Language preservation and agent-based computer simulations

25 March 2021

Katie Mudd
Peter Dekker
Introduction

- Urgency of language preservation
  - Loss of language => loss of culture
  - Language as a window into cognition

https://www.swarthmore.edu/SocSci/langhotspots/old_maps.html
Introduction

ABACUS: Advancing Behavioral and Cognitive Understanding of Speech
Thanks to

Universiteit Leiden
Centre for Linguistics

Lingua Pax

KU Leuven
CIPL

University of Cologne

Society for Endangered Languages

European Free Alliance
Today’s agenda

● Introduction by Bart de Boer
● What is agent-based modeling?
● Break
● Agent-based modelling software
● Case study Kata Kolok, Indonesia (Katie Mudd)
● Participant case study: Yurakaré, Bolivia (Sonja Gipper, GBS)
● Break
● Participant case study: Hadza, Tanzania (Richard Griscom, LUCL)
● Participant case study: Friulian (Stefania Garlatti, EFA)
● Q&A
Introduction by Bart de Boer

● Why is linguistic diversity important?
● For (evolutionary) linguists:
  ○ A link with the past
  ○ A repository of cultural knowledge
  ○ A source of information about what humans are capable of
● For ordinary people
  ○ A link to their past
  ○ A cultural heritage to be proud of
  ○ A way to communicate to your grandkids
  ○ But also: a lot of work to learn and maintain!
● But we should not forget that language death is a fact of life
  ○ One theory says the number of languages has been falling since the invention of agriculture
  ○ And speakers have excellent, respectable reasons why they switch!
Introduction by Bart de Boer

“Language”

Collective Phenomenon

Languages move up the mountain

Individual Behaviors

People move down the mountain
Introduction by Bart de Boer

- Collective behavior is hard to predict
  - Decisions in the best interest of the individual may not lead to optimal collective behavior
  - Outcomes may be counterintuitive

- Agent-based models are tools to help deal with this
  - The VUB AI-lab has been doing this since 1995

- So far: more “fundamental” questions
  - But I am delighted that Katie and Peter are thinking about how to apply them to the worthy cause of language preservation
Flattening the curve

Covid19 Modeling the flattening of the curve (Smaldino)

A generic model of disease transmission: SIR (Susceptible - Infected - Recovered)

Covid19 Modeling the flattening of the curve (Smaldino)
Why model?

- We are bad at understanding complex systems
- Articulate a system and all its parts
- What are the relationships between the parts?

(Verbal model) → (Formal model)

(Images of verbal and formal models)

(Smaldino, 2016)
Why model?

- Predicting
- Explaining
- Guiding data collection
- Discover new questions
- Offer crisis options in near-real time
- Reveal the apparently simple to be complex
- Non-invasive!

(Epstein, 2008)
Agent-based modeling (ABM)

- Only constraint is programming it - very open framework!
- Focal point is the individual

Local interactions

Global phenomenon
ABMs

- Agents
  a. Internal data representations
  b. Means for modifying their internal data representations
  c. A fixed set of rules they must follow
ABM outline

Purpose

Input
ex. number of speakers

State of the world
Initialization
ex. network

Time step
ex. marriage
ex. interactions
Language in ABMs

Language games

Language as a variable

Before interaction

After interaction
(Agent-based) models for endangered languages

Abrams & Strogatz (2003). Modelling the dynamics of language death

Kandler et al. (2010). Language shift, bilingualism and the future of Britain's Celtic languages

de Bie & de Boer (2007). An ABM of linguistic diversity

Civico (2019). The dynamics of language minorities: Evidence from an ABM of language contact
ABMs and the real world

- Agent-based model can describe current situation, to get better insight
- ABM can evaluate effects of policy measures, but:
  - Agent-based model should not be the only evidence
  - Further (empirical) research is needed before changing policy
- We as researchers evaluate policy alternatives
  - Others can choose one alternative and implement/advocate
- We as language/AI researchers see ourselves best fit to model effects that directly influence language (e.g. education, marriage patterns)
  - Dynamics of political/economical systems on language are difficult to model

https://www.youtube.com/watch?v=wKOslhIFt6U
ABMs of specific languages

- We saw how ABMs work in general
- How to apply them to specific case studies of language endangerment?
- Case studies in second half:
  - Kata Kolok, Indonesia (Katie Mudd, VUB)
  - Yurakaré, Bolivia (Sonja Gipper, GBS)
  - Hadza, Tanzania (Richard Griscom, LUCL)
  - Friulian, Italy (Stefania Garlatti, EFA)
Questions?
Break

