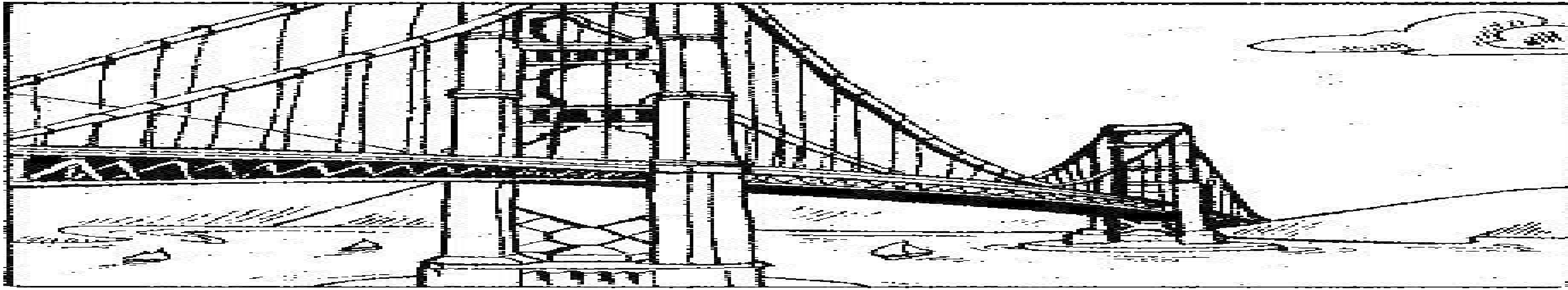
Mathematical models of language shift and reversing language shift

Alcàntera Research Group

Sociolinguistics Symposium 21



Alcàntera Research Group

Manel Perucho, Department of Astronomy and Astrophysics, Universitat de València

Francisco Grimaldo, Department of Informatics, Universitat de València

Emilia López-Iñesta, Department of Informatics, Universitat de València

Clara Miralles, Student Master of Physics, Universität Heildelberg

Ernest Querol, Department of Arts and Humanities, Universitat Oberta de Catalunya

Weaving bridges between Sociolinguistics and mathematical applied methods

Alcàntera Research Group: the bridge with Sociolinguistics

Summary

1. Dynamical models to predict the evolution of the use of languages
2. Dynamical models Analysing the data from Valencian Country and Catalonia
    2.1. Dynamical Model of Language shift: Analysing the data from Valencian Country
    2.2. Dynamical Model of Reversing Language Shift: Analysing the data from Catalonia
3. Conclusions
4. New researchs

1. Dynamical models to predict the evolution of the use of languages

1. Dynamical models to predict the evolution of the use of languages

- **Forerunners**

- 2003: Abrams and Strogatz publish in *Nature* in which they studied Language death Dynamics.

- 2005: Mira and Paredes improved the model of Abrams and Strogatz

1. Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Hendler, G. *Nature* **412**, 819–822 (2001).
2. McPhedran, R. C. *et al. Aust. J. Chem.* **54**, 241–244 (2001).
3. Cattaneo-Vietti, R. *et al. Nature* **383**, 397–398 (1996).
4. Perry, C. C. & Keeling-Tucker, T. *J. Biol. Inorg. Chem.* **5**, 537–550 (2000).
5. Simpson, T. L. *The Cell Biology of Sponges* (Springer, New York, 1984).
6. Sarikaya, M. *et al. J. Mater. Res.* **16**, 1420–1428 (2001).
7. Marcuse, D. *Principles of Optical Fiber Measurements* (Academic, New York, 1981).
8. Clegg, W. J., Kendall, K., Alford, N. M., Button, T. W. & Birchall, J. D. *Nature* **347**, 455–457 (1990).
9. Kamat, S., Su, X., Ballarini, R. & Heuer, A. H. *Nature* **405**, 1036–1040 (2000).
10. Cha, J. N., Stucky, G. D., Morse, D. E. & Deming, T. J. *Nature* **403**, 289–292 (2000).
11. Kroger, N., Lorenz, S., Brunner, E. & Sumper, M. *Science* **298**, 584–586 (2002).

Competing financial interests: declared none.

Linguistics

Modelling the dynamics of language death

Thousands of the world's languages are vanishing at an alarming rate, with 90% of them being expected to disappear with the current generation[1]. Here we should yield the same transition probability as a swap in the fraction of speakers and relative status; thus $P_{xy}(x,s) = P_{yx}(1-x, 1-s)$. We also assume that no one will adopt a language that has no speakers ($P_{yx}(0,s) = 0$) or no status ($P_{yx}(x,0) = 0$), and that P_{yx} is smooth and monotonically increasing in both arguments.

These mild assumptions imply that equation (1) generically has three fixed points. Of these, only $x=0$ and $x=1$ are stable. The model therefore predicts that two languages cannot coexist stably — one will eventually drive the other to extinction.

To test our model, we collected data on the number of speakers of endangered languages in 42 regions of Peru, Scotland, Wales, Bolivia, Ireland and Alsace-Lorraine, four instances of which are shown in Fig. 1. We fit the model's solutions to the data, assuming transition functions of the forms $P_{yx}(x,s) = cx^a s$ and $P_{xy}(x,s) = c(1-x)^a(1-s)$. Unexpectedly, the exponent a was found to be roughly constant across cultures, with $a = 1.31 \pm 0.25$ (mean ± standard deviation; further details are available from the authors).

histories of countries where two languages coexist today generally involve split populations that lived without significant interaction, effectively in separate, monolingual societies. Only recently have these communities begun to mix, allowing language competition to begin.

So what can be done to prevent the rapid disintegration of our world's linguistic heritage? The example of Quebec French demonstrates that language decline can be slowed by strategies such as policy-making, education and advertising, in essence increasing an endangered language's status. An extension to equation (1) that incorporates such control on s through active feedback does indeed show stabilization of a bilingual fixed point.

