Language preservation and agent-based computer simulations

25 March 2021

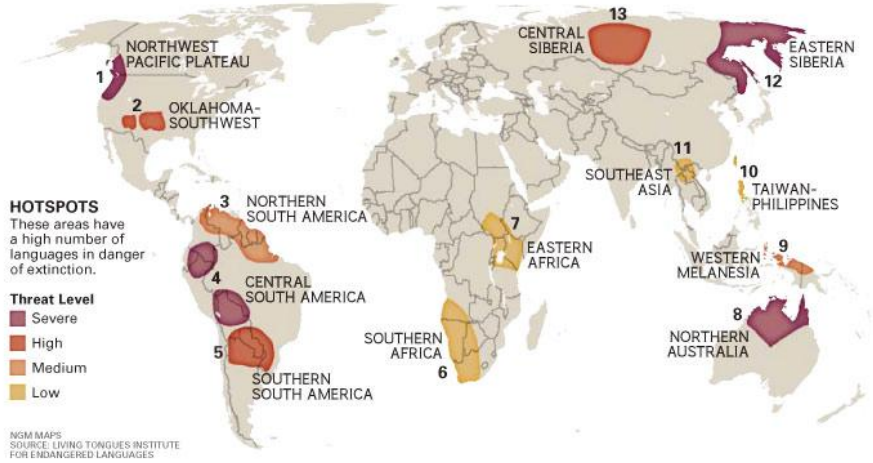
Katie Mudd
Peter Dekker

Introduction

Upcoming Decade of Indigenous Languages (2022 – 2032) to focus on Indigenous language users' human rights

2019 INTERNATIONAL YEAR OF Indigenous Languages

- 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

VUB ARTIFICIAL INTELLIGENCE RESEARCH GROUP

Thanks to

Universiteit Leiden
Centre for Linguistics

Lin gua Pax

KU LEUVEN

CIPL

University of Cologne

Society for Endangered Languages

European Free Alliance

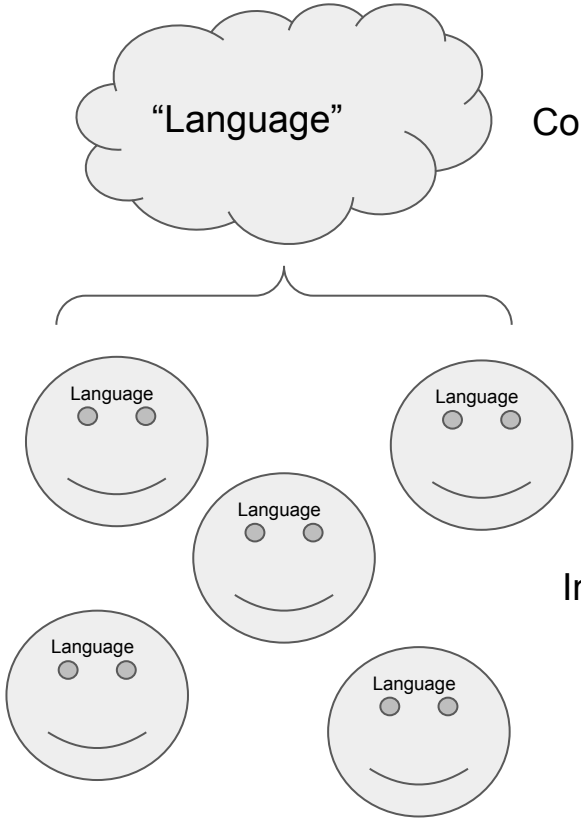
Today’s agenda

- 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

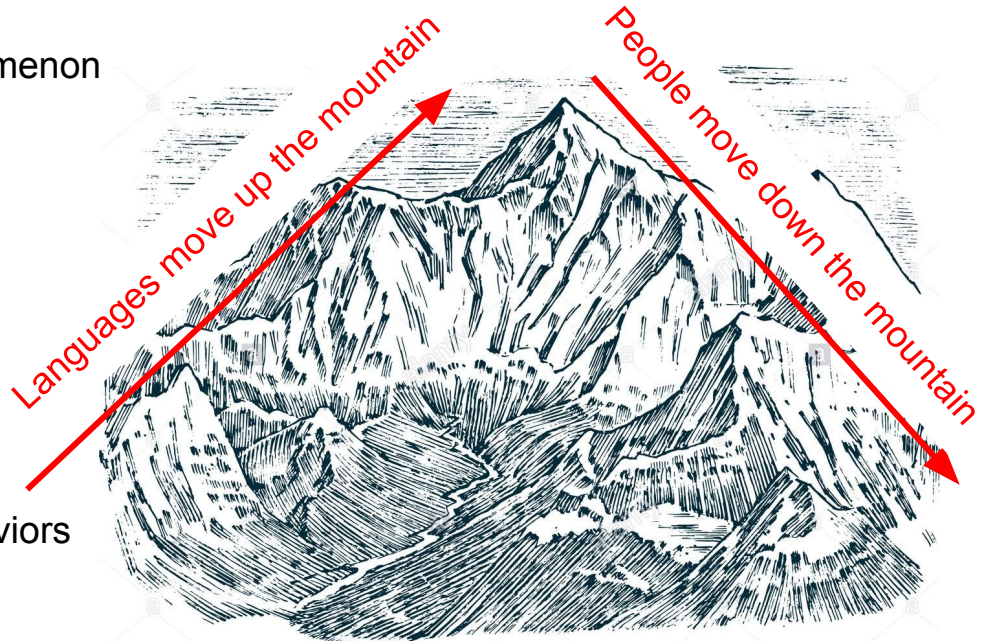
- Why is linguistic diversity important?
- For (evolutionary) linguists:
  - A link with the past
  - A repository of cultural knowledge
  - A source of information about what humans are capable of
- For ordinary people
  - A link to their past
  - A cultural heritage to be proud of
  - A way to communicate 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

