A COMPUTATIONAL MODEL ON THE EMERGENCE OF CANONICAL WORD ORDER IN LANGUAGE AND TRANSFER THROUGH GENERATIONS OVER DIFFERENT COMMUNICATIONAL STRUCTURES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

BURCU ÇINAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCES

SEPTEMBER 2022

A computational model on the emergence of canonical word order in language and transfer through generations over different communicational structures

submitted by **BURCU ÇINAR** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Sciences Department, Middle East Technical University** by,

Prof. Dr. Deniz ZEYREK BOZŞAHİN _____
Dean, **Graduate School of Informatics**

Dr. Ceyhan TEMÜRCÜ _____
Head of Department, **Cognitive Sciences**

Prof. Dr. Hüseyin Cem BOZŞAHİN _____
Supervisor, **Department, School**

Dr. Rabia ERGİN _____
Co-supervisor, **Department, School**

Examining Committee Members:

Prof. Dr. Committee Member 1 _____
Department, School

Assoc. Prof. Dr. Committee Member 2 _____
Department, School

Assoc. Prof. Dr. Committee Member 3 _____
Department, School

Assist. Prof. Dr. Committee Member 4 _____
Department, School

Assist. Prof. Dr. Committee Member 5 _____
Department, School

Date: 28.08.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Burcu ÇINAR

Signature :

ABSTRACT

A COMPUTATIONAL MODEL ON THE EMERGENCE OF CANONICAL WORD ORDER IN LANGUAGE AND TRANSFER THROUGH GENERATIONS OVER DIFFERENT COMMUNICATIONAL STRUCTURES

ÇINAR, Burcu

M.Sc., Department of Cognitive Sciences

Supervisor: Prof. Dr. Hüseyin Cem BOZŞAHİN

Co-Supervisor: Dr. Rabia ERGİN

September 2022, 24 pages

English abstract here

Keywords: canonical word order, emergence of word order, communicational networks, computational model, iterated learning, bayesian learning

ÖZ

DİLDE KANONİK KELİME DİZİLİMİNİN ORTAYA ÇIKIŞI VE FARKLI İLETİŞİM YAPILARI ÜZERİNDEN NESİLLER ARASI AKTARIMI ÜZERİNE BİR HESAPLAMALI MODEL

ÇINAR, Burcu

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Prof. Dr. Hüseyin Cem BOZŞAHİN

Ortak Tez Yöneticisi: Dr. Rabia ERGİN

Eylül 2022, 24 sayfa

Türkçe öz buraya

Anahtar Kelimeler: kanonik kelime dizilimi, kelime diziliminin ortaya çıkışı, iletişim yapıları, hesaplamalı model, tekrarlamalı model

Dedication here

ACKNOWLEDGMENTS

Acknowledgments here

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S	Subject
O	Object
V	Verb
ABSL	Al-Sayyid Bedouin Sign Language
CTSL	Central Taurus Sign Language
ISL	Israeli Sign Language
NSL	Nicaraguan Sign Language
L1	First Language
L2	Second Language
EM	Expectation-Maximization

CHAPTER 1

INTRODUCTION

Most of the sentences we use today are consciously or unconsciously formed in a certain word order. Many languages are sensitive to word orders, and a canonical word order is important as it includes most of the meaning of the sentence spoken as well as different orderings of it can lead to other meanings. For example, the knowledge of who did what to whom can be a knowledge learned by ordering the words appropriately. There is a huge difference in 'man kisses woman" vs. "woman kisses man". The canonical word order of a language is defined by the order of the three main components of basic transitive sentences: subject (S), verb (V), and object (O). Subject is the one who does the action, verb is the action to be done, and the object is the one who is affected by the action. Logically, we can order these three items in 6 different ways: SOV, SVO, VSO, VOS, OVS, OSV. For example, while the dominant word order for Turkish is SOV, for English it is SVO. In addition to these, there are also languages that do not conform to a certain basic element sequence. We can call this as free word order languages. But these seven possible word orderings are not evenly distributed amongst the world languages. According to [1] 41% of languages use SOV as their cannonical word order, like Turkish, Korean, Persian. 35% of the languages use SVO as their canonical word order, like English, Chinese, French. 7% of the languages use VSO, 2% of the languages use VOS, 0.8% of the languages use OVS, again, 0.3% of the languages use OSV as their canonical word order. Also, 13.7% of the languages use free word order to communicate. In these languages there is also a dominant word order, but other orderings are also possible to convey a true meaning while communicating. So the overall distribution looks like this: (SOV, SVO) > free order > VSO > (VOS, OVS) > OSV. However, where does this canonical orders and the agreement in a certain order came from is still a matter of curiosity. To explore this, we need to go back a very long time, to our ancestors which is impossible with today's facility. Although experimental studies on the subject give various ideas, a holistic approach is not satisfactory due to the comprehensiveness of the research.

As mentioned before, it is not possible to observe the exact stages of the languages spoken in the world from their emergence to the present day. Therefore, in the researches on the subject, mostly the findings obtained from the young emerged sign languages (village sign languages like ABSL, CTSL and NSL) or the silent gestures (or pantomimes) of the speakers were used. SOV and SVO are also predominant word orders in sign languages according to [2]. This comparative study of 42 sign languages suggests that cognitive/communicative biases are involved in determining the dominant order in a language. These studies mostly done on basic sentences, complex sentences need more deep researches and experiments.

