Language-Augmented Multi-agent Learning and Communication

Maxime Toquebiau[1,2], Nicolas Bredeche[2], Faïz Ben Amar[2], Jae Yun Jun Kim[1]

[1]ECE Paris
[2]ISIR, Sorbonne Universités

September 12th, 2024

Table of Content

Related works: Language

Language helps learning

i.e., language is a way to understand concepts and how they relate

- ➢ *The Origins of Intelligence in Children*, Piaget, 1952.
- ➢ *Thought and Language*, Vygostky, 1934.

→ Language can be a tool to guide the training of RL agents

- ➢ *Speaker-Follower Models for Vision-and-Language Navigation*, Fried et al., NeurIPS 2018.
- ➢ *Language and culture internalization for human-like autotelic AI*, Colas et al., Nature Machine Intelligence 2022.
- ➢ *Inner Monologue: Embodied Reasoning through Planning with Language Models*, Huang et al., CoRL 2023.
- ➢ *Survey on Large Language Model-Enhanced Reinforcement Learning: Concept, Taxonomy, and Methods*, Cao et al., 2024.

Language helps communicating ideas

Natural languages have evolved into complex mechanisms allowing:

- Use of many different concepts describing the environment and abstract ideas
- Composition of concepts
- Transmission of knowledge

- ➢ *The Synthetic Modeling of Language Origins*, Steels, Evolution of Communication 1997.
- ➢ *The evolution of syntactic communication*, Nowak et al., Nature 2000.

Related works: Multi-agent Communication

Communication in MADRL is mostly "Emergent Communication" (EC)

i.e., agents develop a communication system during training, based on task rewards

- ➢ *Emergence of Grounded Compositional Language in Multi-Agent Populations*, Mordatch and Abbeel, AAAI CoAI 2018.
- ➢ *On the interaction between supervision and self-play in emergent communication*, Lowe et al., ICLR 2020.
- ➢ *Compositionality and Generalization in Emergent Languages*, Chaabouni et al., ACL 2020.
- ➢ *Emergent Communication at Scale*, Chaabouni et al., ICLR 2022.

→ EC can work well but is hard to evaluate and interpret

- ➢ *On the Pitfalls of Measuring Emergent Communication*, Lowe et al., AAMAS 2019.

Learning of language-based communication is mostly limited to text-based tasks

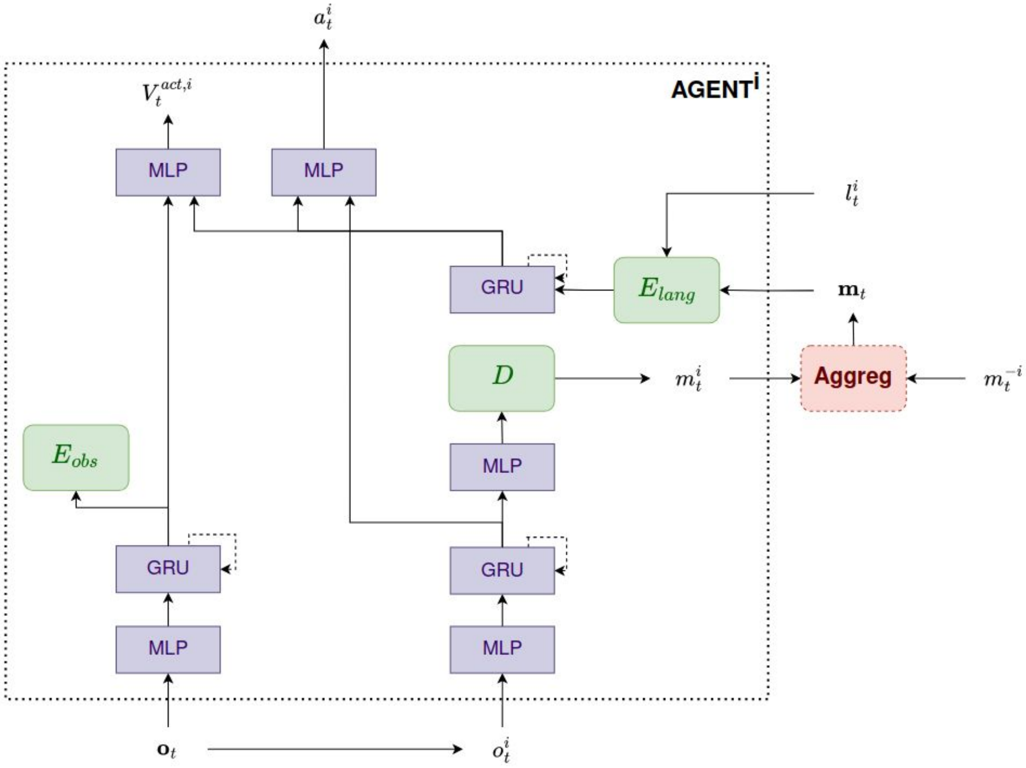
e.g., Visual-Question-Answering, Speaker-Listener