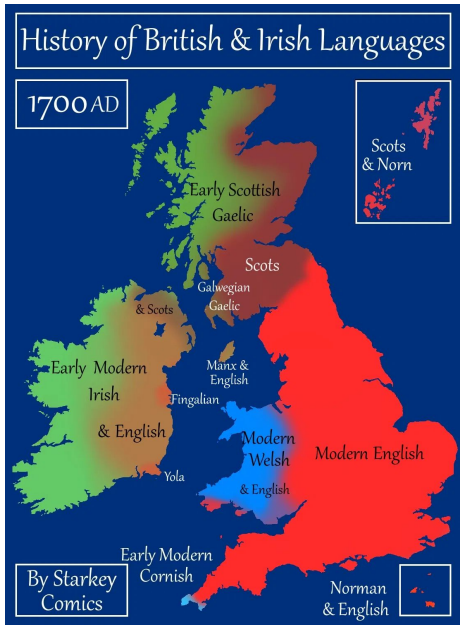


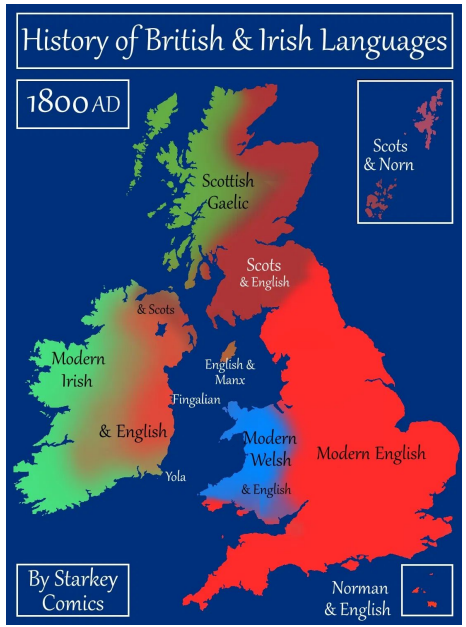
Collective Phenomenon

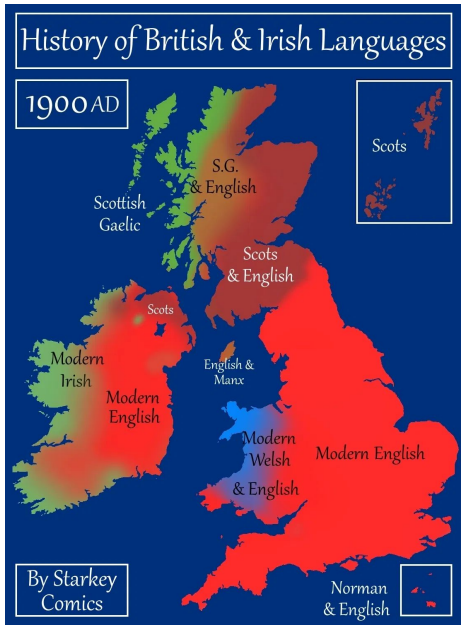
Individual Behaviors

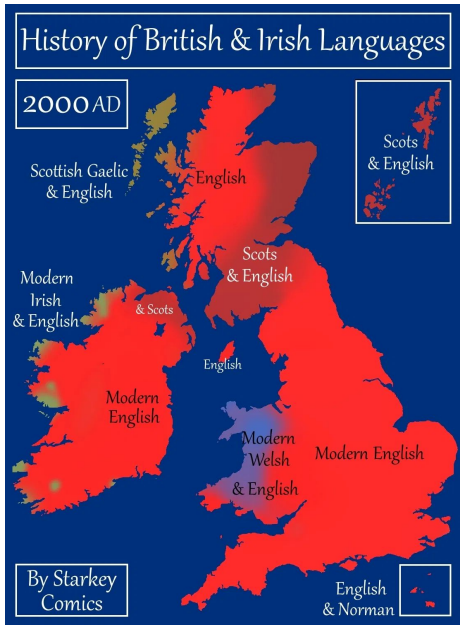


Introduction by Bart de Boer



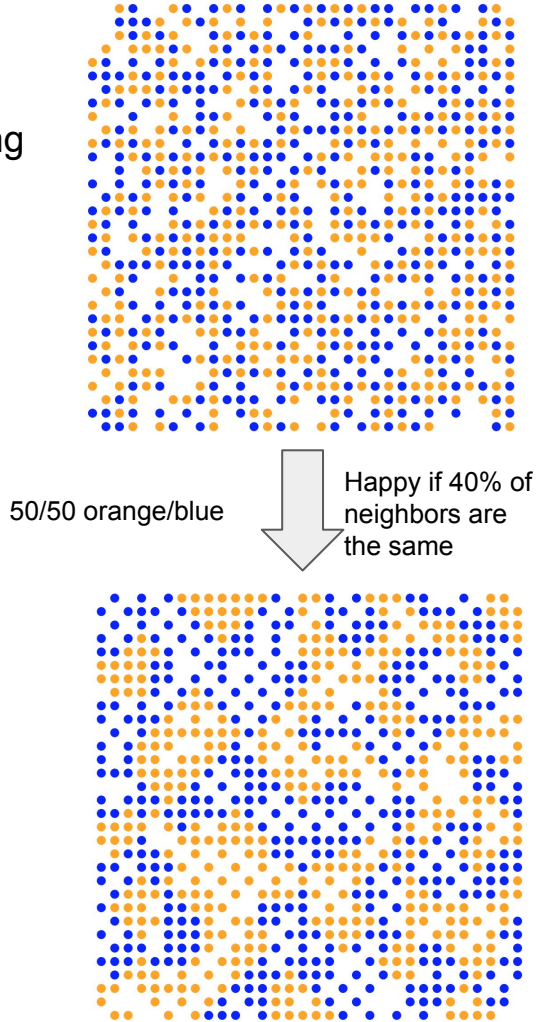






Introduction by Bart de Boer

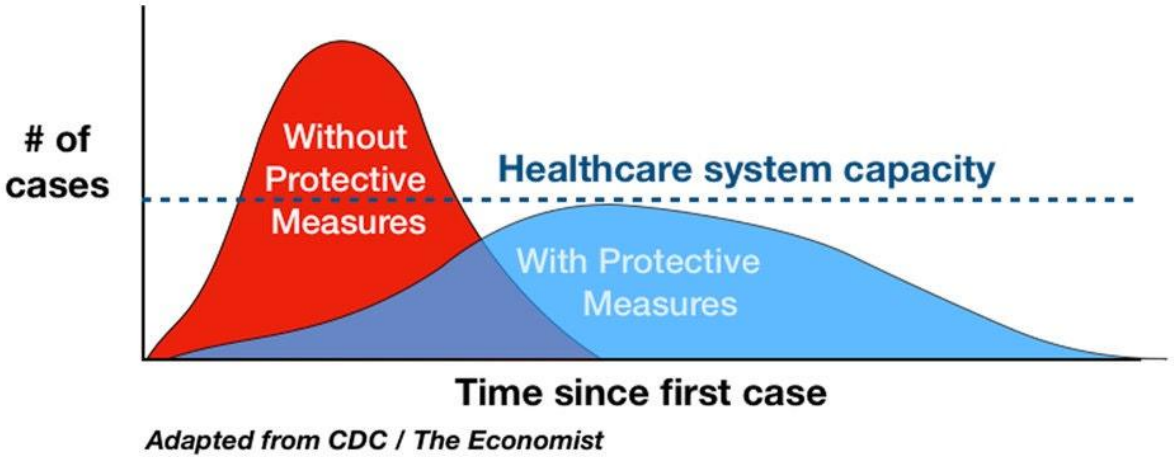
The Schelling game,
After Thomas Schelling



- 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

AI FLANDERS
BUILDING OUR DIGITAL FUTURE

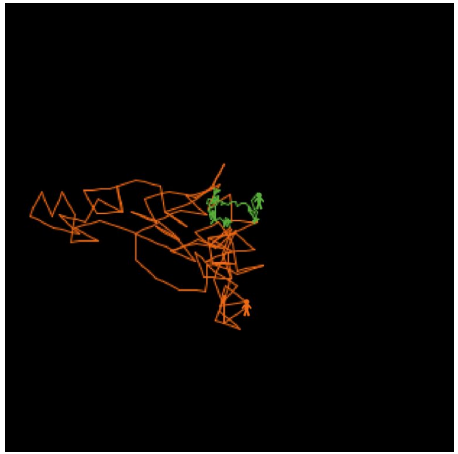
Flattening the curve

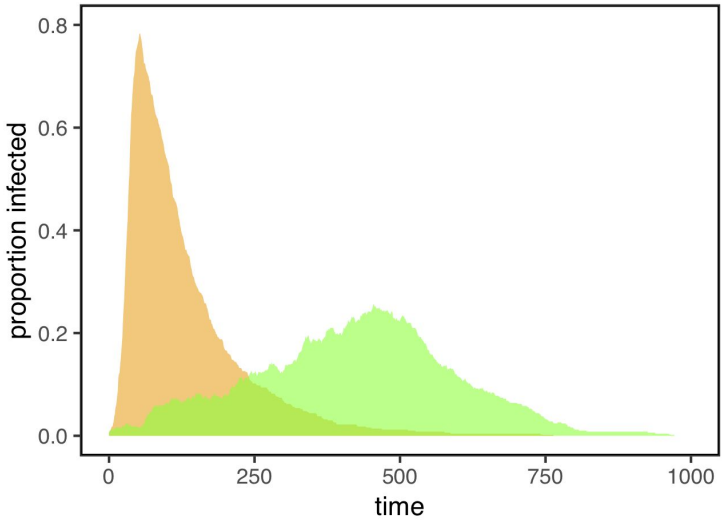


https://www.nytimes.com/article/flatten-curve-coronavirus.html

Covid19 Modeling the flattening of the curve (Smaldino)

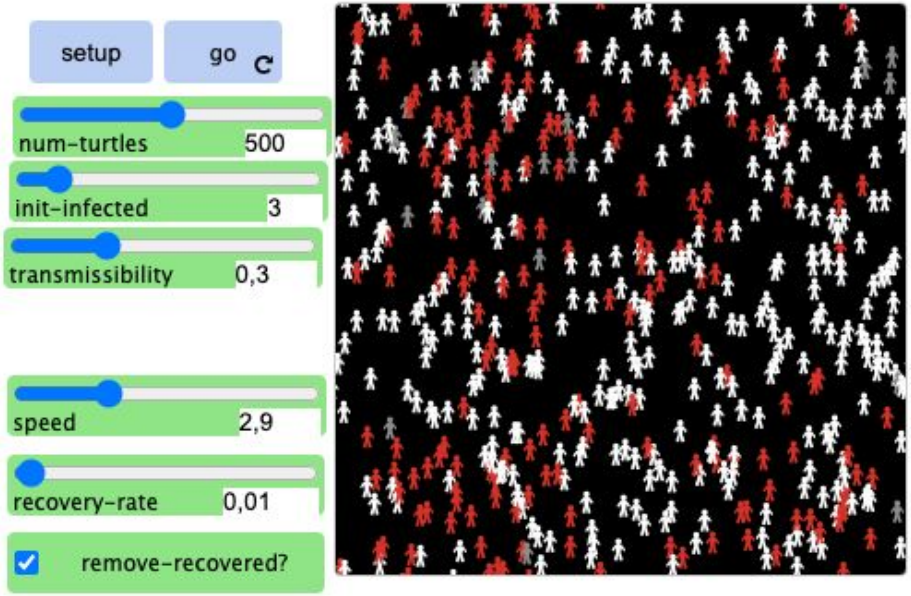
A generic model of disease transmission: SIR (Susceptible - Infected - Recovered)

