a memory based model for the emergence of vocabulary

Galip Ümit Yolcu 2016400249

May 27, 2019

Cmpe 557, Complex Systems Final Presentation

Introduction

Language as a self organizing complex system:

- No intrinsic meaning in words(language as game)
- Trying to understand how everyone else uses words
- Self organization through agent interactions

Some simplifications:
Mapping words(signals, sounds) to limited number of meanings (objects, situations)

Two approaches

There are two main approaches:

1. Sociocultural or semiotic
   - Emergence of vocabulary in a population
   - Interactions between agents change agents
   - Horizontal organisation
2. Sociobiological or evolutionary
   - Evolution of language across generations
   - Language as fitness
   - Strategies of language acquisition

Sociocultural: Naming Game

Emergence of vocabulary for objects

Minimal Naming Game proposed by Baronchelli et al.: common vocabulary for an object

Sociocultural: Naming Game

Agents have inventory of words for the object

Two agents interact:

- Speaker selects a word from inventory
- Listener checks inventory, if word is found, it is a success; if not a failure

Sociocultural: Naming Game

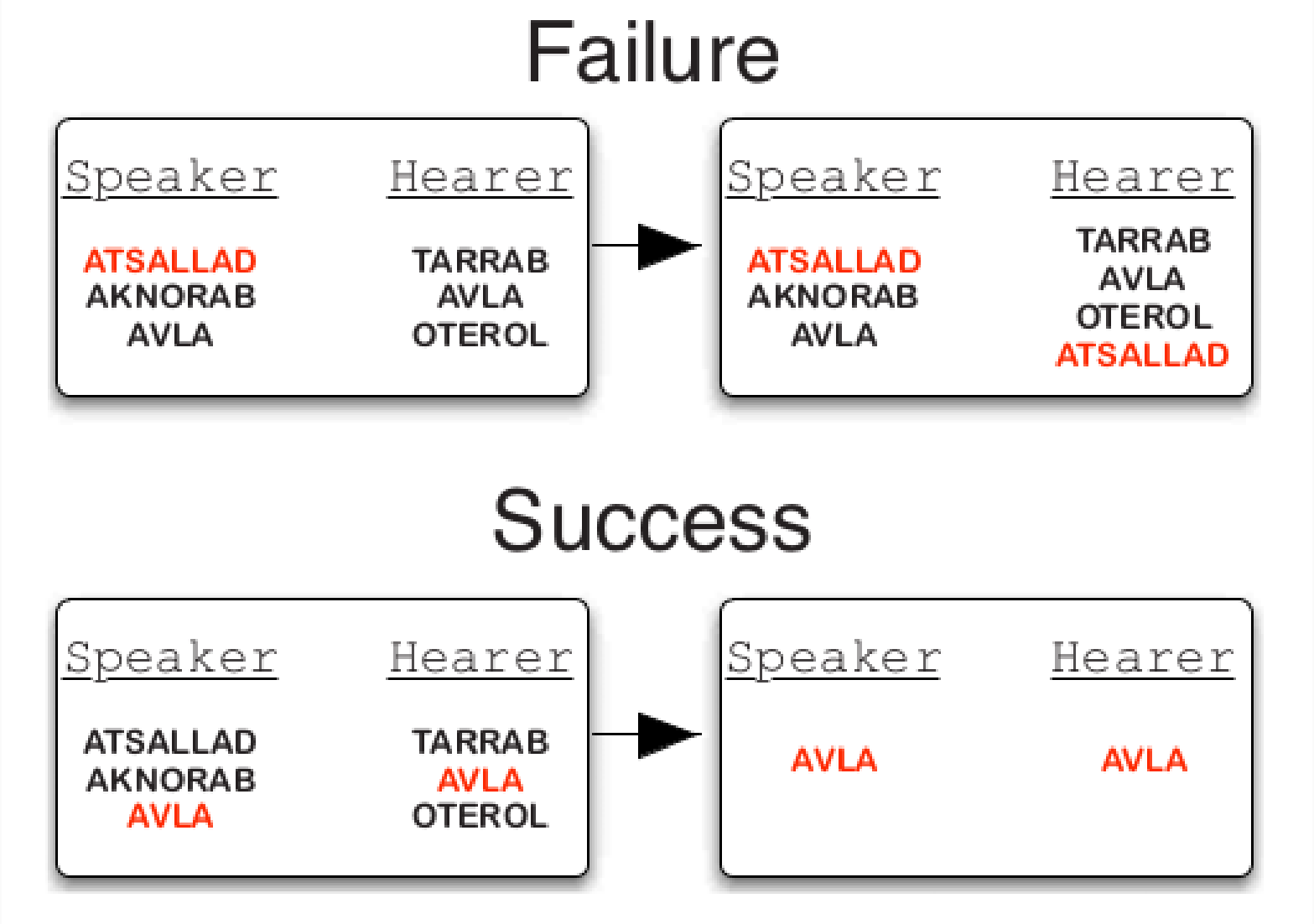
In case of failure:

