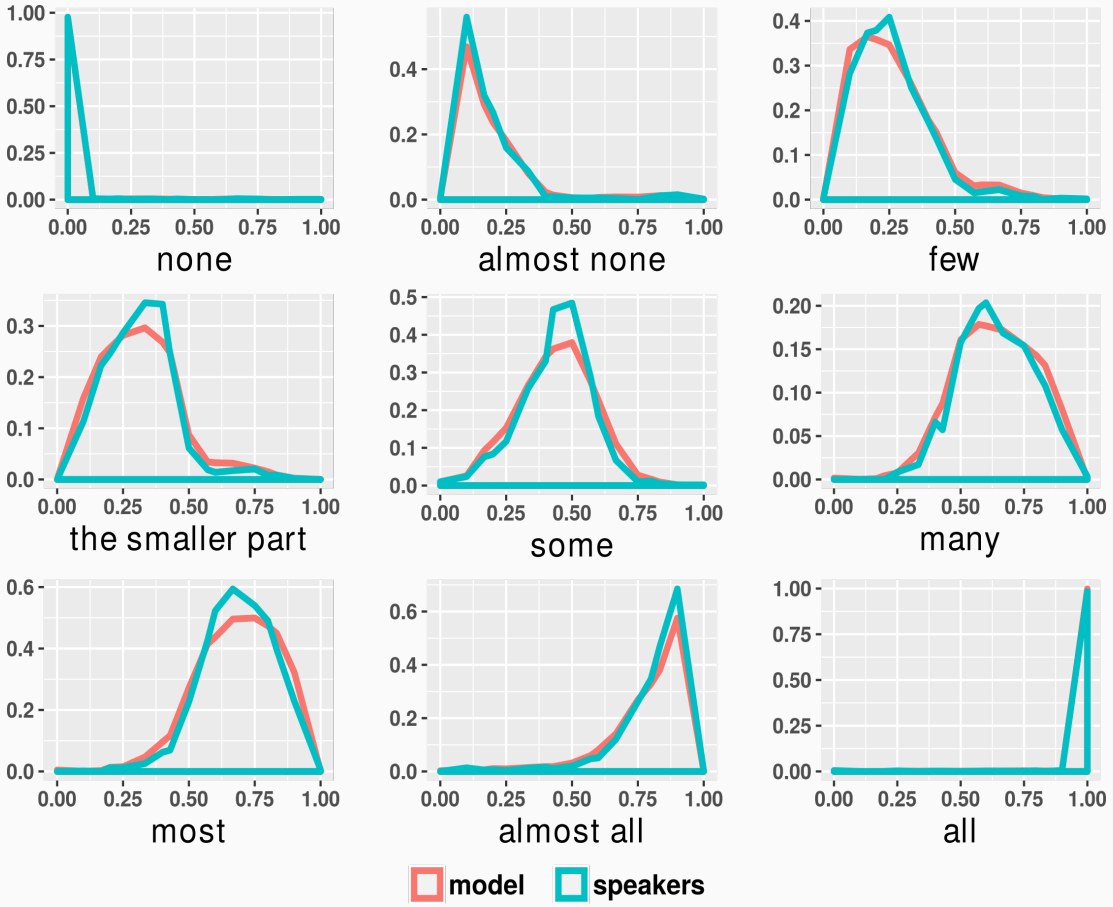
57%

Synthetic: (1) perfect balancing; (2) NO