Compositionality and Generalization in Emergent Languages

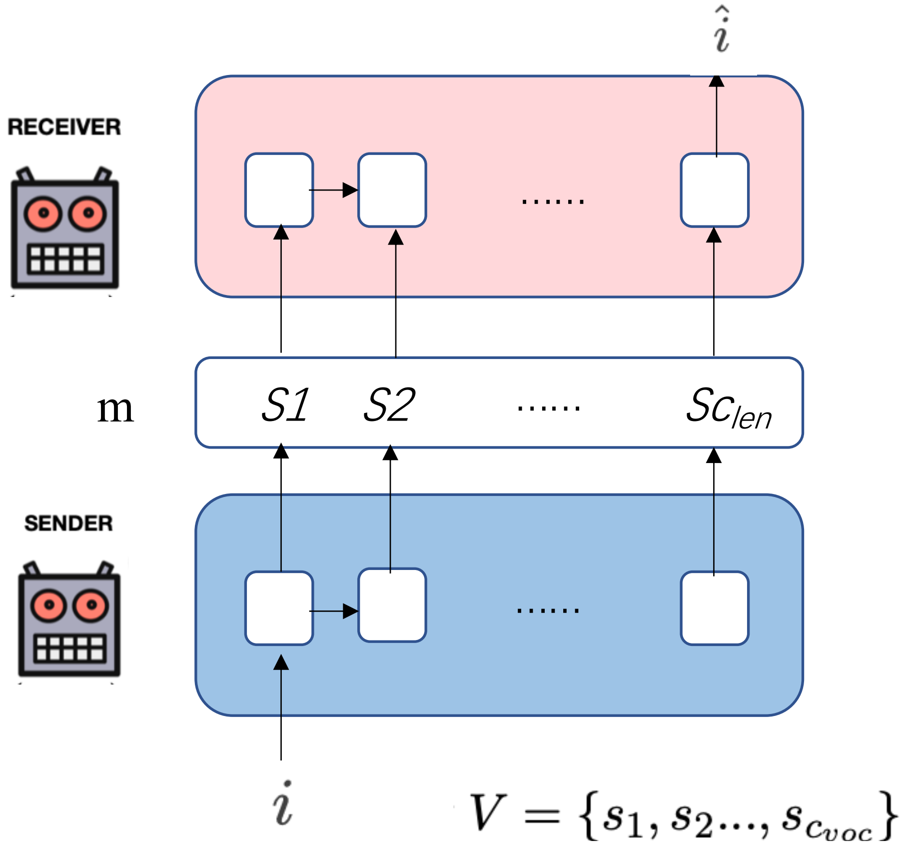
Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, Marco Baroni

Introduction

- Compositionality in human languages
  - Efficient strategy where we need to only encode finite set of words & systematic rules to encode infinite meanings.
  - Enables generalization to novel meanings.
- Compositionality in emergent languages
  - To develop communicating neural network agents, we need to enable them to communicate with compositional languages.
  - Neural network agents are general-purpose learners that we can intervene on to assess what affects the degree of compositionality of their languages. This makes the study of emergent communication important for **language evolution**.

Replication process -- understanding the game

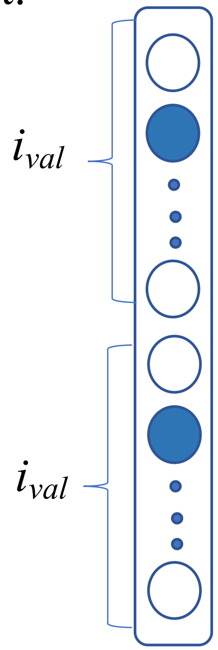
- Sender network receives one input i and chooses a sequence of symbols from its vocabulary V = $\{s1, s2..., s_{cvoc}\}$ of size c_{voc} to construct a message m of fixed length c_{len}.
- Receiver network consumes m and outputs $\hat{i}$.
- Agents are successful if $\hat{i} = i$ (i.e., Receiver reconstructs Sender's input).



Replication process -- understanding the game

- Each input consists of i_{att} attributes, each with i_{val} possible values.
- Messages are of a fixed length c_{len}. At each position j, a symbol s_j is sampled from Sender's vocabulary of size c_{voc}.

