A unified computer model for internal and external constraints in language evolution

5th International Conference on the Evolution of Language
Leipzig, 1-4-2004

Christophe Coupé
Laboratoire Dynamique du Langage, CNRS - Université Lyon 2, France

Overview

Main questions:

What is language change?

How do internal and social constraints interact in language change?

How can we investigate the rate of language change or the linguistic diversity in a community of speakers?

What the model does, and what it doesn’t

- The model investigates the *evolution* of language with general mechanisms (back to basics)…
- It pays equal attention to internal and social factors, and allows various degrees of complexity for both …
- But is not a model of *emergence* (although the framework may capture several existing models of emergence)

Outline

- Underlying theoretical framework
- Description of the model
- First experiments
- Perspectives and work in progress

Outline

- Underlying theoretical framework

Language as a complex dynamical system

- The underlying paradigm:
  - « *A complex dynamical system… consists of a set of interacting elements where the behavior of the total is an indirect, non-hierarchical consequence of the behavior of the different parts* » [Steels, 1997]
- Applied to various levels:
  - Words & lexicon [Steels, 1996],
  - Phonemes & phonological system [Lindblom, 1984]
  - Idiolects & language of a community
- Language is made of structures:
  - A structure is a collection of interactions between elements building a set as a coherent whole and giving it its specific aspect
  - Phonological system (significant oppositions), grammar, lexicon

Space of possible states

- The abstract space defined by all possible states of language
  - a muti-dimensional space, with continuous (eg. formantic values) or discrete dimensions (eg. word-order)

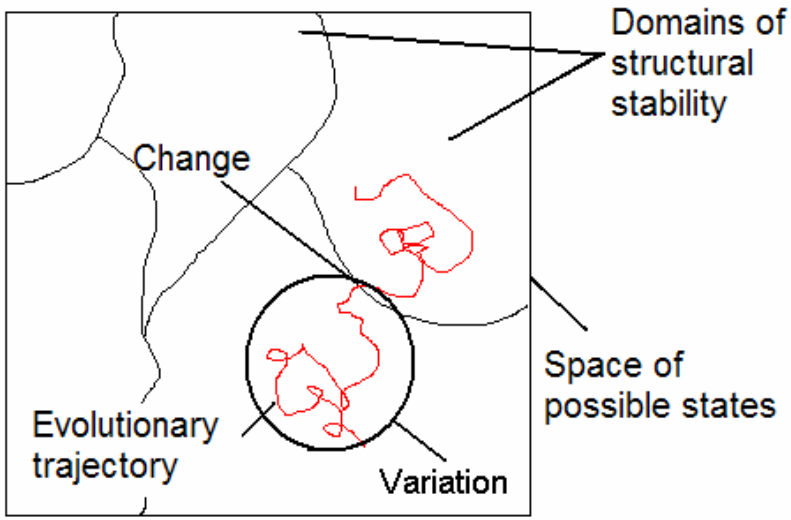
- Divided in domains of « structural stability »:
  - domains defined by a linguistic structure
  - all states in a domain correspond to the same structure
  - eg. allophonic variations for a phoneme

Language change

- Language evolves in the space of possible states according to various constraints
- Internal versus external forces (eg. [Labov, 2001])

