Is compositionality overrated?
The view from language emergence

Marco Baroni

CBMM Virtual Seminar
June 2020

Facebook AI Research

The FAIR Paris/Barcelona EViL team

Emmanuel Dupoux

Rahma Chaabouni

Diane Bouchacourt

Roberto Dessì

Eugene Kharitonov

Outline

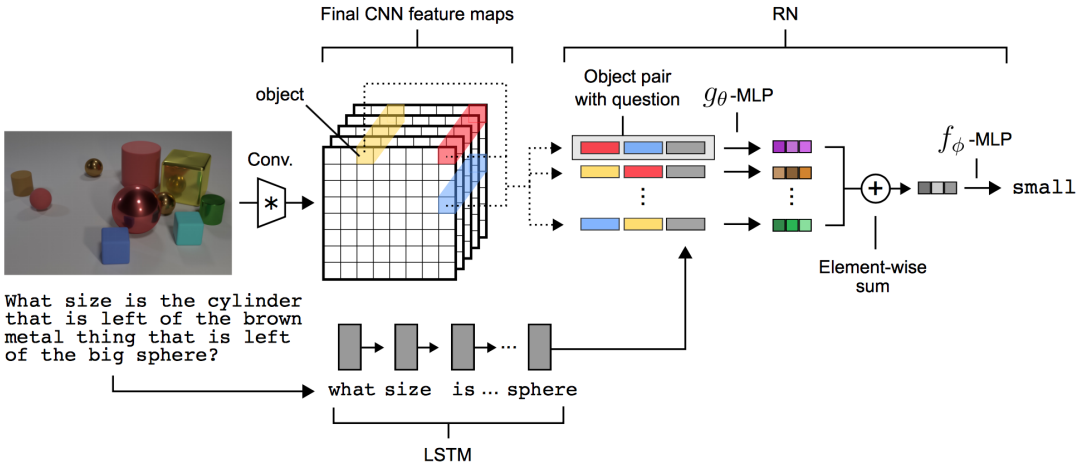
- Emergent language in deep networks: why, how, what?
- Compositionality and generalization in emergent languages
- Take-home messages

Deep networks do amazing things... but just one thing at a time!

https://cs.stanford.edu/people/karpathy/cnnembed/,
https://www.inverse.com/article/12664-google-s-alphago-supercomputer-wins-second-go-match-vs-lee-sedol

How can we harness their powers more flexibly?

- Manual gluing?

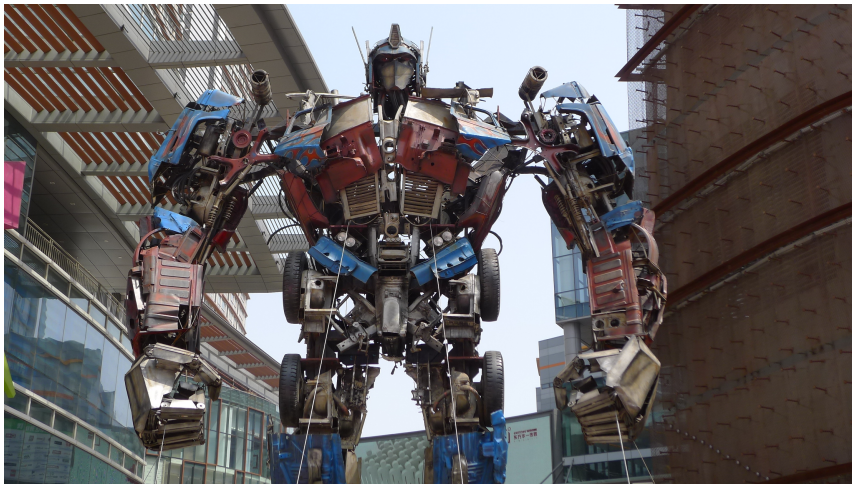



```
$ egrep '^a' in.txt | sort | uniq -c > out.txt
```

https://hackernoon.com/deepmind-relational-networks-demystified-b593e408b643

How can we harness their powers more flexibly?

- Manual gluing?
- Good Old AGI?



Wikipedia

How can we harness their powers more flexibly?

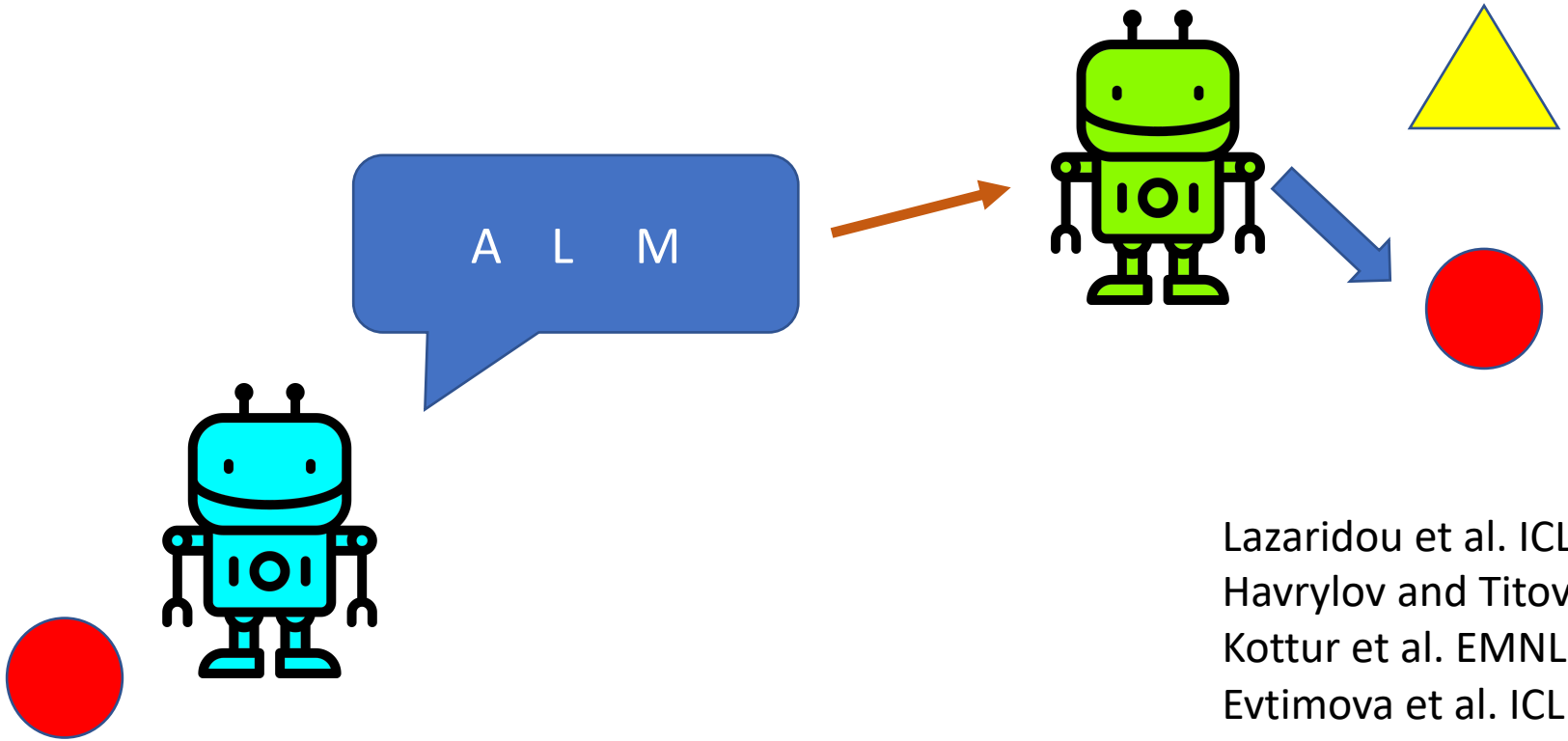
- Manual gluing?
- Good Old AGI?
- Language!

