Capstone I

Neo Alabastro

Neo Alabastro
NYU Shanghai — Undergraduate, 567 West Yangsi Rd, Pu Dong Xin Qu, China, 200124, email: neo.alabastro@nyu.edu

Contents

1 Introduction

This project will explore a single question: *How would artificial intelligence develop a constructed language?* To begin answering this question, we need other questions:

- What is a constructed language?
- What is language?
- What is the interplay between language and the human mind?

We must first understand what human language is before entering the realm of inhuman language (such as those for an artificial intelligence). To begin answering the above questions, let us work backward starting from question three.

A major point of linguistic debate about language is whether language is an innate feature of the human mind or whether it originates from somewhere or something external. To rephrase the dispute in terms of question three, does the interplay between language and the human mind originate innately or externally? The 2002 article "Simulated Evolution of Language: a Review of the Field" by Amy Perfors describes debate in the evolution of language and explores how computational work provides insight into its evolution. Perfors states that some stances about the origin of language argue that language is an innate feature, and that all humans have an innate faculty to acquire full fluency in any language. Other stances argue that language is a byproduct of general intelligence or just something that has adapted to our minds rather than our minds adapting to it (2.4). Perfors also argues that regardless of stance, "both… would agree that the ability to think in complex language helps develop and refine the ability to think" (3.17). Perfors supports the argument drawing upon the research of Derek Bickerton to conclude that: "[Humans] cannot have access to 'things in the world' except as they are filtered through our representation system: as Bickerton states, 'there is not, and cannot in the nature of things ever be a representation without a medium to support it in.'" (4.7).

Creating a representational system is a byproduct of the refinement of the ability to think; the refined ability to think then creates an effective instrument in which to represent and improve such a system that is based on representations—"knowing a language means being able to produce and understand new sentences never spoken before" (Victoria Fromkin et al., 5). In other words, that which lies beyond the representations may come to fruit.

From these arguments, a generalization of what language is for humans can be formulated: Language is a way for humans to organize consciousness—to organize one's being and existence. In effect, it is also the means through which humans communicate. These two facets of language—the first facet being an instrument to pare down the external forces of the sensory world into a kind of internal control, and the second facet being a multi-monocular stereoscope through which humans can convene to share these sensory experiences and delegate consciousness—are intertwined methods that manifest beyond the physical world. Not only can we use

language to perceive what is in front of us, we can also use language to grasp the unknown, the immaterial, the inhuman.

1.1 The Functions of Language

Language functions in two main ways:

1. As an instrument to construct one's inner and outer worlds.
2. As a systematic method to delegate and distribute perception.

1.1.1 As an instrument to construct one's inner and outer worlds

Language manifests as a system of symbols. These symbols are not in the literal sense of alphanumeric characters or drawings, but rather of symbolic concepts. to communicate and construct one's view of the world, one must use representations, qualified stand-ins, to even begin formulating it. After all, we quite literally cannot shove the entire world in our brain. By creating representations, the internal and external world is able to be categorized and manipulated in ways unimaginable. Describing one's internal emotions, like sadness or joy, can be expressed using the same words that can also describe external happenings or concepts like what one ate in the morning or the beauty of the Ko'olau mountains. Therefore, language requires representations defined by an abstract system—the world of vocabulary and grammar. This system dictates one's personal constructions of the inner and outer realities in which they are. External senses like touch, vision, and sound capture the signals of something's existence. Language lets us construct that something's being.

By this effect, this act of construction also buttresses a subconscious symmetry and balance that can rectify and pacify the chaos of the inner and outer, bringing a metaphorical homeostasis in which one's being can exist.

1.1.2 As a systematic method to delegate and distribute perception

These representations can be produced in a physical form: language is expressed in an analog format by producing sounds, gestures, or illustrations through actions like vocalization, signing, or writing (digital communication simply facilitates analog interaction). These methods use human organs like the hands, mouth, or face, to instantiate personal perceptions using a tangible form. Language exists because we can make it so. Therefore, language requires a means to transpose meaning in order for that meaning to be distributed or delegated, which in turn creates connection between others.

In the book *Flow: The Psychology of Optimal Experience*, Mihaly Csikszentmihalyi writes:

> When I say to an acquaintance whom I meet in the morning, 'Nice day,' I do not convey primarily meteorological information—which would be redundant anyway, since he has the same data as I do—but achieve a great variety of other unvoiced goals. For instance, by addressing him I recognize his existence, and express my willingness to be friendly. Second, I reaffirm one of the basic rules of interaction in our culture, which holds that talking about the weather is a safe way to establish contact between people. Finally, by emphasizing that the weather is 'nice' I imply the shared value that 'niceness' is a desirable attribute. So the offhand remark becomes a message that helps keep the content of my acquaintance's mind in its accustomed order. His answer 'Yeah, it's great isn't it?' will help to keep order in mine.
>
> — Csikszentmihalyi 129

The connection that language facilitates is abstract and implicit, even though words themselves are explicit representations of objects or ideas. A simple greeting is able to delegate and distribute one's homeostasis, affecting another person in the process.