- Listener adds word to her inventory

In case of success:

- Both parties delete every word except the successfully used

Sociocultural: Naming Game



Sociobiological: Evolutionary Language Game

Agents have an association matrix $A_{N\times M}$

Production matrix P and comprehension matrix Q:

$$p_{ij} = \frac{a_{ij}}{\sum_{j'}^{m} a_{ij'}} \quad q_{ji} = \frac{a_{ij}}{\sum_{i'}^{n} a_{i'j}} \tag{1}$$

Sociobiological: Evolutionary Language Game

Comprehension between two agents:

$$F(L_1, L_2) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (p_{ij}^{(1)} q_{ji}^{(2)} + p_{ij}^{(2)} q_{ji}^{(1)}) \tag{2}$$

Population fitness:

$$F(C) = \frac{1}{2\binom{|C|}{2}} \sum_{k} \sum_{l} F(L_k, L_l) \text{ where } k \neq l \tag{3}$$

Sociobiological: Evolutionary Language Game

New generation samples from teachers

Language evolves: imperfect sampling and different teacher selection strategies

Why not both?

Sociobiologic approach vs. Sociocultural approach

There are models that try to benefit from both

They are similar: interactions change the matrices

The Model

Agents have m memory sites for each meaning

Interact and update memory

Last m words are kept in memory

There can be duplicates in memory!

The Model: Interaction Rule

Agents are selected: one is speaker, the other is listener

Speaker selects random meaning, selects word from memory

Listener hears word, selects meaning from memory

If the meaning is correct, interaction is a success

If not, a failure

The Model: Memory Update Rule

Agent hold the last m words the was used

Success: both parties update memory

Failure: only listener updates memory

The Model: Relation to other models

Memory $\equiv$ Inventory with constrained repeated elements

Memory $\equiv$ Constrained association matrix:

- Count number of words to find number of columns
- a_{ij} = number of times j is found in memory for i