- Variation: evolution of the system inside a domain of structural stability

- Language change
  - A *structural* change
  - From one domain to another

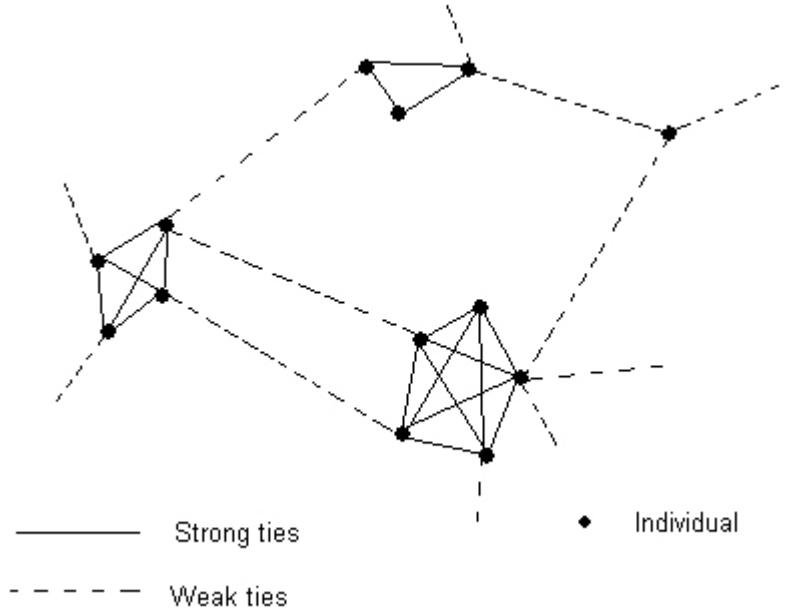


Internal constraints

- Cognitive constraints
  - Minimize the cognitive load during encoding & decoding (eg. working memory, access to data)
- Constraints in perception & production
  - Ease of the production vs. ease of the perception (eg. hypo-/hyper-correction [Ohala, 1993])
- Efficient transmission of information

External constraints (1)

- Sociolinguistics:
  - Language as a tool to assert one's identity [Labov, 1972]
- Individuals interact in complex social networks
  - role of social ties in actuation and implementation of changes
  - Contacts between languages
  - Contacts between dialects or registers



from [Milroy, 1992]

→ **Impact language change & evolution in the communities**

External constraints (2)

- From social ties to linguistic evolution
  - « *from Bloomfield's principle of density, which argues that each communicative act is accompanied by a slight degree of convergence of linguistic systems of speaker and interlocutor*. » [Labov, 2002]

  - « *Anspach in The Why of Fashion (1967) argues that the initiating spark is the need of people to be like others and yet to be distinct from others* » [Labov, 2001]

  → Positive links: convergence of linguistic systems

  → Negative links: (relative) divergence of linguistic systems

From a theoretical framework to a computational model

Task:

Derive a computational model from the preceding abstract descriptions

Why?

1. **Possibility to investigate the dynamics of language change**
2. **Possibility to measure various features of language change given the social structure or the internal constraints**
3. **Few models integrate both internal & external constraints**

How?

...

Outline

- Underlying theoretical framework
- Description of the model

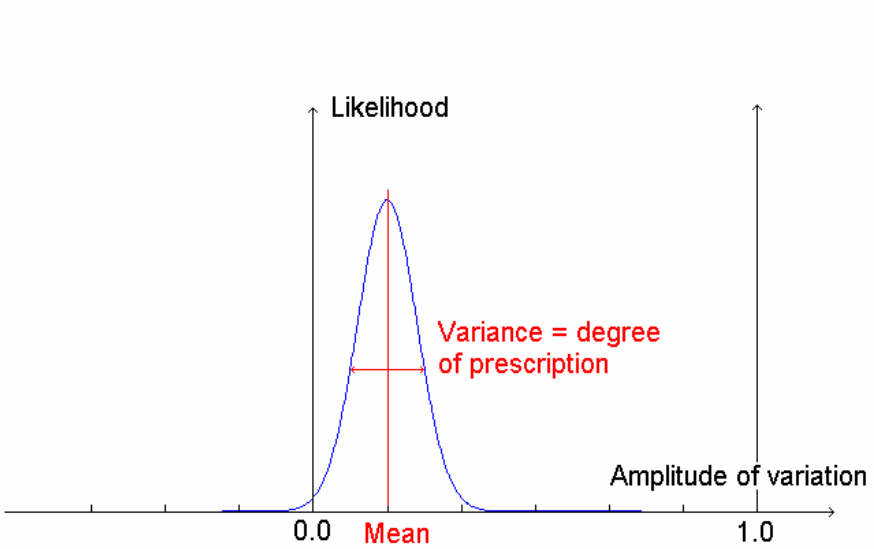
Parameters of the model

- N agents; each agent = one linguistic system
  - (x1, x2, …, xm): linguistic variables of the system
- A social network to tie the agents
  - Links between -1.0 & 1.0 (0.0: no interaction)
- A function describing the natural constraints on each agent (for all variables)
- Assumption: same internal constraints for all agents

A model based on probabilistic distributions

- Constraints modelled by Gaussian probabilistic distributions:
  - A non-deterministic behavior…
  - which follows constraints on average
  - A distribution for each linguistic variable

- Description of the Gaussian:
  - Mean = most likely variation
  - Variance
    - V = 0 → deterministic behavior
    - V = +∞ → fully random behavior

