EMPIRICAL CORROBORATION OF THE CATASTROPHE THEORY MODEL IN CATALONIA (1993 AND 2000), IN THE VALENCIAN COUNTRY (1998), BALEARIC ISLANDS (2001) AND ANDORRA (2002)

Ernest Querol[1]
Institut de Sociolingüística Catalana

This poster presents data empirically establishing the correctness of our proposed catastrophe theory model. In addition to the descriptive data on language use, we needed explanatory data. For this we used statistical analysis: discriminating, multiple regression, Pearson's correlations and a new method utilising induction graphs. The results obtained are spectacular and multiply by five the results obtained using the language attitude concept. We were also able to confirm the accuracy of the predictions obtained using the proposed model, by repeating the same questionnaire survey in Catalonia after a lapse of seven years.

1. Empirical corroboration of the catastrophe theory model

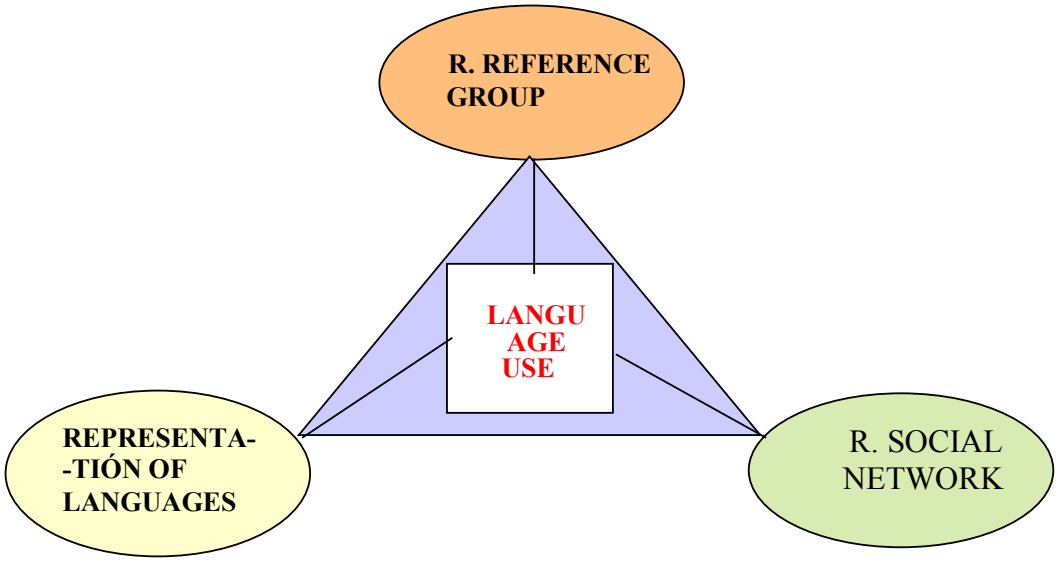
The poster that we present here forms the continuation of our paper entitled "A new theoretical model in the study of bilingual language use: Catastrophe theory" which can be read in the proceedings. In this we reached the conclusion that language use is conditioned by feedback from the following variables (see below Figure1).

Our intention is to give an overall representation of the five research projects referred to in the title, and for that we will divide the data / into the following groups:

- Descriptive
- Explanatory
- Predictive

[1] Passatge de la Banca, 1-3, 08002 Barcelona. Tel: 93 567 10 22; fax: 93 567 10 01; email: equerolp@gencat.net

Figure 1. Relationship between independent variables and the dependent variable (language use).



Source: own work

1.1. Details of the research

1.1.1. The questionnaire

To design the questionnaire we adapted the surveys that Réal Allard and Rodrigue Landry carried out to study the use of French in New Brunswick and which they very kindly made available to us. A detailed explanation of the questionnaires we used can be found in Querol (1999a, 2000).

1.1.2. Administering of the questionnaires

The research was carried out in the same way in all instances, except in the case of the Andorra research which, since it asked for the four most spoken languages, needed special treatment. Below we provide a summary of the details of the other four:

- a *representative* sample of each territory;
- students in their last year of compulsory secondary education (+/- 16);
- questionnaire with *378 questions*;
- the questionnaire was available in *Catalan* or in *Spanish*;
- two investigators were on hand to give the questionnaires: one spoke Catalan and the other Spanish.