1.2 Formulations

By nature, natural languages of human origin, or natlangs, are languages that are unplanned. That is, their origins are not necessarily traceable to a single point in time of conception where many people or one person gathered around and decided to create a language like Chinese or English. These languages have evolved and manifested alongside human evolution. Other examples of natlangs are languages like Arabic, Japanese, French, or Hawaiian. Also, natlangs do not need to just be verbal, as the myriads of sign languages in the world are also natlangs. Communities of deaf individuals, given enough time, develop a linguistic system that organizes consciousness in the two ways described in the previous section (Perfors 3.8).

On the other hand, constructed language, or conlangs, are a type of artificial, human creation. In the book *The Art of Language Invention*, David J. Peterson (18) writes that a conlang is:

> Any language that has been consciously created by one or more individuals in its fullest form…so long as either the intent or the result of the creation process is a fully functional linguistic system

Conlangs therefore fit into the aforementioned definitions of language because they are a fully functional linguistic system that humans can use. The entire linguistic system that is a conlang is oriented toward fulfilling different goals, like self-expression or human-computer interaction. All conlangs have their own raison d'être. Because conlangs are created with a purpose and intent, they have the liberty to take the

organization of consciousness to its extremes, constructing consciousness in novel or unnatural ways.

Since many generative AIs exist that are able to parse and manipulate language fluently, like ChatGPT, their training in natural language has already bootstrapped a significant part of the conlang process. A conlang cannot be created without the bedrock of a natlang. One cannot create a complex language without already having a complex language.

The boundlessness of the Internet has allowed worldwide contributions and broadcast decision-making to catalyze development of conlangs, fortifying the legitimacy of their presence as a meritable living language and not just child's play. And so, conlangs have become a communal enterprise. People rally behind a conlang, and have the possibility to contribute to the community to make it in some way, a piece of their own. Much debate about conlangs come from the communities that form them themselves. Discussions are often about the direction in which the language should move toward. Should we add new words? Should we change the grammar? Are these decisions congruent to the language's purpose?

These topics propose new questions about the scope of conlangs in the context of AI. The fact that conlangs live through speakers (just as any other natural language does), means that a single AI fluent in a natlang, like ChatGPT or Google's Gemini, that may have the capacity to generate its own conlang, may render the conlang that it generates dead on arrival. It may be able to use the conlang to construct its inner and outer world, but who is it to delegate and distribute its pieces of consciousness too?

1.3 Orientation

Conlangs are often categorized into three types: auxiliary, engineered, and artistic. As many things are, these categorizations exist along a spectrum. A conlang may fit exactly into one category or be in-between somewhere amongst them.

An auxiliary language is one developed to facilitate social order in some way, or as Peterson describes it, "a conlang created for international communication" (Peterson 21). Created in 1887, Esperanto is an auxiliary language that was developed with the intent of world peace. It was meant to be an international language, one that everyone could speak and through which could communicate effectively with everyone else.

The creator, L.L. Zamenhof, adapted features like word-building systems from Slavic and German semantics (Krunoslav Puskar, 108). This makes Esperanto an A Posteriori language, or a conlang "whose grammar and vocabulary are drawn from an existing source" (Peterson 22). Esperanto seeks to improve the delegation of consciousness and provides a simple, learnable platform on which to construct one's inner and outer worlds.

An engineered language is one like Ithkuil, the brainchild of John Quijada. Ithkuil is an A Priori language, meaning it has no connection or basis in any natlang, where

the grammar and vocabulary are not based on existing languages (Peterson 22). It is purely contrived. Ithkuil is a specialized method to delegate consciousness to a hyperfine degree. As the introduction to the text describing its grammar states, "the language uses a matrix of grammatical concepts intended to express deeper levels of human cognition more overtly, logically, and precisely than natural languages" (Quijada). Ithkuil has its own intricate and robust written script.

An artistic conlang is one like Toki Pona by Sonja Lang, which is also A Priori. Toki Pona contains only 120 words and 14 letters of the Latin alphabet. The theme of the language is semantic reduction, distilling thoughts into fundamental units, or as Lang puts it, "[it lets us] understand complex relationships in terms of their smaller parts" (7). Toki Pona is a unique instrument to construct one's inner and outer worlds, because, unlike Ithkuil, it "does not strive to convey every single facet and nuance of human communication" (8). Toki Pona also has its own written script, which takes the form of hieroglyphs.

The simulated language evolution experiment conducted by Perfors does not focus on methods of conlang production. The study used genetic algorithms, an approach to machine learning, to study the evolution of syntax and evolution of communication. Perfors says "Genetic programming has multiple advantages over other approaches to machine learning… it is strongly analogous to natural selection… [it also] can discover programs capable of solving given tasks in a remarkably short time" (7.5). At the time of this writing, the study occurred over 22 years ago. Generative AI has achieved a competence of mimicry indistinguishable to genuine human communication, and so, rather than developing algorithms that can be implementations of algorithmic processes to guide machine learning models, an AI conlang could be created atop an already fluent generative model. The conlang would sit atop the natural language processing that exists at the software level. This is a beneficial abstraction, because it allows the conlang to manifest in complicated forms, such as the three conlangs presented above.