In this study we too only cover basic transitive sentences with three components (subject, object and verb). These transitive sentences also split into two kinds of sentences: reversible and irreversible

transitive sentences. In reversible sentences it is semantically logical and possible to change the places of object and subject in the sentence, i.e. "the girl pushes the boy" could be "the boy pushes the girl" depends on the situation. On the other hand, irreversible transitive sentences does not allow us to change the places of subject and object semantically, i.e. "the girl cuts the paper" cannot turn into "the paper cuts the girl".

One of the explanations from [3] is that the observed word order frequencies may be the result of genetically encoded biases towards certain orders, as part of the universal grammar hypothesis. This could be a possibility, some word orders may be easier for us from birth. [4] suggests that the prevalence of SOV order across the world's languages may arise in part because SOV order is most compatible with how we conceptually represent transitive events. Also, [5] proposes that all living languages today descend from a single common ancestor, a proto-language uses the SOV word order. Thereof, it is possible that SOV and SVO have a special cognitive place.

[6] tries to explain the current frequencies of languages with three principles: "theme-first-principle", "verb-object-bonding", and the "animate-first principle". The frequencies are proportional to the number of principles they realised. According to [7] these three principles are realised in SOV and SVO, two are realised in VSO, one in VOS and OVS and none in OSV. It is also seen that the animacy level of subject and object also affected how we structure the sentences in many studies [2][8][4][9] [10][11], which will be explained in the literature review section in detail.

1.1 Research Questions

The current studies proposes some explanations to different questions, like why some orders more preferred than others, which are mentioned before, how network communication frequency and network type helps to make a language more systematic [12][13], how different generations behave while transferring the language by iterated learning[14], and so on.

There are several questions this study seeks to answer. These questions also serve as a source for the creation and testing of the model. First, previously studies suggest that SOV is cognitively most basic word order to start a language. But what if this is not the case? If we start our with another word order bias will the word order converge to one of the most frequent orders today, SOV or SVO? Second, how does size of the community and different network structures affects the dominant word order emergence and evolution speed? We know that not all people are the same. We don't learn the same way, or live the same way. Therefore, there should be an effect of if. We want to observe where the margin of error is different and how these differences will affect the word order of the model. So, what will it change if there are many/few people who change the language? And what does it tell us about real life? Will word orders remain fixed without language changers?

1.2 Contributions of the Study

This study was carried out with the hope of filling a few gaps in the literature. Although there are experimental studies and explanations on the subject, human life and opportunities do not allow a holistic study. For this reason, computers and models can enable us to combine different existing views. This study mainly focused on a computational model that connects the current explanations to

strengthen our valid understandings and/or give another perspective, and of course expects to benefit the literature.

The model produced in this study allows us to see how different and many effects affect the today's dominant word order distributions. It tries to be a playground for many people to try, observe and understand. And the good thing is, they don't need to go back or forth a century or so. This model that has not been created in the literature before and has a deficiency. People can try different network models to communicate, different frequencies of communication, different sizes of population, different language types, different starting bias and different numbers of generations to predict. This model will give us a holistic perspective of what has been told in the literature before.

1.3 Study Limitations

First, this is a study based on inferences made upon the behavior of users of languages that have emerged in the near future. It is difficult to generalize to the whole world and to all times. This study captures the current understandings. Also, the study has built upon experimental data and the way of making sense of this data that there is no holistic information due to the limitations of experiments and setups. Hence, not all sentence types (like complex ones) could not be included we only cover reversible and irreversible transitive sentences.

1.4 Organisation of the Thesis

This master's thesis shows a model that can help us to understand how the word orders of today's languages distribute, while connecting different arguments from different studies. The organisation of the thesis is as follows: This first chapter explains the overall topic and general views about the topic, while presenting the research questions, contributions of this study to the literature and giving the limitations of the study.

Chapter 2 represents literature review on the topic. It will give more in-depth information on how the other previous studies contribute and suggests ideas to our study.

Chapter 3 will explain the overall model. There are different steps and stages in the model, all will be covered in this section.

Chapter 4 provides the different experiments setups on the model. And the results of these experiments will be given. Also, it will include explanations to understand how the model works and what the results tell us.

Chapter 5 is where the discussion will be done and represented, and the links and differences with existing studies also will be presented.

Chapter 6,the last chapter, gives the conclusion of this master's thesis and possible future work will be presented.

Last but not the least, references follows up all the chapters and Appendix A ... Appendix B ...

CHAPTER 2

BACKGROUND INFORMATION

In this chapter, related studies are given in detail. Since our study tries to make a model that connects different parameters of emergence in word orders in languages and transmission to the generations in different circumstances, we need to understand the actual human behaviour. To do that, current literature needs to be examined. There are different studies done on the topic. They will be given to see what has been changed and how we can connect them with our model. This section is divided into three sections in terms of their contribution to the model. First section tries to give the previous studies on the relationship and implications of nonverbal communication (gestures/pantomimes), newly emerged sign languages, and artificial language experiments done labs with word ordering behaviour of humans. Second section examines the studies on iterative learning and reflects the effect of language on the transfer of generations. In third section, studies on the effect of the size of the community and the network structures on the systematicity of the language will be discussed.

