Journal of Autonomous Intelligence (2019) Volume 2 Issue 3.
doi:10.32629/jai.v2i3.56

Original Article

Compositional Grounded Language for Agent Communication in Reinforcement Learning Environment

K. Lannelongu[1*], M. de Milly[1], R. Marcucci[1], S. Selevarangame[1], A. Supizet[1], and A. Grincourt[1]

ECE Paris School of Engineering, France

ABSTRACT

In a context of constant evolution of technologies for scientific, economic and social purposes, Artificial Intelligence (AI) and Internet of Things (IoT) have seen significant progress over the past few years. As much as Human-Machine interactions are needed and tasks automation is undeniable, it is important that electronic devices (computers, cars, sensors…) could also communicate with humans just as well as they communicate together. The emergence of automated training and neural networks marked the beginning of a new conversational capability for the machines, illustrated with chat-bots. Nonetheless, using this technology is not sufficient, as they often give inappropriate or unrelated answers, usually when the subject changes. To improve this technology, the problem of defining a communication language constructed from scratch is addressed, in the intention to give machines the possibility to create a new and adapted exchange channel between them. Equipping each machine with a sound emitting system which accompany each individual or collective goal accomplishment, the convergence toward a common ‘’language’’ is analyzed, exactly as it is supposed to have happened for humans in the past. By constraining the language to satisfy the two main human language properties of being ground-based and of compositionality, rapidly converging evolution of syntactic communication is obtained, opening the way of a meaningful language between machines.

Keywords: *Machine Learning; Reinforcement Learning; Natural Language Processing*

ARTICLE INFO
Received: Oct 12, 2019
Accepted: Oct 31, 2019
Available online: Nov 15, 2019

***CORRESPONDING AUTHOR**
K. Lannelongue, ECE Paris School of Engineering, France;
michel.cotsaftis@ece.fr;

CITATION
K. Lannelongue, M. de Milly, R. Marcucci, S. Selevarangame, A. Supizet and A. Grincourt. Compositional grounded language for agent communication in reinforcement learning environment. Journal of Autonomous Intelligence 2019;2(3): 1-8. doi: 10.32629/jai.v2i3.56

1. Introduction

Development of agents capable to communicate and to use a flexible and meaningful language is one of the long-standing and most ambitious challenges faced by AI[1–2]. Aside development of information systems designed for giving machines always more autonomy[3–6], there also exists a large range of expression channels by which machines are giving their “status” to human supervisors through adapted communication protocols usually made by human operators for interpretation and control[7–10]. In most cases at this level the machines are preprogrammed and controlled to operate with modest autonomy a series of tasks predefined by human operator[11–15]. However, in a coming context where humans and machines are going to work collectively, and machines to become more autonomous “agents”, necessary next step has to be prepared where agents will hold some language capacity to better interact and to productively collaborate or make decisions interpretable by humans[16–18]. With the improvement of speech-understanding technology and voice-input applications, the need for Natural Language Processing (NLP)[19] will increase. Nowadays NLP is very effective for small text translation[20], scenarized chat-bots or filtering spam messages. Question-Answering (QA) is becoming more and more popular through applications such as Siri, OK Google, chat-bots and virtual assistants[21–22]. Even if

they look promising, much of their success is for the moment a result of intelligently designed statistical models based on static, passive, and mainly supervised regimes ultimately trained on large static datasets[23–24]. In this context, the use of NLP for creating a seamless and interactive interface between humans and machines will continue to be a priority for today and tomorrow increasingly cognitive applications. Developing an artificial sophisticated language system[25–27] is mandatory for machines to become more intelligent and to gain the ability to learn like humans[28]. In parallel, it could also open important insights into questions related to development of human language and cognition. It immediately comes out that, if communication is to be created from first principles, the only way to do it is from necessity. In other words, approaches learning to imitate human language from examples, even if useful, only capture structural and/or statistical relationships. They are completely missing language functional aspect and do not provide any answer on why language exists[29–31]. More precisely, they do not relate language as it stands with the reason of its existence, which is a successful coordination mean between humans. Here to replicate as much as possible for machines what was occurring for humans, it is claimed that if such language is created from scratch, it should necessarily develop in an environment giving this emerging language the two main properties of human one, ie to be grounded-based and compositional[32–34], even if other models can be conceived[35–39].