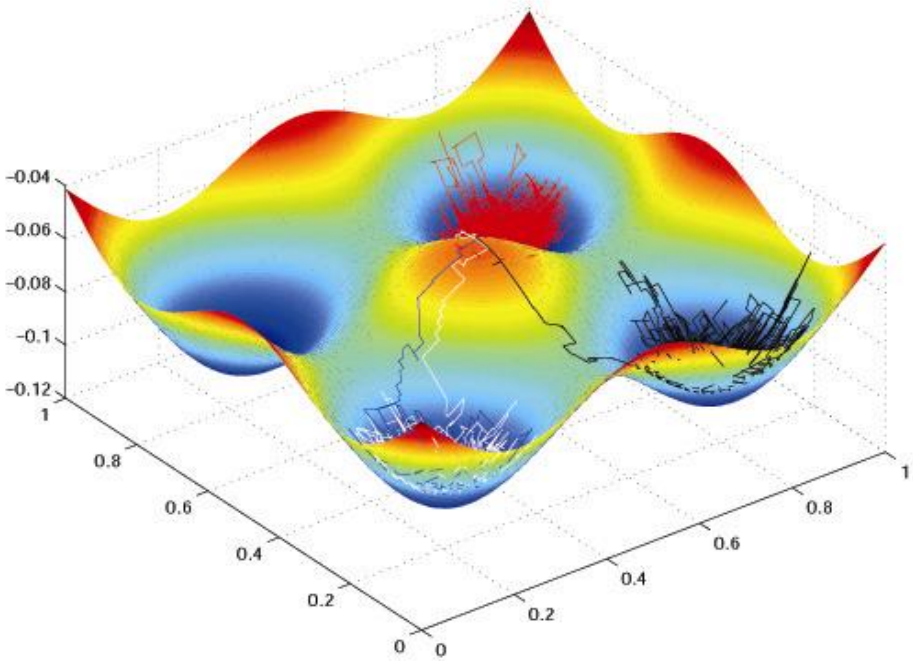


Modelling internal constraints

- Notion of fitness landscape:
  - from domains of structural stability to basins of attractions
  - variation = evolution inside a basin of attraction
  - change = jump from one attractor to another

- Most likely variation
  - $\propto$ slope of the landscape for each variable
  - local topography
    ←→ local view of the landscape
    ←→ unplanned change

    No need to know the whole landscape to compute local evolutions



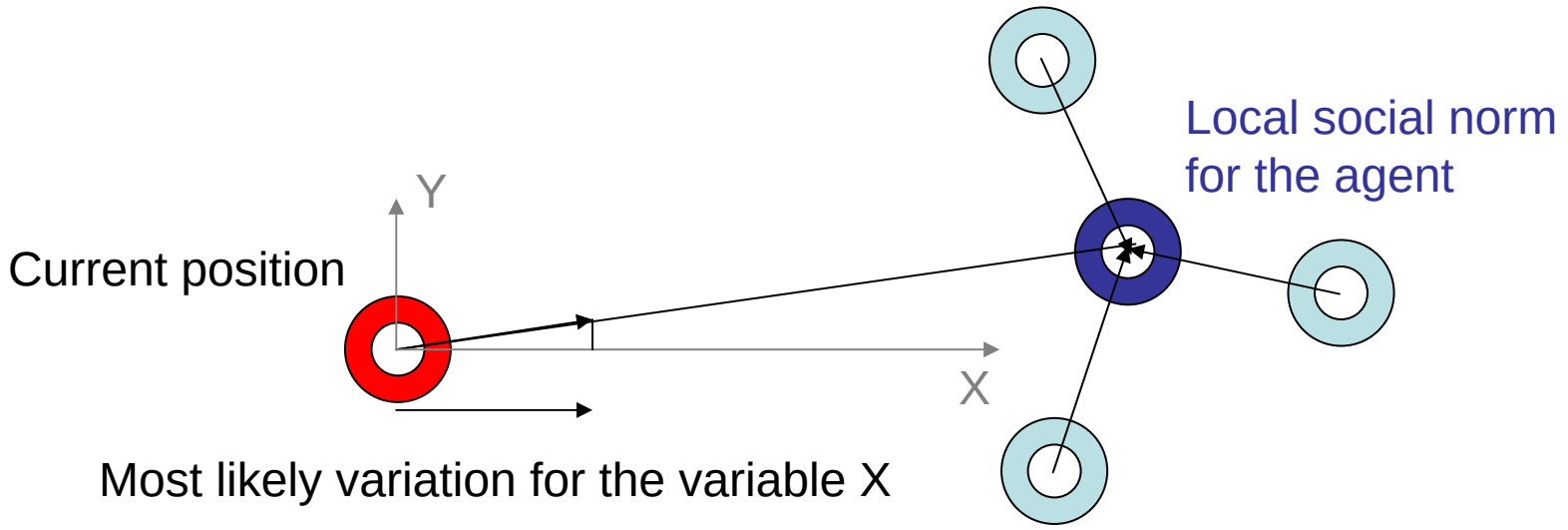
A fitness landscape for two linguistic variables ([0,1]x[0,1])