2 The Whos, Whats, and Hows

Academia is unconcerned with conlangs. In my research, I have found no experiments or case studies that evaluate or explore conlang evolution using simulation, especially with the involvement of AI. Conlangs are often relegated to the boundaries of the closed drawer, hidden away from the external world—or in the words of J.R.R. Tolkien, "a secret vice". It is unfortunate that conlangs are relegated to these positions in academic spaces because they are instruments that unlock unlimited opportunities in exploring ways of constructing thought. Even though research focuses on natlangs, there is much to learn about the approach and implementation of technologies and experimentation techniques related to the evolution of language in general, which can then be applied to the study of conlangs.

2.1 Structures of Approach

This section details structures of approach that others in the field have developed to experiment with language evolution. These are general concepts rather than implementations. They are the "what" and not the "how".

The framework for conducting an experiment as this is called a simulation. Rather than going out and observing people use language over generations of time to verify a hypothesis and perhaps deduce a theory, computers are used to simulate such evolving processes to support a hypothesis and perhaps induce a theory. Angelo Cangelosi and Domenico Parisi state that "simulation models are implementations of theories, and as theories they aim at describing reality at some essential and necessarily simplified level because in science it is simplification that produces understanding" (13). Therefore, simulations of any kind of phenomena are bottom-up rather than top-down experiments; scenarios are created and then observed.

2.1.1 Generational gaming

The common procedure for simulating the evolution of language is to create circumstances, like a problem or conflict, to incite and catalyze such a long-term process by creating something called a language game, which can be played by both humans and machines. Luc Steels (343) describes a language game as a "routinized turn taking interaction". Steels then describes the simple premise:

> There is a shared cooperative goal… The speaker has a specific communicative goal, conceptualizes the world for language, and transforms this conceptualization into an utterance. The hearer must parse the utterance, reconstruct its meaning and map it into his own perceptual experience of the world. Games may fail in which case diagnostics and repair strategies are used by speaker and hearer

> to expand, adjust and align their language systems so that they may have more success in the future

This description has a few implications:

> "There is a shared cooperative goal…"

All participants in a language game share the same goal. This means that to attain a goal, some kind of systematic method to delegate and distribute perception is needed. This is language.

> "The speaker has a specific communicative goal, conceptualizes the world for language, and transforms this conceptualization into an utterance."

The speaker, or transmitter, uses the instrument of language to construct the inner world, their conceptualization of the communicative goal held in their mind, and their outer world, transmitting this conceptualization with a medium.

> "The hearer must parse the utterance, reconstruct its meaning and map it into his own perceptual experience of the world."

The hearer, or receiver, maps the utterances to their symbolic counterparts, effectively using language to construct a bridge between their outer and inner world.

> "Games may fail in which case diagnostics and repair strategies are used by speaker and hearer to expand, adjust and align their language systems so that they may have more success in the future."

Failure is determined by whether or not the aforementioned shared goal has been achieved.

Language games facilitate a type of learning called iterated learning. Iterated learning is "the process by which a behavior arises in one individual through induction on the basis of observations of behavior in another individual who acquired that behavior in the same way" (Simon Kirby et al., 108). Iterated learning is the method through which an individual acquires their native language. So in a sense, the game of real life is the language game for humans.

It is worth noting that the book section "Language Evolution with Deep Learning" by Mathieu Rita et al. writes of the term communication games, which are defined as "a framework used to investigate how perceptual, interactive, or environmental pressures shape the emergence of structured communication protocols" (2). For the purposes of this project, I will use the term language games instead. The term "communication" defines too broad of a discourse space, since communication can span to multiple domains of communication, not just linguistic communication, such as the

gestures of the great apes which indicate intention or attention (Michael Tomasello, 22).

2.1.2 Generational learning

In the setting of a computer simulation: In one generation, a predecessor agent teaches a successor agent, in the same way a parent teaches their child a language. After this teaching, the generation concludes. A new generation starts and the successor, or the child, from the previous generation becomes the predecessor, and now teaches a new successor agent. This generational sequence continues an arbitrary number of times. There does not need to be a single length generation lifetime for a predecessor. Predecessors one to n generations removed can continue to exist in the same generation as the present generation successors. Such variation in lifetime was conducted in experiments conducted in the paper "Understanding Language Evolution in Overlapping Generations of Reinforcement Learning Agents" by Lewys G. Brace and Seth Bullock

The advantage of generational learning is that a simulation can simulate several hundred or thousand generations of iterations of a single language in a short amount of real-life time. This is useful when conducting experiments observing broad changes over long time spans. This project is concerned with the creativity aspects of a conlang as manifested by LLMs. There is no true numerical measurement to be collected or observed by the conlang's change over time, but rather the changes that manifested in the conlang: the final continuous product. Therefore, multi-generational agent learning will not be a focal point of this project.

2.1.3 Agent interactions

The article *Bias Amplification in Language Model Evolution: An Iterated Learning Perspective* by Yi Ren et al. explains an approach to language evolution using iterated learning. Learning can be done with just two agents, and consists of three phases (3):

1. **Imitation**: an ignorant agent learns something from a predecessor.
2. **Interaction**: the same agent uses this new knowledge to execute a task (often conducted as a language game).
3. **Transmission**: after completing the task, the agent generates useful data for the next generation of ignorant agents.

