THE EVOLUTION OF ADJECTIVAL MONOTONICITY

Fausto Carcassi
Marieke Schouwstra
Simon Kirby

Presented by – Devashish Kamble

What are scalar adjectives?

Adjectives that indicate the degree or intensity of a property they modify are called scalar adjectives.

As an example:

- acceptable < good < great < superb
- warm < hot < scorching
- bad < awful < terrible < horrible

Why is monotonicity a general property of scalar adjectives?

A good way to investigate why scalar adjectives are monotonic is by looking at how they evolved

Experimental Design

The authors designed three computational Iterated Learning models that show how the following biases influenced the evolution of monotonicity in scalar adjectives :

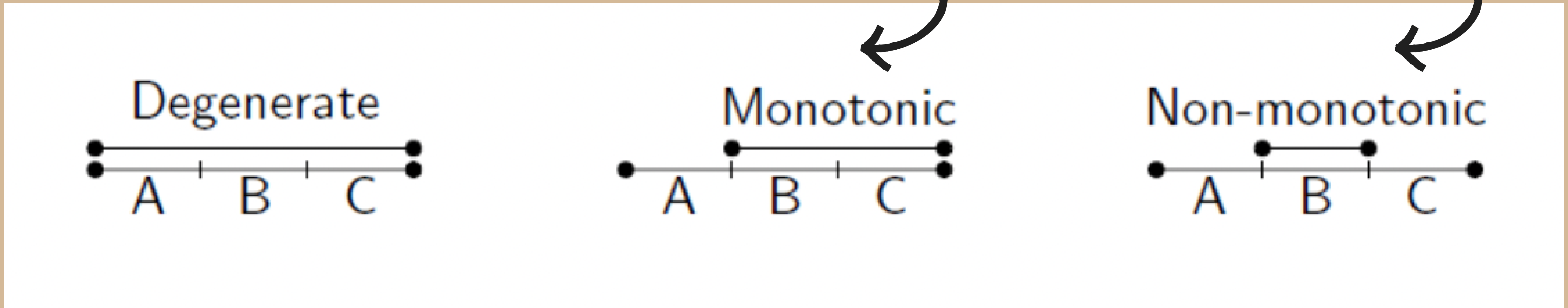
- simplicity
- informativeness
- pragmatic reasoning

Experimental Design

- In all the models, languages encode a set of three degrees on a scale, modelling scalar adjectives.
- These meanings may cover a part of the scale, or the entire scale.
- The complexity of these languages is based on the amount of boundaries that a meaning has.

big

big but not huge



Model 1 : Pressure from learning (Simplicity)