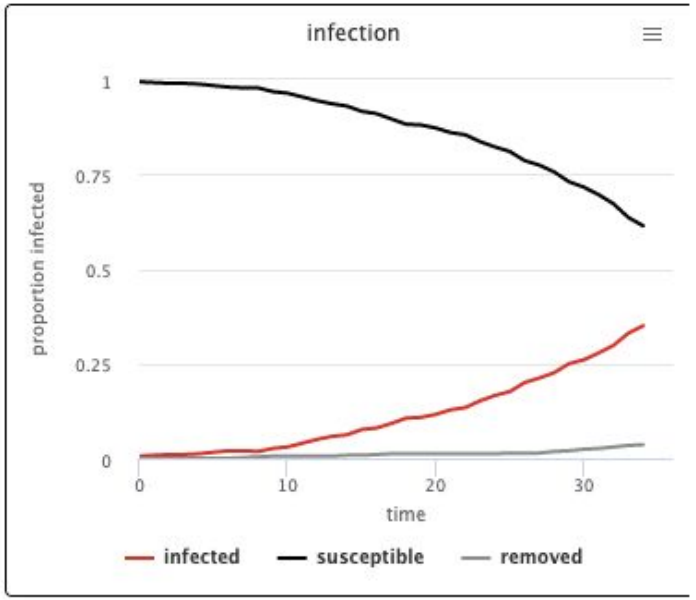




http://smaldino.com/wp/covid-19-modeling-the-flattening-of-the-curve/

Covid19 Modeling the flattening of the curve (Smaldino)

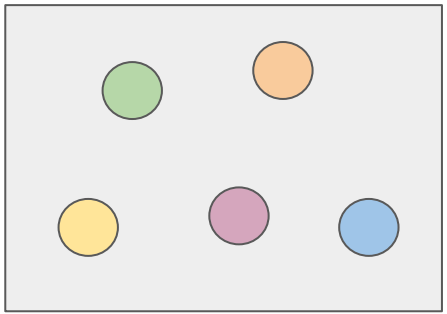


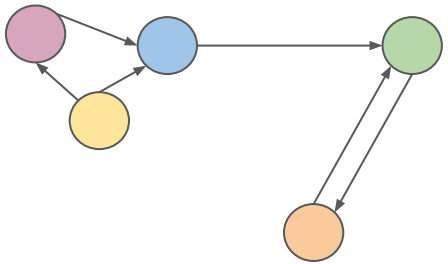


Why model?

Verbal model → Formal model

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





(Smaldino, 2016)

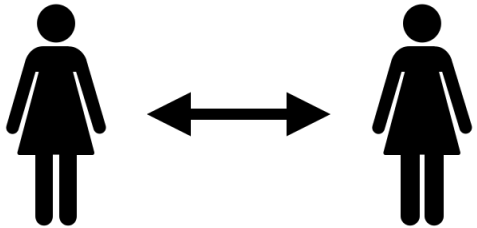
Why model?

- 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

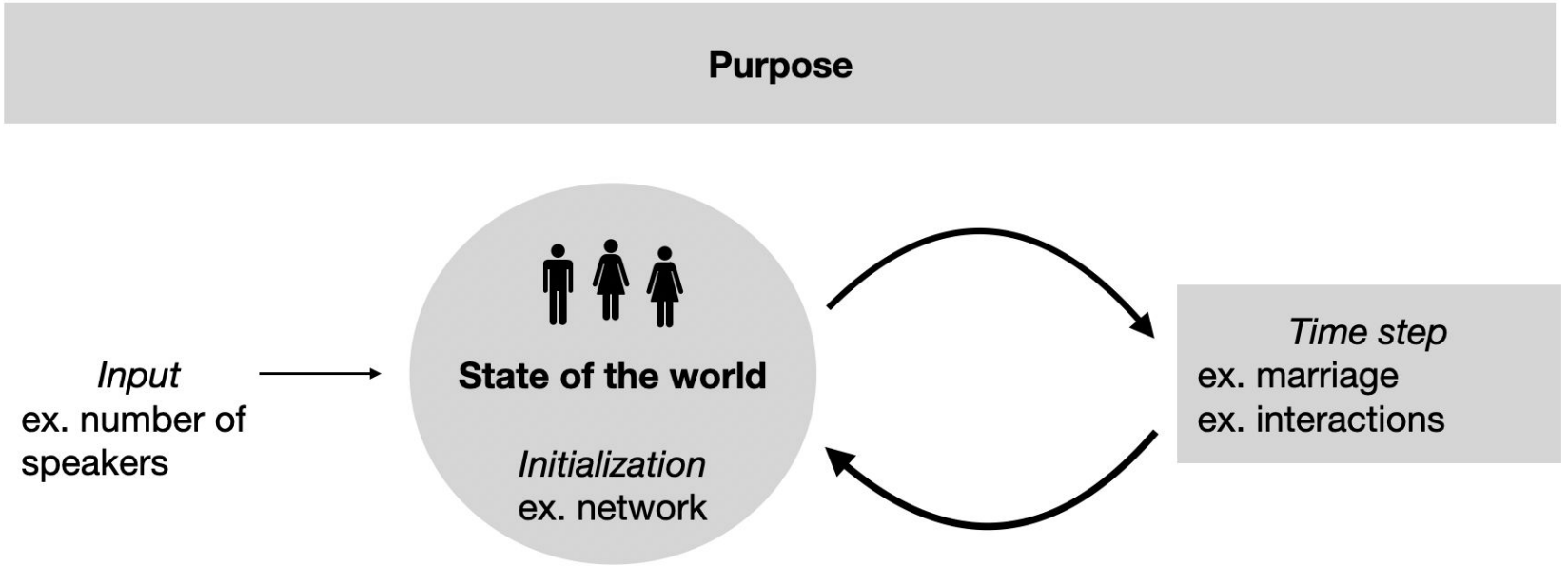
ABMs

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

Before interaction

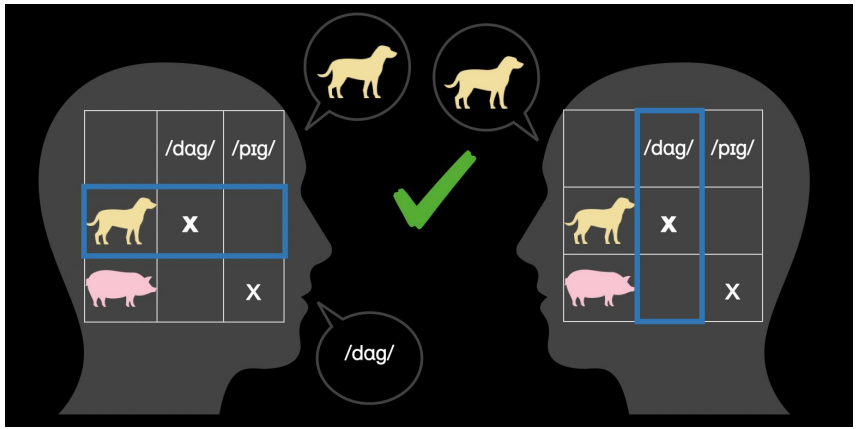
After interaction

ABM outline

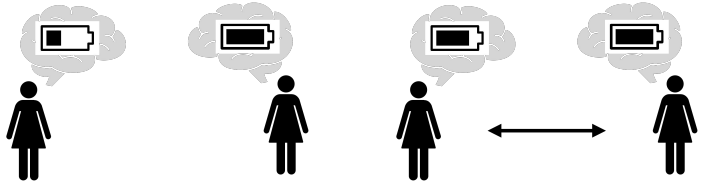

Purpose
Input
ex. number of speakers
State of the world
Initialization
ex. network
Time step
ex. marriage
ex. interactions