To recap, language evolution experiments are often conducted using computer simulations that represent simplified versions of realistic variables. Simulations have a beginning and end, and are divided up into a certain number of epochs. In each epoch, one or multiple instances of a language game takes place. The language game is the scenario that is created for agents in the simulation to exercise the theory that is being tested. Language games facilitate iterated learning. The process of iterated learning is what generates useful data for the next epoch to instantiate the next instance of the

language game. An arbitrary number of epochs may take place before the simulation's termination.

2.2 Implementation of Structures

Before significant advancements in portable processing power in computers, genetic algorithms have been used to simulate language evolution. Algorithms do exactly as stated and are not flexible in terms of high levels of noise in signal to noise ratios. Neural networks (NNs) on the other hand are less terse and rigid, but still lack the agility of algorithmic processes that move quick because they require more computing power. Since neural networks are based on models of the human brain, "they can focus on the influence of both cognitive and neural mechanisms on language development and evolution" (Cangelosi and Parisi, 401). This trait makes NNs favorable choices when modeling smaller levels of detail in language like the complex entwining of phonology and morphology, especially because they are significantly more resilient to noise in data processing. Faults and errors in a simulation have less opportunity and less probability to propagate through generations. However, genetic algorithms can be used to enhance NN neuron evolution. Cangelosi and Parisi used a genetic algorithm to evolve the connection weights of an agent's neural network (404).

2.2.1 Coming into being

The book section Language Evolution with Deep Learning also provides a significant case study into the design of linguistic deep learning models in a simulation, from high level overviews of deep learning concepts to basic implementations. One of those implementations is the communicative agent.

The entities that talk to each other in a language are agents. One type is a communicative agent (CA), which is an agent that has faculty for communication behaviors and interpretation of meaning. CAs are composed of functional modules and an internal map of systematic representations that act like a switchboard for each module's operations (Rita et al. 5):

1. **Perception module**: maps observations of an environment to an internal representation.
2. **Action module**: maps actions to internal mappings.
3. **Generation module**: maps internal mappings to a message.
4. **Understanding module**: maps a message to an internal representation.

CAs fit into a sender-receiver framework. The aforementioned paper by Brace and Bullock explores the concept of this framework using reinforcement learning, in order to "offer insights into how lexical items become established within the context of the cultural evolution of human language in structured populations with overlapping generations" (1). CAs can send and receive lexical items, like words, within a particular

circumstance, like in a problem-solving situation, and amongst or between a milieu, like in an environment where CAs can communicate with one another.

2.2.2 Putting the pieces together

A typical simulation that uses iterated learning is to use CAs. These CAs can exist in a simulated environment.

One experiment done by Ren et al. was to make LLM agents—such as various GPT models, Claude, and Mixtral—"infer and generate [a] shared rule by summarizing several input-output pairs". One generation is given a list of three pairings of input and output, where the input is understood to equate the output. The agent would then have to induce the underlying rules of what the output would be given a new type of pairing with the same inputs in a different configuration. The simulated world or another LLM would give feedback to the initial agent, explaining whether the submitted rules are wrong or not, and asking them to refine it again (6). Similar to what I will be proposing, this paper uses already-trained LLMs as agents in a simulation.

Another experiment done by Rita et al. is a visual discrimination game using reinforcement learning. A sender sees an image and tells a receiver. The receiver then must guess what the image depicts amongst a set of other images. The original image is revealed to receiver and both the sender and receiver are rewarded according to their performance (16). This experiment does not use LLMs.

Furthermore, an experiment done by Brace and Bullock aggregate multiple agents from multiple generations in a single context. Agents are categorized by either being mature or immature. Mature agents have existed for more than one generation. Immature agents have only existed in the current one. Immature agents learn by observing the outcome of a language game in which two mature agents were the participants. After two generations, the mature agent is removed from the simulation. In each generation, agents in the immature group interact amongst themselves after observing the mature group interact. The experiment found that success of evolution is proportional to the number of participants and the more those participants interact with each other (495). This experiment also does not use LLMs.

All such experiments presented above use English or a morphology and syntax that is A Posteriori. Also, any particular interaction that happened follow the interaction framework explained earlier.

3 Project Concept

Much of the content so far has drawn upon technical research, or research that is strictly numerically quantified and analyzed through the language of mathematics and LaTeX. Artistic capacity can only be a byproduct of what is presented and proved. Given that this project is not a technical endeavor, this project will focus on the artistic and philosophical. The technical background and implementations are the means to the end of creating art.

3.1 The Goal

So, the goal of a project like this is not to test and validate whether an AI can create a constructed language. Asking an LLM to create a conlang is a straightforward affair, and there is no doubt that with the right prompt crafting, a workable result can be achieved. Rather, this project will explore language's infinite capacity to symbolize, founded upon new technologies that enable unfathomable, novel expressions. This process will serve to demonstrate language's divergence from the initial layman conceptions as a sole form of human communication, to an emergent process that supplants physical reality in its various dimensional mediums. I seek to enmesh two polar opposite forms of linguistic disciplines, one that is often for the sake of language, that is the creation of conlangs; and the other that is for the sake of seeking the truth, the scientific linguistic explorations of language evolution using cutting-edge technologies and techniques.