- ➢ *Learning Cooperative Visual Dialog Agents With Deep Reinforcement Learning*, Das et al., ICCV 2017.
- ➢ *Countering Language Drift via Visual Grounding*, Lee et al., EMNLP 2019.

⇒ How can language-based communication be learnt in embodied settings ?

Language-Augmented Multi-Agent Learning and Communication

Method

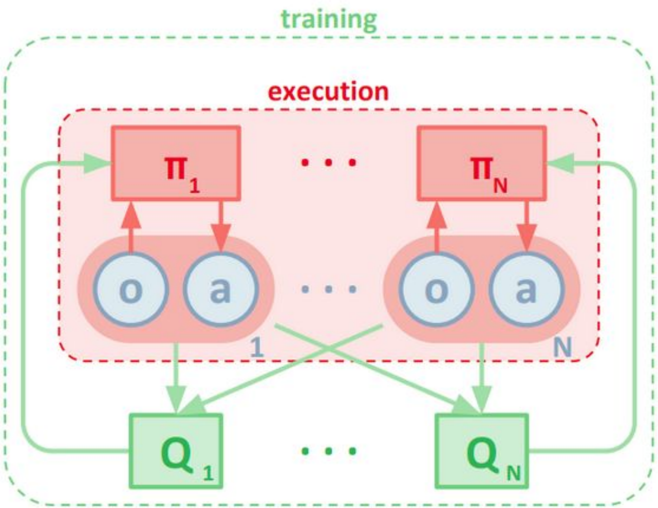


Learning

Policy Learning

MAPPO

(The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games, Yu et al., NeurIPS 2021)



Language Learning

CLIP Loss

With the **Observation Encoder** $\omega : \mathbb{R}^N \rightarrow \mathbb{R}^M$, and the **Language Encoder** $\lambda : \mathbb{R}^{L \times V} \rightarrow \mathbb{R}^M$,

the grounding objective is:

