

Simulation of Usage-Based Language Acquisition in Ernest

Mentor
Dr. Amitabh Mukherjee
Project Proposal by
Diksha Gupta

THE PLAN

Background
Ernest is a software agent that simulates language acquisition in a virtual world. It is designed to learn a language by interacting with other agents in a simulated environment. The goal is to understand how language is acquired in a natural setting, where learners are exposed to a variety of linguistic input and must infer the underlying grammar from this input.

MY PRESENT

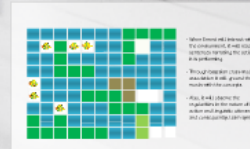


Two phenomena:

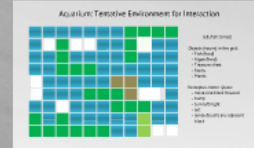
- Lexicon Acquisition
- Syntax Acquisition

Current Status

- Planning the environment and possible interactions
- Building the corpus of training sentences
- Studying algorithms for cross-modal associative learning (Hidden Markov Models)



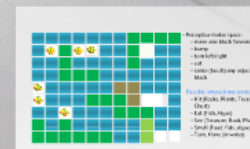
• Ernest can see and hear objects in the environment. It can also hear the agent's utterances.
• Ernest can move around the environment and interact with objects.
• Ernest can learn the location of objects in the environment and use this information to learn the meaning of words.



Acquaint: Tentative Environment for Interaction

Simulation of **grounded**, usage-based grammar acquisition in an artificial agent Ernest

Ernest
• A software agent that simulates language acquisition in a virtual world.
• Designed to learn a language by interacting with other agents in a simulated environment.
• Goal is to understand how language is acquired in a natural setting.



• Ernest can see and hear objects in the environment. It can also hear the agent's utterances.
• Ernest can move around the environment and interact with objects.
• Ernest can learn the location of objects in the environment and use this information to learn the meaning of words.

Theories of Language Acquisition

- Generativist Approach by Noam Chomsky
- Empiricist Approach by Lewis Carroll, MacWhinney, Tomasello
- Usage-based Grammar by Tomasello

Generative Grammar (Chomsky)

- Focuses on innate universal grammar which helps construction of a child's first language.
- Which is the basis, that grammar is then shaped by the input (learning of statistics of grammar).
- Structures, which are learned with no specific language are statistically learnt and modified into a target grammar.

Usage-based Grammar (Tomasello)

- Language acquisition is **learned through statistical observations**.
- Computational rules are specific to each language and are not language specific.
- Focuses on the **input** for language acquisition and not language specific.
- **Learned through usage**
 - Learning through usage
 - Statistical learning
 - Statistical learning through social communication

Simulation of Usage-Based Language Acquisition in Ernest

Mentor

Dr. Amitabh Mukherjee

Project Proposal by

Diksha Gupta

Theories of Language Acquisition

- **Generativist Approach** by Noam Chomsky
- **Empiricist Approach** by Bates, Powells, MacWhinney, Tomasello
 - **Usage-based Grammar** by Tomasello

Generative Grammar (Chomsky)

- Presence of **innate universal grammar** which imposes constraints on infant's hypothesis space.
- Which is to say, that grammar is not ontogenetically learnt (Poverty of Stimulus Argument)

Phonemes, words associated with to a specific language are contextually learnt and mapped onto an innate grammar.

Usage-based Grammar (Tomasello)

- Language acquisition as **context-driven statistical abstractions**
- Grammatical rules partly syntactic but also semantic: stated in terms of their pattern of use and communicative function
- Innate cognitive abilities for language acquisition are not language specific. Instead they are:
 - intention inference
 - pattern finding and
 - predisposition towards social communication

My Proposal

Simulation of **grounded**,
usage-based grammar acquisition in an
artificial agent: Ernest

