Competing random walks behind language change

Jean-Pierre Nadal
CNRS & EHESS
nadal@lps.ens.fr

Laboratoire de Physique Statistique
(LPS, UMR 8550 CNRS - ENS – UPMC – Univ. Paris Diderot)
Ecole Normale Supérieure (ENS)

&

Centre d’Analyse et de Mathématique Sociales
(CAMS, UMR 8557 CNRS - EHESS)
Ecole des Hautes Etudes en Sciences Sociales (EHESS)

Structural Complexity in Natural Language(s) - Paris, 30-31 May 2016

Language change

Social phenomena: learning/adaptation in a population of interacting individuals

Basic mechanism: reinforcement, imitation

Locutors tend to reproduce what they have heard and/or what they have already produced in the recent past.

Such self-reinforcement dynamics (at the scale of a population) is ubiquitous – it is observed for sounds, words, syntactic structures, etc...

This reinforcement/imitation dynamics is at the root of models of language emergence, language learning and language evolution.

Most models focus on the convergence of a population towards a shared language.

Language change: models

Category learning/evolution in linguistics:

- levels: phonemes, lexicon, grammar, semantics…
- time scales: language origin, language evolution, language acquisition, adult learning

Some references (among many others):

Phonetic categories

M. A. Erickson and J. K. Kruschke, « Rules and Exemplars in Category Learning », 1998
P.-Y. Oudeyer, « The self-organization of speech sounds », 2005

Lexicon

L. Steels: Talking heads experiment

Lexical/semantic level

F. Cucker, S. Smale, D.-X. Zhou, « Modelling Language Evolution », 2004
conditions for convergence

Semantics

B. Victorri, « Continuity and Discreteness in Lexical Semantics », 1996, 2004
conditions for polysemy in a permanently evolving system

Language change: models

continuous level
internal representation
object/semantic space
neural activity ?

discrete level
sent/perceived category
phonemes, words…

B. Victorri, C. Fuchs, *La polysémie, construction dynamique du sens*, Hermès, 1996; B. Victorri, « Continu et discret en sémantique lexicale », 2004.

M. A. Erickson and J. K. Kruschke, « Rules and Exemplars in Category Learning » , 1998

J. B. Pierrehumbert , « Exemplar dynamics: Word frequency, lenition and contrast » , 2000

F. Cucker, S. Smale, D.-X. Zhou, « Modelling Language Evolution », 2004

G. J. Baxter, R. A. Blythe, W. Croft & A. J. McKane, « Utterance selection model of language change », 2006

…

Adaptive dynamics

sent signal (category) → noisy transmited signal → perceived signal

perceived signal ↓ updating of internal representation

updating of internal representation → new probability for sending a category

Modeling of the learning/adaptation dynamics

- Work with Janet Pierrehumbert (Northwestern, USA)

    Phonemes: frequencies of use

- Work with Quentin Feltgen (LPS, ENS) & Benjamin Fagard (Lattice, ENS)

    Grammaticalization

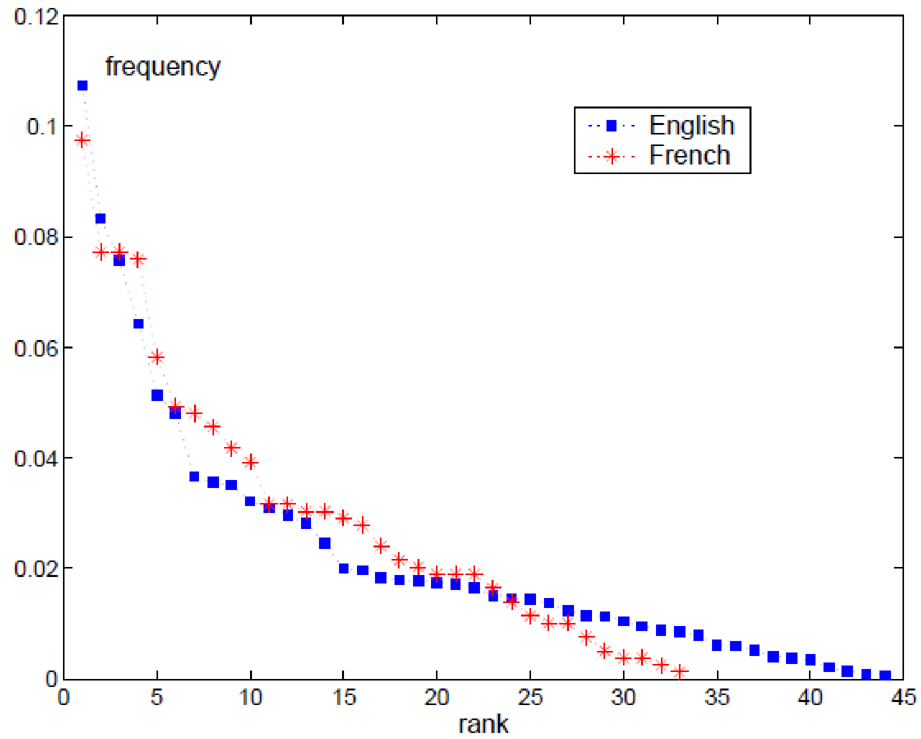
Modeling of the learning/adaptation dynamics

- Work with Janet Pierrehumbert (Northwestern, USA)

Phonemes: frequencies of use

J. B. Pierrehumbert, Exemplar dynamics: Word frequency, lenition and contrast , 2000

J. Bybee, Frequency of use and the organization of language. Oxford Univ. Press, 2007.



Modeling of the evolution of frequency of use

- M categories (phonemes)
- Production: at each time step a randomly chosen agent i sends a signal (category) to another agent, j

probabilistic choice:

$p_m^{(i)}(t) \equiv$ Probability that i sends m at time t $(m = 1,\ldots, M)$

- Perception by agent j

$q_m^{(j)}(t)$ = proba of perceiving category m at t (perception noise)

Then, update of the probabilities for producing each category

* forgetting: for every category,

$$p_m^{(j)}(t+1) = \lambda\, p_m^{(j)}(t) \qquad \text{where} \quad \lambda < 1$$

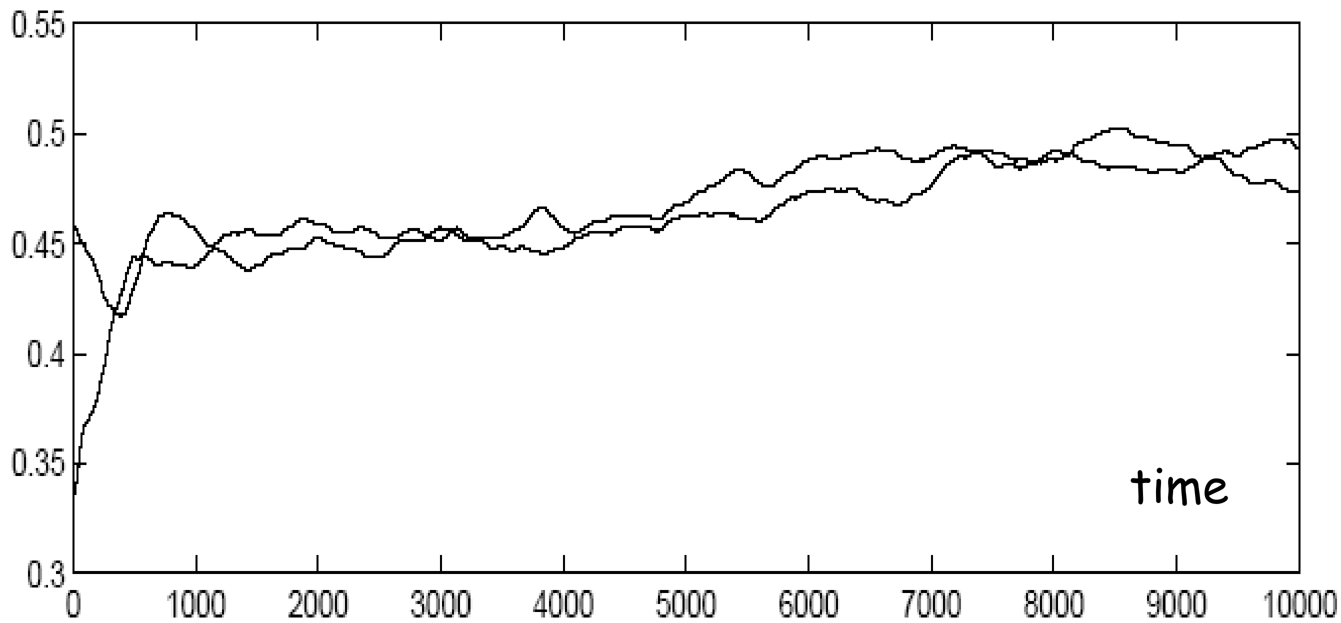
* reinforcement: for the category m perceived at t:

$$p_m^{(j)}(t+1) = \lambda\, p_m^{(j)}(t) + 1 - \lambda$$

• Simple example: 2 agents, 2 categories 'a', 'b'

frequency entrainment

P(sending 'a')



$\varepsilon = 0.02$

Uniform perception noise:

probability ε to misperceive
in which case perception of any category with equal probability

Representative agent (« mean-field » approach)

- $m = 1, \ldots, M$ categories
- $p_m(t)$ = proba producing category m at time t
- $q_m(t)$ = proba perceiving category m at time t

$$= \Sigma_n \, C(m,n)\, p_n(t)$$ where $= C(m,n)$ confusion matrix

Simplest case: uniform noise

$$q_m(t) = (1 - \varepsilon)\, p_m(t) + \varepsilon / M$$

- reinforcement mechanism:

$$p_m(t+1) = \lambda\, p_m(t) + (1 - \lambda)\, \xi_m(t)$$

where: $\xi_m(t) = 1$ if m = perceived category at time t

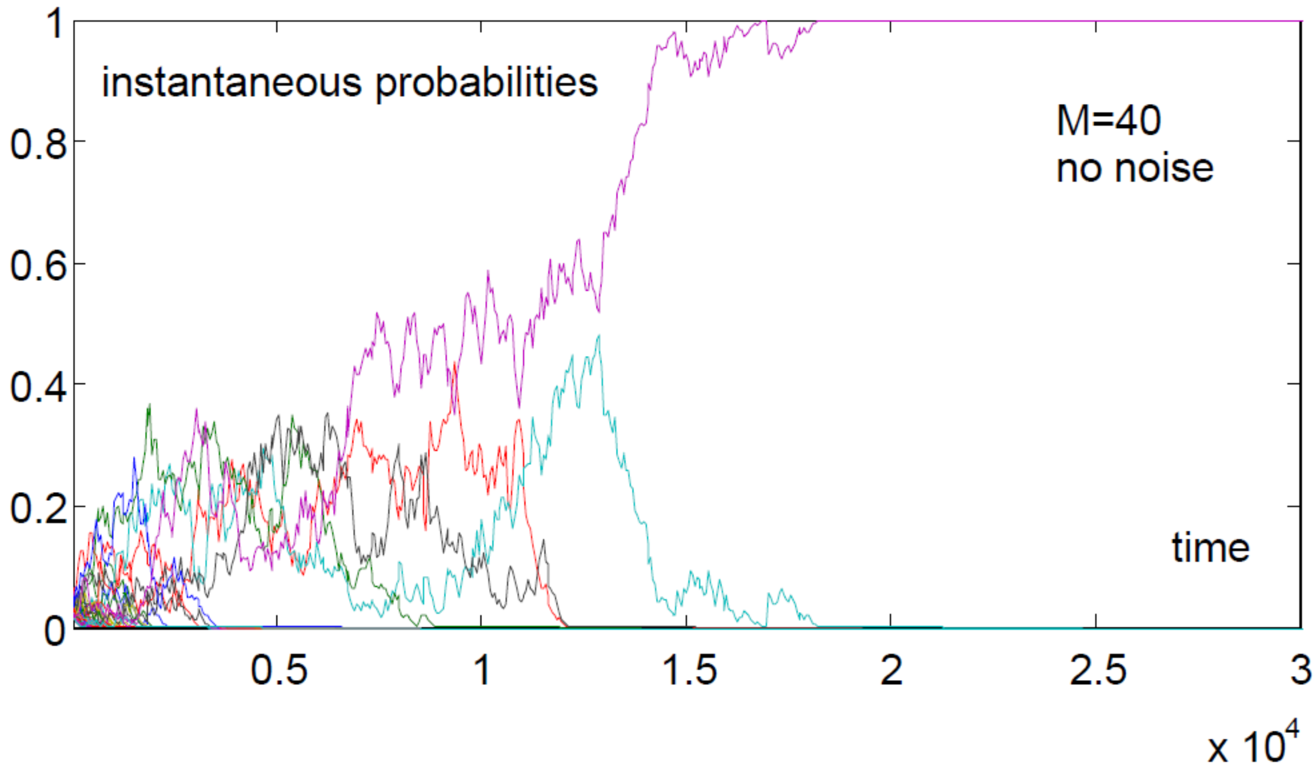
0 otherwise

Rem.: time scale: $\tau = - 1 / \log(\lambda)$

Dynamics: asymptotics

- Reinforcement -> 'winner take all' mechanism

  without noise, only one category survives



Dynamics: asymptotics

- Reinforcement -> 'winner take all' mechanism

  without noise, only one category survives
- Noise (confusion) -> mixing

Average dynamics: $\langle \, . \, \rangle \equiv$ average over all possible histories

$$\langle p_m(t) \rangle = \lambda \langle p_m(t) \rangle + (1-\lambda) \sum_n C(m,n) \langle p_n(t) \rangle$$

$t \to \infty$

$$\langle p_m \rangle_\infty = \sum_n C(m,n) \langle p_n \rangle_\infty$$

asymptotic state

= eigenvector of the confusion matrix C for the largest eigenvalue

(unique if C irreducible)

($\approx$ particular case of Cucker, Smale & Zhou, 2003)

Hence: mean values of category frequencies simply reflects the confusion matrix

Simplest case

- uniform noise

$q_m(t)$ = proba perceiving category m at time t $(m=1,\ldots,M)$

$$= (1 - \varepsilon)\, p_m(t) + \varepsilon / M$$

Average dynamics:

$$\langle p_m(t+1) \rangle = \lambda \langle p_m(t) \rangle + (1-\lambda) [(1 - \varepsilon) \langle p_m(t) \rangle + \varepsilon / M]$$

$$= \langle p_m(t) \rangle - (1-\lambda)\, \varepsilon\, [\langle p_m(t) \rangle - 1/M]$$

$$\langle p_m \rangle_\infty = 1/M$$

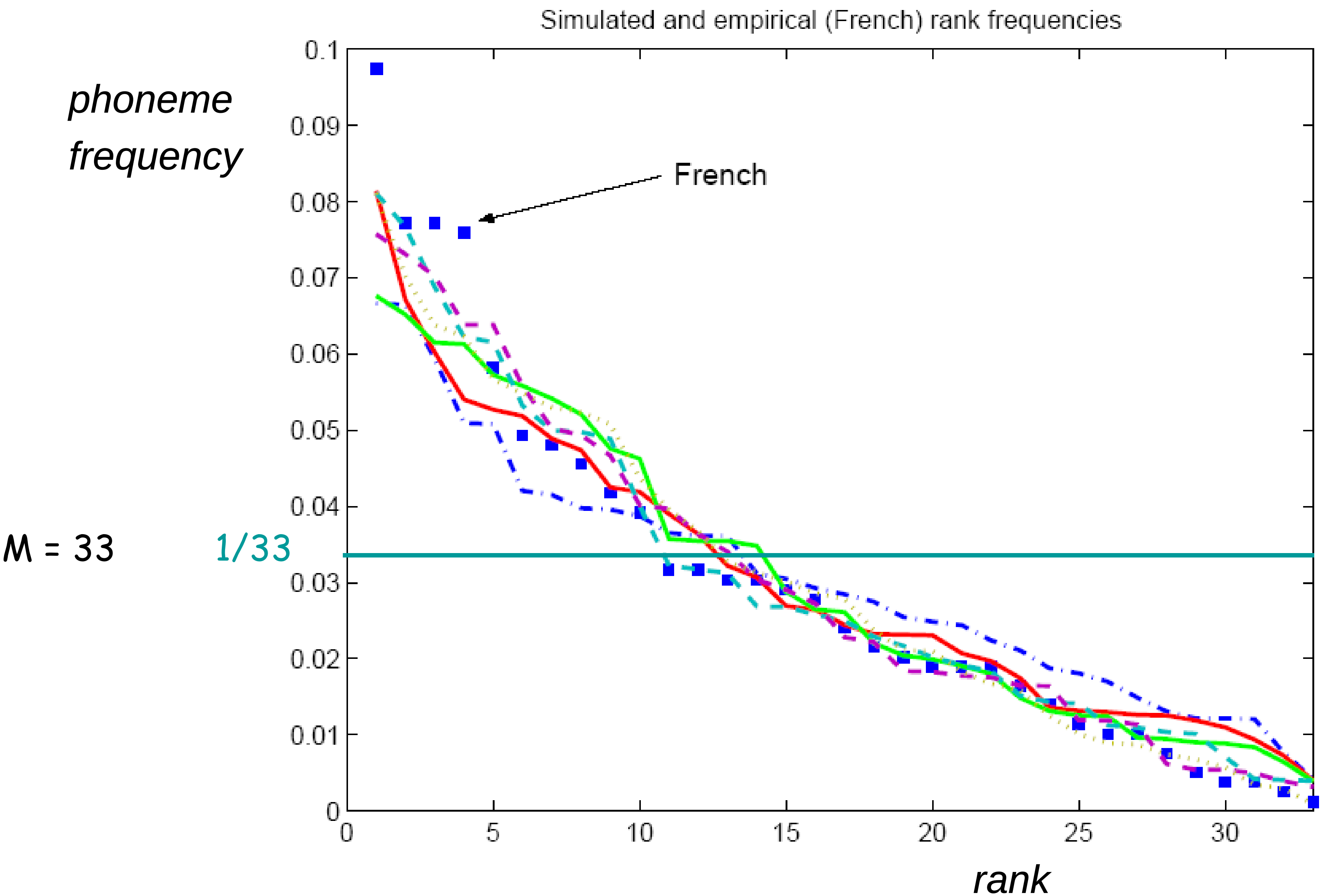
uniform distribution

Predicts typical frequencies all equal

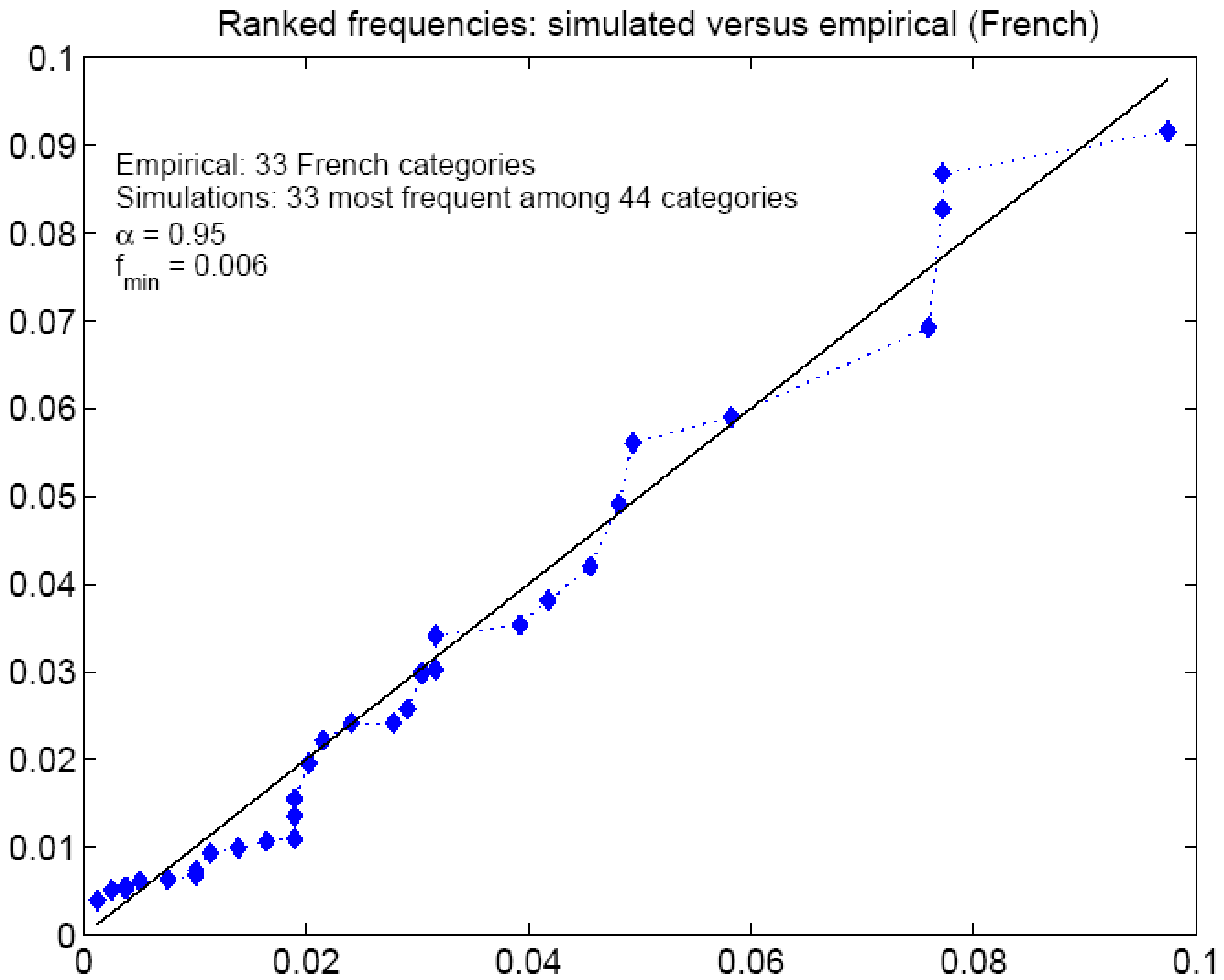
Yet, fluctuations around the mean:

$$\langle (p_m - \langle p_m \rangle_\infty)^2 \rangle_\infty = (1/M)\,(1 - 1/M)\,(1-\lambda) / (1-\lambda + 2 \lambda \varepsilon)$$