$$J(\theta_\omega, \theta_\lambda) = max[cosim(\omega(o_k), \lambda(l_k))]$$

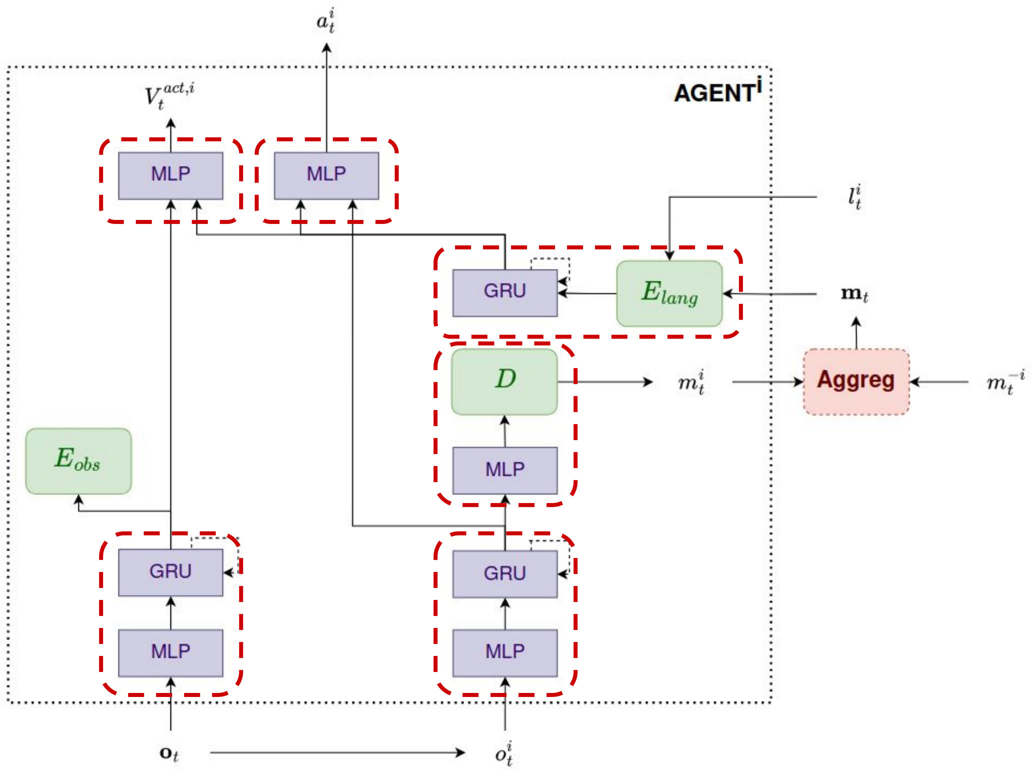
Captioning Loss

With the **Language Encoder** $\lambda : \mathbb{R}^{L \times V} \rightarrow \mathbb{R}^M$, and the **Decoder** $\delta : \mathbb{R}^M \rightarrow \mathbb{R}^{L \times V}$,

the captioning objective is:

$$J(\theta_\lambda, \theta_\delta) = min\left[\frac{1}{N}\sum_{i=0}^{N}(\hat{l}_i - l_i)^2\right]$$

Method



a_t^i
AGENT^i
$V_t^{act,i}$
MLP
MLP
l_t^i
GRU
E_{lang}
$\mathbf{m}_t$
D
m_t^i
Aggreg
m_t^{-i}
E_{obs}
MLP
GRU
MLP
GRU
MLP
$\mathbf{o}_t$
o_t^i

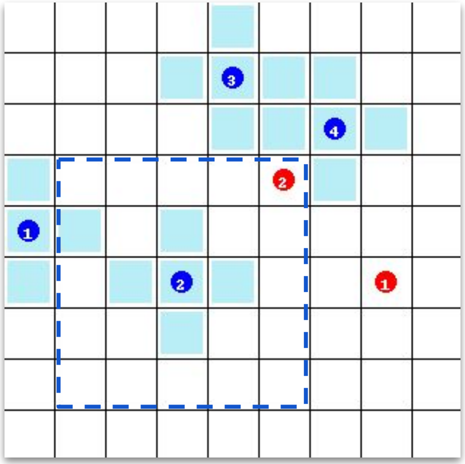
Experiments

Demonstrating advantages of having language

Experiments on Predator-Prey in Magym environment, two versions:

- see_agents: agents see surroundings preys and other agents
- no_see_agents: agents do not see other agents

During training, agents get their observation and the corresponding “perfect message”: e.g., agents 2, 3 and 4 observe “Prey Center”



