A statistical model for the study of dispersal, competition, and evolution of languages

Marco Patriarca
IFISC (CSIC-UIB)
marco.patriarca@ifisc.uib-csic.es
http://ifisc.uib-csic.es

KBFI (NICPB) - Tallinn, Estonia
http://www.kbfi.ee

Summary

- Statistical mechanics and languages
- Language and evolution theory
- Biogeography and languages
- Space-continuous models, an example
- Agent-based models with geography, an example
- Conclusion and future directions

IFISC

Darwin: Evolution of language, hypothesis of a single ancestor language

• The **analogy language-species** (and e.g. words-genes) cannot be close, since *each* language horizontally exchanges e.g. words with a language with which it is in contact. A mathematical model of language evolution and interaction has to take this into account. Therefore the meaning of an apparently simple and standard question such as:
"From which language does a language A *derive*?"
is questioned at a very basic level.

• There are various models of **language dynamics**, which have been inspired by statistical mechanics, ecology, population dynamics, etc., and many Corresponding papers, which have put the Darwin hypothesis into a quantitative framework.
Many other models and corresponding papers are coming.
For an overview see Søren Wichmann, *The Emerging Field of Language Dynamics,* Language and Linguistics Compass 2/3 (2008): 442–455

Alfred Russell Wallace

Wallace's research was very focused on the **biogeographical aspects**.

The importance of **(paleo)climate**, e.g. the (dis)appearance of water channels knowledge of the environment across space and time)

Distance is not a reliable indicator of biological affinity, e.g. due to the extinction in intermediate areas leaving discontinuous ranges and **patches**.

Speciation can occur (more easily) after **geographical isolation** and the consequent adaptation to the local environment (importance of the **biogeography of islands**).

IFISC

MODELS OF LANGUAGE COMPETITION
(Languages as fixed species)

- Model of Abrams and Strogatz, Nature
- AB and Voter models
- Generalization of AB/Voter models which include bilinguals

See eg

Mira & Paredes, Europhys. Lett. 69 (2005) 1031
Wang & Minett
Castello, Eguiluz, S. Miguel, New J. Phys. 8 (2006) 306
Castello, Toivonen, Eguiluz, Saramaki, Kaski, S.Miguel, EPL 79 (2007) 66006

IFISC

LANGUAGE SIZE DISTRIBUTION AND LANGUAGE EVOLUTION

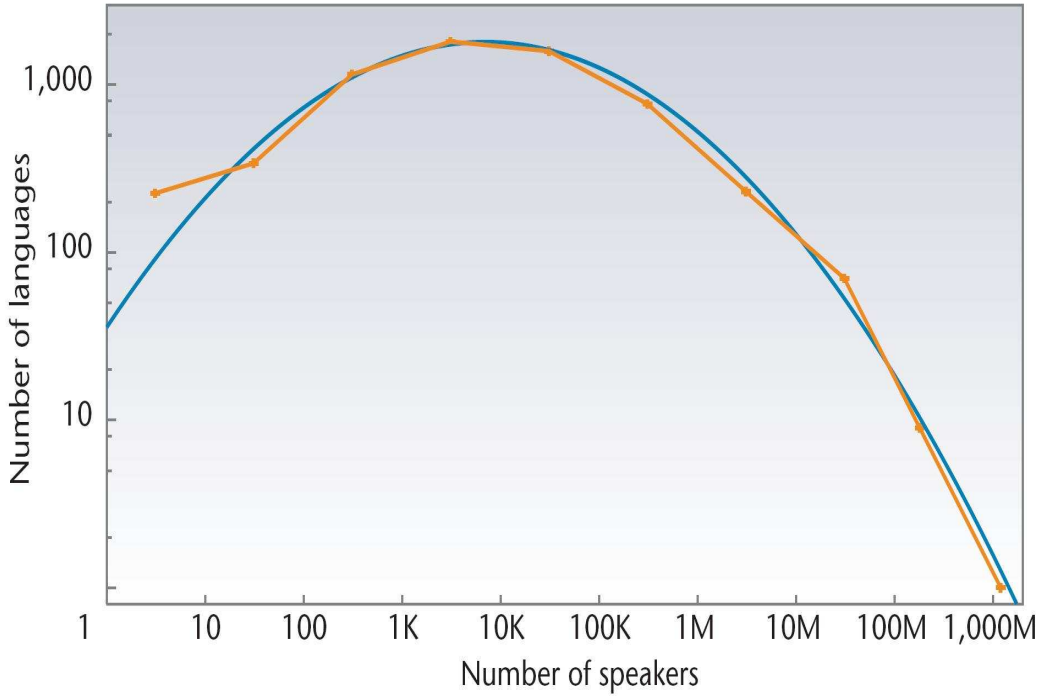
(Phase transition between a single language and fragmentation)

- Bit-string model of Stauffer and Schulze, see e.g. Phys. Life Rev. 2 (2005) 89

- Model of Nowak, see e.g. Proc. Natl. Acad. Sci. 96, 8028 (1999)

- W.J. Sutherland, *Parallel extinction risk and global distribution of languages and species*, Nature 423 (2003) 276

- T. Tesileanu and H. Meyer-Ortmanns, *Competition of Languages and their Hamming Distance,* Int. J. Mod. Phys. C 17 (2006) 256



EVOLUTION, NUMBER-AREA RELATION

- Holman, Schulze, Stauffer, Wichman, Linguistic Typology 11 (2007), 393

- Colonization models *a la* de Oliveira, Physica A 361 (2006) 361

IFISC

GEOGRAPHY & HISTORY

- Lotka-Volterra models in space, reaction-drift-diffusion equations (already used in anthropology: Fisher equations, population growth, etc.) → **Example**

[- Patriarca&Heinsalu, Influence of geography on language competition, Physica A]

- Agent-based models (e.g. Voter model, Naming game, etc.), with agents located on the sites of a regular lattice → **preliminary results** [Patriarca&Heinsalu]

- Agent-based models (same dynamics) with random walkers → **under development**

- Schulze&Stauffer, Competition of languages in the presence of a barrier, Physica A
- Patriarca&Leppänen, Physica A 338 (2004) 296
- Stauffer, Castello, Eguiluz, S.Miguel, *Microscopic Abrams–Strogatz model of language competition*, Physica A 374 (2007) 835
- C. Schulze, D. Stauffer, Language simulation after a conquest, arXiv:0707.0072
- Patriarca&Heinsalu, Influence of geography on language competition, Physica A

The Wallace line: a border in south-east Asia between two regions with very different biological species, due to paleoclimatic features of the region.