bias; (3) NO world knowledge

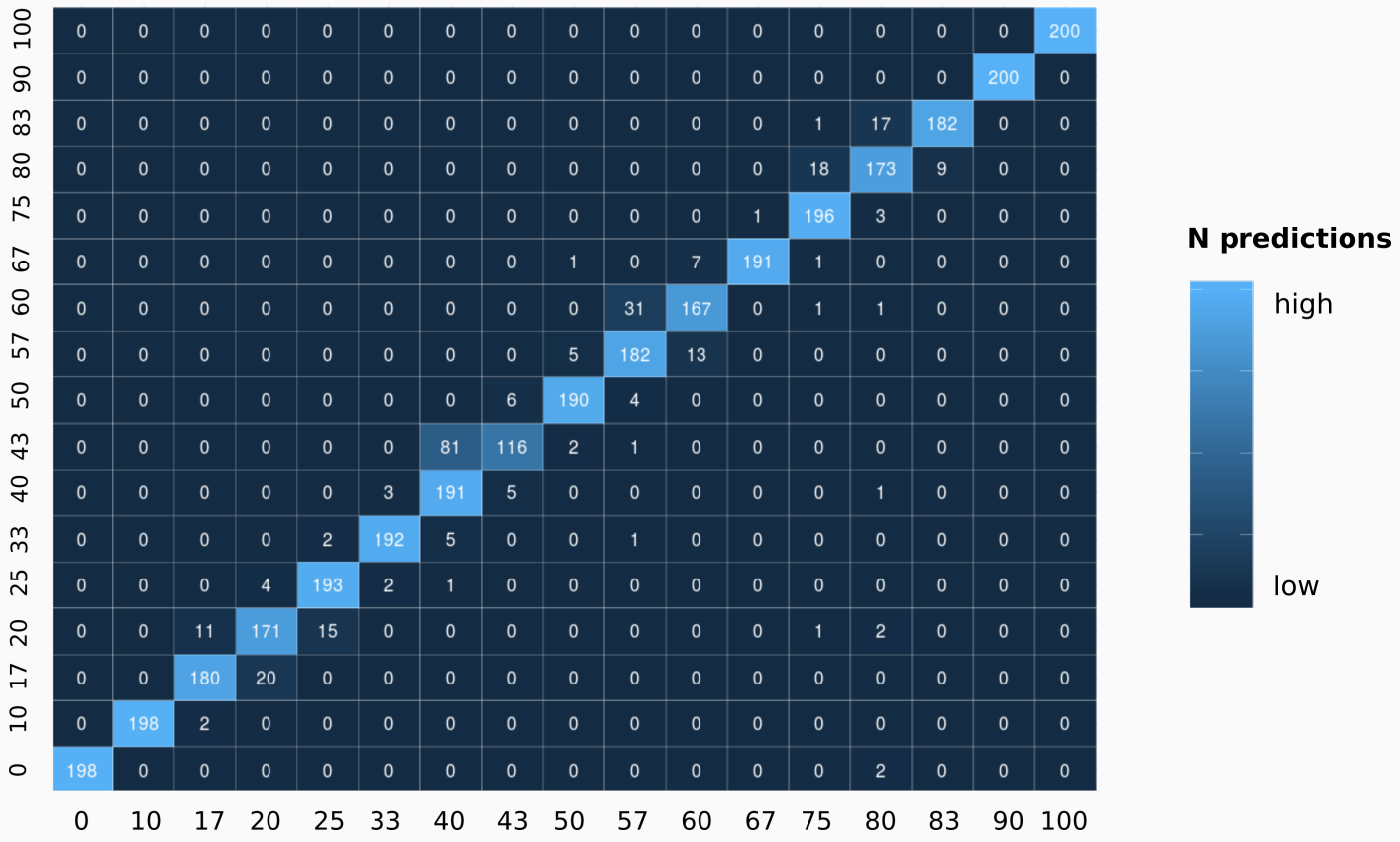
Multi-Task Learning (MTL)