In Andorra, data was obtained in the case of each question on Catalan, Spanish, French and Portuguese. For that reason, the number of questions included was four times less, but the total number of answers obtained was very similar. We administered the questionnaire to all students on roll for the last year of compulsory secondary in all the school systems. The questionnaire was only available in Catalan, the sole official language of the country, and it was administered only by the undersigning author of this poster (who was also involved in all the other areas mentioned) speaking the official language.

1.2. The empirical results

1.2.1. Descriptive data

The questionnaire provides a wealth of information but here we will only mention the most relevant. The bibliography lists publications which contain the results obtained by each survey. Here we provide overall information concerning sampling and reliability.

Table 2. Reliability.

	VALENCIAN COUNTRY 1998	**CATALONIA 1993**	**CATALONIA 2000**	**BALEARIC ISLANDS 2001**	**ANDORRA 2002**
questionnaires completed	452	432	443	447	416
reliable to	0.9762	0.9751	0.9768	0.9808	0.9305

It can be seen that all samples include more than 400 students and reliability exceeds 0.90, which can be considered a high degree of reliability, since anything over 0.75 is acceptable.

Secondly we present data on the languages of the students and the language they chose to answer the questionnaire.

The Table 3 shows the great differences between the territories where Catalan is spoken. The Valencian country is the area where the language is least transmitted and least used, followed in ascending order by Andorra, the Balearic Islands and Catalonia.

Table 3. Languages learned and language used to answer the questionnaire.

	VALENCIAN COUNTRY 1998	CATALONIA 1993	CATALONIA 2000	BALEARIC ISLANDS 2001	ANDORRA 2002
First language learned by the students: Catalan	25.2%	47.1%	59.5%	39.5%	35.1%
Second language learned by the students: Spanish	72.7%	52.9%	40.5%	56.7%	43.2%
Language chosen for replying: Catalan	24%	61.3%	66.2%	34.5%	All
Language chosen for replying: Spanish	76%	38.7%	33.8%	65.4%	Not possible

1.2.2. Explanatory data

To explain the variables influencing the use of Catalan we will use the more usual bivariable and multivariable analyses and also a new method of segmentation by induction graphs. We will start with analysis of the correlations.

Table 4. Correlation with the use of Catalan.

VALENCIAN COUNTRY 1998	BALEARIC ISLANDS 2001	ANDORRA 2002
- social network in Catalan .796 -representation of Spanish .768 - representation of Catalan .713 -social network in Spanish .710	-representation of Spanish ……-.826 - representation of Catalan .817 -social network in Catalan .801 -social network in Spanish -.757	-wish to use Catalan .835 -social network in Catalan .784 -representation of Catalan .753 -representation of Spanish -.515
CATALUÑA 1993	**CATALUÑA 2000**	
- representation of Spanish .846 - representation of Catalan -social network in Catalan .783 -social network in Spanish .536	-representation of Spanish .868 - social network in Spanish .854 - representation of Catalan .833 - social network in Catalan .824	

What is important to note, in the first place, is the marked improvement in the percentage of correlations compared to those obtained by the language attitudes variable which scores no more than 9%. Now we see all five variables reach .70 (more than 50%) and four of them obtain scores in excess of .82 (more than 64%, the highest reaching 75.34%). Accordingly, we can consider our initial objective accomplished. Secondly, note that in the questionnaire, in addition to the variables associated with our

model, we have introduced the main intermediate variables favoured by the sociology of the language: instrumental and integrational motivation, identity, ethnolinguistic vitality, etc. Of these, only identity obtained results similar to (if lower than) those obtained by our model. This fact shows even more clearly the ideal nature of the variables that we have adopted in our hypothesis.

We can now proceed to give a brief overview of the results. It is clear that we would need far more space than we have here, to be able to present a comprehensible and comprehensive analysis, as we did in Querol (1999a, 2000, 2001, 2002b, 2003a, 2003b, 2003c), where the reader can obtain much more information. As regards *correlations*, we observe that the two variables obtaining highest values are the representation of the languages and the social networks. What changes is the relative importance of each of these. In Catalonia and the Balearic Islands we can see that the representation of the language is the most important variable, while in the Valencian Country, the social network in each of the languages occupies first place and fourth place. In Andorra, where the questionnaire was different, the variable that best correlated with the use of Catalan was the desire expressed to use that language.