Daniel M. Abrams, Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA

e-mail: dma32@cornell.edu

1. Krauss, M. *Language* **68**, 4–10 (1992).
2. Nowak, M. A., Komarova, N. L. & Niyogi, P. *Nature* **417**, 611–617 (2002).
3. Chomsky, N. A. *Syntactic Structures* (Mouton, New York, 1957).
4. Kroch, A. *Lang. Variat. Change* **1**, 199–244 (1989).

90% of them being expected to disappear with the current generation[1]. Here we develop a simple model of language competition that explains historical data on the decline of Welsh, Scottish Gaelic, Quechua (the most common surviving indigenous language in the Americas) and other endangered languages. A linguistic parameter that quantifies the threat of language extinction can be derived from the model and may be useful in the design and evaluation of language-preservation programmes.

Previous models of language dynamics have focused on the transmission and evolution of syntax, grammar or other structural properties of a language itself[2–7]. In contrast, the model we describe here idealizes languages as fixed, and as competing with each other for speakers. For simplicity, we also assume a highly connected population, with no spatial or social structure, in which all speakers are monolingual.

Consider a system of two competing languages, X and Y, in which the attractiveness of a language increases with both its number of speakers and its perceived status[8] (a parameter that reflects the social or economic opportunities afforded to its speakers). Suppose an individual converts from Y to X with a probability, per unit of time, of $P_{yx}(x,s)$, where x is the fraction of the population

further details are available from the authors).

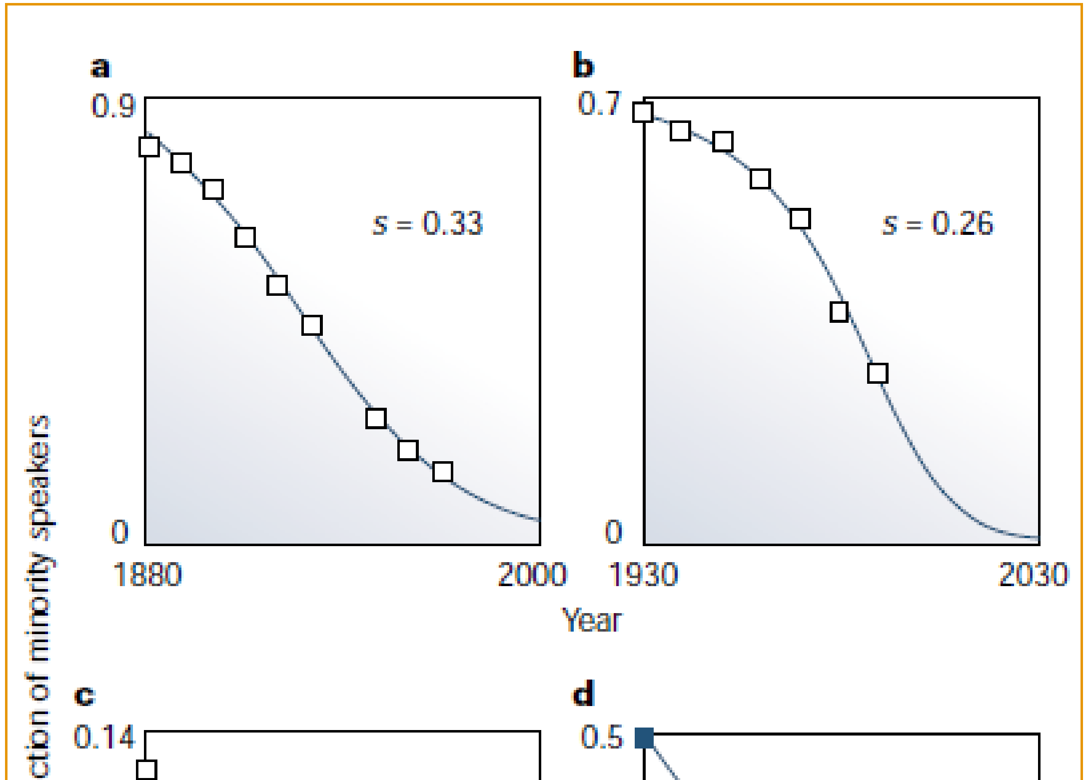
Of the remaining parameters, status, s, is the most relevant linguistically; it could serve as a useful measure of the threat to a given language. Quechua, for example, still has many speakers in Huanuco, Peru, but its low status is driving a rapid shift to Spanish, which leads to an unfortunate situation in which a child cannot communicate with his or her grandparents.

Contrary to the model's stark prediction, bilingual societies do, in fact, exist. But the

New York, 1957).
4. Kroch, A. *Lang. Variat. Change* **1**, 199–244 (1989).
5. Hawkins, J. A. & Gell-Mann, M. *The Evolution of Human Languages* (Addison-Wesley, Reading, Massachusetts, 1992).
6. Niyogi, P. & Berwick, R. C. *Ling. Phil.* **20**, 697–719 (1997).
7. Lightfoot, D. *The Development of Language: Acquisition, Changes and Evolution* (Blackwell, Oxford, 1999).
8. Fishman, J. A. *Reversing Language Shift* (Multilingual Matters, Philadelphia, 1991).
9. Withers, C. W. J. *Gaelic in Scotland 1698–1981: The Geographical History of a Language* (Donald, Edinburgh, 1984).
10. Aitchison, J. & Carter, H. *A Geography of the Welsh Language* (Univ. Wales Press, Cardiff, 1994).
11. UK Office of Population Censuses and Surveys. *1991 Census: Report for Wales, 1* (HMSO, London, 1993).