3.2 Philosophical Considerations

In the design of this project, it is crucial to remember that language games are not something humans intentionally engage in in the real world. Instead of being driven by the simple structures of reinforcement learning, such as identifying an image correctly then receiving a reward, we have extrinsic motivations like wealth and food, as well as intrinsic ones like self-integrity, moral goodness, and just a drive to satiate a need to have a good time. Intrinsic motivations are more abstract and can be less logical, and one can argue, less "scientific". Furthermore, these intrinsic motivations bring about philosophical and psychological questions too, such as, why do we need to have drive and have a moral compass that dictates our actions? How does the manipulation of the truth, the matter of fact and the objective, become an instrument for manipulation, like lying? Those questions direct the way we use language, the way language interacts with us, and most importantly, the way language evolves.

And here is a point of philosophical contention for a project like this one. To explore the "infinite capacity to symbolize", those that do the symbolization, the agents, need a reason to symbolize. If they cannot symbolize or have no reason to do so,

there exists no simulation. Of course, the previous experiments did give a need for the agents to symbolize by making them participate in language games. However, human symbolizations are not just correct associations. They are associations that are correct for a reason. One reason an association is correct is because most speakers agree that that association is true, the ultimate defacto. A cat is a cat because we all agree that that creature has a map-bound representation to the word and its meaning. However, when it comes to conlangs, that reason for a symbol's meaning or existence is not as linear. Interlocutors in online communities can argue over what direction to take a conlang in, which means conlangs have the abstract element of intentionality. Take for instance, a conlang that can symbolize objects, things, or concepts as numbers. The intentionality is the philosophical approach to express the world in a numerical way that is not confined to the logics of mathematics. To simulate such developmental and evolutionary complexity, I argue that language games are not enough for a conlang simulation, especially when using LLMs. A surrounding reality must be simulated too.

3.3 Fundamental Parameters

Here are the fundamental parameters for the implementation of this project:

1. A continuous, simulated world must exist for the agents.
2. The agents must be communicative agents.
3. The agents must have a purpose.
4. The agents must have a constant stream of perceptual input.

3.3.1 A continuous, simulated world must exist

A continuous world means that there exists time and change. Previous events have effects on current and future events. The temporal scale of the world is infinite and there exists no end. A simulated world means that the world has a "physical" dimension. That means there is already a constructed materialization of the world in some form that can be perceived in some way by an agent. Theoretically, an agent also has mobility to traverse this materialization. In such a continuous and simulated world, the question an agent must be able to answer is not just "what do you see?", but also "how do you see it?".

3.3.2 The agents must be communicative agents

As explained, the agents must have all the faculties of a communicative agent. Fortunately, generative AIs like LLMs can combine many of their faculties into a unified package.

3.3.3 The agents must have a purpose

Imbued in each agent is a purpose. This purpose is defined from the outset. This purpose can also be an assigned job or a goal that an agent must execute for some particular purpose in the simulation.

3.3.4 The agents must have a constant stream of perceptual input

The agents must not just be able to communicate but must also be able to perceive continuous changes in the environment on a change-by-change level. For example, if the simulation states that "it is now nighttime", in circumstances that may permit an agent to be aware of a nighttime, the agent must be able to receive this information exactly and immediately. If an agent "dies" in a simulation, this news must also be broadcasted and received. This constant perceptual input is also necessary if an agent is to traverse an environment.

3.4 Implementing a World

Here are two ways in which to implement a world for these agents to exist inside.

3.4.1 A simulation completely removed from the physical world

A simulated world could be generated from a seed. This seed is a prompt, such as:

> You are in a room surrounded by chairs. There is window that shines light upon the chairs. Your interlocutors sit across and around you in those chairs. The day has just started. Outside there are fir trees and an axe in a stub. The room has distinct details, it is twenty-by-twenty meters in length and width. It is ten meters in height.

Each agent is given this seed at the beginning. Further details of the world can be included, such as the name of the world, a simple description of time of day.

As time progresses, the state of the simulation is sent to each agent. For example, if the simulation is to express a season change, there must be a description without explicit labeling, such as:

> The leaves on trees fall to the ground. They take on a different hue.

Notice that there is no labeling for the color of the leaves, but rather just the objective fact that they are now a different color. The labeling is in the hands of the agents.

Relationships between agents must also be defined to have a conversation. Relationships instantiate power dynamics and interaction presets. It defines how the agents are to use language with each other and develop a language with abstract human concepts of the self and the other.

3.4.2 A simulation that is a hybrid of our world and their own

Opening up the simulation to include our reality is another possibility into incorporating a world in which agents can interact. Using a device, like a camera or a microphone, to intercept real world signals and translate them into a form perceptible for an agent may be feasible. Time then is dictated by real life time, and events are dictated by real life events. Perhaps someone moves in front of the camera, something is said by the microphone, or perhaps there is a change of scenery in front of the camera. Regardless of input method, the same distribution of information to the agents must be considered and implemented.

Creating a simulation that is void of the real world means a sanitized view of a conlang's development. The viewer of an experiment can truly take a backseat and simply observe. Bridging the simulation into the real world means the viewer can also become a participant in the simulation.

