An Agent-Based Model to Explain the Emergence of Dominant Word Orders

Burcu Çınar
Middle East Technical University, Ankara, Türkiye
cinarb@metu.edu.tr

MIDDLE EAST TECHNICAL UNIVERSITY

Abstract

A model is proposed to test the existing explanations regarding the emergence of dominant word orders by searching the ways in which language structures are shaped by culture, cognition, and communication. The model comprises three components: agent characteristics (including their biases and personality traits); communication functions to simulate environmental and communal effects; and iterated learning [1] that transfers and updates existing information from one generation to the next. 972 test cases were generated from the various parameters included in the model and all were run.

Introduction

It is known that many languages today have a dominant word order. The overall distribution of word orders in the world languages is as (SOV, SVO) > VSO > (VOS, OVS) > OSV [2] as in Figure 1. But the fact that these word orders are not evenly distributed has been the subject of many research and where these dominant word orders came from is still a matter of curiosity. Various explanations have been given on this subject with the help of newly emerged sign languages which offer a valuable opportunity to study language evolution; artificial language experiments; and computational models.

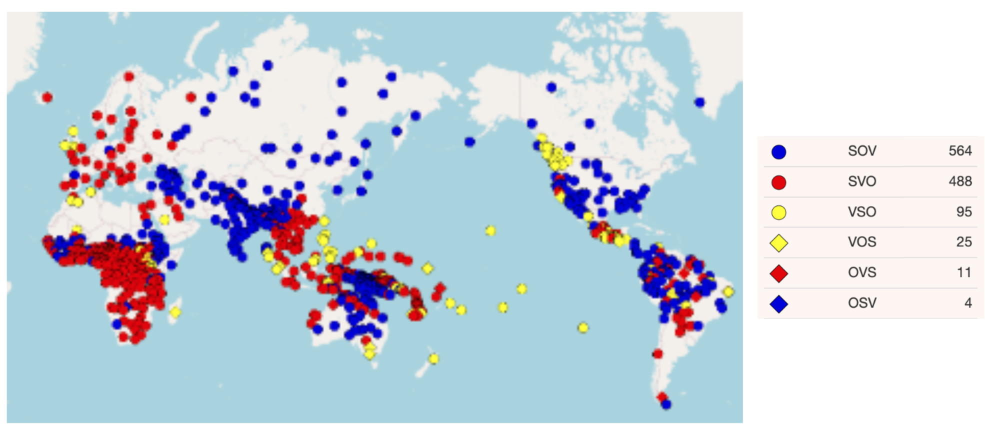

Figure 1: Dominant word order distribution in the world

In Figure 2, even if the words are in the same order, the meanings are different:

(a) if OSV is the dominant word order and it is the girl who kisses the boy.

(b) if SOV is the dominant word order and it is the boy who kisses the girl.

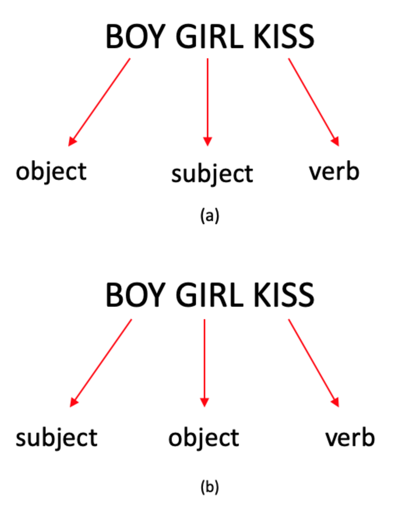

Figure 2: Same order, different meaning

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

- A dominant word order may be absent in the very initial stages of a young system;
- There may be cognitive biases [3];
- Communicative pressures shaping word order preferences [4];
- Considering the prevalence of SOV and SVO, there is so far robust evidence for the S before O pattern [5] [6], displaying a bias for the agent preceding the patient.

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 cycle given in Figure 3 is adopted in this study; real-word observations and lab experiments give an insight to create a computational model.

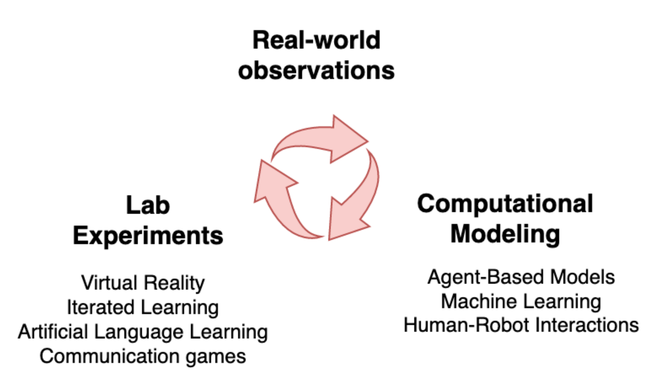

Figure 3: The cycle that contribute the language evolution studies

Model

Word Order Weights of Agents

Each agent produce word orders to communicate with other agent/s. Agents record the word orders as weights that they encounter/hear. These weights represent the corresponding order in the fixed word order list defined. An update example is given in Figure 5.