Competing financial interests: declared none.

Figure 1 The dynamics of language death. Symbols show the proportions of speakers over time of: **a**, Scottish Gaelic in Sutherland, Scotland[9]; **b**, Quechua in Huanuco, Peru; **c**, Welsh in Monmouthshire, Wales[10]; **d**, Welsh in all of Wales, from historical data[10] (blue) and a single modern census[11] (red). Fitted curves show solutions of the model in equation (1), with parameters c, s, a and $x(0)$ estimated by least absolute-values regression. Where possible, data were obtained from several population censuses collected over a long timespan; otherwise a single



tion of syntax, grammar or other structural properties of a language itself[2–7]. In contrast, the model we describe here idealizes languages as fixed, and as competing with each other for speakers. For simplicity, we also assume a highly connected population, with no spatial or social structure, in which all speakers are monolingual.

Consider a system of two competing languages, X and Y, in which the attractiveness of a language increases with both its number of speakers and its perceived status[8] (a parameter that reflects the social or economic opportunities afforded to its speakers). Suppose an individual converts from Y to X with a probability, per unit of time, of $P_{yx}(x,s)$, where x is the fraction of the population speaking X, and $0 \leqslant s \leqslant 1$ is a measure of X's relative status. A minimal model for language change is therefore

$$\frac{dx}{dt} = yP_{yx}(x,s) - xP_{xy}(x,s) \qquad (1)$$

where $y = 1 - x$ is the complementary fraction of the population speaking Y at time t. By symmetry, interchanging languages

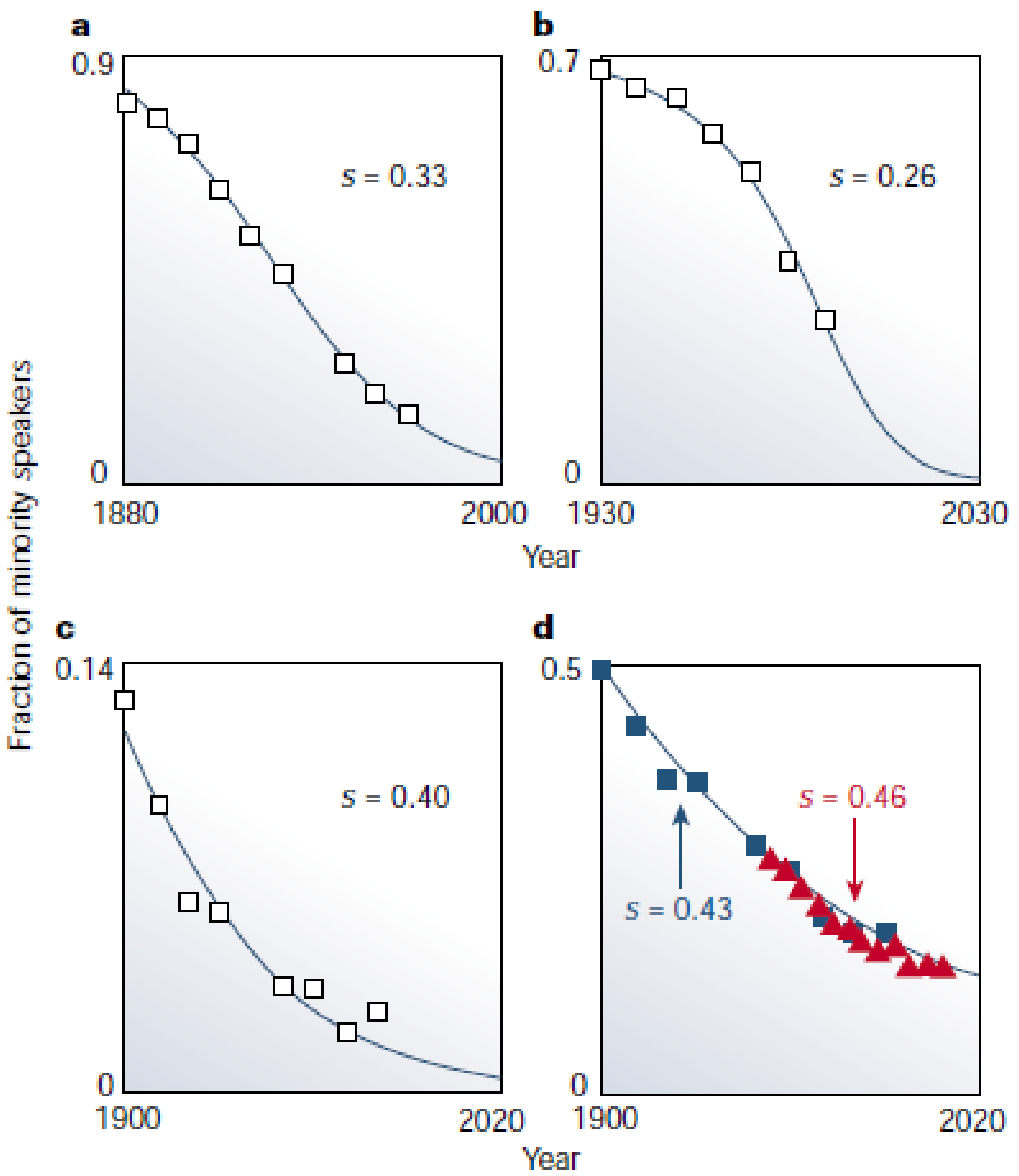


Figure 1 The dynamics of language death. Symbols show the proportions of speakers over time of: **a**, Scottish Gaelic in Sutherland, Scotland[9]; **b**, Quechua in Huanuco, Peru; **c**, Welsh in Monmouthshire, Wales[10]; **d**, Welsh in all of Wales, from historical data[10] (blue) and a single modern census[11] (red). Fitted curves show solutions of the model in equation (1), with parameters c, s, a and $x(0)$ estimated by least absolute-values regression. Where possible, data were obtained from several population censuses collected over a long timespan; otherwise, a single recent census with age-structured data was used (although errors are introduced, the size of which are reflected in the differing fits in **d**). Using the fraction of Catholic masses offered in Quechua in Peru as an indicator, we reconstructed an approximate history of the language's decline.