3.5 Implementing a Purpose

A conlang has a purpose and for that purpose to manifest or have legitimacy, the agents, those who are developing the conlang, must have an intrinsic reason. This intrinsic reason is more of a pseudo-intrinsic reason, because it is not developed by themselves, but is instead imbued by an initial prompt. Here is a prompt that imbues purpose:

> You are a researcher looking to improve the efficiency and communication of language. You work at a laboratory along with ten other colleagues that experiment with the many possibilities of expression.

With this prompt, the agent is "primed" and will generate output accordingly against further input. This given purpose of being a researcher orients the approach of the agent way in which it decides how a conlang should or could evolve.

3.6 Implementing a Goal

A communicative goal is required for the agents to provide focal points of discourse. The previous example stated that the reason for research is "to improve the efficiency and communication of language". Another possible goal in another simulation could

be “to develop a way to communicate complex ideas in a compact form”. The agents will have work together to achieve such goals through conversation.

3.7 Implementing a Conversation

Agents speak English in conversation and perceive the world in English.

Since all agents have the capacity to generate thoughts and content based on preconceived inputs, each will have content that can be developed into arguments for a conversation. The conversation is the engine that produces the creative output. In some fashion, agents must be able to “meet” each other. This will create a circumstance in which agents have opportunities, hallmark to iterated learning processes explored earlier, to imitate other agents, interact with other agents, and transmit or teach data to other agents.

Meeting can occur individually or in groups. Conversations take place between two agents, a sender and a receiver. A group conversation does not necessarily mean that all the agents have the possibility to be senders. It just means that there is more than one receiver. Of course, the agent sending and the agent receiving is a matter of circumstance and social role.

Without making the intricacies of social dynamics too complex, a simple approach to have all agents on the same page is to have all congregate in the same virtual space. In other words, all of what is said by each agent is broadcasted and received by every other agent present. This approach is feasible if all of the agents are primed to be researchers and convene in a meeting.

In such a meeting, there would be an odd number of participants for the purposes of deal breaking. The agents would have to agree upon what parts of the conlang to evolve and why it matters. The meeting can be over after a certain period of time or if an agreement is unanimously accepted or if a majority consider a certain evolution the best path. After a meeting ends, one iteration of the conlang is established. The next iteration comes in the next meeting. Meetings are similar to the idea of a language game, however, the difference is that the purpose is to develop a language in alignment with a purpose regarding external and internal agent circumstances. Multiple variables are in action and affect the outcome, rather than a discrete and controlled sender-receiver dynamic.

3.8 Implementing the Conlang

The conlang is the parameter for agreement in a conversation. The conlang must have a purpose and type of categorization amongst the main three presented earlier: engineered, auxiliary, or artistic. Regardless of conlang type, the conlang needs an orthography, visualization, or vocalization.

3.8.1 Method 1: Creating a conlang from scratch

This method is the most involved and has more leeway for intent and is dependent on the goal. A possible conlang that can be created using technology as an extension of human capacity for language is one that utilizes color.

As mentioned earlier, conlangs like Ithkuil and Toki Pona have their own original orthography. Esperanto has orthography, however, it borrows from the Latin script to render morphology and express phonetics. Ithkuil and Toki Pona's orthography are on the extreme sides of differing complexity. Ithkuil has an orthography that is compact and highly saturated in meaning. Toki Pona has an orthography that is simple and descriptive. A color conlang would activate a new visual cue that adds more information to a symbol.

3.8.2 Method 2: Developing an existing conlang

Using an existing conlang like Toki Pona or Ithkuil means more constraint on what the goal given to the agents would be. The goal would then become the actual goal of the original conlang. The simulation would then reorient itself from developing a unique take on conlangs to one that shows how a conlang can be extended into new realms using digital signals.

Using an existing conlang as a starting point will allow more time to develop the technical aspects of the project, like the CAs or visualizations. The conlangs explained in the introduction and background section all provide useful starting points. Esperanto's use of Latin orthography can make the writing easier to evolve. Toki Pona and Ithkuil present interesting possibilities in terms of further developing a new orthography. Something inherent about these conlangs though is that they are for humans by humans.

3.9 Implementing Human Perception

Humans must be able to perceive the conlang in some way. Since this is a digital medium, there are many avenues to explore the different sensory inputs of language expression. For example, translating a color language into something we can perceive is to use hexadecimal strings as writing and the "vocalization" as an LED light source that expresses that exact color representation. Another would be to use generative voice models to generate vocalizations that are impossible for humans to replicate with our physical vocal tract, but can nevertheless still be heard and perceived by us. Only a machine can reproduce the signal.

3.10 Construction in Context

Where do these pieces fit in relation to previous experiments?

3.10.1 No generations

The predecessors and successors in generational learning as explored by previous language evolution experiments are not present in this simulation. Rather than passing the torch of the language between generations of mature and immature agents, all agents have the same competency of the conlang. The variable that changes is the environment rather than the agent.

