Research project - "Socio-geolinguistic atlas of the Basque language"

Gotzon Aurrekoetxea, Jose Luis Ormaetxea

University of the Basque Country
gotzon.aurrekoetxea@ehu.es

Abstract

This paper presents a new research project, linking geolinguistic and sociolinguistic aspects by investigating language differences between older and younger speakers in the southern Basque Country. To this end data will be collated using a questionnaire covering phonological, morphological, syntactic and lexical characteristics, and questions will be asked of the two groups in 75 locations.

Keywords: dialectology, sociolinguistics, geolinguistics

1. Introduction

The aim of this Basque socio-geolinguistic atlas project is to carry out a dual examination of Basque speech – a sociolinguistic examination, and also a geolinguistic survey.

The research project therefore examines the two disciplines of dialectology, understood in a broad sense: the geolinguistic point of view as the study of language differences in accordance with space, and the sociolinguistic point of view as a study of the differences each speech model has from the point of view of sociolinguistic factors – the differences between each kind of speech and other forms of speech, and the differences found within each type.

There is a single objective of research in relation to these two points of view: to study and measure the effect of the standard language during its development over the last 40 years on normal speech patterns.

We wish to extend our research to the entire Basque-speaking area – we will examine forms of speech in the southern Basque Country, and also in the northern Basque Country.

The main details of the research project are set out below, although neither the data nor the results of research will be shown here. The project began in 2006, and we expect the results to become available for publishing in 2007-2008.

2. Objective and preliminary hypothesis

Since Basque began to be utilised in urban communications and thus since work began on its standard register, new relationships have appeared between traditional and new forms of speech.

The traditional forms of speech which took shape in Basque dialects or literary Basque dialects have had different models throughout the history of Basque, and each model evolved down through the centuries quite freely due to a lack of a unifying force. In this way the dialects developed quite differently and with more or less of their own characteristics. The lack of relations between those with different forms of speech has not been a hindrance – on the contrary, this was a positive feature since it helped each model take its own direction.

This being the case, traditional forms of speech in Basque have shown considerable dialectalisation, and this situation provides an opportunity to examine the geolinguistic variability of the language.

However, along with the creation of *batua* (unified Basque), unifying forces have appeared throughout the dialects, encouraged by the power and attractiveness of the cultured prestigious variety.

These unifying forces are coming on strong, if the small-scale and partial surveys conducted to date are anything to go by.

Since the beginnings of *batua* as could be pinpointed at the Arantzazu Congress almost 40 years ago, it is time to examine and measure the effect of the standard variety within original or traditional dialects, and also examine the state of health of traditional ways of speaking.

The objective of this research project is therefore to examine the language's propensity to change. In this way, we will study the effect of two important factors: age and the geographic factor. In doing so, we intend to examine the language situation of the speakers of two different generations – what we want to examine is how the Basque speakers from the two different generations deal with the language on an everyday basis. The two generations to be studied here are 40-50 years old and 20-25 years old.

Those belonging to the first generation have had a considerable relationship with the standard model, on the one hand because they experienced the revival of Basque first-hand, and on the other because they were subsequently in close contact with the speech used at school in their children's education.

The second generation was brought up from birth totally immersed in standard Basque. They had to

merge the language used by their parents and grandparents and the model used at school in order to get on with their lives and relate to one another.

The partial surveys conducted to date have shown that *batua* has had a substantial effect on traditional dialects and forms of speech.

Thus, the objective of this research project is to confirm this hypothesis, i.e. that the effect of *batua* has been greater than was originally thought, and that the speed at which the similarity phenomenon has been produced is considerable. The intention is also to measure the speed of the effect of standard Basque on dialects, and determine which area of language has been the most influential.

3. Background and the current situation

There has been considerable development in the study of diatopic varieties in relation to Basque: this started with Prince Bonaparte's investigation of the subject, and many and varied projects have been carried out prior to the Basque Popular Speech Atlas project (EHHA) on which the *Euskaltzaindia* Basque Language Academy is working.