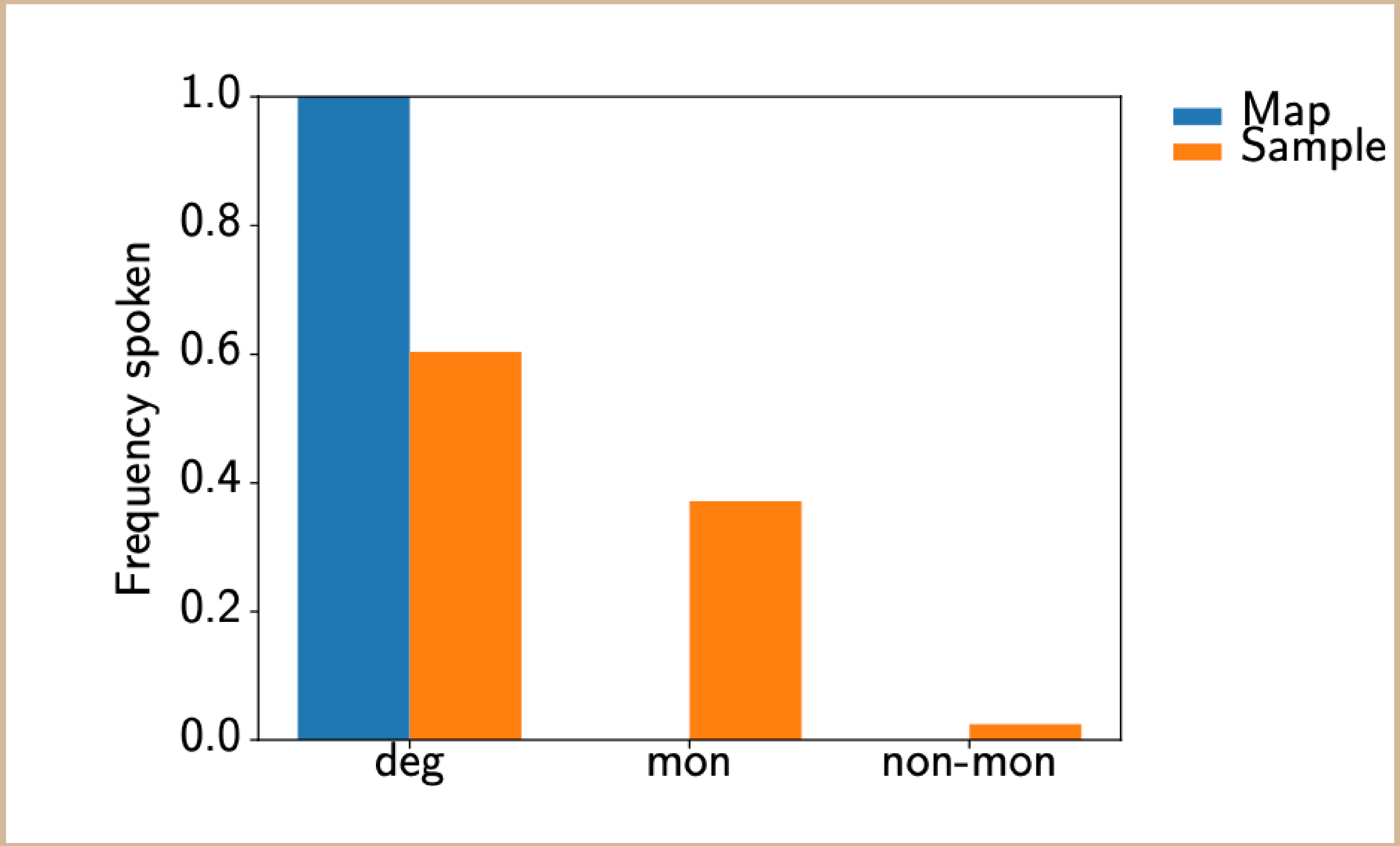
Methods

Maximum a posteriori (MAP) agents

- Each model agent has access to the entire set of possible languages, out of which it chooses one language to use and teach to the next generation.
- The agent chooses the language which has the highest probability based on the language that their teacher used in the previous generation.
- The choice of language is also influenced by the agent's own prior probability distribution over all possible languages, which is based on the language's complexity.
- Agents favour less complex (i.e. simpler) languages, and thus these languages are assigned a greater prior probability

Sampling agents

Results



The authors found that the bias for simplicity alone does not lead to languages encoding monotonic categories, but rather to **degenerate languages.**

These languages are easy to learn, but they are not usable for communication, and thus do not portray natural languages accurately.

Model 2 : Pressure from communication (Informativeness)

