

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. Director of Institute
Dean, **Graduate School of Informatics**

Prof. Dr. Head of Department
Head of Department, **Cognitive Sciences**

Assoc. Prof. Dr. Prof. Dr. Hüseyin Cem Bozşahin
Supervisor, **Department, School**

Assoc. Prof. Dr. Dr. Rabia Ergin
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 Çınar

Signature :

ABSTRACT

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

Çınar, Burcu

M.Sc., Department of Cognitive Sciences

Supervisor: Assoc. Prof. Dr. Prof. Dr. Hüseyin Cem Bozşahin

Co-Supervisor: Assoc. Prof. Dr. Dr. Rabia Ergin

September 2022, 21 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

Çınar, Burcu

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Doç. Dr. Prof. Dr. Hüseyin Cem Bozşahin

Ortak Tez Yöneticisi: Doç. Dr. Dr. Rabia Ergin

Eylül 2022, 21 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

ABSTRACT.....	iv
ÖZ.....	v
DEDICATION.....	vi
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xii
CHAPTERS	
1 INTRODUCTION	1
1.1 Research Questions	2
1.2 Contributions of the Study.....	2
1.3 Study Limitations	3
1.4 Organisation of the Thesis.....	3
2 RELATED WORK	5
2.1 Gesturing and Young Sign Languages	5
2.2 Iterated Learning	5

2.3	Network Structures	5
2.4	Current Computational Models based on Bayes	6
3	THE MODEL	7
3.1	Research Method and Experiment Design	7
4	EXPERIMENT SETUPS AND THEIR RESULTS	9
4.1	Method	9
5	DISCUSSION	11
5.1	Method	11
6	CONCLUSION AND FUTURE WORK	13
	REFERENCES	15
APPENDICES		
A	TABLES FOR RELATED WORK CHAPTER	17
A.1	Summary of the Studies	17
B	EXTRA MATERIAL	19
C	INSTRUMENTS AND ETHICAL CLEARANCE	21

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
NSL	Nicaraguan Sign Language

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 canonical 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

RELATED WORK

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.

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.

2.1 Gesturing and Young Sign Languages

There has been a increase in interest in sign languages to understand language emergence in general, throughout the years. 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. [4]

2.2 Iterated Learning

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 ([16]; [17]; [12]). It was theorized that emerging sign languages that emerge in tiny groups (i.e. village sign languages) had less conventionalized structure. Languages forming in larger groups and/or communities with less common heritage (deaf community sign languages, for example) tend to be more uniform ([18]).

[19] investigates the word order variability to test the hypothesis. They search 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 shown that there is significantly more variance in word order preferences in CTS� (used in a small community in Turkey) as opposed to those in ABSL (used in a bigger community in Israel, both within and across signers: CTS� 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 than that of smaller communities.

2.4 Current Computational Models based on Bayes

CHAPTER 3

THE MODEL

In this chapter, the details of the user experiment are presented.

3.1 Research Method and Experiment Design

CHAPTER 4

EXPERIMENT SETUPS AND THEIR RESULTS

4.1 Method

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. Lopic, 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] G. Lupyan and R. Dale, “Language structure is partly determined by social structure,” *PLOS ONE*, vol. 5, pp. 1–10, 01 2010.
- [17] 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.
- [18] 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.
- [19] 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.

APPENDIX A

TABLES FOR RELATED WORK CHAPTER

A.1 Summary of the Studies

APPENDIX B

EXTRA MATERIAL

APPENDIX C

INSTRUMENTS AND ETHICAL CLEARANCE