Our bet!

https://www.pxfuel.com/en/free-photo-iypmu

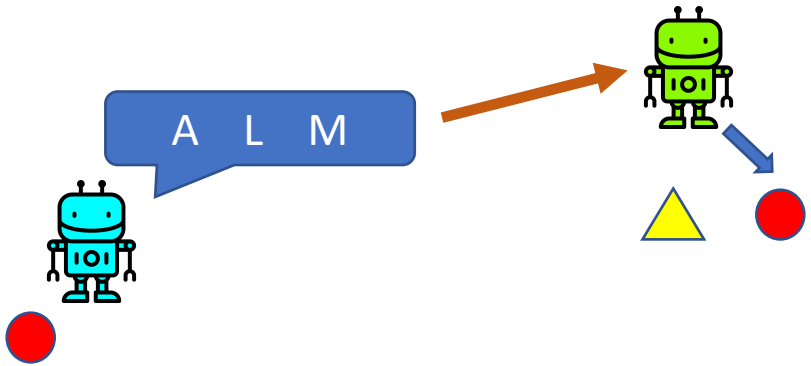
Communication games

http://www.publicdomainfiles.com/



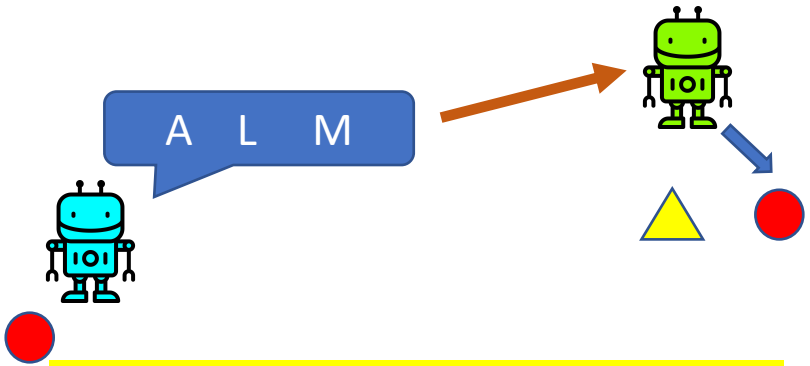
Lazaridou et al. ICLR 2017,
Havrylov and Titov NIPS 2017,
Kottur et al. EMNLP 2017,
Evtimova et al. ICLR 2018,
…

Communication games



- Two networks must jointly solve a task
- *Sender* network sees some input (e.g., a target image) and sends a *message* to *Receiver* network
- Receiver gets some input, including Sender message, and performs an action (e.g., point to target image) to complete the task
- The message is a single ***discrete*** symbol or a sequence of ***discrete*** symbols from a fixed alphabet
- Networks rewarded for task success only, ***no*** supervision on the messages generated by Sender

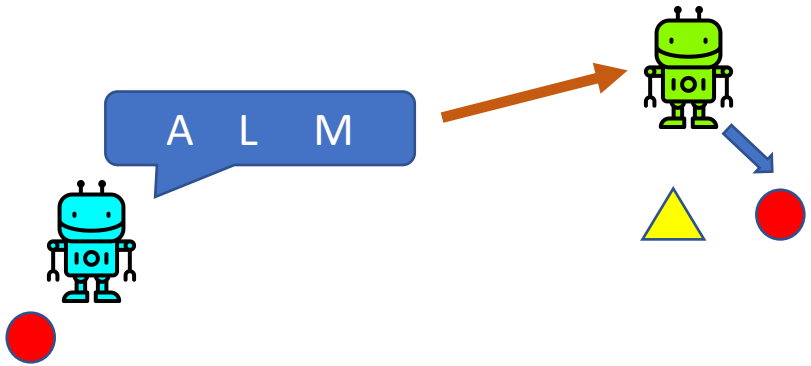
Communication games



this is the only way in which we constrain the emergent system to be human-language-like

- Two networks must jointly solve a task
- *Sender* network sees some input (e.g., a target image) and sends a *message* to *Receiver* network
- Receiver sees some input, including Sender message, and performs an action (e.g., point to target image) to complete the task
- The message is a single ***discrete*** symbol or a sequence of ***discrete*** symbols from a fixed alphabet
- Networks rewarded for task success only, ***no*** supervision on the messages generated by Sender

Communication games



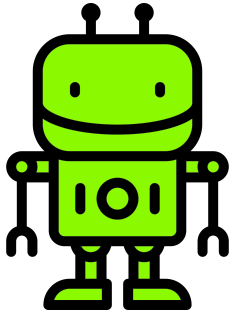
... because... where could we get the training data from?

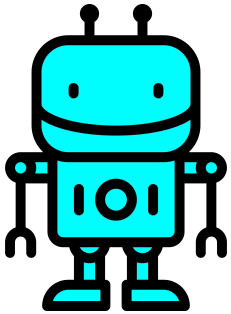
- Two networks must jointly solve a task
- *Sender* network sees some input (e.g., a target image) and sends a *message* to *Receiver* network
- Receiver sees some input, including Sender message, and performs an action (e.g., point to target image) to complete the task
- The message is a single ***discrete*** symbol or a sequence of ***discrete*** symbols from a fixed alphabet
- Networks rewarded for task success only, ***no*** supervision on the messages generated by Sender

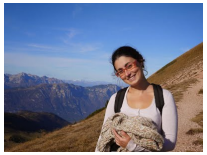
The emergence of words

Pexels

127



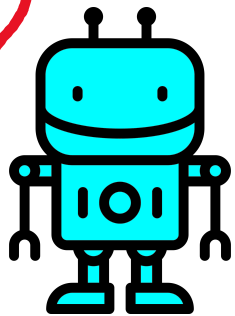




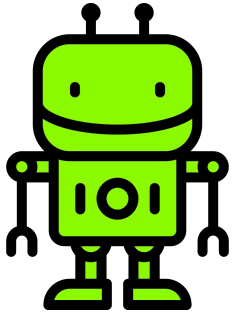
Lazaridou et al. ICLR 2017

The emergence of words

Pexels, Pixabay

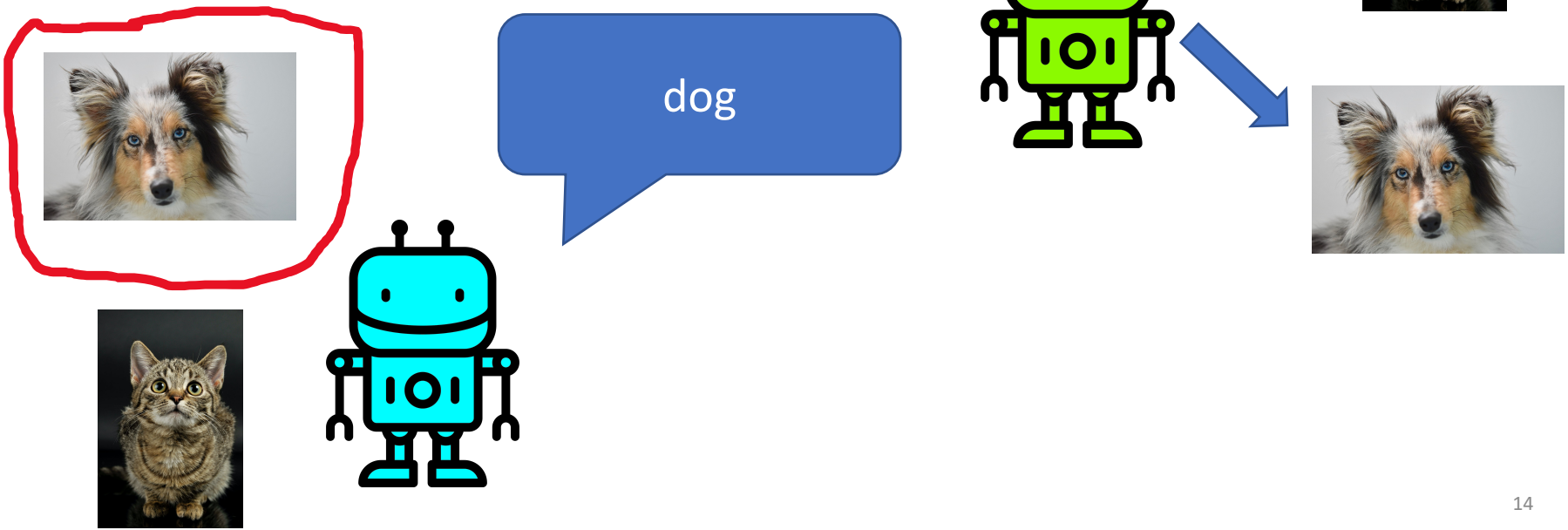


127



The emergence of words

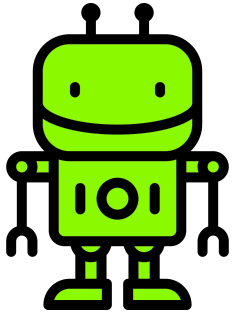
Pexels, Pixabay

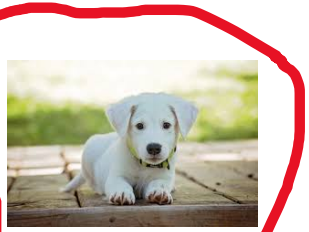


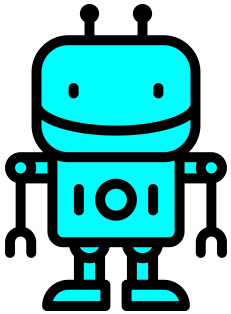
Emergent languages are tricky!