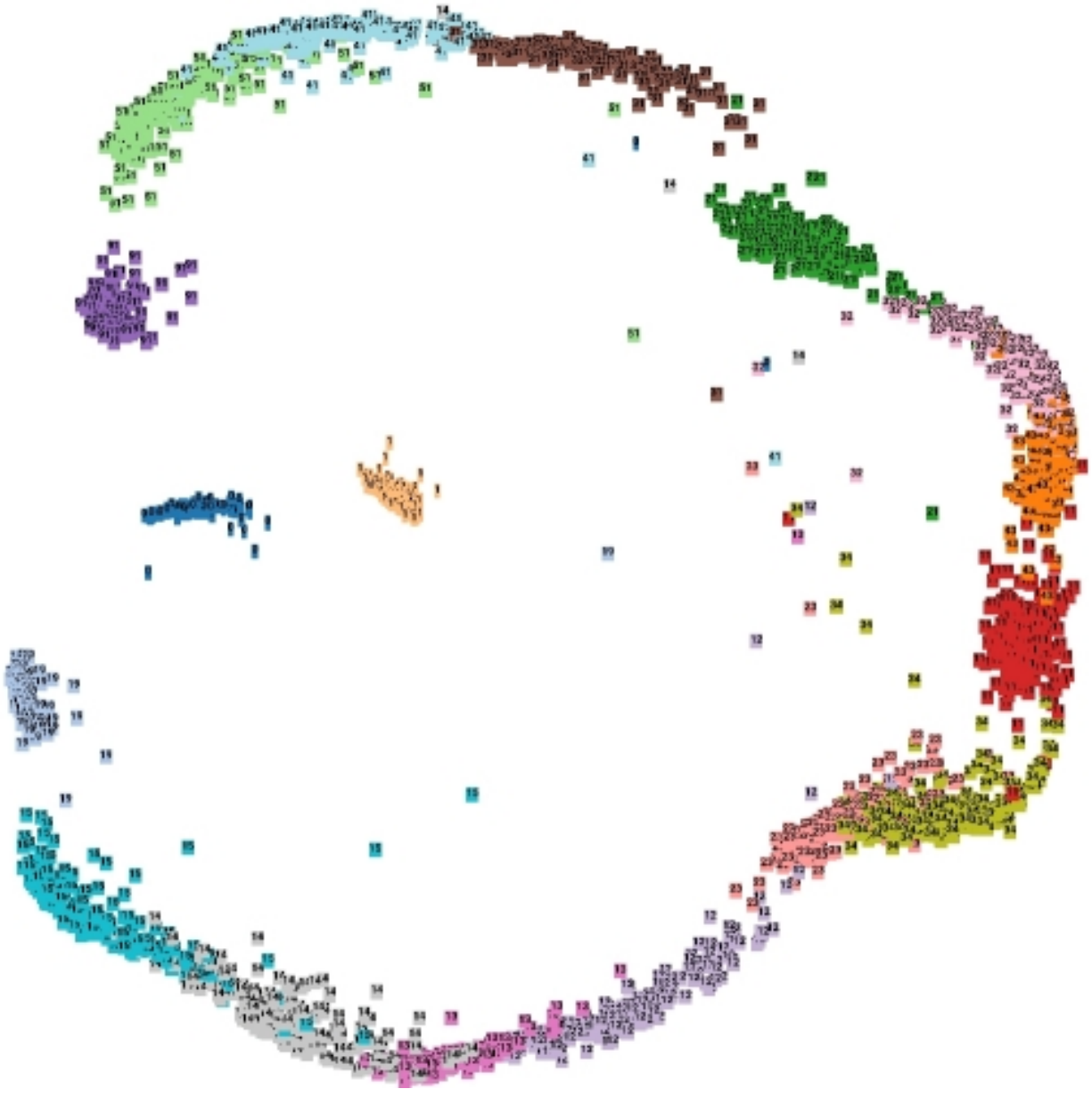
1a. Does the MTL model learn (quantifiers)?



1b. Does the MTL model learn (proportions)?



2a. What does the MTL learn?