Empirical data & numerical simulations



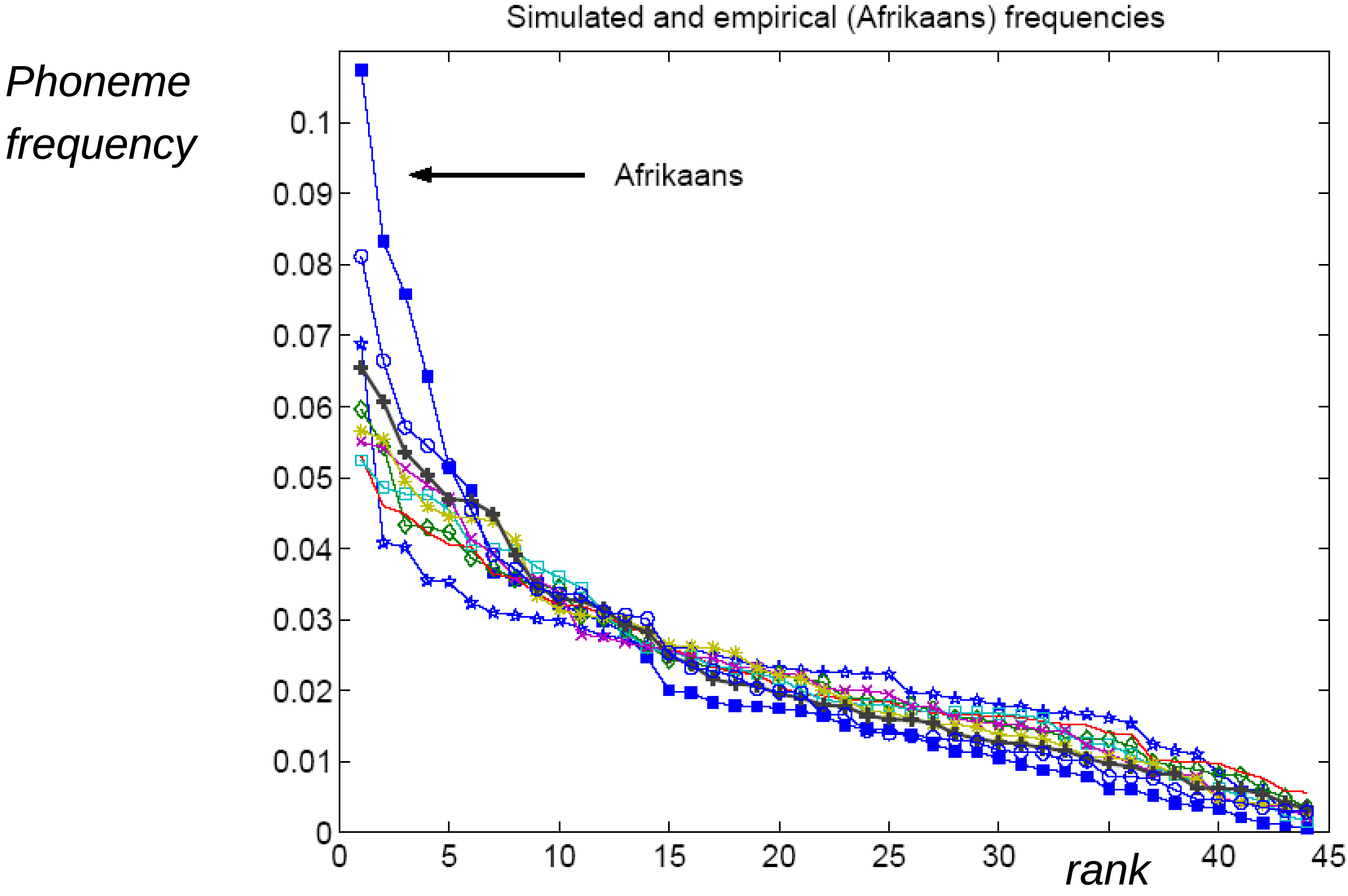
Simulations vs empirical data



phonemes frequencies (simulation)

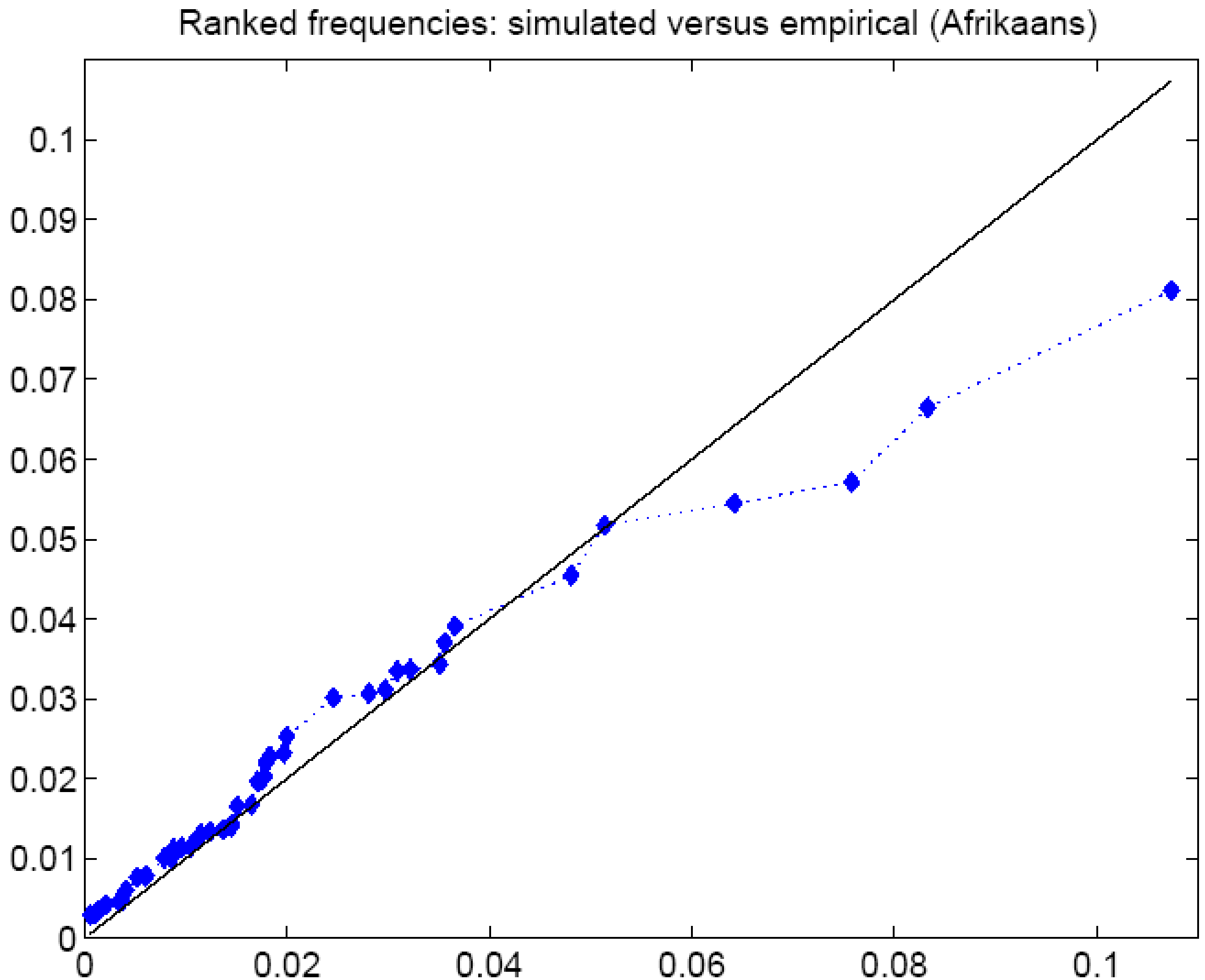
French phonemes frequencies

Simulations vs empirical data



Simulations vs empirical data

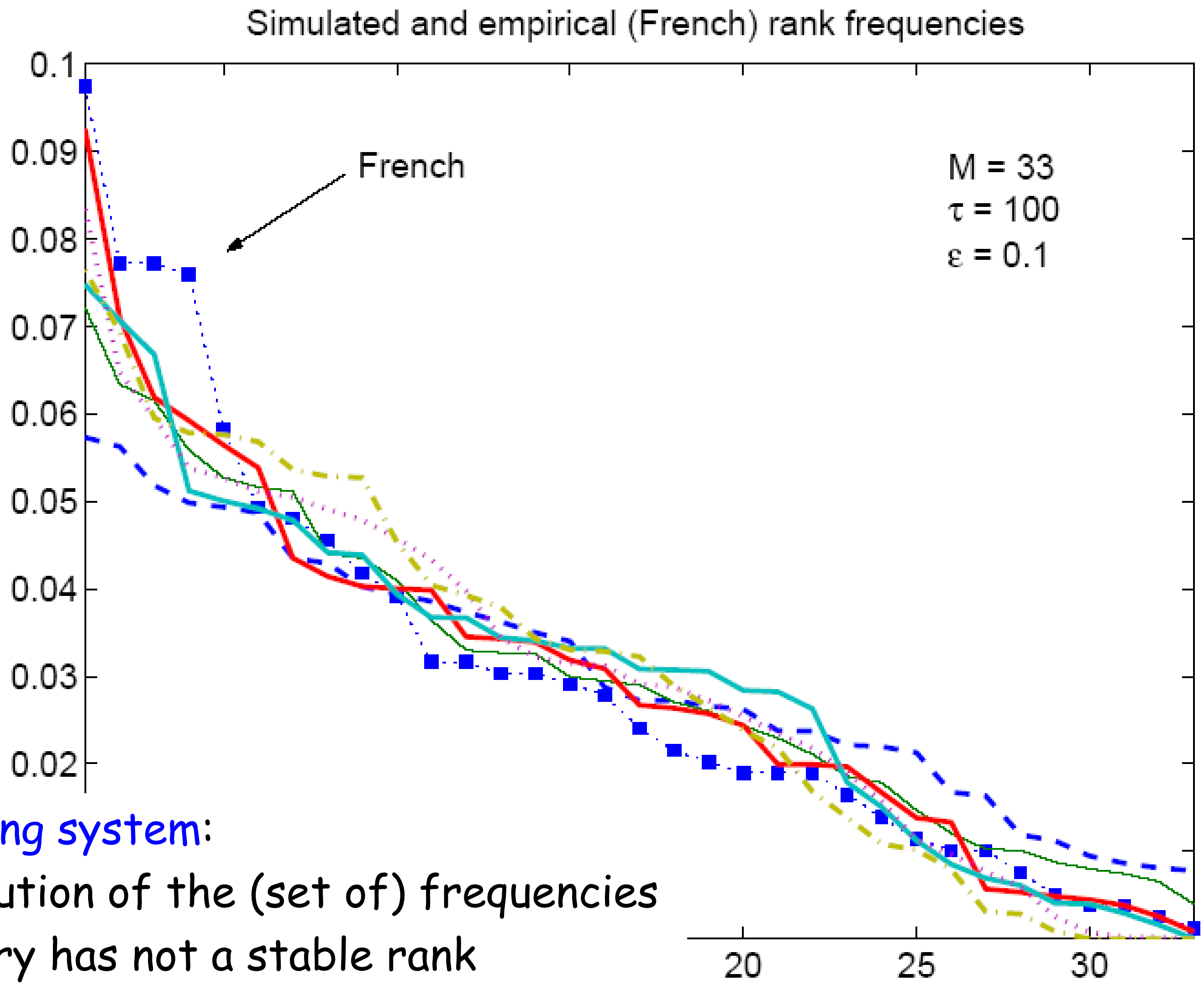
Ranked phonemes frequencies (simulation)



Afrikaans phonemes frequencies

Simulations vs empirical data

phonemes frequencies



Permanently evolving system:

Stationary distribution of the (set of) frequencies but a given category has not a stable rank