At training time...

dog



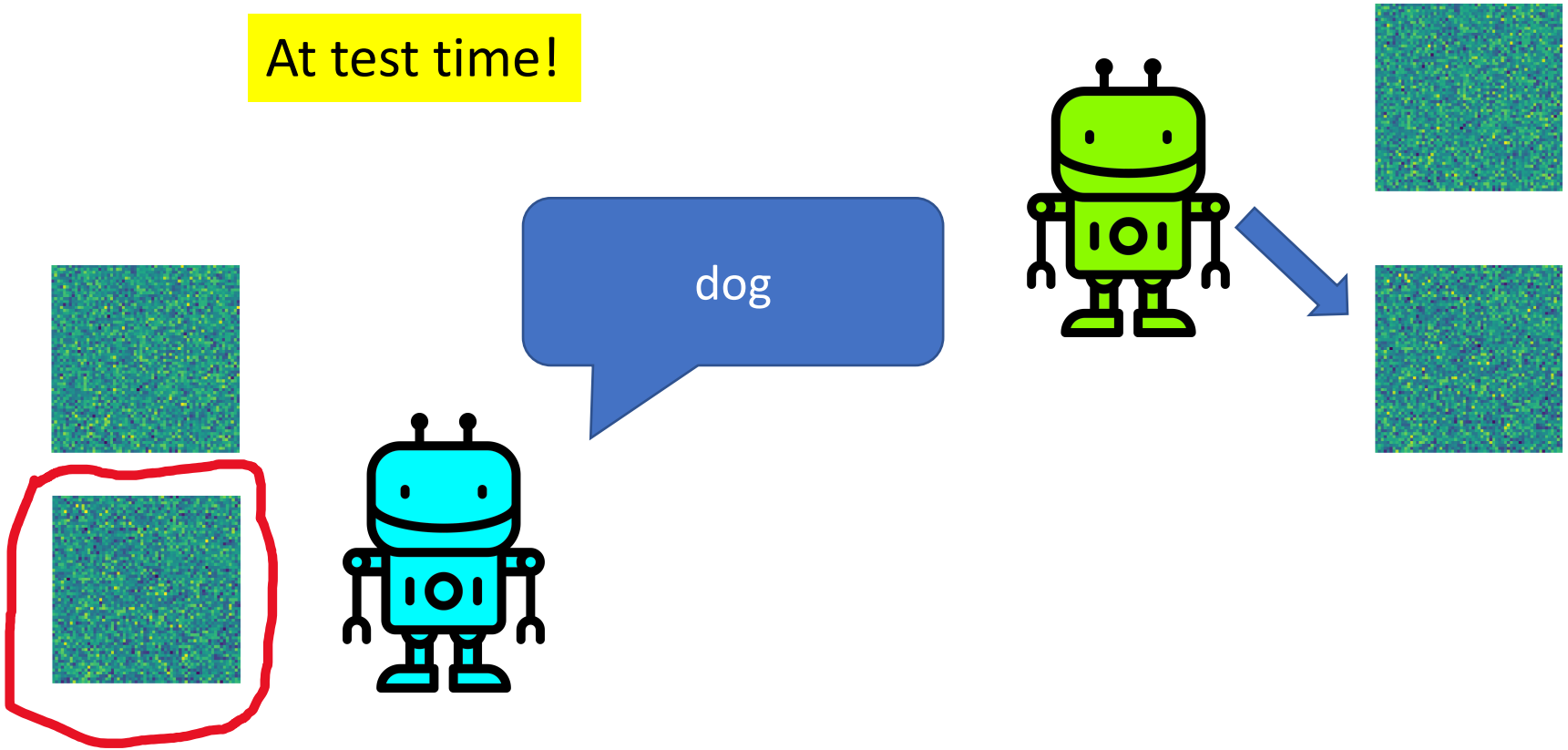




Bouchacourt and Baroni EMNLP 2018

Emergent languages are tricky!

At test time!

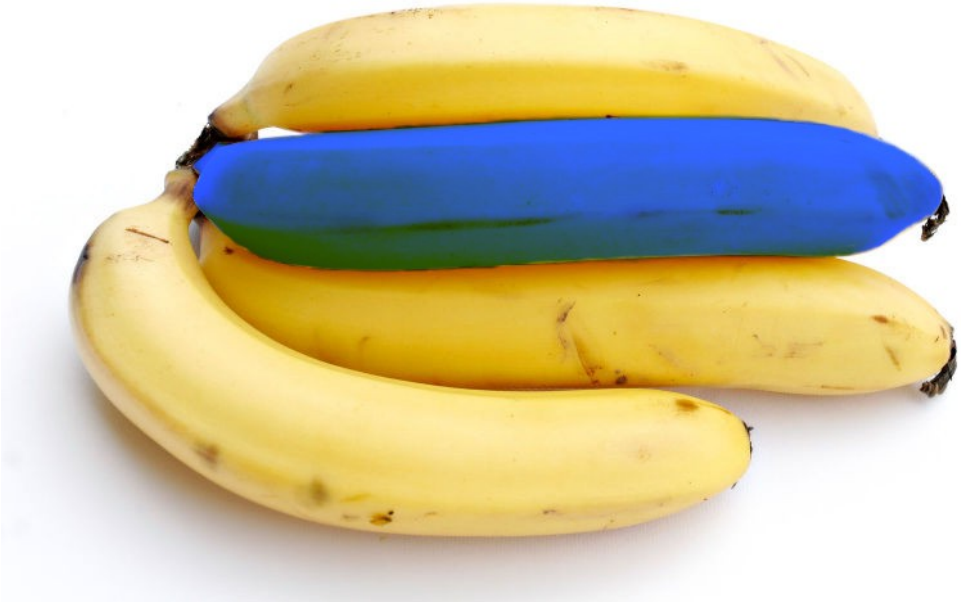


Outline

- Emergent language in deep networks: why, how, what?
- **Compositionality and generalization in emergent languages**
- Take-home messages

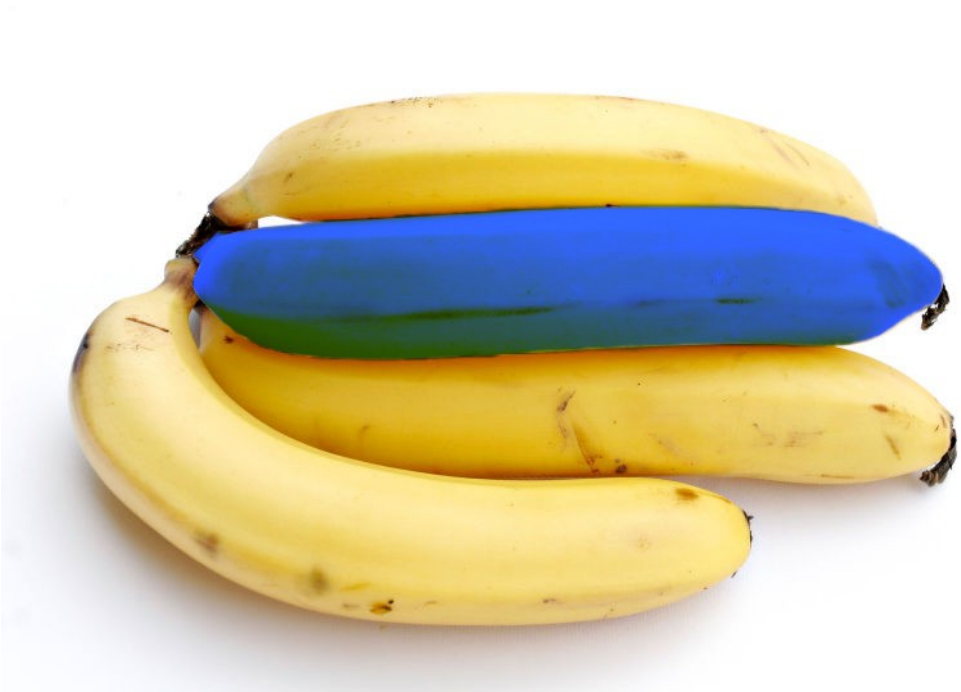
Are emergent languages compositional?

Andreas ICLR 2019, Choi et al ICLR 2018, Havrylov & Titov NIPS 2017, Kottur et al EMNLP 2017, Mordatch & Abbeel AAAI 2018, Resnick et al AAMAS 2020...



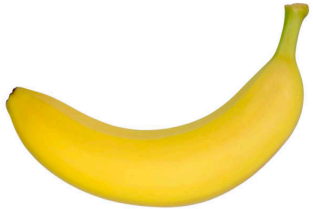
https://commons.wikimedia.org/wiki/File:BlueBanana.jpg

Generalization

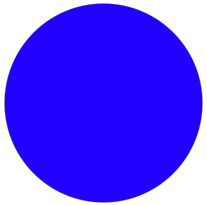


Compositionality and generalization

Train



”banana”



”blue”

”blue banana”

Test

Slide credit: Rahma Chaabouni

What makes an (emergent) language compositional?[1]

[1] Where “compositional” is *not just a synonym* for: “able to generalize”

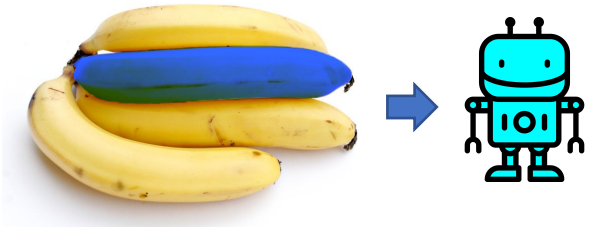
What makes an (emergent) language compositional?