Modelling external constraints

- For each agent:
  - Compute the sum of attractions and repulsions with other agents

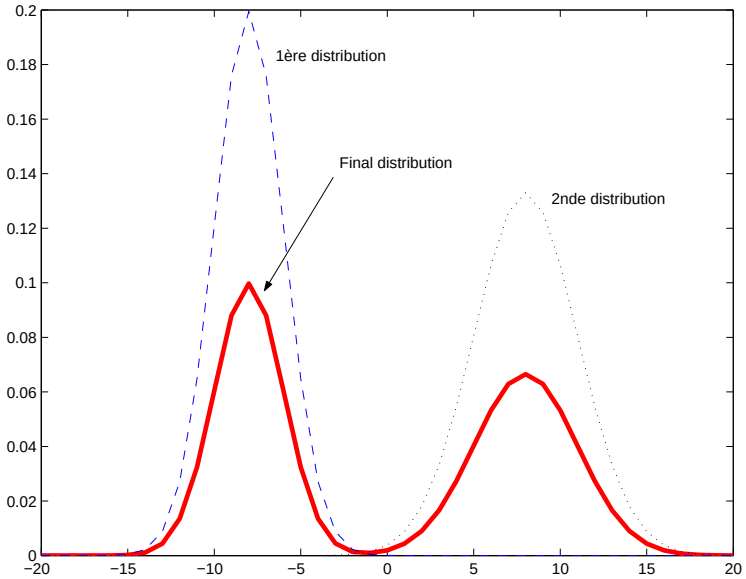
  → local social norm

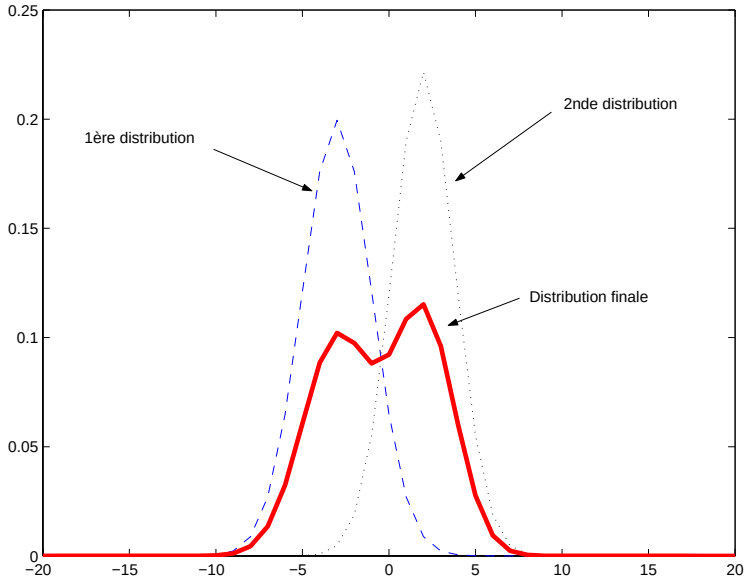
  → most likely variation $\propto$ (local social norm – current position)



Integrating constraints

Mixture of Gaussians





- Constraints can either go in the same or opposite direction(s)
  - social constraints can run counter to internal constraints
  - link with linguistic diversity

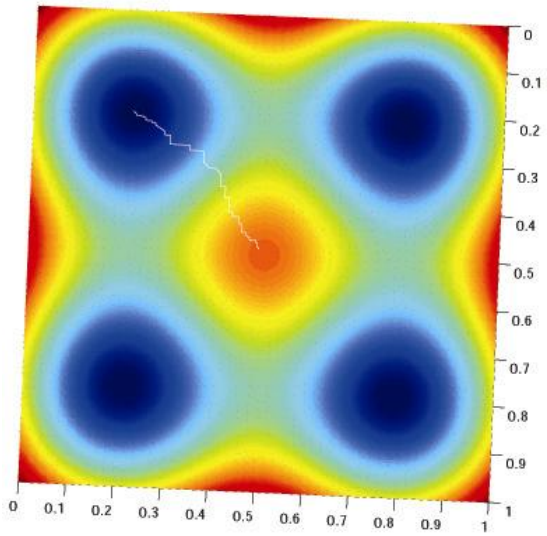
Variables to be measured

- Rate of change:
  - Number of transitions between attractors per time period (flat landscape: length of the trajectory)
  - Evolution of the idiolects / evolution of the barycentre of the idiolects
- Diversity for N agents:
  - Dispersion in various basins of attraction (flat landscape: spatial entropy or, more easily, average distance to the barycentre of the agents)