Evolution of the population of one species

- **Verhulst equation**: in the discret way
- **Lotka-Volterra equations**: in the continous way = **differential equacions**

Differential equations

Allow us:

- to modelize the changes produced according to one determined variable observing the changes in one relevant tangible (= physical) parameter and/or other linked parameters.
- The continuous obervation of the changes of the variables along infinitesimal time intervals (as in the differential concept).
- Describe the behaviour of the system as a whole.

At the last quarter of the 20th century a whole branch of the modern physics has gone a path to try to formalise systems related with the socials sciences, with the ecology studies and other domains (Strogatz, 1994; Kaplan and Glass, 1995).

Abrams and Strogatz (2003)

They develop a simple model for language competition (all speakers are monolinguals)

- **Population evolution of A speakers, determined by variable *x*, would be:**

  $$dx/dt = y\,P_{yx}(x,s) - x\,P_{xy}(x,s),$$

  being

  ***y* the number of speakers in Language B,**

  $P_{yx}(x,s)$ the probability that B speakers shift into group A depends on:

  1. **the number of A speakers (x) and**
  2. **the A language status, determined by parametre s. This parametre reflects the language appeal.**

- **Function of probability being expressed in these terms: $P_{yx}(x,s) = c\,x^{a}\,s$. In the equation *c* shows the estimation of the shift pace from A language to B language, meanwhile the exponent *a* controls the dependency of the probability of the language shift with the number of A speakers, *x*.**

Paredes and Mira Model (2005)

They included:

- **A third speaker group: bilinguals.**

- **A new parametre: likeness, *k*.**

  Modulates the probability to shift among groups.

Analysis of the Galician context: Spanish-Gallician.

2. Dynamical Models: Analysing the data from Valencian Country and Catalonia

Image NASA
© 2007 Tele Atlas
© 2007 Europa Technologies
Image © 2007 TerraMetrics
© 2007
Google
1382 km
Pointer lat 40.621147° lon 9.254831°
Streaming 100%
Eye alt 3435.60 km

Dynamical Models: Analysing the data from Valencian Country and Catalonia

- **Method**

Initial research of the **dynamics of the evolution of the language shift** in:

- the Valencian Country
- Catalonia

using mathematical models of differential equations

We discuss the application of Abrams and Strogatz (2003) and Mira and Paredes (2005) to empirical data obtained by official polls.

Dynamical Model of Language shift: Analysing the data from Valencian Country


Andorra
Catalunya del Nord
Franja de Ponent
Catalunya
Mar Catalana
País Valencià
Mallorca
Menorca
l'Alguer
Illes Balears
Eivissa i Formentera
el Carxe
Els Països Catalans
70.520 km²
13.534.721 hab. (2006)
100 km

Dynamical Models: Analysing the data from Valencian Country

Miralles, Clara; Perucho, Manel and Querol, Ernest (2015): “Models dinàmics de competició entre llengües aplicats al cas català-castellà al País Valencià”, *Anuari de Psicologia de la Societat Valenciana de Psicologia*, Vol. 16, n. 1, p. 31-55.

We discussed the application of Abrams and Strogatz (2003) and Mira and Paredes (2005) to empirical data obtained by official polls:

- **Service of Research and Studies** (SIES) of the Generalitat Valenciana 1989, 1992, 1995, 2005 and 2010.

- **Valencian Academy of the Language** (AVL) in 2004.

To compare two temporary series

- SIES:1989, 1992, 1995, 2005 and 2010.

- AVL only 2004.

We have transformed the data obtained in intervals of age into a temporal distribution.

To be able to apply the mathematical model to the evolution of the number of speakers we had to transfer every data dot (age, speakers' fraction) for a fixed time (year 2004)

to dots (year, speakers' fraction) for any age range in the survey.