We now consider the extent to which discriminatory analysis helps us to explain the use of Catalan. Firstly, here are the results summarised in table form (Table 5).

Table 5. Variables that together constitute the discriminating function with respect to the use of Catalan.

	VALENCIAN COUNTRY 1998	**BALEARIC ISLANDS 2001**	**ANDORRA 2002**
Discriminating analysis **Function in terms of Catalan use.**	- social network in Catalan -social network in Spanish -group belonged to	-representation of Spanish -representation of Catalan social network in Spanish -social network in Spanish	-wish to use Catalan -social network in Catalan - representation of Spanish -social network in Spanish
-percentage correctly classified	64,4%	60%	71.8%

	CATALONIA 1993	**CATALUÑA 2000**
Discriminating analysis **Function in terms of Catalan use.**	-representation of Spanish -social network in Catalan -identity **-social network in Spanish** - representation of Catalan	-representation of Spanish **-social network in Spanish** -social network in Catalan -identity
- percentage correctly classified	83.8%	72.3%

Here once again, with respect to the variables involved in the *discriminating analysis*, we see that social network emerges as most important in the Valencian Country, while in Catalonia and the Balearic Islands it is the representation of Spanish that carries most weight. This analysis will assist us in predicting which language group a speaker belongs to. Thus, in terms of correct clssification of students, in Catalonia we obtained an impressive 83.8% rate of correct prediction, while in the Valencian country the rate was 64.4% and in the Balearic Islands it was 60%.

We shall use one other model of analysis, multiple regression analysis (Table 6).

Table 6. Predictors of use of Catalan according to multiple regression analysis.

VALENCIAN COUNTRY 1998	BALEARIC ISLANDS 2001	ANDORRA 200 2
- social network in Catalan -social network in Spanish -representation of Spanish	-**representation of Spanish** -**social network in Catalan** -social n**etwork in Spanish** -reference group	Since questions were asked about the four most spoken languages this analysis will not produce significant results

CATALONIA 1993	CATALONIA 2000
-representation of Spanish -social network in Catalan -identity -parents' language	-representation of Spanish -social network in Catalan -social network in Spanish -identity

Multiple regression analysis enables us to specify which of the independent variables determines the variation in the use of Catalan. Once again it indicates the importance of social networks for the use of Catalan in the Valencian Country: they occupy the first two places in the ordering of factors. In the case of Catalonia and the Balearic Islands, the two foremost factors are the representation of Spanish and the two social networks. We can see that apart from reference group (not a numerical variable), all the variables we proposed in the hypothesis occupy the initial positions in terms of importance in the analysis.

Indeed, the data in Tables 4, 5 and 6 confirm our proposed model for studying the processes of language shift.

We have been able to corroborate, in the four different contexts, all the hypotheses we had formulated (presented schematically in Figure 1). Furthermore, we have achieved the objective we had set of arriving at a unified solution for what we saw as the principal problems in sociology of language, and we have done this both formally

(establishing isomorphism with butterfly catastrophe) and empirically, enabling us to obtain quintuple corroboration.

1.2.2.1. Induction graph techniques (SIPINA)

We decided to include an additional method of explanatory analysis –knowledge from induction graphs– which has recently become available. This will allow us to explain a qualitative variable (the use of Catalan, in this case) by means of explanatory variables. The latter are themselves qualitative, but take discrete or segmented form. In this method a predictive function is presented in the form of a graph and decision tree, enabling us to explain or foresee the value taken by a particular *endogenous* variable as a function of a series of *exogenous* variables.

The programme we use was developed by D.A. Zighed and Ricco Rakotomalala, members of ERIC (Equipe de Recherche en Ingénierie des Connaissances) of the Université Lumière Lyon 2. This involves "knowledge engineering", which deals with problems relating to the extraction, modelling and transference of knowledge via computer with the aim of achieving so-called intelligent machines. The latter are machines able to assist human beings in carrying out certain complex cognitive tasks such as recognising, decision taking, and so on.