The EHHA is, without a doubt, the deepest and most prolonged study made of different forms of speech in Basque. The academics carrying out this project, Gotzon Aurrekoetxea and Xarles Videgain, have used the normal methodology of language atlases in their collation of data, and set out the progress made in the form of data cartography.

The EHHA covers the entire Basque-speaking area in 145 survey locations on both sides of the Pyrenees.

From the diastratic point of view, however, next to nothing has been studied in relation to Basque. To date, very few surveys have been conducted in relation to the similarity phenomena between different forms of speech or between the standard variety and dialects, or on the effect of standard Basque on traditional ways of speaking. Consultation of the known bibliography on Basque turns up only the work of two researchers: Gotzon Aurrekoetxea and Sonia González.

Aurrekoetxea has carried out the widest research in this area. He worked with different areas and age groups, and started work in this direction at the Euskaltzaindia's 14th Congress in Bilbao in 1999 when he presented "Unifying trends between standard Basque and dialects (from the theoretical point of view)". He then continued to develop this line of research to publish the following: "Bizkaian Basque speakers and standard Basque" (2000), "Basque dialects under a deluge of standard Basque (the case of Arratia)" (2003), and "Unifying trends between standard language and dialects".

Sonia González (2001) has also studied the effect of standard Basque on children ("Variety and standard: the

effect of *batua* Basque on the speech of two children in Bizkaia").

From the sociolinguistic point of view or, to put it another way, taking into account age, social status, studies and different sociolinguistic factors, no prolonged and developed research has been published to date on Basque beyond the items mentioned above.

In relation to the languages of the surroundings, however, research and publications setting out this perspective have become increasingly common over the last 10 years (Auer 1998, Auer & Hinskens 1996, Bellman 1998).

The basic line of research is what is proposed in this project if the intention is to conduct an in-depth survey into the changeability of the language. Knowledge of language changeability factors and parameters is becoming increasingly essential in research into the changes occurring within languages.

4. Theoretical framework of the project

The socio geolinguistic Basque atlas project is based on sociolinguistic and geolinguistic parameters. It is situated between two disciplines, it takes up the methodology of both disciplines and makes its contribution to both.

On the subject of diatopic variations, knowledge of normal Euskara in the territories of the southern Basque Country will be covered in an initial phase. The northern Basque Country will be studied in a second phase, and this will set out Basque's geolinguistic characteristics. The intention is to obtain Basque in a rural atmosphere at the locations chosen. The subject of research will be the language used by Basque speakers who have not lost their ties to tradition and who, although they were born into and grew up within a farmhouse atmosphere, now lead a city life. In other words, traditional culture, the traditional way of speaking a language, and the culture as urbanised, injected with the daily life of the work place, and consequently those with knowledge of new forms of speech.

Among diastratic varieties, the age factor will be the variable used. Two age groups will be studied: mature adults (45-50 years old) and younger people (20-25 years old). In both age groups, the informants will be men. No account will be taken of origins or social and economic factors.

In relation to diaphasic variations, a circle of friends and acquaintances forms the scenario for the information we wish to gather. Although the informants will be told that the intention is to study the register in which they talk to their family and friends, using a formal investigation methodology the research format itself will condition the result to a greater or lesser extent, and even though what we want to do is record colloquial language, we feel there is always the risk of recording a more formal register.

Taking these three varieties into account, different areas of the language will be covered in the investigation: phonology, morphology, syntax and lexicon. The aim is to investigate the language as a whole, in order to study its major characteristics. It is judged essential to examine the language in its entirety, but data will be collected without mixing up the different areas.

Thus, as we approach the results of research and as we discover phonological changeability, we will find out what things are changing and how they are changing in terms of morphology, syntax and lexicon. Similarly, we will observe which linguistic field shows the greatest propensity to change, and whether or not this propensity to change is balanced.

We will be able to make associations between language fields and between forms of speech in a different manner, examining qualitative data and also using quantitative data.