The meaning of a linguistic expression is a function of the meaning of its parts and the rules used to combine them (Boole, Frege, Montague, etc.)

What makes an (emergent) language compositional?

The meaning of a linguistic expression is a function of the meaning of its parts and the rules used to combine them (Boole, Frege, Montague, etc.)

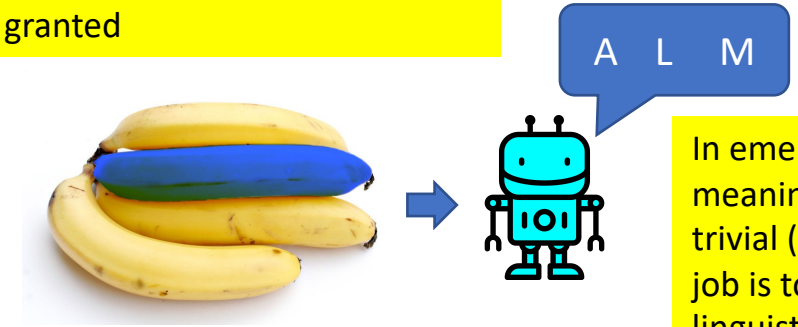
This focuses entirely on the meaning side; the “parts” of the linguistic expression and their combination are taken for granted



What makes an (emergent) language compositional?

The meaning of a linguistic expression is a function of the meaning of its parts and the rules used to combine them (Boole, Frege, Montague, etc.)

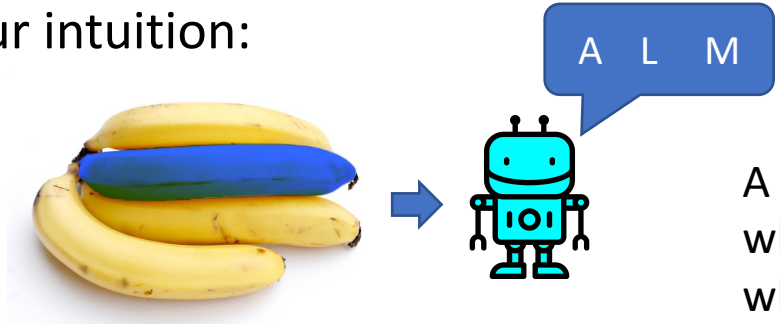
This focuses entirely on the meaning side; the “parts” of the linguistic expression and their combination are taken for granted



In emergent language simulations, meaning composition is typically trivial (blue+banana), and the hard job is to discover the “parts” of the linguistic expression and how they combine

What makes an (emergent) language compositional?

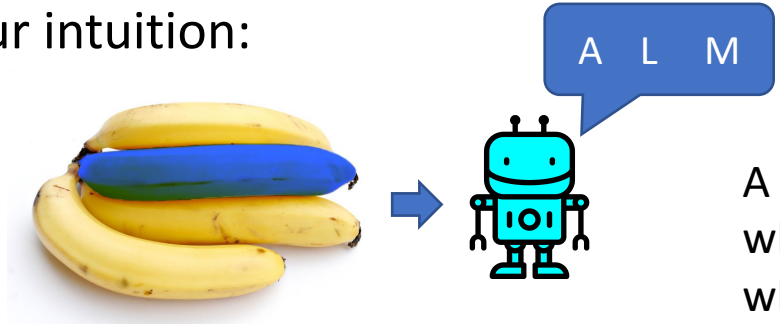
Our intuition:



A compositional language is one where it is easy to read out which parts of a linguistic expression refer to which components of the input

What makes an (emergent) language compositional?

Our intuition:



A compositional language is one where it is easy to read out which parts of a linguistic expression refer to which components of the input

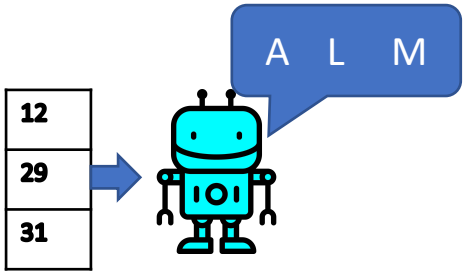
Compositional:
AL -> `blue'
M -> `banana'

Less compositional:
AL -> `blue'
LM -> `banana'

Not compositional:
ALM -> `blue banana'

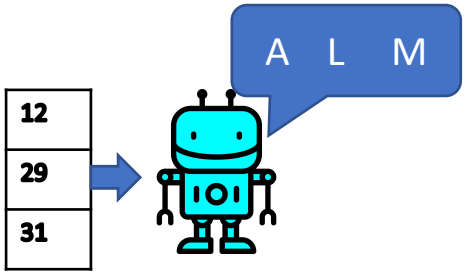
Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:



Naïve compositionality

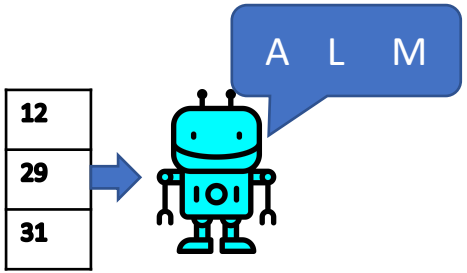
Applies when the only way to combine primitive input elements is to assemble them in a collection:



Examples: a list of attribute-value pairs (equivalently, a vector of values), a set of objects, a list of properties (‘blue’, ‘banana’), ...

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:

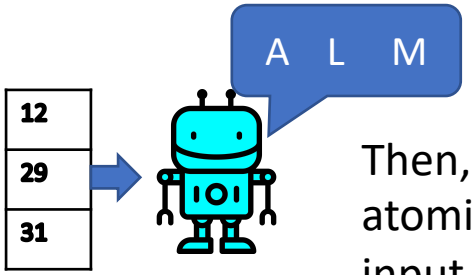


Examples: a list of attribute-value pairs (equivalently, a vector of values), a set of objects, a list of properties (‘blue’, ‘banana’), ...

90% of current emergent language simulations

Naïve compositionality

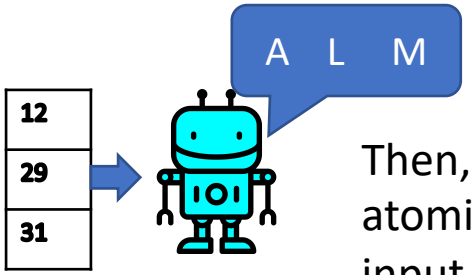
Applies when the only way to combine primitive input elements is to assemble them in a collection:



Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:

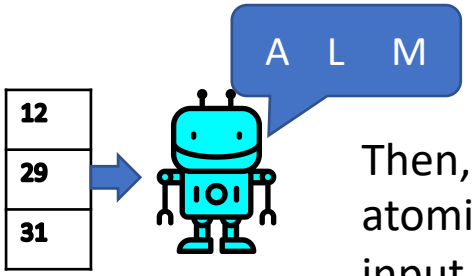


Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

Such language is compositional because a collection of inputs will have to be expressed by the juxtaposition of the corresponding atomic symbols!

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:

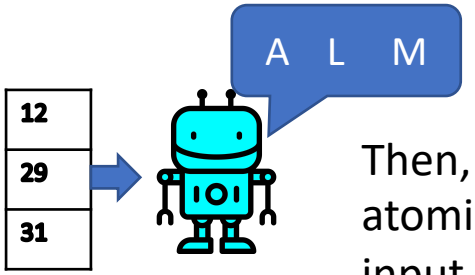


Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

”Naïve” because it only considers ensembling as meaning composition function and atomic-symbol juxtaposition as form composition

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:



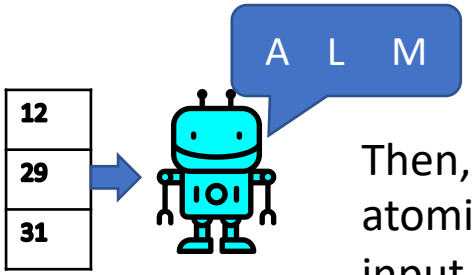
Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

”Naïve” because it only considers ensembling as meaning composition function and atomic-symbol juxtaposition as form composition

This makes for an easy read-out of what is being composed!