This programme is called SIPINA, (Interactive System for Non-tree Processes of Interrogation). It extracts knowledge from data and produces decision graphs. The latter are a generalisation of tree decision diagrams making it possible to handle problems of explanation. Such problem are especially prevalent in the human sciences (sociology, biology, psychology, medicine, etc.). SIPINA has been used, for example, to facilitate diagnosis, in the measuring of risks and in establishing profiles typical of certain population. To obtain more information on induction graphs and on this programme, readers may consult Rakotomalala (1997) and Rakotomalala & Zighed (2000). We will use this to explain the four linguistic groups we have established:

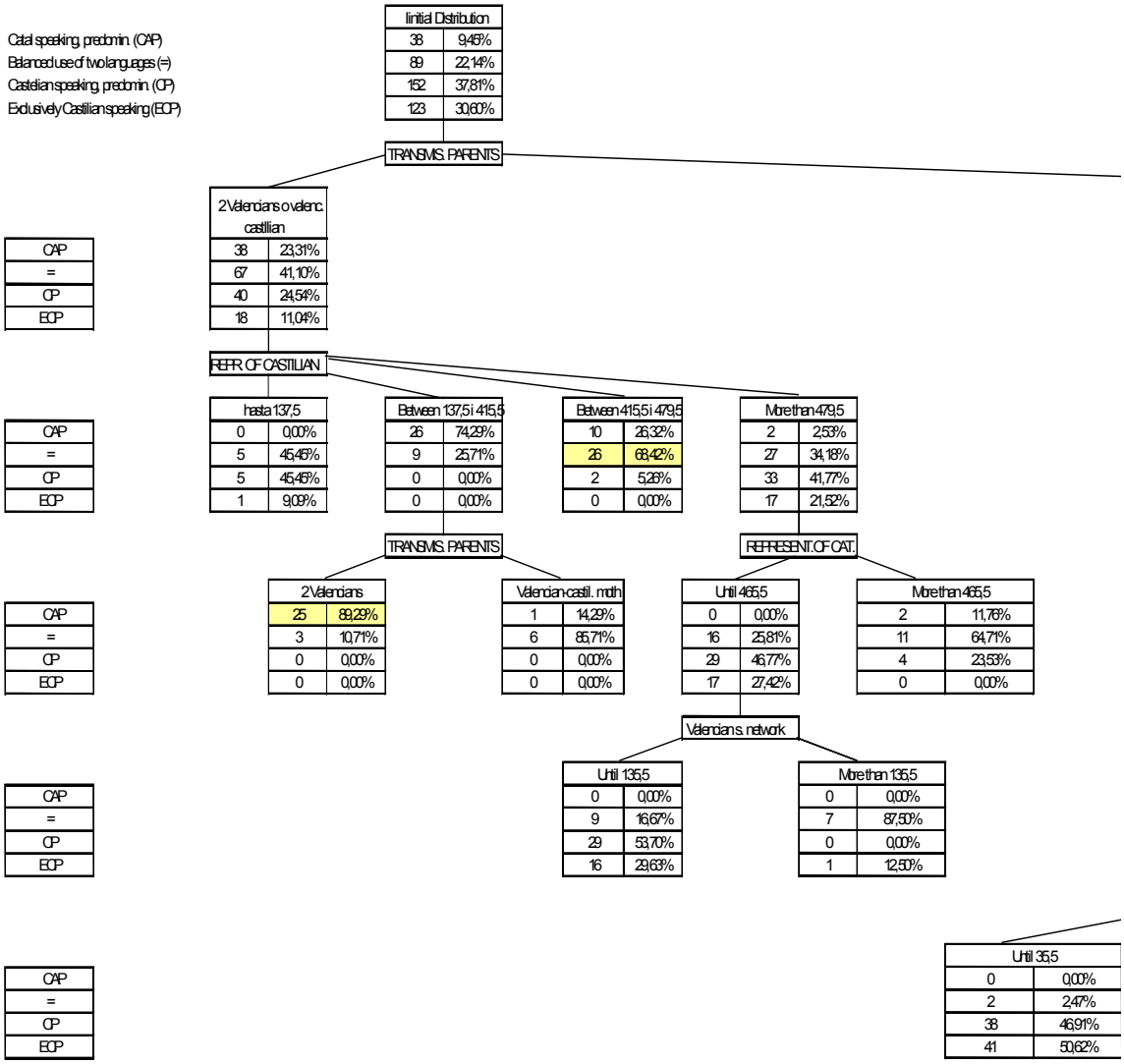
- Catalan speaking, predominantly (CP),
- Balanced use of the two languages (=),
- Castilian-speaking, predominantly (CP),
- Exclusively Castilian-speaking (ECP).