Outline

- Underlying theoretical framework
- Description of the model
- First experiments

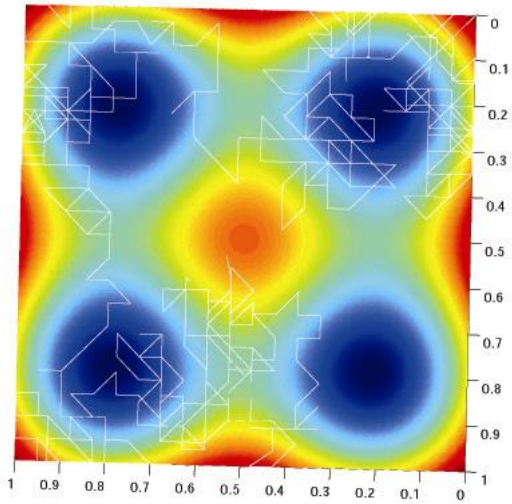
Testing a simple fitness landscape

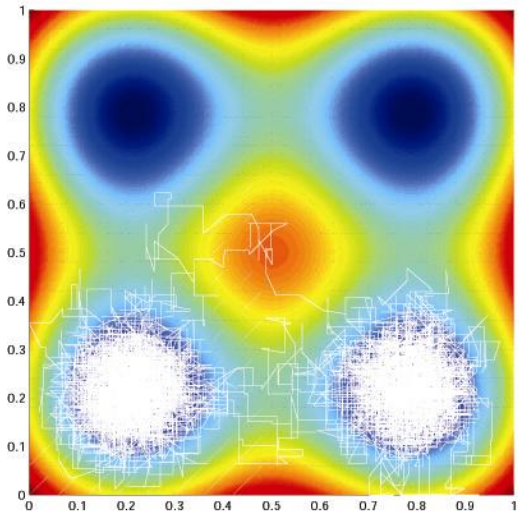


V = 0 → stable state

One agent

V = +∞ → random walk