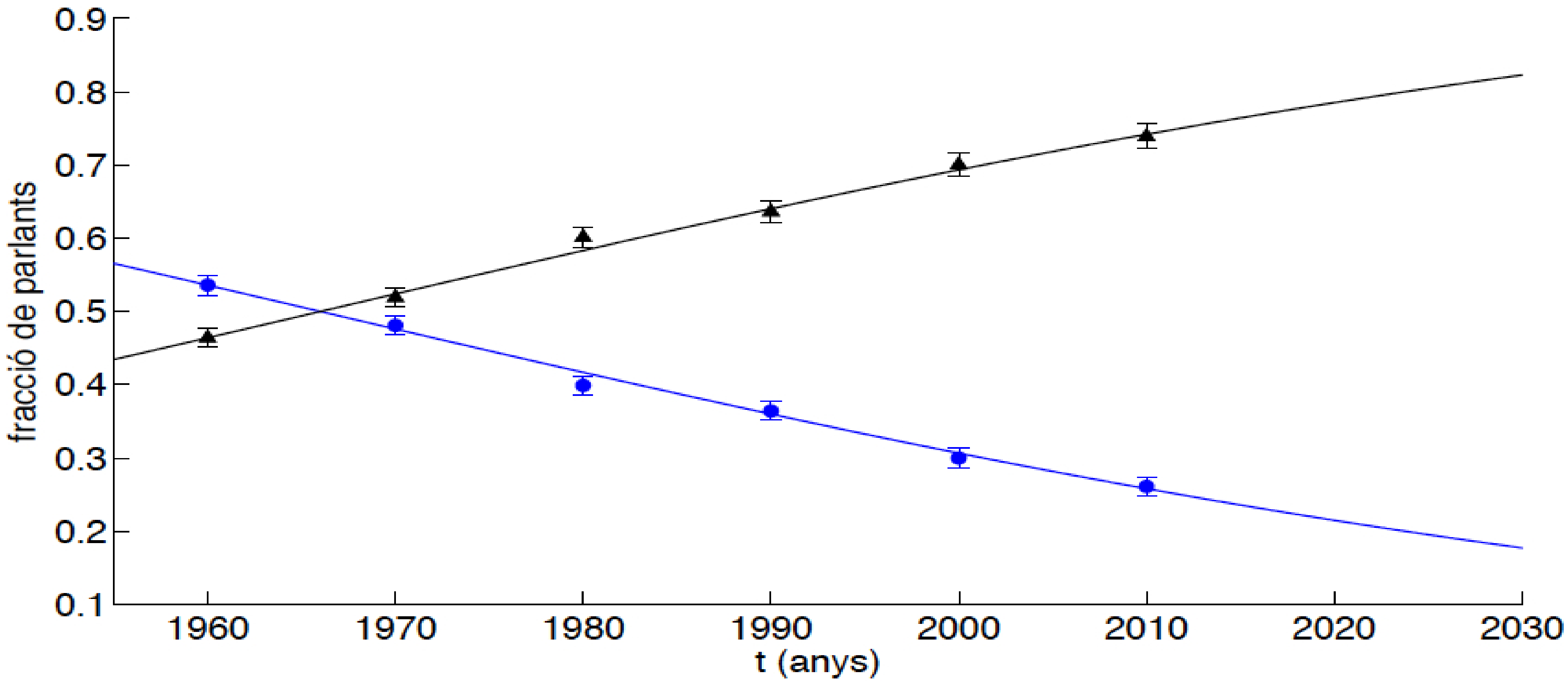
We considered that every age range represented those aged between 20 to 30.

Analysed data with Abrams and Strogatz (2003) Model

The decreasing trend observed in the percentage of Catalan speakers is confirmed.

Values were very similar in both surveys (SIES and AVL).

FIGURA 1. ***Fraccions de valencianoparlants i castellanoparlants en l'àmbit familiar respecte del temps***