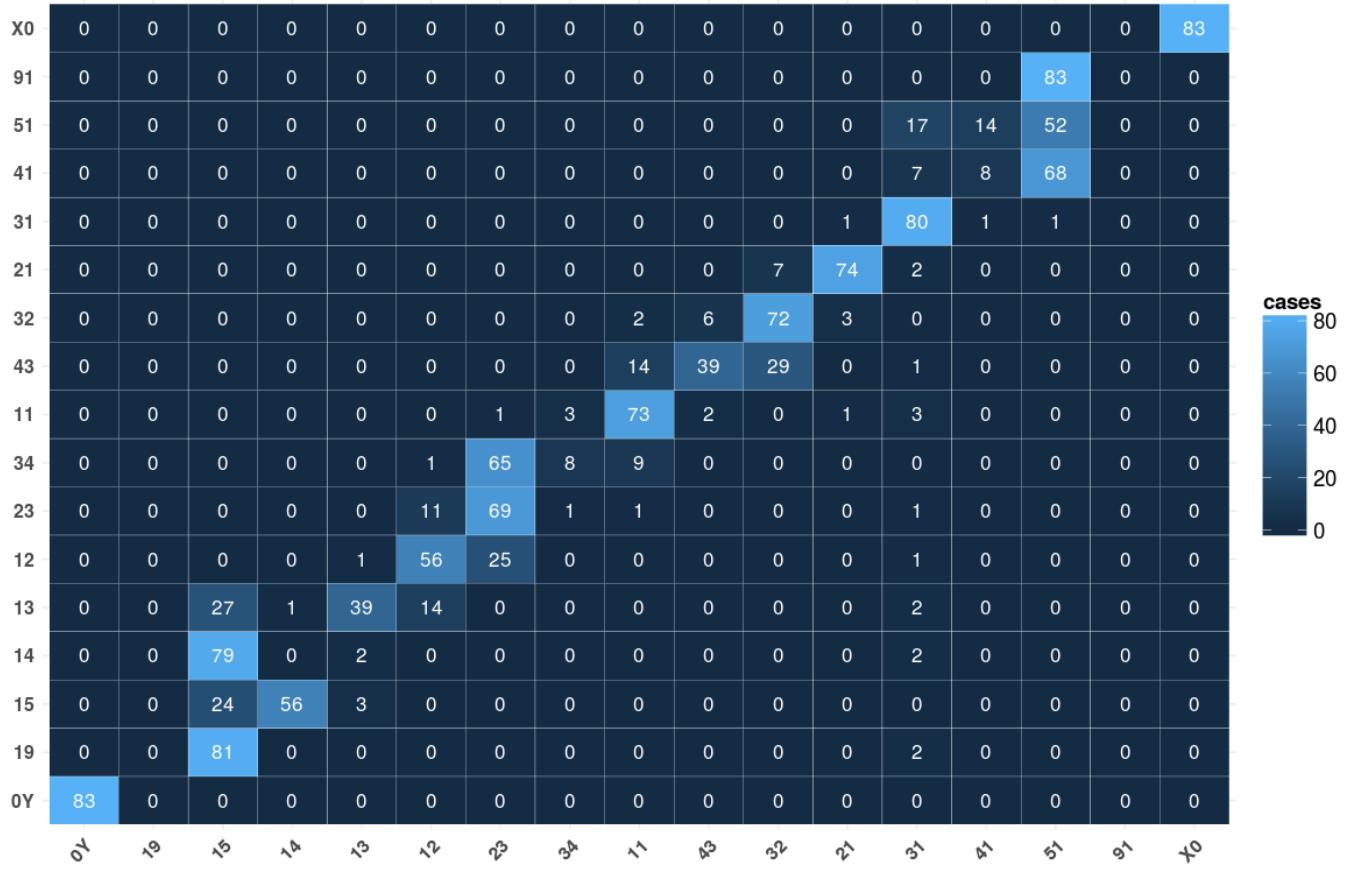
2b. What does the MTL learn?

- **Not about number** of targets! Counting animals hurts performance of previous tasks

2b. What does the MTL learn?

- **Not about number** of targets! Counting animals hurts performance of previous tasks
- Hierarchical ordering of (difficulty of) tasks: **Reversing architecture does not work**