2.1 Gesture Experiments, Emerged Sign Languages and Dominant Word Orders

We can first understand how gestures help us to convey how humans learn languages. [15] supports that we all learn languages which we are exposed after we born. But if we are not able to acquire language as we supposed to be, like what if we cannot hear? Human beings need to communicate, so, they will try alternative ways like gesturing. Furthermore, what if the deaf people are the first ones in a population? They need to be the bridge between hearing ones and future generations. Examining the word-order patterns in communication systems developed in laboratories is a a popular method. Like in the studies, [16], [8] to explore cognitive biases for word order preferences, silent gesture, a practice in which adult hearing individuals transmit information exclusively using their hands and no voice, has been utilized.

Language learners are initially attracted to a language's canonical order (for L1 [17]; for L2 [18], [19], [20], [21]), because they are one of the sources to understand who (subject) did something to whom (object). Throughout the years, to comprehend the appeal of word ordering, there has been a increase in interest in sign languages while understanding language emergence in general. Emerging sign languages allow us to discover language emergence in real time, which goes well beyond the possibilities of spoken language research. Equally valuable, gestures of spoken people also help us to understand the cognitive effects in communication.

Most of the works consist of picture/video setups and experiments. Like [22], they carried out 4 experiments, 2 with English (a subject-initial language)speaking children and 2 with Fijian (a subject-

final language speaking children. Children tried to learn a miniature artificial language based on names for 2 horses and 2 carts either ordering them in horse-cart or cart-horse. Then, they examined the influence of animacy with toy graders and toy boulders.

They showed that grader+boulder sentences is more easily to learn than the reverse. So it is said that agency of an item makes a representation easy to acquire and they distinct agent/patient property then the animate/inanimate property. Also, the results were the same in both English speaking group and Fijian speaking group. Hence, there is no effect of first acquired language's word order, this may be a cognitive bias for putting agents (subjects) before patients (objects).

[4] conveyed a cross-linguistic study. They wanted to see whether the language we speak is influential on how we communicate non-verbally. They asked English, Turkish, Spanish and Chinese participants, where all their native languages have different dominant word orders. Participants carried out two non-verbal tasks: first is describing an event without speech (using gestures only) and second is reconstructing an event using pictures. Their results show that, speaker's word order for their native languages had no effect on their non-verbal communication. They also found out that all of the speakers of four languages adopted the same ordering behaviour, agent-patient-act (or SOV). This is also the mostly used word order in many languages, newly developing gestural languages, and experimental artificial languages, too. The results show that humans impose a natural order on occurrences when we describe and rebuild them non-verbally and when we create new language.

[23] made two gesture-production experiments and one gesture comprehension experiment on native and normally hearing Italian and Turkish participants, where the languages have different word order preferences (Italian - SVO, Turkish - SOV). In the first experiment, they tested whether participants used their gestures by following their native language's structural regularities. In the second experiment, they used a stimuli to get improvised gestures. They wanted to see if there is an evidence for phrase structure in their improvised gestures. In the third experiment, they investigated whether the preferences found in gesture production also appear in gesture comprehension. And finally, they study the preferences of phrase structure by assessing participants' order preferences for prosodically flat sequences of words in their native language. Results show that, in the direct connection between the sensory-motor and the conceptual systems, SOV is the preferred order; the SVO order is favoured by the computational system of grammar.

Observing different word order preferences in reversible (where both agent and patient are animate) and irreversible sentences (where agent is animate and inanimate patient), [24] claims that this state depends on says that the state depends on the communicative pressures of semantic factors. SOV is used for irreversible events since it is clear who is doing what to whom (i.e. "MAN-BREAD-CUT"). On the other hand, in reversible events (i.e. "WOMAN-MAN-KISS"), gesturers prefer SVO over SOV by putting the action(verb) in the middle of two animated participants in the event, maybe to reduce any kind of misunderstanding between the agent and patient roles.

biliblbilbilbiblSOV is the most used word order in the world languages. But where did SVO come from and why? [25] treated the subject differently and wanted to test whether SVO emerged for a reason, i.e. maybe it is not well suited for describing reversible events (i.e. a girl kissing a boy). If we use SOV in reversible event, that sentence will be "a girl a boy kissing", but this will only be understood if we know SOV is the word order we use. If someone using OSV as their canonical word order, that sentence will be understood as it was the boy who is kissing the girl. Another reason is maybe the pressures of efficiency (it is logical to mention agents before patients as a principle, which will

definitely rule out many other alternative word orders). They tested speakers of both English (SVO) and Turkish (SOV), by asking them to use pantomime to describe some reversible and irreversible events. Additionally, they gave some participants the task of teaching the experimenter the form of the gestures while being consistent about them. These restrictions caused SVO to appear in Turkish and English-speaking individuals (SOV). Their results show that being efficient, putting subjects before objects in the sentences, and avoiding SOV order for reversible events are the three requirements that SVO permits language users to achieve, and this is said to be at least part of the reason why SVO arises in the world's languages.