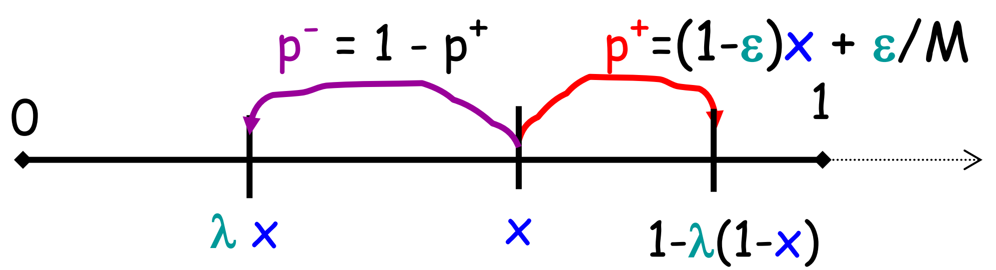
Analysis: ramdom walk process

- reinforcement / imitation mechanism $\rightarrow$ random walk process:

$$x(t+1) = \lambda\, x(t) + (1-\lambda)\, \xi(t)$$

where: $\xi(t) = 1$ with probability: $(1-\varepsilon)\, x(t) + \varepsilon / M$
$\qquad\qquad\quad 0$ otherwise

$0 \leq x(t) \leq 1$



- *Master equation*

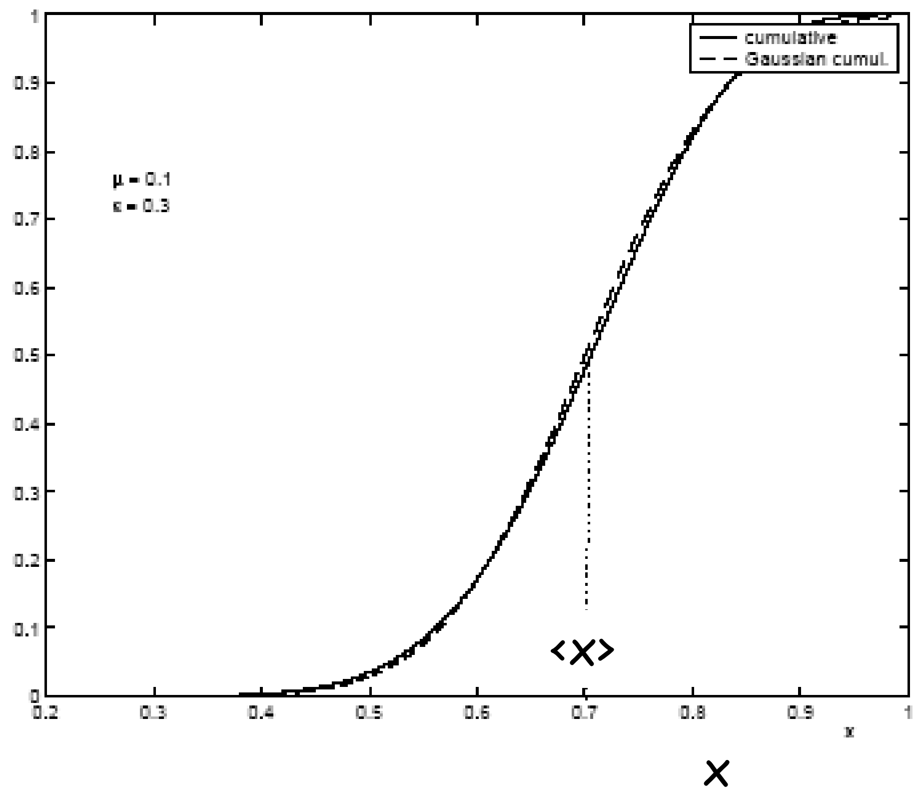
$$P_{t+1}(x) = \Theta(x - 1 + \lambda)\, [(\varepsilon/M) + (1-\varepsilon)\,(x - 1 + \lambda) / \lambda]\, (1/\lambda)\, P_t((x - 1 + \lambda) / \lambda)$$
$$+\, \Theta(\lambda - x)\, [1 - \varepsilon/M - (1-\varepsilon)\, x/\lambda]\, (1/\lambda)\, P_t(x / \lambda)$$

where: $\Theta(x) = 1$ if $x > 0$, and 0 otherwise

•Smooth regime: quasi Gaussian behaviour near the mean value

(for small $\mu = 1 - \lambda$)

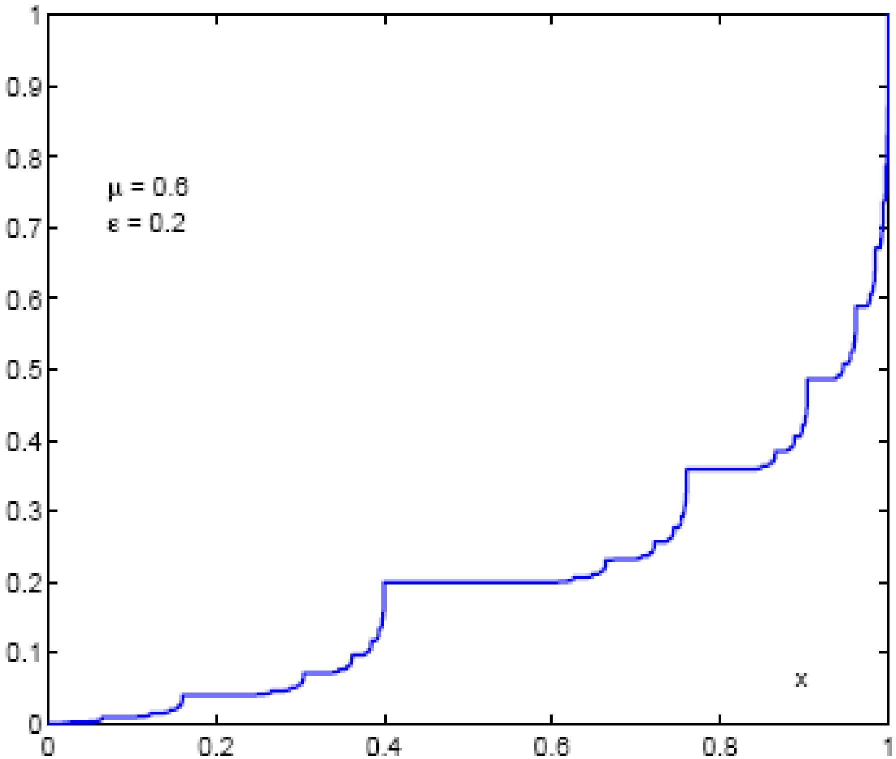
Cumulative distribution of x



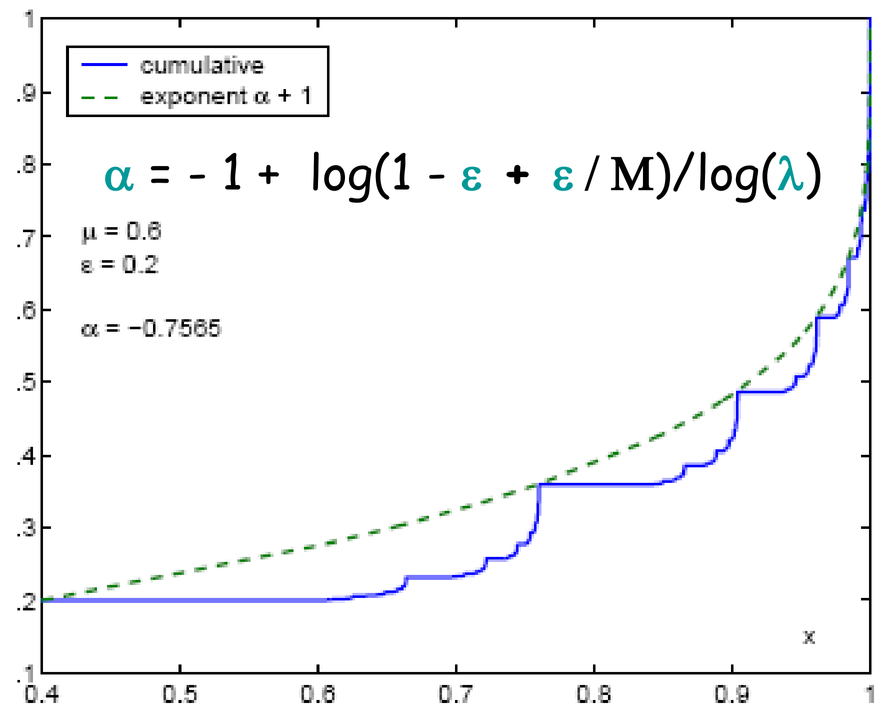
Singular regime: infinite number of singularities $\varepsilon / M < 1 - \lambda$

- Fractal support.
- Power law behaviour near the boundary (singularities)
- Implication: very long time spent near the singularities