3.10.2 Micro-level language games

In this simulation, the overarching shared cooperative goal is to develop the conlang. At the conversation level, the language game depends on the type of conversation being held. There can be multiple types of conversation templates defined, such as, one in which the conlang is used or one in which the conlang is simply discussed. Conversations provide context, history, and data on which to further the evolution of the conlang.

One possible palpable shared goal that the agents may work toward together is translating a text from English into the conlang.

In conversation, agents describe their perception of the world to each other by first using English, then by having constructive discussions about how the interpretation of the world can be translated. Each agent creates an internal guide to the conlang. If an agent is to make an error when producing content in the conlang, other agents may proceed to correct the offending agent based on their own guide. The agent must then correct the output. Anytime the conlang is used, an agent must verify if its internal mapping of the language maps similarly to the other agent's. This correction interaction is an example of a repair strategy.

4 Technicalities

The engine of the simulation is a monolithic apparatus. This apparatus, which I will refer to as the facilitator, manages the state of the simulation, directs communication channels, and, depending on agent implementation, translates the possible output of the conlang to external interfaces for us to perceive.

The facilitator is one part server, one part clock, and one part generative source. It is a server because it routes message data to agents; It is a clock because it keeps track of time, whether it be for conversations or for the simulation of a virtual world; and it is a generative source because it generates situations and circumstances for the agents to take part in. For example, if the world is entirely simulated, the facilitator would, based on time passed and a dash of creativity, determine that the season is now changing. It would then broadcast this circumstance to agents. The facilitator also keeps track of the conlang, storing any type of textual or media progress in a database.

Agents are CAs, and therefore have a perception, action, generation, and understanding module, as well as an internal mapping of meanings. The perception module is an interface that can accept any type of media data, like text or images. The generation and understanding module takes the form of an LLM. Input can be fed into the LLM as serialized text or image data. Output from an LLM can then be transmitted using an action module, or fed into another generative module that can transmute text into some form of multimedia, like image or sound. The internal mapping of meanings can be a simple SQLite or a lightweight NoSQL solution. Vector databases are irrelevant.

Since these agents are LLMs, communication is implemented as textual data. These agents communicate in messages. Messages consist of a sender, a receiver, and content. The sender is the originator of the message, the agent that wrote it, and the receiver is who exactly the message is for. Given the environment circumstances, the facilitator may distribute to message to one or n participants in a conversation (for example, if all agents are in the same shared virtual space).

4.1 Programming

Python and Go will be used to program agents and the facilitator. Python has a rich ecosystem of machine learning libraries, and Go is a solid general purpose programming language for networking, containerization, and concurrent processes. There are a few protocols that can be used to send data between agents, like HTTP with JSON, however one of interest that has robust realtime capabilities is protobuf. A CA's perception and action modules would send and receive data in this format.

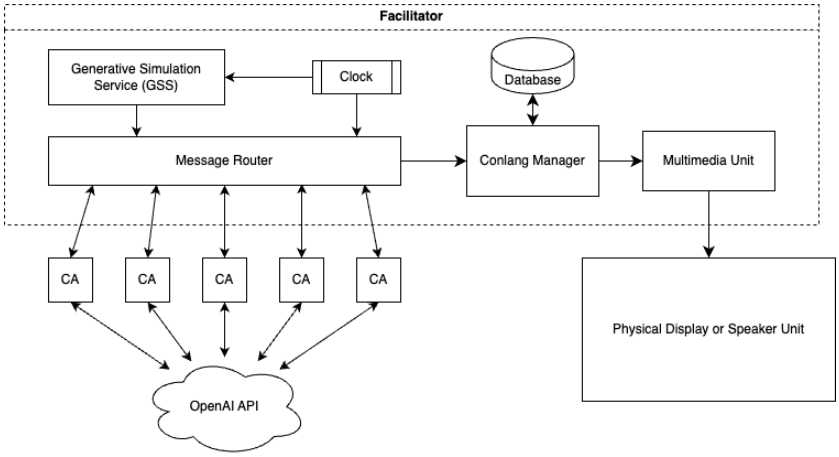

Fig. 1: A possible method for developing the simulation.

Working with external interfaces to display text can be done using a JavaScript based web framework, or using the Nannou creative coding library written in Rust. Nannou is efficient and can run real time visualizations with ease.

4.2 Connections

Docker containers can be used for the agents and facilitators. Communication to interfaces that display data to the physical world would communicate solely through an exposed API of the facilitator. Data to an external LLM source like OpenAI's API for queries can happen on a per-container basis, so as to lessen the burden of facilitating every network interaction through the facilitator.

With this setup, a single computer can be used to run the entire project. An external monitor can be used to project the state of the simulation, which is dependent on whether the simulation is self-contained or connected somehow to physical reality; the state of every agent, such as their conversations; and any data related to the conlang, like a dictionary, various translations, or orthographic representations.

Alternatively, multiple microcomputers like Raspberry Pi's can be used to establish the presence of the agents in real life. These can then be routed via WiFi or Ethernet cable for communication. Each device could have a physical output to a monitor that displays information. If the simulation is a hybrid between digital and physical, then other sensory interfaces for the device can be added as well.