The qualitative examinations of language data will show the influence of different characteristics – which are more resistant in relation to standard Basque, or which characteristics of standard Basque completely enter the forms of speech first or most quickly.

Statistics programmes will be used to examine the quantitative data. In this way, within the different fields of language, we will be able to measure which shows the greatest tendency towards change, which shows greatest development, and which fields most characteristics of standard Basque have entered.

In relation to forms of speech, there will also be an opportunity to make a division between them, and we can measure which has taken in most alien characteristics and which has best retained its linguistic features.

Our research into diastratic changeability will allow us to gauge how the differences between the two generations materialise, and the results will give us the chance to demonstrate several clear parameters.

Lastly, we will also be able to gauge the power of the standard variety, and ascertain whether the trend of similarity is lively and solid or ongoing.

5. Research project

The main topic in the line of research set out in this project is changeability. This line is extremely important, firstly because it can pinpoint how similar trends between different forms of speech are occurring, because we can measure the strength or speed of similarity, and lastly because it will provide us with tools to set out future projections.

There can be no doubt that the results of research will supply us with resources to secure appropriate tools to plan the language corpus.

The forerunner to this project is the EHHA atlas. Starting with what the atlas shows, the aim is to record the language of the younger generation. The EHHA methodology will be applied (in construction of the questionnaire, collation of data, choice of survey locations, profile of informants etc.), but the feature distinguishing this project from the EHHA is its examination of geolinguistic changeability in addition to sociolinguistic changeability.

5.1. Objectives

The following are the objectives we wish to achieve:

1. To conduct an in-depth survey into the geographic varieties of the language.

2. To conduct an in-depth survey into the sociolinguistic varieties of the language.

3. To conduct an in-depth survey into factors relating to levelling or similarity between varieties of the language.

4. To create an oral speech dialectal database.

5. To create a database of the different kinds of speech as transcribed.

5.2. Methodology

5.2.1. Questionnaire

The survey will be a closed questionnaire, with some 200 linguistic variables. Questions will be chosen as the most appropriate for this project on the basis of the EHHA Atlas. The questionnaire will include lexical and grammatical characteristics, broken down as follows:

- Noun morphology: 23 questions

- Verb morphology: 39 questions

- Syntax: 20 questions

- Lexicon: 120 questions

Phonological characteristics will be obtained by examining the data obtained from the answers to questions on other fields, and so there is no need to provide a special questionnaire for phonology.

In relation to grammar, the questionnaire has been established with the aim of setting out the major features in each field, examining the essence of the language in each field.

With regard to lexicon, concepts which have not been lost and show extensive change on the journey from traditional culture to urban life will be selected.

5.2.2. Survey locations

The questionnaire will be applied to 100 locations in the Basque Country (in the southern Basque Country in Araba, Bizkaia, Gipuzkoa and Nafarroa, and in the

northern Basque Country in Lapurdi, Nafarroa Beherea and Zuberoa). As to choice of locations, priority will be given to those set out in the EHHA Atlas project, and the locations will be selected from these.

5.2.3. Informants

There will be two informants in each location, one older informant and one younger informant. The informants will be chosen in accordance with the following characteristics:

- Sex: all the informants will be male. It has been decided to continue the informants' profile as used in the EHHA Atlas project, where these were mostly male. The intention of research is to obtain homogeneity from the point of view of gender, and collation of female data has been reserved for a future project.

- Age: we wish to obtain full data for three generations. The EHHA Atlas project has already collected data from people over 70 years old. In this research project, data will be gathered from the generation composed of their children (who are now between 45 and 50 years old) and grandchildren (now between 20 and 25 years old).

- Education: in this kind of investigation, one important feature is the education or cultural level of the informants.

Since this investigation involves the education of informants in different age groups, the intention is to secure a broad average level for each age group. The EHHA Atlas project has already taken a step in this direction, since most of the subjects lived on farms and had had little education. The intention of this investigation is also, for each age group, to use people with a cultural level which is majority for that age group:

- In relation to the education of older people, our working concept is low to medium – those who were born on country farms and now work in a more urban atmosphere, as far as possible.