Cumulative distribution of x



Cumulative distribution of x near x = 1

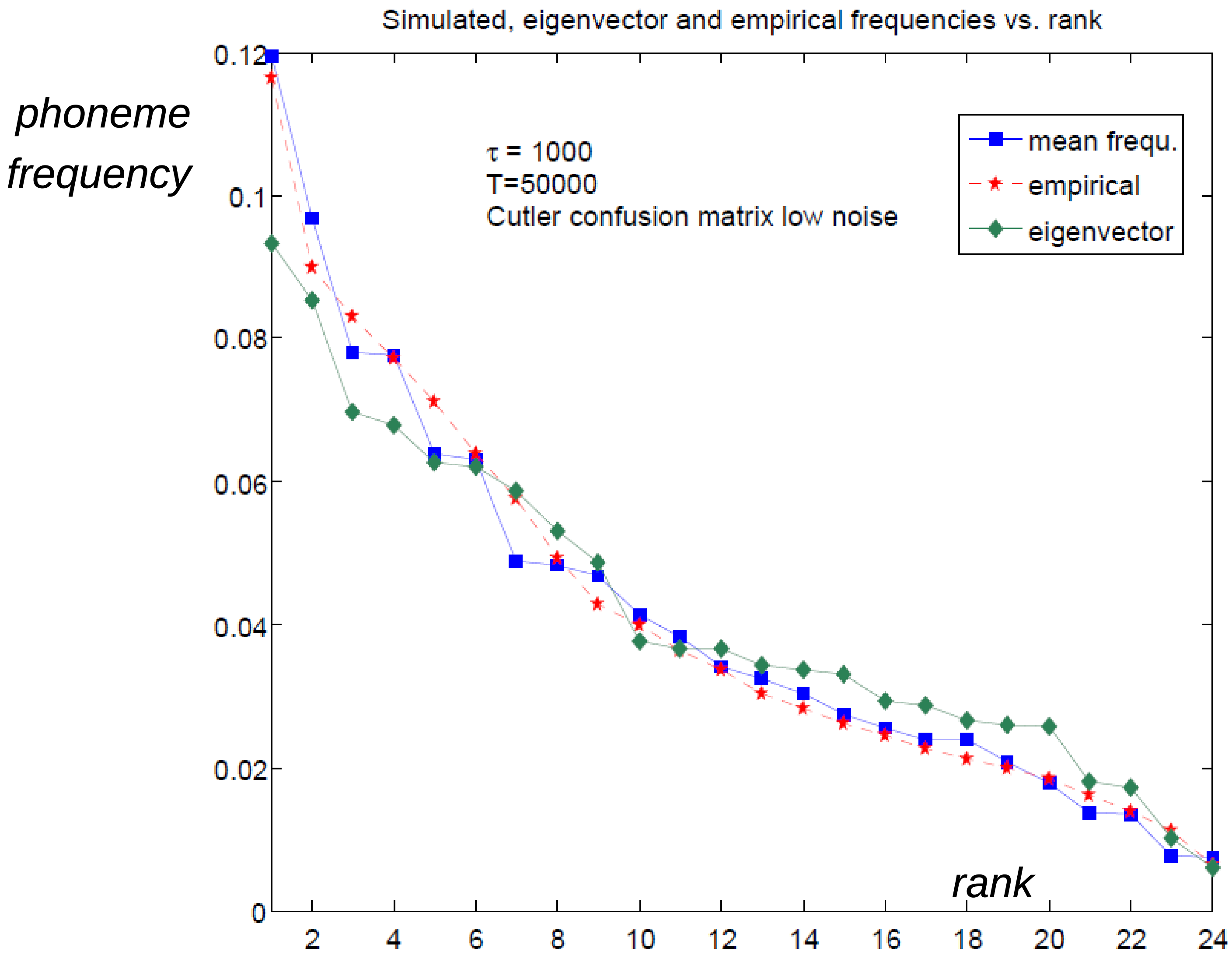


$$\alpha = -1 + \log(1 - \varepsilon + \varepsilon / M)/\log(\lambda)$$

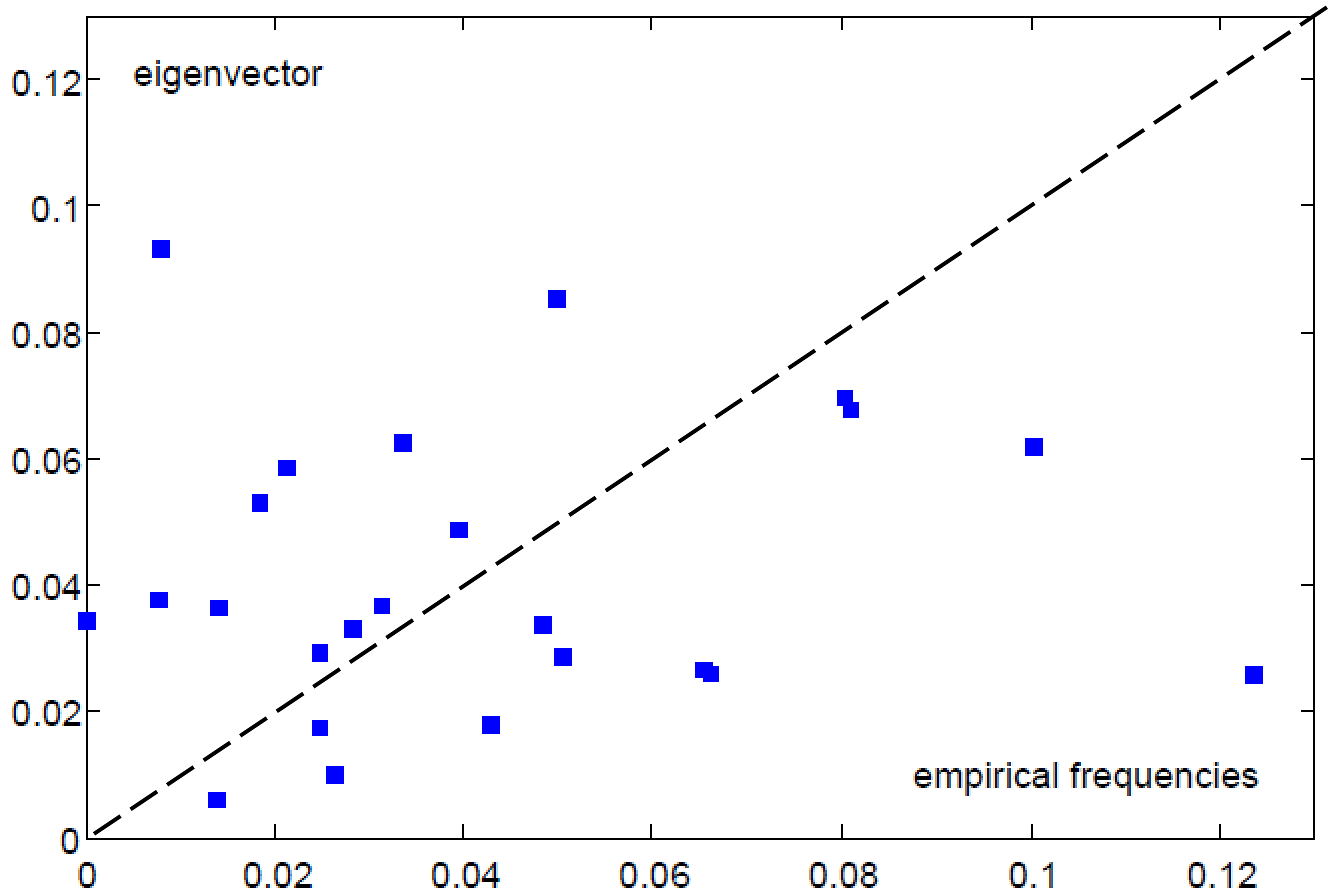
Back to

« mean values of category frequencies reflects the confusion matrix »

Making use of the empirical Confusion matrix (Cutler et al 2004)



Yet, the eigenvector does not give the correct ordered list of phonemes (although positive correlation)



Production confusion matrix? Other constraints? Syllabic or word contexts?

Modeling of the adaptation dynamics

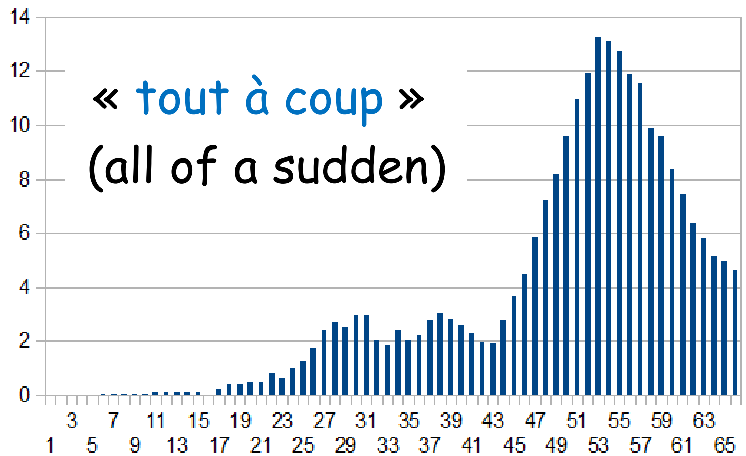
- Work with **Quentin Feltgen** (LPS ENS) & Benjamin Fagard (Lattice, ENS)

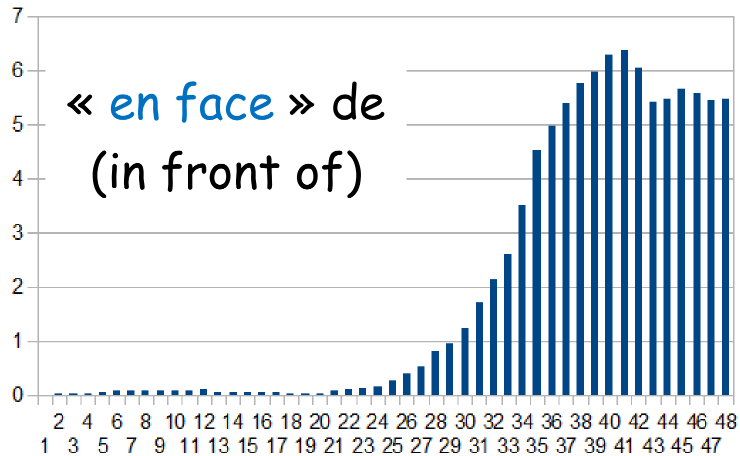
Grammaticalization

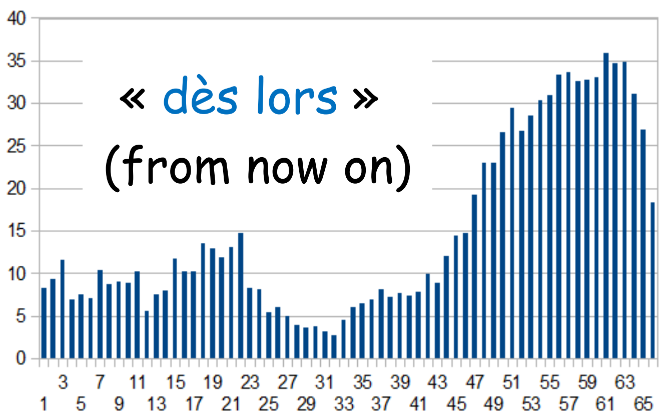
the process by which a non-grammatical item
acquires a grammatical status

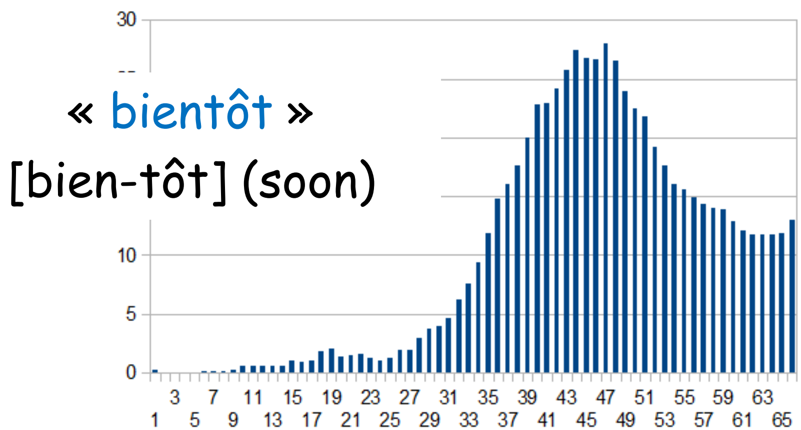
Grammaticalization

Database: Frantext (French corpus, from the 10th to the 21st century)