Input:



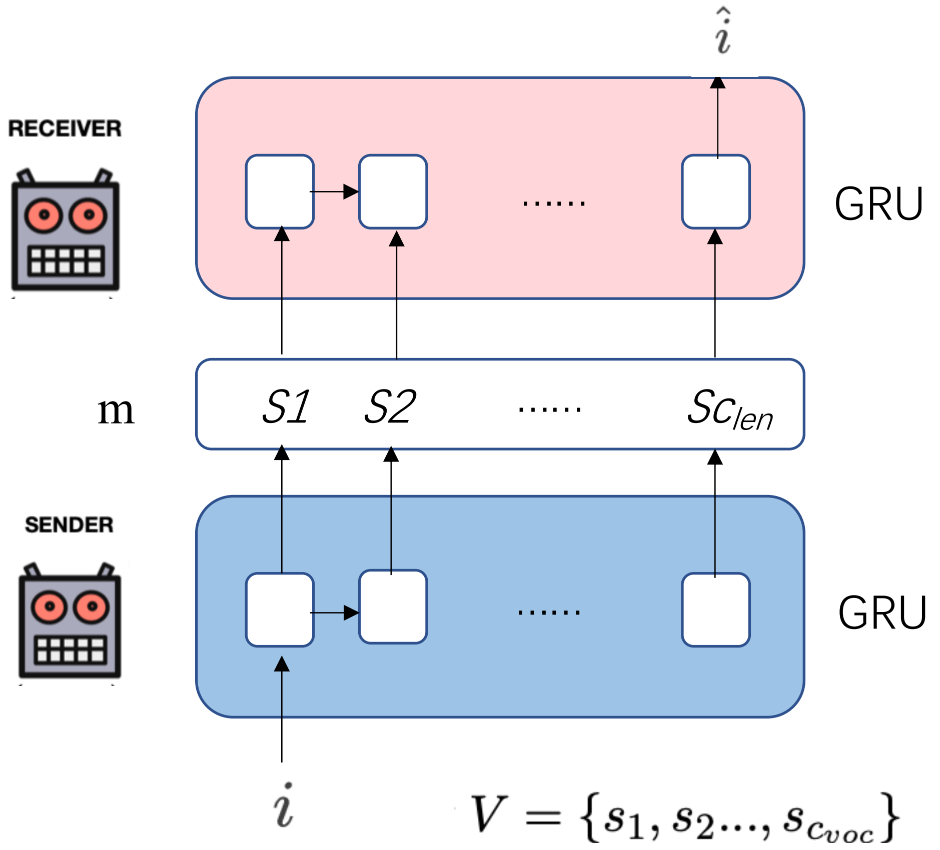
Example of an input i with i_{att}=2. It encodes the second value of the first attribute and the first value of the second attribute.

Message: $S1 \quad S2 \quad \ldots\ldots \quad Sc_{len}$

$$V = \{s_1, s_2 \ldots, s_{c_{voc}}\}$$

Replication process -- understanding agent architecture

- Both agents are implemented as single-layer GRU cells.
  - hidden states of size 500
- Loss: the average cross-entropy between output distributions and Sender's input.



Replication process -- Generalization in emergent languages

- Generalization emerges “naturally” if the input space is large.
  - Input size $|I| = {i_{val}}^{i_{att}}$
  - Channel capacity $|C| = {c_{voc}}^{c_{len}}$
- Sender: $|C| >= |I|$
- Parameters:
  - --n_values=xx --n_attributes=xx --vocab_size=10 --max_len=6 --batch_size=5120 --data_scaler=60 --n_epochs=3000 --sender_hidden=500 --receiver_hidden=500 --sender_entropy_coeff=0.5 --sender_cell="gru" --receiver_cell="gru" --lr=0.001 --receiver_emb=30 --sender_emb=5 --random_seed=9

c_{voc}	5				10				50				100			
(i_{val}, i_{att}) \ c_{len}	2	3	4	{6,8}	2	3	4	{6,8}	2	3	4	{6,8}	2	3	4	{6,8}
(4,4)			X	X		X	X	X	X			X	X			X
(5,2)	X	X	X	X	X	X	X	X	X			X	X			X
(5,3)		X	X	X		X	X	X	X			X	X			X
(5,4)			X	X		X	X	X	X			X	X			X
(10,2)		-	X	X	X	X	X	X	X			X	X			X
(10,3)				X		-	X	X	X	X	X	X	X			X
(10,4)				{-, X}			-	X		X	X	X	-	X	X	X
(16,2)			-	X		X	X	X	X			X	X			X
(25,2)			-	X		-	X	X	X			X	X			X
(50,2)				X			-	X	-	X	X	X	X	X	X	X
(100,2)				{-, X}			-	X		X	X	X	-	X	X	X

Table 3: Grid search. ‘X’ indicates tested settings with at least one successful run. ‘-’ indicates tested settings without any successful run. Finally, blank cells correspond to settings that were not explored for the reasons indicated in the text.