[9] also uses gesture production to see what causes preferences of SOV than SVO, or vice verse. To do that, they focused on the role of the verb in the sentences. They also mention extensional (i.e. throw) and intensional (i.e. think) verbs. Their results for events with extentional events show that SOV is mostly preferred order, which is consistent with the previous works ([4], [23], [24]). But, with event with intentional verbs lead to the SVO order instead. They conclude that meaning of the verb is critical while ordering words in emerging language systems. This results also support the language evolutionary point, where it implies that semantic underlies the early formation of syntactic rules of the language.

[26] argues that the emergence of SVO is possible with exposure of a shared lexicon, which makes it possible to liberate adequate cognitive resources to use syntax. Eventually, they argue that SVO is more efficient word order to express syntactic relationships. They gave Italian (SVO) and Persian (SOV) speakers a set of gestures to learn, and then they asked them to narrate basic events to verify their hypothesis. It was revealed that after enough time, when both groups developed a consistent gesture repertoire, there emerged a coherent usage of SVO. This study is particularly is important for the Persian-speaking participants since they switched their native language word order and started to use SVO once they were exposed to a new lexicon and become confident about it.

One piece of information to support the existence of a cognitive bias for SOV might be that the SOV is still dominant today. For example, a cross-linguistic study of [27] argued that SOV is the oldest word order used, and with different needs and conflicts made other orders to emerge. The geographical distribution of word orders around the world also seems to support this opinion. On every continent but Africa, SOV is the dominant word order. SVO, on the other hand, is mostly limited to Africa and Eurasia, with little to no presence in the Americas and Austronesia.

On the side of sign languages, there are different orientations. Naturally emerged sign languages are said to have little or no influence of existing languages. Previous studies on these systems show that there is a great amount of variation in word order preferences in the initial stages of young sign languages. There is a significant degree of variance in word order preferences in the early stages of young sign languages, according to earlier studies on these systems. An emerged sign language in Colombia is lacked a regular word order preference according to [28]. Likewise, in NSL, in [29], it is reported that there are different word order preferences in different age groups with no clear tendency towards a certain word order.

Like [4], who said that the native language has little/no effect on gestural word order preferences, ABSL is reported to have different tendency of a word order (SOV) in spite of their regional languages (Hebrew and Arabic - both SVO) [30], that is pointing out the word order preferences does not always influenced by the local speaking language. The preference of word orders of ABSL users shows a

systematic variance (OSV-SOV in different event structures) according to a more recent research by [11].

According to [31], ISL has showed an SOV tendency as a consistent word order in younger generations. In another study [32], it was reported that in an American and Chinese homesign system, the 'subject' in an event is frequently ignored and users preferred the OV order. This again shows that these emerged languages are independent of the languages spoken in the surrounding area.

For the young emerged languages, most of them are lack of a clear word order tendencies in either reversible or irreversible events. There are studies that show clear preferences generally after the first generation. [33] and [34] show that CTSL(used in a small community in Turkey) has no plain word order preference in the first generation, but with the second generation SOV seems to be the dominant order for irreversible events and OSV is for the reversible events, and in the third generation, the previous preferences become more common even OSV is the least prevalent word order. It has been reported that SVO, the most common after SOV, has not yet emerged in this population. Correspondingly, similar results were obtained in ABSL, according to [11]. Instead of SVO, OSV has emerged in this community, too. This is an intriguing result, because SVO is preferred over OSV in today's languages. These results make us ask similar questions again. Is SOV the simplest sort or do we have a bias for it? What caused the emergence of SVO?

The issue of who or what comes first when forming a sentence has also been the subject of many studies. We know that in world languages, subject predicate before verb is fire (S>V). To explain this, [35] refers to the "agent-first" principle. This principle says we have a bias that causes us to put the subject before the verb while forming a sentence. This principle is supported by late second language learners [36] and pidgin languages [37], that they prone to use subject before verb to distinguish between roles in a sentence with an agent and a patient. It seems to be defending a cognitive tendency that causes us to use the subject of the sentence before the object naturally. On the other hand, there are also findings that contradict these views. These tendencies may not be observed in newly emerged sign language systems, like dominant OSV order in CTSL and ABSL. This may undermine the validity of "agent-first" principle. The principle alone seems to be not enough, other pressures should have also taken place in these sysyems. In this regard, [11] tries to explain this situation with animacy property of the characters in the event. It argues that regardless of whether it is a subject or an object, the human or the character most similar to a human being is put forward. But these principles, again, due to limited test cases, cannot be fully defended as true or false. As many events left out. For example, sentences with inanimate subjects (with animate/inanimate objects) in new languages have not yet been studied. Likewise, situations where animal subjects and human objects are present, or where human subjects act on animal objects, were not studied.

The approaches to studying word order mentioned so far can be summarized as follows:

- a consistent word order may be absent in the very initial stages of a young natural system,
- there are cognitive biases (independent of surrounding/native languages) and communicative pressures shaping word order preferences, and
- considering the prevalence of SOV and SVO, there is so far robust evidence for the S>O pattern, displaying a bias for the agent preceding the patient.