Thus we can use mathematical framework of Evolutionary Language Game

Parameters and Outputs

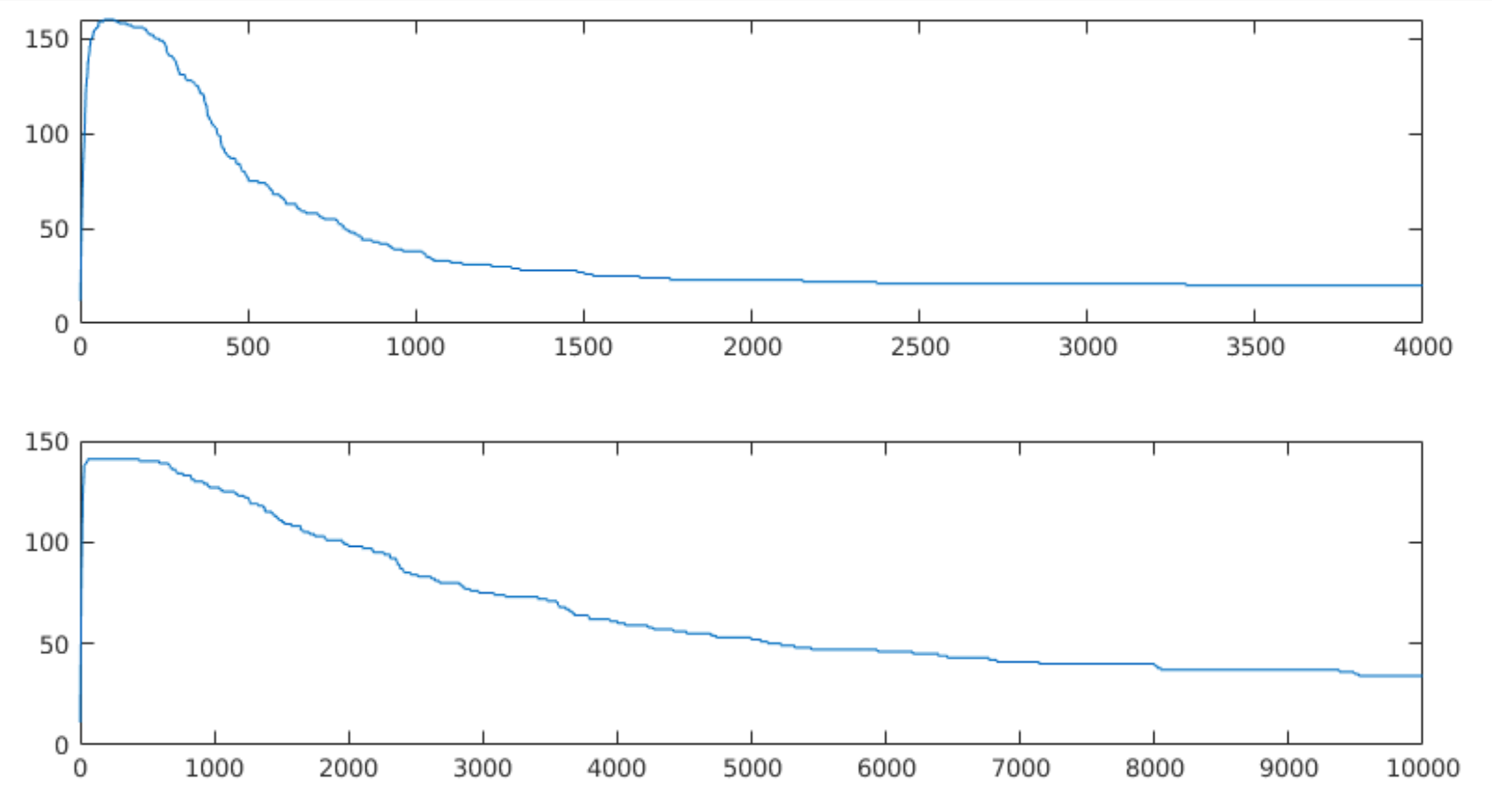
Parameters:

- N: number of agents in the population
- M: number of meanings(objects)
- m: memory size for each meaning
- W: number of words(signals)
- IT: number of iterations before the simulation is ended. Each agent is the speaker once in an iteration.

Outputs:

- D: number of words in use by the population (Naming Game)
- F: population fitness (Evolutionary Language Game)