Example 1: A geographical model of culture transmission

Cultural interaction — Dispersal — Growth

$$\frac{\partial f_1}{\partial t} = R(f_1, f_2) - \nabla \cdot (\mathbf{F} f_1) + \nabla \cdot (D \nabla f_1) + \alpha f_1 \left(1 - \frac{f_1 + f_2}{K}\right),$$

$$\frac{\partial f_2}{\partial t} = -R(f_1, f_2) - \nabla \cdot (\mathbf{F} f_2) + \nabla \cdot (D \nabla f_2) + \alpha f_2 \left(1 - \frac{f_1 + f_2}{K}\right),$$

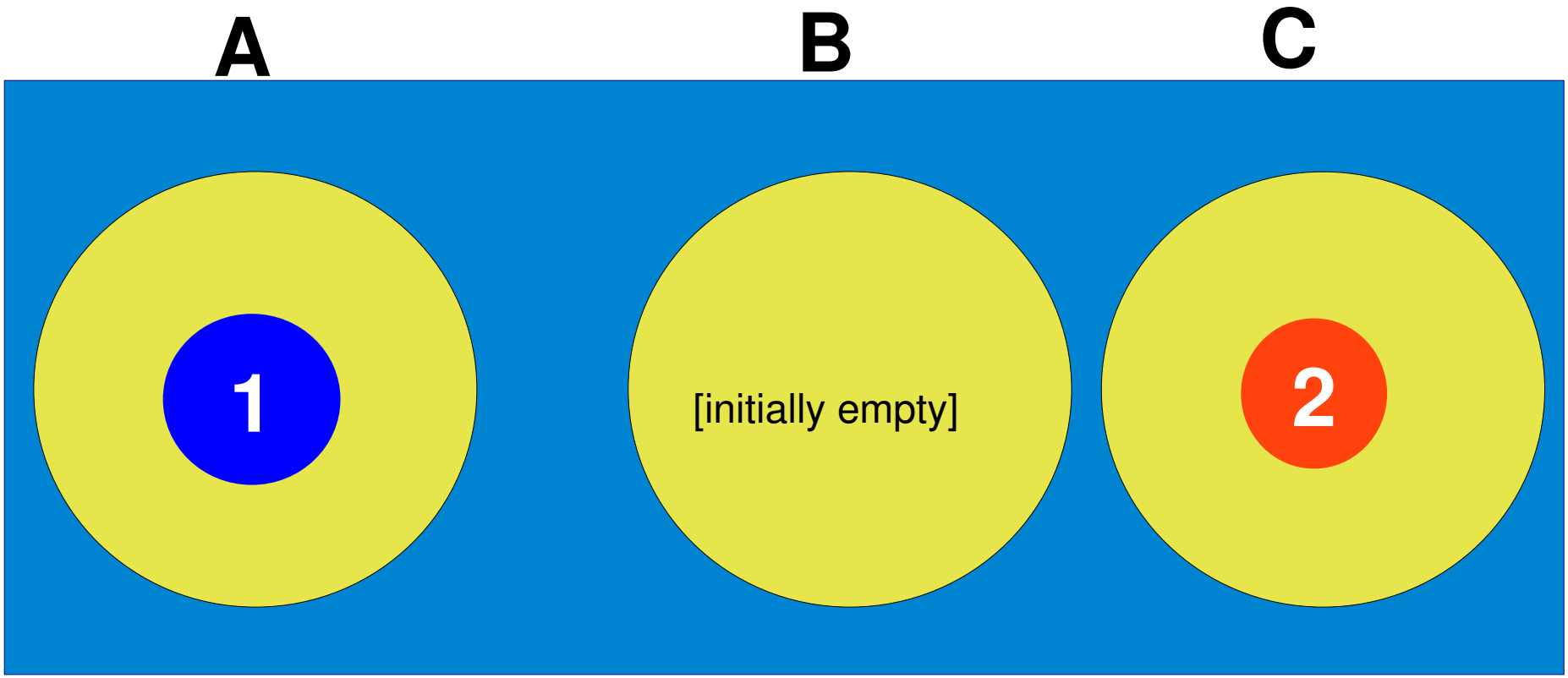
where the reaction term *a la* Abrams and Strogatz represents cultural transmission,