On the other hand, [38] claims in experimental studies (like gestures, pantomimes) that people prefer an unnatural way. In such cases, although people prefer or have chance to choose a certain word order to describe certain kinds of events, it is stated that people actually prefer to use the same order during natural speech. So, regardless of the types of events, we prefer to use the same order while talking. Taking all these into account, they argue that repeated use will eventually lead to regularity rather than naturalness.

2.2 Learning Models and Iterated Learning

There are different proposals on how learning occur as a model. Models that involve learning the language from different aspects, including word learning, frequency learning, have been proposed before. However, a model of word orders does not exist so far. This study is essentially a model proposal to fill this gap. Yet, for our model we are influenced by a few of the studies. Since there is no consensus on how language is learned, only the insights and studies we used for our model will be given here.

[39] proposes a framework based on the principles of Bayesian inference to capture human concept learning from examples. In the study, it is supported that a given collection of instances can provide significantly more information about the concept to the Bayesian learner, who can then use this knowledge to rationally estimate the likelihood that any new item will also be an instance of the concept. The proposed system consists of 3 components: the first is a prior probability distribution across a set of potential concepts in the hypothesis space; the second is the likelihood function to compute the probability of each hypothesis with the help of the provided set of utterances, finally, the the principle for which has the learner calculate the likelihood of applying a concept to new objects by averaging all hypotheses' predictions and weighting them according to their posterior probabilities. This study is important because it shows how it is possible for people to learn and generalize concepts from just one or a few examples which also support the actual human behaviour. Although it may not be compatible fully, we will use the main components of this work. We will map concept learning to word order preferences.

One model that we considered is the iterated learning model where language is treated as a cultural knowledge [40]. Iterated learning is a way of transmitting one information from an individual to the other. An individual's learning is provided by the other individual's output of its learning. At each utterance, each learner sees data, forms a hypothesis, then, produces the data to the next learner (see Figure 1). The aim of this process is simply investigating the cultural evolution of linguistic structure.

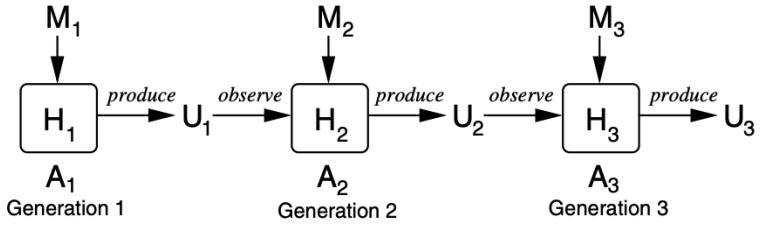

Figure 1: iterated learning model

This model is also related with Bayesian inference. At each learning process each learner update their beliefs based on rational procedure. For this study, for example, the preferred order to describe a reversible/irreversible event will be updated for events in related class.

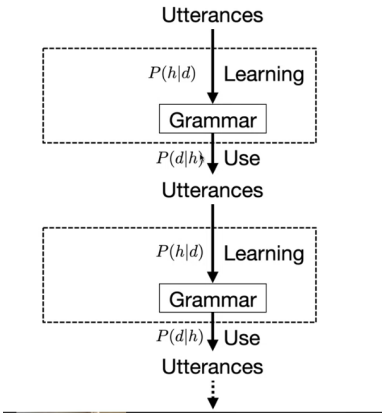

Figure 2: modelling iterated learning based on hypothesis

Another study of [41] combines the ideas of iterated learning and bayesian inference in an agent based model. They create a model where learners compute a posterior distribution by combining a prior knowledge of a language with the provided data. They showed that the prior and the amount of data they provided affected whole model when they applied iterated learning. They argued that the problem is choosing the maximum probable posterior. Then, they suggested a variant of the EM algorithm. They show how iterated learning fits into language acquisition and create a connection between spoken and written languages; this suggests that information transferred through iterated learning will eventually come to represent the learners' ideas.

For learning the meanings of words, [42] proposes a Bayesian framework. Their theory simply describes how learners may draw valid conclusions about a word's referents from a few examples by using reasoning and inductive reasoning that combines prior knowledge of the word's meaning with the statistical structure of the observed cases using Bayes. They included the how adults and children differ when learning. They have done three experiments on adults and children. In three tests, both adults and kids acquire referents for item categories at different levels, testing the predictions of the Bayesian approach. The Bayesian explanation outperforms competing theories in terms of both quantitative model fits and the ability to explain major qualitative events. Then they showed how Bayesian word learning models were more applicable. Of course, we must not forget that this model also only deals with simplified versions of real challenges.

[43] wanted to test the iterated learning paradigm with an artificial language learning experiment with humans in the laboratory. This study mainly support how the previous mathematical and computational models of iterated learning can really be applied in actual world. Iterated learning is said to be the explanation of cultural transmission of language with particular and natural constraints of language. Their experiments showed that iterated learning is a vigorous model for understanding the behavior of humans in transmitting language to generations. Another study [14] also shows how iterated learning helps us to understand origins of a language. They reviewed previous studies of iterated learning. In

this study, they used computational simulations of agents; mathematical modelling; and laboratory experiments to support the power of iterated learning.