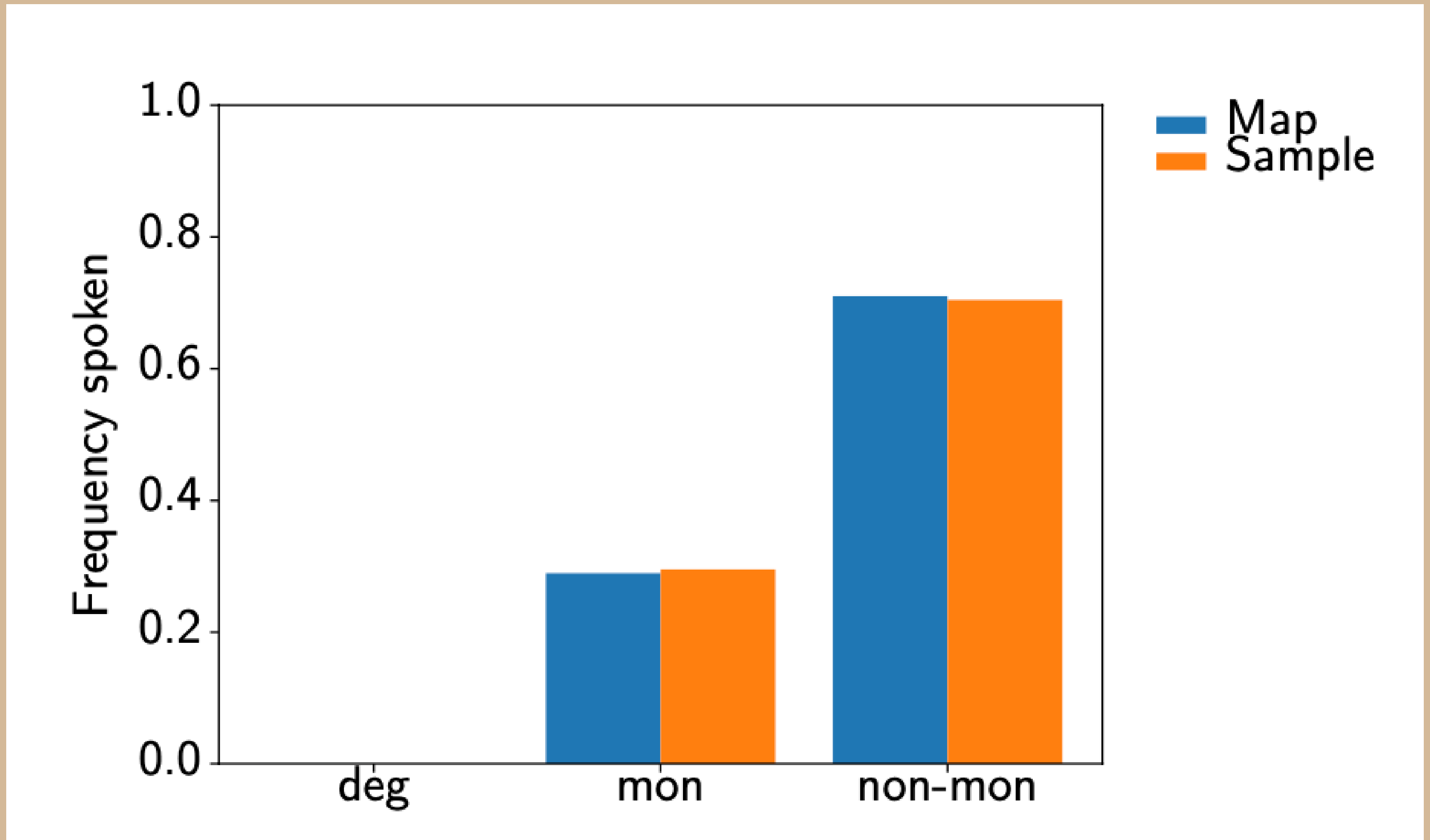
Methods

Agent aS - Speaker

Agent aH - Hearer

$$\text{Communication} = \begin{cases} 1\ ;\ \text{if } dS = dH \\ 0\ ;\ \text{if } dS \mathrel{!=} dH \text{ or } lS \mathrel{!=} lH \end{cases}$$