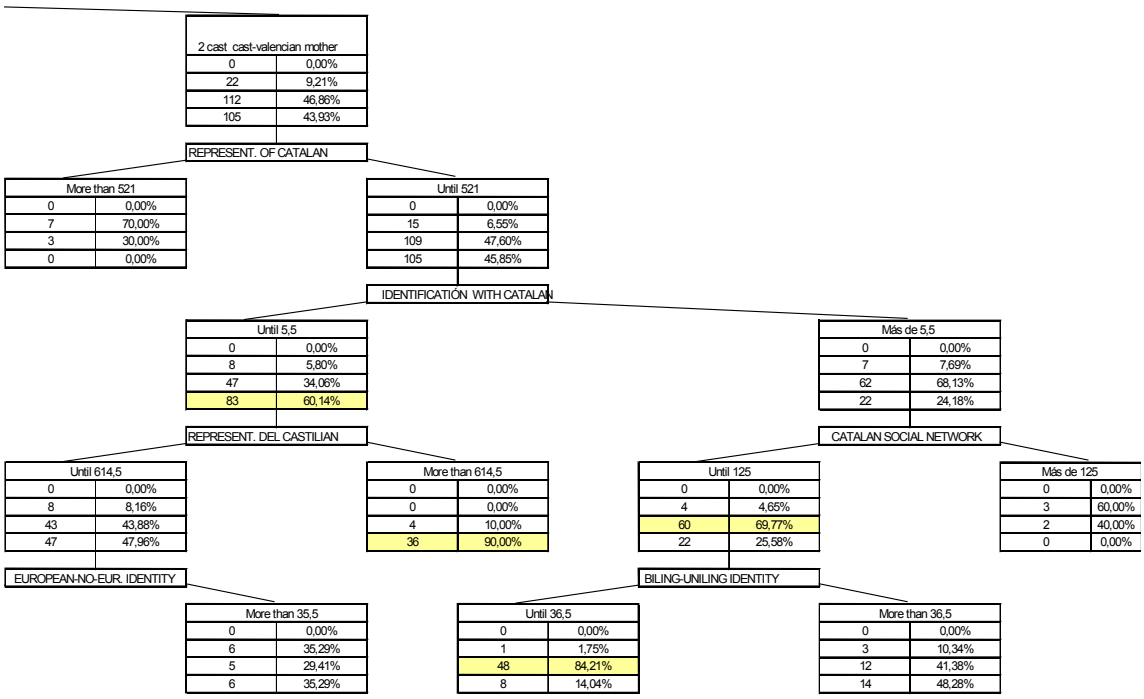
The objective here was to place a student in one of these groups without actually referring to the information that determines the groups10. In fact, the process is similar to discriminate analysis. Firstly, we see in table 12 the overall results provided by this programme, being the solution of the game we described at the outset of our presentation of the theoretical model (see Table 7 below).

This diagram, then, provides us with the solution to table 1 in the communication, from which we had removed those variables that have an effect on the use of the languages. Here we can show in a clear and easily apprehended way that the final objective of our investigations is to determine the main variables influencing or tending to determine language use. It shows us which are the segments that more readily allow us to identify the different groups. The first of these is "transmission from the parents", that is to say, whether the parents spoke to the students in one language or another. This in itself is hardly surprising; on the contrary, it is an observation that anyone can make. And of course it is connected with the importance of social network in the case of Catalan, an aspect which we have already noted in the previous analyses, above.

The second set of factors that discriminate between the two main groups, those in the first part of the table and those below (because there was no room to display all of this information together, horizontally) is the representation of the language which is not the learned one. The third set of factors which divide up the subgroups here are once again the language transmitted by the parents and the representation that the students construct of Catalan, within the group with Spanish speaking parents (or with a Catalan-speaking mother) where we find the variable "identification with Valencians". On the fourth level of importance there is social network in Catalan (in two subgroups) and the representation constructed of Spanish. If we draw even finer distinctions among those with Spanish-speaking parents (or a Catalan-speaking mother) we find the question of whether the students feel European or not, and whether they feel themselves to be bilingual or unilingual.