Some linguists emphasize the importance of frequency of occurrence in language learning. Since our model adopted the idea of frequency learning, studies should be included here. Usage-based theories hold that the acquisition of language is exemplar based. It is said that the frequency with which we encounter information provides learning with what we associate with it.

The article [44] tried to demonstrate how the frequency of input has a profound impact on language acquisition. The way of humans forming sentences and syntax are affected by frequency are provided. It is the result of the learning of many constructs and the frequency-biased abstraction of their internal regularities. The categories and patterns that make up linguistic regularities are said to be formed through given inputs and experience. They advocated that language change is greatly explained by frequency. It is finally concluded that the sensitivity of learners to frequency in many linguistic areas has effects on how implicit and explicit learning theories interact.

Another study [45] that wanted to examine the connection between children's language acquisition and the frequency of morphological, lexical, and syntactic forms as their input with some brief counter-arguments towards the effects of frequency in language learning, too. They have reported that they explored several variables interact with frequency effect, since children do not have a simple way of mapping the input string directly. Their study was done in English. They tested for a frequency-based explanation with this empirical work they have done. At the end, they discussed a relationship between the relative frequency of forms in the input of infants and errors, including morphological errors, optional infinitive errors, and accusative-for-nominative errors.

A study [46] wanted to prove the role that the frequency effect plays in language learning. They wanted to prove that frequency effect must be considered in children's language learning. They studied simple syntactic constructions along with other respects about children's language acquisition and then presented theories. In the study, it was reported that the high frequency provides early acquisition, provides systematization while causes errors against competing forms and also, interacts with pattern learning mechanism. Henceforward, this study documents the importance of frequency, regardless of any other language acquisition explanation.

Finally, a few models that we have influenced and adapted to our own model can be found on GitHub. We were altered by the materials of Lab 2, 3, and 4, used in Simulating Language class, one of which was taught at Edinburgh University [47]. Lab 2 is the simplest way that we could start with. It is a simplified version of the model in [39], which is a simpler version of the model in [42] for word learning. The model in the lab allows us to create meanings for words as referents lists. Then, a hypothesis space is created and a Bayesian rules take care of the rest. Lab 3 model is simple Bayesian model of frequency learning. This model allows us to explore the effects of the prior and the data on frequency learning as in the model of [48]. Finally, Lab 4 presents us the iterated Bayesian learning model. This model combines the replication of the iterated learning model of the evolution of frequency distributions [49] , and is built around the Bayesian model of frequency learning/regularisation from Lab 3 model. This model also allows us to explore the effects of learning biases.

Also, we have get ideas from a tutorial of agent based models tutorial [50]. In this repository, the Part 3 has a complex model to represent a population of agents that interact with each other. Agents

have different personality types (as stubborn or flexible, affects the learning flexibility). Then multiple simulations can be done with different random populations in this model.

2.3 Network Structures

Another issue that this study deals with is the structures of communities. Of course, today there are communities that communicate in different sizes and in different ways. Previous study reveals that the size and social structure of the community may play an important influence in the evolution of language ([51]; [52]; [12]). It was theorized that emerging sign languages that emerge in tiny groups (i.e. village sign languages) had less conventionalized structure. According to [53], languages forming in larger groups and/or communities with less common heritage (deaf community sign languages, for example) tend to be more uniform .

[54] used a laboratory setting to investigate how word order preferences vary with different parameters: learning biases, size of the community and the amount of data that participants are exposed to. Their results showed that the size of the population and the amount of data they are exposed play a significant role for language convergence. If the participants exposed data from one and single speaker, results show that listeners learn with the frequency and variability was seen to persist over three generations. But with the same amount of distributed data from multiple speakers, frequency learning was reported to be failed. They showed that there is a more prominent bias for SOV that OSV and VSO. In addition, it is reported that the effect of the amount of data exposed from multiple speakers contributed greatly to frequency learning.

[55] investigated the word order variability to test the hypothesis. They searched how word order becomes a standard in new communication systems that differ in their social structure and community size by using real data. The results showed that there is significantly more variance in word order preferences in CTSL as opposed to those in ABSL (used in a bigger community in Israel, both within and across signers: CTSL signers show less convergence as a community, and are less consistent in their own productions. These results support the hypothesis that the size of a language community has an effect on conventionalizing in early stages of language emergence: the language of bigger communities is more uniform in structure that that of smaller communities.

CHAPTER 3

THE MODEL

In this chapter we will explain the dynamics of our model and how we adopted the information mentioned in Chapter 2.

The word order preferences in reversible vs. irreversible occurrences are the primary linguistic process under examination in our study. We have given in the previous section that events involving two characters from the same semantic category (animate-animate, or reversible events) present a communicative pressure to people. The roles of the agent and the patient (i.e. male or female) suffer from ambiguity, especially since there is no systematic preference for word order in young emerging communication systems. On the other hand, it is not possible to experience ambiguity in non-reversible situations. As in an earlier example, it will be the man who cuts the bread, since bread will not be the subject here to cut the man, it will be understood correctly by the listener no matter how the sentence is formed.