Results

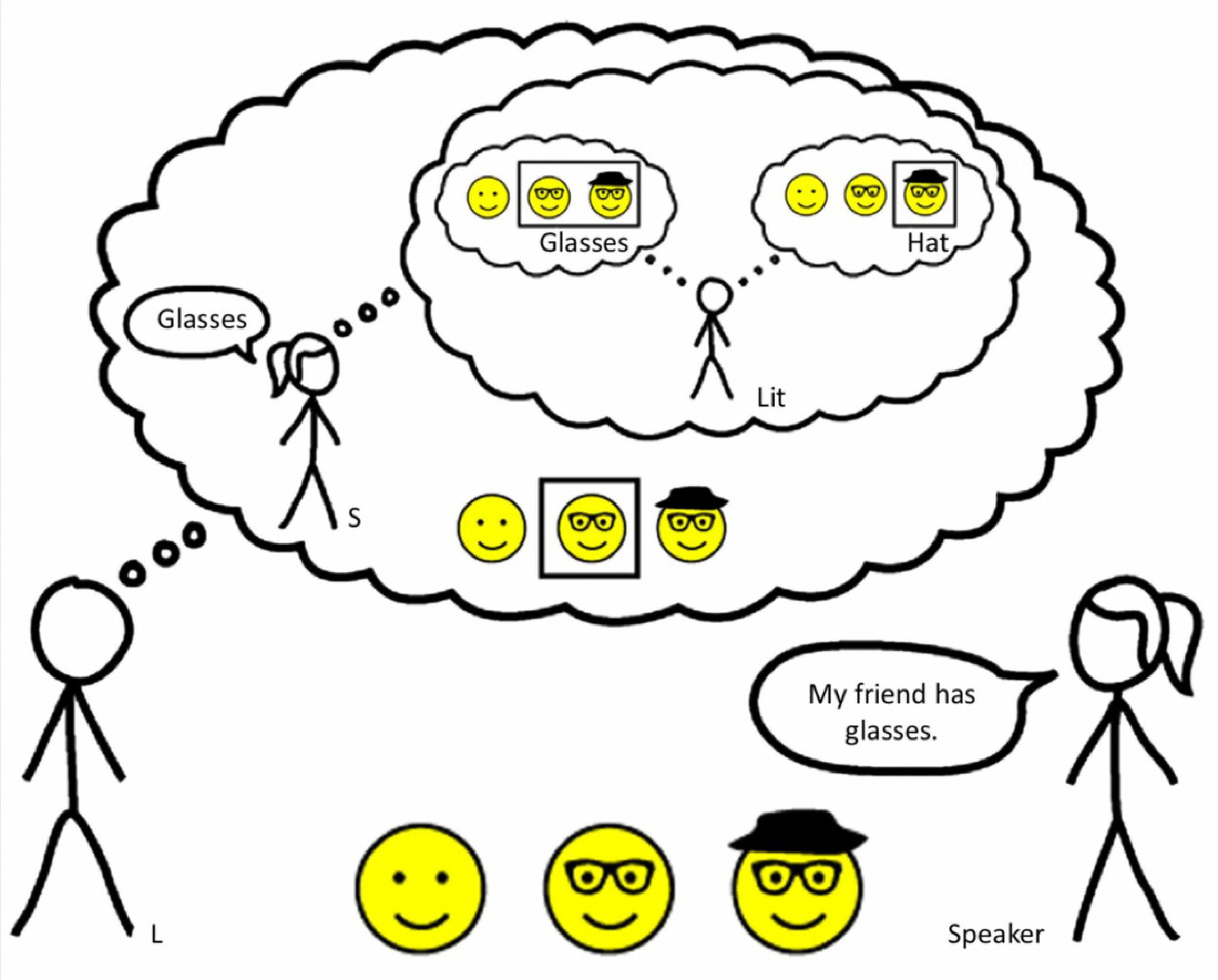


Adding a pressure for communicative accuracy decreases the frequency of monotonic languages as they are communicatively less accurate than **non-monotonic languages.**

Since humans do not use non-monotonic categories, this model did not portray natural languages accurately either..