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:

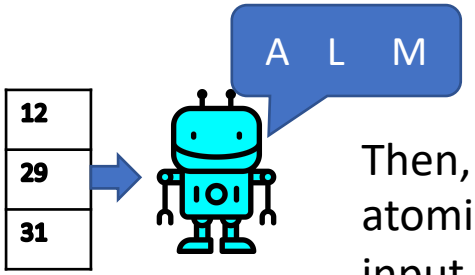


Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

Naïvely compositional:
A -> 29
L -> 12
M -> 31

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:



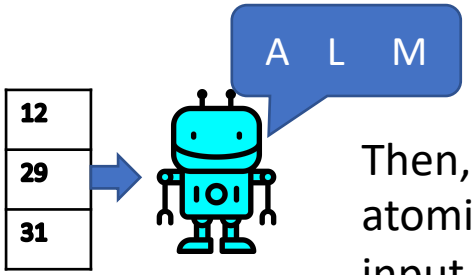
Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

Naïvely compositional:
A -> 29
L -> 12
M -> 31

bag-of-symbols or positional

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:



Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

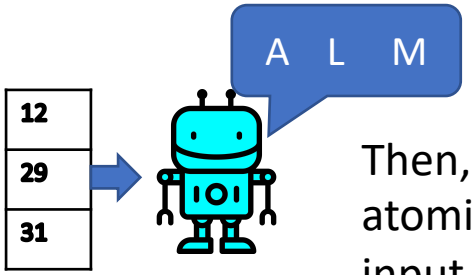
Naïvely compositional:
A -> 29
L -> 12
M -> 31

Non-compositional:
ALM -> [12,29,31]

bag-of-symbols or positional

Naïve compositionality

Applies when the only way to combine primitive input elements is to assemble them in a collection:



Then, a language is **naïvely compositional** if the atomic symbols in its expressions refer to single input elements, *independently of either input or linguistic context*

Naïvely compositional:
A -> 29
L -> 12
M -> 31

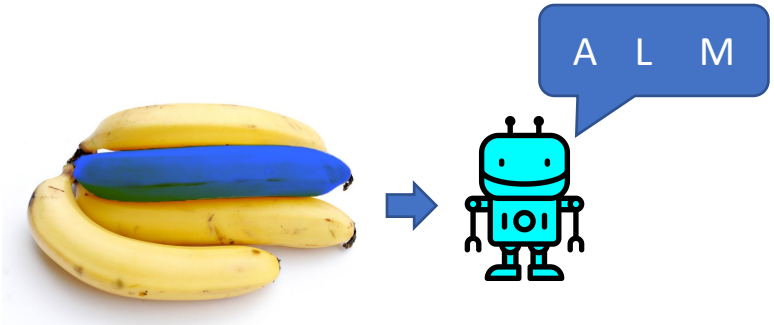
Non-compositional:
ALM -> [12,29,31]

Non-naïvely compositional:
A -> 29 if immediately followed by L
L -> 12 if other input values are odd
M -> 31 if one of previous symbols is A

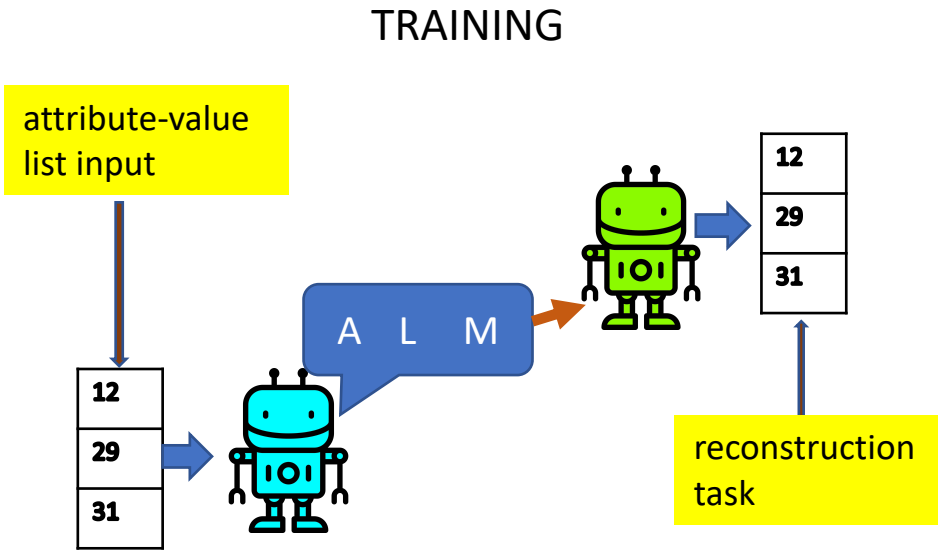
bag-of-symbols or positional

Compositionality and generalization in emergent languages

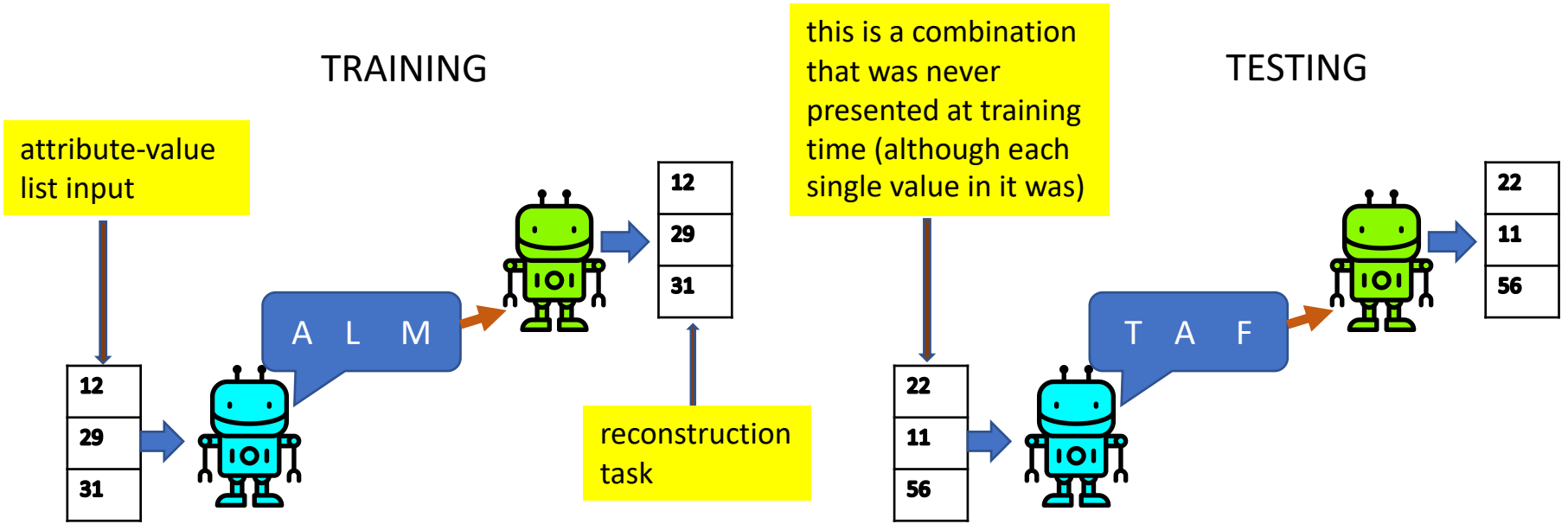
Chaabouni, Kharitonov et al. ACL 2020



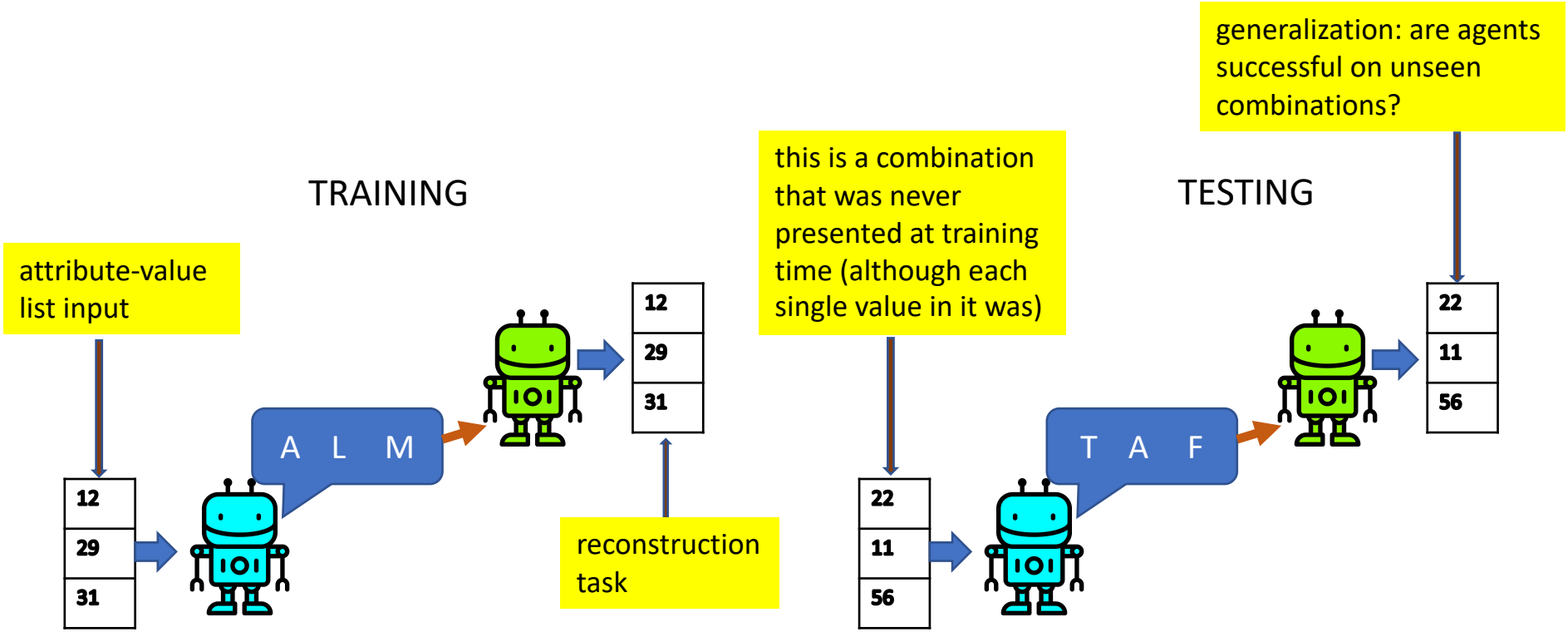
Compositionality and generalization in a communication game