Present project aims at developing a new technique of NLP by fostering the emergence of a compositional and ground-based language amongst the machines exactly like it already exists between humans. Human language is grounded because it is based on experience in the real world. If a dictionary defines words with other words, a human will associate a word with sensory-motor experience (sight, touch etc.). As the agents will use words to describe concepts in their environment, a ground-based language will emerge rather easily. Compositionality in a language consists in that the meaning of a complex expression is determined by the meaning of its constituent expressions and the rules used to combine them[40–41], in the idea of adding up individual words to create a meaningful sentence altogether. The emergence of compositionality in a language only happens if the number of describable concepts (or learning events) is larger than the vocabulary size by following Zipf law which states that the frequency ϕ_i of occurrence of a word is inversely proportional to its position i.[42]

2. Environment Description

In this work, a physically simulated two-dimensional environment consisting of N agents and M landmarks in continuous space and discrete time is considered. Both agent and landmark entities inhabit physical positions p in space and possess descriptive physical characteristics, such as color and shape type, see **Figure 1**.

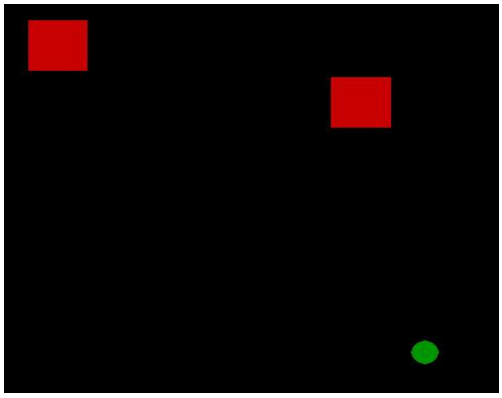

Figure 1. Example of the environment for N=1 (green circle) and M=2 (red square)**.**

In addition, agents can move in the environment and direct their gaze to a location $\boldsymbol{\ell}$. Denote X the physical state of an entity. To facilitate the emergence of previous two language properties, the environment considered here is a cooperative and partially observable Markov game[43], which is a multi-agent extension of a Markov decision process. The cooperative setting allows formulate the problem as a joint minimization across all agents, as opposed to minimization-maximization problems resulting from competitive settings. The reward for each agent i (i=1,…,N) is given at time t by $r(\boldsymbol{s}_i(t),\boldsymbol{a}_i(t))$, where $\boldsymbol{s}_i \in \boldsymbol{S}$ the set of the possible configurations of all N agents, $\boldsymbol{a}_j \in \boldsymbol{A}$ is the set of possible agent actions for all N agents, and $t \in [0,T]$. In a cooperative setting, the problem is to find for each agent i a policy π_i maximizing the expected shared return **R** for all agents:

$$\mathbb{R}_{MAX} = \mathrm{Max}_{\pi}\, \mathbf{R}(\pi) \quad ; \quad \mathbf{R}(\pi) = \mathbb{E}[\Sigma_{t=0}^{T} \Sigma_{i=0}^{N} r(\mathcal{s}_i(t), \boldsymbol{a}_i(t))] \tag{1}$$