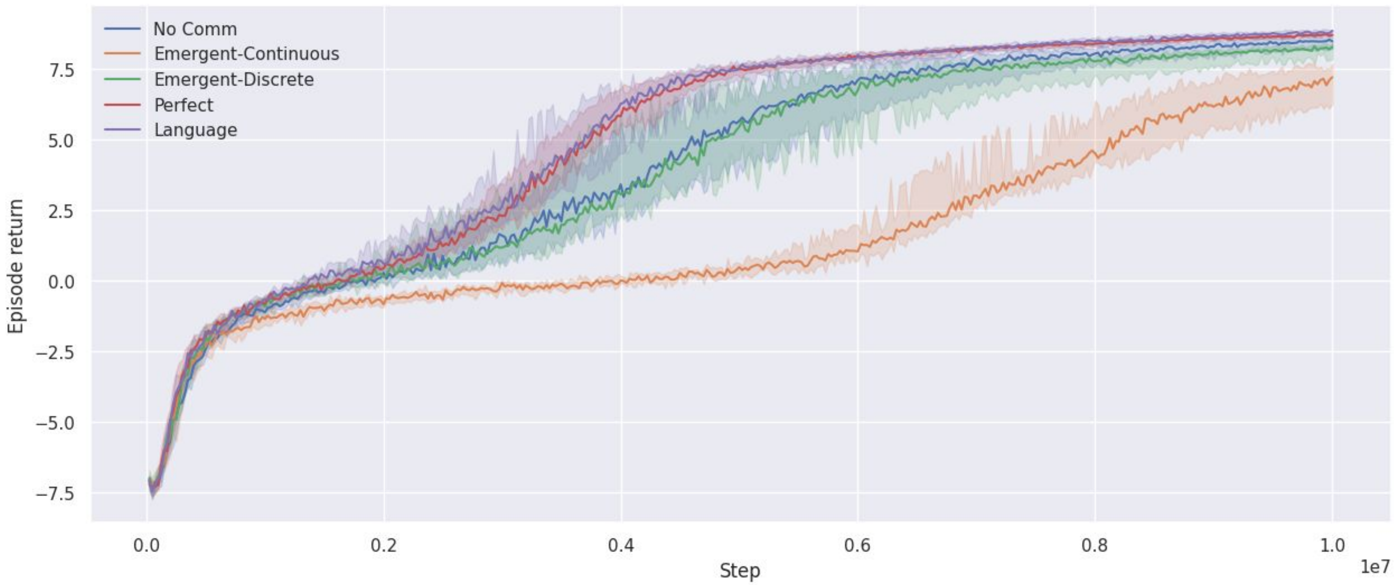
Objective is to show that language allows:

- Better multi-agent learning
- Adaptability
- Zero-shot teaming
- Interpretation
- Interaction

Results

Training on Predator-Prey

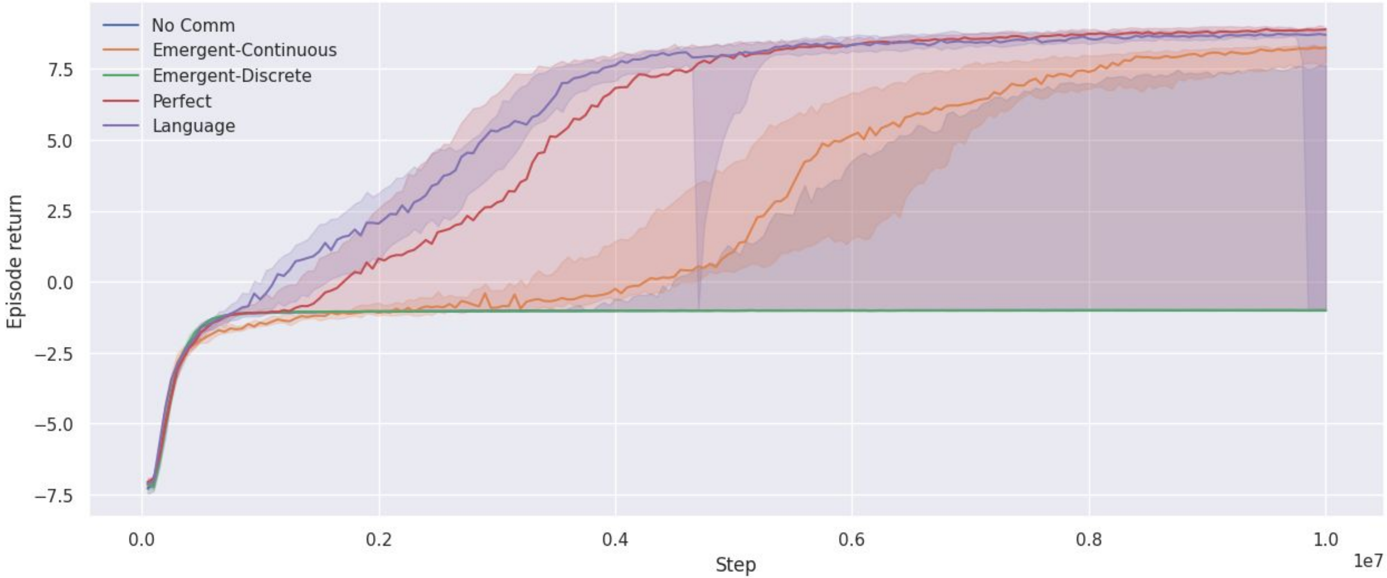
Predator-Prey see_agents (15 runs each, median and ci-95)