Language in ABMs

Language games



Language as a variable

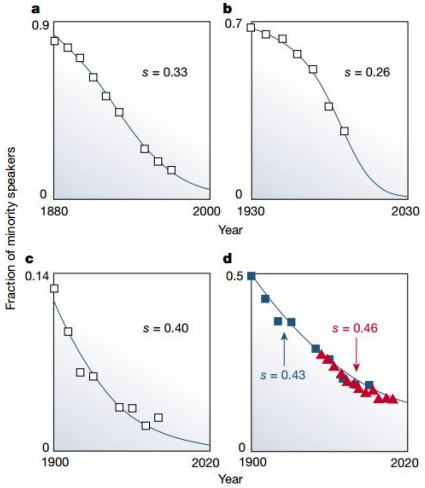


Before interaction

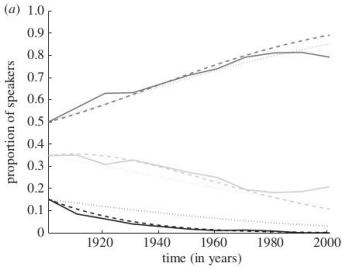
After interaction

(Agent-based) models for endangered languages

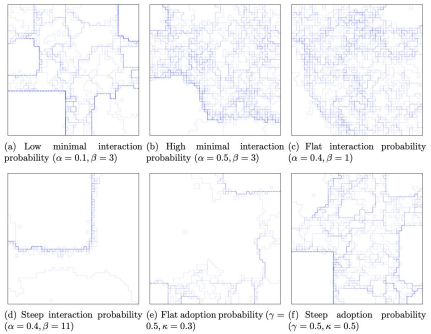
Abrams & Strogatz (2003). Modelling the dynamics of language death



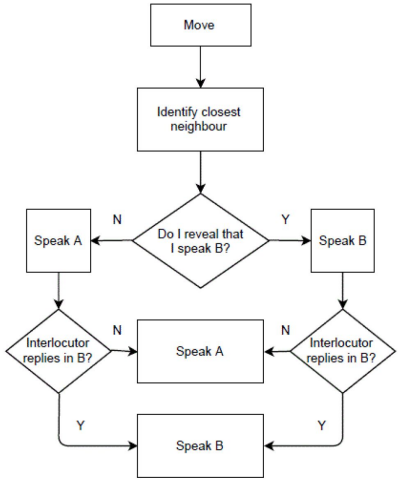
Kandler et a. (2010). Language shift, bilingualism and the future of Britain's Celtic languages



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

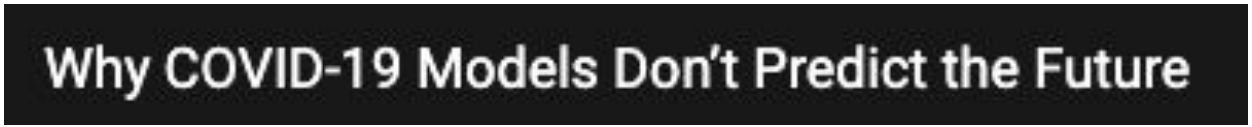


Civico (2019). The dynamics of language minorities: Evidence from an ABM of language contact



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

https://www.sciencelearn.org.nz/images/2041-six-types-of-tea

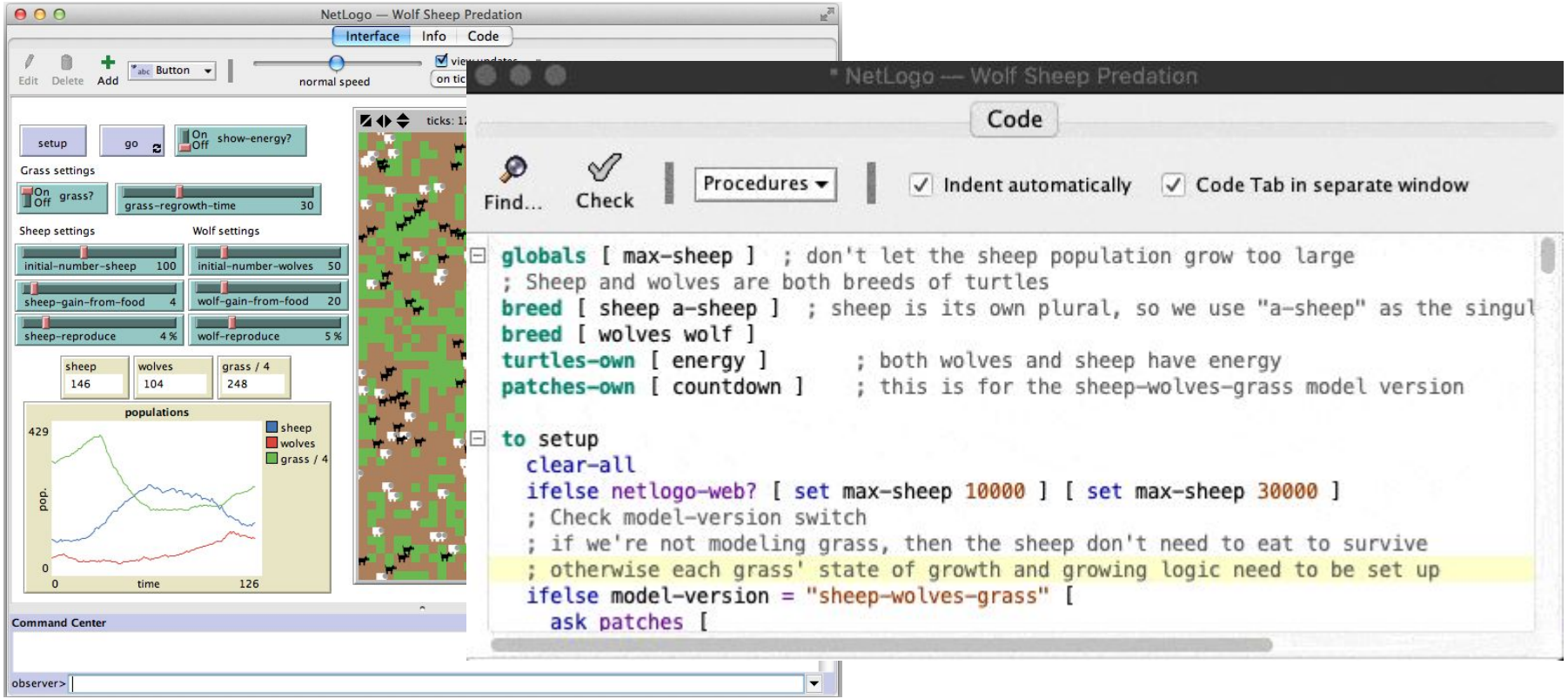
Implementing an agent-based model

- 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)