3. Does it generalize to unseen combinations?



Gradable Adjectives

Adjectives: Gradable vs Non-gradable

Adjectives: Gradable vs Non-gradable

Non-gradable

The **open** car

the **green** car

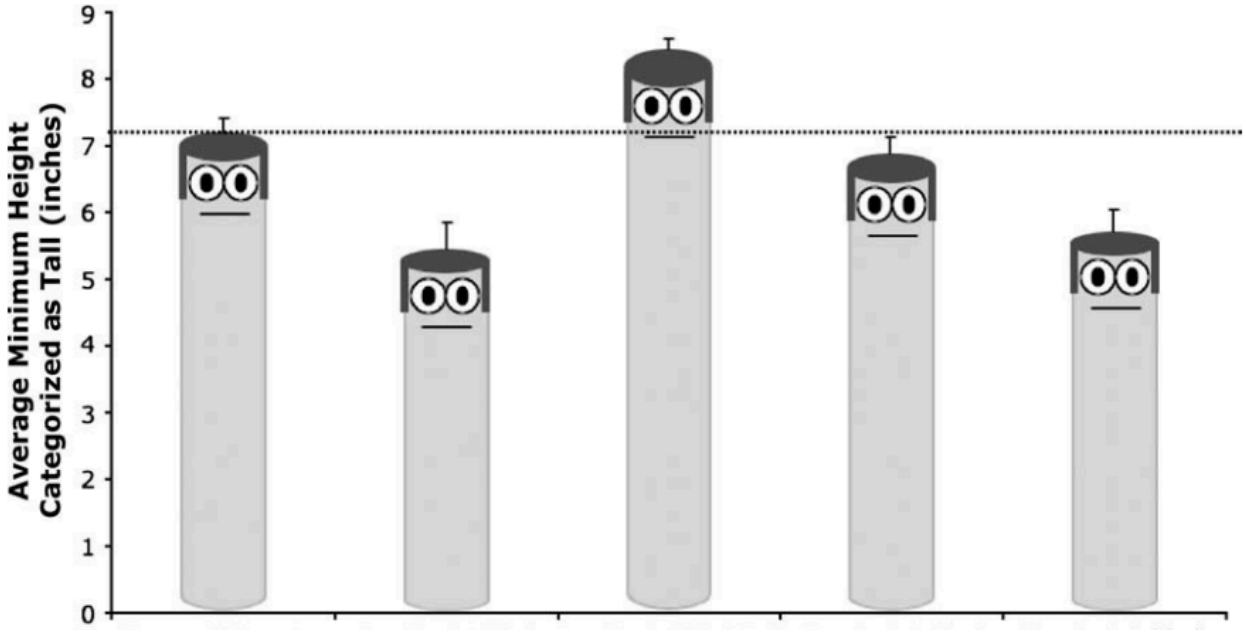
Gradable

The **old/older/oldest** car

The **small/smaller/smallest** car

GAs: A *function* over a set of comparison

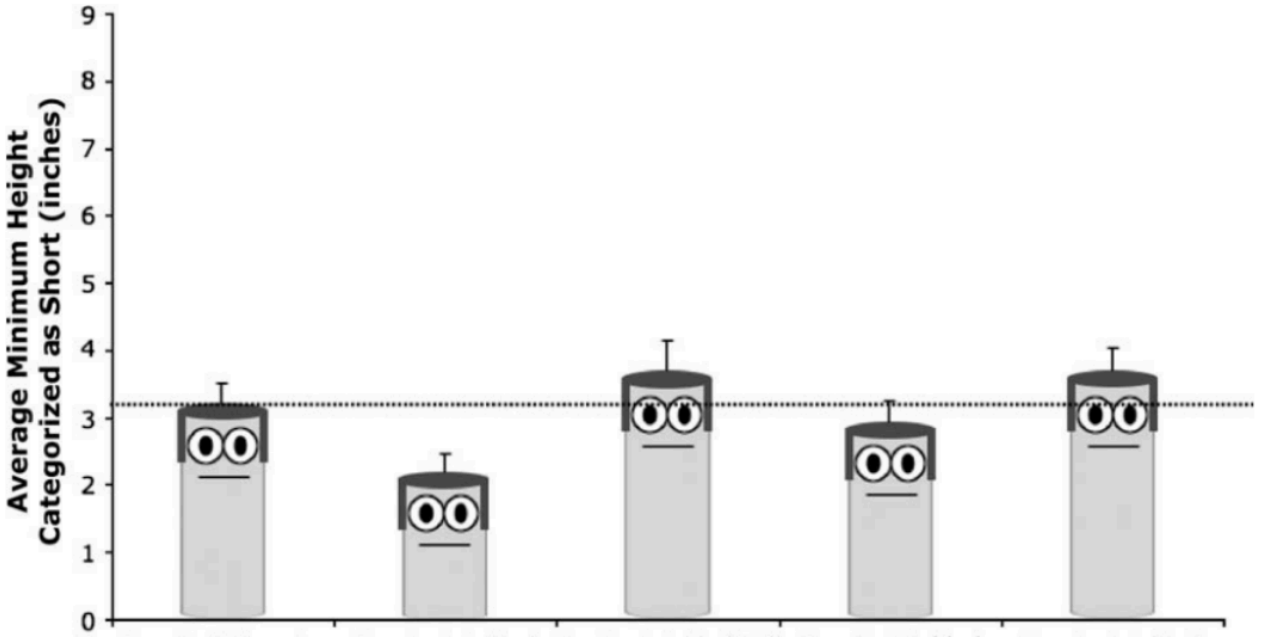
Tall/short use dependent on **height** of comparison set