3.1 Agents & Language

3.2 Iterated Learning

3.3 Network Structures

CHAPTER 4

SIMULATIONS AND THEIR RESULTS

4.1

CHAPTER 5

DISCUSSION

5.1 Method

CHAPTER 6

CONCLUSION AND FUTURE WORK

REFERENCES

[1] M. S. Dryer, "Order of subject, object and verb," 2013.

[2] D. J. Napoli and R. Sutton-Spence, "Order of the major constituents in sign languages: Implications for all language," *Frontiers in psychology*, vol. 5, p. 376, 2014.

[3] T. Briscoe, "Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device," *Language*, vol. 76, no. 2, pp. 245–296, 2000.

[4] S. Goldin-Meadow, W. C. So, A. Özyürek, and C. Mylander, "The natural order of events: How speakers of different languages represent events nonverbally," *Proceedings of the National Academy of Sciences*, vol. 105, no. 27, pp. 9163–9168, 2008.

[5] T. Givón, *On understanding grammar*. Academic Press, 2014.

[6] B. B. Blake, "Russell s. tomlin, basic word order. functional principles. london: Croom helm, 1986. pp. 308.," *Journal of Linguistics*, vol. 24, no. 1, p. 213–217, 1988.

[7] L. Maurits, D. Navarro, and A. Perfors, "Why are some word orders more common than others? a uniform information density account," in *Advances in Neural Information Processing Systems* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds.), vol. 23, Curran Associates, Inc., 2010.

[8] M. L. Hall, R. I. Mayberry, and V. S. Ferreira, "Cognitive constraints on constituent order: Evidence from elicited pantomime," *Cognition*, vol. 129, no. 1, pp. 1–17, 2013.

[9] M. Schouwstra and H. de Swart, "The semantic origins of word order," *Cognition*, vol. 131, no. 3, pp. 431–436, 2014.

[10] A. Kocab, H. Lam, and J. Snedeker, "When cars hit trucks and girls hug boys: The effect of animacy on word order in gestural language creation," *Cognitive science*, vol. 42, no. 3, pp. 918–938, 2018.

[11] I. Meir, M. Aronoff, C. Börstell, S.-O. Hwang, D. Ilkbasaran, I. Kastner, R. Lepic, A. L. Ben-Basat, C. Padden, and W. Sandler, "The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages," *Cognition*, vol. 158, pp. 189–207, 2017.

[12] L. Raviv, A. Meyer, and S. Lev-Ari, "Larger communities create more systematic languages," *Proceedings of the Royal Society B: Biological Sciences*, vol. 286, no. 1907, p. 20191262, 2019.

[13] L. Raviv, A. Meyer, and S. Lev-Ari, "The role of social network structure in the emergence of linguistic structure," *Cognitive Science*, vol. 44, no. 8, p. e12876, 2020.

[14] S. Kirby, T. Griffiths, and K. Smith, "Iterated learning and the evolution of language," *Current opinion in neurobiology*, vol. 28, pp. 108–114, 2014.

[15] S. Goldin-Meadow, *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Psychology Press, 2005.

[16] M. L. Hall, Y. D. Ahn, R. I. Mayberry, and V. S. Ferreira, "Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime," *Journal of memory and language*, vol. 81, pp. 16–33, 2015.

[17] D. I. Slobin and T. G. Bever, "Children use canonical sentence schemas: A crosslinguistic study of word order and inflections," *Cognition*, vol. 12, no. 3, pp. 229–265, 1982.

[18] J. M. Meisel, H. Clahsen, and M. Pienemann, "On determining developmental stages in natural second language acquisition," *Studies in second language acquisition*, vol. 3, no. 2, pp. 109–135, 1981.

[19] M. Pienemann and M. Johnston, "A predictive framework of sla," *Manuscript: University of Sydney*, 1987.

[20] V. Cook and V. J. Cook, *Linguistics and second language acquisition*, vol. 12. Macmillan London, 1993.

[21] V. J. Cook, "Timed grammaticality judgements of the head parameter in l2 learning," *The Dynamics of Language Processes: Essays in Honor of Hans W. Dechert, Gunter Narr Verlag, Tübingen*, pp. 15–31, 1994.

[22] B. Byrne and E. Davidson, "On putting the horse before the cart: Exploring conceptual bases of word order via acquisition of a miniature artificial language," *Journal of Memory and Language*, vol. 24, no. 4, pp. 377–389, 1985.

[23] A. Langus and M. Nespor, "Cognitive systems struggling for word order," *Cognitive psychology*, vol. 60, no. 4, pp. 291–318, 2010.

[24] E. Gibson, S. T. Piantadosi, K. Brink, L. Bergen, E. Lim, and R. Saxe, "A noisy-channel account of crosslinguistic word-order variation," *Psychological science*, vol. 24, no. 7, pp. 1079–1088, 2013.

[25] M. L. Hall, V. S. Ferreira, and R. I. Mayberry, "Investigating constituent order change with elicited pantomime: A functional account of svo emergence," *Cognitive Science*, vol. 38, no. 5, pp. 943–972, 2014.

[26] H. Marno, A. Langus, M. Omidbeigi, S. Asaadi, S. Seyed-Allaei, and M. Nespor, "A new perspective on word order preferences: the availability of a lexicon triggers the use of svo word order," *Frontiers in psychology*, vol. 6, p. 1183, 2015.