Table 7. Segmentation of the sample by linguistic group.





Thus, with last two of these variables and the transmission of the languages carried out by the parents, the rest of the variables that most effectively isolate or determine the two linguistic groups are those that we had proposed in our model. This, then, is another means of confirming our hypothesis, this time using induction graphs.

Thus, this new knowledge technique enables us to capture the characteristics of each group. We can see this more clearly in Table 8.

Table 8. Description of the 4 linguistic groups using induction graphs.

More Catalan-speaking	Students with: -Both **parents** Catalan-speaking, and -Scoring between 137.5 and 415.5 in the **representation of Spanish scale.**
Balanced use	Students with: -Both **parents** Catalan-speaking or with the father Catalan-speaking and the mother Spanish-speaking, and -With a score of between 415.5 and 479.5 in the **representation of Spanish scale.**
More Spanish-speaking	Students with: -**Less** than 521 on the **representation of Spanish scale.** -More than 5.5 on the **Identification with Catalan speakers scale**, and -**Less** than 125 on the social **network in Catalan** Index.
	An even more extreme segment can be observed, if, apart from the characteristics mentioned here, the students evince a score of: -Up to 36.5 on the **Bilingual – Unilingual** identity scale.

Exclusively Spanish-speaking	Students with a score: -**Less** than 521 on the **representation of Catalan** scale, and -**Less** than 5.5 on the **Identification with Valencian speakers** scale.
	An even more extreme subsegment can be seen, if, apart from the above-mentioned conditions, the students evince a score of: -**More** than 614.5 on the **representation of Spanish scale.**

1.2.3. Predictive data

In the analysis of the processes of language shift, as in all social disciplines, it is very important not only to register the situation as it currently presents itself but to be able to predict what will happen in the future. The model that we have sketched in this communication, apart from obtaining very significant correlations between the three variables we have just looked at (social network, by language and by reference group) has enabled us to classify speakers, by means of discriminate analysis, in the four linguistic groups based on degree of use of the languages with a success rate of 83,8%. But, more than that, it has enabled us to make predictions regarding future trends. And in fact we have been able to verify that predictions were fulfilled perfectly (that is, the predicted trend occurred) over a seven year period. Let us look at this at greater length.

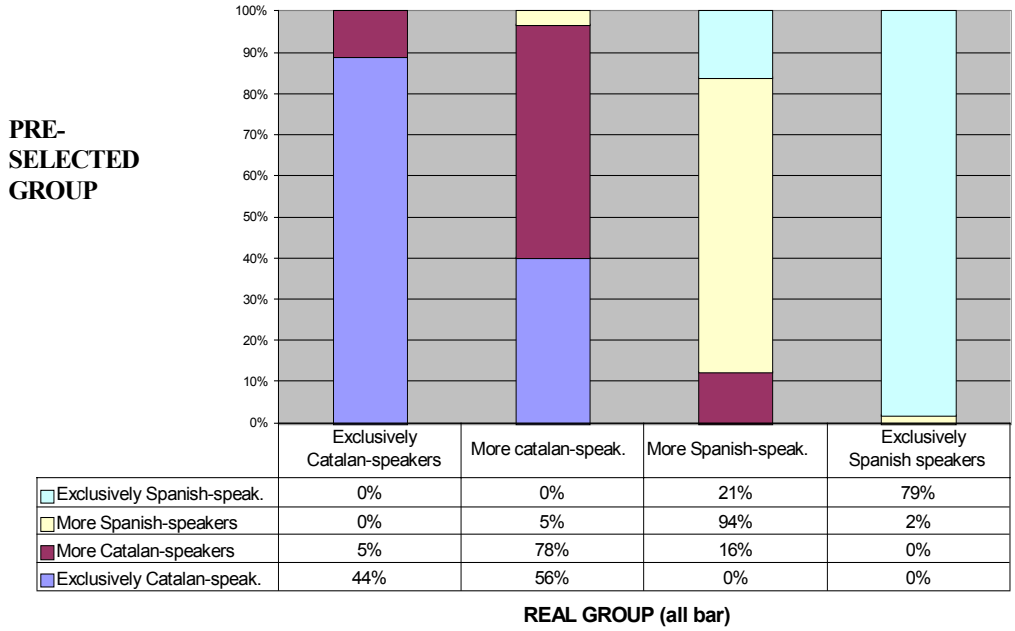
1.2.3.1. Forecasts on pertaining to a particular linguistic group

The model that we have presented here allows us to predict which group a student belongs to, without referring to information that would directly indicate such a classification[2]. With the 1993 Catalonia data it succeeded in *predicting* correctly in 83,8% of cases; in no case was there an erroneous placement in a group other than one which was contiguous to the correct one. We can see this in Table 9.