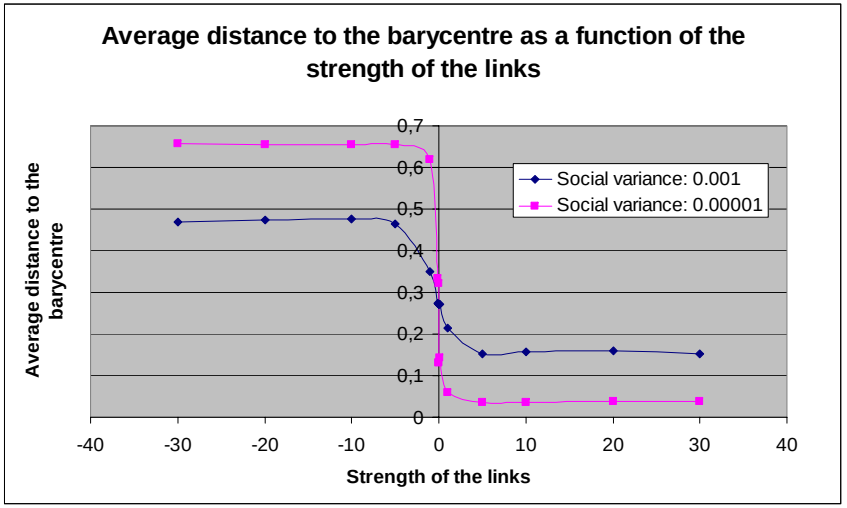


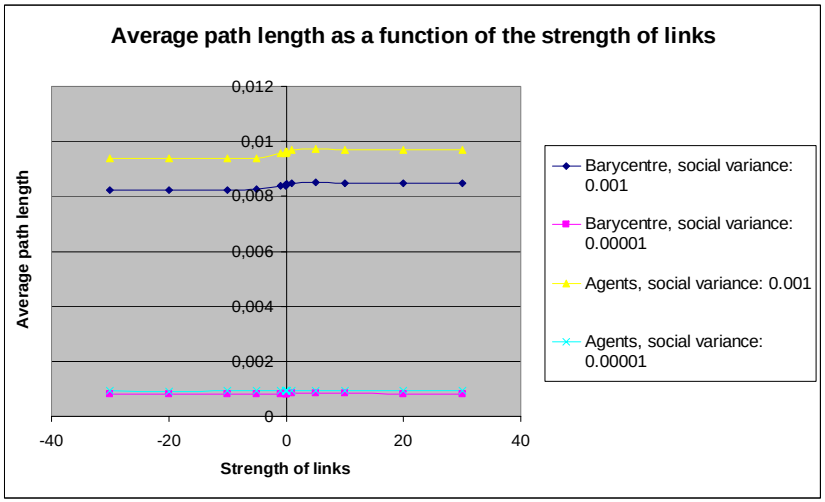
0 < V < +∞ → periods of stability in basins of attraction + changes

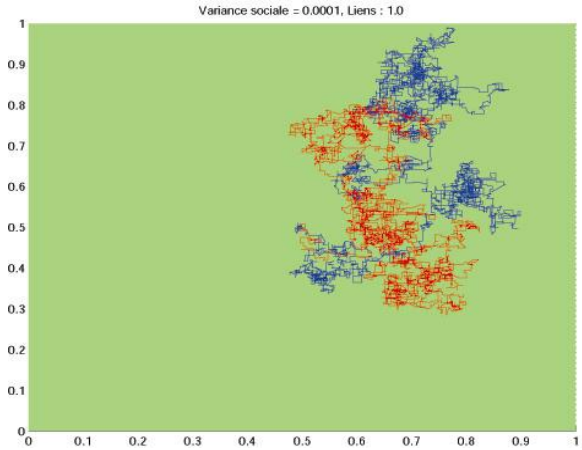
→ *Punctuated equilibriums*

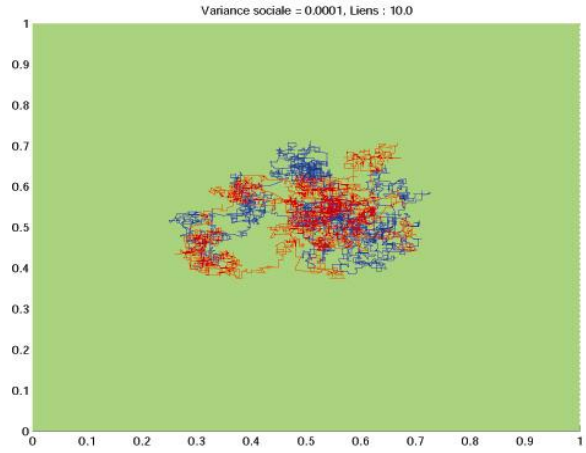
Simple social networks (1)