Implementing an agent-based model

- Coding from scratch (in any programming language)
- Coding with the help of software packages
  - Netlogo
  - Mesa (Python)
- Other software packages (not discussed today):
  - AgentPy (Python)
  - RNetlogo (R programming language)
Netlogo

- Visual design of models + own programming language
- Relatively little programming knowledge needed
- Desktop client (development/viewer) and online (viewer)
Netlogo desktop client

globals [ max-sheep ]; don't let the sheep population grow too large
; Sheep and wolves are both breeds of turtles
breed [ sheep a-sheep ]; sheep is its own plural, so we use "a-sheep" as the singul
breed [ wolves wolf ];
turtles-own [ energy ]; both wolves and sheep have energy
patches-own [ countdown ]; this is for the sheep-wolves-grass model version

to setup
  clear-all
  ifelse netlogo-web? [ set max-sheep 10000 ] [ set max-sheep 30000 ]
  ; Check model-version switch
  ; if we're not modeling grass, then the sheep don't need to eat to survive
  ; otherwise each grass' state of growth and growing logic need to be set up
  ifelse model-version = "sheep-wolves-grass"
    ask patches [
Netlogo Web: Honeycomb
Mesa

- Library for Python programming language
  - Use of other Python libraries (e.g. math, graphs)
  - Data analysis of results in Python
- No visual development: everything programmed
- Interactive visualization in browser
- Can be combined with Jupyter notebook for results analysis

https://mesa.readthedocs.io/
Now, we set up the batch run, with a dictionary of fixed and changing parameters. Let’s hold everything fixed except for homophily:

```python
In [23]: parameters = {"height": 10, "width": 10, "density": 0.8, "minority_pc": 0.2, "homophily": range(1,9)}
In [24]: model_reporters = {'Segregated_Agents': get_segregation}
In [25]: param_sweep = BatchRunner(SchellingModel, parameters, iterations=10, max_steps=200, model_reporters=model_reporters)
In [26]: param_sweep.run_all()
In [27]: df = param_sweep.get_model_vars_dataframe()
In [28]: plt.scatter(df['homophily'], df['Segregated_Agents'])
plt.grid(True)
```
class MoneyAgent(Agent):
    """ An agent with fixed initial wealth."""
    def __init__(self, unique_id, model):
        super().__init__(unique_id, model)
        self.wealth = 1

    def move(self):
        possible_steps = self.model.grid.get_neighborhood(
            self.pos,
            moore=True,
            include_center=False)
        new_position = self.random.choice(possible_steps)
        self.model.grid.move_agent(self, new_position)

    def give_money(self):
        cellmates = self.model.grid.get_cell_list_contents([self.pos])
        if len(cellmates) > 1:
            other = self.random.choice(cellmates)
            other.wealth += 1
            self.wealth -= 1

    def step(self):
        self.move()
        if self.wealth > 0:
            self.give_money()

class MoneyModel(Model):
    """ A model with some number of agents."""
    def __init__(self, N, width, height):
        self.num_agents = N
        self.grid = MultiGrid(width, height, True)
        self.schedule = RandomActivation(self)
        # Create agents
        for i in range(self.num_agents):
            a = MoneyAgent(i, self)
            self.schedule.add(a)
            # Add the agent to a random grid cell
            x = self.random.randrange(self.grid.width)
            y = self.random.randrange(self.grid.height)
            self.grid.place_agent(a, (x, y))

    def step(self):
        self.schedule.step()
from MoneyModel import *
from mesa.visualization.modules import CanvasGrid
from mesa.visualization.ModularVisualization import ModularServer

def agent_portrayal(agent):
    portrayal = {
        "Shape": "circle",
        "Filled": "true",
        "Layer": 0,
        "Color": "red",
        "r": 0.5
    }
    return portrayal

grid = CanvasGrid(agent_portrayal, 10, 10, 500, 500)
server = ModularServer(MoneyModel,
                        [grid],
                        "Money Model",
                        {"N":100, "width":10, "height":10})
server.port = 8521 # The default
server.launch()
Forest fire example
Questions?
Case study: Kata Kolok

- Emerged in rural community in Bali
- 7 generations ago
- Deafness is a recessive trait
- +/- 2200 individuals
- 47 deaf (2008)
- +/- 70% know the sign language
- Adapted to deafness!
  - SL acquisition from birth
Case study: Kata Kolok

Study mechanisms:
- How gene and culture influence
- SL persistence

Predict result if different marriage patterns
Case study: Kata Kolok

Genetic and cultural transmission

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AA</td>
<td>Aa</td>
</tr>
<tr>
<td>a</td>
<td>aA</td>
<td>aa</td>
</tr>
</tbody>
</table>
Case study: Kata Kolok
Yurakaré