Here on top of performing physical actions, each agent utters verbal communication symbols v at every time step. These utterances are discrete elements of an abstract symbol vocabulary $\mathcal{V}$ of size K. The symbols are denuded of any significance nor meaning, and are treated as abstract categorical variables emitted by each agent and observed by all other agents[44]. The vector representing one-hot encoding of symbols v is denoted by $\mathbf{V}$. Each agent has private internal goals, not observed by other agents, specified by vector G. These goals are grounded to the physical environment and consist of moving to a location. Agents are emitting verbal utterances for accomplishment of their cooperative goal. To aid in accomplishing its goals, each agent i has also a private internal recurrent memory bank M_i not observed by other agents. This memory bank has no designed behavior and it is up to the agents to learn how to utilize it appropriately. The full state of the environment is now given by :

$$S_{ext} = \{X(1,...,(N+M)),V(1,...,N),M(1,...,N),G(1,...,N)\} \tag{2}$$

Each agent observes the physical states of all entities in the environment, the verbal utterances of all agents, and its own private memory and goal vector. The observation $\boldsymbol{o}_j \in \mathcal{O}$ (j=1,…,N) for agent i is:

$$\boldsymbol{o}(i)(S) = [\,{}_i X(1,...,(N+M)), V(1,...,N), M(i), G(i)\,] \tag{3}$$

where ${}_i X(j)$ is the observation of entity j physical state in agent i reference frame, and $\mathcal{O}$ is the set of observations made by all agents. Finally system dynamical equations are given in[45].

In present multi-agent environment, each agent for simplicity will act by sampling actions from the same stochastic policy π on the same sets of observations $\mathcal{O}$ and of actions $\mathcal{A}$ (which are the features/parameters in the model).

3. Optimum Policy Determination

The problem is to find the common policy π maximizing their share return $r(.,.)$. Here a cooperative setting is used where the policy is learned by maximizing the expected shared return for all agents. In present case a model capable of fostering both a clear message sending and the right action to take following a received message is required. Traditional model-free reinforcement learning approaches are not optimal in present environment. Indeed, Q-learning method faces scalability issues as it scales quadratically with Q value (parameter defining the model) updated by calculating optimal action for each state (using max function with quadratic complexity). Moreover the model-free policy gradient method uses sampling to estimate the gradient of policy return and can exhibit high variance. This is not suitable when dealing with sequential communication actions. To enable the policy learning process avoid the above mentioned problems, a batch of 100 random environment instantiations is sampled at every optimization iteration and back-propagates their dynamics through time to calculate the total gradient return. As the communication is resting here on discrete symbols emission, there is an issue as backpropagation is made through differentiation (gradient computation during forward propagation) with respect to the parameters and works with continuous variables. A way to differentiate and to back-propagate the environment state dynamics partially defined by categorical features is required. The problem is solved by using an efficient gradient estimator approximating the non-differentiable sample from a categorical distribution by a differentiable sample out of continuous Gumbel-Softmax distribution[46]. This is achieved by using the Gumbel-Softmax categorical re-parameterization, which gives an end-to-end differentiable model.

Assuming a random variable with a categorical distribution with class probabilities $\pi(j)$ (j=1,..k), the Gumbel-Max trick[47] provides a simple and efficient way to draw samples z from a categorical distribution with class probabilities π by applying one-hot encoding to the Argmax function. Instead of drawing samples from a categorical distribution, the distribution is approximated by a continuous one, so that one can compute the gradient and back-propagate the loss to tune accordingly the model parameters. One then gets the

k-dimensional vectors

$$z_i = E(\pi(i),G(i))/\{\Sigma_{i=1,...k}\, E(\pi(i))\} \quad (4)$$

with $E(\pi(i),G(i)) = \exp\{[\log(\pi(i) + G(i)]/\tau\}$. As the softmax temperature τ moves away from 0, the samples are not one-hot and tend to uniform distribution for $\tau \to \infty$. For any $\tau > 0$, the Gumbel-Softmax distribution is smooth and differentiable. Hence categorical samples are replaced by samples from Gumbel-Softmax distribution during the training of the neural network, allowing gradient computation and backpropagation use. The system is built on three processing modules: Communication network, Physical network and Final network. The policy must consolidate multiple incoming communication symbol streams emitted by other agents, as well as incoming observations of physical entities. The outputs of individual processing modules are pooled with a Softmax operation into feature vectors f_c and f_x with 256 components.

Each network has three hidden layers of 256 hidden units. The Communication network inputs are the utterances given by agents, the Physical network ones are observations of the states, and the Final network outputs are the gaze, the acceleration, the new utterances and the memory update.