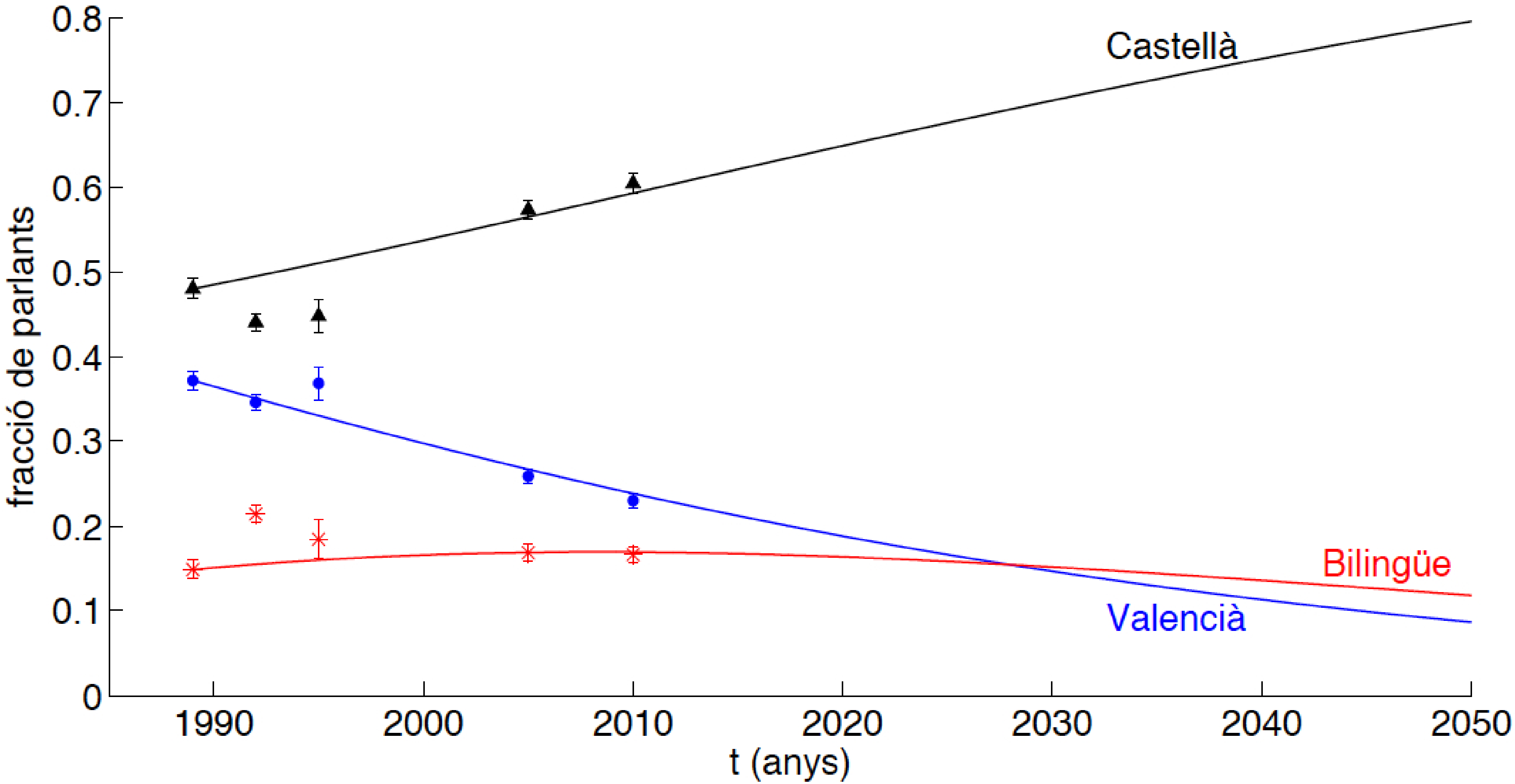


Analysed data

- Paredes i Mira: The likeness parametre between Castilian and Gallician k = 0.8

- In ours : K = 0.35 Castilian -Valencian

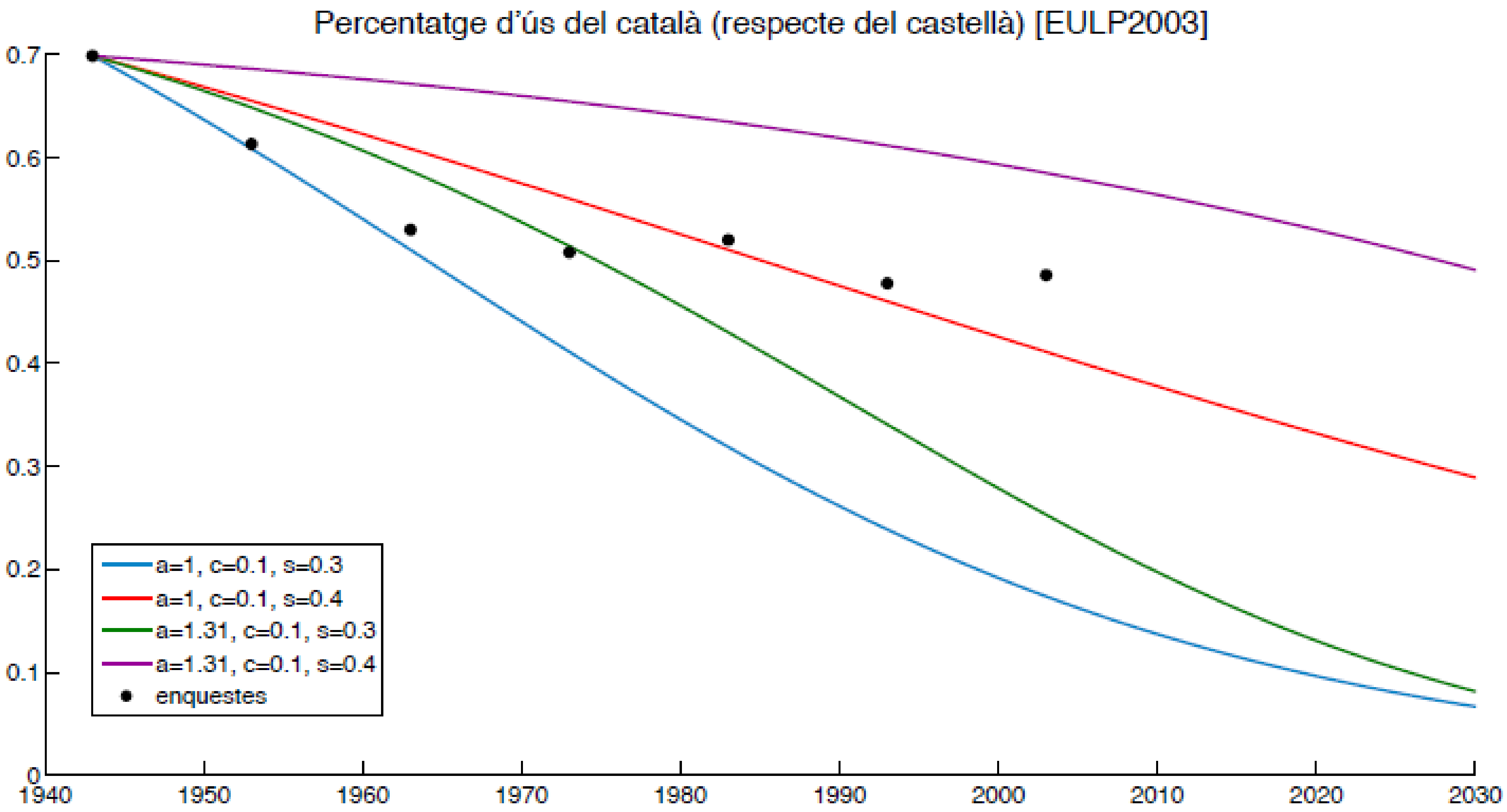
FIGURA 5. ***Fracció de parlants en relació amb el temps i classificades en tres grups: valencianoparlants (punts blaus), castellanoparlants (triangles negres) i bilingües (estreles roges)***



Font: Dades extretes de les enquestes del SIES.

2.2. Dynamical Model of Reversing Language Shift: Analysing the data from Catalonia


Andorra
Catalunya del Nord
Franja de Ponent
Catalunya
Mar Catalana
País Valencià
Mallorca
Menorca
l'Alguer
Illes Balears
Eivissa i Formentera
el Carxe
Els Països Catalans
70.520 km²
13.534.721 hab. (2006)
100 km



