

The first sentence the second sentence

a smaller subtitle

Dewi E. Timman
12419273

Bachelor thesis
Credits: 12 EC

Bachelor *Cognition, Language and Communication*



University of Amsterdam
Faculty of Humanities
Spuistraat 134
1012 VB Amsterdam

Supervisor

Dr. R. G. Alhama

Institute for Logic, Language and Computation
Faculty of Science
University of Amsterdam
Science Park 107
1098 XG Amsterdam

May 15, 2024

Abstract

Contents

1	Introduction	3
1.1	Literature review	3
1.1.1	Agent-based linguistic communication	3
1.1.2	Referential communication games	4
1.1.3	Word embedding vectors	4
1.2	Current research	5
2	Method	6
2.1	Data	6
2.1.1	(Data analysis)	6
2.2	Experimental Design	6
2.3	Model description	6
2.3.1	Hyperparameters	7
2.4	(Optimization)	7
3	Results	8
4	Conclusion	9
5	Discussion	10
	References	11

Chapter 1

Introduction

We are able to talk and communicate with each other. Our Languages are very complex. But how did language originate in the first place? [In this paper...](#)

1.1 Literature review

Structure:

Language derives meaning from its use (Wittgenstein, 1953).

Compositionality

-> agent-based referential communication games

Complex language

-> vectors from LLM

When researching the emergence of language, [usually ... \(more information about what kind of research is done in the field of language emergence\)](#) To do this, it is important to define a few concepts. First, about what kind of communication are we talking? Second, how can we research this communication? And Third, in what state do we start the research?

1.1.1 Agent-based linguistic communication

In this research, there will be looked at agent-based linguistic communication. This means that there is communication between two or more agents with the use of language. Usually, that means the agents need to solve a task together. [bronnen! \(maybe more information about the kind of tasks and/or the relevance\)](#) To solve the task, agents need to succesfully communicate to each other.

Zubek, Korbak, and Rączaszek-Leonardi (2023) provides an overview of different games

Abstract of (Vogt, 2005): The paper confirms previous findings that a transmission bottleneck serves as a pressure mechanism for the emergence of compositionality, and that a communication strategy for guessing the references of utterances aids in the development of qualitatively ‘good’ languages. In addition, the results show that the emerging languages reflect the structure of the world to a large extent and that the development of a semantics, together with a competitive selection mechanism, produces a faster emergence of compositionality than a predefined semantics without such a selection mechanism.

1.1.2 Referential communication games

To research this agent-based linguistic communication, agents can participate in a referential communication game. In such a game, agents talk about objects or other entities in a specified world. To do this, they need to come up with a language to communicate about their world. Usually, the agents do not have any prior linguistic knowledge.

bronnen! (ook meer informatie over de game zelf in relatie met taal, waarop berust de game?) bronnen!

Li, Ponti, Vulić, and Korhonen (2020) researched ref com games for machine translation.

1.1.3 Word embedding vectors

In this research, however, agents do start with prior linguistics knowledge. Agents use the word embedding vectors of a large language model (LLM) in order to ... (more about the relevance of this research. Maybe some papers about research with grounded knowledge.

Why is it relevant to already have some prelinguistic knowledge?

- To see what agents do when they have some already present structure/semantics (structured compositional language is most likely to emerge when agents perceive the world as being structured (Lazaridou, Hermann, Tuyls, & Clark, 2018)))

1.1.3.1 LLMs

The LLM from which the word embedding vectors are chosen, is Because ... (maybe a short overview of the best choices of LLMs for this kind of task and why)

1.2 Current research

Gap: Research with prior linguistic knowledge -> Do the agents do something with the structure/semantics already present in the word embedding vectors and can this say something about language emergence?

RQ: What is the effect of word embedding vectors from a language model on the emergence of agent-based linguistic communication from referential communication games? (simplify?) SubRQ: Do the agents communicate successfully? SubRQ: Do they communicate using compositionality?

Chapter 2

Method

Programming a model and see how it behaves and what makes it behave like that.
Dependent variable: output measured by the model and its evaluation metric -> how often is the communication successful? Compositionality metric
Independent variable: input -> words/sentences

2.1 Data

What data? Data preprocessing? Data split?

2.1.1 (Data analysis)

What does the data look like? Is there a data imbalance?

2.2 Experimental Design

What is the setup?

In this research, agents participate in a referential communication game. The game is adapted from Lazaridou et al. (2018) and is a variant of the Lewis signaling game (Lewis, 1969). More formally, the game works as following:

2.3 Model description

What does the model look like? Which equations are used? Diagrams/pseudocode?

2.3.1 Hyperparameters

What are the hyperparameters? (Rita et al., 2022) shows things about the loss and overfitting resulting in more compositionality.

2.4 (Optimization)

What optimizations were done?

Chapter 3

Results

Chapter 4

Conclusion

Chapter 5

Discussion

References

- Lazaridou, A., Hermann, K. M., Tuyls, K., & Clark, S. (2018). Emergence of linguistic communication from referential games with symbolic and pixel input..
- Lewis, D. D. K. (1969). *Convention: a philosophical study*. Cambridge, Mass: Harvard University Press.
- Li, Y., Ponti, E. M., Vulić, I., & Korhonen, A. (2020). Emergent communication pretraining for few-shot machine translation.
- Rita, M., Tallec, C., Michel, P., Grill, J.-B., Pietquin, O., Dupoux, E., & Strub, F. (2022). Emergent communication: Generalization and overfitting in lewis games.
- Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial intelligence*, 167(1), 206-242.
- Wittgenstein, L. J. J. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.
- Zubek, J., Korbak, T., & Rączaszek-Leonardi, J. (2023). Models of symbol emergence in communication: a conceptual review and a guide for avoiding local minima.