- In relation to the younger generation, the informants selected are those studying at university or have now completed their studies.

- Probability sample or model: our research will be deliberately sampled. The researcher or data collector will decide who will take part in the investigation as an informant in each location from among the characteristics listed above. No stratified samples will be used.

- Speech register: the register we wish to record from informants is informal, as used in their everyday lives.

5.2.4. Survey methodology

In order to create the written and spoken dialectal base composed of the data gathered, the methods and procedures used in international dialectology research will be employed.

The surveys will be conducted using a closed questionnaire, and the methodology to be employed will be based on direct investigations. Data will be accumulated on mini-disk and video camera. It is most important that the video and audio material be high quality. We also feel it is essential to use these materials to collate data. Use of video is proposed for two reasons: firstly, it is extremely useful for transcription of data, and secondly because it is helpful to grasp what is happening in human communications outside the language.

The surveys will be carried out on a solo basis – in other words, the person collecting data will give the questionnaire to the informant only. If the informant takes this very seriously, although there is a risk of recording a more formal register, this provides greater guarantees as to the certainty and quality of data.

Moreover, only one data collector will operate for each location in the survey, who will gather data for that location, and will be solely responsible for the suitability of the data, reporting to the person responsible for the research project.

In relation to the kind of discourse required to secure data, the conversation methodology has been selected. Below we have listed some of the reasons for selection of this method:

- This was the method used to collect the EHHA Atlas data.

- This is the most expedient way of gathering data.

- A large amount of data is to be collected.

- Data has to be collected at many locations.

We attach a great deal of importance to this method, precisely because we wish to collect data from different generations across the entire Basque-speaking area. Following in the footsteps of W. Labov (1972), since we wish to obtain a quantitatively important data corpus, it is absolutely necessary to use the conversation method.

The IPA phonetic alphabet will be used for transcription of the data, adapted to Basque, and in fact this is what is used in the Basque Language Academy's EHHA Atlas project.

5.3. Research plan

The project we propose will be broken down into the following phases:

5.3.1. Preparation of the questionnaire:

To launch the research project, based on the EHHA Atlas questionnaire, the most suitable questions for this project will be chosen (of the 2,875 questions used in the EHHA questionnaire, around 200 will be selected for our research).

This questionnaire includes, as mentioned above, lexical and grammatical characteristics (phonology, morphology and syntax).

5.3.2. Data collation and transcription

Survey locations: as mentioned above, the questionnaire will be applied to 100 locations in the Basque Country.

Informants: two informants will be selected at each location: one will be 45-50 years old, whilst the other will be 20-25 years old.

For collection of data, as mentioned above, minidisk and video cameras will be used. To transcribe the data, in the first instance the audio material will be stored in the computer using a suitable processing software, and subsequently this will be entered in the data base with "transcription-friendly" software. The idea is to use programmes readily available on the market to digitalise and label the audio material: Cooledit, Praat or similar.

In relation to video-processing and storage software, the "home-video" system will be sufficient for this project.

5.3.3. Creating a dialectal Basque database

A cooperation agreement has been signed with *Euskaltzaindia* to create a dialectal Basque database. The Academy has developed a new computer programme to handle the data gathered for the EHHA Atlas, and this programme is well suited to computerise the data transcribed for our research project. *Euskaltzaindia* has left the computer application to this research team through an agreement.

5.3.4. Creating consultation tools

Creating the corpus is just as basic a concern as allowing data to be consulted. The research team's point of view is that the data collected should be available to society as far as possible, to make it accessible to the scientific community.

5.3.5. Using the data

Different usages of the data have been envisaged:

1. A contrastive language atlas. It could be possible to publish data in digital format in addition to book format, and this would provide contrasting data between both generations. The publication would consist of the data collated, with the help of maps.

2. In-depth examination of data could be published for national and international conferences and specialist linguistic magazines *(ASJU, Lapurdum; Dialektologia y Geolinguistika, RIEV, Géolinguistique, Euskalingua, Dialectologia* etc.*)*. Among others, the topics of this research could be: language changeability factors, the behaviour of different linguistic characteristics, and the differences between the two areas.