Figure 4: Example updating weights with a tendency in the language

Producing Word Orders

The probability for each word order type to be chosen is proportional to their areas in the roulette wheel.

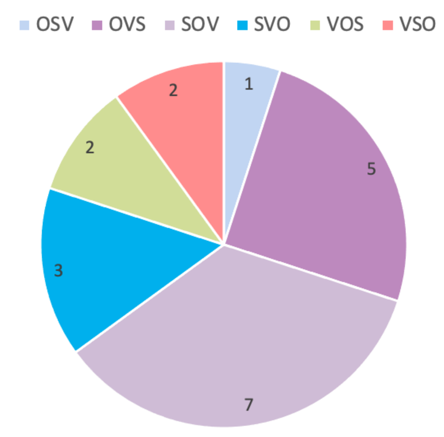

Figure 5: A roulette wheel areas according to weights of [1, 5, 7, 3, 2, 2] of word orders accordingly

Network Structures

Forming a community can be one-on-one. Another way to form a community is where one speaks and others listen (star). Where there are random conversations in a particular group and everyone listens (multiple stars in a group), or where everyone can listen to everyone else speak (totally connected/mesh). These were created from network topology types in computer science as expressed in Figure 6.

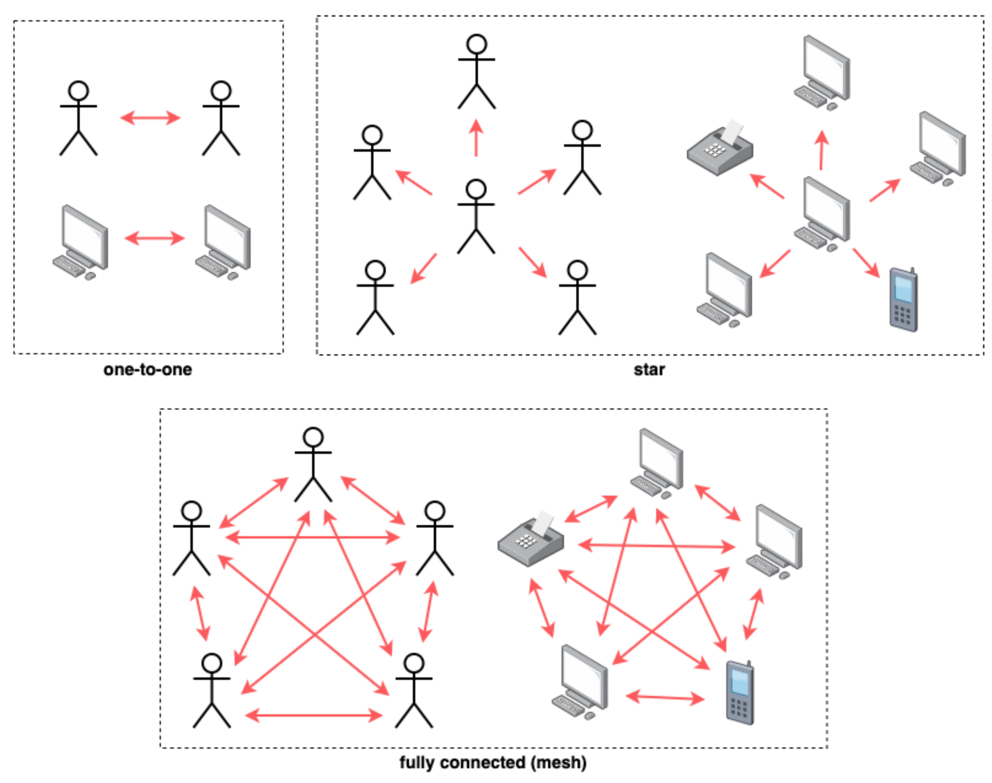

Figure 6: Network topologies and how agents form communication groups according to them

Iterated Learning

Iterated learning is seen as a critical mechanism of language evolution and used to explain the origins of structure in language. In this study, all agents are updating their current knowledge with the other agents' knowledge as in Figure 7.

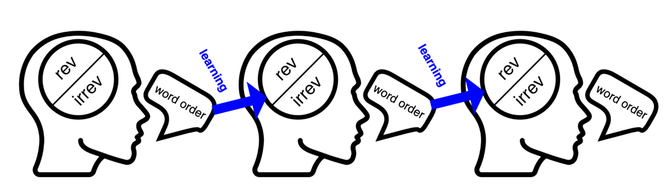

Figure 7: World order transmission

Results & Summary

Parameters used in the model (and default values):

- **starting bias**: uniform
- **number of generations**: 25
- **tendency for a word order**: yes
- **first community size**: 50
- **network type**: mesh
- **data size**: 5000
- **personality distribution**: F=S

Our results show that in some cases, dominant word order can emerge without any initial bias with different probabilities towards one order. We demonstrate that linguistic pressures and the rules of the language can influence a community's preference for specific word order shown in Figure 8.