4. Compositionality with Dirichlet Process

As seen on **Figure 2**, the plot of learning events number vs vocabulary size is closely approximated by Zipf law. The approximation by "Zipfian" law is not surprising as it describes the patterns of many natural situations (word frequencies in language, city populations, websites traffic, ..). As seen on the plot, there are not many learning events (i.e. concepts describable by words) so a large lexicon size is not necessary.

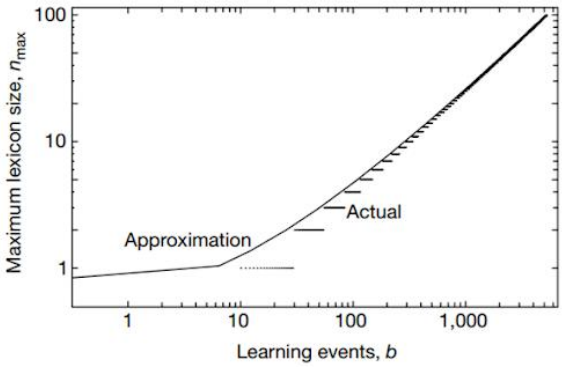

Figure 2. Maximum lexicon size vs learning events (actual calculated value and zipfian law approximation)

On the other hand however, by setting the vocabulary size to a too small number (e.g. 5), the optimization is stuck in local minima because there are too few words available and concepts tend to merge together. For instance, in English language, only thirty different words are needed to count from 0 to 999,999. Compositionality and language evolution created a balance between too few words making the counting difficult due to confusion and too many words making it tedious. Here agents are given a vocabulary size of 20 different words with a penalty if they use too many different words calculated with Dirichlet process, which is a probability distribution the range of which is itself a set of probability distributions[48]. In present case, the word the agent is going to choose in its vocabulary is the base variable, and another distribution is applied on it to describe how the random variable is distributed.

The probabilities P_{nu} for drawing a new utterance from the base distribution and P_d for not drawing a new word from an already picked-up variable are respectively :

$$P_{nu} = \alpha/(\alpha + n - 1) \quad ; \quad P_d = N_x/(\alpha + n - 1) \quad (5)$$

where α is the Dirichlet hyper-parameter determining the probability to pick a new word, N_x is the number of times utterance x has been picked and N the total number of utterances. The reward is given by :

$$r_c = \Sigma_{i,t,k} \mathbb{1}[v_i(t) = v_k]\log(p(v_k) \quad (6)$$

The less a symbol is uttered, the lower is the probability that it will be sampled in the future and the higher is the penalty. This mechanism fosters the use of

most "popular" words, hence it leads to limit the size of used vocabulary implying the emergence of compositionality.

5. Results and Discussion

In present experience, an environment is first built up with one agent and two landmarks, see Figure 1. From one batch to another, the landmarks and agent positions are randomized. A landmark differs from another one by its color characteristics. The goal of the agent is given through an utterance listened by the agent at the beginning of the time step. The evolution of the distance between the agent and the goal indicates how the agent minimizes the distance to the target (the goal is one of the two landmarks) as iterations go on.

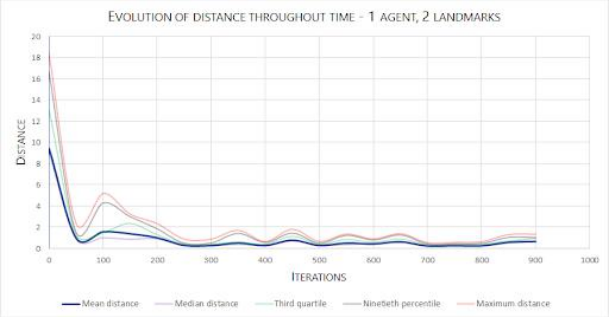

Figure 3. Distance to target vs iteration number

Figure 3 displays the time evolution of the distance and its dispersion between the agent and the target. Though converging, the dispersion continues to oscillate even for large number of iterations and despite mean distance much stronger convergence. This is here due to the very nature of the optimization process resting upon transformation "backward" of discrete observation process into an "enriched" continuous one with usual errors amplification at discreteness frequency. The phenomenon is further amplified when considering two agents and two landmarks allowing now exchange between agents through utterance communication. In this case, each agent is holding the goal of the other at the beginning, and they have to communicate to achieve their goals in cooperation. The distance between agent and to the landmark target is minimized through time, see **Figure 4**.