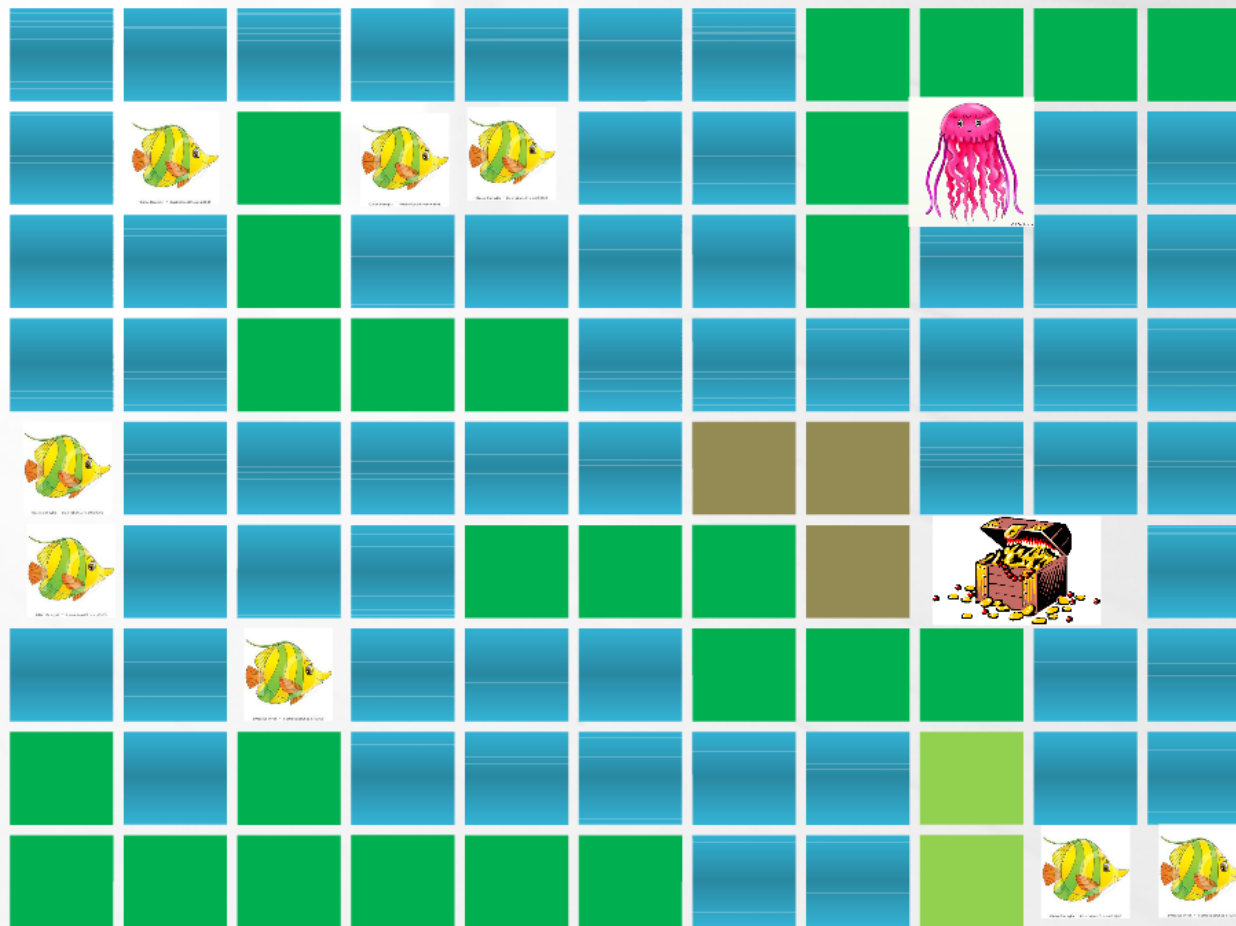
Ernest

Artificial agent simulated by Georgeon, Ritter

It implements emergent learning phenomena as theorized in natural organisms

It has intrinsic motivations and is capable of simulating course of action through inhibition.