The real group is the bar in a single colour. For example, in Table 9, the exclusively Spanish-speaking group is correctly delimited or determined in 79% of cases, only 2% of the "more Spanish-speaking" students are included in the narrow strip at the base of the bar. However, 21% are wrongly placed, indicated in the bar representing the "more Spanish-speaking" students, which is the group contiguous to it.

[2] We also excluded from the analysis questions asking if the students felt themselves to be Catalan-speakers or Spanish-speakers (self-categorisation) in order to achieve a perspective that was not influenced by questions so directly related to the variable under study.

Table 9. Level of successful prediction (83.8%) by the model (1993).



	Exclusively Catalan-speakers	More catalan-speak.	More Spanish-speak.	Exclusively Spanish speakers
Exclusively Spanish-speak.	0%	0%	21%	79%
More Spanish-speakers	0%	5%	94%	2%
More Catalan-speakers	5%	78%	16%	0%
Exclusively Catalan-speak.	44%	56%	0%	0%

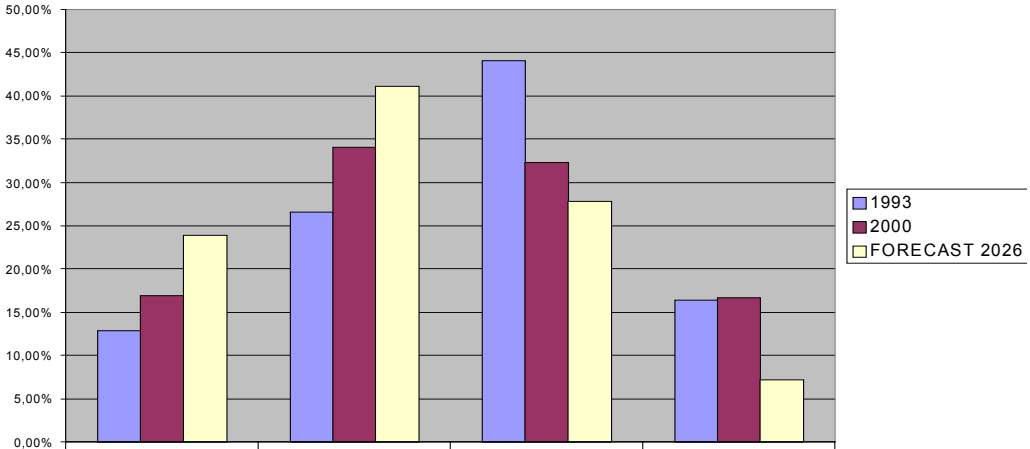
1.2.3.2. Confirmation of predictions made in 1993

The *reference group* (or desired language group), which we situated in the upper vertex of the triangle in our model, plays a decisive role in the attraction of speakers to one group or another (recall that we spoke of *attractors,* after catastrophe theory). Accordingly, the reference group will play a fundamental predictive role in fulfilment of the forecasts made in 1993. We thus proposed this variable as a factor in predicting trends in the composition of the established linguistic groups. We stated that the desired group in 1993 could provide the actual rate of use one generation (33 years) later. We can see this in the next table (see below Table 10).

With the 1993 data we determined the existence of the four linguistic groups (which is the first horizontal line: 12.90%, 26.60%, 44.10% and 16.40%) and the groups to which the students would like to belong (which is the third horizontal line of the same table: 23.90%, 41.10%, 27.80% and 7.20%). With the data we have just harvested (second horizontal line) we can see that the groups confirm the prediction we had made in the course of the 1993 research: the two groups of Catalan speakers increase, and there is a decrease in those who most speak Spanish. The group of exclusively Spanish

speakers, however, goes in the opposite of the expected direction and increases by 0.30%. The latter fact is not very significant, however, since the rest of the changes are of the order of 4%, 7.50% and 11.80%, all in the direction we had predicted.