Replication process -- Generalization in emergent languages

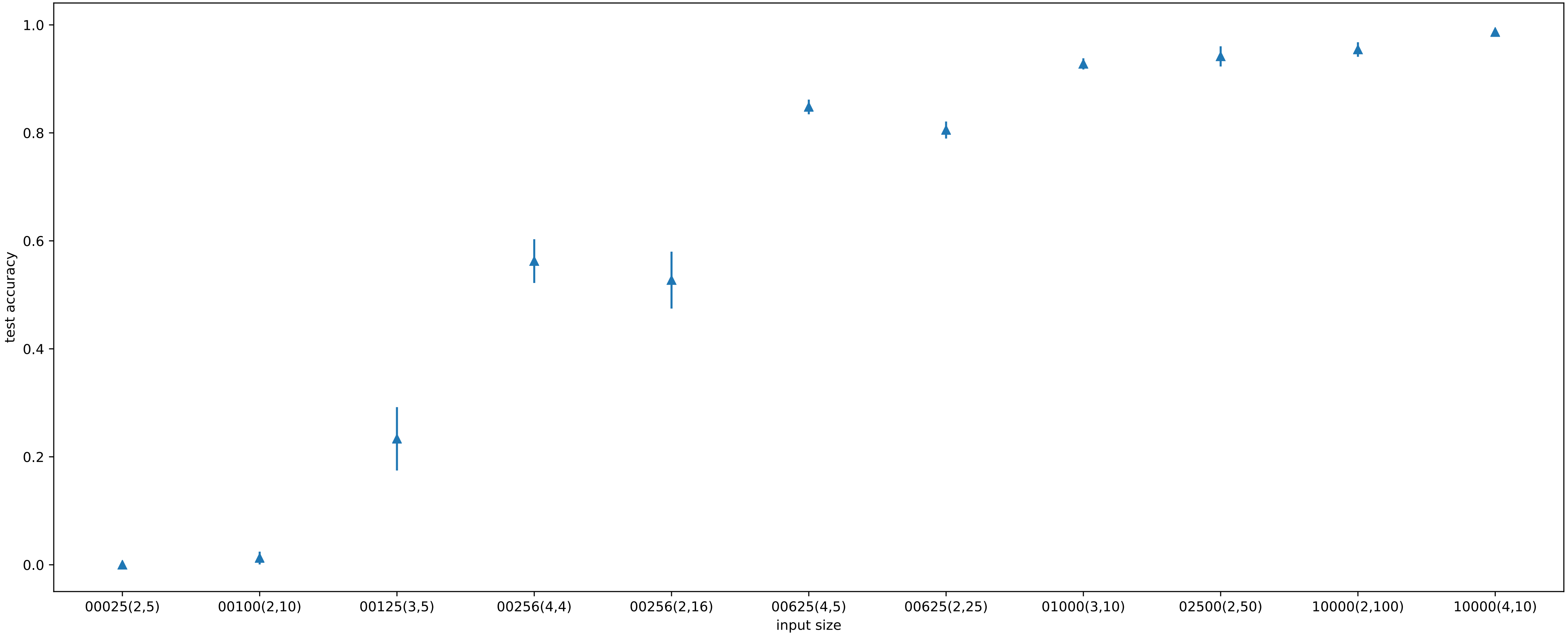
See replication codes & files: https://github.com/yuqing0304/compositionality_replicate

Test accuracy:

```
{'00025(2,5)': {'acc_mean': 0.0, 'acc_sem': 0.0},
'00100(2,10)': {'acc_mean': 0.0125, 'acc_sem': 0.011858541225631423},
'00125(3,5)': {'acc_mean': 0.23333334028720856, 'acc_sem': 0.05868939128794347},
'00256(2,16)': {'acc_mean': 0.527272741496563, 'acc_sem': 0.05269591837361792},
'00256(4,4)': {'acc_mean': 0.5625, 'acc_sem': 0.0405046293650491 3},
'00625(2,25)': {'acc_mean': 0.8052631676197052, 'acc_sem': 0.015779727972030193},
'00625(4,5)': {'acc_mean': 0.8479999780654908, 'acc_sem': 0.013623509400751192},
'01000(3,10)': {'acc_mean': 0.92777778506279, 'acc_sem': 0.010356306263776276},
'02500(2,50)': {'acc_mean': 0.941666716337204, 'acc_sem': 0.018680427008010262},
'10000(2,100)': {'acc_mean': 0.9543877124786377, 'acc_sem': 0.013494911879231615},
'10000(4,10)': {'acc_mean': 0.9865853190422058, 'acc_sem': 0.004432800984419676}}
```

acc_mean: mean accuracy (10 different initializations per setting).
acc_sem: standard error of the mean.

Replication process -- Generalization in emergent languages



Emergent languages are able to almost perfectly generalize to unseen combinations as long as input size $|I|$ is sufficiently large.

My reflection:

- The code for the EGG framework is heavily encapsulated, so it takes time for me to understand its functioning.
- The hyperparameters used in the paper are needed in order to replicate the same results. It may take a long time to try different hyperparameters (I have tried replicating the experiment mentioned in *Anti-efficient encoding in emergent communication* first. But I haven't replicated the results yet after trying many parameters.)
- I think these experiments are innovative and insightful and I love to play these games, though it takes time for me to be skilled.