Aquarium: Tentative Environment for Interaction



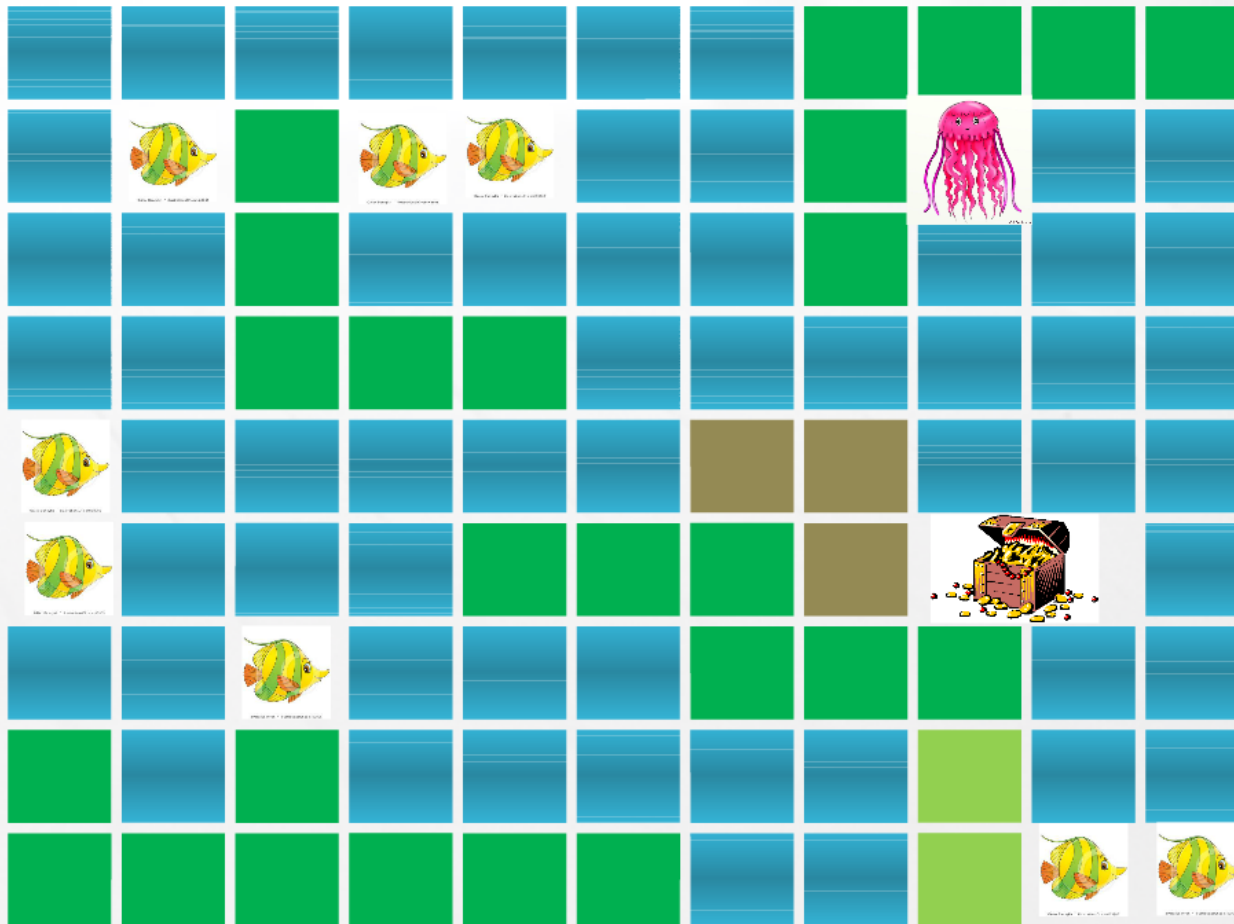
Jellyfish: Ernest

Objects (nouns) in the grid:

- Fish (food)
- Algae (food)
- Treasure chest
- Rocks
- Plants

Perceptuo-motor Space:

- move one block forward
- bump
- turn left/right
- eat
- sense (touch) any adjacent block