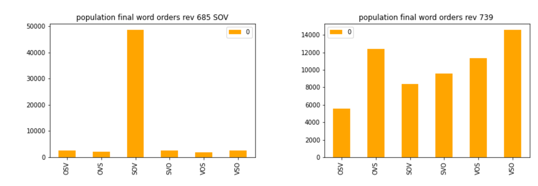

Figure 8: Existence of a preference for a word order

Contrastively, if a word order is not needed thanks to any mechanism in the language (case markings/suffixes) and if this community is biased from the beginning towards a word order, they can accept the bias from the beginning and carry it to the future as in Figure 9.

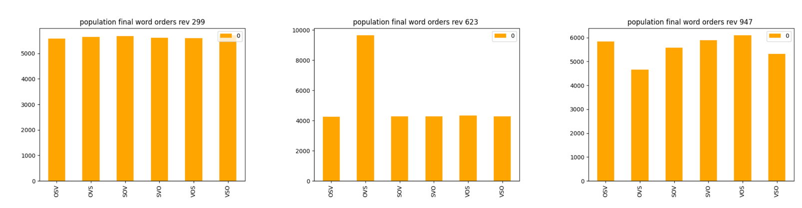

Figure 9: Uniform, Biased & Random bias final word orders

Results also show how the first community's size as in Figure 10 and different network structures affect the dominant word order emergence and evolution speed. As the community grows, the language may accept a word order, since the differences will be difficult to keep track of and keep in mind.

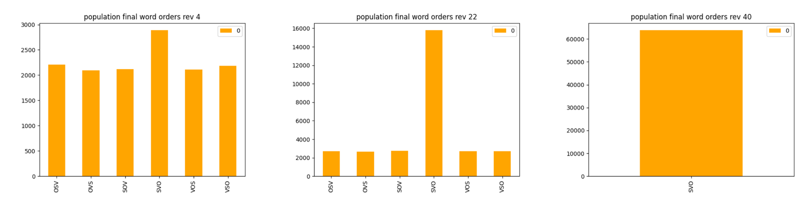

Figure 10: Different first community size final word orders

On the other hand, every member of the community must be connected with each other. For this, the model is tested with different network structure types. It turns out that communities that are more connected to each other reach a regular language faster as in Figure 11.

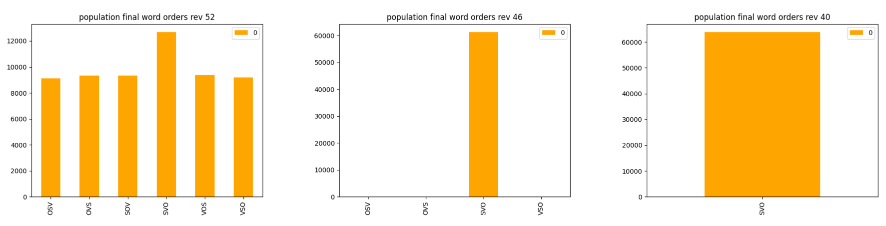

Figure 11: Different network structure final word orders, from left to right one-to-one, star, mesh

Furthermore, communities with more flexible agents tend to change their language more readily and accept a dominant word order, while those with more stubborn individuals show little or no change in language from the beginning as in Figure 12.

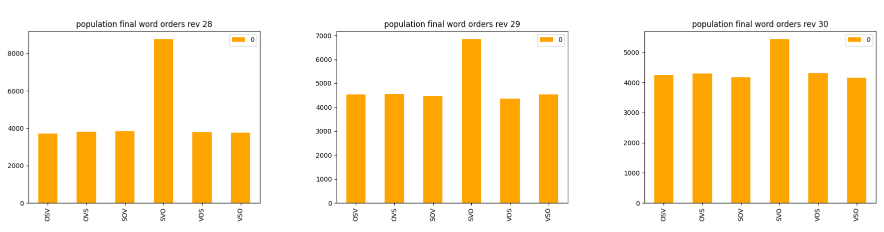

Figure 12: Different personality distributions final word orders

In summary, our model provides insights into the factors that contribute to the emergence and evolution of dominant word orders in natural languages. The results of this study contribute to a deeper understanding of the origins and mechanisms of word order preferences in human language while facilitating the exploration of the mechanisms underlying human learning and knowledge transfer.

References

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

[2] M. S. Dryer, "Order of subject, object and verb," 2013. [Online]. Available: https://wals.info/chapter/81

[3] 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.

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

[5] 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.

[6] 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.