Table 10. Confirmation of predictions.



	Exclusively Catalan-speakers	More Catalan-speakers	More Spanish-speakers	Exclusively Spanish-speakers
1993	12,90%	26,60%	44,10%	16,40%
2000	16,90%	34,10%	32,30%	16,70%
FORECAST 2026	23,90%	41,10%	27,80%	7,20%

The data obtained 2000, then, actually induces us to shorten the period of 33 years that we had settled upon, since 36,36%, 51,72% and 71,08% of the total increases we had predicted over that span of time had already come about in only seven years. The exception was the exclusively Spanish-speaking group, which had gone very slightly in the opposite direction to that predicted, as we have just noted.

1.3. Conclusions

We have shown that the theoretical and methodological model that we put forward is not only able to describe the process of language shift but also to explain it and make prediction regarding its future development. What is more, after the seven years that have passed between the first and second questionnaire surveys undertaken in Catalonia, we can now show the correctness of these predictions. We have detected a reversing of language shift, evident from objective data, and are able to make predictions firmly based on the latter for the immediate future.

We have confirmed all the hypotheses we put forward, in the four main territories in which Catalan is spoken. What is more, we have fulfilled the objective we had set ourselves of arriving at a unified solution for those problems which, as we see it, are the main problems facing the sociology of the language. And we have done it both in formal terms (establishing the isomorphism that there exists with butterfly catastrophe) and empirically, which, as we have just seen, has enabled us to obtain quintuple corroboration of the model.

Bibliographical references

Doise, W. & A. Palmonari (1986). *L'étude des représentations sociales.* Neuchâtel-Paris: Delachaux et Niestlé.

Jodelet, D. (1989). "Représentations sociales: un domaine en expansion". In D. Jodelet, *Les représentations sociales*. Paris: Presses Universitaires de France.

Merton, R.K. & S. Kitt (1950). "Contributions to the theory of reference group behavior". In R.K. Merton & P.F. Lazarsfeld, *Continuities in social resarch: Studies in the scope and method of the American soldier*. Glencoe: Ill., Free Press.

Moscovici, S. (1984). *Psychologie sociale*. Paris: Presses Universitaires de France.

Moscovici, S. (1961) *La psychanalyse, son image et son public*. Paris: Presses Universitaires de France.

Querol, E. (1997). "Un nou model per a l'estudi del comportament lingüístic: la teoria de les catàstrofes". *Caplletra* 21, 161-184.

Querol, E. (1999a). *Cap a un nou marc teòric per a l'estudi de les variables dels processos de substitució lingüística*. Barcelona: Publicacions Universitat de Barcelona, doctoral thesis on microfiche, 3568.

Querol, E. (1999b). "A new theoretical approach to the study of reversing language shift: Catastrophe theory". Paper presented at the *7th International Conference on Minority Languages*. Bilbao (in press).

Querol, E. (2000). *Els valencians i el valencià: usos i representacions socials*. Paiporta: Editorial Denes.

Querol, E. (2001). "Evolució dels usos i de les representacions socials de les llengües a Catalunya (1993-2000)". *Noves SL* 2, Summer.

Querol, E. (2002a). "A new model for the evaluation of language planning. A case Study: Catalonia (1993-2000)". *Sociolingüística* 16, 129-42.

Querol, E. (2002b). "Usos i representacions socials de les llengües a les Balears". *Noves SL* 6, Summer.

Querol, E. (2003a). *Els Catalans i el català: usos i representacions socials* (1993-2000). Barcelona: Editorial Empúries (in press).

Querol, E. (2003b). *Els balears i el català: usos i representacions socials*. Palma de Mallorca: Editorial Documenta Balear (in press).

Querol, E. (2003c). *Els andorrans i el català: usos i representacions socials*. Andorra la Vella: Govern Andorrà (in press).

Rakatomala, R. (1997). *Graphes d'induction*.Lyon. PhD dissertation, Université Claude Bernard-Lyon I.

Rakatomala, R. & D. Zigued (2000). *Les graphes d'induction.* Éd. Hermes.