Results: $M = N = 20$, $W \to \infty$, $m = 8, 30$



Time evolution of D

Results: $M = N = 20$, $W \to \infty$

The results are parallel with Naming Game

3 phases

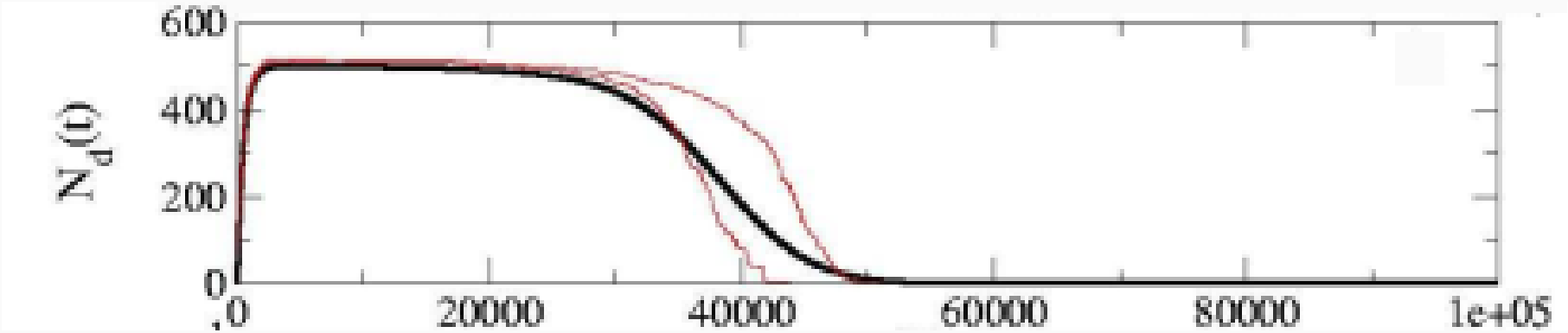
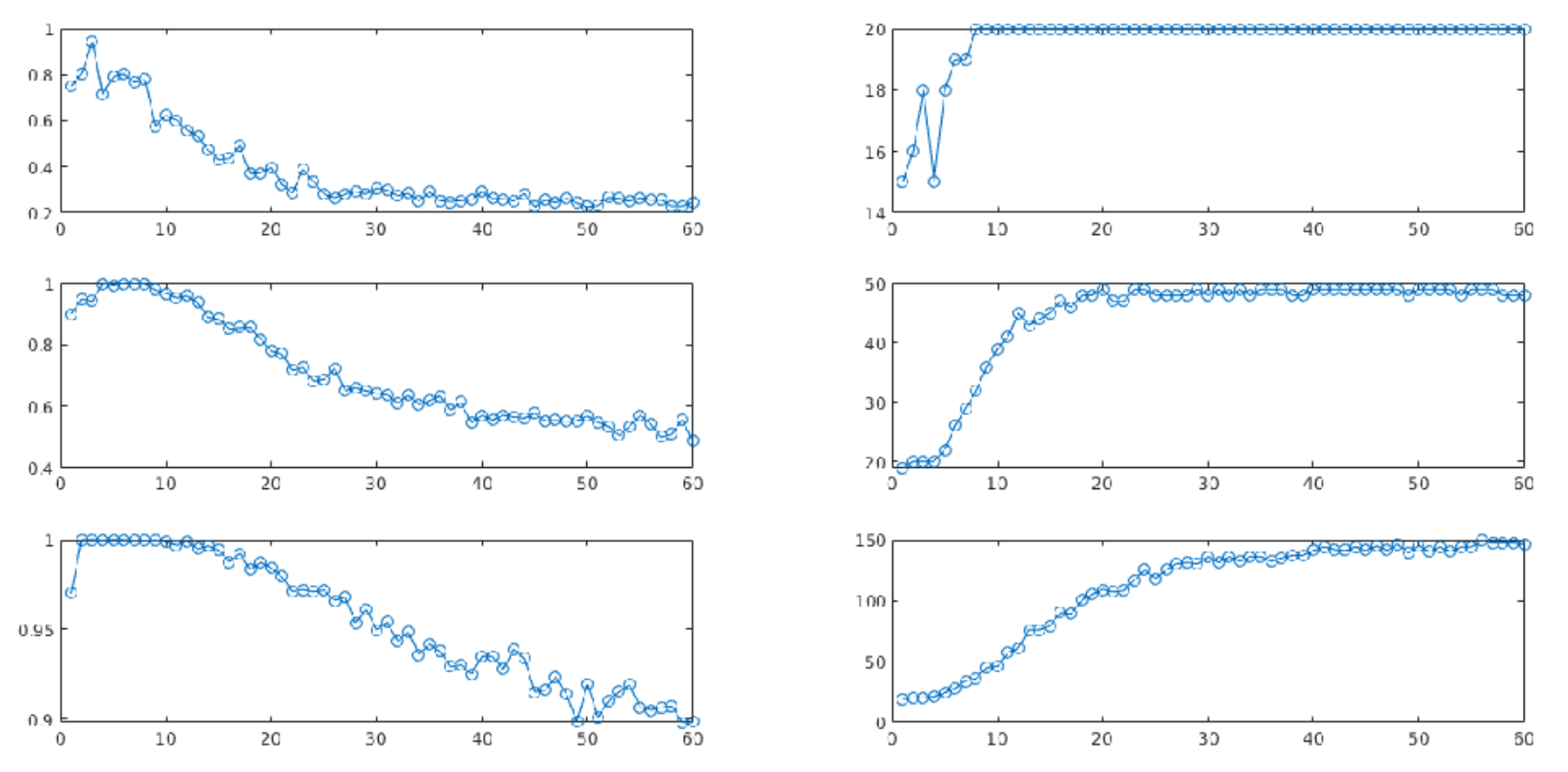


