New Perspectives in Iberian Dialectology / Nouvelles perspectives en dialectologie ibérique

David Heap, Enrique Pato, and Claire Gurski (eds.)

Twelfth International Conference on Methods in Dialectology
(Moncton, New Brunswick, Canada, 1st – 5th August 2005)

The linguistic atlas of basque language (EHHA): *Renewing methods in geolinguistics*

Gotzon Aurrekoetxea
Academy of the Basque Language

The linguistic atlas of basque language (EHHA): Renewing methods in Geolinguistics

Gotzon Aurrekoetxea
Academy of the Basque Language

1. Introduction

The Basque language or Euskara -the way Basque people name it- is one of the oldest languages in Europe and it is spoken in the Eastern side of the Cantabrian Sea, in the Bay of Biscay. It is a pre Indo-European language, but it is surrounded by Romance languages since the Romanization of the Iberian Peninsula, 20 centuries ago. This minorized language is not widely spoken in all the regions known as culturally and ethnically Basque, but it is only spoken in the northern half of the Basque country.

The history of the Linguistic Atlas of the Basque Country has been long and complicated even since the beginning of the project. The idea of starting the Linguistic Atlas arose as soon as the French Linguistic Atlas (ALF) was finished. From then on, the need of such a project was often mentioned by Basque Scholars, but it didn't came true until 1983, when The Royal Academy of the Basque Language designated a group of people to work on it. The Academy took the responsibility of the project due to the lack of interest from university and other research institutions. The objective of this methodology is to know what type of language produces the "ideal" speaker and the type of language he or she is able to understand.

2. Field Methods[1]

Questionnaires were collected between 1987 and 1992. The methodology used in this project has been the same as the one used in conventional linguistic atlases; in most cases, the questionnaire, which compiles 2875 questions, gathers lexicon (specially traditional lexicon), phonology, morphology (which is very important in the Basque language) and syntax. The data have been gathered in 145 places; we have chosen aged people (older than 60 years old), usually one informant in each place, although in some places, there have been more than one. We did not consider sociolinguistic variables in this project.

But other characteristics are not that usual in conventional linguistic atlases:

1. All the questionnaires have been recorded: all the collected data have been recorded in every place.
2. The type of data-gathering used in the Basque Atlas is different from the one usually used in similar project. The methodology of guided interview was chosen, and the 7 gatherers were given two orders: they should not interrupt the speaker

---

[1] The innovating contributions of the project to Language Cartography were presented in the *4th International Congress of Dialectologists and Geolinguistics*, two years ago, in Riga (Latvia).

and they should try to create a leaded interview. Consequently, we also collected longer texts, together with answers to the questions posed. In these texts, we can find the wisdom that is included in the concept of popular wisdom (sayings, tales, songs, beliefs, recipes, definitions...).

3. Renovation in data-gathering: The proposals method

In spite of taking this into account, there is a point related to the methodology of data-gathering which I would like to analyze more at length and more deeply: we call it "the proposal system". This system is an asking-method which is barely used in linguistic atlases. It was first used in geolinguistics by Ravier, in his *Atlas linguistique de Languedoc Occidental (ALOc)* in France.

Although I have mentioned this issue before,[2] I think that it is a very important subject-matter, and that the contribution the method itself makes is also important for the moment of data gathering; that is why I would like to talk about it in more detail.

It is known that every human being is familiar with the different registers of his or her mother tongue. That is to say, we don't use the same type of language when we are at the bar or at home and when we speak with a prestigious person. And it doesn't matter which the scientific training or the cultural formation of the person is, to mention only some examples.

Likewise, it is also known that our passive language competence is greater than our active competence. That is to say, everyone is able to understand language forms or lexical units that we have never produced. In other words, even if we know passively many units of a language, we don't use all of them to communicate with each other. We don't use all the language units that are used by others; and moreover, we know through closer or weaker relationships with people who speak another accent or variety, that there are other language units that are not familiar to us and that we are able to differentiate them.

Being things so, in geolinguistic data-gathering, it is not taken into account, and researchers follow the methodology which was created in the origins of geolinguistics. We have to admit that from a sociolinguistic point of view, Labov's contribution to the field is crucial. But apart from the influence of extra-linguistic problems, the methodology has not much developed with regard to the speaker's language competence.

I think that to examine thoroughly the variability of languages, we need to use an appropriate method in gathering data and bring up to date old methods.

From this point of view, we can talk about the advantages of the methodology used in the Basque Atlas: that is, the proposal system.

According to the methodology chosen in this project, we decided to use two question-systems: first, we made conventional direct questions and then, we asked to the subjects whether they knew the words we proposed them or not.