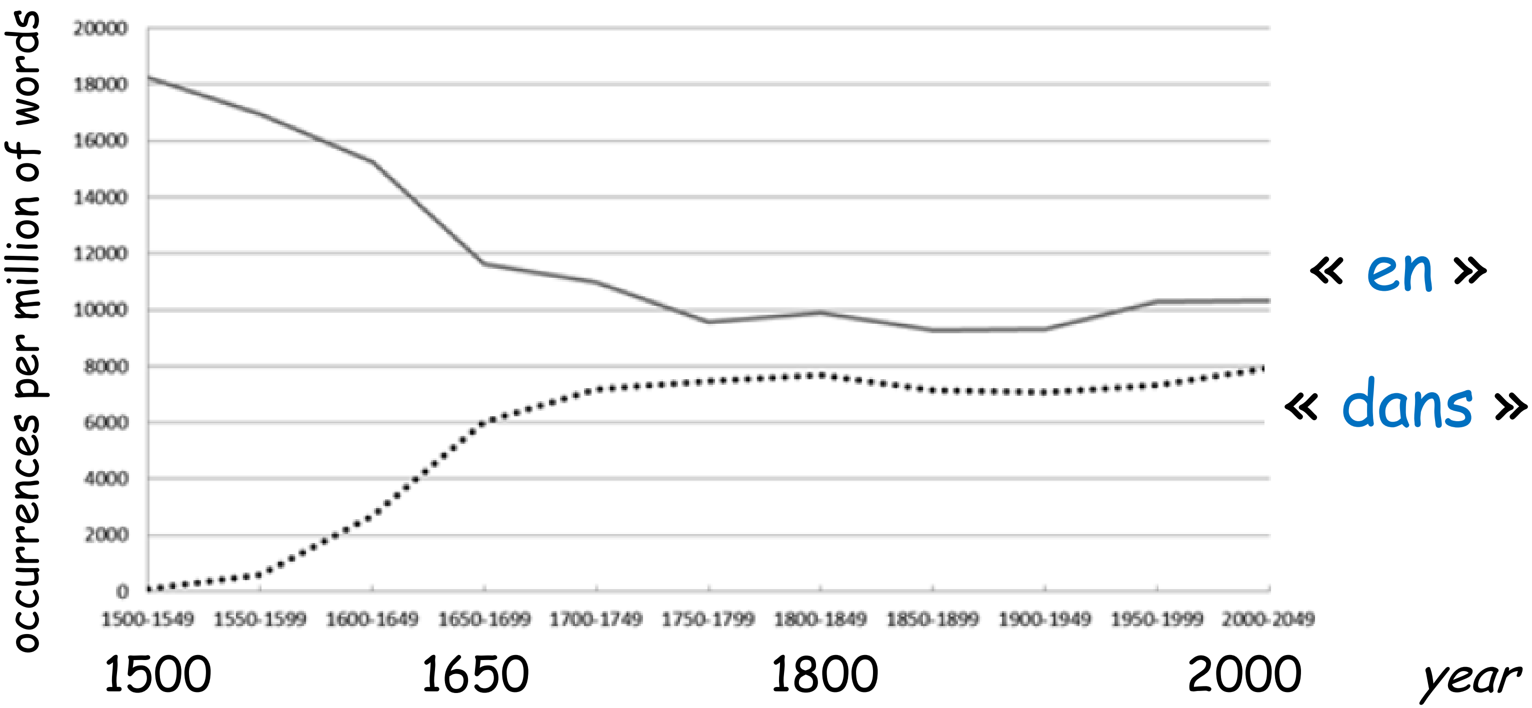






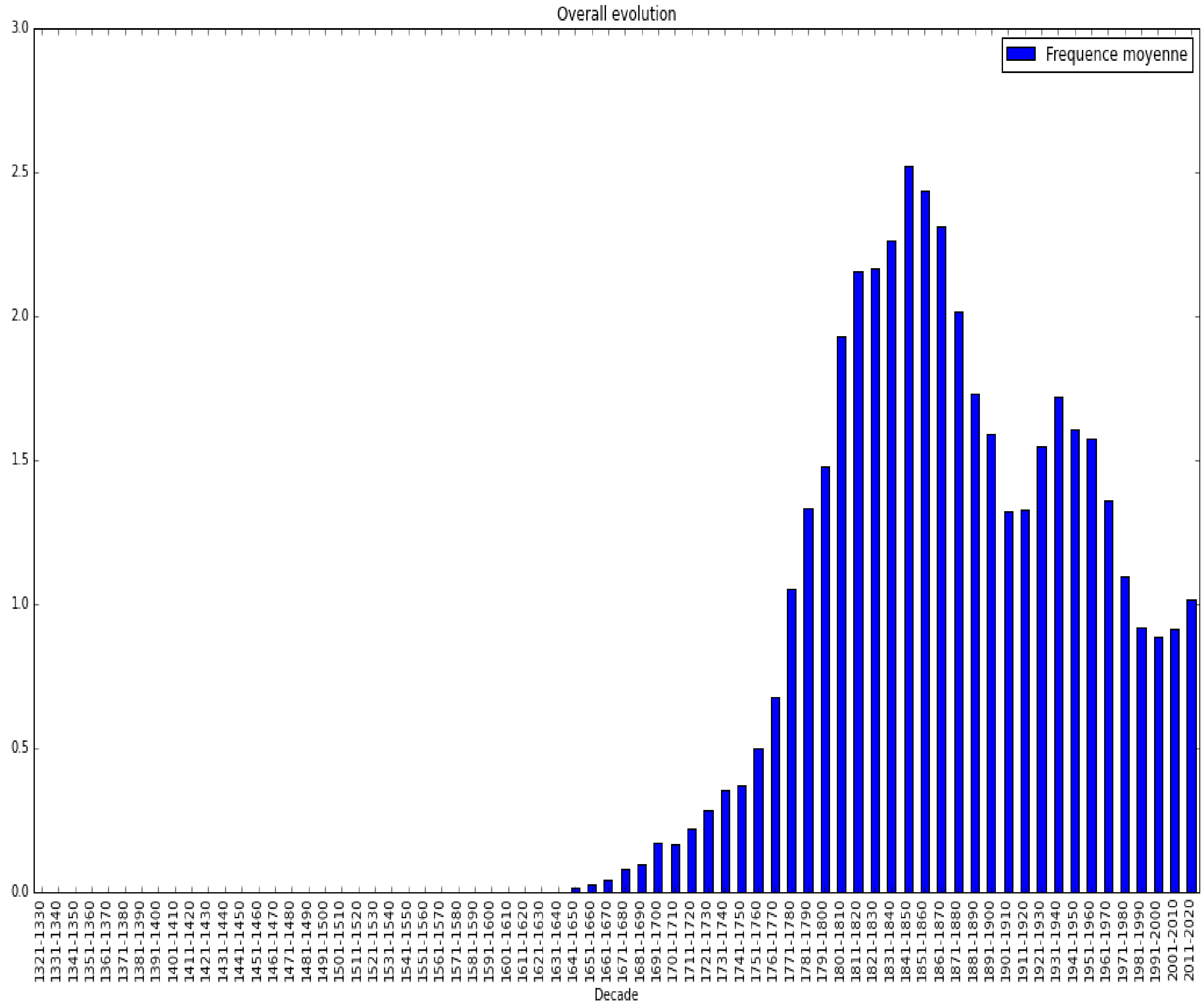
Grammaticalization

- competition



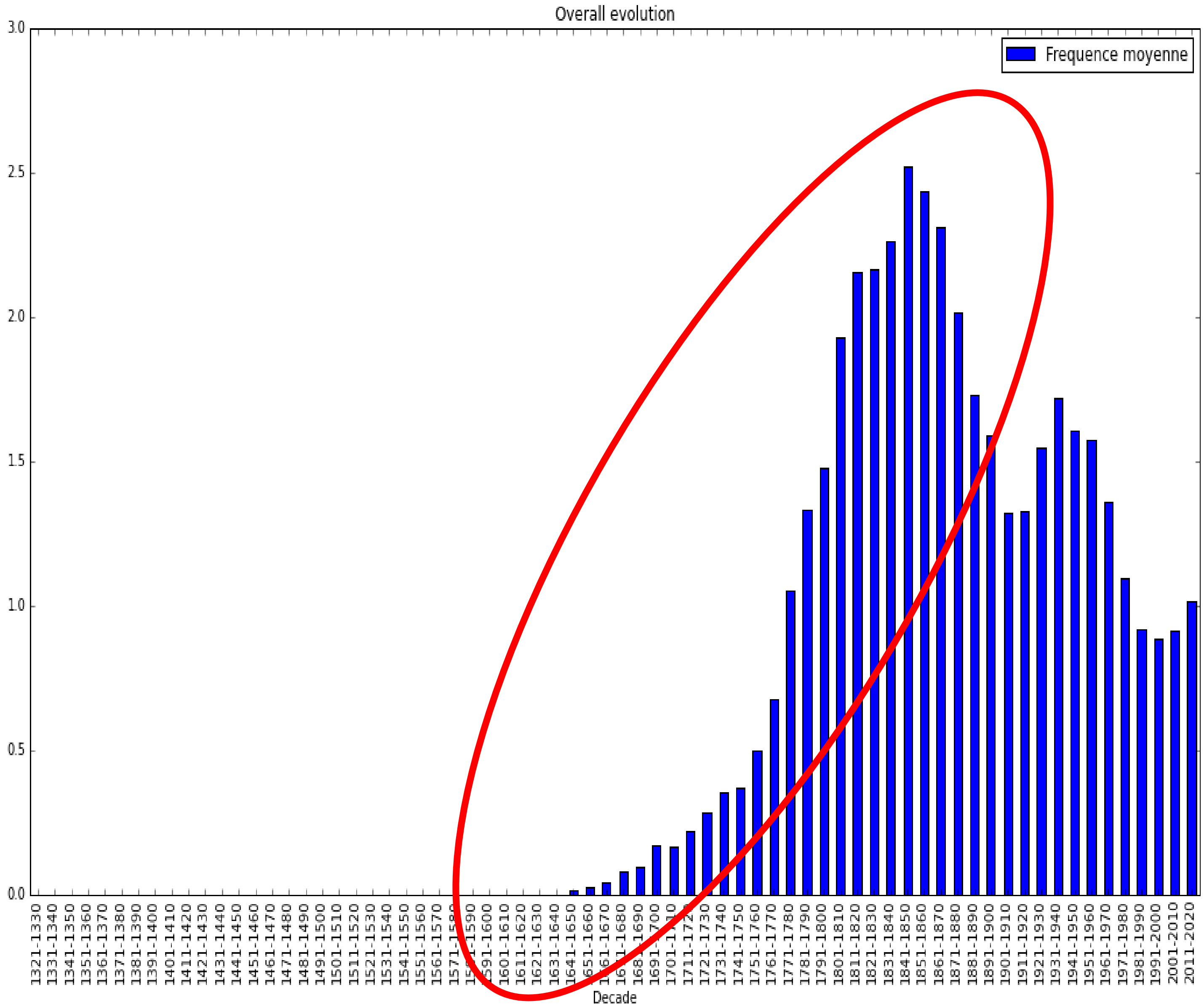
Fagard & Combettes, 2013

A l'insu de

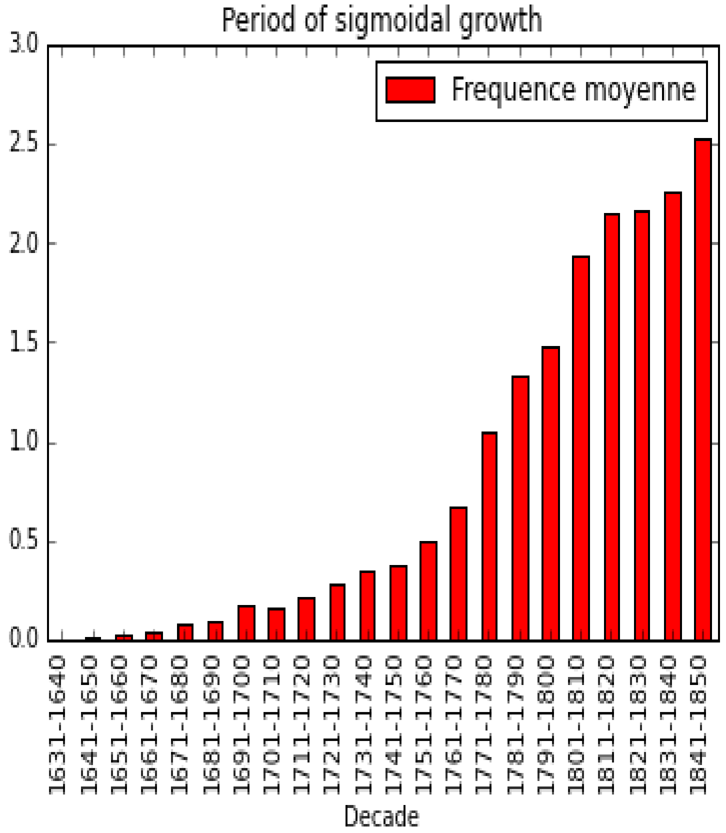


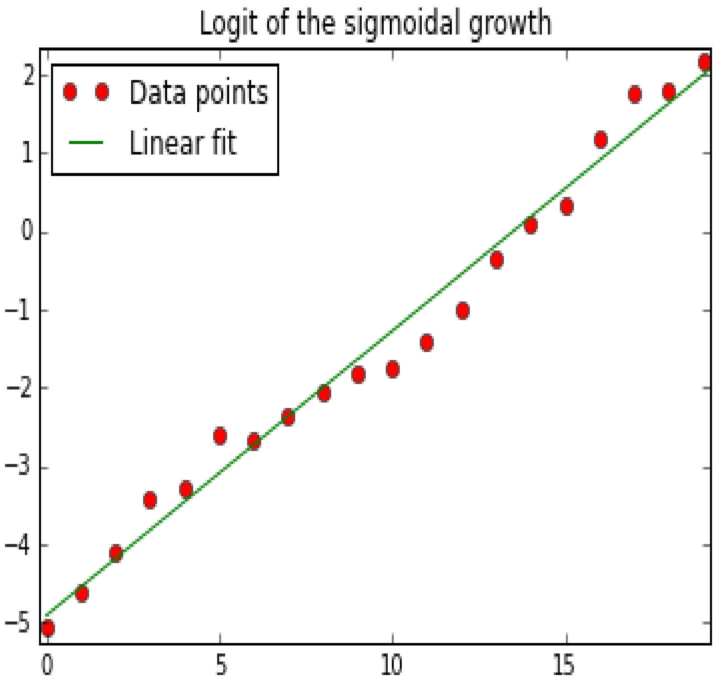
Overall evolution
Frequence moyenne
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Decade
1321-1330
1331-1340
1341-1350
1351-1360
1361-1370
1371-1380
1381-1390
1391-1400
1401-1410
1411-1420
1421-1430
1431-1440
1441-1450
1451-1460
1461-1470
1471-1480
1481-1490
1491-1500
1501-1510
1511-1520
1521-1530
1531-1540
1541-1550
1551-1560
1561-1570
1571-1580
1581-1590
1591-1600
1601-1610
1611-1620
1621-1630
1631-1640
1641-1650
1651-1660
1661-1670
1671-1680
1681-1690
1691-1700
1701-1710
1711-1720
1721-1730
1731-1740
1741-1750
1751-1760
1761-1770
1771-1780
1781-1790
1791-1800
1801-1810
1811-1820
1821-1830
1831-1840
1841-1850
1851-1860
1861-1870
1871-1880
1881-1890
1891-1900
1901-1910
1911-1920
1921-1930
1931-1940
1941-1950
1951-1960
1961-1970
1971-1980
1981-1990
1991-2000
2001-2010
2011-2020

A l'insu de