Results

Training on Predator-Prey

Predator-Prey no_see_agents (15 runs each, median and ci-90)

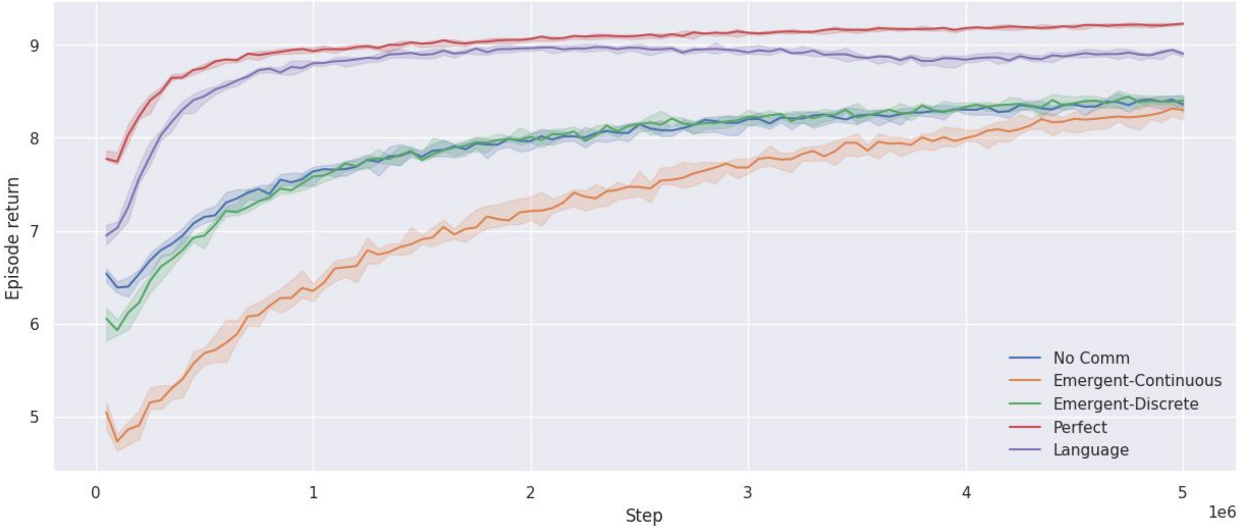


Results

Adaptation

For each run:

- Take best run trained on 9x9 P-P see_agents
- Train on 15x15 P-P see_agents (15 runs each, median and ci-95)



Next:

- Adapt on different task (or different environment?)

Results

Zero-shot Teaming

For each algo:

- Make teams from 2 best runs (2 agents from each)
- Eval on set of 24 "difficult" scenarios

	Mean Return	
	Mean	Std
No Comm	0.14	2.8
Emergent-Continuous	0.1	2.2
Emergent-Discrete	-0.64	1.29
Perfect	**5.95**	2.42
Language	3.03	2.85

Next:

- Same with teams made from 4 different runs (1 agent from each)

Results

Interpretation

Compare messages generated by Language vs. perfect messages:

→ **98%** of generated messages are perfect

Longer messages are harder to generate perfectly:

Generated_Message	Perfect_Message
['Prey', 'Center', 'Prey', 'South']	['Prey', 'East', 'Prey', 'South', 'West']
['Prey', 'West', 'Prey', 'South']	['Prey', 'West', 'Prey', 'South', 'West']
['Prey', 'South', 'West', 'Prey', 'South']	['Prey', 'Center', 'Prey', 'South', 'West']
['Prey', 'North', 'Prey', 'West']	['Prey', 'North', 'Prey', 'North', 'West']

Thank you for you attention !

Questions ?