$$R(f_1, f_2) = k\left(s_1 f_1^a f_2 - s_2 f_2^a f_1\right)$$

Invasion of an island

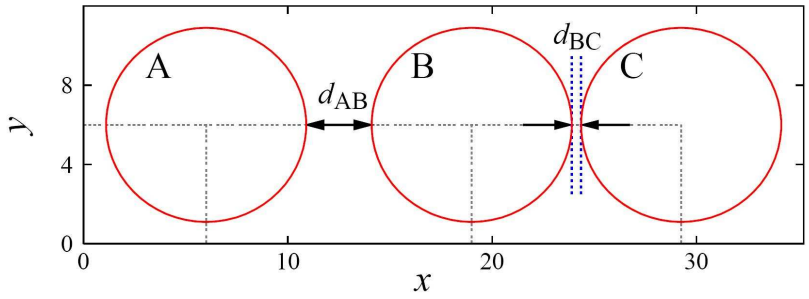
status $s_1 = 0.6$

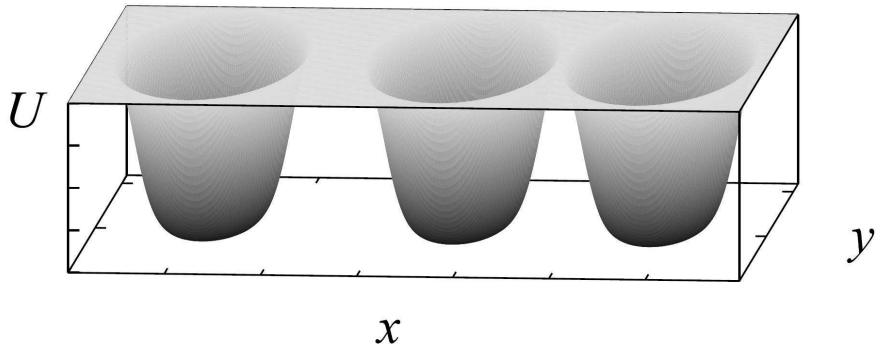
status $s_2 = 0.4$



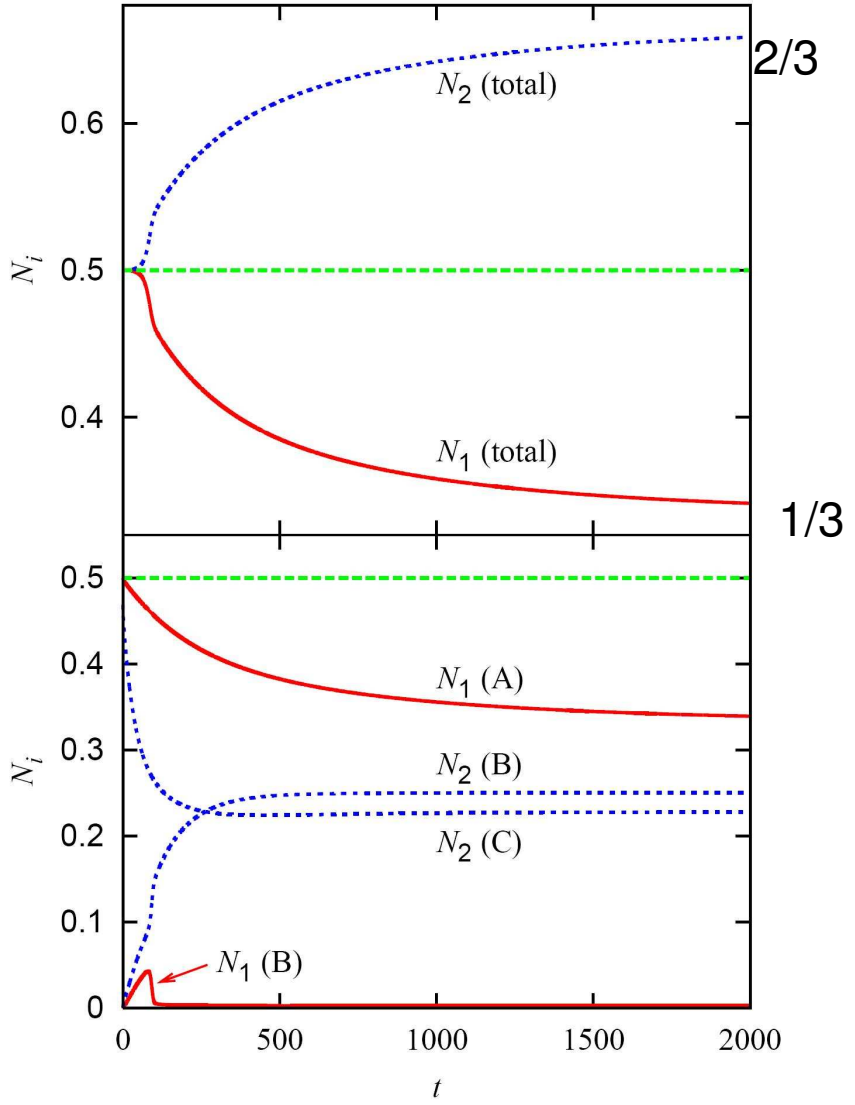
IFISC

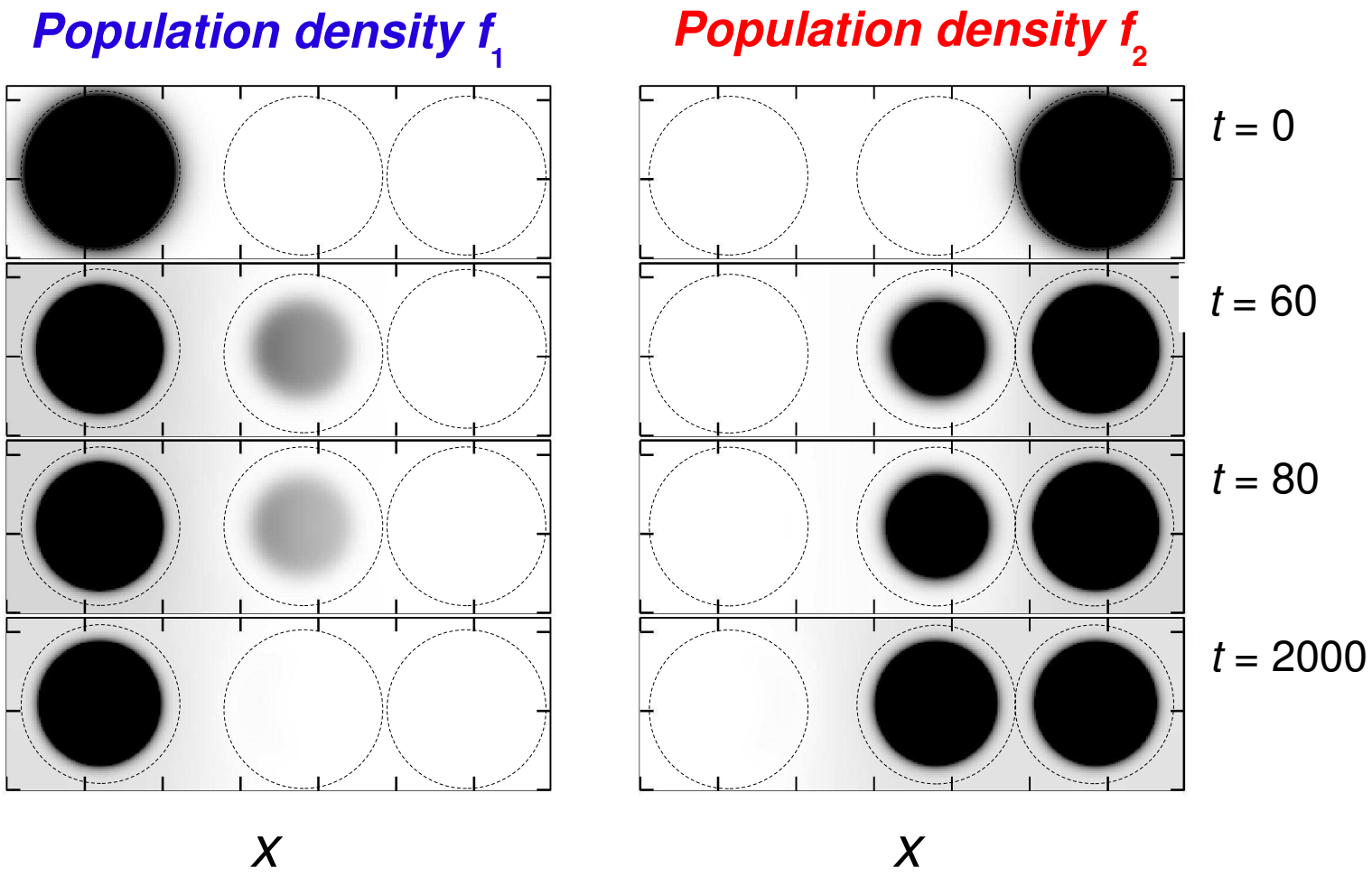
Mathematical model





Population size versus time




Population density f_1
Population density f_2
$t = 0$
$t = 60$
$t = 80$
$t = 2000$
x
x

IFISC

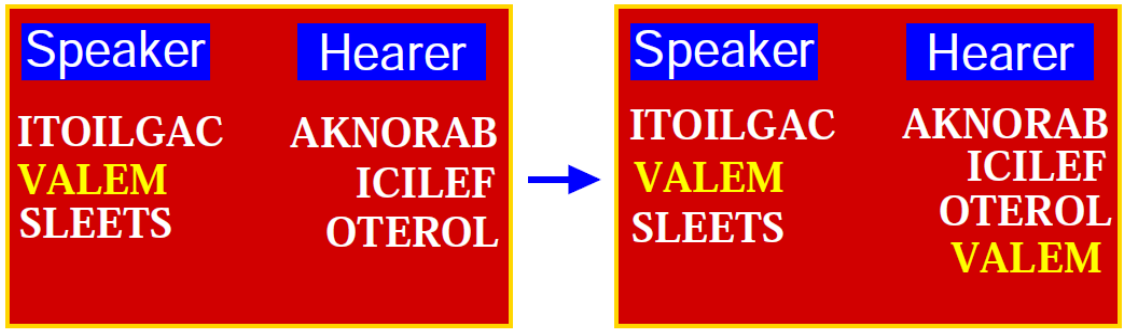
Example 2: A MORE LINGUISTICS-ORIENTED MODEL, THE NAMING GAME (Semiotic Dynamics)

- Model introduced by Luc Steels, Artif. Life 23, 319–332 (1995)
- Model going beyond the voter model or some generalizations of it, since it is inspired by the actual interaction mechanisms between speakers.
- How a language changes and a common consensus is reached is described in a simple way without the need of introducing many generations.
- The model describes how a community self-organizes to become such, i.e., *how convergence toward a common vocabulary takes place.*