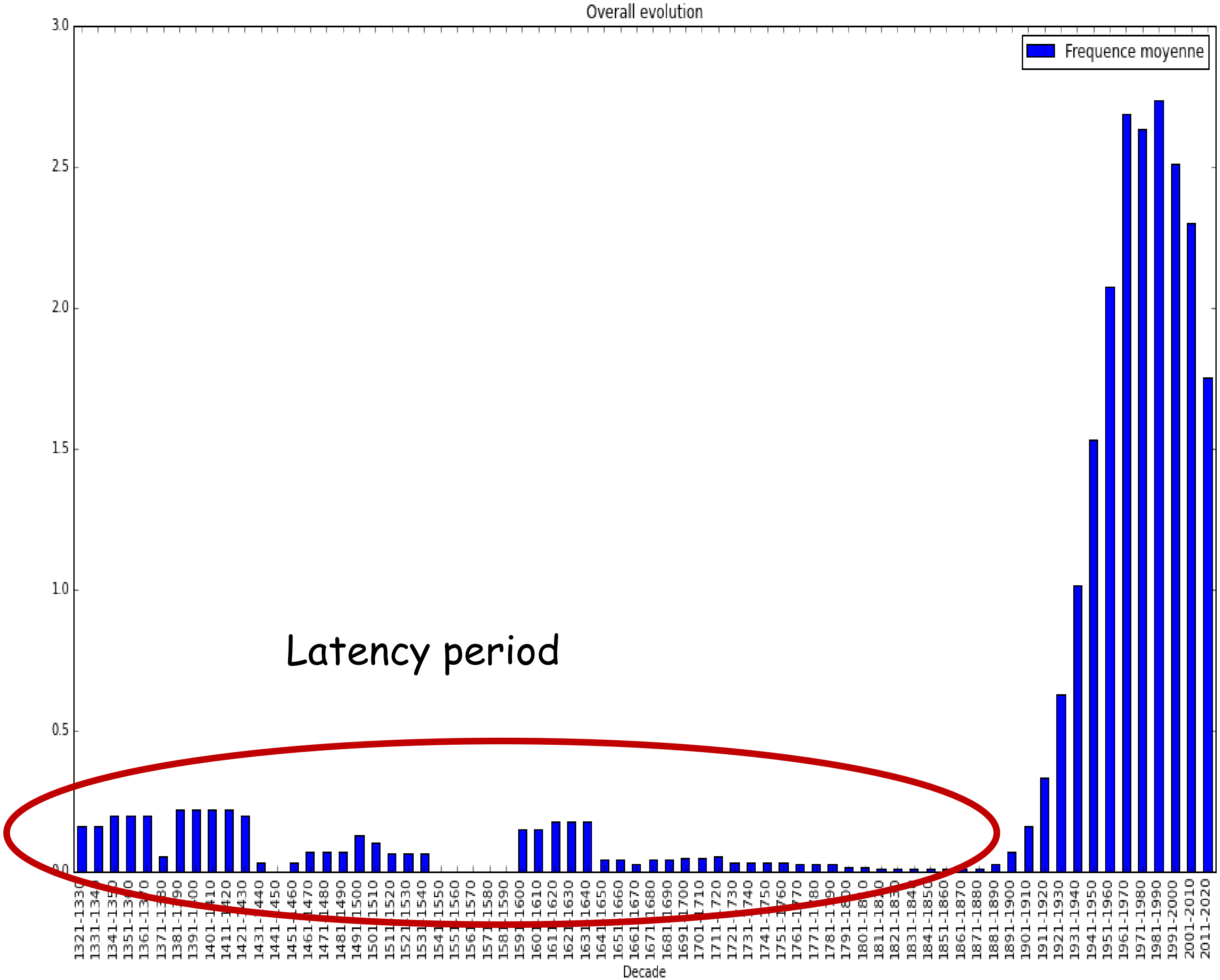


A l'insu de


Period of sigmoidal growth
Frequence moyenne
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1631-1640
1641-1650
1651-1660
1661-1670
1671-1680
1681-1690
1691-1700
1701-1710
1711-1720
1721-1730
1731-1740
1741-1750
1751-1760
1761-1770
1771-1780
1781-1790
1791-1800
1801-1810
1811-1820
1821-1830
1831-1840
1841-1850
Decade


Logit of the sigmoidal growth
Data points
Linear fit
2
1
0
−1
−2
−3
−4
−5
0
5
10
15

Par ailleurs



Basic hypotheses:

- Augmentation of frequency =

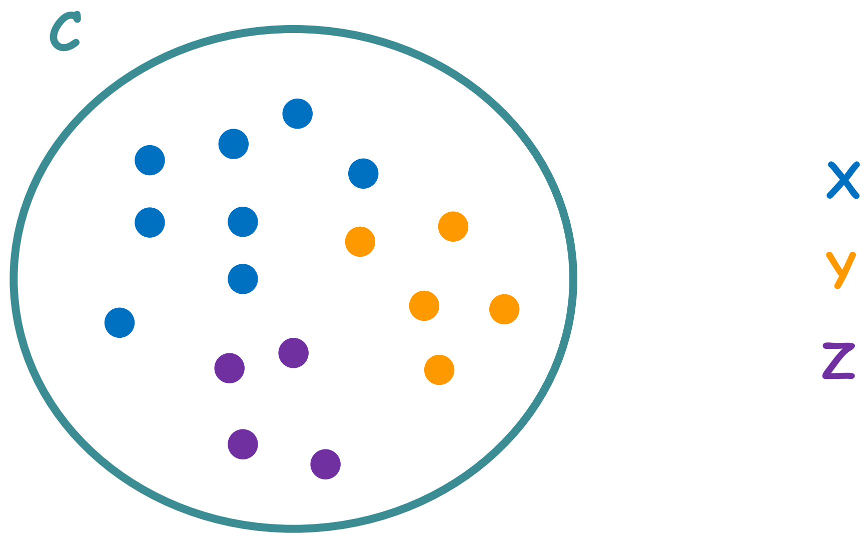
semantic expansion (rather than a more frequent use of a current meaning)

Model:

- Exemplar type approach: for every meaning (semantic context), population of occurrences
- Network in the space of concepts
- Finite memory

Model

- Each concept (or semantic context) C is characterized by a set of exemplars: set of forms used to express this concept, each one with its number of occurences



Basic scheme:

Total number of balls kepf fixed (finite memory)

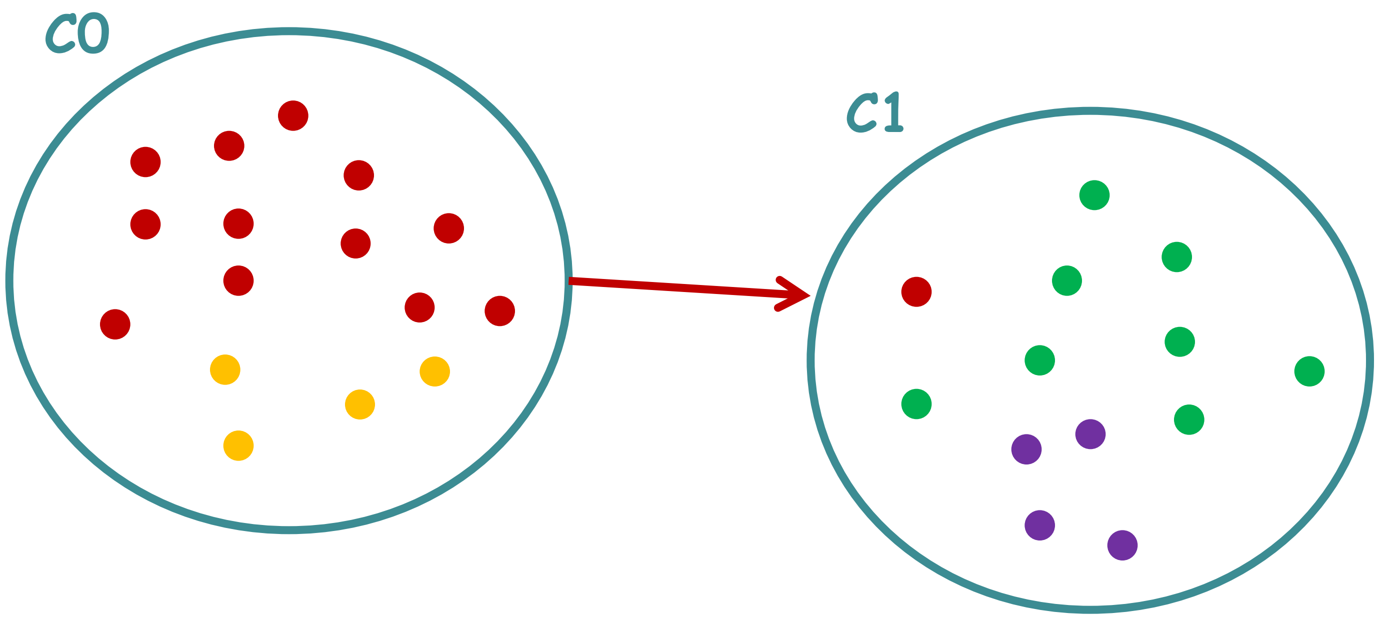
Probability to express C with X = fraction of blue balls