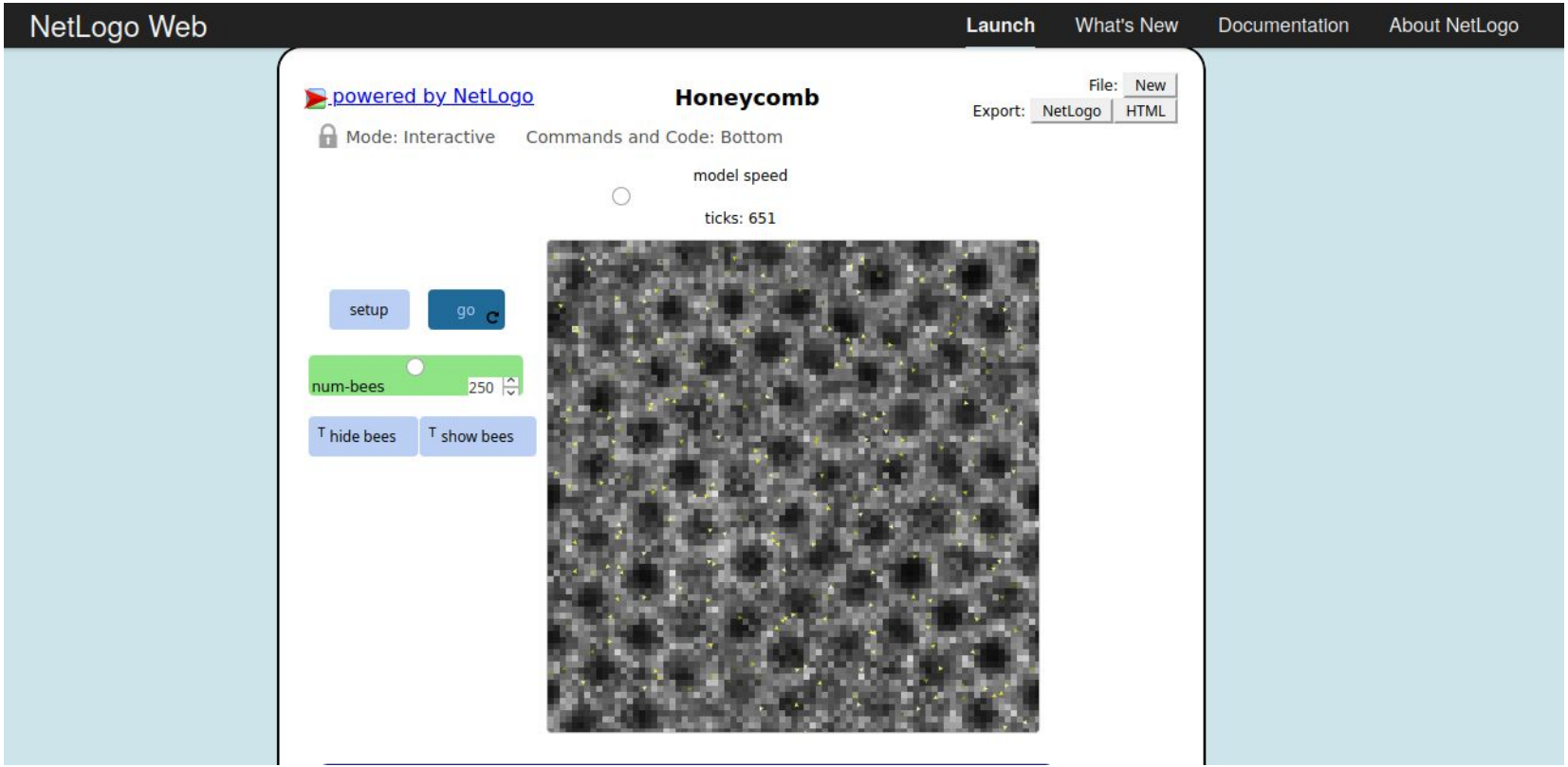
Netlogo

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

Netlogo desktop client



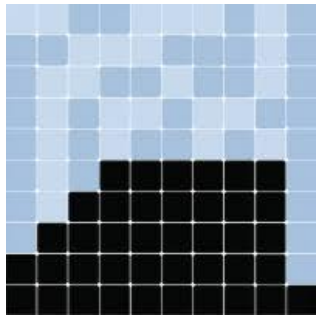
Netlogo Web: Honeycomb



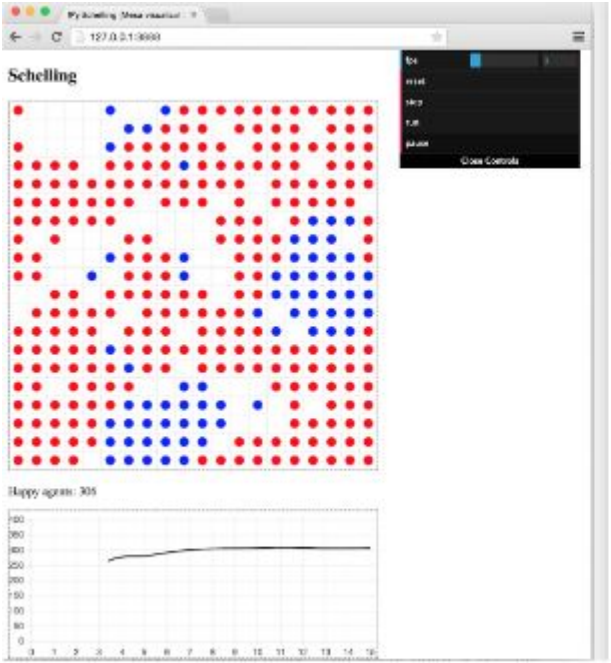
Mesa

- 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/

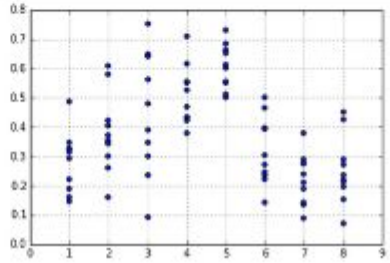


Browser visualization & Jupyter notebook



Now, we set up the batch run, with a dictionary of fixed and changing parameters. Let's hold everything fixed except for Homophily.

```
In [13]: parameters = {'height': 10, 'width': 10, 'density': 0.8, 'minority_pc': 0.2
                       'homophily': range(1,9)}

In [14]: model_reporters = {'Segregated_Agents': get_segregation}

In [24]: param_sweep = BatchRunner(SchellingModel, parameters, iterations=10,
                                   max_steps=200,
                                   model_reporters=model_reporters)

In [25]: param_sweep.run_all()

In [26]: df = param_sweep.get_model_vars_dataframe()

In [28]: plt.scatter(df.homophily, df.Segregated_Agents)
         plt.grid(True)
```



Mesa tutorial: Model code

https://mesa.readthedocs.io/en/stable/tutorials/intro_tutorial.html

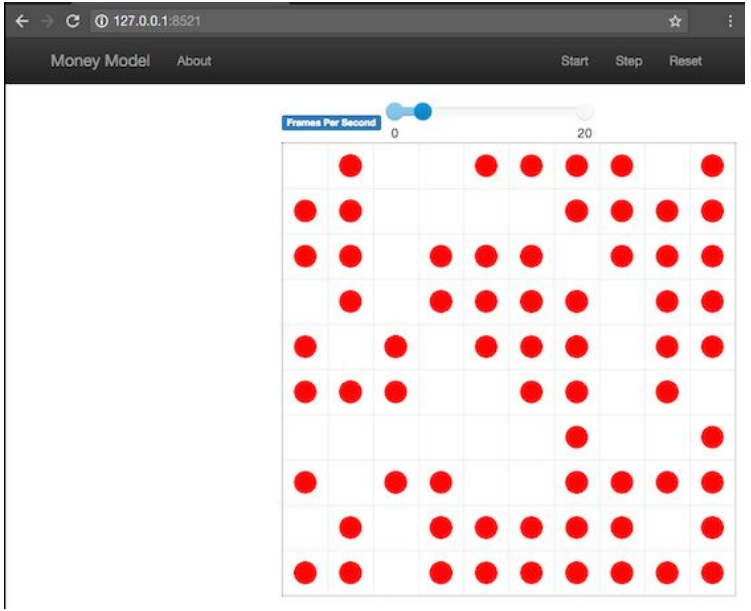
```python
class MoneyAgent(Agent):
    """ An agent with fixed initial wealth."""
    def __init__(self, unique_id, model):
        super().__init__(unique_id, model)
        self.wealth = 1

    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()
```

```python
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()
```