[27] F. J. Newmeyer, *Language form and language function*. MIT press, 2000.

[28] W. Washabaugh, "On the sociality of creole languages," in *Dialect and Language Variation*, pp. 542–557, Elsevier, 1986.

[29] A. Senghas, E. L. Newport, and T. Supalla, "Argument structure in nicaraguan sign language: The emergence of grammatical devices," in *In E. Hughes & A. Greenhill (Eds.), Proceedings of the 21st annual Boston university Conference on Language Development*, Citeseer, 1997.

[30] W. Sandler, I. Meir, C. Padden, and M. Aronoff, "The emergence of grammar: Systematic structure in a new language," *Proceedings of the National Academy of Sciences*, vol. 102, no. 7, pp. 2661–2665, 2005.

[31] I. Meir, W. Sandler, C. Padden, M. Aronoff, *et al.*, "Emerging sign languages," *Oxford handbook of deaf studies, language, and education*, vol. 2, pp. 267–280, 2010.

[32] S. Goldin-Meadow and C. Mylander, "Spontaneous sign systems created by deaf children in two cultures," *Nature*, vol. 391, no. 6664, pp. 279–281, 1998.

[33] R. Ergin, *Central Taurus Sign Language: a unique vantage point into language emergence*. Tufts University, 2017.

[34] R. Ergin, I. Meir, D. Ilkbaşaran, C. Padden, and R. Jackendoff, "The development of argument structure in central taurus sign language," *Sign Language Studies*, vol. 18, no. 4, pp. 612–639, 2018.

[35] R. Jackendoff, "the autonomy of syntax," *Verb-particle explorations*, vol. 1, p. 67, 2002.

[36] W. Klein and C. Perdue, "The basic variety (or: Couldn't natural languages be much simpler?)," *Second language research*, vol. 13, no. 4, pp. 301–347, 1997.

[37] T. Givón, *Functionalism and grammar*. John Benjamins Publishing, 1995.

[38] M. Schouwstra, K. Smith, and S. Kirby, "The emergence of word order conventions: improvisation, interaction and transmission," 2020.

[39] J. B. Tenenbaum, *A Bayesian framework for concept learning*. PhD thesis, Massachusetts Institute of Technology, 1999.

[40] K. Smith, S. Kirby, and H. Brighton, "Iterated learning: A framework for the emergence of language," *Artificial life*, vol. 9, no. 4, pp. 371–386, 2003.

[41] T. L. Griffiths and M. L. Kalish, "Language evolution by iterated learning with bayesian agents," *Cognitive science*, vol. 31, no. 3, pp. 441–480, 2007.

[42] F. Xu and J. B. Tenenbaum, "Word learning as bayesian inference.," *Psychological review*, vol. 114, no. 2, p. 245, 2007.

[43] S. Kirby, H. Cornish, and K. Smith, "Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language," *Proceedings of the National Academy of Sciences*, vol. 105, no. 31, pp. 10681–10686, 2008.

[44] N. C. Ellis, "Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition," *Studies in second language acquisition*, vol. 24, no. 2, pp. 143–188, 2002.

[45] E. Lieven, "Input and first language acquisition: Evaluating the role of frequency," *Lingua*, vol. 120, no. 11, pp. 2546–2556, 2010.

[46] B. AMBRIDGE, E. KIDD, C. F. ROWLAND, and A. L. THEAKSTON, "The ubiquity of frequency effects in first language acquisition," *Journal of Child Language*, vol. 42, no. 2, p. 239–273, 2015.

[47] S. Kirby, K. Smith, and M. Spike, "Simulating language, academic year 2021-2022." https://github.com/smkirby/simlang2022, 2022.

[48] C. L. Hudson Kam and E. L. Newport, "Regularizing unpredictable variation: The roles of adult and child learners in language formation and change," *Language learning and development*, vol. 1, no. 2, pp. 151–195, 2005.

[49] F. Reali and T. L. Griffiths, "The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning," *Cognition*, vol. 111, no. 3, pp. 317–328, 2009.

[50] L. Raviv and B. Thompson, "A short tutorial on agent based modeling in python." https://github.com/Limor-Raviv/Tutorial_Agent_Based_Models, 2018.

[51] G. Lupyan and R. Dale, "Language structure is partly determined by social structure," *PLOS ONE*, vol. 5, pp. 1–10, 01 2010.

[52] I. Meir and W. Sandler, "Variation and conventionalization in language emergence," *Eds.), Language contact, continuity and change in the genesis of modern Hebrew*, pp. 337–363, 2019.

[53] I. Meir, A. Israel, W. Sandler, C. Padden, and M. Aronoff, "The influence of community on language structure: Evidence from two young sign languages," *Linguistic Variation*, vol. 12, 12 2012.

[54] A. Kocab, J. Ziegler, and J. Snedeker, "It takes a village: The role of community size in linguistic regularization," *Cognitive Psychology*, vol. 114, p. 101227, 2019.

[55] R. Ergin, L. Raviv, A. Senghas, C. Padden, and W. Sandler, "Community structure affects convergence on uniform word orders: evidence from emerging sign languages," The Evolution of Language Conferences, 2020.