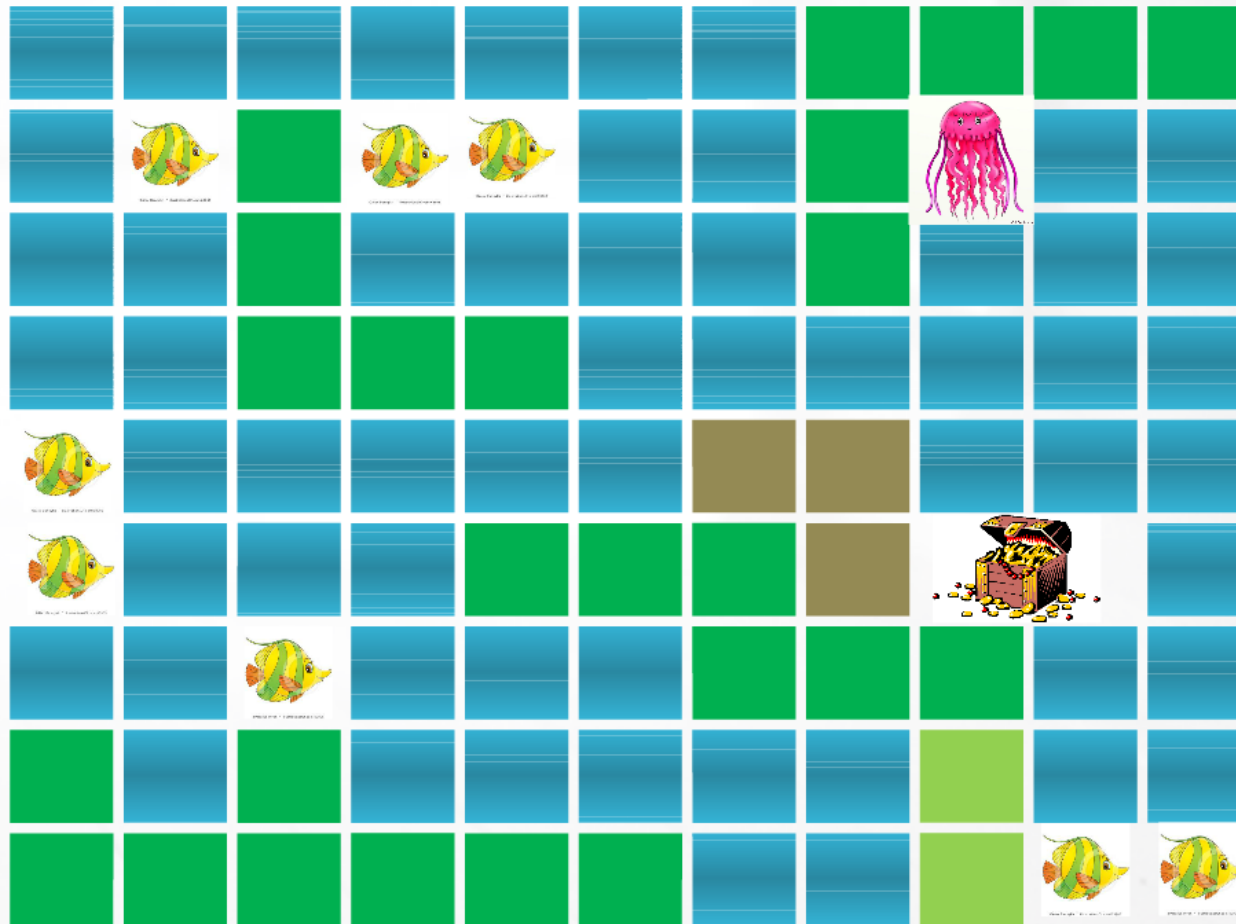
- Perceptuo-motor space:
 - move one block forward
 - bump
 - turn left/right
 - eat
 - sense (touch) any adjacent block

Possible interactions (verbs):

- Hit (Rocks, Plants, Treasure Chest)
- Eat (Fish, Algae)
- See (Treasure, Rock, Plants)
- Smell (Food: fish, algae)
- Turn, Move (in water)

Two phenomena:

- Lexicon Acquisition
- Syntax Acquisition



- When Ernest will interact with the environment, it will read sentences narrating the action it is performing.
- Through bayesian cross-model association it will ground the words with the concepts.
- Also, it will observe the regularities in the nature of its action and linguistic utterances and consequently, learn syntax.

Current Status

- Planning the environment and possible interactions
- Building the corpus of training sentences
- Studying algorithms for cross-modal associative learning.
(Hidden Markov Models)

References

- Georgeon, O. L., & Ritter, F. E. An intrinsically-motivated schema mechanism to model and simulate emergent cognition. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2011.07.003
- <http://e-ernest.blogspot.in/>
- Cowie, Fiona, "Innateness and Language", *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.)
- Tomasello, Michael. "First steps toward a usage-based theory of language acquisition." *Cognitive Linguistics* 11.1/2 (2000): 61-82.
- Cangelosi, Angelo, et al. "Integration of action and language knowledge: A roadmap for developmental robotics." *Autonomous Mental Development, IEEE Transactions on* 2.3 (2010): 167-195.
- Salvi, Giampiero, et al. "Language Bootstrapping: Learning Word Meanings From Perception–Action Association." *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42.3 (2012): 660-671.

Thank You!