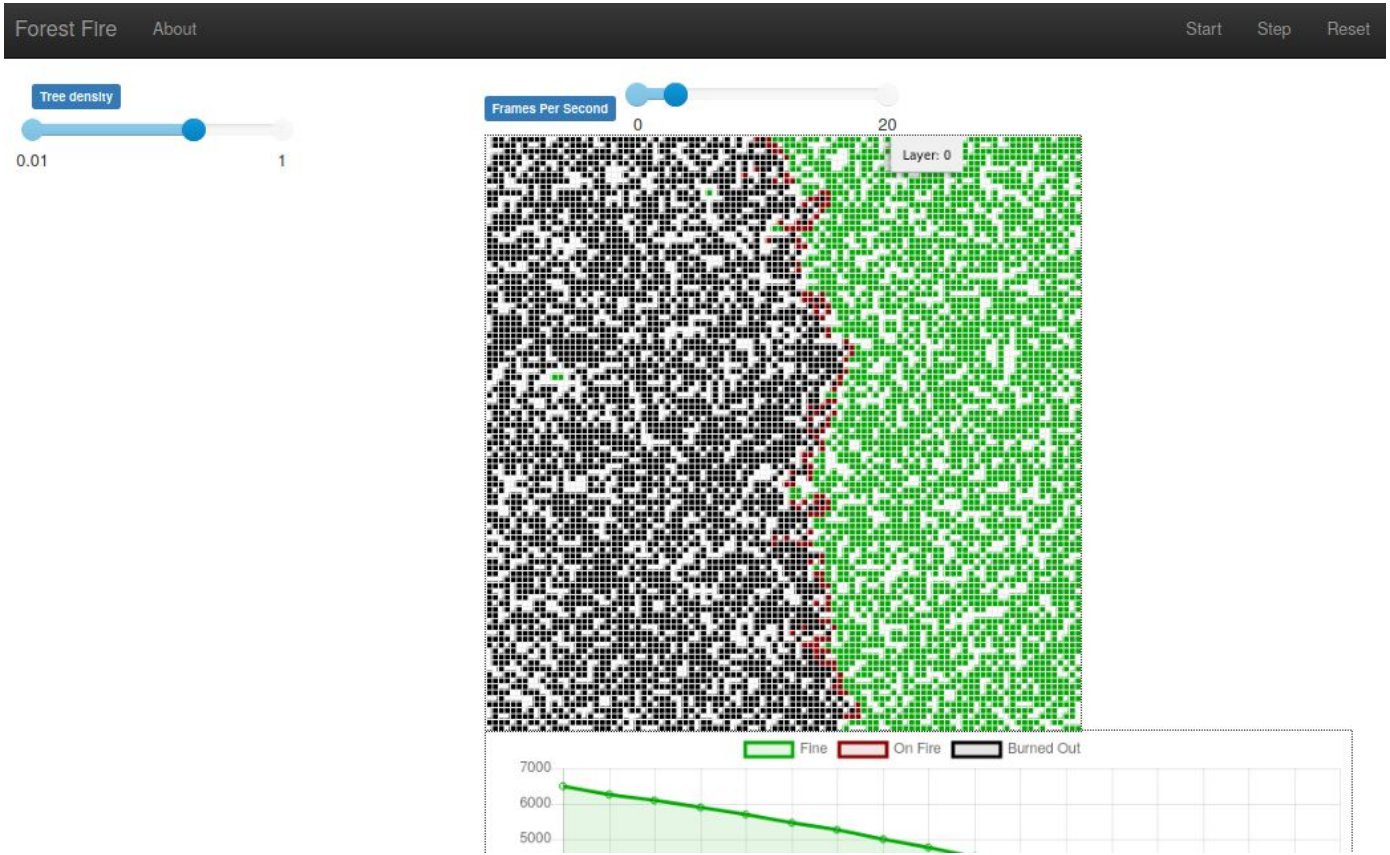
Mesa tutorial: Visualization

https://mesa.readthedocs.io/en/stable/tutorials/adv_tutorial.html

```
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()
```

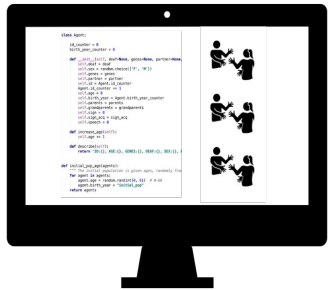


Forest fire example



Questions?

Case study: Kata Kolok



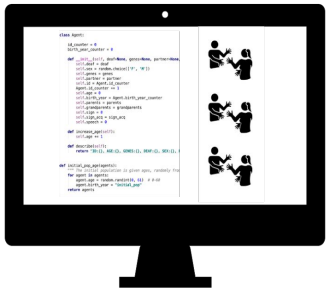
Bart de Boer

Connie de Vos

Hannah Lutzenberger

- 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

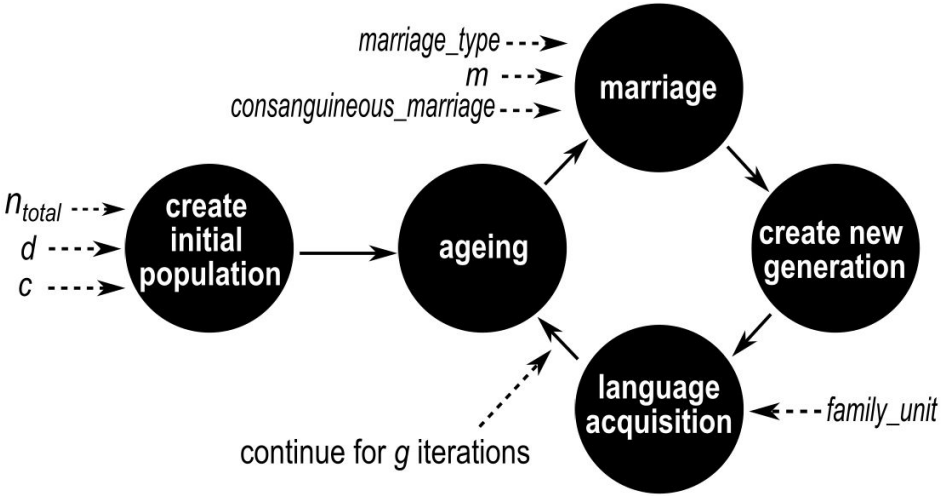
Predict result if different marriage patterns

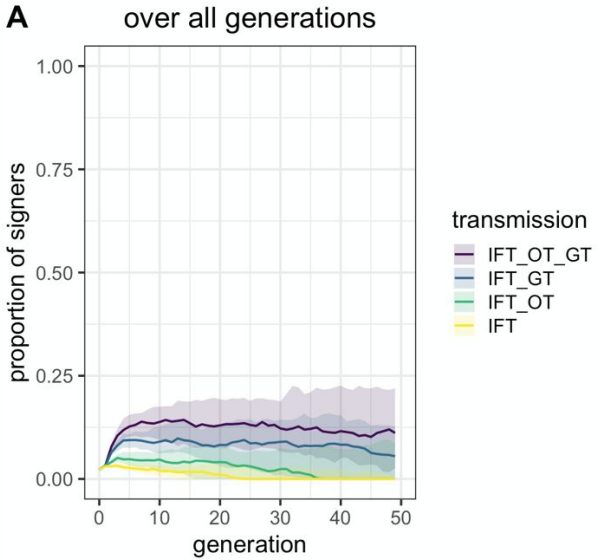
Case study: Kata Kolok

Genetic and cultural transmission

	A	a
A	AA	Aa
a	aA	aa

Case study: Kata Kolok



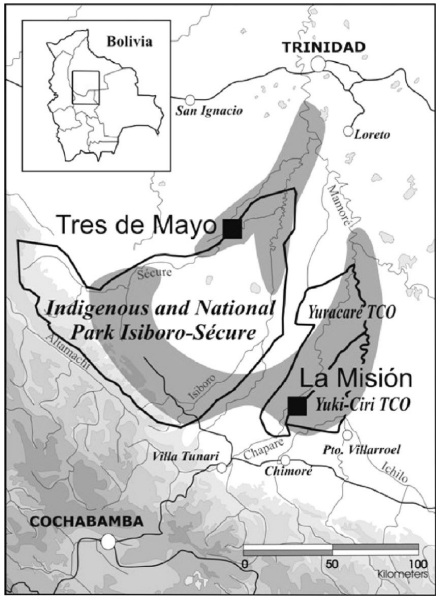


Society for Endangered Languages

Yurakaré

Sonja Gipper

- 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



Van Gijn, Hirtzel & Gipper (2010)

Yurakaré references

Apaza Apaza, Ignacio. 2012. La descolonización cultural, lingüística y educativa en Bolivia. Estudios Bolivianos 17. 156-186.

Choque C., Roberto. 1996. La educación indigenal boliviana: El proceso educativo indígena-rural. Estudios Bolivianos 2. 125–181.

CPE (Constitución Política del Estado). 2009. Constitución Política del Estado de 2009, 7 de Febrero, 2009.

Drange Danbolt, Live. 2011. The challenge of bilingualism in a multilingual society: The Bolivian case. Journal of Intercultural Communication 27. http://immi.se/intercultural/

Gijn, Rik van. 2006. A grammar of Yurakaré. Nijmegen, The Netherlands: Radboud Universiteit.

Gijn, Rik van, Vincent Hirtzel, Sonja Gipper & Jeremías Ballivián Torrico. 2011. The Yurakaré Archive. Online language documentation, DoBeS Archive, MPI Nijmegen. https://archive.mpi.nl/tla.