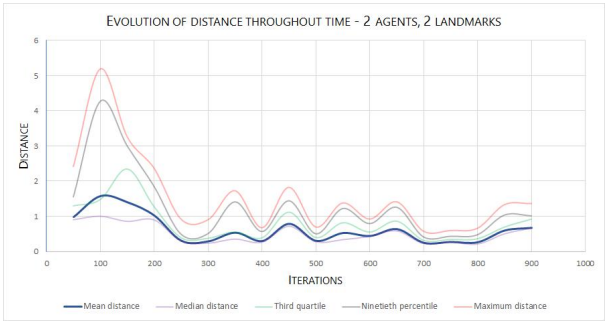

Figure 4. Evolution of syntaxic communication vs iteration number

As for previous case, the distance drops as the agents learn to complete the task, showing convergence of the approach. However, the much larger number of required iterations ($\cong 3$ times) stresses the limitation of the enlarged parameter window by utterances collection. Reason is that when they are started from scratch (ie when the communication language is built up from nothing), the simple and very global character of these new signals makes their gradient poorly directive, inducing higher oscillatory behavior related to more difficult phase mismatch resorption between agents. Consequence is that for overcoming well-identified difficulty to understand fully other participant goal through this channel (still observed at human level!) – here mathematically manifested by flat gradients in the iteration process –, it is necessary to run more iterations on a large enough initial set of utterances for agents to enrich their data bank and create a much finer response filter by redundancy. In present case, convergence curve shows that through the hearing during enough iterations, agents do identify, already with direct on-line training, how to react in order to complete their goal. In other words, such result indicates that there is a true emergence of meaning that can be quantified from the utterances used and heard by the agents. The result is the more interesting as, in a preliminary test of the interest of information enrichment resulting from use of utterances emitted by the agents, the study is developed on purpose here in a relatively "flat" way, where utterances are not related in present simple example to emergency constraints drastically narrowing the landscape explored by the agents. This fundamental problem of information quality will be discussed elsewhere with emergence of "language".

6. Conclusion

To give agents a possible ‘’intelligent’’ relationship similar to human beings, application of optimization techniques in the framework of cooperative and partially observable Markov games has been proposed. For the accomplishment of their goals, and in order to develop naturally and spontaneously a communication process between them, agents are, on top of observation of their environment, given a set of utterances they can emit at each step of their move toward their target. For being efficient this supplementary process is imposed to comply with the two conditions of producing a grounded language and satisfying compositionality, both important characteristic features of human language (ie spoken by human agents). In the simple case of one agent and two landmarks in a plane space, the research of optimum policy is shown, even when starting from scratch, to converge toward the emergence of a meaningful language. With two (communicating) agents, similar but slower convergence is observed, as expectable from specifically chosen weakly directive example here. These results are showing that, basically, utterances do have a meaning for each agent after its training, and exhibit an interesting potentiality as compared to traditionally used statistics based approaches.

Next step is sentences creation by agents with the vocabulary they are given. A way to evaluate the production of sentences is to modulate enough environment complexity for the agents to explore new actions in addition to the spatial movements studied in present paper. This provides a useful study about how much the vocabulary has to be adapted in regards of the actions possibly performed in an environment. Possible final step along this line is to end up with fully human speaking agents. It is already understandable that this step is difficult to implement because the bound on input complexity resulting from the number of letters in current human alphabet is too loose to limit the formation of words and much more of sentences from scratch to a short enough time for manageable communication. Implementation of entire human language grammatical correctness is not scalable in order to make sure that the language developed by the agents is itself correct. Even for a proportional increase of words learned by agents, the number of sentences they could correctly produce rises exponentially with unavoidable consequences on computational power requirements, calling for a specific and more restricted type of language for the agents.