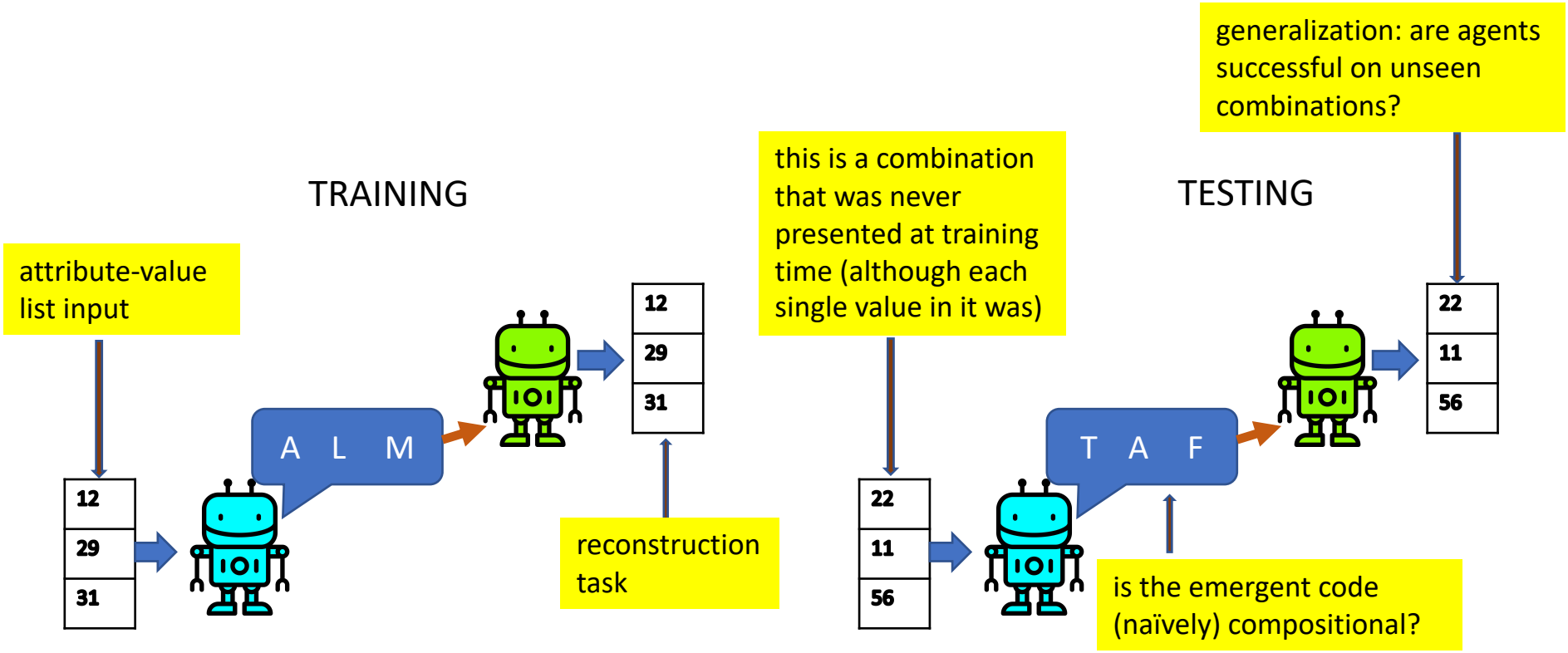
Compositionality and generalization in a communication game



Compositionality and generalization in a communication game

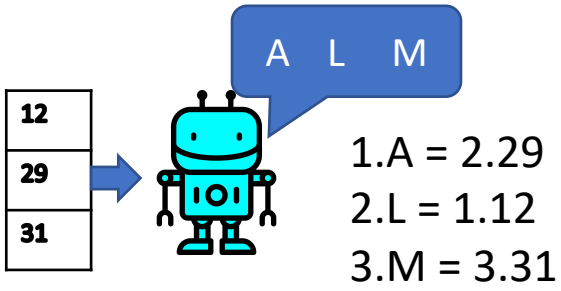


Compositionality and generalization in a communication game



Quantifying (one type of) naïve compositionality

Positional disentanglement measures strong form of naïve compositionality: to what extent do symbols in a certain position univocally refer to different values of the same attribute

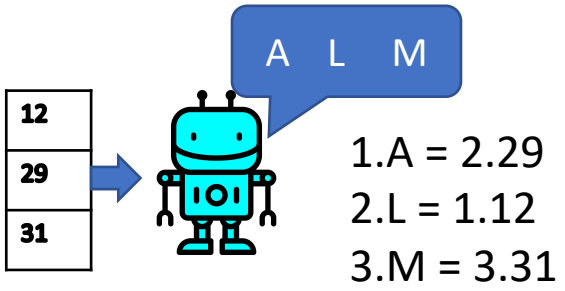


$$posdis = \frac{1}{c_{len}} \sum_{j=1}^{c_{len}} \frac{\mathcal{I}(s_j; a_1^j) - \mathcal{I}(s_j; a_2^j)}{\mathcal{H}(s_j)}$$

$$a_1^j = argmax_a \mathcal{I}(s_j; a); a_2^j = argmax_{a \neq a_1^j} \mathcal{I}(s_j; a)$$

Quantifying (one type of) naïve compositionality

Positional disentanglement measures strong form of naïve compositionality: to what extent do symbols in a certain position univocally refer to different values of the same attribute



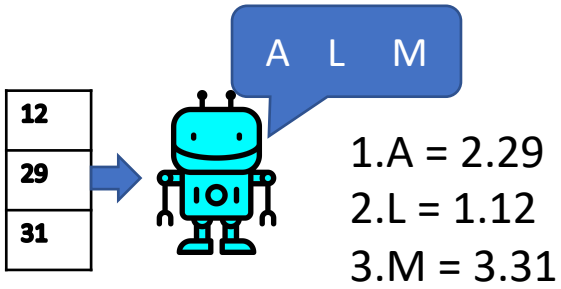
Similar results in experiments with other compositionality measures!

$$posdis = \frac{1}{c_{len}} \sum_{j=1}^{c_{len}} \frac{\mathcal{I}(s_j; a_1^j) - \mathcal{I}(s_j; a_2^j)}{\mathcal{H}(s_j)}$$

$$a_1^j = argmax_a \mathcal{I}(s_j; a); a_2^j = argmax_{a \neq a_1^j} \mathcal{I}(s_j; a)$$

Quantifying (one type of) naïve compositionality

Positional disentanglement measures strong form of naïve compositionality: to what extent do symbols in a certain position univocally refer to different values of the same attribute



$$posdis = \frac{1}{c_{len}} \sum_{j=1}^{c_{len}} \frac{\mathcal{I}(s_j; a_1^j) - \mathcal{I}(s_j; a_2^j)}{\mathcal{H}(s_j)}$$

$$a_1^j = argmax_a \mathcal{I}(s_j; a); a_2^j = argmax_{a \neq a_1^j} \mathcal{I}(s_j$$

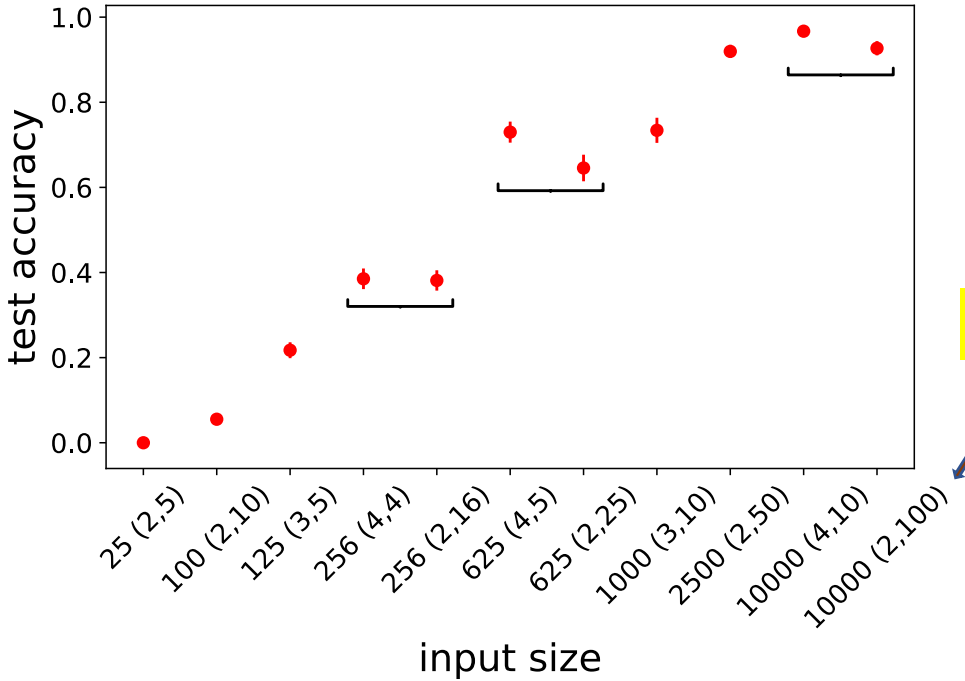
Similar results in experiments with other compositionality measures!