Hirtzel, Vincent. 2010. Le maître à deux têtes: Enquête sur le rapport à soi d'une population d'Amazonie bolivienne, les Yuracaré. Paris: École des Hautes Études en Sciences Sociales.

Hirtzel, Vincent. 2019. Informe sobre la historia de la comunidad yurakaré "LA MISIÓN": Con datos sobre la historia de la instalación de los yuquis sobre la orilla izquierda del río Ichilo. Nanterre: Unpublished manuscript.

INE (Instituto Nacional de Estadística). 2015. Censo de población y vivienda 2012 Bolivia. Características de la población. La Paz: Instituto Nacional de Estadística.

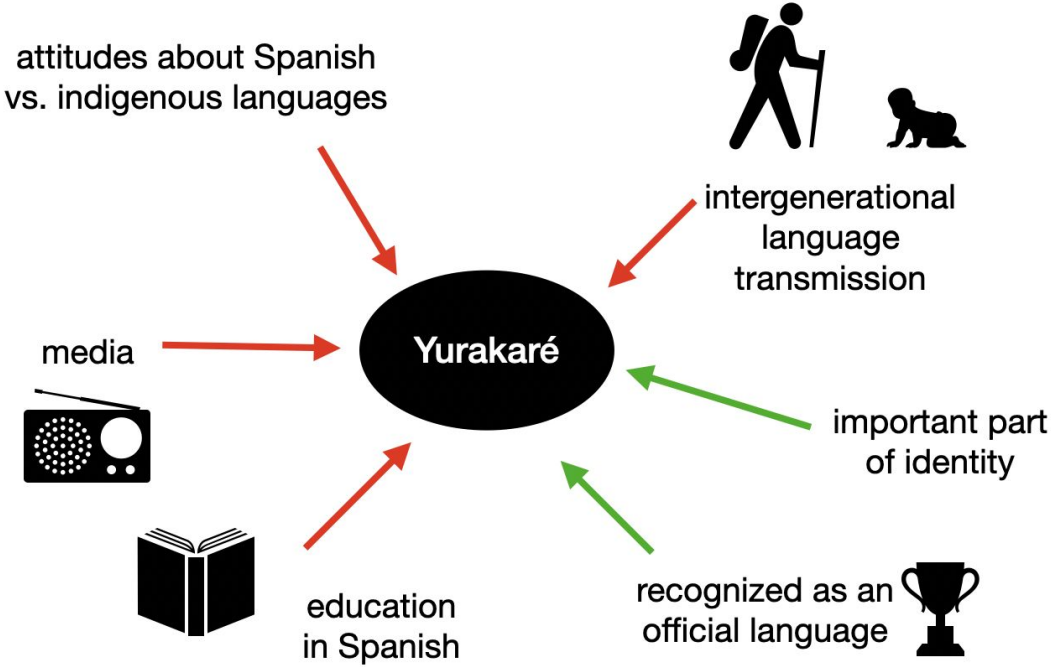
Pinto, Libertad. 2015. Diseño de una serie animada para coadyuvar al proceso de revitalización lingüística y cultural yurakaré. Cochabamba: Universidad Mayor de San Simón MA Thesis.

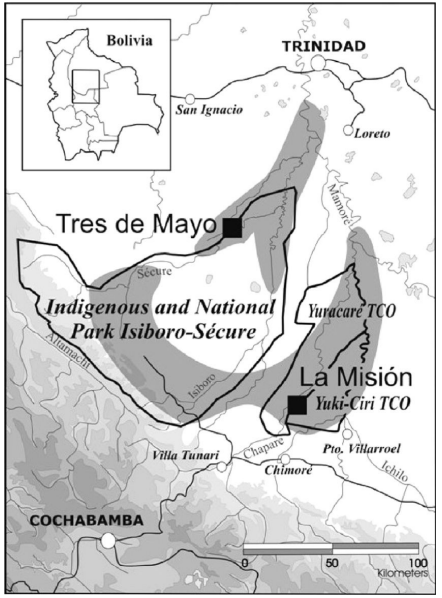
Plaza Martínez, Pedro (coord.). 2011. Historia, lengua, cultura y educación en la nación yurakaré. Cochabamba: FUNPROEIB Andes.

Sánchez Camacho, Arminda Justina. 2005. Teshentala: La educación yurucaré en la práctica cultural de la caza. Cochabamba: Universidad Mayor de San Simón MA Thesis.

Yurakaré

Sonja Gipper



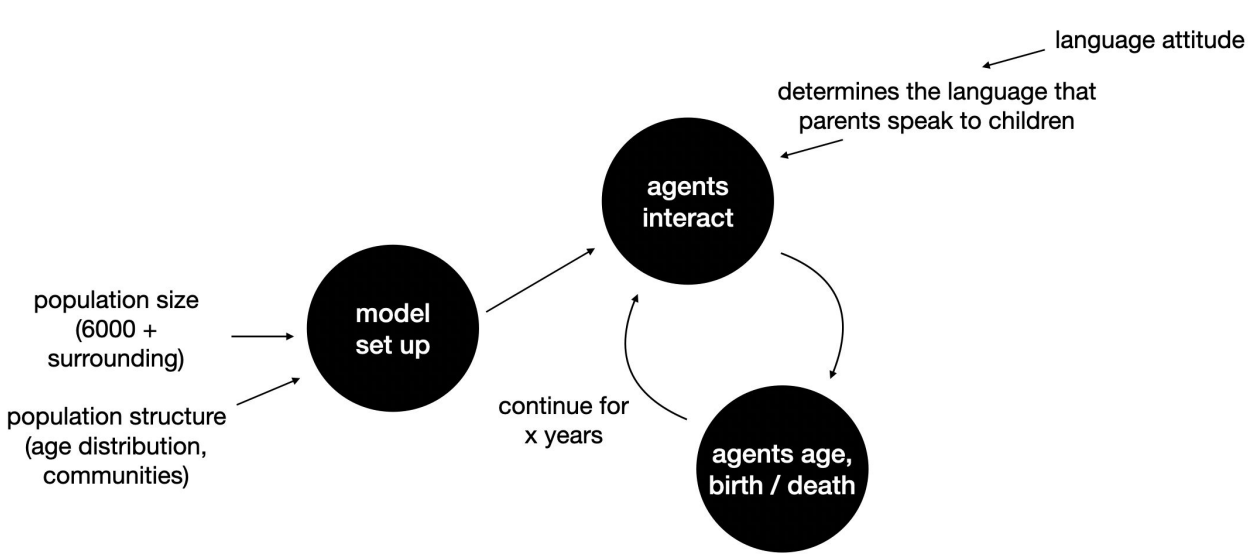


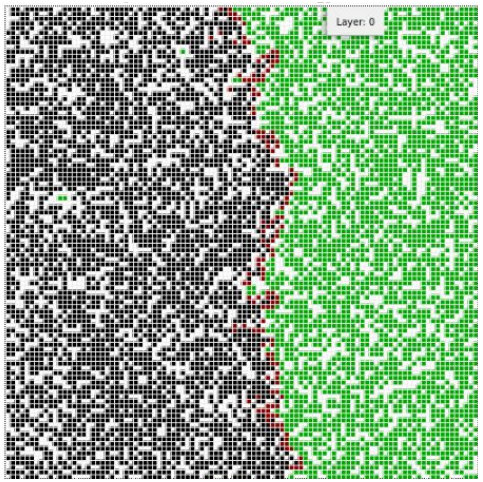
Van Gijn, Hirtzel & Gipper (2010)

Yurakaré model example

Sonja Gipper

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





non-Yurakaré (Spanish speaking)