Acknowledgments

The authors are very much indebted to ECE Paris School of Engineering for having provided the necessary set-up within which the work has been developed, to Mr G. Ducrocq for useful discussions and Pr. M. Cotsaftis for help in the preparation of the manuscript

References

1. S.J. Russell, P. Norvig. Artificial intelligence: A modern approach, 2nd edn. Prentice Hall, NJ, 2003.
2. J. Bratman, M. Shvartsman, R.L. Lewis, *et al*. A new approach to exploring language emergence as boundedly optimal control in the face of environmental and cognitive constraints. Proc. 10th Intern. Conf. on Cognitive Modeling, pp.7–12, 2010.
3. M.P. Deisenroth, G. Neumann, J. Peters. A survey on policy search for robotics. Foundations and Trends in Robotics 2013; (1-2): 1-142.
4. M. Cotsaftis. Toward global complex systems control – the autonomous intelligence challenge. J. of. Autonomous Intelligence 2019; 2(1): 11-27.
5. D.J.C. MacKay. Information theory, inference, and learning algorithms. Cambridge University Press 2003.
6. R.S. Sutton, A.G. Barto. Reinforcement learning. The MIT Press 1998.
7. R. Hafner, M. Riedmiller. Reinforcement learning in feedback control. Machine Learning 2011; 84(1-2): 137-169.
8. J. Kober, J.A. Bagnell, J. Peters. Reinforcement learning in robotics : A survey. Intern. J. Robotic Research 2013; 32(11): 1238–1274.
9. R. Coulom. Reinforcement learning using neural networks, with applications to motor control. PhD thesis, Institut National Polytechnique de Grenoble, 2002.
10. Shi-Xiang Gu, E. Holly, T. Lillicrap, *et al*. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. arXiv : 1610.00633v2 [cs.RO], 2016.
11. C.M. Bishop. Pattern recognition and machine learning. Information Science and Statistics. Springer-Verlag, 2006.
12. K. Doya. Reinforcement learning in continuous