Model 3 : Pragmatic agents

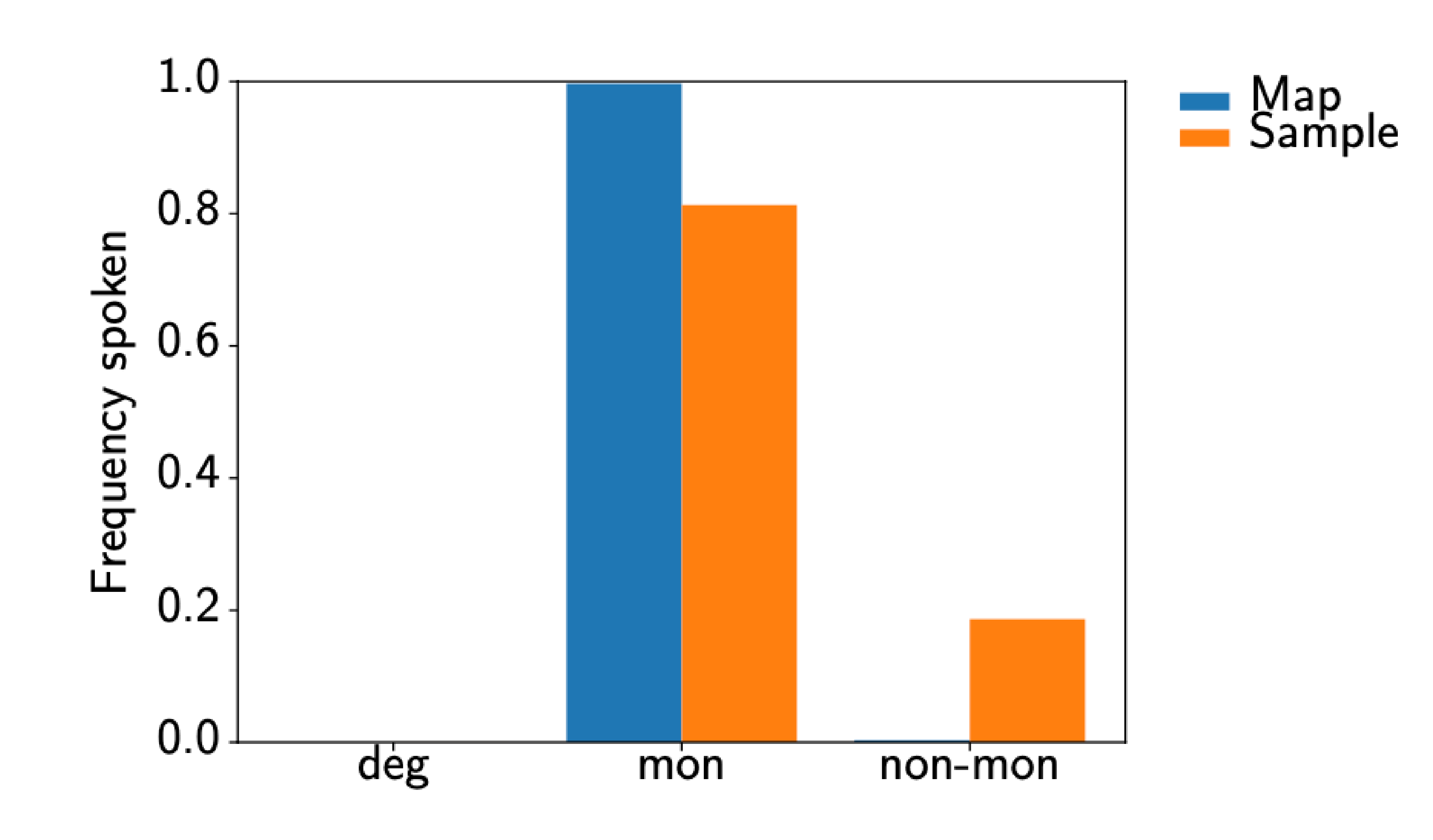
Rational Speech Acts Agents



One with hat and glasses (HG), one with only glasses (G), and one with neither (N)

- The speaker says “my friend has glasses”.
- The listener being rational realises that if the speaker meant to refer to HG, he/she could have said “my friend has a hat” to avoid any ambiguity.
- The fact that he/she did not say so implied that HG was not the referent.
- Thus the listener shares the intuition that the sentence “my friend has glasses” refers to G and not HG or N

Results



Using pragmatic reasoning, the informativeness of a non-monotonic category no longer outweighed the simplicity of a monotonic category.

Agents now favoured languages which encode **monotonic categories**

Conclusion

The models therefore show that monotonicity in scalar adjectives could have emerged due to the combined pressure for simplicity and informativeness, given that humans reason pragmatically.

Glossary

The probability of observing a degree :

$$p(l \mid d, s_p) \propto p(s_p \mid d, l)p(l)$$

The prior probability for each language l is :

$$p(l) \propto 2^{-\gamma \mathscr{L}(l)}$$

Where γ models the strength of the bias for simpler languages and L (l) is the description length of language l.

The probability of a sequence G being produced by a speaker of language l is :

$$p(G \mid l) = \prod_{\langle s_i, d_i \rangle \in G} \frac{1}{|D|} p(s_i \mid d_i, l)$$

The expected communicative success of an agent aH with language lH listening to an agent aS with language lS is :

$$c(a_H, a_S) = \sum_{\langle d_i, s_j \rangle \in D \times S} p(s_j \mid d_i, l_S) p(d_i \mid s_j, l_H)$$

Glossary

The expected communicative success of an agent aH with language lH listening to an agent aS with language lS is :

$$\mathscr{U}_{S_1}(s;d) = \log(p_{L_0}(d \mid s))$$

where PL0 (d | s) is the probability that the literal listener attributes to degree d after having received signal s

The signal to utter with a probability proportional to the utility for the literal listener is :

$$p_{S_1}(s \mid d) \propto \exp(\alpha \mathscr{U}_{S_1}(s;d))$$

where α determines the strength of increase in probability of picking an utterance given an increase in utility.

After receiving a signal s with meaning m from a speaker, the probability of each degree calculated by L1 is:

$$p_{L_1}(d \mid s) \propto p_{S_1}(s \mid d) p_{L_1}(d)$$

where pL1 (d) is the prior probability that the listener attributes to the degree being observed.

Thank You!