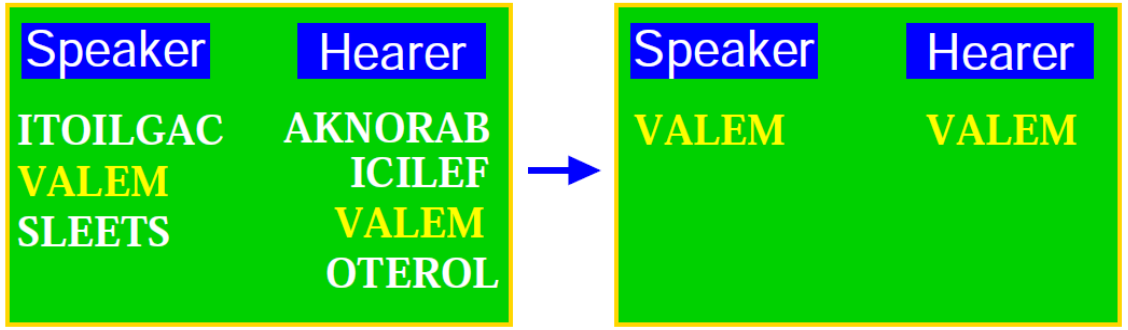
- A. Baronchelli, M. Felici, E. Caglioti, V. Loreto and L. Steels, *Sharp Transition towards Shared Vocabularies in Multi-Agent Systems*
- K. Kosmidis, A. Kalampokis, P. Argyrakis, Language evolution and population dynamics in a system of two interacting species, Physica A 353 (2005) 595.
- A. Baronchelli, L. Dall'Asta, A. Barrat, V. Loreto, Topology induced coarsening in language games
- A. Lipowski, D. Lipowska, arXiv:0801.1658.

IFISC

Failure



Success



From A. Baronchelli et al.

State of the system

- **N agents** (speakers, groups, et
- Each agent has an own **vocabulary** of words used to indicate a specific object (but it may represents words, grammar rules, etc.)

Initial conditions

- each agent gives initially a (random) word to the object

Dynamics

- each time step there is a random meeting/interactions between agents
- one agent is **S** = **speaker** (saying a word), the other is **H** = **hearer**