Figure: Figure from Naming Game paper

Results: $M = N = 20$, $IT = 1000$, $W = 20,50,300$



Evolution of F and D as m varies from 1 to 60 for different W

Results

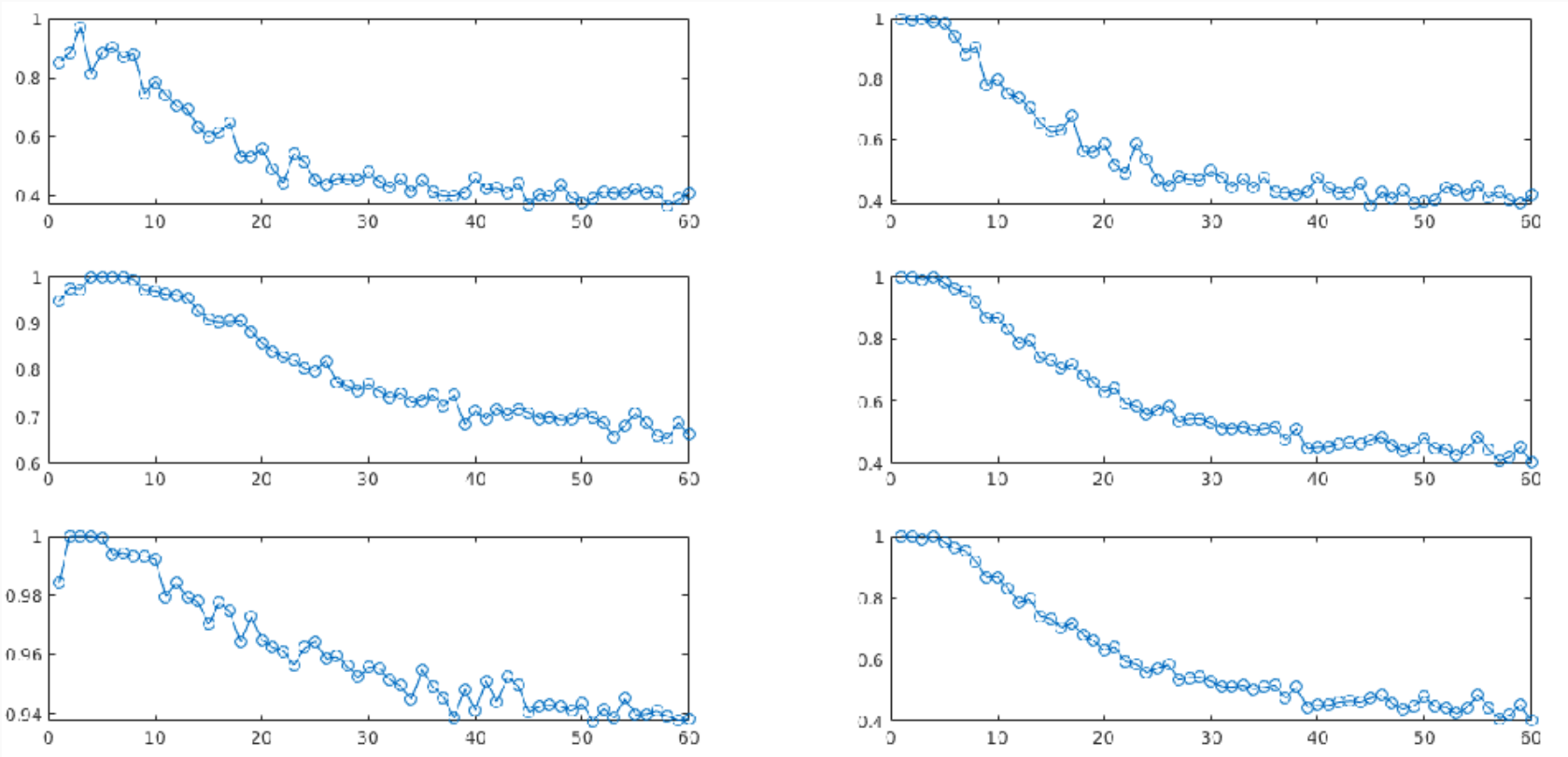
Low memory size means high success

As W get bigger, success raises due to low homonym probability

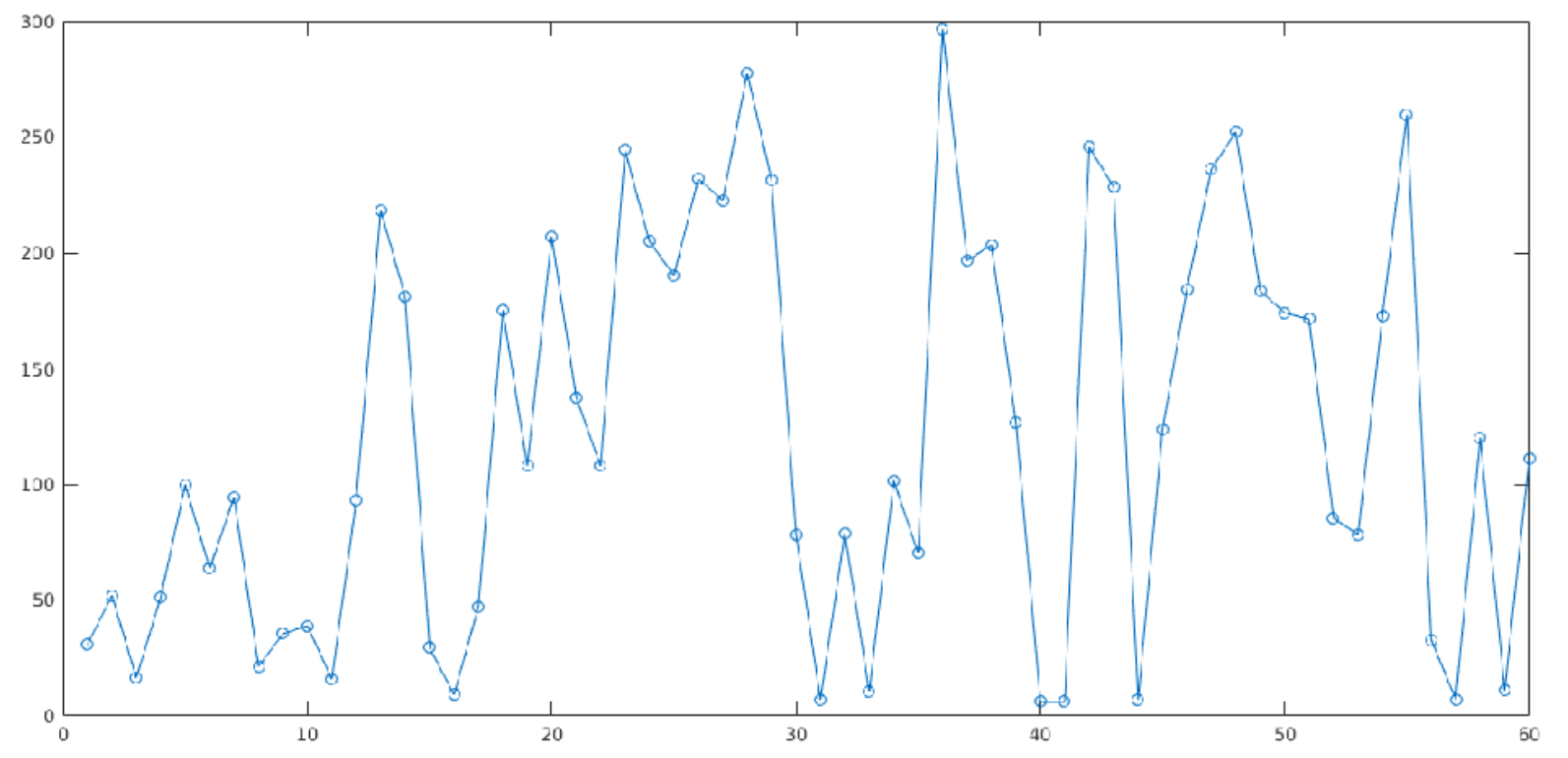
Number of words reach a maximum and stops

High memory size makes it difficult to completely overwrite words: low success(synonyms) and greater number of words

Failed Attempts: Homonyms and Synonyms



Failed Attempts: Time to Emergence



For $m = 30$ values are:

N=[53.2,93.5,42,52,150.1,199.5,150.2,11.6,249.5,333.5,78.9,133.4,10.1,28.5,146]

averaged over 10 iterations

Conclusion

Low memory size increases success

Low memory size decreases self organization time

All results of Evolutionary Language Game apply

Possible further studies:

- Emergence of syntax
- Errors in perception, error limits
- Evolving environments

Questions?