Informativity and efficient compression of sentences

Kyle Mahowald, Melissa Kline, Evelina Fedorenko, Edward Gibson (MIT)

Do people drop arguments efficiently?

Languages vary in how often they drop arguments.

Children vary in how often they drop arguments, too.

Here, we manipulate the *informativity* of arguments and see if people preferentially drop less informative arguments.

Informativity of arguments

Using insights from information theory, we define the informativity of an argument as its predictability

We manipualte that two ways:

-manipulating the predictability of the arguments from the verb

-manipulating the number of arguments

Hypothesis

More predictable arguments are more likely to be dropped.

Experiment 1: Manipulating predictability of arguments

- Participants (n=47) on Amazon's Mechanical Turk were asked to give two-word compressions for simple Subject-Verb-Object sentences

agent predictable from verb, patient not

The **policeman arrested** the **shopkeeper**.

patient predictable from verb, agent not

The **engineer shuffled** the **cards**.

neither argument predictable from verb

The **musician punched** the **plumber**.

- Manipulate: active/passive, animate/inanimate patient.
- Prediction: participants will selectively keep words that are unpredictable (more informative).

Experiment 1 Results

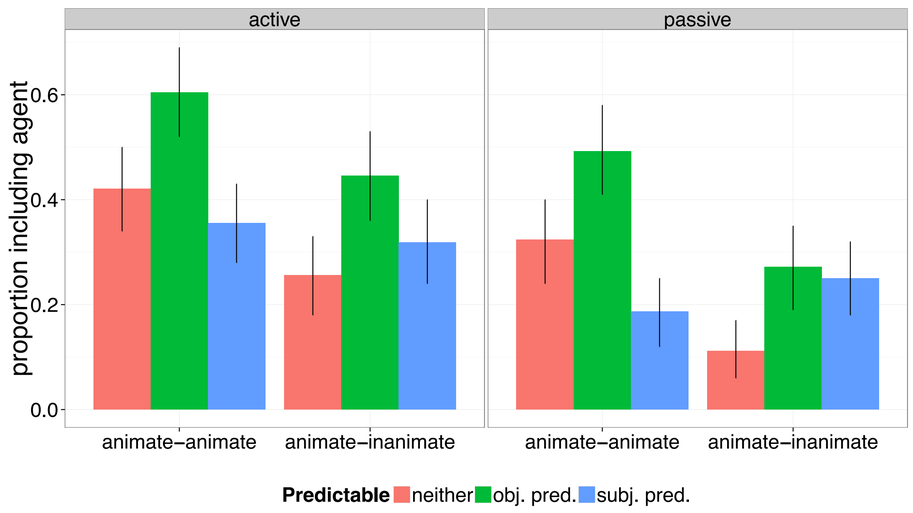


Fig. 1: Proportion of trials in which the agent is included, split up by active/passive and animate/inanimate and predictability. The informative arguments are kept more often. (Green bars higher than blue bars.)

Experiment 2: Manipulating number of arguments

- In this task, participants (n=91) saw a description and image of an action (e.g., "Amy holds the spoon"), and were then shown an array of 7 potential agents (e.g., the two women in Figure 2) and patients (e.g., the 5 objects). The proportion of agents and patients varied, but the total number of objects was constant.
- The prediction is that, as the number of agents in the picture goes up, the agent is more likely to be included. As the number of patients goes up, the agent is more likely to be dropped.

Fig 2: Experiment 2 example image

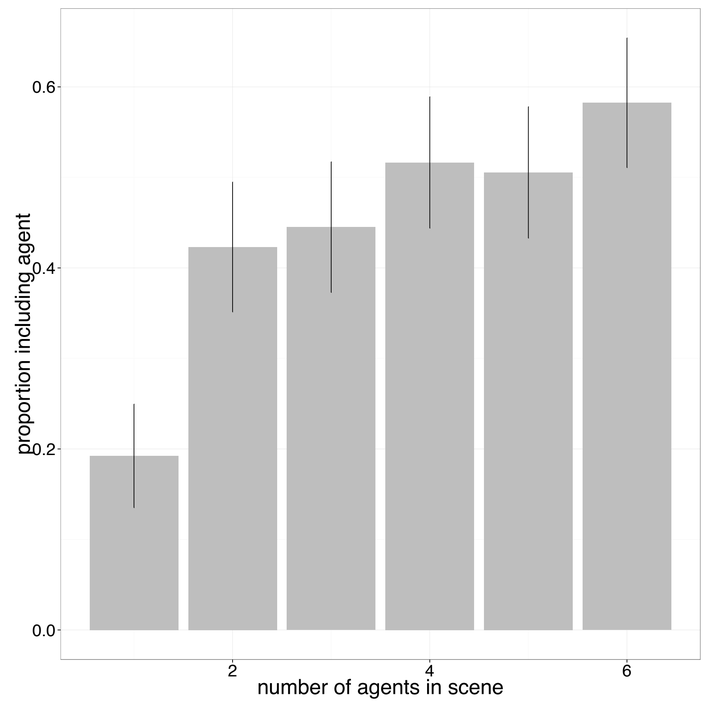
- Assuming a uniform prior over which agent will be referred to, the probability of a given agent being mentioned is

$$P(\frac{1}{n_{agents}})$$

- Therefore, the informativity (surprisal) of an argument is, here, just a function of the number of potential arguments in that role.

Experiment 2 Results

Fig 3: Proportion of trials in which the agent is included as a function of number of agents in the scene. As the number of agents goes up, so too does the proportion of agents included.



Conclusion

- Participants are sensitive to the informativity of arguments in deciding whether to drop them or not in a compression task.
- The effect generalizes to include informativity manipulations affecting the semantic predictability of the argument as well as the number of potential arguments.
- This result is consistent with efficient language usage, but more work is needed to assess how it relates to factors like subject-drop cross-linguistically and in child language learners.

Get the poster here!

Kyle's contact info:
kylemaho@mit.edu

web.mit.edu/kylemaho/www/mahowald_inform_poster_CUNY2015.pdf