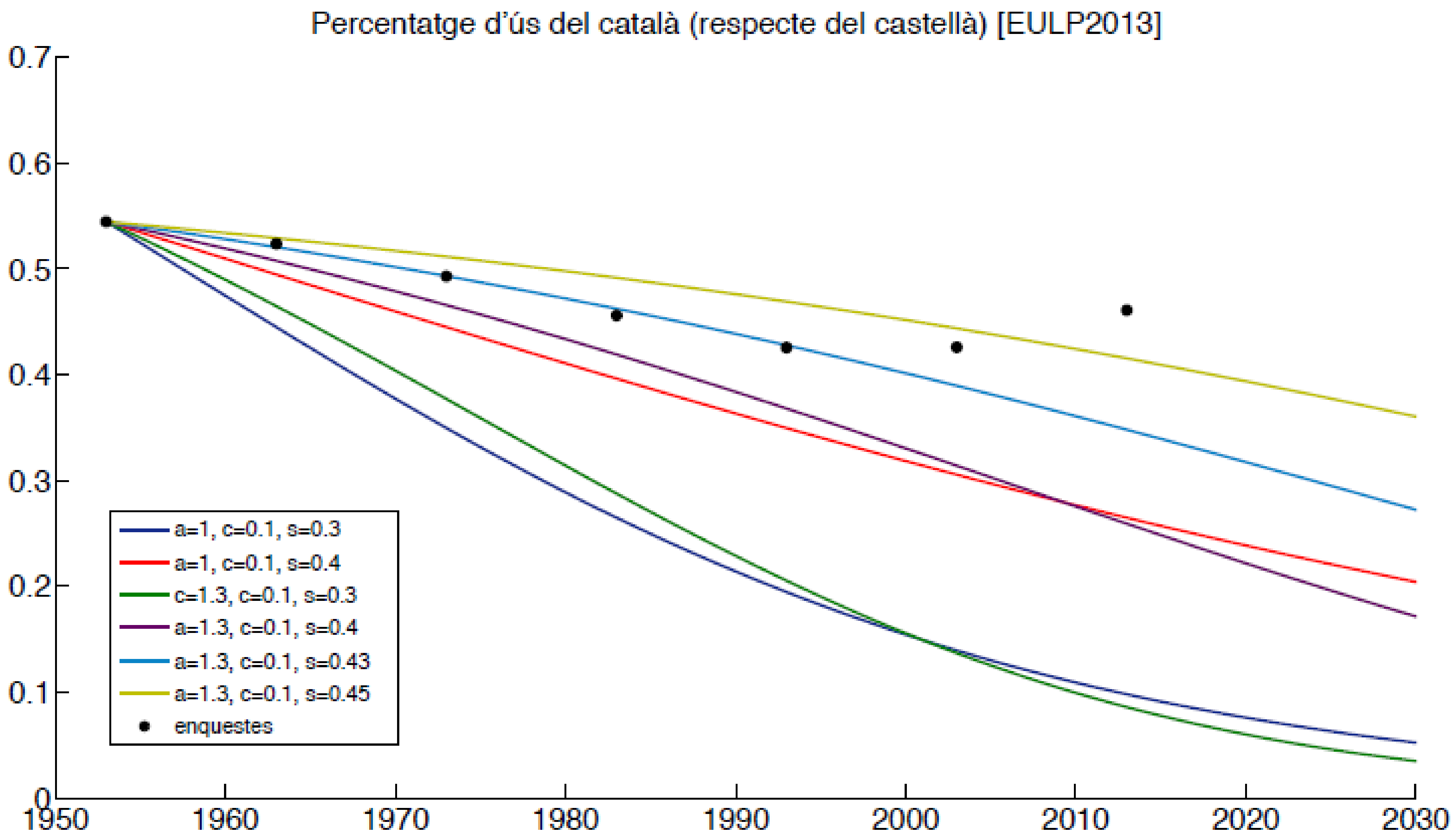
Percentatge d'ús del català (respecte del castellà) [EULP2003]
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
a=1, c=0.1, s=0.3
a=1, c=0.1, s=0.4
a=1.31, c=0.1, s=0.3
a=1.31, c=0.1, s=0.4
enquestes

Can we apply the extinction model to Catalonia?

- As the **extinction** model always moves **downwards** and the parametres make it drop dramatically without observing the modulation of the fall.

- The **failure in the adjustment** let us observe that the dymanics in Catalonia are not those in the Valencian Country.

- The model **can not reproduce a concave evolution graph** (all the adjusted curves are convex).