Still: to the extent that our emergent languages are compositional, they use positional encoding of information

Do emergent languages support generalization?

Do emergent languages support generalization?

Yes, in function of how varied the training input is!

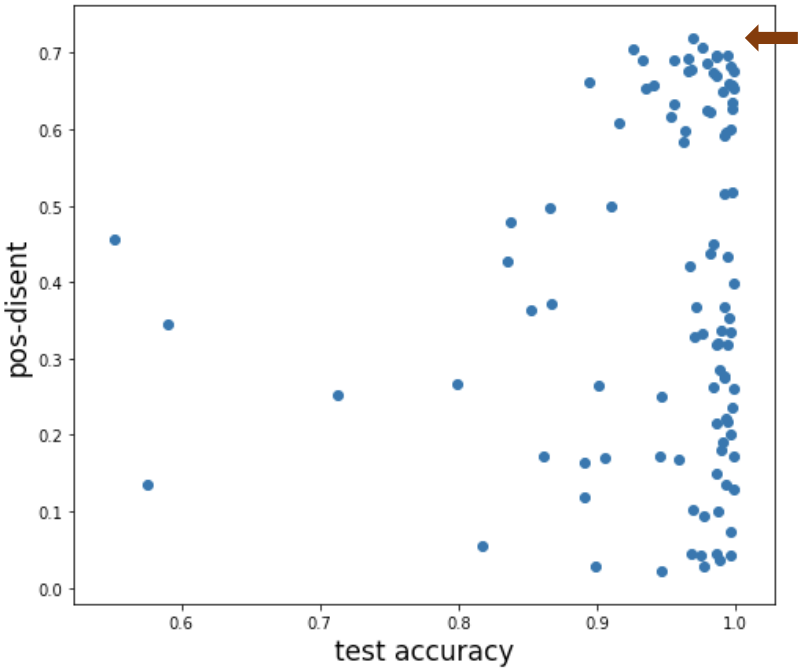


size = $Nvals^{Natts}$

A general pearl of wisdom: do not test neural network generalization capabilities in small toy worlds!

Is compositionality needed for generalization?

No!... no correlation between generalization and compositionality!



top compositionality scores are far from theoretical max (=1)!

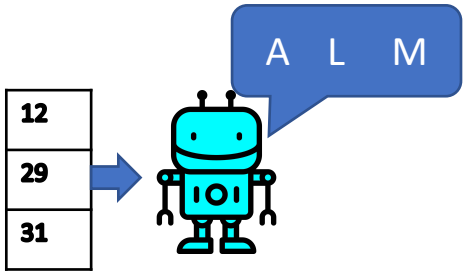
What's going on?
"Leaky" compositionality

- Agents always need more expressive power to converge than would be strictly necessary in a perfectly compositional language
  - for 2 100-valued attributes in input, language needs at 3-symbol strings with 100-symbol vocabulary
  - Emergent languages are not "efficient" (Chaabouni et al NeurIPS 2019)
- The extra leeway is used by languages in non-compositional ways
  - E.g., in one of the most compositional languages, 2 symbols largely disambiguate the values of 2 attributes, but the 3d symbol is often needed to ``fine-tune'' the referent:
    - a c f -> 1.53 2.43
    - a -> 1.53 or 1.72
    - c -> 2.43
    - f -> 1.53 or 2.94

Is compositionality good for something?

Is compositionality good for something?

Interpretability, humans in the loop!



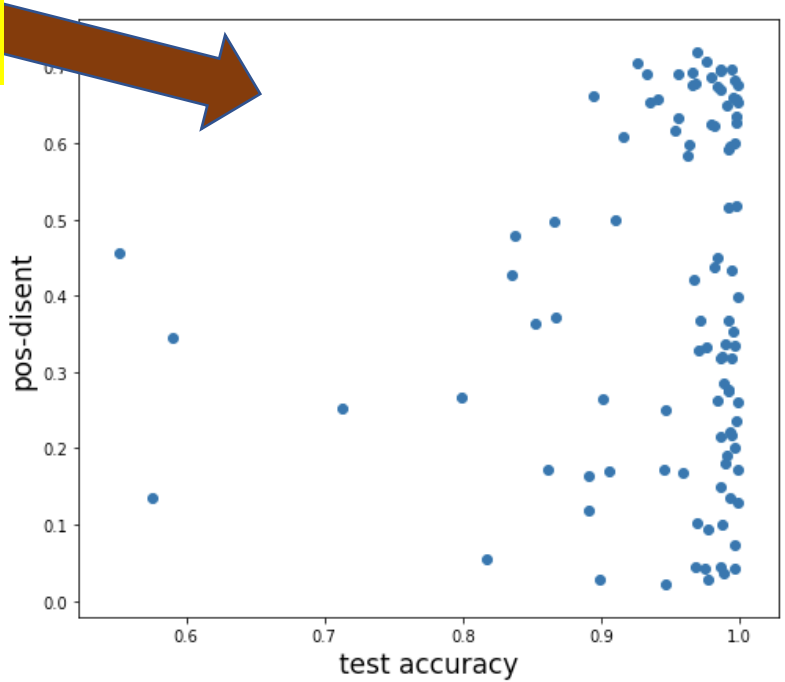
Easy:
A in first position refers to value 12 of attribute 1

Hard:
A in first position and M in third refer to combination of 1.12 and 3.31 if second symbol is L, P or Q

Is compositionality good for something?

It might not be necessary, but it appears to be sufficient for generalization

there's no high-compositionality, low-generalization language!



Is compositionality good for something?

Compositional languages are “viral”!

- More compositional languages are easier for new Receivers (even Receivers with different architectures):
  - .87 Spearman correlation between compositionality and training speed of new Receivers with frozen Sender languages
  - .80 Spearman correlation between compositionality and generalization performance of new Receivers trained with frozen Sender languages
- We knew that cultural transmission favors the emergence of compositionality (iterated learning experiments: Kirby, Smith, etc.)
- It is also the case that compositionality favors cultural transmission!
- This is good, given that our end-goal is to breed large communities of interacting agents

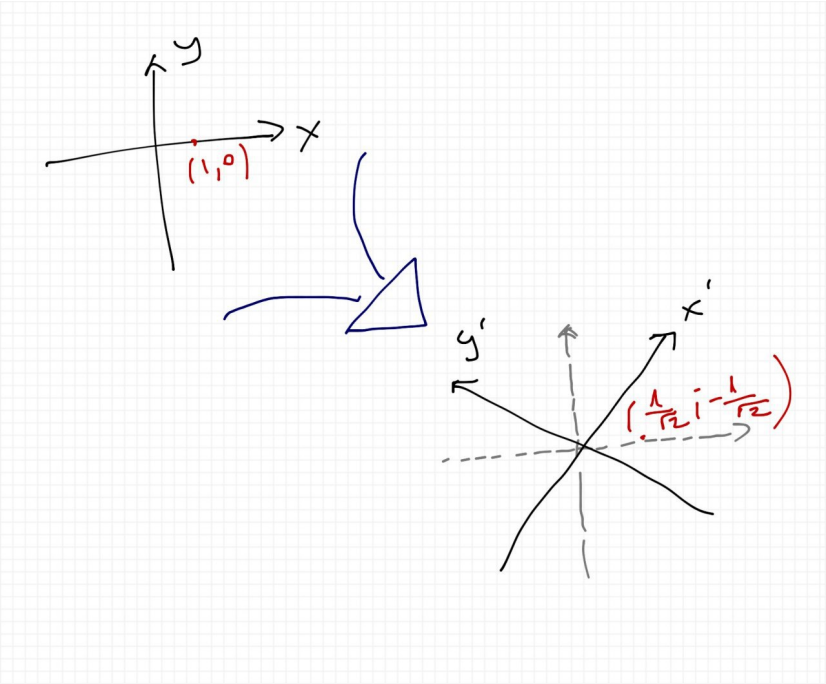
Are compositional languages always the easiest to spread?