The gatherers don't take into account only the answer that the speaker gives to a certain question. Although that is their main task, interviewers go beyond that, and they have to

---

[2] "L'interprétation dans les atlas linguistiques: le cas de l'Atlas linguistique basque (EHHA)" in *the 4th International Congress of Dialectologists and Geolinguistics* (Riga, 2003).

examine carefully the speaker's passive competence by focusing on the speaker's polylectal grammar.

The objective of this methodology is to know what type of language produces the "ideal" speaker and the type of language he or she is able to understand.

It is obvious that even if two speakers live in the same place and even though they have other similar "objective" characteristics, they can have a different active knowledge, but also a different passive competence. Even if this is true, in my opinion, there is no problem in considering the subject representative of his or her town, because we have to understand that everyone is different and consequently, if we considered more subjects for each place we would obtain differing results.

Anyway, following the traditional methodology used in geolinguistics, the speaker choosing system used when compiling atlases wouldn't fail. Nevertheless, it is precisely that what shows the weakness of this method with reference to this point.

The method used in EHHA for data-gathering is based on direct questions in guided interviews. Then, speakers are systematically proposed some words, and they are asked whether they know them or not. The procedure is the following: regardless of whether the speaker had answered the question or not, the data-gatherer proposes to the subject some words and asks him whether he knows them or not, whether those words are used in the place he lives or not, whether they have the same meaning or not, etc.

The words proposed have the following characteristics:

1. They have been gathered in the towns nearby.
2. We have not proposed the same words in all the towns.
3. We have not proposed the same amount of words in all the towns.
4. The more new words we gathered, the more we proposed them; we have proposed less word in the first questionnaires, but the amount of words proposed has increased in the last ones.

On the one hand, to give some examples, in the first 516 questions (they can be found in the two first books that will be published), we have collected 71.370 answers and 54.593 proposals have been made (125.963 words have been collected). 43.43% of the words have been collected by the proposal way. Using this survey technique we have gathered more than half of the words compared to the linguistic data collected by the classical methodology. Some of these words are the answers given to the questions by speakers and the others are the words that have been proposed or suggested to the speakers. We have to take into account that they have been gathered using two different techniques and we can't mix them. However, I think that it is a good way to gather all the data which is in the brain of native speakers. And, in consequence, it can be an excellent aid to explore variability in depth the idiolect of speakers.

On the other hand, and fixing our eyes only on the proposals, the attitude of the speakers to the proposals can be divided in five groups (Picture 1):

1. The speaker knows the word and has recognized it as being used in his town: 10.98 cases (there is the 19%)
2. He knows the word but it doesn't have the same meaning: 4.256 cases (8%)

3. He knows the word but it is not used in his town: 2.656 cases (5%)
4. He doesn't know the word proposed: 35.746 cases (65%)
5. Other answers: 1.736 cases (3%)

Picture 1.
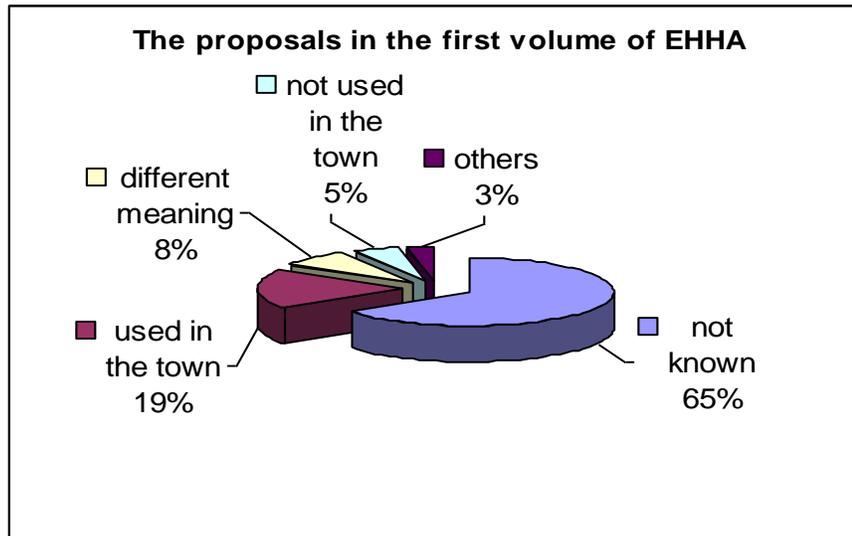
Before any other consideration, we have to say that in the maps and also in the phonetic lists of answers or in the index, as well as the answers, we have gathered only the proposals which were recognized as being used in the town with the same meaning. Nonetheless, the proposals identified as used in the town don't have the same weight as the answers, and both in the list and on the map, they are represented differently: in the lists, proposals are preceded by a star [*] and they are written in italics (Picture 2). On the map, instead of taking the whole area which belongs to the town, they are represented by a little circle (Picture 3).