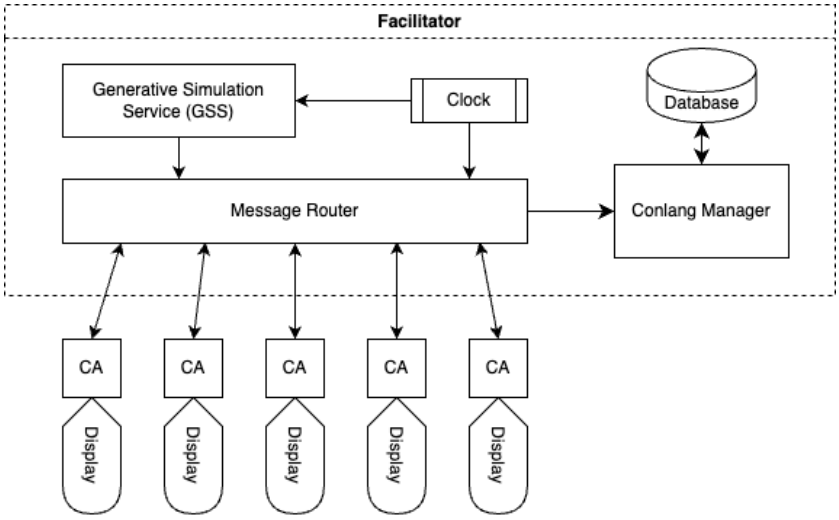

Fig. 2: A method in which each CA has its own output.

4.3 Participants

The total number of participant agents should not be less than three. Having an extremely large pool of participants, like a number of 1001, creates increasing complexity that I find difficult to properly manage. I think five or seven is a healthy, manageable number of agents. Research has shown more agents means more successful evolution, but the agents and environment are more complex. The lack of population breadth is made up for in individual depth.

4.4 Expression

Here is a miscellaneous collection of ideas that can be built upon the technical implementations.

4.4.1 Color language

Color language with form. Color position, and where the color, hue, value, lightness, etc. is an opportunity for development of morphology and syntax. Shape and form may take also add more to the dimension of expression for these constructs. It may be expressed in 2D or 3D format, however, given the nature of LLMs, a 2D ASCII-like text format may be of interest, similar to the random art generated in SSH key generation command line applications. These units of pictograph can then be transformed into meaning through sentences, or one pictograph can express an entire sentence.

4.5 Problems and Pitfalls

4.5.1 Global synchronization

Keeping all the agents in sync according to a global time will be difficult, and may create agent arguments.

4.5.2 Erroneous validations

Generative AI can have a tendency to hallucinate, producing invalid nonsense content. Hallucinations may occur in the process of generating conlang content. As previously mentioned, a mitigation technique to stop runaway hallucinations is for the agents to validate any output of the conlang.

Bibliography

Amy Perfors. "Simulated Evolution of Language: A Review of the Field." *Journal of Artificial Societies and Simulation*, vol. 5, no. 2, Mar. 2002.

Angelo Cangelosi, and Domenico Parisi. *Simulating the Evolution of Language*. Springer London, 2002.

David J. Peterson. *The Art of Language Invention : From Horse-Lords to Dark Elves, The Words Behind World-Building*. Penguin Books, 2015.

J.R.R. Tolkien. "A Secret Vice." *The Monsters and the Critics and Other Essays*, Allen & Unwin, 1983, https://search.worldcat.org/title/9945117.

John Quijada. *Ithkuil*. https://www.ithkuil.net/.

Krunoslav Puskar. "Common Criticism of Esperanto: Facts and Fallacies." *Język. Komunikacja. Informacja*, Oct. 2015, pp. 85–116, https://repozytorium.amu.edu.pl/handle/10593/4078.

Lewys G. Brace, and Seth Bullock. "Understanding Language Evolution in Overlapping Generations of Reinforcement Learning Agents." *Proceedings of the Artificial Life Conference 2016*.

Luc Steels. "Modeling the Cultural Evolution of Language." *Physics of Life Reviews*, no. 8, Oct. 2011, pp. 339–56, https://www.sciencedirect.com/science/article/pii/S1571064511001060.

Mathieu Rita, et al. "Language Evolution with Deep Learning." *Oxford Handbook of Approaches to Language Evolution*, Mar. 2024, https://arxiv.org/abs/2403.11958.

Michael Tomasello. *Origins of Human Communication*. The MIT Press, 2008.

Mihaly Csikszentmihalyi. *Flow : the Psychology of Optimal Experience*. 1st. ed., Harper & Row, 2008, https://willamette.on.worldcat.org/oclc/232961487.

Simon Kirby, et al. "Iterated Learning and the Evolution of Language." *Current Opinion in Neurobiology*, edited by Michael Brainard and Tecumseh Fitch, no. 28, 2014, pp. 108–14, https://doi.org/http://dx.doi.org/10.1016/j.conb.2014.07.014.

Sonja Lang. *Toki Pona: The Language of Good*. 2014.

Victoria Fromkin, et al. *An Introduction to Language*. 10th ed., Cengage Learning Australia Pty Limited, 2022.

Yi Ren, et al. *Bias Amplification in Language Model Evolution: An Iterated Learning Perspective*. Oct. 2024, https://doi.org/https://doi.org/10.48550/arXiv.2404.04286.