No! This will depend on the input (Lazaridou et al ICLR 2018), the task, and *what comes easy to a neural network!*

Kharitonov and Baroni: Emergent Language Generalization and Acquisition Speed are not Tied to Compositionality

https://arxiv.org/abs/2004.03420

Coordinate and rotated languages



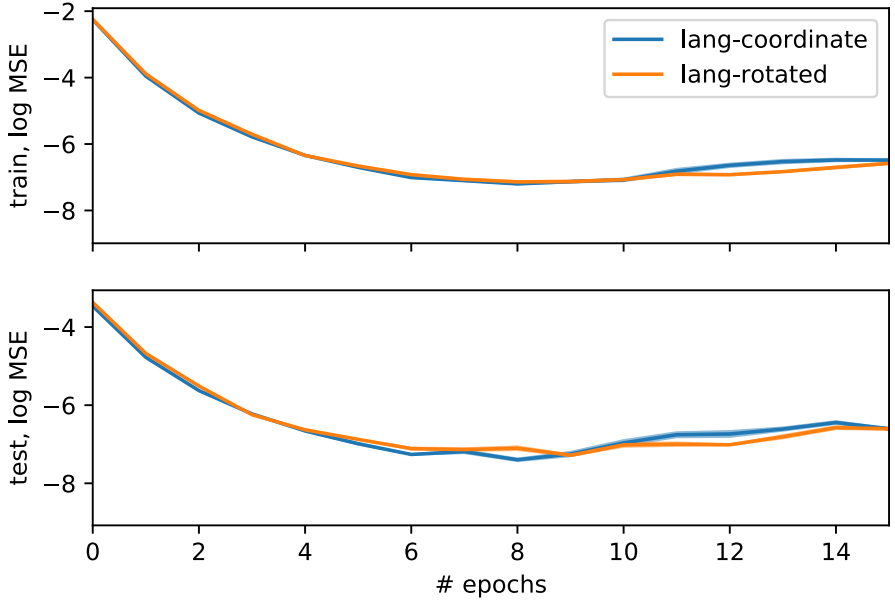
- Inputs: coordinates within unit circle
- Manually-crafted languages instead of trainable Sender
- Naïvely compositional “coordinate” language: two symbols directly corresponding to (discretized) coordinates
- ”Rotated” language: symbols correspond to (discretized) coordinates after rotating the axes by $\pi/4$
  - Identifying either element always requires looking at both symbols, resulting in a very entangled encoding

Image credit: Eugene Kharitonov

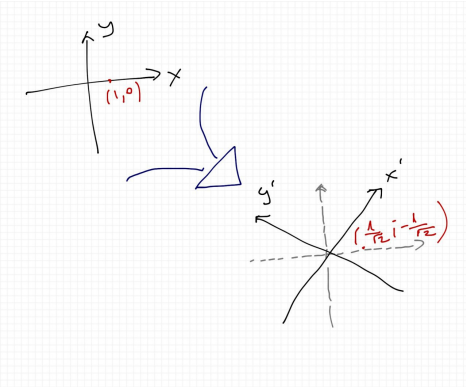
Coordinate and rotated languages

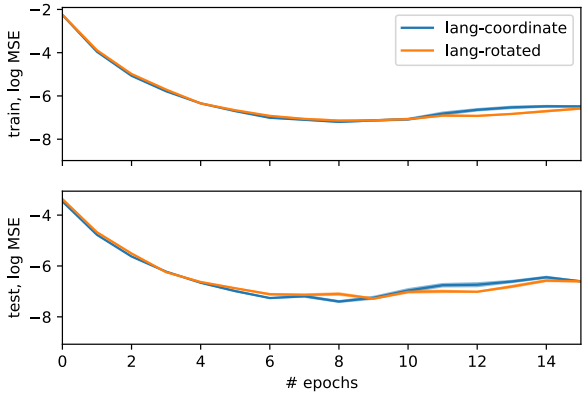
ease of learning

ease of use



Coordinate and rotated languages





- In the rotated language a linear transformation links (values denoted by) symbols and inputs
- Linear transformations are neural networks' favorite sport, so for Listener highly entangled rotated language is as easy as perfectly (naïve) compositional coordinate language
- There is nothing universal about more compositional languages being easier to learn and use!

Outline

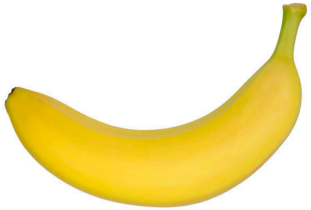
- Emergent language in deep networks: why, how, what?
- Compositionality and generalization in emergent languages
- **Take-home messages**

Take-home messages

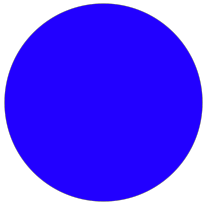
- High-level program: evolve a shared language to empower communities of specialized neural networks
- Emergent neural network languages do tricky things
- Lots of work on emergent language compositionality, however:
  - No widely accepted, useful definition of compositionality
  - If what you want is generalization, ease of learning, viral strength, you might be better off focusing on optimizing those properties directly, without worrying about how “compositional” the emergent languages are
- But surely human language, the queen of communication systems, is very compositional?

Human language: the paragon of compositionality

Train



”banana”



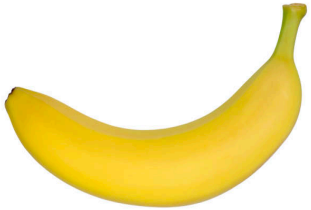
”blue”

”blue banana”

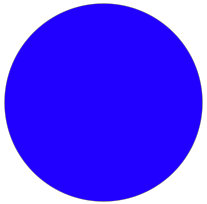
Test

Human language: the paragon of compositionality

Train



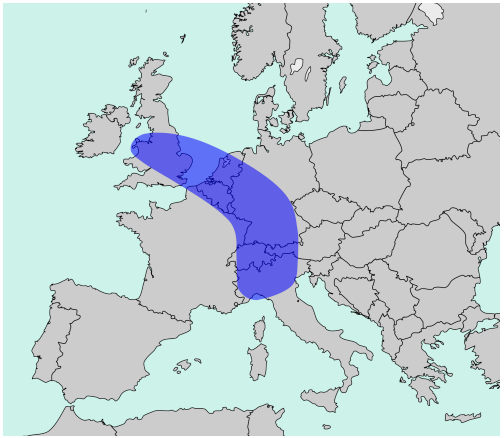
”banana”



”blue”

”blue banana”

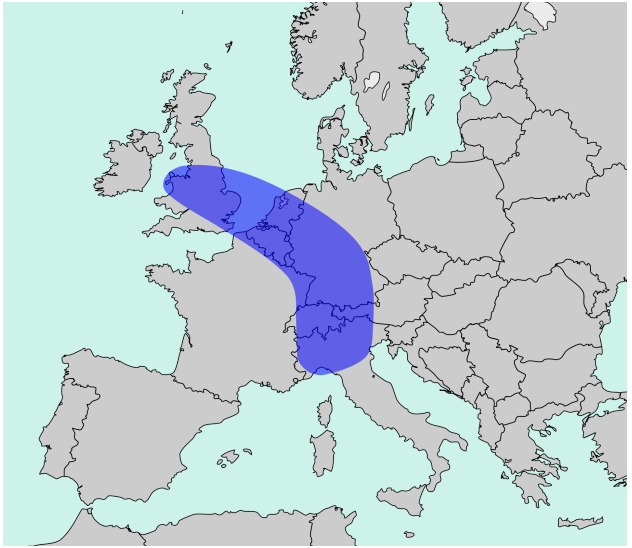
Test



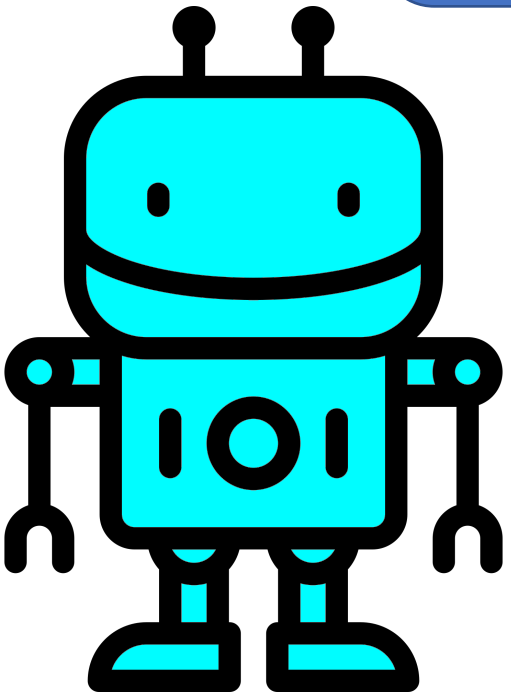
https://en.wikipedia.org/wiki/Blue_Banana

Blue banana communication

- Human language is full of non-compositional expressions: frozen metaphors, idioms, lexicalized constructions, ...
- It is not by chance that, 50 years into the Montagovian program, we are still only able to account for small fragments of English and other languages with a fully compositional semantics
- The right goal for emergent languages (and neural networks in general) might not be full compositionality, but a human-language-like opportunistic, efficient mixture of compositional and non-compositional means of expression



t h a n k y o u



https://github.com/facebookresearch/EGG