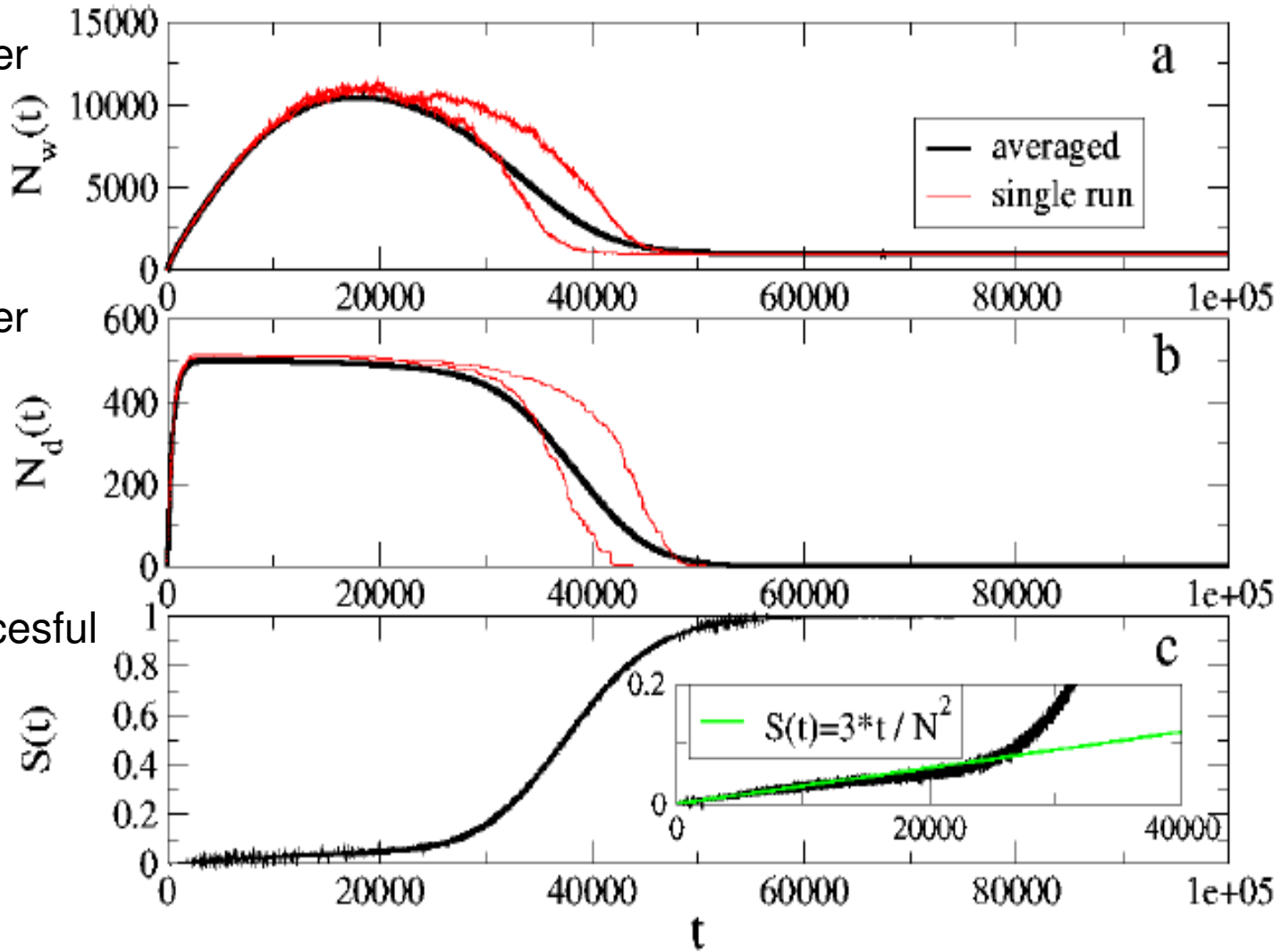
1) if **H** does not know/have the word, the word is added to the vocabulary
2) if both **H** and **S** know/have the word, all the other synonyms are deleted