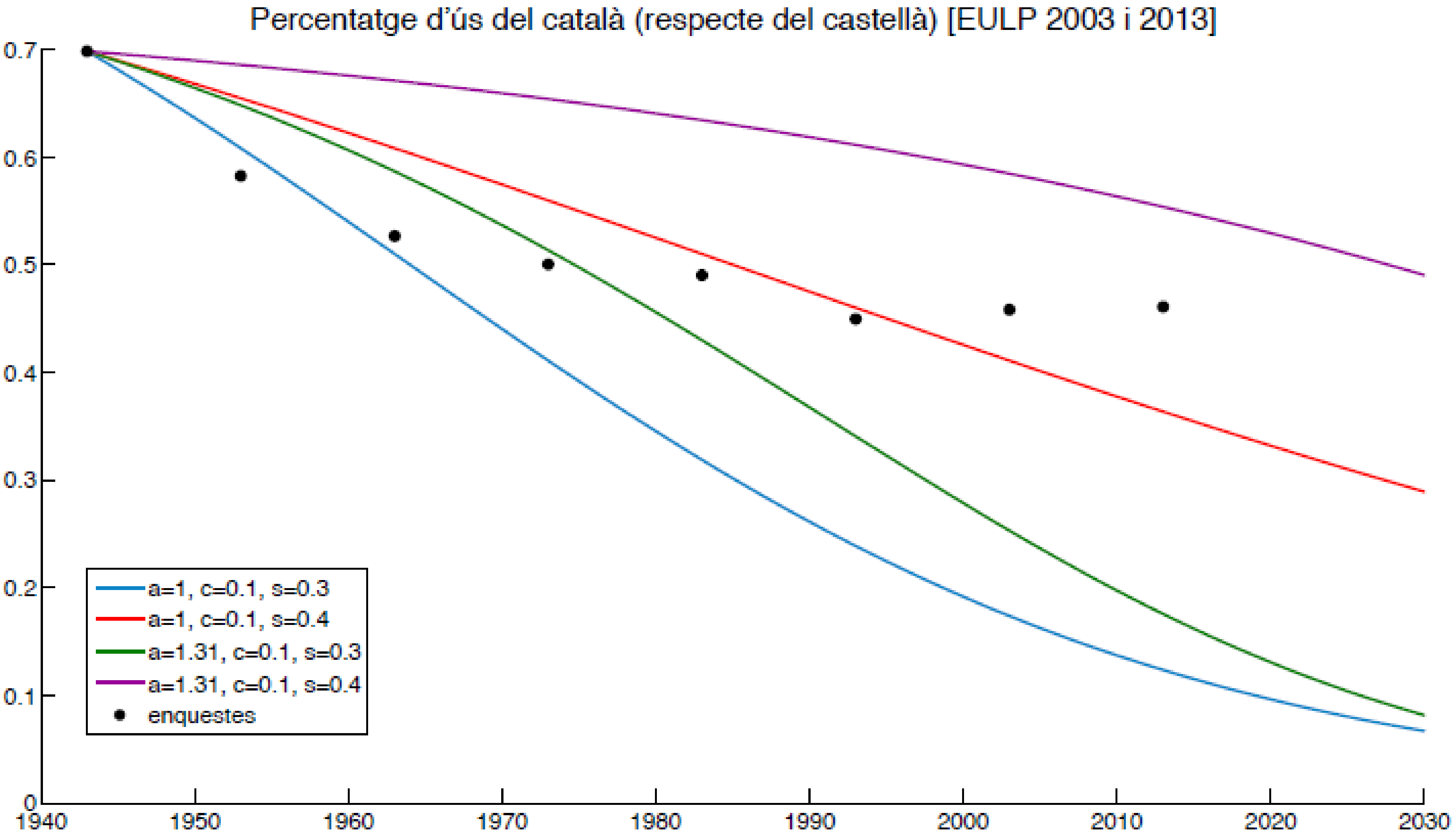

Percentatge d'ús del català (respecte del castellà) [EULP2013]
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1950
1960
1970
1980
1990
2000
2010
2020
2030
a=1, c=0.1, s=0.3
a=1, c=0.1, s=0.4
c=1.3, c=0.1, s=0.3
a=1.3, c=0.1, s=0.4
a=1.3, c=0.1, s=0.43
a=1.3, c=0.1, s=0.45
enquestes

Abrams-Strogatz model can only describe vanishing languages

It will **not be capable of modeling** the trends of recovering as the one shown in the 2013 graph.

Unless there was an inversion in the role of the languages involved in a timeline focal point, that is to say, that the change in its status switches from **$s<0.5$** to $s>0.5$.

- However, that would imply a time shift in the status parameter (a wishful fact, indeed), but the model does not allow that shift, being ***s*** **a fixed parameter**.



Percentatge d'ús del català (respecte del castellà) [EULP 2003 i 2013]
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
a=1, c=0.1, s=0.3
a=1, c=0.1, s=0.4
a=1.31, c=0.1, s=0.3
a=1.31, c=0.1, s=0.4
enquestes

3. Conclusions

Abrams-Strogatz:

dynamic model with vanishing languages

It works well with the evolution in the Valencian Country

It does not work with the description of the evolution in Catalunya

It is not a case of a vanishing language

Differential equations = Differential conclusions

- Lotka-Volterra differential equations has with time three possible solutions called *fixed*:

1) **The population of *B* disappears** and only remainds population of *A*.

2) **The population of *A* disappears** and only remainds population of *B*.

3) **An unstable fixed point** in which there is population of both spices.

Unstable = a small perturbation can make the system tend to one of the stable solutions 1 or 2

Differential equations = Differential conclusions

Linguistic conflict paradigm only considers 2 possibilities: extintion / recovery

Consider a **third** possibility: an **unstable fixed point** for 2 reasons:

1. Theorical: Lotka-Volterra equations includs it.

2. **Empirical**: Analysis of pulls of Catalonia shows this **unstable fixed point.**

From discontinuity to continuity

- **Querol (1998) Catastrophe theory model studies discontinuity**

↓

- Alcàntera Resesarch group studies continuity:

The final aim of this line of research is to **develop a mathematical modeling of the system** to describe it as precisely as possible, **determining the key parameters** that rule language dynamics.

The first challenge rests on the modelisation of the bid data in the **Catalonia** pulls (2003, 2008 i 2013) that **explain the evolution of the uses** in this country.

↓

The search for a working model which allows us to **explain the current reversing Language shift.**

4. New researchs

Further research

- Once the data form the surveys of the language uses from the Populations in **all the territories of the catalophone lands** we will be able to compare them amb the ones from 2003-04, and we will try to modelise the evolution with the rest of territories, excepting Catalunya.

Another line of research

- In our working group has been **Predicting the use of Catalan through supervised classification**, that we have processed in this paper:

Grimaldo, f., López-Iñesta, E, Perucho, M. and Querol, E. (2016): "Predicció de l'ús del català mitjançant la classificació supervisada", *Treballs de Sociolingüística Catalana*, 26. (forthcoming July 2016).

Weaving bridges between Sociolinguistics and mathematical applied methods

- To end, we would like to notice the need of a closeknit cooperation, open and constructive, between the **science/technical domains** and the **sociology of language**, so that we can enhance the working models and reflect upon the meaning of the parameters in the equations (status, likeness, etc.).

- This is, indeed, the aim of our Alcàntera research team: **the bridge with sociolinguistics**.

Thanks for your attention!