6. The consequences of research: improvement and innovative features

The following will be some of the results of the research project:

- the results of sociolinguistic examination in relation to age: linguistic changeability in accordance with age will be studied, for example.

- the relationship between standard language and speech from each area will also be studied: an examination will be carried out to determine any effect of standard language on all forms of speech.

- an examination will also be carried out into how the different parts of the language behave within the relationships between standard language and speech from each area – which factor changes most and most easily? – and an attempt will be made to answer similar questions.

These surveys will constitute progress in relation to knowledge of changeability, and we will be in a position to draw comparisons between Basque and occurrences in other languages.

Moreover, creating an oral and written dialectal database will constitute considerable headway for more linguistic research to be carried out. To date we have a substantial lack in terms of a Basque dialectal database, and this is a much-needed tool for synchronic and diachronic research on the language.

The location of this database at the University of the Basque Country would reinforce the area of research into Euskara, since this would make it an important reference point. To date investigators have been forced to work compiling data for each piece of research. If the database as proposed is compiled in a serious manner, then as it gradually accumulates there will be increasingly less work to do in terms of data-gathering. This will also represent a gain in terms of the reliability of data.

7. Bibliography

Arnold, J. (co-editor), (1996). *Sociolinguistic variation: data, theory, and analysis* (selected papers from NWAV 23 at Stanford/Conference on New Ways of Analyzing Variation, 1994. Stanford), California: Center for the Study of Language and Information, Stanford.

Auer, P. (1998). "Dialect Levelling and the Standard Varieties in Europe", *Folia Linguistica* XXXII/1-2, 1-9.

Auer, P., F. Hinskens (1996). "The Convergence and Divergence of Dialects in Europe. New and not so New Developments in an Old Area", *Sociolinguistica* 10, 1-30.

Aurrekoetxea, G. (1997). "Ahozko testuak transkribatzeko irizpideak" (Criteria for transcription of oral texts), *Uztaro* 23, 87-94.

Aurrekoetxea, G. (2000). "Bizkaieradunok Euskara Estandarraren Aurrean" (Bizkaian Basque speakers and standard Basque), in *Mendebaldeko Berbetearen Formalizazinoa*; Mendebalde Kultur Alkartea, Bilbo, 235-245.

Aurrekoetxea, G. (2003). "Euskalkiak estandarraren uholdepean (Arratiako kasua)" (Basque dialects under a deluge of standard Basque (the case of Arratia), in *Ahozkotasuna aztergai*, Bilbo, Mendebalde Kultur Alkartea, 167-185.

Aurrekoetxea, G. (2004). "Estandar eta dialektoen arteko bateratze-joerak (ikuspuntu teorikotik begirada bat)" (Unifying trends between standard Basque and dialects – (from the theoretical point of view)), *Uztaro*, 50, 45-57.

Aurrekoetxea, G. (pending publication). "Hizkuntza estandarraren eta dialektoen arteko bateratze-joerak" (Unifying trends between standard Basque and dialects).

Aurrekoetxea, G., Videgain, X. (1993). "Euskal Herriko Hizkuntz Atlasa: galdesorta" (Basque Language Atlas: questionnaire), *Euskera* 38, 529-647.

Bellmann, G. (1998). "Convergence and Divergence in the Swiss German Dialects", *Folia Linguistica* XXXII/1-2, 53-67.

González, S. (2001). "Barietatea eta estandarra: batuaren eragina ume bizkaitar biren hizkeran" (Variety and standard: the effect of standard *batua* Basque on the speech of two children in Bizkaia), *Enseiukarrean* 15, 89-144.

Labov, W. (1972). "Some principles of linguistic methodology", *Language in Society* I, 97-120.

Paolillo, J. C. (2002). *Analyzing Linguistic Variation. Statistical Models and Methods*, Stanford, CSLI Publications.