Instead of working on the onomasiologic point of view, it seems that by using these data, it is possible to work also on the semasiologic one, but it is not completely viable, because all the words have not been systematically proposed in all the places. We have made a big effort to collect these words. But the results have been fantastic. We have suggested 54.593 words in the first two books of the Basque linguistic atlas.

On the one hand, we have collected 35.746 negative proposals: there are words that native speakers don't know. Those are considered negative data (the "données négatives" from the Linguistic Atlas of Languedoc). Nowadays, thanks to this study, we know which word is known and which one isn't known in each village. Until now we never had so much information about a place. Despite the fact that negative data is important, the Basque language Atlas doesn't take it into account, even though we keep it in our data-base.

On the other hand, we think that the percentage of 19 % that shows the words recognized as being used in the town is very significant. If this methodology had not been

used, almost the fifth part of the words which we are looking for wouldn't have been collected. With this data on the hand, we can ask about the quality of the data we collect. We can say without doubt that we only collect a part of the language used by native speakers.

4. Conclusions

I hope that my contribution would be interesting to follow researching in the best way to collect linguistic data. I show a different way to investigate in the basic data completing the traditional way used so far in Linguistic Atlases. My aim is to look for survey system that makes easier to study the idiolect of native speakers wholly. The objective of my contribution is to help in this way.

References

Aurrekoetxea, Gotzon. 1986. "Euskal Herriko Hizkuntz Atlasa (EHHA): inkesta metodologia eta ezezko datuak" [The Linguistic Atlas of the Basque Country (EHHA): methodology of questionnaires and negative data], *Euskera* XXXI/2, 413-424.

Aurrekoetxea, Gotzon. 1992. "Euskal Herriko Hizkuntz Atlasaren informatizapena" [The computerization of The Linguistic Atlas of the Basque Country]. In G. Aurrekoetxea & X. Videgain (eds.), *Proceeding of International Congress on Dialectology*. Bilbao: Euskaltzaindia, 319-337.

Aurrekoetxea, Gotzon & Xarles Videgain (eds.). 1989. "L'atlas linguistique du Pays Basque: Euskal Herriko Hizkuntza Atlasa" [The Linguistic Atlas of the Basque Country]. In *Hommage au Musée Basque*. Bayonne: 495-510.

Aurrekoetxea, Gotzon & Xarles Videgain (eds.). 1992. *Acts of the International Congress on Dialectology*, (*Bilbao, 21/25-X-1991*). Bilbao: Euskaltzaindia.

Aurrekoetxea, Gotzon & Xarles Videgain. 1993. "Euskal Herriko Hizkuntz Atlasa (EHHA): galdesorta / cuestionario / questionnaire" [The Linguistic Atlas of the Basque Country (EHHA): questionnaire], *Euskera* XXXVIII/3, 529-647.

Aurrekoetxea, Gotzon & Xarles Videgain. 1994. "Historia y futuro del Atlas Lingüístico Vasco (EHHA)". In P. García Mouton (ed.), *Geolingüística. Trabajos europeos*. Madrid: CSIC, 79-97.

Aurrekoetxea, Gotzon. 2002a. "El atlas lingüístico vasco (EHHA)". In M. Aurnague & M. Roché (eds.), *Hommage à Jacques Allières, I. Domaines basque et pyrénéen*. Biarritz: Atlantica, 63-71.

Aurrekoetxea, Gotzon. 2002b. "Algunas consideraciones sobre la contrapregunta en las encuestas lingüísticas". In L. Rabassa (ed.), *Mélanges offerts à Jean-Louis Fossat*. Toulouse: Cahier d'Etudes Romanes 11/12, 57-65.

Ravier, Xavier. 1965. "Le traitement des données négatives dans l'Atlas linguistique et ethnographique de la Gascogne". *Revue de linguistique romane* XXIX, 262-274.