- Influence of the strength of the link



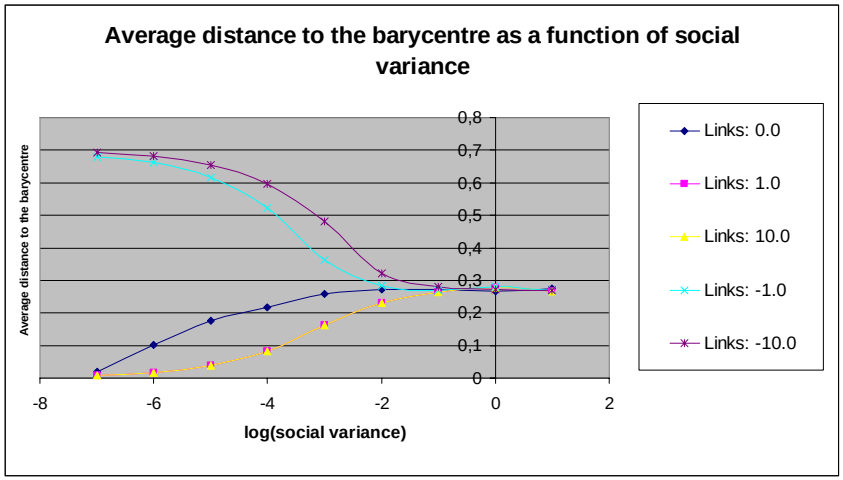


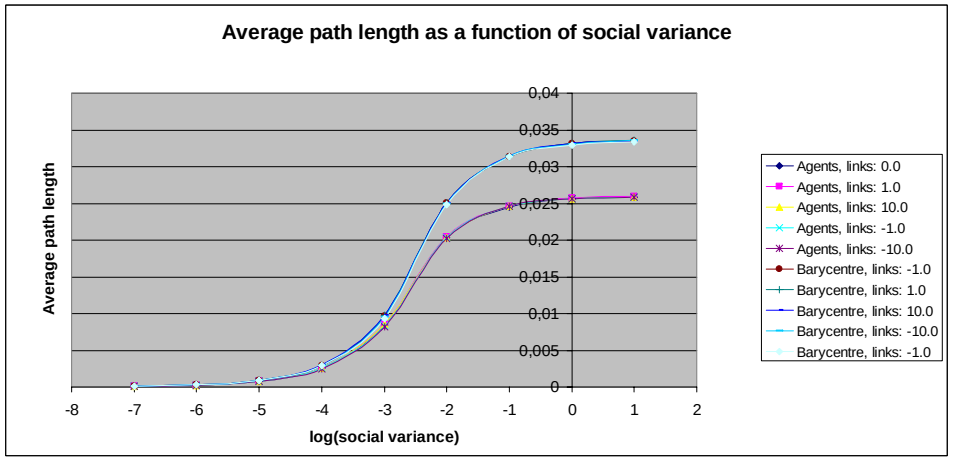




Simple social networks (2)

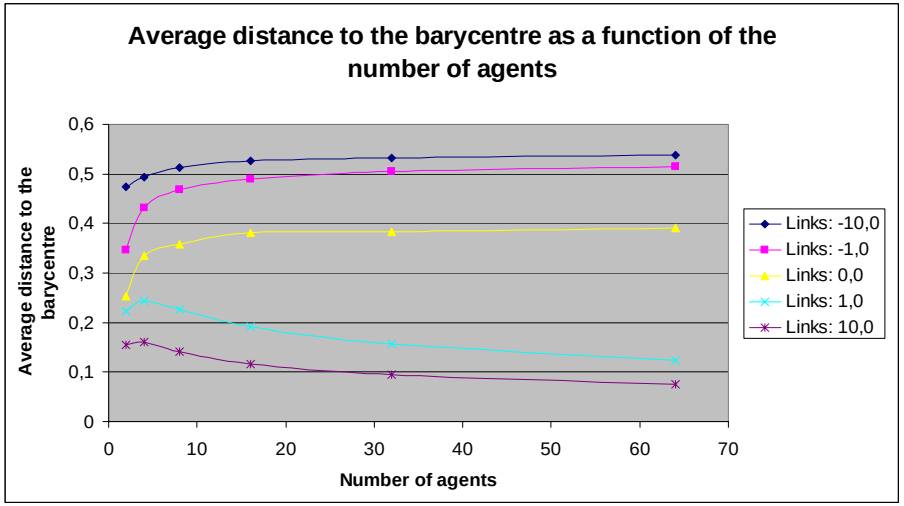
- Weak or strong variance of the Gaussian distributions



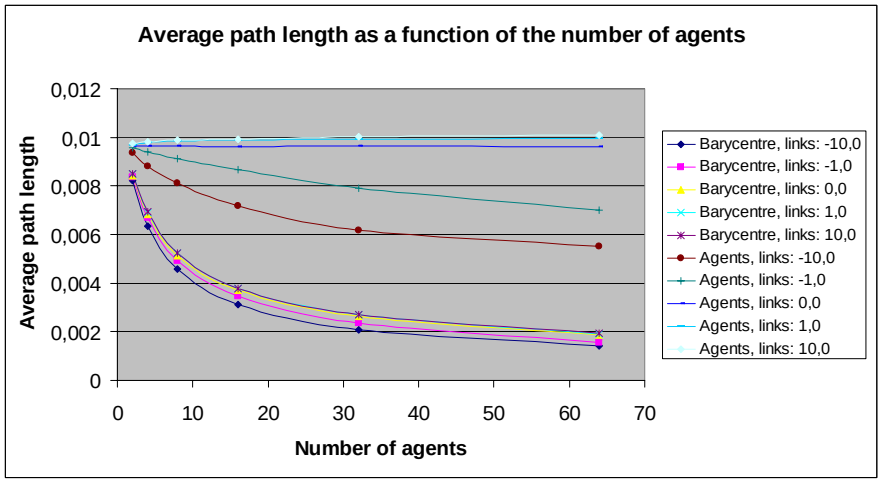


Simple social networks (3)