- 6,000 individuals identify as Yurakaré
  - 1600 speak Yurakaré
  - most bilingual, but most in younger gen monolingual in Spanish
  - Break in intergenerational transmission => endangered
- Official language of Bolivia (of 37)
- Linguistic isolate
- 16th century: contact with Spanish-speaking colonizers
- 1937: Spanish education
- Media and radio in Spanish
- 1994: intercultural bilingual education reform (failed)
- Constitution of 2009: recognizes all indigenous languages as official (but teaching is still in Spanish)
- Yurakaré language is an important part of identity
Yurakaré references


Yurakaré

- attitudes about Spanish vs. indigenous languages
- intergenerational language transmission
- important part of identity
- recognized as an official language
- media
- education in Spanish

Van Gijn, Hirtzel & Gipper (2010)
Yurakaré model example

What is the effect of intergenerational language transmission on the persistence of Yurakaré?

Sonja Gipper
https://www.pinterest.ca/pin/51439620728816348/
Hadza

- **Area:** Rift Valley, Tanzania
- **Language family:** isolate.
- **Status:** not an official language; Swahili and English used in schools and media.
- **Number of speakers:** 1,000-2,000.
- **Society:** non-hierarchical social structure and egalitarian social practices.

https://rgris.com/hadzabe/
Hadza

- **Factors contributing to endangerment:**
  - Shift from nomadic foraging/hunting to sedentary lifestyles in small villages, caused by:
    - loss of natural habitat as others claimed the land
    - loss of wildlife due to poaching and increasing urbanization
    - increase in tourism
  - Disruption of traditional practices associated with specialized language use
  - Language loss (especially northwest)

- **Steps to prevent language extinction:**
  - No formal language preservation organization within community
  - Community access is problematic: scarce access to internet, computers, or mobile phones
  - Creation of large collection of audio-visual recordings, to preserve knowledge of elders
  - Development of educational programming complicated by decentralized geographical distribution ➔ Decentralized education?

https://rgris.com/hadzabe/
Hadza

- loss of habitat
- loss of wildlife
- increase in tourism
- shift from nomadic to sedentary
  - disruption of traditional practice
  - language loss
  - loss of specialized language use

poaching

urbanization
Hadza

- Setting: Language game or "language as a variable" model?
- Environment representation:
  - Spatial grid? Represents habitat loss
  - Network structure? Represents decentralized contact between speakers
- Add interventions:
  - Recordings collection
  - Decentralized educational program

https://rgris.com/hadzabe/
Friulian

- **Area**: Friuli region, northeastern Italy
- **Language family**: Rhaeto-Romance, Romance, Indo-European
  - **Related**: Romansh (Switzerland) and Ladin (Italy)
- **Status**: recognised as minority language. Limited use in schools, by local authorities (place names) and media
- **Number of speakers**: 420,000 daily + 180,000 occasional. But:
  - differing fluency levels
  - not used across all social domains
- **History**:
  - Friulian historically used in in all social strata
  - 19th and 20th century: pressure to switch to Italian
  - Last 20-30 years: minority language status and revitalization efforts
Friulian

- **Threats**
  - Population loss in rural areas; work in Italian-speaking urban areas
  - Low social status → children not brought up in Friulian
  - Popularity Italian-language media

- **Revitalization efforts (20-30 years)**
  - Social status improved
    - Language tools: dictionary, spellchecker, etc.
  - Still decrease in fluency and number of speakers

- **Possible future interventions**
  - Engaging type of schooling (teacher certification); speaking at home not enough
  - Media presence
  - Economic value
    - Friulian product marketing
    - Language requirement for public service
Friulian

- Each speaker has fluency level (basic knowledge vs fluent)
- Each speaker has probability to use language in different domains (e.g. work, family)
- Social status
  - Increase social status of language:
    - Media
    - Education
    - Economic value
  - More social status $\rightarrow$ more parent-child transmission?
- Bilingualism
Resources

**MODELS ARE STUPID, AND WE NEED MORE OF THEM**

*Paul E. Smaldino*

NetLogo comes with a large library of sample models. Click on some examples below.
Wrap up & what next?

● Agent-based models simulate interactions between speakers
● Open framework: you decide what you put in the model
● Can be used to simulate language endangerment & evaluate policy measures
  ○ Further research is needed before implementing measures
● What next?
  ○ Implement models yourself
  ○ Collaborate with AI/computer science/social science researchers
  ○ Contact us if you have questions!
    ▪ katie.mudd@ai.vub.ac.be
    ▪ peter.dekker@ai.vub.ac.be
References


Questions?