Bizkaia
Arrazola (Atxondo): aɾði.
Arrieta: aɾði.
Bakio: aɾði.
Bermeo: aɾði.
Berriz: aɾði.
Bolibar (Markina-Xemein): aɾði.
Busturia: aɾði.
Dima: aɾði.
Elantxobe: aɾði.
Elorrio: aɾði.
Errigoiti: aɾði.
Gamiz-Fika: aɾði.
Getxo: aɾði.
Gizaburuaga: aɾði.
Ibarruri (Muxika): aɾði.
Kortezubi: aɾði.
Larrabetzu: aɾði.
Laukiz: aɾði.
Leioa: aɾði.
Lekeitio: aɾði.
Lemoa: aɾði.
Lemoiz: aɾði.
Mañaria: aɾði.
Markin-Etxebarri: aɾkakośo, aɾði.
Mendata: aɾði.
Mungia: aɾði.
Ondarroa: aɾði, *aɾkakuʃu.
Orozko: aɾði.
Otxandio: aɾði.
Sondika: aɾði.
Uribe-Etxebarri: aɾði.
Zaratamo: aɾði.
Zeanuri: aɾði.
Zeberio: aɾði.
Zollo (Arrankudiaga): aɾði.
Zornotza: aɾði.

Araba
Aramaio: aɾði, aɾkákuʃo.

Gipuzkoa
Beasain: aɾkágʼśó.
Beizama: aɾkakośó.
Bergara: aɾkakóʃo.
Deba: aɾkákośó.
Donostia (Ibaeta): aɾkakuśo.
Eibar: aɾði.
Elduain: aɾkákʷśó.
Elgoibar: aɾkayóśu.
Errezil: aɾkakośo, *kukuśo.
Ezkio-Itsaso: aɾkakośó.
Getaria: aɾkakuśo.
Hernani: aɾkakuśo.
Hondarribia: kukuśo.
Ikaztegieta: aɾkakośo.
Lasarte: aɾkakośo.
Legazpi: aɾkakuśo.
Leintz-Gatzaga: *aɾði, *aɾkakuśo.
Mendaro: aɾkakuśo.
Oiartzun: aɾkakuśo.
Oñati: aɾkakóʃo, *aɾði.
Oñati (Araotz): aɾkakośo, *aɾði.
Orexa: aɾkakośo.
Orio: aɾkakuśo.
Pasaia: aɾkakuśo.
Tolosa: aɾkaʷóśó.
Urretxu: aɾkakośó.
Zegama: aɾkakośó.
Zestoa (Arroa): aɾkakośo.

Nafarroa
Abaurregaina: kukuśo.
Alkotz: kukuśua (mark.), kukuśwa (mark.), kukuśoa (mark.).
Aniz: kukuś, kukuśo.
Arbizu: aɾkakośo.
Beruete: aɾkakośo.
Donamaria: kukuśu.
Dorrao: aɾkakośo.
Eiaurrieta: kukuśo.
Erratzu: kukuśo.
Etxalar: kukuśo.
Etxaleku: aɾkakuśo.

Luzaide: kukuśo.
Mezkiritz (Ureta): kukuśo.
Oderitz: aɾkakuśo.
Suarbe: kukuśo.
Sunbilla: kukuśu, kukuśo.
Urdiain: aɾkakóʃo.
Zilbeti: kukuśu.
Zugarramurdi: kukuśo.

Lapurdi
Ahetze: kukuśo.
Arrangoitze: kukuśu.
Azkaine: kukuśo.
Bardoze: kukuśo.
Beskoitze: *kukuśo.
Donibane-Lohizune: kukuśo.
Hazparne: kukuśo.
Hendaia: kukuśo.
Itsasu: kukuśo.
Makea: kukuśo.
Mugerre: kukuśo.
Sara: kukuśo.
Senpere: kukuśo.
Urketa: kukuśwa (mark.).
Uztaritze: kukuśo.

Nafarroa Beherea
Aldude: kukuśo.
Arboti-Zohota: kukuśo.
Armendaritze: kukuśo.
Arnegi: kukuśo.
Arruta-Sarrikota-Amikuze: kuyuśo.
Baigorri: kukuśo.
Bastida: kukuśo.
Behorlegi: kukuśo.
Bidarrai: kukuśo.
Ezterenzubi: kukuśo.
Gamarte: kukuśo.
Garruze: kukuśo.
Irisarri: kukuśo.
Isturitze: kukuśo.
Jutsi: kukuśo.

Domintxine-Berrueta: khukuśo.
Eskiula: kykyśo.
Larrañe: kykyśo.
Muntori: kykyśo.
Phagola: kykyśo.
Santa-Grazi: kykyśo.
Sohüta: kykyśo.
Urdiñarbe: kykyśo.
Ürrüxtoi-Larrabile: kykyśo.

Mapan sartzen ez diren erantzunak:
Gipuzkoa
Leintz-Gatzaga - G: *arkakuso.

---

30. Mapa: pulga / puce / flea / *(Pulex irritans)*

GALDERA: 01300; ALG: 60

- ardi
- arkakoso
- arkakoxo
- arkakuso
- arkakuxo
- kukuso
- küküso
- khukuso