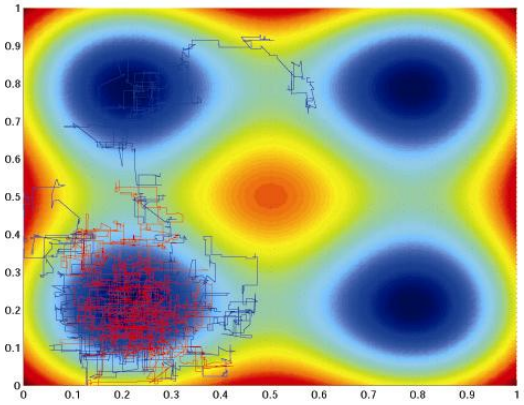
- Influence of the number of agents



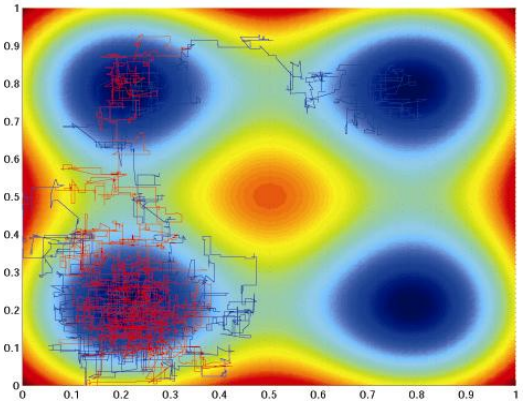
Similarity with Nettle's results (Nettle, 1999)



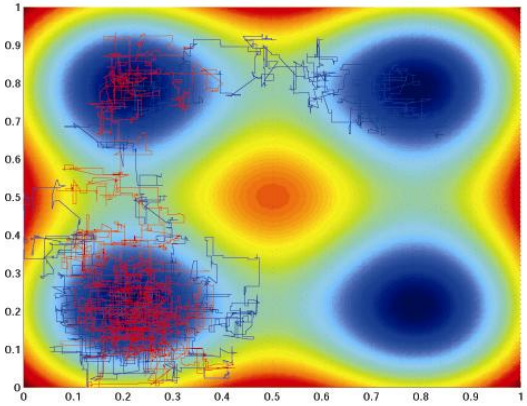
Integrating external & internal constraints



Time = 5,000



Time = 7,000



Time = 8,000

Same results as with separated constraints:

- punctuated equilibriums
- influence of the social structure

Outline

- Underlying theoretical framework
- Description of the model
- First experiments
- Perspectives and work in progress

Toward realistic networks

- Increasing number of studies on specific networks
  - smallworld networks
  - scale-free networks

→ Impact on rate of change & diffusion of innovation?
→ Solving the threshold problem?
→ Evolution of languages during prehistory

- Take new parameters into account:
  - clustering coefficient
  - Average path length
  - See Ke & Gong's talk « Language change and social networks »
  - negative links and diversity

A ‘real’ exemple of internal constraints: phonological systems

- PSs as complex dynamical systems:
  - Interactions between vowels [Lindblom, 1984] [de Boer, 2001]

- Extracting knowledge from a database
  - The UPSID database : 451 PSs as a representative sample of world’s langagues (Maddieson & Precoda 1989)
  - Statistical approach and GA to extract the relevant fitness landscape from the data

Conclusions

- A model integrating internal and external constraints in a single framework
    - quite technical (takes time to describe)
    - but captures a large number of situations (including previous models as naming games etc.)

- Key-points:
    - Investigate the impact of various social structures
    - Deal with highly-dimensional fitness landscapes
    - Possible additional features to the model: other descriptions of the linguistic systems

- Perspectives:
    - Integrate more data from real cases in the model (ex. UPSID)
    - Investigate the prehistory of languages by studying the impact of our predecessors' social structures

Thank you for your attention

Acknowledgements

- Members of the « MEL » team (CNRS OHLL program)
- Members of the Language Engineering Laboratory, City Univ. of HK

Limitations

- No interactions between social and natural constraints
  - only competition for the most likely variation
  - Specific influence of social ties on linguistic structures

Lardil (Queensland, Australie) (Evans, 2003)

Nya-rri	*ngithun*	*thabu*	*waangkur*	*riwur*.
1exc-du.HAR	my	elder.brother	go:FUT	east:FUT

« My elder brother and I will go east »

  - Not a central phenomenon
- Independent computations for the linguistic variables

Prehistoric linguistic diversity and rate of evolution

- With time, increasing contacts between groups, slow increase of population density
- Human density and social networks

  (Jacquesson, 2000)

- An oversimplified scenario for the evolution of languages