IFISC

Total Number of words

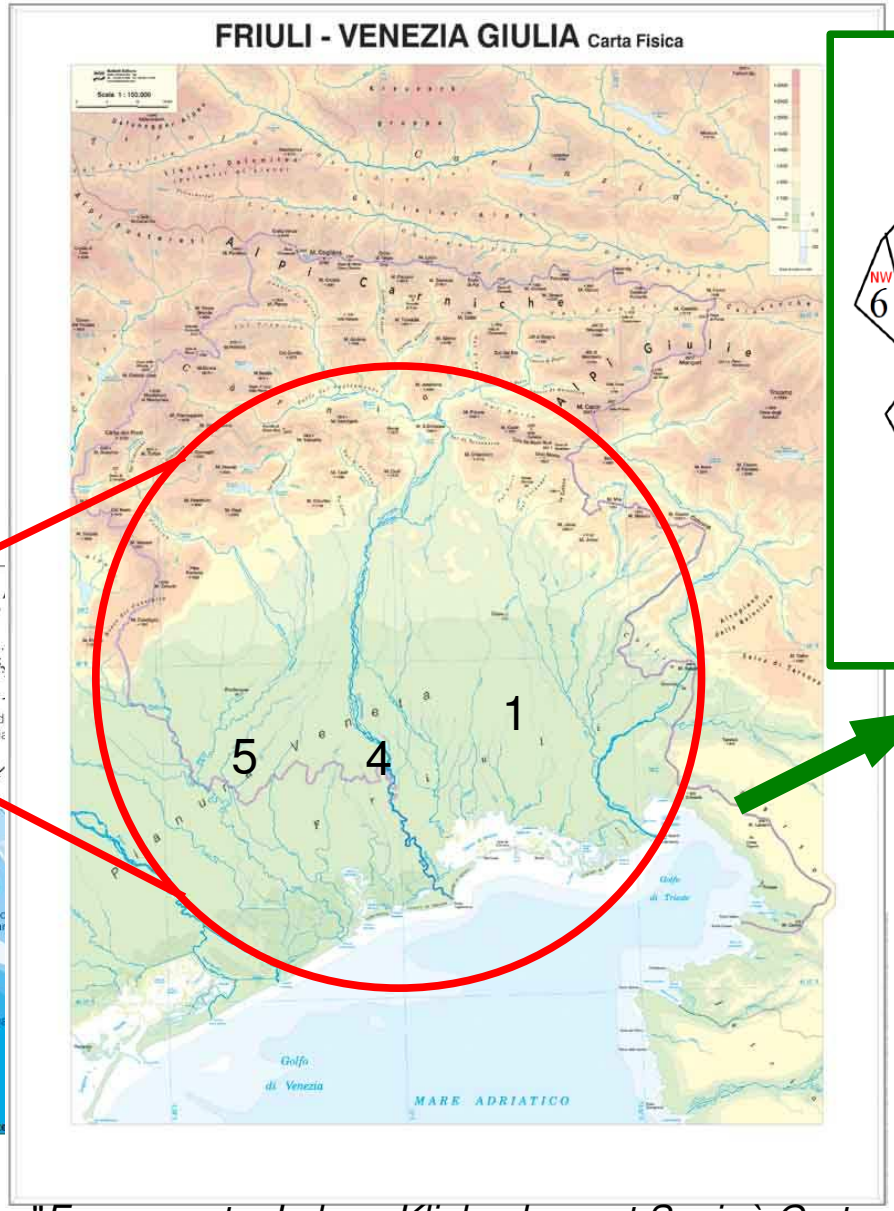
Total Number of different words

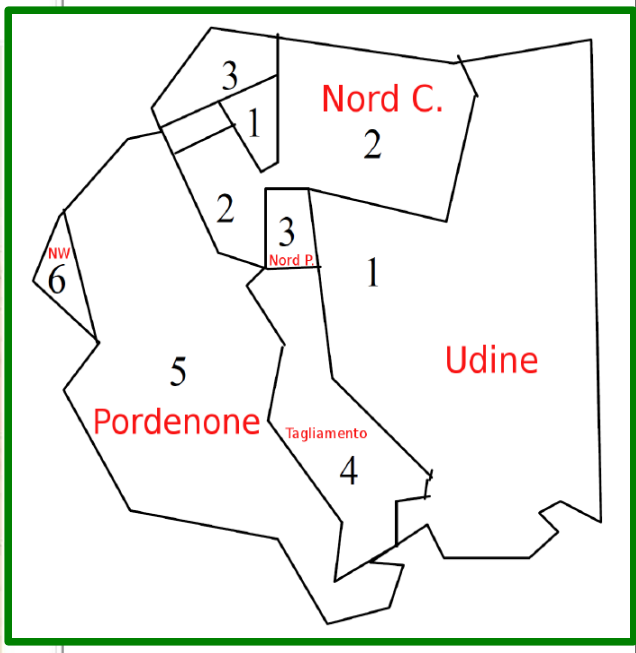
Rate of succesful events



IFISC

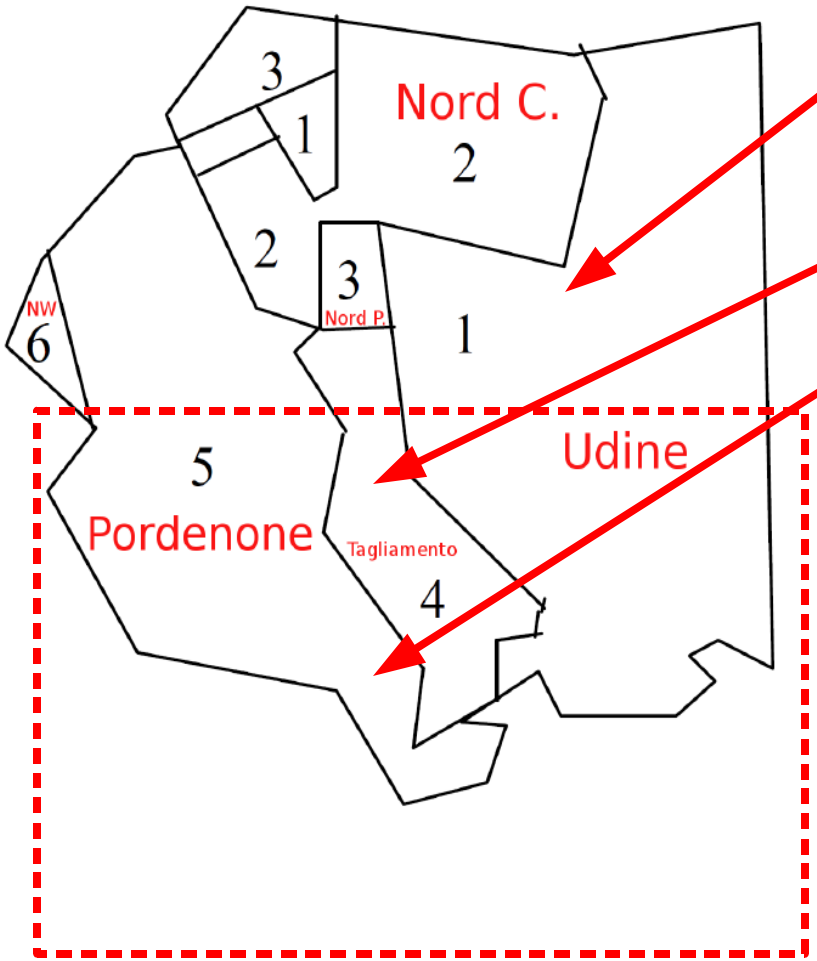
EXAMPLE:
Strong vowels In Friuli





Above: Simplified scheme of southern Friuli. Zones 5 (Pordenone), 4 (river Tagliamento), and 1 (Udine) are shown also on the physical map (left).

See: **Jean Léo Léonard**, "*Francescato, Labov, Klinkenberg et Sapir à Gorto : les voyelles fortes du frioulan dans tous leurs états structuraux, diatopiques, diffusionnistes et sériels*", in Carmen Alén Garabato, Teddy Arnavielle, Christian Camps (éds.) *La romanistique dans tous ses états*, Paris, L'Harmattan, pp. 183-207. Data from the work of Francescato cited therein.



Aires	*e	*o	*ɛ	*ɔ
1.	e:	o:	i:	u:
2.	ei	ou	i:	u:
3.	ia	ua	ei	ou
4.	e:	o:	ei	ou
5.	ei	ou	ei	ou
6.	ei	eu	ia	eu

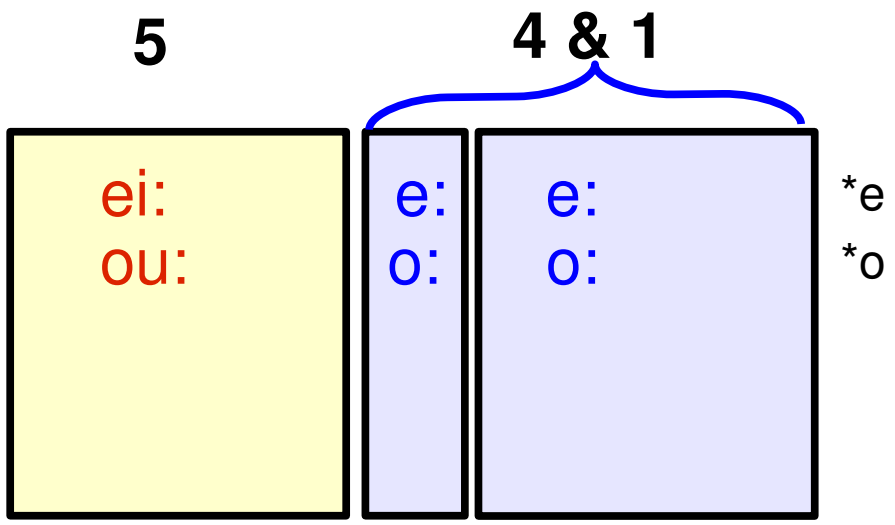
5	4	1	
ei:	e:	e:	*e
ou:	o:	o:	*o
ei	ei	i:	*ɛ
ou	ou	u:	*ɔ