Yurakaré

less remote

more remote

Break

https://www.pinterest.ca/pin/51439620728816348/

Universiteit
Leiden
Centre for Linguistics

Hadza

Universiteit Leiden
Centre for Linguistics
Richard Griscom

- **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

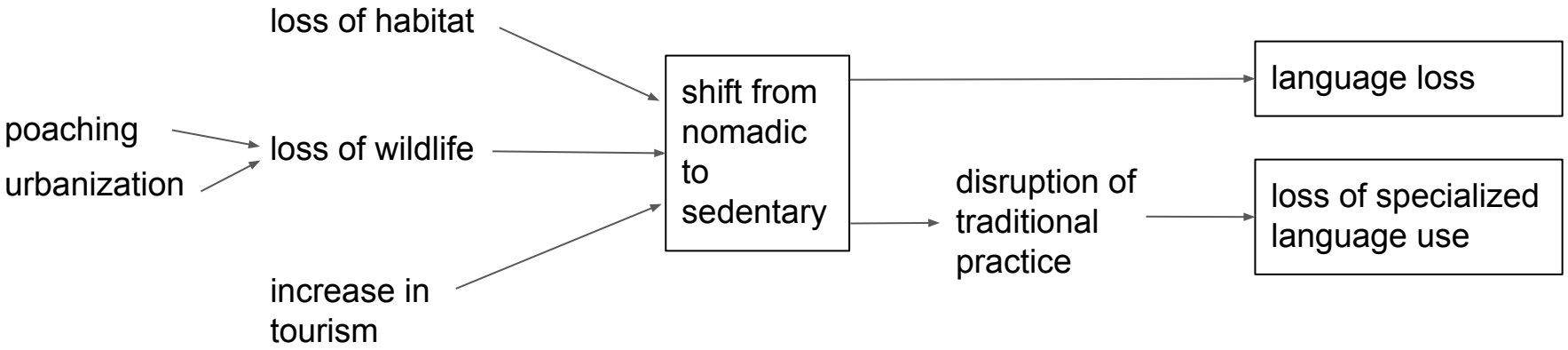
Universiteit Leiden
Centre for Linguistics
Richard Griscom

- 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

Universiteit Leiden
Centre for Linguistics
Richard Griscom

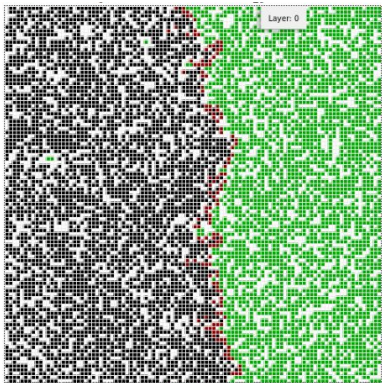


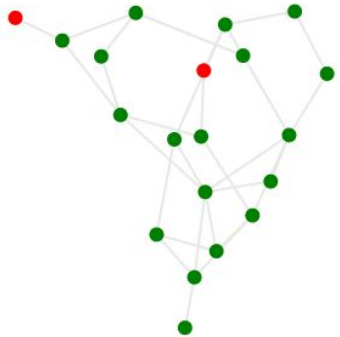
Hadza

Universiteit Leiden
Centre for Linguistics
Richard Griscom

- 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/






European
Free
Alliance

Friulian

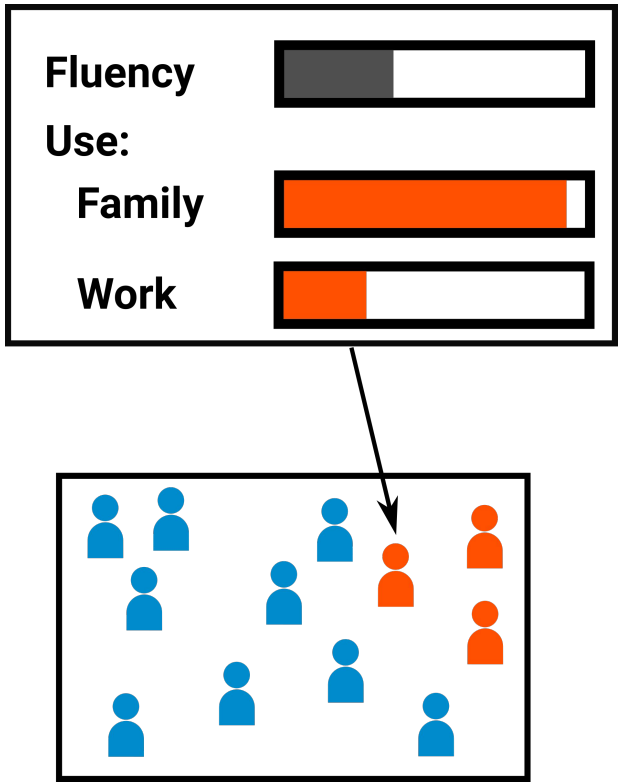
- **Area:** Friuli region, northeastern Italy
- **Language family:** Rhaeto-Romance, Romance, Indo-European
  - **Related:** Romansh (Switzerland) and Ladin (Italy)
- **Status:** recognised as minority language. Limited use in schools, by local authorities (place names) and media
- **Number of speakers:** 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

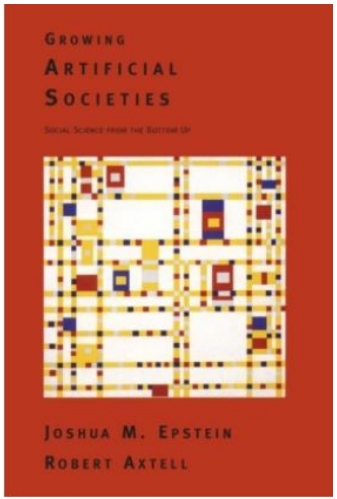
- **Threats**
  - Population loss in rural areas; work in Italian-speaking urban areas
  - Low social status → children not brought up in Friulian
  - Popularity Italian-language media
- **Revitalization efforts (20-30 years)**
  - Social status improved
    - Language tools: dictionary, spellchecker, 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 → more parent-child transmission?
- Bilingualism



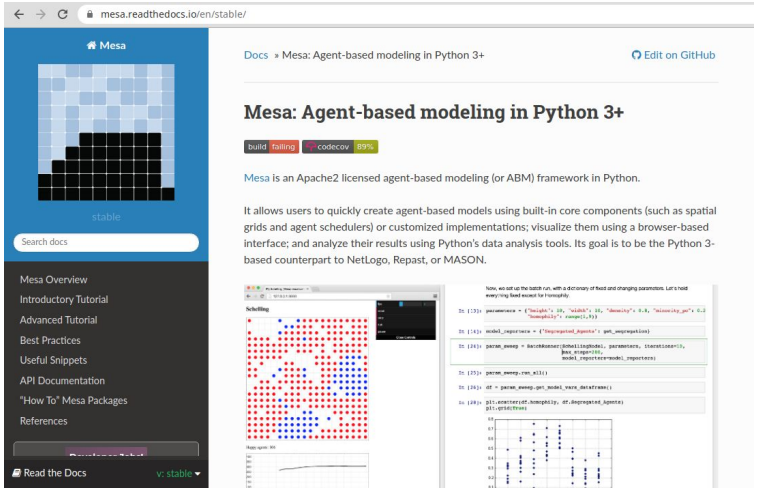
Resources



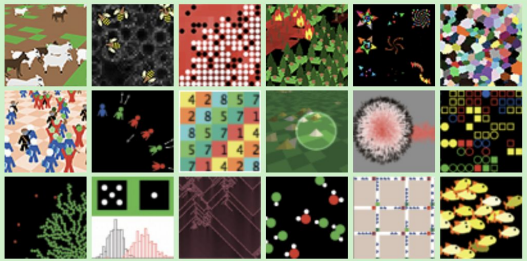
Epstein & Axtell (1996)

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.



Netlogo

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

Abrams, D. M., & Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature*, *424*(6951), 900–900. https://doi.org/10.1038/424900a

Civico, M. (2019). The Dynamics of Language Minorities: Evidence from an Agent-Based Model of Language Contact. *Journal of Artificial Societies and Social Simulation*, *22*(4), 27. https://doi.org/10.18564/jasss.4097

De Bie, P. & de Boer, B. (2007). An agent-based model of linguistic diversity. Language, Games and Evolution, 1.

Epstein, J. M. (2008, October 31). *Why Model?* [Text.Article]. JASSS. http://jasss.soc.surrey.ac.uk/11/4/12.html

Gijn, R. van, Hirtzel, V., & Gipper, S. (2010). Updating and loss of color terminology in Yurakaré: An interdisciplinary point of view. *Language & Communication*, *30*(4), 240–264. https://doi.org/10.1016/j.langcom.2010.02.002

Kandler, A., Unger, R., & Steele, J. (2010). Language shift, bilingualism and the future of Britain's Celtic languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1559), 3855–3864. https://doi.org/10.1098/rstb.2010.0051

Lutzenberger, H. (in press). Threat or natural fluctuation? Revisiting language vitality of Kata Kolok, the sign language of a village in Bali. *UNESCO World Report of Languages 2019*.

Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In *Computational Social Psychology* (pp. 311–331). https://doi.org/10.4324/9781315173726-14

Questions?