time and space. Neural Computation 2000; 12(1): 219-245.
13. R.J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning 1992; 8: 229-256.
14. E. Theodorou, J. Buchli, S. Schaal. A generalized path integral control approach to reinforcement learning. J. Machine Learning Research 2010; 11: 3137-3181.
15. S. Amari. Natural gradient works efficiently in learning. Neural Computation 1998; 10: 251-276.
16. A. Lazaridou, A. Peysakhovich, M. Baroni. Multi-agent cooperation and the emergence of (natural) language. arXiv:1612.07182, 2016.
17. A. Lazaridou, N.T. Pham, M. Baroni. Towards multi-agent communication-based language learning. arXiv: 1605.07133, 2016.
18. B.M. Lake, T.D. Ullman, J.B. Tenenbaum, *et al*. Building machines that learn and think like people. arXiv:1604.00289 [cs.AI], 2016.
19. P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman. Natural language processing: An introduction. J Am Med Inform Assoc. 2011; 18(5): 544–551.
20. D. Bahdanau, K.H. Cho, Y. Bengio. Neural machine translation by jointly learning to align and translate. ArXiv:1412:3555, 2014.
21. G. Durrett, T. Berg-Kirkpatrick, D. Klein. Learning-based single-document summarization with compression and anaphoricity constraints. arXiv:1603.08887, 2016.
22. B. Dhingra, L. Li, X. Li, *et al*. End-to-end reinforcement learning of dialogue agents for information access. arXiv:1609.00777 [Cs], 2016.
23. A. Graves. Generating sequences with recurrent neural networks. ArXiv:1308.0850, 2014.
24. N. Kalchbrenner, E. Grefenstette, P. Blunsom. A convolutional neural network for modelling sentences. Proc. 52th Annual Meeting of the Association for Computational Linguistics 2014; I: 655-665.
25. L. Busoniu, R. Babuska, B. De Schutter, *et al* . Reinforcement learning and dynamic programming using function approximators. Taylor & Francis, CRC Press, 2010.
26. L.P. Kaelbling, M.L., Littman, *et al*. Reinforcement learning: A survey. J. Artificial Intelligence Research 1996; 4: 237–285.
27. L. Bottou. From machine learning to machine reasoning. Machine Learning 2014; 94(2): 133–149.
28. J. Weston, S. Chopra, A. Bordes. Memory networks. Proc. ICLR 2015, arXiv:1410.3916, 2015.
29. R. Jackendoff. Foundations of language. Oxford Univ. Press, 2003.
30. R.C. Berwick, N. Chomsky. Why only us: Language and evolution. Cambridge, MA, MIT Press, 2016.
31. S.I. Reynolds. Reinforcement learning with exploration. PhD Thesis, School of Computer Science, The University of Birmingham, UK, 2002.
32. L. Steels. What triggers the emergence of grammar? Proc. 2nd Intern. Symp. on the Emergence and Evolution of Linguistic Communication (EELC'05), pp.143–150, 1995.
33. R. Socher, A. Perelygin, J.Y. Wu, *et al*. Recursive deep models for semantic compositionality over a sentiment treebank. Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP), Vol.1631, 2013.
34. S.J. Gershman, E.J. Horvitz, J.B. Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. Science 2015 ; 349: 273–278.
35. T. Mikolov. Statistical language models based on neural networks. PhD Thesis, Brno University of Technology, 2012.
36. J.N. Foerster, Y.M. Assael, N. de Freitas, *et al*. Learning to communicate with deep multi-agent reinforcement learning. Proc. Annual Conference on Neural Information Processing Systems, pp.2137-2145, 2016.
37. S. Kirby, T. Griffiths, K. Smith. Iterated learning and the evolution of language. Current Opinion in Neurobiology 2014; 28: 108–114.
38. I. Sutskever, O. Vinyals, Q.V. Le. Sequence to sequence learning with neural networks, Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.D. Weinberger, eds., Vol.27, Curran Associates, Inc. , pp.3104–3112, 2014.
39. K. Beuls, L. Steels. Agent-based models of strategies for the emergence and evolution of grammatical agreement. PloS one 2013; 8(3): e58960.
40. T. Mikolov, I. Sutskever, K. Chen, *et al*. Distributed representations of words and phrases and their compositionality. arXiv:1310.4546 [cs.CL], 2013.
41. I. Mordatch, P. Abbeel. Emergence of grounded compositional language in multi-agent populations. arXiv: 1703.04908, 2018.
42. M.A. Nowak, J.B. Plotkin, V.A.A. Jansen. The evolution of syntactic communication. Nature 2000; 304(6777): 405-498.
43. M.L. Littman. Markov games as a framework for multi-agent reinforcement learning. Proc. XIth Intern. Conf. on Machine Learning 1994; 157: 157–163.
44. C.D. Manning, J. Bauer, M. Surdeanu, *et al*. Stanford core nlp natural language processing toolkit, 2014.
45. The physical state of agent i is $X_i(t)$ = col[p,dp/dt,v,c]$_i$(t) with c the (fixed) color of the agent, and its action space a_i is a_i = col[u_p,u_v,v]. It obeys differential equations $dX_i(t)/dt$ =

col[dp/dt, dp/dt + u_p + f({$X_i(t)$}),u_v,0)] where f(.) are physical interactions between agents and dp/dt a damping term to ease numerical computation with adjustable ∈ [0,1].

46. E. Jang, S. Gu, B. Poole. Categorical reparameterization with gumbel-softmax. arXiv : 1611.01144, [stat], 2016.
47. C.J. Maddison, D. Tarlow, T. Minka. A* sampling. Advances in Neural Information Processing Systems, pp.3086–3094, 2014.
48. Y.W. Teh. Dirichlet process, Encyclopedia of Machine Learning. Springer, pp.280–287, 2011.