❑ Reinforcement:

In case of perception of X: remove a ball taken at random, add a blue ball

Model

- Network of concepts



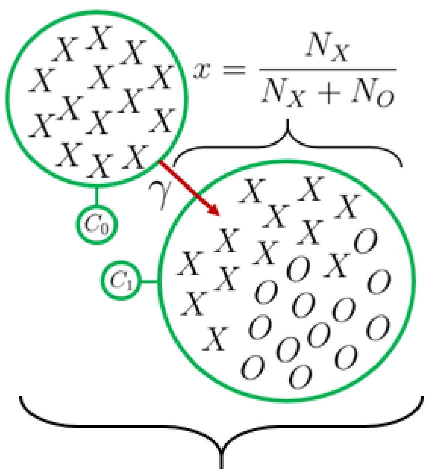
❑ Diffusion to sites with strong conceptual links (adding balls at connected site)

Case of a competition between two variants

- Initially *X* and O are used to express C0 and C1, respectively
- Increase of the conceptual link from C0 to C1:

X enters in competition with O to express the same meaning as O

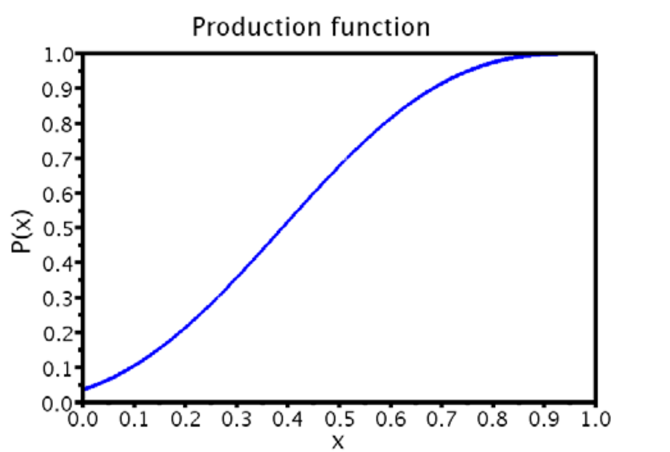
(this may come from a need for expressivity in context C1: individuals may make use of a new way to attract the attention of their interlocutors)

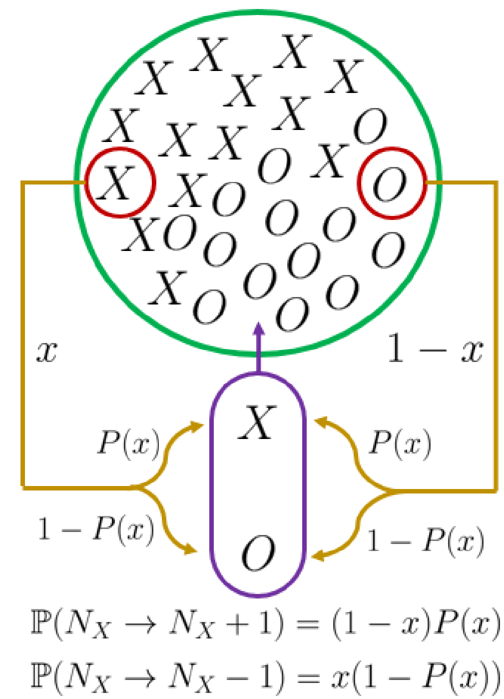


$$f = \frac{x + \gamma}{1 + \gamma}$$

Effective frequency (increases with γ)

- Probability to choose *X*: has to be an increasing function of f, saturating at 1 → nonlinear function of f, hence of x, P(x)



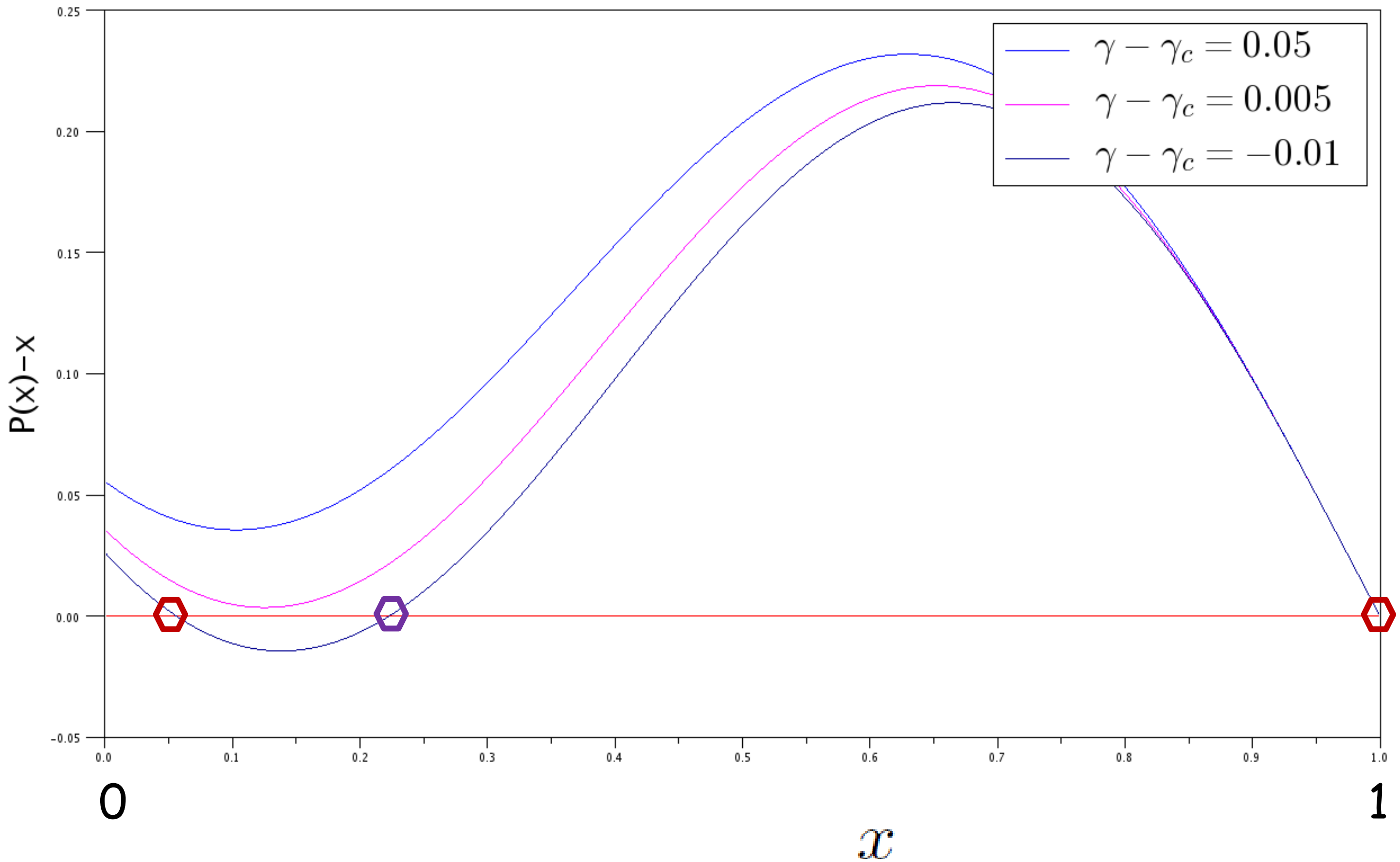


$$\dot{x} = P(x) - x$$

$$\dot{x} = P(x) - x$$

(continuous time limit)

Dynamics near criticality



stable fixed points

unstable fixed point

Model

- **Results**

Depends on the strength of the conceptual link γ.

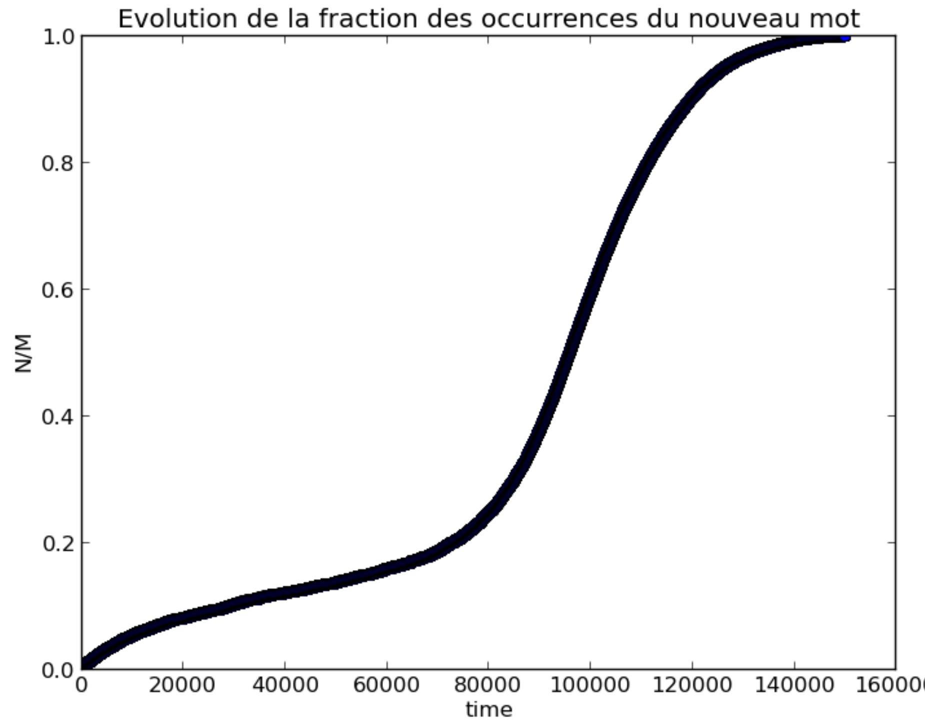
If too small, grammaticalization does not occur.

There is a critical value above which the new item will dominate.

Just above the critical value:

- Long phase with a low frequency for the new item.

-Then fast transition with a sigmoidal shape until the new item dominates.



Model extension

- Vectorial representation of concepts

- Meaning of a word: computed from its use in different contexts

- Diffusion in the network: evolution of meaning

➔ emergence of semantic bleaching

Q Feltgen, B Fagard & JPN, TAL, Volume 55 Num. 3, pp. 47-71

Conclusion

- The model takes into account the reinforcement mechanism + specific features of linguistics & cognitive aspects (finite working memory)
- It reproduces stylized facts

Ongoing work

- More data: other languages
- Mathematical analysis

Collaborators & references

Janet Pierrehumbert,
Northwestern University, Evanston

phonemes, frequency of use
2007 unpublished - J. Pierrehumbert, « Sustaining linguistic complexity », Keynote address, Society for language development, Boston, Nov. 1, 2007.

Quentin Feltgen,
PhD student, Laboratoire de Physique Statistique, (LPS, UMR 8550 CNRS - ENS - UPMC - Univ. Paris Diderot), Ecole Normale Supérieure, Paris

&

Benjamin Fagard
Laboratoire Langues, Textes, Traitements informatique, Cognition (LaTTiCe, UMR 8094 CNRS - ENS - Université Paris 3), Ecole Normale Supérieure, Paris

Grammaticalization
Q Feltgen, B Fagard & JPN, TAL, Volume 55 Num. 3, pp. 47-71, online May 2016
Same authors, Chapter in "Language in Complexity: The Emerging Meaning", Springer, to appear.