Barner & Snedeker, 2008, Child Development

GAs: A *function* over a set of comparison

Entity *short* in one **context** can be *tall* in another one



Barner & Snedeker, 2008, Child Development

Computational study (ongoing)

Question

Can a multimodal model learn to use **positive GAs** (i.e. big/small) in a given **visual context**?

Pezzelle, Fernández, *in preparation*, 2019

Motivation

Size GAs:

- represent functions that map objects onto **scales of degrees**

  [Cresswell, 1976, Kennedy, 2013]

Motivation

Size GAs:

- represent functions that map objects onto **scales of degrees**

  [Cresswell, 1976, Kennedy, 2013]

- subtend a statistical **operation** over a comparison set

  [Barner and Snedeker, 2008, Schmidt et al., 2009]

Motivation

Size GAs:

- represent functions that map objects onto **scales of degrees**

  [Cresswell, 1976, Kennedy, 2013]
- subtend a statistical **operation** over a comparison set

  [Barner and Snedeker, 2008, Schmidt et al., 2009]
- learned through their use in *grounded* contexts, i.e. **vision**

  [Schmidt et al., 2009]

1. Do models learn vague GAs?

Battery of tasks to evaluate **model abilities**:

- (A) **Superlative GAs** (biggest/smallest) –> **measurement/sorting**
- (B) Positive GAs in **same-shape scenes** –> **threshold**
- (C) Positive GAs in **different-shape scenes** –> **threshold**
- (D) Positive GAs in **same-shape sets** –> **comparison set**

3. Do models generalize?

Battery of experiments to test **generalization**:

- train on scene-level tasks, test on **comparison set-level** ones
- train on **mixed-task datasets**: Some shapes in train, different shapes in test
- train on mixed-abilities cases, test on **single-ability** ones

Expected results

We expect **models** to:

- **fail** in almost all tasks (baseline)

Expected results

We expect **models** to:

- **fail** in almost all tasks (baseline)
- being able to easily solve the easier tasks, **make errors** in harder ones (SoA models)

Expected results

We expect **models** to:

- **fail** in almost all tasks (baseline)
- being able to easily solve the easier tasks, **make errors** in harder ones (SoA models)
- show **some generalization** abilities (i.e. shortcuts)

Expected results

We expect **models** to:

- **fail** in almost all tasks (baseline)
- being able to easily solve the easier tasks, **make errors** in harder ones (SoA models)
- show **some generalization** abilities (i.e. shortcuts)
- error patterns in line with **difficulty of cases** (i.e. distance from threshold)

Few/many/big/small questions?

References i

Barner, D. and Snedeker, J. (2008).

Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents.

Child development, 79(3):594–608.

Cresswell, M. J. (1976).

The semantics of degree.

In *Montague grammar*, pages 261–292. Elsevier.

Fabbri, S., Caviola, S., Tang, J., Zorzi, M., and Butterworth, B. (2012).

The role of numerosity in processing nonsymbolic proportions.

The Quarterly Journal of Experimental Psychology, 65(12):2435–2446.

Kennedy, C. (2013).

Projecting the adjective: The syntax and semantics of gradability and comparison.

Routledge.

Schmidt, L. A., Goodman, N. D., Barner, D., and Tenenbaum, J. B. (2009).

How tall is tall? Compositionality, statistics, and gradable adjectives.

In *Proceedings of the 31st annual conference of the cognitive science society*, pages 2759–2764. Citeseer.