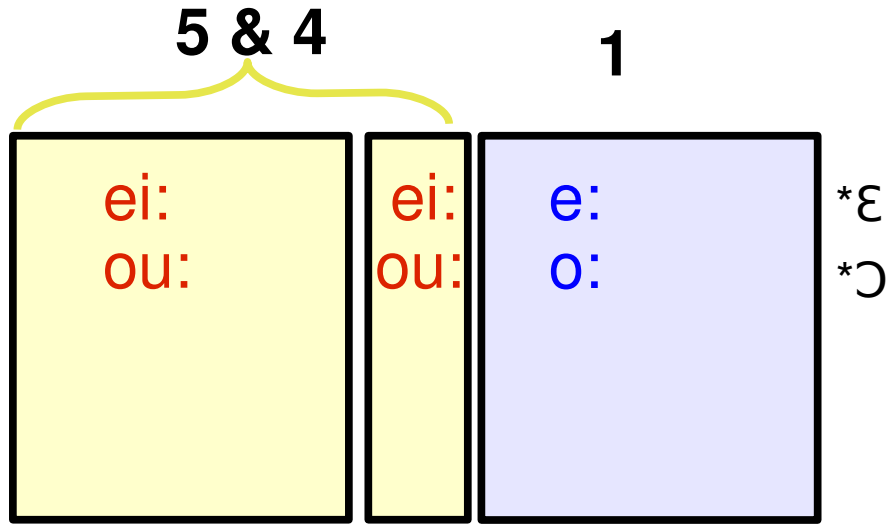
Question >>

IFISC

Why is each vowel observed only on one side of the river and possibly on the opposite side along a narrow zone close to the river?

That is, why only in the region (5)+(4) or in the region (1)+(4) ?





Agent model >>

Examples of use of strong vowels

Mots-témoins	Frioulan oriental		*Frioulan occidental*	
	Graphie	API	Graphie	API
PEDE	*pît*	[pi:t]	*péit*	[peit]
DECE	*dîs*	[di:s]	*déis*	[deis]
*GREVE	*grîf*	[gri:f]	*greif*	[greif]
NOVU	*nûf*	[nu:f]	*nouf*	[nouf]
FOCU	*fûk*	[fu:k]	*fouk*	[fouk]
LOCU	*lûk*	[lu:k]	*louk*	[louk]

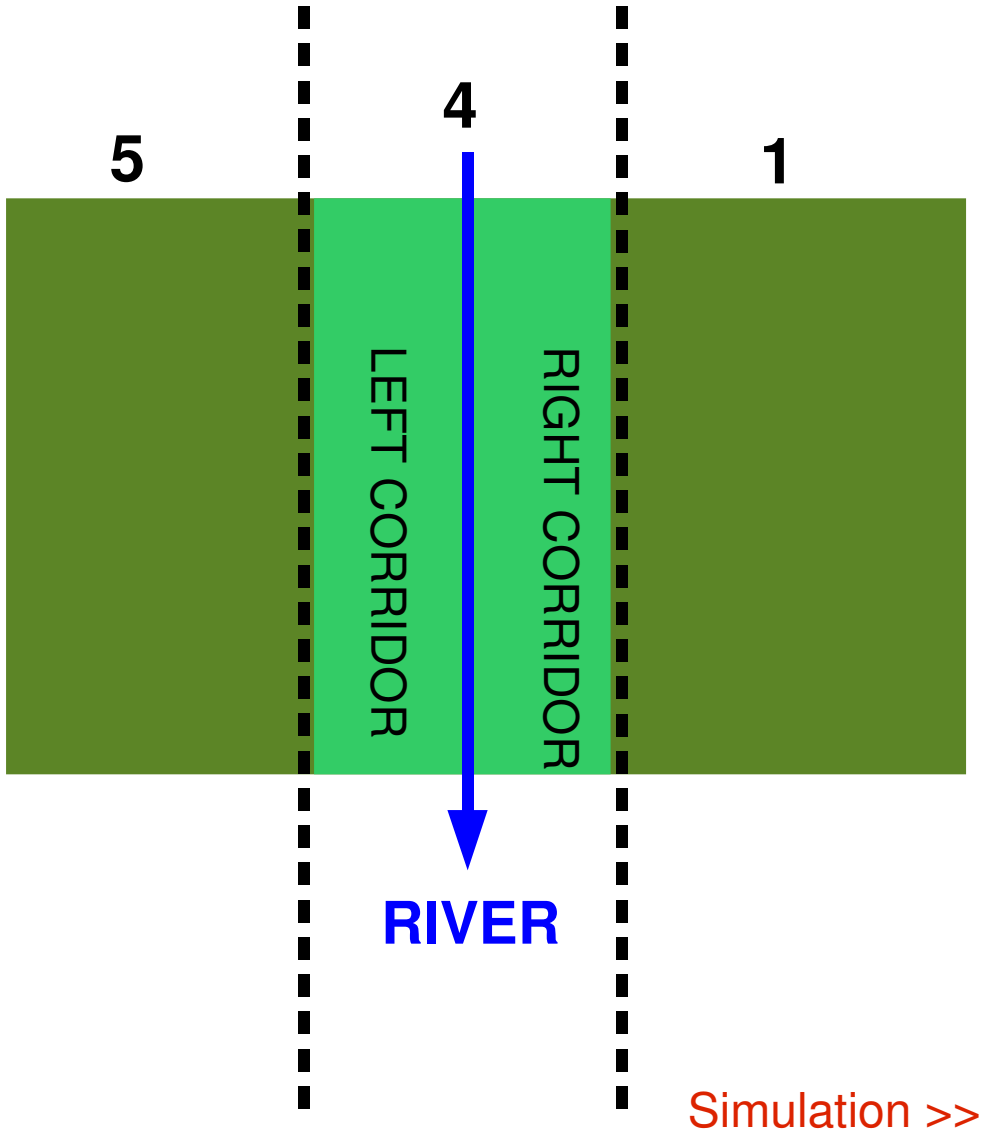
IFISC

Example of spatial naming game, in which agents interact randomly with first neighbor only (square lattice assumed) on the following simplified landscape, consisting of:

Corridor (4 left and right): a region with higher flows (higher rates of interaction between agents)

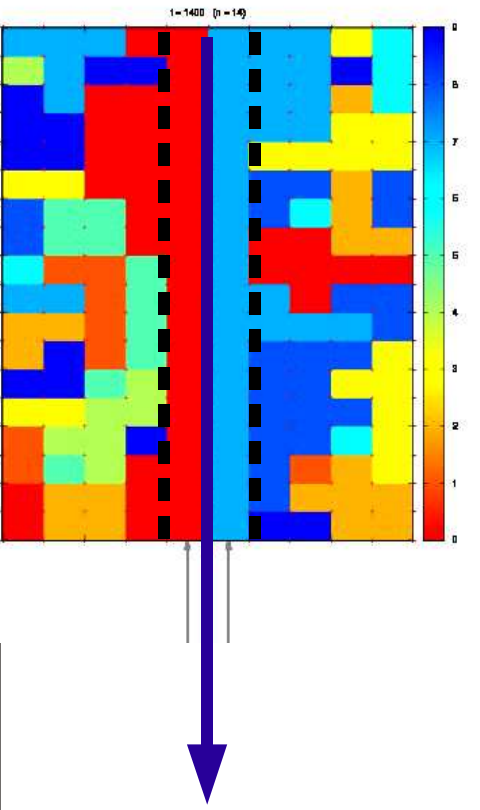
River (4, blue line): it represents a barrier, through which interaction is slower (smaller interaction probability)

Rest of the area (1 and 5): agents interact locally with the neighbors at the reference rate

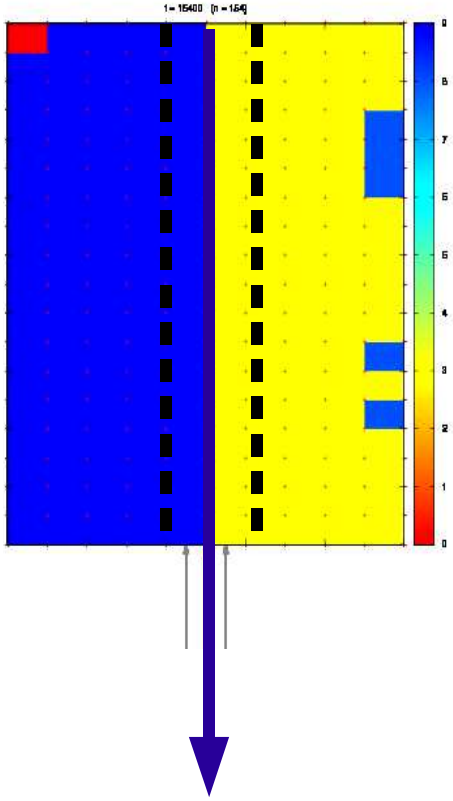


Simulation >>

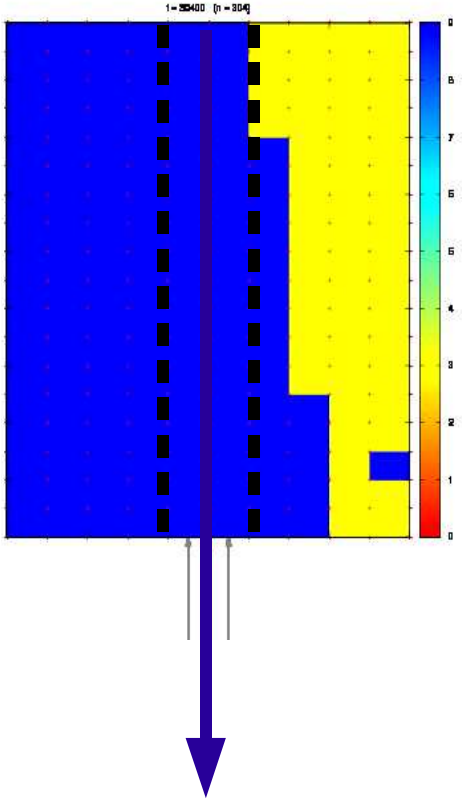
Close to initial conditions



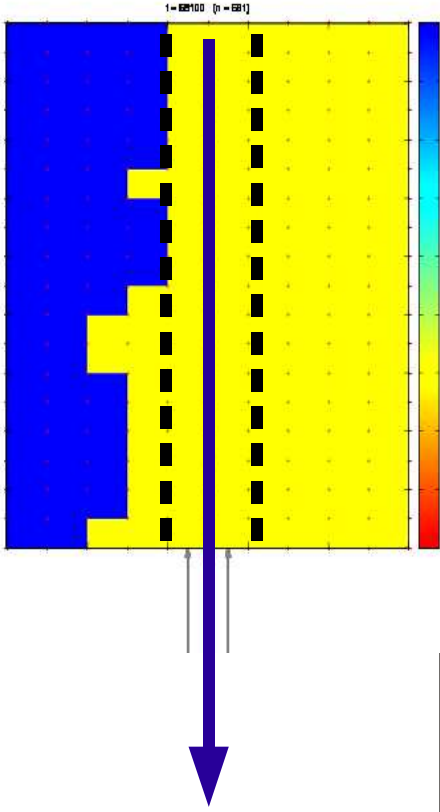
50%/50% sharing



Blue on the right too



Yellow on the left too



- Size 10x20
- Probability to communicate (vertically) along the corridor is 1
- Probability to communicate across the barrier is 1000 times smaller
- Probability to communicate outside the corridor is 1000 times smaller

Only two “words” (strong wovels) survive on the opposite sides of the river, occupying from time to time also the other shore. Shore are characterized by a higher interaction probability respect to the rest, thus representing a corridor for cultural flow. video >>

Final remarks

- Darwin hypothesis and space-time evolution of languages can be out in a quantitative framework

- Many elements of biogeography can/have to be directly exported to describe language

- Suitable linguistics dynamics has to be used

- Many data waiting to be described/explained

http://ifisc.uib-csic.es
marco.patriarca@ifisc-csic.es